Proceedings of the Institute of Mathematics and Mechanics,
National Academy of Sciences of Azerbaijan
Volume 00, Number 0, XXXX, Pages 000–000

ON THE APPROXIMATION BY RBF NEURAL NETWORKS WITH TRANSLATIONS

AIDA KH. ASGAROVA

Abstract. This paper studies the approximation of continuous multivariate functions by radial basis function neural networks with two fixed centers, in which all units share a common smoothing factor and incorporate additional translations. Under the assumption that the max and min functions defined via the distance mappings from these centers are continuous, we derive a formula for the exact computation of the approximation error in the uniform norm. Under this hypothesis, we also obtain a characterization of best approximations from the considered class in terms of extremal paths.

1. Introduction

Radial basis functions (RBFs) form a family of multivariate functions whose values depend only on the distance from a prescribed center. In other words, for a center $\mathbf{c}$ and a radius ρ, an RBF takes the same value at all points $\mathbf{x}$ satisfying $\|\mathbf{x}-\mathbf{c}\| = \rho$. Based on RBFs, Broomhead and Lowe [9] introduced *radial basis function neural networks* (RBFNNs), which have proved to be effective tools for approximation problems. Initially developed for interpolation in high-dimensional spaces, these networks are now widely used in various applications. Typical areas include function approximation (see, e.g., [2, 19, 24, 15, 23, 30]), prediction (see, e.g., [29, 32]), parameter estimation (see, e.g., [27, 28]), pattern recognition (see, e.g., [22, 31]), and control problems (see, e.g., [18]).

A standard RBFNN consists of an input layer, a hidden layer, and an output layer. Each hidden unit is associated with a center, and for an input vector $\mathbf{x} = (x_1, \dots, x_d)$ it computes the distance to this center $\mathbf{c} \in \mathbb{R}^d$. The output of a hidden unit is obtained by applying a nonlinear function to this distance, thus producing a radial signal. The output layer forms a linear combination of all hidden-unit outputs. For simplicity, we restrict attention to the case of a scalar output. The extension to vector-valued outputs is immediate.

With d inputs and a single output, the network function is given by

$$\sum_{i=1}^{m} w_i \, g\left(\frac{\|\mathbf{x}-\mathbf{c}_i\|}{\sigma_i}\right).$$

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

Key words and phrases. RBFNN, path, extremal path, best approximation, approximation error.

Here $m \in \mathbb{N}$ denotes the number of hidden units, $(w_1, \dots, w_m) \in \mathbb{R}^m$ are the output weights, $\mathbf{x} \in \mathbb{R}^d$ is the input vector, $\mathbf{c}_i \in \mathbb{R}^d$ are the centers, and $\sigma_i > 0$ are the smoothing factors. The function $g : \mathbb{R} \to \mathbb{R}$ is the activation function.

In practice, different activations may be used and the smoothing factors may either vary across units or be kept fixed. We consider networks of the form

$$\sum_{i=1}^{m} w_i \, g\left(\frac{\|\mathbf{x} - \mathbf{c}_i\|}{\sigma} - \theta_i \right),$$

where a common smoothing factor σ is used for all hidden units and each term is allowed an additional translation θ_i. For the universal approximation property of such RBFNNs with translations, see [15].

It is worth noting that allowing translations increases the flexibility of the network and improves its approximation capabilities. While a fixed smoothing factor controls the scale of the radial responses, the shifts θ_i make it possible to adjust their location and better capture local features of the target function. Consequently, the class of realizable functions is enlarged and the approximation error may be reduced without increasing the architectural complexity.

Let $f(\mathbf{x}) = f(x_1, \dots, x_d)$ be a continuous function defined on a compact set $Q \subset \mathbb{R}^d$. We approximate f by elements of the class

$$\mathcal{S} = \mathcal{S}(g, \mathbf{c}_1, \mathbf{c}_2, \sigma) = \left\{ \sum_{i=1}^{m} w_i \, g\left(\frac{\|\mathbf{x} - \mathbf{c}_i\|}{\sigma} - \theta_i \right) : \ w_i, \theta_i \in \mathbb{R}, \ \mathbf{c}_i \in \{\mathbf{c}_1, \mathbf{c}_2\} \right\}.$$

Thus, the centers are restricted to two prescribed points $\mathbf{c}_1$ and $\mathbf{c}_2$, while the weights and translations remain free parameters.

