DOI: 10.37863/umzh.v73i5.263

UDC 517.5

R. A. Aliev (Inst. Math. and Mech. Nat. Acad. Sci. Azerbaijan, Baku and Baku State Univ., Azerbaijan),
A. A. Asgarova (Azerbaijan Univ. Languages, Baku),
V. E. Ismailov (Inst. Math. and Mech. Nat. Acad. Sci. Azerbaijan, Baku)

ON THE REPRESENTATION BY BIVARIATE RIDGE FUNCTIONS*

ПРО ЗОБРАЖЕННЯ ГРЕБЕНЕВИМИ ФУНКЦІЯМИ ДВОХ ЗМІННИХ

We consider the problem of representation of a bivariate function by sums of ridge functions. It is shown that if a function of a certain smoothness class is represented by a sum of finitely many arbitrarily behaved ridge functions, then it can also be represented by a sum of ridge functions of the same smoothness class. As an example, this result is applied to a homogeneous constant coefficient partial differential equation.

Розглядається задача зображення двовимірної функції сумами гребеневих функцій. Показано, що коли функцію певного класу гладкості зображено скінченною сумою гребеневих функцій довільної поведінки, її також можна зобразити сумою гребеневих функцій того ж класу гладкості. Як приклад цей результат застосовано до однорідного диференціального рівняння з частинними похідними і зі сталими коефіцієнтами.

1. Introduction. Last 30 years have seen a growing interest in the study of special multivariate functions called ridge functions. This interest is due to applicability of such functions in various research areas. 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 $g:\mathbb{R}\to\mathbb{R}$ and $\mathbf{a}=(a_1,\ldots,a_m)$ is a fixed vector (direction) in $\mathbb{R}^m\backslash\{\mathbf{0}\}$. These functions and their linear combinations find applications in computerized tomography (see, e.g., [16, 21, 24]), in statistics (especially, in the theory of projection pursuit and projection regression; see, e.g., [4, 6]) and in the theory of neural networks (see, e.g., [7, 9, 11, 23, 28]). Ridge functions are also widely used in modern approximation theory as an effective and convenient tool for approximating complicated multivariate functions (see, e.g., [8, 13, 20, 22, 25]). For more on ridge functions and application areas, see the book [26] and survey papers [10, 12, 19].

It should be remarked that ridge functions have been used in the theory of partial differential equations under the name of *plane waves* (see, e.g., [15]). In general, linear combinations of ridge functions with fixed directions occur in the study of hyperbolic constant coefficient partial differential equations. For example, assume that $(\alpha_i,\beta_i),\ i=1,\ldots,r$, are pairwise linearly independent vectors in $\mathbb{R}^2$. Then the general solution to the homogeneous equation

$$\prod_{i=1}^{r}\left(\alpha_i\frac{\partial}{\partial x}+\beta_i\frac{\partial}{\partial y}\right)u(x,y)=0 \tag{1.1}$$

are all functions of the form

* 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).

$$u(x,y) = \sum_{i=1}^{r} v_i(\beta_i x - \alpha_i y) \tag{1.2}$$

for arbitrary univariate functions $v_i,\ i = 1,\ldots,r,$ from the class $C^r(\mathbb{R})$.

Note that the solution of Eq. (1.1) is the sum of bivariate ridge functions. Sums of bivariate ridge functions also occur in basic mathematical problems of computerized tomography. For example, Logan and Shepp [15] (the term "ridge function" was coined by them) considered the problem of reconstructing a given but unknown function $f(x,y)$ from its integrals along certain lines in the plane. More precisely, let D be the unit disk in the plane and a function $f(x,y)$ be square integrable and supported on D. We are given projections $P_f(t,\theta)$ (integrals of f along the lines $x\cos\theta + y\sin\theta = t$) and looking for a function $g = g(x,y)$ of minimum L_2 norm, which has the same projections as f: $P_g(t,\theta_j) = P_f(t,\theta_j),\ j = 0,1,\ldots,n-1,$ where angles θ_j generate equally spaced directions, i.e., $\theta_j = \dfrac{j\pi}{n},\ j = 0,1,\ldots,n-1$. The authors of [15] showed that this problem of tomography is equivalent to the problem of L_2-approximation of the function f by sums of bivariate ridge functions with equally spaced directions $(\cos\theta_j, \sin\theta_j),\ j = 0,1,\ldots,n-1$. They gave a closed-form expression for the unique function $g(x,y)$ and showed that the unique polynomial $P(x,y)$ of degree $n-1$ which best approximates f in $L_2(D)$ is determined from the above n projections of f and can be represented as a sum of n bivariate ridge functions.

