Azerbaijan Journal of Mathematics
V. 3, No 1, 2013, January
ISSN 2218-6816

A review of some results on ridge function approximation

Vugar E. Ismailov

Abstract. This paper reviews some results on approximation of multivariate functions by sums of ridge functions with fixed directions.

Key Words and Phrases: Ridge function; Best approximation; Proximinality; Cycle; Path; Orbit.
2000 Mathematics Subject Classifications: 41A30, 41A50, 41A63

1. Introduction

Last 20 years have seen a tremendous interest to the theory of approximation of multivariate functions by so-called *ridge functions*. This is first due to the great importance of such functions in popular application areas such as computerized tomography (see, e.g., [38-40,47,51,55]), statistics (see, e.g., [7,8,12,16,21]) and neural networks (see, e.g., [9,22,48, 57,59, 62,63,69,70]). A ridge function is a multivariate function of the form $g(\mathbf{a}\cdot\mathbf{x})$, where g is a univariate function, $\mathbf{a}=(a_1,...,a_n)$ is a vector (direction) different from zero, $\mathbf{x}=(x_1,...,x_n)$ is a variable and $\mathbf{a}\cdot\mathbf{x}=\sum_{i=1}^{n}a_ix_i$. The vector $\mathbf{a}$ is called a direction of the function $g(\mathbf{a}\cdot\mathbf{x})$. It should be remarked that long before the appearance of the name "ridge", these functions have been used in the theory of partial differential equations under the name of *plane waves* (see, e.g., [37]). For example, assume that $(\alpha_i,\beta_i),\ i=1,...,r$, are pairwise linearly independent vectors in $\mathbb{R}^2$. Then the general solution to the homogeneous partial differential equation

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

where the derivatives are understood in the sense of distributions, are all functions of the form

$$u(x,y)=\sum_{i=1}^{r}g_i\left(\beta_ix-\alpha_iy\right),$$

for arbitrary continuous univariate functions g_i, $i=1,...,r$. The term "ridge function" was coined by Logan and Shepp in their seminal paper [47] dedicated to the basic mathematical problem of computerized tomography. The problem consists of reconstructing a given multivariate function from values of its integrals along certain lines in the plane. 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 [47] used the criterion of minimizing the L_2 norm of g. That is, they found a function $g(x)$ 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 of [47] 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 [38] solved the above problem of tomography without requiring that the considered directions be equally spaced. Marr [51] 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 match 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 with fixed directions:

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

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.

Ridge function approximations appear also in statistics in Projection Pursuit. This term was introduced by Friedman and Tukey [15] to name a technique for the explanatory analysis of large and multivariate data sets. This technique seeks out "interesting" linear projections of the multivariate data onto a line or a plane. Projection Pursuit algorithms approximate a multivariate function $f(x_1,...,x_n)$ by sums of ridge functions with variable directions, that is, by functions from the set

$$\mathcal{R}_r = \left\{ \sum_{i=1}^{r} g_i\left(\mathbf{a}^i \cdot \mathbf{x}\right) : \mathbf{a}^i \in \mathbb{R}^n \setminus \{\mathbf{0}\},\ g_i : \mathbb{R} \to \mathbb{R}, i = 1,...,r \right\}.$$

Here r is the only fixed parameter, directions $\mathbf{a}^1,...,\mathbf{a}^r$ and functions $g_1,...,g_r$ are free to choose. The first method of such approximation was developed by Friedman and Stuetzle [16]. Their approximation process called projection pursuit regression (PPR) operates in a stepwise and greedy fashion. The process does not find a best approximation from $\mathcal{R}_r$, it algorithmically constructs functions $g_r \in \mathcal{R}_r$, such that $\|g_r - f\|_{L_2} \to 0$ as $r \to \infty$. At stage m, PPR looks for a univariate function g_m and direction $\mathbf{a}^m$ such that the ridge function $g_m(\mathbf{a}^m \cdot \mathbf{x})$ best approximates the residual $f(x) - \sum_{j=1}^{m-1} g_j(\mathbf{a}^j \cdot \mathbf{x})$. Projection pursuit regression has been proposed as an approach to bypass the curse of dimensionality and now is applied to prediction in applied sciences. In [7,8], Candes developed a new approach based not on stepwise construction of approximation but on a new transform called the ridgelet transform. The ridgelet transform represents general functions as integrals of ridgelets - specifically chosen ridge functions.

The significance of approximation by ridge functions can be well understood from its role in the theory of neural networks. Ridge functions appear in the definitions of many central neural network models. It is a broad knowledge that neural networks are being successfully applied across an extraordinary range of problem domains, in fields as diverse as finance, medicine, engineering, geology and physics. Generally speaking, neural networks are being introduced anywhere that there are problems of prediction, classification or control. Thus not surprisingly, there is a great interest to this powerful and very popular area of research (see, e.g., [59] and a great deal of references therein). An artificial neural network is a way to perform computations using networks of interconnected computational units vaguely analogous to neurons simulating how our brain solves them. An artificial neuron, which forms the basis for designing neural networks, is a device with n real inputs and an output. This output is generally a ridge function of the given inputs. In mathematical terms, a neuron may be described by

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

where $\mathbf{x} = (\mathbf{x}_1, ..., x_n) \in \mathbb{R}^n$ are the input signals, $w = (w_1, ..., w_n) \in \mathbb{R}^n$ are the synaptic weights, $\theta \in \mathbb{R}$ is the bias, σ is the activation function and y is the output signal of the neuron. In a layered neural network the neurons are organized in the form of layers. We have at least two layers: an input and an output layer. The layers between the input and the output layer (if any) are called hidden layers, whose computation nodes are correspondingly called hidden neurons or hidden units. The output signals of the first layer are used as inputs to the second layer, the output signals of the second layer are used as inputs to the third layer, and so on for the rest of the network. Neural networks with this kind of architecture is called as multilayer feedforward perceptron (MLP). This is the most popular model among other neural network models. In this model, a neural network with a single hidden layer and one output represents a function of the form

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

Here the weights $\mathbf{w}^i$ are vectors in $\mathbb{R}^n$, the thresholds θ_i and the coefficients c_i are real numbers and the activation function σ is a univariate function. We fix only σ and r. Note that the functions $\sigma(\mathbf{w}^i \cdot \mathbf{x} - \theta_i)$ are ridge functions. Thus, it is not surprising that some approximation theoretic problems concerned with neural nets have strong association with the corresponding problems of approximation by ridge functions.

As indicated above, approximation by members of the set $\mathcal{R}\left(\mathbf{a}^1, ..., \mathbf{a}^r\right)$ is of great importance in mathematical problems of computerized tomography. This type of approximation is also essential in applications, where it is required to approximate complicated multivariate function by functions of simple structure. Note that when the directions coincide with the basic directions of the considered space, the approximation from the set $\mathcal{R}\left(\mathbf{a}^1, ..., \mathbf{a}^r\right)$ turns into the problem of approximation of multivariate functions by sums of univariate functions. The last subject has been within research interests of many authors (see e.g. [41] and a great deal of references therein).

Representability of polynomials by sums of ridge functions is a building block for many results. In many works (see, e.g., [59]), the following fact is fundamental:

Every multivariate polynomial $h(\mathbf{x}) = h(x_1, ..., x_n)$ of degree k can be represented in the form

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

where p_i is a univariate polynomial, $\mathbf{a}^i \in \mathbb{R}^n$, and $r = \binom{n-1+k}{k}$.

For example, for the representation of a bivariate polynomial of degree k, it is needed $k+1$ univariate polynomials and $k+1$ directions (see [47]). The proof of this fact is organized so that the directions $\mathbf{a}^i$, $i = 1, ..., r$, are chosen once for all the multivariate polynomials of k-th degree. At one of the seminars held at the Technion - Israel Institute of Technology, Professor A.Pinkus posed two problems:

1) Can every multivariate polynomial of degree k be represented by less than r ridge functions ?

2) How large is the set of polynomials represented by $r-1$, $r-2$, ... ridge functions ?

Note that for bivariate polynomials the 1-st problem is solved positively, that is, the number $r = k+1$ can be reduced. Indeed, for a bivariate polynomial $P(x,y)$ of k-th degree, there exists a large set of real numbers $c_0, ..., c_k$ such that

$$\sum_{i=0}^{k} c_i \frac{\partial^k}{\partial x^i \partial y^{k-i}} P(x,y) = 0.$$

Further, the numbers c_i, $i = 0, ..., k$, can be selected to enjoy the property that the polynomial $\sum_{i=0}^{k} c_i t^i$ has distinct real zeros. Then it is not difficult to verify that the differential operator $\sum_{i=0}^{k} c_i \frac{\partial^k}{\partial x^i \partial y^{k-i}}$ can be written in the form

$$\prod_{i=1}^{k} \left(b_i \frac{\partial}{\partial x} - a_i \frac{\partial}{\partial y} \right), \tag{1}$$

for some pairwise linearly independent vectors (a_i, b_i), $i = 1, ..., k$. But the operator (1) annihilates a function $f(x,y)$ if and only if $f(x,y)=\sum_{i=1}^{k} g_i(a_i x+b_i y)$. Thus, we obtain that the polynomial $P(x,y)$ can be represented as a sum of k ridge functions.

In connection with the 2-nd problem of Pinkus, V. Maiorov [50] studied certain geometrical properties of the manifold $\mathcal{R}_r$, namely, he estimated the ε-entropy numbers in terms of smaller ε-covering numbers of the compact class formed by the intersection of the class $\mathcal{R}_r$ with the unit ball in the space of polynomials of degree at most s on $\mathbb{R}^n$. Let E be a Banach space and let for $x\in E$ and $\delta>0$, $S(x,\delta)$ denote the ball of radius δ centered at the point x. For any positive number ε, the ε-covering number of a set F in the space E represents the quantity

$$L_{\varepsilon}(F,E)=\min\left\{N:\ \exists x_1,...,x_N\in F \text{ such that } F\subset\bigcup_{i=1}^{N}S(x_i,\varepsilon)\right\}.$$

The ε-entropy of F is defined as the number $H_{\varepsilon}(F,E)\overset{def}{=}\log_2 L_{\varepsilon}(F,E)$. The notion of ε-entropy has been devised by A.N.Kolmogorov (see [44,45,67]) in order to classify compact metric sets according to their massivity.

In order to formulate Maiorov's result here let $\mathcal{R}_r$ be the class of all possible linear combinations of r ridge functions, $\mathcal{P}_s^n$ be the space of all polynomials of degree at most s on $\mathbb{R}^n$, $L_q=L_q(I)$, $1\le q\le\infty$, be the space of q-integrable functions on the unit cube $I=[0,1]^n$ with the norm $\|f\|_q=\left(\int_I|f(x)|^q\,dx\right)^{1/q}$, BL_q be the unit ball in the space L_q, and $B_q\mathcal{P}_s^n=BL_q\cap\mathcal{P}_s^n$ be the unit ball in the space $\mathcal{P}_s^n$ equipped with the L_q metric.

Proposition 1. *(Maiorov [50]). Let $r,s\in\mathbb{N}$, $1\le q\le\infty$, and $0<\varepsilon<1$. Then the ε-entropy of the class $B_q\mathcal{P}_s^n\cap\mathcal{R}_r$ in the space L_q satisfies the inequalities*

1)

$$c_1 rs\le\frac{H_{\varepsilon}(B_q\mathcal{P}_s^n\cap\mathcal{R}_r,L_q)}{\log_2\frac{1}{\varepsilon}}\le c_2 rs\log_2\frac{2es^{n-1}}{r},\tag{2}$$

for $r\le s^{n-1}$.

2)

$$c_1^{'}s^n\le\frac{H_{\varepsilon}(B_q\mathcal{P}_s^n\cap\mathcal{R}_r,L_q)}{\log_2\frac{1}{\varepsilon}}\le c_2^{'}s^n,\tag{3}$$

for $r>s^{n-1}$. In (2) and (3), $c_1,c_2,c_1^{'},c_2^{'}$ are constants depending only on n.

In this paper, we survey some results on the approximation of multivariate functions by ridge functions with fixed directions, that is by functions from the set $\mathcal{R}\left(\mathbf{a}^1,...,\mathbf{a}^r\right)$. Many results concerning the approximation of multivariate functions by linear combinations of ridge functions from the set $\mathcal{R}_r$ and their applications in neural network theory may be found in [9,22,48, 49,56,57, 59,62,63, 66,69,70].

2. Representation of multivariate functions by linear combinations of ridge functions

One of the basic problems concerning the approximation by sums of ridge functions with fixed directions is the problem of verifying if a given function f belongs to the space $\mathcal{R}\left(\mathbf{a}^1, ..., \mathbf{a}^r\right)$. This problem has a simple solution if a dimension of a considered space $n = 2$ and a given function $f(x, y)$ has partial derivatives up to r-th order. For the representation of $f(x, y)$ in the following form

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

it is necessary and sufficient that

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

The last verification equation is valid for all continuous bivariate functions provided that the derivatives are understood in the generalized sense.

Unfortunately, such simple verification does not carry over to the representation $f(\mathbf{x}) = \sum\limits_{i=1}^{r} g_i\left(\mathbf{a}^i \cdot \mathbf{x}\right)$, $\mathbf{x} = (x_1, ..., x_n)$, if the dimension $n > 2$. Below we cite two results on the representation of a given multivariate function as a sum of ridge functions with fixed directions.

Proposition 2. *(Diaconis, Shahshahani [10]). 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 [10]) showing that one cannot simply dispense with the polynomial $P(\mathbf{x})$ in the above proposition (this term is necessary for the sufficient part of the proposition).

Lin and Pinkus [46] obtained more general result on the representation by ridge functions. We need some notation to present their result. 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 3. *(Lin, Pinkus [46]). 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.

In his recent paper [60], A.Pinkus considers the problems of smoothness and uniqueness in ridge function representation. For a given function $f\in \mathcal{R}\left(\mathbf{a}^1,...,\mathbf{a}^r\right)$, he poses and answers the following questions. If f belongs to some smoothness class, what can we say about the smoothness of the functions g_i? How many different ways can we write f as a linear combination of ridge functions ?

Further, we consider the following two problems.

Problem 1. *What conditions imposed on $f:X\rightarrow \mathbb{R}$ are necessary and sufficient for the inclusion $f\in \mathcal{R}(\mathbf{a}^1,...,\mathbf{a}^r;X)$?*

Problem 2. *What conditions imposed on X are necessary and sufficient that every function defined on X belongs to the space $\mathcal{R}(\mathbf{a}^1,...,\mathbf{a}^r;X)$?*

As noticed above, Problem 1 was solved for continuous functions (see [6,46]). It is also noticed that the like but different problem of representation by sums of ridge functions and a polynomial was solved for continuously differentiable functions (see [10]). Problem 2 was solved by Braess and Pinkus [4] for finite subsets X of $\mathbb{R}^d$ (i.e. the problem of interpolation was solved). In [4], it is required to characterize all finite subsets $\{\mathbf{x}^1,...,\mathbf{x}^k\}\subset \mathbb{R}^n$ such that for any data $\{\alpha_1,...,\alpha_k\}\subset \mathbb{R}$ there exists a function $g\in \mathcal{R}\left(\mathbf{a}^1,...,\mathbf{a}^r\right)$ satisfying the equations $g(\mathbf{x}^i)=\alpha_i$, $i=1,...,k$. Such finite sets $\{\mathbf{x}^1,...,\mathbf{x}^k\}$ are said to have the interpolation property. In connection with the problem of interpolation, Braess and Pinkus [4] introduced two additional notions: Given directions $\{\mathbf{a}^j\}_{j=1}^r\subset \mathbb{R}^n\backslash\{\mathbf{0}\}$, we say that a set of points $\{\mathbf{x}^i\}_{i=1}^k\subset \mathbb{R}^n$ has the *NI*-property (non-interpolation property) with respect to $\{\mathbf{a}^j\}_{j=1}^r$, if there exists $\{\alpha_i\}_{i=1}^k\subset \mathbb{R}$ such that we cannot find a function $g\in \mathcal{R}\left(\mathbf{a}^1,...,\mathbf{a}^r\right)$ satisfying $g(\mathbf{x}^i)=\alpha_i$, $i=1,...,k$. We say that the set $\{\mathbf{x}^i\}_{i=1}^k\subset \mathbb{R}^n$ has the *MNI*-property (minimal non-interpolation property) with respect to $\{\mathbf{a}^j\}_{j=1}^r$, if $\{\mathbf{x}^i\}_{i=1}^k$ but no proper subset thereof has the *NI*-property. The following proposition is valid.

Proposition 4. *(Braess and Pinkus [4]). The set* $\{\mathbf{x}^i\}_{i=1}^k \subset \mathbb{R}^n$ *has the NI-property if and only if there is a vector* $\mathbf{m} = (m_1, ..., m_k) \in \mathbb{Z}^k \backslash \{\mathbf{0}\}$ *such that*

$$\sum_{j=1}^{k} m_j g(\mathbf{a}^i \cdot \mathbf{x}^j) = 0, \ i = 1, ..., r,$$

for all functions $g : \mathbb{R} \to \mathbb{R}$. *This set has the MNI-property if and only if the vector* $\mathbf{m}$ *has the additional properties: it is unique up to multiplication by a constant and all its components* $m_1, ..., m_k$ *are different from zero.*

In [28], the author considered Problems 1 and 2 without imposing on f additional conditions of continuity and differentiation and without requiring that X be a finite set. In fact, in [28] these problems were solved for more general set of functions, namely for the set

$$\mathcal{B}(X) = \mathcal{B}(h_1, ..., h_r; X) = \left\{ \sum_{i=1}^{r} g_i(h_i(x)), \ x \in X, \ g_i : \mathbb{R} \to \mathbb{R}, \ i = 1, ..., r \right\},$$

where $h_i : X \to \mathbb{R}, \ i = 1, ..., r$, are arbitrarily fixed functions. In particular, the functions $h_i, \ i = 1, ..., r$, may be equal to scalar products of the variable $\mathbf{x}$ with some vectors $\mathbf{a}^i$, $i = 1, ..., r$. Only in this special case, we have $\mathcal{B}(h_1, ..., h_r; X) = \mathcal{R}(\mathbf{a}^1, ..., \mathbf{a}^r; X)$.

The main idea leading to the solutions rests on exploiting of new objects called cycles with respect to r functions $h_i : X \to \mathbb{R}, \ i = 1, ..., r$ (and in particular, with respect to r directions $\mathbf{a}^1, ..., \mathbf{a}^r$). In the sequel, by δ_A we will denote the characteristic function of a set $A \subset \mathbb{R}$. That is,

$$\delta_A(y) = \begin{cases} 1, \ if \ y \in A, \\ 0, \ if \ y \notin A. \end{cases}$$

Definition 1. *(see [23,27,28,43]). Given a subset* $X \subset \mathbb{R}^d$ *and nonzero functions* $h_i : X \to \mathbb{R}, \ i = 1, ..., r$. *A set of points* $\{x_1, ..., x_n\} \subset X$ *is called to be a cycle with respect to the functions* $h_1, ..., h_r$ *(or, concisely, a cycle if there is no confusion), if there exists a vector* $\lambda = (\lambda_1, ..., \lambda_n)$ *with the nonzero real coordinates* $\lambda_i, \ i = 1, ..., n$, *such that*

$$\sum_{j=1}^{n} \lambda_j \delta_{h_i(x_j)} = 0, \ i = 1, ..., r. \tag{4}$$

If $h_i = \mathbf{a}^i \cdot \mathbf{x}$, $i = 1, ..., r$, where $\mathbf{a}^1, ..., \mathbf{a}^r$ are some nonzero directions in $\mathbb{R}^d$, a cycle with respect to the functions $h_1, ..., h_r$ will be also called a cycle with respect to the directions $\mathbf{a}^1, ..., \mathbf{a}^r$.

