St.Petersburg Institute
of Fine Mechanics and Optics
(Technical University)

Division of Applied Mathematics and Physics

Proceedings of Student Conferences in Mathematics

Edited by V.N.Vasiliev

St.Petersburg
1997

Scientific editor: Professor **V.N.Vasiliev,**
St.Petersburg Institute of Fine Mechanics and Optics (Technical University)

Referee: Professor **A.Marino,** Università di Pisa (Italy)

ISBN 5-7577-0017-3

Published according to the approval of the Scientific Council
of St.Petersburg IFMO of 24.12.96

Foreword of the referee...

I think that the book "Proceedings of students conferences in Mathematics" is a useful and very nice collection of papers on the topics of undoubtful interest for the students of scientific faculties of the universites. On each particular topic the principal results are provided, sometimes even suggesting different ways to prove the same statements, so as to give the students the opportunity to understand better what they are studying and to get an essential idea of the whole area of mathematics of which the exposed theories take part. In some parts the exposition is quite original and unusual. Summing up, I would say that this book can be used in the courses of Analysis at the universities as an excellent supplementary to the existing textbooks.

Antonio Marino,
Professor of Mathematical Analysis,
Università di Pisa, Italy

Preface

The book you are opening now is an attempt to describe some results of an experiment in the courses of Mathematical Analysis to students of of Physics and Mathematics. In fact, a common problem in teaching Mathematical Analysis at the Universities is that the bulk of obligatory information to be communicated to students is gradually increasing: besides "classical" theory of functions, the courses are to include now a deal of general topology, functional analysis and operator theory. Thus, unless one wants the students to spend all their time studying only Analysis, one is, however painful it might be, to neglect some of the topics. But besides being unpleasant for any maths teacher, such omitting of analytical stuff can never be absolutely harmless for students. A possible solution to this problem would be to leave some of the beautiful and important, but usually neglected chapters of Analysis to the students in order that they study these arguments in a profound way by themselves. Such an approach besides being useful for a teacher allowing him to concentrate the attention on fewer central arguments of the course, is also of great help for the student, who in this way is constrained to undertake his own efforts to learn Analysis "manually", not just making textbook excercises but doing some elements of independent research.

Let us now explain in brief the organization of the students' work, one of the results of which is this book. The "research" topics are normally chosen by the students themselves. It is supposed further that any student can under the appropriate scientific supervision develop any chosen argument, however complex it might seem. We hope that from the book it would be clear that such an assumption is justified by an experience. Starting to look at the given subject, the student is first of all to overview the existing literature. Most students however, do not limit themselves by simple compilation of the known facts and making a survey of the literature, they rather rearrange the material by their order of ideas, sometimes even filling it with examples of their own. Surely, this requires great efforts both from the student and its supervisor, but this never comes out to be in vain.

What you find in this book are systemized collections of known results prepared by the students while working at the assigned research topics. They concern rather delicate, but beautiful and very important parts of analysis, which normally are destinated to be neglected in the Analysis courses. Thus we think that the book might be interesting from at least two points of view: first, for everybody, as an "easy readable" supplement to the existing Analysis textbooks, and, second, as a source of ideas for those interested in the above-described didactical approach. We hope that our book be inspiring in this sense.

Acknowledgments

This work was undertaken at the Department of Computer Technology of PIFMO, within the framework of the special educational project.

This project aimed at selection and training of young people talented in the field of exact sciences has been launched in St.Petersburg 6 years ago. Its main intention was to bring up the high level specialists in the sphere of the applied mathematics, computaitonal physics and computer science who would be able to apply the latest achievements of sciences to the concrete problems of industry. Such an extra-class training for a small elitary groups of specially selected students is provided on the basis of the specially founded Division of Applied Mathematics and Physics which comprises both the Faculty of Physics of St.Petersburg University and the Department of Computer Technology of PIFMO. The staff of the Division consist of the experienced teachers and professors of the IFMO and the St.Petersburg State University. The annual admittance to the Division is about thirty young people. The system of student selection is based first of all on the unique St.Petersburg network of special physical and mathematical schools. This network has been fruitfully working for more than forty years. Only during the last two years pupils of these schools took more than a half of the total number of places in the Russian national teams at the International Olympiads in mathematics, physics and computer science and won the grater part of Russians awards.

Development of the research skills is considered to be one of the main goals of education at the Division. Students acquire this experience when carrying out a numerous cycle of course works which have certain research features. These works start at the very beginning of the first course of study and continue during all four years of study. This educational project is mainly intended to integrate Russia into the top levels of the international educational and research system. That is why particular attention is paid here to settting up long term cooperation with the countries of European Community. Now the Division has several commitments with the best European Universities about further training of our students on the master and postgraduate levels. Some of these commitments include the subsequent teaching of students at our and foreign universities according to the specially formed joint study plans.

The project authorities, prof. Vladimir V. Vasiliev (Chief) and prof. Vladimir G. Parfenov are gratefuly acknowledged. The editors would like also to express their thanks to the supervisor of the project prof. Nikolay Y. Dodonov. Last, but not least, we would like to thank prof. Antonio Marino (Università di Pisa), prof. Franco Conti (Scuola Normale Superiore, Pisa) for their interest in the our work and for their particular attention to the preliminary versions of this book.

People who made this book possible

Supervisors of the project

Nikolai Y. Dodonov, Professor of Department of Mathematics, SPb IFMO.

Vladimir N. Vasiliev, Chief of the Department of Computer Technology, SPb IFMO.

Vladimir G. Parfenov, Chief of the Division of Applied Mathematics and Physics, SPb IFMO.

Scientific Consultant

Eugene Stepanov, Associate professor of the Department of Computer Technology, SPb IFMO.

Editors

Dmitry A. Ilchenko, Student of Department of Computer Technology, SPb IFMO.

Anthony N. Likhodedov, Student of Department of Computer Technology, SPb IFMO.

Andrew A. Zdorovtsev, Student of Department of Computer Technology, SPb IFMO.

SPb IFMO stands for the St.Petersburg Institute of Fine Mechanics and Optics (Technical University)

Contents

A Criterion for Absolute Continuity of Induced Measure

A. Zdorovtsev

Introduction

Let a measure μ absolutely continuous with respect to the N-dimensional Lebesgue measure λ be defined on a region $Q \subset \mathbf{R}^N$. As an example one can think of a measure generated by an N-dimensional random vector $\mathbf{X}$ with the finite probability density $p(\mathbf{x})$. Consider the transformation of μ by the map $\mathbf{f}$: $Q \to \mathbf{R}^M$ (in general, $M \neq N$). It is a measure induced by μ by the map $\mathbf{f}$ defined by the relationship

$$\mu_{\mathbf{f}} B = \mu \mathbf{f}^{-1}(B),$$

where $\mathbf{f}^{-1}(B) \subset Q$ is the preimage of the set B. The question arises, when the induced measure is also absolutely continuous with respect to the M-dimensional Lebesgue measure. Translated into terms of the given example, it reads: in which case the transformed random vector $\mathbf{Y} = \mathbf{f}(\mathbf{X})$ has a finite probability density $q(\mathbf{y})$)?

The absolute continuity of μ will be implied by the absolute continuity of $\lambda_{\mathbf{f}}$, the measure induced under the action of $\mathbf{f}$ by the Lebesgue measure λ. In fact, if $\lambda_{\mathbf{f}}$ is absolutely continuous, then for any set Z, $\lambda Z = 0$, holds

$$\lambda_{\mathbf{f}} Z = \lambda \mathbf{f}^{-1}(Z) = 0,$$

which implies by absolute continuity of μ that

$$\mu_{\mathbf{f}} Z = \mu \mathbf{f}^{-1}(Z) = 0.$$

Therefore it is sufficient to solve the problem for the Lebesgue measure λ which we consider in what follows.

Case I: $M = N$

We prove Theorem 1.

Theorem 1 *Given a map* $\mathbf{f}$ *of a region (an open and connected set)* $Q \subset \mathbf{R}^N$ *into* $\mathbf{R}^N$, *and let* D *be its critical set (the set of points at which* $\mathbf{f}$ *is differentiable and has Jacobian* $\det \mathbf{f}' = 0$*). Then the measure of the image of the critical set*

$$\lambda \mathbf{f}(D) = 0.$$

PROOF. We first prove the following auxiliary result.

Lemma 1.1 *Let a map* $\mathbf{f}$ *be differentiable at a point* $\mathbf{x} \in Q$ *and have Jacobian*

$$\det \mathbf{f}'(\mathbf{x}) = \det \mathbf{L} = L.$$

Then for any $\varepsilon > 0$ *there exists a neighborhood* U *of* $\mathbf{x}$ *such that for any finite collection of points* $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N \in U$

$$|V(\langle \mathbf{f}(\mathbf{x}), \mathbf{f}(\mathbf{x}_1), \ldots, \mathbf{f}(\mathbf{x}_N)\rangle) - L \cdot V(\langle \mathbf{x}, \mathbf{x}_1, \ldots, \mathbf{x}_N\rangle)| < \varepsilon \Big(\max_n \|\mathbf{x}_n - \mathbf{x}\|\Big)^N,$$

where $\langle \cdot, \cdot, \ldots, \cdot \rangle$ *denotes the* N*-dimensional simplex with the respective summits and* V *denotes the algebraic volume counting orientation (positive for right simplices and negative for left ones).*

PROOF OF THE LEMMA. Without loss of generality we may consider that

$$\mathbf{x} = \mathbf{0}, \qquad \mathbf{f}(\mathbf{x}) = \mathbf{0}.$$

Then the volumes of simplices are expressed by the determinants:

$$V(\langle \mathbf{0}, \mathbf{x}_1, \ldots, \mathbf{x}_N\rangle) = \frac{1}{N} \det(\mathbf{x}_1, \ldots, \mathbf{x}_N),$$

$$V(\langle \mathbf{0}, \mathbf{f}(\mathbf{x}_1), \ldots, \mathbf{f}(\mathbf{x}_N)\rangle) = \frac{1}{N} \det(\mathbf{f}(\mathbf{x}_1), \ldots, \mathbf{f}(\mathbf{x}_N)).$$

Now take a sufficiently small $\delta > 0$. By definition of a derivative there is a neighborhood U of $\mathbf{x}$ such that at any point $\mathbf{t} \in U$

$$\mathbf{f}(\mathbf{t}) = \mathbf{L}\mathbf{t} + \mathbf{u}, \qquad \|\mathbf{u}\| < \delta \|\mathbf{t}\|.$$

Consider an arbitrary set of points $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N \in U$. For them we have

$$\mathbf{f}(\mathbf{x}_n) = \mathbf{L}\mathbf{x}_n + \mathbf{u}_n, \qquad \|\mathbf{u}_n\| < \delta \|\mathbf{x}_n\|.$$

Then

$$\det(\mathbf{f}(\mathbf{x}_1), \ldots, \mathbf{f}(\mathbf{x}_N)) = \det(\mathbf{L}\mathbf{x}_1 + \mathbf{u}_1, \ldots, \mathbf{L}\mathbf{x}_N + \mathbf{u}_N) =$$
$$= \sum_{i_1 < i_2 < \ldots < i_n} \pm \det(\mathbf{u}_{i_1}, \mathbf{u}_{i_2}, \ldots, \mathbf{u}_{i_n}; \mathbf{L}\mathbf{x}_{\bar{\imath}_1}, \mathbf{L}\mathbf{x}_{\bar{\imath}_2}, \ldots, \mathbf{L}\mathbf{x}_{\bar{\imath}_{N-n}}),$$

where $\bar{\imath}_1 < \bar{\imath}_2 < \ldots < \bar{\imath}_{N-n}$ is the complement of the set of indices $i_1 < i_2 < \ldots < i_n$. Hence

$$\begin{aligned}
&|\det(\mathbf{f}(\mathbf{x}_1), \ldots, \mathbf{f}(\mathbf{x}_N)) - L\det(\mathbf{x}_1, \ldots, \mathbf{x}_N)| = \\
&\quad |\det(\mathbf{f}(\mathbf{x}_1), \ldots, \mathbf{f}(\mathbf{x}_N)) - \det(\mathbf{L}\mathbf{x}_1, \ldots, \mathbf{L}\mathbf{x}_N)| = \\
&\quad \left| \sum_{i_1<i_2<\ldots<i_n;n>0} \pm \det(\mathbf{u}_{i_1}, \mathbf{u}_{i_2}, \ldots, \mathbf{u}_{i_n}; \mathbf{L}\mathbf{x}_{\bar{\imath}_1}, \mathbf{L}\mathbf{x}_{\bar{\imath}_2}, \ldots, \mathbf{L}\mathbf{x}_{\bar{\imath}_{N-n}}) \right| \le \\
&\quad \sum_{i_1<i_2<\ldots<i_n;n>0} |\det(\mathbf{u}_{i_1}, \mathbf{u}_{i_2}, \ldots, \mathbf{u}_{i_n}; \mathbf{L}\mathbf{x}_{\bar{\imath}_1}, \mathbf{L}\mathbf{x}_{\bar{\imath}_2}, \ldots, \mathbf{L}\mathbf{x}_{\bar{\imath}_{N-n}})| \le \\
&\quad \sum_{i_1<i_2<\ldots<i_n;n>0} \|\mathbf{u}_{i_1}\| \cdot \|\mathbf{u}_{i_2}\| \cdot \ldots \cdot \|\mathbf{u}_{i_n}\| \cdot \|\mathbf{L}\mathbf{x}_{\bar{\imath}_1}\| \cdot \|\mathbf{L}\mathbf{x}_{\bar{\imath}_2}\| \cdot \ldots \cdot \|\mathbf{L}\mathbf{x}_{\bar{\imath}_{N-n}}\| \le \\
&\quad \sum_{i_1<i_2<\ldots<i_n;n>0} \delta^n \|\mathbf{x}_{i_1}\| \cdot \|\mathbf{x}_{i_2}\| \cdot \ldots \cdot \|\mathbf{y}_{i_n}\| \cdot \|\mathbf{L}\|^{N-n} \|\mathbf{x}_{\bar{\imath}_1}\| \cdot \|\mathbf{x}_{\bar{\imath}_2}\| \cdot \ldots \cdot \|\mathbf{x}_{\bar{\imath}_{N-n}}\| = \\
&\quad \left(\sum_{i_1<i_2<\ldots<i_n;n>0} \delta^n \|\mathbf{L}\|^{N-n} \right) \|\mathbf{x}_1\| \cdot \|\mathbf{x}_2\| \cdot \ldots \cdot \|\mathbf{x}_N\| \le \\
&\qquad (2^N - 1)\delta \|\mathbf{L}\|^{N-1} \left(\max_n \|\mathbf{x}_n\| \right)^N
\end{aligned}$$

for δ sufficiently small ($\delta \le \|\mathbf{L}\|$). This implies the statement of the lemma: it is sufficient to take

$$\delta < \frac{\varepsilon}{(2^N - 1)\|\mathbf{L}\|^{N-1}}. \qquad \square$$

By Lemma 1.1 if a point $\mathbf{x}$ belongs to the critical set D (in this case $L = 0$), then for any $\varepsilon > 0$ there exists a neighborhood $U(\mathbf{x})$ such that for any set of points $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N \in U(\mathbf{x})$

$$|V(\langle \mathbf{f}(\mathbf{x}), \mathbf{f}(\mathbf{x}_1), \ldots, \mathbf{f}(\mathbf{x}_n) \rangle)| < \varepsilon \left(\max_n \|\mathbf{x}_n - \mathbf{x}\| \right)^n,$$

Take an arbitrary $\varepsilon > 0$ and for every point $\mathbf{x} \in D$ construct an N-dimensional parallelepiped $T_{\mathbf{x}} \ni \mathbf{x}$ with rational summits entirely lying in the neighborhood $U(\mathbf{x})$ corresponding to the given ε (and also in the region Q).

By countability of $\mathbf{Q}$ (the set of rational points) such parallelepipeds constitute a countable covering $\{T_k\}$ of D.

Now construct a system of N-dimensional cubes $\{P_m\}$ on the basis of the system $\{T_k\}$ by induction:

We consider the parallelepipeds $T_1, T_2, \ldots$ one by one and add to the system $\{P_m\}$ the sets of cubes generated by them in the following way. Suppose the elements of $\{P_m\}$ generated by the parallelepipeds $T_1, T_2, \ldots, T_{k-1}$ are already constructed. Then the part of T_k not intersecting with them (denote it by A) can be represented as the union of a finite set of identical N-dimensional cubes because all the dimensions of A are rational and can be reduced to a common denominator. Determine for each of these cubes whether it contains at least one point $\mathbf{x}$ such that $T_{\mathbf{x}} = T_k$. We add to the system $\{P_m\}$ "on behalf of" T_k those cubes which do.

The system $\{P_m\}$ has the following properties:

a) It is disjoint and therefore has total volume not more than the measure of the whole region λQ. This is clear from the construction.

b) It covers the set D. Prove that by contradiction.

Suppose there exists a point $\mathbf{x} \in D$ not belonging to any of the cubes P_m. Consider the parallelepiped $T_k = T_x$. Processing it when constructing $\{P_m\}$ we separated its part not intersecting with the cubes already constructed into a set of disjoint cubes. Since $\mathbf{x}$ is not covered by $\{P_m\}$, it had to lie in one of those cubes. But then we had to add that cube to $\{P_m\}$ as containing the point $\mathbf{x}$ such that $T_x = T_k$, thus covering $\mathbf{x}$. This is a contradiction.

Denote by $\mathbf{t}_m$ one of the points of D due to which the cube P_m entered the system $\{P_m\}$. We have

$$P_m \subset T_{\mathbf{x}_m} \subset U(\mathbf{x}_m).$$

Hence for any set of points $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N \in P_m$

$$|V(\langle \mathbf{f}(\mathbf{t}_m), \mathbf{f}(\mathbf{x}_1), \ldots, \mathbf{f}(\mathbf{x}_N)\rangle)| < \varepsilon\left(\max_n \|\mathbf{x}_n - \mathbf{t}_m\|\right)^N \leq$$

$$\leq \varepsilon\left(\sqrt{N}\,\sqrt[N]{\lambda P_m}\right)^N = \varepsilon N^{N/2}\lambda P_m.$$

Now consider an arbitrary simplex $\langle \mathbf{x}_1, \ldots, \mathbf{x}_{N+1}\rangle \in P_m$. For it we obtain

$$|V(\langle \mathbf{f}(\mathbf{x}_1), \ldots, \mathbf{f}(\mathbf{x}_{N+1})\rangle)| \leq$$

$$\leq \sum_{n=1}^{N+1} |V(\langle \mathbf{f}(\mathbf{x}_1), \ldots, \mathbf{f}(\mathbf{x}_{n-1}), \mathbf{f}(\mathbf{t}_m), \mathbf{f}(\mathbf{x}_{n+1}), \ldots, \mathbf{f}(\mathbf{x}_{N+1})\rangle)| \leq \varepsilon(N+1)N^{N/2}\lambda P_m.$$

Therefore for any set of points $\mathbf{y}_1, \ldots, \mathbf{y}_{N+1} \in \mathbf{f}(P_m)$ the volume of simplex with apices $\mathbf{y}_n$ does not exceed $\varepsilon(N+1)N^{N/2+1}\lambda P_m$. We need the following lemma.

Lemma 1.2 *Given a set $B \subset \mathbf{R}^N$ such that for any finite collection of points $\mathbf{y}_1, \ldots, \mathbf{y}_{N+1} \in B$*

$$|V(\langle \mathbf{y}_1, \ldots, \mathbf{y}_{N+1}\rangle)| \leq C,$$

where C is some constant. Then $\lambda B \leq \theta_N C$, where the constant θ_N depends only on the dimension of the space.

PROOF OF THE LEMMA. For simplicity we restrict ourselves to considering the planar ($N = 2$) case. The proof can be then simply reiterated word-to-word for arbitrary N.

Denote by S the supremum of the areas of the triangles with summits in B. There exists a triangle $\triangle XYZ$, with $X, Y, Z \in B$, which possesses an area

$$\text{Area}(\triangle XYZ) > S/2.$$

For the heights of this triangle H_{XY}, H_{XZ} and H_{YZ} we obtain then

$$\frac{S}{|XY|} < H_{XY} \leq \frac{2S}{|XY|}, \qquad \frac{S}{|XZ|} < H_{XZ} \leq \frac{2S}{|XZ|}, \qquad \frac{S}{|YZ|} < H_{YZ} \leq \frac{2S}{|YZ|}.$$

For any point $W \in B$ the areas of the triangles $\triangle XYW$, $\triangle XZW$ and $\triangle YZW$

$$\text{Area}(\triangle XYW), \text{Area}(\triangle XZW), \text{Area}(\triangle YZW) \leq S.$$

Then for the heights h_{XY}, h_{XZ} and h_{YZ}, dropped from the point W onto the corresponding sides of these triangles, we have

$$h_{XY} \leq \frac{2S}{|XY|}, \qquad h_{XZ} \leq \frac{2S}{|XZ|}, \qquad h_{YZ} \leq \frac{2S}{|YZ|}.$$

Therefore any point $W \in B$ belongs to the meet of the strips, symmetrically enveloping the lines (XY), (XZ) and (YZ), of width $4S/|XY|$, $4S/|XZ|$ and $4S/|YZ|$ respectively. Hence the whole set B lies inside this intersection.

By the relations for H_{XY}, H_{XZ} and H_{YZ} we can increase the widths of the strips, each not more than twice, to make them proportional to the heights of the triangle $\triangle XYZ$. Obviously, the intersection of the enlarged strips contains B as before.

The area of the intersection of the enlarged strips is maximum when the intersection is a centrally symmetric hexagon and is, in that case, $3/2$ of the area of the triangle similar to $\triangle XYZ$, with heights equal to the widths of the new strips. The similarity coefficient does not exceed $4 \cdot 2 : 1 = 8$. Hence the area of the intersection is not more than

$$3/2 \quad 8^3 S = \theta_2 S \leq \theta_2 C,$$

which implies

$$\lambda B \leq \theta_2 C.$$

This completes the proof of the lemma.

Lemma 1.2 implies that

$$\lambda \mathbf{f}(P_m) \leq \theta_N \varepsilon (N+1) N^{N/2+1} \lambda P_m.$$

Therefore, since $\{P_m\}$ is a covering of D,

$$\lambda \mathbf{f}(D) \leq \sum_m \lambda \mathbf{f}(P_m) \leq \theta_N \varepsilon (N+1) N^{N/2+1} \sum_m \lambda P_m \leq \theta_N \varepsilon (N+1) N^{N/2+1} \lambda Q.$$

Hence by arbitrarity of ε we have

$$\lambda \mathbf{f}(D) = 0,$$

which proves the theorem. ∎

We prove Theorem 2.

Theorem 2 *Given a map* $\mathbf{f}$ *of a region* $Q \subset \mathbf{R}^N$ *into* $\mathbf{R}^N$, *let* D *be its critical set. Let* $\mathbf{f}$ *be differentiable everywhere on a set* A, $\lambda A > 0$, *disjoint with* D. *Then the measure of the image of* A

$$\lambda \mathbf{f}(A) > 0.$$

PROOF. Since $\det \mathbf{f}' \neq 0$ on A, by continuity of the measure λ there exists a subset $K \subset A$, $\lambda K > 0$, on which the Jacobian is bounded away from zero:

$$|\det \mathbf{f}'(\mathbf{x})| > C > 0.$$

Approximate K outwards with a closed set G also of positive measure.

Lemma 2.1 *Let* $\mathbf{y} \in \mathbf{f}(G)$. *Then* $\mathbf{f}^{-1}(\mathbf{y}) \cap G$ *is at most countable.*

PROOF OF THE LEMMA. We show that every point of $\mathbf{f}^{-1}(\mathbf{y}) \cap G$ is isolated. Suppose the contrary, i.e. that there exists a sequence of points $\mathbf{x}_k \in \mathbf{f}^{-1}(\mathbf{y}) \cap G$ converging to a point $\mathbf{t} \in \mathbf{f}^{-1}(\mathbf{y}) \cap G$, $\mathbf{t} \neq \mathbf{x}_k$.

For simplicity consider that

$$\mathbf{y} = \mathbf{0}, \qquad \mathbf{t} = \mathbf{0}.$$

Denote $\mathbf{L} = \mathbf{f}'(\mathbf{x})$. Then we can write the definition of derivative as

$$\mathbf{f}(\mathbf{x}) = \mathbf{L}\mathbf{x} + \alpha(\mathbf{x}), \qquad \frac{\|\alpha(\mathbf{x})\|}{\|\mathbf{x}\|} \to 0, \quad \mathbf{x} \to \mathbf{0}.$$

Applying it to the sequence $\mathbf{x}_k$, we obtain

$$\mathbf{L}\mathbf{x}_k = -\alpha(\mathbf{x}_k), \qquad \frac{\|\alpha(\mathbf{x}_k)\|}{\|\mathbf{x}_k\|} \to 0, \quad k \to \infty,$$

because $\mathbf{f}(\mathbf{x}_k) = 0$. Hence for any $\varepsilon > 0$, if k is sufficiently large,

$$\|\mathbf{L}\mathbf{x}_k\| < \varepsilon \|\mathbf{x}_k\|,$$

which implies that

$$\inf_{\|\mathbf{x}\|=1} \|\mathbf{L}\mathbf{x}\| = 0.$$

But the latter is possible only if

$$\det \mathbf{L} = \det \mathbf{f}'(\mathbf{x}) = 0.$$

Therefore we obtain a contradiction with the assumption $\det \mathbf{f}' \neq 0$ on the set G.

Hence all the points of $\mathbf{f}^{-1}(\mathbf{y}) \cap G$ are isolated. Then their quantity is at most countable because it is impossible to allocate an uncountable set of disjoint balls in $\mathbf{R}^N$. □

By Lemma 2.1 for every point $\mathbf{y} \in \mathbf{f}(G)$

$$\mathbf{f}^{-1}(\mathbf{y}) \cap G = \{\mathbf{g}_k(\mathbf{y})\}.$$

Prove continuity of all the maps $\mathbf{g}_k : \mathbf{f}(G) \to G_k = g_k(\mathbf{f}(G))$ on $\mathbf{f}(G)$. Suppose the contrary, i.e. that there exists a sequence $\mathbf{y}_m \in \mathbf{f}(G)$ converging to a point $\mathbf{y} \in \mathbf{f}(G)$ such that $\mathbf{g}_k(\mathbf{y}_m)$, does not converge to $\mathbf{g}_k(\mathbf{y})$. Extract from $\{\mathbf{g}_k(\mathbf{y}_m)\}$ a subsequence separated from $\mathbf{g}_k(\mathbf{y})$ (disjoint with some neighborhood of $\mathbf{g}_k(\mathbf{y})$) and then extract from the latter its converging subsequence

$$\mathbf{g}_k(\mathbf{y}_{m_p}) \to \mathbf{x}, \quad m \to \infty.$$

It can be done because the set G is compact (bounded and closed).

Further by continuity of $\mathbf{f}$ on G ($\mathbf{f}$ is everywhere differentiable on G)

$$\mathbf{y}_{m_p} = \mathbf{f}(\mathbf{g}_k(\mathbf{y}_{m_p})) \to \mathbf{f}(\mathbf{x}).$$

Hence

$$\mathbf{f}(\mathbf{x}) = \mathbf{y}.$$

But then $\mathbf{x} = \mathbf{g}_k(\mathbf{y})$ that implies that

$$\mathbf{g}_k(\mathbf{y}_{m_p}) \to \mathbf{g}_k(\mathbf{y}),$$

that contradicts with separability of the sequence $\mathbf{g}_k(\mathbf{y}_{m_p})$ and the point $\mathbf{g}_k(\mathbf{y})$. Therefore the maps $\mathbf{g}_k$ are continuous.

Now consider the sets G_k. They are closed as the continuous images of the closed set $\mathbf{f}(G)$. Hence they are measurable. As in union they constitute the whole G, there exists k such that

$$\lambda G_k > 0.$$

Lemma 2.2 *The map* $\mathbf{g} = \mathbf{g}_k$ *is differentiable on* $H = \mathbf{f}(G)$ *and its derivative*

$$\mathbf{g}'(\mathbf{y}) = [\mathbf{f}'(\mathbf{g}(\mathbf{y}))]^{-1}.$$

PROOF OF THE LEMMA. For simplicity we prove differentiability at a point $\mathbf{0}$ assuming $\mathbf{g}(\mathbf{0}) = \mathbf{0}$. Denote $\mathbf{L} = \mathbf{f}'(\mathbf{0})$. We have to show that

$$\mathbf{g}(\mathbf{y}) = \mathbf{L}^{-1}\mathbf{y} + \beta(\mathbf{y}), \qquad \frac{\|\beta(\mathbf{y})\|}{\|\mathbf{y}\|} \to 0, \quad \mathbf{y} \to \mathbf{0},$$

i.e. that

$$\frac{\|\mathbf{g}(\mathbf{y}) - \mathbf{L}^{-1}\mathbf{y}\|}{\|\mathbf{y}\|} \to 0, \quad \mathbf{y} \to \mathbf{0}.$$

By continuity of $\mathbf{f}$ and $\mathbf{g}$ the latter is equivalent to

$$\frac{\|\mathbf{x} - \mathbf{L}^{-1}\mathbf{f}(\mathbf{x})\|}{\|\mathbf{f}(\mathbf{x})\|} \to 0, \quad \mathbf{x} \to \mathbf{0}$$

(we have introduced the change of variables $\mathbf{x} = \mathbf{g}(\mathbf{y})$). By definition of a derivative we have

$$\frac{\|\mathbf{x} - \mathbf{L}^{-1}\mathbf{f}(\mathbf{x})\|}{\|\mathbf{f}(\mathbf{x})\|} = \frac{\|\mathbf{x} - \mathbf{L}^{-1}(\mathbf{L}\mathbf{x} + \alpha(\mathbf{x}))\|}{\|\mathbf{L}\mathbf{x} + \alpha(\mathbf{x})\|}, \qquad \frac{\|\alpha(\mathbf{x})\|}{\|\mathbf{x}\|} \to 0, \quad \mathbf{x} \to \mathbf{0},$$

which is equal to

$$\frac{\|-\mathbf{L}^{-1}\alpha(\mathbf{x})\|}{\|\mathbf{L}\mathbf{x}+\alpha(\mathbf{x})\|} \leq \frac{\|\mathbf{L}^{-1}\|\cdot\|\alpha(\mathbf{x})\|}{\left(\inf\limits_{\|\mathbf{x}\|=1}\mathbf{L}\mathbf{x}\right)\cdot\|\mathbf{x}\|+\|\alpha(\mathbf{x})\|} =$$

$$= \frac{\|\mathbf{L}^{-1}\|\cdot\|\alpha(\mathbf{x})\|/\|\mathbf{x}\|}{\inf\limits_{\|\mathbf{x}\|=1}\mathbf{L}\mathbf{x}+\|\alpha(\mathbf{x})\|/\|\mathbf{x}\|} \to 0, \quad \mathbf{x}\to\mathbf{0},$$

This proves the statement of the lemma.

Note that on the set H

$$|\det\mathbf{g}'(\mathbf{y})| = |\det[\mathbf{f}'(\mathbf{g}(\mathbf{y}))]^{-1}| = |[\det\mathbf{f}'(\mathbf{g}(\mathbf{y}))]^{-1}| < 1/C.$$

Now prove that $\lambda H > 0$. Suppose the contrary: $\lambda H = 0$. Then take an arbitrary $\varepsilon > 0$ and cover H with a system of parallelepipeds $\{W_l\}$ of total volume

$$\sum_l \lambda T_l < \varepsilon.$$

Further, we construct around every point $\mathbf{y} \in H$ a parallelepiped $Z_{\mathbf{y}}$ with rational summits entirely lying in the parallelepiped $W_l \ni \mathbf{y}$ and in U, a neighborhood of $\mathbf{y}$ such that for any finite collection of points $\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_n \in U \cap H$

$$|V(\langle\mathbf{g}(\mathbf{y}),\mathbf{g}(\mathbf{y}_1),\ldots,\mathbf{g}(\mathbf{y}_N)\rangle) - \det\mathbf{g}'(\mathbf{y})V(\langle\mathbf{y},\mathbf{y}_1,\ldots,\mathbf{y}_N\rangle)| <$$

$$< |\det\mathbf{g}'(\mathbf{y})|\left(\max_n\|\mathbf{y}_n-\mathbf{y}\|\right)^N.$$

Such U exists according to Lemma 1.1.

The system $\{Z_{\mathbf{y}}\}$ is a countable covering of the set H and may be denoted by $\{Z_k\}$. So by complete analogy with the proof of Theorem 1 we can construct a disjoint system of cubes $\{R_m\}$ covering H and entirely lying in the union of the parallelepipeds W_l (hence, of total volume not more than ε). Then for any finite collection of points $\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_N \in R_m \cap H$

$$|V(\langle\mathbf{g}(\mathbf{z}_m),\mathbf{f}(\mathbf{y}_1),\ldots,\mathbf{f}(\mathbf{y}_N)\rangle) - \det\mathbf{g}'(\mathbf{z}_m)V(\langle\mathbf{z}_m,\mathbf{y}_1,\ldots,\mathbf{y}_N\rangle)| <$$

$$< |\det\mathbf{g}'(\mathbf{z}_m)N^{N/2}|\lambda R_m, \quad z_m \in H.$$

Therefore,

$$|V(\langle\mathbf{g}(\mathbf{z}_m),\mathbf{g}(\mathbf{y}_1),\ldots,\mathbf{f}(\mathbf{y}_N)\rangle)| \leq$$

$$\leq |\det\mathbf{g}'(\mathbf{z}_m)|N^{N/2}\lambda R_m + |\det\mathbf{g}'(\mathbf{z}_m)||V(\langle\mathbf{z}_m,\mathbf{y}_1,\ldots,\mathbf{y}_N\rangle)| \leq$$

$$\leq |\det\mathbf{g}'(\mathbf{z}_m)|N^{N/2}\lambda R_m + |\det\mathbf{g}'(\mathbf{z}_m)|\lambda R_m =$$

$$= |\det\mathbf{g}'(\mathbf{y})|(N^{N/2}+1)\lambda R_m \leq 1/C(N^{N/2}+1)\lambda R_m.$$

This implies, again in analogy with the proof of Theorem 1, that for an arbitrary finite collection of points $\mathbf{y}_1, \dots, \mathbf{y}_{N+1} \in \mathbf{g}(R_m \cap H)$ the volume of the simplex with summits $\mathbf{y}_n$ does not exceed $(N+1)/C(N^{N/2}+1)\lambda R_m$. Then by Lemma 1.2

$$\lambda\mathbf{g}(R_m \cap H) \leq \theta_N(N+1)/C(N^{N/2}+1)\lambda R_m.$$

Since $\{R_m\}$ covers H,

$$\lambda G_k = \lambda\mathbf{g}(H) \leq \sum_m \lambda\mathbf{g}(R_m \cap H) \leq \theta_N(N+1)/C(N^{N/2}+1)\sum_m \lambda R_m \leq$$

$$\leq \theta_N(N+1)/C(N^{N/2}+1)\varepsilon.$$

Therefore, since ε can be chosen arbitrarily,

$$\lambda G_k = 0,$$

which leads to a contradiction. Thus

$$\lambda H > 0.$$

But $H = \mathbf{f}(G_k) \subset \mathbf{f}(G) \subset \mathbf{f}(K) \subset \mathbf{f}(A)$. This implies that

$$\lambda\mathbf{f}(A) > 0. \qquad \blacksquare$$

Theorems 1 and 2 can be combined into a criterion for absolute continuity of the induced measure.