In this paper, we are interested in the problem of smoothness in representation by sums of bivariate ridge functions with finitely many fixed directions. Assume we are given n pairwise linearly independent directions $(a_i,b_i),\ i = 1,\ldots,n,$ in $\mathbb{R}^2$ and a function $F\colon \mathbb{R}^2 \to \mathbb{R}$ of the form

$$F(x,y) = \sum_{i=1}^{n} g_i(a_i x + b_i y). \tag{1.3}$$

Assume in addition that F is of a certain smoothness class, what can we say about the smoothness of g_i? The case $n = 1$ is obvious. In this case, if $F \in C^k(\mathbb{R}^2)$, then for a vector $(c,d) \in \mathbb{R}^2$ satisfying $a_1c + b_1d = 1$ we have that $g_1(t) = F(ct,dt)$ is in $C^k(\mathbb{R})$. The same argument can be carried out for the case $n = 2$. In this case, since the vectors (a_1,b_1) and (a_2,b_2) are linearly independent, there exists a vector $(c,d) \in \mathbb{R}^2$ satisfying $a_1c + b_1d = 1$ and $a_2c + b_2d = 0$. Therefore, we obtain that the function $g_1(t) = F(ct,dt) - g_2(0)$ is in the class $C^k(\mathbb{R})$. Similarly, one can verify that $g_2 \in C^k(\mathbb{R})$.

The picture drastically changes if the number of directions $n \geq 3$. For $n = 3$, there are ultimately smooth functions which decompose into sums of very badly behaved ridge functions. This phenomena comes from the classical Cauchy functional equation. This equation,

$$f(x+y) = f(x) + f(y), \quad f\colon \mathbb{R} \to \mathbb{R}, \tag{1.4}$$

looks very simple and has a class of simple solutions $f(x) = cx,\ c \in \mathbb{R}$. Nevertheless, it easily follows from the Hamel basis theory that the Cauchy functional equation has also a large class of wild solutions. These solutions are called "wild" because they are extremely pathological 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 g be any wild solution of the equation (1.4). Then the zero function can be represented as

$$0 = g(x) + g(y) - g(x + y). \tag{1.5}$$

Note that the functions involved in (1.5) are bivariate ridge functions with the directions $(1, 0)$, $(0, 1)$ and $(1, 1)$, respectively. This example shows that for smoothness of the representation (1.3) one must impose additional conditions on the representing functions g_i, $i = 1, \ldots, n$.

Such additional conditions are recently found by Pinkus [27]. He proved that for a large class of representing functions g_i, the representation is smooth. That is, if apriori assume that in the representation (1.3), the functions g_i belong to a certain class of "well behaved functions", then they have the same degree of smoothness as the function F. As the mentioned class of "well behaved functions" one may take, e.g., the set of functions that are continuous at a point, bounded on one side on a set of positive measure, monotonic at an interval, Lebesgue measurable, etc. (see [27]). Konyagin and Kuleshov [17] proved that in (1.3) the functions g_i inherit smoothness properties of F (without additional assumptions on g_i) if and only if the directions $\mathbf{a}^i$ are linearly independent. Note that the results of Pinkus and also Konyagin and Kuleshov are valid not only in bivariate but also in multivariate case. There are also other results on ridge function representation, which involve certain convex subsets of the m-dimensional space (see [17, 18]).

In this paper, we study a different aspect of the problem of representation by ridge functions. Assume in the representation (1.3) $F \in C^k(\mathbb{R}^2)$ but the functions g_i are arbitrary. That is, we allow very badly behaved functions (for example, not continuous at any point). Can we write F as a sum $\sum_{i=1}^{n} f_i(a_i x + b_i y)$ but with the $f_i \in C^k(\mathbb{R})$, $i = 1, \ldots, n$? We see that the answer to this question is positive as expected. For the sake of convenience we state the result over $\mathbb{R}^2$, but in fact it holds over any open set in $\mathbb{R}^2$.

Note that the above problem is not elementary as it seems. There are cases when representation with good functions is not possible. Such situations happen over closed sets with no interior. In [14], Ismailov and Pinkus presented an example of a function of the form

$$F(x, y) = g_1(a_1 x + b_1 y) + g_2(a_2 x + b_2 y),$$

that is bounded and continuous on the union of two straight lines but such that both g_1 and g_2 are necessarily discontinuous, and thus cannot be replaced with continuous functions f_1 and f_2.

