ON THE REPRESENTATION OF FUNCTIONS OF SEVERAL VARIABLES AS A SUPERPOSITION OF FUNCTIONS OF A SMALLER NUMBER OF VARIABLES*

V.I. Arnol'd

Moscow

translated by Gerald Gould

In this paper we wish to give an account of several recent papers by Moscow mathematicians devoted to the question in the title of this paper. §1 contains the definition of superposition of functions and the statement of Hilbert's 13th problem relating to superpositions. §2 is devoted to superpositions of smooth functions. In §3 we present several very recent papers, in spite of the fact that the content of that section is now perhaps only of historical interest. The principal topic there is the description given by Kronrod of "the tree of components of a function of several variables", which is a concept whose popularization would seem to be very desirable (although the connection between this concept and the problems considered in our paper has proved to be less close than it originally appeared). The reader interested only in the strongest (and, moreover, the simplest in its method of proof) result relating to the representation of continuous functions of several variables as superpositions of functions of a smaller number of variables can, after looking at the introductory §1 go straight to §4, missing out §2–3. In addition, the smaller print in this paper means, as usual, that the corresponding material is auxiliary and omitting it will not affect the reader's understanding of what follows.

1. One of the problems of the famous problem book by Pólya and Szegö[1] begins as follows:

"*Do functions of three variables exist at all?*"

The meaning of this question is as follows. Suppose that we have two functions of two variables $u(x, y)$ and $v(y, z)$. We now consider a new function of two variables $w(u, v)$ and substitute our functions in place of u and v. Then the function $w[u(x, y), v(y, z)]$ now depends on the three variables x, y and z. Thus, by substituting in place of the arguments u and v of the function of two variables $w(u, v)$ the new functions of two variables we obtain

* Mat. Prosveshchenie **3**, 41–61 (1958)

[1] Pólya, G., Szegö, G.: Problems and theorems of analysis, part I. Moscow, Section II, Problems 119 and 119a.

a function of three variables (one can even obtain a function of four variables $w[u(x,y), v(z,t)]$; we call this function a *single superposition* formed from the functions of two variables u, v and w. It is clear that all the properties of this function are determined by our three functions of two variables. Pólya and Szegö's question (which, however, was not formulated in their book in all its breadth) is as follows: can all functions of three variables be reduced to such a superposition (or a somewhat more complicated superposition) of functions of two variables, or do there in fact exist functions that are "essentially of three variables" which cannot be reduced to functions of two variables.

Note first of all that if one can also use *discontinuous* functions, then the answer to Pólya and Szegö's question is clearly negative.[2] Thus the only question of interest is whether or not all *continuous* functions of three variables are representable as superpositions of *continuous* functions of two variables.

In fact, a discontinuous function $u = \phi(x,y)$ enables one to map the (x,y) plane bijectively onto the line u [the fact that the set of pairs (x,y) of numbers and the set of numbers u have the same cardinality means precisely that these sets can be bijectively mapped onto each other]. We now choose *any* function of three variables $F(x,y,z)$ and *define* the function $\psi(u,z)$ by the equality

$$\psi[\phi(x,y), z] = F(x,y,z)\,;$$

this is possible because each pair of values (x,y) corresponds to a unique value $u = \phi(x,y)$ and we can take $\psi(u,z)$ to be equal to the corresponding value of $F(x,y,z)$.[3]

For the *simplest* continuous functions of three variables it is not hard to find representations of them as superpositions of continuous functions of two variables. For example, the function

$$F(x,y,z) = xy + yz$$

can be represented in the form

$$F = w[u(x,y), v(y,z)]\,,$$

where

$$w(u,v) = u + v, \quad u(x,y) = x + y, \quad v(y,z) = yz\,.$$

For the somewhat more complicated function

$$F(x,y,z) = xy + yz + zx$$

it is already impossible to represent it as a *simple* superposition of functions of two variables;[4] However, it is possible to represent it as a *double* superposition of functions of two variables, that is, in the form

[2] See the solution of problem 119 in Pólya and Szegö's book.

[3] It suffices to require that no two distinct pairs (x,y) correspond to the same value $u = \phi(x,y)$; here, for values $\bar{u}$ not belonging to the range of the function $\phi(x,y)$ the function $\psi(\bar{u},z)$ can be defined arbitrarily.

[4] See Pólya and Szegö's book.

$$w\{u[p(x,y),q(y,z)],\ v[r(y,z),s(z,x]\}\,;$$

it suffices merely to set

$$w(u,v)=u+v$$

and

$$u(p,q)=p+q,\quad p(x,y)=xy,\quad q(y,z)=yz,\quad v(r,s)=s,\quad s(z,x)=zx\,.$$

In general, all *entire rational functions* of several variables can by definition be obtained from their arguments by means of a multiple application of the operations of addition and multiplication, that is, they are the result of a multiple superposition of functions of not more than two variables

$$\phi(x,y)=x+y,\quad \psi(x,y)=xy,\quad f(x)=x+a,\quad g(x)=ax\,,$$

that is, the result of a multiple substitution of the arguments of these functions by more complex expressions formed by means of the same functions. By analogy with this, the *rational functions* are obtained as superpositions of six of the simplest functions of not more than two variables:

$$\phi(x,y)=x+y,\quad \psi(x,y)=xy,\quad \chi(x,y)=\frac{x}{y},$$
$$f(x)=x+a,\quad g(x)=ax,\quad h(x)=\frac{a}{x}\,.$$

If a segment of x is a function of known segments $a,b,c,\ldots$, then in order to be able to construct it using a ruler and compasses, it is necessary and sufficient that this function be homogeneous of the first dimension and that it be a superposition of these same simplest functions and the function $y=\sqrt{x}$. All the *elementary functions* can be represented as superpositions obtained via those same functions and in addition certain special functions of one variable, such as

$$\sqrt[n]{x},\quad e^x,\quad \ln(x),\quad \sin(x),\quad \text{and others.}$$

The simplest examples of algebraic functions going outside the limits of the class of elementary functions are provided by the roots of algebraic equations; the arguments of these functions are the values of the coefficients of the equations. But the roots of equations of the first, second, third and fourth degrees are, as is well known, elementary functions of the coefficients obtained as the result of superposition of those same functions of two variables, the sum, the difference, the product and the quotient, and (for equations of these 2nd–4th degrees) functions of the single variable $\sqrt[n]{x}$ (here $n=2$ in the case of a quadratic equation and can be equal to 2 or 3 in the case of equations of the 3rd and 4th degrees). For equations of the 5th and higher degrees such a representation is not possible in general; this was shown by Abel. However, the roots of equations of the 5th and 6th degrees can be expressed in terms of the coefficients by means of superpositions of certain more complex analytic functions of two variables; these representations can be used for the practical calculation of the roots of equations; in particular for nomographic solution of equations.

