**Zeitschrift:** L'Enseignement Mathématique

**Herausgeber:** Commission Internationale de l'Enseignement Mathématique

**Band:** 23 (1977)

**Heft:** 1-2: L'ENSEIGNEMENT MATHÉMATIQUE

**Artikel:** ON REPRESENTATION OF FUNCTIONS BY MEANS OF SUPERPOSITIONS AND RELATED TOPICS

**Autor:** Vitushkin, A. G.

**DOI:** https://doi.org/10.5169/seals-48931

# ON REPRESENTATION OF FUNCTIONS BY MEANS OF SUPERPOSITIONS AND RELATED TOPICS [1]

by A. G. VITUSHKIN

## CONTENTS



---

[1]) Summary of lectures given at the University of California in Los Angeles, in April-May 1977, under the sponsorship of the International Mathematical Union.

## PREFACE

By means of an algebraic substitution, the so-called Tschirnhaus transformation, the general algebraic equation of the $n$-th degree $x^n + a_1 x^{n-1} + a_2 x^{n-2} + \dots + a_{n-1} x + a_n = 0$ may be reduced to the form $y^n + b_4 y^{n-4} + b_5 y^{n-5} + \dots + b_{n-1} y + 1 = 0$. Further attempts by algebraists to reduce the solution of the general algebraic equation to the solution of equations containing a smaller number of parameters remained unsuccessful for a long time (the problem of resolvents).

In his famous Mathematical Problems [1] Hilbert looked at this problem in a new way, formulating it as No. 13 in the following form: the impossibility of solving the general equation of the 7-th degree by means of functions of only two variables. To prove this Hilbert regarded it as possible to show that the equation of the 7-th degree $f^7 + xf^3 + yf^2 + zf + 1 = 0$ is not soluble by means of any continuous functions of only two variables.

Various mathematicians have understood the 13-th Problem differently and have attributed to it results of a different character.

Hilbert [3] found an algebraic substitution reducing the solution of the general algebraic equation of the 9-th degree to the solution of equations with 4 parameters. Hilbert proved also the existence of analytic functions of three variables not representable by superpositions of functions of only two variables. Ostrowski [2] constructed an analytic function of two variables not representable as a superposition of infinitely differentiable functions of one variable and arithmetic operations. The author [4] proved the existence of smooth functions of several variables not representable by superpositions of smooth functions of a smaller number of variables.

Bieberbach [5] attempted to prove that there exist continuous functions of three variables, not representable as a superposition of continuous functions of two variables. Not for nothing did Bieberbach call the 13-th Problem "unfortunate" (see [6]). Many years later, by the combined efforts of Kolmogorov [7], [9] and Arnol'd [8], the opposite was proved. So Hilbert's conjecture was shown to be false. By Kolmogorov's theorem any continuous function of several variables can be represented by means of a superposition of continuous functions of a single variable and the operation of addition.

Hilbert's 13-th problem gave rise to a great number of investigations in algebra and analysis, but the kernel of the problem never the less remains untouched. In this connection Lorentz [12] made an expressive analogy. The example of Peano of a mapping of an interval onto a square does not answer the question about the difference between an interval and a square. In the same way the theorem of Kolmogorov does not close the 13-th problem, but only makes it more interesting. It is known, for example, that superpositions of Kolmogorov's type, composed of smooth functions, do not even represent all analytic functions [48].

Thus, Hilbert's idea of proving the impossibility of solving the general equation of the 7-th degree by means of functions of only two variables can be developed in a more positive way. Results available at present do not contradict, for example, the possibility that the function $f(x, y, z)$ defined by the equation $f^7 + xf^3 + yf^2 + zf + 1 \equiv 0$ is not a finite superposition of analytic functions of two variables. On the other hand nobody has disproved that any algebraic function is a superposition of algebraic functions of a single variable and arithmetic operations.

This paper is a summary of the lectures given at the University of California in Los Angeles in April-May of 1977. Chapter I presents a survey of results, the remaining chapters are devoted to proofs.

## Chapter 1. — Survey of results

The survey presented is based on the surveys [10]-[12], [33]-[35]. It also covers recent results:

*Definition.* We will say that a function $f = f(x_1, ..., x_n)$ is a superposition of the functions

$$\varphi^{(\alpha)}_{\beta_1,\beta_2,\ldots,\beta_\alpha}\left(U^{(\alpha+1)}_{\beta_1,\beta_2,\ldots,\beta_\alpha,1}, U^{(\alpha+1)}_{\beta_1,\beta_2,\ldots,\beta_\alpha,2}, \ldots, U^{(\alpha+1)}_{\beta_1,\beta_2,\ldots,\beta_\alpha,k}\right)$$

$$(\beta_i = 1, 2, \ldots, k, \ i = 1, \ldots, \alpha, \ \alpha = 0, 1, \ldots, s-1)$$

of $k$ variables if $f$ identically equals the function $\varphi$, defined by the equalities

$$\varphi = \varphi^{(0)}(U_1{}^{(1)}, U_2{}^{(1)}, \ldots, U_k{}^{(1)}),$$

$$U^{(\alpha)}_{\beta_1,\ldots,\beta_\alpha} = \varphi^{(\alpha)}_{\beta_1,\ldots,\beta_\alpha}\left(U^{(\alpha+1)}_{\beta_1,\beta_2,\ldots,\beta_\alpha,1}, U^{(\alpha+1)}_{\beta_1,\beta_2,\ldots,\beta_\alpha,2}, \ldots, U^{(\alpha+1)}_{\beta_1,\beta_2,\ldots,\beta_\alpha,k}\right)$$

$$\beta_i = 1, 2, \ldots, k, \ i = 1, 2, \ldots, \alpha, \ \alpha = 1, 2, \ldots, s-1,$$

$$U^{(s)}_{\beta_1,\beta_2,\ldots,\beta_s} = x_{j(\beta_1,\beta_2,\ldots,\beta_s)}.$$

The number $s$ is called the order of superposition.

## § 1. *Superpositions of analytic functions*

In stating the 13-th Problem [1] Hilbert added that he had a rigorous proof of the fact that there exists an analytic function of three variables that cannot be obtained by a finite superposition of functions of only two arguments. Although he did not indicate exactly what kind of functions of two variables he had in mind, Hilbert was apparently thinking of analytic functions of two variables.

The existence of analytic functions of three variables not representable by means of superpositions of analytic functions of two variables is a simple fact and can be obtained from the following considerations. The partial derivatives of order $k$ of a function represented by a superposition are defined by the derivatives of the functions composing the superposition. The number of different partial derivatives of order $p$ of a function of two variables is equal to $\frac{p(p-1)}{1 \cdot 2}$. Consequently, the number of parameters defining the derivatives of order $k$ of the superposition has order $k^3$ ($s$ is fixed). On the other hand the number of different partial derivatives of order not greater than $k$ for a function of three variables is of the order $k^4$. Hence for any $s$ there exists a sufficiently large $k$ such that one can find a polynomial of the $k$-th degree not representable by a superposition of order $s$ of infinitely differentiable functions of two variables. The desired non-representable analytic function can be given as a sum of non-representable polynomials.

More general results in this direction were obtained by Ostrowski [2], who showed, in particular, that the analytic function of two arguments

$\xi(x, y) = \sum_{n=1}^{\infty} \frac{x^n}{n^y}$ is not a finite superposition of infinitely differentiable functions of one variable and algebraic functions of any number of variables.

The proof of this result is based on the fact that the function $\xi(x, y)$ does not satisfy any algebraic partial differential equation, that is, an equation of the form

$$\Phi\left(\xi, \frac{\partial \xi}{\partial x}, \frac{\partial \xi}{\partial y}, \ldots, \frac{\partial^{\mu+\lambda}\xi(x, y)}{\partial x^{\mu}\partial y^{\lambda}}\right) = 0, \quad \text{where} \quad \Phi$$

is a polynomial with constant coefficients in the function $\xi$ and its partial derivatives up to a certain order. At the same, it is comparatively simple to prove that any function of two variables which is a finite superposition of infinitely differentiable functions of one variable and algebraic functions of any number of variables necessarily satisfies some algebraic partial differential equation. In the same paper, Ostrowski conjectured that the function $\xi(x, y)$ is not a superposition of continuous functions of one variable and algebraic functions of any number of variables (see the theorem of Kolmogorov [9]).

## § 2. *The problem of resolvents*

Algebraic equations up to the 4-th degree inclusive are soluble by radicals, that is, the roots of these equations can be represented as functions of the coefficients in the form of a superposition of arithmetic operations and functions of one variable of the form $\sqrt[n]{t}$ $(n = 2, 3)$. The general equation of the 5-th degree, is insoluble by radicals, as Abel and Galois showed. But since the general equation of the 5-th degree may be reduced by algebraic substitutions to the form $x^5 + tx + 1 = 0$, containing a single parameter $t$, we may say that a root of the general equation of the 5-th degree is also represented as a function of the coefficients in the form of superpositions of arithmetic operations and algebraic functions of one variable. The problem of resolvents can be formulated in terms of superpositions in the following way: to find, for any number $n$, the smallest number $k$ such that a root of the general equation of the $n$-th degree as a function of the coefficients is represented in the form of a superposition of algebraic functions of $k$ variables. In [3] Hilbert conjectured that for $n = 6, 7, 8$ the number $k$ is 2, 3, 4, respectively. Hilbert's result [3] for an equation of the 9-th degree was all the more unexpected: a root of the general equation of the 9-th degree is representable as a superposition of algebraic functions of four variables. Wiman [13], generalizing Hilbert's result, proved that $k \leqslant n - 5$ for any $n \geqslant 9$. As G. N. Chebotarev [14] observed, it can be proved by the same method that $k \leqslant n - 6$ for $n \geqslant 21$ and $k \leqslant n - 7$ for $n \geqslant 121$. A number of papers by N. G. Chebotarev [15] was devoted to the problem of resolvents. However, the basic Theorem turned out to be wrong (see [16]).

In correcting Chebotarev's paper Morosov found the right statements but his proofs also were not without essential gaps [17]. Nevertheless, in spite of the mistakes the papers of Chebotarev and Morosov have had a positive influence on subsequent authors.

Arnol'd [18] and Lin [17] have shown that the function $f_n = f(z_1, \ldots, z_n)$ which is the solution of the algebraic equation $f^n + z_1 f^{n-1} + z_2 f^{n-2} + \ldots + z_n = 0$ for $n \geqslant 3$ can not be strictly represented as a superposition of entire algebraic functions of a smaller number of variables and polynomials of any number of variables. Let us recall that a function $f = f(z_1, \ldots, z_k)$ is called an entire algebraic function if it satisfies an equation $f^m + p_1 f^{m-1} + \ldots + p_m = 0$, where $p_1, \ldots, p_m$ are polynomials in $z_1, \ldots, z_k$. The sentence "a function can not be strictly represented as a superposition" means in the case under consideration that every superposition representing the function must have unnecessary branches, i.e. the number of branches of any superposition must be at least $n + 1$. Using that theorem for $n = \{3, 4\}$ we see that in spite of the fact that the equations of degree 3 and 4 are soluble by radicals they do not have strict representations. This explains in a sense why unnecessary roots appear when one uses Cardano's formulas.

Hovanski (see [19] and [20]) has shown that the solution of the equation $f^5 + xf^2 + yf + 1 = 0$ can not be represented by a superposition of entire algebraic functions of a single variable and polynomials in several variables. We recall that the Tschirnhaus transformation reduces the general equation of the 5-th degree to an equation with a single parameter, that is, the function of Hovanski is represented by a superposition of algebraic functions of a single variable and arithmetic operations. This counter example demonstrates that the restriction not to use the operation of division, is really strong.

We conclude the discussion of the problem of resolvents with a formulation of a well-known problem: is it possible to represent any algebraic function by means of a superposition of functions of a single variable and rational functions of any number of variables.

## § 3. *Superpositions of smooth functions and the theory of approximation*

In [4] it was proved that in the class of all $S$ times continuously differentiable functions of $n$ variables there exist some that cannot be represented as a finite superposition of functions for which the ratio of the number of arguments to the number of derivatives they have is strictly less than $n/S$.

This theorem shows that the ratio $n/S$ can serve as a measure of the complexity of $S$ times differentiable functions of $n$ variables. The original proof of this theorem made use of the theory of multi-dimensional variations of sets and estimates of the number of $\varepsilon$-distant smooth functions (see [21], [22]). Kolmogorov [23] showed that the same result can be obtained using only estimates of the number of elements of $\varepsilon$-nets of functional compacts.

We denote by $F_S^n$ the set of functions $f(x_1, x_2, \dots, x_n)$ defined on an $n$-dimensional cube, whose partial derivatives up to order $S$ inclusive are all continuous and bounded by some constant $C$. Let $N_\varepsilon(F_S^n)$ be the minimum number of spheres of radius $\varepsilon$ in the space of all continuous functions by which the set $F_S^n$ can be covered.

It turns out that

$$\lim_{\varepsilon \to 0} \frac{\log \log N_\varepsilon (F_s^n)}{\log \left(\frac{1}{\varepsilon}\right)} = \frac{n}{s}.$$

Hence it follows that if $n/S > n'/S'$, then the set of functions $F_S^n$ is, in a certain sense, "more massive" than $F_{S'}^{n'}$.

If a consideration of the massivity of functional compacts does not give the answer then the problems remain open. For example, there is no answer to the question: is it possible to represent any analytic function of several variables by means of a superposition of smooth functions of a smaller number of variables.

The topic of superpositions led to a large number of papers in approximation theory. Here we formulate two results concerning non-linear approximations.

Let $\mathcal{I}^n$ be a cube $0 \leqslant x_i \leqslant 1$ $(i = 1, \dots, n)$; $C$—the space of all realvalued continuous functions defined on $\mathcal{I}^n$ with the uniform norm; $F$—a compact subset of $C$, $\Phi$—a surface in $C$ which consists of the functions represented in the form

$$\varphi = \frac{\sum\limits_{\alpha_1+\alpha_2+\ldots+\alpha_p \leq k} a_1^{\alpha_1} \cdot a_2^{\alpha_2} \ldots a_p^{\alpha_p} f_{\alpha_1,\ldots,\alpha_p}}{\sum\limits_{\beta_1+\beta_2+\ldots+\beta_p \leq k} b_1^{\beta_1} \cdot b_2^{\beta_2} \ldots b_p^{\beta_p} g_{\beta_1,\ldots,\beta_p}} \cdot$$

where the natural numbers $p$ and $k$ and the collections $\{f_{\alpha_1,\ldots,\alpha_p} \in C\}$ and $\{g_{\beta_1,\ldots,\beta_p} \in C\}$ are fixed in advance and independent of $\varphi$, $\{a_i\}$ and $\{\beta_i\}$ are positive integers and the coefficients $\{a_i\}$ and $\{b_i\}$ defining the function $\varphi$ can take arbitrary real values.

We remark that for $k = 1$ the class $\Phi$ can be turned into any of the usual classes in approximation theory by means of an appropriate choise of the number $p$ and collections $\{f_{\alpha_1,\ldots,\alpha_p}\}$ and $\{g_{\beta_1,\ldots,\beta_p}\}$. For example it can be turned into the classes of polynomials or rational functions of a fixed degree.

We put $e_{pk}(F) = \sup\limits_{f \in F} \inf\limits_{\varphi \in \Phi} \|f - \varphi\|$. Estimates of $e_{pk}$ for some functional compacts can be found in [21], [22], [24], [25]. Here are two examples of such estimates

$$1. \quad e_{pk}(F_s^n) \geqslant a\left(\frac{1}{p \log (k+1)}\right)^{s/n},$$

where $a > 0$ does not depend on $p$ and $k$.

2. For the set $F_{dc}$ consisting of all functions which have an analytic extension to some domain $d$ in $n$-dimensional complex space bounded in modulus by some constant $C$ the following inequality is valid

$$e_{pk}(F_{dc}) \geqslant bq^{n\sqrt{p \log (k+1)}},$$

where $b > 0$ and $0 < q < 1$ are constants independent of $p$ and $k$.

Now there are more elementary proofs of these inequalities for $k = 1$ with precise estimates of the constant (see Erohin [26], Lorentz [24], Tihomirov [27], Shapiro [25]).

Let us clarify the meaning of these inequalities. We agree to characterize the complexity of any algorithm for the approximate calculation of functions firstly by the number of parameters used in the algorithm, and secondly by the complexity of the scheme of the calculation, for example, by the number of arithmetic operations required for the approximate calculation of functions by means of the given algorithm.

In the above-mentioned method of approximation of functions by functions from $\Phi$ the parameters are the numbers $\{a_i\}$ and $\{b_i\}$, and the number of arithmetic operations increases very rapidly as $k$ increases. At the same time, from the inequalities mentioned above it follows that an increase in $k$ leads to an insignificant improvement in the accuracy of the approximation. Hence, in a certain sense, a more economical approximation of functions is by means of expressions of the given form with $k = 1$, that is, by fractions of the form

$$\frac{\sum_{i=0}^{p} a_i f_i(x)}{\sum_{j=0}^{p} b_j g_j(x)}$$

The same inequalities with $k = 1$ show that there are no methods of approximating functions by fractions of the given form essentially better than the standard methods of approximating functions by algebraic (or trigonometric) polynomials.

## § 4. *Superpositions of continuous functions*

Kolmogorov's theorem on the possibility of representing continuous functions of $n$ variables as superpositions of continuous functions of three variables was highly unexpected (see [7]).

In this paper Kolmogorov proves that on the $n$-dimensional cube $\mathscr{I}^n$ we can construct continuous functions $\varphi_i(x)$ $(i = 1, 2, \dots, n+1)$ such that any continuous function $f(x)$, defined on the cube $\mathscr{I}^n$, can be represented in the form

$$f(x) = \sum_{i=1}^{n+1} f_i(d_i(x)),$$

where $d_i(x)$ is a continuous mapping of $\mathscr{I}^n$ onto the one-dimensional tree [1]) $D$ if the components of the level sets of the functions $\varphi_i(x)$, and $f_i(d_i)$ is a continuous function on the tree $D_i$. Since the trees $\{D_i\}$ can be embedded homeomorphically in the plane (see [30]), the functions $\{f_i(d_i(x))\}$ can be thought of as superpositions

$$\{f_i(u_i(x_1, x_2, \dots, x_n),\ v_i(x_1, x_2, \dots, x_n))\}$$

[1]) Kronrod [29] has shown that the components of all possible level sets of any continuous function defined on $\mathscr{I}^n$ in a certain natural topology, form a tree, that is, a one-dimensional locally connected continuum, not containing homeomorphic images of circles. Kronrod calls this the "one-dimensional tree of the function".

where $\{f_i(u_i, v_i)\}$ are continuous functions of two variables, and $\{u_i(x)\}$ and $\{v_i(x)\}$ are fixed continuous functions of $n$ variables. Kolmogorov derived from this the result that for $n \geqslant 4$ any continuous function of $n$ variables can be represented by the following superposition of continuous functions of not more than $n - 1$ variables:

$$\sum_{i=1}^{n} f_i\big(u_i(x_1, x_2, \ldots, x_{n-1}),\ v_i(x_1, x_2, \ldots, x_{n-1}),\ x_n\big).$$

Arnol'd [8], [22] showed that, firstly, in Kolmogorov's construction [7] we can manage with functions $\{\varphi_i(x)\}$ whose one-dimensional trees $\{D_i\}$ have index at each branch point equal to 3, and, secondly, for any compact set $F$ of functions defined on such a tree $D$, the given tree can be so placed in three-dimensional $u$, $v$, $w$-space that any continuous function $f(d) = f(u, v, w) \in F$ can be represented as the sum of functions of the coordinates, $f(u, v, w) = \varphi(u) + \psi(v) + \kappa(w)$. Hence it follows that any continuous function $f(x, y, z)$ of three variables can be represented as a superposition of the form $f(x, y, z) = \sum_{i=1}^{9} f_i(\varphi_i(x, y), z)$, where all the functions are continuous, and the functions $\{\varphi_i(x, y)\}$ can be regarded as fixed, when $f(x, y, z)$ is taken from a compact set. Thus, Arnol'd had the last word in refuting Hilbert's conjecture. At the same time Kolmogorov [9] obtained, in a certain sense, the definitive result in this direction.

Each continuous function of $n$ variables, given on the unit cube in $n$-dimensional space, is representable as a superposition of the form

$$f(x_1, x_2, \ldots, x_n) = \sum_{q=1}^{2n+1} g_q\Big(\sum_{p=1}^{n} \varphi_{p,q}(x_p)\Big)\,, \tag{I}$$

where all the functions are continuous, and moreover the functions $\{\varphi_{p,q}(x_p)\}$ are standard and monotonic.

In particular, each continuous function of two variables is representable in the form

$$f(x, y) = \sum_{i=1}^{5} f_i(\alpha_i(x) + \beta_i(y))\,. \tag{II}$$

Kolmogorov's theorem can be supplemented by the following result of Bari, which was obtained in connection with problems of Fourier series: any continuous function of one variable $f(t)$ can be represented in the form $f(t) = f_1(\varphi_1(t)) + f_2(\varphi_2(t)) + f_3(\varphi_3(t))$, where all the functions $\{f_i\}$ and $\{\varphi_i\}$ are absolutely continuous [32].

From the theorems of Kolmogorov and Bari it follows that each continuous function of $n$ variables can be represented as a superposition of absolutely continuous monotonic functions of one variable and the operation of addition.

A detailed account of Kolmogorov's theorem is to be found in the surveys [9], [33]-[36]. The proof presented by Kahane is of special interest [36]. He does not attempt to construct the functions $\{ \varphi_{p,q} \}$ (as the proof of Kolmogorov does) but instead he shows by means of Baire's theorem, that most selections of increasing functions $\{ \varphi_{pq} \}$ will do. This approach also lead to other interesting results.

Fridman [37] showed that the inner functions $\{ \varphi_{pq} \}$ can be chosen from the class Lip 1. Kahane noticed that this follows directly from Kolmogorov's theorem. For any finite collection of continuous and monotone functions $\{ f_k(x) \}$ on the segment [0, 1] there exists a homeomorphism $x = \varphi(s)$ of the segment [0, 1] onto itself such that the functions $\{ g_k(s) = f_k(\varphi(s)) \}$ belong to the class Lip 1. The homeomorphism is taken as

$$s = \varphi^{-1}(x) = \varepsilon \big( x + \sum_k |f_k(x) - f_k(0)| \big) .$$

The constant $\varepsilon$ is chosen to satisfy the condition $\varphi^{-1}(1) = 1$. By means of such homeomorphisms all inner functions in Kolmogorov's formula can be turned into functions satisfying the condition Lip 1.

There are some other improvements of Kolmogorov's theorem: Doss [38], Bassalygo [39], Lorentz [34], Sprecher [40] (see chapter 3, § 1). There are also many results concerning special types of superpositions (see [21], [33], [41]-[44]).

## § 5. *Linear superpositions*

We return again to superpositions of smooth functions.

One of the most interesting current problems on the subject of superpositions is the following: does there exist an analytic function of two variables that cannot be represented as a finite superposition of continuously differentiable (smooth) functions of one variable and the operation of addition?