The result of this paper can be applied to a higher order partial differential equation in two variables if its solution is given by a sum of sufficiently smooth plane waves (see, for example, Eq. (1.1)). Based on our theorem below, in this case, one can demand only smoothness of the sum and dispense with smoothness of the plane wave summands.

2. Main results. We start this section with the following theorem.

Theorem 2.1. *Assume that (a_i, b_i), $i = 1, \ldots, n$, are pairwise linearly independent vectors in $\mathbb{R}^2$ and a function $F \in C^k(\mathbb{R}^2)$ has the form*

$$F(x, y) = \sum_{i=1}^{n} g_i(a_i x + b_i y), \tag{2.1}$$

where g_i are arbitrary univariate functions and $k \geq n - 2$. Then F can be represented also in the form

$$F(x,y)=\sum_{i=1}^{n} f_i(a_ix+b_iy).$$

Here, the functions $f_i\in C^k(\mathbb{R}),\ i=1,\ldots,n.$

Proof. Since the vectors (a_{n-1},b_{n-1}) and (a_n,b_n) are linearly independent, there is a nonsingular linear transformation $S\colon (x,y)\to(x',y')$ such that $S\colon (a_{n-1},b_{n-1})\to(1,0)$ and $S\colon (a_n,b_n)\to(0,1)$. Thus, without loss of generality we may assume that the vectors (a_{n-1},b_{n-1}) and (a_n,b_n) coincide with the coordinate vectors $e_1=(1,0)$ and $e_2=(0,1)$, respectively. Therefore, to prove the theorem it is enough to show that if a function $F\in C^k(\mathbb{R}^2)$ is expressed in the form

$$F(x,y)=\sum_{i=1}^{n-2} g_i(a_ix+b_iy)+g_{n-1}(x)+g_n(y)$$

with arbitrary g_i, then there exist functions $f_i\in C^k(\mathbb{R}),\ i=1,\ldots,n$, such that F is expressed also in the form

$$F(x,y)=\sum_{i=1}^{n-2} f_i(a_ix+b_iy)+f_{n-1}(x)+f_n(y).$$

By $\Delta_l^{(\delta)}f$ we denote the increment of a function f in a direction $l=(l',l'')$. That is,

$$\Delta_l^{(\delta)}f(x,y)=f(x+l'\delta,y+l''\delta)-f(x,y).$$

We also use the notation $\dfrac{\partial f}{\partial l}$ which denotes the derivative of f in the direction l.

It is easy to check that the increment of a ridge function $g(ax+by)$ in a direction perpendicular to (a,b) is zero. Let $l_1,\ldots,l_{n-2}$ be unit vectors perpendicular to the vectors $(a_1,b_1),\ldots,(a_{n-2},b_{n-2})$ correspondingly. Then for any set of numbers $\delta_1,\ldots,\delta_{n-2}\in\mathbb{R}$ we have

$$\Delta_{l_1}^{(\delta_1)}\ldots\Delta_{l_{n-2}}^{(\delta_{n-2})}F(x,y)=\Delta_{l_1}^{(\delta_1)}\ldots\Delta_{l_{n-2}}^{(\delta_{n-2})}\big[g_{n-1}(x)+g_n(y)\big]. \tag{2.2}$$

Denote the left-hand side of (2.2) by $S(x,y)$. That is, set

$$S(x,y)\overset{\text{def}}{=}\Delta_{l_1}^{(\delta_1)}\ldots\Delta_{l_{n-2}}^{(\delta_{n-2})}F(x,y).$$

Then from (2.2) it follows that, for any real numbers δ_{n-1} and δ_n,

$$\Delta_{e_1}^{(\delta_{n-1})}\Delta_{e_2}^{(\delta_n)}S(x,y)=0,$$

or, in expanded form,

$$S(x+\delta_{n-1},y+\delta_n)-S(x,y+\delta_n)-S(x+\delta_{n-1},y)+S(x,y)=0.$$

Putting in the last equality $\delta_{n-1}=-x,\ \delta_n=-y$, we obtain

$$S(x,y)=S(x,0)+S(0,y)-S(0,0).$$

This means that

$$\Delta_{l_1}^{(\delta_1)} \ldots \Delta_{l_{n-2}}^{(\delta_{n-2})} F(x,y) = \Delta_{l_1}^{(\delta_1)} \ldots \Delta_{l_{n-2}}^{(\delta_{n-2})} F(x,0) +$$

$$+ \Delta_{l_1}^{(\delta_1)} \ldots \Delta_{l_{n-2}}^{(\delta_{n-2})} F(0,y) - \Delta_{l_1}^{(\delta_1)} \ldots \Delta_{l_{n-2}}^{(\delta_{n-2})} F(0,0).$$

