Proceedings of the Institute of Mathematics and Mechanics,
National Academy of Sciences of Azerbaijan
Volume 45, Number 1, 2019, Pages 31–40

ON THE HÖLDER CONTINUITY IN RIDGE FUNCTION REPRESENTATION

RASHID A. ALIEV, AYSEL A. ASGAROVA, AND VUGAR E. ISMAILOV

Abstract. In this paper, we show that if a multivariate function is locally Hölder continuous of some degree and represented by a sum of arbitrarily behaved ridge functions, then, under a suitable condition, it can be also represented by a sum of ridge functions, which are locally Hölder continuous of the same degree.

1. Introduction

A *ridge function* is a multivariate function of the form

$$g(\mathbf{a}\cdot\mathbf{x}) = g(a_1x_1 + \ldots + a_mx_m),$$

where $g : \mathbb{R} \to \mathbb{R}$, $\mathbf{a} = (a_1, ..., a_m)$ is a fixed vector (direction) in $\mathbb{R}^m \setminus \{\mathbf{0}\}$, $\mathbf{x} = (x_1, ..., x_m)$ is the variable and $\mathbf{a}\cdot\mathbf{x}$ is the usual inner product in $\mathbb{R}^m$. In the theory of partial differential equations, ridge functions have been known under the name of *plane waves (*see, e.g., [14]). The term "ridge function" was devised by Logan and Shepp in their pioneering paper [22] dedicated to the mathematics of computerized tomography (see also [15, 16, 24, 25]). After a 1981 paper by Friedman and Stuetzle [8] ridge functions started to appear also in statistics, especially, in the theory of projection pursuit and projection regression (see, e.g., [7, 8, 9]). The general idea therein was to reduce "dimension" and thus bypass the "curse of dimensionality".

Ridge functions are used in many models in neural network theory. For example, in one of the popular models called MLP (multilayer feedforward perceptron) model, the simplest case considers functions of the form

$$\sum_{i=1}^{r} c_i\sigma(\mathbf{w}^i\cdot\mathbf{x} - \theta_i).$$

Here the weights $\mathbf{w}^i$ are vectors in $\mathbb{R}^m$, the thresholds θ_i and the coefficients c_i are real numbers and the activation function σ is a univariate function. Note that for each $\theta \in \mathbb{R}$ and $\mathbf{w} \in \mathbb{R}^m \setminus \{\mathbf{0}\}$ the function

$$\sigma(\mathbf{w}\cdot\mathbf{x} - \theta)$$

2010 *Mathematics Subject Classification.* 26B40, 39B22.
Key words and phrases. ridge function; Hölder continuity; Cauchy functional equation.

is a ridge function. For a detailed survey of approximation theory of the MLP model see [30].

Ridge functions are interesting also to approximation theorists. In approximation theory, these functions are implemented as an effective and convenient tool for approximating complicated multivariate functions (see, e.g., [10, 11, 12, 13, 19, 21, 23, 26, 29]).

In this paper, we consider the problem of representation by sums of ridge functions with r, $r \geq 1$, fixed directions. Let the directions $\mathbf{a}^i \in \mathbb{R}^m \backslash \{\mathbf{0}\}$, $i = 1, ..., r$, be given and pairwise linearly independent. The first problem arising here is about the representability of a given multivariate function $f : \mathbb{R}^m \to \mathbb{R}$ as a sum of ridge functions with the directions $\mathbf{a}^i$, $i = 1, ..., r$. In other words, we want to know when the function f can be written in the form

$$f(\mathbf{x}) = \sum_{i=1}^{r} g_i(\mathbf{a}^i \cdot \mathbf{x}). \tag{1.1}$$

This problem has a simple solution if f depends on two variables and has partial derivatives up to r-th order. For the representation of $f(x, y)$ in the following form

$$f(x, y) = \sum_{i=1}^{r} g_i(a_i x + b_i y),$$

it is necessary and sufficient that

$$\prod_{i=1}^{r} \left(b_i \frac{\partial}{\partial x} - a_i \frac{\partial}{\partial y} \right) f = 0. \tag{1.2}$$

Note that the last assertion is valid also for continuous functions of two variables provided that the derivatives are understood in the sense of distributions. It should be remarked that this simple assertion is not directly generalized to the case when f depends on more than two variables.