Linear superpositions arise as a result of the following argument. Suppose that a function of two variables $f(x, y)$ is an $s$-fold superposition of certain smooth functions of one variable $\{ f_i(t) \}$ and the operation of addition. We vary this superposition, that is, we consider a superposition

$\tilde{f}(x, y)$ of the same form, but composed of the functions $\{f_i(t) + \varphi_i(t)\}$, where $\{\varphi_i(t)\}$ are small perturbations that are also smooth functions of one variable. Then the difference of these superpositions can be written in the form

$$\tilde{f}(x, y) - f(x, y) = \sum_{i=1}^{N} p_i(x, y)\, \varphi_i(q_i(x, y)) + o\,(\max_i \sup_t |\varphi_i(t)|)\,, \tag{III}$$

where the functions $\{p_i(x, y)\}$ are expressed in terms of the original functions $\{f_i(t)\}$ and their derivatives, so that we can only say of them that they are continuous; $\{q_i(x, y)\}$ are expressed only in terms of the functions $\{f_i(t)\}$, hence they are continuously differentiable; the remainder term $o\,(\max_i \sup_t |\varphi_i(t)|)$ is an infinitely small quantity compared with $\max_i \sup_t |\varphi_i(t)|$, provided only that the functions $\left\{\frac{d\varphi_i}{dt}\right\}$ have some fixed modulus of continuity. Equation (III) gives some hope of reducing the general problem of superpositions of smooth functions to the determination of analytic functions not representable by superpositions of the form

$$\sum_{i=1}^{N} p_i(x, y)\, \varphi_i(q_i(x, y))\,, \tag{IV}$$

where $\{p_i(x, y)\}$ are preassigned continuous functions, $\{q_i(x, y)\}$ are preassigned continuously differentiable functions, and $\{\varphi_i(t)\}$ are arbitrary continuous functions of one variable.

Such superpositions are called linear, to emphasize the fact that the functions $\{p_i(x, y)\}$ and $\{q_i(x, y)\}$ are fixed and the superposition depends linearly on the variable functions $\{\varphi_i(t)\}$. We note that Kolmogorov's superpositions (I), (II) are also linear, since all $p_i \equiv 1$ and $q_i(x, y) = \alpha_i(x) + \beta_i(y)$ $(i = 1, 2, 3, 4, 5)$ are fixed continuous functions.

It is proved in [47], [48] that for any continuous functions $\{p_i(x, y)\}$ and continuously differentiable functions $\{q_i(x, y)\}$ there exists an analytic function of two variables not representable as a superposition of the form (IV). Henkin showed that the set of superpositions of the form (IV) is closed and consequently nowhere dense in the space of all continuous functions of two variables. Hence, in particular, it follows that there exists even a polynomial not representable as a superposition of the form (IV).

A comparison of these results with Kolmogorov's theorem leads to the conclusion that the inner functions of Kolmogorov's formula, although continuous, must inevitably be essentially non-smooth.

We note that the results mentioned above can be extended without any essential difficulties to superpositions of the form

$$\sum_{i=1}^{N} p_i(x_1, \ldots, x_n) f_i(q_i(x_1, \ldots, x_n)),$$

where $\{p_i\}$ are preassigned continuous functions, $\{q_i\}$ are preassigned smooth functions and $\{f_i\}$ are arbitrary continuous functions of one variable. But as it turns out this does not apply to superpositions of the form

$$\sum_{i=1}^{N} p_i(x_1, \ldots, x_n) f_i(q_{1,i}(x_1, \ldots, x_n), \ldots, q_{k,i}(x_1, \ldots, x_n)),$$

where $\{p_i\}$ are fixed continuous functions of $n$ variables; and $\{q_{1i}\}, \ldots, \{q_{ki}\}$ are fixed smooth functions of $n$ variables ($k<n$). Fridman answered that question only for $n = 3, 4$, $k = 2$ and $\{p_i\} \equiv 1$.

Also it is not known to what extent the problem of superpositions of smooth functions can be reduced to that of linear superpositions. "Such a reduction is proved only in the case of the so called stable" superpositions [10]. It turns out that not every analytic function of $n$ variables can be represented by means of superpositions of smooth functions of a smaller number of variables it is assumed that the scheme is stable, i.e. for a small perturbation of a function represented the perturbations of the functions composing the superposition are comparatively small.

## Chapter 2. — Superpositions of smooth functions

In this chapter we prove the existence of smooth functions of $n$ variables ($n \geqslant 2$), not representable by superpositions of smooth functions of a smaller number of variables.

### § 1. *The notion of entropy*

We will denote by $C(\mathscr{I})$ the space of all functions defined on a set $\mathscr{I}$ and continuous on $\mathscr{I}$ (the norm is the maximum of the absolute value of the function). We fix a compact $F \subset C(\mathscr{I})$ and a positive number $\varepsilon$. A set $F^* \subset C(\mathscr{I})$ is called an $\varepsilon$-net of $F$ if for any $f \in F$ there exists $f^* \subset F^*$

such that $\|f - f^*\| \leqslant \varepsilon$. We denote by $N_\varepsilon(F)$ the number of elements of a minimal $\varepsilon$-net of $F$. The number $H_\varepsilon(F) = \log_2 N_\varepsilon(F)$ is called the $\varepsilon$-entropy of the set $F$.

The notion of entropy arises in a natural way in connection with various problems of analysis. We consider an example.

Let $f$ be a function. It is known only that $f$ belongs to a compact $F$. For example a smoothness condition of $f$ and estimates of derivatives are given. We consider the problem of tabulating the function $f$. The first part of the problem is to write down in a table some number (parameters of $f$). For example, the values of $f$ at certain points or the Taylor coefficients of $f$ can be taken as such parameters. The second part of the problem is to present a decoding algorithm universal for all $f \in F$ which allows $f$ to be calculated at any point with the accuracy $\varepsilon$.

The complexity of a table is usually characterized by two factors—its volume (the total number of binary digits required to write down all the parameters of the table) and the complexity of the decoding algorithm. It is easy to see that the volume of the most economical table presenting $f$ with the accuracy $\varepsilon$ equals $H_\varepsilon(F)$. Moreover it is possible to characterize the decoding algorithm too in terms of the entropy [21], [22], [24], [25].

It will be shown in paragraphs 2 and 3 that the number of $\varepsilon$-distant smooth functions depends in an essential way on the number of variables. This enables us to construct smooth functions of $n$ variables not representable by smooth functions of a smaller number of variables.

We present here estimates of the entropy for a few concrete classes.

1. Let $F_S^n$ be the class of all real valued functions, defined on a cube $\mathscr{I}: \{0 \leqslant x_i \leqslant 1, i = 1, \dots, n\}$ whose partial derivatives of order up to $S$ are bounded in modulas by a constant $C$. Then

$$c'\left(\frac{1}{\varepsilon}\right)^{n/s} \leqslant H_\varepsilon(F_s^n) \leqslant c''\left(\frac{1}{\varepsilon}\right)^{n/s},$$

where $C' > 0$, $C'' > 0$ are independent of $\varepsilon$.

2. Let $F^c_{\rho_1,\rho_2,\dots,\rho_n}$ be the space of functions analytic on the $n$-dimensional cube $\{-1 \leqslant x_k \leqslant 1\}$ $(k = 1, 2, \dots, n)$ having analytic continuations in the region $E_\rho = E_{\rho_1} \times E_{\rho_2} \times \dots \times E_{\rho_n}$ which are bounded in modulus in this region by the constant $C > 0$, where $E_{\rho_k}$ is the region of the complex plane $z_k = x_k + iy_k$ bounded by the ellipse with semi-major axis $\rho_k$ and with foci at the points $-1, 1$ of the real axis $(k = 1, 2, \dots, n)$. Then

$$H_\varepsilon(F^c_{\rho_1,\rho_2,\dots,\rho_n}) - \frac{1}{(n+1)!}\prod_{k=1}^{n}\frac{1}{\log\ \rho_k}\left(\log\frac{c}{\varepsilon}\right)^{n+1} + O\left[\left(\log\frac{c}{\varepsilon}\right)^n \log\log\frac{c}{\varepsilon}\right].$$

3. Let $F^n_{s,c}$ be the class of real valued functions on the cube $\{-1 \leqslant x_k \leqslant 1\}$ $(k=1, \dots, n)$, bounded in modulus on that cube by the constant $s_k$ and such that their analytic extensions are entire functions of order $s_k$ with respect to $z_k = x_k + iy_k$ $(k=1, \dots, n)$. Then

$$H_\varepsilon(F^n_{s,c}) = \frac{1}{(n+1)!}\prod_{k=1}^{n} s_k \left(\log\frac{c}{\varepsilon}\right)^{n+1}\left(\log\log\frac{c}{\varepsilon}\right)^{-n} +$$
$$= O\left[\left(\log\frac{c}{\varepsilon}\right)^{n+1}\left(\log\log\frac{c}{\varepsilon}\right)^{-n-1}\right].$$

These estimates and other results connected with estimates of entropy and applications are to be found for example in [49]-[53].

## § 2. *The entropy of the space of smooth functions*

Here we give an estimate of the entropy of the class of $S$ times differentiable functions of $n$ variables. The lower estimate was obtained in [4], the upper one—in [23].

We fix integers $n \geqslant 1$ and $p \geqslant 0$ and numbers $0 \leqslant \alpha \leqslant 1$, $L > 0$, $C > 0$, $\rho > 0$. We will denote by $\mathscr{I}$ the cube $0 \leqslant x_i \leqslant \rho$ $(i=1, \dots, n)$ and by $F = F^{\rho,n}_{S,L,c}$ $(S = p+\alpha)$ the set of all real valued functions defined on $\mathscr{I}$ such that their partial derivatives of order $p$ satisfy the condition Lip $\alpha$ with the constant $L$ and

$$\left|\frac{\partial^{k_1+\dots+k_n} f(0)}{\partial^{k_1}x_1 \dots \partial^{k_n}x_n}\right| \leqslant c \quad \left(\sum_{i=1}^{n} k_i \leqslant p\right)$$

We say that the function $g(x)$ satisfies the condition Lip $\alpha$ with the constant $L$ if for any $x'$ and $x''$

$$|g(x') - g(x'')| \leqslant L(r(x', x''))^\alpha,$$

where $r(x', x'')$ is the distance between $x'$ and $x''$.

THEOREM 2.2.1. *If* $\varepsilon > 0$ *is sufficiently small then*

$$A\rho^n\left(\frac{L}{\varepsilon}\right)^{n/s} \leqslant H_\varepsilon(F) \leqslant B\rho^n\left(\frac{L}{\varepsilon}\right)^{n/s},$$

*where* $A$ *and* $B$ *are positive constants depending only on* $s$ *and* $n$.

We choose $\delta > 0$ such that the number $\rho/\delta$ is an integer. We divide the cube $\mathscr{I}$ into $\left(\frac{\rho}{\delta}\right)^n$ cubes $P_i$ $\left(i = 1, 2, \ldots, \left(\frac{\rho}{\delta}\right)^n\right)$ by hyperplanes, parallel to its $(n-1)$-dimensional edges. Each of the cubes $P_i$ has side of length $\delta$, and the edges of these cubes are parallel to those on $\mathscr{I}$. Let $C_i$ denote the centre of the cube $P_i$ and $S_i$ the $n$-dimensional closed sphere (inscribed in $P_i$) of radius $\delta/2$ and centre at the point $C_i$. Put

$$\varphi_i(x) = \varphi_i(x_1, x_2, \ldots, x_n) = \begin{cases} 0, & \text{if } x \in \mathscr{I} - S_i \\ A\left(1 + \cos\left(\frac{2\pi}{\delta} r\,(C_i, x)\right)\right)^p & \text{if } x \in S_i, \end{cases}$$

where $r\,(C_i, x)$ is the distance from the point $x$ to the centre $C_i$ of the sphere $S_i$. Put, further,

$$\varphi_{\eta_1, \eta_2, \ldots, \eta_h}(x) = \sum_{i=1}^{h} \eta_i \varphi_i(x)$$

$$\left(\eta_i = \pm 1;\ i = 1, 2, \ldots, h;\ h = \left(\frac{\rho}{\delta}\right)^n\right).$$

LEMMA 2.2.1. *We can find a positive number* $A\,(s, L, n)$, *such that when* $A = A\,(s, L, n)\,\delta^s$ *and given any set of numbers* $\eta_i\,(i = 1, 2, \ldots, h)$*-the corresponding function* $\varphi_{\eta_1, \eta_2, \ldots, \eta_h}(x)$ *belongs to* $F$.

*Proof.* By differentiating $\varphi_i(x)$ it is not difficult to see that inside the sphere $S_i$ its partial derivatives of all orders exist. And the modulus of any partial derivative of order $k$ is bounded inside $S_i$ by $AB\,(s, k, n)\,\delta^{-k}$, where $B\,(s, k, n)$ is some constant, depending only on $s, k, n$. In particular, any derivative of the function $\varphi_i(x)$ of order $p + 1$ is bounded in the sphere $S_i$ by the constant

$$AB\,(s, p+1, n)\,\delta^{-p-1} = \frac{A\,(s, L, n)\,B\,(s, p+1, n)}{\delta^{1-\alpha}}.$$

Let $g\,(x)$ be any $p$-th order partial derivative of the functions $\varphi_i(x)$. We take two points $a$ and $b$ belonging to the sphere $S_i$. Then $g\,(b) - g\,(a) = r\,(a, b)\,\frac{\partial g\,(c)}{\partial r}$, where $\frac{\partial g\,(c)}{\partial r}$ is the derivative of $g\,(x)$ along the direction $(a, b)$, taken at some point $c$ of $[a, b]$. Since any $p$ + 1-th order partial derivative of $\varphi_i(x)$ is bounded inside the sphere by the constant

$\frac{A\,(s, L, n)\,B\,(s, p+1, n)}{\delta^{1-\alpha}}$, we have $\left|\frac{\partial g\,(c)}{\partial r}\right| \leqslant n\,\frac{A\,(s, L, n)\,B\,(s, p+1, n)}{\delta^{1-\alpha}}$

And then

$$| g(b) - g(a) | \leqslant \left| \rho \frac{\partial g(c)}{\partial r} \right| \leqslant \rho n \frac{A(s, L, n) B(s, p+1, n)}{\delta^{1-\alpha}}$$

$$\leqslant \rho^{\alpha} n A(s, L, n) B(s, p+1, n) .$$

Put

$$A(s, L, n) = \frac{L}{2n B(s, p+1, n)} .$$

Then

$$| g(b) - g(a) | \leqslant \frac{1}{2} L \rho^{\alpha} .$$

Now let $\Psi(x)$ be any of the $p$-th partial derivatives of the function $\varphi_{\eta_1, \eta_2, \ldots, \eta_h}(x)$. We choose two points $x'$ and $x''$ of $\mathscr{I}$ ($x' \in S_i$, $x'' \in S_j$) and let $g_1(x)$ and $g_2(x)$ be the partial derivatives of the same kind as $\Psi(x)$ of the functions $\varphi_i(x)$ and $\varphi_j(x)$ (respectively). It is easy to verify that $g_1(x)$ and $g_2(x)$ are continuous on $\mathscr{I}$ and identically equal to zero on the sets $\mathscr{I} - S_i$ and $\mathscr{I} - S_j$ (respectively). We select some point $x_0$ belonging to the boundary of the sphere $S_i$ and lying on the segment $[x', x'']$. Then

$$| \psi(x'') - \psi(x') | \leqslant | g_1(x'') - g_1(x') | + | g_2(x'') - g_2(x') |$$

$$\leqslant | g_1(x') - g_1(x_0) | + | g_2(x'') - g_2(x_0) | \leqslant | g(b) - g(a) |$$

$$\leqslant \frac{1}{2} L (r(x', x_0))^{\alpha} + \frac{1}{2} L (r(x'', x_0))^{\alpha} \leqslant L (r(x', x''))^{\alpha} .$$

If one of the points $x'$, $x''$ (or both) belongs to the set $\mathscr{I} - \sum_{i=1}^{h} S_i$, then we can prove similarly that

$$| \varphi(x'') - \varphi(x') | \leqslant L (r(x', x''))^{\alpha} .$$

Q.E.D.

LEMMA 2.2.2. *There exists a positive constant* $A$, *depending only on* $s, L, n$ *such that for sufficiently small* $\varepsilon$

$$H_{\varepsilon}(F) \geqslant A \rho^{n} \left( \frac{1}{\varepsilon} \right)^{n/s} .$$

*Proof.* We choose some positive number $k > 1$ such that when $\delta = \left( \frac{k \varepsilon}{A(s, L, n)} \right)^{1/s}$ is an integer.

We choose two different functions of the type $\varphi_{\eta 1,\dots,\eta h}(x)$ and $\varphi_{\tau 1,\tau 2,\dots,\tau h}(x)$, $A = A(s, L, n)\,\delta^s$ and $A(s, L, n)$ is taken so small that both functions belong to the family $F$. Since the functions we have chosen are assumed to be different, for some $i$ $\tau_i \neq \eta_i$. And therefore

$$| \varphi_{\eta 1,\eta 2,\dots,\eta h}(c_i) - \varphi_{\tau 1,\tau 2,\dots,\tau_n}(c_i) |$$
$$= 2A = 2A(s, L, n)\,\delta^s = 2k\varepsilon > 2\varepsilon\,.$$

Hence

$$H_\varepsilon(F) \geqslant \log\, 2^h = \left(\frac{\rho}{\delta}\right)^n = \left(\frac{A(s, L, n)}{k}\right)^{\frac{n}{s}} \rho^n \left(\frac{1}{\varepsilon}\right)^{\frac{n}{s}}$$

Q.E.D.

LEMMA 2.2.3. *There exists a constant* $B > 0$ *such that for sufficiently small* $\varepsilon > 0$

$$H_\varepsilon(F) \leqslant B\rho^n \left(\frac{1}{\varepsilon}\right)^{\frac{n}{s}}$$

*Proof.* Let us choose some $\delta > 0$ such that the ratio $\rho/\delta$ is an integer. In the cube $\mathscr{I}$ consider the uniform lattice with step $\delta$, consisting of the points $d_i\,(i = 1, 2, \dots, h; h = \left(\frac{\rho}{\delta} + 1\right)^n)$.

We shall assume the corners of the lattice to be numbered so that the point $d_1$ coincides with the origin of co-ordinates, and for any $i$

$$r(d_{i-1}, d) = \delta\,.$$

We now choose some function $f(x)$ of the family $F$ and we shall show a method of constructing a table for this function the volume of which is less than $B\rho^n \left(\frac{1}{\varepsilon}\right)^{n/s}$.

Let $h_p$ denote the number of different kinds of partial derivative (of all orders up to and including the $p$-th) of a function of $n$ variables. It is not difficult to verify that $h_p \leqslant (p+1)^n$. Let $\{ \tau_1^{j,k} \}$ $(\tau_1^{j,k} = 0, 1)$ be the coefficients of the binary representation of the numbers

$$\frac{\partial^{k_1+k_2+\dots+k_n} f(d_1)}{\partial x_1^{k_1}\, \partial x_2^{k_2} \dots \partial x_n^{k_n}} \qquad (k_1 + k_2 + \dots + k_n) \leqslant p$$

written in some order ($k$ is the order of the derivative, $j = 1, 2, \dots, h_1^k$). Then the numbers

$$\left\{\frac{\partial^{k_1+k_2+\ldots+k_n} f(d_1)}{\partial x_1^{k_1}\,\partial x_2^{k_2}\ldots\partial x_n^{k_n}}\right\} \qquad (k_1+k_2+\ldots+k_n=k)$$

are represented in the table to an accuracy of $\delta^{s-k}$, i.e.

$$h_1^k \leqslant \left(\left[\log\frac{c}{\delta^{s-k}}\right]+1\right)(k+1)^n$$

binary digits $\tau_1^{s,k}$ $(j=1, 2, ..., h_1^k)$ are sufficient to represent them in binary. Thus, to represent all partial derivatives of $f(x)$ at the point $x = d_1$ in binary we need

$$h_1 = \sum_{k=0}^{p} h_1^k \leqslant (p+1)^{n+1}\left(1+\log\frac{c}{\delta^s}\right)$$

binary digits

$$\tau_1^{j,k}\,(j=1, 2, \ldots, h_1^k\,, \quad k=0, 1, 2, \ldots, p)\,.$$

Let us assume now that we have found a method for selecting the digits $\{\tau_1^{j,k}\}$ $(i=1, 2, ..., q-1)$ together with a rule for calculating from these digits the values of the numbers

$$\left\{\frac{\partial^{k_1+k_2+\ldots+k_n} f(d_i)}{\partial x_1^{k_1}\,\partial x_2^{k_2}\ldots\partial x_n^{k_n}}\right\} \qquad (k_1+k_2+\ldots+k_n=k)$$

$(i=1, 2, ..., q-1)$ to an accuracy of $\delta^{s-k}$ $(k=0, 1, ..., p)$. We examine the subsequent procedure for constructing the table for $f(x)$. Let $g_k(x)$ be one of the $k$-th order partial derivatives of $f(x)$. According to the induction hypothesis, the values of all partial derivatives of order $m \leqslant p-k$ of $g_k(x)$ at the point $x = d_{q-1}$ can be calculated to an accuracy of $\delta^{s-k-m}$ $(m=0, 1, ..., p-k)$ from that part of the table already constructed. From Lagrange's formula, the value of $g_k(d_q)$ is found sufficiently accurately from the approximate values of the derivatives of $g(x)$ at $d_{q-1}$. Therefore, to represent the numbers $g_k(d_q)$ to an accuracy of $\delta^{s-k}$ we need only a small number of binary digits. Since $r(d_{q-1}, d_q) = \delta$ all the corresponding co-ordinates (except one) of the points $d_{q-1}, d_q$ are equal. For definiteness, we shall suppose that

$$x_1(d_q) = x_1(d_{q-1}) + \delta \quad \text{and} \quad x_i(d_q) = x_i(d_{q-1})$$

for $i = 2, 3, ..., n$. Then

$$g_k(d_q) = \sum_{m=0}^{p-m-1} \frac{\partial^m g_k(d_{q-1})}{\partial x_1^m}\cdot\frac{\delta^m}{m!}$$

$$+\frac{1}{(p-1)!}\,\frac{\partial^{p-k} g_k(d_{q-1}+\theta\delta)}{\partial x_1^{p-k}}\,\delta^{p-k}$$

$$=\sum_{m=0}^{p-k}\frac{\partial^m g_k(d_{q-1})}{\partial x_1^m}\cdot\frac{\delta^m}{m!}+\frac{L}{(p-1)!}\,\theta\delta^{s-k},$$

where $0 \leqslant \theta \leqslant 1$. But since $\dfrac{\partial^m g_k(d_{q-1})}{\partial x_1^m}$ is given by the table only to an accuracy of $\delta^{s-k-m}$ $(m=0, 1, \ldots, p-k)$ $g_k(d_q)$ is determined by the constructed part of the table only to an accuracy of

$$\sum_{m=0}^{p-1}\delta^{s-k-m}\frac{\delta^m}{m!}+\frac{L\delta^{s-k}}{(p-k)!}=\delta^{s-k}\left(\sum_{m=0}^{p-k}\frac{1}{m!}+\frac{L}{(p-k)!}\right)\leqslant e(L+1)^{s-k}$$

Therefore, in order to represent the value of $g_k(d_q)$ in the table to an accuracy of $\delta^{s-k}$, it is sufficient to put another $h_q^{j,k}=[\log((L+1)\,e)]+1$ binary digits in the table. Hence, to determine the values of all $k$ th order partial derivatives of $f(x)$ it is sufficient to add $h_q^k \leqslant (k+1)^n\, h_q^{j,k}$ binary digits to the table $(k=0, 1, \ldots, p)$. Thus, the approximate representation of the values of all partial derivatives of the functions $f(x)$ at the point will use only

$$h_q=\sum_{k=0}^{p} h_q^k \leqslant (p+1)^{n+1}\,(1+\log\,[e(L+1)])$$

binary digits.

The volume of the table $T$ which we have constructed is equal to

$$P(T)=\sum_{q=1}^{k} h_q \leqslant (p+1)^{n+1}\left(1+\log\frac{c}{\delta^s}\right)$$
$$+(h-1)(p+1)^{n+1}\,(1+\log\,[e(L+1)])\,.$$

We shall now describe the rule we use to enable us to compute the value of $f(x)$ at any point of the cube $\mathscr{I}$ from the parameters of the table. To do this, we divide the cube $\mathscr{I}$ in some way into sets $\omega_q$ $(\omega_q \ni d_q)$ the diameter of each set not exceeding $\delta\sqrt{n}$, and such that $\sum_{q=1}^{h}\omega_q=\mathscr{I}$. The approximate value of the function $f(x)$ is calculated using the parameters $\tau_q^{j,k}$ of $T$ in the following way.

