Proceedings of the Institute of Mathematics and Mechanics,
National Academy of Sciences of Azerbaijan
Volume 41, Number 2, 2015, Pages 106–118

ON THE REPRESENTATION BY SUMS OF RIDGE FUNCTIONS

RASHID A. ALIEV, VUGAR E. ISMAILOV, AND TARANA M. SHAHBALAYEVA

Abstract. In the current paper, we review some results on the representation by sums of ridge functions with finitely many directions. In the special case of three directions, we prove that if a function of a certain smoothness class is represented by sums of arbitrarily behaved ridge functions, then it can also be represented by sums of ridge functions of the same smoothness class.

1. Introduction

A *ridge function,* in its simplest format, is a multivariate function of the form

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

where $g:\mathbb{R}\rightarrow\mathbb{R}$ and $\mathbf{a}=(a_1,...,a_n)$ is a fixed vector (direction) in $\mathbb{R}^n\backslash\left\{\mathbf{0}\right\}$. These functions and their linear combinations appear in various areas. They appear in the theory of PDE's (where they are called *plane waves*, see, e.g., [17]), in mathematical problems of computerized tomography (see, e.g., [18, 19, 24, 26, 27]), in the theory of projection pursuit and projection regression (see, e.g., [5, 6, 8, 9, 10]), and in neural networks (see [31] and many related references therein). Ridge functions are also extensively used in modern approximation theory as an effective tool for approximating complicated multivariate functions (see, e.g., [11, 12, 13, 14, 15, 16, 22, 23, 25, 28, 30]).

In the current paper, we consider the problem of representation of multivariate functions by sums of ridge functions with finitely many fixed directions. Assume we are given r pairwise linearly independent directions $\mathbf{a}^i$, $i=1,...,r$, in $\mathbb{R}^n\backslash\left\{\mathbf{0}\right\}$. The first problem arising here is about the representability of a given multivariate function $f:\mathbb{R}^n\rightarrow\mathbb{R}$ as a linear combination of ridge functions with the directions $\mathbf{a}^i$, $i=1,...,r$. That is, we want to know when 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 the dimension $n=2$ and the given function $f(x,y)$ has partial derivatives up to r-th order. For the representation of $f(x,y)$

2010 *Mathematics Subject Classification.* 26B40, 39B22.

Key words and phrases. ridge function; closed path.

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 bivariate functions provided that the derivatives are understood in the generalized sense. It should be remarked that this simple assertion does not carry over to the general case when the space dimension $n > 2$. In this case, there known several results, which will be analyzed in Section 2.

Let a function $f(\mathbf{x})$ can be represented in the form (1.1). Let, besides, f be of a certain smoothness class. What can we say about the smoothness of g_i? The case $r = 1$ is clear. In this case, if $f \in C^k(\mathbb{R}^n)$, then for $\mathbf{c} \in \mathbb{R}^n$ 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 $\mathbf{c} \in \mathbb{R}^n$ 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 $C^k(\mathbb{R})$. By the same way, one can verify that $g_2 \in C^k(\mathbb{R})$. Note that the picture is completely different if $r \geq 3$. For $r = 3$, it was shown in [29] that, there are smooth functions which decompose into sums of not only nonsmooth but very badly behaved ridge functions.

It was first proved by Buhman and Pinkus [4] that if in (1.1) $f \in C^k(\mathbb{R}^n)$, $k \geq r - 1$ and $g_i \in L^1_{loc}(\mathbb{R})$ for each i, then $g_i \in C^k(\mathbb{R})$ for $i = 1, ..., r$. Pinkus [29] generalized this result and showed that the solution is quite simple and natural if the functions g_i are taken from a certain class of "reasonably well behaved functions" As the mentioned class of "reasonably well behaved functions" one may take, e.g., the set of functions that are continuous at a point, the set of Lebesgue measurable functions, etc. In [29], such classes are denoted by $\mathcal{B}$ (for the rigorous definition see [29]). The result of Pinkus states that if in (1.1) $f \in C^k(\mathbb{R}^n)$ and each $g_i \in \mathcal{B}$, then necessarily $g_i \in C^k(\mathbb{R})$ for $i = 1, ..., r$. This result gives rise to a natural and, in our opinion, important question. Assume in the representation (1.1) $f \in C^k(\mathbb{R}^n)$ but the functions g_i are badly behaved. Can we write f as a sum $\sum_{i=1}^{r} f_i(\mathbf{a}^i \cdot \mathbf{x})$ but with the smooth f_i, $i = 1, ..., r$? For the case $r = 3$, this interesting question will be discussed in Section 3.

This paper is organized as follows. It consists of this Introduction and two more sections. In Section 2, we give a brief historical review of the known results on the representation by ridge functions. Finally, in Section 3, we discuss the possibility of smooth representation with the proviso that nonsmooth representation is possible. More precisely, for $r = 3$, we prove that if representation (1.1) holds for $f \in C^k(\mathbb{R}^n)$, then the functions g_i can be replaced by some functions from the class $C^k(\mathbb{R})$.