Assume we know that a function $f(\mathbf{x})$ can be represented in the form (1.1). Assume in addition that f is of the class $C^k(\mathbb{R}^m)$. What can we say about g_i? Can we say that $g_i \in C^k(\mathbb{R})$? The case $r = 1$ is obvious. In this case, if $f \in C^k(\mathbb{R}^m)$, then for $\mathbf{c} \in \mathbb{R}^m$ satisfying $\mathbf{a}^1 \cdot \mathbf{c} = 1$ we have that $g_1(t) = f(t\mathbf{c})$ is in $C^k(\mathbb{R})$. The same argument can be carried out for the case $r = 2$. In this case, since the vectors $\mathbf{a}^1$ and $\mathbf{a}^2$ are linearly independent, there exists a vector $\mathbf{c} \in \mathbb{R}^m$ satisfying $\mathbf{a}^1 \cdot \mathbf{c} = 1$ and $\mathbf{a}^2 \cdot \mathbf{c} = 0$. Therefore, we obtain that the function $g_1(t) = f(t\mathbf{c}) - g_2(0)$ is in the class $C^k(\mathbb{R})$. Similarly, one can verify that $g_2 \in C^k(\mathbb{R})$ (see [3]).

The above question becomes quite difficult if the number of directions $r \geq 3$. For $r = 3$, there are many smooth functions which decompose into sums of very badly behaved ridge functions. This is a consequence of the classical Cauchy Functional Equation (CFE). This equation is defined as

$$h(x + y) = h(x) + h(y), \ \ h : \mathbb{R} \to \mathbb{R}, \tag{1.3}$$

which has a class of simple solutions $h(x) = cx$, $c \in \mathbb{R}$. However, it easily follows from the Hamel basis theory that CFE has also a large class of badly behaved solutions. These solutions are called “badly behaved” because they are weird over reals. They are, for example, not continuous at a point, not monotone at an interval, not bounded at any set of positive measure (see, e.g., [1]). Let h_1 be any such solution of the equation (1.3). Then the zero function can be written as

$$0 = h_1(x) + h_1(y) - h_1(x+y). \tag{1.4}$$

Note that the functions involved in (1.4) are ridge functions with the directions $(1,0)$, $(0,1)$ and $(1,1)$ respectively. This particular example shows that for smoothness of the representation (1.1) one must impose additional conditions on the functions g_i, $i = 1, ..., r$.

It was first proved by Buhman and Pinkus [6] that if in (1.1) $f \in C^k(\mathbb{R}^m)$, $k \geq r-1$ and $g_i \in L^1_{loc}(\mathbb{R})$ for each i, then $g_i \in C^k(\mathbb{R})$ for $i = 1, ..., r$. In [28] Pinkus found a strong relationship between CFE and the problem of smoothness in ridge function representation. He generalized extensively the previous result of Buhman and Pinkus [6]. He showed that the solution is quite simple and natural if the functions g_i are taken from a class $\mathcal{B}$ of real-valued functions u defined on $\mathbb{R}$. By definition, u is in $\mathcal{B}$ if for any function $v \in C(\mathbb{R})$ for which $u - v$ satisfies CFE, $u - v$ is linear, i.e. $u(x) - v(x) = cx$, where $c \in \mathbb{R}$ (see [28]). The result of Pinkus states that if in (1.1) $f \in C^k(\mathbb{R}^m)$ and each $g_i \in \mathcal{B}$, then necessarily $g_i \in C^k(\mathbb{R})$ for $i = 1, ..., r$.

The above representation problem was also considered by Konyagin and Kuleshov [17, 18] and by Kuleshov [20]. They mainly analyze the continuity of the representation, that is, the question if and when continuity of f in (1.1) guarantees the continuity of g_i. There are also other results concerning the smoothness of ridge function representation generalizing the above result of Pinkus (see [20]). The results in [17, 18, 20] involve certain subsets (convex open sets, convex bodies, etc.) of $\mathbb{R}^m$ instead of only $\mathbb{R}^m$ itself.

The results of Pinkus [28] gave rise to the following natural and important problem. Assume in the representation (1.1) $f \in C^k(\mathbb{R}^m)$, but the functions g_i are arbitrarily behaved (that is, we allow very badly behaved functions). Can we write f as a sum $\sum_{i=1}^r f_i(\mathbf{a}^i \cdot \mathbf{x})$ but with the $f_i \in C^k(\mathbb{R})$, $i = 1, ..., r$? This problem was posed in [6] and [27]. In [3], Aliev and Ismailov obtained a partial solution to this problem. Their solution comprises the cases in which $k > 1$ and $r - 1$ directions of given r directions are linearly independent. Note that such condition on directions is satisfied by default if we are given three directions, as it is assumed that all the directions are pairwise linearly independent. The representation problem in the case of three directions was initially considered in [5]. For bivariate functions having the degree of smoothness $k \geq r-2$, the problem was completely solved in [4].

Kuleshov [20] generalized Aliev and Ismailov's result [3] to the other possible cases of k. That is, he proved that if a function $f \in C^k(\mathbb{R}^m)$, where $k \geq 0$, is of the form (1.1) and $r-1$-tuple of the given set of r directions $\mathbf{a}^i$ forms a linearly independent system, then in (1.1) the functions g_i can be replaced with functions $f_i \in C^k(\mathbb{R})$, $i = 1, ..., r$ (see [20, Theorem 3]). In [2], we proved this result using completely different ideas. Note that our proof contains a theoretical method for constructing the mentioned functions $f_i \in C^k(\mathbb{R})$ (see [2, Theorem 2.1, Theorem 2.2]). Using this method, we also estimated the modulus of continuity of f_i in terms of the modulus of continuity of f (see [2, Remark 2]).

