Electronic Colloquium on Computational Complexity, Report No. 50 (1998)

A Discrete Approximation and Communication Complexity Approach to the Superposition Problem

Farid Ablayev* Svetlana Ablayeva†

Abstract

The superposition (or composition) problem is a problem of representation of a function f by a superposition of "simpler" (in a different meanings) set Ω of functions. In terms of circuits theory this means a possibility of computing f by a finite circuit with 1 fan-out gates Ω of functions.

Using a discrete approximation and communication approach to this problem we present an *explicit* continuous function f from Deny class, that can not be represented by a superposition of a lower degree functions of the same class on the first level of the superposition and arbitrary Lipshitz functions on the rest levels. The construction of the function f is based on particular Pointer function g (which belongs to the uniform AC^0) with linear one-way communication complexity.

1 Introduction

In complexity theory the superposition approach provides a new proof of the separating of monotone NC^1 from monotone P [KaRaWi]. In classic mathematic the problem of representation of functions by functions of "simpler" (in some sense) quality has a long history and is based on the following problem. It is known that a common equation $a_1x^n + a_2x^{n-1} + \cdots + a_nx + a_{n+1} = 0$ for $n \leq 4$ can be solved over radicals. In terms of the superposition problem this means that the roots of the equation can be represented by a superposition of arithmetic operations and one variable function of the form $\sqrt[n]{a}$ $(n = 2,\ 3)$ of coefficients of the equation. Galua and Abel proved that a common equation of the 5-th order can not be solved in radicals (can not be represented as a superposition of this special form). Hilbert [Hi], formulated the 13-th problem the problem of *representing a solution of a common equation of the 7-th order as a superposition of functions of two variables.* The importance of the 13-th Hilbert problem is that it demonstrates one of the points of growth of function theory: it motivated an investigation of different aspects of the superposition problem.

*Dept. of Theoretical Cybernetics, Kazan State University. Email: ablayev@ksu.ru. Work done in part while visiting the University of Bonn. The research supported partially by the Volkswagen-Stiftung and Russia Fund for Basic Research 96-01-01962

†Dept. of Differential Equations of Kazan State University. Research supported partially by Russia Fund for Basic Research 96-01-01962

ISSN 1433-8092

Arnold [Ar] and Kolmogorov [Ko] proved that *arbitrary continuous function* $f(x_1, \ldots, x_k)$ *on* $[0,1]^k$ *can be represented as a superposition of continuous functions of one variable and sum operation:*

$$f(x_1, \ldots, x_k) = \sum_{i=1}^{2k+1} f_i \left(\sum_{j=1}^{k} h_{ij}(x_j) \right) \tag{1}$$

Note that the functions h_{ij} are chosen independently from f and as it is proved in [Lo], the functions h_{ij} belong to Hölder class.

Vitushkin made an essential advance in the investigation of the superposition problem. Let $\mathcal{F}_p^k$ denote the class of all continuous functions of k variables which has restricted continuous partial derivatives up to the p-th order. Vitushkin (see a survey [Vi]) proved the following theorem

Theorem 1 *There exists a function from* $\mathcal{F}_p^k$ *which can not be represented by a superposition of functions from* $\mathcal{F}_q^t$ *if* $\frac{k}{p} > \frac{t}{q}$.

Later Kolmogorov gave a proof of Theorem 1 that was based on comparing complexity characteristics (entropy of discrete approximation of functional spaces) of classes $\mathcal{F}_p^k$ and $\mathcal{F}_q^t$. Kolmogorov's proof shows that the classic superposition problem has a complexity background. The notion of entropy of functional spaces that was introduced by Kolmogorov was a result of the influence of Shannon's ideas. *Note that Kolmogorov's and Vitushkin's proofs show only the existence of the functions of Theorem 1 and do not present an example of a function from* $\mathcal{F}_p^k$ *that can not be represented by a superposition of functions from* $\mathcal{F}_q^t$. See the survey [Vi] and [Lo] for more information on the subject.

Further advance in presenting "constructive" continuous function which is not presented by certain superposition of simpler functions ("hard continuous function") was made in the paper [Ma]. It was proved that a function $f_G \in \mathcal{F}_p^k$ that is defined by a most hard (in terms of general circuits complexity) boolean function G can not be represented by a superposition of functions from $\mathcal{F}_q^t$ if $\frac{k}{p} > \frac{t}{q}$. Remind that almost all boolean functions are hard, but an explicit example of hard boolean function is not known, yet.