2. A brief overview of the problem of representation by ridge functions

The problem of representation by ridge functions appeared in connection with the understanding of mathematics of computerized tomography. In tomography, the reconstruction of a given multivariate function from values of its integrals along certain lines in the plane is essential. The integrals along parallel lines can be considered as a ridge function. Thus, the problem is to reconstruct f from some set of ridge functions generated by the function f itself. In practice, one can consider only a finite number of directions along which the above integrals are taken. Obviously, reconstruction from such data needs some additional conditions to be unique, since there are many functions g having the same integrals. For uniqueness, Logan and Shepp in their pioneering paper [24] used the criterion of minimizing the L_2 norm of g. That is, they found a function $g(x,y)$ with the minimum L_2 norm among all functions, which has the same integrals as f. 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 : $P_g(t,\theta_j) = P_f(t,\theta_j)$, $j = 0,1,...,n-1$, where angles θ_j generate equally spaced directions, i.e. $\theta_j = \frac{j\pi}{n}$, $j = 0,1,...,n-1$. The authors showed that this problem of tomography is equivalent to the problem of L_2-approximation of a given function f by sums of ridge functions with equally spaced directions $(\cos\theta_j, \sin\theta_j)$, $j = 0,1,...,n-1$. 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 ridge functions.

Kazantsev [18] solved the above problem of tomography without requiring that the considered directions be equally spaced. Marr [26] considered the problem of finding a polynomial of degree $n-2$ whose projections along lines joining each pair of n equally spaced points on the circumference of D best matches the given projections of f in the sense of minimizing the sum of squares of the differences. Thus we see that the problems of tomography give rise to an independent study of approximation-theoretic properties of the following set of ridge functions:

$$\mathcal{R}\left(\mathbf{a}^1,...,\mathbf{a}^r\right) = \left\{ \sum_{i=1}^{r} g_i\left(\mathbf{a}^i\cdot\mathbf{x}\right) : \mathbf{x}\in\mathbb{R}^n,\ g_i:\mathbb{R}\to\mathbb{R},\ i=1,...,r \right\},$$

where directions $\mathbf{a}^1,...,\mathbf{a}^r$ are fixed and belong to $n-$dimensional Euclidean space. Note that the set $\mathcal{R}\left(\mathbf{a}^1,...,\mathbf{a}^r\right)$ is a linear space.

Besides $\mathcal{R}\left(\mathbf{a}^1,...,\mathbf{a}^r\right)$, we will also use the notation $\mathcal{R}\left(\mathbf{a}^1,...,\mathbf{a}^r;X\right)$ which stands for the set of functions from $\mathcal{R}\left(\mathbf{a}^1,...,\mathbf{a}^r\right)$ but restricted to a set $X\subset\mathbb{R}^n$.

One of the basic problems concerning sums of ridge functions with fixed directions is the problem of verifying if a given function f belongs to the space $\mathcal{R}\left(\mathbf{a}^1,...,\mathbf{a}^r\right)$. In introduction, we have seen that this problem has a simple solution if the space dimension $n=2$ and the given function $f(x,y)$ has partial derivatives up to r-th order. But if $n\geq 3$, the picture drastically changes. Note that in the left hand side of Eq. (1.2), the differential operator

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

involves vectors $(b_i, -a_i)$, which are perpendicular to the directions (a_i, b_i), $i = 1, ..., r$. For the case $n \geq 3$, there are many vectors perpendicular to $\mathbf{a}^i$, for each $i = 1, ..., r$. Therefore, the resulting operator should involve all these vectors. The corresponding theorem, highlighting this idea, belongs to Diaconis and Shahshahani [7].

Proposition 2.1 (Diaconis, Shahshahani [7]). *Let $\mathbf{a}^1, ..., \mathbf{a}^r$ be pairwise independent vectors in $\mathbb{R}^n$. Let for $i = 1, 2, ..., r$, H^i denote the hyperplane $\{\mathbf{c} \in \mathbb{R}^n\colon \mathbf{c} \cdot \mathbf{a}^i = 0\}$. Then a function $f \in C^r(\mathbb{R}^n)$ can be represented as*

$$f(\mathbf{x}) = \sum_{i=1}^{r} g_i \left(\mathbf{a}^i \cdot \mathbf{x} \right) + P(\mathbf{x}),$$

where $P(\mathbf{x})$ is a polynomial of degree not more than r, if and only if

$$\prod_{i=1}^{r} \sum_{s=1}^{n} c_s^i \frac{\partial f}{\partial x_s} = 0,$$

for all vectors $\mathbf{c}^i = (c_1^i, c_2^i, ..., c_n^i) \in H^i$, $i = 1, 2, ..., r$.

The main drawback of this proposition is the "unwanted term" $P(\mathbf{x})$ in the representation formula. There are examples (see [7]) showing that one cannot simply dispense with the polynomial $P(\mathbf{x})$ in Proposition 2.1.

Lin and Pinkus [23] obtained more general result on the representation by ridge functions. We need some notation to present their result. First note that each polynomial $p(x_1, ..., x_n)$ generates the differential operator $p(\frac{\partial}{\partial x_1}, ..., \frac{\partial}{\partial x_n})$. Let $P(\mathbf{a}^1, ..., \mathbf{a}^r)$ denote the set of polynomials which vanish on all the lines $\{\lambda \mathbf{a}^i, \lambda \in \mathbb{R}\}$, $i = 1, ..., r$. Obviously, this is an ideal in the ring of all polynomials. Let Q be the set of polynomials $q = q(x_1, ..., x_n)$ such that $p(\frac{\partial}{\partial x_1}, ..., \frac{\partial}{\partial x_n})q = 0$, for all $p(x_1, ..., x_n) \in P(\mathbf{a}^1, ..., \mathbf{a}^r)$.

Proposition 2.2 (Lin, Pinkus [23]). *Let $\mathbf{a}^1, ..., \mathbf{a}^r$ be pairwise linearly independent vectors in $\mathbb{R}^n$. A function $f \in C(\mathbb{R}^n)$ can be expressed in the form*

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

if and only if f belongs to the closure of the linear span of Q.

