DE GRUYTER
OPEN

DOI 10.1515/tmj-2017-0030

On the error of approximation by ridge functions with two fixed directions

Aida Kh. Asgarova[1], Arzu M-B. Babayev[2] and Ibrahim K. Maharov[3]

Institute of Mathematics and Mechanics, National Academy of Sciences of Azerbaijan, Az-1141, Baku, Azerbaijan
E-mail: aidaasgarova@gmail.com[1], arzumb.babayev@gmail.com[2], ibrahimmaharov@gmail.com[3]

Abstract

We consider the problem of approximation of a continuous multivariate function by sums of two ridge functions in the uniform norm. We obtain a formula for the approximation error in terms functionals generated by closed paths.

2010 Mathematics Subject Classification. **41A30**. 41A50, 41A63
Keywords. Ridge function, path, extremal element, approximation error.

1 Introduction

In modern approximation theory, ridge functions play an essential role. A *ridge function* is a multivariate function of the form

$$G(\mathbf{x}) = g(\mathbf{a} \cdot \mathbf{x}) = g(a_1 x_1 + \ldots + a_d x_d),$$

where $g : \mathbb{R} \to \mathbb{R}$ and $\mathbf{a} = (a_1, ..., a_d)$ is a fixed vector (direction) in $\mathbb{R}^d \setminus \{\mathbf{0}\}$. In other words, a ridge function is a multivariate function constant on the parallel hyperplanes $\mathbf{a} \cdot \mathbf{x} = c,\ c \in \mathbb{R}$. These functions and their linear combinations arise naturally in problems of computerized tomography (see, e.g., [26, 31]), statistics (see, e.g., [5, 9, 10, 15]), partial differential equations [24] (where they are called *plane waves*), neural networks (see, e.g., [6, 16, 33, 35] and references therein), and approximation theory (see, e.g., [6, 7, 13, 19, 21, 25, 27, 32, 33, 34, 37]).

Consider the following set of 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 obtain a formula for computation of the approximation error

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

Recall that if there exists $g_0 \in \mathcal{R}$ such that

$$\|f - g_0\| = E(f),$$

then g_0 is called an extremal element.

Tbilisi Mathematical Journal 10(2) (2017), pp. 111–120.
Tbilisi Centre for Mathematical Sciences.

Received by the editors: 03 November 2016.
Accepted for publication: 15 March 2017.

The approximation problem concerning the set $\mathcal{R}(\mathbf{a},\mathbf{b})$ arises in other problems too. Buck [4] considered the classical functional equation: given $s \in C[0,1]$, $0 \leq s(t) \leq 1$, for which $u \in C[0,1]$ does there exist $\varphi \in C[0,1]$ such that

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

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

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

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 $K = \{(x,y) : y = x \text{ or } y = s(x), \ 0 \leq x \leq 1\}$.

One can observe that if $d = 2$, $\mathbf{a}$ and $\mathbf{b}$ coincide with the coordinate directions, then the functions $g_1(\mathbf{a} \cdot \mathbf{x})$ and $g_2(\mathbf{b} \cdot \mathbf{x})$ are univariate. We see that the approximation of a bivariate function by sums of univariate functions is a special case of the approximation problem considered in this paper. It should be remarked that there are many papers devoted to this subject (see, e.g., [2, 8, 11, 14, 22, 23, 28, 29, 36] and references therein).

2 The approximation error formula

Suppose Q is a compact set in $\mathbb{R}^d$ and $\mathbf{a}, \mathbf{b} \in \mathbb{R}^d \backslash \{\mathbf{0}\}$ are fixed directions.

Definition 2.1. *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* $\mathbf{b}$.

Paths with respect to two directions in $\mathbb{R}^2$ were first considered by Braess and Pinkus [3]. They showed that paths give geometric means of deciding if a set of points $\{\mathbf{x}^i\}_{i=1}^m \subset \mathbb{R}^2$ has the "non-interpolation property" (for this terminology see [3]). Ismailov and Pinkus [17] used these objects to solve the problem of interpolation on straight lines by ridge functions with two fixed directions. If $\mathbf{a}$ and $\mathbf{b}$ are the coordinate vectors in $\mathbb{R}^2$, then the objects in Definition 2.1 turn into "bolts of lightning" (see, e.g., [1, 5, 29]). It is well known that the idea of bolts was first introduced by Diliberto and Straus [8] and played an essential role in problems of approximation by sums of univariate functions (see, e.g., [8, 11, 14, 22, 23, 28, 29]). Note that the name "bolt of lightning" is due to Arnold [1]. Ismailov [18, 20] generalized paths to those with respect to a finite set of functions. Paths with respect to n arbitrarily fixed functions turned out to be very useful in problems of representation by linear superpositions.