Let for $i = 1, ..., r$, the set $\{h_i(x_j), \ j = 1, ..., n\}$ have k_i different values. Then it is not difficult to see that Eq. (4) stands for a system of $\sum_{i=1}^r k_i$ homogeneous linear equations in unknowns $\lambda_1, ..., \lambda_n$. If this system has any solution with the nonzero components, then the given set $\{x_1, ..., x_n\}$ is a cycle. In the last case, the system has also a solution $m = (m_1, ..., m_n)$ with the nonzero integer components $m_i,\ i = 1, ..., n$. Thus, in Definition 1, the vector $\lambda = (\lambda_1, ..., \lambda_n)$ can be replaced by a vector $m = (m_1, ..., m_n)$ with $m_i \in \mathbb{Z}\backslash\{0\}$.

For example, the set $l = \{(0, 0, 0),\ (0, 0, 1),\ (0, 1, 0),\ (1, 0, 0),\ (1, 1, 1)\}$ is a cycle in $\mathbb{R}^3$ with respect to the functions $h_i(z_1, z_2, z_3) = z_i,\ i = 1, 2, 3$. The vector λ in Definition 1 can be taken as $(2, 1, 1, 1, -1)$.

In the case $r = 2$, the picture of cycles becomes more clear. Let, for example, h_1 and h_2 be the coordinate functions on $\mathbb{R}^2$. In this case, a cycle is the union of some sets A_k with the property: each A_k consists of vertices of a closed broken line with the sides parallel to the coordinate axis. These objects (sets A_k) 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, [41, Chapter 2]). If X and the functions h_1 and h_2 are arbitrary, the sets A_k can be described as a trace of some point traveling alternatively in the level sets of h_1 and h_2, and then returning to its primary position. It should be remarked that in the case $r > 2$ cycles do not admit such a simple geometric description. We refer the reader to Braess and Pinkus [1] for the description of cycles when $r = 3$ and $h_i(\mathbf{x}) = \mathbf{a}^i{\cdot}\mathbf{x}$, $\mathbf{x} \in\mathbb{R}^2,\ \mathbf{a}^i \in \mathbb{R}^2\backslash\{\mathbf{0}\},\ i = 1, 2, 3$.

Let $T(X)$ denote the set of all functions on X. With each pair $\langle p, \lambda\rangle$, where $p = \{x_1, ..., x_n\}$ is a cycle in X and $\lambda = (\lambda_1, ..., \lambda_n)$ is a vector known from Definition 1, we associate the functional

$$G_{p,\lambda} : T(X) \to \mathbb{R},\quad G_{p,\lambda}(f) = \sum_{j=1}^{n} \lambda_j f(x_j).$$

In the following, such pairs $\langle p, \lambda\rangle$ will be called *cycle-vector pairs* of X. It is clear that the functional $G_{p,\lambda}$ is linear and $G_{p,\lambda}(g) = 0$ for all functions $g \in \mathcal{B}(h_1, ..., h_r; X)$.

Lemma 1. *[28]. Let X have cycles $h_i(X) \cap h_j(X) = \varnothing$, for all $i, j \in \{1, ..., r\},\ i \neq j$. Then a function $f : X \to \mathbb{R}$ belongs to the set $\mathcal{B}(h_1, ..., h_r; X)$ if and only if $G_{p,\lambda}(f) = 0$ for any cycle-vector pair $\langle p, \lambda\rangle$ of X.*

Proof. The necessity is obvious, since the functional $G_{p,\lambda}$ annihilates all members of the set $\mathcal{B}(h_1, ..., h_r; X)$. Let us prove the sufficiency. Introduce the notation

$$\begin{aligned} Y_i &= h_i(X),\ i = 1, ..., r; \\ \Omega &= Y_1 \cup ... \cup Y_r. \end{aligned}$$

Consider the following subsets of Ω:

$$\mathcal{L} = \{Y = \{y_1, ..., y_r\} : \text{if there exists } x \in X \text{ such that } h_i(x) = y_i,\ i = 1, ..., r\}. \tag{5}$$

In what follows, all the points x associated with Y by (5) will be called $(*)$-points of Y. It is clear that the number of such points depends on Y as well as on the functions

$h_1, ..., h_r$, and may be greater than 1. But note that if any two points x_1 and x_2 are $(*)$-points of Y, then necessarily the set $\{x_1, x_2\}$ forms a cycle with the associated vector $\lambda_0 = (1; -1)$. Indeed, if x_1 and x_2 are $(*)$-points of Y, then $h_i(x_1) = h_i(x_2)$, $i = 1, ..., r$, whence

$$1 \cdot \delta_{h_i(x_1)} + (-1) \cdot \delta_{h_i(x_2)} \equiv 0, \; i = 1, ..., r.$$

The last identity means that the set $p_0 = \{x_1, x_2\}$ forms a cycle and $\lambda_0 = (1; -1)$ is an associated vector. Then by the condition of the sufficiency, $G_{p_0,\lambda_0}(f) = 0$, which yields that $f(x_1) = f(x_2)$.

Let now Y^* be the set of all $(*)$-points of Y. Since we have already known that $f(Y^*)$ is a single number, we can define the function

$$t : \mathcal{L} \to \mathbb{R}, \; t(Y) = f(Y^*).$$

Or, equivalently, $t(Y) = f(x)$, where x is an arbitrary $(*)$-point of Y.

Consider now a class $\mathcal{S}$ of functions of the form $\sum_{j=1}^{k} r_j \delta_{D_j}$, where k is a positive integer, r_j are real numbers and D_j are elements of $\mathcal{L}$, $j = 1, ..., k$. We fix neither the numbers k, r_j, nor the sets D_j. Clearly, $\mathcal{S}$ is a linear space. Over $\mathcal{S}$, we define the functional

$$F : \mathcal{S} \to \mathbb{R}, \; F\left(\sum_{j=1}^{k} r_j \delta_{D_j}\right) = \sum_{j=1}^{k} r_j t(D_j).$$

First of all, we must show that this functional is well defined. That is, the equality

$$\sum_{j=1}^{k_1} r_j' \delta_{D_j'} = \sum_{j=1}^{k_2} r_j'' \delta_{D_j''},$$

always implies the equality

$$\sum_{j=1}^{k_1} r_j' t(D_j') = \sum_{j=1}^{k_2} r_j'' t(D_j'').$$

In fact, this is equivalent to the implication

$$\sum_{j=1}^{k} r_j \delta_{D_j} = 0 \Longrightarrow \sum_{j=1}^{k} r_j t(D_j) = 0, \text{ for all } k \in \mathbb{N}, \, r_j \in \mathbb{R}, \, D_j \subset \mathcal{L}. \tag{6}$$

Suppose that the left-hand side of the implication (6) be satisfied. Each set D_j consists of r real numbers $y_1^j, ..., y_r^j$, $j = 1, ..., k$. By the hypothesis of the lemma, all these numbers are different. Therefore,

$$\delta_{D_j} = \sum_{i=1}^{r} \delta_{y_i^j}, \; j = 1, ..., k. \tag{7}$$

Eq. (7) together with the left-hand side of (6) gives

$$\sum_{i=1}^{r}\sum_{j=1}^{k} r_j \delta_{y_i^j} = 0. \tag{8}$$

Since the sets $\{y_i^1, y_i^2, ..., y_i^k\}$, $i = 1, ..., r$, are pairwise disjoint, we obtain from (8) that

$$\sum_{j=1}^{k} r_j \delta_{y_i^j} = 0, \ i = 1, ..., r. \tag{9}$$

Let now $x_1, ..., x_k$ be some $(*)$-points of the sets $D_1, ..., D_k$, respectively. Since by (5) $y_i^j = h_i(x_j)$, for $i = 1, ..., r$ and $j = 1, ..., k$, it follows from (9) that the set $\{x_1, ..., x_k\}$ is a cycle. Then by the condition of the sufficiency, $\sum_{j=1}^{k} r_j f(x_j) = 0$. Hence, $\sum_{j=1}^{k} r_j t(D_j) = 0$. We have proved the implication (6) and hence the functional F is well defined. Note that the functional F is linear (this can be easily seen from its definition).

Consider now the following space:

$$\mathcal{S}' = \left\{ \sum_{j=1}^{k} r_j \delta_{\omega_j} \right\},$$

where $k \in \mathbb{N}$, $r_j \in \mathbb{R}$, $\omega_j \subset \Omega$. As in the above, we do not fix the parameters k, r_j and ω_j. Clearly, the space $\mathcal{S}'$ is larger than $\mathcal{S}$. Let us prove that the functional F can be linearly extended to the space $\mathcal{S}'$. So, we must prove that there exists a linear functional $F' : \mathcal{S}' \to \mathbb{R}$ such that $F'(x) = F(x)$, for all $x \in \mathcal{S}$. Let H denote the set of all linear extensions of F to subspaces of $\mathcal{S}'$ containing $\mathcal{S}$. The set H is not empty, since it contains a functional F. For each functional $v \in H$, let $dom(v)$ denote the domain of v. Consider the following partial order in H: $v_1 \le v_2$, if v_2 is a linear extension of v_1 from the space $dom(v_1)$ to the space $dom(v_2)$. Let now P be any chain (linearly ordered subset) in H. Consider the following functional u defined on the union of domains of all functionals $p \in P$:

$$u : \bigcup_{p \in P} dom(p) \to \mathbb{R}, \ u(x) = p(x), \text{ if } x \in dom(p).$$

Obviously, this functional is well defined and linear. Besides, the functional u provides an upper bound for P. We see that the arbitrarily chosen chain P has an upper bound. Then by Zorn's lemma, there is a maximal element $F' \in H$. We claim that the functional F' must be defined on the whole space $\mathcal{S}'$. Indeed, if F' is defined on a proper subspace $\mathcal{D} \subset \mathcal{S}'$, then it can be linearly extended to a space larger than $\mathcal{D}$ by the following way: take any point $x \in \mathcal{S}' \backslash \mathcal{D}$ and consider the linear space $\mathcal{D}' = \{\mathcal{D} + \alpha x\}$, where α runs through all real numbers. For an arbitrary point $y + \alpha x \in \mathcal{D}'$, set $F''(y + \alpha x) = F'(y) + \alpha b$, where b is any real number considered as the value of F'' at x. Thus, we constructed a linear functional $F'' \in H$ satisfying $F' \le F''$. The last contradicts the maximality of F'. This means that the functional F' is defined on the whole $\mathcal{S}'$ and $F \le F'$ (F' is a linear extension of F).

Define the following functions by means of the functional F':

$$g_i : Y_i \to \mathbb{R},\ g_i(y_i) \overset{def}{=} F'(\delta_{y_i}),\ i = 1, ..., r.$$

Let x be an arbitrary point in X. Obviously, x is a $(*)$-point of some set $Y = \{y_1, ..., y_r\} \subset \mathcal{L}$. Thus,

$$\begin{aligned} f(x) &= t(Y) = F(\delta_Y) = F\left(\sum_{i=1}^{r} \delta_{y_i}\right) = F'\left(\sum_{i=1}^{r} \delta_{y_i}\right) = \\ \sum_{i=1}^{r} F'(\delta_{y_i}) &= \sum_{i=1}^{r} g_i(y_i) = \sum_{i=1}^{r} g_i(h_i(x)). \end{aligned}$$

◀

Definition 2. *A cycle $p = \{x_1, ..., x_n\}$ is said to be minimal if p does not contain any cycle 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 cycle with respect to the functions $h_i(z_1, z_2, z_3) = z_i,\ i = 1, 2, 3$. Adding the point $(0,1,1)$ to l, we will have a cycle, but not minimal. The vector λ associated with $l \cup \{(0,1,1)\}$ can be taken as $(3, -1, -1, -2, 2, -1)$.

A minimal cycle $p = \{x_1, ..., x_n\}$ has the following obvious properties:

(a) *The vector λ associated with p by Eq. (4) is unique up to multiplication by a constant;*

(b) *If in (4), $\sum_{j=1}^{n} |\lambda_j| = 1$, then all the numbers $\lambda_j,\ j = 1, ..., n,$ are rational.*

Thus, a minimal cycle p uniquely defines the functional

$$G_p(f) = \sum_{j=1}^{n} \lambda_j f(x_j), \quad \sum_{j=1}^{n} |\lambda_j| = 1.$$

Lemma 2. *[28]. The functional $G_{p,\lambda}$ is a linear combination of functionals $G_{p_1}, ..., G_{p_k}$, where $p_1, ..., p_k$ are minimal cycles in p.*

Proof. Let $\langle p, \lambda \rangle$ be a cycle-vector pair of X, where $p = \{x_1, ..., x_n\}$ and $\lambda = (\lambda_1, ..., \lambda_n)$. Let $p_1 = \{y_1^1, ..., y_{s_1}^1\}$, $s_1 < n$, be a minimal cycle in p and

$$G_{p_1}(f) = \sum_{j=1}^{s_1} \nu_j^1 f(y_j^1), \ \ \sum_{j=1}^{s_1} \left|\nu_j^1\right| = 1.$$

Without loss of generality, we may assume that $y_1^1 = x_1$. Put

$$t_1 = \frac{\lambda_1}{\nu_1^1}.$$

Then the functional $G_{p,\lambda} - t_1 G_{p_1}$ has the form

$$G_{p,\lambda} - t_1 G_{p_1} = \sum_{j=1}^{n_1} \lambda_j^1 f(x_j^1),$$

where $x_j^1 \in p$, $\lambda_j^1 \neq 0$, $j = 1, ..., n_1$. Clearly, the set $l_1 = \{x_1^1, ..., x_{n_1}^1\}$ is a cycle in p with the associated vector $\lambda^1 = (\lambda_1^1, ..., \lambda_{n_1}^1)$. Besides, $x_1 \notin l_1$. Thus, $n_1 < n$ and $G_{l_1,\lambda^1} = G_{p,\lambda} - t_1 G_{p_1}$. If l_1 is minimal, then the proof is completed. Let l_1 be not minimal. Let $p_1 = \{y_1^2, ..., y_{s_2}^2\}$, $s_2 < n_1$, be a minimal cycle in l_1 and

$$G_{p_2}(f) = \sum_{j=1}^{s_2} \nu_j^2 f(y_j^2), \quad \sum_{j=1}^{s_2} \left|\nu_j^2\right| = 1.$$

Without loss of generality, we may assume that $y_1^2 = x_1^1$. Put

$$t_2 = \frac{\lambda_1^1}{\nu_1^2}.$$

Then the functional $G_{l_1,\lambda^1} - t_2 G_{p_2}$ has the form

$$G_{l_1,\lambda^1} - t_2 G_{p_2} = \sum_{j=1}^{n_2} \lambda_j^2 f(x_j^2),$$

where $x_j^2 \in l_1$, $\lambda_j^2 \neq 0$, $j = 1, ..., n_2$. Clearly, the set $l_2 = \{x_1^2, ..., x_{n_2}^2\}$ is a cycle in l_1 with the associated vector $\lambda^2 = (\lambda_1^2, ..., \lambda_{n_2}^2)$. Besides, $x_1^1 \notin l_2$. Thus, $n_2 < n_1$ and $G_{l_2,\lambda^2} = G_{l_1,\lambda^1} - t_2 G_{p_2}$. If l_2 is minimal, then the proof is completed. Let l_2 be not minimal. Repeating the above process for l_2, then for l_3, etc., after some $k-1$ steps we will come to a minimal cycle l_{k-1} and the functional

$$G_{l_{k-1},\lambda^{k-1}} = G_{l_{k-2},\lambda^{k-2}} - t_{k-1} G_{p_{k-1}} = \sum_{j=1}^{n_{k-1}} \lambda_j^{k-1} f(x_j^{k-1}).$$

Since the cycle l_{k-1} is minimal,

$$G_{l_{k-1},\lambda^{k-1}} = t_k G_{l_{k-1}}, \quad \text{where } t_k = \sum_{j=1}^{n_{k-1}} \left|\lambda_j^{k-1}\right|.$$

Now putting $p_k = l_{k-1}$ and considering the above chain relations between the functionals G_{l_i,λ^i}, $i = 1, ..., k-1$, we obtain that

$$G_{p,\lambda} = \sum_{i=1}^{k} t_i G_{p_i}.$$

◄

Theorem 1. *[28]. Let $X \subset \mathbb{R}^d$ and $h_1, ..., h_r$ be any nonzero real functions defined on X.*

1) Let X have cycles with respect to the functions $h_1, ..., h_r$. A function $f : X \to \mathbb{R}$ belongs to the space $\mathcal{B}(h_1, ..., h_r; X)$ if and only if $G_p(f) = 0$ for any minimal cycle $p \subset X$.

2) Let X have no cycles. Then $\mathcal{B}(h_1, ..., h_r; X) = T(X)$.

Proof. 1) The necessity is clear. Let us prove the sufficiency. On the strength of Lemma 2, it is enough to prove that if $G_{p,\lambda}(f) = 0$ for any cycle-vector pair $\langle p, \lambda \rangle$ of X, then $f \in \mathcal{B}(X)$.

Consider a system of intervals $\{(a_i, b_i) \subset \mathbb{R}\}_{i=1}^r$ such that $(a_i, b_i) \cap (a_j, b_j) = \varnothing$ for all the indices $i, j \in \{1, ..., r\}$, $i \neq j$. For $i = 1, ..., r$, let τ_i be one-to-one mappings of $\mathbb{R}$ onto (a_i, b_i). Introduce the following functions on X:

$$h_i^{'}(x) = \tau_i(h_i(x)), \ i = 1, ..., r.$$

It is clear that any cycle with respect to the functions $h_1, ..., h_r$ is also a cycle with respect to the functions $h_1^{'}, ..., h_r^{'}$, and vice versa. Besides, $h_i'(X) \cap h_j'(X) = \varnothing$, for all $i, j \in \{1, ..., r\}$, $i \neq j$. Then by Lemma 1

$$f(x) = g_1'(h_1'(x)) + \cdots + g_r'(h_r'(x)),$$

where $g_1', ..., g_r'$ are univariate functions depending on f. From the last equality we obtain that

$$f(x) = g_1'(\tau_1(h_1(x))) + \cdots + g_r'(\tau_r(h_r(x))) = g_1(h_1(x)) + \cdots + g_r(h_r(x)).$$

That is, $f \in \mathcal{B}(X)$.

2) Let $f : X \to \mathbb{R}$ be an arbitrary function. First suppose that $h_i(X) \cap h_j(X) = \varnothing$, for all $i, j \in \{1, ..., r\}$, $i \neq j$. In this case, the proof is similar to and even simpler than that of Lemma 1. Indeed, the set of all $(*)$-points of Y consists of a single point, since otherwise we would have a cycle with two points, which contradicts the hypothesis of the 2-nd part of our theorem. Further, the well definition of the functional F becomes obvious, since the left-hand side of (6) also contradicts the nonexistence of cycles. Thus, as in the proof of Lemma 1, we can extend F to the space $\mathcal{S}'$ and then obtain the desired representation for the function f. Since f is arbitrary, $T(X) = \mathcal{B}(X)$.

Using the techniques from the proof of the 1-st part of our theorem, one can easily generalize the above argument to the case when the functions $h_1, ..., h_r$ have arbitrary ranges.◄

Theorem 2. *[28]. $\mathcal{B}(h_1, ..., h_r; X) = T(X)$ if and only if X has no cycles with respect to the functions $h_1, ..., h_r$.*

Proof. The sufficiency immediately follows from Theorem 1. To prove the necessity, assume that X has a cycle $p = \{x_1, ..., x_n\}$. Let $\lambda = (\lambda_1, ..., \lambda_n)$ be a vector associated with p by Eq. (4). Consider a function f_0 on X with the property: $f_0(x_i) = 1$, for indices i such that $\lambda_i > 0$ and $f_0(x_i) = -1$, for indices i such that $\lambda_i < 0$. For this function, $G_{p,\lambda}(f_0) \neq 0$. Then by Theorem 1, $f_0 \notin \mathcal{B}(X)$. Hence, $\mathcal{B}(X) \neq T(X)$. The contradiction shows that X does not admit cycles.◄

From Theorems 1 and 2 we obtain the following corollaries for the ridge function representation.