What can we say about the representation of arbitrary (not necessarily continuous or differentiable) multivariate function $f : \mathbb{R}^n \to \mathbb{R}$ by sums of ridge functions with the given directions $\mathbf{a}^1, ..., \mathbf{a}^r$. This problem was considered in [14]. The corresponding solution uses such objects as "paths" and "path functionals".

Definition 2.1. *A set of points* $l = \{\mathbf{x}^1, \ldots, \mathbf{x}^m\} \subset \mathbb{R}^n$ *is called a closed path with respect to the directions* $\mathbf{a}^1, \mathbf{a}^2, ..., \mathbf{a}^r$ *if there exists a vector* $\lambda = (\lambda_1, \ldots, \lambda_m) \in \mathbb{Z}^m \setminus \{\mathbf{0}\}$ *such that*

$$\sum_{j=1}^{m} \lambda_j \delta_{\mathbf{a}^i \cdot \mathbf{x}^j} = 0, \quad \textit{for all } i = 1, \ldots, r. \tag{2.1}$$

Here δ_a is the characteristic function of a set $\{a\}$ and $\mathbf{a}^i \cdot \mathbf{x}^j$ is the inner product of these two vectors.

The idea of closed paths with respect to r directions in $\mathbb{R}^n$ was first considered in the paper by Braess and Pinkus [3]. Klopotowski, Nadkarni, Rao [21] defined these objects with respect to canonical projections. In our paper [14], which deals with linear superpositions and the Kolmogorov superposition theorem, closed paths have been generalized to those having association with r arbitrary functions. In these three works, it was shown that nonexistence of closed paths of the respective form is both necessary and sufficient for

1) interpolation by ridge functions [3];
2) representation of multivariate functions by sums of univariate functions [21];
3) representation by linear superpositions [14].

Closed paths are also appeared in duality relations in approximation by sums of univariate functions. They are necessary for description of extreme points of the set of measures orthogonal to such sums (see [12]).

Let for $i = 1, ..., r$, the set $\{\mathbf{a}^i \cdot \mathbf{x}^j,\ j = 1, ..., m\}$ have s_i different values. Then it is not difficult to see that Eq. (2.1) stands for a system of $\sum_{i=1}^{r} s_i$ homogeneous linear equations in unknowns $\lambda_1, ..., \lambda_m$. If this system has any solution with nonzero integer components, then the given set $\{\mathbf{x}^1, \ldots, \mathbf{x}^m\}$ is a closed path.

For example, the set $l = \{(0,0,0),\ (0,0,1),\ (0,1,0),\ (1,0,0),\ (1,1,1)\}$ is a closed path in $\mathbb{R}^3$ with respect to the basic directions. The vector λ in Definition 2.1 can be taken as $(-2,1,1,1,-1)$.

In the case $r = 2$, the picture of closed paths becomes more clear. Let, for example, $\mathbf{a}^1$ and $\mathbf{a}^2$ be the basic directions in $\mathbb{R}^2$. In this case, a closed path is the union of some sets A with the property: each A consists of vertices of a closed broken line with the sides parallel to the coordinate axis. These objects (sets A) have been exploited in practically all works devoted to the approximation of bivariate functions by univariate functions, although under the different names (see, for example, [20, Chapter 2]). If X and the directions $\mathbf{a}^1$ and $\mathbf{a}^2$ are arbitrary, the sets A can be described as a trace of some point traveling alternatively in two hyperplanes perpendicular to these directions and then returning to its primary position. It should be remarked that in the case $r > 2$, closed paths do not admit such a simple geometric description.

Let X be a subset of $\mathbb{R}^n$ and $T(X)$ denote the set of all functions on X. With each pair $\langle l, \lambda \rangle$, where $l = \{\mathbf{x}^1, \ldots, \mathbf{x}^m\}$ is a closed path in X and $\lambda = (\lambda_1, ..., \lambda_m)$ is a vector known from Definition 2.1, we associate the functional

$$G_{l,\lambda} : T(X) \to \mathbb{R},\ G_{l,\lambda}(f) = \sum_{j=1}^{m} \lambda_j f(\mathbf{x}^j).$$

It is clear that the functional $G_{l,\lambda}$ is linear and $G_{l,\lambda}(g) = 0$ for all functions $g \in \mathcal{R}\left(\mathbf{a}^1, ..., \mathbf{a}^r; X\right)$.

Definition 2.2. *A closed path $l = \{\mathbf{x}^1, \ldots, \mathbf{x}^m\}$ is said to be minimal if l does not contain any closed path as its proper subset.*

For example, the set $l = \{(0,0,0),\ (0,0,1),\ (0,1,0),\ (1,0,0),\ (1,1,1)\}$ considered above is a minimal closed path with respect to the coordinate directions in $\mathbb{R}^3$. Adding the point $(0,1,1)$ to l, we will have a closed path, but not minimal. The vector λ associated with $l \cup \{(0,1,1)\}$ can be taken as $(3,-1,-1,-2,2,-1)$.

It is not difficult to verify that a minimal closed path l uniquely defines the functional

$$G_l(f) = \sum_{j=1}^{m} \lambda_j f(\mathbf{x}^j), \quad \sum_{j=1}^{m} |\lambda_j| = 1.$$

The following proposition is valid.

Proposition 2.3 (see [14]). *Let $X \subset \mathbb{R}^n$ and $\mathbf{a}^1, ..., \mathbf{a}^r$ be arbitrarily fixed directions in $\mathbb{R}^n$.*

1) Let X have closed paths with respect to the directions $\mathbf{a}^1, ..., \mathbf{a}^r$. A function $f : X \to \mathbb{R}$ belongs to the space $\mathcal{R}\left(\mathbf{a}^1, ..., \mathbf{a}^r; X\right)$ if and only if $G_l(f) = 0$ for any minimal closed path $l \subset X$.