Theorem 3 *Given an almost everywhere differentiable map* $\mathbf{f}$ *of a region* $Q \subset \mathbf{R}^N$ *into* $\mathbf{R}^N$, *let* D *be its critical set. Then for absolute continuity of the induced measure* $\lambda_\mathbf{f}$ *it is necessary and sufficient that*

$$\lambda D = 0.$$

PROOF. First prove the necessity. Let the measure $\lambda_\mathbf{f}$ be absolutely continuous. By Theorem 1

$$\lambda\mathbf{f}(D) = 0.$$

Then the absolute continuity implies that

$$\lambda D \leq \lambda\mathbf{f}^{-1}(\mathbf{f}(D)) = \lambda_\mathbf{f}(\mathbf{f}(D)) = 0,$$

i.e.

$$\lambda D = 0.$$

Now prove sufficiency. Let $\lambda D = 0$. Suppose the induced measure $\lambda_\mathbf{f}$ is not absolutely continuous, i.e. there exists a set B, $\lambda B = 0$, such that

$$\lambda_\mathbf{f}(B) = \lambda\mathbf{f}^{-1}(B) > 0.$$

Since $\mathbf{f}$ is almost everywhere differentiable and $\lambda D = 0$, the part K of the set $\mathbf{f}^{-1}(B)$ on which $\mathbf{f}$ is differentiable and Jacobian zero is also of positive measure

$$\lambda K = \lambda \mathbf{f}^{-1}(B) > 0.$$

But then by Theorem 2

$$\lambda \mathbf{f}(K) > 0,$$

which implies that

$$\lambda B \geq \lambda \mathbf{f}(K) > 0.$$

This is a contradiction, hence the measure $\lambda_{\mathbf{f}}$ is absolutely continuous. ■

Case II: $M > N$

We prove Theorem 4.

Theorem 4 *Given a map* $\mathbf{f}$ *of a region* $Q \subset \mathbf{R}^N$ *into* $\mathbf{R}^M$, $M > N$, *let* $\mathbf{f}$ *be everywhere differentiable on a set* $A \subset Q$. *Then*

$$\lambda \mathbf{f}(A) = 0.$$

Proof. Introduce the map $\mathbf{g} : Q \times \mathbf{R}^{M-N} \to \mathbf{R}^M$:

$$\mathbf{g}((x^1, \ldots, \mathbf{x}^M)) = \mathbf{f}((x^1, \ldots, x^N)).$$

Obviously the map $\mathbf{g}$ is differentiable on the set $A \times \mathbf{R}^{M-N}$ (by differentiability of $\mathbf{f}$ on A) and its Jacobian $\det \mathbf{g}' = 0$ as including several zero columns. Therefore the set $A \times \mathbf{R}^{M-N}$ is contained in the critical set of the map $\mathbf{g}$ and hence by Theorem 1

$$\lambda \mathbf{f}(A) = \lambda \mathbf{g}(A \times \mathbf{R}^{M-N}) = 0. \quad ■$$

Theorem 4 implies that the induced measure may not be absolutely continuous in the case of $M > N$ if the map $\mathbf{f}$ is differentiable on a set of positive measure.

Case III: $M < N$

We prove Theorem 5.

Theorem 5 *Given a map* $\mathbf{f}$ *of a region* $Q \subset \mathbf{R}^N$ *into* $\mathbf{R}^M$, $M < N$, *let* D *be its critical set (the set of points at which* $\mathbf{f}$ *is differentiable and with* $\operatorname{rank} \mathbf{f}' < M$*). Let* $\mathbf{f}$ *be everywhere differentiable on a set* $A \subset Q$, $\lambda A > 0$, *disjoint with* D. *Then the measure of the image of* A

$$\lambda \mathbf{f}(A) > 0.$$

PROOF. Suppose the contrary, i.e. that

$$\lambda\mathbf{f}(A) = 0.$$

Choose an arbitrary set of axes in the space $\mathbf{R}^N$: $n_1 < n_2 < \ldots < n_{N-M}$. Consider the section of the set A

$$A(a_1, \ldots, a_{N-M}) = \{(x^{\bar{n}_1}, \ldots, x^{\bar{n}_M}) \in \mathbf{R}^M \mid \quad \mathbf{x} \in A, x^{n_m} = a_m\},$$

where $\bar{n}_1 < \bar{n}_2 < \ldots < \bar{n}_M$ is a collection of indices complementary to $n_1 < n_2 < \ldots < n_{N-M}$. Denote by $\mathbf{g}_{a_1,\ldots,a_{N-M}} : Q(a_1, \ldots, a_{N-M}) \to \mathbf{R}^M$ the map

$$\mathbf{g}_{a_1,\ldots,a_{N-M}}((x^{\bar{n}_1}, \ldots, x^{\bar{n}_M})) = \mathbf{f}(\mathbf{x}).$$

Obviously it is differentiable on $A(a_1, \ldots, a_{N-M})$ and its Jacobian $\det \mathbf{g}'_{a_1,\ldots,a_{N-M}}$ is equal to $(\bar{n}_1, \ldots, \bar{n}_M)$-minor of the derivative $\mathbf{f}'$.

The measure of the image of the section of A

$$\lambda\mathbf{g}_{a_1,\ldots,a_{N-M}}(A(a_1, \ldots, a_{N-M})) \leq \lambda\mathbf{f}(A) = 0.$$

Therefore the Jacobian $\det \mathbf{g}'_{a_1,\ldots,a_{N-M}} = 0$ almost everywhere on $A(a_1, \ldots, a_{N-M})$ (otherwise the measure of the image of that part of $A(a_1, \ldots, a_{N-M})$ on which $\det \mathbf{g}'_{a_1,\ldots,a_{N-M}} \neq 0$ would have been positive, which contradicts with the fact that the image $\mathbf{g}_{a_1,\ldots,a_{N-M}}(A(a_1, \ldots, a_{N-M}))$ is a null-set). But then $(\bar{n}_1, \ldots, \bar{n}_M)$-minor of $\mathbf{f}'$ $\det \mathbf{f}'_{\bar{n}_1,\ldots,\bar{n}_M} = 0$ almost everywhere on A, because

$$\lambda(\{\det \mathbf{f}'_{\bar{n}_1,\ldots,\bar{n}_M} = 0\} \cap A) =$$

$$= \int\limits_{\mathbf{R}^{N-M}} \lambda(\{\det \mathbf{g}'_{a_1,\ldots,a_{N-M}} = 0\} \cap A(a_1, \ldots, a_{N-M}))\lambda(da_1)\ldots\lambda(da_{N-M}) =$$

$$= \int\limits_{\mathbf{R}^{N-M}} 0 \cdot \lambda(da_1)\ldots\lambda(da_{N-M}) = 0.$$

So any minor of the derivative $\mathbf{f}'$ is zero almost everywhere on A. Hence $\operatorname{rank}\mathbf{f}' < M$ almost everywhere on A as the countable union of null-sets is a null-set. We come to a contradiction with the fact that $\operatorname{rank}\mathbf{f}' = M$ everywhere on A. Hence

$$\lambda\mathbf{f}(A) > 0. \qquad \blacksquare$$

The version of Theorem 5 for the case of $N = \infty$ can be proved in a similar way.

Theorem 5 can be reformulated as a sufficient condition for absolute continuity of the induced measure:

Theorem 6 *Given an almost everywhere differentiable map* $\mathbf{f}$ *of a region* $Q \subset \mathbf{R}^N$ *(resp.* $\mathbf{R}^\infty$*) into* $\mathbf{R}^M$, $M < N$, *let* D *be its critical set. Then for absolute continuity of the induced measure* $\lambda_\mathbf{f}$ *it is sufficient that*

$$\lambda D = 0.$$

Bibliography

1. A.N. Kolmogorov, S. V. Fomin. Elements of the Theory of Functions and Functional Analysis. Moscow, "Nauka", 1968. in Russian.
2. J.C. Oxtoby. Measure and Category, Moscow, "Mir", 1971. in Russian.

Andrew Zdorovtsev.
Graduated from the physical and mathematical school No. 239 in 1993. Student of the Dept. of Computer Technology since 1993. Winner of St.Petersburg city school olympiads in physics and mathematics in 1987–1993 and students mathematical olympiads in 1993-1996. Absolute winner of the national students olympiad in mathematics in 1995. "Soros student" in 1995 and 1996.

Convergence Types of Series of Functions

D. Ilchenko, A. Zdorovtsev

Series of functions

Definition 1 An infinite series $\sum\limits_{i=1}^{\infty} u_i(x)$ where $u_1, u_2, \dots u_n, \dots$ are the given functions of an independent variable $x \in \mathbf{R}$, called terms of series, is called *series of functions*.

Fixed x, series of functions becomes an ordinary numerical series. Therefore it is possible to consider functional series as a mapping of a set of admissible x into a set of numerical series.

Definition 2 A finite sum of functions $\sum\limits_{i=1}^{N} u_i(x)$ is called a *partial sum* $S_N(x)$ of series of functions.

Definition 3 If for each $x \in [a, b]$ the given series of functions converges (as a numerical series) we say that it *converges everywhere on a segment* $[a, b]$. In this case by the sum of such series we mean the sum of a corresponding numerical series considered as function of x, adopting the notation

$$f(x) = \sum_{i=1}^{\infty} u_i(x).$$

Clearly,

$$f(x) = \lim_{N \to \infty} S_N(x).$$

Definition 4 The difference $f(x) - S_N(x)$ of series,convergent everywhere on $[a, b]$ is called the *remainder* $R_N(x)$ of such series.

The most important question there reads as follows: does sum of a series $f(x)$ preserve the properties of partial sums $S_N(x)$? E.g. we know that a sum of a finite number of continuous functions is a continuous function itself. But can we state the continuity of this sum $f(x)$ on a segment $[a, b]$, knowing that each $u_i(x)$ of a given series is continuous on this segment, or should one require something extra?

Convergence types of the series of functions

Definition 5 Series of functions $\sum\limits_{i=1}^{\infty} u_i(x)$, is called *tamely convergent* on a segment $[a,b]$, if there exist a numerical series with positive terms $\sum\limits_{i=1}^{\infty} \varepsilon_i$, which majorate the absolute values of the corresponding terms of series of functions for all $x \in [a,b]$, i.e.

$$\exists \{\varepsilon_n\}_{n=1}^{\infty} \quad \sum_{n=1}^{\infty} \varepsilon_n \quad \text{is convergent and} \quad \forall n \in \mathbf{N} \quad |u_n(x)| < \varepsilon_n \quad \text{on} \quad [a,b].$$

Definition 6 *Grouping of the series of functions* $\sum\limits_{i=1}^{\infty} u_i(x)$ is called an operation that maps this series into the series

$$\sum_{i=1}^{\infty} U_i(x),$$

in which $U_k(x) = \sum\limits_{i=c_{k-1}+1}^{c_k} u_i(x)$, where $c_0 = 0$, and $\{c_1\}_{j=1}^{\infty}$ is increasing sequence of natural numbers.

Definition 7 Series of functions $\sum\limits_{i=1}^{\infty} u_i(x)$, convergent everywhere on a segment $[a,b]$ is called *generalized tamely convergent* on a segment $[a,b]$, if there is a grouping that maps them into tamely convergent series $\sum\limits_{i=1}^{\infty} U_i(x)$ on a segment $[a,b]$.

Definition 8 Series of functions $\sum\limits_{i=1}^{\infty} u_i(x)$ is called *uniformly convergent* on segment $[a,b]$, if for each positive ε there exist $N \in \mathbf{N}$, such that each part $\sum\limits_{i=p}^{q} u_i(x), \quad p \leq q$ of the considered series is less in absolute value than ε, for $p > N$, i. e.

$$\forall \varepsilon > 0 \quad \exists N \in \mathbf{N} \quad \forall p,q : N < p \leq q \quad \left| \sum_{i=p}^{q} u_i(x) \right| < \varepsilon \quad \text{on} \quad [a,b].$$

Definition 9 Series of functions $\sum\limits_{i=1}^{\infty} u_i(x)$, convergent everywhere on a segment $[a,b]$ is called *generalized uniformly convergent* on a segment $[a,b]$, if for each positive $\varepsilon > 0$ there exist an infinite set of $N \in \mathbf{N}$, for which the remainders $R_N(x)$ of the series for all $x \in [a,b]$ is less than ε in absolute value, i.e.

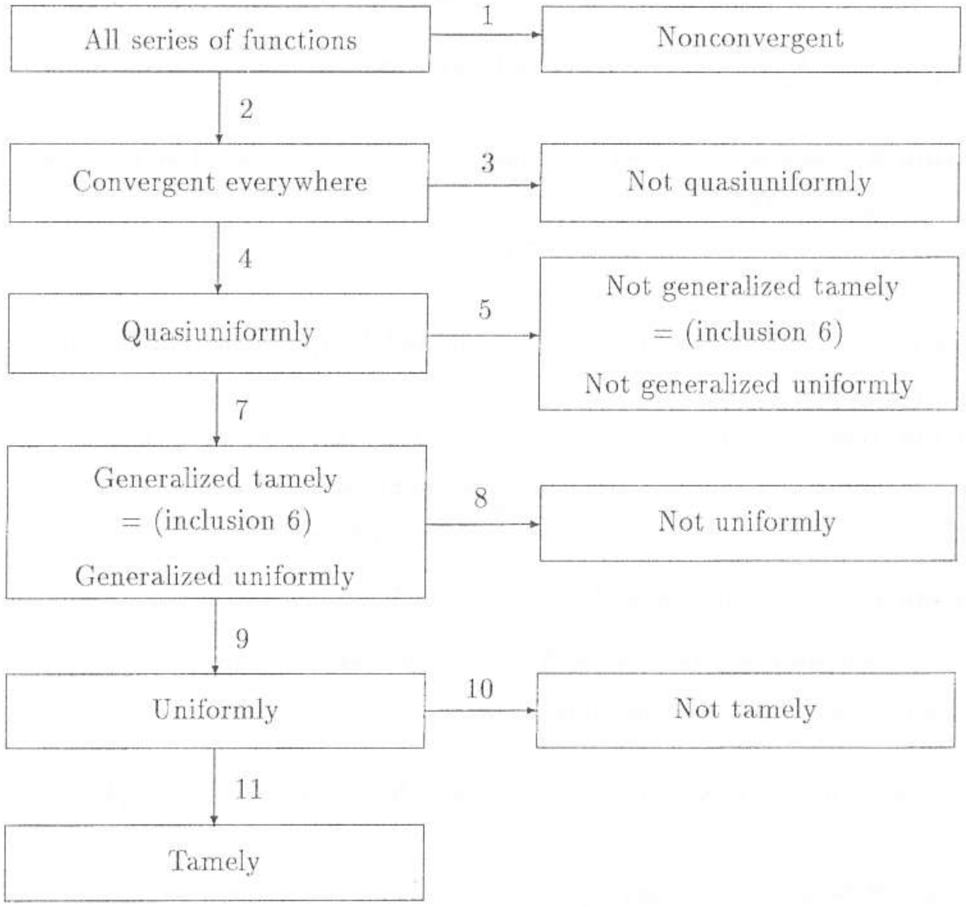
$$\forall \varepsilon > 0 \quad \{N \quad : \quad |R_N(x)| < \varepsilon \quad \text{on} \quad [a,b]\} \quad \text{is denumerable.}$$

Definition 10 Series of functions $\sum\limits_{i=1}^{\infty} u_i(x)$, convergent everywhere on a segment $[a,b]$ is called *quasiuniformly convergent* on a segment $[a,b]$, if for each $\varepsilon > 0$ and for each $m \in \mathbf{N}$ there exist such $M\ in\mathbf{N}, M > m$, that for each $x \in [a,b]$, there is $N \in \mathbf{N}, m \leq N \leq M$, for which the remainder $R_N(x)$ is less than ε in absolute value, i. e.

$$\forall \varepsilon > 0 \quad \forall m \in \mathbf{N} \quad \exists M > m, M \in \mathbf{N} \quad \forall x : a \leq x \leq b$$
$$\exists N \in \mathbf{N} \quad m \leq N \leq M, \quad |R_N(x)| < \varepsilon.$$

The relations between the different types of convergence of series of functions

The relations between the different types of series of functions convergence are represented on a scheme below.



To prove the above scheme we state some theorems and examples. The inclusions 2, 4 follow from the definitions. The inclusions 1 and 12 are self-evident.

Theorem 1 (inclusion 6) *Series of functions* $\sum\limits_{i=1}^{\infty} u_i(x)$ *is generalized tamely convergent on a segment* $[a,b]$, *if and only if it is generalized uniformly convergent on this segment.*

Proof. 1) **"if"** Because of the generalized tame convergence there exists tamely convergent series $\sum\limits_{i=1}^{\infty} U_i(x)$, obtained by grouping of the given series. Thus there exists a series of positive terms $\sum\limits_{i=1}^{\infty} \varepsilon_i$, satisfying $|U_k(x)| < \varepsilon_k$. Suppose $\varepsilon > 0$. The remainders of the convergent series $\sum\limits_{i=K+1}^{\infty} \varepsilon_i \to 0$, as $K \to \infty$, hence

$$\exists M : \quad \sum_{i=K+1}^{\infty} \varepsilon_i < \varepsilon \quad \forall K > M.$$

Denote the remainders $\sum\limits_{i=K+1}^{\infty} U_i(x)$ by $r_K(x)$. Then for $K > M$

$$|r_K(x)| \leq \sum_{i=K+1}^{\infty} |U_i| \leq \sum_{i=K+1}^{\infty} \varepsilon_i < \varepsilon.$$

It it clear that $r_K(x) = R_{c_K}(x)$. Hence for all $K > M$ $\quad |R_{c_K}(x)| < \varepsilon$. It follows that $\{N : |R_N(x)| < \varepsilon \quad \text{on} \quad [a,b]\}$ includes $\{c_{M+1}, c_{M+2}, \ldots\}$, i. e. is infinite. Thus $\sum\limits_{i=1}^{\infty} u_i(x)$ is generalized uniformly convergent on $[a,b]$.

2) "only if" Consider an arbitrary convergent numerical series with positive terms $\sum\limits_{i=1}^{\infty} \varepsilon_i$ Because of the generalized uniform convergence there exist an infinite set of natural numbers j_{kK}, where $k, K = 1, 2, \ldots$, such that $|R_{j_{kK}}(x)| < \varepsilon_k \quad \text{on} \quad [a,b]$. Assume $c_0 = 0, c_1 = j_{11}$, and

$$c_k = \min\{j_{kK} \quad : \quad j_{kK} > c_{k-1}\} \quad \forall k > 1.$$

Then $|R_{c_k}(x)| < \varepsilon_k$. We group the terms of the series according to the increasing sequence $c_0, c_1, c_2, \ldots$ In addition

$$|U_k(x)| = |R_{c_{k-1}}(x) - R_{c_k}(x)| \leq |R_{c_{k-1}}(x)| + |R_{c_k}(x)| < \varepsilon_{k-1} + \varepsilon_k.$$

The series with terms $v_k = \varepsilon_{k-1} + \varepsilon_k$, is evidently convergent, because so is $\sum\limits_{i=1}^{\infty} \varepsilon_i$. Therefore the series $\sum\limits_{i=1}^{\infty} U_i(x)$ is tamely convergent on a segment $[a,b]$, and the given series is generalized tamely convergent. ■

Theorem 2 (inclusion 7) *If a series of functions $\sum\limits_{i=1}^{\infty} u_i(x)$ is generalized uniformly convergent on a segment $[a,b]$, then it is quasiuniformly convergent on this segment.*

Proof. Suppose $\varepsilon > 0$, $m \in \mathbf{N}$. Because of the generalized uniform convergence on $[a,b]$ there exist an infinite number of $N \in \mathbf{N}$, satisfying $|R_N(x)| < \varepsilon$. In particular, $\exists N > m$. Assume $M = N$. We will have:

$$\exists M \in \mathbf{N} \quad \forall x : a \leq x \leq b \quad |R_N(x)| < \varepsilon, \quad m \leq N$$

and therefore

$$\exists M > m, M \in \mathbf{N} \quad \forall x : a \leq x \leq b \quad \exists N \in \mathbf{N} \quad m \leq N \leq M \quad |R_N(x)| < \varepsilon.$$

Because of the arbitrariness of ε and m, the series is quasiuniformly convergent on $[a,b]$. ■

Theorem 3 (inclusion 9) *If a series of functions $\sum\limits_{i=1}^{\infty} u_i(x)$ uniformly converge on $[a,b]$, then it is generalized uniformly convergent on this segment.*

PROOF. Suppose $\varepsilon > 0$. Because of the uniform convergence

$$\exists N \in \mathbf{N} \quad \forall p,q \in \mathbf{N} : N < p \leq q \quad \left|\sum_{i=p}^{q} u_i(x)\right| < \varepsilon \quad \text{on} \quad [a,b].$$

The series is convergent according to the Cauchy criterion, thus we can speak about the remainders. Passing to the limit for $q \to \infty$, we obtain: $|R_{p-1}(x)| \leq \varepsilon$. Because there are infinite number of p, such that $p > N$, the series is clearly generalized uniformly convergent on a segment $[a,b]$. ■

Theorem 4 (inclusion 11) *If a series of functions $\sum\limits_{i=1}^{\infty} u_i(x)$ is tamely convergent on a segment $[a,b]$, then it is uniformly convergent on this segment.*

PROOF. Suppose $\varepsilon > 0$. Because of the tame convergence there exists a convergent series $\sum\limits_{i=1}^{\infty} \varepsilon_i$ with positive terms, such that $|u_n(x)| < \varepsilon_n$. The remainders of convergent series $\sum\limits_{i=M+1}^{\infty} \varepsilon_i \to 0$ as $M \to \infty$, therefore,

$$\exists N \in \mathbf{N} \quad : \quad \sum_{i=M+1}^{\infty} \varepsilon_i < \varepsilon, \quad \forall M > N, M \in \mathbf{N}.$$

Now assume $N < p \leq q, p, q \in \mathbf{N}$. Then

$$\left|\sum_{i=p}^{q} u_i(x)\right| \leq \sum_{i=p}^{q} |u_i(x)| < \sum_{i=p}^{q} \varepsilon_i \leq \sum_{i=p}^{\infty} \varepsilon_i < \varepsilon.$$

Hence the series uniformly converges on a segment $[a,b]$. ■

Definition 11 Consider a function $W(s,t,w)$ defined on a segment $[a,b]$, so that $W(s,t,w) = 0$ for each $x \notin (s,t)$, $W(s,t,w) = w$ for $\frac{s+t}{2}$ (the middle of the interval) and is linear on both left and right halves of the segment $[s,t]$.
Thus the graph of $W(s,t,w)(x)$ has a shape of an isosceles triangle, built on the segment $[s,t]$. Clearly this function is continuous everywhere on the set of its definition.

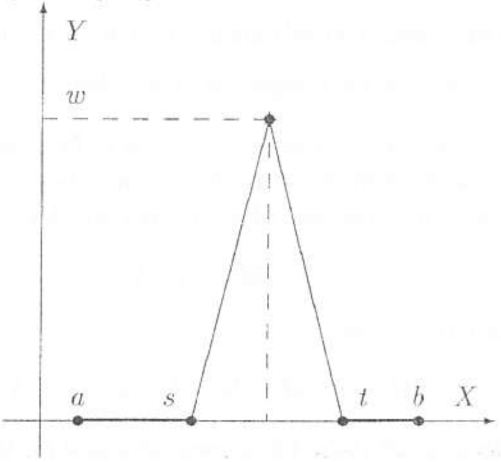


Fig. 1. The graph of $W(s,t,w)$.

Example (inclusion 3) Consider the series $\sum\limits_{i=1}^{\infty} u_i(x)$ on a segment $[0,1]$ where

$$u_n(x) = W\left(\frac{1}{n+1}, \frac{1}{n}, 1\right)(x).$$

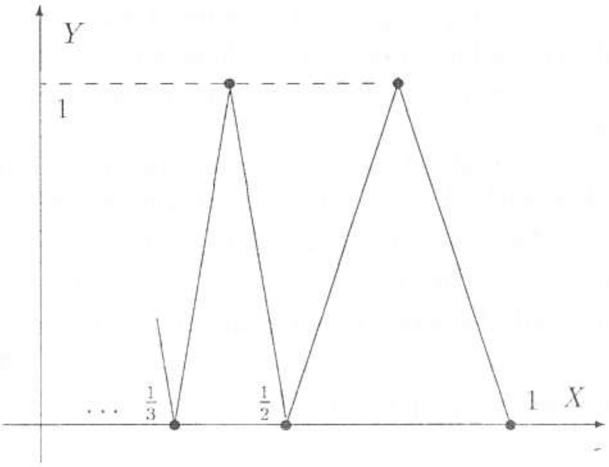


Fig.2. The graph of the sum of the series. Example for inclusion 3.

1) We prove that it is convergent everywhere on this segment. Each $x \in [0,1]$ belongs to no more than one interval of the type $\left(\frac{1}{n+1}, \frac{1}{n}\right)$, $n = 1, 2, \ldots$ Thus for each $x \in [a,b]$ in $\sum\limits_{i=1}^{\infty} u_i(x)$ series will contain no more than one nonzero term. Therefore it is convergent on $[0,1]$.

2) We now prove that it does not converge quasiuniformly on a given segment. Suppose $\varepsilon = \frac{1}{2}, m = 1$, Take an arbitrary $M > m$. Set $x = \frac{1}{2}\left(\frac{1}{2M+1} + \frac{1}{2M}\right)$. Then

$$u_1(x) = 0, u_2(x) = 0, \ldots u_{2M-1}(x) = 0, u_{2M}(x) = 1, u_{2M+1}(x) = 0, \ldots$$

Hence the sum of series is equal to 1, and the remainders

$$R_1(x) = 1, R_2(x) = 1, \ldots R_{2M-1}(x) = 1, R_{2M}(x) = 0, R_{2M+1}(x) = 0, \ldots$$

Then it is evident that for no N between m and M will $|R_N(x)|$ be less than $\varepsilon = \frac{1}{2}$, for $|R_N(x)| = 1$. Since M is arbitrary, $\sum\limits_{i=1}^{\infty} u_i(x)$ is not quasiuniformly convergent. ■

Example (inclusion 5) Consider the series $\sum\limits_{i=1}^{\infty} u_i(x)$ on a segment [0,1], where

$$u_n(x) = W\left(\frac{1}{n+1}, \frac{1}{n}, 1\right)(x) - W\left(\frac{1}{n+2}, \frac{1}{n+1}, 1\right)(x).$$

1) We prove that it converges quasiuniformly on this segment. The partial sums

$$S_N(x) = \left[W\left(\frac{1}{2}, 1, 1\right)(x) - W\left(\frac{1}{3}, \frac{1}{2}, 1\right)(x)\right] + \left[W\left(\frac{1}{3}, \frac{1}{2}, 1\right)(x) - \right.$$

$$\left. W\left(\frac{1}{4}, \frac{1}{3}, 1\right)(x)\right] + \ldots + \left[W\left(\frac{1}{N+1}, \frac{1}{N}, 1\right)(x) - W\left(\frac{1}{N+2}, \frac{1}{N+1}, 1\right)(x)\right] =$$

$$W\left(\frac{1}{2}, 1, 1\right)(x) - W\left(\frac{1}{N+2}, \frac{1}{N+1}, 1\right)(x).$$

Each $x \in [0,1]$ belongs to no more than one interval of the type $\left(\frac{1}{N+1}, \frac{1}{N}\right)$, $N \in \mathbf{N}$. Thus for each x in sequence $\{S_j(x)\}_{j=1}^{\infty}$ we will encounter no more than one term not equal to $W(\frac{1}{2}, 1, 1)(x)$. Hence this sequence tends to $f(x) = W(\frac{1}{2}, 1, 1)(x)$. Then the remainders $R_N(x) = f(x) - S_N(x) = W(\frac{1}{N+2}, \frac{1}{N+1}, 1)(x)$. Now suppose $\varepsilon > 0$, $m \in \mathbf{N}$. Set $M = m+1$. Assume $x \in [0,1]$. Then x will get into no more than to one interval of the type $\left(\frac{1}{m+2}, \frac{1}{m+1}\right)$, $\left(\frac{1}{m+3}, \frac{1}{m+2}\right) = \left(\frac{1}{M+2}, \frac{1}{M+1}\right)$. Therefore at least one of $R_m(x), R_{m+1}(x) = R_M(x)$ will be equal to 0, and less than ε, i. e. there will exist such $N \in \mathbf{N}$, $n < N < M$ that $|R_N(x)| < \varepsilon$. Since m, x and ε are arbitrary, the series is quasiuniformly convergent on a segment $[0,1]$.

2) We show that the series is not generalized uniformly convergent on the given segment. Suppose $\varepsilon = \frac{1}{2}$. For each $N \in \mathbf{N}$ the remainder $R_N(x) = W\left(\frac{1}{N+2}, \frac{1}{N+1}, 1\right)\left(\frac{1}{2}\left(\frac{1}{N+2} + \frac{1}{N+1}\right)\right) = 1$. Therefore it can not be less than $\varepsilon = \frac{1}{2}$ (in absolute value) everywhere on the segment $[0,1]$. Therefore, $\{N : |R_N(x)| < \varepsilon \quad \text{on} \quad [0,1]\}$ is numerable (it is void), and the series is not generalized uniformly convergent on $[0,1]$. ■

Example (inclusion 8) Consider the series $\sum\limits_{i=1}^{\infty} u_i(x)$, on $[0,1]$, where

$$u_{2k-1}(x) = W\left(\frac{1}{k+1}, \frac{1}{k}, 1\right)(x), u_{2k}(x) = -u_{2k-1}(x), \qquad k = 1, 2, \ldots$$

1) We prove that it is generalized uniformly convergent on this segment. The partial sums

$$S_{2K-1}(x) = u_1(x) + u_2(x) + \ldots + u_{2K-1}(x) =$$

$$u_1(x) - u_1(x) + u_3(x) - u_3(x) + \ldots + u_{2K-1}(x) = u_{2K-1}(x) = W\left(\frac{1}{K+1}, \frac{1}{K}, 1\right)(x),$$

$$S_{2K}(x) = u_1(x) + u_2(x) + \ldots + u_{2K-1}(x) + u_{2K}(x) =$$

$$u_1(x) - u_1(x) + u_3(x) - u_3(x) + \ldots + u_{2K-1}(x) - u_{2K-1}(x) = 0.$$

Each $x \in [0,1]$ belongs to no more than to one interval of the type $(\frac{1}{K+1}, \frac{1}{K})$, $K \in \mathbf{N}$. Thus for each x in sequence $S_1(x), S_2(x), \ldots, S_{2K-1}(x), S_{2K}(x), \ldots$ there exists no more than one nonzero term. Therefore this sequence tends to $f(x) \equiv 0$. Then the remainders

$$R_{2K-1}(x) = f(x) - S_{2K-1}(x) = -W(\frac{1}{K+1}, \frac{1}{K}, 1)(x),$$

$$R_{2K}(x) = f(x) - S_{2K}(x) = 0.$$

Suppose now $\varepsilon > 0$. Then $|R_{2K}(x)| = |0| = 0 < \varepsilon$ everywhere on $[0,1]$ for all K. I. e. the set $\{N : |R_N(x)| < \varepsilon \quad \text{on} \quad [0,1]\}$ is infinite because it includes all even numbers. since $\varepsilon > 0$ can be chosen arbitrarily, the given series is generalized uniformly convergent on $[0,1]$.

2) Now we prove that it does not converge uniformly on the given segment. Suppose $\varepsilon = \frac{1}{2}$, N is arbitrary. Choose $p = q = N+1 > N$. For all k

$$\left|u_{2k-1}\left(\frac{1}{2}\left(\frac{1}{k+1} + \frac{1}{k}\right)\right)\right| = \left|u_{2k}\left(\frac{1}{2}\left(\frac{1}{k+1} + \frac{1}{k}\right)\right)\right| = 1,$$

hence $\left|\sum\limits_{i=p}^{q} u_i(x)\right| = |u_{N+1}(x)|$ is equal to 1 in a certain point, and therefore is greater than ε. Then it is not true that $\left|\sum\limits_{i=p}^{q} u_i(x)\right| < \varepsilon$ everywhere on $[0,1]$. Thus, by the arbitrariness of N, we can conclude that the series does not converge uniformly on $[0,1]$. ■

Example (inclusion 10) Consider the series $\sum\limits_{i=1}^{\infty} u_i(x)$ on $[0,1]$, where

$$u_n(x) = W\left(\frac{1}{n+1}, \frac{1}{n}, \frac{1}{n}\right)(x), \qquad n = 1, 2, \ldots$$

1) We prove that it converges uniformly on this segment. Suppose $\varepsilon > 0$. Chose $N > \frac{1}{\varepsilon}$, assuming $N < p \leq q$. Each x on $[0,1]$ belongs to no more than one interval of the type $(\frac{1}{n+1}, \frac{1}{n})$ $n \in \mathbf{N}$. Thus for each $x \in [0,1]$ in sequence $u_{N+1}(x), u_{N+2}(x), \ldots$ there exists no more than one nonzero term. Furthermore, if there exist such a nonzero term, it is equal to $W(\frac{1}{m+1}, \frac{1}{m}, \frac{1}{m})$ for a certain $m > N$. Hence it can not be greater than $\frac{1}{m} < \frac{1}{N} < \varepsilon$. Thus

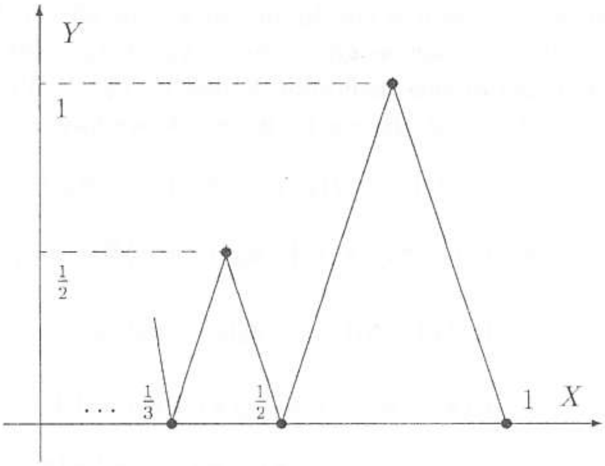


Fig. 3. The sum of the series. Example (inclusion 10).

$$\left|\sum_{i=p}^{q} u_i(x)\right| = \sum_{i=p}^{q} u_i(x) < \varepsilon$$

for each x, and according to the arbitrarity of ε, the given series are uniformly convergent on $[0,1]$.

2) Now we show that it does not converge tamely on the given segment. Suppose $\varepsilon = \frac{1}{2}$. For all n $\left|u_k\left(\frac{1}{2}\left(\frac{1}{n+1} + \frac{1}{n}\right)\right)\right| = u_k\left(\frac{1}{2}\left(\frac{1}{n+1} + \frac{1}{n}\right)\right) = \frac{1}{k}$, hence the series $\sum\limits_{i=1}^{\infty} |u_i(x)|$ can not be majorized by the numerical series less than nonconvergent harmonic series $\sum\limits_{i=1}^{\infty} \frac{1}{i}$ on $[0,1]$. Therefore the series $\sum\limits_{i=1}^{\infty} u_i(x)$ can not be tamely convergent on $[0,1]$.

Continuity criterion for the sum of series of functions

The following important result is known as the Arzelà–Borel theorem.

Theorem 5 (Arzelà–Borel) *Let the series of continuous functions $\sum\limits_{i=1}^{\infty} u_i(x)$ converge everywhere on $[a,b]$. Then its sum is continuous, if and only if the convergence is quasiuniform.*

PROOF.

