Proceedings of the Institute of Mathematics and Mechanics,
National Academy of Sciences of Azerbaijan
Volume 50, Number 1, 2024, Pages 152–161
https://doi.org/10.30546/2409-4994.2024.50.1.152

ON A FORMULA FOR THE APPROXIMATION BY RBF NEURAL NETWORKS WITH TWO HIDDEN NODES

AIDA KH. ASGAROVA AND IBRAHIM K. MAHAROV

Abstract. In the context of approximating a continuous multivariate function using radial basis function (RBF) neural networks with two hidden nodes, we focus on evaluating the approximation error in the uniform norm. We derive a formula for this error by utilizing functionals generated by closed paths.

1. Introduction

The pioneering work of Broomhead and Lowe [6] introduced Radial Basis Function (RBF) neural networks, which have gained a reputation as universal approximators due to their exceptional performance in function approximation problems. Initially developed for data interpolation in higher-dimensional spaces, RBF networks have found applications across various engineering domains, serving as a valuable tool for function approximation, prediction, estimation, and system control (see, e.g., [9, 23, 17, 21, 22, 26, 27, 28, 29]).

One of the key advantages of RBF neural networks lies in the simplicity of computing their network parameters. These networks possess the ability to handle intricate nonlinear mappings and offer a swift and reliable learning mechanism, all while maintaining computational efficiency. Consequently, they strike a balance between complexity and computational cost. The set of RBF neural networks considered in this paper consists of the following functions

$$\sum_{i=1}^{m} w_i g\left(\frac{\|\mathbf{x}-\mathbf{c_i}\|}{\sigma_i}-\theta_i\right). \tag{1.1}$$

Here $m \in \mathbb{N}$ is the number of nodes in the hidden layer, $(w_1, ..., w_m) \in \mathbb{R}^m$ is the vector of weights, $\mathbf{x} \in \mathbb{R}^d$ is an input vector, $\mathbf{c_i} \in \mathbb{R}^d$ and $\sigma_i \in \mathbb{R}$ are the centroids and smoothing factor (or width) of the i-th node, $1 \leq i \leq m$, respectively, $\theta_i \in \mathbb{R}$ are thresholds and $g : \mathbb{R} \to \mathbb{R}$ is the so-called activation function.

Radial basis functions (RBFs) are a type of multivariate functions that are constant on spheres centered at a particular point. In other words, for a given center point $\mathbf{c}$ and radius α, the RBF evaluates to a constant value on the sphere defined by the equation $\|\mathbf{x} - \mathbf{c}\| = \alpha$, $\alpha \in \mathbb{R}$, where $\mathbf{x}$ represents the input variables.

2010 *Mathematics Subject Classification.* 41A30, 41A63, 92B20.

Key words and phrases. radial basis function, path, extremal element, approximation error.

Let $g(x)$ be a continuous activation function on $\mathbb{R}$, consider the approximation of the continuous function $f(\mathbf{x}) = f(x_1, ..., x_d)$ on a compact subset $Q \subset \mathbb{R}^d$ using a set of radial basis function (RBF) neural networks:

$$\mathcal{G} = \mathcal{G}(g, \mathbf{c_1}, \mathbf{c_2}) = \left\{ \sum_{i=1}^{m} w_i g\left(\frac{\|\mathbf{x} - \mathbf{c_i}\|}{\sigma_i} - \theta_i \right) : w_i, \sigma_i, \theta_i \in \mathbb{R}; \ \mathbf{c_i} = \mathbf{c_1} \, or \, \mathbf{c_i} = \mathbf{c_2} \right\} \quad (1.2)$$

In equation (1.2), $\mathbf{c}_1$ and $\mathbf{c}_2$ are fixed center points of the radial basis functions used in the RBF neural network. On the other hand, the numbers w_i, σ_i, and θ_i are variables that can vary and need to be determined during the process.

The approximation error is defined as follows:

$$E\,(f) = E(f, \mathcal{G}) \stackrel{def}{=} \inf_{u \in \mathcal{G}} \|f - u\|\,,$$

where

$$\|f - u\| = \max_{\mathbf{x} \in Q} |f(\mathbf{x}) - u(\mathbf{x})|.$$

In this paper, we aim to derive a formula for computing the approximation error in the context of the discussed RBF neural networks. We propose that $E(f)$ can be obtained by evaluating values of specially constructed functionals at the function f.

2. Formula for the approximation error

Assume $Q \subset \mathbb{R}^d$ and $\mathbf{c}_1, \mathbf{c}_2 \in \mathbb{R}^d$ are fixed center points.