For example, when $\sigma = 1$, the set $\mathcal{S}(g, \mathbf{c}_1, \mathbf{c}_2, \sigma)$ coincides with the following class of RBFNNs:

$$\mathcal{A} = \mathcal{A}(g, \mathbf{c}_1, \mathbf{c}_2) = \left\{ \sum_{i=1}^{m} w_i \, g(\|\mathbf{x} - \mathbf{c}_i\| - \theta_i) : \ w_i, \theta_i \in \mathbb{R}, \ \mathbf{c}_i = \mathbf{c}_1 \text{ or } \mathbf{c}_i = \mathbf{c}_2 \right\}.$$

Certain approximation properties of this class were studied in [7, 15].

The error of approximation is defined by

$$E(f) = E(f, \mathcal{S}) = \inf_{h \in \mathcal{S}} \|f - h\|, \qquad \|f - h\| = \max_{\mathbf{x} \in Q} |f(\mathbf{x}) - h(\mathbf{x})|.$$

An element $u \in \mathcal{S}$ is called a best approximation to f from $\mathcal{S}$ if

$$\|f - u\| = E(f, \mathcal{S}).$$

The main goal of this paper is to derive a formula for the approximation error for the class $\mathcal{S}(g, \mathbf{c}_1, \mathbf{c}_2, \sigma)$ and to obtain a characterization of best approximations from this class. We show that the quantity $E(f)$ can be expressed in terms of suitably defined functionals evaluated at the function f, and that best approximants in $\mathcal{S}$ are characterized in terms of extremal paths.

2. Main results

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

Definition 2.1. *A finite or infinite ordered set*

$$p = (\mathbf{p}_1, \mathbf{p}_2, \dots) \subset \mathbb{R}^d, \qquad \mathbf{p}_i \neq \mathbf{p}_{i+1},$$

is called a path *with respect to the centers* $\mathbf{c}_1$ *and* $\mathbf{c}_2$ *if it satisfies one of the following conditions:*

(1) $\|\mathbf{p}_1 - \mathbf{c}_1\| = \|\mathbf{p}_2 - \mathbf{c}_1\|$, $\|\mathbf{p}_2 - \mathbf{c}_2\| = \|\mathbf{p}_3 - \mathbf{c}_2\|$, $\|\mathbf{p}_3 - \mathbf{c}_1\| = \|\mathbf{p}_4 - \mathbf{c}_1\|$, *and so on, alternating between the centers.*

(2) $\|\mathbf{p}_1 - \mathbf{c}_2\| = \|\mathbf{p}_2 - \mathbf{c}_2\|$, $\|\mathbf{p}_2 - \mathbf{c}_1\| = \|\mathbf{p}_3 - \mathbf{c}_1\|$, $\|\mathbf{p}_3 - \mathbf{c}_2\| = \|\mathbf{p}_4 - \mathbf{c}_2\|$, *and so on, alternating in the opposite order.*

In Definition 2.1, the distances are measured alternately from two fixed points. There are, however, various generalizations of this concept. For example, instead of alternating distances from two fixed points, one may consider two fixed vectors $\mathbf{a}^1, \mathbf{a}^2 \in \mathbb{R}^d$ and alternate the scalar products $\mathbf{a}^i \cdot \mathbf{x}$ with these vectors. Paths defined in terms of two fixed vectors in $\mathbb{R}^2$ were first studied by Braess and Pinkus [8], who used them to analyze whether a set of points $\{\mathbf{x}^i\}_{i=1}^m \subset \mathbb{R}^2$ possesses the non-interpolation property for so-called *ridge functions*. Detailed discussions on ridge functions and their properties can be found in [13, 26]. Paths involving two fixed vectors in $\mathbb{R}^d$ have been further studied in several other works (see, e.g., [12, 14]).

In $\mathbb{R}^2$, when the two vectors are taken as the coordinate directions, the sequence of points $(\mathbf{p}_1, \mathbf{p}_2, \ldots)$ is called a "bolt of lightning" (see [1]). This concept, originally referred to as "permissible lines," was introduced by Diliberto and Straus [10] and has since played a central role in the approximation of multivariate functions by sums of univariate functions and sums of two algebras (see, for instance, [4, 5, 11, 20, 21]). The term "bolt of lightning" is attributed to Arnold [1]. Ismailov [13] later generalized this concept by introducing paths defined with respect to a finite set of functions, extending both the idea of bolts of lightning and the notion of paths associated with two fixed vectors. These generalized paths have proved to be highly effective in problems concerning representation by linear superpositions (see, e.g., [13]).