In this paper, we continue investigations on ridge function representations. The question considered here is as follows. Assume f is Hölder continuous on compact subsets of $\mathbb{R}^m$ and possesses the representation (1.1). Can we replace g_i with Hölder continuous functions of the same degree as f? We answer this question positively in the case when $r-1$ directions of given r directions are linearly independent. First we prove this when the space dimension $m=2$ and the number of directions $r=3$. Then based on the multidimensional techniques exploited in our previous paper [2], we formulate the main result in the general case. Note that the main result of this paper (see Theorem 2.2) cannot be obtained directly from the above mentioned results.

2. Main result

We start this section with the following well-known definition.

Definition 2.1. We say that a function $F:\mathbb{R}^m\to\mathbb{R}$ is locally Hölder continuous with degree α, $0<\alpha\leq 1$, if for any compact set $K\subset\mathbb{R}^m$ there is a number $M=M\left(F;\alpha;K\right)>0$ such that for any $\mathbf{x}=(x_1,...,x_m)\in K$ and $\mathbf{y}=(y_1,...,y_m)\in K$ the inequality $\left|F\left(\mathbf{x}\right)-F\left(\mathbf{y}\right)\right|\leq M\cdot\sum_{i=1}^m\left|x_i-y_i\right|^\alpha$ holds.

The class of locally Hölder continuous functions with degree α is denoted by $H_\alpha^{(loc)}\left(\mathbb{R}^m\right)$. For a function $F\in H_\alpha^{(loc)}\left(\mathbb{R}^m\right)$ and a compact set $K\subset\mathbb{R}^m$ put

$$H\left(F;\alpha;K\right)=\sup\left\{\left|F\left(\mathbf{x}\right)-F\left(\mathbf{y}\right)\right|\cdot\left[\sum_{i=1}^m\left|x_i-y_i\right|^\alpha\right]^{-1}:\ \mathbf{x},\mathbf{y}\in K,\ \mathbf{x}\neq\mathbf{y}\right\}.$$

The following lemma plays a key role in the proof of our main result.

Lemma 2.1. *Assume a function* $F\in H_\alpha^{(loc)}\left(\mathbb{R}^2\right)$, $0<\alpha\leq 1$, *has the form*

$$F\left(x,y\right)=h\left(x\right)+h\left(y\right)-h\left(x+y\right),$$

where $h:\mathbb{R}\to\mathbb{R}$ *is an arbitrarily behaved function. Then there exists a function* $f\in H_\alpha^{(loc)}\left(\mathbb{R}\right)$ *such that*

$$F\left(x,y\right)=f\left(x\right)+f\left(y\right)-f\left(x+y\right).$$

Proof. Consider the functions

$$G\left(x,y\right)=F\left(x,y\right)-F\left(0,0\right),\ g\left(x\right)=h\left(x\right)-F\left(0,0\right).$$

Then the function $G\left(x,y\right)$ also belongs to the class $H_\alpha^{(loc)}\left(\mathbb{R}^2\right)$ and we have the equalities

$$G\left(x,y\right)=g\left(x\right)+g\left(y\right)-g\left(x+y\right),\ G\left(0,0\right)=0. \tag{2.1}$$

It follows from (2.1) that

$$g\left(0\right)=g\left(0\right)+g\left(0\right)-g\left(0\right)=G\left(0,0\right)=0$$

and for any $x\in\mathbb{R}$

$$G\left(x,0\right)=g\left(x\right)+g\left(0\right)-g\left(x\right)=g\left(0\right)=0, \tag{2.2}$$

$$G\left(0,x\right)=g\left(0\right)+g\left(x\right)-g\left(x\right)=g\left(0\right)=0, \tag{2.3}$$

$$G\left(x,x\right)=g\left(x\right)+g\left(x\right)-g\left(2x\right)=2g\left(x\right)-g\left(2x\right). \tag{2.4}$$

Since the equality (2.4) is equivalent to the equality

$$g(x) = \frac{1}{2} g(2x) + \frac{1}{2} G(x,x),$$

we can write that

$$g\left(\frac{1}{2}\right) = \frac{1}{2} g(1) + \frac{1}{2} G\left(\frac{1}{2}, \frac{1}{2}\right),$$

$$g\left(\frac{1}{4}\right) = \frac{1}{4} g(1) + \frac{1}{4} G\left(\frac{1}{2}, \frac{1}{2}\right) + \frac{1}{2} G\left(\frac{1}{4}, \frac{1}{4}\right),$$