In the sequel, we use the term "path" instead of the long expression "path with respect to the directions $\mathbf{a}$ and $\mathbf{b}$". 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 also a path. A path $(\mathbf{p}_1, ..., \mathbf{p}_n)$ in a set Q is called extensible if there exist points $\mathbf{y}, \mathbf{z} \in Q$ such that $(\mathbf{y}, \mathbf{p}_1, ..., \mathbf{p}_n, \mathbf{z})$ is a path. For example, in a square $ABCD$ with the vertices $(1,0)$, $(0,1)$, $(-1,0)$, $(0,-1)$, the set joining middle points of the sides AB, BC, CD and AD forms a closed path. Any path $(\mathbf{p}_1, ..., \mathbf{p}_n) \subset ABCD$ with $\mathbf{p}_1$ and $\mathbf{p}_n$ different from A, B, C, D, is extensible.

We associate a 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 $\|G_p\| = 1$.

To prove our main result we need two auxiliary lemmas from [21].

Lemma 2.1. *Let a compact set Q have closed paths. Then*

$$\sup_{p \subset Q} |G_p(f)| \leq E(f), \tag{2.1}$$

where the sup is taken over all closed paths. Moreover, inequality (2.1) is sharp, i.e. there exist functions for which (2.1) turns into equality.

Lemma 2.2. *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*

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

Then the functions

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

are defined and continuous on T_h.

The following theorem is valid.

Theorem 2.1. *Let $Q \subset \mathbb{R}^d$ be a convex compact set and $f \in C(Q)$. Assume the following conditions hold.*
1) there exists an extremal element $g_0 \in \mathcal{R}(\mathbf{a},\mathbf{b})$ for the function f;
2) for any extensible path $q = (\mathbf{q}_1, ..., \mathbf{q}_n) \subset Q$ there exist points $\mathbf{q}_{n+1}, \mathbf{q}_{n+2}, ..., \mathbf{q}_{n+s} \in Q$ such that $(\mathbf{q}_1, ..., \mathbf{q}_n, \mathbf{q}_{n+1}, ..., \mathbf{q}_{n+s})$ is a closed path and s is not more than some positive integer n_0 independent of q.
Then the approximation error can be computed by the formula

$$E(f) = \sup_{p \subset Q} |G_p(f)|,$$

where the sup is taken over all closed paths.

Proof. For brevity of the exposition, in the sequel, we use the concept of "an extremal path". A finite or infinite path $(\mathbf{p}_1, \mathbf{p}_2, ...)$ is said to be extremal for a function $u \in C(Q)$ if $u(\mathbf{p}_i) = (-1)^i \|u\|, i = 1, 2, ...$ or $u(\mathbf{p}_i) = (-1)^{i+1} \|u\|, i = 1, 2, ...$ (see [21]). Regarding extremal paths for the function $f_1 = f - g_0$, there are only two possible options. The first option is when there exists a closed path $p_0 = (\mathbf{p}_1, ..., \mathbf{p}_{2n})$ extremal for the function f_1. In this case, it is easy to see that

$$|G_{p_0}(f)| = |G_{p_0}(f - g_0)| = \|f - g_0\| = E(f).$$

Considering this, the assertion of the theorem follows from (2.1). The second option is when there does not exist a closed path extremal for the function f_1. Let us prove that in this case, there exists an infinite path extremal for f_1. 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. Here by length of a path we mean its number of points. Define the following functions:

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

where

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

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

Note that by Lemma 2.2, all the above functions $f_n(\mathbf{x})$, $n = 2, 3, ...,$ are continuous on Q. Since g_0 is an extremal element for f, the equality $\|f_1\| = E\,(f)$ holds. Let us show that $\|f_2\| = E\,(f)$. Indeed, for any $\mathbf{x} \in Q$

$$f_1(\mathbf{x}) - g_{1,1}(\mathbf{a}\cdot\mathbf{x}) \le \frac{1}{2}\left(\max_{\substack{\mathbf{y}\in Q\\ \mathbf{a}\cdot\mathbf{y}=\mathbf{a}\cdot\mathbf{x}}} f_1(\mathbf{y}) - \min_{\substack{\mathbf{y}\in Q\\ \mathbf{a}\cdot\mathbf{y}=\mathbf{a}\cdot\mathbf{x}}} f_1(\mathbf{y})\right) \le E(f) \tag{2.6}$$

and

$$f_1(\mathbf{x}) - g_{1,1}(\mathbf{a}\cdot\mathbf{x}) \ge \frac{1}{2}\left(\min_{\substack{\mathbf{y}\in Q\\ \mathbf{a}\cdot\mathbf{y}=\mathbf{a}\cdot\mathbf{x}}} f_1(\mathbf{y}) - \max_{\substack{\mathbf{y}\in Q\\ \mathbf{a}\cdot\mathbf{y}=\mathbf{a}\cdot\mathbf{x}}} f_1(\mathbf{y})\right) \ge -E(f). \tag{2.7}$$

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

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

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

and

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

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

Using (2.6) and (2.7) in the last two inequalities, we obtain that for any $\mathbf{x} \in Q$

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

Thus,

$$\|f_2\| \leq E(f). \tag{2.8}$$

Since $f_2 - f \in \mathcal{R}(\mathbf{a},\mathbf{b})$, it follows from (2.8) that $\|f_2\| = E(f)$.

Similarly, one can show that $\|f_3\| = E(f)$, $\|f_4\| = E(f)$, and so on. Thus, $\|f_n\| = E(f)$ for all $n = 1, 2, ...$

Let us now prove the following implications

$$f_1(\mathbf{p}_0) < E(f) \Rightarrow f_2(\mathbf{p}_0) < E(f) \tag{2.9}$$

and

$$f_1(\mathbf{p}_0) > -E(f) \Rightarrow f_2(\mathbf{p}_0) > -E(f), \tag{2.10}$$

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

$$f_1(\mathbf{p}_0) < E(f) \Rightarrow f_1(\mathbf{p}_0) - g_{1,1}(\mathbf{a}\cdot\mathbf{p}_0) < E(f). \tag{2.11}$$

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, $g_{1,1}(\mathbf{a}\cdot\mathbf{p}_0) = 0$. Therefore,

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

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 $\varepsilon_1,\ \varepsilon_2 \geq 0$ and $\varepsilon_1 + \varepsilon_2 \neq 0$. In this case,

$$f_1(\mathbf{p}_0) - g_{1,1}(\mathbf{a}\cdot\mathbf{p}_0) \leq \max_{\substack{\mathbf{y}\in Q\\ \mathbf{a}\cdot\mathbf{y}=\mathbf{a}\cdot\mathbf{p}_0}} f_1(\mathbf{y}) - g_{1,1}(\mathbf{a}\cdot\mathbf{p}_0) = \frac{1}{2}\left(\max_{\substack{\mathbf{y}\in Q\\ \mathbf{a}\cdot\mathbf{y}=\mathbf{a}\cdot\mathbf{p}_0}} f_1(\mathbf{y}) - \min_{\substack{\mathbf{y}\in Q\\ \mathbf{a}\cdot\mathbf{y}=\mathbf{a}\cdot\mathbf{p}_0}} f_1(\mathbf{y})\right)$$

$$= E(f) - \frac{\varepsilon_1 + \varepsilon_2}{2} < E(f).$$

Thus we have proved (2.11). Using the same method, we can also prove that

$$f_1(\mathbf{p}_0) - g_{1,1}(\mathbf{a}\cdot\mathbf{p}_0) < E(f) \Rightarrow f_1(\mathbf{p}_0) - g_{1,1}(\mathbf{a}\cdot\mathbf{p}_0) - g_{2,1}(\mathbf{b}\cdot\mathbf{p}_0) < E(f). \tag{2.12}$$

Implications (2.11) and (2.12) yield (2.9). By the same way one can prove the validity of (2.10). From implications (2.9) and (2.10) it follows 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 is extremal for f_1.