In the following discussion, for simplicity, we will use the term "path" instead of the longer phrase "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 considered closed if $(\mathbf{p}_1, \mathbf{p}_2, ..., \mathbf{p}_{2n}, \mathbf{p}_1)$ also forms a path.

Let us also consider the following class of radial functions, denoted by $\mathcal{D}$.

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

In the following, we use the proximinality of $\mathcal{D}$. Recall that $\mathcal{D}$ is said to be proximinal in $C(Q)$ if for every $f \in C(Q)$ there exists $v \in \mathcal{D}$ such that

$$\|f - v\| = \inf_{h \in \mathcal{D}} \|f - h\|.$$

It is important to note that in the definition of our class $\mathcal{S}(g, \mathbf{c}_1, \mathbf{c}_2, \sigma)$, each term $w_i g\left(\frac{\|\mathbf{x}-\mathbf{c}_i\|}{\sigma} - \theta_i\right)$ can be interpreted as a function $h_\sigma(\|\mathbf{x} - \mathbf{c}_i\|)$, where $\mathbf{c}_i$ is either $\mathbf{c}_1$ or $\mathbf{c}_2$. The function h_σ is dependent on the parameters w_i and θ_i. It is clear that any element $u \in \mathcal{S}(g, \mathbf{c}_1, \mathbf{c}_2, \sigma)$ belongs to the class of radial functions $\mathcal{D}$. In other words, we have $\mathcal{S}(g, \mathbf{c}_1, \mathbf{c}_2, \sigma) \subset \mathcal{D}$.

For every closed path $p = (\mathbf{p}_1, \mathbf{p}_2, ..., \mathbf{p}_{2n})$ we define the following functional:

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

This functional, corresponding to the closed path p, possesses 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 $\|G_p\| = 1$.

Now consider the concept of *extremal paths.*

Definition 2.2. *A finite or infinite path* $(\mathbf{p}_1, \mathbf{p}_2, \cdots)$ *is said extremal for a function* $f \in C(Q)$ *if it satisfies one of the following conditions:*

(1) $f(\mathbf{p}_i) = (-1)^i \|f\|, i = 1, 2, \cdots$, or

(2) $f(\mathbf{p}_i) = (-1)^{i+1} \|f\|, i = 1, 2, \cdots$.

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 $h \in C(Q)$, let us define the following real-valued functions:

$$s_1(a) = \max_{\substack{\mathbf{x} \in Q \\ \|\mathbf{x} - \mathbf{c}_1\| = a}} h(x), \ s_2(a) = \min_{\substack{\mathbf{x} \in Q \\ \|\mathbf{x} - \mathbf{c}_1\| = a}} h(x), \ a \in X_1, \tag{2.1}$$

$$w_1(b) = \max_{\substack{\mathbf{x} \in Q \\ \|\mathbf{x} - \mathbf{c}_2\| = b}} h(x), \ w_2(b) = \min_{\substack{\mathbf{x} \in Q \\ \|\mathbf{x} - \mathbf{c}_2\| = b}} h(x), \ b \in X_2. \tag{2.2}$$

The following theorem, which implies the continuity of the above functions, is taken from [3]:

Theorem 2.1 (see [3]). *Let* $Q \subset \mathbb{R}^d$ *be a compact set. The functions* s_1 *and* s_2 *are continuous on* X_1 *(and* w_1 *and* w_2 *are continuous on* X_2*) for any* $h \in C(Q)$*, provided that the following condition holds:*

for any two points $\mathbf{x}, \mathbf{y} \in Q$ *satisfying*

$$\|\mathbf{x} - \mathbf{c}_1\| = \|\mathbf{y} - \mathbf{c}_1\| \quad (\text{respectively } \|\mathbf{x} - \mathbf{c}_2\| = \|\mathbf{y} - \mathbf{c}_2\|),$$

and for any sequence $\{\mathbf{x}_n\} \subset Q$ *converging to* $\mathbf{x}$*, there exists a sequence* $\{\mathbf{y}_n\} \subset Q$ *converging to* $\mathbf{y}$ *such that*

$$\|\mathbf{x}_n - \mathbf{c}_1\| = \|\mathbf{y}_n - \mathbf{c}_1\| \quad (\text{respectively } \|\mathbf{x}_n - \mathbf{c}_2\| = \|\mathbf{y}_n - \mathbf{c}_2\|)$$

for all $n = 1, 2, \ldots$.

