ON THE REPRESENTATION BY BIVARIATE RIDGE FUNCTIONS

R. A. Aliev, A. A. Asgarova, and V. E. Ismailov

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

1. Introduction

The last three decades have been characterized by a growing interest in the investigation of special multivariate functions called ridge functions. This interest is explained by the applicability of these functions in various research areas. A ridge function is a multivariate function of the form

\[ g(a \cdot x) = g(a_1 x_1 + \ldots + a_m x_m), \]

where \( g : \mathbb{R} \to \mathbb{R} \) and \( a = (a_1, \ldots, a_m) \) is a fixed vector (direction) in \( \mathbb{R}^m \setminus \{0\}. \) 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 extensively used in modern approximation theory as an efficient and convenient tool for the approximation of complicated multivariate functions (see, e.g., [8, 13, 20, 22, 25]). For more detail about ridge functions and the fields of their application, see the book [26] and the surveys [10, 12, 19].

It is worth noting 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 partial differential equations with constant coefficients. 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 of the homogeneous equation

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

are all functions of the form

\[ u(x, y) = \sum_{i=1}^{r} v_i (\beta_i x - \alpha_i y) \] (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. The sums of bivariate ridge functions are also encountered in basic mathematical problems of computerized tomography. Thus, Logan and Shepp [15] (they coined the term “ridge function”) studied the problem of reconstruction of a given but unknown function \( f(x, y) \) according to its integrals along certain lines in the plane. More precisely, let \( D \) be the unit disk in the plane and let \( f(x, y) \) be a square integrable function supported on \( D \). Given projections \( P_f(t, \theta) \) (the integrals of \( f \) along the lines \( x \cos \theta + y \sin \theta = t \)), we seek a function \( g = g(x, y) \) with the minimum \( L_2 \) norm, which has the same projections as \( f \):

\[
P_g(t, \theta_j) = P_f(t, \theta_j), \quad j = 0, 1, \ldots, n - 1,
\]

where the angles \( \theta_j \) generate equally spaced directions, i.e.,

\[
\theta_j = \frac{j \pi}{n}, \quad j = 0, 1, \ldots, n - 1.
\]

The authors of [15] showed that the posed 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

\[
(\cos \theta_j, \sin \theta_j), \quad j = 0, 1, \ldots, n - 1.
\]

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 is the best approximation of \( f \) in \( L_2(D) \), is determined according to the above \( n \) projections of \( f \) and can be represented in the form of a sum of \( n \) bivariate ridge functions.

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

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

In addition, assume that \( F \) belongs to a certain class of smoothness. Then what can we said 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 \( a_1 c + b_1 d = 1 \), we conclude that \( g_1(t) = F(ct, dt) \) is in \( C^k(\mathbb{R}) \). The same argument can be used in 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 \( a_1 c + b_1 d = 1 \) and \( a_2 c + b_2 d = 0 \). Therefore, we conclude that the function \( g_1(t) = F(ct, dt) - g_2(0) \) belongs to the class \( C^k(\mathbb{R}) \). Similarly, one can show 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 decomposing into the sums of ridge functions with very bad behaviors. This phenomenon comes from the classical Cauchy functional equation. This equation,

\[
f(x + y) = f(x) + f(y), \quad f: \mathbb{R} \to \mathbb{R}, \tag{1.4}
\]

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 also has a large class of wild solutions. These solutions are called “wild” because they are extremely pathological over reals. They are, e.g., not continuous at a point, not monotone in an interval, and not bounded for any set of positive measure (see, e.g., [1]). Let \( g \) be any wild solution of 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 directions \((1, 0), (0, 1),\) and \((1, 1),\) respectively. This example shows that, for the smoothness of representation (1.3), it is necessary to impose additional conditions on the representing functions \(g_i, i = 1, \ldots, n.\)