1) "if" We prove the continuity of $f(x)$ in an arbitrary point of $x_0 \in [a,b]$, i. e. that $f(x) \to f(x_0)$ as $x \to x_0$. Suppose $\varepsilon > 0$. Because of the convergence of the series as $x \to x_0$, remainders $R_N(x_0)$ become less than $\frac{\varepsilon}{3}$. Since convergence is quasiuniform, there exist such $M > m$, that

$$\forall x : a \le x \le b \quad \exists N_x \in \mathbf{N} \quad m \le N \le M \quad |R_N(x)| < \frac{\varepsilon}{3}.$$

Since each term is continuous in x_0, for all $n \in \mathbf{N}$ $u_n(x) \to u_n(x_0)$, $\quad x \to x_0$. Therefore in a certain neighbourhood V_n of x_0 holds $|u_n(x) - u_n(x_0)| < \frac{\varepsilon}{3M}$. Denote the minimum of such neighbourhoods $V_1, V_2, \ldots, V_M$ by V. $|u_n(x) - u_n(x_0)| < \frac{\varepsilon}{3M}$ for all $n = 1, 2, \ldots, M$. Hence for all $x \in V$ we have:

$$|f(x) - f(x_0)| = |S_{N_x}(x) + R_{N_x}(x) - S_{N_x}(x_0) - R_{N_x}(x_0)| \le$$

$$|S_{N_x}(x) - S_{N_x}(x_0)| + |R_{N_x}(x)| + |R_{N_x}(x_0)| < |S_{N_x}(x) - S_{N_x}(x_0)| + \frac{\varepsilon}{3} + \frac{\varepsilon}{3} =$$

$$|u_1(x) - u_1(x_0) + u_2(x) - u_2(x_0) + \ldots + u_{N_x}(x) - u_{N_x}(x_0)| + \frac{2\varepsilon}{3} \le$$

$$|u_1(x) - u_1(x_0)| + |u_2(x) - u_2(x_0)| + \ldots + |u_{N_x}(x) - u_{N_x}(x_0)| + \frac{2\varepsilon}{3} \le$$

$$|u_1(x) - u_1(x_0)| + |u_2(x) - u_2(x_0)| + \ldots +$$

$$|u_{N_x}(x) - u_{N_x}(x_0)| + \ldots + |u_M(x) - u_M(x_0)| + \frac{2\varepsilon}{3} <$$

$$\frac{M\varepsilon}{3M} + \frac{2\varepsilon}{3} = \varepsilon.$$

Thus we found the neighbourhood V of x_0, where $|f(x) - f(x_0)| < \varepsilon$. Therefore, since ε is arbitrary, $f(x)$ indeed tends to $f(x_0)$. Thus $f(x)$ is continuous on $[a,b]$.

2) "only if" Suppose $\varepsilon > 0$, $m \in \mathbf{N}$. Since the series is convergent everywhere on $[a,b]$, for each z on $[a,b]$ there exist such $N_z > m$, that $|R_{N_z}(z)| < \varepsilon$. f is continuous in z, hence R_{N_z} is, too, continuous in z. Therefore there exists a certain neighbourhood V_z of z, in which $|R_{N_z}| < \varepsilon$. Consider the class of neighbourhoods V_z for various z of $[a,b]$. Evidently, each z of this class is covered by at least one of the members of this class (e.g., V_z). This class, therefore, forms the covering of a segment $[a,b]$. According to the Borel lemma, one can select a finite subcovering $V_{z_1}, V_{z_2}, \ldots, V_{z_k}$ of $[a,b]$. Denote $M = \max\{N_{z_1}, N_{z_2}, \ldots, N_{z_k}\}$. Suppose now that $x \in [a,b]$. Then it is covered by a certain element V_{z_i} of this subcovering. Therefore, $|R_{N_{z_i}}(x)| < \varepsilon$. But, clearly, $m < N_{z_i} \le M$. So we have found such N from the Definition 10 between m and M, that $|R_N(x)| < \varepsilon$. Since ε, m and x are arbitrary, the series is quasiuniformly convergent on $[a,b]$. ■

The Arzelà–Borel theorem solves completely the problem of continuity conditions for a sum of series of continuous functions. The relations between the different types of convergence provide a set of evident sufficient conditions on continuity of a sum of a series of continuous functions.

Corollary 5.1 *A sum of a series of continuous on $[a,b]$ functions is continuous, provided that one of the following 4 conditions holds:*

1) The series is generalized uniformly convergent on $[a,b]$.

2) The series is generalized tamely convergent on $[a,b]$.

3) The series is uniformly convergent on $[a,b]$.

4) The series is tamely convergent on $[a,b]$.

One must note, however, that none of these conditions is necessary for the sum to be continuous. This is shown by an example for inclusion 5. In fact, the series of continuous terms $u_n(x) = W(\frac{1}{n+1}, \frac{1}{n}, 1)(x) - W(\frac{1}{n+2}, \frac{1}{n+1}, 1)(x)$, on $[0,1]$ considered there tends to the continuous sum $f(x) = W(\frac{1}{2}, 1, 1)(x)$. Nevertheless, this series on $[0,1]$ does not converge even generalized uniformly (or, as was proven above, generalized tamely). Of course, one can not speak about uniform or tame convergence in this case.

The case of series with constant signs

Observe that in the above examples the series had alternating signs. The sign of term $u_n(x)$ was changing not even with the change of n, but also of x. It turns out that there does not exist any suitable example for series of constant signs, which is claimed by the following theorem.

Theorem 6 *Let the series of continuous nonnegative (nonpositive) functions $\sum\limits_{i=1}^{\infty} u_i(x)$ converge everywhere on $[a,b]$. Then its sum is continuous, if and only if the convergence is uniform.*

Proof. For nonpositive terms, changing signs for all terms we come to nonnegative series. Thus it is sufficient to consider the case $u_n(x) \geq 0$ on $[a,b]$.

1) "if" This part follows directly from the Arzelà–Borel theorem, since uniformly convergent series converge quasiuniformly.

2) "only if" Suppose $\varepsilon > 0$. Partial sums S_N are continuous as finite sums of continuous functions. Therefore, so are remainders $R_N = f - S_N$, as difference of two continuous functions. Remainders of convergent series tend to 0 for all z from a segment $[a,b]$. Hence we can find such N_z, that $R_{N_z}(z) < \varepsilon$. Since functions R_{N_z} are all continuous, there exists such neighbourhood V_z of z, where $R_{N_z}(x) < \varepsilon$. Consider the class of such neighbourhoods V_z for various z from $[a,b]$. Evidently, each z of this class is covered by at least one of the members of this class (e.g., V_z). This class, therefore, forms the covering of a segment $[a,b]$. According to the Borel lemma, one can select a finite subcovering of $V_{z_1}, V_{z_2}, \ldots, V_{z_k}$ of $[a,b]$. Denote $M = \max\{N_{z_1}, N_{z_2}, \ldots, N_{z_k}\}$. Suppose now that $x \in [a,b]$. Then it is covered by a certain element V_{z_i} of this subcovering. Therefore, $|R_{N_{z_i}}(x)| < \varepsilon$. It is clear that sequence R_N is decreasing: $R_{N+1} = R_N - u_{N+1}$. Since $N_{z_i} \leq N$, $R_N < \varepsilon$. Thus, $R_N < \varepsilon$ on $[a,b]$. Then

$$\forall p, q \in \mathbf{N} \quad : \quad N < p \leq q \quad \sum_{i=p}^{q} u_i(x) < \varepsilon \quad \text{on} \quad [a,b],$$

since

$$\sum_{i=p}^{q} u_i(x) < \sum_{i=N+1}^{\infty} u_i(x) = R_N(x).$$

Thus there exists $N \in \mathbf{N}$ from the uniform convergence definition, since we can choose ε arbitrarily, the series is uniformly convergent on $[a,b]$. ■

The theorem proven above looks similar to Arzelà–Borel theorem. Comparing them, we come to an obvious corollary.

Corollary 6.1 *Uniform and quasiuniform convergence definitions coincide in the case of constant-signed series.*

It is clear that in this case all four definitions (quasiuniform, uniform, generalized tame and generalized uniform convergence) coincide.

The following particular case of the corollary considered in the previous chapter can be examined as a corollary of this theorem.

Particular case *The sum of continuous terms series of constant signs on $[a,b]$ is continuous, if the series is tamely convergent on $[a,b]$.*

Again, we must note that this condition is not necessary for the sum to be continuous. This is shown by the example for inclusion 5. In fact, constant-signed series with continuous terms $u_n(x) = W(\frac{1}{n+1}, \frac{1}{n}, \frac{1}{n})(x)$ on the segment $[0,1]$ considered there is uniformly convergent and hence have the continuous sum. However, this series is not tamely convergent on $[0,1]$.

Convergence of series with constant signs

As we have noted in the corollary in above paragraph, the scope of the possible convergence types is highly shrunk for series with constant signs at least for the series with continuous terms. We prove the following more general statement.

Theorem 7 *Quasiuniform, generalized uniform, generalized tame and uniform convergence coincide for series with constant signs.*

PROOF. Without loss of generality, we can make the proof for nonnegative terms series only. Furthermore, according to the scheme of relations between the different types of convergence, it is sufficient to prove only one inclusion: that nonnegative and quasiuniformly convergent series $\sum\limits_{i=1}^{\infty} u_i(x)$ converge uniformly. Suppose $\varepsilon > 0$ and chose $m = 1$. Then thanks to the quasiuniform convergence,

$$\exists M > 1 \quad : \quad \forall x, a \leq x \leq b \quad \exists N_x \leq M, \quad R_{N_x}(x) < \varepsilon.$$

As we noted above, the remainders $R_{N_x}(x)$ of nonnegative series decrease. Therefore, $R_M(x) < \varepsilon$, for each $x \in [a,b]$ since $N_x < M$. Then

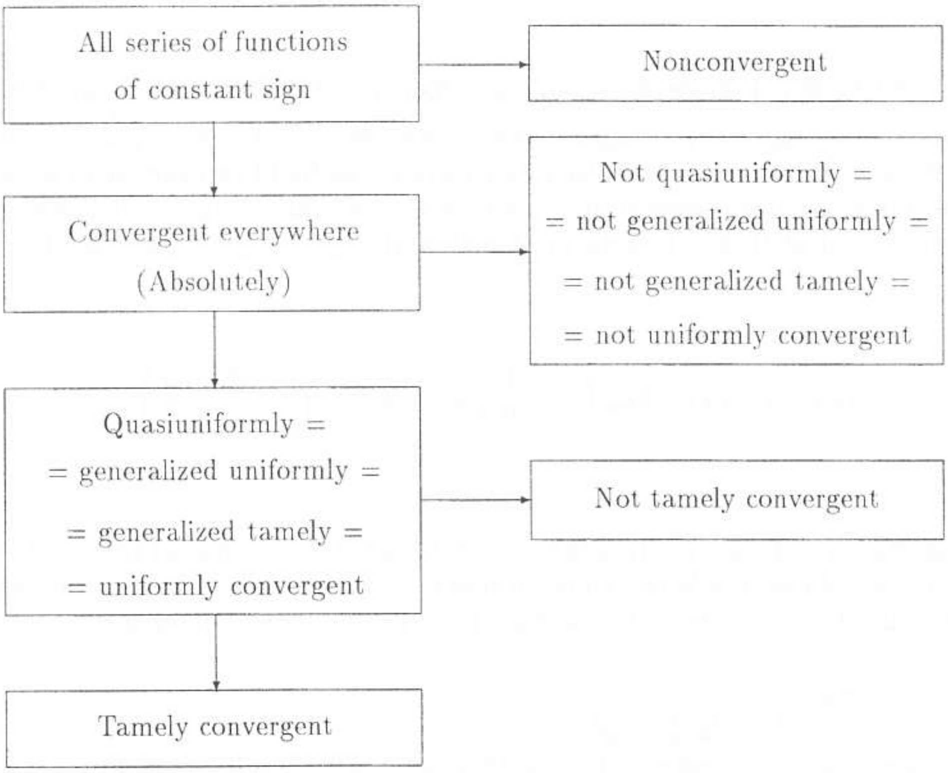
$$\forall p, q \in \mathrm{N} \quad : \quad M < p \leq q \quad \text{holds} \quad \sum_{i=p}^{q} u_i(x) < \varepsilon \quad \text{on} \quad [a,b],$$

because

$$\sum_{i=p}^{q} u_i(x) < \sum_{i=M+1}^{\infty} u_i(x) = R_m(x)$$

Thus according to the definition, since we choose ε arbitrarily, the series is uniformly convergent on $[a,b]$. ■

Nevertheless, the definition of tame convergence even for series of constant signs remains unique. It can be concluded from the Example to inclusion 10. And the Example for inclusion 3 demonstrates that for series with constant signs quasiuniform convergence (as well as the generalized tame and the generalized uniform) does not coincide with an ordinary pointwise convergence. Thus the scheme of relations is modified in the following way for series of constant signs.


All series of functions of constant sign
Nonconvergent
Convergent everywhere (Absolutely)
Not quasiuniformly = = not generalized uniformly = = not generalized tamely = = not uniformly convergent
Quasiuniformly = = generalized uniformly = = generalized tamely = = uniformly convergent
Not tamely convergent
Tamely convergent

Conclusion

It was proven in the above paragraphs that none of the sufficient conditions for continuity of sum of series of functions, derived from the Arzelà–Borel theorem and the

Theorem 6 is necessary. Namely, it is not necessary for the series with continuous terms to be even generalized uniformly (tamely) convergent, for its sum to be continuous. However, here we come to a question: if a series does not have the required type of convergence on the whole interval, maybe it is uniformly (tamely) convergent on a certain subinterval of this interval? It turns out to be sometimes wrong. To illustrate this we will state two examples—for arbitrary series and for series with constant signs.

Example 1 Here we apply an unusual way to construct the series of functions. Instead of presenting each term $u_n(x)$ explicitly we will provide the pointwisely infinitesimal consequence of remainders $R_n(x)$ and show that this consequence in fact define the series of functions convergent everywhere. Define the terms of the series as $u_n(x) = R_{n-1}(x) - R_n(x)$. Then the partial sums

$$S_N(x) = \sum_{i=1}^{N} u_i(x) = \sum_{i=1}^{N}(R_{i-1}(x) - R_i(x)) = R_0(x) - R_N(x)$$

evidently tend to $R_0(x)$, given $N \to \infty$, since $R_N(x) \to 0$. Therefore $\sum\limits_{i=1}^{\infty} u_i(x)$ tends pointwisely to the sum $f(x) = R_0(x)$. The remainders $f(x) - S_N(x)$ are equal to $R_0(x) - (R_0(x) - R_N(x)) = R_N(x)$. Using this method, we build an example of series of continuous functions on an interval $(0, 1)$ which is not generalized uniformly convergent on any subinterval of $(0, 1)$. Denote the function $W(\frac{1}{2^{m-k+1}}, \frac{1}{2^{m-k}}, \frac{1}{2^k})(x)$ by $Y_{km}(x)$, setting

$$I_{km}(x) = Y_{km}(x) + Y_{km}\left(x - \frac{1}{2^k}\right) + \ldots + Y_{km}\left(x - \frac{2^k - 1}{2^k}\right).$$

At last, let $R_K(x) = I_{0,2K}(x) + I_{1,2K}(x) + \ldots + I_{K,2K}(x)$. Then $R_K(x)$ represents the sum of functions of the type $W(x)$, equal to zero everywhere, except for a certain number of intervals. Using the definition of $R_K(x)$, we write out these intervals:

$$\begin{array}{ll}
I_{0,2K}: & (\frac{1}{2^{2K+1}}, \frac{1}{2^{2K}}) \\
I_{1,2K}: & (\frac{1}{2^{2K}}, \frac{1}{2^{2K-1}})(\frac{1}{2} + \frac{1}{2^{2K}}, \frac{1}{2} + \frac{1}{2^{2K-1}}) \\
I_{2,2K}: & (\frac{1}{2^{2K-1}}, \frac{1}{2^{2K-2}})(\frac{1}{4} + \frac{1}{2^{2K-1}}, \frac{1}{4} + \frac{1}{2^{2K-2}})(\frac{1}{2} + \frac{1}{2^{2K-1}}, \frac{1}{2} + \frac{1}{2^{2K-2}})(\frac{3}{4} + \frac{1}{2^{2K-1}}, \frac{3}{4} + \frac{1}{2^{2K-2}}) \\
\ldots & \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\
I_{K,2K}: & (\frac{1}{2^{K+1}}, \frac{1}{2^K})(\frac{1}{2^K} + \frac{1}{2^{K+1}} \frac{1}{2^K} + \frac{1}{2^K}) \quad \ldots \quad (\frac{2^K-1}{2^K} + \frac{1}{2^{K+1}}, \frac{2^K-1}{2^K} + \frac{1}{2^K})
\end{array}$$

We rewrite them in the reverse order:

$(\frac{1}{2^{K+1}}, \frac{1}{2^K})$	$(\frac{1}{2^K}+\frac{1}{2^{K+1}}, \frac{1}{2^K}+\frac{1}{2^K})$	$\dots$	$(\frac{1}{4}+\frac{1}{2^{K+1}}, \frac{1}{4}+\frac{1}{2^K})$	$\dots$	$(\frac{1}{2}+\frac{1}{2^{K+1}}, \frac{1}{2}+\frac{1}{2^K})$	$\dots$	$(\frac{3}{4}+\frac{1}{2^{K+1}}, \frac{3}{4}+\frac{1}{2^K})$
$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
$(\frac{1}{2^{2K-1}}, \frac{1}{2^{2K-2}})$		$\dots$	$(\frac{1}{4}+\frac{1}{2^{2K-1}}, \frac{1}{4}+\frac{1}{2^{2K-2}})$	$\dots$	$(\frac{1}{2}+\frac{1}{2^{2K-1}}, \frac{1}{2}+\frac{1}{2^{2K-2}})$	$\dots$	$(\frac{3}{4}+\frac{1}{2^{2K-1}}, \frac{3}{4}+\frac{1}{2^{2K-2}})$
$(\frac{1}{2^{2K}}, \frac{1}{2^{2K-1}})$		$\dots$		$\dots$	$(\frac{1}{2}+\frac{1}{2^{2K}}, \frac{1}{2}+\frac{1}{2^{2K-1}})$	$\dots$	
$(\frac{1}{2^{2K+1}}, \frac{1}{2^{2K}})$		$\dots$		$\dots$		$\dots$	

Intervals of the first group belong to $[0, \frac{1}{2^K}]$, of the second group belong to $[\frac{1}{2^K}, \frac{2}{2^K}]$, and so further so that the intervals of 2^K-th group belong $[\frac{2^K-1}{2^K}, 1]$. Thus intervals of different groups do not intersect each other. It is also easy to note that within each group the intervals are mutually disjoint. Hence all the intervals are mutually disjoint. We show now that $R_K(x) \to 0$ given $K \to \infty$. We take a $z \in [0,1]$. Suppose $\varepsilon > 0$ and choose N, such that $\frac{1}{2^N} < \varepsilon$. Then z hits no more than one interval for all K on which the functions of the type $W(x)$ that form the remainder $R_K(x)$ are not equal to zero. Consider the functions the maximal values of which are greater than ε. Obviously, they all have to be included in the expression for functions $I_{0,2K}(x), I_{1,2K}(x), \dots, I_{N-1,2K}(x)$. We write out the intervals on which these functions are not equal to zero, as above:

$(\frac{1}{2^{2K-N+2}}, \frac{1}{2^{2K-N+1}})$	$\dots$	$(\frac{1}{4}+\frac{1}{2^{2K-N+2}}, \frac{1}{4}+\frac{1}{2^{2K-N+1}})$	$\dots$	$(\frac{1}{2}+\frac{1}{2^{2K-N+2}}, \frac{1}{2}+\frac{1}{2^{2K-N+1}})$	$\dots$	$(\frac{3}{4}+\frac{1}{2^{2K-N+2}}, \frac{3}{4}+\frac{1}{2^{2K-N+1}})$	$\dots$
$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
$(\frac{1}{2^{2K-1}}, \frac{1}{2^{2K-2}})$	$\dots$	$(\frac{1}{4}+\frac{1}{2^{2K-1}}, \frac{1}{4}+\frac{1}{2^{2K-2}})$	$\dots$	$(\frac{1}{2}+\frac{1}{2^{2K-1}}, \frac{1}{2}+\frac{1}{2^{2K-2}})$	$\dots$	$(\frac{3}{4}+\frac{1}{2^{2K-1}}, \frac{3}{4}+\frac{1}{2^{2K-2}})$	$\dots$
$(\frac{1}{2^{2K}}, \frac{1}{2^{2K-1}})$	$\dots$		$\dots$	$(\frac{1}{2}+\frac{1}{2^{2K}}, \frac{1}{2}+\frac{1}{2^{2K-1}})$	$\dots$		$\dots$
$(\frac{1}{2^{2K+1}}, \frac{1}{2^{2K}})$	$\dots$		$\dots$		$\dots$		$\dots$

As we see, there is only finite number of them, namely $1+2+\ldots+2^{N-1}$. With unlimited increase of K the structure remains, and intervals of each group shrink to their limit point: intervals of the first group shrink to 0, intervals of the second group shrink to $\frac{1}{2^{N-1}}$, and so further, so that intervals of the 2^{N-1}-th group shrink to $\frac{2^{N-1}-1}{2^{N-1}}$. Therefore, since the number of intervals is finite, if z does not coincide with no point of limit we always can find M, such that for all $K > M$ the intervals will be closer to their limit point than z, and hence z will not hit any of them. For z coinciding with the point of limit it is easily seen that z will belong to none of the intervals for sufficiently large K, since it is away from any of the limit points and intervals shrinking to it will never reach it. Anyway, starting from a certain $M \in \mathbf{N}$, values of functions $I_{0,2K}, I_{1,2K}, \dots, I_{N-1,2K}$ on z will be equal to zero. The sum of $I_{N,2K}(x), \dots, I_{K,2K}(x)$, which are the remaining components of $R_{2K}(x)$ is less than ε (for each K only one of two $W(z)$ pikes that are included in these functions, is not equal to 0, hence

$$\forall K > M \quad R_K(z) < \varepsilon.$$

Since ε is arbitrary, $R_K(x) \to 0$ where $K \to \infty$. Therefore, $R_K(x)$ define the convergent series on $[0,1]$.

1) Given the evident continuity of functions $R_K(x)$, the terms of the series $u_n(x) = R_{n-1}(x) - R_n(x)$ are also continuous. The series tends to function $f(x) = R_0(x)$, continuous in $[0,1]$.

2) We prove that this series is not generalized uniformly convergent on any subinterval of $[0,1]$. Choose the arbitrary subinterval (a,b). Clearly it contains a certain interval of the type $(\frac{p}{2^q}, \frac{p+1}{2^q})$. Let $\varepsilon = \frac{1}{2^q}$. Consider now the remainder $R_K(x)$, $\quad K > q$. One has

$$R_K(x) = I_{0,2K}(x) + I_{1,2K}(x) + \ldots + I_{q,2K}(x) + \ldots + I_{K,2K}(x) \geq$$

$$I_{q,2K}(x) = Y_{q,2K}(x) + Y_{q,2K}\left(x - \frac{1}{2^q}\right) + \ldots + Y_{q,2K}\left(x - \frac{p}{2^q}\right) + \ldots + Y_{q,2K}\left(x - \frac{2^q-1}{2^q}\right) \geq$$

$$Y_{q,2K}\left(x - \frac{p}{2^q}\right) = W\left(\frac{1}{2^{2K-q+1}}, \frac{1}{2^{2K-q}}, \frac{1}{2^q}\right)\left(x - \frac{p}{2^q}\right) =$$

$$W\left(\frac{p}{2^q} + \frac{1}{2^{2K-q+1}}, \frac{p}{2^q} + \frac{1}{2^{2K-q}}, \frac{1}{2^q}\right)(x).$$

$K > q$, $2K - q > q$, thus

$$\frac{p}{2^q} + \frac{1}{2^{2K-q+1}} < \frac{p}{2^q} + \frac{1}{2^{2K-q}} < \frac{p}{2^q} + \frac{1}{2^q} = \frac{p+1}{2^q}.$$

Therefore the whole interval on which

$$W\left(\frac{p}{2^q} + \frac{1}{2^{2K-q+1}}, \frac{p}{2^q} + \frac{1}{2^{2K-q}}, \frac{1}{2^q}\right)(x) \neq 0,$$

belongs to $[\frac{p}{2^q}, \frac{p+1}{2^q}]$, and consequently also to $[a,b]$. Therefore there is a certain point on $[a,b]$ in which $W(\frac{p}{2^q} + \frac{1}{2^{2K-q+1}}, \frac{p}{2^q} + \frac{1}{2^{2K-q}}, \frac{1}{2^q})(x)$ reaches its maximum value equal to $\frac{1}{2^q}$. In this point, then, $R_K(x)$ has its value not less than $\frac{1}{2^q}$ (actually, this value is exactly equal to $\frac{1}{2^q}$, since we have proven the intervals where the functions forming $R_K(x)$ are nonzero are mutually disjoint). Thus we obtain

$$\forall K > q \quad \exists z \quad a \leq z \leq b \quad R_K(x) \geq \frac{1}{2^q} = \varepsilon.$$

Therefore, $\{K : R_K(x) < \varepsilon\}$ is finite, and the series is not generalized uniformly convergent on $[a,b]$. ■

Example 2 Consider nonnegative series on $[0,1]$ $\sum\limits_{i=1}^{\infty} u_i(x)$, where

$$u_n(x) = W\left(\frac{1}{2^n} - \frac{1}{4^n}, \frac{1}{2^n} + \frac{1}{4^n}, \frac{1}{n}\right)(x) + W\left(\frac{3}{2^n} - \frac{1}{4^n}, \frac{3}{2^n} + \frac{1}{4^n}, \frac{1}{n}\right)(x) +$$

$$W\left(\frac{5}{2^n}-\frac{1}{4^n},\frac{5}{2^n}+\frac{1}{4^n},\frac{1}{n}\right)+\ldots+W\left(\frac{2^n-1}{2^n}-\frac{1}{4^n},\frac{2^n-1}{2^n}+\frac{1}{4^n},\frac{1}{n}\right)(x).$$

1) We prove that it converges uniformly on this interval. Suppose $\varepsilon > 0$. Choose $N > \frac{2}{\varepsilon}$, taking an arbitrary $z \in [0,1]$.

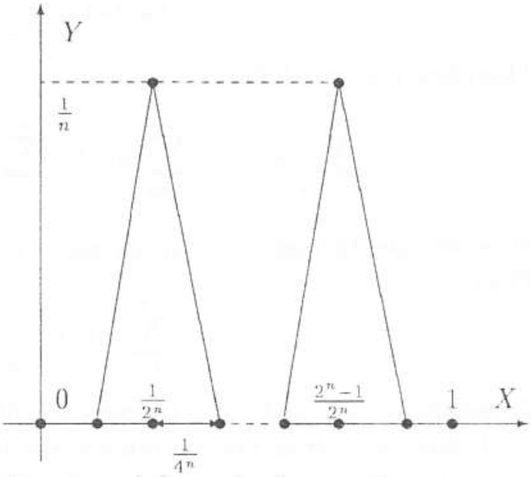


Fig. 4. $u_n(x)$ graph of example.

Lemma *If $u_n(x) \neq 0$, then all $u_{n+1}(x), u_{n+2}(x), \ldots, u_{2n-1}(x)$ are equal to zero.*

Proof of lemma. We prove the lemma on contradiction. Suppose that the statement is wrong for a certain m, $n < m < 2n$, i. e. $u_m(x) \neq 0$. Since $u_n(x) \neq 0$, then $x \in (\frac{2k+1}{2^n} - \frac{1}{4^n}, \frac{2k+1}{2^n} + \frac{1}{4^n})$, and since $u_m(x) \neq 0$, then $x \in (\frac{2l+1}{2^m} - \frac{1}{4^m}, \frac{2l+1}{2^m} + \frac{1}{4^m})$. The above intervals intersect each other, hence the distance between their centers is less than the sum of their radii, so we have

$$\left|\frac{2k+1}{2^n}-\frac{2l+1}{2^m}\right| < \frac{1}{4^n}+\frac{1}{4^m} < \frac{2}{4^n}.$$

Hence

$$\frac{|(2k+1)2^n-(2l+1)2^{2n-m}|}{2^{2n}} < \frac{2}{2^{2n}},$$

$$|(2k+1)2^n-(2l+1)2^{2n-m}| < 2,$$

$$|(2k+1)2^{m-n}-(2l+1)| < \frac{2}{2^{2n-m}} = \frac{1}{2^{2n-m-1}}.$$

Since $m > n$, then 2^{m-n} is an even integer. Hence $(2k+1)2^{m-n}-(2l+1)$ is also an integer, and, moreover, is odd. Then $|(2k+1)2^{m-n}-(2l+1)| \geq 1$. Meanwhile, for $m \leq 2n-1$ holds $\frac{1}{2^{2n-m-1}} \leq \frac{1}{2^0} = 1$. Thus we come to a contradiction: $1 < 1$, having proven the lemma. □

Consider the example further. We select out the nonzero terms of the remainder $\sum\limits_{i=N+1}^{\infty} u_i(z)$. Denote them by $u_{k_0}(z), u_{k_1}(z), \ldots$ mindless of finiteness of their number. According to the lemma, the distance between the two terms $u_{k_m}(z), u_{k_{m+1}}(z)$ is not less than K_m. Then we can write:

$$K_0 \geq N+1,$$

$$K_1 \geq K_0 + K_0 \geq 2(N+1),$$

$$K_2 \geq K_1 + K_1 \geq 4(N+1),$$

$$\ldots$$

$$K_m \geq 2^m(N+1), \dots$$

$u_{k_m}(z)$, obviously, is not greater than the maximum value of the respective $W(x)$, i. e. $\frac{1}{K_m}$. Then

$$u_{k_m}(z) \leq \frac{1}{K_m} \leq \frac{1}{2^m(N+1)}.$$

Therefore the remainder

$$\sum_{i=N+1}^{\infty} u_i(z) = \sum_{i=k_0}^{\infty} u_i(z) \leq \sum_{i=0}^{\infty} \frac{1}{2^i(N+1)} = \frac{2}{N+1} < \frac{2}{N} < \varepsilon,$$

since we have the geometric progression with denominator $\frac{1}{2}$. Suppose now $N < p \leq q$. Then

$$\sum_{i=p}^{q} u_i(z) \leq \sum_{i=N+1}^{\infty} u_i(z) < \varepsilon.$$

Therefore, since z and ε can be chosen arbitrarily, the series is uniformly convergent on $[0,1]$. Since the terms are continuous, the sum $f(x)$ is also continuous in $[0,1]$ according to the Theorem 6.

2) We prove now that this series tamely converges on no subsegment of $[0,1]$. We take a certain subsegment $[a,b]$, and chose such N, that 2^{N-1} is greater than the length of $[a,b]$. The distance between the two "pikes" of $u_n(x)$ is equal to

$$\frac{2k+1}{2^n} - \frac{2k-1}{2^n} = \frac{2}{2^n} = \frac{1}{2^{n-1}}.$$

Thus for $n > N$ it will be less than the length of $[a,b]$. Therefore, starting from N, at least one maximum point of $u_n(x)$ will hit the segment $[a,b]$. The value of $u_n(x)$ there is equal to $\frac{1}{n}$. Then even numerical series, consisting of suprema of $u_n(x)$ over $[a,b]$, starting from N, is harmonic series, that is nonconvergent. Therefore this series does not tamely converge on $[a,b]$.

The examples given show that in fact from all types of convergence of series of functions, only quasiuniform convergence is related to continuity of a. Namely, even in case of continuity of sum the quasiuniform convergence turns out to be very much "isolated" and "separated" from any other type of convergence, excluding types of convergence that coincide with it. Thus the quasiuniform convergence turns out to be the "well guessed" criterion for the continuity of a sum of a series of continuous functions.

Bibliography

1. N.N. Luzin. The Theory of Functions of Real Variable. Moscow, "Uchpedgiz", 1948. in Russian.

2. F.A. Medvedev. History Surveys of the Theory of Real Variable Functions. Moscow, "Nauka", 1975. in Russian.

Dmitry Ilchenko.
Graduated from the physical and mathematical school No. 239 in 1993. Student of the Dept. of Computer Technology since 1993. Winner of St.Petersburg city school olympiads in physics and mathematics in 1987–1993 and students mathematical olympiads in 1993–1996. Absolute 5th place in the Russian national students mathematical olympiad in 1995. "Soros student" in 1995.

The Weierstrass Theorem on the Uniform Approximation of the Continuous Functions by Polinomials

A. Likhodedov, M. Sinitsyn

Introduction

Very often, while solving different problems, it is necessary to study the characters of a certain function, that is hard or even impossible to do, starting from its original definition.

In these cases, one usually tries to find a new analytical expression for the investigated function, representing it as the limit of a sequence of more simple functions (the partial sums of a functional series, converging uniformly to the original function). Then, if the values of the partial sum are close enough to the values of the function on the examined region, it is possible to investigate some properties of the function, starting from the properties of this partial sum.

There are several types of functions, suitable for using as the members of a series, representing the given function, where in every particular case it is more convenient to use its own kind of functional series. Here, the representations of a function as the series of polynomials, will be examined.

Necessity of continuity

Which conditions must be imposed upon the function, in order to approximate it with polynomials with any preassigned accuracy? The answer is given by the Weierstrass Theorem: for the representability of a function as the sum of polynomial series (it is equivalent to the fact, that the function could be approximated with a polynomial with any accuracy), continuity of the function is sufficient. Thus, since the necessity of continuity is almost obvious (it will be proved later), continuity is the necessary and sufficient condition.

Since before it was only known that the functions, decomposable in power series, could be represented as a convergent polynomial series, this theorem significantly increases the class of these functions (for decomposability of a function in power series, even infinite differentiability is insufficient).

We prove the necessity of continuity for the approximability of the given function with polynomials.

Theorem 1 *Let $F_1(x) + F_2(x) + \ldots + F_n(x) + \ldots$ be a uniformly convergent series, and $S(x)$ — its sum, where all $F_i(x)$ are continuous functions on $[a, b]$. Then $S(x)$ is also continuous on $[a, b]$.*

PROOF. Take an arbitrary small $\varepsilon > 0$. Since the series converges uniformly, there exists such n, that $|S_n(x) - S(x)| < \frac{\varepsilon}{3}$ for all x from $[a, b]$ (S_n is the partial sum of the series).

$S_n(x)$ is continuous, as the sum of continuous functions. Then, for any z from $[a, b]$ there exists its vicinity U, such that for any x from U $|S_n(x) - S_n(z)| < \frac{\varepsilon}{3}$. Thus, for any x from U the following is true:

$$|S(x) - S(z)| < |S(x) - S_n(x)| + |S_n(x) - S_n(z)| + |S_n(z) - S(z)| <$$
$$< \frac{\varepsilon}{3} + \frac{\varepsilon}{3} + \frac{\varepsilon}{3} = \varepsilon.$$

With respect to the definition of continuity, that proves the statement of the theorem. ■

Examine the example, illustrating the necessity of uniform convergence of the series $F_1(x) + F_2(x) + \ldots + F_n(x) + \ldots$. Consider $U_n(x)$ — the continuous functions, equal to zero outside of the segment $[\frac{1}{n}; \frac{1}{n+1}]$, equal to 1 in the middle of this segment and linear in its left and right halves. The series $U_1(x) + U_2(x) + \ldots + U_n(x) + \ldots$ converges to the function $f(x)$ (see fig.1).