In this paper we generalize results of [Ab] where first example of *explicit* "hard continuous function" was presented. We use a discrete approximation of continuous functions and the communication complexity technique for the investigation of the superposition problem. Using certain Pointer boolean function g from the uniform AC^0 with the linear one-way communication complexity we define an explicit continuous function that can not be represented by a superposition of a lower degree functions of the same class on the first level of the superposition and arbitrary Lipshitz functions on the rest levels.

Informally speaking our method based on the following. Having continuous function f we suppose that it is presented by a superposition S of some kind of continuous functions. We consider their proper discrete approximations df and DS and compare the communication complexity C_{df} and C_{DS} of df and DS respectively. By showing $C_{DS} < C_{df}$ we prove that f can not be presented by the superposition S.

The theoretical model for the investigation of communication complexity of computation was introduced by Yao [Yao]. We refer to the book [KuNi] for more information on the subject.

2 The Function $f_{\omega,g}$

We define explicit continuous function $f_{\omega,g}$ of k arguments on the cube $[0,1]^k$by explicit boolean function g (more precisely sequence $g = \{g_n\}$ of explicit boolean functions). Informally speaking our construction of $f_{\omega,g}$ can be described as follows. We partition cube $[0,1]^k$to the infinite number of cubes (to 2^{kn} cubes for each $n > n_0$). The function $f_{\omega,g}$ in each of 2^{kn} cubes is defined by boolean function g_n of kn arguments. Now tern to the formal definition of $f_{\omega,g}$.

We consider $n = 2^j - 1,\ j \geq 1$ throughout the paper in order not use ceiling and floor brackets. Let $I_n = [\frac{1}{n+1}, \frac{2}{n+1}]$ be a closed interval, $I_n^k = \underbrace{I_n \times \cdots \times I_n}_{k}$, and $I^k = \bigcup_{n \geq 1} I_n^k$. Let $\Sigma = \{0,1\}$. We consider the following mapping $a : \Sigma^* \to [0,1]$. For a word $v = \sigma_1 \ldots \sigma_n$ we define

$$a(v) = \frac{1}{n+1}\left(1 + \sum_{i=1}^{n} \sigma_i 2^{-i} + \frac{1}{2^{n+1}}\right).$$

Denote $A_n = \{a(v) : v \in \Sigma^n\}$. For a number $a(v) \in A_n$ denote

$$I_n(a(v)) = \left[a(v) - \frac{1}{(n+1)2^{n+1}},\ a(v) + \frac{1}{(n+1)2^{n+1}}\right].$$

a closed interval of real numbers of size $\delta(n) = \frac{1}{(n+1)2^n}$. From the definitions of A_n and $I_n(a(v))$ it holds that:

1. For $a(v),\ a(v') \in A_n$ and $a(v) \neq a(v')$ segments $I_n(a(v))$, and $I_n(a(v'))$ can intersect only by boundary.

2. $\bigcup_{a(v) \in A_n} I_n(a(v)) = I_n$

Let us define the function $\Psi_{n,a(v)}(x)$ on the segment $I_n(a(v)),\ a(v) \in A_n$ as follows

$$\Psi_{n,a(v)}(x) = \begin{cases} 1 + \frac{2}{\delta(n)}(x - a(v)), & a(v) - \frac{\delta(n)}{2} \leq x \leq a(v) \\ 1 - \frac{2}{\delta(n)}(x - a(v)), & a(v) \leq x \leq a(v) + \frac{\delta(n)}{2} \\ 0, & x \notin \left[a(v) - \frac{\delta(n)}{2}, a(v) + \frac{\delta(n)}{2}\right] \end{cases} \qquad (2)$$

From the definition it follows that the function $\Psi_{n,a(v)}(x)$ reaches the maximum value 1 in the center of the segment $I_n(a(v)),\ a(v) \in A_n$ and value 0 in the border points of this segment.

For a sequence $v = (v_1, \ldots, v_k)$, where $v_i \in \Sigma^n$, $1 \leq i \leq k$, denote $I_n^k(b(v)) = I_n(a(v_1)) \times \cdots \times I_n(a(v_k))$ a k-dimension cube of size $\delta(n)$, where $b(v) = (a(v_1), \ldots, a(v_k))$.