These additional conditions have been recently established by Pinkus [27]. He proved that, for a large class of representing functions \(g_i,\) the representation is smooth. This means that if we \textit{a priori} assume that the functions \(g_i\) in representation (1.3) belong to a certain class of functions with “good behavior,” then they have the same degree of smoothness as the function \(F.\) As the indicated class of functions with “good behavior,” we can take, e.g., the set of functions continuous at a point, bounded on one side in a set of positive measure, monotonic in an interval, Lebesgue measurable, etc. (see [27]). Konyagin and Kuleshov [17] proved that the functions \(g_i\) in (1.3) inherit the smoothness properties of \(F\) (without additional assumptions imposed on \(g_i\)) if and only if the directions \(a_i\) are linearly independent. Note that the results obtained by Pinkus and also by Konyagin and Kuleshov are valid not only in the bivariate case but also in the multivariate case. There are also some other results obtained for ridge-function representations and involving certain convex subsets of the \(m\)-dimensional space (see [17, 18]).

In the present paper, we study a different aspect of the problem of representations by ridge functions. Assume that \(F \in C^k(\mathbb{R}^2)\) in representation (1.3) but the functions \(g_i\) are arbitrary. This means that we allow the presence of functions with very bad behaviors (e.g., discontinuous at any point). Is it possible to represent \(F\) in the form of a sum \(\sum_{i=1}^{n} f_i(a_i x + b_i y)\) but with \(f_i \in C^k(\mathbb{R}), i = 1, \ldots, n?\) It is easy to see that the answer to this question is positive, as could be expected. For the sake of convenience, we formulate the result over \(\mathbb{R}^2.\) However, in fact, it is true for any open set in \(\mathbb{R}^2.\)

Note that the above problem is not so elementary as it seems to be. Thus, there are cases in which representations with good functions are impossible. These situations happen for closed sets without interior. In [14], Ismailov and Pinkus presented an example of a function of the form

\[
F(x, y) = g_1(a_1 x + b_1 y) + g_2(a_2 x + b_2 y).
\]

This function is bounded and continuous in the union of two straight lines but such that both \(g_1\) and \(g_2\) are necessarily discontinuous and, hence, cannot be replaced with continuous functions \(f_1\) and \(f_2.\)

The result of the present paper can be applied to a higher-order partial differential equation in two variables if its solution can be represented as the sum of sufficiently smooth plane waves [see, e.g., Eq. (1.1)]. According to the theorem presented in what follows, in this case, one can demand solely the smoothness of the sum and dispense with the smoothness of plane-wave terms.

### 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

\[
F(x, y) = \sum_{i=1}^{n} g_i(a_i x + b_i y),
\]

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

\[
F(x, y) = \sum_{i=1}^{n} f_i(a_i x + b_i y).
\]

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 exists a nonsingular linear transformation \(S: (x, y) \to (x', y')\) such that

\[
S: (a_{n-1}, b_{n-1}) \to (1, 0) \quad \text{and} \quad S: (a_n, b_n) \to (0, 1).
\]

Thus, without loss of generality, we can 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 sufficient to show that if a function \(F \in C^k(\mathbb{R}^2)\) can be expressed in the form

\[
F(x, y) = \sum_{i=1}^{n-2} g_i(a_i x + b_i y) + g_{n-1}(x) + g_n(y)
\]

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

\[
F(x, y) = \sum_{i=1}^{n-2} f_i(a_i x + b_i y) + f_{n-1}(x) + f_n(y).
\]

By \(\Delta^{(\delta)} f\) we denote the increment of a function \(f\) in the direction \(l = (l', l'')\), i.e.,

\[
\Delta^{(\delta)} f(x, y) = f(x + l'\delta, y + l''\delta) - f(x, y).
\]

We also use the notation \(\frac{\partial f}{\partial l}\) to denote the derivative of \(f\) in the direction \(l\).