We supposed above that any path extremal for f_1 has the length not more than N. Let us show that in his case, 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 this path 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 we have shown above, the path q is extremal for f_1. Assume $f_1(\mathbf{q}_N) = E(f)$. Then there is not a point $\mathbf{q}_0 \in Q$ 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 $\mathbf{q}_0 \notin q$, then 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. On the other hand, if there was such $\mathbf{q}_0$ and $\mathbf{q}_0 \in q$, then from points of q we could form a closed extremal path for f_1, which would contradict our assumption that there does not exist a closed extremal path for f_1. Hence we conclude that

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

Therefore,

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

From the last inequality, by the similar way as above, one can obtain that

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

This means that the path $(\mathbf{q}_1, \mathbf{q}_2, ..., \mathbf{q}_N)$ can not be extremal for f_2. Thus 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 obtain that there is not a path extremal for f_{N+1}. Then there is not a point $\mathbf{p}_0 \in Q$ such that $|f_{N+1}(\mathbf{p}_0)| = \|f_{N+1}\|$. But the norm $\|f_{N+1}\|$ must be attained, since by Lemma 2.2, all the functions f_2, $f_3, ..., f_{N+1}$ are continuous on the compact set Q. The obtained contradiction means that there exists an infinite path extremal for f_1.

Let a path $p = (\mathbf{p}_1, \mathbf{p}_2, ..., \mathbf{p}_n, ...)$ be infinite and extremal for f_1. Note that all the points $\mathbf{p}_i$ must be distinct, otherwise we could form a closed extremal path, contrary to our assumption. Without loss of generality we may assume that this path is extensible (if it is not, we may start with the path $(\mathbf{p}_2, ..., \mathbf{p}_n, ...)$). Consider the sequence $p_n = (\mathbf{p}_1, \mathbf{p}_2, ..., \mathbf{p}_n)$, $n = 1, 2, ...$, of finite paths. Since there exists an extremal element By condition (2) of the theorem, for each path p_n there exists a closed path $p_n^{m_n} = (\mathbf{p}_1, \mathbf{p}_2, ..., \mathbf{p}_n, \mathbf{q}_{n+1}, ..., \mathbf{q}_{n+m_n})$, where $m_n \leq n_0$. The functional $G_{p_n^{m_n}}$ obeys the inequalities

$$\left|G_{p_n^{m_n}}(f)\right| = \left|G_{p_n^{m_n}}(f - g_0)\right| \leq \frac{n\,\|f - g_0\| + m_n\,\|f - g_0\|}{n + m_n} = \|f - g_0\| \tag{2.13}$$

and

$$\left|G_{p_n^{m_n}}(f)\right| \geq \frac{n\,\|f - g_0\| - m_n\,\|f - g_0\|}{n + m_n} = \frac{n - m_n}{n + m_n}\,\|f - g_0\|\,. \tag{2.14}$$

We obtain from (2.13) and (2.14) that

$$\sup_{p_n^{m_n}} \left|G_{p_n^{m_n}}(f)\right| = \|f - g_0\|\,. \tag{2.15}$$

Since g_0 is an extremal element, it follows from (2.15) and Lemma 2.1 that

$$E\,(f) = \sup_{p\subset Q} |G_p(f)|\,,$$

where the sup is taken over all closed paths of Q. The theorem has been proved.

Remark 1. Theorem 2.1 generalizes the result of Diliberto and Straus (see [8, Theorem 1]) from the sum of univariate functions to the sum of ridge functions.

Remark 2. The question if there exists an extremal element $g_0 \in \mathcal{R}$ for f is far from trivial. Some conditions on Q sufficient for the existence of an extremal element for each f in $C(Q)$ may be found in [19].

Note that the hypothesis on the set Q "for any extensible 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, take the unit square $\mathbb{I}^2 = [0;1]^2$ and fix the directions $\mathbf{a} = \left(1; \frac{1}{2}\right)$ and $\mathbf{b} = \left(\frac{1}{2}; \frac{1}{2}\right)$. In this case, the vertex $(1;1)$ is not reached with any of the directions orthogonal to $\mathbf{a}$ and $\mathbf{b}$ respectively. Therefore, for any positive integer n_0 and any point $\mathbf{q}_0$ in $\mathbb{I}^2$ one can chose a point $\mathbf{q}_1 \in \mathbb{I}^2$ from a sufficiently small neighborhood of $(1;1)$ so that any path containing $\mathbf{q}_0$ and $\mathbf{q}_1$ has the length more than n_0. In general, if a compact convex set $Q \subset \mathbb{R}^2$ satisfies the second condition of the theorem, then any point in the boundary of Q must be reached with at least one of the two directions orthogonal to $\mathbf{a}$ and $\mathbf{b}$ respectively. In the $d-$dimensional space, $d > 2$, there are many directions orthogonal to $\mathbf{a}$ and $\mathbf{b}$. In this case, the condition requires that any point in the boundary of Q should be reached with at least one direction orthogonal to $\mathbf{a}$ or $\mathbf{b}$. It should be remarked that if the space $\mathcal{R}(\mathbf{a},\mathbf{b})$ is proximinal in $C(Q)$ (that is, if any function in $C(Q)$ has an extremal element from $\mathcal{R}(\mathbf{a},\mathbf{b})$), then the second condition can be removed, which shows the following corollary.