Corollary 1. *Let $X \subset \mathbb{R}^d$ and $\mathbf{a}^1, ..., \mathbf{a}^r \in \mathbb{R}^d \backslash \{\mathbf{0}\}$.*

1) Let X have cycles with respect to the directions $\mathbf{a}^1, ..., \mathbf{a}^r$. A function $f : X \to \mathbb{R}$ belongs to the space $\mathcal{R}(\mathbf{a}^1, ..., \mathbf{a}^r; X)$ if and only if $G_p(f) = 0$ for any minimal cycle $p \subset X$.

2) Let X have no cycles. Then every function $f : X \to \mathbb{R}$ belongs to the space $\mathcal{R}(\mathbf{a}^1, ..., \mathbf{a}^r; X)$.

Corollary 2. *$\mathcal{R}(\mathbf{a}^1, ..., \mathbf{a}^r; X)= T(X)$ if and only if X has no cycles with respect to the directions $\mathbf{a}^1, ..., \mathbf{a}^r$.*

Note that solutions to problems 1 and 2 are given correspondingly by corollaries 1 and 2. Although it is not always easy to find all cycles of a given set X and even to know if X possesses a single cycle at least, corollaries 1 and 2 carry more practical than theoretical character in them. Particular cases of problems 1 and 2 are evidence in favor of our opinion. For example, for the problem of representation by sums of two ridge functions, the picture of cycles is completely describable (see the beginning of this section). A geometric description of cycles with respect to 3 and more directions is quite complicated and inquires deep techniques from geometry and graph theory. This is not within the aim of our study.

From the last corollary, it follows that if representation by sums of ridge functions with fixed directions $\mathbf{a}^1, ..., \mathbf{a}^r$ is valid in the class of continuous functions (or in the class of bounded functions), then such representation is valid in the class of all functions. For a rigid mathematical formulation of this result, let us introduce the notation:

$$\mathcal{R}_c(\mathbf{a}^1, ..., \mathbf{a}^r; X) = \left\{ \sum_{i=1}^{r} g_i(\mathbf{a}^i \cdot \mathbf{x}),\ \mathbf{x} \in X,\ g_i(\mathbf{a}^i \cdot \mathbf{x}) \in C(X),\ i = 1, ..., r \right\},$$

$$\mathcal{R}_b(\mathbf{a}^1, ..., \mathbf{a}^r; X) = \left\{ \sum_{i=1}^{r} g_i(\mathbf{a}^i \cdot \mathbf{x}),\ \mathbf{x} \in X,\ g_i(\mathbf{a}^i \cdot \mathbf{x}) \in B(X),\ i = 1, ..., r \right\}.$$

Here $C(X)$ and $B(X)$ denote the spaces of continuous and bounded functions, respectively, defined on $X \subset \mathbb{R}^d$ (for the first space, the set X is supposed to be compact). It follows from the results of Sternfeld [64,65] that the equality $\mathcal{R}_c(\mathbf{a}^1,...,\mathbf{a}^r;X)=C(X)$ implies the equality $\mathcal{R}_b(\mathbf{a}^1,...,\mathbf{a}^r;X)=B(X)$ (see the next section). In other words, if every continuous function is represented by sums of ridge functions (with fixed directions !), then every bounded function also enjoys such representation (naturally, with bounded summands). Corollaries 1 and 2 allows us to obtain the following result:

Corollary 3. *Let X be a compact subset of $\mathbb{R}^d$ and $\mathbf{a}^1,...,\mathbf{a}^r$ be given directions in $\mathbb{R}^d\backslash\{\mathbf{0}\}$. If $\mathcal{R}_c(\mathbf{a}^1,...,\mathbf{a}^r;X)=C(X)$, then $\mathcal{R}(\mathbf{a}^1,...,\mathbf{a}^r;X)=T(X)$.*

Proof. If every continuous function defined on $X\subset\mathbb{R}^d$ is represented by sums of ridge functions with the directions $\mathbf{a}^1,...,\mathbf{a}^r$, then it can be shown by applying the same idea (as in the proof of Theorem 2) that the set X has no cycles with respect to the given directions. Only, because of continuity, Urysohn's great lemma should be taken into account. That is, it should be taken into account that, by assuming the existence of a cycle $p_0=\{x_1,...,x_n\}$ with an associated vector $\lambda_0=(\lambda_1,...,\lambda_n)$, we can deduce from Urysohn's great lemma the existence of *a continuous function* $u:X\to\mathbb{R}$ satisfying:

1) $u(x_i)=1$, for indices i such that $\lambda_i>0$,
2) $u(x_j)=-1$, for indices j such that $\lambda_j<0$,
3) $-1<u(x)<1$, for all $x\in X\backslash p_0$.

These properties would mean that $G_{p_0,\lambda_0}(u)\neq 0\Longrightarrow u\notin\mathcal{R}_c(\mathbf{a}^1,...,\mathbf{a}^r;X)\Longrightarrow$ $\mathcal{R}_c(\mathbf{a}^1,...,\mathbf{a}^r;X)\neq C(X)$.

But if X has no cycles with respect to the directions $\mathbf{a}^1,...,\mathbf{a}^r$, then by Corollary 2, $\mathcal{R}(\mathbf{a}^1,...,\mathbf{a}^r;X)=T(X)$. ◀

Let us now give some examples of sets over which the representation by linear combinations of ridge functions is possible.

(1) Let $r=2$ and X be the union of two parallel lines not perpendicular to the given directions $\mathbf{a}^1$ and $\mathbf{a}^2$. Then X has no cycles with respect to $\{\mathbf{a}^1,\mathbf{a}^2\}$. Therefore, by Corollary 2, $\mathcal{R}\left(\mathbf{a}^1,\mathbf{a}^2;X\right)=T(X)$.

(2) Let $r=2$, $\mathbf{a}^1=(1,1)$, $\mathbf{a}^2=(1,-1)$ and X be the graph of the function $y=\arcsin(\sin x)$. Then X has no cycles and hence $\mathcal{R}\left(\mathbf{a}^1,\mathbf{a}^2;X\right)=T(X)$.

(3) Let's now assume we are given r directions $\{\mathbf{a}^j\}_{j=1}^r$ and $r+1$ points $\{\mathbf{x}^i\}_{i=1}^{r+1}\subset\mathbb{R}^d$ such that

$$\begin{aligned}
\mathbf{a}^1\cdot\mathbf{x}^i &= \mathbf{a}^1\cdot\mathbf{x}^j\neq\mathbf{a}^1\cdot\mathbf{x}^2,\ \text{ for } 1\le i,j\le r+1,\ i,j\neq 2,\\
\mathbf{a}^2\cdot\mathbf{x}^i &= \mathbf{a}^2\cdot\mathbf{x}^j\neq\mathbf{a}^2\cdot\mathbf{x}^3,\ \text{ for } 1\le i,j\le r+1,\ i,j\neq 3,\\
&\quad\dots\dots\dots\dots\dots\dots\\
\mathbf{a}^r\cdot\mathbf{x}^i &= \mathbf{a}^r\cdot\mathbf{x}^j\neq\mathbf{a}^r\cdot\mathbf{x}^{r+1},\ \text{ for } 1\le i,j\le r.
\end{aligned}$$

The simplest data realizing these equations are the basis directions in $\mathbb{R}^d$ and the points $(0,0,...,0)$, $(1,0,...,0)$, $(0,1,...,0)$,..., $(0,0,...,1)$. From the first equation we obtain that $\mathbf{x}^2$ cannot be a point of any cycle in $X = \{\mathbf{x}^1, ..., \mathbf{x}^{r+1}\}$. Sequentially, from the second, third, ..., r-th equations it follows that the points $\mathbf{x}^3, \mathbf{x}^4, ..., \mathbf{x}^{r+1}$ also cannot be points of cycles in X, respectively. Thus, the set X does not contain cycles at all. By Corollary 2, $\mathcal{R}\left(\mathbf{a}^1, ..., \mathbf{a}^r; X\right) = T(X)$.

(4) Let's assume we are given directions $\{\mathbf{a}^j\}_{j=1}^r$ and a curve γ in $\mathbb{R}^d$ such that for any $c \in \mathbb{R}$, γ has at most one common point with at least one of the hyperplanes $\mathbf{a}^j \cdot \mathbf{x} = c$, $j = 1, ..., r$. By Definition 1, the curve γ has no cycles and hence $\mathcal{R}\left(\mathbf{a}^1, ..., \mathbf{a}^r; \gamma\right) = T(\gamma)$.

It follows from Corollary 2 that a set $\{\mathbf{x}^i\}_{i=1}^k$ has the *NI*-property if and only if $\{\mathbf{x}^i\}_{i=1}^k$ contains a cycle with respect to the functions $h_i = \mathbf{a}^i \cdot \mathbf{x}$, $i = 1, ..., r$ (or, simply, to the directions $\mathbf{a}^i$, $i = 1, ..., r$) and the *MNI*-property if and only if the set $\{\mathbf{x}^i\}_{i=1}^k$ itself is a minimal cycle with respect to the given directions. Taking into account this argument and Definitions 1 and 2, we obtain that the set $\{\mathbf{x}^i\}_{i=1}^k$ has the *NI*-property if and only if there is a vector $\mathbf{m} = (m_1, ..., m_k) \in \mathbb{Z}^k \backslash \{\mathbf{0}\}$ such that

$$\sum_{j=1}^{k} m_j g(\mathbf{a}^i \cdot \mathbf{x}^j) = 0,$$

for $i = 1, ..., r$ and all functions $g : \mathbb{R} \to \mathbb{R}$. This set has the *MNI*-property if and only if the vector $\mathbf{m}$ has the additional properties: it is unique up to multiplication by a constant and all its components are different from zero. This special consequence of Corollary 2 was proved in [4].

3. Best approximating ridge functions and some density questions

The approximation problem considered in this section is to approximate a continuous multivariate function $f(\mathbf{x}) = f(x_1, ..., x_d)$ by sums of two ridge functions in the uniform norm. We give a necessary and sufficient condition for a sum of two ridge functions to be a best approximation to $f(\mathbf{x})$. This main result is next used in a special case to obtain an explicit formula for the approximation error and to construct one best approximation. The problem of well approximation by such sums is also considered.

Consider the following set of sums of ridge functions

$$\mathcal{R} = \mathcal{R}(\mathbf{a},\mathbf{b}) = \{g_1(\mathbf{a}\cdot\mathbf{x}) + g_2(\mathbf{b}\cdot\mathbf{x}) : g_i \in C(\mathbb{R}), i = 1, 2\}.$$

That is, we fix directions $\mathbf{a}$ and $\mathbf{b}$ and consider linear combinations of ridge functions with these directions.

Let $f(\mathbf{x})$ be a given continuous function on some compact subset Q of $\mathbb{R}^d$. We want to find conditions that are necessary and sufficient for a function $g_0 \in \mathcal{R}(\mathbf{a},\mathbf{b})$ to be an extremal element (or a best approximation) to f. In other words, we want to characterize such sums $g_0(\mathbf{x}) = g_1(\mathbf{a}\cdot\mathbf{x}) + g_2(\mathbf{b}\cdot\mathbf{x})$ of ridge functions that

$$\|f - g_0\| = \max_{\mathbf{x}\in Q}|f(\mathbf{x}) - g_0(\mathbf{x})| = E(f),$$

where

$$E(f) = E(f,\mathcal{R}) \overset{def}{=} \inf_{g\in\ \mathcal{R}(\mathbf{a},\mathbf{b})}\|f - g\|,$$

is the error in approximating from $\mathcal{R}(\mathbf{a},\mathbf{b})$. The other related problem is how to construct these sums of ridge functions. We also want to know if we can approximate well, i.e. for which compact sets Q, $\mathcal{R}(\mathbf{a},\mathbf{b})$ is dense in $C(Q)$ in the topology of uniform convergence. It should be remarked that solutions to these problems may be useful in connection with the study of partial differential equations. For example, assume that (a_1, b_1) and (a_2, b_2) are linearly independent vectors in $\mathbb{R}^2$. Then the general solution of the homogeneous partial differential equation

$$\left(a_1\frac{\partial}{\partial x} + b_1\frac{\partial}{\partial y}\right)\left(a_2\frac{\partial}{\partial x} + b_2\frac{\partial}{\partial y}\right)u(x,y) = 0, \tag{10}$$

are all functions of the form

$$u(x,y) = g_1(b_1x - a_1y) + g_2(b_2x - a_2y), \tag{11}$$

for arbitrary g_1 and g_2. In [19], Golitschek and Light described an algorithm that computes the error of approximation of a continuous real – valued function $f(x,y)$ by solutions of equation (10), provided that $a_1 = b_2 = 1, a_2 = b_1 = 0$. Using Theorem 3 below, one can characterize those solutions (11) that are extremal to a given function $f(x,y)$. For certain class of functions $f(x,y)$, one can also easily calculate the approximation error and construct one extremal solution (see Theorems 4 and 5).

The problem of approximating by functions from the set $\mathcal{R}(\mathbf{a},\mathbf{b})$ arises in other contexts, too. Buck [5] studied the classical functional equation: given $\beta(t) \in C[0,1]$, $0 \le \beta(t) \le 1$, for which $u \in C[0,1]$ does there exist $\varphi \in C[0,1]$ such that

$$\varphi(t) = \varphi(\beta(t)) + u(t)?$$

He proved that the set of all u satisfying this condition is dense in the set

$$\{v \in C[0,1]:\ v(t) = 0 \text{ whenever } \beta(t) = t\},$$

if and only if $\mathcal{R}(\mathbf{a},\mathbf{b})$ with the unit directions $\mathbf{a} = (1;0)$ and $\mathbf{b} = (0,1)$ is dense in $C(K)$, where $K = \{(x,y) : y = x \text{ or } y = \beta(x),\ 0 \le x \le 1\}$.

Note that if the space dimension $d = 2$, $\mathbf{a}$ and $\mathbf{b}$ are the unit directions, then the functions $g_1(\mathbf{a}\cdot\mathbf{x})$ and $g_2(\mathbf{b}\cdot\mathbf{x})$ are univariate. Thus the approximation of a bivariate function by sums of univariate functions is a special case of the approximation problem considered here.

Although there are enough reasons to consider approximation problems associated with the set $\mathcal{R}(\mathbf{a},\mathbf{b})$ in an independent way, one may ask why sums of only two ridge functions are considered instead of sums with an arbitrary number of terms. We will try to answer this fair question at the end of this section.

Definition 3. *[33]. A finite or infinite ordered set $p = (\mathbf{p}_1, \mathbf{p}_2, ...) \subset Q$ with $\mathbf{p}_i \neq \mathbf{p}_{i+1}$, and either $\mathbf{a} \cdot \mathbf{p}_1 = \mathbf{a} \cdot \mathbf{p}_2, \mathbf{b} \cdot \mathbf{p}_2 = \mathbf{b} \cdot \mathbf{p}_3, \mathbf{a} \cdot \mathbf{p}_3 = \mathbf{a} \cdot \mathbf{p}_4, ...$ or $\mathbf{b} \cdot \mathbf{p}_1 = \mathbf{b} \cdot \mathbf{p}_2,\ \mathbf{a} \cdot \mathbf{p}_2 = \mathbf{a} \cdot \mathbf{p}_3, \mathbf{b} \cdot \mathbf{p}_3 = \mathbf{b} \cdot \mathbf{p}_4, ...$is called a path with respect to the directions $\mathbf{a}$ and $\mathbf{b}$.*

This notion (in two-dimensional case) was introduced by Braess and Pinkus [4]. They showed that paths give geometric means of deciding if a set of points $\left\{\mathbf{x}^i\right\}_{i=1}^m \subset \mathbb{R}^2$ has the *NI*-property (for this terminology see the previous section).

If $\mathbf{a}$ and $\mathbf{b}$ are the unit vectors in $\mathbb{R}^2$, then Definition 3 defines an ordinary path (or a bolt of lightning in a number of papers, e.g. [17,19, 34,35,41,42,52]). It is well known that the idea of ordinary paths, first introduced by Diliberto and Straus [11], played significant role in many problems of the approximation of bivariate functions by sums of univariate functions (see, for example, [11, 14,17,19, 34,35, 41,42, 52]). Paths with respect to two directions are used also in neural network theory (see [24,25]).

For the sake of brevity, we use the term "path" instead of the expression "path with respect to the directions $\mathbf{a}$ and $\mathbf{b}$".

The length of a path is the number of its points. A single point is a path of the unit length. A finite path $(\mathbf{p}_1, \mathbf{p}_2, ..., \mathbf{p}_{2n})$ is said to be closed if $(\mathbf{p}_1, \mathbf{p}_2, ..., \mathbf{p}_{2n}, \mathbf{p}_1)$ is a path.

We associate each closed path $p = (\mathbf{p}_1, \mathbf{p}_2, ..., \mathbf{p}_{2n})$ with the functional

$$G_p(f) = \frac{1}{2n} \sum_{k=1}^{2n} (-1)^{k+1} f(\mathbf{p}_k).$$

This functional has the following obvious properties:

(a) If $g \in \mathcal{R}(\mathbf{a},\mathbf{b})$, then $G_p(g) = 0$.

(b) $\|G_p\| \leq 1$ and if $p_i \neq p_j$ for all $i \neq j$, $1 \leq i, j \leq 2n$, then $\|G_p\| = 1$.

Lemma 3. *[33]. Let a compact set Q have closed paths. Then*

$$\sup_{p \subset Q} |G_p(f)| \leq E(f), \tag{12}$$

where the sup is taken over all closed paths. Moreover, inequality (12) is sharp, i.e. there exist functions for which (12) turns into equality.

Proof. Let p be a closed path of Q and g be any function from $\mathcal{R}(\mathbf{a},\mathbf{b})$. Then by the linearity of G_p and properties (a) and (b):

$$|G_p(f)| = |G_p(f-g)| \leq \|f-g\|. \tag{13}$$

Since the left-hand and the right-hand sides of (13) do not depend on g and p respectively, it follows from (13) that

$$\sup_{p\subset Q}|G_p(f)| \leq \inf_{\mathcal{R}(\mathbf{a},\mathbf{b})} \|f-g\|. \tag{14}$$

Now we prove the sharpness of (12). By assumption Q has closed paths. Then Q has closed paths $p' = (\mathbf{p}_1', ..., \mathbf{p}_{2m}')$ such that all points $\mathbf{p}_1, ..., \mathbf{p}_{2m}$ are distinct. In fact, such special paths can be obtained from any closed path $p = (\mathbf{p}_1, ..., \mathbf{p}_{2n})$ by the following simple algorithm: if the points of the path p are not all distinct, let i and $k > 0$ be the minimal indices such that $\mathbf{p}_i = \mathbf{p}_{i+2k}$; delete from p the subsequence $\mathbf{p}_{i+1}, ..., \mathbf{p}_{i+2k}$ and call p the obtained path; repeat the above step until all points of p are all distinct; set $p' := p$. On the other hand, there exist continuous functions $h = h(\mathbf{x})$ on Q such that $h(\mathbf{p}_i') = 1$, $i = 1, 3, ..., 2m-1$, $h(\mathbf{p}_i') = -1$, $i = 2, 4, ..., 2m$ and $-1 < h(\mathbf{x}) < 1$ elsewhere. For such functions we have

$$G_{p'}(h) = \|h\| = 1, \tag{15}$$

and

$$E(h) \leq \|h\|, \tag{16}$$

where the last inequality follows from the fact that $0 \in \mathcal{R}(\mathbf{a},\mathbf{b})$. From (14)-(16) it follows that

$$\sup_{p\subset Q}|G_p(h)| = E(h).$$

◄

Lemma 4. *[33]. Let Q be a convex compact subset of $\mathbb{R}^d$, $f(\mathbf{x}) \in C(Q)$. For a vector $\mathbf{e} \in \mathbb{R}^d\backslash\{\mathbf{0}\}$ and a real number t set*

$$Q_t = \{\mathbf{x} \in Q : \mathbf{e}\cdot\mathbf{x} = t\}, \quad T_h = \{t \in \mathbb{R} : Q_t \neq \emptyset\}.$$

Then the functions

$$g_1(t) = \max_{\mathbf{x}\in Q_t} f(\mathbf{x}), \quad t \in T_h \quad \text{and} \quad g_2(t) = \min_{\mathbf{x}\in Q_t} f(\mathbf{x}), \quad t \in T_h,$$

are defined and continuous on T_h.