Attempts to obtain a representation of roots of 7th-degree equations as a superposition of suitable functions have not been crowned with success. Using algebraic transformations the general 7th-degree equation

$$x^7 + a_1x^6 + a_2x^5 + a_1x^6 + a_3x^4 + a_4x^3 + a_6x + a_7 = 0$$

can be reduced to the form

$$x^7 + ax^3 + bx^2 + cx + d = 0$$

where a, b and c are elementary (algebraic) functions of the coefficients $a_1, a_2, \ldots, a_7$ of the original equation, therefore they are expressed in terms of these coefficients as superpositions composed of simple functions of two variables. Thus, the question of the possibility of representing the roots of a 7th-degree equation by superpositions of functions of two variables reduces to the problem of finding such a representation for the special function of three variables a, b, c of the roots of the equation written above.

To date it is not known whether this function of three variables (which is easily seen to be analytic) can be represented as a superposition of analytic functions of two variables. Nevertheless, Hilbert managed to show that *certain* analytic functions of three variables are not such superpositions.

Hilbert's result is in connection with the following situation. If a function of *three* variables is a superposition of functions of two variables, then among the partial derivatives of the superposition and the functions of which it is composed there exist fully determined analytic relations. Therefore if *all* analytic functions of three variables are representable in such a form, then the space of coefficients of the series expansion of the functions of two variables involved in this superposition can be mapped analytically onto the space of coefficients of the expansion of functions of three variables; but this is not possible, since the latter space has a greater dimension (here we are restricted by the definite but large number of first coefficients of the expansion, that is, the first partial derivatives).

In his lecture at the 1900 International Mathematical Congress held in Paris the celebrated German mathematician David Hilbert posed 23 problems awaiting solution.[5] The thirteenth of these "Mathematical problems" of Hilbert's was as follows:

Can the roots of the equation

$$x^7 + ax^3 + bx^2 + cx + 1 = 0$$

be represented as superpositions of continuous functions of two variables?

[5] Hilbert, D.:Mathematische Problemen; Gesammelte Abhandlungen, vol.3, No.17 (1935). [*Editor's note*: Translation of this work of Hilbert's will appear in the next issues of Mat. Prosveshch.]

Hilbert conjectured that the answer to this question would turn out to be negative; in that case the more general question of whether all functions of three variables are superpositions of continuous functions of two variables would be solved at the same time.

2. The first results touching on Hilbert's 13th problem were obtained under the assumptions that the superpositions have some special form, for example, under conditions restricting the 'single' superpositions; they all supported Hilbert's conjecture.[6] The strongest result here is the result of A.G. Vitushkin who succeeded in constructing a polynomial such that neither the polynomial itself nor any function sufficiently close to it (in the sense of uniform convergence) can be decomposed into a simple superposition of continuous functions of two variables in any region or in any system of coordinates.

Further results are in connection with restrictions imposed on the functions involved in the superposition. As already recalled, Hilbert had noted earlier that it was impossible to obtain all the *analytic* functions of three variables as superpositions of *analytic* functions of two variables. Important results in this direction were obtained by Vitushkin,who by using his theory of multidimensional variations of functions and sets showed that not all l times continuously differentiable functions of three variables can be represented as superpositions of $\left[\frac{2}{3}l\right]$ times[7] differentiable functtons of two variables all of whose derivatives of order $\left[\frac{2}{3}l\right]$ satisfy Lipschitz condititons.[8]

In Kolmogorov's interpretation[9] Vitushkin's results are connected with the difference of the 'dimensions' of the corresponding function spaces. As Pontryagin and Shnirel'man had already proved in 1932, the dimension of a compact metric space (for example, a cube in Euclidean space) can be defined in the following way. We cover our space with 'small' sets of diameter ε. Clearly, the number $N(\varepsilon)$ of sets required to do this will increase as ε gets smaller; here it can be shown that $N(\varepsilon)$ increases as $\frac{1}{\varepsilon^n}$, where n is the

[6] The simplest examples of this kind already appear in the book of Pólya and Szegö; a number of other examples (due to A.Ya. Dubovitskiĭ and R.A. Minlos) are given in the book:Vitushkin, A.G.: On multidimensional variations. Moscow (1955).

[7] Here the square brackets indicate the integer part.

[8] It also follows from this result that there exists in a three-dimensional cube an analytic function (of three variables) satisfying a Lipschitz condition with Lipschitz constant 1 such that no functions close to it (including the function itself) can be represented as an s-fold superposition of two variables satisfying a Lipschitz condition with some constant L_1 (s and L_1 are given in advance), and there exists an unbounded differentiable function satisfying a Lipschitz condition with Lipschitz constant 1 which is not a superposition of functions of two variables satisfying a Lipschitz condition. See Vitushkin's book referred to in footnote 6.

[9] Kolmogorov, A.N.: Estimates of the minimum number of elements of ε-nets in various function classes and their application to the question of the representation of functions of several variables as superpositions of functions of a smaller number of variables. Usp. Mat. Nauk **10**, No.1, 192–195 (19??).

dimension of the space; thus the dimension n can be *defined* as the limit

$$\liminf_{\varepsilon\to 0}\left[-\frac{\log N(\varepsilon)}{\log\varepsilon}\right].$$

For infinite-dimensional spaces this limit is equal to infinity. However, in a number of cases the number $N(\varepsilon)$ can increase as the function $1:\exp(z^k)$, where k is some constant which one can provisionally call the "dimension of the infinite-dimensional space". Thus for infinite-dimensional spaces the role of dimension is played by the limit[10]

$$\liminf_{\varepsilon\to 0}\left[-\frac{\log\log N(\varepsilon)}{\log\varepsilon}\right].$$

For the space of functions $f(x_1,x_2,\ldots,x_n)$ of n arguments defined on an n-cube, where the functions are l times differentiable in all their arguments and are such that all the partial derivatives of order l satisfy a Hölder condition of order α,[11] the above-defined dimension can be considered to be equal to

$$\frac{n}{l+\alpha}.$$

Hence it follows, in particular, that the set of l times differentiable functions of three arguments is in a certain sense 'richer in its elements' than the set of $\left[\frac{2}{3}l\right]$ times differentiable functions of two arguments satisfying a Lipschitz condition (that is, a Hölder condition of order 1); hence it follows that it is impossible to express all the first functions as superpositions of the last ones.

---

[10] Instead of the number $N(\varepsilon)$ of sets of diameter ε completely covering the (compact) space one could choose the number $M(\varepsilon)$ of points of an ε-net, that is, the smallest number of points such that each point of the space is at a distance of at most ε from at least one of the chosen points, or the maximum number $K(\varepsilon)$ of points such that the distance between any two of them is greater than ε. It is curious to note that the same definition of the dimension of function spaces was arrived at (almost at the same time) by Shannon [Shannon, C.E.: The mathematical theory of communication, Urbana (1949); in the Russian translation of Shannon's work (in the collection "Theory of transmission of electric signals in the presence of noise". Inost. Lit., Moscow (1953)) the corresponding place was omitted for some reason] which started from arguments relating to "the theory of information": in the space of the transmitted information $K(\varepsilon)$ is the maximum number of 'ε-different signals' that cannot be confused by the receiver provided that the distortion of the information in the transmitter does not exceed ε.