Definition 2.1. (see [4]) *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 $\|\mathbf{p}_1 - \mathbf{c}_1\| = \|\mathbf{p}_2 - \mathbf{c}_1\|, \|\mathbf{p}_2 - \mathbf{c}_2\| = \|\mathbf{p}_3 - \mathbf{c}_2\|$, $\|\mathbf{p}_3 - \mathbf{c}_1\| = \|\mathbf{p}_4 - \mathbf{c}_1\|, ...$ or $\|\mathbf{p}_1 - \mathbf{c}_2\| = \|\mathbf{p}_2 - \mathbf{c}_2\|, \|\mathbf{p}_2 - \mathbf{c}_1\| = \|\mathbf{p}_3 - \mathbf{c}_1\|$, $\|\mathbf{p}_3 - \mathbf{c}_2\| = \|\mathbf{p}_4 - \mathbf{c}_2\|, ...$is called a path with respect to the centers $\mathbf{c}_1$ and $\mathbf{c}_2$.*

In the given definition, the distances are alternated between two fixed points. However, paths have many different variations. For example, instead of using points, one can consider two hyperplanes for alternating distances. Specifically, the two hyperplanes are defined as $\mathbf{a}^i \cdot \mathbf{x} = \alpha_i$, $i = 1, 2$. The notation "$\cdot$" denotes the standard scalar product in $\mathbb{R}^d$, and the distances from these hyperplanes can be alternated instead of distances from points. Certainly, in $\mathbb{R}^2$, hyperplanes can be represented by straight lines. Consequently, one can consider distances from straight lines in $\mathbb{R}^2$. Paths involving distances from two straight lines in $\mathbb{R}^2$ were initially explored by Braess and Pinkus in their work [5]. They investigated these paths as a means of determining if a set of points $\{\mathbf{x}^i\}_{i=1}^m \subset \mathbb{R}^2$ possesses the non-interpolation property for "ridge functions". For detailed discussions and analysis regarding these functions, their properties, and their applications see [5, 11, 14, 25]. Ismailov and Pinkus [16] utilized paths with respect to two directions $\mathbf{a}^1$ and $\mathbf{a}^2$ in $\mathbb{R}^d$ to address the problem of interpolation on straight lines using ridge functions. Paths with respect to two directions in $\mathbb{R}^d$ were also employed in other studies [3, 12, 13, 15]. In the context of $\mathbb{R}^2$, if two straight lines are considered as the coordinate lines, then the set of points $(\mathbf{p}_1, \mathbf{p}_2, ...)$ can be visualized as "bolts of lightning" (see, e.g., [1, 10]). The introduction of bolts of lightning under the name of "permissible lines" was initially attributed to Diliberto and Straus in their work [7] and they have been widely employed in problems related to approximation using sums of univariate functions (see, e.g., [7, 8, 10, 19, 20]). Note that the name “bolt of lightning” is attributed to Arnold in his work [1]. Additionally, Ismailov [14] introduced paths with respect to a finite set of functions. That is, he extended lightning bolts and paths with respect to two directions to paths involving n arbitrarily fixed functions. These last objects have proven to be highly valuable in problems of representation by linear superpositions (see, e.g., [14]).

In the following discussion, we will simplify the terminology by using the term “path” instead of the longer expression “path with respect to the centers $\mathbf{c}_1$ and $\mathbf{c}_2$”. 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.

Let's consider also the following class of functions denoted as $\mathcal{D}$.

$$\mathcal{D} = \{r_1(\|\mathbf{x} - \mathbf{c}_1\|) + r_2(\|\mathbf{x} - \mathbf{c}_2\|) : \ r_i \in C(\mathbb{R}), \ i = 1, 2\}.$$

Note that, in the definition of $\mathcal{G} = \mathcal{G}(g, \mathbf{c_1}, \mathbf{c_2})$ each term $w_i g\left(\frac{\|\mathbf{x}-\mathbf{c_i}\|}{\sigma_i} - \theta_i\right)$ can be interpreted as a function $h(\mathbf{x} - \mathbf{c_i})$ with $\mathbf{c_i} = \mathbf{c_1}$ or $\mathbf{c_i} = \mathbf{c_2}$. The function h depends on the parameters w_i, σ_i and θ_i. It is evident that an element $v \in \mathcal{G} = \mathcal{G}(g, \mathbf{c_1}, \mathbf{c_2})$ also belongs to the class $\mathcal{D}$. In other words, $\mathcal{G} = \mathcal{G}(g, \mathbf{c_1}, \mathbf{c_2})$ is a subset of the class $\mathcal{D}$.

For each closed path $p = (\mathbf{p}_1, \mathbf{p}_2, ..., \mathbf{p}_{2n})$ let us consider the following functional:

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

This functional is associated with the closed path p and exhibits the following obvious properties:

(a) If $r \in \mathcal{D}$, then $G_p(r) = 0$.

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

Let us consider the concept of *extremal paths* (see [13]).