Let $x \in \omega_q$. Then, for the approximate value of $f(x)$ we take

$$f^*(x)=\sum_{k_1+k_2+\ldots+k_n\leqq p} a_{k_1,k_2,\ldots,k_n}\prod_{i=1}^{n}\frac{(x_i-x_i(d_q))^{k_i}}{k_i!}$$

where $a_{k_1,k_2,\ldots,k_n}$ is the approximate value (to an accuracy of $\delta^{s-k}$, $k = \sum_{i=1}^{n} k_i$)

of partial derivative $\dfrac{\partial^{k_1+k_2+\ldots+k_n} f(d_q)}{\partial x_1^{k_1}\, \partial x_2^{k_2} \ldots \partial x_n^{k_n}}$. Since $f(x) \in F$

$$\| f(x) - f^*(x) \| \leqslant \delta^s \big((p+1)^m + L + 1\big) = B(s, L, n)\, \delta^s = \varepsilon' .$$

Therefore,

$$H_{\varepsilon'}(F) \leqslant (p+1)^{n+1} \left(1 + \log \frac{c}{\delta^s}\right) + (h-1)(p+1)^{n+1} \big(1 + \log\, (e(L+1))\big) .$$

We now define $\delta$ in the form

$$\delta = \left(\frac{k\varepsilon}{B(s, L, n)}\right)^{1/s}$$

We choose $k < 1$ so that the ratio $\rho/\delta$ is an integer. Then

$$H_\varepsilon(F) \leqslant H_{\varepsilon'}(F) \leqslant (p+1)^{n+1} \left(1 + \log \frac{c}{\delta^s}\right)$$
$$+ (h-1)(p+1)^{n+1} \big(1 + \log\, (e(L+1))\big),$$

i.e. for sufficiently small $\varepsilon$ $H_\varepsilon(F) \geqslant B\rho^n \left(\dfrac{1}{\varepsilon}\right)^{n/s}$, where $B > 0$ is a constant which can be taken to depend on $s, L, n$ only.

Q.E.D.

*Proof of the Theorem* 2.2.1. First let $L = 1$. Then from lemmas 2.2.2. and 2.2.3 we have

$$A\rho^n \left(\frac{1}{\varepsilon}\right)^{n/s} \leqslant H_\varepsilon(F) \leqslant B\rho^n \left(\frac{1}{\varepsilon}\right)^{n/s}$$

where $A$ and $B$ are positive constant, depending only on $s$ and $n$, since in this case $L = 1$. But since

$$H_{\frac{\varepsilon}{L}}(F_{s,1,\frac{c}{L}}) = H_\varepsilon(F)$$

for sufficiently small $\varepsilon$

$$A(s, n)\, \rho^n \left(\frac{L}{\varepsilon}\right)^{n/s} \leqslant H_\varepsilon(F) \leqslant B(s, n)\, \rho^n \left(\frac{L}{\varepsilon}\right)^{n/s}$$

Q.E.D.

## § 3. *Theorem on superpositions of smooth functions*

We will denote by $C_s(\mathcal{J}^n)$ the space of $n$ times differentiable functions of $n$ variables defined on the cube $\mathcal{J}^n$ with the norm

$$\|f\| = \sum_{p=1}^{s} \sum_{k_1+k_2+\ldots+k_n=p} \max_{x\in\mathcal{J}^n} \left| \frac{\partial^{k_1+\ldots+k_n} f(x)}{\partial x_1^{k_1} \ldots \partial x_n^{k_n}} \right|$$

THEOREM 2.3.1. *Let the numbers* $s \geqslant 1$, $s' \geqslant 1$ *and natural* $n$ *and* $n'$ *be such that* $\frac{n}{s} > \frac{n'}{s'}$. *Then the set of functions from* $C_s(\mathcal{J}^n)$ *not representable on* $\mathcal{J}^n$ *by superpositions of* $S'$ *times differentiable functions of* $n'$ *variables is a set of second category.*

The space $C_s(\mathcal{J}^n)$ is complete and consequently the set mentioned in the theorem is not empty. The theorem is true for any $s \geqslant 1$, $s' \geqslant 1$ but we will assume for simplicity that $s$ and $s'$ are integers.

LEMMA 2.3.1. *Let* $f$ *and* $\tilde{f}$ *be* $q$*-fold superpositions composed of the functions* $\{\varphi^p_{\alpha_1,\ldots,\alpha_p}\}$ *and* $\{\tilde{\varphi}^p_{\alpha_1,\ldots,\alpha_p},\}$ *where all functions composing the superpositions satisfy the condition* Lip 1 *with the constant* $L$ *and for any collection* $p, \alpha_1, \ldots, \alpha_p$

$$\max \left| \varphi^p_{\alpha_1,\ldots,\alpha_p} - \tilde{\varphi}^p_{\alpha_1,\ldots\alpha_p} \right| \leqslant \varepsilon$$

*Then*

$$\max_{x\in\mathcal{J}^n} \left| f(x) - \tilde{f}(x) \right| \leqslant (L+1)^q \;\; \varepsilon$$

The lemma can easily be proved by induction in $q$.

LEMMA 2.3.2. *Let* $\Omega$ *be an open subset of* $C_s(\mathcal{J}^n)$ *and* $\Omega^* \subset C(\mathcal{J}^n)$. *If every* $f \in \Omega$ *allows uniform approximations on* $\mathcal{J}^n$ *with any accuracy by functions from* $\Omega^*$, *i.e. the closure of* $\Omega^*$ *contains* $\Omega$, *then* $H_\varepsilon(\Omega^*) \geqslant C\left(\frac{1}{\varepsilon}\right)^{n/s}$, *where* $C > 0$ *is independent of* $\varepsilon$.

The lemma is easily reduced to lemma 2.2.1 and lemma 2.2.2.

We denote by $\Omega_k$ the set of all functions of $C(\mathcal{J}^n)$ which are $k$-fold superpositions composed of $s'$ times differentiable functions of $n'$ variables with partial derivatives bounded by the same constant $k$.

LEMMA 2.3.3. *If* $\frac{n}{s} > \frac{n'}{s'}$ *then for any natural* $k$ *the set* $\Omega_k \cap C_s(\mathcal{I}^n)$ *is nowhere dense in* $C_s(\mathcal{I}^n)$.

By lemma 2.3.1 and the theorem 2.2.1 for any natural $k$ $H_\varepsilon(\Omega_k) \leqslant C\left(\frac{1}{\varepsilon}\right)^{n'/s'}$, where $C$ does not depend on $\varepsilon$. Hence, it follows from the inequality $\frac{n}{s} > \frac{n'}{s'}$ and lemma 2.3.2 that the set $\Omega_k \cap C_s(\mathcal{I}^n)$ is nowhere dense in $C_s(\mathcal{I}^n)$.

Now to prove the theorem we have to notice only that the set of functions from $C_s(\mathcal{I}^n)$ representable by superpositions coincides with $\bigcup_{k=1}^{\infty} (\Omega_k \cap C_s(\mathcal{I}^n))$. By lemma 2.3.3 the sets $\{\Omega_k \cap C_s(\mathcal{I}^n)\}$ are nowhere dense and consequently the set of not representable functions is a set of second category.

## CHAPTER 3. — SUPERPOSITIONS OF CONTINUOUS FUNCTIONS

In this chapter we present the proof of the theorem of Kolmogorov given by Kahane [36]. This proof which is based on Baire's theory contains a minimum of concrete constructions and shows that there exists a wide choice of inner functions for Kolmogorov's formula.

### § 1. *Certain improvements of Kolmogorov's theorem*

By the theorem of Kolmogorov any function defined and continuous on the cube $\mathcal{I}^n$ can be represented as

$$f(x_1, \dots, x_n) = \sum_{q=1}^{2n+1} g_q\Big(\sum_{p=1}^{n} \varphi_{p,q}(x_p)\Big),$$

where $\{\varphi_{p,q}\}$ are specially chosen continuous and monotonic functions which do not depend on $f$, and where $\{g_q\}$ are continuous functions.

Lorentz [12] has noticed that in the theorem of Kolmogorov the functions $\{g_q\}$ can be chosen independently of $q$. In fact, by adding constants to the functions $t_q = \sum_{p=1}^{n} \varphi_{p,q}(x_p)$ $(q = 1, \dots, 2n+1)$ one can make the ranges of the functions pairwise disjoint and consequently the functions $\{ t_q \}$ can be considered as the restrictions of a single function $\{ g_q \}$.

Sprecher [40] has shown that the functions $\{ \varphi_{p,q} \}$ can be chosen in the form $\varphi_{p,q}(x_p) = \lambda_p \varphi_q(x_p)$ where $\{ \lambda_p \}$ are constants and $\{ \varphi_q \}$-are continuous monotonic functions.

Thus any continuous function can be represented as

$$f(x_1, \ldots, x_n) = \sum_{q=1}^{2n+1} g\left( \sum_{p=1}^{n} \lambda_p \varphi_q(x_p) \right),$$

where the constants $\{ \lambda_p \}$ and the continuous monotone functions $\{ \varphi_q \}$ do not depend on $f$, and where $g$ is a continuous function.

Kahane [36] has shown that such a representation is possible with almost every collection of constants $\{ \lambda_p \}$ and "quasi every" collection of continuous functions $\{ \varphi_q \}$. The precise statement of this theorem will be given below. Here we consider some further results concerning the formula of Kolmogorov.

Doss [38] has shown that for any continuous monotonic functions $\varphi_{p,q}$ $(p=1, 2;\ q=1, 2, 3, 4)$ there exists a continuous function $f(x_1, x_2)$ of two variables not representable as a superposition of the form $\sum_{q=1}^{4} g_q \left( \sum_{p=1}^{2} \varphi_{p,q}(x_p) \right)$, where $\{ g_q \}$ are continuous functions.

Bassalygo [39] succeeded in showing that for any continuous functions $\varphi_i(x_1, x_2)$ $(i=1, 2, 3)$ there exists a continuous function $f(x_1, x_2)$ that is not equal to any superposition of the form $\sum_{i=1}^{3} g_i(\varphi_i(x_1, x_2))$, where $\{ g_i \}$ are continuous functions.

Tihomirov showed that Kolmogorov's theorem can be generalized as follows: for any compact $K$ of dimension $n$ there exists a homeomorphic embedding $\Psi(x) = \{ \Psi_1(x), \ldots, \Psi_{2n+1}(x) \}$, $x \in K$ into $(2n+1)$-dimensional euclidean space such that any continuous function $f(x)$ on $K$ can be represented in the form $f(x) = \sum_{i=1}^{2n+1} g_i(\Psi_i(x))$, where $\{ g_i \}$ are continuous functions of one variable.

In the same paper [36] Kahane has shown that there exist complex numbers $\lambda_p$ $(p=1, \ldots, n)$ and complex valued functions $\varphi_q$ $(q=1, \ldots, 2n+1)$ possessing the following properties.

1. The function $\varphi_q$ is a monotonic continuous transformation of the real axis onto the circle $|t| = 1$ $(q=1, \ldots, 2n+1)$.

2. The function $t_q = \sum_{p=1}^{n} \lambda_p \varphi_q(x_p)$ maps the cube $\mathscr{I}^n$ into the circle $|t| = 1$.

3. The transformation $\Psi$ given by the equalities $t_q = \sum_{p=1}^{n} \lambda_p \varphi_q(x_p)$ $(q = 1, \ldots, 2n+1)$ is one-to-one on $\mathscr{I}^n$.

4. For any function $f$ continuous on $\mathscr{I}^n$ there exists a function $g(z)$ continuous on the disk $|z| \leqslant 1$, holomorphic inside that disk, and such that $f = \sum_{q=1}^{2n+1} g\left(\sum_{p=1}^{n} \lambda_p \varphi_q(x_p)\right)$.

The transformation $\Psi$ gives an embedding of the cube $\mathscr{I}^n$ into the torus $|t| = 1$ $(q = 1, \ldots, 2n+1)$ such that any function continuous on the cube $\tilde{\mathscr{I}}^n = \Psi(\mathscr{I}^n)$ is represented in the form $f(t_1, \ldots, t_{2n+1}) = \sum_{q=1}^{2n+1} g(t_q)$, where $g$ is a function holomorphic in the unit disk. This means in particular that any function continuous on $\tilde{\mathscr{I}}^n$ has an analytic extension to the polydisk $|t_q| \leqslant 1$ $(q = 1, \ldots, 2n+1)$.

## § 2. *The theorem of Kahane*

Let $M$ be a complete metric space. We recall that a set is called a set of second category if it is the intersection of a countable family of open sets which are everywhere dense in $M$. By the theorem of Baire in a complete metric space no set of second category is empty. The massivity of such sets is characterized by the fact that the intersection of a countable family of sets of second category is again a set of second category and consequently is not empty.

We will say that a statement is true for quasi every element of $M$ if it is true for a set of elements of second category.

Let us consider an example. Let $\Phi$ be the space with uniform norm consisting of all functions continuous and non-decreasing on the segment $\mathscr{I}^1$ $(0 \leqslant t \leqslant 1)$. It can be shown easily that quasi every element of $\Phi$ is a strictly increasing function.

In fact, any strictly increasing function belongs to any set defined as $\varphi(r') < \varphi(r'')$, where $r' < r''$ are fixed rational numbers. Any set defined by an inequality of that type is open and everywhere dense in $\Phi$, and the set of all such sets is countable.

Let $\mathscr{I}^n$ be the cube $\{ 0 \leqslant x_i \leqslant 1, i = 1, \ldots, n \}$; $C(\mathscr{I}^n)$-the space of all functions continuous on $\mathscr{I}^n$ with the uniform norm; $\Phi$-the space of functions continuous and non-decreasing on the segment $\mathscr{I}^1$ (with the uniform norm); $\Phi^k = \Phi \times \ldots \times \Phi$ the $k$-th power of the space $\Phi$.

THEOREM 3.2.1. *Let* $\lambda_p$ $(p=1, \ldots, n)$ *be a collection of rationally independent constants. Then for quasi every collection* $\{ \varphi_1, \ldots, \varphi_{2n+1} \} \in \Phi^{2n+1}$ *it is true that any function* $f \in C(\mathscr{I}^n)$ *can be represented on* $\mathscr{I}^n$ *in the form*

$$f(x) = \sum_{q=1}^{2n+1} g\Big( \sum_{p=1}^{n} \lambda_p \varphi_q(x_p)\Big),$$

*where* $g$ *is a continuous function.*

## § 3. *The main lemma*

We fix a function $f \in C(\mathscr{I}^n)$, positive numbers $\lambda_p$ $(p=1, \ldots, n)$ and a positive $\varepsilon$. We will denote by $\Omega_f$ the set of all collections $\{ \varphi_1, \ldots, \varphi_{2n+1} \} \in \Phi^{2n+1}$ for each of which there exists a continuous function $h$ such that $\| h \| \leqslant \| f \|$ and $\| f(x) - \sum_{q=1}^{2n+1} h\big( \sum_{p=1}^{n} \lambda_p \varphi_q(x_p)\big) \| < (1-\varepsilon) \| f \|$. The latter inequality is strict and consequently the set $\Omega_f$ is open.

The idea of the construction is contained in the following statement.

LEMMA 3.3.1. *If* $\| f \| \neq 0$, *the numbers* $\{ \lambda_p \}$ *are rationally independent, and* $0 < \varepsilon < \dfrac{1}{2n+2}$, *than the corresponding set* $\Omega_f$ *is everywhere dense in* $\Phi^{2n+1}$.

*Proof.* Let us fix an open set $\Omega \subset \Phi^{2n+1}$ and prove that $\Omega \cap \Omega_f$ is not empty. This will imply that $\Omega_f$ is everywhere dense in $\Phi^{2n+1}$.

We choose a number $\delta > 0$ and denote by $\mathscr{I}_q(j)$ the segment defined by the inequality

$$q \cdot \delta + (2n+1) j \cdot \delta \leqslant t \leqslant q \cdot \delta + (2n+1) j\delta + 2n\delta$$
$$(q=1, \ldots, 2n+1, j \text{ is an integer}).$$

The value $\delta$ will be determined below. Now we notice, firstly, that for any $q$ the segments $\mathscr{I}_q(j)$ $(j=0, \pm 1, \pm 2)$ are pairwise desjoint and every two consecutive segments are separated by an interval of length $\delta$ and, secondly, that, every point of the real axis belongs to at least $2n$ of the sets $\sum_j \mathscr{I}_q(j)$, $(q=1, \ldots, 2n+1)$.

We denote by $P_q(j_1, \ldots, j_n)$ the cube

$$q\delta + (2n+1)j_k\delta \leqslant x_k \leqslant q \cdot \delta + (2n+1)j_k\delta + 2n\delta \ (k=1, \ldots, n).$$

We emphasise that every point $x \in \mathscr{I}^n$ belongs to at least $n+1$ of the sets $\sum_{j_1, \ldots, j_n} P_q(j_1, \ldots, j_n)$ $(q=1, \ldots, 2n+1)$. We also remark that for any $q$ the cubes $\{P_q(j_1, \ldots, j_n)\}$ are pairwise disjoint.

We denote by $\Omega^*$ the subset of $\Phi^{2n+1}$ consisting of the collections $\varphi_1, \ldots, \varphi_{2n+1}$ such that for every $q$ the function $\varphi_q$ is constant on every one of the segments $\{\mathscr{I}_q(j)\}$. We will assume that $\delta$ is so small that $\Omega^* \cap \Omega$ is not empty.

We choose a collection $\{\varphi_1, \ldots, \varphi_{2n+1}\} \in \Omega^* \cap \Omega$. We will show that this collection belongs to $\Omega_f$. We put $t_q \equiv \sum_{p=1}^{n} \lambda_p \varphi_q(x_p)$. Since the numbers $\{\lambda_p\}$ are rationally independent we can change the constants $\{\varphi_q(\mathscr{I}_q(j))\}$ slightly, so that the new values of $t_q(p_q(j_1, \ldots, j_n))$ are pairwise different and the collection $\varphi_1, \ldots, \varphi_{2n+1}$ remains in $\Omega^* \cap \Omega$.

We denote by $f_q(j_1, \ldots, j_n)$ the value of the function $f$ at the center of $P_q(j_1, \ldots, j_n)$ and by $h$ the function defined in the following way: $h(t_q(j_1, \ldots, j_n)) = \frac{1}{2n+1} f_q(j_1, \ldots, j_n)$ outside the set $\bigcup_{q, j_1, \ldots, j_n} t_q(j_1, \ldots, j_n)$ the function $h$ is defined in such a way that it is continuous on the whole real axis and $\| h \| \leqslant \frac{1}{2n+1} \| f \|$.

Now we estimate the function $\left| f - \sum_{q=1}^{2n+1} h(t_q) \right| = \left| \sum_{q=1}^{2n+1} \frac{f}{2n+1} - h(t_q) \right|$. For any $x \in \mathscr{I}^n$, $q, j_1, \ldots, j_n$

$$\left| \frac{f}{2n+1} - h(t_q) \right| \leqslant \frac{1}{2n+1} \| f \| + \| h \| \leqslant \frac{1}{2n+1} \| f \| + \frac{1}{2n+1} \| f \|$$
$$= \frac{2}{2n+1} \| f \| .$$

If $x \in P_q(j_1, \ldots, j_n)$, then

$$\left| \frac{f}{2n+1} - h(t_q) \right|$$
$$\leqslant \max_{q, j_1, \ldots, j_n} \left| \max_{x \in p_q(j_1, \ldots, j_n)} \frac{f(x)}{2n+1} - \min_{x \in p_q(j_1, \ldots, j_n)} \frac{f(x)}{2n+1} \right| = \rho .$$

We recall that every $x \in \mathscr{I}^n$ belongs to at least $n+1$ of the cubes $\{P_q(j_1, \ldots, j_n)\}$. Hence

$$|f - \sum_{q=1}^{2n+1} h(t_q)| \leqslant (n+1)\rho + n\frac{2}{2n+1}\|f\| .$$

But $\lim\limits_{\delta\to 0} \rho = 0$, consequently for sufficiently small $\delta$ and $\varepsilon < \dfrac{1}{2n+2}$

$$|f - \sum_{q=1}^{2n+1} h(t_q)| < (1-\varepsilon)\|f\| .$$

The lemma is proved.

## § 4. *The proof of the theorem*

We denote by $F$ a countable set, everywhere dense in $C(\mathscr{I}^n)$. We choose $\varepsilon$ satisfying the condition of lemma 3.3.1 and consider $\Omega_{f_k}$ $(f_k \in F)$ corresponding to this $\varepsilon$ and the collection $\lambda_p$ mentioned in the theorem. The sets $\{\Omega_{fk}\}$ are open and by lemma 3.3.1 they are everywhere dense in $\Phi^{2n+1}$. Consequently, according to the definition, almost every element of $\Phi^{2n+1}$ belongs to $\Phi^* = \bigcap\limits_{f_k \in F} \Omega_{f_k}$.

We fix a collection $\{\varphi_1, \ldots, \varphi_{2n+1}\} \in \Phi^*$ and a function $f \in C(\mathscr{I}^n)$ and show that the desired representation of $f$ takes place. If $f \equiv 0$ then as the function $g$ we can take $g \equiv 0$. We will assume below that $f \not\equiv 0$. According to the definition of $\Omega_{f_k}$ there exists for any $f_k \in F$ a function $h_k$ such that $|f_k - \sum\limits_{q=1}^{2n+1} h_k(\sum\limits_{p=1}^{n} \lambda_p\varphi_q(x_p))| \leqslant (1-\varepsilon)\|f_k\|$. The set $F$ is everywhere dense in $C(\mathscr{I}^n)$. Consequently for any $f \in C(\mathscr{I}^n)$ $(f \not\equiv 0)$ there exists $h = \gamma(f)$ such that

$$|f - \sum_{q=1}^{2n+1} h(\sum_{p=1}^{n} \lambda_p\varphi_q(x_p))| < \left(1 - \frac{\varepsilon}{2}\right)\|f\| .$$

We define the sequence of functions $\chi_0, \chi_1, \chi_2, \ldots$ by the recurrent equalities

$$\chi_0 = f, \quad \chi_{k+1} = \chi_k - \sum_{q=1}^{2n+1} g_k(\sum_{p=1}^{n} \lambda_p\varphi_q(x_p)),$$

where $g_k = \gamma(\chi_k)$. The series $\sum\limits_{k=0}^{\infty} g_k$ converges uniformly and consequently the function $g = \sum\limits_{k=0}^{\infty} g_k$ is continuous and

$$f - \sum_{q=1}^{2n+1} g(\sum_{p=1}^{n} \lambda_p\varphi_q(x_p)) = 0 .$$

The theorem is proved.

# CHAPTER 4. — LINEAR SUPERPOSITIONS

In this chapter we prove that there exist analytic functions which are not representable by means of linear superpositions of smooth functions of one variable.

## § 1. *Notation*

Throughout we assume that all the functions are defined and continuous for all values of the arguments. If we say that a function is continuously differentiable, we mean by this that its first partial derivatives are defined and continuous for all values of the arguments; $z = (x, y)$ is the point of the plane with coordinates $x$ and $y$; grad $[q(z)]$ is the gradient of the function $q(z)$, that is, the vector-function with coordinates $\frac{\partial q}{\partial x}$ and $\frac{\partial q}{\partial y}$;

$$D\left(\frac{q_1, q_2}{x, y}\right) = \begin{vmatrix} \frac{\partial q_1}{\partial x} & \frac{\partial q_1}{\partial y} \\ \frac{\partial q_2}{\partial x} & \frac{\partial q_2}{\partial y} \end{vmatrix}$$

is the Jacobian of the pair of functions $q_1$ and $q_2$.

$q(D)$ is the image of the set $D$ under the mapping effected by the function $q(x, y)$; $q^{-1}(\delta)$ is the complete inverse image of the interval $\delta$ on the axis of values of the function $q(x, y)$.