....

$$g\left(\frac{1}{2^n}\right) = \frac{1}{2^n} g(1) + \frac{1}{2^n} G\left(\frac{1}{2}, \frac{1}{2}\right) + \frac{1}{2^{n-1}} G\left(\frac{1}{4}, \frac{1}{4}\right) + ... + \frac{1}{2} G\left(\frac{1}{2^n}, \frac{1}{2^n}\right), n \in N. \tag{2.5}$$

We obtain from (2.2) that for any $k \in \mathbb{N}$

$$\left|G\left(\frac{1}{2^k}, \frac{1}{2^k}\right)\right| = \left|G\left(\frac{1}{2^k}, \frac{1}{2^k}\right) - G\left(\frac{1}{2^k}, 0\right)\right| \leq \left(\frac{1}{2^k}\right)^\alpha \cdot H\left(G; \alpha; [-1,1]^2\right)$$
$$= \frac{1}{2^{k\alpha}} \cdot H\left(F; \alpha; [-1,1]^2\right) = \frac{c_0}{2^{k\alpha}}, \tag{2.6}$$

where $c_0 = H\left(F; \alpha; [-1,1]^2\right)$. It follows from (2.5) and (2.6) that

$$\left|g\left(\frac{1}{2^n}\right)\right| \leq \frac{|g(1)|}{2^n} + \frac{1}{2^n} \cdot \frac{c_0}{2^\alpha} + \frac{1}{2^{n-1}} \cdot \frac{c_0}{2^{2\alpha}} + ... + \frac{1}{2} \cdot \frac{c_0}{2^{n\alpha}}$$
$$\leq \frac{|g(1)|}{2^n} + \frac{c_0}{2^{n\alpha+1}} \left[1 + \frac{1}{2^{\alpha-1}} + ... + \frac{1}{2^{(n-2)(\alpha-1)}} + \frac{1}{2^{(n-1)(\alpha-1)}}\right]$$
$$\leq \frac{|g(1)|}{2^n} + \frac{c_1}{2^{n\alpha+1}} \leq \frac{c_2}{2^{n\alpha}}, \tag{2.7}$$

where c_1 and c_2 are some constants independent of $n \in \mathbb{N}$. Consider a regular irreducible dyadic fraction $\frac{m}{2^n} \in \left[\frac{1}{2^k}, \frac{1}{2^{k-1}}\right)$ and represent it in the binary system:

$$\frac{m}{2^n} = \frac{1}{2^k} + \frac{1}{2^{k+p_1}} + ... + \frac{1}{2^{k+p_s}} + \frac{1}{2^n}.$$

We obtain from (2.1) that

$$g\left(\frac{m}{2^n}\right) = g\left(\frac{1}{2^k} + \frac{1}{2^{k+p_1}} + ... + \frac{1}{2^{k+p_s}} + \frac{1}{2^n}\right)$$
$$= g\left(\frac{1}{2^k}\right) + g\left(\frac{1}{2^{k+p_1}} + ... + \frac{1}{2^{k+p_s}} + \frac{1}{2^n}\right) - G\left(\frac{1}{2^k}, \frac{1}{2^{k+p_1}} + ... + \frac{1}{2^{k+p_s}} + \frac{1}{2^n}\right)$$
$$= g\left(\frac{1}{2^k}\right) + g\left(\frac{1}{2^{k+p_1}}\right) + ... + g\left(\frac{1}{2^{k+p_s}}\right) + g\left(\frac{1}{2^n}\right)$$
$$-G\left(\frac{1}{2^k}, \frac{1}{2^{k+p_1}} + ... + \frac{1}{2^{k+p_s}} + \frac{1}{2^n}\right)$$
$$-G\left(\frac{1}{2^{k+p_1}}, \frac{1}{2^{k+p_2}} + ... + \frac{1}{2^{k+p_s}} + \frac{1}{2^n}\right) - ... - G\left(\frac{1}{2^{k+p_s}}, \frac{1}{2^n}\right). \tag{2.8}$$

It follows from (2.2) and (2.3) that for any $a > 0$ and for any $x, y \in [-a, a]^2$

$$|G(x,y)| = |G(x,y) - G(0,y)| \leq H\left(F; \alpha; [-a,a]^2\right) \cdot |x|^\alpha, \tag{2.9}$$

$$|G(x,y)| = |G(x,y) - G(x,0)| \leq H\left(F;\alpha;[-a,a]^2\right) \cdot |y|^\alpha . \tag{2.10}$$