Definition 2.2. *A finite or infinite path* $(\mathbf{p_1}, \mathbf{p_2}, \cdots)$ *is said to be extremal for a function* $u \in C(Q)$ *if* $u(\mathbf{p_i}) = (-1)^i \|u\|, i = 1, 2, \cdots,$ *or* $u(\mathbf{p_i}) = (-1)^{i+1} \|u\|, i = 1, 2, \cdots$.

The following lemma is valid.

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

$$\sup_{p \subset Q} |G_p(f)| \leq \inf_{u \in \mathcal{G}} \|f - u\|, \tag{2.1}$$

where the sup is taken over all closed paths.

Proof. Let p be a closed path of Q and r be any function from $\mathcal{D}$. Based on the linearity of the functional G_p and the properties (a) and (b) mentioned earlier,

$$|G_p(f)| = |G_p(f - r)| \leq \|f - r\|. \tag{2.2}$$

Since the left-hand and the right-hand sides of (2.2) do not depend on the choice of the function r and p respectively, it follows from (2.2) that

$$\sup_{p\subset Q}|G_p(f)| \leq \inf_{r\in\mathcal{D}}\|f-r\|. \tag{2.3}$$

Since $\mathcal{G} = \mathcal{G}(g,\mathbf{c_1},\mathbf{c_2})$ is a subset of $\mathcal{D}$, we have $\inf\limits_{r\in\mathcal{D}}\|f-r\| \leq \inf\limits_{u\in\mathcal{G}}\|f-u\|$. Therefore $\sup\limits_{p\subset Q}|G_p(f)| \leq \inf\limits_{u\in\mathcal{G}}\|f-u\|$. This confirms that the lemma holds.

The images of the distance functions $\|\mathbf{x}-\mathbf{c}_1\|$ and $\|\mathbf{x}-\mathbf{c}_2\|$ on the compact set Q are denoted by X_1 and X_2, respectively. For any function $h\in C(Q)$, consider the real functions

$$\begin{aligned}
s_1(a) &= \max_{\substack{\mathbf{x}\in Q\\ \|\mathbf{x}-\mathbf{c}_1\|=a}} h(x),\ s_2(a) = \min_{\substack{\mathbf{x}\in Q\\ \|\mathbf{x}-\mathbf{c}_1\|=a}} h(x),\ a\in X_1,\\
g_1(b) &= \max_{\substack{\mathbf{x}\in Q\\ \|\mathbf{x}-\mathbf{c}_2\|=b}} h(x),\ g_2(b) = \min_{\substack{\mathbf{x}\in Q\\ \|\mathbf{x}-\mathbf{c}_2\|=b}} h(x),\ b\in X_2.
\end{aligned}$$

When are these functions continuous on the appropriate sets X_1 and X_2? The following lemma, which is essential for proving our main result, Theorem 2.1, answers this question.

Lemma 2.2. (see [2]). *Let $Q\subset\mathbb{R}^d$ be a compact set. Then the functions s_1 and s_2 are continuous on X_1 (g_1 and g_2 are continuous on X_2) for any $h\in C(Q)$ if the following condition, which we call the condition A, holds:*

(A) for any two points $\mathbf{x}$ and $\mathbf{y}$ in Q with $\|\mathbf{x}-\mathbf{c}_1\| = \|\mathbf{y}-\mathbf{c}_1\|$ ($\|\mathbf{x}-\mathbf{c}_2\| = \|\mathbf{y}-\mathbf{c}_2\|$) and any sequence $\{\mathbf{x}_n\}_{n=1}^{\infty}$ tending to $\mathbf{x}$, there exists a sequence $\{\mathbf{y}_n\}_{n=1}^{\infty}$ tending to $\mathbf{y}$ such that $\|\mathbf{x}_n-\mathbf{c}_1\| = \|\mathbf{y}_n-\mathbf{c}_1\|$ ($\|\mathbf{x}_n-\mathbf{c}_2\| = \|\mathbf{y}_n-\mathbf{c}_2\|$) for all $n=1,2,...$

The following theorem is true.

Theorem 2.1. *Let $Q\subset\mathbb{R}^d$ be a compact set and $f\in C(Q)$. Suppose the following conditions hold.*

1) f has a best approximation in $\mathcal{D}$;

2) The above condition (A) holds;

3) there exists a positive integer N_0 such that any path $(\mathbf{p}_1,...,\mathbf{p}_n)\subset Q$, $n>N_0$, or a subpath of it can be made closed by adding not more than N_0 points of Q.

Then for any continuous nonpolynomial activation function $g:\mathbb{R}\to\mathbb{R}$ the approximation error by RBF neural networks with two hidden nodes $\mathcal{G}=\mathcal{G}(g,\mathbf{c_1},\mathbf{c_2})$ can be computed by the formula

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

where the sup is taken over all closed paths.