The following theorem establishes a formula for the approximation error for the class $\mathcal{S}$ in terms of closed-path functionals.

Theorem 2.2. *Let* $Q \subset \mathbb{R}^d$ *be a compact set such that* $\mathcal{D}$ *is proximinal in* $C(Q)$*. Let the activation function* $g \in C(\mathbb{R})$ *be bounded, nonconstant, and have a finite limit at* $+\infty$ *or* $-\infty$*. Assume further that the functions* (2.1) *and* (2.2) *are continuous. Then, for any* $f \in C(Q)$*, the approximation error with respect to the RBFNN class* $\mathcal{S} = \mathcal{S}(g, \mathbf{c}_1, \mathbf{c}_2, \sigma)$ *is given by the formula*

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

where the supremum is taken over all closed paths in Q*.*

Proof. We begin by establishing the inequality

$$\sup_{p \subset Q} |G_p(f)| \leq \inf_{h \in \mathcal{S}} \|f - h\|, \tag{2.3}$$

By the linearity of the functional G_p and its properties (a) and (b), for any closed path $p \subset Q$ and any function $r \in \mathcal{D}$, we have

$$|G_p(f)| = |G_p(f-r)| \leq \|f-r\|.$$

Since p and r are arbitrary, it follows that

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

Because $\mathcal{S} \subset \mathcal{D}$, we obtain

$$\inf_{r \in \mathcal{D}} \|f-r\| \leq \inf_{h \in \mathcal{S}} \|f-h\|,$$

and thus (2.3) holds.

Next, let $v \in \mathcal{D}$ be a best approximation to f in $\mathcal{D}$. Put

$$v(\mathbf{x}) = r_1\left(\|\mathbf{x}-\mathbf{c}_1\|\right) + r_2\left(\|\mathbf{x}-\mathbf{c}_2\|\right) = r_1\left(\sigma\frac{\|\mathbf{x}-\mathbf{c}_1\|}{\sigma}\right) + r_2\left(\sigma\frac{\|\mathbf{x}-\mathbf{c}_2\|}{\sigma}\right)$$
$$= v_1\left(\frac{\|\mathbf{x}-\mathbf{c}_1\|}{\sigma}\right) + v_2\left(\frac{\|\mathbf{x}-\mathbf{c}_2\|}{\sigma}\right),$$

where $r_i, v_i \in C(\mathbb{R})$, $i = 1, 2$ and each v_i depends on the corresponding r_i.

Set $f_1 := f - v$. Suppose first that there exists a closed path $p_0 = (\mathbf{p}_1, \ldots, \mathbf{p}_n)$ in Q that is extremal for f_1. Then, for the functional G_{p_0}, we can write that:

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

Since $g \in C(\mathbb{R})$ is bounded, nonconstant, and has a finite limit at $+\infty$ or $-\infty$, it follows from the classical results on the density of translates (see [25]) that the set of finite linear combinations

$$\left\{\sum_{i=1}^{m} c_i g(t-\theta_i) : m \in \mathbb{N},\ c_i, \theta_i \in \mathbb{R}\right\}$$

is dense in $C(\mathbb{R})$ with respect to uniform convergence on compact sets.

Therefore, for any $\varepsilon > 0$, there exist $m_1, m_2 \in \mathbb{N}$ and real numbers c_{ij}, θ_{ij} such that for all $t \in [a, b]$, where $[a, b]$ contains the sets

$$\left\{\frac{\|\mathbf{x}-\mathbf{c}_1\|}{\sigma} : \mathbf{x} \in Q\right\}, \quad \left\{\frac{\|\mathbf{x}-\mathbf{c}_2\|}{\sigma} : \mathbf{x} \in Q\right\},$$

we have

$$\left|v_1(t) - \sum_{j=1}^{m_1} c_{1j} g(t-\theta_{1j})\right| < \frac{\varepsilon}{2} \tag{2.5}$$

and

$$\left|v_2(t) - \sum_{j=1}^{m_2} c_{2j} g(t-\theta_{2j})\right| < \frac{\varepsilon}{2}. \tag{2.6}$$