By the hypothesis of the theorem, the derivative $\dfrac{\partial^{n-2}}{\partial l_1 \ldots \partial l_{n-2}} F(x,y)$ exists at any point $(x,y) \in$ $\in \mathbb{R}^2$. Thus, it follows from the above formula that

$$\frac{\partial^{n-2} F}{\partial l_1 \ldots \partial l_{n-2}}(x,y) = h_{1,1}(x) + h_{2,1}(y), \tag{2.3}$$

where $h_{1,1}(x) = \dfrac{\partial^{n-2}}{\partial l_1 \ldots \partial l_{n-2}} F(x,0)$ and $h_{2,1}(y) = \dfrac{\partial^{n-2}}{\partial l_1 \ldots \partial l_{n-2}} F(0,y) - \dfrac{\partial^{n-2}}{\partial l_1 \ldots \partial l_{n-2}} F(0,0)$. Note that $h_{1,1}$ and $h_{2,1}$ belong to the class $C^{k-n+2}(\mathbb{R})$.

By $h_{1,2}$ and $h_{2,2}$ denote the antiderivatives of $h_{1,1}$ and $h_{2,1}$ satisfying the condition $h_{1,2}(0) =$ $= h_{2,2}(0) = 0$ and multiplied by the numbers $1/(e_1 \cdot l_1)$ and $1/(e_2 \cdot l_1)$ correspondingly. That is,

$$h_{1,2}(x) = \frac{1}{e_1 \cdot l_1} \int_0^x h_{1,1}(z) dz,$$

$$h_{2,2}(y) = \frac{1}{e_2 \cdot l_1} \int_0^y h_{2,1}(z) dz.$$

Here, $e \cdot l$ denotes the scalar product between vectors e and l. Obviously, the function

$$F_1(x,y) = h_{1,2}(x) + h_{2,2}(y)$$

obeys the equality

$$\frac{\partial F_1}{\partial l_1}(x,y) = h_{1,1}(x) + h_{2,1}(y). \tag{2.4}$$

From (2.3) and (2.4) we obtain

$$\frac{\partial}{\partial l_1} \left[\frac{\partial^{n-3} F}{\partial l_2 \ldots \partial l_{n-2}} - F_1 \right] = 0.$$

Hence, for some ridge function $\varphi_{1,1}(a_1 x + b_1 y)$,

$$\frac{\partial^{n-3} F}{\partial l_2 \ldots \partial l_{n-2}}(x,y) = h_{1,2}(x) + h_{2,2}(y) + \varphi_{1,1}(a_1 x + b_1 y). \tag{2.5}$$

Here, all the functions $h_{2,1},\ h_{2,2}(y),\ \varphi_{1,1} \in C^{k-n+3}(\mathbb{R})$.

Set the functions

$$h_{1,3}(x) = \frac{1}{e_1 \cdot l_2} \int_0^x h_{1,2}(z) dz,$$

$$h_{2,3}(y) = \frac{1}{e_2 \cdot l_2} \int\limits_0^y h_{2,2}(z)dz,$$

$$\varphi_{1,2}(t) = \frac{1}{(a_1, b_1) \cdot l_2} \int\limits_0^t \varphi_{1,1}(z)dz.$$

Note that the function

$$F_2(x, y) = h_{1,3}(x) + h_{2,3}(y) + \varphi_{1,2}(a_1x + b_1y)$$

obeys the equality

$$\frac{\partial F_2}{\partial l_2}(x, y) = h_{1,2}(x) + h_{2,2}(y) + \varphi_{1,1}(a_1x + b_1y). \tag{2.6}$$

From (2.5) and (2.6) it follows that

$$\frac{\partial}{\partial l_2}\left[\frac{\partial^{n-4}F}{\partial l_3 \ldots \partial l_{n-2}} - F_2\right] = 0.$$

The last equality means that, for some ridge function $\varphi_{2,1}(a_2x + b_2y)$,

$$\frac{\partial^{n-4}F}{\partial l_3 \ldots \partial l_{n-2}}(x, y) = h_{1,3}(x) + h_{2,3}(y) + \varphi_{1,2}(a_1x + b_1y) + \varphi_{2,1}(a_2x + b_2y). \tag{2.7}$$

Here, all the functions $h_{1,3},\ h_{2,3},\ \varphi_{1,2},\ \varphi_{2,1} \in C^{k-n+4}(\mathbb{R})$.