[11] A function $f(x)$ satisfies a Hölder condition of order α if there exists a number C such that for each x_1,x_2 in the domain of the function

$$|f(x_1)-f(x_2)|<C|x_1-x_2|^2.$$

A function of several variables is said to satisfy a Hölder condition if it satisfies this condition as a function of each of its variables.

3. However, in the domain of all continuous functions Hilbert's conjecture has proved to be false.

In the spring of 1956 Kolmogorov succeeded in showing that every continuous function of n variables ($n \geqslant 4$) defined on an n-cube is a superposition of continuous functions of the three variables.[12] The main tool in his construction is the one-dimensional *tree of components of level sets of a function introduced by Kronrod.*[13]

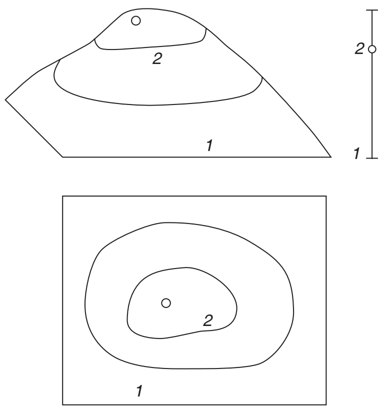

Fig. 1.

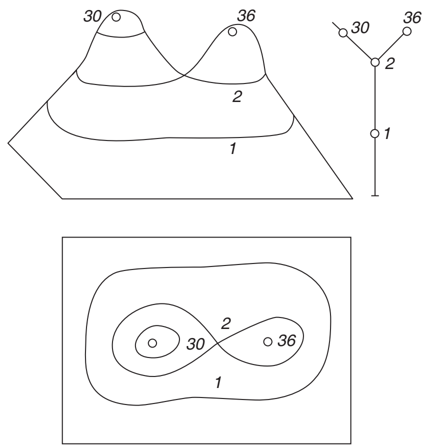

Fig. 2.

By the *level set* of a function we mean the collection of all points in the domain of the function at which the function takes some fixed value. For example, if the function of a point of part of the land surface represents the height at this point above sea level, then the level set will consist of all points of the locality having the same height above sea level; in topography these level sets are called *contour lines*. In Figs. 1 and 2 we have depicted simple functions of two variables and the 'maps' of the level sets of these functions (that is, a partition of the squares on which the functions are defined into

[12] Kolmogorov, A.N.: On the representation of continuous functions of several variables by superpositions of continuous functions of a smaller number of variables. Dokl. Akad. Nauk SSSR **108**, 179–182(1956); English transl. in Amer. Math Soc. transl. Ser. 2, vol. 17, 369–373 (1961).

[13] Kronrod, A.S.: On functions of two variables. Usp. Mat. Nauk **5**, No.1, 24–134 (1950).

their separate level sets). A level set can consist of a single piece (for example, all the level sets of the function depicted in Fig. 1 or the 1-level set of the function depicted in Fig 2; or it may consist of several connected pieces or components (for example, the 3-level set in Fig. 2 consists of the two pieces 3a and 3b). To study the structure of the set of components of a level set of a continuous function Kronrod proposed that one use the notion of a tree.

In topology, by a *tree* we mean a curve ('one-dimensional locally connected continuum') not containing any closed arcs ('homeomorphic images of a circle'). A tree can be represented in the following way. From the base of the 'trunk' of the tree there emerge 'branches' at the 'branch points' (the number of branch points can be denumerable and from each such point there can be denumerably many branches coming out of it); in turn, from each branch there can emerge new branches (we can call them 'twigs'), and so on (Fig. 3). In general a tree can be somewhat complex; however, as the celebrated Austrian (now American) mathematician Karl Menger showed, there exists in the plane a *universal tree* such that any other tree is a part of it (more precisely, such that any tree is homeomorphic to a part of the universal tree).[14]

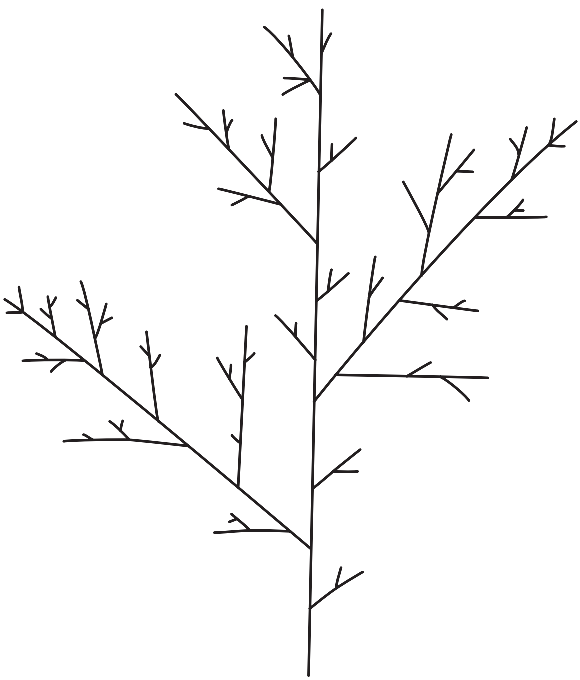

Fig. 3.

[14] Menger, K.: Kurventheorie, Ch. X. Berlin–Leipzig (1932).

Kronrod showed that *the set of components of all level sets of a continuous function of several variables is naturally representable as a tree.*

Thus, for example, the set of components of the level set of the functions depicted in Fig. 1 corresponds to a segment (the set of level 1 corresponds to the point 1 of this segment and the set of level 2 corresponds to the point 2); the set of components of the level sets of the somewhat more complicated function depicted in Fig. 2 corresponds to a Y-shaped tree (the set of level 1 corresponds to point 1 of the tree, the "figure 8" set of level 2 corresponds to the branch point 2: the components $3a$ and $3b$ of the set of level 3 correspond to the points $3a$ and $3b$ of the tree).

In more precise terms, one can introduce on the set of components a natural topology after which it becomes a topological space T which Kronrod called the *one-dimensional tree of the function.*

A study of the structure of this space can be carried out in the following way. First, T is the continuous image of an n-dimensional cube and therefore T is a locally connected continuum. Second, under the map d of the cube onto T the inverse image of each point of T is a component, that is, a closed connected set. We call such maps *monotone.*[15] Visually they can be represented as a contraction without gluing: for example, the projection of a square onto one of its sides is a monotone map, while the formation of a cylinder from a square by gluing is not a monotone map. One can prove that simple connectedness is preserved under a monotone map; therefore T, which is the monotone image of a cube, is *a simply connected set.* Finally, under a mapping of T onto a segment different components of the same level are taken to each point of the segment, that is, a zero-dimensional subset of T (not containing connected pieces) and, as is well known, under a map with zero-dimensional inverse images the dimension is not lowered. Therefore T is *one-dimensional.* Thus T is a one-dimensional and simply connected locally connected continuum. Hence T is a *tree.*

We can regard each function $f(x_1, x_2, \ldots, x_n)$ as a superposition of two maps: 1) a map $d(x_1, x_2, \ldots, x_n)$ of the domain of definition *onto the tree of components of the level sets* of f; under the map d the image of each point of the domain of definition is the component of the level set containing this point; 2) the map $f(d)$ of the set of components onto the segment that is the range of the function $f(x_1, x_2, \ldots, x_n)$. Under this map all the components of the level set $f(x_1, x_2, \ldots, x_n) = t$ are taken to the point t.