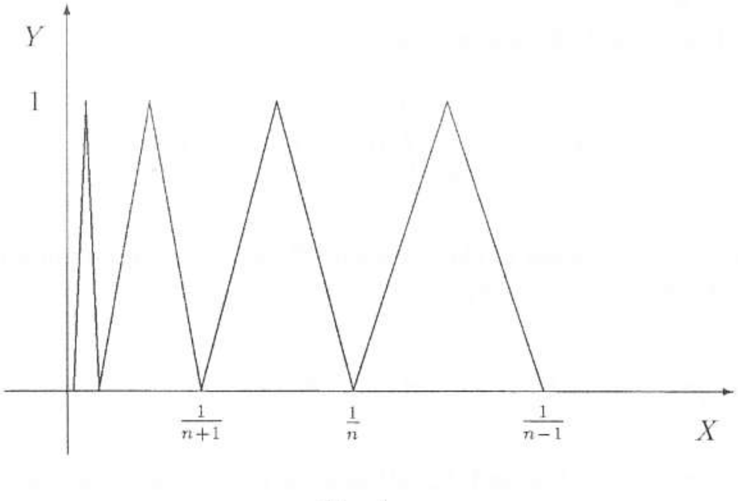


Fig. 1

Consider the behavior of the function $f(x)$, as x tends to 0. If $X_n = \frac{1}{n}$, then $f(X_n)$ is equal to 0, and if X_n is the midpoint of the segment $[\frac{1}{n}; \frac{1}{n-1}]$, then $f(X_n) = 1$; therefore the right-side limit in 0 does not exist and $f(x)$ is not a continuous function.

The Weierstrass theorem

Now, we prove the sufficiency of continuity, i.e. the Weierstrass Theorem.

For every continuous on $[a, b]$ function $f(x)$, and any $\varepsilon > 0$ there exists a polynomial $P(x)$, such that $|P(x) - f(x)| < \varepsilon$. (Sometimes this theorem is formulated differently: Every continuous function $f(x)$ could be represented as the sum of a uniformly convergent polynomial series).

There will be three proves of this theorem.

The Landau Proof

We prove the theorem for the segment $[0, 1]$ first (the polynomial, satisfying the conditions of the Weierstrass theorem on the $[0, 1]$ segment will be found).

We introduce the following notation:

$$I_n = \int\limits_{-1}^{1} (1 - u^2)^n \, du; \qquad K_n = \int\limits_{-n^{-\frac{1}{3}}}^{n^{\frac{1}{3}}} (1 - u^2)^n \, du;$$

$$L_n = \int\limits_{n^{-\frac{1}{3}} \leq |u| \leq 1} (1 - u^2)^n \, du.$$

Thus, $I_n = K_n + L_n$;

Now, we prove, that the polynomial

$$P_n(x) = \frac{1}{I_n} \cdot \int\limits_{0}^{1} f(v) \cdot (1 - (v - x^2))^n \, dv$$

complies with the hypothesis of the theorem ($P_n(x)$ — is the polynomial of x, since the integrand is the polynomial of x).

Lemma 1.1 *$\frac{L_n}{I_n}$ tends to 0, as n tends to infinity.*

Proof. Since the integrand L_n attains its maximum when $|u| = n$ (see fig. 2), in accordance with the theorem about monotonicity of the integral, we have $L_n < 2 \cdot (1 - n^{-\frac{2}{3}})^n$.

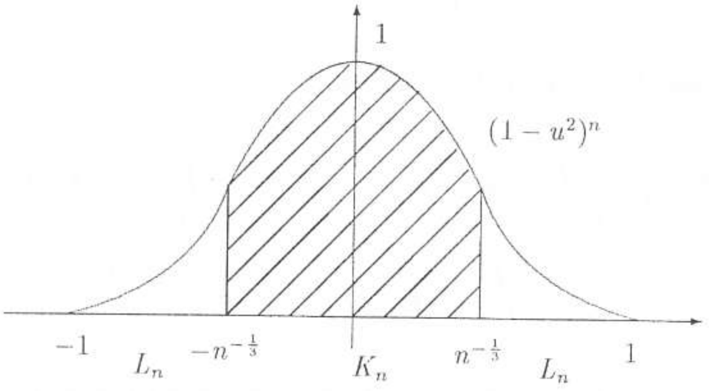


Fig. 2

In addition, $I_n > \int\limits_{-\frac{1}{2}n^{-\frac{1}{3}}}^{\frac{1}{2}n^{-\frac{1}{3}}} (1-u^2)^n\,du$, which, by the same theorem about monotonicity of the integral, is greater then $(1-\frac{1}{4}n^{-\frac{2}{3}})^n \cdot n^{-\frac{1}{3}}$. Thus,

$$\frac{L_n}{I_n} < 2n^{\frac{1}{3}} \cdot \left(\frac{1-n^{-\frac{2}{3}}}{1-\frac{1}{4}n^{-\frac{2}{3}}}\right)^n = 2n^{\frac{1}{3}} \cdot \left(1 - \frac{\frac{3}{4}n^{-\frac{2}{3}}}{1-\frac{1}{4}n^{-\frac{2}{3}}}\right)^n <$$

$$< 2n^{\frac{1}{3}} \cdot \left(\frac{4-\frac{3}{4}n^{-\frac{2}{3}}}{4}\right)^n = 2n^{\frac{1}{3}} \cdot \left(1-\frac{3}{4}n^{-\frac{2}{3}}\right)^n$$

and, since $1 + x < e^x$, as $x \neq 0$, we have

$$2n^{\frac{1}{3}} \cdot (1-\frac{3}{4}n^{-\frac{2}{3}})^n < 2n^{\frac{1}{3}} \cdot e^{-\frac{3}{4}n^{-\frac{2}{3}} \cdot n} = 2n^{\frac{1}{3}} \cdot e^{-\frac{3}{4}n^{\frac{1}{3}}}$$

Denote $\frac{3}{4}n^{\frac{1}{3}}$ by z. Then, since $\frac{z}{e^z}$ tends to 0, as z tends to infinity, we obtain that $\frac{L_n}{I_n} < \frac{8}{3}\frac{z}{e^z}$ also tends to 0. Thus, the limit of the ratio $\frac{L_n}{I_n}$ is equal to 0. That proves the lemma. □

PROOF OF THE THEOREM. Prove, that $P_n(x)$ is the desired polynomial. We make the change of variable: $v = u + x$; then $P_n(x) = \frac{1}{I_n} \cdot \int\limits_{-x}^{1-x} f(u+x) \cdot (1-u^2)^n\,du$. Let $0 < a < b < 1$ and $x \in [a,b]$, then at sufficiently great n $(n < min\{a, 1-b\})$ $-x < -n < n < 1-x$.

Then

$$P_n(x) = \frac{1}{I_n} \cdot (\int\limits_{-x}^{-n^{-\frac{1}{3}}} f(u+x) \cdot (1-u^2)^n\,du + \int\limits_{-n^{-\frac{1}{3}}}^{n^{-\frac{1}{3}}} f(u+x) \cdot (1-u^2)^n\,du +$$

$$+ \int\limits_{n^{-\frac{1}{3}}}^{1-x} f(u+x) \cdot (1-u^2)^n \, du)$$

and, if $M = max|f(u)|$ on $[0,1]$ we obtain:

$$|R_n(x)| = \frac{1}{I_n} \cdot \left| \int\limits_{-x}^{-n^{-\frac{1}{3}}} f(u+x) \cdot (1-u^2)^n \, du + \int\limits_{n^{-\frac{1}{3}}}^{1-x} f(u+x) \cdot (1-u^2)^n \, du \right| \leq$$

$$\leq \frac{M}{I_n} \cdot \int\limits_{n^{\frac{1}{3}} \leq |u| \leq 1} (1-u^2)^n \, du = \frac{M \cdot L_n}{I_n}.$$

By the previous lemma, $\frac{L_n}{I_n}$ tends to 0, as n tends to infinity. That implies, that $\frac{M \cdot L_n}{I_n}$ also tends to 0, and, since $\frac{M \cdot L_n}{I_n}$ does not depend on x, $R_n(x)$ converges uniformly to 0 on $[a,b]$, as n tends to infinity.

$P_n(x) = R_n(x) + \frac{W_n(x)}{I_n}$, where $W_n(x) = \int\limits_{-n^{-\frac{1}{3}}}^{n^{-\frac{1}{3}}} f(u+v) \cdot (1-u^2)^n \, du$.

Now, we show the proximity of $W_n(x)$ to $f(x)$ on $[a,b]$ (since the uniform smallness of R_n is proved, it will prove the statement of the theorem).

By the theorem on the mean value, $W_n(x) = f(x + q \cdot n^{-\frac{1}{3}}) \cdot K_n$, where $|q| < 1$. Then

$$|f(x) - P_n(x)| = |f(x) - f(x + q \cdot n^{-\frac{1}{3}}) \cdot \frac{K_n}{I_n} - R_n(x)| <$$

$$< |f(x) - f(x + q \cdot n^{-\frac{1}{3}}) \cdot \frac{K_n}{I_n}| + |Rn(x)|.$$

Now take an arbitrary small $\varepsilon > 0$. Since $R_n(x)$ converges uniformly to 0, there exists such N_ε, that for any $n > N_\varepsilon$, $|R_n(x)| < \frac{\varepsilon}{2}$. $f(x)$ is continuous on $[0,1]$, and, by the Cantor's theorem, it is uniformly continuous on $[0,1]$. Besides, the ratio $\frac{K_n}{I_n}$ tends to 1, as n tends to infinity. Then, there exists M_ε, such that for any $n > M_\varepsilon$, the following is true

$$|f(x) - f(x + q \cdot n^{-\frac{1}{3}}) \cdot \frac{K_n}{I_n}| < \frac{\varepsilon}{2}.$$

Consider $N = max\{N_\varepsilon, M_\varepsilon\}$. Then, when $n > N$

$$|f(x) - f(x + q \cdot n^{-\frac{1}{3}}) \cdot \frac{K_n}{I_n}| + |R_n| < \varepsilon.$$

Thus, we have proved, that $P_n(x)$ converges uniformly to $f(x)$ on $[a,b]$, where $0 < a < b < 1$.

Now prove this theorem for an arbitrary segment $[g,h]$. Introduce the new variable $y = \frac{x-g}{h-g}$; $f(x) = f(g + (h-g) \cdot y) = F(y)$. Evidently, $g < x < h$ implies, that $0 < y < 1$; and $F(y)$ is continuous, since $f(x)$ is continuous. Let

$$g < j < o < h, \quad a = \frac{j-g}{h-g}, \quad b = \frac{o-g}{h-g}.$$

It is obvious, that $0 < a < b < 1$. Then, by the above theorem, for any $\varepsilon > 0$, there exists a polynomial $P_n(y)$, such that $|F(y) - P_n(y)| < \varepsilon$, i.e. $|f(x) - P_n(\frac{x-g}{h-g})| < \varepsilon$. Thus, $A(x) = P_n(\frac{x-g}{h-g})$ is the polynomial,approximating $f(x)$ on $[j, o]$. To extend the proved on the whole segment $[g, h]$, we should define $f(x)$ on the extended segment, containing $[g, h]$, and use the already proved statement. That completes the proof of the theorem.

This proof seems to be rather artificial, but it is not correct — the Landau's method is based on the idea of consideration of a function $G_n(x, v)$, with the following characters:

1) $G_n(x, v)$ is integrable by v and nonnegative.

2) For all $d > 0$

$$\lim_{n\to\infty} \int\limits_{|x-v|<d} G_n(x, v)\, dv = 1$$

and

$$\lim_{n\to\infty} \int\limits_{|x-v|>d} G_n(x, v)\, dv = 0.$$

Then,

$$\lim_{n\to\infty} P_n(x) = \lim_{n\to\infty} \int\limits_0^1 G_n(x, v) \cdot f(v)\, dv = f(x).$$

The last statement can be proved by analogy with the previous theorem (while proving we did not use any properties of the function $\frac{(1-(v-x)^2)^n}{I_n}$, except those, that any function G_n possesses). Thus, it is only left to find a function, possessing all the characters of the function G_n. Landau offered $\frac{(1-(v-x)^2)^n}{I_n}$ in the capacity of this function. The analogous idea is in the base of the Bernstein's proof, but he offered another function, as $G_n(x, v)$.

The Lebesgue's Proof

Lebesgue has proved the Weierstrass theorem in somewhat different statement, using the tame convergence concept: Any continuous on the segment $[a, b]$ function $f(x)$ is the sum of a tame convergent (on $[a, b]$) polynomial series $P_1(x) + P_2(x) + \ldots + P_n(x) + \ldots$

Consider the idea of tame convergence and its connection with uniform convergence: The functional series $U_1(x) + U_2(x) + \ldots + U_n(x) + \ldots$ is called *tame convergent* on the segment $[a, b]$, if there exists a numerical positive convergent series $E_1 + E_2 + \ldots + E_n + \ldots$, so that for all n greater than a certain N $|U_n(x)| < E_n$, for all x on the segment $[a, b]$.

Theorem 2 *Any tame convergent functional series on the segment $[a, b]$ is also uniformly convergent on this segment.*

PROOF. Let the series $U_1(x) + U_2(x) + \ldots + U_n(x) + \ldots$ be tame convergent. Then, for the sufficiently great M $|U_{M+1}(x)| < E_{M+1}$; $|U_{M+2}(x)| < E_{M+2}$; ... for all $x \in [a, b]$. According to the definition of tame convergence, the series $E_1 + E_2 + \ldots + E_n + \ldots$ is convergent, then, according to the Cauchy's criterion $|E_p + E_{p+1} + \ldots + E_q| < \varepsilon$ for all

p and q, greater then some number K. Now, haven taken N, equal to $max\{M, K\}$, we obtain, that for all p and q greater then N the following inequalities are true:

$$|U_p(x) + \ldots + U_q(x)| < |U_p(x)| + |U_{p+1}(x)| + \ldots + |U_q(x)| < E_p + ... + E_q < \varepsilon$$

Therefore, the given series is uniformly convergent, according to the Cauchy's criterion. ■

Corollary 2.1 *The sum of a tame convergent series of continuous functions is a continuous function.*

Proof. Since by the Weierstrass theorem, the above statement is true for uniform convergence, it is true for tame convergence, according to the previous theorem. □

Thus, the Weierstrass theorem in its usual formulation is the corollary of the theorem in Lebesgue's formulation. Now, we prove their equivalence - it will be shown, that by arranging the members of the uniformly convergent series into groups, tame convergence can be achieved. let the series $f_1(x) + \ldots + f_n(x) + \ldots$ be uniformly convergent. According to the Cauchy criterion for any $\varepsilon > 0$ there exists N_ε, so that for all $m, n > N_\varepsilon, (m < n)$ the following is true $|f_m(x) + \ldots + f_n(x)| < \varepsilon$. Consider such a sequence ε_i, that the series $\varepsilon_1 + \ldots + \varepsilon_n + \ldots$ converges. Also, we select such N_{ε_i}, that they form the increasing sequence. Then

$$f_1(x) + \ldots + f_n(x) + \ldots =$$
$$= [f_1(x) + ... + f_{N_{\varepsilon_1}}(x)] + \ldots [f_{N_{\varepsilon_1}+1}(x) + \ldots + f_{N_{\varepsilon_2}}(x)] + \ldots (*).$$

Since in this transformation we do not change the positions, but only arrange the members of the series into groups, its sum does not change. According to the selection of ε_i,

$$[f_1(x) + \ldots + f_{N_{\varepsilon_1}}(x)] \leq |[f_1(x) + \ldots + f_{N_{\varepsilon_1}}(x)|,$$
$$[f_{N_{\varepsilon_i}+1}(x) + \ldots + f_{N_{\varepsilon_{i+1}}}(x)] \leq \varepsilon_i.$$

Every member of the series $(*)$ does not exceed the appropriate member of the series

$$|[f_1(x) + \ldots + f_{N\varepsilon_1}(x)]| + \varepsilon_1 + \ldots + \varepsilon_n + \ldots,$$

Thus, the series $(*)$ is tame convergent, and its sum is equal to the sum of the original series. In other words, the existence of the series of polynomials, converging uniformly to the given function, implies the existence of the tame convergent polynomial series with the same sum. I.e., the usual formulation of the Weierstrass Theorem and the formulation of Lebesgue are equivalent.

Despite the fact, that in the formulation of the Weierstrass theorem it does not matter how exactly - uniformly or tame the polynomial series converges, these concepts are different. Let on the $[a, b]$ $f_n(x) = \frac{(-1)^n}{n}$. It is easy to understand, that the series $f_1(x) + \ldots + f_n(x) + \ldots$ is uniformly convergent on the $[a, b]$: for all x its sum is equal to the sum of the convergent series

$$1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \frac{1}{5} - \ldots$$

Also, it is obvious, that this series is not tame convergent: it does not converge absolutely ($1 + \frac{1}{2} + \frac{1}{3} + \ldots + \frac{1}{n} + \ldots$ tends to infinity, while n tends to infinity), and that implies, that every series $\varepsilon_1 + \ldots + \varepsilon_n + \ldots$, where $|f_n(x)| = \frac{1}{n} < \varepsilon_n$, diverges.

Now, consider the concept of polygonal functions, used during the proof of the theorem.

Let function $L(x)$ be continuous on a segment $[a, b]$. $L(x)$ is called *polygonal*, if the segment $[a, b]$ could be divided on finite number of segments $[a, X_1], [X_1, X_2], \ldots, [X_n, b]$, where $L(x)$ is linear.

The polygonal function $l(x)$, equal to 0 for all $x \leq x_0$, where x_0 is a certain point of the segment $[a, b]$, and is linear for all $x \geq x_0$ is called *the elementary polygonal function.*

Theorem 3 *Any polygonal function $L(x)$, defined on a segment $[a, s]$ could be represented as the sum of a linear function and limited number of elementary polygonal functions - $L(x) = kx + b + l_1(x) + \ldots + l_n(x)$.*

Proof.

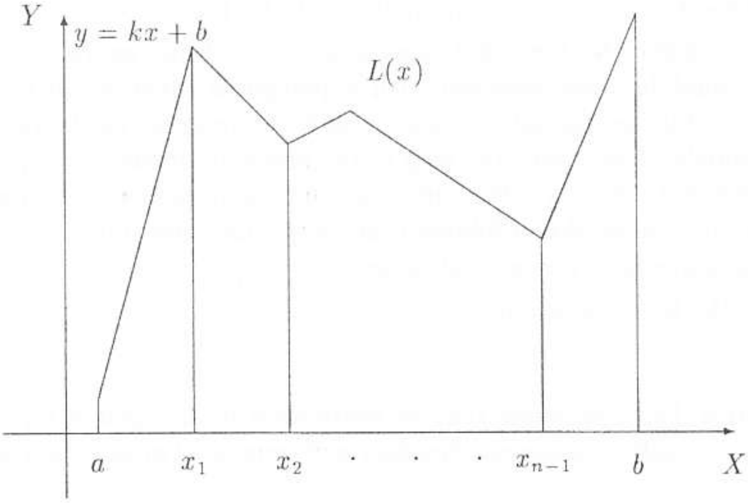


Fig. 3

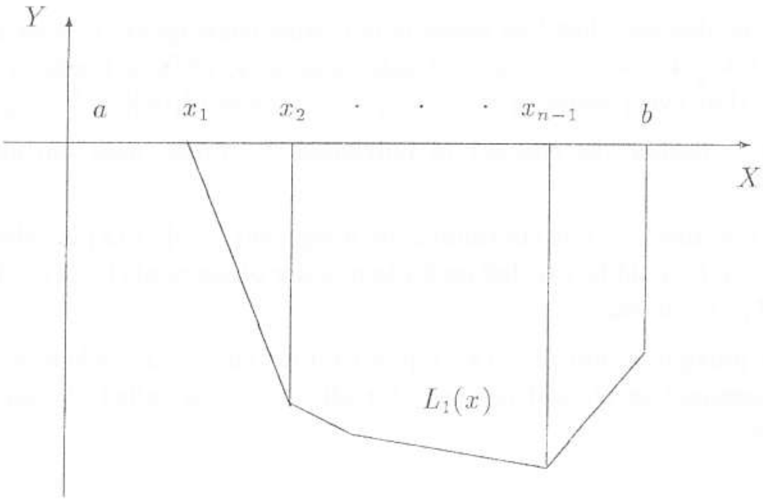


Fig. 4

Take the first link of the polygonal function $L(x)$: it is the line, defined with the equation $y = kx + b$. Now, consider the function $L_1(x) = L(x) - kx - b$: it is a polygonal function, equal to 0 on the first segment (see fig.4). Then, subtracting the elementary polygonal function $l_1(x)$, equal to 0 on the first segment and coinciding with the second link of $L_1(x)$ on the second segment from $L_1(x)$, we obtain $L_2(x)$ — the polygonal function, equal to 0 on the first two segments. And so on. Continuing this process, we obtain the function $L_n(x) = L(x) - kx - b - l_1(x) - l_2(x) - \ldots - l_{n-1}(x)$, which is the elementary polygonal function itself. Designating it as $l_n(x)$ we receive, that $L(x) = kx + b + l_1(x) + \ldots + l_n(x)$, as was to be proved.

Now consider the idea of Lebesgue's proof. First prove, that every continuous function could be approximated with a polygonal function with any exactness, and every elementary polygonal function could be represented as the tame convergent series of polynomials. Then, with the help of the previous theorem it is possible to show that every polygonal function is the sum of a tame convergent polynomial series. From this, using the first assumption it follows that every continuous function could be represented as a tame convergent polynomial series.

Prove the first statement:

Theorem 4 *Let a function $f(x)$ be continuous on a segment $[a, b]$. Then it could be approximated with a polygonal function with any level of accuracy on $[a, b]$.*

PROOF.

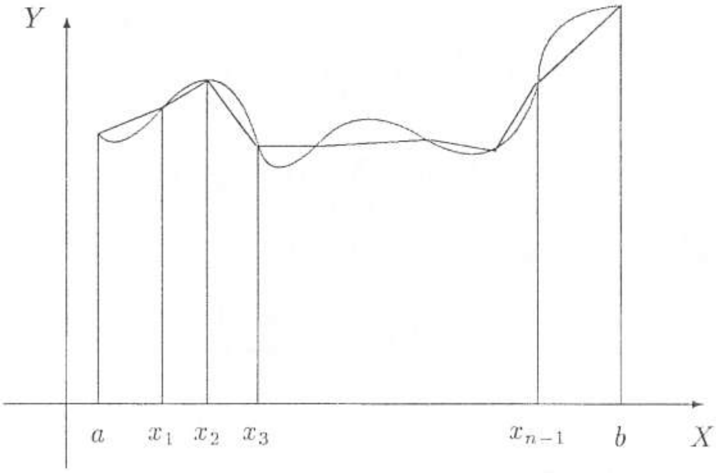


Fig. 5

This theorem follows obviously from the Cantor's theorem. Since $f(x)$ is continuous on the segment, it is uniformly continuous on it. Then for an arbitrary $\varepsilon > 0$ there exists such $d > 0$, that for all x_1, x_2, $|x_1 - x_2| < d$ it is true that $|f(x_1) - f(x_2)| < \varepsilon$. Now, perform the partitioning X_i of the $[a, b]$, so that the distance between every two neighboring points does not exceed d. Consider the polygonal function $L(x)$, having fractures in the points of intersections of perpendiculars to abscissa, drawn through the points of partitioning and the graph of the function (see fig.5). Evidently, $min\{f(X_i), f(X_{i+1})\} < L(x) < max\{f(X_i), f(X_{i+1})\}$. Then, for all x from the arbitrary segment of partitioning $[X_i, X_{i+1}]$ $L(x)$ does not exceed the maximum value of f on this segment and is not smaller than the minimum value. Thus, for all x from $[X_i, X_{i+1}]$ it is true, that $|f(x) - L(x)| < \varepsilon$. That proves the statement of the theorem. ■

Now prove, that any elementary polygonal function could be represented as a tame convergent series of polynomials.

Lemma 4.1 *For any $\varepsilon > 0$ and an arbitrary segment $[a, b]$ such a polynomial $P(x)$ could be found, that for all x from $[a, b]$ $||x| - P(x)| < \varepsilon$.*

Proof. Examine the decomposition of $(1 + Z)^{\frac{1}{2}}$ using the binomial formula:

$$(1+z)^{\frac{1}{2}} = 1 + C^1_{\frac{1}{2}} \cdot Z + \ldots + C^k_{\frac{1}{2}} \cdot Z^k + \ldots$$

or

$$(1+Z)^{\frac{1}{2}} = 1 + \frac{1}{2}Z - \frac{1}{8}Z^2 + \ldots + (-1)^{k-1} \cdot \frac{1}{2k} \cdot \frac{(2k-3)!!}{(2k-2)!!} \cdot Z^k + \ldots$$

or

$$(1+z)^{\frac{1}{2}} = 1 + \ldots + (-1)^{k-1} \cdot \frac{1}{2k} \cdot Q_k \cdot Z^k + \ldots \quad (**),$$

where

$$Q_k = \frac{(2k-3)!!}{(2k-2)!!} = \frac{1}{2} \cdot \frac{3}{4} \cdot \ldots \cdot \frac{2k-3}{2k-2} =$$

$$= (1-\frac{1}{2})\cdot(1-\frac{1}{4})\cdot\ldots\cdot(1-\frac{1}{2k-2}).$$

$$\ln Q_k = \sum_{m=1}^{k-1} \ln(1-\frac{1}{2m}),$$

but $ln(1-\frac{1}{2m}) = -\frac{1}{2}m - \frac{1}{2m} - \ldots < -\frac{1}{2}m$, so

$$\ln Q_k < -\frac{1}{2}\cdot\sum_{m}^{k-1}\frac{1}{m}.$$

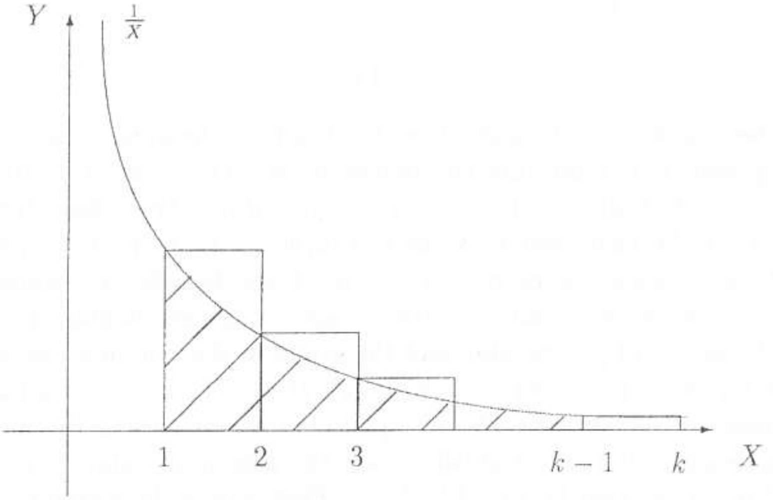


Fig. 6

Since $(\ln X)' = \frac{1}{X}$, $1+\frac{1}{2}+\ldots+\frac{1}{K-1} > \ln K$ (see fig.6); thus $\ln Q_k < -\frac{1}{2}\ln k$, from where $Q_k < \frac{1}{\ln\sqrt{k}} < \frac{1}{\sqrt{k}}$

Therefore, when $|Z| < 1$ every member of the series $(**)$ does not exceed the corresponding member of the series $1+\frac{1}{2\sqrt{2}}+\frac{1}{3\sqrt{3}}+\ldots+\frac{1}{k\sqrt{k}}+\ldots$by the absolute value. This series converges, and that implies that the series $(**)$ is tame convergent. When $|Z| < 1$, the series $(**)$ has a nonnegative sum, i.e. $(1+Z)^{\frac{1}{2}} = \sqrt{1+Z}$. If $Z = x^2-1$, while $|x| < 1$, we obtain:

$$\sqrt{x^2} = |x| = 1+\frac{1}{2}(x^2-1)-\ldots+(-1)^{k-1}\cdot\frac{1}{2k}\cdot Q_k\cdot(x^2-1)^k+\ldots.$$

In order to extend the proved on the arbitrary segment $[a,b]$, we consider the segment $[-N,N]$, containing [a,b]. Let t be equal to $N\cdot x$; then $|x| < 1$ implies $|t| < N$ and

$$|t| = N\cdot|x| = N\cdot\left(1+\frac{1}{2}(x^2-1)-\ldots+(-1)^{k-1}\cdot\frac{1}{2k}Q_k(x^2-1)^k+\ldots\right) =$$

$$= N \cdot \left(1 + \frac{1}{2}\left(\frac{t^2}{N^2} - 1\right) - \ldots + (-1)^k \cdot \frac{1}{2k} \cdot Q_k \cdot \left(\frac{t^2}{N^2} - 1\right)^k + \ldots\right) \quad (***)$$

Since $|\frac{t}{N}| < 1$, we obtain the tame convergent series - every member of this series does not exceed the corresponding member of the series $N \cdot (1 + \frac{1}{2\sqrt{2}} + \frac{1}{3\sqrt{3}} + \ldots + \frac{1}{k\sqrt{k}} + \ldots$ by the absolute value.

The series $N \cdot (1 + \frac{1}{2\sqrt{2}} + \frac{1}{3\sqrt{3}} + \ldots + \frac{1}{k\sqrt{k}} + \ldots) = N + \frac{N}{2\sqrt{2}} + \ldots + \frac{N}{k\sqrt{k}} + \ldots$ converges, so there exists its partial sum S_w, such that the remainder of the series does not exceed $\frac{1}{N^2}$: Take such a partial sum of the series $1 + \frac{1}{2\sqrt{2}} + \frac{1}{3\sqrt{3}} + \ldots + \frac{1}{k\sqrt{k}} + \ldots$, that its remainder does not exceed $\frac{1}{N^3}$ by modulus. Then, since

$$N + \frac{N}{2\sqrt{2}} + \ldots + \frac{N}{k\sqrt{k}} + \ldots = N \cdot \left(1 + \frac{1}{2\sqrt{2}} + \ldots + \frac{1}{k\sqrt{k}} + \ldots\right),$$

the corresponding partial sum S_w of the series $N \cdot (1 + \frac{1}{2\sqrt{2}} + \frac{1}{3\sqrt{3}} + \ldots + \frac{1}{k\sqrt{k}})$ is not greater than $\frac{1}{N^2}$ by modulus. Consider w-th partial sum of the series $(***)$. Since the series $(***)$ is tame convergent (its every member does not exceed the corresponding member of the series $N \cdot \left(1 + \frac{1}{2\sqrt{2}} + \frac{1}{3\sqrt{3}} + \ldots + \frac{1}{k\sqrt{k}} + \ldots\right)$), then its remainder also does not exceed $\frac{1}{N^2}$ by modulus. Thus, for the partial sum of this series $\tilde{S_w}(x)$ it is true, that

$$||x| - \tilde{S_w}(x)| < \frac{1}{N^2}.$$

We choose such N, that $\frac{1}{N^2} < \varepsilon$. Then, $||x| - \tilde{S_w}(x)| < \varepsilon$ on the segment $[a, b]$, and that means, that $\tilde{S_w}(x)$ is the desired polynomial. □

Corollary 4.1 *On every segment $[a, b]$ there exists a series of polynomials, converging tamely to $|x|$.*

PROOF. By the previous lemma, for every n there exists such a polynomial $\tilde{S_n}(x)$, that $||x| - \tilde{S_n}(x)| < \frac{1}{n^2}$). Let

$$P_1(x) = \tilde{S_1}(x), P_2(x) = \tilde{S_2}(x) - \tilde{S_1}(x), \ldots, P_n(x) = \tilde{S_n}(x) - \tilde{S}_{n-1}(x), \ldots$$

On the segment $[-n, n]$

$$|P_n(x)| = ||x| + \tilde{S_n}(x) - \tilde{S}_{n-1}(x) - |x|| < ||x| - \tilde{S}_{n-1}(x)| + |\tilde{S_n}(x) - |x|| <$$

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

Obviously, there exists n, such that the segment $[-n, n]$ contains $[a, b]$. $|P_n(x)| < \frac{2}{(n-1)^2}$, so the series $P_1(x) + \ldots + P_n(x) + \ldots$ is tame convergent on $[a, b]$. Also, $\lim_{n\to\infty} [P_1(x) + \ldots + P_n(x)] = \lim_{n\to\infty} \tilde{S_n}(x) = |x|$. □

Thus, we have proved, that the function $|x|$ could be represented as a tame convergent series of polynomials. Obviously, this statement could be spread on every function $|x - q|$:

$$|x - q| = \hat{P}_1(x) + \ldots + \hat{P}_n(x) + \ldots,$$

where $\hat{P}_k(x) = P_k(x-q)$.

Since every polygonal function $l(x)$ could be represented in the form of $m(|x-q| + x-q)$, it also could be decomposed in the tame convergent series of polynomials. Hence, according to the Theorem 3, every polygonal function could be represented as a tame convergent series of polynomials. ■

By the Theorem 4, for any $\varepsilon > 0$ there exists such a polygonal function $L(x)$, that

$$|f(x) - L(x)| < \frac{\varepsilon}{2}$$

for all x from $[a,b]$. Conversely, according to the above conclusions for $L(x)$ there exists such a polynomial $P(x)$, that

$$|L(x) - P(x)| < \frac{\varepsilon}{2}$$

Then $|f(x) - P(x)| < \varepsilon$ on the segment $[a,b]$. Hence, for any natural number n, there exists such a polynomial $S_n(x)$, that $|f(x) - S_n(x)| < \frac{1}{2n^2}$ on $[a,b]$. Take $P_n(x) = S_n(x) - S_{n-1}(x)$.

$$|P_n(x)| = |S_n(x) - S_{n-1}(x)| \leq |f(x) - S_{n-1}(x)| + |f(x) - S_n(x)| \leq \frac{1}{(n-1)^2}.$$

since $1 + \frac{1}{4} + \ldots + \frac{1}{n^2} + \ldots$ converges,

$$P_1(x) + \ldots + P_n(x) + \ldots, where(P_n(x) = S_n(x) - S_{n-1}(x))$$

is tame convergent. Also,

$$P_1(x) + \ldots + P_n(x) + \ldots = f(x),$$

i.e. $f(x)$ could be represented in the form of the tame convergent series of polynomials on the segment $[a,b]$. That completes the proof of the main statement of the Weierstrass theorem.

The Landau's proof does not give us an indication, how to find the polynomial, approximating the given function $f(x)$ on the segment $[a,b]$ with the given accuracy (to construct that polynomial it is necessary, to find n, such that the modulus of the remainder

$$\frac{K_n}{I_n} \cdot \int\limits_{-n^{-\frac{1}{3}}}^{n^{-\frac{1}{3}}} f(u+x) \cdot (1-u^2)^n \, du - f(x)$$

is not greater than $\frac{\varepsilon}{2}$ for all x from $[a,b]$, and this could be very difficult, or even impossible). The Lebesgue's proof, as will be shown further, gives us a universal (although extraordinary awkward) method of constructing of such a polynomial.

Now, we try to find the degree of the power of the Lebesgue's polynomial, approximating function $f(x)$ on the segment $[a,b]$ with an accuracy up to ε.

To construct the approximating polynomial, first a polygonal function $L(x)$, approaching $f(x)$ with accuracy up to $\frac{\varepsilon}{2}$, should be found. Then it is necessary to decompose this polygonal function in the sum of linear and elementary polygonal functions, and each of them must be approximated with a polynomial with an accuracy up to $\frac{\varepsilon}{2}$.

The power of the Lebesgue polynomial k is determined by the following inequality:

$$R = \frac{1}{k\sqrt{k}} + \frac{1}{(k+1)\sqrt{k+1}} + \ldots < \frac{1}{N^3},$$

where $\frac{1}{N^2} < \frac{\varepsilon}{2}$. That is $\frac{1}{N^3} < \sqrt{\frac{\varepsilon^3}{8}}$, that implies $R < \sqrt{\frac{\varepsilon^3}{8}}$. Now, we bound R from below by the integral $\int\limits_k^\infty x^{-\frac{3}{2}}\,dx$, that is equal to $\frac{1}{2}k^{-\frac{1}{2}}$. Therefore, to make R less, than $\sqrt{\frac{\varepsilon^3}{8}}$, it is necessary, that $\frac{1}{2}k^{-\frac{1}{2}} < \sqrt{\frac{\varepsilon^3}{8}}$, that is $k > \frac{2}{\varepsilon^3}$. Thus, although in theory it is possible to construct the Lebesgue's polynomial, approximating $f(x)$ with the given accuracy ε, in practice it could be realized only when ε is sufficiently great (the power of the Lebesgue polynomial, approximating $f(x)$ with accuracy 0.01, is more than two millions).