Substituting $t = \frac{\|\mathbf{x}-\mathbf{c}_1\|}{\sigma}$ in (2.5) and $t = \frac{\|\mathbf{x}-\mathbf{c}_2\|}{\sigma}$ in (2.6) gives

$$\left|v_1\left(\frac{\|\mathbf{x}-\mathbf{c}_1\|}{\sigma}\right) + v_2\left(\frac{\|\mathbf{x}-\mathbf{c}_2\|}{\sigma}\right) - \sum_{i=1}^{m} w_i g\left(\frac{\|\mathbf{x}-\mathbf{c}_i\|}{\sigma} - \theta_i\right)\right| < \varepsilon,$$

where $m = m_1 + m_2$ and each $\mathbf{c}_i$ equals either $\mathbf{c}_1$ or $\mathbf{c}_2$.

Hence,

$$\left\| f - \sum_{i=1}^{m} w_i g\left(\frac{\|\mathbf{x}-\mathbf{c}_i\|}{\sigma} - \theta_i\right)\right\| \le \|f-v\| + \left\| v - \sum_{i=1}^{m} w_i g\left(\frac{\|\mathbf{x}-\mathbf{c}_i\|}{\sigma} - \theta_i\right)\right\| < \|f-v\| + \varepsilon.$$

Taking the infimum on the left-hand side and using (2.4), we obtain

$$E(f,\mathcal{S}) < |G_{p_0}(f)| + \varepsilon,$$

and by letting $\varepsilon \to 0$,

$$E(f,\mathcal{S}) \le |G_{p_0}(f)|.$$

Together with (2.3), this implies

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

proving the theorem in the case where an extremal closed path exists.

Now consider the case where no closed extremal path for f_1 exists. We will show that for any $n \in \mathbb{N}$, there exists an extremal path of length n for f_1.

Suppose, for contradiction, that there is a positive integer N such that every extremal path for f_1 has length at most N.

Define sequences of functions f_n for $n \ge 2$ by

$$f_n = f_{n-1} - u_{1,n-1} - u_{2,n-1},$$

where

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

and

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

By the assumption of the theorem each f_n is continuous, since the defining formulas involve only the max and min functions. By continuity of f_n and properties of $u_{1,n-1}, u_{2,n-1}$, it can be shown that

$$\|f_n\| = E(f,\mathcal{S}) \quad \text{for all } n.$$

Moreover, one can show that if every extremal path for f_1 has length at most N, then every extremal path for f_2 has length at most $N-1$, and similarly for $f_3, f_4, \ldots$. After $N+1$ steps we reach a contradiction, since f_{N+1} would have no extremal path, while $\|f_{N+1}\| = E(f)$ and the norm of a continuous function on a compact set must be attained. Therefore, for every n, there exists an extremal path of length n for f_1.

Consider now the sequence of extremal paths $p_n = (\mathbf{p}_1^n, ..., \mathbf{p}_n^n)$, $n = 1, 2, ...$ By assumption, every function in $C(Q)$ has a best approximation in $\mathcal{D}$. Consequently, the lengths of irreducible paths in Q (i.e., paths of minimal length joining their first and end points) are bounded by some positive integer N_0 (see [17, p.58]). Hence, any path in Q whose length exceeds N_0 can be made closed by adding at most N_0 points. We say that a path $(\mathbf{p}_1, ..., \mathbf{p}_n) \subset Q$ can be made closed if there exist points $\mathbf{q}_i \in Q$, $i = 1, ..., m$, such that $(\mathbf{p}_1, ..., \mathbf{p}_n, \mathbf{q}_1, ..., \mathbf{q}_m)$ is a closed path.

Therefore, any extremal path $p_n = (\mathbf{p}_1^n, \ldots, \mathbf{p}_n^n)$ with $n > N_0$ can be extended to a closed path

$$p_n^{m_n} = (\mathbf{p}_1^n, \ldots, \mathbf{p}_n^n, \mathbf{q}_{n+1}^n, \ldots, \mathbf{q}_{n+m_n}^n),$$

where $m_n \leq N_0$.