Thus, for example, the function of two variables $f(x, y) = xy$ defined on the square $-1 \leqslant x \leqslant 1, -1 \leqslant y \leqslant 1$ can be represented as a superposition of two maps: the map of the square onto the X-shaped tree of the components of the level sets of this function (Fig. 4) [under which all the points of the 'cross' $xy = 0$ or one of the branches of the hyperbola $xy = const$ are taken to a single point of the tree], and the map of this tree onto the segment $-1 \leqslant t \leqslant 1$ [under which two points of the tree corresponding to branches of the same hyperbola (or one branch point corresponding to the cross $xy = 0$)are taken to the same point of the segment.

[15] *Editor's note*: Since (non-strictly) monotone continuous functions of a single variable have this property [see the remark by Keldysh on p.261 of the current issue].

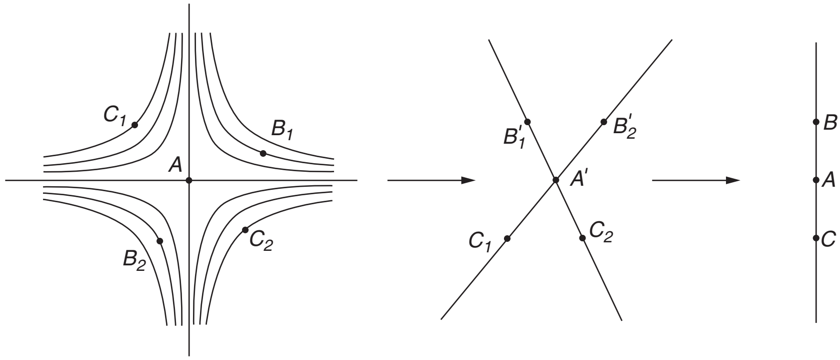

Fig. 4.

Thus, each function $f(x_1, x_2, \ldots, x_n)$ of n variables can be represented as a superposition of two new functions: the function $d(x_1, x_2, \ldots, x_n)$, which defines a map of the domain of definition of $f(x_1, x_2, \ldots, x_n)$ onto the tree of components of the level sets of this function, and $f(d)$, which is the map of the tree onto a segment (since each point d of the tree belonging to a given level set corresponds to a single value of $f(d)$ of the function f. Since a tree can be embedded in a plane, the points of this plane can be defined by the coordinates $u(d)$ and $v(d)$; this means that the second map $f(d)$ can be regarded as a function of two variables $f(u, v)$, while the first map $d(x_1, x_2, \ldots, x_n)$ can be regarded as two functions of n variables $u(x_1, x_2, \ldots, x_n)$ and $v(x_1, x_2, \ldots, x_n)$.

Kolmogorov managed to represent each function of n variables as a sum of $n+1$ functions each of which has *standard* (that is, not dependent on the function in question) components of the level sets:

$$f(x_1, x_2, \ldots, x_n) = \sum_{r=1}^{n+1} f^r(x_1, x_2, \ldots, x_n)\,;$$

thus, each function of two variables $f(x, y)$ can be represented as a function of three functions $f^1(x, y), f^2(x, y)$ and $f^3(x, y)$ where the 'maps' of the level sets of these three functions do not depend on f, but have some predetermined form, as illustrated in Fig. 5. Here for each function $f^r(x_1, x_2, \ldots, x_n)$ ($r = 1, 2, \ldots, n+1$) the map $d^r(x_1, x_2, \ldots, x_n)$ of the domain of definition onto the tree will not depend on the function f; on the other hand, the second map $f^r(d)$ of the tree onto the range of f^r does depend on f.

We now regard the function of n variables $f(x_1, x_2, \ldots, x_n)$ as a *one-parameter* (depending on the parameter x_n!) *family of functions of* $n-1$ *variables*:

$$f(x_1, x_2, \ldots, x_n) = f_{x_n}(x_1, x_2, \ldots, x_{n-1})\,.$$

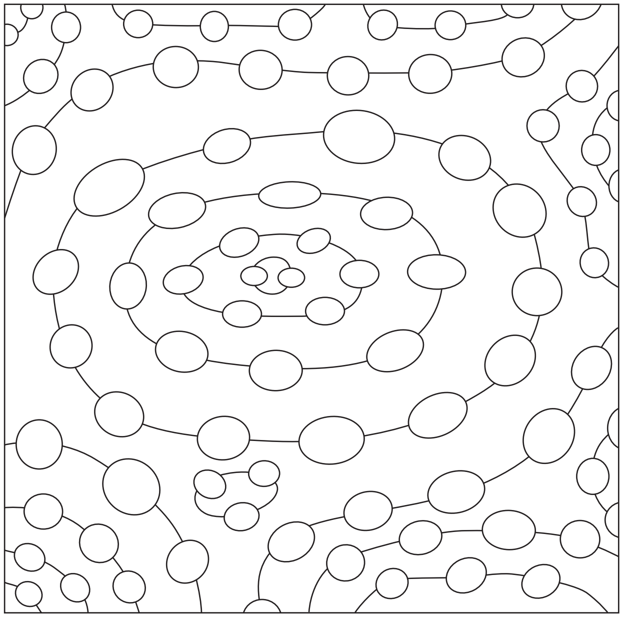

Fig. 5.

In this case we have

$$
\begin{aligned}
f(x_1, x_2, \ldots, x_n) &= f_{x_n}(x_1, x_2, \ldots, x_{n-1}) \\
&= \sum_{r=1}^{n+1} f^r_{x_n}(x_1, x_2, \ldots, x_{n-1}) \\
&= \sum_{r=1}^{n+1} f^r_{x_n}(d^r(x_1, x_2, \ldots, x_{n-1})) \\
&= \sum_{r=1}^{n+1} f^r(d^r(x_1, x_2, \ldots, x_{n-1}), x_n)\,, \qquad (1)
\end{aligned}
$$

where $d^r(x_1, x_2, \ldots, x_{n-1})$ is a map of the domain of definition of the function $f^r_{x_n}(x_1, x_2, \ldots, x_{n-1})$ which, as we have said, is *independent of the value of the parameter* x_n (the components of the level sets of the function f^r are standard!) and $f^r_{x_n}(d^r) = f^r(d^r, x^n)$ is the map of the point of the 'standard tree' d^r onto the range of f^r (which now depends on x_n). By introducing the system of coordinates (u^r, v^r) onto the plane of the tree d^r we obtain:

$$
f(x_1, x_2, \ldots, x_n) \ = \ \sum_{r=1}^{n} f^r(u^r(x_1, x_2, \ldots, x_{n-1}),\ v^r(x_1, x_2, \ldots, x_{n-1})\,, x_n)\,; \qquad (2)
$$

in other words, we have a *representation of an arbitrary function f of n variables as a sum of n functions each of which can be represented as a superposition of a function of three variables $f^r(u^r, v^r, x_n)$ and two functions $u^r(x_1, x_2, \ldots, x_{n-1})$ and $v^r(x_1, x_2, \ldots x_{n-1})$ of $n-1$ variables.* In the case when $n > 3$ we can apply the same process to the functions u^r and v^r of $n-1$ variables, so that we can eventually represent *a function of n variables $f(x_1, x_2, \ldots, x_n)$ as a superposition of functions of three variables.* Thus, the function $f(x_1, x_2, x_3, x_4)$ can now be represented in the form

$$f(x_1, x_2, x_3, x_4) = \sum_{r=1}^{4} f^r(u^r(x_1, x_2, x_3), v^r(x_1, x_2, x_3), x_4)\,; \qquad (2a)$$