Note that in the left-hand sides of (2.3), (2.5) and (2.7) we have the mixed directional derivatives of F and the order of these derivatives is decreased by one in each consecutive step. Continuing the above process, until it reaches the function F, we obtain the desired result.

Theorem 2.1 is proved.

Theorem 2.1 can be applied to Eq. (1.1) as follows.

***Corollary* 2.1.** *Assume a function $u \in C^r(\mathbb{R}^2)$ is of the form* (1.2) *with arbitrarily behaved v_i. Then u is a solution to the equation* (1.1).

Note that the method exploited in the proof of Theorem 2.1 enables us to construct the functions f_i, $i = 1, \ldots, n$, by induction. First let us accept some notation. By $(\widetilde{a}_p, \widetilde{b}_p)$, $p = 1, \ldots, n-2$, denote the images of vectors (a_p, b_p) under the linear transformation which takes the vectors (a_{n-1}, b_{n-1}) and (a_n, b_n) to the unit vectors $e_1 = (1, 0)$ and $e_2 = (0, 1)$, respectively. Clearly,

$$\widetilde{a}_p = \frac{a_pb_n - a_nb_p}{a_{n-1}b_n - a_nb_{n-1}}, \qquad \widetilde{b}_p = \frac{a_{n-1}b_p - a_pb_{n-1}}{a_{n-1}b_n - a_nb_{n-1}}, \quad p = 1, \ldots, n-2.$$

Consider the vectors

$$l_p = \left(\frac{\widetilde{b}_p}{\sqrt{\widetilde{a}_p^2 + \widetilde{b}_p^2}}, \frac{-\widetilde{a}_p}{\sqrt{\widetilde{a}_p^2 + \widetilde{b}_p^2}}\right), \quad p = 1, \ldots, n-2.$$

Note that for $p = 1, \ldots, n-2$, the vectors l_p are perpendicular to the vectors $(\widetilde{a}_p, \widetilde{b}_p)$, respectively. Consider also the function, which is generated by the above liner transformation

$$F^*(x, y) = F\left(\frac{b_nx - b_{n-1}y}{a_{n-1}b_n - a_nb_{n-1}}, \quad \frac{a_nx - a_{n-1}y}{a_nb_{n-1} - a_{n-1}b_n}\right).$$

***Corollary* 2.2.** *The functions* f_i, $i = 1, \ldots, n$, *in Theorem* 2.1 *can be constructed inductively by the formulas*

$$f_p = \varphi_{p,n-p-1}, \quad p = 1, \ldots, n-2,$$

$$f_{n-1} = h_{1,n-1}, \qquad f_n = h_{2,n-1}.$$

Here,

$$h_{1,1}(t) = \frac{\partial^{n-2}}{\partial l_1 \ldots \partial l_{n-2}} F^*(t,0),$$

$$h_{2,1}(t) = \frac{\partial^{n-2}}{\partial l_1 \ldots \partial l_{n-2}} F^*(0,t) - \frac{\partial^{n-2}}{\partial l_1 \ldots \partial l_{n-2}} F^*(0,0),$$

$$h_{1,k+1}(t) = \frac{1}{e_1 \cdot l_k} \int\limits_0^t h_{1,k}(z)dz, \quad k = 1, \ldots, n-2,$$

$$h_{2,k+1}(t) = \frac{1}{e_2 \cdot l_k} \int\limits_0^t h_{2,k}(z)dz, \quad k = 1, \ldots, n-2,$$

and

$$\varphi_{p,1}(t) = \frac{\partial^{n-p-2} F^*}{\partial l_{p+1} \ldots \partial l_{n-2}} \left(\frac{\widetilde{a}_p t}{\widetilde{a}_p^2 + \widetilde{b}_p^2}, \frac{\widetilde{b}_p t}{\widetilde{a}_p^2 + \widetilde{b}_p^2} \right) - h_{1,p+1} \left(\frac{\widetilde{a}_p t}{\widetilde{a}_p^2 + \widetilde{b}_p^2} \right) -$$

$$-h_{2,p+1} \left(\frac{\widetilde{b}_p t}{\widetilde{a}_p^2 + \widetilde{b}_p^2} \right) - \sum_{s=1}^{p-1} \varphi_{s,p-s+1} \left(\frac{\widetilde{a}_s \widetilde{a}_p + \widetilde{b}_s \widetilde{b}_p}{\widetilde{a}_p^2 + \widetilde{b}_p^2} t \right),$$

$$p = 1, \ldots, n-2 \quad \left(\textit{for } p = n-2, \frac{\partial^{n-p-2} F^*}{\partial l_{p+1} \ldots \partial l_{n-2}} := F^* \right),$$

$$\varphi_{p,k+1}(t) = \frac{1}{(\widetilde{a}_p, \widetilde{b}_p) \cdot l_{k+p}} \int\limits_0^t \varphi_{p,k}(z)dz, \quad p = 1, \ldots, n-3, \quad k = 1, \ldots, n-p-2.$$