The proof of this lemma is not difficult and can be obtained by the well-known elementary methods of mathematical analysis.

Definition 4. *[33]. A finite or infinite path $(\mathbf{p}_1, \mathbf{p}_2, ...)$ is said to be extremal for a function $u(\mathbf{x}) \in C(Q)$ if $u(\mathbf{p}_i) = (-1)^i \|u\|, i = 1, 2, ...$ or $u(\mathbf{p}_i) = (-1)^{i+1} \|u\|$, $i = 1, 2, ...$*

Theorem 3. *[33]. Let $Q \subset \mathbb{R}^d$ be a convex compact set with the property: for any path $q = (\mathbf{q}_1, \mathbf{q}_2, ..., \mathbf{q}_n) \subset Q$ there exist points $\mathbf{q}_{n+1}, \mathbf{q}_{n+2}, ..., \mathbf{q}_{n+s} \in Q$ such that $(\mathbf{q}_1, \mathbf{q}_2, ..., \mathbf{q}_{n+s})$ is a closed path and s is not more than some positive integer N_0 independent of q. Then a necessary and sufficient condition for a function $g_0 \in \mathcal{R}(\mathbf{a},\mathbf{b})$ to be extremal for the given function $f(\mathbf{x}) \in C(Q)$ is the existence of a closed or infinite path $l = (\mathbf{p}_1, \mathbf{p}_2, ...)$ extremal for the function $f_1(\mathbf{x}) = f(\mathbf{x}) - g_0(\mathbf{x})$.*

It should be remarked that the hypothesis on the convex compact set Q "for any path $q = (\mathbf{q}_1, \mathbf{q}_2, ..., \mathbf{q}_n) \subset Q$ ... independent of q" strongly depends on the fixed directions $\mathbf{a}$ and $\mathbf{b}$. For example, in the familiar case of a square $S \subset \mathbb{R}^2$ there are many directions which are not allowed. If it is possible to reach a corner of the square with not more than one of the two directions orthogonal to $\mathbf{a}$ and $\mathbf{b}$ respectively (we don't differentiate between directions $\mathbf{c}$ and $-\mathbf{c}$), the triple $(S, \mathbf{a}, \mathbf{b})$ does not satisfy the hypothesis of the theorem. Here are simple examples: Let $S = [0;1]^2$, $\mathbf{a} = (1;0)$, $\mathbf{b} = (1;1)$. Then the ordered set $\{(0;1),(1;0),(1;1)\}$ is a path in S which can not be made closed. In this case, $(1;1)$ is not reached with the direction orthogonal to $\mathbf{b}$. Let now $\mathbf{a} = \left(1; \frac{1}{2}\right)$, $\mathbf{b} = (1;1)$. Then the corner $(1;1)$ is reached with none of the directions orthogonal to $\mathbf{a}$ and $\mathbf{b}$ respectively. In this case, for any positive integer N_0 and any point $\mathbf{q}_0$ in S one can chose a point $\mathbf{q}_1 \in S$ from a sufficiently small neighborhood of the corner $(1;1)$ so that any path containing $\mathbf{q}_0$ and $\mathbf{q}_1$ has the length more than N_0. These examples and a little geometry show that if a compact convex set $Q \subset \mathbb{R}^2$ satisfies the hypothesis of the theorem, then any point in the boundary of Q must be reached with each of the two directions orthogonal to $\mathbf{a}$ and $\mathbf{b}$ respectively. If $Q \subset \mathbb{R}^d, \mathbf{a}, \mathbf{b} \in \mathbb{R}^d \backslash \{\mathbf{0}\}$, $d > 2$, there are many directions orthogonal to $\mathbf{a}$ and $\mathbf{b}$. In this case, the hypothesis of the theorem requires that any point in the boundary of Q should be reached with at least two directions orthogonal to $\mathbf{a}$ and $\mathbf{b}$, respectively.

Proof. **Necessity.** Let $g_0 = g_{1,0}(\mathbf{a}{\cdot}\mathbf{x}) + g_{2,0}(\mathbf{b}{\cdot}\mathbf{x})$ be an extremal element from $\mathcal{R}(\mathbf{a},\mathbf{b})$ to f. We must show that if there is not a closed path extremal for f_1, then there exists a path extremal for f_1 with the infinite length (number of points). Suppose the contrary. Suppose that there exists a positive integer N such that the length of each path extremal for f_1 is not more than N. Set the following functions:

$$f_n = f_{n-1} - g_{1,n-1} - g_{2,n-1}, \quad n = 2, 3, ...,$$

where

$$g_{1,n-1} = g_{1,n-1}(\mathbf{a}{\cdot}\mathbf{x}) = \frac{1}{2}\left(\max_{\substack{\mathbf{y}\in Q \\ \mathbf{a}\cdot\mathbf{y}=\mathbf{a}\cdot\mathbf{x}}} f_{n-1}(\mathbf{y}) + \min_{\substack{\mathbf{y}\in Q \\ \mathbf{a}\cdot\mathbf{y}=\mathbf{a}\cdot\mathbf{x}}} f_{n-1}(\mathbf{y}) \right),$$

$$g_{2,n-1} = g_{2,n-1}(\mathbf{b}{\cdot}\mathbf{x}) = \frac{1}{2}\left(\max_{\substack{\mathbf{y}\in Q \\ \mathbf{b}\cdot\mathbf{y}=\mathbf{b}\cdot\mathbf{x}}} (f_{n-1}(\mathbf{y}) - g_{1,n-1}(\mathbf{a}{\cdot}\mathbf{y})) \right.$$

$$\left. + \min_{\substack{\mathbf{y}\in Q \\ \mathbf{b}\cdot\mathbf{y}=\mathbf{b}\cdot\mathbf{x}}} (f_{n-1}(\mathbf{y}) - g_{1,n-1}(\mathbf{a}{\cdot}\mathbf{y})) \right).$$

By lemma 4, all the functions $f_n(\mathbf{x})$, $n = 2, 3, ...$, are continuous on Q. By assumption, g_0 is a best approximation to f. Hence, $\|f_1\| = E(f)$. Now we show that $\|f_2\| = E(f)$. Indeed, for any $\mathbf{x} \in Q$:

$$f_1(\mathbf{x}) - g_{1,1}(\mathbf{a}\cdot\mathbf{x}) \leq \frac{1}{2}\left(\max_{\substack{\mathbf{y}\in Q\\ \mathbf{a}\cdot\mathbf{y}=\mathbf{a}\cdot\mathbf{x}}} f_1(\mathbf{y}) - \min_{\substack{\mathbf{y}\in Q\\ \mathbf{a}\cdot\mathbf{y}=\mathbf{a}\cdot\mathbf{x}}} f_1(\mathbf{y})\right) \leq E(f), \tag{17}$$

and

$$f_1(\mathbf{x}) - g_{1,1}(\mathbf{a}\cdot\mathbf{x}) \geq \frac{1}{2}\left(\min_{\substack{\mathbf{y}\in Q\\ \mathbf{a}\cdot\mathbf{y}=\mathbf{a}\cdot\mathbf{x}}} f_1(\mathbf{y}) - \max_{\substack{\mathbf{y}\in Q\\ \mathbf{a}\cdot\mathbf{y}=\mathbf{a}\cdot\mathbf{x}}} f_1(\mathbf{y})\right) \geq -E(f). \tag{18}$$

Using the definition of $g_{2,1}(\mathbf{b}\cdot\mathbf{x})$, for any $\mathbf{x} \in Q$ we have

$$f_1(\mathbf{x}) - g_{1,1}(\mathbf{a}\cdot\mathbf{x}) - g_{2,1}(\mathbf{b}\cdot\mathbf{x})$$

$$\leq \frac{1}{2}\left(\max_{\substack{\mathbf{y}\in Q\\ \mathbf{b}\cdot\mathbf{y}=\mathbf{b}\cdot\mathbf{x}}} (f_1(\mathbf{y}) - g_{1,1}(\mathbf{a}\cdot\mathbf{y})) - \min_{\substack{\mathbf{y}\in Q\\ \mathbf{b}\cdot\mathbf{y}=\mathbf{b}\cdot\mathbf{x}}} (f_1(\mathbf{y}) - g_{1,1}(\mathbf{a}\cdot\mathbf{y}))\right),$$

and

$$f_1(\mathbf{x}) - g_{1,1}(\mathbf{a}\cdot\mathbf{x}) - g_{2,1}(\mathbf{b}\cdot\mathbf{x})$$

$$\leq \frac{1}{2}\left(\min_{\substack{\mathbf{y}\in Q\\ \mathbf{b}\cdot\mathbf{y}=\mathbf{b}\cdot\mathbf{x}}} (f_1(\mathbf{y}) - g_{1,1}(\mathbf{a}\cdot\mathbf{y})) - \max_{\substack{\mathbf{y}\in Q\\ \mathbf{b}\cdot\mathbf{y}=\mathbf{b}\cdot\mathbf{x}}} (f_1(\mathbf{y}) - g_{1,1}(\mathbf{a}\cdot\mathbf{y}))\right).$$

Using (17) and (18) in the last two inequalities, we obtain that for any $\mathbf{x} \in Q$

$$-E(f) \leq f_2(\mathbf{x}) = f_1(\mathbf{x}) - g_{1,1}(\mathbf{a}\cdot\mathbf{x}) - g_{2,1}(\mathbf{b}\cdot\mathbf{x}) \leq E(f).$$

Therefore

$$\|f_2\| \leq E(f). \tag{19}$$

Since $f_2(\mathbf{x}) - f(\mathbf{x})$ belongs to $\mathcal{R}(\mathbf{a},\mathbf{b})$, we deduce from (19) that

$$\|f_2\| = E(f).$$

By the same way, one can show that $\|f_3\| = E(f)$, $\|f_4\| = E(f)$, and so on. Thus we can write

$$\|f_n\| = E(f), \text{ for any } n.$$

Let us now prove the implications

$$f_1(\mathbf{p}_0) < E(f) \Rightarrow f_2(\mathbf{p}_0) < E(f), \tag{20}$$

and

$$f_1(\mathbf{p}_0) > -E(f) \Rightarrow f_2(\mathbf{p}_0) > -E(f), \tag{21}$$

where $\mathbf{p}_0 \in Q$. First, we are going to prove the implication

$$f_1(\mathbf{p}_0) < E(f) \Rightarrow f_1(\mathbf{p}_0) - g_{1,1}(\mathbf{a} \cdot \mathbf{p}_0) < E(f). \tag{22}$$

There are two possible cases.

1) $\max\limits_{\substack{\mathbf{y}\in Q \\ \mathbf{a}\cdot\mathbf{y}=\mathbf{a}\cdot\mathbf{p}_0}} f_1(\mathbf{y}) = E(f)$ and $\min\limits_{\substack{\mathbf{y}\in Q \\ \mathbf{a}\cdot\mathbf{y}=\mathbf{a}\cdot\mathbf{p}_0}} f_1(\mathbf{y}) = -E(f)$. In this case, $g_{1,1}(\mathbf{a} \cdot \mathbf{p}_0) = 0$. Hence,

$$f_1(\mathbf{p}_0) - g_{1,1}(\mathbf{a} \cdot \mathbf{p}_0) < E(f).$$

2) $\max\limits_{\substack{\mathbf{y}\in Q \\ \mathbf{a}\cdot\mathbf{y}=\mathbf{a}\cdot\mathbf{p}_0}} f_1(\mathbf{y}) = E(f) - \varepsilon_1$ and $\min\limits_{\substack{\mathbf{y}\in Q \\ \mathbf{a}\cdot\mathbf{y}=\mathbf{a}\cdot\mathbf{p}_0}} f_1(\mathbf{y}) = -E(f) + \varepsilon_2$,
where ε_1, ε_2 are nonnegative real numbers with the sum $\varepsilon_1 + \varepsilon_2 \neq 0$. In this case,

$$\begin{aligned} f_1(\mathbf{p}_0) - g_{1,1}(\mathbf{a}\cdot\mathbf{p}_0) &\leq \max_{\substack{\mathbf{y}\in Q \\ \mathbf{a}\cdot\mathbf{y}=\mathbf{a}\cdot\mathbf{p}_0}} f_1(\mathbf{y}) - g_{1,1}(\mathbf{a}\cdot\mathbf{p}_0) = \\ &= \frac{1}{2}\left(\max_{\substack{\mathbf{y}\in Q \\ \mathbf{a}\cdot\mathbf{y}=\mathbf{a}\cdot\mathbf{p}_0}} f_1(\mathbf{y}) - \min_{\substack{\mathbf{y}\in Q \\ \mathbf{a}\cdot\mathbf{y}=\mathbf{a}\cdot\mathbf{p}_0}} f_1(\mathbf{y}) \right) = \\ &= E(f) - \frac{\varepsilon_1 + \varepsilon_2}{2} < E(f). \end{aligned}$$

Thus we have proved (22). Using this method, we can also prove that

$$f_1(\mathbf{p}_0) - g_{1,1}(\mathbf{a}\cdot\mathbf{p}_0) < E(f) \Rightarrow f_1(\mathbf{p}_0) - g_{1,1}(\mathbf{a}\cdot\mathbf{p}_0) - g_{2,1}(\mathbf{b}\cdot\mathbf{p}_0) < E(f). \tag{23}$$

Now (20) follows from (22) and (23). By the same way we can prove (21). It follows from implications (20) and (21) that if $f_2(\mathbf{p}_0) = E(f)$, then $f_1(\mathbf{p}_0) = E(f)$ and if $f_2(\mathbf{p}_0) = -E(f)$, then $f_1(\mathbf{p}_0) = -E(f)$. This simply means that each path extremal for f_2 will be extremal for f_1.

Now we show that if any path extremal for f_1 has the length not more than N, then any path extremal for f_2 has the length not more than $N - 1$. Suppose the contrary. Suppose that there is a path extremal for f_2 with the length equal to N. Denote it by $q = (\mathbf{q}_1, \mathbf{q}_2, ..., \mathbf{q}_N)$. Without loss of generality we may assume that $\mathbf{b} \cdot \mathbf{q}_{N-1} = \mathbf{b} \cdot \mathbf{q}_N$. As it has been shown above, the path q is also extremal for f_1. Assume that $f_1(\mathbf{q}_N) = E(f)$. Then there is not any $\mathbf{q}_0 \in Q$ such that $\mathbf{q}_0 \neq \mathbf{q}_N$, $\mathbf{a} \cdot \mathbf{q}_0 = \mathbf{a} \cdot \mathbf{q}_N$ and $f_1(\mathbf{q}_0) = -E(f)$. Indeed, if there was such $\mathbf{q}_0$ and $\mathbf{q}_0 \notin q$, the path $(\mathbf{q}_1, \mathbf{q}_2, ..., \mathbf{q}_N, \mathbf{q}_0)$ would be extremal for f_1. But this would contradict our assumption that any path extremal for f_1 has the length not more than N. Besides, if there was such $\mathbf{q}_0$ and $\mathbf{q}_0 \in q$, we could form some closed path extremal for f_1. This also would contradict our assumption that there does not exist a closed path extremal for f_1.

Hence,

$$\max_{\substack{\mathbf{y}\in Q \\ \mathbf{a}\cdot\mathbf{y}=\mathbf{a}\cdot\mathbf{q}_N}} f_1(\mathbf{y}) = E(f), \qquad \min_{\substack{\mathbf{y}\in Q \\ \mathbf{a}\cdot\mathbf{y}=\mathbf{a}\cdot\mathbf{q}_N}} f_1(\mathbf{y}) > -E(f).$$

Therefore

$$|f_1(\mathbf{q}_N) - g_{1,1}(\mathbf{a}\cdot\mathbf{q}_N)| < E(f).$$

From the last inequality it is easy to obtain that (see the proof of implications (20) and (21))

$$|f_2(\mathbf{q}_N)| < E(f).$$

This means, on the contrary to our assumption, that the path $(\mathbf{q}_1, \mathbf{q}_2, ..., \mathbf{q}_N)$ can not be extremal for f_2. Hence, any path extremal for f_2 has the length not more than $N-1$.

By the same way, it can be shown that any path extremal for f_3 has the length not more than $N-2$, any path extremal for f_4 has the length not more than $N-3$ and so on. Finally, we will obtain that there is not a path extremal for f_{N+1}. Hence, there is not a point $\mathbf{p}_0 \in Q$ such that $|f_{N+1}(\mathbf{p}_0)| = \|f_{N+1}\|$. But by lemma 4, all the functions $f_2, f_3, ..., f_{N+1}$ are continuous on the compact set Q; hence, the norm $\|f_{N+1}\|$ must be attained. This contradiction means that there exists a path extremal for f_1 with the infinite length.

Sufficiency. Let a path $l = (\mathbf{p}_1, \mathbf{p}_2, ..., \mathbf{p}_{2n})$ be closed and extremal for f_1. Then

$$|G_l(f)| = \|f - g_0\|. \tag{24}$$

By Lemma 3

$$|G_l(f)| \leq E(f). \tag{25}$$

It follows from (24) and (25) that g_0 is a best approximation.

Let now a path $l = (\mathbf{p}_1, \mathbf{p}_2, ..., \mathbf{p}_n, ...)$ be infinite and extremal for f_1. Consider the sequence $l_n = (\mathbf{p}_1, \mathbf{p}_2, ..., \mathbf{p}_n)$, $n = 1, 2, ...$, of finite paths. By the property of the set Q defined in theorem's statement, for each l_n there exists a closed path $l_n^{m_n} = (\mathbf{p}_1, \mathbf{p}_2, ..., \mathbf{p}_n, \mathbf{q}_{n+1}, ..., \mathbf{q}_{n+m_n})$, where $m_n \leq N_0$. Then for any positive integer n:

$$\left|G_{l_n^{m_n}}(f)\right| = \left|G_{l_n^{m_n}}(f - g_0)\right| \leq \frac{n\,\|f - g_0\| + m_n\,\|f - g_0\|}{n + m_n} = \|f - g_0\|, \tag{26}$$

and

$$\left|G_{l_n^{m_n}}(f)\right| \geq \frac{n\,\|f - g_0\| - m_n\,\|f - g_0\|}{n + m_n} = \frac{n - m_n}{n + m_n}\,\|f - g_0\|. \tag{27}$$

It follows from (26) and (27) that

$$\sup_{l_n^{m_n}} \left|G_{l_n^{m_n}}(f)\right| = \|f - g_0\|. \tag{28}$$

Now we deduce from (28) and Lemma 3 that

$$\|f - g_0\| \leq E(f).$$

Hence, g_0 is a best approximation.◄

It is well known that characterization theorems of this type are very essential in approximation theory. Chebyshev was the first to prove the like result for polynomial approximation. Khavinson [42] characterized extremal elements in the special case of the problem considered here. His case allows the approximation of a continuous bivariate function $f(x, y)$ by functions of the type $\varphi(x) + \psi(y)$. It should be noted that the techniques used in the proof of Theorem 3 are completely different from those used in [42].

Now we want to deal with the error of approximation. The value of the approximation error depends not only on the approximated function f but also on a geometrical structure of the set X. For example, if X has an interior point, then the error of approximation cannot equal to zero for a function $f \notin \mathcal{R}\left(\mathbf{a}^1, ..., \mathbf{a}^r\right)$ (see [46]). This fact gives rise to some problems on approximate or exact computations of the approximation error and algorithms for constructing best approximating ridge sums. The main difficulties arise when such problems are considered in continuous function spaces endowed with the uniform norm. In literature, there is one essential algorithm called Diliberto and Straus algorithm. The essence of this algorithm is the following. Let X be a compact subset of $\mathbb{R}^n$ and A_i be a best approximation operator from the space of continuous functions $C(X)$ to the subspace of ridge functions $G_i = \{g_i\left(\mathbf{a}^i \cdot \mathbf{x}\right) : \ g_i \in C(\mathbb{R}), \ \mathbf{x} \in X\}$, $i = 1, ..., r$.