[we recall once more that the function of four variables $f = f^1+f^2+f^3+f^4$ can be obtained as a superposition consisting of a single function of two variables $\phi(f^1, f^2) = f^1 + f^2$]. For $n = 3$ we obtain in this way only the representation

$$f(x, y, z) = \sum_{r=1}^{3} f^r(d^r(x, y)\ \ z)\,, \qquad (3)$$

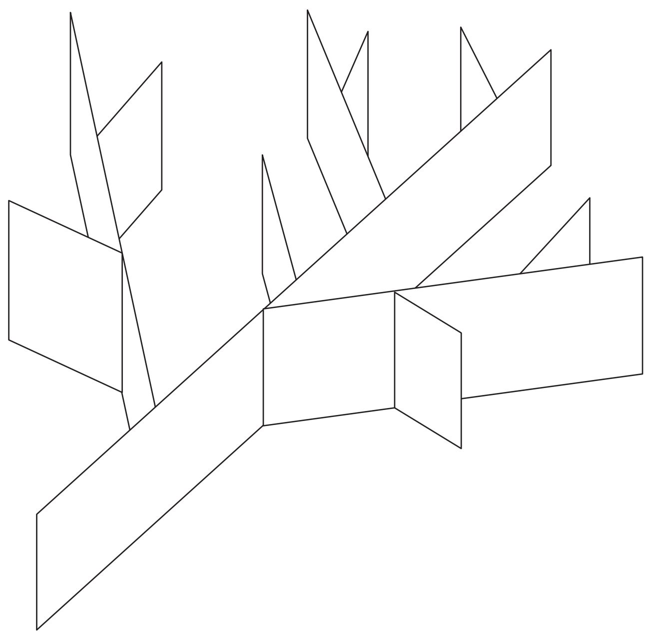

Fig. 6.

where $d^r(x, y)$ is a map of the square (x, y) onto the tree (which can be defined by two functions of two variables) and the $f^r(d^r, z)$ are defined on the set of pairs (d^r, z) where z ranges over the segment and d^r ranges over the tree, that is, functions of three variables that can, however, be defined on some special *two-dimensional* set, which is the direct product of the tree and the segment (see Fig. 6).

Recently it became clear[16] that the results of Kolmogorov can be improved: *any continuous function of n variables can be represented as a sum of $3n$ functions each of which can be represented as a superposition obtained by substituting in the function of two variables in place of one of the arguments the function of $n-1$ variables.*

The proof of this result is based on the fact that the trees d^r featuring above can be located in a three-dimensional cube (u, v, w) so that each function defined on any of them can be decomposed into a sum of three functions depending only on one of the coordinates

$$f^r(d^r) = \phi^r(u^r) + \psi^r(v^r) + \chi^r(w^r)\,. \tag{4}$$

Hence from (1) we obtain:

$$\begin{aligned} f(x_1, x_2, \ldots, x_n) &= \sum_{r=1}^{n} f^r_{x_n}(d^r(x_1, x_2, \ldots, x_n)) \\ = \sum_{r=1}^{n}[\phi^r_{x_n}(u^r(x_1, x_2, \ldots, x_{n-1})) &+ \psi^r_{x_n}(v^r(x_1, x_2, \ldots, x_{n-1})) \\ &+ \chi^r_{x_n}(w^r(x_1, x_2, \ldots, x_{n-1})) \\ = \sum_{r=1}^{n}[\phi^r(u^r(x_1, x_2, \ldots, x_{n-1}), x_n) &+ \psi^r(v^r(x_1, x_2, \ldots, x_{n-1}), x_n) \\ &+ \chi^r(w^r(x_1, x_2, \ldots, x_{n-1}), x_n)\,. \end{aligned}$$

In particular, as applied to functions of three variables we obtain instead of (3):

$$f(x, y, z) = \sum_{r=1}^{3}[\phi^r(u^r(x, y), z) + \psi^r(v^r(x, y), z) + \chi^r(w^r(x, y), z)]\,. \tag{5}$$

Thus, *each continuous function of three variables can be represented as a sum of 9 functions each of which is a single superposition of functions of two variables.* This then is the answer to the question posed by Hilbert.

In the proof of the decomposition (4) an essential role is played by the fact that in Kolmogorov's construction one can, as it turns out, avoid only trees having exceptional branch points of the third order (that is, points at which a single branch emerges from the main 'trunk'). Next it is easy to see that the simplest 'Y-shaped' tree can be arranged in the square (u, v) so that any function $f(u, v)$ defined on it

[16] Arnold, V.I.: On functions of three variables. Dokl. Akad. Nauk SSSR **114**, 679–681 (1957).

can be represented as the sum of two functions of a single variable: in fact, if in Fig. 7a we define arbitrarily the function $\phi(u)$ on the interval $\left(0, \frac{1}{2}\right)$, then we can define the function $\psi(v)$ on the interval $\left(0, \frac{1}{2}\right)$ since the sum $\phi(u)+\psi(v)$ on OA of the tree coincides with $f(u, v)$; next, we define the function $\psi(v)$ on the interval $\left(\frac{1}{2}, 1\right)$ so that the sum of $\phi(u)+\psi(v)$ on the interval AB of the tree coincides with $f(u, v)$; finally, we can define $\phi(u)$ on the interval $\left(\frac{1}{2}, 1\right)$ so that the sum $\phi(u) + \psi(v)$ on the interval AC of the tree coincides with $f(u, v)$; thus the function $f(u, v)$ defined on the tree can be represented as the sum $\phi(u) + \psi(v)$. If the tree has two branch points, that is, it has the form depicted in Fig. 7b, then the function $f(u, v)$ defined on it can also be represented as a sum $\phi(u) + \psi(v)$; we merely need to start from the definitions of the functions $\phi(u)$ and $\psi(v)$ on the interval $\left(\frac{3}{4}, 1\right)$, assuming that on the interval DC of the tree the sum $\phi(u) + \psi(v)$ coincides with $f(u, v)$, and then define the functions $\phi(u)$ and $\psi(v)$ in the same way as before, so that the sum $\phi(u) + \psi(v)$ on the entire tree coincides with the function $f(u, v)$. In the same way, any function defined on a tree with *finitely many* third-order branch points can be represented as a sum of two functions of one variable. For functions defined on a tree with *infinitely many* branch points, the above procedure fails; nevertheless, such a tree can be located in a *three-dimensional* cube such that a function defined on it can be represented as a sum of three functions depending on the separate coordinates.