Proof. By assumption, the function f has a best approximation within the class $\mathcal{D}$, which we denote as $r_0(\mathbf{x}) = r_{10}(\|\mathbf{x}-\mathbf{c}_1\|)+r_{20}(\|\mathbf{x}-\mathbf{c}_2\|)$, where $r_{i0}\in$

$C(\mathbb{R})$ for $i = 1, 2$. Now, let's focus on extremal paths specifically for the function $f_1 = f - r_0$. When considering such paths, we can distinguish between two possible cases.

Case 1. There exists a closed path $p_0 = (\mathbf{p}_1, \ldots, \mathbf{p}_n)$ extremal for the function f_1. In this case, according to Definition 2.2, we can express this as follows:

$$|G_{p_0}(f)| = |G_{p_0}(f - r_0)| = \|f - r_0\|. \tag{2.4}$$

Based on the condition imposed on g, we can conclude that the class of functions $\sum_{i=1}^{m} c_i g(w_i t - \theta_i)$ is dense in $C(\mathbb{R})$ in the topology of uniform convergence on compact sets (as discussed in [18]). It is important to note that the values of w_i can vary on any subset of the real line that contains a sequence converging to a finite limit point (as stated in Proposition 3.11 in [24]). Therefore, for any $\varepsilon > 0$, there exist natural numbers m_1 and m_2, and real numbers c_{ij}, w_{ij}, and θ_{ij}, where $i = 1, 2$ and $j = 1, \ldots, m_i$, satisfying the following inequalities:

$$\left| r_{10}(t) - \sum_{j=1}^{m_1} c_{1j} g(w_{1j} t - \theta_{1j}) \right| < \frac{\varepsilon}{2} \tag{2.5}$$

and

$$\left| r_{20}(t) - \sum_{j=1}^{m_2} c_{2j} g(w_{2j} t - \theta_{2j}) \right| < \frac{\varepsilon}{2} \tag{2.6}$$

for all $t \in [a, b]$. Here $[a, b]$ is a sufficiently large interval which contains both the sets $\{||\mathbf{x} - \mathbf{c_1}|| : \mathbf{x} \in Q\}$ and $\{||\mathbf{x} - \mathbf{c_2}|| : \mathbf{x} \in Q\}$.

By substituting $t = \|\mathbf{x} - \mathbf{c}_1\|$ into (2.5) and $t = \|\mathbf{x} - \mathbf{c}_2\|$ into (2.6), we can deduce that

$$\left| r_{10}(\|\mathbf{x} - \mathbf{c_1}\|) + r_{20}(\|\mathbf{x} - \mathbf{c_2}\|) - \sum_{i=1}^{m} w_i g\left(\frac{\|\mathbf{x} - \mathbf{c_i}\|}{\sigma_i} - \theta_i \right) \right| < \varepsilon \tag{2.7}$$

holds for all $\mathbf{x} \in Q$, and some $w_i, \sigma_i, \theta_i \in \mathbb{R}$ and $\mathbf{c_i} = \mathbf{c_1}$ or $\mathbf{c_i} = \mathbf{c_2}$. Obviously,

$$\left\| f - \sum_{i=1}^{m} w_i g\left(\frac{\|\mathbf{x} - \mathbf{c_i}\|}{\sigma_i} - \theta_i \right) \right\| \leq$$

$$\leq \|f - r_{10} - r_{20}\| + \left\| r_{10} + r_{20} - \sum_{i=1}^{m} w_i g\left(\frac{\|\mathbf{x} - \mathbf{c_i}\|}{\sigma_i} - \theta_i \right) \right\|. \tag{2.8}$$

It follows from (2.8) that

$$E(f, \mathcal{G}(g)) \leq \|f - r_{10} - r_{20}\| + \left\| r_{10} + r_{20} - \sum_{i=1}^{m} w_i g\left(\frac{\|\mathbf{x} - \mathbf{c_i}\|}{\sigma_i} - \theta_i \right) \right\|. \tag{2.9}$$

The last inequality together with (2.4) and (2.7) yield

$$E(f, \mathcal{G}) \leq |G_{p_0}(f)| + \varepsilon.$$

Since ε can be arbitrarily small, we can write that

$$E(f, \mathcal{G}) \leq |G_{p_0}(f)|.$$

By using Lemma 2.1, we can conclude that

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

where the sup is taken over all closed paths.