That is, for each function $f \in C(X)$, the function $A_i f$ is a best approximation to f from G_i. Set

$$Tf = (I - A_r)(I - A_{r-1}) \cdots (I - A_1)f,$$

where I is the identity operator. It is clear that

$$Tf = f - g_1 - g_2 - \cdots - g_r,$$

where g_k is a best approximation from G_k to the function $f - g_1 - g_2 - \cdots - g_{k-1}$, $k = 1, ..., r$. Consider powers of the operator T: T^2, T^3 and so on. Is the sequence $\{T^n f\}_{n=1}^{\infty}$ convergent? In case of an affirmative answer, which function is the limit of $T^n f$,as $n \to \infty$? One may expect that the sequence $\{T^n f\}_{n=1}^{\infty}$ converges to $f - g^*$, where g^* is a best approximation from $\mathcal{R}\left(\mathbf{a}^1, ..., \mathbf{a}^r\right)$ to f. This conjecture was first formulated by Diliberto and Straus [11] in 1951 for the approximation of multivariate functions by sums of univariate functions (that is, sums of ridge functions with basic directions). But later it was shown by Aumann [1] that the algorithm may not converge at all for $r > 2$ (even for the sum of univariate functions). For $r = 2$, the sequence $\{T^n f\}_{n=1}^{\infty}$ converges to $f - g_0$, where g_0 is a best approximation from $\mathcal{R}\left(\mathbf{a}^1, \mathbf{a}^2\right)$ (see [58]). In general case, when $r > 2$ an algorithm for finding a best uniform approximation (or L_p-approximation, $p \neq 2$) from the space $\mathcal{R}\left(\mathbf{a}^1, ..., \mathbf{a}^r\right)$ to f is not yet known. In L_2 metric, the Diliberto and Straus algorithm converges to the desired resulting function for an arbitrary number of distinct directions (see [58]).

In [11], Diliberto and Straus also established a formula for the error in approximating bivariate functions by sums of univariate functions. Their formula contains the supremum over all closed ordinary paths. Although the formula is valid for all continuous functions, it is not easily calculable. Therefore, it does not give the desired effect if one is interested in the precise value of the approximation error. After this general result some authors started to seek easily calculable formulas for the approximation error by considering not the whole space, but some subsets of continuous functions (see, for example, [2,3,34,35,36,42,61]).These subsets were chosen so that they could provide precise and easy computation of the approximation error. Since the set of ridge functions contain univariate functions as its proper subset, one may ask for explicit formulas for the error in approximating by sums of ridge functions.

In this section, we see how with the use of Theorem 3 it is possible to find the error and an extremal element in approximating a continuous function by sums of ridge functions. We restrict ourselves to $\mathbb{R}^2$. To make the problem more precise, let Ω be a compact set in $\mathbb{R}^2$, $f(x_1,x_2) \in C(\Omega)$, $\mathbf{a} = (a_1, a_2), \mathbf{b} = (b_1, b_2)$ be linearly independent vectors. We want, in some conditions on f and Ω, to establish a formula for an easy and direct computation of the error in approximating from $\mathcal{R}(\mathbf{a},\mathbf{b})$.

Theorem 4. *Let*

$$\Omega = \left\{\mathbf{x} \in \mathbb{R}^2 : c_1 \le \mathbf{a}\cdot\mathbf{x} \le d_1, \quad c_2 \le \mathbf{b}\cdot\mathbf{x} \le d_2\right\},$$

where $c_1 < d_1$ and $c_2 < d_2$. Let a function $f(\mathbf{x}) \in C(\Omega)$ have the continuous partial derivatives $\frac{\partial^2 f}{\partial x_1^2}, \frac{\partial^2 f}{\partial x_1 \partial x_2}, \frac{\partial^2 f}{\partial x_2^2}$ and for any $\mathbf{x} \in \Omega$:

$$\frac{\partial^2 f}{\partial x_1 \partial x_2}(a_1 b_2 + a_2 b_1) - \frac{\partial^2 f}{\partial x_1^2} a_2 b_2 - \frac{\partial^2 f}{\partial x_2^2} a_1 b_1 \ge 0.$$

Then

$$E(f) = \frac{1}{4}\left(f_1(c_1, c_2) + f_1(d_1, d_2) - f_1(c_1, d_2) - f_1(d_1, c_2)\right),$$

where

$$f_1(y_1, y_2) = f\left(\frac{y_1 b_2 - y_2 a_2}{a_1 b_2 - a_2 b_1}, \frac{y_2 a_1 - y_1 b_1}{a_1 b_2 - a_2 b_1}\right). \tag{29}$$

Proof. Introduce the new variables

$$y_1 = a_1 x_1 + a_2 x_2, \quad y_2 = b_1 x_1 + b_2 x_2. \tag{30}$$

Since the vectors (a_1, a_2) and (b_1, b_2) are linearly independent, for any $(y_1, y_2) \in Y$, where $Y = [c_1, d_1] \times [c_2, d_2]$, there exists only one solution $(x_1, x_2) \in \Omega$ of the system (30). The coordinates of this solution are

$$x_1 = \frac{y_1 b_2 - y_2 a_2}{a_1 b_2 - a_2 b_1}, \qquad x_2 = \frac{y_2 a_1 - y_1 b_1}{a_1 b_2 - a_2 b_1}. \tag{31}$$

The linear transformation (31) transforms the function $f(x_1, x_2)$ to the function $f_1(y_1, y_2)$. Consider the approximation of $f_1(y_1, y_2)$ from the set

$$\mathcal{Z} = \{z_1(y_1) + z_2(y_2) : z_i \in C(\mathbb{R}),\ i = 1, 2\}.$$

It is easy to see that

$$E(f, \mathcal{R}) = E(f_1, \mathcal{Z}). \tag{32}$$

With each rectangle $S = [u_1, v_1] \times [u_2, v_2] \subset Y$ we associate the functional

$$L(h, S) = \frac{1}{4}\left(h(u_1, u_2) + h(v_1, v_2) - h(u_1, v_2) - h(v_1, u_2)\right), \quad h \in C(Y).$$

This functional has the following obvious properties:

(i) $L(z, S) = 0$ for any $z \in \mathcal{Z}$ and $S \subset Y$.

(ii) For any point $(y_1, y_2) \in Y$, $L(f_1, Y) = \sum_{i=1}^{4} L(f_1, S_i)$, where $S_1 = [c_1, y_1] \times [c_2, y_2]$, $S_2 = [y_1, d_1] \times [y_2, d_2]$, $S_3 = [c_1, y_1] \times [y_2, d_2]$, $S_4 = [y_1, d_1] \times [c_2, y_2]$.

By the conditions of the theorem, it is not difficult to verify that

$$\frac{\partial^2 f_1}{\partial y_1 \partial y_2} \geq 0 \text{ for any } (y_1, y_2) \in Y.$$

Integrating both sides of the last inequality over arbitrary rectangle $S = [u_1, v_1] \times [u_2, v_2] \subset Y$, we obtain that

$$L(f_1, S) \geq 0. \tag{33}$$

Set the function

$$f_2(y_1, y_2) = L(f_1, S_1) + L(f_1, S_2) - L(f_1, S_3) - L(f_1, S_4). \tag{34}$$

It is not difficult to verify that the function $f_1 - f_2$ belongs to $\mathcal{Z}$. Hence,

$$E(f_1, \mathcal{Z}) = E(f_2, \mathcal{Z}). \tag{35}$$

Calculate the norm $\|f_2\|$. From the property (ii), it follows that

$$f_2(y_1, y_2) = L(f_1, Y) - 2(L(f_1, S_3) + L(f_1, S_4)),$$

and

$$f_2(y_1, y_2) = 2(L(f_1, S_1) + L(f_1, S_2)) - L(f_1, Y).$$

From the last equalities and (33), we obtain that

$$|f_2(y_1, y_2)| \leq L(f_1, Y), \text{ for any } (y_1, y_2) \in Y.$$

On the other hand, one can check that

$$f_2(c_1, c_2) = f_2(d_1, d_2) = L(f_1, Y), \tag{36}$$

and

$$f_2(c_1, d_2) = f_2(d_1, c_2) = -L(f_1, Y). \tag{37}$$

Therefore

$$\|f_2\| = L(f_1, Y). \tag{38}$$

Note that the points $(c_1, c_2), (c_1, d_2), (d_1, d_2), (d_1, c_2)$ in the given order form a closed path with respect to the directions $(0;1)$ and $(1;0)$. We conclude from (36)-(38) that this path is extremal for f_2. By Theorem 3, $z_0 = 0$ is a best approximation to f_2. Hence,

$$E(f_2, \mathcal{Z}) = L(f_1, Y). \tag{39}$$

Now from (32),(35) and (39) we finally conclude that

$$E(f, \mathcal{R}) = L(f_1, Y) = \frac{1}{4}\left(f_1(c_1, c_2) + f_1(d_1, d_2) - f_1(c_1, d_2) - f_1(d_1, c_2)\right),$$

which is the desired result.◀

Theorem 5. *Let all the conditions of the previous theorem hold and $f_1(y_1, y_2)$ is the function defined in (29). Then the function $g_0(y_1, y_2) = g_{1,0}(y_1) + g_{2,0}(y_2)$, where*

$$g_{1,0}(y_1) = \frac{1}{2}f_1(y_1, c_2) + \frac{1}{2}f_1(y_1, d_2) - \frac{1}{4}f_1(c_1, c_2) - \frac{1}{4}f_1(d_1, d_2),$$

$$g_{2,0}(y_2) = \frac{1}{2}f_1(c_1, y_2) + \frac{1}{2}f_1(d_1, y_2) - \frac{1}{4}f_1(c_1, d_2) - \frac{1}{4}f_1(d_1, c_2),$$

and $y_1 = a_1x_1 + a_2x_2$, $y_2 = b_1x_1 + b_2x_2$, is a best approximation from the set $\mathcal{R}(a, b)$ to the function f.

Proof. It is not difficult to verify that the function $f_2(y_1, y_2)$ defined in (34) has the form

$$f_2(y_1, y_2) = f_1(y_1, y_2) - g_{1,0}(y_1) - g_{2,0}(y_2).$$

On the other hand, we know from the proof of Theorem 4 that

$$E(f_1, \mathcal{Z}) = \|f_2\|.$$

Therefore, the function $g_{1,0}(y_1) + g_{2,0}(y_2)$ is a best approximation to f_1. Then the function $g_{1,0}(\mathbf{a} \cdot \mathbf{x}) + g_{2,0}(\mathbf{b} \cdot \mathbf{x})$ is an extremal element from $\mathcal{R}(\mathbf{a}, \mathbf{b})$ to $f(\mathbf{x})$.◀

Remark 1. *Rivlin and Sibner [61], and Babaev [3] proved Theorem 4 for the case in which* **a** *and* **b** *are the unit vectors. Our proof of Theorem 4 is different, short and elementary. Moreover, it has turned out to be useful in constructing of an extremal element (see the proof of Theorem 5).*

Obviously, the set $\mathcal{R}\left(\mathbf{a}^1, ..., \mathbf{a}^r\right)$ is not dense in $C(\mathbb{R}^n)$ in the topology of uniform convergence on compact subsets of $\mathbb{R}^n$. Density here does not hold because the number of considered directions is finite. If consider all the possible directions, then the set $\mathcal{R} = span\{g(\mathbf{a} \cdot \mathbf{x}) : \ g \in C(\mathbb{R}), \ \mathbf{a} \in \mathbb{R}^n \backslash \{\mathbf{0}\}\}$ will be certainly dense in the space $C(\mathbb{R}^n)$ in the above mentioned topology. In order to be sure, it is enough to consider only the functions $e^{\mathbf{a} \cdot \mathbf{x}} \in \mathcal{R}$, which have dense linear span in $C(\mathbb{R}^n)$. For density it is not necessary to comprise all directions. The following result shows how many directions should be taken to satisfy the density requirements.

Proposition 5. *(Vostrecov and Kreines [68], Lin and Pinkus [46]). For density of the set*

$$\mathcal{R}(\mathcal{A}) = span\{g(\mathbf{a}\cdot\mathbf{x}):\ g\in C(\mathbb{R}),\ \mathbf{a}\in\mathcal{A}\subset\mathbb{R}^n\}$$

in $C(\mathbb{R}^n)$, (in the topology of uniform convergence on all compacta) it is necessary and sufficient that the only homogeneous polynomial which vanishes identically on $\mathcal{A}$ is the zero polynomial.

Since in the definition of $\mathcal{R}(\mathcal{A})$ we vary over all univariate functions g, allowing a direction $\mathbf{a}$ is equivalent to allowing all directions $k\mathbf{a}$ for every real k. Thus it is sufficient to consider only the set $\mathcal{A}$ of directions normalized to lie on the unit sphere S^{n-1}. For example, if $\mathcal{A}$ is a subset of the sphere S^{n-1}, which contains an interior point (interior point with respect to the induced topology on S^{n-1}), then $\mathcal{R}(\mathcal{A})$ is dense in the space $C(\mathbb{R}^n)$. The proof of the above proposition highlights an important fact that the set $\mathcal{R}(\mathcal{A})$ is dense in $C(\mathbb{R}^n)$ in the topology of uniform convergence on compact subsets if and only if $\mathcal{R}(\mathcal{A})$ contains all the polynomials (see [46]).

One may ask the following question: are there cases in which the set $\mathcal{R}\left(\mathbf{a}^1,...,\mathbf{a}^r\right)$ is dense in the space of all continuous functions? Undoubtedly, a positive answer depends on the geometrical structure of compact sets over which all the considered functions are defined. Let us first consider the case $r=2$. This case may be interesting in the theory of partial differential equations. Take, for example, equation (10). A positive answer to the problem means that for any continuous function f there exist solutions of the given equation uniformly converging to f.

It should be remarked that the problem of density $\mathcal{R}\left(\mathbf{a}^1,\mathbf{a}^2\right)$ is a special case of the problem considered by Marshall and O'Farrell. In [53], they obtained necessary and sufficient conditions for a sum A_1+A_2 of two subalgebras to be dense in $C(U)$, where $C(U)$ denotes the space of real-valued continuous functions on a compact Hausdorff space U. Understanding the great interest to the approximation by ridge functions, we like to describe Marshall and O' Farrell's solution applied to the problem considered here.

Let X be a compact subset of $\mathbb{R}^d$. The relation on X, defined by setting $\mathbf{x}\approx\mathbf{y}$ if $\mathbf{x}$ and $\mathbf{y}$ belong to some path in X, is an equivalence relation. The equivalence classes are called orbits.

Theorem 6. *(see [53]). Let X be a compact subset of $\mathbb{R}^d$ with all its orbits closed. Then the set $\mathcal{R}\left(\mathbf{a}^1,\mathbf{a}^2\right)$ is dense in $C(X)$ if and only if X contains no closed paths with respect to the directions $\mathbf{a}^1$ and $\mathbf{a}^2$.*

The proof immediately follows from proposition 2 in [53] established for the sum of two algebras. Since that proposition was given without proof, we give the proof of Theorem 6 for completeness of the exposition.

Proof. Necessity. If X has closed paths, then X has closed paths $p'=(\mathbf{p}_1',...,\mathbf{p}_{2m}')$ such that all points $\mathbf{p}_1',...,\mathbf{p}_{2m}'$ are distinct. In fact, such special paths can be obtained from any closed path $p = (\mathbf{p}_1, ..., \mathbf{p}_{2n})$ by the following simple algorithm: if the points of the path p are not all distinct, let i and $k > 0$ be the minimal indices such that $\mathbf{p}_i = \mathbf{p}_{i+2k}$; delete from p the subsequence $\mathbf{p}_{i+1}, ..., \mathbf{p}_{i+2k}$ and call p the obtained path; repeat the above step until all points of p are all distinct; set $p' := p$. By Urysohn's great lemma, there exist continuous functions $h = h(\mathbf{x})$ on X such that $h(\mathbf{p}_i') = 1$, $i = 1, 3, ..., 2m-1$, $h(\mathbf{p}_i') = -1$, $i = 2, 4, ..., 2m$ and $-1 < h(\mathbf{x}) < 1$ elsewhere. Consider the measure

$$\mu_{p'} = \frac{1}{2m} \sum_{i=1}^{2m} (-1)^{i-1} \delta_{\mathbf{p}_i'} ,$$

where $\delta_{\mathbf{p}_i'}$ is a point mass at $\mathbf{p}_i'$. For this measure, $\int_X h d\mu_{p'} = 1$ and $\int_X g d\mu_{p'} = 0$ for all functions $g \in \mathcal{R}\left(\mathbf{a}^1, \mathbf{a}^2\right)$. Thus the set $\mathcal{R}\left(\mathbf{a}^1, \mathbf{a}^2\right)$ cannot be dense in $C(X)$.

Sufficiency. We are going to prove that the only annihilating regular Borel measure for $\mathcal{R}\left(\mathbf{a}^1, \mathbf{a}^2\right)$ is the zero measure. Suppose, contrary to this assumption, there exists a nonzero annihilating measure on X for $\mathcal{R}\left(\mathbf{a}^1, \mathbf{a}^2\right)$. The class of such measures with total variation not more than 1 we denote by S. Clearly, S is weak-* compact and convex. By the Krein-Milman theorem, there exists an extreme measure μ in S. Since the orbits are closed, μ must be supported on a single orbit. Denote this orbit by T.

For $i = 1, 2$, let X_i be the quotient space of X obtained by identifying the points $\mathbf{y}$ and $\mathbf{z}$ whenever $\mathbf{a}^i \cdot \mathbf{y} = \mathbf{a}^i \cdot \mathbf{z}$. Let π_i be the natural projection of X onto X_i. For a fixed point $t \in X$ set $T_1 = \{t\}$, $T_2 = \pi_1^{-1}(\pi_1 T_1)$, $T_3 = \pi_2^{-1}(\pi_2 T_2)$, $T_4 = \pi_1^{-1}(\pi_1 T_3)$, ... Obviously, $T_1 \subset T_2 \subset T_3 \subset \cdots$. Therefore, for some $k \in \mathbb{N}$, $|\mu|(T_{2k}) > 0$, where $|\mu|$ is a total variation measure of μ. Since μ is orthogonal to every continuous function of the form $g\left(\mathbf{a}^1 \cdot \mathbf{x}\right)$, $\mu(T_{2k}) = 0$. From the Haar decomposition $\mu(T_{2k}) = \mu^+(T_{2k}) - \mu^-(T_{2k})$ it follows that $\mu^+(T_{2k}) = \mu^-(T_{2k}) > 0$. Fix a Borel subset $S_0 \subset T_{2k}$ such that $\mu^+(S_0) > 0$ and $\mu^-(S_0) = 0$. Since μ is orthogonal to every continuous function of the form $g\left(\mathbf{a}^2 \cdot \mathbf{x}\right)$, $\mu(\pi_2^{-1}(\pi_2 S_0)) = 0$. Therefore, one can chose a Borel set S_1 such that $S_1 \subset \pi_2^{-1}(\pi_2 S_0) \subset T_{2k+1}$, $S_1 \cap S_0 = \varnothing$, $\mu^+(S_1) = 0$, $\mu^-(S_1) \geqslant \mu^+(S_0)$. By the same way one can chose a Borel set S_2 such that $S_2 \subset \pi_1^{-1}(\pi_1 S_1) \subset T_{2k+2}$, $S_2 \cap S_1 = \varnothing$, $\mu^-(S_2) = 0$, $\mu^+(S_2) \geqslant \mu^-(S_1)$, and so on.

The sets $S_0, S_1, S_2, ...$are pairwise disjoint. For otherwise, there would exist positive integers n and m, with $n < m$ and a path $(y_n, y_{n+1}, ..., y_m)$ such that $y_i \in S_i$ for $i = n, ..., m$ and $y_m \in S_m \cap S_n$. But then there would exist paths $(z_1, z_2, ..., z_{n-1}, y_n)$ and $(z_1, z_2', ..., z_{n-1}', y_m)$ with z_i and z_i' in T_i for $i = 2, ..., n-1$. Hence, the set

$$\{z_1, z_2, ..., z_{n-1}, y_n, y_{n+1}, ..., y_m, z_{n-1}', ..., z_2', z_1\},$$

would contain a closed path. This would contradict our assumption on X.