For the functional $G_{p_n^{m_n}}$, the following bounds hold:

$$\left|G_{p_n^{m_n}}(f)\right| = \left|G_{p_n^{m_n}}(f - v)\right| \leq \frac{n \left\|f - v\right\| + m_n \left\|f - v\right\|}{n + m_n} = \left\|f - v\right\|, \tag{2.7}$$

and

$$\left|G_{p_n^{m_n}}(f)\right| \geq \frac{n \left\|f - v\right\| - m_n \left\|f - v\right\|}{n + m_n} = \frac{n - m_n}{n + m_n} \left\|f - v\right\|. \tag{2.8}$$

From (2.7) and (2.8), we have

$$\sup_{p_n^{mn}} |G_{p_n^{m_n}}(f)| = \|f - v\|,$$

and consequently, by applying the method involving the function $f - v = f - v_1 - v_2$ in the case where a closed extremal path exists, we obtain the inequality

$$E(f, \mathcal{S}) \leq \sup_{p_n^{mn}} |G_{p_n^{m_n}}(f)| \leq \sup_{p \subset Q} |G_p(f)|. \tag{2.9}$$

Combining (2.3) and (2.9) yields

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

where the supremum is taken over all closed paths in Q.

This completes the proof. □

In approximation theory, best approximations u from some class of approximants are often characterized by an alternation principle describing the structure of the difference $f - u$. In the present setting, this role is played by extremal paths. The closed path functionals G_p annihilate the class $\mathcal{D}$ and therefore depend only on the difference $f - u$. This makes it possible to formulate a criterion for best approximations in the class $\mathcal{S}$ in terms of extremal paths.

Theorem 2.3. *Assume the hypotheses of Theorem 2.2 hold. Let $f \in C(Q)$ and let $u \in \mathcal{S} = \mathcal{S}(g, \mathbf{c}_1, \mathbf{c}_2, \sigma)$. Then u is a best approximation to f from $\mathcal{S}$ (that is, $\|f - u\| = E(f, \mathcal{S})$) if and only if one of the following two conditions holds:*

1. *There exists a closed path $p \subset Q$ that is extremal for $f - u$.*
2. *For every $N \in \mathbb{N}$ there exists a (not necessarily closed) path $p \subset Q$ consisting of N points that is extremal for $f - u$.*

Proof. Let $u \in \mathcal{S}$ and set $r := f - u$.

"If" part. Assume that one of the conditions (1) or (2) holds.

Case (1): existence of a closed extremal path for r. Let $p_0 = (\mathbf{p}_1, \ldots, \mathbf{p}_{2n}) \subset Q$ be a closed path that is extremal for $r = f - u$. Then, by Definition 2.2, either

$r(\mathbf{p}_k) = (-1)^k\|r\|$ for all k, or $r(\mathbf{p}_k) = (-1)^{k+1}\|r\|$ for all k. In either case, substituting into the definition of G_{p_0} gives

$$G_{p_0}(r) = \frac{1}{2n}\sum_{k=1}^{2n}(-1)^{k+1}r(\mathbf{p}_k) = \pm\|r\|,$$

and therefore

$$|G_{p_0}(r)| = \|r\| = \|f-u\|. \tag{2.10}$$

Since $u \in \mathcal{S} \subset \mathcal{D}$ and G_{p_0} annihilates $\mathcal{D}$ (property (a)), we have $G_{p_0}(u) = 0$, and thus

$$G_{p_0}(f) = G_{p_0}(f-u) = G_{p_0}(r).$$

Combining this with (2.10) yields

$$|G_{p_0}(f)| = \|f-u\|. \tag{2.11}$$

For any $h \in \mathcal{S}$, Theorem 2.2 implies

$$\sup_{p\subset Q}|G_p(f)| = E(f,\mathcal{S}) \le \|f-h\|. \tag{2.12}$$

Applying (2.12) with $h = u$ and using (2.11), we obtain

$$E(f,\mathcal{S}) \le \|f-u\| = |G_{p_0}(f)| \le \sup_{p\subset Q}|G_p(f)| = E(f,\mathcal{S}).$$

Hence $\|f-u\| = E(f,\mathcal{S})$, i.e. u is a best approximation to f from $\mathcal{S}$.

Case (2): arbitrarily long extremal paths for r. Assume that for every $N \in \mathbb{N}$ there exists a (not necessarily closed) path $p_N = (\mathbf{p}_1^N, \dots, \mathbf{p}_N^N) \subset Q$ that is extremal for $r = f - u$.