It is easy to see that the increment of a ridge function \(g(\alpha x + \beta y)\) in the direction perpendicular to \((a, b)\) is equal to zero. Let \(l_1, \ldots, l_{n-2}\) be unit vectors perpendicular to the vectors \((a_1, b_1), \ldots, (a_{n-2}, b_{n-2})\), respectively. Then, for any set of numbers \(\delta_1, \ldots, \delta_{n-2} \in \mathbb{R}\), we find

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

We denote the left-hand side of (2.2) by \(S(x, y)\), i.e., we set

\[
S(x, y) \overset{\text{def}}{=} \Delta^{(\delta_1)} \ldots \Delta^{(\delta_{n-2})} F(x, y).
\]

Then it follows from (2.2) that, for any real numbers \(\delta_{n-1}\) and \(\delta_n\),

\[
\Delta^{(\delta_{n-1})} \Delta^{(\delta_n)} S(x, y) = 0,
\]

or, in the expanded form,

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

Setting in the last equality \(\delta_{n-1} = -x\) and \(\delta_n = -y\), we obtain

\[
S(x, y) = S(x, 0) + S(0, y) - S(0, 0).
\]
This means that
\[
\Delta^{(\delta_1)}_{l_1} \ldots \Delta^{(\delta_{n-2})}_{l_{n-2}} F(x, y) = \Delta^{(\delta_1)}_{l_1} \ldots \Delta^{(\delta_{n-2})}_{l_{n-2}} F(x, 0)
\]
\[
+ \Delta^{(\delta_1)}_{l_1} \ldots \Delta^{(\delta_{n-2})}_{l_{n-2}} F(0, y) - \Delta^{(\delta_1)}_{l_1} \ldots \Delta^{(\delta_{n-2})}_{l_{n-2}} F(0, 0).
\]

By the hypothesis of the theorem, the derivative \( \partial^{n-2} F / \partial l_1 \ldots \partial l_{n-2} \) exists at any point \((x, y) \in \mathbb{R}^2\). Thus, it follows from the above formula that
\[
\frac{\partial^{n-2} F}{\partial l_1 \ldots \partial l_{n-2}}(x, y) = h_{1,1}(x) + h_{2,1}(y),
\] (2.3)

where
\[
h_{1,1}(x) = \frac{\partial^{n-2}}{\partial l_1 \ldots \partial l_{n-2}} F(x, 0)
\]
and
\[
h_{2,1}(y) = \frac{\partial^{n-2}}{\partial l_1 \ldots \partial l_{n-2}} F(0, y) - \frac{\partial^{n-2}}{\partial l_1 \ldots \partial l_{n-2}} F(0, 0).
\]

Note that \( h_{1,1} \) and \( h_{2,1} \) belong to the class \( C^{k-n+2}(\mathbb{R}) \).

By \( h_{1,2} \) and \( h_{2,2} \) we 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 \( 1/(e_2 \cdot l_1) \), respectively, i.e.,
\[
h_{1,2}(x) = \frac{1}{e_1 \cdot l_1} \int_0^x h_{1,1}(z) dz,
\]
\[
h_{2,2}(y) = \frac{1}{e_2 \cdot l_1} \int_0^y h_{2,1}(z) dz.
\]

Here, \( e \cdot l \) denotes the scalar product of the vectors \( e \) and \( l \). Obviously, the function
\[
F_1(x, y) = h_{1,2}(x) + h_{2,2}(y)
\]
obeyes the equality
\[
\frac{\partial F_1}{\partial l_1}(x, y) = h_{1,1}(x) + h_{2,1}(y).
\] (2.4)

It follows from (2.3) and (2.4) that
\[
\frac{\partial}{\partial l_1} \left[ \frac{\partial^{n-3} F}{\partial l_2 \ldots \partial l_{n-2}} - F_1 \right] = 0.
\]
Hence, for some ridge function \( \varphi_{1,1}(a_1 x + b_1 y) \), we get

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

(2.5)

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

We introduce the functions

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

\[ h_{2,3}(y) = \frac{1}{e_2 \cdot l_2} \int_0^y h_{2,2}(z)dz, \]

\[ \varphi_{1,2}(t) = \frac{1}{(a_1, b_1) \cdot l_2} \int_0^t \varphi_{1,1}(z)dz. \]

Note that the function

\[ F_2(x, y) = h_{1,3}(x) + h_{2,3}(y) + \varphi_{1,2}(a_1 x + b_1 y) \]

obeys the equality

\[
\frac{\partial F_2}{\partial l_2^2}(x, y) = h_{1,2}(x) + h_{2,2}(y) + \varphi_{1,1}(a_1 x + b_1 y).
\]  