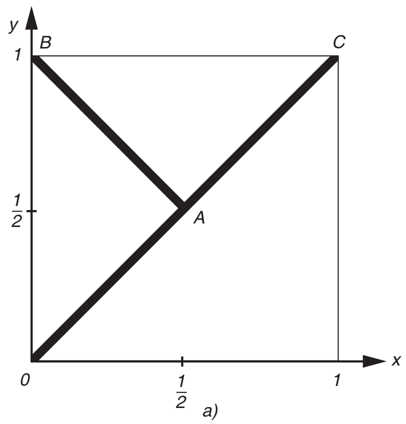

Fig. 7a.

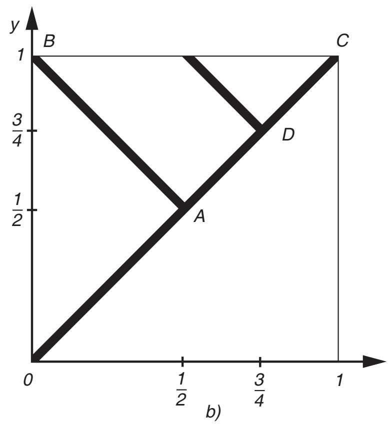

Fig. 7b.

It turns out that the complicated constructions that we have been talking about are superfluous for obtaining the final result. In the next section we give a much more direct route enabling one to obtain stronger theorems.

4. The above discussion enables one to answer *in the negative* the question posed by Pólya and Szegö whether there exist functions of three variables; more precisely this means that all continuous functions of three variables can be reduced to superpositions of continuous functions of two variables and all the properties of a function of three variables $f(x_1, x_2, x_3)$ are completely determined by certain functions of two variables, namely, the nine functions $u^r, v^r, w^r (r = 1, 2, 3)$ and the nine functions ϕ^r, ψ^r, χ^r $(r = 1, 2, 3)$ featuring in the representation (5). It is now natural to pose the question: *do there exist functions of two variables ?*

The precise meaning of the latter question is as follows. A superposition

$$F[f_1(f_2(\dots f_{n-1}(f_n))) \dots]$$

of any number of functions of one variable is, of course, a function of one variable and one cannot obtain functions of more than one variable in this way. However, if we add to our supply of functions of one variable just one 'standard' function of two variables, say, the sum

$$g(x, y) = x + y,$$

then superpositions composed of $g(x, y)$ and functions of one variable can now be functions of any number of variables; thus, for example, the $(n-1)$-fold superposition of the function g

$$g(g(g \dots g(g(x_1, x_2), x_3), \dots, x_{n-1}), x_n) = x_1 + x_2 + \dots + x_{n-1} + x_n$$

is a function of n variables. Here there arises the question: *can all continuous functions of two or more variables be represented as superpositions of this kind ?* This is the question we have in mind when we ask whether there exist ('artificial') functions of two variables. [More precisely, here we could ask: is our supply of functions of two variables essentially exhausted by one such function $g(x, y) = x + y$?]

If we restrict ourselves to the simplest representations of functions of two variables as a superposition of the function $g(x, y)$ and continuous functions of one variable, then the answer to the question of the possibility of obtaining all functions of two variables will be negative; thus, one can show by quite elementary means that the set of functions defined on a square that are representable in the form $f[\phi(x) + \psi(y)]$ $(f, \phi, \psi$ are continuous functions of one variable) not only fails to coincide with the set of all continuous functions, but is even nowhere dense and non-closed.[17] On the other hand, Kolmogorov had proved even before he had obtained the representation (2) that any continuous function of n variables can be *approximated to within any degree of accuracy* by a superposition of continuous functions of one variable and the sum $g = x + y$; thus, for example, any function $f(x, y)$ of two variables can be approximated arbitrarily closely by an expression

$$P_1(x) \cdot Q[R_1(x) + y] + P_2(x) \cdot Q[R_2(x) + y]\,,$$

[17] See Arnold, V.I.: On the representation of functions of two variables in the form $\chi[\phi(x) + \psi(y)]$. Usp. Mat. Nauk **12**, No.2, 119–121 (1957).

where $P_1(x), P_2(x); R_1(x), R_2(x); Q(u)$ are specially chosen polynomials of one variable.[18]

In more recent times, in his attempts to simplify the methods by which he had obtained the representations (2) and (5), Kolmogorov turned his attention to more elementary considerations that led him to the above result. Along these lines he succeeded in proving by extraordinary elementary and elegant means that *each continuous function of* n *variables defined on the unit cube of* n*-dimensional space* E_n *can be represented in the form*

$$f(x_1, x_2, \ldots, x_n) = \sum_{r=1}^{2n+1} h_q \left[\sum_{p=1}^{n} \phi_q^p(x_p) \right], \tag{6}$$

where the $h_q(u)$ are continuous and the $\phi_p^q(x_p)$ are, in fact, standard, that is, they do not depend on the choice of the function f; in particular, each continuous function of two variables can be represented in the form

$$f(x, y) = \sum_{q=1}^{5} h_q[\phi_q(x) + \psi_q(y)]. \tag{6a}$$

For $n = 3$ it follows from (6) that

$$f(x, y, z) = \sum_{q=1}^{7} h_q[\phi_q(x) + \psi_q(y) + \chi_q(z)]] = \sum_{q=1}^{7} F_q[g_q(x, y)z],$$

where we have set

$$F_q(u, z) = h_q[u + \chi_q(z)], \quad g_q(x, y) = \phi_q(x) + \psi_q(y).$$

This last formula is even stronger than (5), since here the function of three variables $f(x, y, z)$ is representable in the form of seven (and not nine, as in (5)) terms that are single superpositions of functions of two variables; here these functions of two variables themselves have a special simple structure, and the inner function $g_q(x, y)$ (and the functions $\chi_q(z)$ occurring in the definition of $F_q(u, z)$) are, moreover, standard [so that all the properties of the functions $f(x, y, z)$ are completely determined by the seven functions of one variable $h_q(v)$].

The proof of (6) is so simple and beautiful that we shall reproduce it here almost in its entirety, referring those interested in the details to the author's more formalized account.[19] Since all the ideas of the proof occur quite clearly already in the case $n = 2$, we shall merely talk about the representation (6a)

[18] See Kolmogorov's paper mentioned in the footnote 9 on page 5.

[19] See Kolmogorov, A.N.: On the representation of continuous functions of several variables as superpositions of continuous functions of one variable. Dokl. Akad. Nauk SSSR **114**, 953–956 (1957).

of an arbitrary continuous function $f(x, y)$ of two variables x and y. The possibility of such a representation is proved in several stages.

1°. The 'inner' functions $\phi_q(x)$ and $\psi_q(y)$ of the representation (6a) are completely independent of the function $f(x, y)$ to be decomposed.