2) Let X have no closed paths. Then $\mathcal{R}\left(\mathbf{a}^1, ..., \mathbf{a}^r; X\right) = T(X)$.

The above problem of representation of a multivariate function by ridge functions gives rise to the problem of representation of some classes of functions by such sums. For example, one may consider the following problem. Let X be a subset of the $n-$dimensional Euclidean space. Let $C(X)$, $B(X)$, $T(X)$ denote the set of continuous, bounded and all real functions defined on X correspondingly. In the first case, we additionally suppose that X is a compact set. Let $\mathcal{R}_c\left(\mathbf{a}^1, ..., \mathbf{a}^r; X\right)$ and $\mathcal{R}_b\left(\mathbf{a}^1, ..., \mathbf{a}^r; X\right)$ denote the subspaces of $\mathcal{R}\left(\mathbf{a}^1, ..., \mathbf{a}^r; X\right)$ comprising only sums of continuous and bounded terms $g_i\left(\mathbf{a}^i \cdot \mathbf{x}\right)$, $i = 1, ..., r$, correspondingly. The following questions naturally arise: For which sets X, one can claim that $\mathcal{R}_c\left(\mathbf{a}^1, ..., \mathbf{a}^r; X\right) = C(X)$, $\mathcal{R}_b\left(\mathbf{a}^1, ..., \mathbf{a}^r; X\right) = B(X)$, and $\mathcal{R}\left(\mathbf{a}^1, ..., \mathbf{a}^r; X\right) = T(X)$? The first two problems in more general setting were solved by Sternfeld [32, 33]. The third problem was solved in [14]. Let us consider the corresponding results. Assume we are given directions $\mathbf{a}^1, ..., \mathbf{a}^r \in \mathbb{R}^n \backslash \{\mathbf{0}\}$ and some set $X \subseteq \mathbb{R}^n$. The family $F = \{\mathbf{a}^1, ..., \mathbf{a}^r\}$ *uniformly separates points* of X if there exists a number $0 < \lambda \leq 1$ such that for each pair $\{\mathbf{x}_j\}_{j=1}^m$, $\{\mathbf{z}_j\}_{j=1}^m$ of disjoint finite sequences in X, there exists some direction $\mathbf{a}^k \in F$ so that if from the two sequences $\{\mathbf{a}^k \cdot \mathbf{x}_j\}_{j=1}^m$and $\{\mathbf{a}^k \cdot \mathbf{z}_j\}_{j=1}^m$ we remove a maximal number of pairs of points $\mathbf{a}^k \cdot \mathbf{x}_{j_1}$ and $\mathbf{a}^k \cdot \mathbf{z}_{j_2}$ with $\mathbf{a}^k \cdot \mathbf{x}_{j_1} = \mathbf{a}^k \cdot \mathbf{z}_{j_2}$, then there remains at least λm points in each sequence (or , equivalently, at most $(1-\lambda)m$ pairs can be removed). Sternfeld [33], in particular, proved that a finite family of directions $F = \{\mathbf{a}^1, ..., \mathbf{a}^r\}$ uniformly separates points of X if and only if

$\mathcal{R}_b\left(\mathbf{a}^1,...,\mathbf{a}^r;X\right) = B(X)$. In [33], Sternfeld also obtained a practically convenient sufficient condition for the equality $\mathcal{R}_b\left(\mathbf{a}^1,...,\mathbf{a}^r;X\right) = B(X)$.To describe this condition, define the set functions

$$\tau_i(Z) = \{\mathbf{x} \in Z : \ |p_i^{-1}(p_i(\mathbf{x})) \bigcap Z| \geq 2\},$$

where $Z \subset X$, $p_i(\mathbf{x}) = \mathbf{a}^i \cdot \mathbf{x}$, $i = 1, \ldots, r$, and $|Y|$ denotes the cardinality of a considered set Y. Define $\tau(Z)$ to be $\bigcap_{i=1}^k \tau_i(Z)$ and define $\tau^2(Z) = \tau(\tau(Z))$, $\tau^3(Z) = \tau(\tau^2(Z))$ and so on inductively.

Proposition 2.4 (Sternfeld [33]). *If $\tau^k(X) = \emptyset$ for some k, then $\mathcal{R}_b\left(\mathbf{a}^1,...,\mathbf{a}^r;X\right) = B(X)$. If X is a compact subset of $\mathbb{R}^n$, and $\tau^k(X) = \emptyset$ for some k, then $\mathcal{R}_c\left(\mathbf{a}^1,...,\mathbf{a}^r;X\right) = C(X)$.*