The validity of above formulas for the functions $h_{1,k}$ and $h_{2,k}$, $k = 1, \ldots, n-1$, is obvious. The formulas for $\varphi_{p,1}$ and $\varphi_{p,k+1}$ can be obtained from (2.5), (2.7) and the subsequent (assumed but not written) equations if we put $x = \widetilde{a}_p t / (\widetilde{a}_p^2 + \widetilde{b}_p^2)$ and $y = \widetilde{b}_p t / (\widetilde{a}_p^2 + \widetilde{b}_p^2)$.

***Remark* 2.1.** If in Theorem 2.1 $k \geq n-1$, then the functions f_i, $i = 1, \ldots, n$, can be constructed (up to polynomials) by the method discussed in Buhmann and Pinkus [3]. This method is based on the fact that for a direction $\mathbf{c} = (c_1, \ldots, c_m)$ orthogonal to a given direction $\mathbf{a} \in \mathbb{R}^m \backslash \{\mathbf{0}\}$, the operator

$$D_{\mathbf{c}} = \sum_{s=1}^m c_s \frac{\partial}{\partial x_s}$$

acts on m-variable ridge functions $g(\mathbf{a}\cdot\mathbf{x})$ as follows:

$$D_{\mathbf{c}}g(\mathbf{a}\cdot\mathbf{x}) = (\mathbf{c}\cdot\mathbf{a})\,g'(\mathbf{a}\cdot\mathbf{x}).$$

Thus, if in our case for fixed $r\in\{1,\ldots,n\}$, vectors $l_k,\ k\in\{1,\ldots,n\},\ k\neq r$, are perpendicular to the vectors (a_k,b_k), then

$$\prod_{\substack{k=1\\k\neq r}}^{n} D_{l_k}F(x,y) = \prod_{\substack{k=1\\k\neq r}}^{n} D_{l_k}\sum_{i=1}^{n} f_i(a_ix+b_iy) =$$

$$= \sum_{i=1}^{n}\left(\prod_{\substack{k=1\\k\neq r}}^{n}((a_i,b_i)\cdot l_k)\right) f_i^{(n-1)}(a_ix+b_iy) =$$

$$= \prod_{\substack{k=1\\k\neq r}}^{n}((a_r,b_r)\cdot l_k)\,f_r^{(n-1)}(a_rx+b_ry).$$

Now f_r can be easily constructed from the above formula (up to a polynomial of degree at most $n-2$). Note that this method is not valid if in Theorem 2.1 the function F is of the class $C^{n-2}(\mathbb{R}^2)$. However, in this case, Corollary 2.2 is applicable.

***Remark* 2.2.** Some polynomial terms appear while attempting to obtain a smoothness result in multivariate case. In [2], we proved that if a function $f(x_1,\ldots,x_n)$ of a certain smoothness class is represented by a sum of r arbitrarily behaved ridge functions, then, under suitable conditions, it can be represented by a sum of ridge functions of the same smoothness class and some n-variable polynomial of a certain degree. The appearance of a polynomial term is mainly related to the fact that in $\mathbb{R}^n$, $n\geq 3$, there are many directions orthogonal to a given direction. Note that a polynomial term also appears in verifying if a given function of n variables $(n\geq 3)$ is a sum of ridge functions (see [5]). However, paralleling the above theorem, we conjecture that if a multivariate function of a certain smoothness class is represented by a sum of arbitrarily behaved ridge functions, then it can also be represented by a sum of ridge functions of the same smoothness class.

***Remark* 2.3.** For $k\geq n-1$, Theorem 2.1 can be obtained from Theorem 3.1 of [2]. Indeed, according to Theorem 3.1 [2], the function F in Theorem 2.1 can be represented in the form

$$F(x,y) = \sum_{i=1}^{n} f_i(a_ix+b_iy) + P(x,y), \tag{2.8}$$

where the functions $f_i\in C^k(\mathbb{R})$ and $P(x,y)$ is a polynomial of total degree at most $n-1$. But it is known that a bivariate polynomial of degree $n-1$ is decomposed into a sum of ridge polynomials with any given n pairwise linearly independent directions (see, e.g., [21]). That is, in (2.8)

$$P(x,y) = \sum_{i=1}^{n} p_i(a_ix+b_iy),$$

where p_i are univariate polynomials of degree at most $n-1$. However, in the setting considered here, Theorem 2.1 is more informative. It covers the extra case $k=n-2$ and allows one to construct the representing functions f_i by induction.