It should be noted, that the power of the polynomial, necessary for the approximation of the function with the given accuracy ε does not depend on the form of the function. In fact, the form of the given function determines only the approximating polygonal function $L(x)$, but the power of polynomials, approximating every elementary polygonal function (forming the $L(x)$ decomposition) depends, as we have seen, only on ε.

The Bernstein's Proof

Now consider the Bernstein's proof the Weierstrass Theorem. As it was mentioned before, it is based on the same idea, as the Landau's proof: First, the function G_n, with special properties (see above) is selected and then, using this function, we built a polynomial, that satisfies the conditions of the theorem. In Bernstein's method we take $\hat{G}_n(\frac{m}{n}, x) = C_n^m x^m \cdot (1-x)^{n-m}$, instead of $G_n(v,x) = \frac{((1-(v-x^2))^n}{I_n}$, as before. In the capacity of an approximating polynomial we consider $B_n = \sum\limits_{m=0}^{n} \hat{G}_n(\frac{m}{n}, x) \cdot f(\frac{m}{n})$, instead of $\int\limits_{a^b} f(v) \cdot G_n(x,v)\,dv$.

Thus, the character of a function G, that $\int\limits_{|x-v|<d} G_n(v,x)\,dv$ tends to 0 for all $d > 0$, is replaced with the following character: for all $d > 0$ $\sum C_n^m x^m \cdot (1-x)^{n-m} \to 0$, as $n \to \infty$, provided that $|\frac{m}{n} - x| > d$. Now, we prove that functions $\hat{G}_n$ possess this character (first, we prove the correctness of the statement for the segment $[0,1]$).

Lemma 4.2 *Given $0 < x < 1$ and m, n — integer numbers. Then, for any $\varepsilon > 0$ and $s > 0$, there exists such N, depending on ε and s, that for all $n > N$ the following inequality is true:*

$$\sum_{|\frac{m}{n}-x|>s} C_n^m \cdot x^m \cdot (1-x)^{n-m} < \varepsilon.$$

Proof. Multiply $\sum\limits_{|\frac{m}{n}-x|>s} C_n^m \cdot x^m \cdot (1-x)^{n-m}$ by the number $\frac{(\frac{m}{n}-x)^2}{s^2}$, greater than

1:

$$\sum_{|\frac{m}{n}-x|>s} C_n^m \cdot x^m \cdot (1-x)^{n-m} < \frac{1}{s^2} \cdot \sum_{|\frac{m}{n}-x|>s} \left(\frac{m}{n}-x\right)^2 \cdot C_n^m \cdot x^m \cdot (1-x)^{n-m} \leq$$

$$\leq \frac{1}{(s\cdot n)^2} \cdot \sum_{m=0}^{n} (m-n\cdot x)^2 \cdot C_n^m \cdot x^m \cdot (1-x)^{n-m}$$

Now, differentiate the binomial formula

$$\sum_{m=0}^{n} C_n^m \cdot p^m \cdot q^{n-m} = (p+q)^n \tag{1}$$

by p and multiply the resulting equality by p:

$$\sum_{m=0}^{n} m \cdot C_n^m \cdot p^m \cdot q^{n-m} = n \cdot p \cdot (p+q)^{n-1} \tag{2}$$

Repeating differentiation and multiplying by p once more we obtain:

$$\sum_{m=0}^{n} m^2 \cdot C_n^m \cdot p^m \cdot q^{n-m} = n \cdot p \cdot (n \cdot p + q) \cdot (p+q)^{n-2} \tag{3}$$

Substituting $p = x$ and $q = 1 - x$ in the identities (2), (3), (4), we multiply them by $(n \cdot x)^2$, $-2 \cdot n \cdot x$ and 1 respectively, and add. We obtain:

$$\sum_{m=0}^{n} (m - n\cdot x)^2 \cdot C_n^m \cdot x^m \cdot (1-x)^{n-m} = n \cdot x \cdot (1-x) \leq \frac{1}{4} n.$$

Noticing, that the sum in the left side of this inequality is the same as the sum in the right side of the inequality (1), we obtain:

$$\sum_{|\frac{m}{n}-x|>s} C_n^m \cdot x^m \cdot (1-x)^{n-m} < \frac{1}{4n^2 \cdot s^2}.$$

Now, select n, such that

$$\frac{1}{4n^2 \cdot s^2} < \varepsilon$$

and the inequality

$$\sum_{|\frac{m}{n}-x|>s} C_n^m \cdot x^n \cdot (1-x)^{n-m} < \varepsilon$$

becomes true. That proves the statement of the lemma. □

Now we prove the Weierstrass theorem using the Bernstein's method.

Proof of the theorem. Let $f(x)$ be a continuous function on the segment $[0, 1]$. Let M be the maximum of its modulus, $o > 0$, $(\varepsilon = \frac{o}{4M})$. Take such s, that if

$|x_1 - x_2| < s$, the following is true: $|f(x_1) - f(x_2)| < \frac{o}{2}$. Prove, that $|B_n(x) - f(x)| < o$, at sufficiently great n, where

$$B_n(x) = \sum_{m=0}^{n} f(\frac{m}{n}) \cdot C_n^m \cdot x^m \cdot (1-x)^{n-m}.$$

Examine the sum $\sum\limits_{m=0}^{n} C_n^m \cdot x^m \cdot (1-x)^{n-m}$. According to the binomial formula this sum is equal to 1. We divide it in two sums $\sum\limits_1$ and $\sum\limits_2$, where

$$\sum_1 = \sum_{|\frac{m}{n}-x| \leq s} f(\frac{m}{n}) \cdot C_n^m \cdot x^m \cdot (1-x)^{n-m},$$

$$\sum_2 = \sum_{|\frac{m}{n}-x| > s} f(\frac{m}{n}) \cdot C_n^m \cdot x^m \cdot (1-x)^{n-m}.$$

Then, by the preceding lemma, we obtain, that

$$\sum_1 C_n^m \cdot x^n \cdot (1-x)^{n-m} < 1,$$

$$\sum_2 C_n^m \cdot x^n \cdot (1-x)^{n-m} < \varepsilon$$

as n is sufficiently great.

Now consider the function $f(x)$. We will show, that the sequence

$$B_n(x) = \sum_{m=0}^{n} f(\frac{m}{n}) \cdot C_n^m \cdot x^m \cdot (1-x)^{n-m}$$

converges uniformly to $f(x)$ on the segment $[0,1]$:

$$|B_n(x) - f(x)| = |\sum_{m=0}^{n} f(\frac{m}{n}) \cdot C_n^m \cdot x^m \cdot (1-x)^{n-m} -$$

$$-f(x) \cdot \sum_{m=0}^{n} C_n^m \cdot x^m \cdot (1-x)^{n-m}| = |\sum_{m=0}^{n} (f(\frac{m}{n}) - f(x)) \cdot C_n^m \cdot x^m \cdot (1-x)^{n-m}| <$$

$$< \sum_1 |f(\frac{m}{n}) - f(x)| \cdot C_n^m \cdot x^m \cdot (1-x)^{n-m} +$$

$$+ \sum_2 (|f(\frac{m}{n})| + |f(x)|) \cdot C_n^m \cdot x^m \cdot (1-x)^{n-m} <$$

/we use continuity and boundedness of the function/

$$< \frac{o}{2} \sum_1 C_n^m \cdot x^m \cdot (1-x)^{n-m} + 2M \cdot \sum_2 C_n^m \cdot x^m \cdot (1-x)^{n-m} <$$

/using boundedness of both sums, obtained in the beginning of the proof/

$$< \frac{o}{2} \cdot 1 + 2M \cdot \varepsilon = \frac{o}{2} + \frac{o}{2} = o.$$

Now extend this result on the arbitrary segment $[a, b]$. Let

$$L = b - a, \quad F(t) = f(a + t \cdot L), \quad \varepsilon > 0.$$

By the preceding theorem, since $t \in [0, 1]$, there exists N, such that for all $n > N$

$$B_n(t) - F(t) < \varepsilon,$$

where

$$B_n(t) = \sum_{m=0}^{n} F(\frac{m}{n}) \cdot C_n^m \cdot t^m \cdot (1-t)^{n-m}.$$

Make a substitution $t = \frac{x-a}{L}$. We obtain, that

$$f(x) - B_n'(x) < \varepsilon,$$

where $B_n'(x) = B_n(\frac{x-a}{L})$. Therefore, $B_n'(x)$ is a Bernstein's polynomial for $f(x)$.

From previously mentioned the main statement of the Weierstrass theorem follows:

$$f(x) = B_1(x) + [B_2(x) - B_1(x)] + \ldots + [B_n(x) - B_{n-1}(x)] + \ldots \qquad (*)$$

where $(*)$ is the uniformly convergent series.

Now, we verify, that with the help of Bernstein's polynomials the problem of approximating of the given function could be solved. Estimate the degree of an error, arising while replacing the function $f(x)$ with the polynomial $B_n(x)$. Assume, that $f(x)$ is continuous and three times differentiable on the segment $[0, 1]$; moreover its third derivative is continuous. The equality $\sum\limits_{m=0}^{n} C_n^m \cdot x^m \cdot (1-x)^{n-m} = 1$ implies, that $B_n(x) - f(x) = [f(\frac{m}{n}) - f(x)] \cdot C_n^m \cdot x^m \cdot (1-x)^{n-m}$. By the Taylor's formula

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

$$= \left(\frac{m}{n} - x\right) \cdot f'(x) + \frac{1}{2} \cdot \left(\frac{m}{n} - x\right)^2 \cdot f''(x) + \frac{1}{2} \cdot \left(\frac{m}{n} - x\right)^2 \cdot [f''(q_n^m) - f''(x)],$$

where $q_n^m \in [\frac{m}{n}, x]$. Then

$$B_n(x) - f(x) = f'(x) \cdot \sum_{m=0}^{n} \left(\frac{m}{n} - x\right) \cdot C_n^m \cdot x^m \cdot (1-x)^{n-m} +$$

$$+ \frac{1}{2} f''(x) \cdot \sum_{m=0}^{n} \left(\frac{m}{n} - x\right) \cdot C_n^m x^m \cdot (1-x)^{n-m} +$$

$$+\frac{1}{2}\sum_{m=0}^{n}\left(\frac{m}{n}-x\right)^2\cdot[f''(q_n^m)-f''(x)]\cdot C_n^m\cdot x^m\cdot(1-x)^{n-m}=S_1+S_2+S_3,$$

that implies

$$|B_n(x)-f(x)-S_2|\leq|S_1|+\frac{1}{2}\sum_{m=0}^{n}\left(\frac{m}{n}-x\right)^2\cdot|f''(q_n^m)-f''(x)|\cdot$$

$$\cdot C_n^m\cdot x^m\cdot(1-x)^{n-m}=S_1+S_3.$$

Now, we prove, that $S_1=0$. In fact, differentiating the identity

$$(1-x+x)^n=\sum_{m=0}^{n}C_n^m\cdot x^m\cdot(1-x)^{n-m}// \tag{4}$$

and multiplying both sides by x, we obtain:

$$n\cdot x=\sum_{m=0}^{n}m\cdot C_n^m\cdot x^m\cdot(1-x)^{n-m}. \tag{5}$$

Dividing the last equation by n, we receive

$$x=\sum_{m=0}^{n}\frac{m}{n}\cdot C_n^m\cdot x^m\cdot(1-x)^{n-m},$$

that implies, that $S_1=0$. Repeating he operation of differentiating the identity (5) and multiplying by x, we obtain:

$$\sum_{m=0}^{n}m^2\cdot C_n^m\cdot x^m\cdot(1-x)^{n-m}=n\cdot(n-1)\cdot x^2+n\cdot x. \tag{6}$$

(1) and (2) implies, that $S_2=x\cdot\frac{1-x}{n}$.

Show, that S_3 tends to 0, as $n^{1+\varepsilon}$, where $\varepsilon>0$. Divide S_3 in two sums — $\sum\limits_1$ and $\sum\limits_2$, where $\sum\limits_1$ includes all x, satisfying the inequality $|\frac{m}{n}-x|<\frac{1}{n}$, and $\sum\limits_2$ - all others.

Since $f'''(x)$ is continuous on $[0,1]$, it is bounded on it. Since $f''(x)$ is differentiable on $[0,1]$, it is continuous on this segment, and also bounded on it. Let M — be the maximum of the modulus of $f''(x)$, and A — be the maximum of the modulus of $f'''(x)$. Then, obviously, there exists such $C>0$, that $w(d)<C\cdot d$, where w is the modulus of continuity of $f''(x)$. By the definition of the modulus of the continuity, we have

$$|f''(q)-f''(x)|\leq w(|q-x|),$$

and for any value x, whose $\sum\limits_1$ includes, it is true, that

$$w(|q-x|)<\frac{C}{n^{\frac{2}{5}}}$$

Thus:

$$\sum_1 \le \frac{C}{n^{\frac{2}{3}}} \cdot \sum_1 \left(\frac{m}{n}-x\right)^2 \cdot C_n^m \cdot x^m \cdot (1-x)^{n-m} <$$

/since $|\frac{m}{n}-x| < \frac{1}{n^{-\frac{2}{5}}}$/

$$< \frac{C}{n^{\frac{6}{5}}} \sum_1 C_n^m x^m (1-x)^{n-m} \le \frac{C}{n^{\frac{6}{5}}} \sum_2 < 2M \sum_2 \left(\frac{m}{n}-x\right)^2 C_n^m x^m (1-x)^{n-m} <$$

/since $|\frac{m}{n}-x| \ge \frac{1}{n^{-\frac{2}{5}}}$/

$$< 2M \cdot \sum_2 \cdot n^{\frac{4}{5}} \cdot \left(\frac{m}{n}-x\right)^4 \cdot C_n^m \cdot x^m \cdot (1-x)^{n-m} \le$$

$$\le 2M \cdot n^{\frac{4}{5}} \cdot \sum_{m=0}^{n} \left(\frac{m}{n}-x\right)^4 \cdot C_n^m \cdot x^m \cdot (1-x)^{n-m}.$$

Consider the identity (5) and differentiate it three times, multiplying each time by x. We obtain:

$$\sum_{m=0}^{n} m^3 \cdot C_n^m \cdot x^m \cdot (1-x)^{n-m} = n \cdot (n-1)(n-2) \cdot x^3 + 3n \cdot (n-1) \cdot x^2 + n \cdot x \quad (7)$$

Repeating the same operation, we obtain:

$$\sum_{m=0}^{n} m^4 \cdot C_n^m \cdot x^m \cdot (1-x)^{n-m} = \quad (8)$$

$$n \cdot (n-1) \cdot (n-2) \cdot (n-3) \cdot x^4 + 6n \cdot (n-1) \cdot (n-2) \cdot x^3 + \quad (9)$$

$$+7n \cdot (n-1)x^2 + n \cdot x. \quad (10)$$

The relationships (6),(7),(8) and (9) imply that:

$$\sum_{m=0}^{n} (m - n \cdot x)^4 \cdot C_n^m \cdot x^m \cdot (1-x)^{n-m} =$$

$$= 3n^2 \cdot x^2 \cdot (1-x)^2 + n \cdot x \cdot (1-x)(1-6x-6x^2) < K \cdot n,$$

where K — is a constant. Then,

$$2M \cdot n^{\frac{4}{5}} \cdot \sum_{m=0}^{n} (\frac{m}{n}-x)^4 \cdot C_n^m \cdot x^m \cdot (1-x)^{n-m} < 2M \cdot K \cdot n^{\frac{4}{5}} \cdot \frac{n^2}{n^4} = \frac{2M \cdot K}{n^{\frac{6}{5}}}.$$

Thus, $S_3 = \sum_1 + \sum_2 < \frac{C}{n^{\frac{6}{5}}} + \frac{2M \cdot K}{n^{\frac{6}{5}}}$, from where

$$|B_n(x) - f(x) - S_2| < \frac{C}{n^{\frac{6}{5}}} + \frac{2M \cdot K}{n^{\frac{6}{5}}},$$

i.e. $|B_n(x)-f(x)-S_2|$ is tending to 0 not slower, than n. Then, since $S_2 = \frac{1}{2}f''(x)\cdot x\cdot\frac{1-x}{n}$, we have

$$B_n(x) - f(x) \sim \frac{1}{2n}x\cdot(1-x)\cdot f''(x). \tag{11}$$

Therefore, for all x, where $f''(x)$ distincts from zero, the error of approximation of $f(x)$ with Bernstein's polynomials has a degree $\frac{1}{n}$ (if $f''(x) = 0$, then a degree of approximation is higher). As in the case of Lebesgue's polynomials, the degree of approximation of the given function by Bernstein's polynomials does not depend on f.

A Bernstein's polynomial, approximating $f(x)$ with the accuracy ε could be built much easier, than a Lebesgue's polynomial (the degree of Bernstein's polynomial is $\frac{1}{\varepsilon}$, while the degree of Lebesgue's polynomial is $\frac{1}{\varepsilon^3}$. Thus, Bernstein's polynomials are more convenient for practical use, then others, considered in that paper.

The proof of the Weierstrass theorem is the proof of the fact, that every continuous function, regardless of the way of its representing, has its own analytic expression. We could study the properties of any continuous function, with the help of approximating polynomials, that could be built, using the Bernstein's method.

Bibliography

1. A. I. Khinchin, Eight Lectures on Mathematical Analysis, Moscow, "Nauka", 1977. in Russian.

2. N. N. Luzin, Theory of Real-valued Functions, Moscow, "Uchpedgiz", 1948. in Russian.

3. V. L. Goncharov, Theory of Interpolation and Approximation of Functions, Moscow, "ONTI-GTTI", 1934. in Russian.

Anthony Likhodedov.

Graduated from the physical and mathematical school No. 239 in 1993. Student of the Dept. of Computer Technology since 1993. Winner of school olympiads in physics and mathematics in 1987–1993 and students' mathematical olympyads in 1993–1996. Absolute 2nd place in the nationwide students' olympiad in physics in 1995. Winner of the the title "Soros student" in 1995.

Max Sinitsyn.
Graduated from the physical and mathematical school No. 239 in 1993. Student of the Dept. of Computer Technology from 1993 to 1994. Winner of school olympiads in mathematics and informatics in 1987–1993. He now studies in Central Methodist College, Missouri, USA.

About Equivalence of Strong and Weak Convexity of Functions

M. Matveev

Introduction

Convexity is a concept of great importance in analysis. Consider the condition of convexity of an arbitrary real function: for all $x, y \in \mathbf{R}$, $\alpha \in [0, 1]$

$$F(\alpha x + (1-\alpha)y) \leq \alpha F(x) + (1-\alpha)F(y) \tag{1}$$

This condition could be simply verified for $\alpha = 1/2$:

$$F(\frac{1}{2}x + \frac{1}{2}y) \leq \frac{1}{2}F(x) + \frac{1}{2}F(y) \tag{2}$$

It is well known, that for any continuous function the conditions (1) and (2) are equivalent. In the current paper we consider the question of their equivalence for arbitrary functions. We will show, that the previous conditions are not equivalent in general, but the condition (1) is sufficient for convexity of measurable functions.

Strong and weak convexity

Basic definitions

Now, we give the basic definitions:

Definition 1 Function $F : \mathbf{R} \to \mathbf{R}$ is called strongly convex, if for all $x, y \in \mathbf{R}$, $\alpha \in [0, 1]$

$$F(\alpha x + (1-\alpha)y) \leq \alpha F(x) + (1-\alpha)F(y)$$

Definition 2 Function $F : \mathbf{R} \to \mathbf{R}$ is called strongly concave, if for all $x, y \in \mathbf{R}$, $\alpha \in [0, 1]$

$$F(\alpha x + (1-\alpha)y) \geq \alpha F(x) + (1-\alpha)F(y)$$

Definition 3 Function $F : \mathbf{R} \to \mathbf{R}$ is called weakly convex, if for all $x, y \in \mathbf{R}$

$$F((x+y)/2) \leq (F(x) + F(y))/2$$

The weak concavity could be defined similarly.

Evidently, strong convexity implies weak convexity (substituting $\alpha = 0.5$ in the definition of strong convexity, we get the definition of weak convexity).

An example of nonequivalence of strong and weak convexity

Now, we prove the existence of a weakly convex, but not strongly convex function.
Remark Further we will give only the proofs for convex functions (the same facts for concave functions could be proved by analogy with proofs for convex functions).

Theorem 1 *Any continuous weakly convex function is strongly convex.*

PROOF. We prove the inequality (1) for a binary rational α by induction on the power of 2 in denominator. If it is equal to 1, we get the definition of weak convexity.

Let α be equal to $\frac{k}{2^n}$, where n is an natural number, and k is an odd natural number. Then

$$F(\frac{k}{2^n}x + (1 - \frac{k}{2^n})y) \leq$$
$$\frac{1}{2}(F(\frac{k+1}{2^n}x + (1 - \frac{k+1}{2^n})y) + F(\frac{k-1}{2^n}x + (1 - \frac{k-1}{2^n})y)) \leq$$
$$\frac{1}{2}(\frac{k+1}{2^n}F(x) + (1 - \frac{k+1}{2^n})F(y) + \frac{k-1}{2^n}F(x) + (1 - \frac{k-1}{2^n})F(y)) \leq$$
$$\frac{k}{2^n}F(x) + (1 - \frac{k}{2^n})F(y)$$

The first of these inequalities is valid since $k+1$ and $k-1$ are even natural numbers.

Now, prove that fact for an arbitrary real α. There exists a sequence $\alpha_n \to \alpha$, where all α_n are binary rational. Realizing a passage to the limit at $n \to \infty$ in the proved inequality

$$F(\alpha_n x + (1 - \alpha_n)y) \leq \alpha_n F(x) + (1 - \alpha_n)F(y)$$

and using that F is continuous, we obtain (1), as was to be proved. ■

Thus, the concepts of weak and strong convexity (concavity) are equivalent for continuous functions. Consider the question of their equivalence for arbitrary functions. It is found, that it depends essentially on the extra axioms, added to the usual set of axioms.

There exist several assertions, independent on the standard set-theoretic axioms. Sometimes, it is reasonable to consider them as extra axioms. Most well-known and useful from them is an axiom of choice.

Axiom of choice may be formulated in three equal ways:

Axiom of choice *For any set of nonempty sets $\{E_\alpha\}$ there exists $\phi : \{\alpha\} \to E_\alpha$ such that $\phi(\alpha) \in E_\alpha$.*

Lemma (Zorn) *Let M be a partially ordered set. Then, existence of the supreme element for every its linear ordered subset implies existence of the supreme element for the whole M.*

Theorem (Zermelo) *Any set can be totally ordered.*

Definitions of partial, linear and total order on a set and proof of equivalence of these assertions can be founded in [1]. Further we use some more conceptions and facts from there (such as transfinite induction).

Now we show, that if we accept the axiom of choice, then weak convexity does not imply strong convexity.

Theorem 2 (Hamel basis) *In any vector space there exists a basis.*

Proof. Order a set of linear independent systems by inclusion:

$$X \leq Y \Longleftrightarrow X \subset Y$$

Obviously, every chain has the supreme element – the join of its elements. Therefore the whole set has the supreme element, that, evidently, will be a basis. Really, if $\{e_\alpha\}$ is a supreme element and x is a vector, which is out of linear cover of $\{e_\alpha\}$, we can join x to $\{e_\alpha\}$ and this will be linear independent system, containing $\{e_\alpha\}$. ■

Remark Any linear independent system may be completed to basis. This fact, containing previous theorem, can be proved similarly. We should only examine a set of linear independent systems, containing given system.

Definition 4 Function $F : R \to R$ is called additive, if for all $x, y \in R$

$$F(x+y) = F(x) + F(y)$$

Theorem 3 *There exists a nonlinear additive function.*

Proof. Let $\{e_\alpha\}$ be a basis in R over Q. We define F on the basis and then on the whole space R, by additivity. We set such values of F on the first three basis elements, that F is not linear, for example, $F(e_{\alpha_0}) = F(e_{\alpha_1}) = 0$, $F(e_{\alpha_2}) = 1$. Other values could be chosen in arbitrary way. ■

Remark Thus we can construct 2^c functions (while there exist only c continuous functions).

Corollary 3.1 *There exists a weakly convex, but not strongly convex function.*

Proof. Consider the function, constructed in the previous theorem. It is both weakly convex and concave. But if it is strongly convex and concave, then it is linear. But we have constructed it non-linear, so it is not strong convex and concave. □

(If it is strong concave and not strong convex, we can change its sign to obtain needed example.)

The concepts of strong and weak convexity for measurable functions

We see, that we can not assert the equivalence of two definitions for arbitrary functions. However, consider the largest "sensible" class of functions, studied in analysis – measurable. Further we will show, that it would be impossible to construct a measurable function with the, described above properties.

Theorem 4 *Let $F : \mathbf{R} \to \mathbf{R}$ be measurable and additive. Then $F(x) = ax$, where $a \in R$.*

Proof. We divide the proof into two parts: first we prove summability of the function, and then, that summability of the function F implies its continuity.

Prove the first part. Consider a segment $[a, b]$. We will prove, that it has a subsegment, where F is bounded above.(Then we shall choose analogously a subsegment of the chosen subsegment, where F is bounded below; then F is summable on this subsegment). Let $A_n = \{x \in [a, b] : F(x) > n\}, n \in Z$. Prove that by contradiction: suppose, that F is not bounded on any subsegment of [a,b]. Then, evidently, every subsegment of $[a, b]$ intersects with any A_n; that means, that every A_n is everywhere dense on $[a, b]$. Obviously, $\cap A_n = \emptyset$ and $\cup A_n = [a, b]$, so, according to the continuity of measure, we have $\lim_{n\to\infty} m(A_n) = 0$, $\lim_{n\to-\infty} m(A_n) = b - a$, where m is a Lebesgue measure. Hence, there exists $n \in Z : m(A_n) > m(A_{n+1})$. Choose $x_0 \in A_n \backslash A_{n+1}$ and $x \in A_{n+2} : 0 < x - x_0 < \epsilon$, where $\epsilon = m(A_n) - m(A_{n+1})$. It could be done, since A_{n+2} is everywhere dense. $y \in A_n \Rightarrow F(y) > n \Rightarrow F(y + x - x_0) = F(y) + F(x) - F(x_0) > n + (n + 2) - (n + 1) = n + 1$. So, we have $y \in A_n \Rightarrow (y + x - x_0) \in A_{n+1}$. $A_n + (x - x_0) \subset A_{n+1}$, therefore $m(A_n + (x - x_0)) \leq m(A_{n+1})$. Since m is translation-invariant, $m(A_n + (x - x_0)) \geq m(A_n) - (x - x_0)$ (since, $m(A_n + (x - x_0)\backslash[a, b])$ is not greater, than $(x - x_0)$). Thus $m(A_{n+1}) \geq m(A_n) - (x - x_0) > m(A_n) - \epsilon > m(A_{n+1})$ This is a contradiction.

We have proved the summability of the function F on a certain segment $[a, b]$. In order to prove the summability on any finite segment, we cover it with segments of the length $b - a$. If $[c, d]$ is such a segment, then $F(x) = F(x + a - c) + F(c - a)$, where $x + a - c \in [a, b]$, $x \in [c, d]$. Hence F is bounded and summable on $[c, d]$. Thus F is bounded on any finite segment and, hence, is summable on it.

Now, prove the continuity of the function. Notice, oddness of the function F: $F(0) = F(0 + 0) = 2 \cdot F(0)$, i.e. F(0) = 0; $F(-x) = F(0 - x) = F(0) - F(x) = -F(x)$. Let $a \in \mathbf{R}_+$.

$\int\limits_{-a}^{0} F(x)dm = \int\limits_{0}^{a} -F(x)dm = -\int\limits_{0}^{a} F(x)dm$

$\int\limits_{-a}^{0} F(x)dm = \int\limits_{0}^{a} F(x - a)dm = \int\limits_{0}^{a} F(x)dm - F(a)dm = \int\limits_{0}^{a} F(x)dm - aF(a)$, where $\int\limits_{a}^{b} F(x)dm$ is a Lebesgue integral.

Comparing the first and the second equalities, we obtain $\int\limits_{0}^{a} F(x)dm = a \cdot F(a)/2$ Let $x \to a$. Then $m([a, x]) \to 0$, and hence, $\int\limits_{a}^{x} F(y)dm \to 0$ (because of absolute continuity of Lebesgue integral). But $\int\limits_{a}^{x} Fdm = \int\limits_{0}^{x} Fdm - \int\limits_{0}^{a} Fdm = x \cdot F(x)/2 - a \cdot F(a)/2$ That implies continuity of the function $x \cdot F(x)/2$. Hence, F is continuous everywhere except 0 (and hence in 0). ∎

So, an additive nonlinear measurable function does not exists. Now we will prove the equivalence of strong and weak convexity (concavity) for measurable functions.

(This contains, certainly, result of previous theorem.)

Theorem 5 *Every measurable weakly convex function is strongly convex.*

PROOF. Let F be measurable, and $F((x+y)/2) \leq (F(x)+F(y))/2$ (weak convexity). We prove the following assertions:

1.Every segment contains a subsegment, where F is bounded.

2. If there exist subsegments to the right and to the left of point x_0, where F is bounded, then F is continuous in x_0.

Remark (on measurability) Measurability is required only in the first assertion.

PROOF. Now prove the first assertion. Examine an arbitrary segment. Without loss of generality, we may consider its measure to be equal to 1. Let $A_n = F^{-1}(n, +\infty)$. Evidently, $A_{n+1} \subset A_n$ and $mA_n \to 0 (n \to +\infty)$. If F is unbounded above on every subsegment, then every A_n is everywhere dense. Take $n : mA_n < 1/3$. Choose $a \in A_{n+1}$ near the center of the segment (so that the distance between a and the center is not greater, than 0.1). $F(x) < n$ implies $F(2 \cdot x - a) > n + 2$, since $F(x) + F(2 \cdot x - a) \geq 2 \cdot F(a) > (n+1) \cdot 2$. Therefore $x \in A_n \Rightarrow (2 \cdot x - a) \in A_{n+2}$. Thus the image of the complement, being of measure greater, than 2/3, under the central symmetry transformation is embedded in A_{n+2}, being of measure smaller, than 1/3. However, the measure of the "lost" part could not be greater than 0.1. This is a contradiction. Hence, F is bounded above on a certain subsegment. Let x be the center of this subsegment, $c = \sup F$ on it.

Then for all y $F(y) \geq 2 \cdot F(x) - F(2 \cdot x - y) \geq 2 \cdot F(x) - c$. Thus F is bounded below on this segment.

Prove the second assertion. Prove it by contradiction. Then, there exists a sequence $x_n \to x_0, \epsilon > 0 : |F(x_n) - F(x_0)| > \epsilon$ for all n (x_0 is an interior point of the domain of definition). Further, $F(x_n) \geq F(x)$ is required. Therefore, we change the sequence x_n to the sequence $x'_n : x'_n = x_n$, if $F(x_n) \geq F(x_0)$, and $x'_n = 2 \cdot x_0 - x_n$, if $F(x_n) < F(x_0)$. Then, $x'_n - x_0$ and $F(x'_n) > F(x_0) + \epsilon$ for all $n \in N$.

Notice, that $F(x + 2 \cdot d) - F(x + d) \geq F(x + d) - F(x)$. Therefore, if we move from x_0 by d, then one more time by d and so on, then image moves at least by (number of moves)$\cdot$ (the first step of image). Now we move from x_0 to our interval (where F is bounded) with the step $x_n - x_0$. We can make the step d as small as we want to, holding the first step of image greater then ϵ; therefore, when we reach the interval, the value can become as great as we want (since, the image always moves at least by ϵ). Thus, F could not be bounded on the interval. This is a contradiction. ■

To clarify the relations between the examined classes of functions, we prove following theorem.

Theorem 6 *Every strongly convex function is continuous.*

PROOF. Let F be strongly convex, x_0 – a point of its domain of definition. We will prove, that F is bounded in a certain vicinity of x_0. Take an arbitrary $\epsilon > 0$. On the segment $(x_0 - \epsilon, x_0 + \epsilon)$ we have $F(x) \leq \max\{F(x_0 - \epsilon), F(x_0 + \epsilon)\}$. Really, there exists $\alpha \in [0, 1]$ such that

$$x = \alpha(x_0 - \epsilon) + (1 - \alpha)(x_0 + \epsilon).$$

Substituting to definition of strong convexity, we see that

$$F(x) \leq \alpha F(x_0 - \epsilon) + (1 - \alpha)F(x_0 + \epsilon)$$

and this implies desired inequality. Hence, F is bounded above. The inequality $F(x) \geq 2F(x_0) - F(2x_0 - x) \geq 2F(x_0) - \sup F$ implies, that F is also bounded below. Thus, F is bounded and weakly convex on $(x_0 - \epsilon, x_0 + \epsilon)$. Thus, by the second assertion of the theorem 4, F is continuous. ■

Thus, on an arbitrary interval we have:

measurability + weak convexity⟺strong convexity

continuity + weak convexity⟺strong convexity

We have proved the equivalence of strong and weak convexity for measurable functions. However, it is impossible to prove the existence of non-measurable sets (and, hence, functions) within the framework of the standard theory of sets. It could be proved with the help of the axiom of choice, and accepting another axiom (the axiom of determination) it could be shown, that all sets are measurable.

Definition 5 Denote $T = \{f : \mathbf{N} \to \mathbf{N}\}$ Let $A \subset T$.

Examine the following game: player 1 defines $f(1)$, player 2 defines $f(2)$, and so on. The first player wins, if the built sequence $f \in A$; otherwise the second player wins. Define strategy as a map σ of the set of finite sequences of natural numbers into the set of natural numbers, if one's moving in accordance with it : $f(n) = \sigma(f(1), f(2), .., f(n-1))$, results in a victory (regardless of the rival's actions).

We call a set A determined, if there exists a strategy for one of the players.

The axiom of determination Every set $A \subset T$ is determined.

Accepting the axiom of determination, we obtain, that all R subsets are measurable (here we do not adduce the proof of that fact, see [2] for it). Hence, strong and weak convexity are equal.

Supplement

In this chapter we generalize a concept of weak convexity and prove that all obtained results are still true. Then we study some properties of the graph of nonlinear additive function, which show its "insensibility".

Definition 6 Let $0 \leq \alpha \leq 1$. Function $F : \mathbf{R} \to \mathbf{R}$ is called weakly convex with a constant α, if for all $x, y \in \mathbf{R}$

$$F(\alpha x + (1 - \alpha)y) \leq \alpha F(x) + (1 - \alpha)F(y).$$

Definition 7 Let $0 \leq \alpha \leq 1$. Function $F : \mathbf{R} \to \mathbf{R}$ is called weakly convex with a constant α, if for all $x, y \in \mathbf{R}$

$$F(\alpha x + (1 - \alpha)y) \geq \alpha F(x) + (1 - \alpha)F(y)$$

Theorem 7 *All the facts, proved for weak convexity are also true for weak convexity with a constant α.*

PROOF. The theorem could be proved by the absolute analogy with the previous proofs–instead of the field Q, we consider its extension $Q(\alpha)$. It is countable and dense everywhere; thus it has the properties, required from the basic field in the above proofs. ■

We'll need some definitions and theorems for our next result. See [3] and [1] for details.

Definition 8 A set is said to be of the first category, if it could be represented as at most countable union of nowhere dense sets.

Definition 9 A set is said to have the Baire property, if it could be represented as the symmetric difference of an opened set and a set of the first category.

Definition 10 Let $E \subset X \times Y$, $x \in X$. $E_x = \{y \in Y : (x, y) \in X \times Y\}$ is called a projection of E on Y along x. Projection on X is defined similarly.