The sufficient condition "$\tau^n(X) = \emptyset$ for some n" turns out to be also necessary for the case $r = 2$. In this case the equality $\mathcal{R}_b\left(\mathbf{a}^1,\mathbf{a}^2;X\right) = B(X)$ is equivalent to the equality $\mathcal{R}_c\left(\mathbf{a}^1,\mathbf{a}^2;X\right) = C(X)$. In another work [32], Sternfeld obtained a measure-theoretic necessary and sufficient condition for the equality $\mathcal{R}_c\left(\mathbf{a}^1,...,\mathbf{a}^r;X\right) = C(X)$. Let $\mathbf{a}^1,...,\mathbf{a}^r \in \mathbb{R}^n\backslash\{\mathbf{0}\}$, $p_i(\mathbf{x}) = \mathbf{a}^i \cdot \mathbf{x}$, $i = 1, \ldots, r$, X be a compact set in $\mathbb{R}^n$ and $M(X)$ be a class of measures defined on some field of subsets of X. The family $F = \{\mathbf{a}^1,...,\mathbf{a}^r\}$ *uniformly separates measures* of the class $M(X)$ if there exists a number $0 < \lambda \leq 1$ such that for each measure μ in $M(X)$ the equality $\left\|\mu \circ p_k^{-1}\right\| \geq \lambda \left\|\mu\right\|$ holds for some direction $\mathbf{a}^k \in F$. Sternfeld [32], in particular, proved that the equality$\mathcal{R}_c\left(\mathbf{a}^1,...,\mathbf{a}^r;X\right) = C(X)$ holds if and only if the family of directions $\{\mathbf{a}^1,...,\mathbf{a}^r\}$ uniformly separates measures of the class $C(X)^*$ (that is, the class of regular Borel measures). Besides, he proved that $\mathcal{R}_b\left(\mathbf{a}^1,...,\mathbf{a}^r;X\right) = B(X)$ if and only if the family of directions $\{\mathbf{a}^1,...,\mathbf{a}^r\}$ uniformly separates measures of the class $l_1(X)$ (that is, the class of finite measures defined on countable subsets of X). Since $l_1(X) \subset C(X)^*$, the first equality $\mathcal{R}_c\left(\mathbf{a}^1,...,\mathbf{a}^r;X\right) = C(X)$ implies the second equality $\mathcal{R}_b\left(\mathbf{a}^1,...,\mathbf{a}^r;X\right) = B(X)$. The inverse is not true (see [32]).

For the problem of representation $\mathcal{R}\left(\mathbf{a}^1,...,\mathbf{a}^r;X\right) = T(X)$, Ismailov [14] obtained the following necessary and sufficient condition in terms of closed paths.

Proposition 2.5 (see [14]). *$\mathcal{R}\left(\mathbf{a}^1,...,\mathbf{a}^r;X\right) = T(X)$ if and only if X has no closed paths with respect to the directions $\mathbf{a}^1,...,\mathbf{a}^r$.*

It should be remarked that the above results of Sternfeld and Ismailov were obtained for more general functions, than linear combinations of ridge functions, namely for functions of the form $\sum_{i=1}^r g_i(h_i(x))$, where h_i are arbitrarily fixed functions defined on X.

3. Smoothness in representation by three ridge functions

In this section we consider and solve partially the following problem. Let a function $f \in C^k(\mathbb{R}^n)$ can be represented in the form $\sum_{i=1}^r g_i(\mathbf{a}^i \cdot \mathbf{x})$ and we know nothing about the behavior of the functions g_i. Can we represent f in the form $\sum_{i=1}^r f_i(\mathbf{a}^i \cdot \mathbf{x})$ but with sufficiently smooth functions f_i? For the case $r \leq 2$, as we have already seen in Introduction, such representation is possible with $f_i \in C^k(\mathbb{R})$, $i = 1, 2$. That is, we can chose representing ridge functions $f_i(\mathbf{a}^i \cdot \mathbf{x})$ from the class of smoothness, to which the represented function f belongs. This is an ideal situation. For $r \geq 3$, the problem is quite difficult to have such a simple solution. In this section, we prove that if $r = 3$ and the representation (1.1) holds for $f \in C^k(\mathbb{R}^n)$, $k \geq 2$, then the functions g_i can be replaced by some functions f_i from the class $C^k(\mathbb{R})$. In the case $r = 3$ and $k = 1$, this statement is valid under a mild assumption that the first order partial derivatives of f are Hölder continuous.

Theorem 3.1. *Assume $\mathbf{a}^i$, $i = 1, 2, 3$, are pairwise linearly independent directions in $\mathbb{R}^n \backslash \{\mathbf{0}\}$. Assume that a function $f \in C^k(\mathbb{R}^n)$ is of the form*

$$f(\mathbf{x}) = \sum_{i=1}^{3} g_i(\mathbf{a}^i \cdot \mathbf{x}). \tag{3.1}$$

1) If $k \geq 2$, then f can be represented also in the form

$$f(\mathbf{x}) = \sum_{i=1}^{3} f_i(\mathbf{a}^i \cdot \mathbf{x}),$$

where the functions $f_i \in C^k(\mathbb{R})$, $i = 1, 2, 3$.

2) If $k = 1$, then the above statement holds with the proviso that $f \in C^{1,\alpha}(\Omega)$ for any bounded subset Ω of $\mathbb{R}^n$ (α is not fixed and depends on Ω).

Remark. Theorem 3.1 in a more general form, involving any number of $r \geq 3$ directions, will appear in [2].

To prove the above result we need the following lemmas, which can be proven by using two basic theorems of calculus (namely, the mean value theorem and the term by term differentiation theorem).

Lemma 3.1. *Let $h \in C^1(\mathbb{R})$, $h(0) = h'(0) = 0$ and h' is Hölder continuous on any finite interval $[a, b]$. Then the function*

$$H(t) = \sum_{k=1}^{\infty} 2^{k-1} h(\frac{t}{2^k}) \tag{3.2}$$

is well defined and continuously differentiable on the real axis.

Lemma 3.2. *Let $h \in C^p(\mathbb{R})$, $p \geq 2$ and $h(0) = h'(0) = 0$. Then the function $H(t)$ (defined by (3.2)) is in the class $C^p(\mathbb{R})$.*

Proof of Theorem 3.1. We start with the second part of the theorem. That is, let $f \in C^{1,\alpha}(\Omega)$ for any bounded subset Ω of $\mathbb{R}^n$ (here α is different for different sets Ω) and the formula (3.1) holds.