Consider the following continuous function $\Psi_{n,b(v)}(x)$ inside each cube $I_n^k(b(v))$, $v = (v_1, \ldots, v_k) \in S_n,\ \ b(v) = (a(v_1), \ldots, a(v_k))$.

$$\Psi_{n,b(v)}(x) = \prod_{i=1}^{k} \Psi_{n,a(v_i)}(x_i),$$

Function $\Psi_{n,b(v)}(x)$ has following important properties: it reaches the maximum value 1 in the center of the cube $I_n^k(b(v))$; for all border points x of cube $I_n^k(b(v))$ it holds that $\Psi_{n,b(v)}(x) = 0$.

Let $g = \{g_n(v)\}$, where

$$g_n : \underbrace{\Sigma^n \times \cdots \times \Sigma^n}_{k} \to \{0, 1\}$$

be the sequence of the following Pointer boolean functions. For a sequence $v = (v_1, \ldots, v_k)$, where $v_i \in \Sigma^n$, $1 \le i \le k$, we will consider the following partition $pat(n, k)$: each word v_i of the sequence v is divided into two parts: the beginning u_i and the end w_i of length $l(n, k) = n - d(n, k)$ and $d(n, k) = \lceil(\log kn)/k\rceil$ respectively. We will write $v = (u, w)$ and call $u = (u_1, \ldots, u_k)$ the first part of the input sequence v and $w = (w_1, \ldots, w_k)$ the second part of the input sequence v.

Function $g_n(u, w) = 1$ iff $(ord(w_1 \ldots w_k) + 1)$-th bit in the word $u_1 \ldots u_k$ is one ($ord(\bar\sigma)$ denotes the integer whose binary representation is $\bar\sigma$. The numeration of bits in the words starts from 1). We will use both notation $g_n(v)$ and $g_n(u, w)$ for the boolean function g_n.

The function g_n can be formally described by the following formula:

$$g_n(u, w) = \bigvee_{\substack{\sigma \\ 0 \le ord(\sigma) \le |u|-1}} \bigwedge_{i=1}^{kd(n,k)} y_i^{\sigma_i} \wedge x_{ord(\sigma)},$$

where y_j (x_j) is the j-th symbol of the sequence w (u) in the common numeration of its elements. Clear that g is in the uniform class Σ_2 (AC^0).

Let $\omega(\delta)$ be a continuous function such that $\lim_{\delta\to 0} \omega(\delta) = 0$. Define a continuous function $f_{\omega,g}$ on cube $[0, 1]^k$ as follows:

$$f_{\omega,g}(x) = \sum_{\substack{n=2^j-1, \\ j \ge k}} \sum_{v \in \Sigma^n} (2g_n(v) - 1)\omega(\delta(n))\Psi_{n,b(v)}(x), \quad (3)$$

3 The Result

Remind definitions from functions theory. Denote $\mathcal{C}$ to be a class of continuous functions of $k \ge 1$ variables which are defined on closed cube $[0, 1]^k$. It is known that functions from $\mathcal{C}$ are uniformly continuous. Following functions theory for each $f(x_1, \ldots, x_k) \in \mathcal{C}$ define *modulus of continuous* $\omega_f(\delta)$. That is, $\omega_f(\delta)$ is a least upper bound of $|f(x) - f(x')|$, for all $x, x' \in [0, 1]^k$ such that $|x - x'| = \max_{1\le i\le k} |x_i - x'_i| \le \delta$.

Use the following standard definitions. Denote

$$\mathcal{H}_\omega = \{f \in \mathcal{C} : \omega_f(\delta) \le M\omega(\delta), \text{ for some } M > 0\},$$

Denote $\widehat{\mathcal{H}}_\omega$ an essential subset of $\mathcal{H}_\omega$. That is,

$$\widehat{\mathcal{H}}_\omega = \{f \in \mathcal{C} : M_1\omega(\delta) \le \omega_f(\delta) \le M_2\omega(\delta), \text{ for some } M_1, M_2 > 0\}.$$

The following classes are known as Hölder classes in functions theory:

$$\mathcal{H}_\gamma = \{f \in \mathcal{C} : \omega_f(\delta) \le M\delta^\gamma, \text{ for some } M > 0\} \qquad (\gamma \in (0,1])).$$

The following properties are known as classic properties:

1. $\mathcal{H}_{\gamma'} \subset \mathcal{H}_\gamma$ if $\gamma < \gamma'$.