Corollary 2.1. *Let $Q \subset \mathbb{R}^d$ be a convex compact set and $\mathcal{R}(\mathbf{a},\mathbf{b})$ is proximinal in $C(Q)$. Then the assertion of Theorem 2.1 holds for each function $f \in C(Q)$.*

The proof immediately follows from the result that if $\mathcal{R}(\mathbf{a},\mathbf{b})$ is proximinal in $C(Q)$, then the lengths of irreducible paths are uniformly bounded (see [19]). Note that a path $(\mathbf{p}_1, ..., \mathbf{p}_n)$ is irreducible if any path connecting $\mathbf{p}_1$ and $\mathbf{p}_n$ has equal to or more than n points. If the lengths of irreducible paths are uniformly bounded by some positive integer n_0, then for $n > n_0$ extremal paths $p_n = (\mathbf{p}_1, \mathbf{p}_2, ..., \mathbf{p}_n)$, known from the proof of Theorem 2.1, must be made closed by adding not more than n_0 points. But as we see above, this leads to the assertion of Theorem 2.1.

Acknowledgement

This work was supported by the Research Program Competition launched by the National Academy of Sciences of Azerbaijan (Program: Approximation by neural networks and some problems of frames).

References

[1] V. I. Arnold, On functions of three variables, *Dokl. Akad. Nauk SSSR* **114** (1957), 679-681; English transl, *Amer. Math. Soc. Transl.* **28** (1963), 51-54.

[2] M-B. A. Babaev, 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.

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

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

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

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

[7] W. Dahmen and C. A. Micchelli, Some remarks on ridge functions, *Approx. Theory Appl.* **3** (1987), 139-143.

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

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

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

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

[12] M. Golomb, 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.

[13] Y. Gordon, V. Maiorov, M. Meyer, S. Reisner, On the best approximation by ridge functions in the uniform norm, *Constr. Approx.* **18** (2002), 61-85.

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

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

[16] V. E. Ismailov, Approximation by ridge functions and neural networks with a bounded number of neurons, *Appl. Anal.* **94** (2015), no. 11, 2245-2260.

[17] V. E. Ismailov and A. Pinkus, Interpolation on lines by ridge functions, *J. Approx. Theory* **175** (2013), 91-113.

[18] V. E. Ismailov, A note on the representation of continuous functions by linear superpositions, *Expo. Math.* **30** (2012), 96-101.

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

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

[21] V. E. Ismailov, Characterization of an extremal sum of ridge functions. *J. Comput. Appl. Math.* **205** (2007), no. 1, 105–115.

[22] V. E. Ismailov, On error formulas for approximation by sums of univariate functions, *Int. J. Math. and Math. Sci.*, volume **2006** (2006), Article ID 65620, 11 pp.

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

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

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

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

[27] V. Maiorov, R.Meir and J.Ratsaby, On the approximation of functional classes equipped with a uniform measure using ridge functions, *J.Approx. Theory* **99** (1999), 95-111.

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

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

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

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

[32] B. Pelletier, Approximation by ridge function fields over compact sets, *J.Approx. Theory* **129** (2004), 230-239.

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

[34] A. Pinkus, Approximating by ridge functions, in: *Surface Fitting and Multiresolution Methods,* (A.Le Méhauté, C.Rabut and L.L.Schumaker, eds), Vanderbilt Univ.Press (Nashville),1997,279-292.

[35] A. Pinkus, Approximation theory of the MLP model in neural networks, *Acta Nume-rica.* **8** (1999), 143-195.

[36] T. J. Rivlin and R. J. Sibner, 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.

[37] X. Sun and E. W. Cheney, The fundamentality of sets of ridge functions, *Aequationes Math.* **44** (1992), 226-235.