$e(q, t)$ is the set of level $t$ of the function $q = q(x, y)$.

$\tau(e, z)$ is the unit tangent vector to the curve $e$ at the point $z \in e$.

$\gamma(\tau_1, \tau_2)$ is the absolute value of the acute angle between the vectors $\tau_1$ and $\tau_2$.

$h_1(e)$ is the length of the set $e$.

$d_1(e)$ is the one-dimensional diameter of the set $e$.

$0(\gamma)$ is a quantity bounded by a constant depending only on $\gamma$.

$\rho(A_1, A_2)$ is the distance between the sets $A_1$ and $A_2$ in the sense of deviation, more precisely

$$\rho(A_1, A_2) = \max\{\sup_{z_1 \in A_1} \inf_{z_2 \in A_2} \rho(z_1, z_2), \sup_{z_2 \in A_2} \inf_{z_1 \in A_1} \rho(z_1, z_2)\},$$

where $\rho(z_1, z_2)$ is the distance between the points $z_1$ and $z_2$.

## § 2. *Estimate of the difference of the integrals of one term of a superposition along nearby level curves*

Let $G$ be a region of the plane of the variables $x$ and $y$, and $q_1(x, y)$ and $q_2(x, y)$ continuously differentiable functions satisfying in this region the following conditions: a) the partial derivatives with respect to $x$ and with respect to $y$ have modulus of continuity $\omega(\delta)$; b) the inequalities

$$0 < \gamma \leqslant |\operatorname{grad} [q_i(x, y)]| \leqslant \frac{1}{\gamma} < \infty \qquad (i = 1, 2)$$

are satisfied everywhere in $G$, where $\gamma$ is a constant; c) for any point $(x, y) \in G$ the absolute value of the acute angle formed by the level curves of the functions $q_1(x, y)$ and $q_2(x, y)$ which pass through this point is greater than some positive constant $\gamma$.

LEMMA 4.2.1. *Let $e'_{q2}$ and $e''_{q2}$ be two level curves of the function $q_2$ and $e'_{q1}$ and $e''_{q1}$ level curves of the function $q_1$; $[a', a''] \subset G$ the segment of the curve $e'_{q1}$ with end-points $a' \in e'_{q2}$ and $a'' \in e''_{q2}$; $[b', b'']$ the segment of the curve $e''_{q1}$ with end-points $b' \in e'_{q2}$ and $b'' \in e''_{q2}$. Then*

$$h_1([b', b'']) \leqslant h_1([a', a'']) \times (1 + c_1(\gamma)\,\omega(\delta)),$$

*where $\delta = d_1([a', a''] \cup [b', b''])$ and $c_1(\gamma)$ depends only on $\gamma$.*

*Proof.* Since $q_2(a'') - q_2(a') = q_2(b'') - q_2(b')$, we have

$$\int\limits_{s \in [a', a'']} \frac{\partial q_2}{\partial s}\, ds = \int\limits_{s \in [b', b'']} \frac{\partial q_2}{\partial s}\, ds .$$

Consequently, $\dfrac{\partial q_2(a^*)}{\partial s} h_1([a', a'']) = \dfrac{\partial q_2(b^*)}{\partial s} h_1([b', b''])$, where $\dfrac{\partial q_2(a^*)}{\partial s}$ and $\dfrac{\partial q_2(b^*)}{\partial s}$ are the derivatives at the points $a^* \in [a', a'']$ and $b^* \in [b', b'']$ along the curves $[a', a'']$ and $[b', b'']$, respectively. We show that $\dfrac{\partial q_2(a^*)}{\partial s} = \dfrac{\partial q_2(b^*)}{\partial s} + O(\gamma)\,\omega(\delta)$. We denote by $q_2^*$ the derivative of $q_2$ at the point $b^*$ in the direction of $\tau(e'_{q1}, a^*)$ and put $\alpha = \gamma\{\tau[e''_{q1}, b^*], \tau[e'_{q1}, a^*]\}$. From conditions a) and b) it follows that $\dfrac{\partial q_2(a^*)}{\partial s} = q_2^* + O(1)\,\omega(\delta)$ and $\alpha$

$= O(\gamma)\,\omega(\delta)$. We denote by $\beta_1$ and $\beta_2$ the values of the angles formed by the vectors $\tau\,[e''_{q_1}, b^*]$ and $\tau\,[e'_{q_1}, a^*]$ with the vector grad $[q_2(b^*)]$. We have

$$\left| q_2^* - \frac{\partial q_2(b^*)}{\partial s} \right| = |\operatorname{grad}[q_2(b^*)]|\;|\cos\beta_2 - \cos\beta_1| = O(\gamma)\,\alpha$$
$$= O(\gamma)\,\omega(\delta).$$

Thus,

$$\frac{\partial q_2(a^*)}{\partial s} = q_2^* + O(1)\,\omega(\delta) = \frac{\partial q_2(b^*)}{\partial s}$$
$$+ O(1)\left\{ \left| q_2^* - \frac{\partial q_2(b^*)}{\partial s} \right| + \omega(\delta) \right\} = \frac{\partial q_2(b^*)}{\partial s} + O(\gamma)\,\omega(\delta).$$

Consequently,

$$h_1([b', b'']) = h_1([a', a''])\,\frac{\partial q_2(a^*)}{\partial s}\left(\frac{\partial q_2(b^*)}{\partial s}\right)^{-1}$$
$$= h_1([a', a''])\left(1 + O(\gamma)\,\omega(\delta)\left(\frac{\partial q_2(b^*)}{\partial s}\right)^{-1}\right)$$
$$= h_1([a', a''])\,(1 + O(\gamma)\,\omega(\gamma)),$$

since by virtue of b) $\dfrac{\partial q_2(b^*)}{\partial s} > |\operatorname{grad}[q_2(b^*)]|\sin\gamma$. This, proves the lemma.

LEMMA 4.2.2. *Let* $q_m(x, y)$ $(m = 1, 2, \dots, N)$ *be continuously differentiable functions. In any region* $D$ *we can find a subregion* $G \subset D$, *determine a constant* $\gamma > 0$, *and renumber the functions* $\{q_m(x, y)\}$ *with two indices so that the functions*

$$q_i^k(x, y) = q_m(x, y) \quad (i = 0, 1, 2, \dots, n;\; k = 1, 2, \dots, m_i;\; \sum_{i=0}^{n} m_i = N)$$

*obtained after the renumbering satisfy the following conditions:*

(1) *when* $i = 0$, $q_i^k = \text{const}$ *in* $G$, *and when* $i > 0$, $\gamma \leqslant |\operatorname{grad}[q_i^k(x, y)]| \leqslant \frac{1}{\gamma}$ *for every point* $(x, y) \in G$;

(2) *the functions* $q_i^k(x, y)$ $(i > 0$ *fixed,* $k = 1, 2, \dots, m_i)$ *have in the region* $G$ *identical sets of level curves, more precisely, in the region* $G$, $q_i^k(x, y) \equiv \varphi_i^{k,l}(q_i^l(x, y))$, *where* $\varphi_i^{k,l}(t)$ *is a strictly monotonic continuously differentiable function of* $t$;

(3) *when* $i \neq j\,(i, j \neq 0)$, *then for any* $k$ *and* $l$ *the absolute value of the acute angle formed by the level curves of the functions* $q_i^k(x, y)$ *and* $q_j^l(x, y)$ *which pass through an arbitrary point* $(x, y) \in G$ *is greater than* $\gamma$.

*Proof.* By the continuity of the partial derivatives of the functions $\{ q_m(x, y) \}$ there exists a subregion $G^* \subset D$ inside which for any function $q_m(x, y)$ either grad $q_m(x, y) \equiv 0$ or $|$ grad $q_m(x, y) |$ is greater than some positive constant. From the continuity of the partial derivatives of the functions $\{ q_m(x, y) \}$ it follows also that there exists a subregion $G^{**} \subset G^*$ inside which for any pair of functions $q_r(x, y)$ and $q_s(x, y)$ one of two conditions holds: either $D\left(\frac{q_r, q_s}{x, y}\right) \equiv 0$ in $G^{**}$, or for every point of $G^{**}$ the level curves of $q_r(x, y)$ and $q_s(x, y)$ that pass through this point intersect at a non-zero angle $\left(D\left(\frac{q_r, q_s}{x, y}\right) \neq 0 \text{ in } G^{**}\right)$. From the implicit function theorem it follows that there exists a subregion $G \subset G^{**}$ in which condition (2) is satisfied for every pair of functions $q_r(x, y)$ and $q_s(x, y)$ with gradients different from zero and with determinant $D\left(\frac{q_r, q_s}{x, y}\right) \equiv 0$.

We now renumber the functions $\{ q_m(x, y) \}$ with two indices in such a way that only functions constant in $G$ have lower index zero, and the same lower index is assigned to those functions whose level curves coincide identically in $G$. This proves the lemma.

We consider in the region $G$ a superposition of the form $\sum\limits_{i=0}^{n} \sum\limits_{k=1}^{m_i} p_r(x, y)$ $f_i^k(q_i^k(x, y))$, where $\{ f_i^k(t) \}$ are continuous functions of one variable, $\{ p_i^k(x, y) \}$ are continuous functions satisfying in $G$ the condition $| p_i^k(x, y) | \leqslant \frac{1}{\gamma}$ and $\{ q_i^k(x, y) \}$ are continuously differentiable functions satisfying in $G$ conditions (1), (2), (3) of Lemma 4.2.2. Let $\omega(\delta)$ be the common modulus of continuity in $G$ of the functions $\left\{ p_i^k(x, y)\,;\ \frac{\partial q_i^k(x, y)}{\partial x}\,;\ \frac{\partial q_i^k(x, y)}{\partial y} \right\}$. Let $[a', a'']$ and $[b', b'']$ be segments of the level curves of the functions $\{ q_i^k(x, y) \}$ ($i > 0$ fixed) lying in $G$. Let

$$\alpha = h_1([a', a''])\,;\ \delta = \rho([a', a''],\ [b', b''])\,;$$

$$\varepsilon = \sup \left| \sum_{i=0}^{n} \sum_{k=1}^{m_i} p_i^k(x, y) f_i^k(q_i^k(x, y)) \right|;$$

$$m = \max_{i,k} \sup \left| f_i^k (q_i^k (x, y)) \right| ,$$

where sup is taken over all points $(x, y) \in [a', a''] \cup [b', b'']$.

LEMMA 4.2.3. *If* $\delta$ *is sufficiently small* $(\omega(\delta) \leqslant C_2(\gamma))$, *then for any* $i > 0$

$$\left| \int\limits_{s \in [a', a'']} \sum_{k=1}^{m_i} p_i^k (s) f_i^k (q_i^k (s))\, ds - \int\limits_{s \in [b', b'']} \sum_{k=1}^{m_i} p_i^k (s) f_i^k (q_i^k (s))\, ds \right|$$
$$\leqslant C_3(\gamma)(\alpha\varepsilon + m\alpha\omega(\delta) + m\delta),$$

*where the constants* $C_2(\gamma)$, $C_3(\gamma)$ *depend only on* $\gamma$.

*Proof.* By (1), (2), (3) there exists a sufficiently small constant $C_2(\gamma)$ and a sufficiently large constant $C_3(\gamma)$ such that if $\omega(\delta) \leqslant C_2(\gamma)$ and for a point $a \in [a', a'']$ the inequalities $h_1([a', a]) \geqslant C_3(\gamma)\,\delta$; $h_1([a, a'']) \geqslant C_3(\gamma)\delta$ are satisfied, then for any $j \neq i$ $(j > 0)$ the level curve of the function $q_j^k$ that passes through $a$ intersects $[b', b'']$ of the level curve of $q_i^k$. Suppose that $\alpha > 2C_3(\gamma)\,\delta$ (if $\alpha \leqslant 2\, C_3(\gamma)\,\delta$, then the assertion of the lemma is trivial) and suppose that the segment $[\tilde{a}', \tilde{a}'']$ of the level curve of $q_i^k$ is such that $[\tilde{a}', \tilde{a}''] \subset [a', a'']$ and $h_1([a', \tilde{a}']) = h_1([\tilde{a}'', a'']) = C_3(\gamma)\,\delta$. On the arc $[\tilde{a}', \tilde{a}'']$ we fix a system of points $a_1, a_2, \ldots, a_\nu$ $(\tilde{a}' = a_1,\ \tilde{a}'' = a_\nu)$, uniformly distributed along the length of this arc, and denote by $b_r$ the point of intersection of $[b', b'']$ with the level curve of $q_j^k$ that passes through $a_r$ (here $j \neq i$ should for the time being be regarded as fixed). Using Lemma 4.2.1 we have

$$\left| \int\limits_{s \in [a', a'']} p_j^k (s) f_j^k (q_j^k (s))\, ds - \int\limits_{s \in [b', b'']} p_j^k (s) f_j^k (q_j^k (s))\, ds \right|$$

$$= \left| \int\limits_{s \in [a_1, a_\nu]} p_j^k (s) f_j^k (q_j^k (s))\, ds - \int\limits_{s \in [b_1, b_\nu]} p_j^k (s) f_j^k (q_j^k (s))\, ds \right|$$

$$+ O(\gamma)\, m\delta$$

$$= \lim_{\nu \to \infty} \Big| \sum_{r=1}^{\nu} p_j^k (a_r) f_j^k (q_j^k (a_r))\, h_1([a_r, a_{r+1}])$$

$$- \sum_{r=1}^{\nu} p_j^k (b_r) f_j^k (q_j^k (b_r))\, h_1([b_r, b_{r+1}]) \Big| + O(\gamma)\, m\delta$$

$$= \lim_{\nu\to\infty} \Big| \sum_{r=1}^{\nu} p_j^k(a_r) f_j^k(q_j^k(a_r))\, h_1([a_r, a_{r+1}])$$

$$- \sum_{r=1}^{\nu} p_j^k(a_r) f_j^k(q_j^k(a_r))\, h_1([a_r, a_{r+1}])\,(1 + O(\gamma)\,\omega(\delta))$$

$$+ \sum_{r=1}^{\nu} (p_j^k(a_r) - p_j^k(b_r)) f_j^k(q_j^k(a_r))\, h_1([b_r, b_{r+1}]) \Big| + O(\gamma)\, m\delta$$

$$= \lim_{\nu\to\infty} \Big| \sum_{r=1}^{\nu} p_j^k(a_r) f_j^k(q_j^k(a_r))\, h_1([a_r, a_{r+1}])\, O(\gamma)\,\omega(\delta)$$

$$+ \sum_{r=1}^{\nu} f_j^k(q_j^k(a_r))\, h_1([b_r, b_{r+1}])\, O(\gamma)\,\omega(\delta) \Big| + O(\gamma)\, m\delta$$

$$= O(\gamma)\, m\alpha\omega(\delta) + O(\gamma)\, m\alpha\omega(\delta) + O(\gamma)\, m\delta = O(\gamma)\, m\,(\delta + \alpha\omega(\delta)).$$

Then

$$\Big| \int\limits_{s\in[a',a'']} \sum_{k=1}^{m_i} p_i^k(s) f_i^k(q_i^k(s))\, ds - \int\limits_{s\in[b',b'']} \sum_{k=1}^{m_i} p_i^k(s) f_i^k(q_i^k(s))\, ds$$

$$\leqslant \Big| \int\limits_{s\in[a',a'']} \sum_{i=0}^{n} \sum_{k=1}^{m_i} p_i^k(s) f_i^k(q_i^k(s))\, ds - \int\limits_{s\in[b',b'']} \sum_{i=0}^{n} \sum_{k=1}^{m_i} p_i^k(s) f_i^k(q_i^k(s))\, ds \Big|$$

$$+ \sum_{j\neq i} \Big| \int\limits_{s\in[a',a'']} \sum_{k=1}^{m_j} p_j^k(s) f_j^k(q_j^k(s))\, ds - \int\limits_{s\in[b',b'']} \sum_{k=1}^{m_j} p_j^k(s) f_j^k(q_j^k(s))\, ds \Big|$$

$$\leqslant C_4(\gamma)\,\alpha\varepsilon + n\,(\max_{j\neq i}\, m_j)\, C_5(\gamma)\, m\,(\delta + \alpha\omega(\delta))$$

$$\leqslant C_3(\gamma)\,(\alpha\varepsilon + m\delta + m\alpha\omega(\delta)).$$

This proves the lemma.

## § 3. *Deletion of dependent terms*

On a bounded closed set $D$ we consider the space of linear superpositions of the form $\sum_{k=1}^{m} p_k(x, y) f_k(q(x, y))$, $(x, y) \in D$. Here the functions $\{p_k(x, y)\}$ and $q(x, y)$ are continuous and fixed, and $\{f_k(t)\}$ are arbitrary continuous functions of one variable. We assume that the function $q(x, y)$ is such that for any sequence $t_n \in q(D) \to t \in q(D)$ we have $\rho\,[e(q, t_n) \cap D, e(q, t) \cap D] \to 0$. We put

$$\lambda(t, D, q, p_1, \dots, p_m) = \inf_{\{c_k\}} \sup_{(x,y)\in e(q,t)\cap D} \Big| \sum_{k=1}^{m} c_k p_k(x, y) \Big|,$$

where inf is taken over all sets of numbers $\{ c_k \}$ for which $\max\limits_k | c_k | = 1$. The function $\lambda (t, D, q, \{ p_k \})$, as a function of $t$, is defined only on the set $q(D)$.

LEMMA 4.3.1. *The function* $\lambda (t, D, q, \{ p_k \})$ *depends continuously on* $t$.

*Proof.* The linear combinations $\sum\limits_{k=1}^{m} c_k p_k (x, y)$ for all possible systems of numbers $\{ c_k \}$ for which $\max\limits_k | c_k | \leqslant 1$, form an equicontinuous set of functions, considered on the bounded closed set $D$. Consequently, for any $\varepsilon > 0$ there is a $\delta > 0$ such that if $| t_1 - t_2 | < \delta$, then

$$\Big| \sup_{(x, y) \in e(q, t_1)} \Big| \sum_{k=1}^{m} c_k p_k (x, y) \Big| - \sup_{(x, y) \in e(q, t_2)} \Big| \sum_{k=1}^{m} c_k p_k (x, y) \Big\| < \varepsilon$$

simultaneously for all systems of numbers $\{ c_k \}$ such that $\max\limits_k | c_k | \leqslant 1$. For definiteness, suppose that $\lambda (t_2, D, q, \{ p_k \}) \geqslant \lambda (t_1, D, q, \{ p_k \})$. Since the expression $\sup\limits_{(x, y) \in e(q, t_1)} \Big| \sum\limits_{k=1}^{m} c_k p_k (x, y) \Big|$ depends continuously on the coefficients $\{ c_k \}$, there exists a system of numbers $\{ c_k^1 \}$ such that $\max\limits_k | c_k^1 | = 1$ and

$$\lambda (t_1, D, q, \{ p_k \}) = \sup_{(x, y) \in e(q, t_1)} \Big| \sum_{k=1}^{m} c_k^1 p_k (x, y) \Big| .$$

Since

$$\lambda (t_2, D, q, \{ p_k \}) \leqslant \sup_{(x, y) \in e(q, t_2)} \Big| \sum_{k=1}^{m} c_k^1 p_k (x, y) \Big| ,$$

we have

$$0 \leqslant \lambda (t_2) - \lambda (t_1) \leqslant \sup_{(x, y) \in e(q, t_2)} \Big| \sum_{k=1}^{m} c_k^1 p_k (x, y) \Big|$$

$$- \sup_{(x, y) \in e(q, t_1)} \Big| \sum_{k=1}^{m} c_k^1 p_k (x, y) \Big| < \varepsilon .$$

This proves the lemma.

LEMMA 4.3.2. *The function* $\lambda (t, D, q, \{ p_k \})$ *depends continuously on* $D$ *in the sense that there exists a function* $\mu (\varepsilon) \to 0$ *as* $\varepsilon \to 0$, *having the property: if the set* $D_\varepsilon \subset D$ *is such that, for any* $t$, $D_\varepsilon \cap e (q, t)$ *forms an* $\varepsilon$*-net in the set* $e (q, t) \cap D$, *then*

$$\max_{t \in q(D)} \left| \lambda(t, D, q, \{p_k\}) - \lambda(t, D_\varepsilon, q, \{p_k\}) \right| \leqslant \mu(\varepsilon) .$$

*Proof.* Using the equicontinuity of the set of functions $\sum_{k=1}^{n} c_k\, p_k\,(x, y)$ where $\max_k |c_k| \leqslant 1$, we conclude that there exists a function $\mu(\varepsilon) \to 0$ as $\varepsilon \to 0$ such that the inequality

$$0 \leqslant \sup_{(x,y) \in e(q,t) \cap D} \left| \sum_{k=1}^{m} c_k p_k(x, y) \right| - \sup_{(x,y) \in e(q,t) \cap D_\varepsilon} \left| \sum_{k=1}^{m} c_k p_k(x, y) \right| \leqslant \mu(\varepsilon).$$

uniformly over all $t \in q(D)$ and over all systems of numbers $\{c_k\}$ for which $\max_k |c_k| \leqslant 1$. For any $\varepsilon > 0$ there exists a system of numbers $\{c_k^\varepsilon\}$ such that $\max_k |c_k^\varepsilon| = 1$ and

$$\lambda(t, D_\varepsilon, q, \{p_k\}) = \sup_{(x,y) \in e(q,t) \cap D_\varepsilon} \left| \sum_{k=1}^{m} c_k^\varepsilon p_k(x, y) \right| .$$

Since for any $\varepsilon$

$$\lambda(t, D, q, \{p_k\}) \leqslant \sup_{(x,y) \in e(q,t) \cap D} \left| \sum_{k=1}^{m} c_k^\varepsilon p_k(x, y) \right|$$

and, on the other hand, $\lambda(t, D, q, \{p_k\}) \geqslant \lambda(t, D_\varepsilon, q, \{p_k\})$ (we recall that $D_\varepsilon \subset D$), we have

$$0 \leqslant \lambda(t, D, q, \{p_k\}) - \lambda(t, D_\varepsilon, q, \{p_k\}) \leqslant \sup_{(x,y) \in e(q,t) \cap D} \left| \sum_{k=1}^{m} c_k^\varepsilon p_k(x, y) \right|$$
$$- \sup_{(x,y) \in e(q,t) \cap D_\varepsilon} \left| \sum_{k=1}^{m} c_k^\varepsilon p_k(x, y) \right| < \mu(\varepsilon) .$$

This proves the lemma.

LEMMA 4.3.3. *Let $F$ be a closed set on the $t$-axis; $F \subset q(D)$. For every $t \in F$, suppose that there exists one and only one system of numbers $\{C_k\}$ ($\max_k |C_k| = 1$) such that $\sum_{k=1}^{m} C_k p_k(x, y) \equiv 0$ on the set $e(q, t) \cap D$. Then each of the functions $\{C_k(t)\}$ depends continuously on $t$ on the set $F$.*

*Proof.* Suppose that $t_n \in F$, $t \in F$ and $t_n \to t$. We put $\overline{\lim_{n\to\infty}}\, C_k(t_n) = \tilde{C}_k$ and $\underline{\lim_{n\to\infty}}\, C_k(t_n) = \approx{C}_k$. Since $\sum_{k=1}^{m} C_k(t_n)\, p_k(x, y) \equiv 0$ on the set $e(q, t_n) \cap D$ and $\rho\,[e(q, t) \cap D, e(q, t_n) \cap D] \to 0$ as $n \to \infty$, we have $\sum_{k=1}^{m} \tilde{C}_k\, p_k(x, y)$

$\equiv 0 \equiv \sum_{k=1}^{m} \approx{C}_k p_k(x, y)$ on the set $e(q, t) \cap D$. Consequently, by the condition of the lemma, $\tilde{C}_k = \approx{C}_k = C_k(t)$. This proves the lemma.