Unfortunately, we do not yet know if the lower assumed degree of smoothness $n-2$ in Theorem 2.1 can be reduced. We think that the final and complete solution to the smoothness problem considered here requires essentially different approach. Nevertheless, we can strengthen our result by considering Hölder continuous functions.

We say that a function $F:\mathbb{R}^m\to\mathbb{R},\ m\geq 1,$ 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 $M=M\left(K\right)>0$ such that for any $\mathbf{x}=(x_1,\ldots,x_m)\in K$ and $\mathbf{y}=(y_1,\ldots,y_m)\in K$ 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. By $C^{k,\alpha}(\mathbb{R}^m)$ we denote the class of functions in $C^k(\mathbb{R}^m)$, kth order partial derivatives of which are locally Hölder continuous with degree α.

The following theorem is valid.

Theorem 2.2. *Assume that* $(a_i,b_i),\ i=1,\ldots,n,$ *are pairwise linearly independent vectors in* $\mathbb{R}^2$ *and a function* $F\in C^{k,\alpha}(\mathbb{R}^2)$ *has the form*

$$F(x,y)=\sum_{i=1}^{n}g_i(a_ix+b_iy),$$

where g_i *are arbitrary univariate functions and* $k\geq n-2$. *Then* F *can be represented also in the form*

$$F(x,y)=\sum_{i=1}^{n}f_i(a_ix+b_iy).$$

Here, the functions $f_i\in C^{k,\alpha}(\mathbb{R}),\ i=1,\ldots,n.$

The proof of Theorem 2.2 can be easily obtained from Theorem 2.1. Indeed, to prove Theorem 2.2 it is only needed to repeat the proof of Theorem 2.1, emphasizing that the functions appearing in the right-hand sides of formulas (2.3), (2.5), (2.7), etc. belong to certain classes $C^{s,\alpha}(\mathbb{R})$ with step by step increasing indicator of smoothness s. More precisely, the functions $h_{1,1}$, $h_{2,1}\in C^{k-n+2,\alpha}(\mathbb{R})$, the functions $h_{1,2},\ h_{2,2},\ \varphi_{1,1}\in C^{k-n+3,\alpha}(\mathbb{R})$, the functions $h_{1,3},\ h_{2,3},\ \varphi_{1,2}$, $\varphi_{2,1}\in C^{k-n+4,\alpha}(\mathbb{R})$, etc.

Theorem 2.1 implies that if a function $F\in C^k(\mathbb{R}^2)$ has the form (2.1) and $k\geq n-2$,then all the partial derivatives of F up to order k are representable as a sum of ridge functions with the given directions $(a_i,b_i),\ i=1,\ldots,n$. Note that the validity of Theorem 2.1 for other possible k strongly depends on answers to the following two questions.

***Question* 1.** Assume a function $F\in C^k(\mathbb{R}^2)$ is of form (2.1). Are the first order partial derivatives $\partial F/\partial x$ and $\partial F/\partial y$ representable as a sum of arbitrarily behaved ridge functions with the directions (a_i,b_i)?

***Question* 2.** Assume a function $F\in C(\mathbb{R}^2)$ is of form (2.1). Is it true that F can be represented also in the form $\sum_{i=1}^{n}f_i(a_ix+b_iy)$ with continuous f_i?

Indeed, a positive answer to Question 1 would mean, by induction, that all the partial derivatives up to order k and hence any mixed directional derivative $\partial^kF/\partial l_1\ldots\partial l_k$ are represented by a sum $\sum_{i=1}^{n}g_i(a_ix+b_iy)$, where g_i are arbitrary univariate functions. Once we could answer Question 2 affirmatively, we would immediately obtain that any derivative $\partial^k F/\partial l_1 \ldots \partial l_k$ is also written in the form $\sum_{i=1}^n f_i(a_i x + b_i y)$ with continuous f_i. Then by choosing the directions $l_1, \ldots, l_k$ orthogonal to the first k directions (a_i, b_i), $i = 1, \ldots, n$, and applying the above method (see the proof of Theorem 2.1) we could conclude that F has representation $\sum_{i=1}^n f_i(a_i x + b_i y)$, where $f_i \in C^k(\mathbb{R})$. It should be remarked that Question 2 is a part of the more general question posed in [26, p. 14].