Theorem 8 (Kuratowski-Ulam) *Let X, Y be topological spaces and $E \subset X \times Y$. If E has the Baire property and for all $x \in X$ (except, may be, a set of first category) E_x has the first category, then E has the first category.*

Theorem 9 (Fubini) *Let X, Y be spaces with measure and $E \subset X \times Y$. If E is measurable and for almost all $x \in X$ the section E_x has measure zero, then E has measure zero.*

The Continuum hypothesis (Cantor) There are no intermediate cardinals between continuum and the power of a countable set.

The last assertion does not depend on the set-theoretic axioms and can be used as the individual axiom.

Theorem 10 (properties of the graph of an additive nonlinear function) *The graph of an additive nonlinear function has the following properties.*

1. It is everywhere dense in $\mathbf{R}^2$; more, its intersection with any opened nonempty subset is continual.

2. It can be of measure zero and the first category simultaneously.

3. Assuming the validity of the continuum hypothesis it can be of the second category and not of measure zero simultaneously (and more, any measurable subset of its complement will be of measure zero).

4. If it is measurable, it is of measure zero; if it has the Baire property, then it is of the first category.

PROOF. 1. Examine an opened $A \subset \mathbf{R}^2$; there exists a rectangle $D = [a, b] \times [c, d] \subset A$, and F is not bounded on $[a, b]$ (otherwise, as it was proved, F is linear); if D does not contain continuum graph points, there are continuum of them above D or below D. Without loss of generality, we may consider, that there are continuum of them above D.

Prove that below there is also a certain point of the graph x_0. If there isn't such point, F is bounded below on $[a,b]$. Hence F is bounded above on $[-b,-a]$ (since $F(-x) = -F(x)$) and F is bounded above on $[a,b]$ (since $F(x) = F(x-a-b)+F(a+b)$ and $x - a - b \in [-b,-a]$ when $x \in [a,b]$) – contradiction.

If $G(x)$ is above D, then $G(\alpha \cdot x + (1-\alpha) \cdot x_0) \in D$ (where $G(t)$ is a point of the graph with abscissa t) at a certain $\alpha \in Q$. We can make continuum such pairs $(x, x_0$. Thus, $G \cap D$ is continual (G – graph), and hence, $G \cap A$ is also continual.

2. Let $\{e_\alpha\}$ be a Hamel basis. To define F, we can define its arbitrary values on $\{e_\alpha\}$. Set all the values to be rational (preserving nonlinearity). Then, G is contained in $R \cdot Q$, that, obviously, is of measure zero and first category.

3. Build a nonlinear additive function, whose graph intersects with every G_δ-set of the second category in R^2. Then it is of the second category, since the complement of a set of the first category contains an everywhere dense G_δ-set of the second category). There are continuum of opened sets, hence,there are continuum of G_δ-sets. By the continuum hypothesis, they could be totally ordered, so that every initial segment is be countable. Let $\{E_\alpha\}$ be this ordering. To build desired F it's sufficient to build a basis R over Q $\{e_\alpha\}$ with set F values on it such that $G(e_\alpha) \in E_\alpha$. Build it by transfinite induction. Suppose, that we have $\{e_\beta : \beta < \alpha\}$ – a linear independent system in R over Q (it is sufficient for it, that for all $\beta < \alpha$ $\{e_\gamma : \gamma < \beta\}$ is linearly independent), and $G(e_\beta) \in E_\beta$. The projection E_α onto the axis X is not countable, since E_α is of the second category; and $span\{e_\beta : \beta < \alpha\}$ is countable (here *span* is linear cover); therefore we can choose $e_\alpha \in R$, that is contained in the projection, but is not contained in $span\{e_\beta\}$. Set F value on e_α by the y-projection of the corresponding point E_α. Then, $G(e_\alpha) \in E_\alpha$ and $\{e_\beta : \beta \leq \alpha\}$ is linearly independent. Thus, we obtain a linearly independent vector system in R over Q with set values. Completing the basis, we obtain the desired function. By analogy with the above, the statement for measure could be proved. The class of closed sets with positive measure E'_α (such set is contained in any measurable set with positive measure) could be considered as the corresponding class of sets. It is continual, and projection of every its element onto the axis X is uncountable.

To build F with both properties we can simply choose on every step two vectors e_α and e'_α such that $G(e_\alpha) \in E_\alpha$, $G(e'_\alpha) \in E'_\alpha$ and $\{e'_\beta : \beta \leq \alpha\} \cup \{e'_\beta : \beta \leq \alpha\}$ is linearly independent.

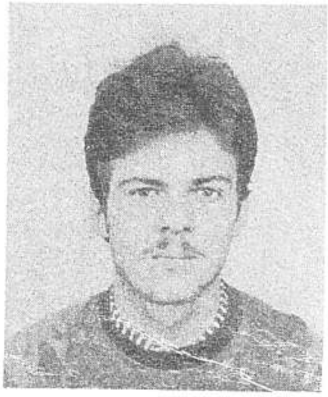
4. These facts could be obviously proved by the Fubini theorem and the Kuratowski-Ulam theorem respectively. ■

Bibliography

1. A.N. Kolmogorov, S.W.Fomin. Elements of Theory of Functions and Functional Analysis. Moscow, "Nauka", 1968. in Russian.

2. W.G. Canoway. Axiom of Choice and Axiom of Determination. Moscow, "Nauka", 1984. in Russian.

3. J.C. Oxtoby. Measure and Category. Moscow, "Mir", 1974. in Russian.



Michael Matveev.
Graduated from the physical and mathematical school No. 239 in 1994. Student of the Dept. of Computer Technology from 1994 to 1995. In 1995 entered St.Petersburg State University, Faculty of Mathematics and Mechanics. Winner of school olympiads in mathematics, physics and programming in 1988–1994 and students' mathematical olympiads in 1994 and 1995. Absolute winner of the nationwide school mathematical olympiad in 1993. Winner of the title "Soros student" in 1995 and 1996. He now leads a mathematical section in St.Petersburg Youth Creation Palace.

Uniformly Distributed Sequences

S. Egorov

Important theorems on uniform distribution

Consider a sequence x_k with a counter $A([a,b])\,(0 \le a < b \le 1)$, equal to the quantity of elements of the sequence $x_k(1 \le k \le k)$, such that the condition $\{x_n\} \in [a,b)$ is true (where $\{x\}$ is the fractional part of x).

Definition 1 The sequence x_k is called *uniformly distributed modulo 1 (u.d mod 1)*, if for all a and $b(0 \le a < b \le 1)$ the following is true:

$$\lim_{n\to\infty} \frac{A([a,b),n)}{n} = b - a \tag{1}$$

Obviously, the formula (1) is equal to the next one:

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} C_{[a,b)}(\{x_n\}) = b - a, \tag{2}$$

where $C_{[a,b)}(x)$ is a function, equal to 1 if $x \in [a,b)$, otherwise - to 0.

Theorem 1 *The sequence x_k u.d mod 1 if and only if for any real-valued continuous function $f(x)$ the condition*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} f(\{x_n\}) = \int_0^1 f(x)dx \tag{3}$$

is true.

Proof. Necessity:

First, we prove (3) for step-functions of the following kind: $\sum\limits_{i=0}^{m} d_i C_{[a_i,a_{i+1})}(x)$, where $0 = a_0 < a_1 < \ldots < a_{m+1} = 1$ is a partitioning of $[0,1)$.

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \sum_{i=0}^{m} d_i C_{[a_i,a_{i+1})}(\{x_n\}) = \sum_{i=0}^{m} \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} d_i C_{[a_i,a_{i+1})}(\{x_n\}) =$$

$$\sum_{i=0}^{m} d_i(a_{i+1} - a_i) = \int_0^1 \sum_{i=0}^{m} d_i C_{[a_i,a_{i+1})}(x)dx$$

The Darboux theorem implies, that for any continuous function $f(x)$ and $\forall\varepsilon > 0$ there exist such two step-functions $\phi_1(x)$ and $\phi_2(x)$, that $\phi_1(x) \leq f(x) \leq \phi_2(x) on[0,1]$ and

$$\int_0^1 (\phi_2(x) - \phi_1(x))dx < \varepsilon.$$

Then,

$$\int_0^1 f(x)dx - \varepsilon \leq \int_0^1 \phi_1(x)dx = \lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} \phi_1(\{x_k\}) \leq \underline{\lim}_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} f(\{x_k\}) \leq$$

$$\overline{\lim}_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} f(\{x_k\}) \leq \lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} \phi_2(\{x_k\}) = \int_0^1 \phi_2(x)dx \leq \int_0^1 f(x)dx + \varepsilon$$

Let $\varepsilon \to 0$. We have

$$\lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} f(\{x_k\}) = \int_0^1 f(x)dx$$

Sufficiency:

Assume the correctness of (3) for any continuous function. Consider $C_{[a,b)}$ where $[a,b) \subset [0,1)$. In this case, $\forall\varepsilon > 0$ there exist two continuous functions $g_1(x)$ and $g_2(x)$ such that $g_1(x) \leq C_{[a,b)}(x) \leq g_2(x)on[0;1]$ and

$$\int_0^1 (g_2(x) - g_1(x))dx < \varepsilon.$$

Then,

$$\int_0^1 C_{[a,b)}(x)dx - \varepsilon \leq \int_0^1 g_1(x)dx = \lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} g_1(\{x_k\}) \leq \underline{\lim}_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} C_{[a,b)}(\{x_k\}) \leq$$

$$\overline{\lim}_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} C_{[a,b)}(\{x_k\}) \leq \lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} g_2(\{x_k\}) = \int_0^1 g_2(x)dx \leq \int_0^1 C_{[a,b)}(x)dx + \varepsilon$$

Let $\varepsilon \to 0$. Thus, we obtain:

$$\int_0^1 C_{[a,b)}(x)dx = \lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} C_{[a,b)}(\{x_k\})$$

Since this expression's is true for all $[a,b) \subset [0,1)$ the sequence x_n is uniformly distributed modulo 1. ■

Corollary 1.1 *A sequence $\{x_n\}$ is uniformly distributed modulo 1, if and only if for any complex-valued continuous function $f(x)$, defined on R with a period 1, the following is true:*

$$\lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N} f(x_n)=\int_0^1 f(x)dx \tag{4}$$

PROOF. Having applied the theorem 1 to real and imaginary parts of f, we obtain that (3) is true for complex-valued functions. Since $f(\{x_n\}) = f(x_n)$, we obtain (4). To prove sufficiency, we notice that in the second part of the proof of the theorem 1, the functions g_1 and g_2 may be chosen so that the following additional conditions are true: $g_1(0) = g_1(1)$ $g_2(0) = g_2(1)$. Then, (4) may be applied to the periodic extensions of g_1 and g_2 to R. □

Theorem 2 (Weil criterion) *The sequence $\{x_n\}$ u.d mod 1 if and only if*

$$\forall h\in Z\backslash\{0\}:\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}e^{2\pi i h x_k}=0 \tag{5}$$

PROOF. Necessity obviously follows from the previous corollary. Now we prove sufficiency, i.e. that (5) implies the correctness of (4) for any complex-valued continuous function with period 1. According to the Weierstrass theorem $\forall\varepsilon > 0$ there exists such a finite linear combination of functions $e^{2\pi ihx}$ $(h \in Z)$ (denote it as $\psi(x)$), that $\sup|f(x) - \psi(x)| < \varepsilon$. That implies:

$$\left|\int_0^1 f(x)dx-\frac{1}{n}\sum_{k=1}^{n}f(x_k)\right|\le\left|\int_0^1(f(x)-\psi(x))dx\right|+\left|\int_0^1\psi(x)dx-\frac{1}{n}\sum_{k=1}^{n}\psi(x_k)\right|+$$

$$+\left|\frac{1}{n}\sum_{k=1}^{n}(\psi(x_k)-f(x_k)\right|\le 2\varepsilon+\left|\int_0^1\psi(x)dx-\frac{1}{n}\sum_{k=1}^{n}\psi(x_k)\right|$$

Since $\varepsilon\to 0$ and $n\to\infty$, we have:

$$\left|\int_0^1 f(x)dx-\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}f(x_k)\right|=0\Longleftrightarrow$$

$$\Longleftrightarrow\int_0^1 f(x)dx=\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}f(x_k)$$

That proves the statement of the theorem. ■

Example. Let x be an irrational number. Then $\{nx\}$ u.d mod 1.

PROOF.

$$\left|\frac{1}{n}\sum_{k=1}^{n}e^{2\pi ihkx}\right|=\frac{1}{n}\left|e^{2\pi ihx}\frac{e^{2\pi ihnx}-1}{e^{2\pi ihx}-1}\right|=$$

$$= \frac{1}{n}\left|\frac{e^{2\pi ihnx}-1}{e^{2\pi ihx}-1}\right| \leq \frac{1}{n|e^{2\pi ihx}-1|}$$

Since x is an irrational number and the denominator of the last fraction is nonzero everywhere, we obtain that the fraction tends to zero, as $n \to \infty$. ■

Quantitative characteristic of uniform distribution

As is easily seen, some uniformly distributed sequences are distributed "better", while others - "worse". Introduce a quantitative characteristic of deflection of the distribution of first n elements of a sequence from the uniform one.

Definition 2 Let $\{x_1, \ldots, x_n\}$ be a finite real-valued sequence. Then, the number

$$D_n^*(x_1, \ldots, x_n) = \sup_{\alpha \in (0,1]} \left|\frac{A([0,\alpha);n)}{n} - \alpha\right| \tag{6}$$

is called *the deflecting* of the sequence $\{x_1, x_2, \ldots, x_n\}$ ($X \subset [0,1)$ and $n \in \mathbf{N}$ meaning is the same as in the definition of a uniformly distributed sequence).

Theorem 3 *The sequence $\{x_n\}$ u.d mod 1 if and only if*

$$\lim_{n\to\infty} D_n^* = 0 \tag{7}$$

Proof. First prove sufficiency. Let $\lim_{n\to\infty} D_n^* = 0$. In this case

$$\forall \alpha \in (0,1] : \lim_{n\to\infty} \frac{A([0,\alpha);n)}{n} = \alpha,$$

that implies

$$\forall \alpha, \beta \in [0,1)(\alpha < \beta) : \lim_{n\to\infty} \frac{A([\alpha,\beta);n)}{n} = \beta - \alpha$$

Thus the sequence is u.d mod 1.

Now prove necessity: If a sequence is u.d mod 1, then, according to the definition, we obtain

$$\forall \alpha \in (0,1] : \lim_{n\to\infty} \frac{A([0,\alpha);n)}{n} - \alpha = 0.$$

Hence, $\lim_{n\to\infty} D_n^* = 0$. The theorem is proved. ■

Theorem 4 *Given n real numbers ($0 \leq x_1 \leq x_2 \leq \ldots \leq x_n < 1$). Then,*

$$D_n^* = \max_{1\leq i\leq n} \max\left(\left|x_i - \frac{i}{n}\right|, \left|x_i - \frac{i-1}{n}\right|\right) = \frac{1}{2n} + \max_{1\leq i\leq n}\left|x_i - \frac{2i-1}{2n}\right| \tag{8}$$

PROOF. To simplify designations, we denote $x_0 = 0$, $x_{n+1} = 1$. In this case

$$D_n^* = \sup_{0<\alpha\leq 1} \left| \frac{A([0,\alpha);n)}{n} - \alpha \right| = \max_{\substack{0 \leq i \leq n \\ x_i < x_{i+1}}} \sup_{x_i<\alpha\leq x_{i+1}} \left| \frac{A([0,\alpha);n)}{n} - \alpha \right| =$$

$$\max_{\substack{0 \leq i \leq n \\ x_i < x_{i+1}}} \sup_{x_i<\alpha\leq x_{i+1}} \left| \frac{i}{n} - \alpha \right| = \max_{\substack{0 \leq i \leq n \\ x_i < x_{i+1}}} \max \left(\left| \frac{i}{n} - x_i \right|, \left| \frac{i}{n} - x_{i+1} \right| \right) \quad (9)$$

Assume, that $x_i < x_{i+1} = \ldots = x_{i+r} < x_{i+r+1}$. Then

$$\forall j (1 \leq j < r) : \max \left(\left| \frac{i+j}{n} - x_{i+j} \right|, \left| \frac{i+j+1}{n} - x_{i+j+1} \right| \right) \leq$$

$$\max \left(\left| \frac{i+1}{n} - x_{i+1} \right|, \left| \frac{i+r}{n} - x_{i+r} \right| \right) \leq$$

$$\max \left(\max \left(\left| \frac{i}{n} - x_i \right|, \left| \frac{i+1}{n} - x_{i+1} \right| \right), \max \left(\left| \frac{i+r}{n} - x_{i+r} \right|, \left| \frac{i+r+1}{n} - x_{i+r+1} \right| \right) \right)$$

If $0 = x_{i+1} = \ldots = x_{i+r} < x_{i+r+1}$, we have:

$$\forall j (1 \leq j < r) : \max \left(\left| \frac{i+j}{n} - x_{i+j} \right|, \left| \frac{i+j+1}{n} - x_{i+j+1} \right| \right) \leq$$

$$\left| \frac{i+r}{n} - x_{i+r} \right| \leq\leq \max \left(\left| \frac{i+r}{n} - x_{i+r} \right|, \left| \frac{i+r+1}{n} - x_{i+r+1} \right| \right)$$

Thus, we may omit the condition $x_i < x_{i+1}$ in the first max of the formula (9). Then

$$D_n^* = \max_{0\leq i\leq n} \max \left(\left| \frac{i}{n} - x_i \right|, \left| \frac{i}{n} - x_{i+1} \right| \right) = \max_{1\leq i\leq n} \max \left(\left| \frac{i}{n} - x_i \right|, \left| \frac{i-1}{n} - x_i \right| \right) =$$

$$\frac{1}{2n} + \max_{1\leq i\leq n} \left| x_i - \frac{2i-1}{2n} \right|$$

That proves the theorem. ■

Now consider the practical method of creating sequences with small deviation.

Definition 3 A segment $[\frac{j-1}{2^{m-1}}, \frac{j}{2^{m-1}})$, where $j = 1, 2, \ldots, 2^{m-1}$, $m = 1, 2, \ldots$ is called *the binary segment* l_{mj}.

Definition 4 A set of 2^ν numbers $r_k \, (0 \leq r_k < 1)$ is called *the* Π_0*-net*, if and only if every binary segment $l_{\nu i}$ contains a number from that set.

Definition 5 A sequence $\{x_n\}$ is called *the* LPi_0*-sequence*, if and only if every its binary part is the Π_0-net where the binary part of a sequence is the set of such its elements x_i, that for each x_i, there exist $k, s \in \mathbf{N}$, such that $k \cdot 2^s \leq i < (k+1) \cdot 2^s$.

Remark Representing i in the binary form $i = \overline{c_\mu c_{\mu-1} \ldots c_1 c_0 e_s e_{s-1} \ldots e_1}$, we obtain that for all x_i of a binary part, the numbers $c_0, c_1, \ldots, c_\mu$ are fixed (here and further an over-line means, that a number is represented in the binary form).

Theorem 5 *Let $\{x_n\}$ be a LPi_0-sequence. Then, for the deviation D^* of the first N numbers of that sequence, the following is true:*

$$D^*(x_0, \dots, x_{N-1}) \leq \frac{r}{N}$$

where r is the number of ones in the binary form of the number N.

PROOF. We represent N in the following form: $N = 2^{\nu_1} + \dots + 2^{\nu_r}$, where $\nu_1 > \nu_2 > \dots > \nu_r \geq 0$. We select r binary parts of the sequence:

$$0 \leq i < 2^{\nu_1}; 2^{\nu_1} \leq i < 2^{\nu_1} + 2^{\nu_2}; \dots; 2^{\nu_1} + \dots + 2^{\nu_{r-1}} \leq i < N$$

where each is the Π_0-net (since $\{x_n\}$ is the LPi_0-sequence). For all these nets $D^{*(j)} \leq \frac{1}{N}$ (since $|A^{(j)}([0;\alpha), 2^{\nu_j}) - 2^{\nu_j}\alpha| \leq 1$)). As

$$A([0;\alpha), N) = \sum_{j=1}^{r} A^{(j)}([0;\alpha), 2^{\nu_j})N = \sum_{j=1}^{r} 2^{\nu_j},$$

we obtain

$$|A([0;\alpha), N) - N\alpha| \leq \sum_{j=1}^{r} |A^{(j)}([0;\alpha), 2^{\nu_j}) - 2^{\nu_j}\alpha| \leq r$$

Hence, $D^*(x_0, \dots, x_{N-1}) \leq \frac{r}{N}$. The theorem is proven. ■

Corollary 5.1 $D^*(x_0, \dots, x_{N-1}) \leq \frac{\log_2(N+1)}{N}$.

Thus, for LPi_0 sequences $D^* = O(\frac{\log N}{N})$, i.e. LPi_0-sequences have only a small deviation and are uniformly distributed very well (the sequences with $D^* = o(\frac{\log N}{N})$ are not found yet, though it wasn't proven, that they do not exist). Now, we investigate the question of building an LPi_0-sequence.

First we introduce the notation for the "exclusive or" operation. We denote it by "xor".

Consider now the concept of Binary-Distributed sequences (BD-sequences). Take an arbitrary sequence $\{V_s\}$ of binary rational numbers. We call the elements of that sequence the direction numbers.

Definition 6 Consider a sequence $\{r_i\}$, built, according to the following rule: if $i = \overline{e_m e_{m-1} \dots e_2 e_1}$, then $r_i = e_1 V_1 \text{ xor } e_2 V_2 \text{ xor } \dots \text{ xor } e_m V_m$. The sequence $\{r_i\}$is called *the BD-sequence* with the direction numbers $\{V_s\}$.

This definition is equivalent to the following:

a) $r(0) = 0; r(2^s) = V_{s+1}$

b) if $2^s < i < 2^{s+1}$, then $r(i) = r(2^s) xor r(i - 2^s)$

We represent V_s in the form of binary fractions: $V_s = \overline{0, v_{s1} v_{s2} \dots v_{sj} \dots}$. Then, $\{V_s\}$ may be defined as the infinite matrix v_{sj}, where for all elements of that matrix, the following is true: $v_{sj} \in \{0; 1\}$. The matrix v_{sj} is called the direction matrix of the sequence $\{r_i\}$.

Theorem 6 *If the direction matrix V_{sj}, corresponding to the BD-sequence is a lower-triangular matrix (i.e. $v_{ss} = 1$, and $v_{sj} = 0$ for all $j > s$), this BD-sequence is a LPi_0-sequence.*

Proof. Take an arbitrary part of the BD-sequence - $\{r_i\}$, of length 2^{-m}. Then, the numbers of the elements of that binary part can be represented as:

$$i = \overline{c_\mu c_{\mu-1} \dots c_{m+1} e_m e_{m-1} \dots e_2 e_1} \tag{10}$$

where $c_\mu, c_{\mu-1}, \dots, c_{m+1}$- are fixed, and $e_m e_{m-1} \dots e_2 e_1$ are arbitrary.
Take now an arbitrary binary segment $l = l_{m+1,j}$ of length 2^{-m}. In the binary system, that segment is defined by the following inequation:

$$\overline{0, a_1 a_2 \dots a_m} \leq x < \overline{0, a_1 a_2 \dots a_m} + \overline{0, 0 \dots 01}$$

where a_k are fixed. It should be proven, that for all sets a_k, c_k, there exists such a number i in the set of numbers, satisfying the equation 10, that $r_i \in l$.

To prove that, we represent r_i in the binary form: $r_i = \overline{0, g_{i1} g_{i1} \dots g_{ij} \dots}$. then, since

$i = \overline{0, e_\mu e_{\mu-1} \dots e_2 e_1}$ $\{r_i\}$ is a BD-sequence, the following is true:

$$g_{ij} = e_1 v_{1j} \mathbf{xor} e_2 v_{2j} \mathbf{xor} \dots \mathbf{xor} e_\mu v_{\mu j}$$

The condition $r_i \in l$ is equivalent to the condition $g_{ij} = a_j$, as $j = 1, 2, \dots m$, and the condition of belonging of i to the binary part 10 means, that $e_j = c_j$, as $j = m+1, \dots, \mu$.

Thus, granting the properties of the operation "xor" we obtain the following system of equations:

$$e_1 v_{1j} \mathbf{xor} \dots \mathbf{xor} e_m v_{mj} = a_j \mathbf{xor} c_{m+1} v_{m+1,j} \mathbf{xor} \dots \mathbf{xor} c_\mu v_{\mu j} \quad (1 \leq j \leq m)$$

In accordance with the conditions of the theorem the above system is triangular and the solutions may be consequently found. That proves the theorem. ■

Thus, the simple way of building the LPi_0-sequences is found.

Numerical integration

Let $\{x_n\}$ be a real-valued sequence u.d mod 1. In this case for any Riemann integrable function on $[0, 1)$ the following equality is true:

$$\int_0^1 f(x) dx = \lim_{n \to \infty} \sum_{n=1}^{N} f(\{x_n\})$$

(This assertion may be proved by complete analogy with the theorem 1). This formula may be used for approximation of the integral $\int_0^1 f(x) dx$. The advantage of this method (Monte-Carlo method) in comparison with quadrature methods is in absence of any conditions, imposed on the function f, except integrability, while the quadrature methods demand continuity, differentiability etc.

Estimate now an error of this method.

Lemma 6.1 *Given $0 \leq x_1 \leq x_2 \leq \ldots \leq x_n < 1$ - n points and $f(x)$ - a finite variation function on $[0,1]$. Then,*

$$\frac{1}{n}\sum_{k=1}^{n} f(x_k) - \int_{0}^{1} f(t)dt = \sum_{k=0}^{n} \int_{x_k}^{x_{k+1}} (t - \frac{k}{n})df(t) \tag{11}$$

where $x_0 = 0$ $x_{n+1} = 1$.

Proof.

$$\sum_{k=0}^{n} \int_{x_k}^{x_{k+1}} (t - \frac{k}{n})df(t) = \int_{0}^{1} tdf(t) - \sum_{k=0}^{n} \int_{x_k}^{x_{k+1}} \frac{k}{n}df(t) =$$

$$= f(1) - \int_{0}^{1} f(t)dt - \sum_{k=0}^{n} \frac{k}{n}(f(x_{k+1}) - f(x_k)) =$$

Using Abel's summing formula we obtain:

$$= f(1) - \int_{0}^{1} f(t)dt + (\frac{1}{n}\sum_{k=0}^{n-1} f(x_{k+1})) - f(1) = \frac{1}{n}\sum_{k=1}^{n} f(x_k) - \int_{0}^{1} f(t)dt$$

The lemma is proved.

Theorem 7 *Given f - a function with a finite variation $V(f)$ on $[0,1]$ and n points - $x_1, \ldots, x_n$ $[0,1)$ with deflection D_n^*. Then*

$$\left| \frac{1}{n}\sum_{k=1}^{n} f(x_k) - \int_{0}^{1} f(t)dt \right| \leq V(t)D_n^* \tag{12}$$

Proof. Without loss of generality we may consider $x_1 \leq x_2 \leq \ldots \leq x_n$. Then, by the previous lemma we have:

$$\left| \frac{1}{n}\sum_{k=1}^{n} f(x_k) - \int_{0}^{1} f(t)dt \right| \leq \sum_{k=0}^{n} \left| \int_{x_k}^{x_{k+1}} \left| t - \frac{k}{n} \right| df(t) \right| \leq$$

$$\sum_{k=0}^{n} \left| \int_{x_k}^{x_{k+1}} \max\left(\left| x_k - \frac{k}{n} \right|, \left| x_{k+1} - \frac{k}{n} \right| \right) df(t) \right| \leq$$

$$\leq \sum_{k=0}^{n} \left| \int_{x_k}^{x_{k+1}} D_n^* df(t) \right| \leq D_n^* \sum_{k=0}^{n} |f(x_{k+1}) - f(x_k)| \leq D_n^* V(f)$$

■

The described above method of the estimation of an error is correct only for functions with a finite variation, while the Monte-Carlo method is correct for all Riemann integrable functions. Moreover, some problems with estimation of the variation $V(f)$ may arise. Therefore we adduce another method of the estimation of an error in terms of the probability theory. Consider a random sequence as $\{x_n\}$. Uniform distribution of this sequence means that its distribution function is constant. Introduce the following notation:

$$I^* = \frac{1}{n}\sum_{k=1}^{n} f(x_k) \qquad I = \int\limits_0^1 f(x)dx$$

Assume the square integrability of $f(x)$ on $[0,1]$. Then, the variance of $f(x_i)$ is $\int\limits_0^1 (f(x) - J)^2dx$, where $J = \int\limits_0^1 f(x)dx$ is its mathematical expectation. From the probability theory, it is well known. that the mean square deviation of I_n^* is equal to $\frac{\sigma}{\sqrt{n}}$, where σ^2 is the variance. In this case it could be easily shown, that at sufficiently big n, the probability of J inclusion in the interval $(I_n^* - \lambda\frac{\sigma}{\sqrt{n}}, I_n^* + \lambda\frac{\sigma}{\sqrt{n}})$ is approximately equal to $2\Phi(\lambda) - 1$, where $\Phi(\lambda)$ is a normalized function of the normal distribution.

Thus, the Monte-Carlo method also has some serious demerits:
a) $|I - I_n^*| = O\left(1\sqrt{n}\right)$, while even in the method of rectangles the approximation error is $O\left(\frac{1}{n}\right)$,
b) the boundaries of errors are determined with a certain probability (not exactly), c) practically we deal with pseudorandom sequences.

Consider an example of a function, such that the Monte-Carlo method is more effective than the method of rectangles. Take

$$f(x) = \sum_{k=0}^{\infty}\left(\frac{3}{4}\right)^k \phi(4^k x),$$

where

$$\phi(x) = \begin{cases} x - 2n, & 2n < x \le 2n+1, \quad n \in Z \\ 2n + 2 - x, & 2n + 1 < x \le 2n+2, \quad n \in Z \end{cases}$$

It could be easily shown, that $f(x)$ is continuous everywhere and differentiable nowhere.

Now we find an approximate value of the integral $\int\limits_0^1 f(x)dx$ using the method of rectangles, partitioning $[0,1]$ on 4^m parts where $m \in \mathbf{N}$.

$$\int\limits_0^1 f(x)dx \approx 4^{-m}\sum_{n=0}^{4^m-1} f(n4^{-m}) = 4^{-m}\sum_{n=0}^{4^m} f(n4^{-m}) - 4^{-m} =$$

$$= 4^{-m}\sum_{n=0}^{4^m}\sum_{k=0}^{\infty}\left(\frac{3}{4}\right)^k \phi(n4^{k-m}) - 4^{-m} = 4^{-m}\sum_{n=0}^{4^m}\sum_{k=0}^{m}\left(\frac{3}{4}\right)^k \phi(n4^{k-m}) - 4^{-m} =$$

$$= 4^{-m}\sum_{k=0}^{m}\left(\frac{3}{4}\right)^k\left(4^k\sum_{n=0}^{4^{m-k}} n4^{k-m} - \frac{1}{2}4^k\right) + \frac{1}{2}4^{-m} - 4^{-m} =$$

$$= 4^{-m}\sum_{k=0}^{m} 3^k\left(4^{k-m}\frac{4^{m-k}(4^{m-k}+1)}{2}-\frac{1}{2}\right)-\frac{1}{2}4^{-m}=4^{-m}\sum_{k=0}^{m}\frac{1}{2}3^k4^{m-k}-\frac{1}{2}4^{-m}=$$

$$=\frac{1}{2}\sum_{k=0}^{m}\left(\frac{3}{4}\right)^k-\frac{1}{2}4^{-m}=\frac{1}{2}\frac{1-\left(\frac{3}{4}\right)^{m+1}}{1-\frac{3}{4}}-\frac{1}{2}4^{-m}=$$

$$=2\left(1-\left(\frac{3}{4}\right)^{m+1}\right)-\frac{1}{2}4^{-m}$$

The exact value of this integral is equal to:

$$\int\limits_0^1 f(x)dx=\frac{1}{2}\sum_{k=0}^{\infty}\left(\frac{3}{4}\right)^k=2$$

Hence the value of the error is

$$\Delta_{4^m}=\frac{1}{4^m}+2\left(\frac{3}{4}\right)^{m+1}$$

Let $n=4^m$. In this case

$$\Delta_n=\frac{1}{n}+\frac{3}{4}2\left(\frac{3}{4}\right)^{\log_4 n}=\frac{1}{n}+\frac{3}{2}\left(\frac{3}{4}\right)^{\frac{\log_{0,75} n}{\log_{0,75} 4}}=$$

$$=\frac{1}{n}+\frac{3}{2}n^{\frac{1}{\log_{0,75} 4}}=\frac{1}{n}+\frac{3}{2}n^{\frac{\ln 0,75}{\ln 4}}$$

As n is sufficiently big, the first term is small and

$$\Delta_n\sim n^{\frac{\ln 0,75}{\ln 4}}\qquad\left(\frac{\ln 0,75}{\ln 4}\approx -0,21\right)$$

Hence, when the choice of the quantity of partitioning points is so unsuccessful, the degree of approximation is about $n^{-0,21}$, while the degree of approximation of the Monte-Carlo method is $n^{-0,5}$. Hence for substantially nonsmooth functions (that change their values significantly on partitioning segments) the Monte-Carlo method can be more effective than quadrature ones.

Bibliography

1. L.Keipers, G.Niderraiter. Uniform Distribution of Sequences. Moscow, "Mir", 1991. in Russian.

2. J.Grenander, V.Franeberger. The Short Course of Computing Probability and Statistics. Moscow, "Mir", 1981. in Russian.

3. W.Rudin. The Principles of Mathematical Analysis. Moscow, "Mir", 1988. in Russian.

Serge Egorov.

Graduated from the physical and mathematical school No. 239 in 1994. Student of the Dept. of Computer Technology since 1994. Winner of school olimpiads in physics and mathematics 1988–1994 and students mathematical olympiads in 1994 and 1995. Absolute 4th place in the nationwide students' olympiad in physics in 1995. Winner of the title "Soros student" in 1995.

Measure and Category

M. Kondratjev

Introduction

This paper is devoted the theory of Lebesgue measure and Baire category of sets as method to prove theorems of existence. The existence of a mathematical object with some property is managed to prove demonstrating that the set of objects not having the prescribed properties is in some sense "small". There are many concepts of "smallness". Two of them are considered in details here: zero Lebesgue Measure and Baire's first category.

In the first part we deal with principal concepts of measure and category in Euclidean, metric and topological spaces. The similarities and distinctions between the classes of sets are discussed in the second part. These are sets of measure zero and sets of first category. Their properties are established there and the examples are given. The third part of the paper is devoted adducing some examples proving theorems of existence by means of measure and category. The problems of continuity (first class functions, integrability), differentiability (nondifferentiable everywhere functions), measurability of functions, equivalence of any set of first category to a set of measure zero are considered. All these examples provide versatile demonstration of application of measure and category to prove theorems of existence.

Measure and category: main concepts and theorems

In what follows we suppose all the properties of Lebesgue measure are known and we introduce only following definition:

Definition 1 A set $A \subset \mathbf{R}^n$ is *the set of Lebesgue measure zero*, if for any $\varepsilon > 0$ such a sequence of intervals $\{I_n\}$, that $A \subset \bigcup\limits_n I_n$ and $\sum\limits_{n=1}^{\infty} I_n < \varepsilon$ exists.

To introduce the concept of category we state the following:

Definition 2 A set A is *dense* in the interval I, if it has nonempty intersection with each subinterval of I ($\forall I_1 \subset I$: $I_1 \bigcap A \neq \emptyset$).

Definition 3 A set A is nowhere dense, if it is not dense in any interval.

It is possible to introduce two definitions, which are equivalent to Definition 3, and will be used later on.