Then considering (2.7) and (2.9) we obtain from (2.8) that

$$\left|g\left(\frac{m}{2^n}\right)\right| \leq \left|g\left(\frac{1}{2^k}\right)\right| + \left|g\left(\frac{1}{2^{k+p_1}}\right)\right| + ... + \left|g\left(\frac{1}{2^{k+p_s}}\right)\right| + \left|g\left(\frac{1}{2^n}\right)\right|$$

$$+\left|G\left(\frac{1}{2^k}, \frac{1}{2^{k+p_1}} + ... + \frac{1}{2^{k+p_s}} + \frac{1}{2^n}\right)\right| + \left|G\left(\frac{1}{2^{k+p_1}}, \frac{1}{2^{k+p_2}} + ... + \frac{1}{2^{k+p_s}} + \frac{1}{2^n}\right)\right| + ...$$

$$+\left|G\left(\frac{1}{2^{k+p_s}}, \frac{1}{2^n}\right)\right| \leq \frac{c_2}{2^{k\alpha}} + \frac{c_2}{2^{(k+p_1)\alpha}} + ... + \frac{c_2}{2^{(k+p_s)\alpha}} + \frac{c_2}{2^{n\alpha}}$$

$$+\frac{c_0}{2^{k\alpha}} + \frac{c_0}{2^{(k+p_1)\alpha}} + ... + \frac{c_0}{2^{(k+p_s)\alpha}} \leq \frac{c_3}{2^{k\alpha}} \leq c_3 \cdot \left(\frac{m}{2^n}\right)^\alpha, \tag{2.11}$$

where c_3 is some constant.

Now we construct an auxiliary function $g^* = g^*(x)$ as follows. For any $m \in \mathbb{Z}$ and $n \in \mathbb{N}$ set $g^*\left(\frac{m}{2^n}\right) = g\left(\frac{m}{2^n}\right)$; for a number x having in the binary system the representation $x = p_0 + \frac{p_1}{2} + ... + \frac{p_n}{2^n} + ...$ set

$$g^*(x) = \lim_{n\to\infty} g\left(p_0 + \frac{p_1}{2} + ... + \frac{p_n}{2^n}\right). \tag{2.12}$$

We must prove that there exists a finite limit on the right-hand side of (2.12). It follows from (2.1), (2.10) and (2.11) that for any $n \in \mathbb{N}$ and $k \in \mathbb{N}$

$$\left|g\left(p_0 + \frac{p_1}{2} + ... + \frac{p_n}{2^n}\right) - g\left(p_0 + \frac{p_1}{2} + ... + \frac{p_n}{2^n} + ... + \frac{p_{n+k}}{2^{n+k}}\right)\right|$$

$$\leq \left|g\left(\frac{p_{n+1}}{2^{n+1}} + ... + \frac{p_{n+k}}{2^{n+k}}\right)\right| + \left|G\left(p_0 + \frac{p_1}{2} + ... + \frac{p_n}{2^n}, \frac{p_{n+1}}{2^{n+1}} + ... + \frac{p_{n+k}}{2^{n+k}}\right)\right|$$

$$\leq (c_3 + c_0) \cdot \left|\frac{p_{n+1}}{2^{n+1}} + ... + \frac{p_{n+k}}{2^{n+k}}\right|^\alpha \leq \frac{c_3 + c_0}{2^{n\alpha}}. \tag{2.13}$$

We obtain from (2.13) that the sequence $\left\{g\left(p_0 + \frac{p_1}{2} + ... + \frac{p_n}{2^n}\right)\right\}_{n=1}^\infty$ is fundamental and, therefore, there exists a finite limit on the right-hand side of (2.12).

Now we show that for any $x, y \in \mathbb{R}$

$$G(x,y) = g^*(x) + g^*(y) - g^*(x+y). \tag{2.14}$$

Let in the binary system

$$x = p_0 + \frac{p_1}{2} + ... + \frac{p_n}{2^n} + ..., y = q_0 + \frac{q_1}{2} + ... + \frac{q_n}{2^n} + ...,$$

$$x + y = r_0 + \frac{r_1}{2} + ... + \frac{r_n}{2^n} + ...$$

Denote

$$x^{(n)} = p_0 + \frac{p_1}{2} + ... + \frac{p_n}{2^n}, y^{(n)} = q_0 + \frac{q_1}{2} + ... + \frac{q_n}{2^n} + ...,$$

$$(x+y)^{(n)} = r_0 + \frac{r_1}{2} + ... + \frac{r_n}{2^n} + ..., \ n \in \mathbb{N}.$$

Considering (2.1) we can write that

$$G\left(x^{(n)}, y^{(n)}\right) = g\left(x^{(n)}\right) + g\left(y^{(n)}\right) - g\left(x^{(n)} + y^{(n)}\right). \tag{2.15}$$

Passing to the limit in (2.15) and taking into account the continuity of the function $G(x,y)$, we have

$$G(x,y) = g^*(x) + g^*(y) - \lim_{n\to\infty} g\left(x^{(n)} + y^{(n)}\right). \tag{2.16}$$