First assume that the vectors $\mathbf{a}^1, \mathbf{a}^2, \mathbf{a}^3$ form a linearly independent system. Complete this system to a linearly independent system consisting of n vectors $\mathbf{a}^1, \mathbf{a}^2, ..., \mathbf{a}^n$. Consider the linear transformation $y_i = \mathbf{a}^i \cdot \mathbf{x}$, $i = 1, ..., n$. Let A

be matrix of this transformation. Using this transformation, we can write the formula (3.1) in the form

$$f(A^{-1}\mathbf{y}) = \sum_{i=1}^{3} g_i(y_i). \tag{3.3}$$

Here, $\mathbf{y} = (y_1, ..., y_n)^T$. In (3.3), taking sequentially $y_2 = y_3 = 0$, $y_1 = y_3 = 0$ and $y_1 = y_2 = 0$, we obtain correspondingly that $g_1 \in C^1(\mathbb{R})$, $g_2 \in C^1(\mathbb{R})$ and $g_3 \in C^1(\mathbb{R})$.

Now assume that the vectors $\mathbf{a}^1, \mathbf{a}^2, \mathbf{a}^3$ are linearly dependent. Since these vectors are pairwise linearly independent, there exist numbers $\lambda_1, \lambda_2 \in \mathbb{R}\backslash\{0\}$ such that $\mathbf{a}^3 = \lambda_1\mathbf{a}^1 + \lambda_2\mathbf{a}^2$. Complete the system $\{\mathbf{a}^1, \mathbf{a}^2\}$ to a linearly independent system $\{\mathbf{a}^1, \mathbf{a}^2, \mathbf{b}^3, ..., \mathbf{b}^n\}$ and consider the linear transformation $\mathbf{y} = B\mathbf{x}$, where $x = (x_1, ..., x_n)^T$, $y = (y_1, ..., y_n)^T$ and B is the matrix, rows of which are formed by the coordinates of the vectors $\mathbf{a}^1, \mathbf{a}^2, \mathbf{b}^3, ..., \mathbf{b}^n$. Using this transformation, we can write (3.1) in the form

$$f(A^{-1}\mathbf{y}) = g_1(y_1) + g_2(y_2) + g_3(\lambda_1 y_1 + \lambda_2 y_2). \tag{3.4}$$

In (3.4), taking sequentially $y_2 = 0$ and $y_1 = 0$, we obtain that

$$g_1(y_1) = f(A^{-1}\mathbf{y})|_{y_2=0} - g_2(0) - g_3(\lambda_1 y_1), \tag{3.5}$$

and

$$g_2(y_2) = f(A^{-1}\mathbf{y})|_{y_1=0} - g_1(0) - g_3(\lambda_2 y_2). \tag{3.6}$$

Considering (3.5) and (3.6) in (3.4), we obtain the equality

$$\begin{aligned} g_3(\lambda_1 y_1 + \lambda_2 y_2) - g_3(\lambda_1 y_1) - g_3(\lambda_2 y_2) = \\ = f(A^{-1}\mathbf{y}) - f(A^{-1}\mathbf{y})|_{y_1=0} - f(A^{-1}\mathbf{y})|_{y_2=0} + g_1(0) + g_2(0). \end{aligned} \tag{3.7}$$

One can easily observe that the right hand side of (3.7) depends only on the variables y_1 and y_2. Denote the right hand side of (3.7) by $F(y_1, y_2)$. That is, set the following function

$$F(y_1, y_2) \stackrel{def}{=} f(A^{-1}\mathbf{y}) - f(A^{-1}\mathbf{y})|_{y_1=0} - f(A^{-1}\mathbf{y})|_{y_2=0} + g_1(0) + g_2(0). \tag{3.8}$$

Along with (3.8) we repeatedly use the following identity, which follows from (3.7) and (3.8)

$$F(y_1, y_2) = g_3(\lambda_1 y_1 + \lambda_2 y_2) - g_3(\lambda_1 y_1) - g_3(\lambda_2 y_2). \tag{3.9}$$

By the hypothesis of the theorem, the partial derivatives of $F(y_1, y_2)$ is Hölder continuous over any bounded set in $\mathbb{R}^2$. Consider the function

$$h(t) = F(\frac{t}{\lambda_1}, \frac{t}{\lambda_2}) - F(0, 0). \tag{3.10}$$

Let us check that the function $h(t)$ satisfies all the conditions of Lemma 3.1. First note that h' is Hölder continuous on any finite interval $[a, b]$. Besides,

$$h(0) = F(0, 0) - F(0, 0) = 0.$$

To prove $h'(0) = 0$, note that by (3.9),

$$\begin{aligned} F(\Delta t,0)-F(0,0) &= g_3(\lambda_1\Delta t)-g_3(\lambda_1\Delta t)-g_3(0)+g_3(0)=0,\\ F(0,\Delta t)-F(0,0) &= g_3(\lambda_2\Delta t)-g_3(\lambda_2\Delta t)-g_3(0)+g_3(0)=0. \end{aligned}$$

From the above equalities, we obtain that

$$\frac{\partial F}{\partial y_1}(0,0)=\frac{\partial F}{\partial y_2}(0,0)=0$$

and hence

$$h^{'}(0)=\frac{1}{\lambda_1}\frac{\partial F}{\partial y_1}(0,0)+\frac{1}{\lambda_2}\frac{\partial F}{\partial y_2}(0,0)=0.$$

Now by Lemma 3.1, the $H(t)$ defined in (3.2) is continuously differentiable on the real axis.