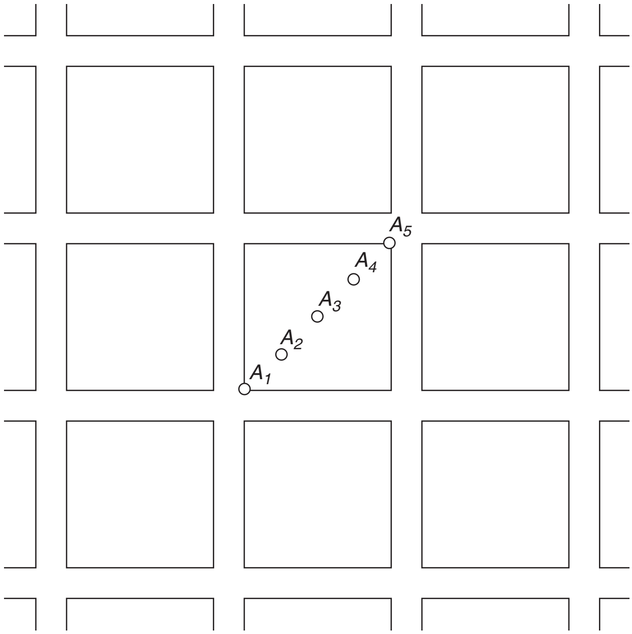

Fig. 8.

To define these functions we require certain preliminary constructions. We consider a 'town' consisting of a system of identical 'blocks' (non-intersecting closed squares) separated by narrow 'streets' all of the same width; see Fig. 8. We homothetically reduce our 'town' N times; for the centre of the homothety we can take, for example, the point A_1; we obtain a new 'town', which we call 'a town of rank 2'. The 'town of rank 3' is obtained in exactly the same way from the 'town of rank 2' by a homothetic reduction with homothety coefficient $\frac{1}{N}$: the 'town of rank 4' is obtained by a homothetic N-fold reduction by the 'town of rank 3', and so on. In general, the 'town of rank k' is obtained from the original 'town' (which we call 'the town of the first rank') by an N^k-fold reduction (with the centre of the homothety at A_1; incidentally the choice of the centre of the homothety is of no importance in what follows).

We call the system of 'towns' constructed above the 1st system. The 'town of the first rank of the qth system' ($q = 2, \ldots, 5$) is obtained from the 'town' depicted in Fig. 7† by moving the point A_1 to the point A_q by a parallel translation. It is not difficult to see that the 'streets' of the 'town' can be chosen sufficiently narrow so that a point of the plane will be covered by at least three blocks of our five 'towns of the first rank'. In the same way, the 'town of the kth rank' of the qth system ($k = 2, 3, \dots : q = 2, \dots, 5$) is obtained from the 'town of the kth rank of the first system' by a parallel translation taking the point A_1^k to the point A_q^k, where A_1^k and A_q^k are obtained from the points A_1 and A_q by a homothety taking the 'town of the first rank' of the first system (that is, our original 'town') to the 'town of the kth rank' of the same first system; here *each point of the plane will belong to 'blocks' of at least three of the five 'towns' of any fixed rank k.*

We define the function

$$\Phi_q(x, y) = \phi_q(x) + \psi_q(y) \quad (q = 1, 2, \dots, 5)$$

so that it divides any two 'blocks' of each 'town' of the system q, that is, so that the set of values taken by $\Phi_q(x, y)$ on a certain 'block' of the 'town of kth rank' (here k is an arbitrary fixed number) of the qth system does not intersect the set of values taken by $\Phi_q(x, y)$ on any other 'block' of the same 'town'. Here, of course, it suffices to consider the function $\Phi_q(x, y)$ on the unit square (and not on the entire plane).

In order that the function $\Phi_q(x, y) = \phi_q(x) + \psi_q(y)$ divide the 'blocks' of the 'town of the first rank' we can require, for instance, that on the projections of the 'blocks' of the 'town' onto the x axis $\phi_q(x)$ differs very slightly from the various integers and on the projections of the 'blocks' on the y axis $\psi_q(x)$ differs very slightly from the various multiples of $\sqrt{2}$ (because $m + n\sqrt{2} = m' + n'\sqrt{2}$ for integers m, n, m', n', only if $m' = m, n' = n$). Here, these conditions do not, of course, determine the functions $\phi_q(x)$ and $\psi_q(y)$ (on the 'streets' the function $\Phi_q = \phi_q + \psi_q$ can in general be defined completely arbitrarily for the moment); using this we can select limits on the values of $\phi_q(x)$ and $\psi_q(y)$ on the 'blocks' of the 'town of the second rank' so that the function $\Phi_q(x, y) = \phi_q(x) + \psi_q(y)$ divides not only the 'blocks' of the 'town of the first rank' but also the 'blocks' of the 'town of the second rank'.[20] In similar fashion, by bringing into consideration 'towns' of subsequent ranks and refining each time the values of the functions $\phi_q(x)$ and $\psi_q(y)$, in the limit we obtain continuous functions $\phi_q(x)$ and $\psi_q(y)$ (one can even require that they be monotone) satisfying the conditions in question.

2° By contrast, the functions $h_q(u)$ of the decomposition (6a) depend essentially on the original function $f(x, y)$.

To construct these functions we prove first of all that *any continuous function $f(x, y)$ of two variables x and y defined on the unit square can be represented in the form*

† *Translator's note*: This should be Fig. 8.

[20] The designated programme can be carried out if N is sufficiently large (so that the blocks of subsequent ranks do not join on to blocks of the previous ones). Kolmogorov chose $N = 18$.

$$f(x,y)=\sum_{q=1}^{5} h_q^{(1)} * \Phi_q(x,y)) + f_1(x,y), \tag{7}$$

where the $\Phi_q(x,y)=\phi_q(x)+\psi_q(y)$ *are the functions constructed above, and*

$$M_1=\max|f_1(x,y)|\leqslant\frac{5}{6}\max|f(x,y)|=\frac{5}{6}M\,, \tag{7a}$$

$$\max|h_q^{(1)}(|\Phi_q(x,y)|\leqslant\frac{1}{3}M,\quad q=1,\ldots,5\,. \tag{7b}$$

We choose the rank k sufficiently large so that the oscillation[21] of the function $f(x,y)$ on each 'block' of any of the 'towns of rank k' does not exceed $\frac{1}{6}M$; this, of course, is possible since as k increases the 'blocks' decrease without limit. Next, let $p_1^{(ij)}$ be a certain 'block' of a 'town of the first system' (and of the chosen rank k); then on this 'block' the (continuous) function $\Phi_1(x,y)$ takes values belonging to a certain segment $\Delta_1^{(ij)}$ of the real line (where, in view of the definition of the function Φ_1, this segment does not intersect segments of values taken by Φ_1 on any of the other 'blocks'). We now define the function $h_1^{(1)}$ on the segment $\Delta_1^{(ij)}$ to be a constant equal to one third of the value taken by $f(x,y)$ on any interior point $M_1^{(ij)}$ of the block $p_1^{(ij)}$ (it does not matter which). (We call this point the 'centre of the block'.) In similar fashion we define the function $h_1^{(1)}$ on each of the other segments defined by the values of $\Phi_1(x,y)$ on the 'block' of the 'town of rank k' of the first system; here all the values of $h_1^{(1)}$ will be at most $\frac{1}{3}M$ in modulus (since the value of $f(x,y)$ at the 'centre' of any 'block' will not exceed M in modulus). We now define in arbitrary fashion the function $h_1^{(1)}(u)$ at those values of the argument u at which it has not already been defined, with the proviso that it be continuous and that inequality (7b) should hold; we define all the other functions $h_q^{(1)}(u)(q=2,\ldots,5)$ in similar fashion.