Now, since the sets $S_0, S_1, S_2, ...$ are pairwise disjoint, and $|\mu|(S_i) \geqslant \mu^+(S_0) > 0$ for each $i = 1, 2, ...$, it follows that the total variation of μ is infinite. This contradiction completes the proof.◄

The following corollary concerns the problem considered by Colitschek and Light in [19].

Corollary 4. *[32]. Let D be a compact subset of $\mathbb{R}^2$ with all its orbits closed. Let W denote the set of all solutions of the wave equation*

$$\frac{\partial^2 w}{\partial s \partial t}(s,t) = 0, \qquad (s,t) \in D.$$

Then

$$\inf_{w \in W} \|f - w\| = 0,$$

for any continuous function $f(s,t)$ on D if and only if D contains no closed classical path (path with respect to the basic directions).

Proof. Let π_1 and π_2 denote the usual coordinate projections, viz: $\pi_1(s,t) = s$ and $\pi_2(s,t) = t$, $(s,t) \in \mathbb{R}^2$. Set $S = \pi_1(D)$ and $T = \pi_2(D)$. It is easy to see that

$$W = \left\{ w \in C(D) : w(s,t) = x(s) + y(t), \quad x \in C^2(S), \ \ y \in C^2(T) \right\}.$$

Set

$$\widetilde{W} = \{ w \in C(D) : w(s,t) = x(s) + y(t), \quad x \in C(S), \ \ y \in C(T) \}.$$

Since the set W is dense in $\widetilde{W}$,

$$\inf_{w \in W} \|f - w\| = \inf_{w \in \widetilde{W}} \|f - w\|.$$

But by Theorem 6, the equality

$$\inf_{w \in \widetilde{W}} \|f - w\| = 0,$$

holds for any $f \in C(D)$ if and only if D contains no closed path with respect to the basic directions.◄

Finally, we indicate the difficulties with the sum of more than two ridge functions. Consider the set $\mathcal{R}\left(\mathbf{a}^1, ..., \mathbf{a}^r\right)$, where $r \geq 3$. How can we define a path? Recall that in the case when $r = 2$, a path is an ordered set of points $(\mathbf{p}_1, \mathbf{p}_2, ..., \mathbf{p}_n)$ in $\mathbb{R}^d$ with edges $\mathbf{p}_i\mathbf{p}_{i+1}$ in alternating hyperplanes. The first, the third, the fifth,... hyperplanes (also the second, the fourth, the sixth,... hyperplanes) are parallel. If not differentiate between parallel hyperplanes, the path $(\mathbf{p}_1, \mathbf{p}_2, ..., \mathbf{p}_n)$ can be considered as a trace of some point traveling in two alternating hyperplanes. In this case, if the point starts and stops at the same location (i.e., if $\mathbf{p}_n = \mathbf{p}_1$) and n is an odd number, then the path functional

$$G(f) = \frac{1}{n-1} \sum_{i=1}^{n-1} (-1)^{i+1} f(\mathbf{p}_i),$$

annihilates each sum of ridge functions with the two fixed directions. The picture becomes more complicated when the number of directions is more than two. The simple generalization of the above-mentioned arguments demands a point traveling in three or more alternating hyperplanes. But in this case the appropriate generalization of the functional G does not annihilate functions from $\mathcal{R}\left(\mathbf{a}^1,...,\mathbf{a}^r\right)$.

There were several attempts to fill this gap in the special case when $r=d$ and $\mathbf{a}^1,...,\mathbf{a}^r$ are the unit vectors. Unfortunately, all these attempts failed (see, for example, the attempts in [11,20] and the refutations in [1,13,54]).

Although approximation techniques for the problems considered here are much less developed, there are some interesting results for the set $\mathcal{R}\left(\mathbf{a}^1,...,\mathbf{a}^r\right)$. For example, Lin and Pinkus [46] characterized $\mathcal{R}\left(\mathbf{a}^1,...,\mathbf{a}^r\right)$, i.e. they found means of determining if a continuous function f (defined on $\mathbb{R}^d$) is of the form $\sum\limits_{i=1}^{r} g_i(\mathbf{a}^i\cdot\mathbf{x})$ for some given $\mathbf{a}^1,...,\mathbf{a}^r\in\mathbb{R}^d\backslash\{\mathbf{0}\}$, but unknown continuous $g_1,...,g_r$. Buhmann and Pinkus [6] solved the problem: Assume we know that a function $f(\mathbf{x})$ is of the form $\sum\limits_{i=1}^{r} g_i(\mathbf{a}^i\cdot\mathbf{x})$. How do we determine the functions g_i?

At the end we want to draw the readers' attention to the following problems (all these problems are general and not solved by the methods introduced in this section).

Let Q be a compact subset of $\mathbb{R}^d$. Consider the approximation of a continuous function defined on Q by functions from $\mathcal{R}\left(\mathbf{a}^1,...,\mathbf{a}^r\right)$. Let $r\geq 3$.

Problem A. *Characterize those functions from $\mathcal{R}\left(\mathbf{a}^1,...,\mathbf{a}^r\right)$ that are extremal to a given continuous function.*

Problem B. *Establish explicit formulas for the error in approximating from $\mathcal{R}\left(\mathbf{a}^1,...,\mathbf{a}^r\right)$ and construct a best approximation.*

Problem C. *Find necessary and sufficient geometrical conditions for the set $\mathcal{R}\left(\mathbf{a}^1,...,\mathbf{a}^r\right)$ to be dense in $C(Q)$.*

It should be remarked that in [53], Problem C was set up for the sum of r subalgebras of $C(Q)$. Lin and Pinkus [46] proved that the set $\mathcal{R}\left(\mathbf{a}^1,...,\mathbf{a}^r\right)$ (r may be very large) is not dense in $C(\mathbb{R}^d)$ in the topology of uniform convergence on compact subsets of $\mathbb{R}^d$. That is, there are compact sets $Q\subset\mathbb{R}^d$ such that $\mathcal{R}\left(\mathbf{a}^1,...,\mathbf{a}^r\right)$ is not dense in $C(Q)$. In the case $r=2$, Theorem 6 complements this result, by describing compact sets $Q\subset\mathbb{R}^2$, for which $\mathcal{R}\left(\mathbf{a}^1,\mathbf{a}^2\right)$ is dense in $C(Q)$.

4. Sums of continuous ridge functions

The problem of representation of a fixed 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 $\mathcal{R}_c\left(\mathbf{a}^1,...,\mathbf{a}^r\right)$ and $\mathcal{R}_b\left(\mathbf{a}^1,...,\mathbf{a}^r\right)$ denote the subspaces of $\mathcal{R}\left(\mathbf{a}^1,...,\mathbf{a}^r\right)$ 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 $C(X)=\mathcal{R}_c\left(\mathbf{a}^1,...,\mathbf{a}^r\right)$, $B(X)=\mathcal{R}_b\left(\mathbf{a}^1,...,\mathbf{a}^r\right)$, and $T(X)=\mathcal{R}\left(\mathbf{a}^1,...,\mathbf{a}^r\right)$? The first two problems in more general setting were solved by Sternfeld [64,65]. The third problem was solved in the previous section. Let us cite the results of Sternfeld for the case of representation by sums of ridge functions. Let 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 [65], 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 $\mathcal{R}_b\left(\mathbf{a}^1,...,\mathbf{a}^r\right)=B(X)$. In [65], Sternfeld also obtained a practically convenient sufficient condition for the equality $\mathcal{R}_b\left(\mathbf{a}^1,...,\mathbf{a}^r\right)=B(X)$.To describe his 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 6. *(Sternfeld [65]). If $\tau^n(X)=\emptyset$ for some n, then $\mathcal{R}_b\left(\mathbf{a}^1,...,\mathbf{a}^r\right)=B(X)$. If X is a compact subset of $\mathbb{R}^n$, and $\tau^n(X)=\emptyset$ for some n, then $\mathcal{R}_c\left(\mathbf{a}^1,...,\mathbf{a}^r\right)=C(X)$.*

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 $\mathcal{R}_b\left(\mathbf{a}^1,\mathbf{a}^2\right)=B(X)$ is equivalent to the equality $\mathcal{R}_c\left(\mathbf{a}^1,\mathbf{a}^2\right)=C(X)$. In another work [64], Sternfeld obtained a measure-theoretic necessary and sufficient condition for the equality $\mathcal{R}_c\left(\mathbf{a}^1,...,\mathbf{a}^r\right)=C(X)$. 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 [64], in particular, proved that the equality$\mathcal{R}_c\left(\mathbf{a}^1,...,\mathbf{a}^r\right)=C(X)$ 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 $\mathcal{R}_b\left(\mathbf{a}^1,...,\mathbf{a}^r\right)=B(X)$ 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 $\mathcal{R}_c\left(\mathbf{a}^1,...,\mathbf{a}^r\right)=C(X)$ implies the second equality $\mathcal{R}_b\left(\mathbf{a}^1,...,\mathbf{a}^r\right)=B(X)$. The inverse is not true (see [64]). We stress again that the above results of Sternfeld 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 arbitrarily fixed functions (bounded or continuous) defined on X.

Consider the following representation problem associated with the set $\mathcal{R}\left(\mathbf{a}^1, ..., \mathbf{a}^r\right)$:

Let X be a compact subset of $\mathbb{R}^n$. Give geometrical conditions that are necessary and sufficient for

$$\mathcal{R}\left(\mathbf{a}^1, ..., \mathbf{a}^r\right) = C\left(X\right),$$

where $C\left(X\right)$ is the space of continuous functions on X furnished with the uniform norm.

We are going to show how this problem is solved for $r = 2$ and indicate some difficulties related to the case $r \geq 3$. In the sequel, we will use the following notation:

$$H_1 = H_1\left(X\right) = \left\{g_1\left(\mathbf{a}^1 \cdot \mathbf{x}\right) : g_1 \in C\left(\mathbb{R}\right)\right\},$$

$$H_2 = H_2\left(X\right) = \left\{g_2\left(\mathbf{a}^2 \cdot \mathbf{x}\right) : g_2 \in C\left(\mathbb{R}\right)\right\}.$$

Note that by this notation, $\mathcal{R}\left(\mathbf{a}^1, \mathbf{a}^2\right) = H_1 + H_2$.

Theorem 7. *(see [32,41]). Let X be a compact subset of $\mathbb{R}^n$. The equality*

$$H_1\left(X\right) + H_2\left(X\right) = C\left(X\right),$$

holds if and only if X contains no closed path and there exists a positive integer n_0 such that the lengths of paths in X are bounded by n_0.

Proof. **Necessity.** Let $H_1 + H_2 = C\left(X\right)$. Consider the linear operator

$$A : H_1 \times H_2 \to C\left(X\right), \quad A\left[\left(g_1, g_2\right)\right] = g_1 + g_2,$$

where $g_1 \in H_1, g_2 \in H_2$. We define the norm on $H_1 \times H_2$

$$\left\|\left(g_1, g_2\right)\right\| = \left\|g_1\right\| + \left\|g_2\right\|.$$

It is obvious that the operator A is continuous with respect to this norm. Besides, since $C\left(X\right) = H_1 + H_2$, A is a surjection. Consider the conjugate operator

$$A^* : C\left(X\right)^* \to \left[H_1 \times H_2\right]^*, \quad A^*\left[G\right] = \left(G_1, G_2\right),$$

where the functionals G_1 and G_2 are defined as follows:

$$G_1\left(g_1\right) = G\left(g_1\right), g_1 \in H_1; \quad G_2\left(g_2\right) = G\left(g_2\right), g_2 \in H_2.$$

An element $\left(G_1, G_2\right)$ from $\left[H_1 \times H_2\right]^*$ has the norm

$$\left\|\left(G_1, G_2\right)\right\| = \max\left\{\left\|G_1\right\|, \left\|G_2\right\|\right\}. \tag{40}$$

Let now $p = \left(p_1, ..., p_m\right)$ be any path with different points: $p_i \neq p_j$ for any $i \neq j$, $1 \leq i, \ j \leq m$. We associate with p the following functional over $C\left(X\right)$:

$$L\left[f\right] = \frac{1}{m}\sum_{i=1}^{m}\left(-1\right)^{i-1} f\left(p_i\right).$$

Since $|L(f)| \leq \|f\|$ and $|L(g)| = \|g\|$ for a continuous function $g(\mathbf{x})$ such that $g(p_i) = 1$, for odd indices i, $g(p_j) = -1$, for even indices j and $-1 < g(\mathbf{x}) < 1$ elsewhere, we obtain that $\|L\| = 1$. Let $A^*[L] = (L_1, L_2)$. One can easily verify that

$$\|L_i\| \leq \frac{2}{m}, i = 1, 2.$$

Therefore, from (40) we obtain that

$$\|A^*[L]\| \leq \frac{2}{m}. \tag{41}$$

Since A is a surjection, there exists $\delta > 0$ such that

$$\|A^*[G]\| \geq \delta \|G\| \qquad \text{for any functional } \; G \in C(X)^*.$$

Hence,

$$\|A^*[L]\| \geq \delta. \tag{42}$$

Now from (41) and (42) we conclude that

$$m \leq \frac{2}{\delta}.$$

This means that for a path with different points, n_0 can be chosen as $\left[\frac{2}{\delta}\right] + 1$.

Let now $p = (p_1, ..., p_m)$ be a path with at least two coinciding points. Then we can form a closed path with different points. This may be done by the following way: let i and j be indices such that $p_i = \; p_j$ and $j - i$ takes its minimal value. Note that in this case all the points $p_i, p_{i+1}, ..., p_{j-1}$ are distinct. Now if $j - i$ is an even number, then the path $(p_i, p_{i+1}, ..., p_{j-1})$, and if $j - i$ is an odd number, then the path $(p_{i+1}, ..., p_{j-1})$ is a closed path with different points. It remains to show that X can not possess closed paths with different points. Indeed, if $q = (q_1, ..., q_{2k})$ is a path of this type, then the functional L, associated with q, annihilates all functions from $H_1 + H_2$. On the other hand, $L[f] = 1$ for a continuous function f on X satisfying the conditions $f(t) = 1$ if $t \in \{q_1, q_3, ..., q_{2k-1}\}$; $f(t) = -1$ if $t \in \{q_2, q_4, ..., q_{2k}\}$; $f(t) \in (-1; 1)$ if $t \in X \backslash q$. This implies on the contrary to our assumption that $H_1 + H_2 \neq C(X)$. The necessity has been proved.

Sufficiency. Let X contains no closed path and the lengths of all paths are bounded by some positive integer n_0. We may suppose that any path has different points. Indeed, in other case we can form a closed path, which contradicts our assumption.

For $i = 1, 2$, let X_i be the quotient space of X obtained by identifying the points a and b whenever $g(a) = g(b)$ for each g in H_i. Let π_i be the natural projection of X onto X_i. For a point $t \in X$ set $T_1 = \pi_1^{-1}(\pi_1 t), T_2 = \pi_2^{-1}(\pi_2 T_1), \ldots$. By $O(t)$ denote the orbit of X containing t. Since the length of any path in X is not more than n_0, we conclude that $O(t) = T_{n_0}$. Since X is compact, the sets $T_1, T_2, ..., T_{n_0}$, hence $O(t)$, are compact. By Theorem 6 (see Section 1.3), $\overline{H_1 + H_2} = C(X)$.

Now show that $H_1 + H_2$ is closed in $C(X)$. Set

$$H_3 = H_1 \cap H_2.$$

Let X_3 and π_3 be the associated quotient space and projection, respectively. Fix some $a \in X_3$. Show, within conditions of our theorem, that if $t \in \pi_3^{-1}(a)$, then $O(t) = \pi_3^{-1}(a)$. The inclusion $O(t) \subset \pi_3^{-1}(a)$ is obvious. Suppose that there exists a point $t_1 \in \pi_3^{-1}(a)$ such that $t_1 \notin O(t)$. Then $O(t) \cap O(t_1) = \emptyset$. By $X|O$ denote the factor space generated by orbits of X. $X|O$ is a normal topological space with its natural factor topology. Hence, we can construct a continuous function $u \in C(X|O)$ such that $u(O(t)) = 0$, $u(O(t_1)) = 1$. The function $v(x) = u(O(x))$, $x \in X$, is continuous on X and belongs to H_3 as a function being constant on each orbit. But, since $O(t) \subset \pi_3^{-1}(a)$ and $O(t_1) \subset \pi_3^{-1}(a)$, the function $v(x)$ can not take different values on $O(t)$ and $O(t_1)$. This contradiction means that there is not a point $t_1 \in \pi_3^{-1}(a)$ such that $t_1 \notin O(t)$. Thus

$$O(t) = \pi_3^{-1}(a), \tag{43}$$

for any $a \in X_3$ and $t \in \pi_3^{-1}(a)$.

Now prove that there exists a positive real number c such that

$$\sup_{z \in X_3} \underset{\pi_3^{-1}(z)}{var} f \le c \sup_{y \in X_2} \underset{\pi_2^{-1}(y)}{var} f, \tag{44}$$

for all f in H_1. Note that for $Y \subset X$, $\underset{Y}{var} f$ is the variation of f on the set Y. That is,

$$\underset{Y}{var} f = \sup_{x,y \in Y} |f(x) - f(y)|.$$

Due to (43), inequality (44) can be written in the following form:

$$\sup_{t \in X} \underset{O(t)}{var} f \le c \sup_{t \in X} \underset{\pi_2^{-1}(\pi_2(t))}{var} f, \tag{45}$$

for all $f \in H_1$.

Let $t \in X$ and t_1, t_2 be arbitrary points of $O(t)$. Then there is a path $(b_1, b_2, ..., b_m)$ with $b_1 = t_1$ and $b_m = t_2$. Besides, by the condition, $m \le n_0$. Let first $\mathbf{a}^2 \cdot b_1 = \mathbf{a}^2 \cdot b_2$, $\mathbf{a}^1 \cdot b_2 = \mathbf{a}^1 \cdot b_3, ..., \mathbf{a}^2 \cdot b_{m-1} = \mathbf{a}^2 \cdot b_m$. Then for any function $f \in H_1$:

$$|f(t_1) - f(t_2)| = |f(b_1) - f(b_2) + ... - f(b_m)| \le$$

$$\le |f(b_1) - f(b_2)| + ... + |f(b_{m-1}) - f(b_m)| \le \frac{n_o}{2} \sup_{t \in X} \underset{\pi_2^{-1}(\pi_2(t))}{var} f. \tag{46}$$

It is not difficult to verify that inequality (46) holds in all other possible cases of the path $(b_1, ..., b_m)$. Now from (46) we obtain (45), hence (44), where $c = \frac{n_0}{2}$. In [108], Marshall and O'Farrell proved the following result (see Proposition 4 in [108]): Let A_1 and A_2 be closed subalgebras of $C(X)$ that contain the constants. Let (X_1, π_1), (X_2, π_2) and (X_3, π_3) be the quotient spaces and projections associated with the algebras A_1, A_2 and $A_3 = A_1 \cap A_2$ respectively. Then $A_1 + A_2$ is closed in $C(X)$ if and only if there exists a positive real number c such that

$$\sup_{z \in X_3} \underset{\pi_3^{-1}(z)}{var} f \le c \sup_{y \in X_2} \underset{\pi_2^{-1}(y)}{var} f,$$

for all f in A_1.

By this proposition, (44) implies that $H_1 + H_2$ is closed in $C(X)$. Thus, we finally obtain that $H_1 + H_2 = C(X)$.