Definition 4 A set A is *nowhere dense*, if its complement A' contains a dense open set.

Definition 5 A set A is *nowhere dense*, if its closure $\overline{A}$ has no internal points.

The class of nowhere dense sets is closed under some operations:

Theorem 1 *Any subset of nowhere dense set is nowhere dense. The union of finite number of nowhere dense sets is nowhere dense. The completion of nowhere dense set is nowhere dense.*

The proof of this theorem is obvious.

A denumerable union of nowhere dense set can be, generally speaking, not nowhere dense one; it can be even dense one. For example, the set of rational points is dense in $\mathbf{R}$, but it is union of denumerable number of points, each of them is nowhere dense set in $\mathbf{R}$. However, it is useful to introduce following definitions.

Definition 6 If a set can be represented of union of at most countable number of nowhere dense sets, it is *the set of first category.*

Definition 7 A set which is not the set of first category is *the set of second category.*

The following theorems are true for these classes of sets:

Theorem 2 (Baire theorem about category) *Complement of any set of first category is dense one. Intersection of any sequence of dense open sets is dense set.*

Theorem 3 *Any subset of a set of first category is also the set of first category. Union at most countable family of sets of first category is the set of first category.*

The proof of theorem 3 is obvious; proofs of both statements of theorem 2 are similar. Let us prove, for example, the first one:

Proof. let $A = \bigcup\limits_n A_n$ be a representation A in the form of denumerable union of nowhere dense sets. For arbitrary interval I we choose the interval $I_1 \subset I \backslash A_1$. Let I_2 is a segment: $I_2 \subset I_1 \backslash A_2$ and etc. Then $\bigcap\limits_n I_n$ is nonempty subset of set $I \backslash A$, hence, A is dense. ■

It follows from theorem 2 that any interval is set of second category.

Above we dealt only with Euclidean spaces. But the concept of category is topological one; if intervals in our definitions replace by open nonempty sets, we will introduce the concept of category in arbitrary topological space, and, obviously, Theorem 1 and 3 remain true there. The Baire theorem is correct under certain conditions: if a space is complete metric space, that is, if it is homeomorphic to space, where any Cauchy sequences converge. Thus, the following theorem is valid.

Theorem 4 *If (X, ρ) is a complete metric space and A is a set of first category in X then $X \backslash A$ dense in X.*

Proof. Let $A = \bigcup_n A_n$, where A_n are nowhere dense sets, and let S_0 is a nonempty open set. Let choose the sequence S_n of balls with radii $r_n < 1/n$ so, that $\overline{S}_n \subset S_{n-1}\backslash A_n (n \geq 1)$. It is possible to make step by step, choosing S_n as ball of sufficiently small radius with centre $x_n \in S_{n-1}\backslash\overline{A}_n$ (it is not empty, because A_n are nowhere dense). Then $\{x_n\}$ is a Cauchy sequence, because

$$\rho(x_i, x_j) < \rho(x_i, x_n) + \rho(x_n, x_j) < 2r_n \qquad \text{for} \quad \forall i, j \geq n.$$

Hence, $\exists x \in X$: $x_n \to x$. So that $x_i \in \overline{S}_n$ for $i \geq n$, then $x \in \bigcap_n \overline{S}_n \subset S_0\backslash A$. Hence, $X\backslash A$ dense in X. ■

It is possible to prove the extended version of the Theorem 3 using topological concepts.

Theorem 5 (Banach) *In a topological space X a union of any family of open sets of first category also is the set of first category.*

Proof. Let G denote a union of a family J of nonempty open sets of first category. Let $F = \{U_\alpha; \alpha \in A\}$ be maximum family of nonintersecting nonempty open sets, each of which is contained in set from J. Each of the sets U_α can be written as a union of nowhere dense sets: $U_\alpha = \bigcup_{n=1}^{\infty} N_{\alpha,n}$. Let $N_\alpha = \bigcup_{\alpha\in A} N_{\alpha,n}$. If open set U intersects N_n, it intersects some set from $N_{\alpha,n}$ as well, and therefore there is an open nonempty set $V \subset (U \cap U_\alpha)\backslash N_{\alpha,n}$. Thus $V \subset U\backslash N_n$, hence N_n is nowhere dense one. We note that $\overline{G}\backslash\bigcup F$ is a dense set nowhere (otherwise, F is not maximal). Hence, $G \subset (\overline{G}\backslash\bigcup F) \cup \bigcup_{\alpha\in A} U_\alpha = (\overline{G}\backslash\bigcup F) \cup \bigcup_{n=1}^{\infty} N_n$ is a set of first category. ■

As is obvious from what has been said that a concept of category is applicable not only to Euclidean space but also to any topological space moreover, the Baire theorem on category holds in any complete metric space. This permits to prove some interesting facts.

Sets of first category and zero measure

Here we study in more detailed manner the classes of sets of first category and measure zero.

Definition 8 A class of sets, which contains all possible denumerable union and any subset of its members, is called *σ-ideal.*

We note, that both the class of sets of first category and the class of sets of measure zero are σ-ideals.

Point and any denumerable sets are both the sets of measure zero and first category; on the other hand, no interval is included in any of these classes. Are there nondenumerable sets belonging to both these classes? The elementary example of such a set is Cantor set C, consisting of the points of $[0, 1]$, not containing the Fig. 1 in its ternary representation. This set can be got by throwing out from $[0, 1]$ the open middle interval $(1/3, 2/3)$, then throwing out the open middle parts from each of sections [0,1/3], [2/3,1], etc. If F_n is union of 2^n intervals with length $(1/3)^n$, remaining on step n, then $=\bigcap\limits_n F_n$. It happens that F_n does not contain any intervals with length more than $(1/3)^n$, hence does not contain any intervals, and, therefore, nowhere dense. On the other hand, sum of lengths of intervals, the components of F_n, is equal to $(2/3)^n$, hence, $\mu() = \mu(\bigcap\limits_n F_n) = \inf\limits_n \mu(F_n) = 0$so C is the set of measure zero. Nondenumerability of C is obvious.

Both the classes of the sets of measure zero and of the first category are σ–ideals, which contain all denumerable sets and some nondenumerable sets. The sets of these classes are "small" in certain sense. A nowhere dense set is small in the sense that it is "full of holes", while a set of first category can be approximated by such a set, and always has a dense set of discontinuities. A set of measure zero is "small" in metric sense, it can be covered by a sequence of intervals, is arbitrarily small total lengths. Maybe one of these classes of sets includes the other one? The following theorem gives a negative answer.

Theorem 6 **R** *is a union of two complementary sets A and B, where A is a set of first category and B is a set of measure zero.*

PROOF. Let $a_1, a_2, \ldots$ stands for an indexed set of rational numbers and let I_{ij} be an interval with length $(i+j)/2$ and centre a_i. Write $G_j = \bigcup\limits_{i=1}^{\infty} I_{ij}$ and $B = \bigcap\limits_{j=1}^{\infty} G_j$, $\forall\varepsilon > 0$: $\exists j$: $\frac{1}{2^j} < \varepsilon$, then $B \subset \bigcup\limits_{i=1}^{\infty} I_{ij}$ and $\sum\limits_{i=1}^{\infty} \mid I_{ij} \mid = \sum\limits_{i=1}^{\infty} \frac{1}{2^{i+j}} = \frac{1}{2^j} < \varepsilon$. Hence, B is the set of measure zero. On the other hand, G_j is the dense open subset of R, as it is a union of intervals and contains all rational points. Hence, its complement G'_j is nowhere dense one, $A = B' = \bigcup\limits_j G'_j$ is the set of first category. ■

As is obvious from what has been said, the set that is "small" in one sense, is not always "small" in other.

Consider Liouville numbers as one more example.

Definition 9 $z \in$ **R** *is Liouville number*, if $z \notin Q, foralln \in N$: $existp, q \in Z$: $\mid z - \frac{p}{q} \mid < \frac{1}{q^n}, q > 1$.

It is possible to prove, that any Liouville number is transcendental. Let E is the set of Liouville numbers. From Definition 9 follows, that

$$E = Q' \cap (\bigcap_{n=1}^{\infty} G_n),$$

where Q' is the set of irrational numbers, and

$$G_n = \bigcup_{q=2}^{\infty} \bigcup_{p=-\infty}^{\infty} (\frac{p}{q} - \frac{1}{q^n}, \frac{p}{q} + \frac{1}{q^n}).$$

The set G_n is a union of intervals. Besides $Q \subset G_n$ (it is obvious), hence G_n is the dense open set. Therefore its complement is the nowhere dense one and:

$$E' = Q \cup (\bigcup_{n=1}^{\infty} G'_n),$$

hence E' is a set of first category because it is a union of denumerable number of the sets of first category; hence, E is a set of second category.

On the other hand, is it possible to find measure of E? Obviously, that $E \subset G_n$ for any n. Consider

$$G_{i,j} = \bigcup_{p=-\infty}^{\infty} (\frac{p}{q} - \frac{1}{q^n}, \frac{p}{q} + \frac{1}{q^n}) \qquad (q = 1, 2, \ldots).$$

$\forall m, n \in N$:

$$E \cap (-m, m) \subset G_n \cap (-m, m) = \bigcup_{q=2}^{\infty} [C_{n,q} \cap (-m, m)] \subset \bigcup_{q=2}^{\infty} \bigcup_{p=-mq}^{\infty} (\tfrac{p}{q} - \tfrac{1}{q^n}, \tfrac{p}{q} + \tfrac{1}{q^n}).$$

Hence, $E \cap (-m, m)$ can be covered by sequence of intervals, the sum of lengths of which if $n > 2$ is equal to

$$\sum_{q=2}^{\infty} \sum_{p=-mq}^{\infty} \frac{2}{q^n} = \sum_{q=2}^{\infty} (2mq + 1) \frac{2}{q^n} \leq \sum_{q=2}^{\infty} (4mq + q) \frac{1}{q^n} =$$

$$(4m + 1) \sum_{q=2}^{\infty} \frac{1}{q^{n-1}} \leq (4m + 1) \int_{1}^{\infty} \frac{dx}{x^{n-1}} = \frac{4m + 1}{n - 2}.$$

Hence, $E \cap (-m, m)$ is the set of measure zero for any m. Therefore, E is a set of measure zero. Thus, E is "small" in sense of measure, but it is "large" in sense of category. Sets E and E' are an example corresponding to Theorem 6.

Theorems of existence

The existence of mathematical object with some property can proved using the following method. In the family of objects we consider a set of objects without this property, and prove that this set is in some sense "small" compared with the rest of the family. Hence, some objects (at least a single one) with the necessary property exist. This method will be considered in more detail in the following examples; we shall use concepts of category and measure as a degree of "smallness".

Nondenumerability of interval

The simplest example of the described above method is the proof of nondenumerability of interval. It was proven, that any denumerable set is a set of measure zero and first category; on the other hand, interval is the set of positive measure and of second category. Therefore, it can not be denumerable. Thus, we have proven the nondenumerability of an interval with help of both measure and category. Below we dwell upon more interesting examples of application of category for proofs of existence.

Functions of first class of Baire

Definition 10 Function f is *the function of first class (Baire)*, if it has the form of limit of convergent everywhere sequence of continuous functions.

Simple examples demonstrate that functions of first class are not always continuous everywhere. So, for example, $f_n(x) = \max(0, 1 - n \mid x \mid)$ are continuous, but they converge to a discontinuous function, which is equal to 1 if $x = 0$ and equal to 0 if $x \neq 0$. However, this function can not be discontinuous everywhere, as the following theorem asserts.

Theorem 7 (Baire) *If f is a function of first class, it is continuous everywhere, except a set of first category.*

Proof. Sufficient condition for proving this theorem is: $\forall \varepsilon > 0 : F_\varepsilon = \{x : \omega(x) \geq 5\varepsilon\}$ is nowhere dense (where $\omega(x)$ is an oscillation of a function f in x).

Let $f(x) = \lim\limits_{n\to\infty} f_n(x)$, where f_n are continuous, and let

$$E_n = \bigcap_{i,j\geq n} \{x; \mid f_i(x) - f_j(x) \mid \leq \varepsilon\} \qquad (n = 1, 2, \ldots).$$

Then E_n are close sets, $E_n \subset E_{n+1}$, $\bigcup\limits_n E_n = \mathbf{R}$. Consider any section $I \subset R$. As $I = \bigcup\limits_n (E_n \cap I)$, then all of $E_n \cap I$ can not be nowhere dense simultaneously. Hence, $\exists n \in N$: $E_n \cap I \ni J$, where J is interval. Then $\forall x \in J; i, j \geq n$: $\mid f_i(x) - f_j(x) \mid \leq \varepsilon$. By setting $j = n, i \to \infty$, deduce that $\forall x \in J$: $\mid f(x) - f_n(x) \mid \leq \varepsilon$. $\forall x_0 \in J$: $\exists$ vicinity $I(x_0) \subset J$: $\forall x \in I(x_0)$: $\mid f_n(x) - f_n(x_0) \mid \leq \varepsilon$. That means, $\forall x \in I(x_0)$: $\mid f(x) - f_n(x_0) \mid \leq 2\varepsilon$. Hence, $\omega(x_0) \leq 4\varepsilon$ and $J \cap F_\varepsilon = \emptyset$. Thus, $\forall I$: $\exists J$: $J \subset I \backslash F_\varepsilon$, hence, F_ε is nowhere dense. The set of discontinuity $F = \bigcup\limits_{n=1}^{\infty} F_{\frac{1}{n}}$ is the set of first category. ■

Thus, the functions of first class are continuous almost everywhere in sense of category. Theorem 7 is very important and useful result and it permits to answer some questions.

It is well known that trigonometrical series can pointwise converge to a discontinuous function. Is this function discontinuous everywhere? Theorem 7 gives us the answer, that its set of discontinuity is the set of first category, so a function is not discontinuity everywhere, and it is even almost continuous in sense of category.

It is also known that the derivative of differentiable everywhere function f can be not continuous everywhere. Is it possible that the derivative be everywhere discontinuous? The answer is negative, because the derivative $f'(x) = \lim\limits_{h\to\infty} \frac{f(x+\frac{1}{h})-f(x)}{\frac{1}{h}}$ is function of first class (if it is defined and finite).

So, we have found out when a function is continuous almost everywhere in sense of category. It is interesting to know when a function is continuous almost everywhere in sense of measure.

Riemann integrability

The next theorem states the answer this question

Theorem 8 *If f is Riemann integrable, its set of discontinuity has zero measure.*

This theorem is known from course of calculus (see [3], p. 566) and we do not prove it.

So, an integrable function is continuous almost everywhere in sense of measure; theorem 8 answer the some problems: an integrable function can not be discontinuous everywhere and etc.

However it is known that Theorem 8 is invertible (if function f is bounded). Theorem 7, generally speaking, is not invertible (counterexample see [1], p.61), however, if a function is considered on a perfect set P, it is invertible.

The inverse theorems are also the theorems of existence: if a function is almost continuous, it is Riemann integrable (it is a limit of everywhere convergent sequence of continuous functions).

Consider one more example connecting continuity and measurability.

Luzin's theorem

Theorem 9 (Luzin) *Function f is measurable if and only if, when for any $\varepsilon > 0$: there is E such that $\mu(E) < \varepsilon$ and restriction on $R\backslash E$ is continuous.*

This theorem is proved in course of functional analysis (see [2], p. 112). Note that it gives one more example of application of measure existence. If a function is continuous almost everywhere (except a set of arbitrarily small measure), then it is measurable. A similar theorem can be formulated for the concept of category, substituting measurability by Baire property:

Definition 11 A set A has *Baire property* if $A = F\Delta Q$, where F is close set, Q is the set of first category. Here Δ stands for symmetric difference between the sets.

The class of sets having Baire property is similar to the class of measurable sets. It is also σ–ideal and contains the class of Borel sets. Moreover, in the class of sets having Baire property the sets of first category play the role of sets of measure zero.

Definition 12 Function f *has Baire property* if for all $U \subset R$ such that U is open, $f^{-1}(U)$ has Baire property.

Theorem 10 *Function f has Baire property if and only if, when such set P of first category, that the restriction f on $R\backslash P$ continuously, exists.*

The proof of this theorem requires to study in detail the classes of sets and functions having Baire property and we do not adduce it here (see [1], p. 66). However, Theorem 10 itself gives brilliant example of proving theorem of existence with help of category. If function continuous almost everywhere (except a set of first category), it has Baire property.

As one more example of application of measure we would like to state the following.

Theorem 11 (Egorov) *If a sequence of measurable functions $f_n(x)$ pointwise converges to $f(x)$ on a set E of finite measure, then for all $\varepsilon > 0$: there is $F \subset E$ such that $\mu(F) < \varepsilon$ and $f_n \rightrightarrows f$ on $E \backslash F$.*

This theorem is also known from course of analysis (see [2], p. 110); it demonstrates, as the concept of measure helps to connect pointwise and uniform convergence: if $f_n \to f$ pointwisely, then $f_n \rightrightarrows f$ everywhere but on a set of arbitrarily small measure.

Notice that a similar theorem for category (in contrast to Luzin theorem) cannot be stated (example see [1], p. 69).

Everywhere nondifferentiable functions

The following example is especially interesting because the proof of the theorem is based on concept of category (in all the above-mentioned examples category is used in formulations of theorems, it is indicator of existence, but it is not a method of proving).
Definition 13 Metric ρ in $C[a,b]$ is *uniform metric*, if $\rho(f,g) = \sup\limits_{a \leq x \leq b} \mid f(x) - g(x) \mid$.

Consider space $C[0,1]$ with uniform norm. Consider

$$E_n = \{f : \exists x \in [0, 1 - 1/n] : \forall h : 0 < h < 1 - x \mid f(x+h) - f(x) \mid \leq nh\}.$$

Let us prove that E_n is close. Take any function from the closure of E_n and any sequence $\{f_k\} \in E_n$: $f_k \to f$. There exists subsequence $\{x_k\}$: $\forall k$, $0 < x_k < 1 - 1/n$ and $\mid f_k(x_k + h) - f_k(x_k) \mid \leq nh, \forall h$: $0 < h < 1 - x_k$. Assume that $x_k \to x$, where $0 \leq x \leq 1 - 1/n$, because it is possible to do if substitute $\{f_k\}$ by its subsequence. If $0 < h < 1 - x$, then for enough large k such that $0 < h < 1 - x_k$:

$$\mid f(x+h) - f(x) \mid \leq \mid f(x+h) - f(x_k + h) \mid + \mid f(x_k + h) - f_k(x_k + h) \mid +$$

$$+ \mid f_k(x_k + h) - f_k(x_k) \mid + \mid f_k(x_k) - f(x_k) \mid + \mid f(x_k) - f(x) \mid \leq$$

$$\mid f(x+h) - f(x_k + h) \mid + \rho(f, f_k) + nh + \rho(f_k, f) + \mid f(x_k) - f(x) \mid .$$

Let $k \to \infty$. For f is continuous in x and $x + h$, we deduce: $\mid f(x+h) - f(x) \mid \leq nh$ for $\forall h$: $0 < h < 1 - x$. Thus, E_n is closed.

Note that any function f from C can be uniform arbitrarily close approximate by a piecewise linear continuous function g. To demonstrate that E_n is nowhere dense in $C[0,1]$, sufficiently to prove, that for any such function g and any $\varepsilon > 0$ function $h \in C \backslash E_n$ exists and $\rho(g, f) < \varepsilon$.

Let M is maximum of slopes of linear segments of g. Take m such that $m\varepsilon > n + M$. Let $\varphi(x) = \min(x - [x], [x] + 1 - x)$ (saw-tooth function), $h(x) = g(x) + \varepsilon\varphi(mx)$. Then in every point of $[0, 1)$ function h has right derivative which is greater than n (The derivative $\varepsilon\varphi(mx)$ is equal to $\pm\varepsilon m$, the derivative of g is not greater than M). Hence, $h \in C \backslash E_n$. As $\rho(g, h) = \varepsilon/2$, then E_n is nowhere dense in $C[0,1]$. Hence, $E = \bigcup\limits_n E_n$ is a set of first category in $C[0,1]$. This set consists of continuous functions, with bounded right difference quotient in some point of $[0, 1)$. Similarly, the set of functions, having bounded left difference quotient in some point from $[0, 1)$, is the set of first category.

The union of such sets gives the set of functions, having finite unilateral derivative in some points of $(0,1)$.

Also it is possible to show that a set of functions, having infinite bilateral derivative in some point from $[0,1)$, is a set of first category. Thus, continuous function has totally disconnected derivative almost always in the sense of category. Moreover, the above–mentioned proof gives the constructive method to obtaine nondifferentiable everywhere function as a sum of uniform convergent series $\sum\limits_{n=1}^{\infty} \varepsilon_n \varphi(m_n x)$.

Generally speaking, application of Baire theorem about category to prove existence (nonemptyness) set is reduced to proof that some element of this set can be obtain as limit of constructed sequence. When this method is applicable, example is constructed by successive approximations.

Returning to our topic, pose another problem: does a function, which has no derivative at all, exist? Yes, it does, but method of category is already unapplicable. Moreover, set of such functions is the set of first category in $C[0,1]$.

The following example is connected with generality of concepts of measure and category. We already observed some common properties of these classes: they contain denumerable sets and does not contain any intervals, they are σ–ideals. Besides, some theorems, is formulated in terms of category, can be formulated in terms of measure, and vice versa. Examples of such duality are theorems 7 and 8, 9 and 10. In fact, duality of measure and category is more wider and deeply, but we do not touch it in this paper.

Mapping of linear sets of first category into sets of measure zero

Let H be the set of automorphisms $I \subset \mathbf{R}$. The following result holds:

Theorem 12 *For any set A of the first category in $I = [0,1]$ such $h \in H$ exists, that $h(A)$ is a set of measure zero. Moreover, set of such automorphisms has second category in H.*

We use completeness of H to prove this theorem.

PROOF. Let $A = \bigcup\limits_n A_n$, where A_n are nowhere dense. Let

$$E_{n,k} = \{h \in H : \mu(h(\overline{A}_n)) < 1/k\}.$$

For all $h \in E_{n,k}$: $h(\overline{A}_n)$ can be covered by open set $G \subset R$: $\mu(G) < 1/k$. There is $\delta > 0$: G contains δ–vicinity of any point of $h(\overline{A}_n)$. If $\rho(g,h) < \delta$, then $g(\overline{A}_n) \subset G$, and, hence, $g \in E_{n,k}$. Hence, $E_{n,k}$ is the open subset of H for all n,k. For all $g \in H, \varepsilon > 0$ divide I into finite number of close subintervals $I_1, I_2, \ldots, I_N$ with length less than ε. We take a close interval $J_i \subset J_i^0 \backslash g(\overline{A}_n)$, where J_i^0 is the interior of I_i. Let h_i is piecewise–linear homeomorphism I_i on itself, which leave fixed endpoints and map J_i onto the interval with length exceeding $\mid I_i \mid -\frac{1}{kN}$ (the graph of such function h_i can be constructed from three sections). Together these h_i define a map $h \in H$, that $\mu(h \circ g(\overline{A}_n)) < 1/k$. It means, $h \circ g \in E_{n,k}$. As $\rho(h \circ g, g) < \varepsilon$, then $E_{n,k}$ is dense in H. Thus, the set

$E = \bigcap_{n,k} E_{n,k}$ is a set of second category in H. If $h \in E$, then $h(\overline{A}_n)$ is the set of measure zero for $\forall n \in N$. As far as $h(A) \subset \bigcup_h h(\overline{A}_n)$, then $h(A)$ is a set of measure zero. ■

This theorem gives one more example when category is applied directly in the proof. The similar theorem can be proven for n-dimensional space.

So, we considered some examples of how the concepts of measure and category are applied to prove the theorems of existence. In one kind of theorems these concepts apply as condition of existence of objects, in another—directly for proof of existence.

Bibliography

1. J.C. Oxtoby. Measure and Category, Springer-Verlag, N.Y., 1971.
2. F. Riesz, B. Sz.-Nagy, Lecons d'analyse Fonctionnelle, Akademiai Kiado, Budapest, 1972. in French.
3. L.D. Kudryavcev. The Course of Calculus. Vol. I, Moscow, Vysshaya Shkola, 1988.

Michael Kondratjev.
Graduated from the physical and mathematical school No. 30 in 1994. Student of the Dept. of Computer Technology since 1994. Winner of school olimpiads in physics and mathematics in 1988–1994 and students' mathematical olympiads in 1994–1996. Winner of the title "Soros student" in 1995.

Theory of Vector Fields on the Plane with Applications

D. Raskin

In this paper the foundations of the theory of vector fields on the plane, a powerful instrument of the modern calculus, are stated and some its applications, including those in complex calculus, are considered.

Definition We say that in the flat area $\Omega \subset \mathbf{R}^2$ a vector field (or, simpler, a field) *is defined*, if a vector corresponds to each point of this area.

Definition We say that the field *turns to zero* in a given point, if the zero-vector corresponds to this point.

Consider a continuous field $\Phi(M)$ on the Jordan curve Γ without self-intersections. Introduce the parameter t, so that the curve Γ is defined by the system

$$\begin{cases} x = x(t) \\ y = y(t) \end{cases}$$

where $t \in (a; b]$ and $x(t)$,$y(t)$ are continuous. In this case the vector field $\Phi(M)$, or $\Phi(x, y)$ where x and y are the coordinates of M, on Γ can be represented as the function $\Phi(t)$.

Definition Let $\Phi(t)$ be a continuous function turning to zero nowhere. We'll call the angle function of the field Φ *a continuous branch of the multifunction:* the angle between $\Phi(t)$ and $\Phi(a)$, which turns to zero at $t = a$.

Denote the angle function by $\Theta(t)$.

Here are some obvious properties of the angle function:

1) The angle function does not change when we turn the whole vector field on the same angle.

2) The angle function does not change when we go over from the given vector field to the normalized one:

$$\Phi_1(M) = \frac{\Phi(M)}{\| \Phi(M) \|},$$

where $\| \Phi(M) \|$—is the norm of the vector $\Phi(M)$.

3) The angle function depends on the way of entering the parameter on Γ.

Now we define one of the basic terms of vector field theory —rotation.

Definition The *rotation* of the continuous vector field Φ on the Jordan curve Γ without self-intersections is the number

$$\gamma(\Phi,\Gamma)=\frac{1}{2\pi}(\Theta(b)-\Theta(a))=\frac{1}{2\pi}\Theta(b).$$

Here is an example of the field whose rotation on the given curve is 2:

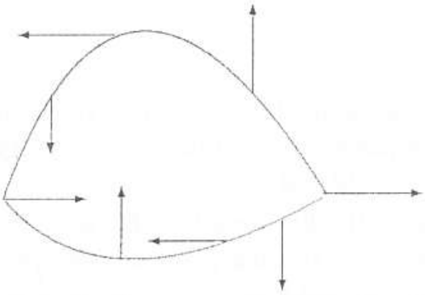


Fig. 1

Here are some obvious properties of rotation:

1) Rotation does not change when we turns the vector field or normalize it (see properties of the angle function).

2) Rotation does not depend on the way of entering the parameter on Γ, it depends only on the orientation of Γ: when we alter the orientation rotation changes sign.

3) Rotation of vector field on a curve, which is a union of some other curves is the sum of the field's rotations on these curves.

4) Rotation can be any real number: for getting a field with the given rotation you need only choose the angle function in the proper way. The fields of infinite rotation also exist, but we do not consider them here. Notice, not taking such fields into account, that if A and B are the ends of the curve Γ and $\Phi(A)$ and $\Phi(B)$ have the same direction, then $\gamma(\Phi,\Gamma)$ is an integer number, and if $\Phi(A)$ and $\Phi(B)$ are directed in an opposite way, then $\gamma(\Phi,\Gamma)=n+\frac{1}{2}$, where $n\in Z$.

There are different methods for calculation of rotation. We derive the Poincaré formula, which is one of these methods.

Let a field Φ be defined on the curve Γ:

$$\Phi(x(t),y(t))=\{\phi(x(t),y(t)),\psi(x(t),y(t))\}$$

with the angle function $\Theta(t)$. We introduce the function $\alpha(t)$ equal to the angle which the corresponding vector of the field forms with the x-axis. Then it is obvious that $d\Theta=d\alpha$ and we have

$$2\pi\gamma(\Phi,\Gamma)=\int\limits_{\Gamma}d\Theta=\int\limits_{\Gamma}d\alpha$$

$$\gamma(\Phi,\Gamma)=\frac{1}{2\pi}\int_a^b d\arctan\frac{\psi}{\phi}=\frac{1}{2\pi}\int_a^b\frac{\phi(t)\psi'(t)-\psi(t)\phi'(t)}{\phi^2(t)+\psi^2(t)}\,dt$$

Thus, we obtain

$$\gamma(\phi,\Gamma)=\frac{1}{2\pi}\int_a^b\frac{\phi(t)\psi'(t)-\psi(t)\phi'(t)}{\phi(t)\psi'(t)-\psi(t)\phi'(t)}\,dt,$$

which is the Poincaré formula.

We can define the rotation on a closed curve as the sum of the rotations on the two curves composing the initial curve in unification. Here the notion of positive circuit direction is used:

Definition A closed curve's circuit direction is called *positive*, if the inside area lies on the left of the curve. The opposite circuit direction is called *negative*.

Evidently, the rotation of the closed curve is integer and does not depend on the way of dividing it into two curves.

We need not sometimes know the exact value of the rotation, it is important for us only to know that it is not zero. The following criterion is useful for such tasks.

Theorem 1 *Let A be a continuous transformation of the closed curve Γ into itself without fixed point, a field Φ is defined on the curve Γ and*

$$\forall M\in\Gamma\quad\frac{\Phi(A(M))}{\parallel\Phi(A(M))\parallel}\neq\frac{\Phi(M)}{\parallel\Phi(M)\parallel}$$

Then

$$\gamma(\Phi,\Gamma)\neq 0.$$

Proof. Let $\gamma(\Phi,\Gamma)=0$. Let the parameter t be defined for Γ and let the transformation $\chi:[a,b]\to[a,b]$ (here a and b are ranges of t) correspond to the transformation A of the curve Γ into itself as a transformation of the parameter. As Γ is closed, we can continuously extend the angle functions $\Theta(t)$ and $\chi(t)$ periodically onto the whole real axis. The function $\Theta(t)$ must have a maximum and a minimum on the segment $[a,b]$—at the points t_1 and t_2 correspondingly. Then $\alpha(t)=\Theta(t)-\Theta(\chi(t))$ has a positive value at t_1 and a negative one at t_2. Thus, $\alpha(t)$ has a zero value between t_1 and t_2, and the contradiction with the condition

$$\forall M\in\Gamma\quad\frac{\Phi(A(M))}{\parallel\Phi(A(M))\parallel}\neq\frac{\Phi(M)}{\parallel\Phi(M)\parallel}$$

proves the theorem. ■

The next theorem with its corollary allows to calculate the rotations of some fields.

Theorem 2 *The rotation of the field of tangents to a smooth curve on this curve is 1. (We call the field which at every point has the value of the tangent to the curve the field of tangents. We can also define the field of inside and outside normals to the curve analogously though we do not turn our attention to those definitions.)*

PROOF. Smoothness of the curve and uniform continuity of the field's angle function $\Theta(t)$, where $t \in [a, b]$ is the curve's parameter, allows us to make divide the curve with points

$$M(t_1), M(t_2), \ldots, M(t_n)$$

so that:

1) $a = t_1 < t_2 < \ldots < t_n$;

2) the increment $\Delta\Theta_i$ of the angle function on every segment is equal to the angle between the vectors at the ends of the segment and less than π;

3) the tangents at $M(t_i)$ in continuation until intersecting with the tangents at $M(t_{i-1})$ and $M(t_{i+1})$ form a closed convex polygon:

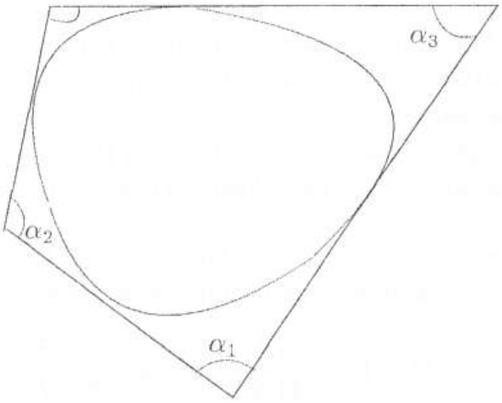


Fig. 2

This polygon as well as the curve does not have self-intersections. The rotation $\gamma(\Phi, \Gamma)$ is the sum of the angles between the positive directions of the polygon's sides a_i and the positive directions of a_{i+1}, taken in the interval $(-\pi, \pi)$, divided by 2π.

On the other hand, it is obvious that

$$\sum_{i=1}^{n} \alpha_i = 2\pi.$$

Hence, we have

$$\gamma(\Phi, \Gamma) = 1,$$

that finishes the proof. ■

Corollary 2.1 *The rotations of the fields of inside and outside normals are equal to 1.*

PROOF. The both fields can be obtained from field of tangents with a turn, and their rotations coincide with the rotations of the field of tangents, i.e. are 1. □

Definition Let closed curves $\Gamma_1 \ldots \Gamma_\nu$ without common points lie within the area Ω_0, bounded with a curve Γ_0. The area Ω consisting of points, which are inside Γ_0 and outside the others curves is called $\nu + 1$-*connected area.* The positive circuit direction of the curve Γ is the direction on which we have Ω on the left: counterclockwise for Γ_0 and clockwise for $\Gamma_1 \dots \Gamma_\nu$. By definition, the rotation of the field Φ on the curve Γ is the difference between the field Φ's rotation on Γ and the sum of its rotations on the curves Γ_i:

$$\gamma(\Phi, \Gamma) = \gamma(\Phi, \Gamma_0) - \sum_{i=1}^{\nu} \gamma(\Phi, \Gamma_i).$$

Now we prove a criterion of zero rotation on the boundary of a multiconnected area.

Theorem 3 *Let Ω be a $\nu + 1$-connected area and let its boundary consist of $\nu + 1$ Jordan curves. Let a continuous vector's field Φ turn to zero nowhere in the area Ω. Then the rotation of Φ on the area Ω's boundary Γ is zero.*

PROOF. By continuity of Φ there exists $\delta > 0$ such that

$$\forall M_1, M_2 \in \Omega \quad | M_1 M_2 |< \delta \Longrightarrow \angle(\overrightarrow{\Phi(M_1)}, \overrightarrow{\Phi(M_2)}) < \frac{\pi}{2}.$$

We divide Ω into areas $\sigma_1, \dots \sigma_n$ whose diameters (the largest distance between their points) is less than δ. We can do it with a finite set of Jordan arcs. Let $L_1, \dots L_m$ be the Jordan's arcs of which boundaries $\Pi_1, \dots \Pi_n$ consist. Here if L_i is on the boundary of Ω, then L_i is the boundary of only one area and L_i's circuit directions of the boundaries of both Ω and Ω_i coincide; if L_i is inside Ω, then it is the boundary of two areas and has opposite circuit directions on them:

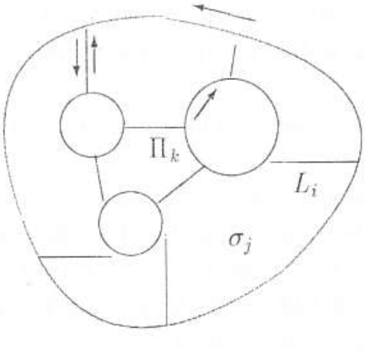


Fig. 3

Hence, obviously, the total rotation for two opposite circuit directions on the arcs inside Ω is zero, and only the arcs on the boundary of Ω does not have zero rotation. Then

$$\sum_{i=1}^{n} \gamma(\Phi, \Pi_i) = \gamma(\Phi, \Gamma).$$

On the other hand, $\gamma(\Phi, \Pi_i) = 0$ for any natural i from 1 through n because there are no oppositely directed vectors on any of the other arcs (otherwise the point would have existed of zero value of the field Φ). Therefore we obtain

$$\gamma(\Phi, \Gamma) = 0,$$

and so we have proved the theorem. ■

Definition Let a vector field Φ be defined on the whole Ω and continuous at all its points, except maybe just some of them. The *singular points* of the field Φ are the points at which it is not defined, not continuous or has zero-vector value. If there is a circle centered at such a point, within which there are no other singular points, we call this singular point *isolated.*

Definition Let M be an isolated singular point of the field Φ. Consider two circles: S_1 and S_2 centered at M of so small a radius that there are no other singular points inside the bigger circle (let it be S_2) or on it. Let Γ be the boundary consisting of the points of S_1 and S_2:

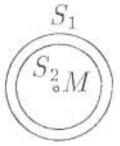


Fig. 4

By the previous theorem (Φ is continuous, defined inside and on S_2 and does not turn to zero there) $\gamma(\Phi, \Gamma) = 0$. So,

$$\gamma(\Phi, S_1) = \gamma(\Phi, S_2) = k.$$

This number k is called the *index of the singular point M*.