Case 2. In the second option, we encounter the situation where there is no closed path extremal for the function $f_1=f-r_0$. To address this case, we aim to demonstrate that for any given natural number n, there exists an extremal path for f_1 comprising precisely n points. Assuming the opposite, let's suppose that there exists a positive integer N such that the length of any extremal path for f_1 does not exceed N. Here, the length of a path denotes the count of points it encompasses. We define the following functions:

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

where

$$u_{1,n-1}=u_{1,n-1}\left(\|\mathbf{x}-\mathbf{c}_1\|\right)=\frac{1}{2}\left(\max_{\substack{\mathbf{y}\in Q\\ \|\mathbf{y}-\mathbf{c}_1\|=\|\mathbf{x}-\mathbf{c}_1\|}}f_{n-1}(\mathbf{y})+\min_{\substack{\mathbf{y}\in Q\\ \|\mathbf{y}-\mathbf{c}_1\|=\|\mathbf{x}-\mathbf{c}_1\|}}f_{n-1}(\mathbf{y})\right)$$

$$u_{2,n-1}=u_{2,n-1}(\|\mathbf{x}-\mathbf{c}_2\|)=\frac{1}{2}\left(\max_{\substack{\mathbf{y}\in Q\\ \|\mathbf{y}-\mathbf{c}_2\|=\|\mathbf{x}-\mathbf{c}_2\|}}(f_{n-1}(\mathbf{y})-u_{1,n-1}(\|\mathbf{y}-\mathbf{c}_1\|))+\right.$$

$$\left.+\min_{\substack{\mathbf{y}\in Q\\ \|\mathbf{y}-\mathbf{c}_2\|=\|\mathbf{x}-\mathbf{c}_2\|}}(f_{n-1}(\mathbf{y})-u_{1,n-1}(\|\mathbf{y}-\mathbf{c}_1\|))\right).$$

By Lemma 2.2, all the functions $f_n(\mathbf{x})$, $n=2,3,...$, are continuous on Q. Since r_0 is an extremal element for f, we have $\|f_1\|=E\left(f\right)$. We now aim to show that $\|f_2\|=E\left(f\right)$. Indeed, for any $\mathbf{x}\in Q$ we have

$$f_1(\mathbf{x})-u_{1,1}(\|\mathbf{x}-\mathbf{c}_1\|)\leq$$

$$\leq\frac{1}{2}\left(\max_{\substack{\mathbf{y}\in Q\\ \|\mathbf{y}-\mathbf{c}_1\|=\|\mathbf{x}-\mathbf{c}_1\|}}f_1(\mathbf{y})-\min_{\substack{\mathbf{y}\in Q\\ \|\mathbf{y}-\mathbf{c}_1\|=\|\mathbf{x}-\mathbf{c}_1\|}}f_1(\mathbf{y})\right)\leq E(f) \tag{2.10}$$

and

$$f_1(\mathbf{x})-u_{1,1}(\|\mathbf{x}-\mathbf{c}_1\|)\geq$$

$$\geq\frac{1}{2}\left(\min_{\substack{\mathbf{y}\in Q\\ \|\mathbf{y}-\mathbf{c}_1\|=\|\mathbf{x}-\mathbf{c}_1\|}}f_1(\mathbf{y})-\max_{\substack{\mathbf{y}\in Q\\ \|\mathbf{y}-\mathbf{c}_1\|=\|\mathbf{x}-\mathbf{c}_1\|}}f_1(\mathbf{y})\right)\geq -E(f). \tag{2.11}$$

Considering the definition of $u_{2,1}(\|\mathbf{x}-\mathbf{c}_2\|)$, for any $\mathbf{x}\in Q$ we can write the following inequality:

$$f_1(\mathbf{x})-u_{1,1}(\|\mathbf{x}-\mathbf{c}_1\|)-u_{2,1}(\|\mathbf{x}-\mathbf{c}_2\|)\leq$$

$$\leq\frac{1}{2}\left(\max_{\substack{\mathbf{y}\in Q\\ \|\mathbf{y}-\mathbf{c}_2\|=\|\mathbf{x}-\mathbf{c}_2\|}}(f_1(\mathbf{y})-u_{1,1}(\|\mathbf{y}-\mathbf{c}_1\|))-\min_{\substack{\mathbf{y}\in Q\\ \|\mathbf{y}-\mathbf{c}_2\|=\|\mathbf{x}-\mathbf{c}_2\|}}(f_1(\mathbf{y})-u_{1,1}(\|\mathbf{y}-\mathbf{c}_1\|))\right)$$

and

$$f_1(\mathbf{x})-u_{1,1}(\|\mathbf{x}-\mathbf{c}_1\|)-u_{2,1}(\|\mathbf{x}-\mathbf{c}_2\|)\geq$$

$$\geq \frac{1}{2}\left(\min_{\substack{\mathbf{y}\in Q\\ \|\mathbf{y}-\mathbf{c}_2\|=\|\mathbf{x}-\mathbf{c}_2\|}} (f_1(\mathbf{y}) - u_{1,1}(\|\mathbf{y}-\mathbf{c}_1\|)) - \max_{\substack{\mathbf{y}\in Q\\ \|\mathbf{y}-\mathbf{c}_2\|=\|\mathbf{x}-\mathbf{c}_2\|}} (f_1(\mathbf{y}) - u_{1,1}(\|\mathbf{y}-\mathbf{c}_1\|)) \right).$$