As in the proof of Theorem 2.2, the assumption that every function in $C(Q)$ admits a best approximation in $\mathcal{D}$ implies that the lengths of irreducible paths in Q are bounded by some integer N_0 (see [17, p. 58]). Consequently, any path in Q of length exceeding N_0 can be extended to a closed path by adding at most N_0 points.

Fix $N > N_0$ and extend the extremal path p_N to a closed path

$$\widetilde{p}_N = (\mathbf{p}_1^N, \dots, \mathbf{p}_N^N, \mathbf{q}_{N+1}^N, \dots, \mathbf{q}_{N+m_N}^N), \qquad m_N \le N_0,$$

with $\mathbf{q}_{N+1}^N, \dots, \mathbf{q}_{N+m_N}^N \in Q$ chosen so that $\widetilde{p}_N$ is a closed path.

Since $u \in \mathcal{S} \subset \mathcal{D}$, we have $G_{\widetilde{p}_N}(u) = 0$ and therefore

$$G_{\widetilde{p}_N}(f) = G_{\widetilde{p}_N}(f-u) = G_{\widetilde{p}_N}(r). \tag{2.13}$$

We estimate $|G_{\widetilde{p}_N}(r)|$ using the same argument as in (2.7)–(2.8). Along the first N points of $\widetilde{p}_N$, the values of r alternate between $\pm\|r\|$, hence the contribution of these N points to the alternating sum has absolute value exactly $N\|r\|$. For the added points we only know $|r| \le \|r\|$, so their contribution is bounded in absolute value by $m_N\|r\|$. Consequently,

$$\frac{N-m_N}{N+m_N}\,\|r\| \le |G_{\widetilde{p}_N}(r)| \le \|r\|.$$

Combining this with (2.13) yields

$$\frac{N-m_N}{N+m_N}\,\|f-u\| \le |G_{\widetilde{p}_N}(f)| \le \|f-u\|.$$

Since $m_N \leq N_0$ while $N \to \infty$, the factor $\frac{N-m_N}{N+m_N}$ tends to 1. Hence, for every $\varepsilon > 0$ there exists N such that

$$|G_{\widetilde{p}_N}(f)| > \|f-u\| - \varepsilon,$$

and therefore

$$\sup_{p \subset Q} |G_p(f)| \geq \|f-u\|.$$

On the other hand, applying (2.12) with $h = u$ gives

$$\sup_{p \subset Q} |G_p(f)| \leq \|f-u\|.$$

Thus $\sup_{p\subset Q} |G_p(f)| = \|f-u\|$, and by Theorem 2.2 we conclude

$$E(f, \mathcal{S}) = \sup_{p \subset Q} |G_p(f)| = \|f-u\|.$$

Hence u is a best approximation to f from $\mathcal{S}$ in Case (2) as well.

"Only if" part. Assume that $u \in \mathcal{S}$ is a best approximation to f from $\mathcal{S}$, that is,

$$\|f-u\| = E(f, \mathcal{S}). \tag{2.14}$$

By Theorem 2.2,

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

Combining with (2.14) yields

$$\sup_{p \subset Q} |G_p(f)| = \|f-u\| = \|r\|.$$

Since $u \in \mathcal{S} \subset \mathcal{D}$ and G_p annihilates $\mathcal{D}$, we have $G_p(f) = G_p(f-u) = G_p(r)$ for every closed path $p \subset Q$, and thus

$$\sup_{p \subset Q} |G_p(r)| = \|r\|. \tag{2.15}$$

We now show that u is also a best approximation to f from $\mathcal{D}$. Since $\mathcal{S} \subset \mathcal{D}$, we have $E(f, \mathcal{D}) \leq E(f, \mathcal{S})$. On the other hand, by the argument used in the proof of (2.3) (with $\mathcal{D}$ in place of $\mathcal{S}$),

$$\sup_{p \subset Q} |G_p(f)| \leq E(f, \mathcal{D}).$$

Hence, by (2.14) and Theorem 2.2,

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

which implies $\|f-u\| = E(f, \mathcal{D})$. Thus u is a best approximation to f from $\mathcal{D}$ as well.