LEMMA 4.3.4. *Suppose that* $\lambda(t, D, q, \{p_k\}) \equiv 0$ *on some non-empty portion* $\delta$ *of the set* $q(D)$. *Then there is a non-empty portion* $\delta^* \subset \delta$ *and an index* $l$ *such that for any continuous functions* $\{f_k(t)\}$ *there are continuous functions* $\{f_k^*(t)\}$ *such that*

$$\sum_{k \neq l} f_k^*(q(x, y))\, p_k(x, y) = \sum_{k=1}^{m} f_k(q(x, y))\, p_k(x, y)$$

*on the set* $q^{-1}(\delta^*) \cap D$.

We recall that a portion $\delta$ of a set $E$ is that part of it which lies in the interval $\delta$.

*Proof.* We prove the lemma by induction on $m$. For $m = 1$ the assertion of the lemma is obvious. We denote by $\delta_k$ the set of all points $t$ of the portion $\delta$ for which $\lambda(t, D, q, p_1, \ldots, p_{k-1}, p_{k+1}, \ldots, p_m) = 0$. By Lemma 4.3.1, the set is closed. Two cases are possible.

1) For some $k$ the set $\delta_k$ contains a non-empty portion $\delta_k'$ of the set $q(D)$. Since $\lambda(t, D, q, p_1, \ldots, p_{k-1}, p_{k+1}, \ldots, p_m) = 0$ for every $t \in \delta_k'$, then by the inductive hypothesis there is a non-empty portion $\delta^* \subset \delta_k'$ and an index $l \neq k$ such that for any continuous functions $f_1(t), \ldots, f_{k-1}(t)$, $f_{k+1}(t), \ldots, f_m(t)$ there are continuous functions $f_1^*(t), \ldots, f_{k-1}^*(t), f_{k+1}^*(t), \ldots, f_m^*(t)$ such that

$$\sum_{i \neq k} f_i(q(x, y))\, p_i(x, y) = \sum_{i \neq k, l} f_i^*(q(x, y))\, p_i(x, y).$$

on the set $q^{-1}(\delta^*) \cap D$. Putting $f_k^*(t) = f_k(t)$, we obtain

$$\sum_{i=1}^{m} f_i(q(x, y))\, p_i(x, y) = \sum_{i \neq l} f_i^*(q(x, y))\, p_i(x, y).$$

So in case 1) the lemma is proved.

2) None of the sets $\delta_k$ contains non-empty portions of the set $q(D)$, that is, $\bigcup_{k=1}^{m} \delta_k$ is nowhere dense in $q(D)$. Therefore there exists a non-empty portion $\delta^* \subset \delta \setminus \bigcup_{k=1}^{m} \delta_k$. Since $\lambda(t, D, q, \{p_k\}) \equiv 0$ on $\delta^*$, for every $t \in \delta^*$ there are numbers $\{C_k(t)\}$ $(\max_k |C_k(t)| = 1)$ such that $\sum_{k=1}^{m} C_k$

$(q(x, y))\, p_k(x, y) \equiv 0$ on $e(q, t) \cap D$. If we had $C_k(t) = 0$ for some $k$, then it would turn out that $t \in \delta_k$. Consequently, $C_k(t) \neq 0$ for any $k$. We show that for every $t \in \delta^*$ the numbers $\{ C_k(t) \}$ are uniquely determined. Assume the contrary. Then there are numbers $\{ C_k'(t) \}$ $(\max | C_k'(t) | = 1)$ such that $\sum_{k=1}^{m} C_k'(q(x, y))\, p_k(x, y) = 0$ on $e(q, t) \cap D$ and $C_k \neq C_k'$ for some $k$. Then

$$\sum_{k \neq 1} [C_k(t)\, C_1'(t) - C_k'(t)\, C_1(t)]\, p_k(x, y) = \sum_{k \neq 1} C_k''(t)\, p_k(x, y) \equiv 0$$

on $e(q, t) \cap D$ and in addition, $C_k'' \neq 0$ for some $k$. Consequently, $t \in \delta_1$. So we have obtained a contradiction, and the uniqueness of the choice of the numbers $C_k(t)$ is proved. Further, we may regard $\{ C_k(t) \}$ as single-valued functions of $t$ on the portion $\delta^*$. By Lemma 4.3.3, the functions $C_k(t)$ are continuous and, as noted above, $C_k(t) \neq 0$ for any $t \in \delta^*$. Then

$$p_1(x, y) = \sum_{k=2}^{m} - \frac{C_k(q(x, y))}{C_1(q(x, y))}\, p_k(x, y)\,, \quad (x, y) \in q^{-1}(\delta^*) \cap D\,.$$

Putting $f(t) = f_k(t) - \dfrac{C_k(t)}{C_1(t)} f_1(t)$, $t \in \delta^*$, we have $\sum_{k=2}^{m} f_k^*(q(x, y))\, p_k(x, y)$

$$= \sum_{k=1}^{m} f_k(q)\, p_k(x, y) - \sum_{k=2}^{m} \frac{C_k(q)}{C_1(q)}\, p_k(x, y)$$

$$= \sum_{k=2}^{m} f_k(q)\, p_k(x, y) + f_1(q)\, p_1(x, y)$$

$$= \sum_{k=1}^{m} f_k(q(x, y))\, p_k(x, y)\,, \quad (x, y) \in q^{-1}(\delta^*) \cap D\,.$$

This proves the lemma.

## § 4. *Reduction of linear superpositions to a form with independent terms*

We fix the continuous functions $p_i^k(x, y)$ and continuously differentiable functions $q_i(x, y)$ $(i = 0, 1, 2, \ldots, n;\ k = 1, 2, \ldots, m_i)$ $n \geqslant 2$, where $\{ q_i(x, y) \}$ satisfy in $D$ conditions (1) and (3) of Lemma 4.2.2, and we consider in $D$ superpositions of the form

$$\sum_{i=0}^{n} \sum_{k=1}^{m_i} p_i^k(x, y)\, f_i^k(q_i(x, y))\,,$$

where $\{ f_i^k(t) \}$ are arbitrary continuous functions of one variable.

We call a bounded closed region $G \subset D$ polyhedral if the boundary of $G$ consists of a finite number of mutually non-intersecting simple closed contours that are unions of a finite number of segments of level curves of the functions $q_i(x, y)$ $(i=1, 2, \ldots, n)$. Let $G \subset D$ be a polyhedral region. We denote by $\Gamma_i$ the set of those $t \in q_i(G)$ for which the set $e(q_i, t) \cap G$ contains a segment of a level curve belonging to the boundary of $G$. For any $i$ the set $\Gamma_i$ consists of a finite number of points. By property (1) of the functions $\{ q_i(x, y) \}$ for every $i$ and for all points $t_0 \in q_i(G) \backslash \Gamma_i$ there exists $\lim\limits_{t \to t_0} e(q_i, t) \rightleftharpoons e(q_0, t_0)$. If $t_0 \in \Gamma_i$, then the last assertion need not hold, but in any case there exists $\lim\limits_{t \to +t_0} e(q_i, t) \subset e(q_i, t_0)$ and $\lim\limits_{t \to -t_0} e(q_i, t) \subset e(q_i, t_0)$ where the limit is taken over the points $t \in q_i(G)$. Here the limit is understood in the sense of the distance $\rho\left(e(q_i, t), e(q_i, t_0)\right)$.

LEMMA 4.4.1. *There is a region* $G \subset D$ *and a system of numbers* $\tau_i^k = 0$ *or* $1$ $(i=0, 1, 2, \ldots, n;\ k=1, 2, \ldots, m_i)$ *such that*

(4) *for any* $i$ *and for any continuous functions* $\{ \varphi_i^k(t) \}$ *there exist continuous functions* $\{ f_i^k(t) \}$ *such that in* $G$

$$\sum_{k=1}^{m_i} p_i^k(x, y)\, \varphi_i^k(q_i(x, y)) \equiv \sum_{k=1}^{m_i} \tau_i^k p_i^k(x, y) f_i^k(q_i(x, y))\,;$$

(5*) *for any polyhedral region* $G^* \subset G$ *and any* $i$, *the set*

$$\{ t : \lambda(t, G^*, q_i, p_i^{k_1}, \ldots, p_i^{k_s}) = 0 \}$$

*is nowhere dense in* $q_i(G^*)$, *where*

$$k_1 = k_1(i), k_2 = k_2(i), \ldots, k_s = k_s(i)$$

*is the set of all values of* $k$ *for which* $\tau_i^k = 1$.

*Proof.* If $i = 0$, then by (1) the set $q_0(D)$ consists of only one point. We choose a region $G_0 \subset D$ and number $\tau_0^k$ $(k=1, 2, \ldots, m_0)$ such that in $G_0$ the functions $p_0^{k_1}, \ldots, p_0^{k_s}$ are a basis for the linear hull of the functions $\{ p_0^k \}$ (condition (4) for $i=0$) and in any region $G^* \subset G_0$ these functions are linearly independent (condition (5*) for $i=0$). Let $G^* \subset D$ be an arbitrary polyhedral region. Then $\lambda(t, G^*, q, \{ p_i^k \})$ as a function of $t$ has, for any $i > 0$, a finite number of points of discontinuity (of the first kind) on the set $q_i(G^*)$, which consists of a finite number of segments (see Lemma 4.3.1). Hence it follows that if the set $\{ t : \lambda(t, G^*, q_i, \{ p_i^k \}) = 0 \}$ is not nowhere dense on $q_i(G^*)$, then the function $\lambda(t) \equiv 0$ on some segment $\delta \subset q_i(G^*)$ not containing points of $\Gamma_i$. By Lemma 4.3.4, there is a segment $\delta^* \subset \delta$ such that in the expression $\sum_{k=1}^{m_i} p_i^k(x, y) f_i^k(q_i(x, y))$ one of the terms can be deleted, without narrowing the class of the functions representable in the region $q^{-1}(\delta^*) \cap G^*$ as superpositions of the given form. Carrying out all possible deletions we can find a region $G \subset G_0 \subset D$ for which the assertion of the lemma is satisfied.

A region $G \subset D$ is called regular if, firstly, it is polyhedral and, secondly, there is a number $\gamma_G > 0$ such that for every $i > 0$ and every $t \in q_i(G)$ the set $e(q_i, t) \cap G$ is the union of a finite number of simple arcs, each of which has length not less than $\gamma_G$. A point $A$ of the boundary of the polyhedral region $G$ is called a vertex if it belongs simultaneously to two segments of the level curves of $q_i(x, y)$ and $q_j(x, y)$ $(i \neq j)$ on the boundary of $G$. Every polyhedral region has a finite number of vertices.

LEMMA 4.4.2. *For every polyhedral region $G$ and every neighbourhood $U$ of the vertices of this region we can construct a regular region $G^* \subset G$ such that $G \backslash U \subset G^*$.*

*Proof.* Let $A_1, A_2, \ldots, A_r$ be the vertices of the polyhedral region $G$; $U_1, U_2, \ldots, U_r$ suitably small neighbourhoods of these vertices. Let $k_m = k_m(A_m)$ be the number of all those functions $\{q_i(x, y)\}$ for each of which the level curve passing through the point $A_m$ does not contain any other points of the set $U_m \cap G$. Let $q_{im}(x, y)$ be one of these functions. We put $k(G) \in q_i(G)$. If $k(G) = 0$, then for any $i$ and any $t \in q_i(G)$ the length of any component of the set $e(q_i, t) \cap G$ is greater than zero and consequently the region $G$ is regular. Suppose that $k(G) > 0$ and $m$ such that $k_m \neq 0$.

We fix $\varepsilon > 0$ and put

$$G_{1m}^* = G \,|\, \{(x, y): |q_{i_m}(x, y) - q(A_m)| < \varepsilon\} \cap U_m .$$

If $U_m$ and $\varepsilon$ are sufficiently small, then inside $U_m$ the region $G_{1m}^*$ has two vertices $A_m'$ and $A_m''$, while the region $G$ has only one vertex $A_m$ there, but $k_m(A_m') = k_m(A_m'') = k_m(A_m) - 1$. We now put $G_1^* = \cap\, G_{1m}^*$, where the intersection is taken over all $m$ such that $k_m \neq 0$. Then $k(G_1^*) = k(G) - 1$. Repeating this construction $k(G)$ times, we obtain a polyhedral region $G^*$ for which $G \backslash G^* \subset U$ and $k(G^*) = 0$. Consequently, $G^*$ is regular. This proves the lemma.

LEMMA 4.4.3. *There exists a set* $G \subset D$, *a number* $\lambda > 0$, *and a set of numbers* $\tau_i^k = 0$ *or* $1$ $(i = 0, 1, \ldots, n;\ k = 1, 2, \ldots, m_i)$ *such that condition* (4) *of Lemma* 4.4.1 *is satisfied, and also the conditions*

(5) *for every* $i$ *and* $t \in q_i(G)$ *and for any functions* $\{f_i^k(t)\}$

$$\max_{(x,y)\in e(q_i,t)\cap G} \left| \sum_{k=1}^{m_i} \tau_i^k p_i^k(x,y) f_i^k(q_i(x,y)) \right| \geqslant \lambda \max_k \left| \tau_i^k f_i^k(t) \right|;$$

(6) $G$ *is a regular region.*

*Proof.* By Lemma 4.4.1 there exists a region $G^* \subset D$ and a set of numbers $\tau_i^k$ such that for every polyhedral subregion $G^{**} \subset G^*$ and for every $i$ the set $\{t : \lambda(t, G^{**}, q_i, p_i^{k_1}, \ldots, p_i^{k_s}) = 0\}$ is nowhere dense in $q_i(G^{**})$, where $k_1, k_2, \ldots, k_s$ is the set of all values of $k$ for which $\tau_i^k = 1$; moreover, on the set $G^*$, for any $i$ the property (4) of Lemma 4.4.1 is satisfied. In order not to change the notation unnecessarily, we assume that all $\tau_i^k = 1$. We now construct a system of regular regions $G_0 \supset G_1 \supset G_2 \supset \ldots \supset G_n = G$, having the following property: for every $j \leqslant i$, $\inf_{t \in q_j(G_i)} \lambda(t, G_i, q_j, \{p_j^k\}) \geqslant \lambda_i > 0$. For $G_0$ we choose any regular region $G_0 \in G^*$. Suppose that the regular regions $G_0, G_1, \ldots, G_{i-1}$ have been constructed. We now construct the set $G_i$. We denote by $\alpha_\delta$ the set $\{t : \lambda(t, q_i, G_{i-1}, \{p_i^k\}) > \delta\}$. Since the functions $\lambda(t, q_i, G_{i-1}, \{p_i^k\})$, have only finitely many points of discontinuity (of the first kind) on the set $q_i(G_{i-1})$, which consists of a finite number of segments (see Lemma 4.3.1), any component of $\alpha_\delta$ is either an interval, or a half-interval, or a segment, or a point. Suppose that the set $\alpha_\delta^N \subset \alpha_\delta$ consists of the $N$ longest components of non-zero length of the set $\alpha_\delta$ (if $\alpha_\delta$ has only $N_0$ $(< N)$ components of non-zero length, then let $\alpha_\delta^N = \alpha_\delta^{N_0}$). We denote by $\bar{\alpha}_\delta^N$ the closure of the set $\alpha_\delta^N$. We put $G_{i-1}^* = G_{i-1} \cap q_i^{-1}(\bar{\alpha}_\delta^N)$. We fix $\varepsilon > 0$. Since $G_{i-1}$ is regular, for every $j$ the length of any component of $e(q_j, t) \cap G_{i-1}$ is greater than $\gamma_G > 0$. And since the set $\{t : \lambda(t, q, G_{i-1}, \{p_i^k\}) = 0\}$ is nowhere dense in $q_i(G_{i-1})$, for sufficiently small $\delta$ and sufficiently large $N$ the set $G_{i-1}^*$ forms a $\varepsilon/2$-net on every set $e(q_j, t) \cap G_{i-1}$, $j < i$. The set $G_{i-1}^*$ is a polyhedral region. We denote by $U(\varepsilon)$ the set of points $(x, y)$ each of which is at a distance of no more than $\varepsilon/4$ from one of the vertices of the set $G_{i-1}^*$. By Lemma 4.4.2 there exists a regular region $G_i \subset G_{i-1}^*$ such that $G_{i-1}^* \backslash G_i \subset U(\varepsilon)$. The set $G_i$ forms an $\varepsilon$-net on every set $e(q_j, t) \cap G_{i-1}, j < i$ and forms an $\varepsilon/2$-net on every set $e(q_i, t) \cap G_{i-1}^*$. By Lemma 4.3.2, for sufficiently small $\varepsilon$,

$$\lambda_i = \min_{j \leq i} \inf_{t \in q_j(G_i)} \lambda(t, G_i, q_j, \{p_i^k\}) > \frac{1}{2}\min\left\{\frac{\delta}{2}, \min_{j<i} \lambda_j\right\}.$$

Thus, the regular regions $G_1, G_2, \ldots, G_n$ can be constructed. The regular region $G = G_n$ satisfies all the requirements of our lemma ($\lambda = \lambda_n$), which is now proved.

## § 5. *The set of linear superpositions in the space of continuous functions is closed*

THEOREM 4.5.1. *Suppose that continuous functions $p_m(x, y)$ and continuously differentiable functions $q_m(x, y)$ $(m = 1, 2, \ldots, N)$ are fixed. Then in any region $D$ of the plane of the variables $x, y$. there exists a closed subregion $G \subset D$ such that the set of superpositions of the form*

$$\sum_{m=1}^{N} p_m(x, y) f_m(q_m(x, y)),$$

*where $\{f_m(t)\}$ are arbitrary continuous functions, is closed (in the uniform metric) in the set of all functions continuous on the set $G$.*

By Lemma 4.2.2 and 4.4.3 we can find a subset $G \subset D$, determine constants $\gamma > 0$ and $\lambda > 0$, and renumber the functions $\{p_m(x, y)\}$ and $\{q_m(x, y)\}$ with two indices so that the functions obtained after the renumbering, $\{p_i^k(x, y)\}$ and $\{q_i^k(x, y)\}$ $(i = 0, 1, 2, \ldots, n;\ k = 1, 2, \ldots, m_i;\ \sum_{i=0}^{n} m_i \leqslant N)$ that is, some functions may be omitted in the renumbering) satisfy conditions (1), (2), (3) of Lemma 4.2.2, and also the conditions:

(4′) for any continuous functions $\{f_m(t)\}$ there exists continuous functions $\{f_i^k(t)\}$ such that on $G$

$$\sum_{m=1}^{N} p_m(x, y) f_m(q_m(x, y)) = \sum_{i=0}^{n} \sum_{k=1}^{m_i} p_i^k(x, y) f_i^k(q_i^k(x, y));$$

(5′) for every $i$ and $t \in q_i^1(G)$ and for any functions $\{f_i^k(t)\}$

$$\max_{(x,y) \in e\left(q_i^1, t\right) \cap G} \left| \sum_{k=1}^{m_i} p_i^k(x, y) f_i^k(q_i^1(x, y)) \right| \leqslant \lambda \max_k |f_i^k(t)|;$$

(6′) $G$ is a regular region with respect to the functions $\{q_i^k(x, y)\}$.

LEMMA 4.5.1. *In the sets* $\{ q_i^1 (G) \}$ *we can select subsets consisting of a finite number of points* $t_{i,j} \in q_i^1 (G)$ $(i = 0, 1, 2, ..., n;\ j = 1, 2, ..., s_i)$ *such that for any continuous functions* $\{ f_i^k (t) \}$

$$\max_{i,k} \max_{t \in q_i^1 (G)} |f_i^k (t)| \leqslant c \left( \max_{(x,y) \in G} \left| \sum_{i=0}^{n} \sum_{k=1}^{m_i} p_i^k (x, y) f_i^k (q_i^1 (x, y)) \right| + \max_{k} |f_i^k (t_{i,j})| \right),$$

*where* $C$ *is a constant not depending on the functions* $\{ f_i^k (t) \}$.

*Proof.* Since $G$ is polyhedral, for each $i$ we can choose in $q_i (G)$ a finite set of points $\{ t_{i,j} \}$ so dense that the components of the level curves $e (q_i^1, t_{i,j}) \cap G$ form a $\delta$-net in the set of all components of the level curves $e (q_i^1, t) \cap G$, $t \in q_i^1 (G)$. A sufficiently small $\delta$, not depending on the functions $\{ f_i^k (t) \}$, will be chosen below. We put

$$\mu = \max_{i,k} \max_{(x,y) \in G} |f_i^k (q_i^1 (x, y))|;$$

$$\varepsilon_1 = \max_{(x,y) \in G} \left| \sum_{i=0}^{n} \sum_{k=1}^{m_i} p_i^k (x, y) f_i^k (q_i^1 (x, y)) \right|; \quad \varepsilon_2 = \max_{k,i,j} |f_i^k (t_{i,j})|.$$

For definiteness, let $f_1^1 (q_1^1 (a)) = \mu$ at the point $a \in G$. By (5′) there exists a point $a' \in G$ such that $\left| \sum_{k=1}^{m_i} p_1^k (a') f_1^k (q_1^1 (a')) \right| \geqslant \lambda \mu$. Let $[a', a^*]$ be a segment of the level curve of the function $q_1^1 (x, y)$ with end-points at $a'$ and $a^*$ such that $h_1 ([a', a^*]) \geqslant \gamma G/2$ (see the definition of a regular region in § 4). On the arc $[a', a^*]$ we fix a point $a''$ such that $\omega (\alpha) \leqslant \dfrac{\lambda}{2m_1}$, where $\alpha = h_1 ([a', a''])$. Then on the segment $[a', a'']$ the function $\varphi_1 (x, y) = \sum_{k=1}^{m_1} p_1^k (x, y) f_1^k (q_1^1 (x, y))$ keeps'a constant sign and satisfies the inequality $|\varphi_1 (x, y)| \geqslant \lambda\mu/2$. In fact, $|\varphi_1 (a')| \geqslant \lambda\mu$ at the point $a'$, and for any point $s \in [a', a'']$

$$|\varphi_1 (s) - \varphi_1 (a')| = \left| \sum_{k=1}^{m_1} (p_1^k (s) - p_1^k (a')) f_1^k (a') \right| \leqslant m_1 \mu \omega (\alpha) \leqslant \frac{\lambda\mu}{2}.$$

Consequently,

$$\left| \int_{s \in [a', a'']} \varphi_1 (s)\, ds \right| \geqslant \frac{1}{2} \lambda \mu \alpha .$$

By construction there is an index $j$ and a segment $[b', b'']$ of the level curve $e (q_1', t_{1,j}) \cap G$ such that $\rho ([a', a''], [b', b'']) < \delta$. We have

$$\Big| \int_{s\in[b',\,b'']} \varphi_1(s)\,ds \Big| \leqslant c_1\varepsilon_2\beta\,,$$

where $\beta = h_1([b', b''])$, $C_1 = m_1 \max\limits_k \max\limits_{(x,y)\in G} |p_1^k(x, y)|$. And since $\alpha$ and $\beta$ are commensurable ($\delta$ will be chosen small in comparison with $\alpha$),

$$\Big| \int_{s\in[a',\,a'']} \varphi_1(s)\,ds - \int_{s\in[b',\,b'']} \varphi_1(s)\,ds \Big| \geqslant \frac{1}{2}\lambda\mu\alpha - c_1'\varepsilon_2\alpha\,.$$

By Lemma 4.2.3

$$\Big| \int_{s\in[a',\,a'']} \varphi_1(s)\,ds - \int_{s\in[b'b'']} \varphi_1(s)\,ds \Big| \leqslant c_3(\alpha\varepsilon_1 + \mu\alpha\omega(\delta) + \mu\delta)\,.$$

Thus, $c_3(\alpha\varepsilon_1 + \mu\alpha\omega(\delta) + \mu\delta) \geqslant \lambda\mu\alpha/2 - c_1'\alpha\cdot\varepsilon_2$. If $\delta$ is taken sufficiently small in comparison with $\alpha$ (in order that $c_3(\alpha\omega(\delta)+\delta) < \lambda\alpha/2$), then we have $\mu \leqslant C(\varepsilon_1 + \varepsilon_2)$. This proves the lemma.