Since $(x+y)^{(n)} - x^{(n)} - y^{(n)}$ equals to 0 or $\frac{1}{2^n}$, it follows from (2.1), (2.7) and (2.10) that

$$\left|g\left((x+y)^{(n)}\right) - g\left(x^{(n)} + y^{(n)}\right)\right| \leq \left|g\left((x+y)^{(n)} - x^{(n)} - y^{(n)}\right)\right|$$

$$+\left|G\left(x^{(n)} + y^{(n)}, (x+y)^{(n)} - x^{(n)} - y^{(n)}\right)\right| \leq \left|g\left(\frac{1}{2^n}\right)\right| + \left|G\left(x^{(n)} + y^{(n)}, \frac{1}{2^n}\right)\right|$$

$$\leq \frac{c_2}{2^{n\alpha}} + H\left(F; \alpha; [-|x| - |y| - 1, |x| + |y| + 1]^2\right) \cdot \frac{1}{2^{n\alpha}}.$$

It follows that

$$\lim_{n\to\infty} g\left(x^{(n)} + y^{(n)}\right) = \lim_{n\to\infty} g\left((x+y)^{(n)}\right) = g^*(x+y).$$

Thus from (2.16) we obtain (2.14).

Now we show that $g^* \in H_\alpha^{(loc)}(\mathbb{R})$. First we prove that for any $\delta \in (0,1)$

$$|g^*(\delta)| \leq c_3 \cdot \delta^\alpha. \tag{2.17}$$

Indeed, let $\delta = \frac{p_1}{2} + ... + \frac{p_n}{2^n} + ... \in (0,1)$. Then it follows from (2.11) that

$$\left|g\left(\frac{p_1}{2} + ... + \frac{p_n}{2^n}\right)\right| \leq c_3 \cdot \left(\frac{p_1}{2} + ... + \frac{p_n}{2^n}\right)^\alpha. \tag{2.18}$$

Passing to the limit in (2.18) we obtain (2.17).

It follows from (2.14), (2.17) and (2.10) that for any $a > 0$, $\delta \in (0,1)$ and for any $x, y \in [-a,a]^2$, $0 \leq x - y \leq \delta$, we have

$$|g^*(x) - g^*(y)| \leq |g^*(x-y)| + |G(y, x-y)|$$

$$\leq c_3 \cdot |x-y|^\alpha + H\left(F; \alpha; [-a,a]^2\right) \cdot |x-y|^\alpha.$$

This means that $g^* \in H_\alpha^{(loc)}(\mathbb{R})$.

Consider the function

$$f(x) = g^*(x) + F(0,0).$$

Note that $f(x)$ also belongs to $H_\alpha^{(loc)}(\mathbb{R})$ and we have the equality

$$F(x,y) = G(x,y) + F(0,0) = g^*(x) + g^*(y) - g^*(x+y) + F(0,0)$$

$$= f(x) + f(y) - f(x+y).$$

The lemma has been proved.

Now we are ready to prove the following theorem.

Theorem 2.1. *Assume a function $F \in H_\alpha^{(loc)}\left(\mathbb{R}^2\right)$, $0 < \alpha \leq 1$, has the form*

$$F(x,y) = h_1(a_1x + b_1y) + h_2(a_2x + b_2y) + h_3(a_3x + b_3y), \tag{2.19}$$

where the (a_i, b_i), $i = 1,2,3$, are pairwise linearly independent directions in $\mathbb{R}^2$ and $h_i : \mathbb{R} \to \mathbb{R}$, $i = 1,2,3$. Then there exist functions $f_i \in H_\alpha^{(loc)}(\mathbb{R})$, $i = 1,2,3$, such that

$$F(x,y) = f_1(a_1x + b_1y) + f_2(a_2x + b_2y) + f_3(a_3x + b_3y).$$

Proof. Since the vectors (a_i, b_i), $i = 1, 2, 3$ are pairwise linearly independent, we can apply a nonsingular linear transformation $S : \mathbb{R}^2 \to \mathbb{R}^2$ of the coordinates such that $S((a_1, b_1)) = (1, 0)$ and $S((a_2, b_2)) = (0, 1)$. Therefore without loss of generality we may assume that $(a_1, b_1) = (1, 0)$, $(a_2, b_2) = (0, 1)$ and herewith $a_3 \neq 0$, $b_3 \neq 0$. Thus we prove the theorem if we prove that for any function $F \in H_\alpha^{(loc)}(\mathbb{R}^2)$ of the form

$$F(x, y) = h_1(x) + h_2(y) + h_3(a_3x + b_3y) \tag{2.20}$$

there exists $f_i \in H_\alpha^{(loc)}(\mathbb{R})$, $i = 1, 2, 3$, such that

$$F(x, y) = f_1(x) + f_2(y) + f_3(a_3x + b_3y). \tag{2.21}$$

Since $a_3 \neq 0$, $b_3 \neq 0$, it follows from (2.20) that

$$F\left(\frac{x}{a_3}, \frac{y}{b_3}\right) = h_1\left(\frac{x}{a_3}\right) + h_2\left(\frac{y}{b_3}\right) + h_3(x + y), \tag{2.22}$$

$$F\left(\frac{x}{a_3}, 0\right) = h_1\left(\frac{x}{a_3}\right) + h_2(0) + h_3(x), \tag{2.23}$$

$$F\left(0, \frac{y}{b_3}\right) = h_1(0) + h_2\left(\frac{y}{b_3}\right) + h_3(y). \tag{2.24}$$