(2.6)

It follows from (2.5) and (2.6) 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 \( \varphi_{2,1}(a_2 x + b_2 y) \), we can write

\[
\frac{\partial^{n-4} F}{\partial l_3 \ldots \partial l_{n-2}}(x, y) = h_{1,3}(x) + h_{2,3}(y) + \varphi_{1,2}(a_1 x + b_1 y) + \varphi_{2,1}(a_2 x + b_2 y).
\]  

(2.7)

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

Note that, on 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 decreases by one in each consecutive step. Continuing the process outlined above 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 that a function \( u \in C^r(\mathbb{R}^2) \) has the form (1.2) with arbitrary behaviors of \( v_i \). Then \( u \) is a solution of equation (1.1).
Note that the method used in the proof of Theorem 2.1 enables us to construct the functions $f_i$, $i = 1, \ldots, n$, by induction. First, we accept some notation. By $(\bar{a}_p, \bar{b}_p)$, $p = 1, \ldots, n - 2$, we denote the images of the vectors $(a, b)$ under the linear transformation, which maps the vectors $(a_{n-1}, b_{n-1})$ and $(a_n, b_n)$ into the unit vectors $e_1 = (1, 0)$ and $e_2 = (0, 1)$, respectively. Clearly,

$$\bar{a}_p = \frac{a_pb_p - a_{n-1}b_{n-1}}{a_{n-1}b_p - a_nb_{n-1}}, \quad \bar{b}_p = \frac{a_{n-1}b_p - a_pb_{n-1}}{a_{n-1}b_n - a_nb_{n-1}}, \quad p = 1, \ldots, n - 2.$$ 

Consider the vectors

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

Note that, for $p = 1, \ldots, n - 2$, the vectors $l_p$ are perpendicular to the vectors $(\bar{a}_p, \bar{b}_p)$, respectively. We also consider the function generated by the above linear transformation

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

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

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

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

Here,

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

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

$$h_{1,k+1}(t) = \frac{1}{e_1 \cdot l_k} \int_0^t h_{1,k}(z) dz, \quad k = 1, \ldots, n - 2,$$

$$h_{2,k+1}(t) = \frac{1}{e_2 \cdot l_k} \int_0^t h_{2,k}(z) dz, \quad k = 1, \ldots, n - 2,$$

and

$$\varphi_{p,1}(t) = \frac{\partial^{n-p-2} F^*}{\partial l_{p+1} \cdots \partial l_{n-2}} \left( \frac{\bar{a}_pt}{\bar{a}_p^2 + \bar{b}_p^2}, \frac{\bar{b}_pt}{\bar{a}_p^2 + \bar{b}_p^2} \right) - h_{1,p+1}\left( \frac{\bar{a}_pt}{\bar{a}_p^2 + \bar{b}_p^2} \right).$$
$$-h_{2,p+1} \left( \frac{\tilde{b}_pt}{\tilde{a}_p^2 + \tilde{b}_p^2} \right) - \sum_{s=1}^{p-1} \varphi_{s,p-s+1} \left( \frac{\tilde{a}_s \tilde{a}_p + \tilde{b}_s \tilde{b}_p t}{\tilde{a}_p^2 + \tilde{b}_p^2 t} \right),$$

for \( p = 1, \ldots, n - 2 \)

$$\varphi_{p,k+1}(t) = \frac{1}{(\tilde{a}_p, \tilde{b}_p) \cdot l_{k+p}} \int_{0}^{t} \varphi_{p,k}(z) dz, \quad p = 1, \ldots, n - 3, \quad k = 1, \ldots, n - p - 2.$$ 

The validity of these formulas for the functions \( h_{1,k} \) and \( h_{2,k}, \quad k = 1, \ldots, n - 1, \) 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 set

$$x = \tilde{a}_p t / (\tilde{a}_p^2 + \tilde{b}_p^2) \quad \text{and} \quad y = \tilde{b}_p t / (\tilde{a}_p^2 + \tilde{b}_p^2).$$