2. Well known class $\mathcal{F} \subset \mathcal{C}$ of continuous functions which have continuous derivatives is a proper subclass of $\mathcal{H}_1$. The class $\mathcal{H}_1$ is known also as Lipshitz class.

3. Class $\mathcal{H}_\gamma$ — is a class of constant functions if $\gamma > 1$.

More common class of functions

$$\mathcal{D} = \left\{ f \in \mathcal{C} : \lim_{\delta \to 0} \omega_f(\delta) \log \frac{1}{\delta} = 0 \right\}$$

is called Deny class. Deny class contains Hölder classes properly.

Let $p > 1$, $a = 1/(e^{p+1})$, and

$$\omega_p(\delta) = \begin{cases} 1/(\ln 1/\delta)^p & \text{if } \ 0 < x \le a \\ 1/(\ln 1/a)^p & \text{if } \ x > a, \end{cases}$$

Class $\mathcal{H}_{\omega_p}$ is a subclass of Deny class $\mathcal{D}$.

Let Ω be some set of functions. We define the superposition of functions of Ω as a function computable by a leveled circuit with a constant number of 1 fan-out gates from the set Ω.

Theorem 2 *Function $f_{\omega_p,g}(x)$ over $k \ge 4$ variables belongs to the class $\widehat{\mathcal{H}^k_{\omega_p}}$ and is not represented by a following superposition of functions:*

1. *Superposition contains on the first level functions of t, $t < k$, variables from the class $\widehat{\mathcal{H}^t_{\omega_p}}$.*
2. *Superposition contains arbitrary continuous functions from $\mathcal{H}_1$ on the remaining levels of superposition.*

Below we present more general theorem 3. Theorem 2 is a corollary of it. Let $\mathbf{A}^t$, $\mathbf{B}^s$ be some classes of continuous functions of t and s variables. Define $Sp^k[\mathbf{A}^t, \mathbf{B}^s]$ class of continuous functions of k variables that can be represented by a superposition of the following form

$$F\left(h_1(x^1_1, \ldots, x^1_t), \ldots, h_s(x^s_1, \ldots, x^s_t)\right),$$

where $F(y_1, \ldots, y_s)$ is a function from class $\mathbf{B}^s$, and $\{h_i(x_1, \ldots, x_t) : 1 \le i \le s\} \subseteq \mathbf{A}^t$.

From the definition it holds that for modules of continuous $\omega_1(\delta)$, $\omega_2(\delta)$ it holds that the function $\omega(\delta) = \omega_2(\omega_1(\delta))$ is a modules of continuous and $Sp^k[\mathcal{H}^t_{\omega_1}, \mathcal{H}^s_{\omega_2}] \subseteq \mathcal{H}^k_\omega$.

Theorem 3 *Let* $\omega_1(\delta)$ *be an increasing function such that* $\frac{\omega_1(\delta)}{\delta}$ *does not increase when* δ *increase and*

$$\log \frac{1}{\omega_1(\delta)} = o\left(\left(\log \frac{1}{\delta}\right)^{1-t/k}\right). \qquad (4)$$

Then for $s \geq 1$, $M > 0, \gamma \in (0, 1]$, $\omega_2(\delta) = M\delta^\gamma$, $\omega(\delta) = \omega_2(\omega_1(\delta))$ *function* $f_{\omega,g}(x)$ *belongs to* $\mathcal{H}_\omega^k \backslash Sp^k[\widehat{\mathcal{H}}_{\omega_1}^t, \mathcal{H}_{\omega_2}^s]$.

The proof of general theorem 3 we present in the next section.

Proof of theorem 2. First. Function $\omega_p(\delta)$ satisfy the conditions for the ω_1 of theorem 3 for arbitrary constant $c > 0$ and especially for $c = 1 - \frac{t}{k}$, $t < k$. Next. Superposition of arbitrary functions from the class $\mathcal{H}_1$ is again a function from $\mathcal{H}_1$. From theorem 3 results the statement of theorem 2. □

4 The proof

The proof of the fact that $f_{\omega,g} \in \widehat{\mathcal{H}}_\omega^k$ results from the following property.