Consider the function

$$H(x, y) = F\left(\frac{x}{a_3}, 0\right) + F\left(0, \frac{y}{b_3}\right) - F\left(\frac{x}{a_3}, \frac{y}{b_3}\right) - h_1(0) - h_2(0),$$

Note that the function $H(x, y)$ belongs to the class $H_\alpha^{(loc)}(\mathbb{R}^2)$ and from (2.22) - (2.24) it follows that

$$H(x, y) = h_3(x) + h_3(y) - h_3(x + y).$$

Applying Lemma 2.1, we obtain that there exists a function $f \in H_\alpha^{(loc)}(\mathbb{R})$ such that

$$H(x, y) = f(x) + f(y) - f(x + y).$$

Introduce the following functions

$$f_1(x) = F(x, 0) - h_1(0) - f(a_3x), \; f_2(x) = F(0, x) - h_2(0) - f(b_3x), \; f_3(x) = f(x).$$

It is not difficult to see that $f_i \in H_\alpha^{(loc)}(R)$, $i = 1, 2, 3$, and

$$\begin{gathered} f_1(x) + f_2(y) + f_3(a_3x + b_3y) \\ = [F(x, 0) - h_1(0) - f(a_3x)] + [F(0, y) - h_2(0) - f(b_3y)] + f(a_3x + b_3y) \\ = F(x, 0) + F(0, y) - h_1(0) - h_2(0) - [f(a_3x) + f(b_3y) - f(a_3x + b_3y)] \\ = F(x, 0) + F(0, y) - h_1(0) - h_2(0) - H(a_3x, b_3y) \\ = F(x, 0) + F(0, y) - h_1(0) - h_2(0) - [F(x, 0) + F(0, y) - F(x, y) - h_1(0) - h_2(0)] \\ = F(x, y). \end{gathered}$$

Thus we obtain that (2.21) holds. The theorem has been proved.

Using the multidimensional techniques exploited in our previous paper [2], Theorem 2.1 can be proven in a more general case. Since the proof for such a generalization is purely technical, we only formulate the final result.

Theorem 2.2. *Assume we are given r directions $\mathbf{a}^i$, $i = 1, ..., r$, in $\mathbb{R}^m \backslash \{\mathbf{0}\}$ and $r - 1$ of them are linearly independent. Assume that a function $F \in H_\alpha^{(loc)}(\mathbb{R}^m)$, $0 < \alpha \leq 1$, is of the form*

$$F(\mathbf{x}) = \sum_{i=1}^{r} g_i(\mathbf{a}^i \cdot \mathbf{x}).$$

Then F can be represented also in the form

$$F(\mathbf{x}) = \sum_{i=1}^{r} f_i(\mathbf{a}^i \cdot \mathbf{x}),$$

where the functions $f_i \in H_\alpha^{(loc)}(\mathbb{R})$, $i = 1, ..., r$.

Acknowledgements. This research was supported by the Science Development Foundation under the President of the Republic of Azerbaijan (grant no. EIF/MQM/Elm-Tehsil-1-2016- 1(26)-71/08/01).