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

$$D_c = \sum_{s=1}^{m} c_s \frac{\partial}{\partial x_s}$$

acts upon \( m \)-variable ridge functions \( g(a \cdot x) \) as follows:

$$D_c g(a \cdot x) = (c \cdot a) g'(a \cdot x).$$

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

$$\prod_{k=1}^{n} D_{l_k} F(x, y) = \prod_{k=1}^{n} D_{l_k} \sum_{i=1}^{n} f_i(a_i x + b_i y)$$

$$= \sum_{i=1}^{n} \left( \prod_{k=1}^{n} ((a_i, b_i) \cdot l_k) \right) f_i^{(n-1)}(a_i x + b_i y)$$

$$= \prod_{k=1}^{n} ((a_r, b_r) \cdot l_k) f_r^{(n-1)}(a_r x + b_r y).$$

Hence, \( f_r \) can be easily constructed from the relation presented above (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 from 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 smoothness results in the multivariate case. In [2], we proved that if a function $f(x_1, \ldots, x_n)$ from a certain smoothness class is represented as the sum of $r$ ridge functions with arbitrary behaviors, then, under suitable conditions, it can be represented as the 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 explained by 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 whether a given function of $n$ variables ($n \geq 3$) is a sum of ridge functions (see [5]). However, parallelling the above theorem, we can also conjecture that if a multivariate function from a certain smoothness class is represented as the sum of ridge functions with arbitrary behaviors, then it can also be represented as the sum of ridge functions from the same smoothness class.

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

$$F(x, y) = \sum_{i=1}^{n} f_i(a_i x + b_i y) + P(x, y),$$

where the functions $f_i \in C^k(\mathbb{R})$ and $P(x, y)$ is a polynomial of total degree that does not exceed $n - 1$. However, it is known that a bivariate polynomial of degree $n - 1$ can be decomposed into a sum of ridge polynomials with any given $n$ pairwise linearly independent directions (see, e.g., [21]). Hence, in (2.8), we have

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

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

Unfortunately, we do not yet know whether the lower accepted 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 in the present paper requires an 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 of degree $\alpha$, $0 < \alpha \leq 1$, if, for any compact set $K \subset \mathbb{R}^m$, there is a number $M = M(K) > 0$ such that, for any $x = (x_1, \ldots, x_m) \in K$ and $y = (y_1, \ldots, y_m) \in K$, the inequality

$$|F(x) - F(y)| \leq M \cdot \sum_{i=1}^{m} |x_i - y_i|^\alpha$$

holds. By $C^{k,\alpha}(\mathbb{R}^m)$ we denote the class of functions in $C^k(\mathbb{R}^m)$ whose $k$th order partial derivatives are locally Hölder continuous with degree $\alpha$.

The following theorem is true:

**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

$$F(x, y) = \sum_{i=1}^{n} g_i(a_i x + b_i y),$$
where $g_i$ are arbitrary univariate functions and $k \geq n - 2$. Then $F$ can be also represented 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 necessary to repeat the proof of Theorem 2.1 emphasizing the fact that the functions appearing on the right-hand sides of relations (2.3), (2.5), (2.7), etc. belong to certain classes $C^{s,\alpha}(\mathbb{R})$ with step-by-step increasing indicators 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 $h_{1,3}$, $h_{2,3}$, $\varphi_{2,1} \in C^{k-n+4,\alpha}(\mathbb{R})$, 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 partial derivatives of $F$ up to the $k$th order can be represented in the form of a sum of ridge functions with 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 the answers to the following two questions:

**Question 1.** Assume that a function $F \in C^{k}(\mathbb{R}^2)$ has the form (2.1). Is it possible to represent the first-order partial derivatives $\partial F/\partial x$ and $\partial F/\partial y$ as sums of ridge functions with arbitrary behaviors and the directions $(a_i, b_i)$?