Notice that the notion of the index is correctly defined only for isolated singular points.

Now we go over to homotopy—one of the most important notions the theory of vector fields.

Definition Let the one-parameter set of vector fields $\Phi(M, \lambda)$, where $M \in \aleph$, $0 \leq \lambda \leq 1$, be defined on a closed set $\aleph$ and be continuous by the whole union of the variables and have no zero values. We say in this case that the set of the vector fields $\Phi(M, \lambda)$ *homotopically connects* the fields $\Phi_0(M) = \Phi(M, 0)$ and $\Phi_1(M) = \Phi(M, 1)$.

Definition We call two fields *homotopic* if we can connect them homotopically with a set of vector fields.

As a simple example, consider the fields Φ and $-\Phi$ in some plane area Ω. Notice that the set of vector fields

$$\Phi(M, t) = \Phi(M)(1 - 2t)$$

connects these fields homotopically and therefore they are homotopic.

Theorem 4 *Let Γ be the boundary of a $\nu + 1$-connected flat area Ω and let homotopic vector field Φ and Ψ be defined on it. Then*

$$\gamma(\Phi, \Gamma) = \gamma(\Psi, \Gamma).$$

PROOF. This theorem is reduced to the case of closed curve if we use the definition of vector field rotation on the boundary of the multiconnected area. So let Γ be a closed curve. Consider the set of the vector fields $\Phi(\lambda)$ which connects Φ and Ψ homotopically. Let $\Theta(t, \lambda_0)$ be the angle function of the field $\Phi(t, \lambda_0)$ where $\lambda_0 \in [0, 1]$. Then the function $\Theta(t, \lambda)$ is continuous by the whole union of the variables. On the other hand, the rotation of any field $\Phi(\lambda)$ is an integer number because Γ is a closed curve. It obviously implies that $\gamma(\lambda)$ is a constant and

$$\gamma(\Phi, \Gamma) = \gamma(\Psi, \Gamma).$$

The theorem is proved. ■

Notice that this theorem gives a positive answer for the question of existence of non-homotopic fields: any two fields of different rotations on the same curve are non-homotopic.

Now we prove an inverse theorem.

Theorem 5 *Let Γ be a closed curve and let vector fields Φ and Ψ be defined on it. Let it also be known that*

$$\gamma(\Phi, \Gamma) = \gamma(\Psi, \Gamma).$$

Then the fields Φ and Ψ are homotopic.

PROOF. Let the curve Γ be defined parametrically and the parameter t varies on the segment $[a, b]$. We normalize the fields Φ and Ψ and turn one of them so that at the point corresponding to the parameter's value $t = a$ values of the fields will be the same and its vectors will be directed like the positive part of the x-axis. Obviously, such a transformation does not have any influence on homotopy of the fields. Let $\Theta_0(t), \Theta_1(t)$ be the angle functions of the fields Φ and Ψ coinciding with the angle between the positive direction of the x-axis and the corresponding vector of one of the fields. Consider the set of the fields depending on the parameter $t \in [a, b]$:

$$\Phi(t, \lambda) = \{\cos[(1 - \lambda)\Theta_0(t) + \lambda\Theta_1(t)], \sin[(1 - \lambda)\Theta_0(t) + \lambda\Theta_1(t)]\}.$$

It obviously connects the fields Φ and Ψ homotopically, and the theorem has been proved. ■

We can conclude a characteristic property of homotopy from the two proved theorems:

Two continuous fields defined on a closed curve Γ and having no zero-vector values on Γ are homotopic if and only if their rotations are the same.

Now we state two other simple criteria of vector fields's homotopy:

1) If continuous vector fields defined on a curve Γ have no zero-vector values and nowhere vectors of these fields are oppositely (identically) directed, then the fields are homotopic.

PROOF. Let Φ and Ψ be the given fields. Let them be nowhere directed oppositely. Evidently the set of vector fields turning to zero nowhere

$$\Phi(M, \lambda) = (1 - \lambda)\Phi(M) + \lambda\Psi(M)$$

connects Φ and Ψ homotopically.

If the fields Φ and Ψ are nowhere identically directed, it is enough to notice that the fields Φ and $-\Phi$ are homotopic and to use obvious transitivity of homotopy. The proof is finished.

Before formulating the second criterion of homotopy it is necessary to give the following definition:

Definition A vector field Ψ is called the *main part* of a vector field Ψ if at any point M, where the field Φ is defined, the following formula takes place:

$$\Phi(M) = \Psi(M) + \omega(M),$$

where ω is a vector field defined on the field Φ's domain of definition and at all its points satisfying:

$$\| \omega(M) \| \leq \| \Psi(M) \| .$$

Now we go over to the second criterion of homotopy:

2) A continuous vector field is homotopic to its main part.

PROOF. Evidently, the field cannot be directed oppositely to its main part and so we can use the previous criterion to prove this one.

Now we touch upon some questions connected with application of vector field theory to complex variable function theory.

We will denote a point of the complex plane $x + iy \in \mathbf{C}$ by the letter z. The transformation

$$f(z) = U(x,y) + iV(x,y)$$

defines the vector field on the plane:

$$f(x,y) = \{U(x,y), V(x,y)\}$$

(sometimes we will call such a field as the field $f(z)$). The singular point of this field (or its index) we will call the singular point of the function f (or its index).

Lemma *The index of the singular point* 0 *of the function z^n is n for an integer n.*

PROOF. Consider the transformation $\omega = z^n$ and the corresponding vector field. Notice that if t is the angle between x-axis and the vector corresponding to z, then

$$z = \rho e^{it},$$

where $\rho = \mid z \mid$ and so

$$\omega = \rho^n e^{int}.$$

Consider this field on the unit circle. The angle function on it has the form

$$\Theta(t) = nt.$$

Evidently, the rotation of the vector field on this circle and the index of the singular point 0 are n, that proves the lemma. □

Let $f(z)$ be analytic at z_0 and $f(z_0) = 0$. Then we can represent $f(z)$ in the form

$$f(z) = a_0(z - z_0)^n \phi(z),$$

where $a_0 \in \mathbf{C}$, $\phi(z)$ is analytic at z_0 and $\phi(z_0) = 1$. The number n is called the order of the null of the function f.

Theorem 6 *The index of a null of an analytic function of order n is n.*

PROOF. Let $f(z) = a_0(z - z_0)^n \phi(z)$ be the investigated function ($f(z_0) = 0$) which is analytic at z_0 and let $h(z) = (z - z_0)^n \phi(z)$. The vectors $\overrightarrow{h(z)}$ and $\overrightarrow{(z - z_0)^n}$ are not directed oppositely on the circle centered at z_0 of radius small enough because:

$$\lim_{z \to z_0} (\arg(h(z)) - \arg((z - z_0)^n)) = \lim_{z \to z_0} \arg \frac{h(z)}{(z - z_0)^n} = \lim_{z \to z_0} \arg \phi(z) = 0.$$

Hence, the rotations of the field given by the function $h(z)$ on these circles are n just as the index of the singular point z_0 of the function $(z - z_0)^n$. The function $f(z)$ differs from $h(z)$ with multiplying by the complex constant a_0 which is reduced to multiplication of the absolute values of the vector field by the constant a_0 and addition of z to their arguments. Such a transformation will not tell on the rotation, so the index of the singular point z_0 of the function f is n. The theorem has been proved. ■

Definition Let $f(z)$ be not defined at the point z_0 and is represented in a neighborhood of z_0 in the form

$$f(z) = \frac{\phi(z)}{(z - z_0)^n},$$

where $\phi(z)$ is analytic in the neighborhood of z_0. Then z_0 is called a *pole of the function* f and the number n is called its *order*.

Theorem 7 *The index of a pole of a function f is equal to its order taken with the minus sign.*

PROOF. Notice that the field corresponding to the function $\frac{\phi(z_0)}{(z-z_0)^n}$ is the main part of the field corresponding to the function $f(z)$ in the neighborhood of z_0 of radius small enough because

$$f(z) = \frac{\phi(z_0)}{(z - z_0)^n} + \frac{\phi(z) - \phi(z_0)}{(z - z_0)^n}$$

and as $\lim\limits_{z \to z_0} \phi(z) - \phi(z_0) = 0$, if z is near enough to z_0, we have

$$\left| \frac{\phi(z_0)}{(z - z_0)^n} \right| > \left| \frac{\phi(z) - \phi(z_0)}{(z - z_0)^n} \right|.$$

The rotation of the field $f(z)$ is equal to the rotation of the field $\frac{\phi(z_0)}{(z-z_0)^n}$ which in turn is obtained from the field $\frac{1}{(z-z_0)^n}$ with multiplying the vectors by the real constant and turning them on the given angle that has no influence on the rotation. And the rotation of the field $\frac{1}{(z-z_0)^n}$ is $-n$. Therefore the index of the singular point z_0 is $-n$. The theorem has been proved. ■

Now, when we have the necessary theoretical base, we can prove the Bohl–Brower theorem which is one of the most important theorems of fixed point. This theorem sometimes allows, as we will see later, to solve the question of existence of the solution of a system of equations.

Theorem 8 (Bohl–Brower) *Let F be a continuous transformation of a circle into itself. Then F has at least one fixed point.*

PROOF. Consider the vector field

$$\Phi(M) = F(M) - M.$$

If there are no fixed points of the transformation F on the boundary circumference S of the circle K, then the field Φ does not turn to zero-vector on S and the field F is nowhere directed oppositely to the field of inside normals to the circumference because the field Φ is directed from the point M to the point $F(M)$. Hence the rotation of the field F on S coincides with the rotation of the field of inside normals to the circumference and

$$\gamma(\Phi, S) = 1.$$

Thus, $\gamma(\Phi, S) \neq 0$ and therefore the filed Φ must turn to zero at least once on the circle K according to one of the proved theorems. So the transformation F has at least one fixed point. The theorem is proved. ■

It is easy to prove the following generalization of the Bohl–Brower theorem:

Theorem 9 *Any continuous transformation of a set T homeomorphic to a circle into itself has at least one fixed point.*

PROOF. Let F be the given transformation, let K be a circle and let B be a homeomorphism of K into the set T. Then $B^{-1}FB$ is a transformation of the circle K into itself which has a fixed point because of its continuity. Let $B^{-1}FB(M_0) = M_0$. Then evidently $B(M_0)$ is the fixed point of the given transformation. The theorem has been proved. ■

As an example of using the Bohl–Brower theorem we affect the question of existence of solutions of equation systems. Consider an equation system of the form:

$$\begin{cases} x = P(x,y) \\ y = Q(x,y) \end{cases}$$

where P and Q are defined and continuous at all real x and y. Define a transformation of the plane into itself:

$$F(x,y) = \{P(x,y), Q(x,y)\}.$$

Then the given system is equivalent to the vector equation

$$M = F(M).$$

The question of existence of the system's solution turns into the question of existence of fixed points of the transformation F. Therefore we can obtain different criteria of existence of the system's solutions. For example, let functions P and Q be bounded:

$$\forall\, x, y \in \mathbf{R} \left\{ \begin{array}{l} \mid P(x,y) \mid\leq c_0 \\ \mid Q(x,y) \mid\leq c_0 \end{array} \right.$$

Then the transformation F transfers the circle defined by the inequality

$$x^2 + y^2 \leq 2c_0^2 b$$

into itself and so it has at least one fixed point and the given system has at least one solution.

We can easily prove the basic algebra theorem using of vector field theory.

Theorem 10 *Any nonconstant polynomial*

$$f(z) = z^n + a_1 z^{n-1} + \ldots + a_n,$$

where $a_1, \ldots, a_n$ are complex coefficients has at least one complex root.

Proof. Let $R >\mid a_0 \mid + \mid a_1 \mid + \ldots + \mid a_{n-1} \mid +1$. Then on the circle defined by the equation $\mid z \mid= R$ the following inequality takes place:

$$\mid f(z) - z^n \mid\leq\mid a_{n-1} \mid R^{n-1} + \mid a_{n-2} \mid R^{n-2} + \ldots + \mid a_1 \mid R + \mid a_0 \mid< R^n =\mid z^n \mid.$$

So the rotation of the field $f(z)$ on this circle is equal to the rotation of its main part z^n, i.e. n. Therefore the field $f(z)$ turns to zero at least at one point and the polynomial $f(z)$ has at least one complex root. The theorem has been proved. ■

Since the index of the null of any polynomial is equal to its order, the sum of the orders of all the roots of a polynomial of degree n is n.

It was shown that the index of any of the function's nulls is equal to its order and the index of a pole is equal to its order taken with the minus sign.

Let $f(z)$ be an analytic function having a finite number of nulls and poles in the area $\Omega \in \mathbf{C}$ and continuous at all the points of the bounded closed area $\overline{\Omega}$ except the poles, and let function f have no nulls on the boundary Γ of the area Ω. Let $N_1(f, \Omega)$ be the number of the nulls of the function f in the area Ω, let $N_2(f, \Omega)$ be the number of the poles, counting every null and pole as many times as great its order is. If we consider a pole as a zero of a negative order, it is natural to call the difference

$$N_1(f, \Omega) - N_2(f, \Omega)$$

the algebraic number of nulls. Let the initial value of the argument of the function $f(z)$ be ϕ_0 when going around the contour Γ and let the final value of the argument after the whole round be ϕ_1. Then

$$N_1(f, \Omega) - N_2(f, \Omega) = \frac{1}{2\pi}(\phi_1 - \phi_0).$$

This formula is called the argument principle in the theory of analytic functions. If a function $g(z)$ analytic on Γ satisfies the condition

$$\forall z \in \Gamma \quad | g(z) <| f(z) |,$$

then the fields $f(z)$ and $f(z)+g(z)$ have the same numbers of nulls because the field $f(z)$ is the main part of the field $f(z)+g(z)$ and their rotations are equal on Γ. Therefore if the functions f and g have no poles in the area Γ, then f and $f+g$ have the same number of nulls in the area Ω. This statement is called the Rouche theorem in the theory of analytic functions.

Let Γ be a closed piecewise smooth curve and let f be a function, analytic on the area Ω bounded by the curve Γ. Then $| f(z) |$ alters continuously and does not turn to zero. The function $\ln | f(z) |$ does not have an increment on the whole round along the curve Γ. So

$$\oint_\Gamma \ln | f(z) | = 0 \ (*).$$

If γ is the rotation of the field $f(z)$ on the curve Γ, then

$$\gamma = \frac{1}{2\pi}(\phi_1 - \phi_0) = \frac{1}{2\pi} \oint_\Gamma d \arg(f(z)) = \frac{1}{2\pi i} \oint_\Gamma d \ln e^{i \arg(f(z))}.$$

In union with $(*)$, it gives

$$\gamma = \frac{1}{2\pi i} \oint_\Gamma d \ln(| f(z) | e^{i \arg(f(z))}.$$

Hence

$$\gamma = \frac{1}{2\pi i} \oint_\Gamma \frac{f'(z)}{f(z)} dz.$$

This expression is called the logarithmic residue of the function $f(z)$ with respect to the curve Γ. The basic theorem about logarithmic residue derives from the previous reasoning and is expressed by the following formula:

$$\frac{1}{2\pi i} \oint_\Gamma \frac{f'(z)}{f(z)} dz = N_1(f, \Omega) - N_2(f, \Omega).$$

We return to systems of equations now. Let the following system be given:

$$\begin{cases} P(x, y) = 0 \\ Q(x, y) = 0 \end{cases}$$

where the functions P and Q are continuous by the union of variables in a closed area Ω with a boundary Γ. Consider the continuous vector field

$$\Phi(x, y) = \{P(x, y), Q(x, y)\}$$

on Ω. If the rotation of this field on Γ is not zero, then at least one solution of the system exists in the area Ω. Otherwise we can try to find subareas $\Omega_1 \subset \Omega$ on whose boundaries the rotation is not zero.

We can, for example, find the rotation approximately by the Poincaré formula with precision 0.5 and after that round the result to an integer number. Since the rotation of the vector field is always integer (if it exists) we will obtain the exact value of the rotation. We can use this method also when we estimate the error of the approximate solution $\{x_0, y_0\}$. If the rotation of the field Φ on the circle defined by the equation $(x-x_0)^2 + (y-y_0)^2 = \rho^2$, where ρ is a positive number, is not zero, then the found approximate value of the solution is at a distance is less than ρ from some exact solution.

Now let one solution of the system be known and let its index be known too. If this index differs from the rotation of the field Φ on Γ, then some other solution or solutions exist.

Another method of proving existence of the solution is based on non-homotopy of vector fields of different rotations. Look at this method in detail.

Given a system of equations:

$$\begin{cases} P(x,y) = 0 \\ Q(x,y) = 0 \\ R(x,y) = 0 \end{cases}$$

where the functions P, Q and R are continuous. The equation $R(x,y) = 0$ defines some flat line. Suppose this line contains a closed Jordan curve Γ. Consider the set of the vector fields Φ_t:

$$\Phi_t(M) = \Phi_t(x,y) = \{P(x,y,t), Q(x,y,t)\}$$

on Γ. Let $\gamma(\Phi_{t_1}, \Gamma) \neq \gamma(\Phi_{t_2}, \Gamma)$. Then there exists some $t \in (t_1, t_2)$ at which field Φ_t turns to zero somewhere on Γ—otherwise Φ_{t_1} and Φ_{t_2} would have been homotopic on Γ and had the same rotation on it. The zero-vector value of the field Φ_t on Γ defines the solution of the given system.

Bibliography

1. M.A. Krasnoselskivı. Vector Fields on the Plane. Moscow, "Nauka", 1963. in Russian.

Daniel Raskin.

Graduated from the physical and mathematical school No. 239 in 1994. Student of the Dept. of Computer Technology since 1994. Winner of school olimpiads in physics and mathematics 1990–1994. Winner of the title "Soros student" in 1995.

Symmetric Transformations from Group-theoretic Point of View

P. Viewkova

The concept of symmetry in the most general sense means that an object or a phenomenon under consideration has something permanent or invariant with respect to some transformations. In this article point symmetry, i.e. the symmetry of isolated figures of finite dimension, is considered.

We introduce the notion of a movement in a metric space X. A bijective map f of a the space into itself conserving the distance between points is called a movement:

$$\forall\, a, b \in X \quad \rho(a,b) = \rho(f(a), f(b)),$$

We call the superposition of two movements, i.e. their sequential application to the given space, the product of the two movements:

$$fg(x) = g(f(x)).$$

Since the product of movements is also a movement, it is natural to put the question concerning the algebraic structure of the movements set Γ. Three important properties take place:

1) Associativity of multiplication:

$$\forall\, a, b, c \in \Gamma \quad a(bc) = (ab)c.$$

2) Existence of an element neutral with respect to multiplication, i.e. of an "identity" movement leaving every point of the space fixed:

$$\exists\, e \in \Gamma : \quad \forall\, a \in \Gamma \quad ae = ea = a.$$

3) Existence of an inverse element:

$$\forall\, a \in \Gamma \quad \exists\, a^{-1} \in \Gamma : \quad aa^{-1} = a^{-1}a = e.$$

These three properties and the completeness with respect to multiplication define on the set Γ a group structure.

A subset H of a group is a subgroup if it is complite with respect to multiplication and satisfies the group axioms:

$$H \subset \Gamma : \quad \forall\, a, b \in H \quad ab, a^{-1}, e \in H.$$

Under these conditions there is no need to verify multiplication associativity because this local property is automatically inherited by H from the group Γ.

We introduce the notion of a generative collection. A collection $\{a_i\} \in H$ generates the group H if every element of the group can be represented as the product of a finite number of powers of the elements of collection:

$$\forall x \in H \quad \exists \{s_i\} \subset \mathbf{N} \cup \{0\} \quad x = \prod_i a_i^{s_i}$$

where almost all the elements s_i must be zero ("almost all" means all but a finite number).

It is possible to extract from the movements group Γ the group of symmetry operation generated by a collection of symmetry elements. The symmetry elements for the bounded figures are rotation and mirror axes. The first will be denoted by $a_1, a_2, \ldots, a_\infty$, the latter by $b_1, b_2, \ldots, b_\infty$ where the indices show the axes' orders.

The presence of a rotation axis of order n in a figure means that turned by angle $2\pi/n$ it coincides with itself. That implies $a_n^n = e$.

The axis of the first order corresponds to the identity operation, the axis of the infinite order corresponds to the operation that "spreads" a point over the circle lying in the plane orthogonal to the axis.

The presence of an inversion (or mirror) axis shows that the figure coincides with itself after turning by the corresponding angle and then inverting, i.e. reflecting of the space in the plane perpendicular to the axis and as a rule passing through the centroid of figure. That implies that when n is even, $b_n^n = e$, when n is odd, $b_n^{2n} = e$.

The mirror axis of the first order corresponds to the simple reflection of the space in the plane perpendicular to the axis, the mirror axis of the second order corresponds to the operation of symmetry center.

The symmetry elements do not form any remarkable algebraic structure, however, selecting some set of them we can generate different subgroups of the symmetry operations group (or simply symmetry group) S. It is necessary to notice that this collection is not minimum, i.e. selecting different sets of symmetry elements as generative collections we can obtain the same group (e.g. $\langle b_3 \rangle = \langle a_3, b_1 \rangle$). The natural reaction would be excluding some redundant elements out of this collection. But firstly, it is not so clear for what purpose (perhaps just for beauty), secondly, nobody knows which elements are redundant, thirdly, that would result only in complication of the further reasoning.

Thus we can attach to every figure some symmetry group (we may speak of the group for if a figure has two symmetry operations, then its has also their product) that is the set of all the symmetry movement after whose action the figure coincides with itself. It is also possible to select at least one set of elements generating this group. Suppose Φ is a figure, i.e. a set of points in X. Then S_Φ is its symmetry group if

$$\forall f \in S_\Phi \quad f(\Phi) = \Phi,$$

Quite different figures may have the same symmetry group (e.g. all the symmetry operations for the n-angular pyramid and the n-angular prism coincide).

If the identity operation is a only symmetric transformation of a figure, then the figure is called asymmetric. Its symmetry group is trivial: $\{e\}$.

To define the symmetry operations for the combination of a set of bodies the following principle is used. According to it the combination inherits the symmetry elements common for all the bodies of the set. In other words, we just take the intersection of the symmetry groups of all the bodies. It is easy to show that the intersection of subgroups of one group is also its subgroup. To do it one just has to verify the group axioms.

We demonstrate this principle on the example of various mutual situation of a sphere and a cylinder. First we find the generative collection for the both bodies separately. The cylinder has: all kinds of simple and inversion axes parallel to its own axis and passing through its center; the elements a_2 and b_2 passing through its center in all the directions orthogonally to its own axis; all the reflection planes (b_1) containing its own axis; the reflection plane perpendicular to its own axis and passing through its center.

For the sphere the generative collection contains all sorts of axes passing through its center without any restriction for the direction (including the symmetry planes defined, as it was said above, by the inversion axes of the first order symmetry).

Consider three variants of the mutual situation of these two bodies (see Fig. 1 *a, b, c*). In Case *(a)*, when the centers of the cylinder and the sphere coincide, all the symmetry elements are common for the both bodies, so the symmetry group of their combination coincides with the symmetry group of cylinder. In Case *(b)*, when the cylinder's axis passes through the center of the sphere, the combination inherits only the cylinder's simple axis and the reflection planes containing the its own axis. In the most general case [Case *(c)*] only one reflection plane containing the cylinder's axis and the sphere's axis remains.

We proceed to consideration of the polyhedrons' symmetry groups. Is there any restrictions for them? Certainly, yes. E.g. there is no way for a polyhedron to have a symmetry group of a sphere. Obviously all the symmetry axes of a polyhedron have to pass through one point that is its center. We want to learn how to verify possibility of the construction of a polyhedron with a given symmetry group.

Suppose we are given such a group and a symmetry elements collection generating it. Take any plane, a face of the future polyhedron we want to construct, and put it "near" some center through which all the given symmetry elements pass. Further we apply the symmetry elements to it. The plane reflects and rotates in every way given by the generative collection. As a result we obtain several more planes which start also reflecting and rotating. And so on, and so forth. Does such a process ever stop? Perhaps, yes. Then we get a finite number of planes coinciding with each other under the action of the symmetry elements. Do they bound a polyhedron? Possibly, yes. Then we are lucky. Really we can get a pair of parallel planes, a dihedron angle or a cylindrical or conic surface over a regular polygon. If the process does not stop, we obtain an infinite number of planes bounding the rotation body of a straight line, a broken line or a circle.

E.g. we take the unidirectional simple symmetry axes of the fifth and third orders. As a result of their application the plane turns by all the angles $2\pi(m/3 +$

$n/5$), where $m, n \in \mathbf{Z}$, i.e. by all the fifteenth parts. So it bounds a conic surface over a regular 15-angle or, if we take a plane orthogonal to the axes, a half-space whose boundary coincides with itself at every turn.

The described above process results in dividing the space into several connected components. We select from them the one containing the center, i.e. the intersection set of all the symmetry elements (a point or a line), and call it the figure generated by the given symmetry group. Thus we associate with every pair *symmetry group—plane* a set of points of the space X.

It is possible to describe all the symmetry groups defining polyhedra. That ultimately reduces to the item-by-item examination of the all possible variants of the symmetry elements combination satisfying the group axioms.

Among the symmetry groups of the bounded figures the crystallographic symmetric groups, i.e. the symmetry groups of crystals, stand separate. The crystal symmetry elements of polyhedra are only simple and inversion axes of the second, third, fourth and sixth orders in one or another combination. The rest axes is forbidden for them as the presence of such axes in a crystal is incompatible with the notion of crystal lattice.

The general number of the crystallographic groups are 32. They generate 44 polyhedra (depending on selection of the initial face for the given symmetry group, one or another polyhedron can be formed). We must notice that here the definition of a polyhedron as a bounded figure suffers because this number contains also 7 prisms and 7 pyramids not bounded "from above" and "from below".

All the crystallographic groups are divided into 7 syngonies. With each of them its own symmetry group whose subgroups are all the groups entering the syngony is associated.

We must notice that the first two syngonies (monocline and tricline) do not generate their own groups because of their poverty.

A computer program demonstrating all the existing crystallographic polyhedra is worked out. The input data is the description of the generative collections of all the syngonies, the description of all their subgroups able to generate polyhedra and the initial face for every crystallographic polyhedra. Applying to this face all the symmetry operations of the given generative collection the program gets the set of faces from the given polyhedron, so far only as the normals to the planes. The point is that any polyhedron generated by a symmetry group is described since all its faces are formed as a result of rotating one of them (the initial face) and hence their distances to the center are the same. That is why the normals to them are sufficient for their univalent definition. Further after time of order n^4 (where n is the number of faces) the program finds all its summits and edges according to the following algorithm.

The item-by-item examination of all triples of planes takes place. For each of the triples the point of intersection of the planes is found (if it exists) and verification of its belonging to the polyhedron is done (more simply, the point must lie to the same side of every plane as the center of the polyhedron or belong to the plane). All the "good" points are selected and those of them that have two common planes are connected with edges. Further definition of the visible and invisible faces is done depending on the directions of their normals.

The program outputs the image on the screen allowing to rotate it in different directions for better visualization. One of the menu options gives a possibility to look at all the symmetry elements for the given polyhedron. The program allows to alter the normal to the initial face and to see how the polyhedron changes because of that. It also allows to regulate the rotation speed and the size of polyhedron. The images of several crystallographic polyhedra formed by the program is given below.

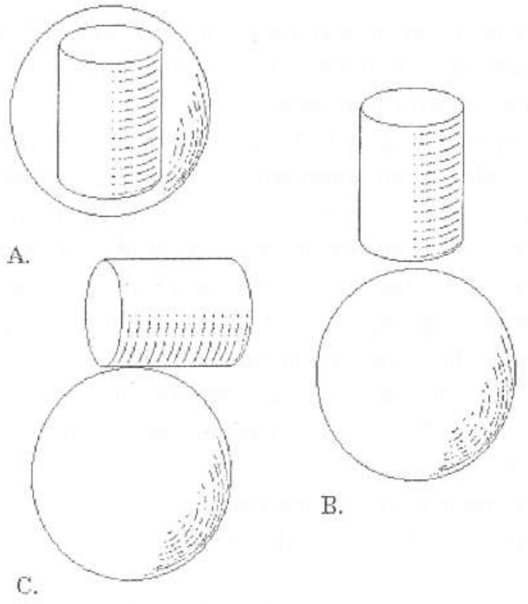
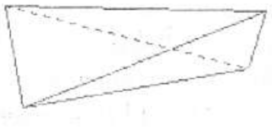
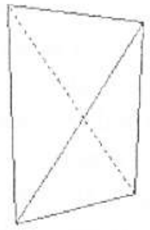


Fig. 1.

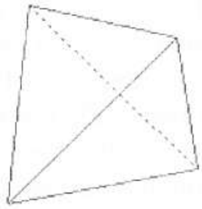
Some crystallographic polyhedrons



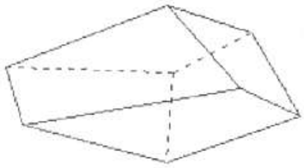
1. Rhombic tetrahedron



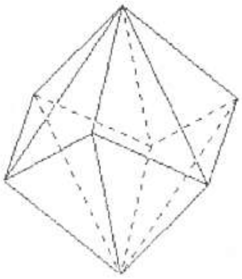
2. Tetragonal discphenoid



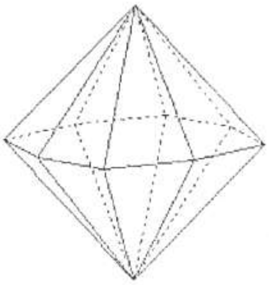
3. Cubic tetrahedron



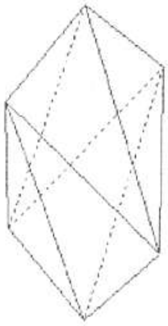
4. Trigonal trapezohedron



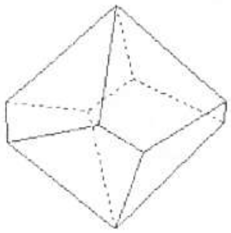
5. Hexagonal scalenohedron



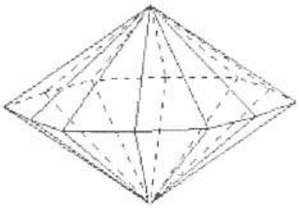
6. Ditetragonal dipyramid



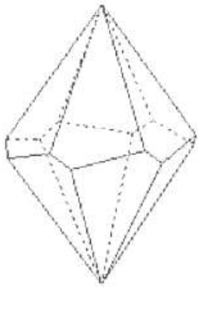
7. Tetragonal scalenohedron



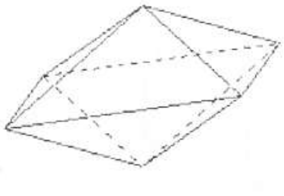
8. Tetragonal trapezohedron



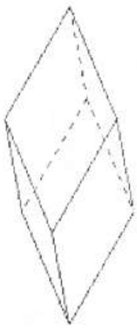
9. Dihexagonal dipyramid



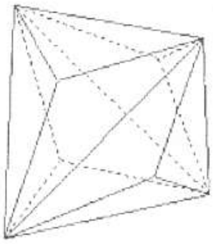
10. Hexagonal trapezohedron



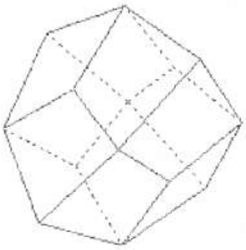
11. Tetragonal dipyramid



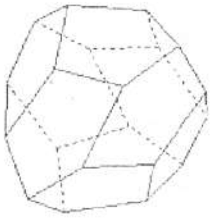
12. Rhombic prism



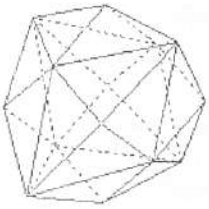
13. Trigon-tristetrahedron



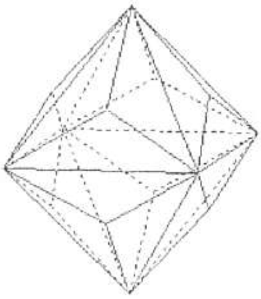
14. Tetragon-tristetrahedron



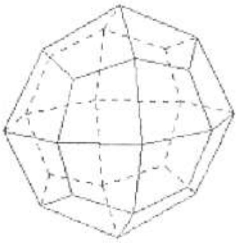
15. Pentagon-tristetrahedron



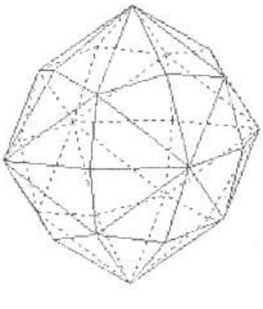
16. Hexatetrahedron



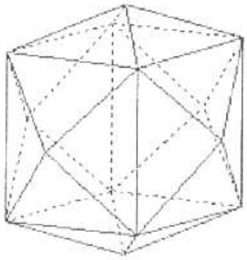
17. Trigon-tris-octahedron



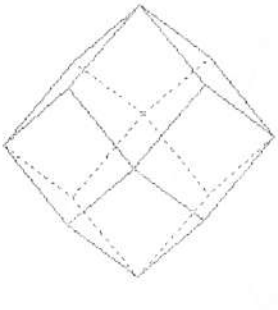
18. Tetragon-tris-octahedron



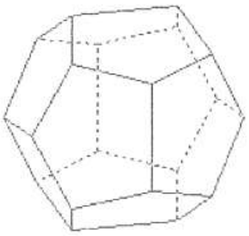
19. Hexaoctahedron



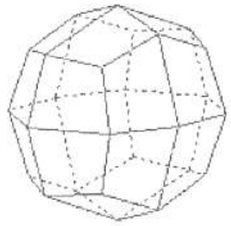
20. Tetrahexahedron



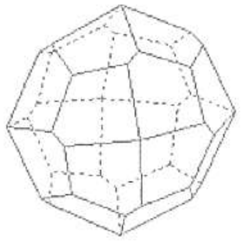
21. Rhomb-dodecahedron



22. Pentagon-dodecahedron



23. Deltoid



24. Pentagon-trioctahedron

Pauline Viewkova.
Graduated from the physical and mathematical school No. 470 in 1994. Studied at Mathematical and Mechanical Dept. of St.Petersburg University since 1994 to 1995. Student of the Dept. of Computer Technology since 1996.

Под общей редакцией Васильева В.Н.
Редакция авторская
Компьютерный набор, верстка и дизайн авторские
Центр распределенных издательских систем ИТМО
Лицензия ЛР №020945 от 29.11.94.
Подписано к печати 06.10.97.
Тираж 200 экз.