Let $B$ be the Banach space consisting of all systems of functions $\{f_i^k(t)\}$, defined and continuous on the sets $\{q_i^1(G)\}$, with the norm

$$\|\{f_i^k(t)\}\|_B = \max_{i,k} \max_{t\in q_i^1(G)} |f_i^k(t)| \quad (i=0, 1, 2, \ldots, n;\ k=1, 2, \ldots, m_i)\,.$$

We denote by $C(G)$ the space of all functions $f(x, y)$ continuous on $G$ with the uniform metric:

$$\|f(x, y)\|_{C(G)} = \max_{(x,y)\in G} |f(x, y)|\,.$$

LEMMA 4.5.2. *The linear operator* $T: B \to C(G)$ *acting by the formula*

$$T(\{f_i^k(t)\}) = f(x, y) = \sum_{i=0}^{n} \sum_{k=1}^{m_i} p_i^k(x, y) f_i^k(q_i^1(x, y))\,,$$

*maps bounded closed sets of* $B$ *onto closed sets of* $C(G)$.

*Proof.* Let $F \subset B$ be a closed and bounded set of elements of $B$. Suppose that $f_n(x, y)$ is a sequence of functions in $T(F) \subset C(G)$, and that $f(x, y) \in C(G)$, where $\|f(x, y) - f_n(x, y)\|_{C(G)} \to 0$ as $n \to \infty$. We show that then $f(x, y) \in T(F)$. Since $f_n(x, y) \in T(F)$, there exists a sequence of elements $\{f_{i,n}^k(t)\} \in F$ such that $T(\{f_{i,n}^k(t)\}) = f_n(x, y)$. By Lemma 4.5.1 we can select in the sets $\{q_i^1(G)\}$ subsets consisting of a finite number of points $t_{i,j} \in q_i'(G)$ $(i=0, 1, \ldots, n;\ j=1, 2, \ldots, s_i)$ such that for each element $\{f_i^k(t)\} \in B$ the inequality

$$\|\{f_i^k(t)\}\|_B \leqslant c\big(\|f(x, y)\|_{C(G)} + \max_{k,j,i} |f_i^k(t_{i,j})|\big)\,,$$

is satisfied, where the constant $C$ does not depend on the functions $\{f_i^k(t)\}$. Since $F$ is a bounded set, there exists a subsequence of suffixes $n_1, n_2, \ldots$ such that for any $i = 0, 1, \ldots, n$; $k = 1, 2, \ldots, m_i$; $j = 1, 2, \ldots, s_i$ the numerical sequence $f_{i,n_\nu}^k \to C_{k,i,j}$ as $\nu \to \infty$. From this and the previous inequality it follows that $\{f_{i,n_\nu}^k(t)\} \in F$ $(\nu = 1, 2, \ldots)$ is a Cauchy sequence, because it is known that the sequence $f_n(x, y) \in T(F)$ is Cauchy sequence. Consequently there exists an element $\{f_i^k(t)\} \in B$ such that $\| \{f_i^k(t) - f_{i,n_\nu}^k(t)\} \|_B \to 0$. Since $F$ is a closed set, $\{f_i^k(t)\} \in F$. The operator $T \colon B \to C(G)$ is bounded. Therefore $T(\{f_i^k(t)\}) = f(x, y)$. Consequently $f(x, y) \in T(F)$. This proves the lemma.

The following lemma from the theory of linear operators [28] turns out to be useful.

LEMMA 4.5.3. *Let $B_1$ and $B_2$ be Banach spaces. If a linear operator $T \colon B_1 \to B_2$ maps bounded closed sets of $B_1$ onto closed sets of $B_2$, then its domain of values is closed.*

*Proof of Theorem* 4.5.1. The set of superpositions of the form $\sum_{m=1}^{N} p_m(x, y) f_m(g_m(x, y))$ coincides on $G$ with the set of superpositions of the form $\sum_{i=0}^{n} \sum_{k=1}^{m_i} p_i^k(x, y) f_i^k(q_i^1(x, y))$. By Lemma 4.5.2 and 4.5.3 the set of the latter superpositions is closed in the space $C(G)$. This proves the theorem.

## § 6. *The set of linear superpositions in the space of continuous functions is nowhere dense*

THEOREM 4.6.1. *For any continuous functions $p_m(x, y)$ and continuously differentiable functions $q_m(x, y)$ $(m = 1, 2, \ldots, N)$ and any region $D$ of the plane of the variables $x, y$ the set of superpositions of the form*

$$\sum_{m=1}^{N} p_m(x, y) f_m(q_m(x, y)),$$

*where $\{f_m(t)\}$ are arbitrary continuous functions, is nowhere dense in the space of all functions continuous in $D$ with uniform convergence.*

By Lemma 4.2.2 we can find a subregion $G^* \subset D$, determine a constant $\gamma^* > 0$, and renumber the functions $\{q_m(x, y)\}$, with two indices so that the functions $\tilde{q}_i^k(x, y)$ $(i=0, 1, 2, \ldots, \tilde{n}; k=1, 2, \ldots, \tilde{m}_i; \sum_{i=0}^{\tilde{n}} \tilde{m}_i = N)$ obtained after the renumbering satisfy conditions (1), (2), (3) of Lemma 4.2.2. We now fix the point $(x_0, y_0) \in G^*$ and the number $v$ so that the line $(y - y_0) + v(x - x_0) = 0$ does not touch at any of the level curves of the functions $\tilde{q}_i^k(x, y)$ $(i = 1, 2, \ldots, \tilde{n})$ that pass through $(x_0, y_0)$. Let $G^{**} \subset G^*$ be a disc with centre at $(x_0, y_0)$ and radius small enough so that the $\{ \tilde{q}_i^k(x, y) \}$ and $q_{N+1}(x, y) = y + vx$ satisfy condition (3) of Lemma 4.2.2 with some constant $\gamma^{**} > 0$. We put $p_{N+1}(x, y) = 1$. By Lemma 4.4.3 we can find a set $G \subset G^{**}$, determine a constant $\lambda > 0$, and again renumber the functions $p_m(x, y)$ and $q_m(x, y)$ $(m = 1, 2, \ldots, N+1)$ with two indices so that the functions $p_i^k(x, y)$ and

$$q_i^k(x, y) \quad (i = 0, 1, 2, \ldots, n+1\ ;\ k = 1, 2, \ldots, m_i\ ;\ \sum_{i=0}^{n+1} m_i \leqslant N+1)$$

that is, some functions may be omitted in the renumbering) obtained after the renumbering satisfy conditions (1)-(3) of Lemma 4.2.2, conditions (4′)-(6′) of § 5, and the condition

$$7 \quad m_{n+1} = 1\,, \quad p_{N+1}^1 = p_{N+1}(x, y) = 1\,, \quad q_{N+1}^1 = q_{N+1}(x, y) = y + vx\,.$$

Let $L$ be the linear space consisting of all system of functions $\{ f_i^k(t) \}$ defined and continuous on the sets $\{ q_i^1(G) \}$ and satisfying the condition

$$\sum_{i=0}^{n+1} \sum_{k=1}^{m_i} p_i^k(x, y) f_i^k(q_i^1(x, y)) \equiv 0 \quad \text{in} \quad G\,.$$

LEMMA 4.6.1. *$L$ is a finite-dimensional linear space.*

*Proof.* By Lemma 4.5.1, in the sets $\{ q_i^1(G) \}$ we can select a subset consisting of a finite number of points $\{ t_{i,j} \}$ such that, if $\{ f_i^k(t) \} \in L$ and $f_i^k(t_{i,j}) = 0$ for all $k, i, j$ then $f_i^k(t) \equiv 0$ on $q_i^1(G)$ for all $i, k$. Thus, the set of functions $\{ f_i^k(t) \}$ is completely determined by a finite set of parameters $\{ f_i^k(t_{i,j}) \}$. Consequently the dimension of the space $L$ is finite. This proves the lemma.

LEMMA 4.6.2. *There exists a natural number $\mu$ such that in $D$ the polynomial $(y + vx)^\mu = Q(x, y)$ is not equal to any superposition of the form $\sum_{m=1}^{N} p_m(x, y) f_m(q_m(x, y))$, where $\{ f_m(t) \}$ are arbitrary continuous functions.*

*Proof.* We denote by $\Phi$ the space of functions of the form $f(y+vx) = f^{1}_{n+1}(q^{1}_{n+1}(x, y))$ that are representable on $G$ by superpositions of the form $[\sum_{m=1}^{N} p_m(x, y) f_m(q_m(x, y))]$. Or, what comes to the same thing (see properties (4′) and (7)), of the form $[\sum_{i=0}^{n} \sum_{k=1}^{m_i} p_i^k(x, y) f_i^k(q_i^k(x, y))]$. Thus, functions of $\Phi$ satisfy the relation $\sum_{i=0}^{n+1} \sum_{k=1}^{m_i} p_i^k(x, y) f_i^k(q_i^k(x, y)) \equiv 0$ in $G$. Consequently the linear space $\Phi$ is naturally embedded in $L$. Since $L$ is finite-dimensional (Lemma 4.6.1), $\Phi$ is also finite-dimensional. Let $l$ be the dimension of $\Phi$. Since the polynomials $(y+vx), (y+vx)^2, \ldots, (y+vx)^{l+1}$ are linearly independent, at least one of them $Q(x, y) = (y+vx)^{\mu}$ is not equal to any superposition of the form under discussion on $G$ or, consequently, in $D$. This proves the lemma.

*Proof of Theorem* 4.6.1. By Lemma 4.6.2 the set of superpositions of the form given in Theorem 4.6.1 does not exhaust all continuous functions on $G$. Consequently, by Theorem 4.5.1, the set of these superpositions is a closed linear subspace of $C(G)$. Hence we conclude that the set of superpositions under discussion is nowhere dense in $C(G)$, nor consequently in $C(D)$. This proves the theorem.

COROLLARY 4.6.1. *For any continuous functions* $p_m(x_1, x_2, \ldots, x_n)$ *and continuously differentiable functions* $q_m(x_1, x_2, \ldots, x_n)$ $(m = 1, 2, \ldots, N)$ *and any region* $D$ *of the space of the variables* $(x_1, x_2, \ldots, x_n)$ *the set of superpositions of the form*

$$\sum_{m=1}^{N} p_m(x_1, x_2, \ldots, x_n) f_m(q_m(x_1, x_2, \ldots, x_n), x_2, x_3, , x_{n-1}),$$

*where* $\{f_m(t, x_2, x_3, \ldots, x_{n-1})\}$ *are arbitrary continuous functions of* $(n-1)$ *variables, is nowhere dense in the space of all functions continuous in* $D$ *with uniform convergence.*

## Chapter 5. — Dimension of the space of linear superpositions

In this chapter we present a calculation of the functional dimension of the space of functions representable by means of linear superpositions and prove that a representation of analytic functions by means superpositions of smooth functions can not be stable.

### § 1. ($\varepsilon$, $\delta$)-*entropy and the "dimension" of function spaces*

Let $G_n$ be a closed region of $n$-dimensional euclidean space, and $C(G_n)$ the space of all functions continuous in $G_n$. Two functions $f_1(x), f_2(x) \in C(G_n)$ are called ($\varepsilon$, $\delta$)-distinguishable if there exists an $n$-dimensional closed sphere $S_\delta \subset G_n$ of radius $\delta$ such that

$$\min_{x \in S_\delta} |f_1(x) - f_2(x)| \geqslant \varepsilon .$$

Let $F \subset C(G_n)$ be a set of continuous functions. A subset $K \subset F$ is called ($\varepsilon$, $\delta$)-distinguishable if any two of its elements are ($\varepsilon$, $\delta$)-distinguishable. We denote by $N_{\varepsilon,\delta}(F)$ the maximum number of elements in an ($\varepsilon$, $\delta$)-distinguishable subset of $F$.

*Definition* 5.1.1. The number $H_{\varepsilon,\delta}(F) = \log_2 N_{\varepsilon,\delta}(F)$, by analogy with the definition of $\varepsilon$-entropy, is called the ($\varepsilon$, $\delta$)-entropy of $F$.

Let $f_0 \in F$. We denote by $F_{\lambda\varepsilon}(f_0)$ the set of functions $f \in F$ such that $|f(x) - f_0(x)| \leqslant \lambda\varepsilon$. It follows immediately from the definition that the expression $\overline{\lim_{\delta \to 0}}\ \overline{\lim_{\varepsilon \to 0}} - \dfrac{\log_2 H_{\varepsilon,\delta}(F_{\lambda\varepsilon}(f_0))}{\log_2 \delta}$ as a function of $\lambda$ does not decrease as $\lambda \to \infty$.

*Definition* 5.1.2. The number

$$r(F, f_0) = \lim_{\lambda \to \infty} \overline{\lim_{\delta \to 0}}\ \overline{\lim_{\varepsilon \to 0}} - \frac{\log_2 H_{\varepsilon,\delta}(F_{\lambda\varepsilon}(f_0))}{\log_2 \delta}$$

is called the functional "dimension" of $F$ at $f_0$. The number $r(F) = \sup (F, f_0)$ is called the functional "dimension" of $F$.

The functional "dimension" $r(F)$ of a set of functions $F \subset C(G_n)$ has the following properties.

5.1.1. Let $\Phi \subset F$ be a set of functions. Then $r(\Phi) \leqslant r(F)$. Moreover, if $\Phi$ is everywhere dense in $F$ in the uniform metric, then $r(\Phi) = r(F)$.

*Proof.* The first part of the assertion follows immediately from the definition. For a proof of the second part it is sufficient to show that $r(\Phi, \varphi_0) \geqslant r(F, \varphi_0)$ for any element $\varphi_0 \in \Phi$. Suppose that the functions $f_1, \ldots, f_N$ from a $(2\varepsilon, \delta)$-distinguishable subset of $F_{\lambda\varepsilon}(\varphi_0)$. Since $\Phi$ is everywhere dense in $F$, there exist functions $\varphi_1, \ldots, \varphi_N \in \Phi$ such that $\max\limits_{x \in G_n} |f_i(x) - \varphi_i(x)| \leqslant \min\left(\frac{\varepsilon}{2}, \lambda\varepsilon\right)$ $(i = 1, 2, \ldots, N)$. These functions form an $(\varepsilon,\delta)$-distinguishable subset of $F_{2\lambda\varepsilon}(\varphi_0)$. Consequently $N_{\varepsilon,\delta}(\Phi_{2\lambda\varepsilon}(\varphi_0)) \geqslant N_{2\varepsilon,\delta}(F_{\lambda\varepsilon}(\varphi_0))$. Hence $r(\Phi, \varphi_0) \geqslant r(F, \varphi_0)$.

5.1.2. For any set $F \subset C(G_n)$ we have $r(F) \leqslant n$.

*Proof.* Suppose that $f_0 \in F$ and $f_1, f_2, \ldots, f_p$ is a maximal set (with respect to $p$) of pairwise $(\varepsilon, \delta)$-distinguishable functions of $F_{\lambda\varepsilon}(f_0)$. Let $\sigma_1, \sigma_2, \ldots, \sigma_q$ be a maximal set (with respect to $q$) of spheres of radius $\delta/3$ in $G_n$, such that no two of them have common interior points. Then any pair of functions $f_i(x)$ and $f_j(x)$ of the given set satisfies on at least one of the spheres $\sigma_l$ the inequality $\min\limits_{x \in \sigma_l} |f_i(x) - f_j(x)| \geqslant \varepsilon$. For the functions $f_i(x)$ and $f_j(x)$ satisfy on some sphere $S_\delta \subset G_n$ the inequality $\min\limits_{x \in S_\delta} |f_i(x) - f_j(x)| \geqslant \varepsilon$. Since $q$ is maximal, it follows that one of the spheres $\sigma_l \subset S_\delta$. Consequently on this sphere the inequality we need is satisfied. We denote by $a_l$ the centre of the sphere $\sigma_l$ $(l = 1, 2, \ldots, q)$. Every set of functions $f_{i_1}, f_{i_2}, \ldots, f_{i_r}$ each pair of which has values differing by not less than $\varepsilon$ at one and the same point consists of a number $r \leqslant 2\lambda + 1$ of functions. (All functions are taken from the set indicated above.) Since every pair of functions $f_i(x)$ and $f_j(x)$ has values differing by not less than $\varepsilon$ at one of the points $a_l$ at least, we have $p \leqslant 2\lambda + 1$. But since the spheres $\{\sigma_i\}$ do not intersect, $q \leqslant C/\delta^n$, where $C$ is a constant depending only on $n$. Consequently,

$$r(F, f_0) \leqslant \lim_{\lambda \to \infty} \lim_{\delta \to 0} \lim_{\varepsilon \to 0} -\frac{\log_2 \log_2 (2\lambda + 1)^{\frac{C}{\delta^n}}}{\log_2 \delta} = n.$$

5.1.3. If $F$ is everywhere dense (in the uniform metric) in the space $C(G_n)$, then $r(F) = n$. In particular $r(C(G_n)) = n$.

*Proof.* By 5.1.1 and 5.1.2 it is sufficient to show that $r\left(C(G_n)\right) \geqslant n$. We denote by $C_\varepsilon(G_n)$ the set of all $f(x) \in C(G_n)$ for which $\max\limits_{x \in G_n} |f(x)| \leqslant \varepsilon$. Let $\theta > 0$ be a constant such that for any $\delta > 0$ we can find $H = [\theta/\delta^n]$ closed and pairwise non-intersecting spheres $\sigma_1, \sigma_2, \ldots, \sigma_H$ of radius $\delta$ in $G_n$. For any system of numbers $\{\alpha_i\}$ $(\alpha_i = \pm 1,\ i = 1, 2, \ldots, H)$ we construct a function $f_{\{\alpha_i\}}(x) \in C_\varepsilon(G_n)$ such that $f_{\{\alpha_i\}}(x) = a_i\varepsilon$ for $x \in \sigma_i$ $(i = 1, 2, \ldots, H)$. These functions are obviously pairwise $(\varepsilon,\delta)$-distinguishable. The number of functions $f_{\{\alpha_i\}}(x)$ for all possible sets $\{\alpha_i\}$ is equal to $2^H$. Consequently $H_{\varepsilon,\delta}\left(C_\varepsilon(G_n)\right) \geqslant H = [\theta/\delta^n]$. Hence $r\left(C(G)\right) \geqslant n$.

COROLLARY 5.1.1. *The space of all polynomials in $n$ variables has functional "dimension" $n$.*

In the same way, the following properties are easily proved.

5.1.4. Let $G_n^1$ and $G_n^2$ be two non-intersecting closed regions in $n$- dimensional space, and $F(G_n^1 \cup G_n^2)$ a space of functions, defined and continuous on $G_n^1 \cup G_n^2$. Denote by $F(G_n^1)$ the space of all functions $\varphi(x)$, defined on the set $G_n^1$, for which there exists a function $\Phi(x) \in F(G_n^1 \cup G_n^2)$ such that $\varphi(x) \equiv \Phi(x)$ for $x \in G_n^1$. The space $F(G_n^2)$ is defined similarly. Then

$$r\left(F(G_n^1 \cup G_n^2)\right) = \max\left\{r\left(F(G_n^1)\right);\ r\left(F(G_n^2)\right)\right\}.$$

5.1.5. If $F$ is a linear space, then $r(F) = r(F, f_0)$ for any function $f_0 \in F$. If $F$ is a finite-dimensional linear space, then $r(F) = 0$.

5.1.6. Let $F$ be a linear metric space with metric $\rho(\varphi, \psi)$ between a pair of functions $\varphi, \psi \in F$. We denote by $F(\rho_0)$ the set of all those functions $\varphi \in F$ for which $\rho(\varphi, 0) \leqslant \rho_0$. Then $r(F) = r\left(F(\rho_0)\right)$.

COROLLARY 5.1.2. *The set of all polynomials in $n$ variables whose partial derivatives of order $p$, for any $p = 1, 2, \ldots,$ are bounded by a constant $0 < K_p < \infty$ has functional "dimension" $n$.*

5.1.7. Let $F$ be a complete linear metric space and $F = \bigcup\limits_{i=1}^{\infty} F_i$, where $\{F_i\}$ are sets of continuous functions. Then $r(F) = \max\limits_i r(F_i)$.

We now write down the main result on the functional "dimension" of a set of linear superpositions.

5.1.8. Let $q_i = q_i(x_1, x_2, \ldots, x_n)$ be continuously differentiable functions of $n$ variables, and $p_i = p_i(x_1, x_2, \ldots, x_n)$ continuous functions of $n$ variables $(i = 1, 2, \ldots, N)$. We denote by $F(G_n, \{p_i\}, \{q_i\})$ the set of superpositions of the form $\sum_{i=1}^{N} p_i(x_1, x_2, ..., x_n) f_i(q_i(x_1, x_2, ..., x_n))$, where $(x_1, x_2, ..., x_n) \in G_n$, and $\{f_i(t)\}$ are arbitrary continuous functions of one variable. Then in any region $D_n$ there exists a closed subregion $G_n \subset D_n$ such that

$$r(F(G_n, \{p_i\}, \{q_i\})) \leqslant 1.$$

For ease of presentation we limit the proof to the case $n = 2$ (§ 3). It is interesting to compare the result 5.1.8 with the following proposition.

5.1.9. Let $\alpha_i(x_1, x_2, ..., x_n) = \sum_{j=1}^{n} \alpha_{ij}(x_j) \quad (i = 1, 2, ..., 2n+1)$

be the continuous functions involved in Kolmogorov's formula (I). We denote by $\psi(G_n, \alpha_i)$ the space of all functions of the form $\psi(\alpha_i(x_1, x_2, ..., x_n))$, where $\psi(t)$ is an arbitrary continuous function of one variable and $(x_1, x_2, ..., x_n) \in G_n$. Then for any $i$ and every region $G_n$, $r(\psi(G_n, \alpha_i)) = n$ (see 5.1.7).

Let $p_i(x_1, x_2, ..., x_n)$ be fixed continuous functions of $n$ variables, $q_{1,i}(x_1, x_2, ..., x_n)$, $q_{2,i}(x_1, x_2, ..., x_n)$, ..., $q_{k,i}(x_1, x_2, ..., x_n)$ fixed continuously differentiable functions of $n$ variables, and $f_i(t_1, t_2, ..., t_k)$ arbitrary continuous functions of $k$ variables, $k < n$ $(i = 1, 2, ..., N)$. One would expect that the set of superpositions of the form (V) (see Chapter I) has functional "dimension" not greater than $k$. However, in this direction, only the following partial result has so far been proved.

5.1.10. Denote by $F(\lambda, G_n, \{p_i\}, \{q_{1,i}\}, ..., \{q_{k,i}\})$ the set of all those continuous functions $\varphi(x_1, x_2, ..., x_n)$ for which there exist continuous functions $\{f_i(t_1, t_2, ..., t_k)\}$ such that in $G_n$.

$$\varphi(x_1, x_2, ..., x_n) = \sum_{i=1}^{N} p_i(x_1, x_2, ..., x_n) f_i(q_{1,i}(x_1, x_2, ..., x_n), ..., q_{k,i}(x_1, x_2, ..., x_n))$$

and

$$\max_i \sup_{(t_1, t_2, ..., t_k)} |f_i(t_1, t_2, ..., t_k)| \leqslant \lambda \sup_{(x_1, x_2, ..., x_n) \in G_n} |\varphi(x_1, x_2, ..., x_n)|$$

Then, for any $\lambda < \infty$, in any region $D_n$ there exists a closed subregion $G_n \subset D_n$ such that

$$r(F(\lambda, G_n, \{p_i\}, \{q_{1,i}\}, ..., \{q_{k,i}\}), 0) \leqslant k.$$

From the last result and Banach's open mapping theorem there follows

COROLLARY 5.1.3. *For any continuous functions* $p_i$ *and continuously differentiable functions* $q_{1,i}, q_{2,i}, \ldots, q_{k,i}$, $k < n$ $(i = 1, 2, \ldots, N)$ *and every region* $G_n$ *there exists a continuous function that is not equal in* $G_n$ *to any superposition of the form* (V).