We now prove that the difference

$$f_1(x,y)=f(x,y)-\sum_{q=1}^{5} h_q^{(1)}(\Phi_q(x,y))$$

satisfies condition (7a), that is,

$$|f_1(x_0,y_0)|\leqslant\frac{5}{6}M\,,$$

where (x_0,y_0) is an arbitrary point of the unit square. This point belongs (as indeed do all the points of the plane) to at least three blocks of 'towns of rank k'; therefore there certainly exist three of the five functions $h_1^{(1)}(\Phi_q(x,y))$ taking at the point (x_0,y_0) a value equal to one third of the value of $f(x,y)$

[21] that is, the difference between the largest and smallest values

at the 'centre' of the corresponding 'block', that is, differing from $\frac{1}{3}f(x_0, y_0)$ by not more than $\frac{1}{18}M$ (since the oscillation of $f(x, y)$ on each block does not exceed $\frac{1}{6}M$); the sum of these three values $h_q^{(1)}(\Phi_q(x_0, y_0))$ differs from $f(x_0, y_0)$ in modulus by at most $\frac{1}{6}M$. But since each of the remaining two numbers $h_q^{(1)}(\Phi_q(x_0, y_0))$ does not exceed $\frac{1}{3}M$ in modulus (in view of (7)), we obtain

$$|f_1(x_0, y_0)| = \left|f(x_0, y_0) - \sum_{q=1}^{5} h_q^{(1)}(\Phi(x_0, y_0))\right| \leqslant \frac{1}{6}M + \frac{2}{3}M = \frac{5}{6}M\,,$$

which proves (7a).

We now apply the same representation (7) to the function $f_1(x, y)$ featuring in (7); we obtain

$$f_1(x, y) = \sum_{q=1}^{5} h_q^{(2)}(\Phi_q(x, y)) + f_2(x, y)$$

or

$$f(x, y) = \sum_{q=1}^{5} h_q^{(1)}(\Phi_q(x, y)) + \sum_{q=1}^{5} h_q^{(2)}(\Phi_q(x, y) + f_2(x, y))\,,$$

where

$$M_2 = \max|f_2(x, y)| \leqslant \frac{5}{6}M_1 \leqslant \left(\frac{5}{6}\right)^2 M$$

and

$$\max|h^{(2)}(\Phi_q(x, y)) \leqslant \frac{1}{3}M_1 \leqslant \frac{1}{3}\cdot\frac{5}{6}M \quad (q = 1, 2, \dots, 5)\,.$$

Next we apply the decomposition (7) to the function $f_2(x, y)$ so obtained, and so on; after an n-fold application of this decomposition we obtain

$$f(x, y) = \sum_{q=1}^{5} h_q^{(1)}(\Phi_q(x, y)) + \sum_{q=1}^{5} h_q^{(2)}(\Phi_q(x, y)) + \cdots$$
$$+ \sum_{q=1}^{5} h_q^{(n-1)}(\Phi_q(x, y)) + f_2(x, y)\,,$$

where

$$M_2 = \max|f_n(x, y)| \leqslant \left(\frac{5}{6}\right)^n M$$

and

$$\max|h_q^{(s)}(\Phi_q(x, y)| \leqslant \frac{1}{3}\left(\frac{5}{6}\right)^{s-1} M \quad (q = 1, 2, \dots, 5;\ s = 1, 2, \dots, n-1))\,.$$

The last estimates show that as $n \to \infty$

$$f(x,y) = \sum_{q=1}^{5} h_q^{(1)}(\Phi_q(x,y)) + \sum_{q=1}^{5} h_q^{(2)}(\Phi_q(x,y)) + \cdots + \sum_{q=1}^{5} h_q^{(n)}(\Phi_q(x,y)) + \cdots ,$$

where the infinite series on the right hand side converges uniformly, as does each of the five series

$$h_q^{(1)}(\Phi_q(x,y)) + h_q^{(2)}(\Phi_q(x,y)) + \cdots + h_q^{(n)}(\Phi_q(x,y)) + \cdots \quad (q = 1, 2, \ldots, 5) .$$

This enables us to introduce the notation

$$h_q(u) = h_q^{(1)}(u) + h_q^{(2)}(u) + \cdots + h_q^{(n)}(u) + \cdots \quad (q = 1, 2, \ldots, 5) .$$

Thus, we finally obtain

$$f(x,y) = \sum_{q=1}^{5} h_q(\Phi_q(x,y)) = \sum_{q=1}^{5} h_q[\phi_q(x) + \psi_q(y)] ,$$

which is the required decomposition (6).

In conclusion we note that the representations (2), (5) and (6) are of purely theoretical interest, since they use essentially non-smooth functions such as the Weierstrass function;[22] therefore for practical purposes these representations are, it would seem, useless (in contrast with the representations (recalled earlier) of roots of equations of the 5th and 6th degrees as superpositions of functions of two variables). Thus the results that we have obtained do not remove the problem of finding convenient representations of, say, roots of 7th degree equations.

It is also unclear to what extent the decomposition (6) can be further improved; for example, the question of the uniqueness of the choice of the function h has not been solved. Also there are no methods enabling one to represent a given smooth function as a superposition of functions that are also relatively smooth; the strongest result in this direction remains the purely negative results of Vitushkin. Positive results of this kind would be of enormous interest.

We note one further result of Kolmogorov that goes in another direction. He proved that for each function of two variables defined on a square there exists a sum

[22] In view of the results of Bari (see Bari, N.K.: Mémoire sur la représentation finie des fonctions continues. Math. Ann. **103**, 145-248 and 590–653 (1930), one can represent each continuous function of one variable as a superposition of absolutely continuous functions. It therefore follows from (6) that each continuous function of n variables can be represented as a superposition of *monotone* functions of one variable and the sum function $g(x,y) = x + y$; however, these monotone functions are also essentially non-smooth.

of the form $\phi(x) + \psi(y)$ that *best approximates* this function. It can also be shown that for any (even everywhere discontinuous) bounded real function f defined on a compact set and any continuous function g defined on the same set there is a continuous function ϕ such that the deviation of $\phi(g)$ from f is a minimum. In particular, for each bounded function $f(x)$ there is a continuous function $\phi(x)$ best approximating it (in the sense of uniform convergence).