Using (2.10) and (2.11) in the last two inequalities, we obtain that for any $\mathbf{x} \in Q$ the following inequality holds:

$$-E(f) \leq f_2(\mathbf{x}) = f_1(\mathbf{x}) - u_{1,1}(\|\mathbf{x}-\mathbf{c}_1\|) - u_{2,1}(\|\mathbf{x}-\mathbf{c}_2\|) \leq E(f).$$

Thus,

$$\|f_2\| \leq E(f). \tag{2.12}$$

Since $f_2 - f \in \mathcal{D}$, it follows from (2.12) that $\|f_2\| = E(f)$.

Similarly, it can be shown that $\|f_3\| = E(f)$, $\|f_4\| = E(f)$, and so on. Therefore, $\|f_n\| = E(f)$ for all $n = 1, 2, ...$

We will now demonstrate the following implications

$$f_1(\mathbf{p}_0) < E(f) \Rightarrow f_2(\mathbf{p}_0) < E(f) \tag{2.13}$$

and

$$f_1(\mathbf{p}_0) > -E(f) \Rightarrow f_2(\mathbf{p}_0) > -E(f), \tag{2.14}$$

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

$$f_1(\mathbf{p}_0) < E(f) \Rightarrow f_1(\mathbf{p}_0) - u_{1,1}(\|\mathbf{p}_0-\mathbf{c}_1\|) < E(f). \tag{2.15}$$

There are two distinct situations that we need to consider.

1) $\max\limits_{\substack{\mathbf{y}\in Q\\ \|\mathbf{y}-\mathbf{c}_1\|=\|\mathbf{p}_0-\mathbf{c}_1\|}} f_1(\mathbf{y}) = E(f)$ and $\min\limits_{\substack{\mathbf{y}\in Q\\ \|\mathbf{y}-\mathbf{c}_1\|=\|\mathbf{p}_0-\mathbf{c}_1\|}} f_1(\mathbf{y}) = -E(f)$.

In this case, $u_{1,1}(\|\mathbf{p}_0-\mathbf{c}_1\|) = 0$. Thus,

$$f_1(\mathbf{p}_0) - u_{1,1}(\|\mathbf{p}_0-\mathbf{c}_1\|) < E(f).$$

2) $\max\limits_{\substack{\mathbf{y}\in Q\\ \|\mathbf{y}-\mathbf{c}_1\|=\|\mathbf{p}_0-\mathbf{c}_1\|}} f_1(\mathbf{y}) = E(f) - \varepsilon_1$ and $\min\limits_{\substack{\mathbf{y}\in Q\\ \|\mathbf{y}-\mathbf{c}_1\|=\|\mathbf{p}_0-\mathbf{c}_1\|}} f_1(\mathbf{y}) = -E(f) + \varepsilon_2$,

where ε_1, $\varepsilon_2 \geq 0$ and $\varepsilon_1 + \varepsilon_2 \neq 0$.

In this case,

$$f_1(\mathbf{p}_0) - u_{1,1}(\|\mathbf{p}_0-\mathbf{c}_1\|) \leq \max_{\substack{\mathbf{y}\in Q\\ \|\mathbf{y}-\mathbf{c}_1\|=\|\mathbf{p}_0-\mathbf{c}_1\|}} f_1(\mathbf{y}) - u_{1,1}(\|\mathbf{p}_0-\mathbf{c}_1\|) =$$

$$= \frac{1}{2}\left(\max_{\substack{\mathbf{y}\in Q\\ \|\mathbf{y}-\mathbf{c}_1\|=\|\mathbf{p}_0-\mathbf{c}_1\|}} f_1(\mathbf{y}) - \min_{\substack{\mathbf{y}\in Q\\ \|\mathbf{y}-\mathbf{c}_1\|=\|\mathbf{p}_0-\mathbf{c}_1\|}} f_1(\mathbf{y}) \right)$$

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

Therefore, we have successfully demonstrated the validity of equation (2.15). By employing a similar approach, we can also establish the proof for that

$$f_1(\mathbf{p}_0) - r_{1,1}(\|\mathbf{p}_0-\mathbf{c}_1\|) < E(f) \Rightarrow f_1(\mathbf{p}_0) -$$
$$-r_{1,1}(\|\mathbf{p}_0-\mathbf{c}_1\|) - r_{2,1}(\|\mathbf{p}_0-\mathbf{c}_2\|) < E(f). \tag{2.16}$$

Implications (2.15) and (2.16) lead to the conclusion stated in equation (2.13). Likewise, using similar reasoning, we can also establish the validity of the equation (2.14). Consequently, based on the implications (2.13) and (2.14) we can deduce 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 implies that any path extremal for f_2 is also an extremal for f_1.