References

1. J. Aczél, *Functional equations and their applications*, Acad. Press, New York (1966).
2. R. A. Aliev, V. E. Ismailov, *On a smoothness problem in ridge function representation*, Adv. Appl. Math., **73**, 154 – 169 (2016).
3. M. D. Buhmann, A. Pinkus, *Identifying linear combinations of ridge functions*, Adv. Appl. Math., **22**, 103 – 118 (1999).
4. E. J. Candès, *Ridgelets: estimating with ridge functions*, Ann. Statist., **31**, 1561 – 1599 (2003).
5. P. Diaconis, M. Shahshahani, *On nonlinear functions of linear combinations*, SIAM J. Sci. and Stat. Comput., **5**, 175 – 191 (1984).
6. J. H. Friedman, W. Stuetzle, *Projection pursuit regression*, J. Amer. Statist. Assoc., **76**, 817 – 823 (1981).
7. N. J. Guliyev, V. E. Ismailov, *On the approximation by single hidden layer feedforward neural networks with fixed weights*, Neural Networks, **98**, 296 – 304 (2018).
8. V. E. Ismailov, *Characterization of an extremal sum of ridge functions*, J. Comput. and Appl. Math., **205**, 105 – 115 (2007).
9. V. E. Ismailov, *Approximation by neural networks with weights varying on a finite set of directions*, J. Math. Anal. and Appl., **389**, 72 – 83 (2012).
10. V. E. Ismailov, *A review of some results on ridge function approximation*, Azerb. J. Math., **3**, № 1, 3 – 51 (2013).
11. V. E. Ismailov, *Approximation by ridge functions and neural networks with a bounded number of neurons*, Appl. Anal., **94**, 2245 – 2260 (2015).
12. V. E. Ismailov, *Approximation by sums of ridge functions with fixed directions* (in Russian), Algebra i Analiz, **28**, № 6, 20 – 69 (2016).
13. V. E. Ismailov, *A note on the equioscillation theorem for best ridge function approximation*, Expo. Math., **35**, № 3, 343 – 349 (2017).
14. V. E. Ismailov, A. Pinkus, *Interpolation on lines by ridge functions*, J. Approx. Theory, **175**, 91 – 113 (2013).
15. F. John, *Plane waves and spherical means applied to partial differential equations*, Intersci., New York (1955).
16. I. Kazantsev, *Tomographic reconstruction from arbitrary directions using ridge functions*, Inverse Problems, **14**, 635 – 645 (1998).
17. S. V. Konyagin, A. A. Kuleshov, *On the continuity of finite sums of ridge functions* (in Russian), Mat. Zametki, **98**, 308 – 309 (2015).
18. S. V. Konyagin, A. A. Kuleshov, *On some properties of finite sums of ridge functions defined on convex subsets of* $\mathbb{R}^n$ (in Russian), Tr. Mat. Inst. Steklova, **293** (2016).
19. S. V. Konyagin, A. A. Kuleshov, V E. Maiorov, *Some problems in the theory of ridge functions* (in Russian), Tr. Mat. Inst. Steklova, **301** (2018).
20. A. Kroó, *On approximation by ridge functions*, Constr. Approx., **13**, 447 – 460 (1997).
21. B. F. Logan, L. A. Shepp, *Optimal reconstruction of a function from its projections*, Duke Math. J., **42**, 645 – 659 (1975).
22. V. E. Maiorov, *On best approximation by ridge functions*, J. Approx. Theory, **99**, 68 – 94 (1999).
23. V. Maiorov, A. Pinkus, *Lower bounds for approximation by MLP neural networks*, Neurocomputing, **25**, 81 – 91 (1999).
24. F. Natterer, *The mathematics of computerized tomography*, Wiley, New York (1986).
25. P. P. Petrushev, *Approximation by ridge functions and neural networks*, SIAM J. Math. Anal., **30**, 155 – 189 (1998).
26. A. Pinkus, *Ridge functions*, Cambridge Tracts Math., **205**, Cambridge Univ. Press, Cambridge (2015).
27. A. Pinkus, *Smoothness and uniqueness in ridge function representation*, Indag. Math. (N.S.), **24**, № 4, 725 – 738 (2013).
28. A. Pinkus, *Approximation theory of the MLP model in neural networks*, Acta Numer., **8**, 143 – 195 (1999).

Received 11.09.18