## § 2. (ε, δ)-*entropy of the set of linear superpositions*

We denote by $S(\delta, z)$ the disc of radius $\delta$ with centre at $z$. Let $p(z) = p(x, y)$ and $q(z) = q(x, y)$ be functions defined in a closed region $G$ of the $x, y$-plane and having the properties:

a) $p(x, y), \dfrac{\partial q(x, y)}{\partial x}, \dfrac{\partial q(x, y)}{\partial y}$ are continuous in $G$ and have modulus of continuity $\omega(\delta)$,

b) the inequalities $0 < \gamma \leqslant |\operatorname{grad}[q(r)]| \leqslant \dfrac{1}{\gamma}$ and $|p(z)| \leqslant \dfrac{1}{\gamma}$, where $\gamma$ is some constant, are satisfied everywhere in $G$.

LEMMA 5.2.1. *Let* $S(\delta, z) \subset G$ *and let* $\mu_q(t)$ *be the function equal to* $2\sqrt{\delta^2 - (t - q(z))^2 |\operatorname{grad}[q(z)]|^{-2}}$ *on*

$$q(z) - \delta |\operatorname{grad}[q(z)]| \leqslant t \leqslant q(z) + \delta |\operatorname{grad}[q(z)]|$$

*and equal to zero elsewhere. Then*

$$\int_{-\infty}^{\infty} |\mu_q(t) - h_1(e(q, t) \cap S(\delta, z))| \, dt \leqslant c_1(\gamma)\, \omega(\delta)\, \delta^2 ,$$

*where* $c_1(\gamma)$ *is a constant depending only on* $\gamma$.

*Proof.* Let $[a, b] \subset e(q, t) \cap S(\delta, z)$ be the segment of the level curve $e(q, t)$, endpoints $a$ and $b$, lying on the boundary of $S(\delta, z)$; $[z, a]$ and $[z, b]$ the vectors with origin at $z$ and endpoints at $a$ and $b$, respectively;

$$\alpha_1 = \gamma(\overrightarrow{[z, a]}, \operatorname{grad}[q(z)]), \quad \alpha_2 = \gamma(\overrightarrow{[z, b]}, \operatorname{grad}[q(z)]).$$

We have

$$|t - q(z)| = |q(a) - q(z)| = \left| \int_{s \in [z, a]} \frac{\partial q}{\partial s} ds \right|$$
$$= \delta \cos \alpha_1 |\operatorname{grad}[q(z)]| (1 + 0(1)\, \omega(\delta))$$

Hence

$$\delta \sin \alpha_1 = \sqrt{\delta^2 - (t - q(z) + 0(\gamma)\,\delta\omega(\delta))^2 \,|\operatorname{grad}\ [q(z)]|^{-2}}$$

and similarly

$$\delta \sin \alpha_2 = \sqrt{\delta^2 - (t - q(z) + 0(\gamma)\,\delta\omega(\delta))^2 \,|\operatorname{grad}\ [q(z)]|^{-2}}$$

By b) the size of the angle swept out by the tangent vector to the level curve $e(q, t)$ on moving along $[a, b]$ does not exceed $C_2(\gamma)\,\omega(\delta)$. Therefore

$$h_1([a, b]) = \delta(\sin \alpha_1 + \sin \alpha_2)(1 + 0(\gamma)\,\omega(\delta))$$
$$= 2\sqrt{\delta^2 - (t - q(z) + 0(\gamma)\,\delta\omega(\delta))^2 \,|\operatorname{grad}\ [q(z)]|^{-2}} + 0(\gamma)\,\delta\omega(\delta).$$

If $\alpha_1 \geqslant C_3(\gamma)\,\omega(\delta)$ ($C_3$ is a sufficiently large constant), then $[a, b] = e(q, t) \cap S(\delta, z)$. Consequently, for

$$|t - q(z)| \leqslant \theta = \delta \cos [C_3\,\omega(\delta)]\,|\operatorname{grad}[q(z)]| \times (1 + 0(1)\,\omega(\delta))$$

we have $h_1(e(q, t) \cap S(\delta, z)) = h_1([a, b])$. Since for every $t$ (by b))

$$h_1(e(q, t) \cap S(\delta, z)) \leqslant C_4(\gamma)\,\delta\,(1 + \omega(\delta)),$$

we have

$$\int_{-\infty}^{\infty} |h_1(e(q, t) \cap S(\delta, z)) - \mu_q(t)|\,dt =$$
$$= \int_{q(z)-\Theta}^{q(z)+\Theta} |h_1(e(q, t) \cap S(\delta, z)) - \mu_q(t)|\,dt + 0(\gamma)\,\delta^2\omega(\delta).$$

We now estimate

$$\int_{q(z)-\Theta}^{q(z)+\Theta} |h_1(e(q, t) \cap S(\delta, z)) - \mu_q(t)|\,dt =$$
$$= \int_{q(z)-\Theta}^{q(z)+\Theta} |h_1([a, b]) - \mu_q(t)|\,dt \leqslant$$
$$\leqslant 2 \int_{q(z)-\Theta}^{q(z)+\Theta} \Big(\sqrt{\delta^2 - (t - q(z) + 0(\gamma)\,\delta\omega(\delta))^2 \,|\operatorname{grad}\ [q(z)]|^{-2}}$$
$$- \sqrt{\delta^2 - (t - q(z))^2 \,|\operatorname{grad}\ [q(z)]|^{-2}}\Big)\,dt + 0(\gamma)\,\delta^2\omega(\delta)$$
$$= 0(\gamma)\,\delta^2\omega(\delta) \int_{-1}^{1} \frac{d\tau}{\sqrt{1 - \tau^2}} + 0(\gamma)\,\delta^2\omega(\delta) = 0(\gamma)\,\delta^2\omega(\delta).$$

Here we have the mean value theorem. This proves the lemma.

LEMMA 5.2.2. *Let* $p(z), q(z)$ *satisfy conditions a) and b);* $S(\delta, z) \subset G$; *let* $f(t)$ *be an arbitrary continuous function, uniformly bounded in modulus by the constant* $m$. *Then*

$$\iint\limits_{(u,v)\in S(\delta,z)} p(u,v) f(q(u,v))\, du dv$$

$$= p(z) \left| \text{grad } [q(z)] \right|^{-1} \int_{-\infty}^{\infty} f(t)\, \mu_q(t)\, dt + \lambda(z)\, m\delta^2 \omega(\delta),$$

*where* $|\lambda(z)| \leqslant C_5(\gamma)$.

*Proof.* Using a) and b) and Lemma 5.2.1 we have

$$\int\limits_{S(\delta,z)} p(u,v) f(q(u,v))\, du dv$$

$$= p(z) \iint\limits_{(u,v)\in S(\delta,z)} f(q(u,v))\, du dv + 0(1)\, m\delta^2 \omega(\delta)$$

$$= p(z) \int_{-\infty}^{\infty} \{ f(t) \int\limits_{s\in e(q,t)\cap S(\delta,z)} \left| \text{grad } [q(s)] \right|^{-2} ds \}\, dt + 0(1)\, m\delta^2 \omega(\delta)$$

$$= p(z) \left| \text{grad } [q(z)] \right|^{-1} \int_{-\infty}^{\infty} \{ f(t) \int\limits_{s\in e(q,t)\cap S(\delta,z)} ds \}\, dt + 0(\gamma)\, m\delta^2 \omega(\delta)$$

$$= p(z) \left| \text{grad } [q(z)] \right|^{-2} \int_{-\infty}^{\infty} f(t)\, h_1(e(q,t)\cap S(\delta,z))\, dt + 0(\gamma)\, m\delta^2 \omega(\delta)$$

$$= p(z) \left| \text{grad } [q(z)] \right|^{-1} \int_{-\infty}^{\infty} f(t)\, \mu_q(t)\, dt + 0(\gamma)\, m\delta^2 \omega(\delta).$$

This proves the lemma.

LEMMA 5.2.3. *Suppose that a number* $\alpha > 0$ *and functions* $p(z)$, $q(z), f(t)$ *satisfying the conditions of Lemma* 5.2.2. *are given. If for every integer* $k$ *such that*

$$\min_{z\in G} q(z) \leqslant t_k = k\delta \frac{\alpha}{m} \leqslant \max_{z\in G} q(z)$$

*and any integer* $l$ *such that*

$$\min_{z\in G} \left| \text{grad } [q(z)] \right| \leqslant t_l' = l \frac{\alpha}{m} \leqslant \max_{z\in G} \left| \text{grad } [q(z)] \right|,$$

*the inequality*

$$\left| \int_{t_k - t_l'\delta}^{t_k + t_l'\delta} f(t) \sqrt{\delta^2 - \left( \frac{t - t_k}{t_l'} \right)^2}\, dt \right| \leqslant \alpha\delta^2$$

*is satisfied, then for every* disc $S(\delta, z) \subset G$

$$\Big| \iint_{(u,v) \in S(\delta, z)} p(u,v) f(q(u,v))\, du dv \Big| \leqslant c_6(\gamma)\left(\alpha\delta^2 + m\delta^2\omega(\delta)\right).$$

*Proof.* Suppose that a disc $S(\delta, z) \subset G$ is given. By the condition of the lemma there are integers $k$ and $l$ such that $|q(z) - t_k| \leqslant \delta\alpha/m$ and $\big| |\operatorname{grad}[q(z)]| - t_l' \big| \leqslant \alpha/m$. From Lemma 5.2.2 we obtain

$$\Big| \int_{(u,v) \in S(\delta, z)} p(u,v) f(q(u,v))\, du\, dv \Big| \leqslant \frac{|p(z)|}{|\operatorname{grad}[q(z)]|} \Big| \int_{-\infty}^{\infty} f(t)\mu_q(t)\, dt \Big|$$

$$+ c_5(\gamma)\, m\delta^2\omega(\delta) \leqslant \frac{2}{\gamma^2} \left| \int_{q(z) - \delta|\operatorname{grad}[q(z)]|}^{q(z) + \delta|\operatorname{grad}[q(z)]|} f(t) \sqrt{\delta^2 - \frac{(t - q(z))^2}{|\operatorname{grad}[q(z)]|^2}}\, dt \right.$$

$$\left. - \int_{t_k - t_l'\delta}^{t_k + t_l'\delta} f(t) \sqrt{\delta^2 - \left(\frac{t - t_k}{t_l'}\right)^2}\, dt \right| + \frac{2}{\gamma^2}\alpha\delta^2 + c_5(\gamma)\, m\delta^2\omega(\delta) \leqslant$$

(by the mean value theorem)

$$\leqslant \frac{2}{\gamma^2}\alpha\delta^2 + c_5(\gamma)\, m\delta^2\omega(\delta) + \frac{2}{\gamma^2}\left(\int_{-1}^{1} \frac{\delta m\, d\tau}{\sqrt{1 - \tau^2}}\right)\delta\frac{\alpha}{m}$$

$$+ \frac{2}{\gamma^2}\left(\int_{-1}^{1} \frac{\delta^2 m\, d\tau}{\sqrt{1 - \tau^2}}\right)\frac{\alpha}{m} \leqslant c_6(\gamma)\left(\alpha\delta^2 + m\delta^2\omega(\delta)\right).$$

This proves the lemma.

We denote by $F_m = F_m(D; p_1, p_2, \ldots, p_N; q_1, q_2, \ldots, q_N)$ the set of superpositions of the form

$$f(x, y) = \sum_{i=1}^{N} p_i(x, y) f_i(q_i(x, y)), \text{ where } \{p_i(x, y)\}$$

and $\{q_i(x, y)\}$ are fixed functions, defined in the closed region $D$ of the $x, y$ plane and satisfying conditions a) and b) with a constant $\gamma$ not depending on $i$ and $\{f_i(t)\}$ are arbitrary continuous functions, defined on $\{[a_i, b_i]\} = \{[\min_{z \in D} q_i(z); \max_{z \in D} q_i(z)]\}$ and uniformly bounded in modulus by the constant $m$.

THEOREM 5.2.1. *There exist constants* $A$ *and* $B$ *such that if* $\varepsilon > Am\omega(\delta)$ *then for the* $(\varepsilon, \delta)$*-entropy of the set of functions* $F_m$, $H_{\varepsilon,\delta}(F_m) \leqslant \frac{B}{\delta}\left(\frac{m}{\varepsilon}\right)^2$, *where* $A$ *and* $B$ *depend only on* $\gamma$, $N$ *and* $D$.

*Proof.* We put

$$R(f(z), \delta) = \max_{S(\delta,z)\subset D} \left| \frac{1}{\pi\delta^2} \iint_{(u,v)\in S(\delta,z)} f(u,v)\, du dv \right| .$$

We denote by $\mathscr{H}_{\varepsilon,\delta}(F_m)$ the $\varepsilon$-entropy of the space $F_m$, taking as the distance between the functions $f_1(z), f_2(z) \in F_m$ the number $R(f_1(z) - f_2(z), \delta)$. The inequality $H_{2\varepsilon,\delta}(F_m) \leqslant \mathscr{H}_{\varepsilon,\delta}(F_m)$ holds owing to the fact that if two functions $f_1(z)$ and $f_2(z)$ are $(\varepsilon, \delta)$-distinguishable, then they are $\varepsilon$-distinguishable also in the sense of the metric $R(f_1(z) - f_2(z), \delta)$. We now estimate the value of $\mathscr{H}_{\varepsilon,\delta}(F_m)$. Let $k$ and $l$ be integers such that

$$\min_{z\in D} q_i(z) \leqslant t_k = k\delta\frac{\alpha}{m} \leqslant \max_{z\in D} q_i(z)$$

and

$$\min_{z\in D} |\operatorname{grad}[q_i(z)]| \leqslant t_l' = l\frac{\alpha}{m} \leqslant \max_{z\in D} |\operatorname{grad}[q_i(z)]| .$$

To compute the function

$$f_\delta(z) = \frac{1}{\pi\delta^2} \iint_{(u,v)\in S(\delta,z)} f(u,v)\, du dv ,$$

where $f(x, y) \in F_m$, $S(\delta, z) \subset D$ to within $\varepsilon$, it is sufficient by Lemma 5.2.3 to give the values of

$$v_i(t_k, t_l') = \frac{1}{\pi\delta^2} \int_{t_k - t_l'\delta}^{t_k + t_l'\delta} f_i(t) \sqrt{\delta^2 - \left(\frac{t - t_k}{t_l'}\right)^2}\, dt$$

to within $\alpha = \pi\varepsilon / (2NC_B(\gamma))$ and to assume that $\delta$ is small enough so that

$$\varepsilon > \frac{2NC_B(\gamma)\, m\omega(\delta)}{\pi} = A(\gamma, N)\, m\omega(\delta) .$$

Since $|v_i(t_k, t_l')| \leqslant C_1 m$, to write the numbers $v_i(t_k, t_l')$ ($i, k, l$ fixed) $\log_2(C_1 m/\alpha)$ binary digits are sufficient. Since

$$\left| v_i(t_{k+1}, t_l') - v_i(t_k, t_l') \right| \leqslant c_8 \frac{1}{\delta^2} \left( \int_{-1}^{1} \frac{\delta m d\tau}{\sqrt{1-\tau^2}} \right) \delta \frac{\alpha}{m} = c_9(\gamma)\,\alpha$$

(here we again use the mean value theorem), to store the numbers $v_i(t_{k+1}, t_l') - v_i(t_k, t_l')$ to within $\alpha$, $\log_2 C_9$ binary digits are sufficient. Therefore to write the numbers $v_i(t_k, t_l')$ ($i$, $l$ fixed; $k$ any admissible number) $C_{10}(\gamma)\left[\log_2 \frac{m}{\alpha} + (b_i - a_i)\frac{m}{\delta\alpha}\right] = \mathscr{H}_{i,l}$ binary digits are sufficient. Consequently the total number of digits sufficient to store all the numbers $v_i(t_k, t_l')$ to within $\alpha$, that is, to store the functions $f_\delta(z)$ to within $\varepsilon$, is

$$\mathscr{H} = \sum_{i,l} \mathscr{H}_{i,l} \leqslant N c_{10}(\gamma)\left[\log_2 \frac{m}{\alpha} + (b_i - a_i)\frac{m}{\delta\alpha}\right]\frac{1}{\gamma}\frac{m}{\alpha} \leqslant \frac{B(\gamma, N, D)}{\delta}\left(\frac{m}{\varepsilon}\right)^2.$$

This proves the theorem.

## § 3. *Functional "dimension" of the space of linear superpositions*

Suppose that continuous functions $p_i(x, y)$ and continuously differentiable functions $q_i(x, y)$ $(i=1, 2, \ldots, N)$ are fixed. Let $G$ be a closed region of the $x, y$ plane. We denote by $F = F(G, \{p_i\}, \{q_i\})$ the set of superpositions of the form $f(x, y) = \sum_{i=1}^{N} p_i(x, y) f_i(q_i(x, y))$, where $(x, y) \in G$ and $\{f_i(t)\}$ are arbitrary continuous functions of one variable. We are interested in the functional dimension of the set $F$.

THEOREM 5.3.1. *In every region $D$ of the $x, y$ plane there exists a closed subregion $G \subset D$ such that*

$$r(F(G, \{p_i\}, \{q_i\})) \leqslant 1.$$

*Proof.* By Theorem 4.5.1, in $D$ there exists a closed subregion $G^* \subset D$ such that the set of superpositions $F(G^*, \{p_i\}, \{q_i\})$ is closed (in the uniform metric) in $C(G^*)$, and the functions $\{q_i(x, y)\}$ satisfy the condition: for any $i$, either grad $[q_i(x, y)] \neq 0$ on $G^*$ or $q_i(x, y) \equiv \text{const}$ on $G^*$. We show that $r(F(G^*, \{p_i\}, \{q_i\})) \leqslant 1$. By Banach's open mapping theorem, there exists a constant $K$ such that for any superposition $\sum_{i=1}^{N} p_i(x, y) f_i(q_i(x, y)) = f(x, y) \in F(G^*, \{p_i\}, \{q_i\})$ there are continuous functions $\{f_i^*(t)\}$, defined on the sets $\{q_i(G^*)\}$ and satisfying the conditions

8) $$f(x, y) = \sum_{i=1}^{N} p_i(x, y) f_i^*(q_i(x, y)) \text{ for all } (x, y) \in G^*;$$

9) $$\max_i \max_{t \in q_i(G^*)} |f_i^*(t)| \geqslant K \max_{(x,y) \in G^*} |f(x, y)|.$$

Denote by $F_{\lambda\varepsilon} = F_{\lambda\varepsilon}(G^*, \{p_i\}, \{q_i\})$ the set of superpositions $f(x,y) \in F(G^*, \{p_i\}, \{q_i\})$ such that $\max_{(x,y) \in G^*} |f(x, y)| \leqslant \lambda\varepsilon$. By Theorem 5.2.1 and (8), (9), there exist constants $A$ and $B$ such that if $\omega(\delta) \leqslant (\lambda AK)^{-1}$ then $H_{\varepsilon,\delta}(F_{\lambda\varepsilon}) \leqslant B(\lambda K)^2/\delta$. Hence the functional dimension

$$r(F_i(G^*, \{p_i\}, \{q_i\})) \leqslant \lim_{\lambda\to\infty} \lim_{\delta\to 0} \lim_{\varepsilon\to 0} \frac{\log_2 \log_2 \frac{B(\lambda K)^2}{\delta}}{\log_2 \delta} = 1$$

This proves the theorem.

From Theorem 5.3.1 and the properties of functional dimension (§ 1) we have the following result, which is a stronger form of Theorem 4.6.1.

COROLLARY 5.3.1. *For any continuous functions* $\{p_i(x, y)\}$ *and continuously differentiable functions* $\{q_i(x, y)\}$ *and every region* $D$ *the set of linear superpositions* $F(D, \{p_i\}, \{q_i\})$ *is nowhere dense in any space of functions that has in every region* $G \subset D$ *functional "dimension" greater than* 1.

*Remark* 5.3.1. All the results about linear superpositions of the form $\sum_{i=1}^{N} p_i(x, y) f_i(q_i(x, y))$ remain valid if we assume that $\{f_i(t)\}$ are arbitrary bounded measurable functions.

## § 4. *Variation of superpositions of smooth functions*

Let $G_n$ be a closed region of the space of the variables $x_1, x_2, \ldots, x_n$ $(n \geqslant 2)$. A function $F(x) = F(x_1, x_2, \ldots, x_n)$ is called a superposition of order $s$ generated by the functions of $k$ $(k > 1)$ variables

$$f_{\beta_1,\beta_2\ldots,\beta_\alpha}(t_1, t_2, \ldots, t_k) \quad (\alpha = 0, 1, 2, \ldots, s;\ \beta_i = 1, 2, \ldots, k)$$

if it is defined in $G$ by relations

$$\begin{cases} F = f(q_1, q_2, \ldots, q_k)\,, \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ q_{\beta_1,\beta_2,\ldots,\beta_\alpha} = f_{\beta_1,\ldots,\beta_\alpha}(q_{\beta_1,\ldots,\beta_\alpha,1} q_{\beta_1,\ldots,\beta_\alpha,2}, \ldots, q_{\beta_1,\ldots,\beta_\alpha,k}), \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ q_{\beta_1,\beta_2,\ldots,\beta_{s+1}} = x_{\gamma(\beta_1,\beta_2,\ldots,\beta_{s+1})}\,, \end{cases} \tag{VI}$$

where $\gamma(\beta_1, \beta_2, \ldots, \beta_{s+1})$ is a function of the indices $\beta_1, \beta_2, \ldots, \beta_{s+1}$ and takes one of the values $1, 2, \ldots, n$. As before, we assume that the functions $\{ \varphi_{\beta_1,\beta_2,\ldots,\beta_\alpha}(t_1, t_2, \ldots, t_k) \}$ are defined for all values of the arguments.

A superposition of any order, generated by functions of one variable, is again a function of one variable. Therefore in this case ($k=1$) we consider superpositions of functions of one variable and the operation of addition, that is, superpositions definable in the following way.

A function $F(x) = F(x_1, x_2, \ldots, x_n)$ $(n>1)$ is called a superposition of order $s$ of the functions $f_{\beta_1,\ldots,\beta_\alpha}(t)$ $(\alpha=0, 1, 2, \ldots, s;\ \beta_i=1, 2)$ if the following relations are satisfied:

$$\left. \begin{array}{l} F = f(q_1+q_2)\,, \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ q_{\beta_1,\beta_2,\ldots,\beta_\alpha} = f_{\beta_1,\beta_2,\ldots,\beta_\alpha}(q_{\beta_1,\beta_2,\ldots,\beta_\alpha,1} + q_{\beta_1,\beta_2,\ldots,\beta_\alpha,2}) \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ q_{\beta_1,\beta_2,\ldots,\beta_{s+1}} = x_{\gamma(\beta_1,\beta_2,\ldots,\beta_{s+1})}\,, \end{array} \right\} \tag{VII}$$

where $\gamma(\beta_1, \beta_2, \ldots, \beta_{s+1})$ takes one of the values $1, 2, \ldots, n$.

Note that we can represent as superpositions of the form (VII), for example, all rational functions of $x_1, x_2, \ldots, x_n$ since we can write any arithmetic operation by such superpositions, for example, $u \cdot v = e^{\ln u + \ln v} = f(f_1(u) + f_2(v))$.