Let us now prove that

$$F(y_1,y_2)-F(0,0)=H(\lambda_1 y_1+\lambda_2 y_2)-H(\lambda_1 y_1)-H(\lambda_2 y_2). \tag{3.11}$$

Define the following functions

$$H_n(t)=\sum_{k=1}^{n}2^{k-1}h(\frac{t}{2^k}),\ n=1,2,... \tag{3.12}$$

It can be shown that the sequence $\{H_n(t)\}_{n=1}^{\infty}$ uniformly converges to the function $H(t)$ on any finite interval $[-T,T]$. Consider the difference

$$\Delta_n=[F(y_1,y_2)-F(0,0)]-[H_n(\lambda_1 y_1+\lambda_2 y_2)-H_n(\lambda_1 y_1)-H_n(\lambda_2 y_2)]. \tag{3.13}$$

Considering (3.12) in (3.13) and then using (3.10), we can write the following equalities

$$\begin{aligned} \Delta_n=F(y_1,y_2)-F(0,0)-\textstyle\sum_{k=1}^{n}2^{k-1}\left[h\left(\frac{\lambda_1 y_1+\lambda_2 y_2}{2^k}\right)-h\left(\frac{\lambda_1 y_1}{2^k}\right)-h\left(\frac{\lambda_2 y_2}{2^k}\right)\right]=\\ =F(y_1,y_2)-F(0,0)-\textstyle\sum_{k=1}^{n}2^{k-1}S_k, \end{aligned} \tag{3.14}$$

where

$$S_k=F\left(\frac{\lambda_1 y_1+\lambda_2 y_2}{2^k\lambda_1},\frac{\lambda_1 y_1+\lambda_2 y_2}{2^k\lambda_2}\right)-F\left(\frac{\lambda_1 y_1}{2^k\lambda_1},\frac{\lambda_1 y_1}{2^k\lambda_2}\right)-F\left(\frac{\lambda_2 y_2}{2^k\lambda_1},\frac{\lambda_2 y_2}{2^k\lambda_2}\right)+F(0,0),$$

for the indices $k=1,...,n$. Considering (3.9) in (3.14) we obtain that

$$\Delta_n=g_3(\lambda_1 y_1+\lambda_2 y_2)-g_3(\lambda_1 y_1)-g_3(\lambda_2 y_2)-F(0,0)-\sum_{k=1}^{n}2^{k-1}B_k,$$

where

$$\begin{aligned} B_k=g_3\left(\tfrac{\lambda_1 y_1+\lambda_2 y_2}{2^{k-1}}\right)-2g_3\left(\tfrac{\lambda_1 y_1+\lambda_2 y_2}{2^k}\right)-g_3\left(\tfrac{\lambda_1 y_1}{2^{k-1}}\right)+\\ +2g_3\left(\tfrac{\lambda_1 y_1}{2^k}\right)-g_3\left(\tfrac{\lambda_2 y_2}{2^{k-1}}\right)+2g_3\left(\tfrac{\lambda_2 y_2}{2^k}\right)+F(0,0), \end{aligned} \tag{3.15}$$

for $k = 1, ..., n$. Now from (3.15) it is easy to see that

$$\Delta_n = 2^n \left[g_3 \left(\frac{\lambda_1 y_1 + \lambda_2 y_2}{2^n} \right) - g_3 \left(\frac{\lambda_1 y_1}{2^n} \right) - g_3 \left(\frac{\lambda_2 y_2}{2^n} \right) - F(0,0) \right].$$

This formula together with (3.9) yield

$$\Delta_n = 2^n \left[F \left(\frac{y_1}{2^n}, \frac{y_2}{2^n} \right) - F(0,0) \right]. \tag{3.16}$$

By the mean value theorem, it follows from (3.16) that

$$\Delta_n = \frac{\partial F}{\partial y_1} \left(\frac{\theta y_1}{2^n}, \frac{\theta y_2}{2^n} \right) y_1 + \frac{\partial F}{\partial y_2} \left(\frac{\theta y_1}{2^n}, \frac{\theta y_2}{2^n} \right) y_2, \tag{3.17}$$

where $\theta \in (0,1)$. Now (3.11) follows from (3.17) after taking limits on both sides of (3.17) as $n \to \infty$.

Set the following functions

$$\begin{aligned} f_1(y_1) &= f(A^{-1}\mathbf{y})|_{y_2=0} - H(\lambda_1 y_1) - g_2(0), \\ f_2(y_2) &= f(A^{-1}\mathbf{y})|_{y_1=0} - H(\lambda_2 y_2) - g_1(0), \\ f_3(t) &= H(t) + F(0,0). \end{aligned}$$

Note that by (3.4) the function $f(A^{-1}\mathbf{y})$ depends only on the two variables y_1 and y_2 , hence the functions f_1 and f_2 are well defined. Clearly, $f_i \in C^1(\mathbb{R})$, for $i = 1, 2, 3$. Besides, using (3.8) and (3.11), we can write that

$$\begin{aligned} \textstyle\sum_{i=1}^{3} f_i(\mathbf{a}^i \cdot \mathbf{x}) = f_1(y_1) + f_2(y_2) + f_3(\lambda_1 y_1 + \lambda_2 y_2) = f(A^{-1}\mathbf{y})|_{y_2=0} - H(\lambda_1 y_1) \\ -g_2(0) + +f(A^{-1}\mathbf{y})|_{y_1=0} - H(\lambda_2 y_2) - g_1(0) + H(\lambda_1 y_1 + \lambda_2 y_2) + F(0,0) = \\ = F(y_1, y_2) + f(A^{-1}\mathbf{y})|_{y_1=0} + f(A^{-1}\mathbf{y})|_{y_2=0} - g_1(0) - g_2(0) = f(A^{-1}\mathbf{y}) = f(\mathbf{x}). \end{aligned}$$

The second part of the theorem has been proved. The first part is proved analogously by using Lemma 3.2 instead of Lemma 3.1.