Property 4.1 *For the function* $f_{\omega,g}$ *it holds that*

1. *In each cube* $I_n^k(b(v)) \in I^k$ *function* $f_{\omega,g}$ *gets its maximum (minimum) value* $\omega(\delta(n))$ $(-\omega(\delta(n)))$ *in the center and value zero in the border of the border of the cube* $I_n^k(b(v))$.
2. *If in addition function* $\omega(\delta)$ *is such that* $\frac{\omega(\delta)}{\delta}$ *does not increase when* δ *increase then for modules of continuous* ω_f *of the function* $f_{\omega,g}$ *it holds that*
   - *(a)* $\omega(\delta(n)) \leq \omega_f(\delta(n)) \leq 2k\omega(\delta(n))$.
   - *(b) for arbitrary* δ $\omega_f(\delta) \leq 2k\omega(\delta)$.

Proof. The proof use standard arguments from functions theory. It will be presented in a complete paper. □

The proof of the second part of the theorem 3 use communication complexity arguments and is based on computing communication complexity of discrete approximations of the function $f_{\omega,g}$.

Let f be an arbitrary continuous function defined on the cube $[0, 1]^k$. Denote

$\alpha(n) = \min\{f(x) : x \in I_n^k = [\frac{1}{n}, \frac{2}{n}]^k\}$, and $\beta(n) = \max\{f(x) : x \in I_n^k = [\frac{1}{n}, \frac{2}{n}]^k\}$.

Definition 4.1 *Let* $f(x_1, \ldots, x_k)$ *be a continuous function on the cube* $[0, 1]^k$. *Call a discrete function* $df : \underbrace{\Sigma^n \times \cdots \times \Sigma^n}_{k} \to [\alpha(n), \beta(n)]$ *an* $\varepsilon(n)$*-approximation of the function* $f(x_1, \ldots, x_k)$, *if for arbitrary* $v = (v_1, \ldots, v_k) \in \Sigma^n \times \cdots \times \Sigma^n$ *it holds that*

$$|f(b(v)) - df(v)| \leq \varepsilon(n).$$

We will use the standard one-way communication computation for computing the boolean function $g_n \in g$. That is, two processors P_u and P_w obtain inputs in accordance with the partition $pat(n,k)$of input v. First part, $u = (u_1, \dots, u_k)$, of input sequence v is known to P_u and second part $w = (w_1, \dots, w_k)$ of input v is known to P_w.

The communication computation of a boolean function g_n is performed in accordance with a one-way protocol ψ as follows. P_u sends message m (binary word) to P_w. Processor P_w computes and outputs the value $g_n(u,w)$. The communication complexity C_ψ of the communication protocol ψ for the partition $pat(n,k)$of an inputs $v = (v_1, \dots, v_k)$ is the length $|m|$ of the message m.

The communication complexity $C_{g_n}(pat(n,k))$ of a boolean function g_n is $\min\{C_\psi : \psi \text{ computes } g_n\}$.

Lemma 4.1 *For the boolean function $g_n \in g$ it holds that*

$$C_{g_n}(pat(n,k)) \geq k(n-1) - \log kn.$$

Proof. With the function $g_n(u,w)$ we associate a $2^{kl(n,k)} \times 2^{kd(n,k)}$ communication matrix CM_{g_n} whose (u,w) entry is $g_n(u,w)$.

Using the fact that $C_{g_n}(pat(n,k)) = \lceil \log nrow(CM_{g_n}) \rceil$, where $nrow(CM_{g_n})$ is the number of distinct rows of communication matrix CM_{g_n} (see [Yao]) and the fact that for the g_n it holds that $nrow(CM_{g_n}) = 2^{kl(n,k)} \geq \frac{2^{k(n-1)}}{kn}$ we obtain the statement of the lemma. □

We will use the same one-way communication computation for computing a discrete function $df(v)$. Let $pat(n,k)$be a partition of input v, $v = (u,w)$. Let P_u and P_w be processors which receive inputs according to $pat(n,k)$. Let $\phi(pat(n,k))$ be a one-way communication protocol, which compute $df(u,w)$. The communication complexity C_ϕ of the $\phi(pat(n,k))$ is the total number of bits transmitted among processors P_u and P_w.

The communication complexity $C_{df}(pat(n,k))$ of a discrete function df we define as follows

$$C_{df}(pat(n,k)) = \min\{C_\phi : \phi(pat(n,k)) \text{ compute } df(v)\}.$$

Definition 4.2 *Define a communication complexity $C_f(pat(n,k), \varepsilon(n))$ of an $\varepsilon(n)$-approximation of the function f as follows:*

$C_f(pat(n,k), \varepsilon(n)) = \min\{C_{df}(pat(n,k)) : df(v)$ *— $\varepsilon(n)$-approximation of* $f\}$.