Let $F(x_1, x_2, \ldots, x_n)$ be a superposition of order $s$ of the continuously differentiable functions $\{ f_{\beta_1,\beta_2,\ldots,\beta_\alpha}(t_1, t_2, \ldots, t_k) \}$ and $\tilde{F}(x_1, x_2, \ldots, x_n)$ the superposition of the same form of the continuously differentiable functions $\{ \tilde{f}_{\beta_1,\beta_2,\ldots,\beta_\alpha}(t_1, t_2, \ldots, t_k) \}$. We put

$$\varphi_{\beta_1,\beta_2,\ldots,\beta_\alpha} = \tilde{f}_{\beta_1,\ldots,\beta_\alpha} - f_{\beta_1,\ldots,\beta_\alpha} \quad (\alpha=0, 1, 2, \ldots, s;\ \beta_i=1, 2, \ldots, k)$$

$$\mu = \max_{\alpha, \beta_1, \ldots, \beta_\alpha} \sum_{i=1}^{k} \sup_t \left| \frac{\partial f_{\beta_1,\ldots,\beta_\alpha}(t_1, \ldots, t_k)}{\partial t_i} \right|,$$

$$\varepsilon = \max_{\alpha, \beta_1, \ldots, \beta_\alpha} \sup_t \left| \varphi_{\beta_1,\ldots,\beta_\alpha}(t_1, t_2, \ldots, t_k) \right|$$

LEMMA 5.4.1. *The inequality*

$$\sup_{x \in G} \left| \tilde{F}(x_1, x_2, \ldots, x_n) - F(x_1, x_2, \ldots, x_n) \right| \leqslant A(\mu, s)\, \varepsilon .$$

*holds, where the constant* $A(\mu, s)$ *depends only on* $\mu$ *and* $s$.

*Proof.* We proceed by induction on $s$. For definiteness suppose that $k < 1$. Having verified the statement of the lemma for $s = 1$ and having made an appropriate inductive assumption for superpositions of order $s - 1$, we have

$$\sup_{x \in G} \left| \tilde{F}(x_1, x_2, \ldots, x_n) - F(x_1, x_2, \ldots, x_n) \right|$$

$$\leqslant \left| f(\tilde{q}_1, \ldots, \tilde{q}_k) - f(q_1, \ldots, q_k) \right| + \left| \varphi(\tilde{q}_1, \tilde{q}_2, \ldots, \tilde{q}_k) \right|$$

$$\leqslant \mu \max_{\beta_1} \sup_{x \in G} \left| \tilde{q}_{\beta_1} - q_{\beta_1} \right| + \varepsilon \leqslant \mu \cdot A(\mu, s-1)\, \varepsilon + \varepsilon = A(\mu, s)\, \varepsilon .$$

(the last by the indictive assumption). This proves the lemma.

Further, let $\omega(\delta)$ be the common modulus of continuity of all the functions $\left\{ \dfrac{\partial f_{\beta_1, \ldots, \beta_\alpha}(t_1, \ldots, t_k)}{\partial t_i} \right\}$ and, in addition, put

$$\varepsilon' = \max_{\alpha, \beta_1, \ldots, \beta_\alpha} \sum_{i=1}^{k} \sup_{t} \left| \frac{\partial \varphi_{\beta_1, \ldots, \beta_\alpha}(t_1, \ldots, t_k)}{\partial t_i} \right|$$

LEMMA 5.4.2. *We have (for case* $k > 1$*)*

$$\tilde{F}(x_1, \ldots, x_n) - F(x_1, \ldots, x_n) = \sum_{\alpha, \beta_1, \ldots, \beta_\alpha} p_{\beta_1, \ldots, \beta_\alpha}(x_1, x_2, \ldots, x_n)$$

$$\times\, \varphi_{\beta_1, \ldots \beta_\alpha,} \left( q_{\beta_1, \ldots, \beta_\alpha, 1}(x_1, \ldots, x_n), \ldots, q_{\beta_1, \ldots, \beta_\alpha, k}(x_1, \ldots, x_n) \right)$$

$$+\, R(x_1, x_2, \ldots, x_n),$$

*where*

$$\left| R(x_1, x_2, \ldots, x_n) \right| \leqslant B(\mu, s, k) \left[ \varepsilon' + \omega\left( A(\mu, s)\, \varepsilon \right) \right] \varepsilon ,$$

$$p_{\beta_1, \ldots, \beta_\alpha}(x_1, x_2, \ldots, x_n) = \prod_{i=0}^{\alpha-1} \frac{\partial f_{\beta_1, \ldots, \beta_i}}{\partial q_{\beta_1, \ldots, \beta_{i+1}}}$$

(for $\alpha = 0$ $p(x_1, x_2, \ldots, x_n) \equiv 1$),

$B(\mu, s, k)$ *is a constant depending only on* $\mu, s, k$. *For* $k = 1$ *the corresponding equation is slightly different (see Chapter I, (III)) :*

$$\tilde{F}(x_1, \ldots, x_n) - F(x_1, \ldots, x_n)$$
$$= \sum_{\alpha, \beta_1, \ldots, \beta_\alpha} p_{\beta_1, \ldots, \beta_\alpha}(x_1, x_2, \ldots, x_n)\, \varphi_{\beta_1, \ldots, \beta_\alpha}(q_{\beta_1, \ldots, \beta_\alpha, 1}(x_1, \ldots, x_n)$$
$$+ q_{\beta_1, \ldots, \beta_\alpha, 2}(x_1, \ldots, x_n)) + R(x_1, \ldots, x_n)\,.$$

*Proof.* As in the preceding lemma we proceed by induction on $s$. Again for definiteness we limit ourselves to the case $k > 1$. For $s = 1$ the assertion of the lemma is easily verified. We assume that it is true for superpositions of order $s - 1$. By Lemma 5.4.1, for superpositions of order $s$ we have

$$\tilde{F}(x_1, \ldots, x_n) - F(x_1, \ldots, x_n) = f(\tilde{q}_1, \tilde{q}_2, \ldots, \tilde{q}_k) - f(q_1, q_2, \ldots, q_k)$$
$$+ \varphi(\tilde{q}_1, \tilde{q}_2, \ldots, \tilde{q}_k) = \varphi(q_1, q_2, \ldots, q_k) + \sum_{\beta_1 = 1}^{k} \frac{\partial f}{\partial q_{\beta_1}}(\tilde{q}_{\beta_1} - q_{\beta_1})$$
$$+ A(\mu, s)\, \varepsilon' \cdot \varepsilon + k \cdot A(\mu, s)\, \omega\big(A(\mu, s)\, \varepsilon\big)\, \varepsilon\,.$$

Since $\tilde{q}_{\beta_1}$ and $q_{\beta_1}$ ($\beta_1 = 1, 2, \ldots, k$) are superpositions of order $s - 1$, by the inductive hypothesis we have

$$\tilde{q}_{\beta_1} - q_{\beta_1} = \sum_{\substack{\alpha > 0 \\ \beta_2, \beta_3, \ldots, \beta_\alpha}} \hat{p}_{\beta_1, \ldots, \beta_\alpha}(x_1, x_2, \ldots, x_n)$$
$$\times \varphi_{\beta_1, \ldots \beta_\alpha}, \big(q_{\beta_1, \ldots, \beta_\alpha, 1}(x_1, x_2, \ldots, x_n), \ldots, q_{\beta_1, \ldots, \beta_\alpha, k}(x_1, x_2, \ldots, x_n)\big)$$
$$+ \acute{R}(x_1, x_2, \ldots, x_n)\,,$$

where

$$\big|\hat{R}(x_1, x_2, \ldots, x_n)\big| \leqslant B(\mu, s - 1, k)\big[\varepsilon' + \omega\big(A(\mu, s - 1)\, \varepsilon\big)\big]\, \varepsilon\,,$$
$$\hat{p}_{\beta_1, \ldots, \beta_\alpha}(x_1, \ldots, x_n) = \prod_{i=1}^{\alpha - 1} \frac{\partial f_{\beta_1, \beta_2, \ldots, \beta_i}}{\partial q_{\beta_1, \ldots, \beta_{i+1}}}$$

(for $\alpha = 1$, $\hat{p}_{\beta_1}(x_1, \ldots, x_n) \equiv 1$).

When we now substative the expressions for the differences $\tilde{q}_{\beta_1} - q_{\beta_1}$ in the formula for $\tilde{F} - F$ above, we obtain the required representation of the difference of two superpositions $\tilde{F} - F$. This proves the lemma.

## § 5. *Instability of the representation of functions as superpositions of smooth functions*

Let $A$ be a set of functions of $n$ variables and $B$ a set of functions of $k$ variables ($k<n$). Suppose that a function $F(x_1, ..., x_n) \in A$ is in a region $G_n$ of the space $x_1, x_2, ..., x_n$ an $s$-fold superposition, generated by a system of functions $\{f_{\beta_1,...,\beta_\alpha}(t_1, ..., t_k)\}$ of $B$.

We say that this superposition is $(A, B)$-stable in $G_n$ if every function $\tilde{F}(x_1, ..., x_n) \in A$ can be represented in $G_n$ as the $s$-fold superposition of the same form of functions $\{\tilde{f}_{\beta_1,...,\beta_\alpha}(t_1, t_2, ..., t_k)\}$ of $B$ such that

$$\max_{\alpha;\beta_1,...,\beta_\alpha} \sup_t \left|\tilde{f}_{\beta_1,...,\beta_\alpha}(t_1, \ldots, t_k) - f_{\beta_1,...,\beta_\alpha}(t_1, \ldots, t_k)\right|$$
$$\leqslant \lambda \sup_{x \in G_n} \left|\tilde{F}(x_1, \ldots, x_n) - F(x_1, \ldots, x_n)\right|,$$

where $\lambda$ is a constant not depending either on $\tilde{F}$ or on the $\{\tilde{f}_{\beta_1,...,\beta_\alpha}\}$.

We denote by $C^{(1)}_{\omega(\delta)}$ the space of all continuously differentiable functions of $k$ variables whose partial derivatives have modulus of continuity $\omega(\delta)$ ($\omega(\delta) \to 0$ as $\delta \to 0$).

THEOREM 5.5.1. *Suppose that each function $F(x_1, ..., x_n) \in A$ is in some region $D_n$ of the space $x_1, ..., x_n$ a superposition of order $s$ of functions of $k$ variables $\{f_{\beta_1,...,\beta_\alpha}(t_1, ..., t_k)\}$ belonging to $C^{(1)}_{\omega(\delta)}$ ($k<n$). If for any subregion $G_n \subset D_n$ the functional "dimension" of $A$ at $F(x_1, ..., x_n) \in A$ is greater than $k$, then the function $F(x_1, ..., x_n)$ cannot be an $(A, C^{(1)}_{\omega(\delta)})$-stable superposition in any such region $G \subset D_n$.*

*Proof.* Assume the contrary, that is, in a region $G_n \subset D_n$ the function $F(x_1, ..., x_n) \in A$ is an $(A, C^{(1)}_{\omega(\delta)})$-stable $s$-fold superposition of functions $\{f_{\beta_1,...,\beta_\alpha}(t_1, ..., t_k)\}$ of $C^{(1)}_{\omega(\delta)}$. Then any function $\tilde{F}(x_1, ..., x_n) \in A$ can be represented as the superposition of the same form of functions $\{\tilde{f}_{\beta_1,...,\beta_\alpha}(t_1, ..., t_k)\}$ of $C^{(1)}_{\omega(\delta)}$ such that

$$\max_{\alpha;\beta_1,...,\beta_\alpha} \sup_t \left|\varphi_{\beta_1,...,\beta_\alpha}(t_1, \ldots, t_k)\right| \leqslant \lambda \sup_{x \in G_n} \left|\tilde{F} - F\right|,$$

where $\varphi_{\beta_1,...,\beta_\alpha} = \tilde{f}_{\beta_1,...,\beta_\alpha} - f_{\beta_1,...,\beta_\alpha}$. By Lemma 5.4.2 we have (for definiteness, $k > 1$)

$$\tilde{F} - F = \sum_{\alpha;\,\beta_1,\ldots,\beta_\alpha} p_{\beta_1,\ldots,\beta_\alpha}(x_1, \ldots, x_n)$$

$$\times\ \varphi_{\beta_1,\ldots,\beta_\alpha}(q_{\beta_1,\ldots,\beta_\alpha,1}(x_1, \ldots, x_n), \ldots, q_{\beta_1,\ldots,\beta_\alpha,k}(x_1, \ldots, x_n)) + R(x_1, \ldots, x_n),$$

where $|R(x_1, \ldots, x_n)| \leqslant \gamma(\varepsilon)\,\varepsilon$, $\gamma(\varepsilon) \to 0$ as $\varepsilon \to 0$, and

$$\varepsilon = \max_{\alpha;\,\beta_1,\ldots,\beta_\alpha} \sup_t |\varphi_{\beta_1,\ldots,\beta_\alpha}(t_1, \ldots, t_k)|$$

$$\leqslant \lambda \sup_{x \in G_n} |\tilde{F}(x_1, \ldots, x_n) - F(x_1, \ldots, x_n)|.$$

That $\gamma(\varepsilon) \to 0$ as $\varepsilon \to 0$ follows from the fact that as $\varepsilon \to 0$ the quantity

$$\varepsilon' = \max_{\alpha;\,\beta_1,\ldots,\beta_\alpha} \sum_{i=1}^{k} \sup \left| \frac{\partial \varphi_{\beta_1,\ldots,\beta_\alpha}(t_1, \ldots, t_k)}{\partial t_i} \right| \to 0,$$

provided only that the modulus of continuity of the partial derivatives of the functions $\{\varphi_{\beta_1,\ldots,\beta_\alpha}(t_1, \ldots, t_k)\}$ is fixed. By 5.1.10 it follows that $r(A, F) \leqslant k$ in some subregion $G_n \subset D_n$. So we have obtained a contradiction to the assumption that $r(A, F) > k$ in any subregion $G_n \subset D_n$ and this proves the theorem.

## REFERENCES


[1] Hilbert, D. Mathematische Probleme. *Nachr. Akad. Wiss. Göttingen* (1900), 253-297; *Gesammelte Abhandlungen*, Bd. 3 (1935), 290-329.

[2] Ostrowski, A. Über Dirichletsche Reihen und algebraische Differentialgleichungen. *Math. Z. 8* (1920), 241-298.

[3] Hilbert, D. Über die Gleichung neunten Grades. *Math. Ann. 97* (1927), 243-250; *Gesammelte Abhandlungen*, Bd. 2 (1933), 393-400.

[4] Vitushkin, A. G. On Hilbert's thirteenth problem. *Dokl. Akad. Nauk SSSR 95* (1954), 701-704.

[5] Bieberbach, L. Bemerkung zum dreizehnten Hilbertschen Problem. *J. Reine Angew. Math. 165* (1931), 89-92.

[6] —— Einfluss von Hilberts Pariser Vortrag über „Mathematische Probleme". *Naturwissenschaften 51* (1930), 1101-1111.

[7] Kolmogorov, A. N. On the representation of continuous functions of several variables by superpositions of continuous functions of fewer variables. *Dokl. Akad. Nauk SSSR 108* (1956), 179-182. *Amer. Math. Soc. Transl.* (2) 17 (1961), 369-373.

[8] Arnol'd, V. I. On functions of three variables. *Dokl. Akad. Nauk SSSR 114* (1957), 679-681.

[9] Kolmogorov, A. N. On the representation of continuous functions of several variables by superpositions of continuous functions of one variable and addition. *Dokl. Akad. Nauk SSSR 114* (1957), 953-956. *Amer. Math. Soc. Transl. (2) 28* (1963), 55-59.

[10] VITUSHKIN, A. G. and G. M. HENKIN. Linear superpositions of functions. *Uspekhi Mat. Nauk 22*, N 1 (1967), 77-124.

[11] VITUSHKIN, A. G. On 13-th Problem of Hilbert. *Hilbert's problem, Izdat. Nauka*, 1969.

[12] LORENTZ, G. G. On the 13-th problem of Hilbert. "*Mathematical Developments Arising from the Hilbert Problems*", Proceedings of Symposia in Pure Mathematics, Vol. 28 (De Kalb, Illinois, May 1974), American Mathematical Society Providence, RI, 1975.

[13] WIMAN, A. Über die Anwendung der Tschirnhausen-Transformation auf die Reduktion algebraischer Gleichungen. *Nova Acta Soc. Sci. Upsal.* (1928), 3-8.

[14] CHEBOTAREV, G. N. On the problem of the resolvent. *Kazan. Gos. Univ. Uch. Zap. 114* (1954), 189-193.

[15] —— On certain questions of the problem of resolvents. *Sobranie sochinenii* (collected works), Vol. 1 (1949), 255-340.

[16] MOROZOV, V. V. On certain questions of the problem of resolvents. *Kazan. Gos. Univ. Uch. Zap. 114* (1954), 173-187.

[17] LIN, V. Superpositions of algebraic functions. *Functional analysis and its applications 10*, N 1 (1976), 37-45.

[18] ARNOL'D, V. I. Topological invariants of algebraic functions, *Functional analysis and its applications 4*, N 2 (1970), 1-9.

[19] HOVANSKI, A. G. On superpositions of holomorphic functions with radicals. *Uspehi Mat. Nauk 26*, N 2 (1971).

[20] —— On a representation of functions by quadrature. *Uspehi Mat. Nauk 26*, N 4 (1971), 251-252.

[21] VITUSHKIN, A. G. *O mnogomernykh variatsiyakh* (On multidimensional variations), Gostehizdat, Moscow 1955.

[22] IVANOV, L. D. *Variatsii mnogestv i funktsii* (Variations of sets and functions). Izdat. Nauka, Moscow 1975.

[23] KOLMOGOROV, A. N. Estimates of the minimal number of elements of $\varepsilon$-nets in various functional classes and their application of representability of functions of several variables by superpositions of functions of fewer variables. *Uspehi Mat. Nauk 10*, N 1 (1955), 192-193; *Dokl. Akad. Nauk SSSR 101* (1955), 192-194.

[24] LORENTZ, G. G. Lower bounds for the degree of approximation. *Trans. Amer. Math. Soc. 97*, N 1 (1960), 25-34.

[25] SHAPIRO, H. S. Some negative theorems of approximation theory. *Michigan Math. 11*, N 3 (1964), 211-217.

[26] EROKHIN, V. D. On the test approximation of analytic functions by rational functions with free poles. *Dokl. Akad. Nauk SSSR 128* (1959), 29-32.

[27] TIHOMIROV, V. M. Diameters of sets in functional spaces and the theory of test approximations. *Uspehi Mat. Nauk 15*, N 3 (1960), 81-120. *Russian Math. Surveys 15*, N 3 (1960), 75-111.

[28] DUNFORD, N. and J. T. SCHWARTZ. *Linear Operators I, General Theory.* Interscience Publishers, New York and London, 1958. Translation: *Lineinye operatory, Obshchaya teoriya*, Izdat. Inost. Lit. Moscow 1962, 526-527.

[29] KRONROD, A. S. On functions of two variables. *Uspehi Mat. Nauk 5*, N 1 (1950), 24-134.

[30] MENGER, K. *Kurventheorie.* Berlin-Leipzig, 1932, Chap. X.

[31] ARNOL'D, V. I. On the representation of continuous functions of three variables by superpositions of continuous functions of two variables. *Math. Sb. 48* (90) (1959), 3-74; *56* (98) (1962, 392. *Amer. Math. Soc. Transl. (2) 28* (1963), 61-147.

[32] BARI, N. Mémoire sur la représentation finie des fonctions continues. *Math. Ann. 103* (1930), 145-248, 598-653.

[33] ARNOL'D, V. I. On the representability of functions of several variables in the form of superpositions of functions of fewer variables. *Matem. Prosv. 3* (1958).

[34] LORENTZ, G. G. Metric entropy, widths and superpositions of functions. *Amer. Math. Monthly 69* (1962), 469-485.

[35] TIHOMIROV, V. M. Kolmogorov's work on the ε-entropy of functional classes and superpositions of functions. *Uspehi Mat. Nauk 18*, N 5 (1963), 55-92. *Russian Mathematical Surveys 18*, N 5 (1963), 51-87.

[36] KAHANE, J.-P. Sur le théorème de superposition de Kolmogorov. *Journal of Approximation theory 13*, N 3 (1975), 229-234.

[37] FRIDMAN, B. Improvement in the smoothness of functions in the Kolmogorov superposition theorem. *Dokl. Akad. Nauk SSSR 177* (1967), 5.

[38] DOSS, R. On the representation of the continuous of two variables by means of addition and continuous functions of one variable. *Colloq. Math. 10* (1963), 249-259.

[39] BASSALYGO, L. A. On the representation of continuous functions of two variables by means of continuous functions of one variable. *Vestnik Moskov. Univ. Ser. I, Mat. Meh. 21* (1966), 58-63.

[40] SPRECHER, D. A. On the structure of continuous functions of several variables. *Trans. Amer. Math. Soc. 115* (1965), 340-355.

[41] ARNOL'D, V. I. On the representability of functions of two variables in the form $\kappa(\varphi(x) + \psi(y))$. *Uspehi Mat. Nauk 12*, N 2 (1957), 119-121.

[42] OFMAN, Yu. P. On the best approximation of functions of two variables by functions of the form $\varphi(x) + \psi(y)$. *Izv. Akad. Nauk SSSR Ser. Mat. 25* (1961), 239-252.

[43] MOTORNYI, V. P. On the question of the best approximation of functions of two variables by functions of the form $\varphi(x) + \psi(y)$. *Izv. Akad. Nauk SSSR Ser. Mat. 27* (1963), 1211-1214.

[44] SPRECHER, D. A. On the structure of representations of continuous functions of several variables as finite sums of continuous functions of one variable. *Proc. Amer. Math. Soc. 17* (1966), 98-105.

[45] HENKIN, G. M. On linear superpositions of continuously differentiable functions. *Dokl. Akad. Nauk SSSR 157* (1964), 288-290. *Soviet Math. Dokl. 5* (1964), 948-950.

[46] FRIDMAN, B. The set of linear superpositions is nowhere dense in the space of continuous functions of several variables. *Izv. Akad. Nauk SSSR Ser. Mat. 36*, N 4 (1972), 814-846.

[47] VITUSHKIN, A. G. Some properties of linear superpositions of smooth functions. *Dokl. Akad. Nauk SSSR 156* (1964), 1003-1006. *Soviet Math. Dokl. 5* (1964), 741-744.

[48] —— Proof of the existence of analytic functions of several variables not representable by linear superpositions of continuously differentiable functions of fewer variables. *Dokl. Akad. Nauk SSSR 156* (1964), 1258-1261. *Soviet Math. Dokl. 5* (1964), 793-796.

[49] —— *Otsenka slozhnosti zadachi tabulirovaniya* (Estimation of the complexity of the tabulation problem). Fizmatgiz, Moscow 1959. *Theory of the transmission and processing of information*. Pergamon press, 1961.

[50] KOLMOGOROV, A. N. and V. M. TIHOMIROV. ε-entropy and ε-capacity of sets in function spaces. *Uspehi Mat. Nauk 14*, N 2 (1959), 3-86.

[51] BUSLAEV, V. I. and A. G. VITUSHKIN. An estimate of the code length of signals with finite spectrum in connection with sound recording problems. *Izv. Akad. Nauk SSSR Ser. Mat. 38*, N 4 (1974), 867-895.

[52] VITUSHKIN, A. G. Coding of signals with finite spectrum and sound recording problems. *Proceedings of the International Congress of Mathematicians*, Vancouver, 1974.

[53] SMUSHKO, V. V. Entropy of the class of entire functions with the frequency dependent metric. *Izv. Akad. Nauk SSSR Ser. Mat. 40*, N 5 (1976), 1173-1186.

[54] HEDBERG, T. Sur les réarrangements de fonctions de la classe $A$ et les ensembles d'interpolation pour $A(D^2)$. *C. R. Acad. Sc. Paris 270* (1970), 1491-1494.

*(Reçu le 28 avril 1977)*

A. G. Vitushkin

Steklov Mathematical Institute
Ul. Vavilova, 42
Moscow 117333, USSR