Acknowledgment. The research of the second author was supported by the Science Development Foundation under the President of the Republic of Azerbaijan - Grant EIF-2013-9(15)-46/11/1.

References

[1] J. Aczél, Functional Equations and their Applications, Academic Press, New York, 1966.
[2] R. A. Aliev, V. E. Ismailov, On a smoothness problem in ridge function representation, Adv. in Appl. Math. (to appear), http://dx.doi.org/10.1016/j.aam.2015.11.002.
[3] D. Braess, A. Pinkus, Interpolation by ridge functions, J. Approx. Theory 73 (1993), 218-236.
[4] M. D. Buhmann, A. Pinkus, Identifying linear combinations of ridge functions, Adv. in Appl. Math. 22 (1999), 103-118.
[5] E. J. Candes, Ridgelets: estimating with ridge functions, Ann. Statist. 31 (2003), 1561-1599.
[6] E. J. Candes, Ridgelets: theory and applications. Ph.D. Thesis, Technical Report, Department of Statistics, Stanford University.

[7] P. Diaconis, M. Shahshahani, On nonlinear functions of linear combina-tions, SIAM J. Sci. Stat. Comput. 5 (1984), 175-191.

[8] D. L. Donoho, I. M. Johnstone, Projection-based approximation and a duality method with kernel methods, Ann. Statist. 17 (1989), 58-106.

[9] J. H. Friedman, W. Stuetzle , Projection pursuit regression, J.Amer. Statist. Assoc. 76 (1981), 817-823.

[10] P. J. Huber, Projection pursuit, Ann. Statist. 13 (1985), 435-475.

[11] V. E. Ismailov, Approximation by ridge functions and neural networks with a bounded number of neurons, Appl. Anal. 94 (2015), 2245-2260.

[12] V. E. Ismailov, A review of some results on ridge function approximation. Azerb. J. Math. **3** (2013), no. 1, 3–51.

[13] V. E. Ismailov, On the proximinality of ridge functions. Sarajevo J. Math. 5(17) (2009), no. 1, 109–118.

[14] V. E. Ismailov, On the representation by linear superpositions, J. Approx. Theory 151 (2008), 113-125.

[15] V. E. Ismailov, On the best L_2 approximation by ridge functions, Appl. Math. E-Notes, 7 (2007), 71-76.

[16] V. E. Ismailov, Characterization of an extremal sum of ridge functions, J. Comp. Appl. Math. 205 (2007), 105-115.

[17] F. John, Plane Waves and Spherical Means Applied to Partial Differential Equations, Interscience, New York, 1955.

[18] I. Kazantsev, Tomographic reconstruction from arbitrary directions using ridge functions, Inverse Problems 14 (1998), 635-645.

[19] I. Kazantsev, Lemahieu I., Reconstruction of elongated structures using ridge functions and natural pixels, Inverse Problems 16 (2000), 505-517.

[20] S. Ya. Khavinson, Best approximation by linear superpositions (approximate nomography), Translated from the Russian manuscript by D. Khavinson. Transla-tions of Mathematical Monographs, 159. American Mathematical Society, Provid-ence, RI, 1997, 175 pp.

[21] A. Klopotowski, M. G. Nadkarni, Rao K. P. S. Bhaskara, When is $f(x_1, x_2, ..., x_n) = u_1(x_1) + u_2(x_2) + \cdots + u_n(x_n)$?, Proc. Indian Acad. Sci. Math. Sci. 113 (2003), 77–86.

[22] A. Kroo, On approximation by ridge functions, Constr. Approx. 13 (1997), 447-460.

[23] V. Ya. Lin, A. Pinkus, Fundamentality of ridge functions, J.Approx. Theory 75 (1993), 295-311.

[24] B. F. Logan, L. A. Shepp, Optimal reconstruction of a function from its projections, Duke Math. J. 42 (1975), 645-659.

[25] V. E. Maiorov, On best approximation by ridge functions, J. Approx. Theory 99 (1999), 68-94.

[26] 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.

[27] F. Natterer, The Mathematics of Computerized Tomography, Wiley, New York, 1986.

[28] P. P. Petrushev, Approximation by ridge functions and neural networks, SIAM J.Math. Anal. 30 (1998), 155-189.

[29] A. Pinkus, Smoothness and uniqueness in ridge function representation, Indag. Math. (N.S.) 24 (2013), no. 4, 725–738.

[30] 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.

[31] A. Pinkus, Approximation theory of the MLP model in neural networks, Acta Numerica. 8 (1999), 143-195.
[32] Y. Sternfeld, Uniform separation of points and measures and representation by sums of algebras. Israel J. Math. 55 (1986), 350–362.
[33] Y. Sternfeld, Uniformly separating families of functions. Israel J. Math. 29 (1978), 61–91.

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

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

Tarana M. Shahbalayeva
Institute of Mathematics and Mechanics, NAS of Azerbaijan, Baku, Azerbaijan
E-mail address: sh.terane@gmail.com

Received: June 15, 2015; Accepted: November 25, 2015