References

[1] J. Aczél, *Functional Equations and their Applications*, Academic Press, New York, 1966.
[2] R.A. Aliev, A.A. Asgarova and V.E. Ismailov, A note on continuous sums of ridge functions, *J. Approx. Theory* **237** (2019), 210–221.
[3] R.A. Aliev and V.E. Ismailov, On a smoothness problem in ridge function representation, *Adv. in Appl. Math.* **73** (2016), 154–169.
[4] R.A. Aliev and V.E. Ismailov, On the representation by bivariate ridge functions, arXiv preprint https://arxiv.org/abs/1606.07940
[5] R.A. Aliev, V.E. Ismailov, T.M. Shahbalayeva, On the representation by sums of ridge functions, *Proc. Inst. Math. Mech. Natl. Acad. Sci. Azerb.* **41** (2015), no. 2, 106–118.
[6] M.D. Buhmann and A. Pinkus, Identifying linear combinations of ridge functions, *Adv. in Appl. Math.* **22** (1999), 103-118.
[7] D.L. Donoho and I.M. Johnstone, Projection-based approximation and a duality method with kernel methods, *Ann. Statist.* **17** (1989), 58-106.
[8] J.H. Friedman and W. Stuetzle, Projection pursuit regression, *J.Amer. Statist. Assoc.* **76** (1981), 817-823.
[9] P.J. Huber, Projection pursuit, *Ann. Statist.* **13** (1985), 435-475.
[10] V.E. Ismailov, Approximation by sums of ridge functions with fixed directions. (Russian); translated from *Algebra i Analiz* **28** (2016), no. 6, 20–69 *St. Petersburg Math. J.* **28** (2017), no. 6, 741-772.
[11] V.E. Ismailov, A note on the equioscillation theorem for best ridge function approximation, *Expo. Math.* **35** (2017), no. 3, 343–349.
[12] V.E. Ismailov, Approximation by ridge functions and neural networks with a bounded number of neurons. *Appl. Anal.* **94** (2015), no. 11, 2245–2260.
[13] V.E. Ismailov, On the best L_2 approximation by ridge functions, *Appl. Math. E-Notes* **7** (2007), 71-76.
[14] F. John, *Plane Waves and Spherical Means Applied to Partial Differential Equations*, Interscience, New York, 1955.
[15] I. Kazantsev, Tomographic reconstruction from arbitrary directions using ridge functions, *Inverse Problems* **14** (1998), 635-645.

[16] I. Kazantsev, I. Lemahieu, Reconstruction of elongated structures using ridge functions and natural pixels, *Inverse Problems* **16** (2000), 505-517.
[17] S.V. Konyagin, A.A. Kuleshov, On the continuity of finite sums of ridge functions. (Russian); translated from *Mat. Zametki* **98** (2015), no. 2, 308–309 *Math. Notes* **98** (2015), no. 1-2, 336–338.
[18] S.V. Konyagin, A.A. Kuleshov, On some properties of finite sums of ridge functions defined on convex subsets of $\mathbb{R}^n$. (Russian) *Tr. Mat. Inst. Steklova* **293** (2016), Funktsionalnye Prostranstva, Teoriya Priblizheniĭ, Smezhnye Razdely Matematicheskogo An., 193–200.
[19] A. Kroo, On approximation by ridge functions, *Constr. Approx.* **13** (1997), 447-460.
[20] A.A. Kuleshov, On some properties of smooth sums of ridge functions. (Russian) *Tr. Mat. Inst. Steklova* **294** (2016), Sovremennye Problemy Matematiki, Mekhaniki i Matematicheskoĭ Fiziki. II, 99–104.
[21] V.Ya. Lin and A. Pinkus, Fundamentality of ridge functions, *J. Approx. Theory* **75** (1993), 295-311.
[22] B.F. Logan and L.A. Shepp, Optimal reconstruction of a function from its projections, *Duke Math.J.* **42** (1975), 645-659.
[23] V.E. Maiorov, On best approximation by ridge functions, *J.Approx. Theory* **99** (1999), 68-94.
[24] R.B. Marr, On the reconstruction of a function on a circular domain from a sampling of its line integrals. *J. Math. Anal. Appl.* **45** (1974), 357–374.
[25] F. Natterer, *The Mathematics of Computerized Tomography*, Wiley, New York, 1986.
[26] P.P. Petrushev, Approximation by ridge functions and neural networks, *SIAM J.Math. Anal.* **30** (1998), 155-189.
[27] A. Pinkus, *Ridge Functions*, Cambridge Tracts in Mathematics, 205. Cambridge University Press, Cambridge, 2015.
[28] A. Pinkus, Smoothness and uniqueness in ridge function representation, *Indag. Math. (N.S.)* **24** (2013), no. 4, 725–738.
[29] A. Pinkus, Approximating by ridge functions, in: *Surface Fitting and Multiresolution Methods*, (A.Le Méhauté, C.Rabut and L.L.Schumaker, eds), Vanderbilt Univ.Press (Nashville), 1997, 279-292.
[30] A. Pinkus, Approximation theory of the MLP model in neural networks, *Acta Numerica* **8** (1999), 143-195.

Rashid A. Aliev
Baku State University, Baku, AZ 1148, Azerbaijan,
Institute of Mathematics and Mechanics, NAS of Azerbaijan, Baku, AZ 1141, Azerbaijan
E-mail address: `aliyevrashid@mail.ru`

Aysel A. Asgarova
Azerbaijan University of Languages, Baku, AZ 1014, Azerbaijan
E-mail address: `asgarova2016@mail.ru`

Vugar E. Ismailov
Institute of Mathematics and Mechanics, NAS of Azerbaijan, Baku, AZ 1141, Azerbaijan
E-mail address: `vugaris@mail.ru`

Received: October 10, 2018; Accepted: February 5, 2019