Paths with respect to two directions are explicit objects and give geometric means of deciding if $H_1 + H_2 = C(X)$. Let us show this in the example of the bivariate ridge functions $g_1 = x_1 + x_2$ and $g_2 = x_1 - x_2$. If X is the union of two parallel line segments in $\mathbb{R}^2$, not parallel to any of the lines $x_1 + x_2 = 0$ and $x_1 - x_2 = 0$, then Theorem 7 holds. If X is any bounded part of the graph of the function $x_2 = \arcsin(\sin x_1)$, then Theorem 7 also holds. Let now X be the set

$$\{(0,0),(1,-1),(0,-2),(-1\tfrac{1}{2},-\tfrac{1}{2}),(0,1),(\tfrac{3}{4},\tfrac{1}{4}),(0,-\tfrac{1}{2}),\\ (-\tfrac{3}{8},-\tfrac{1}{8}),(0,\tfrac{1}{4}),(\tfrac{3}{16},\tfrac{1}{16}),...\}.$$

In this case, there is no positive integer bounding lengths of all paths. Thus, Theorem 7 fails. Note that since orbits of all paths are closed, Theorem 6 from the previous section shows $H_1 + H_2$ is dense in $C(X)$.

If X is any set with interior points, then both Theorem 6 and Theorem 7 fail, since any such set contains the vertices of some parallelogram with sides parallel to the directions $\mathbf{a}^1$ and $\mathbf{a}^2$, that is a closed path.

To solve the above geometrical problem of representation for the general case in which $r \geq 3$ is more difficult than to solve it for $r = 2$. In this case, we even don't know what objects will be an appropriate generalization of paths. Representation by sums of continuous ridge functions requires more complicated relations between points of X than relations induced by paths with respect to only two directions. If disregard continuity, we have seen in Section 2 that cycles with respect to n directions are able to solve the representation problem. But when some topology is involved, the picture is quite different. No one knows a geometrically explicit solution to the problem of representation of continuous multivariate functions by sums of continuous ridge functions. Nevertheless, it should be noted that this problem in quite abstract (and not geometrical) form was solved by Sternfeld. His solution involves a family of functions that separates regular Borel measures on a given compact set X. A family $F = \{h\} \subset C(X)$ is said to be a measure separating family (*m.s.f.*) if there exists a number $0 < \lambda \leq 1$ such that for any measure μ in $C(X)^*$, the inequality $\left\|\mu \circ h^{-1}\right\| \geq \lambda \|\mu\|$ holds for some $h \in F$. Sternfeld [64], in particular, proved that $\mathcal{R}\left(\mathbf{a}^1,...,\mathbf{a}^r\right) = C(X)$ if and only if the family $\{\mathbf{a}^i \cdot \mathbf{x},\ i = 1,...,r\}$ is a *m.s.f.*

Theorem 7 admits a direct generalization to the representation by sums $g_1(h_1(\mathbf{x})) + g_2(h_2(\mathbf{x}))$, where $h_1(\mathbf{x})$ and $h_2(\mathbf{x})$ are fixed continuous functions on X. This generalization needs consideration of new objects – paths with respect to two continuous functions.

Definition 5. *(see [30,32]). Let X be a compact subset of $\mathbb{R}^n$ and $h_i \in C(X), i = 1,2$. A finite ordered subset $(p_1, p_2, ..., p_m)$ of X with $p_i \neq p_{i+1}\,(i = 1,...,m-1)$, and either $h_1(p_1) = h_1(p_2)$, $h_2(p_2) = h_2(p_3)$, $h_1(p_3) = h_1(p_4)$, ..., or $h_2(p_1) = h_2(p_2)$, $h_1(p_2) = h_1(p_3)$, $h_2(p_3) = h_2(p_4)$, ... is called a path with respect to the functions h_1 and h_2 or shortly an h_1-h_2 path.*

Theorem 8. *(see [32,41]). Let X be a compact subset of $\mathbb{R}^n$. Every function $f(x) \in C(X)$ admits a representation*

$$f(x) = g_1(h_1(\mathbf{x})) + g_2(h_2(\mathbf{x})), \ g_1, g_2 \in C(\mathbb{R}),$$

if and only if the set X contains no closed h_1-h_2 path and there exists a positive integer n_0 such that the lengths of h_1-h_2 paths in X are bounded by n_0.

The proof can be carried out by the same arguments as above.

5. Existence of extremal ridge functions

Let E be a normed linear space and F be its subspace. We say that F is proximinal in E if for any element $e \in E$ there exists at least one element $f_0 \in F$ such that

$$\|e - f_0\| = \inf_{f \in F} \|e - f\|.$$

In this case, the element f_0 is said to be extremal to e.

In the following, we are going to deal with the problem of proximinality of the set of linear combinations of ridge functions in the spaces of bounded and continuous functions, respectively. This problem will be considered in the simplest case when the class of approximating functions is the set

$$\mathcal{R} = \mathcal{R}\left(\mathbf{a}^1, \mathbf{a}^2\right) = \left\{g_1\left(\mathbf{a}^1 \cdot \mathbf{x}\right) + g_2\left(\mathbf{a}^2 \cdot \mathbf{x}\right) : g_i : \mathbb{R} \to \mathbb{R}, i = 1, 2\right\}.$$

Here $\mathbf{a}^1$ and $\mathbf{a}^2$ are fixed directions and we vary over g_i. It is clear that this is a linear space. Consider the following three subspaces of $\mathcal{R}$. The first is obtained by taking only bounded sums $g_1\left(\mathbf{a}^1 \cdot \mathbf{x}\right) + g_2\left(\mathbf{a}^2 \cdot \mathbf{x}\right)$ over some set X in $\mathbb{R}^n$. We denote this subspace by $\mathcal{R}_a(X)$. The second and the third are subspaces of $\mathcal{R}$ with bounded and continuous summands $g_i\left(\mathbf{a}^i \cdot \mathbf{x}\right), \ i = 1, 2$, on X respectively. These subspaces will be denoted by $\mathcal{R}_b(X)$ and $\mathcal{R}_c(X)$. In the case of $\mathcal{R}_c(X)$, the set X is considered to be compact.

Let $B(X)$ and $C(X)$ be the spaces of bounded and continuous multivariate functions over X, respectively. What conditions must one impose on X in order that the sets $\mathcal{R}_a(X)$ and $\mathcal{R}_b(X)$ be proximinal in $B(X)$ and the set $\mathcal{R}_c(X)$ be proximinal in $C(X)$? We are also interested in necessary conditions for proximinality. It follows from one result of Garkavi, Medvedev and Khavinson (see Theorem 1 [18]) that $\mathcal{R}_a(X)$ is proximinal in $B(X)$ for all subsets X of $\mathbb{R}^n$. There is also an answer (see Theorem 2 [18]) for proximinality of $\mathcal{R}_b(X)$ in $B(X)$. Is the set $\mathcal{R}_b(X)$ always proximinal in $B(X)$? There is an an example of a set $X \subset \mathbb{R}^n$ and a bounded function f on X for which there does not exist an extremal element in $\mathcal{R}_b(X)$.

We want to draw the readers' attention to the more general case in which the number of directions is more than two. In this case, the set of approximating functions is

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

In a similar way as above, one can define the sets $\mathcal{R}_a(X)$, $\mathcal{R}_b(X)$ and $\mathcal{R}_c(X)$. Using the results of [18], one can obtain sufficient (but not necessary) conditions for proximinality of these sets. This needs, besides paths, the consideration of some additional and more complicated relations between points of X. The case $r \geq 3$ will not be considered in the current section, since our main purpose is to draw readers' attention to the arisen problems of proximinality in the simplest case of approximation. For the existing open problems connected with the set $\mathcal{R}\left(\mathbf{a}^1,...,\mathbf{a}^r\right)$, where $r \geq 3$, see [32] and [58].

Let $\mathbf{a}^1$ and $\mathbf{a}^2$ be two different directions in $\mathbb{R}^n$. Let us recall that a path with respect to the directions $\mathbf{a}^1$ and $\mathbf{a}^2$ is a finite or infinite ordered set of points $(\mathbf{x}^1,\mathbf{x}^2,...)$ in $\mathbb{R}^n$ with the units $\mathbf{x}^{i+1}-\mathbf{x}^i$, $i=1,2,...$, in the directions perpendicular alternatively to $\mathbf{a}^1$ and $\mathbf{a}^2$. In the sequel, we simply use the term "path" instead of the long expression "path with respect to the directions $\mathbf{a}^1$ and $\mathbf{a}^2$ ". The length of a path is the number of its points and can be equal to ∞ if the path is infinite. A singleton is a path of the unit length. We say that a path $\left(\mathbf{x}^1,...,\mathbf{x}^m\right)$ belonging to some subset X of $\mathbb{R}^n$ is irreducible if there is not another path $\left(\mathbf{y}^1,...,\mathbf{y}^l\right)\subset X$ with $\mathbf{y}^1=\mathbf{x}^1$, $\mathbf{y}^l=\mathbf{x}^m$ and $l<m$. If in a path $\left(\mathbf{x}^1,...,\mathbf{x}^m\right)$ m is an even number and the set $\left(\mathbf{x}^1,...,\mathbf{x}^m,\mathbf{x}^1\right)$ is also a path, then the path $\left(\mathbf{x}^1,...,\mathbf{x}^m\right)$ is called to be closed.

The following theorem follows from Theorem 2 of [18]:

Theorem 9. *Let $X\subset\mathbb{R}^n$ and the lengths of all irreducible paths in X be uniformly bounded by some positive integer. Then each function in $B(X)$ has an extremal element in $\mathcal{R}_b(X)$.*

There is a large number of sets in $\mathbb{R}^n$ satisfying the hypothesis of this theorem. For example, if a set X has a cross section according to one of the directions $\mathbf{a}^1$ or $\mathbf{a}^2$, then the set X satisfies the hypothesis of Theorem 9. By a cross section according to the direction $\mathbf{a}^1$ we mean any set $X_{\mathbf{a}^1}=\{x\in X:\ \mathbf{a}^1\cdot\mathbf{x}=c\}, c\in\mathbb{R}$, with the property: for any $\mathbf{y}\in X$ there exists a point $\mathbf{y}^1\in X_{\mathbf{a}^1}$ such that $\mathbf{a}^2\cdot\mathbf{y}=\mathbf{a}^2\cdot\mathbf{y}^1$. By the similar way, one can define a cross section according to the direction $\mathbf{a}^2$. Regarding Theorem 9, one may ask if the condition of the theorem is necessary for proximinality of $\mathcal{R}_b(X)$ in $B(X)$. While we do not know a complete answer to this question, we are going to give an example of a set X for which Theorem 9 fails. Let $\mathbf{a}^1=(1;-1)$, $\mathbf{a}^2=(1;1)$. Consider the set

$$\begin{aligned} X \;=\; &\{(2;\frac{2}{3}),(\frac{2}{3};-\frac{2}{3}),(0;0),(1;1),(1+\frac{1}{2};1-\frac{1}{2}),(1+\frac{1}{2}+\frac{1}{4};1-\frac{1}{2}+\frac{1}{4}),\\ &(1+\frac{1}{2}+\frac{1}{4}+\frac{1}{8};1-\frac{1}{2}+\frac{1}{4}-\frac{1}{8}),...\}. \end{aligned}$$

In what follows, the elements of X in the given order will be denoted by $\mathbf{x}^0,\mathbf{x}^1,\mathbf{x}^2,...$. It is clear that X is a path of the infinite length and $\mathbf{x}^n\rightarrow\mathbf{x}^0$, as $n\rightarrow\infty$. Let $\sum_{n=1}^{\infty}c_n$

be any divergent series with the terms $c_n > 0$ and $c_n \to 0$, as $n \to \infty$. Besides, let f_0 be a function vanishing at the points $\mathbf{x}^0, \mathbf{x}^2, \mathbf{x}^4, ...$, and taking values $c_1, c_2, c_3, ...$ at the points $\mathbf{x}^1, \mathbf{x}^3, \mathbf{x}^5, ...$, respectively. It is obvious that f_0 is continuous on X. The set X is compact and satisfies all the conditions of Proposition 2 of [53]. By this proposition, $\overline{\mathcal{R}_c(X)} = C(X)$. Therefore, for any continuous function on X, thus for f_0,

$$\inf_{g \in \mathcal{R}_c(X)} \|f_0 - g\|_{C(X)} = 0. \tag{47}$$

Since $\mathcal{R}_c(X) \subset \mathcal{R}_b(X)$, we obtain from (47) that

$$\inf_{g \in \mathcal{R}_b(X)} \|f_0 - g\|_{B(X)} = 0. \tag{48}$$

Suppose that f_0 has an extremal element $g_1^0\left(\mathbf{a}^1 \cdot \mathbf{x}\right) + g_2^0\left(\mathbf{a}^2 \cdot \mathbf{x}\right)$ in $\mathcal{R}_b(X)$. By the definition of $\mathcal{R}_b(X)$, the ridge functions $g_i^0, i = 1, 2$, are bounded on X. From (48) it follows that $f_0 = g_1^0\left(\mathbf{a}^1 \cdot \mathbf{x}\right) + g_2^0\left(\mathbf{a}^2 \cdot \mathbf{x}\right)$. Since $\mathbf{a}^1 \cdot \mathbf{x}^{2n} = \mathbf{a}^1 \cdot \mathbf{x}^{2n+1}$ and $\mathbf{a}^2 \cdot \mathbf{x}^{2n+1} = \mathbf{a}^2 \cdot \mathbf{x}^{2n+2}$, for $n = 0, 1, ...$, we can write

$$\sum_{n=0}^{k} c_{n+1} = \sum_{n=0}^{k} \left[f(\mathbf{x}^{2n+1}) - f(\mathbf{x}^{2n})\right]$$

$$= \sum_{n=0}^{k} \left[g_2^0(\mathbf{x}^{2n+1}) - g_2^0(\mathbf{x}^{2n})\right] = g_2^0(\mathbf{a}^2 \cdot \mathbf{x}^{2k+1}) - g_2^0(\mathbf{a}^2 \cdot \mathbf{x}^0). \tag{49}$$

Since $\sum_{n=1}^{\infty} c_n = \infty$, we deduce from (49) that the function $g_2^0\left(\mathbf{a}^2 \cdot \mathbf{x}\right)$ is not bounded on X. This contradiction means that the function f_0 does not have an extremal element in $\mathcal{R}_b(X)$. Therefore, the space $\mathcal{R}_b(X)$ is not proximinal in $B(X)$.

Let us now give sufficient conditions and also a necessary condition for proximinality of $\mathcal{R}_c(X)$ in $C(X)$.

Theorem 10. *[26]. Let the system of independent vectors $\mathbf{a}^1$ and $\mathbf{a}^2$ have a complement to a basis $\{\mathbf{a}^1, ..., \mathbf{a}^n\}$ in $\mathbb{R}^n$ with the property: for any point $\mathbf{x}^0 \in X$ and any positive real number δ there exist a number $\delta_0 \in (0, \delta]$ and a point $\mathbf{x}^\sigma$ in the set*

$$\sigma = \{\mathbf{x} \in X : \mathbf{a}^2 \cdot \mathbf{x}^0 - \delta_0 \leq \mathbf{a}^2 \cdot \mathbf{x} \leq \mathbf{a}^2 \cdot \mathbf{x}^0 + \delta_0\},$$

such that the system

$$\begin{cases} \mathbf{a}^2 \cdot \mathbf{x}' = \mathbf{a}^2 \cdot \mathbf{x}^\sigma, \\ \mathbf{a}^1 \cdot \mathbf{x}' = \mathbf{a}^1 \cdot \mathbf{x}, \\ \sum_{i=3}^{n} \left|\mathbf{a}^i \cdot \mathbf{x}' - \mathbf{a}^i \cdot \mathbf{x}\right| < \delta, \end{cases} \tag{50}$$

has a solution $\mathbf{x}' \in \sigma$ for all points $\mathbf{x} \in \sigma$. Then the space $\mathcal{R}_c(X)$ is proximinal in $C(X)$.

Proof.
Introduce the following mappings and sets:

$$\pi_i : X \to \mathbb{R},\ \pi_i(\mathbf{x}) = \mathbf{a}^i \cdot \mathbf{x},\ Y_i = \pi_i(X),\ i = 1, ..., n.$$

Since the system of vectors $\{\mathbf{a}^1, ..., \mathbf{a}^n\}$ is linearly independent, the mapping $\pi = (\pi_1, ...\pi_n)$ is an injection from X into the Cartesian product $Y_1 \times ... \times Y_n$. Besides, π is linear and continuous. By the open mapping theorem, the inverse mapping π^{-1} is continuous from $Y = \pi(X)$ onto X. Let f be a continuous function on X. Then the composition $f \circ \pi^{-1}(y_1, ...y_n)$ will be continuous on Y, where $y_i = \pi_i(\mathbf{x})$, $i = 1, ..., n$, are the coordinate functions. Consider the approximation of the function $f \circ \pi^{-1}$ by elements from

$$G_0 = \{g_1(y_1) + g_2(y_2) :\ g_i \in C(Y_i),\ i = 1, 2\},$$

over the compact set Y. Then one may observe that the function f has an extremal element in $\mathcal{R}_c(X)$ if and only if the function $f \circ \pi^{-1}$ has an extremal element in G_0. Thus, the problem of proximinality of $\mathcal{R}_c(X)$ in $C(X)$ is reduced to the problem of proximinality of G_0 in $C(Y)$.

Let $T, T_1, ..., T_{m+1}$ be metric compact spaces and $T \subset T_1 \times ... \times T_{m+1}$. For $i = 1, ..., m$, let φ_i be the continuous mappings from T onto T_i. In [35], the author obtained sufficient conditions for proximinality of the set

$$C_0 = \{\sum_{i=1}^{n} g_i \circ \varphi_i :\ g_i \in C(T_i),\ i = 1, ...m\},$$

in the space $C(T)$ of continuous functions on T. Since $Y \subset Y_1 \times Y_2 \times Z_3$, where $Z_3 = Y_3 \times ... \times Y_n$, we can use this result in our case for the approximation of the function $f \circ \pi^{-1}$ by elements from G_0. By this theorem, the set G_0 .is proximinal in $C(Y)$ if for any $y_2^0 \in Y_2$ and $\delta > 0$ there exists a number $\delta_0 \in (0, \delta)$ such that the set $\sigma(y_2^0, \delta_0) = [y_2^0 - \delta_0, y_2^0 + \delta_0] \cap Y_2$ has $(2, \delta)$ maximal cross section. The last means that there exists a point $y_2^\sigma \in \sigma(y_2^0, \delta_0)$ with the property: for any point $(y_1, y_2, z_3) \in Y$, with the second coordinate y_2 from the set $\sigma(y_2^0, \delta_0)$, there exists a point $(y_1', y_2^\sigma, z_3') \in Y$ such that $y_1 = y_1'$ and $\rho(z_3, z_3') < \delta$, where ρ is a metrics in Z_3. Since these conditions are equivalent to the conditions of Theorem 10, the space G_0 is proximinal in the space $C(Y)$. Then by the above conclusion, the space $\mathcal{R}_c(X)$ is proximinal in $C(X)$. ◄

Let us give some simple examples of compact sets satisfying the hypothesis of Theorem 10. For the sake of brevity, we restrict ourselves to the case $n = 3$.

(a) Let X be a closed ball in $\mathbb{R}^3$, a^1 and a^2 be two arbitrary orthogonal directions. Then Theorem 10 holds. Note that in this case we can take $\delta_0 = \delta$ and a^3 as an orthogonal vector to both the vectors a^1 and a^2.

(b) Let X be the unite cube, $a^1 = (1; 1; 0)$, $a^2 = (1; -1; 0)$. Then Theorem 10 also holds. In this case, we can take $\delta_0 = \delta$ and $a^3 = (0; 0; 1)$. Note that the unit cube does not satisfy the hypothesis of the theorem for many directions (take, for example, $a^1 = (1; 2; 0)$ and $a^2 = (2; -1; 0)$).

In the following example, one can not always chose δ_0 as equal to δ.