Lemma 4.2 *For $\varepsilon(n) < \omega(\delta(n))$, for arbitrary $\varepsilon(n)$-approximation df of the function $f_{\omega,g}$ it holds that*

$$C_{g_n}(pat(n,k)) \leq C_f(pat(n,k), \varepsilon(n)).$$

Proof. Suppose that

$$C_{g_n}(pat(n,k)) > C_f(pat(n,k), \varepsilon(n)). \tag{5}$$

This means that there exists an $\varepsilon(n)$-approximation df of the function $f(x_1, \dots, x_k)$ such that for $2^{kl(n,k)} \times 2^{kd(n,k)}$ communication matrices CM_{g_n} and CM_{df} of functions g_n and df it holds that

$$nrow(CM_{g_n}) > nrow(CM_{df}).$$

From the last inequality it follows that there exist two inputs u and u' such that two rows $row_{g_n}(u)$ and $row_{g_n}(u')$ are different but two rows $row_{df}(u)$ and $row_{df}(u')$ are equal. This means that there exists an input sequence w for which it holds that

$$g_n(u, w) \neq g_n(u', w), \quad (6)$$
$$df(u, w) = df(u', w). \quad (7)$$

Let $g_n(u, w) = 1,\ g_n(u', w) = 0$. Let us denote $v = (u, w),\ v' = (u', w)$.Then form the definition of the $f_{\omega,g}$ we have:

$$f_{\omega,g}(b(v)) = \omega(\delta(n)), \quad (8)$$
$$f_{\omega,g}(b(v')) = -\omega(\delta(n)). \quad (9)$$

From the definition of the $\varepsilon(n)$-approximation of the $f_{\omega,g}$ and the property (4.1) it holds that

$$|f_{\omega,g}(b(v)) - df(v)| \leq \varepsilon(n) < \omega(\delta(n)), \quad (10)$$
$$|f_{\omega,g}(b(v')) - df(v')| \leq \varepsilon(n) < \omega(\delta(n)). \quad (11)$$

From our conjunction that $df(v) = df(v')$, from (8), (9), (10), and (11) we have that

$$2\omega(\delta(n)) = |f_{\omega,g}(b(v)) - f_{\omega,g}(b(v'))| \leq$$

$$\leq |f_{\omega,g}(b(v)) - df(v)| + |f_{\omega,g}(b(v')) - df(v')| < 2\omega(\delta(n)).$$

The contradiction proofs that $df(v) \neq df(v')$. $\square$

Let $dh_i : \underbrace{\Sigma^n \times \cdots \times \Sigma^n}_{t} \to \mathcal{Z}$, $1 \leq i \leq t$, be a discrete functions and $DF : \underbrace{\Sigma^n \times \cdots \times \Sigma^n}_{k} \to \mathcal{Z}$, here $\mathcal{Z}$ denote the set of real numbers, be a following discrete function:

$$DF = F(dh_1(v_1^1, \dots, v_t^1), \dots, dh_s(v_1^s, \dots, v_t^s)),$$

where function $F(y_1, \dots, y_s)$ is an arbitrary continuous function.

Lemma 4.3 *For a discrete function $DF(v_1, \dots, v_k)$it holds that*

$$C_{DF}(pat(n,k)) \leq \sum_{i=1}^{s} C_{dh_i}(pat(n,k))$$

Proof Communication protocol $\phi^*(pat(n,k))$ for the function DF consists of processors P_u^* and P_w^*. Given an input u, w $\phi^*(pat(n,k))$ simulate in parallel protocols $\phi_1(pat(n,k))$, $\phi_2(pat(n,k))$,..., $\phi_s(pat(n,k))$ which computes dh_1, $dh_2, \ldots, dh_s$, respectively. The processor P_w^* on received a message from P_u^* and the input w computes outputs y_1,...,y_s of protocols $\phi_1(pat(n,k))$, $\phi_2(pat(n,k))$, $\ldots \phi_s(pat(n,k))$ and then computes and outputs a value $F(y)$, $y = (y_1, \ldots, y_s)$. $\square$

Lemma 4.4 *Let functions ω_1, ω_2 satisfy conditions of the theorem 3. Let for the function $\omega(\delta) = \omega_2(\omega_1(\delta))$ the function $f_{\omega,g}(x_1, \ldots, x_k)$ can be represented as a superposition of the form*

$$F\left(h_1(x_1^1, \ldots, x_t^1), \ldots, h_s(x_1^s, \ldots, x_t^s)\right),$$

where $F \in \mathcal{H}_{\omega_2}^s$ and $\{h_i(x_1, \ldots, x_t) : 1 \le i \le s\} \subset \widehat{\mathcal{H}}_{\omega_1}^t$.