We assumed earlier that any path extremal for f_1 has a length of no more than N. Now we will demonstrate that, in this case, any path extremal for f_2 must have a length of no more than $N-1$. Let's suppose the opposite, assuming the existence of a path extremal for f_2 with a length of N. We denote this path as $q = (\mathbf{q}_1, \mathbf{q}_2, ..., \mathbf{q}_N)$. Without loss of generality, we can assume that $\|q_{N-1} - \mathbf{c}_2\| = \|q_N - \mathbf{c}_2\|$. As we have previously shown, path q is an extremal for f_1. Let's assume $f_1(\mathbf{q}_N) = E(f)$. In that case, there does not exist a point $\mathbf{q}_0 \in Q$ such that $\mathbf{q}_0 \neq \mathbf{q}_N$, $\|q_0 - \mathbf{c}_1\| = \|q_N - \mathbf{c}_1\|$ and $f_1(\mathbf{q}_0) = -E(f)$. To clarify, if such a point $\mathbf{q}_0$ 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. However, this contradicts our assumption that any path extremal for f_1 has a length of no more than N. On the other hand, if there were such a point $\mathbf{q}_0$ and $\mathbf{q}_0 \in q$, then we could form a closed extremal path for f_1 using points from q. This would also contradict our assumption that there is no closed extremal path for f_1. Hence, we can conclude that

$$\max_{\substack{\mathbf{y}\in Q \\ \|\mathbf{y}-\mathbf{c}_1\|=\|\mathbf{q}_N-\mathbf{c}_1\|}} f_1(\mathbf{y}) = E(f), \qquad \min_{\substack{\mathbf{y}\in Q \\ \|\mathbf{y}-\mathbf{c}_1\|=\|\mathbf{q}_N-\mathbf{c}_1\|}} f_1(\mathbf{y}) > -E(f).$$

Therefore,

$$|f_1(\mathbf{q}_N) - u_{1,1}(\|\mathbf{q}_N - \mathbf{c}_1\|)| < E(f).$$

Using a similar approach as described earlier, we can obtain from the last inequality 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. Therefore, we can conclude that any path extremal for f_2 has a length that is not more than $N-1$.

Using the same reasoning, we can extend this result to show that any path extremal for f_3 has a length not more than $N-2$, any path extremal for f_4 has a length not more than $N-3$ and so on. Ultimately, we reach the conclusion that there is no path extremal for f_{N+1}. Consequently, there is no point $\mathbf{p}_0 \in Q$ such that $|f_{N+1}(\mathbf{p}_0)| = \|f_{N+1}\|$. However, since all the functions $f_2, f_3, ..., f_{N+1}$ are continuous on the compact set Q (as per Lemma 2.2), the norm $\|f_{N+1}\|$ must be attained. This contradiction demonstrates that for any n there exists a path $(\mathbf{p}_1, \mathbf{p}_2, ..., \mathbf{p}_n)$ extremal for f_1.

Consider the sequence of extremal paths $p_n = (\mathbf{p}_1^n, \mathbf{p}_2^n, ..., \mathbf{p}_n^n)$, $n = 1, 2, ...$ According to condition (3) of the theorem, for each path p_n there exists a closed path $p_n^{m_n} = (\mathbf{p}_1^n, \mathbf{p}_2^n, ..., \mathbf{p}_n^n, \mathbf{q}_{n+1}^n, ..., \mathbf{q}_{n+m_n}^n)$, with $m_n \leq N_0$. We observe that the functional $G_{p_n^{m_n}}$ satisfies the following inequalities:

$$\left|G_{p_n^{m_n}}(f)\right| = \left|G_{p_n^{m_n}}(f - r_0)\right| \leq \frac{n\,\|f - r_0\| + m_n\,\|f - r_0\|}{n + m_n} = \|f - r_0\| \tag{2.17}$$

and

$$\left|G_{p_n^{m_n}}(f)\right| \geq \frac{n\,\|f - r_0\| - m_n\,\|f - r_0\|}{n + m_n} = \frac{n - m_n}{n + m_n}\,\|f - r_0\|. \tag{2.18}$$

From equations (2.17) and (2.18), we can deduce the following:

$$\sup_{p_n^{m_n}} \left|G_{p_n^{m_n}}(f)\right| = \|f - r_0\|.$$

By employing the approach involving the function $f - r_0 = f - r_{10} - r_{20}$ in Case 1, we can derive the following inequality:

$$E\left(f,\mathcal{G}\right) \leq \sup_{p_n^{mn}} \left|G_{p_n^{mn}}(f)\right| \leq \sup_{p\subset Q} \left|G_p(f)\right|. \tag{2.19}$$

From equation (2.19) and Lemma 2.1, we conclude that:

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

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

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] A.Kh. Asgarova, V.E. Ismailov, Diliberto-Straus algorithm for the uniform approximation by a sum of two algebras, *Proc. Indian Acad. Sci. Math. Sci.* **127** (2017), no. 2, 361–374.