(c) Let $X = \{(x_1, x_2, x_3) : (x_1, x_2) \in Q, \ 0 \leq x_3 \leq 1\}$, where Q is the union of two triangles $A_1B_1C_1$ and $A_2B_2C_2$ with the vertices $A_1 = (0;0)$, $B_1 = (1;2)$, $C_1 = (2;0)$, $A_2 = (1\frac{1}{2};1)$, $B_2 = (2\frac{1}{2};-1)$, $C_2 = (3\frac{1}{2};1)$. Let $a^1 = (0;1;0)$ and $a^2 = (1;0;0)$. Then it is easy to see that Theorem 10 holds (the vector a^3 can be chosen as $(0;0;1)$). In this case, δ_0 can not be always chosen as equal to δ. Take, for example, $\mathbf{x}^0 = (1\frac{3}{4};0;0)$ and $\delta = 1\frac{3}{4}$. If $\delta_0 = \delta$, then the second equation of the system (50) has not a solution for a point $(1;2;0)$ or a point $(2\frac{1}{2};-1;0)$. But if we take δ_0 not more than $\frac{1}{4}$, then for $\mathbf{x}^\sigma = \mathbf{x}^0$ the system has a solution. Note that the last inequality $\left|\mathbf{a}^3 \cdot \mathbf{x}' - \mathbf{a}^3 \cdot \mathbf{x}\right| < \delta$ of the system can be satisfied with the equality $\mathbf{a}^3 \cdot \mathbf{x}' = \mathbf{a}^3 \cdot \mathbf{x}$ if $a^3 = (0;0;1)$.

It should be remarked that the results of [18] tell nothing about necessary conditions for proximinality of the spaces considered there. To fill this gap in our case, we want to give a necessary condition for proximinality of $\mathcal{R}_c(X)$ in $C(X)$. Our result will be based on the result of Marshall and O'Farrell given below. First, let us introduce some notation. By $\mathcal{R}_c^i$, $i = 1, 2$, we will denote the set of continuous ridge functions $g\left(\mathbf{a}^i \cdot \mathbf{x}\right)$ on the given compact set $X \subset \mathbb{R}^n$. Note that $\mathcal{R}_c = \mathcal{R}_c^1 + \mathcal{R}_c^2$. Besides, let $\mathcal{R}_c^3 = \mathcal{R}_c^1 \cap \mathcal{R}_c^2$. For $i = 1, 2, 3$, let X_i be the quotient space obtained by identifying points y_1 and y_2 in X whenever $f(y_1) = f(y_2)$ for each f in $\mathcal{R}_c^i$. By π_i denote the natural projection of X onto X_i, $i = 1, 2, 3$. Note that we have already dealt with the quotient spaces X_1, X_2 and the projections π_1, π_2 in the previous section (see the proof of Theorem 9). The relation on X, defined by setting $y_1 \approx y_2$ if y_1 and y_2 belong to some path, is an equivalence relation. According to Marshall and O'Farrell [52,53], the equivalence classes are called orbits. By $O(t)$ denote the orbit of X containing t. For $Y \subset X$, let $var_Y \ f$ be the variation of a function f on the set Y. That is,

$$\underset{Y}{var} f = \sup_{x,y \in Y} |f(x) - f(y)|.$$

Theorem 11. *[26]. Suppose that the space $\mathcal{R}_c(X)$ is proximinal in $C(X)$. Then there exists a positive real number c such that*

$$\sup_{t \in X} \underset{O(t)}{var} f \leq c \sup_{t \in X} \underset{\pi_2^{-1}(\pi_2(t))}{var} f, \tag{51}$$

for all f in $\mathcal{R}_c^1$.

The proof is simple. In [53], Marshall and O'Farrell proved the following result (see Proposition 4 in [53]): Let A_1 and A_2 be closed subalgebras of $C(X)$ that contain the constants. Let (X_1, π_1), (X_2, π_2) and (X_3, π_3) be the quotient spaces and projections associated with the algebras A_1, A_2 and $A_3 = A_1 \cap A_2$, respectively. Then $A_1 + A_2$ is closed in $C(X)$ if and only if there exists a positive real number c such that

$$\sup_{z \in X_3} \underset{\pi_3^{-1}(z)}{var} f \leq c \sup_{y \in X_2} \underset{\pi_2^{-1}(y)}{var} f, \tag{52}$$

for all f in A_1.

If $\mathcal{R}_c(X)$ is proximinal in $C(X)$, then it is necessarily closed and therefore, by the above proposition, (52) holds for the algebras $A_1^i = \mathcal{R}_c^i,\ i = 1, 2, 3$. The right-hand side of (52) is equal to the right-hand side of (51). Let t be some point in X and $z = \pi_3(t)$. Since each function $f \in \mathcal{R}_c^3$ is constant on the orbit of t (note that f is both of the form $g_1\left(\mathbf{a}^1 \cdot \mathbf{x}\right)$ and of the form $g_2\left(\mathbf{a}^2 \cdot \mathbf{x}\right)$), $O(t) \subset \pi_3^{-1}(z)$. Hence,

$$\sup_{t \in X} \underset{O(t)}{var} f \leq c \sup_{z \in X_3} \underset{\pi_3^{-1}(z)}{var} f. \tag{53}$$

From (52) and (53) we obtain (51).

Note that the inequality (52) provides not worse but less practicable necessary condition for proximinality than the inequality (51) does. On the other hand, there are many cases in which both the inequalities are equivalent. For example, let the lengths of irreducible paths of X are bounded by some positive integer n_0. In this case, it can be shown that the inequality (52), hence (51), holds with the constant $c = \frac{n_0}{2}$ and, moreover, $O(t) = \pi_3^{-1}(z)$ for all $t \in X$, where $z = \pi_3(t)$ (see the proof of Theorem 5 in [32]). Therefore, the inequalities (51) and (52) are equivalent for the considered class of sets X. The last argument shows that all the compact sets $X \subset \mathbb{R}^n$ over which $\mathcal{R}_c(X)$ is not proximinal in $C(X)$ should be sought in the class of sets having irreducible paths consisting of sufficiently large number of points. For example, let $I = [0; 1]^2$ be the unit square, $a^1 = (1; 1),\ a^2 = (1; \frac{1}{2})$. Consider the path

$$l_k = \{(1; 0), (0; 1), (\frac{1}{2}; 0), (0; \frac{1}{2}), (\frac{1}{4}; 0), ..., (0; \frac{1}{2^k})\}.$$

It is clear that l_k is an irreducible path with the length $2k + 2$, where k may be very large. Let g_k be a continuous univariate function on $\mathbb{R}$ satisfying the conditions: $g_k(\frac{1}{2^{k-i}}) = i,\ i = 0, ..., k,\ g_k(t) = 0$ if $t < \frac{1}{2^k},\ i - 1 \leq g_k(t) \leq i$ if $t \in (\frac{1}{2^{k-i+1}}, \frac{1}{2^{k-i}}),\ i = 1, ..., k$, and $g_k(t) = k$ if $t > 1$. Then it can be easily verified that

$$\sup_{t \in X} \underset{\pi_2^{-1}(\pi_2(t))}{var} g_k(\mathbf{a}^1 \cdot \mathbf{x}) \leq 1. \tag{54}$$

Since $\max_{\mathbf{x} \in I} g_k(\mathbf{a}^1 \cdot \mathbf{x}) = k$, $\min_{\mathbf{x} \in I} g_k(\mathbf{a}^1 \cdot \mathbf{x}) = 0$ and $var_{\mathbf{x} \in O(t_1)} g_k(\mathbf{a}^1 \cdot \mathbf{x}) = k$ for $t_1 = (1; 0)$, we obtain that

$$\sup_{t \in X} \underset{O(t)}{var} g_k(\mathbf{a}^1 \cdot \mathbf{x}) = k. \tag{55}$$

Since k may be very large, from (54) and (55) it follows that the inequality (51) cannot hold for the function $g_k(\mathbf{a}^1 \cdot \mathbf{x}) \in \mathcal{R}_c^1$. Thus, the space $\mathcal{R}_c(I)$ with the directions $a^1 = (1; 1)$ and $a^2 = (1; \frac{1}{2})$ is not proximinal in $C(I)$.

It should be remarked that if a compact set $X \subset \mathbb{R}^n$ satisfies the hypothesis of Theorem 10, then the length of all irreducible paths are uniformly bounded (see the proof of Theorem 10 and lemma in [18]). We have already seen that if the last condition does not hold, then the proximinality of both $\mathcal{R}_c(X)$ in $C(X)$ and $\mathcal{R}_b(X)$ in $B(X)$ fail for some sets X. Besides the examples given above, one can easily construct many other examples of such sets. All these examples, Theorems 9, 10, 11 and the following remarks justify the statement of the following conjecture:

Conjecture. *Let X be some subset of $\mathbb{R}^n$. The space $\mathcal{R}_b(X)$ is proximinal in $B(X)$ and the space $\mathcal{R}_c(X)$ is proximinal in $C(X)$ (in this case, X is considered to be compact) if and only if the lengths of all irreducible paths of X are uniformly bounded.*

After completion of our work [26], Medvedev's result came to our attention (see [41, p.58]). His result, in particular, states that the set $R_c(X)$ is closed in $C(X)$ if and only if the lengths of all irreducible paths of X are uniformly bounded. Thus, in the case of $C(X)$, the necessity of the above conjecture was proved by Medvedev.

Note that there are situations in which a continuous function (a specially chosen function on a specially constructed set) has an extremal element in $\mathcal{R}_b(X)$, but not in $\mathcal{R}_c(X)$ (see [41, p.73]). One subsection of [41] (see p.68) is devoted to the proximinality of sums of two univariate functions with continuous and bounded summands in the spaces of continuous and bounded bivariate functions, respectively. If $X \subset \mathbb{R}^2$ and $\mathbf{a}^1, \mathbf{a}^2$ are linearly independent directions in $\mathbb{R}^2$, then the linear transformation $y_1 = \mathbf{a}^1 \cdot \mathbf{x}$, $y_2 = \mathbf{a}^2 \cdot \mathbf{x}$ reduces the problems of proximinality of $\mathcal{R}_b(X)$ in $B(X)$ and $\mathcal{R}_c(X)$ in $C(X)$ to the problems considered in that subsection. But in general, when $X \subset \mathbb{R}^n$, $n > 2$, our case cannot be obtained from that of [41].

References

[1] Aumann G., Uber approximative nomographie, II, Bayer. Akad. Wiss. Math.-Nat. Kl. S.-B. (1959), 103-109.

[2] Babaev M-B.A., Approximation of polynomials in two variables by functions of the form $\varphi(x) + \psi(y)$, (Russian) Dokl. Akad. Nauk. SSSR 193 (1970), 967-969; English transl. Soviet. Math. Dokl. 11 (1970), 1034-1036.

[3] Babaev M-B.A., Sharp estimates for the approximation of functions of several variables by sums of functions of a lesser number of variables, (Russian) Mat. zametki, 12 (1972), 105-114.

[4] Braess D. and Pinkus A., Interpolation by ridge functions, J.Approx. Theory 73 (1993), 218-236.

[5] Buck R.C., On approximation theory and functional equations, J.Approx. Theory, 5(1972),228-237.

[6] Buhmann M.D. and Pinkus A., Identifying linear combinations of ridge functions, Advances in Applied Math. 22 (1999), 103-118.

[7] Candes E. J., Ridgelets: estimating with ridge functions, Ann. Statist. 31 (2003), 1561-1599.

[8] Candes E. J., Ridgelets: theory and applications. Ph.D. Thesis, Technical Report, Department of Statistics, Stanford University.

[9] Chui C. K. and Li X., Approximation by ridge functions and neural networks with one hidden layer, J.Approx. Theory. 70 (1992), 131-141.

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

[11] Diliberto S.P. and Straus E.G., On the approximation of a function of several variables by the sum of functions of fewer variables, Pacific J.Math. 1 (1951), 195-210.

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

[13] Dyn N., A straightforward generalization of Diliberto and Straus algorithm does not work, J.Approx. Theory, 30(1980),247-250.

[14] Dyn N., Light W.A. and Cheney E.W., Interpolation by piecewise-linear radial basis functions, J.Approx. Theory. 59 (1989), 202-223.

[15] Friedman J. H. and Tukey J. W., A Projection Pursuit Algorithm for Exploratory Data Analysis, IEEE Transactions on Computers C-23 (1974), 881–890.

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

[17] Garkavi A.L., Medvedev V.A., Khavinson S.Ya., On the existence of a best uniform approximation of functions of two variables by sums of the type $\varphi(x)+\psi(y)$, (Russian) Sibirskii Mat. Zh., 36 (1995),819-827; English transl. in Siberian Math. J., 36 (1995), 707-713.

[18] Garkavi A.L., Medvedev V.A., Khavinson S.Ya., On the existence of a best uniform approximation of a function of several variables by the sum of functions of fewer variables, (Russian) Matematicheski Sbornik 187 (1996), 3–14; English transl. in Sbornik Mathematics 187 (1996), 623–634.

[19] Golitschek M.v. and Light W.A., Approximation by solutions of the planar wave equation, SIAM J.Numer. Anal. 29 (1992), 816-830.

[20] Golomb M., Approximation by functions of fewer variables, On numerical approximation. Proceedings of a Symposium. Madison 1959. Edited by R.E.Langer. The University of Wisconsin Press, 275-327.

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

[22] Ismailov V. E., Approximation by neural networks with weights varying on a finite set of directions, Journal of Mathematical Analysis and Applications, 389 (2012), 72-83.

[23] Ismailov V. E., A note on the representation of continuous functions by linear superpositions, Expositiones Mathematicae 30 (2012), 96-101.

[24] Ismailov V. E., Approximation by Neural Networks with a Restricted Set of Weights, book chapter (chapter 6), in: Advances in Mathematics Research (published by NOVA Science Publishers, USA), Volume 16, 2011, pp. 193-206.

[25] Ismailov V. E., Approximation capabilities of neural networks with weights from two directions. Azerbaijan Journal of Mathematics 1 (2011), no. 1, 122–128.

[26] Ismailov V. E., On the proximinality of ridge functions, Sarajevo Journal of Mathematics 5(17) (2009), no. 1, 109-118.

[27] Ismailov V. E., On the theorem of M. Golomb, Proceedings of the Indian Academy of Sciences (Math. Sci.), 119 (2009), no. 1, 45-52.

[28] Ismailov V. E., On the representation by linear superpositions, Journal of Approximation Theory 151 (2008), 113-125.

[29] Ismailov V. E., On the approximation by weighted ridge functions. Analele Universitatii de Vest din Timisoara, Ser. Mat.-Inform. 46 (2008), 75–83.

[30] Ismailov V. E., On the approximation by compositions of fixed multivariate functions with univariate functions, Studia Mathematica 183 (2007), 117-126.

[31] Ismailov V. E., On the best L_2 approximation by ridge functions, Applied Mathematics E-Notes, 7 (2007), 71-76.

[32] Ismailov V. E., Representation of multivariate functions by sums of ridge functions, Journal of Mathematical Analysis and Applications 331 (2007), 184-190.

[33] Ismailov V. E., Characterization of an extremal sum of ridge functions, Journal of Computational and Applied Mathematics 205 (2007), 105-115.

[34] Ismailov V. E., On error formulas for approximation by sums of univariate functions, International Journal of Mathematics and Mathematical Sciences, volume 2006 (2006), Article ID 65620, 11 pp.

[35] Ismailov V. E., Methods for computing the least deviation from the sums of functions of one variable, (Russian) Sibirskii Matematicheskii Zhurnal 47 (2006), 1076–1082; translation in Siberian Mathematical Journal 47 (2006), 883-888.

[36] Ismailov V. E., On some classes of bivariate functions characterized by formulas for the best approximation. Radovi Matematicki 13 (2004), 53-62.

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

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

[39] Kazantsev I., Tomographic reconstruction using ridge functions, Proceedings of 1st World Congress on Industrial Process Tomography, Buxton, Derbishyre, UK, April 14-17, 1999, pp. 433-437.

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

[41] Khavinson S. Ya., Best approximation by linear superpositions (approximate nomography), Translated from the Russian manuscript by D. Khavinson. Translations of Mathematical Monographs, 159. American Mathematical Society, Providence, RI, 1997, 175 pp.

[42] Khavinson S.Ya., A Chebyshev theorem for the approximation of a function of two variables by sums of the type $\varphi(x) + \psi(y)$, (Russian) Izv. Acad. Nauk. SSSR Ser. Mat. 33 (1969), 650-666; English tarnsl. Math. USSR Izv. 3 (1969), 617-632.

[43] Klopotowski A., Nadkarni M. G., Bhaskara Rao K. P. S., 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.

[44] Kolmogorov A. N., Asymptotic characteristics of some completely bounded metric spaces, (Russian) Dokl. Akad. Nauk SSSR 108 (1956), 585-589.

[45] Kolmogorov A. N. and Tikhomirov V. M., ε-entropy and ε-capacity of sets in function spaces, (Russian) Uspehi Mat. Nauk no. 2 (86), 14 (1959), 3-86; English transl., Amer. Math. Soc. Transl. (2) 17 (1961), 277-364.

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

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

[48] Maiorov V. and Pinkus A., Lower bounds for approximation by MLP neural networks, Neurocomputing 25 (1999), 81-91.

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

[50] Maiorov V.E., Geometric properties of the ridge function manifold. Adv. Comput. Math. 32 (2010), 239–253.

[51] Marr R. B., 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.

[52] Marshall D.E. and O'Farrell A.G., Approximation by a sum of two algebras. The lightning bolt principle, J. Funct. Anal. 52 (1983), 353-368.

[53] Marshall D.E. and O'Farrell A.G., Uniform approximation by real functions, Fund. Math. 104 (1979),203-211.

[54] Medvedev V.A., Refutation of a theorem of Diliberto and Straus, Mat. zametki, 51(1992), 78-80; English transl. Math. Notes 51(1992), 380-381.

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

[56] Oskolkov K. I., Ridge approximation, Fourier-Chebyshev analysis, and optimal quadrature formulas. (Russian), Tr. Mat. Inst. Steklova 219 (1997), 269–285; translation in Proc. Steklov Inst. Math. 219 (1997), 265–280.

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

[58] Pinkus A., 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.

[59] Pinkus A., Approximation theory of the MLP model in neural networks, Acta Numerica. 8 (1999), 143-195.

[60] Pinkus A., Smoothness and uniqueness in ridge function representation, Indagationes Mathematicae
(to appear, see http://dx.doi.org/10.1016/j.indag.2012.10.004)

[61] Rivlin T.J. and Sibner R.J., The degree of approximation of certain functions of two variables by a sum of functions of one variable, Amer. Math. Monthly 72 (1965), 1101-1103.

[62] Sanguineti M., Universal Approximation by Ridge Computational Models and Neural Networks: A Survey, The Open Applied Mathematics Journal , vol. 2 (2008), no. 1, pp. 31-58.

[63] Shin Y., Ghosh J., Ridge polynomial networks, IEEE Transactions on Neural Networks 6 (1995), 610-22.

[64] Sternfeld Y., Uniform separation of points and measures and representation by sums of algebras. Israel J. Math. 55 (1986), 350–362.

[65] Sternfeld Y., Uniformly separating families of functions. Israel J. Math. 29 (1978), 61–91.

[66] Temlyakov V. N., On approximation by ridge functions, Preprint. Department of Mathematics, University of South Carolina, 1996.

[67] Tikhomirov V. M., The works of A. N. Kolmogorov on ε-entropy of function classes and superpositions of functions. (Russian) Uspehi Mat. Nauk 18 (1963) no. 5 (113), 55–92.

[68] Vostrecov, B. A., Kreĭnes M. A., Approximation of continuous functions by superpositions of plane waves (Russian), Dokl. Akad. Nauk SSSR 140 (1961), 1237-1240; English translation in Soviet Math. Dokl. 2 (1961), 1326-1329.

[69] Wu W., Feng G., Li X., Training multilayer perceptrons via minimization of sum of ridge functions, Adv., Comput. Math. 17 (2002), 331-347.

[70] Xie T.F., Cao F.L., The ridge function representation of polynomials and an application to neural networks, Acta Math. Sin. (Engl. Ser.) 27 (2011), no. 11, 2169–2176.

Vugar E. Ismailov
Institute of Mathematics and Mechanics, National Academy of Sciences of Azerbaijan, Az-1141, Baku, Azerbaijan
E-mail: vugaris@mail.ru

Received 23 December 2012
Published 07 January 2013