Then there exists an $\varepsilon'(n) < \omega(\delta(n))$, such that

$$C_{f_{\omega,g}}(pat(n,k), \varepsilon'(n)) = o(n).$$

Proof. We will denote f our function $f_{\omega,g}$ in the proof of the theorem.

Let $\varepsilon = \omega_1(\delta(n)) / \log \frac{1}{\omega_1(\delta(n))}$. Consider arbitrary function $h \in \{h_1, \ldots, h_s\}$. Let $\alpha(n) = \min\{h(x) : x \in I_n^t = [\frac{1}{n}, \frac{2}{n}]^t\}$, and $\beta(n) = \max\{h(x) : x \in I_n^t = [\frac{1}{n}, \frac{2}{n}]^t\}$. Let

$$\mathcal{R}_{\varepsilon(n)} = \left\{\alpha_i : \alpha_i = \alpha(n) + \varepsilon(n)i, \; i \in \{0, 1, \ldots, \lfloor \tfrac{\beta(n)-\alpha(n)}{\varepsilon(n)} \rfloor\right\} \cup \{\beta(n)\}.$$

Due to selection of the value $\varepsilon(n)$, from the condition (4) of the theorem 3, and from the equality $\delta(n) = \frac{1}{(n+1)2^n}$ we have that:

$$|\mathcal{R}_{\varepsilon(n)}| = 2^{o(n^{1-t/k})} \tag{12}$$

Or $|\mathcal{R}_{\varepsilon(n)}| < 2^{nt}$. This means that there exists an $\varepsilon(n)$-approximator dh of continuous function h,

$$dh : \underbrace{\Sigma^n \times \cdots \times \Sigma^n}_{t} \to \mathcal{R}_{\varepsilon(n)}.$$

For the prove of the statement of the lemma we show that the discrete function

$$DF(v_1, \ldots, v_k) = F(dh_1(v_1^1, \ldots, v_t^1), \ldots, dh_s(v_1^s, \ldots, v_t^s))$$

is the $\varepsilon'(n)$-approximation of the function f and

$$C_{DF}(pat(n,k)) = o(n). \tag{13}$$

Let $v = (v_1, \ldots, v_k) \in \Sigma^n \times \cdots \times \Sigma^n$. First we prove that for some $\varepsilon'(n) < \omega(\delta)$ it holds that

$$|f(b(v)) - DF(v_1, \ldots, v_k)| \le \varepsilon'(n). \tag{14}$$

Denote $x = (x_1, \ldots, x_k) = b(v) = (a(v_1), \ldots, a(v_k))$. Due to the fact that for each $i \in \{1, 2, \ldots, s\}$ function dh_i is $\varepsilon(n)$-approximation of the continuous function h_i it holds that

$$|h_i(x_1^i, \ldots, x_t^i) - dh_i(v_1^i, \ldots, v_t^i)| \leq \varepsilon(n).$$

As function $\omega(\delta)$ decreases when δ decreases then we have

$$|F\left(h_1(x_1^1, \ldots, x_t^1), \ldots, h_s(x_1^s, \ldots, x_t^s)\right) -$$

$$F(dh_1(v_1^1, \ldots, v_t^1), \ldots, dh_s(v_1^s, \ldots, v_t^s))| \leq$$

$$\leq \omega_F(\varepsilon(n)) \leq \frac{M_1}{\left(\log \frac{1}{\omega_1(\delta(n))}\right)^\gamma} \left(\omega_1\left(\delta(n)\right)\right)^\gamma = \varepsilon'(n).$$

From some n_0 for $n > n_0$ it holds that

$$\varepsilon'(n) < \omega(\delta(n)).$$

Last inequality proves (14).