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

Indeed, the affirmative answer to Question 1 would mean, by induction, that all partial derivatives up to the $k$th order and, hence, any mixed directional derivative $\partial^k F/\partial l_1 \ldots \partial l_k$ can be represented in the form of a sum $\sum_{i=1}^{n} g_i(a_ix + b_iy)$, where $g_i$ are arbitrary univariate functions. If we can give affirmative answer to Question 2, then we immediately conclude that any derivative $\partial^k F/\partial l_1 \ldots \partial l_k$ can be also written in the form $\sum_{i=1}^{n} f_i(a_ix + b_iy)$ with continuous $f_i$. Thus, 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 method proposed above (see the proof of Theorem 2.1), we conclude that $F$ admits a representation $\sum_{i=1}^{n} f_i(a_ix + b_iy)$, where $f_i \in C^{k}(\mathbb{R})$. It is worth noting that Question 2 is a part of a more general question posed in [26, p. 14].

The present research was supported by the Science Development Foundation under 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-London (1966).
2. R. A. Aliev and V. E. Ismailov, “On a smoothness problem in ridge function representation,” *Adv. Appl. Math.*, 73, 154–169 (2016).
3. M. D. Buhmann and A. Pinkus, “Identifying linear combinations of ridge functions,” *Adv. Appl. Math.*, 22, 103–118 (1999).
4. E. J. Candés, “Ridgelets: estimating with ridge functions,” *Ann. Statist.*, 31, 1561–1599 (2003).
5. P. Diaconis and M. Shahshahani, “On nonlinear functions of linear combinations,” *SIAM J. Sci. Statist. Comput.*, 5, 175–191 (1984).
6. J. H. Friedman and W. Stuetzle, “Projection pursuit regression,” *J. Amer. Statist. Assoc.*, 76, 817–823 (1981).
7. N. J. Guliyev and V. E. Ismailov, “On the approximation by single hidden layer feedforward neural networks with fixed weights,” *Neural Netw.*, 98, 296–304 (2018).
8. V. E. Ismailov, “Characterization of an extremal sum of ridge functions,” *J. Comput. 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. Appl.*, 389, 72–83 (2012).
10. V. E. Ismailov, “A review of some results on ridge function approximation,” Azerb. J. Math., 3, No. 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,” Alg. Anal., 28, No. 6, 20–69 (2015); English translation: St. Petersburg Math. J., 28, No. 6, 741–772 (2017).
13. V. E. Ismailov, “A note on the equioscillation theorem for best ridge function approximation,” Expo. Math., 35, No. 3, 343–349 (2017).
14. V. E. Ismailov and 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, Interscience Publ., New-York–London (1955).
16. I. Kazantsev, “Tomographic reconstruction from arbitrary directions using ridge functions,” Inverse Probl., 14, 635–645 (1998).
17. S. V. Konyagin and A. A. Kuleshov, “On the continuity of finite sums of ridge functions,” Mat. Zametki, 98, 308–309 (2015); English translation: Math. Notes, 98, No. 1-2, 336–338 (2015).
18. S. V. Konyagin and A. A. Kuleshov, “On some properties of finite sums of ridge functions defined on convex subsets of $\mathbb{R}^n$,” Tr. Mat. Inst. Steklova, 293 (2016); English translation: Proc. Steklov Inst. Math., 293, No. 1, 186–193 (2016).
19. S. V. Konyagin, A. A. Kuleshov, and V. E. Maiorov, “Some problems in the theory of ridge functions,” Tr. Mat. Inst. Steklova, 301 (2018); English translation: Proc. Steklov Inst. Math., 301, No. 1, 144–169 (2018).
20. A. Kroó, “On approximation by ridge functions,” Constr. Approx., 13, 447–460 (1997).
21. B. F. Logan and 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 and A. Pinkus, “Lower bounds for approximation by MLP neural networks,” Neurocomputing, 25, 81–91 (1999).
24. F. Natterer, The Mathematics of Computerized Tomography, Teubner, Stuttgart; Wiley & Sons, Chichester (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 University Press, Cambridge (2015).
27. A. Pinkus, “Smoothness and uniqueness in ridge function representation,” Indag. Math. (N.S.), 24, No. 4, 725–738 (2013).
28. A. Pinkus, “Approximation theory of the MLP model in neural networks,” Acta Numer., 8, 143–195 (1999).