[3] A.Kh. Asgarova, A.M-B. Babayev, I.K. Maharov, On the error of approximation by ridge functions with two fixed directions, *Tbilisi Math. J.* **10** (2017), no. 2, 111-120.

[4] A.Kh. Asgarova, A.M-B. Babayev, I.K. Maharov, On the error of approximation by radial basis functions with fixed centers, *Trans. Natl. Acad. Sci. Azerb. Ser. Phys.-Tech. Math. Sci.* **38** (2018), no. 1, 22-29.

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

[6] D.S. Broomhead, D. Lowe, Multivariable function interpolation and adaptive networks, *Complex Systems* **2** (1988), 321-355.

[7] S.P. Diliberto, 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.

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

[9] Q.P. Ha, H.Wahid, H. Duc, M. Azzi, Enhanced radial basis function neural networks for ozone level estimation, *Neurocomputing* **155** (2015), 62-70.

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

[11] V.E. Ismailov, A note on the best L_2 approximation by ridge functions, *Appl. Math. E-Notes* **7** (2007), 71–76.

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

[13] V.E. Ismailov, A note on the equioscillation theorem for best ridge function approximation, *Expo. Math.* **35** (2017), no. 3, 343-349.

[14] V.E. Ismailov, *Ridge functions and applications in neural networks.* Mathematical Surveys and Monographs, **263**. American Mathematical Society, 2021, 186 pp.

[15] V.E. Ismailov, A formula for the approximation of functions by single hidden layer neural networks with weights from two straight lines, *Proc. Inst. Math. Mech. Natl. Acad. Sci. Azerb.* **48** (2022), no. 1, 140–149.

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

[17] C.-N. Ko, Reinforcement radial basis function neural networks with an adaptive annealing learning algorithm, *Appl. Math. Comput.* **221** (2013), 503-513.
[18] M. Leshno, V. Ya. Lin, A. Pinkus and S. Schocken, Multilayer feedforward networks with a non-polynomial activation function can approximate any function, *Neural Networks* **6** (1993), 861-867.
[19] D.E. Marshall, A.G. O'Farrell, Uniform approximation by real functions, *Fund. Math.* **104** (1979), 203-211.
[20] 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.
[21] G.A. Montazer, D. Giveki, An improved radial basis function neural network for object image retrieval, *Neurocomputing* **168** (2015), 221-233.
[22] H. Muzhou, H. Xuli, The multidimensional function approximation based on constructive wavelet RBF neural network, *Appl. Soft Comput.* **11** (2011), 2173-2177.
[23] J. Park, I.W. Sandberg, Approximation and Radial-Basis-Function Networks, *Neural Computation* **5** (1993), no.2, 305-316.
[24] A. Pinkus, Approximation theory of the MLP model in neural networks, *Acta Numerica* **8** (1999), 143-195.
[25] A. Pinkus, *Ridge functions.* Cambridge Tracts in Mathematics, 205. Cambridge University Press, 2015, 207 pp.
[26] Y. Xie, Y. Xing, A prediction method for the wax deposition rate based on a radial basis function neural network, *Petroleum* **3** (2017), 237-241.
[27] S. Yang, T.O. Ting, K.L. Man, S.-U. Guan, Investigation of neural networks for function approximation, *Procedia Computer Science* **17** (2013), 586-594.
[28] S.-H. Yoo, S.-K. Oh, W. Pedrycz, Optimized face recognition algorithm using radial basis function neural networks and its practical applications, *Neural Networks* **69** (2015), 111-125.
[29] H. Yuan, X. Shen, X. Chen, Prediction of fishing ground based on RBF neural network, *Procedia Engineering* **15** (2011), 3240-3244.

Aida Kh. Asgarova
Institute of Mathematics and Mechanics, Ministry of Science and Education of Azerbaijan, B. Vahabzade str. 9, AZ1148, Baku, Azerbaijan
Khazar University, Baku, Azerbaijan
E-mail address: aidaasgarova@gmail.com

Ibrahim K. Maharov
Institute of Mathematics and Mechanics, Ministry of Science and Education of Azerbaijan, B. Vahabzade str. 9, AZ1148, Baku, Azerbaijan
E-mail address: ibrahimmaharov@gmail.com

Received: December 25, 2023; Accepted: May 10, 2024