Consider now an arbitrary discrete function dh from $\{dh_1, \ldots, dh_s\}$.

It is sufficient to prove that

$$C_{dh}(pat(n, k)) = o(n) \tag{15}$$

Then using lemma 4.3 the (13) results.

With the function dh we associate a $2^{tl(n,k)} \times 2^{td(n,k)}$ communication matrix $CM_{dh}(n)$ whose (u, w) entry is $dh(u, w)$.

$$C_{dh}(pat(n, k), \varepsilon(n)) = \lceil \log nrow(CM_{dh}(n)) \rceil, \tag{16}$$

and

$$nrow(CM_{dh}(n)) \leq \min\left\{2^{tl(n,k)}, \left|\mathcal{R}_{\varepsilon(n)}\right|^{2^{td(n,k)}}\right\}. \tag{17}$$

From the definition of the partition $pat(n, k)$ we have that $l(n, k) = n - d(n.k)$, $d(n, k) = \lceil \frac{\log nk}{k} \rceil$. Using (17), (12) for the equality (16) we obtain inequality (15). □

Finally combining statements of lemmas 4.4, 4.2, and 4.1 we obtain the proof of the theorem 3.

5 Concluding remarks

The communication technique in this paper gives a clear information explanation of the statements of theorems 2 and 3. That is, functions h from the class $\widehat{\mathcal{H}}_{\omega_1}^t$ which satisfies the condition (4) of the theorem 3 can be approximated by discrete functions dh with small communication complexity $o(n)$ (see (15)). Such discrete functions dh on the first level of superposition "can mix" some different inputs during transformation and no functions on the remaining levels can reconstruct this information. Note that in contradiction functions h_{ij} from formula (1) are from the Hölder class $\mathcal{H}_* = \cup_{\gamma>0}\mathcal{H}_\gamma$. These functions do not "lost the information of inputs". They just reorganizing these information.

We conclude with open problems. Whether using discrete approximation together with communication technique is possible to present an explicit function

(i) from $\mathcal{H}_1^k$ which could not be presented by a superposition of functions from $\mathcal{H}_1^t$ if $t < k$;

(ii) from $\mathcal{F}_p^k$ which could not be represented by a superposition of functions from $\mathcal{F}_q^t$, if $\frac{k}{p} > \frac{t}{q}$?

Acknowledgment

We are grateful to Marek Karpinski for helpful discussions on the subject of the paper.

References

[Ab] F. Ablayev, Communication method of the analyses of superposition of continuous functions, *in Proceedings of the international conference "Algebra and Analyses part II. Kazan*, 1994, 5-7 (in Russian). See also F. Ablayev, Communication complexity of probabilistic computations and some its applications, Thesis of doctor of science dissertation, Moscow State University, 1995, (in Russian).

[Ar] V. Arnold, On functions of Three Variables, *Dokladi Akademii Nauk*, 114, 4, (1957), 679-681.

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

[KaRaWi] M. Karchmer, R. Raz, and A. Wigderson, Super-logarithmic Depth Lower Bounds Via the Direct Sum in Communication Complexity, *Computational Complexity*, 5, (1995), 191-204.

[Ko] A. Kolmogorov, On Representation of Continuous Functions of Several Variables by a superposition of Continuous Functions of one Variable and Sum Operation. *Dokladi Akademii Nauk*, 114, 5, (1957), 953-956.

[KuNi] E. Kushilevitz and N. Nisan, Communication complexity, *Cambridge University Press*, 1997.

[Lo] G. Lorenz, Metric Entropy, Widths and Superpositions Functions, *Amer. Math. Monthly* 69, 6, (1962), 469-485.

[Ma] S. Marchenkov, On One Method of Analysis of superpositions of Continuous Functions, *Problemi Kibernetici*, 37, (1980), 5-17.

[Vi] A. Vitushkin, On Representation of Functions by Means of Superpositions and Related Topics, *L'Enseignement mathematique*, 23, fasc.3-4, (1977), 255-320.

[Yao] A. C. Yao, Some Complexity Questions Related to Distributive Computing, *in Proc. of the 11th Annual ACM Symposium on the Theory of Computing*, (1979), 209-213.

ECCC ISSN 1433-8092
http://www.eccc.uni-trier.de/eccc
ftp://ftp.eccc.uni-trier.de/pub/eccc
ftpmail@ftp.eccc.uni-trier.de, subject 'help eccc'