If there exists a closed path $p_0 \subset Q$ such that $|G_{p_0}(r)| = \|r\|$, then, as in Case (1), equality in the alternating average implies that $r(\mathbf{p}_k) = \pm(-1)^k \|r\|$ along p_0, i.e. p_0 is extremal for r. Hence condition (1) holds.

If no such closed path exists, then the supremum in (2.15) is not attained. Since u is also a best approximation to f from $\mathcal{D}$, the same extremal-path alternative used in the proof of Theorem 2.2 applies to the difference $r = f - u$. In this situation, for every $N \in \mathbb{N}$ there exists an extremal path of length N for r. Thus condition (2) holds. This completes the proof. □

Remark 2.1. Theorem 2.3 provides a characterization of best approximations from the class $\mathcal{S}$, not an existence statement. In particular, it does not assert that $\mathcal{S}$ is proximinal in $C(Q)$. Even when $E(f,\mathcal{S})$ admits the closed path formula of Theorem 2.2, the infimum defining $E(f,\mathcal{S})$ may fail to be attained by any element of $\mathcal{S}$. Equivalently, a best approximation from $\mathcal{S}$ may not exist for a given f. Theorem 2.3 states that whenever a best approximant $u \in \mathcal{S}$ exists, it is exactly characterized by the extremal-path alternative (1)–(2).

Remark 2.2. In [6], a similar but distinct class of RBFNNs was considered. That class incorporates varying smoothing factors, which make its construction and implementation more complex than in the case studied here, where all smoothing factors are fixed. More precisely, the following set of RBFNNs was considered:

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

Although some of the ideas in this paper are inspired by those in [6], the main result of this work (Theorem 2.2) cannot be derived directly from the results of [6], since allowing an arbitrary non-polynomial activation function there inevitably requires the presence of varying smoothing factors. In addition, if a non-polynomial activation function turns out to be the activation function considered here, then the main result of [6] follows directly from Theorem 2.2 as a corollary, since $\mathcal{S}(g, \mathbf{c}_1, \mathbf{c}_2, \sigma) \subset \mathcal{G}(g, \mathbf{c_1}, \mathbf{c_2})$ (hence $E(f,\mathcal{G}) \leq E(f,\mathcal{S})$) and $\sup_{p \subset Q} |G_p(f)| \leq E(f,\mathcal{G})$. The latter inequality can be proved in the same manner as the proof of (2.3).

Remark 2.3. In [7], the approximation error of the following class of RBFNNs was evaluated:

$$\mathcal{A} = \mathcal{A}(g, \mathbf{c}_1, \mathbf{c}_2) = \left\{ \sum_{i=1}^{m} w_i \, g(\|\mathbf{x} - \mathbf{c}_i\| - \theta_i) : \; w_i, \theta_i \in \mathbb{R}, \; \mathbf{c}_i = \mathbf{c}_1 \text{ or } \mathbf{c}_i = \mathbf{c}_2 \right\},$$

where g is a non-mean periodic function (for mean periodic functions, see [16]). It is clear from the proof technique that Theorem 2.2 remains valid not only for the practically useful functions considered here, but also for any non-mean periodic activation function. Moreover, the approximating class $\mathcal{S}(g, \mathbf{c}_1, \mathbf{c}_2, \sigma)$ is more general, as it allows an arbitrary fixed smoothing factor.

The present paper extends the results of [7] in two directions. First, it treats a more general class of RBFNNs by allowing an arbitrary fixed smoothing factor. Second, it provides a characterization of best approximations in terms of extremal paths, which was not addressed in [7].

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] C. Arteaga, I. Marrero, Universal approximation by radial basis function networks of Delsarte translates, *Neural Netw.* **46** (2013), 299–305.

[3] 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.

[4] A.Kh. Asgarova, V.E. Ismailov, A Chebyshev-type theorem characterizing best approximation of a continuous function by elements of the sum of two algebras, (Russian) translated from *Mat. Zametki* **109** (2021), no. 1, 19–26, *Math. Notes* **109** (2021), no. 1-2, 15–20.

[5] A.Kh. Asgarova, A.A. Huseynli, V.E. Ismailov, A Chebyshev-type alternation theorem for best approximation by a sum of two algebras, *Proc. Edinb. Math. Soc.* **(2) 66** (2023), no. 4, 971–978.

[6] A.Kh. Asgarova, I.K. Maharov, On the error of approximation by RBF neural networks with two hidden nodes, *Proc. Inst. Math. Mech. Natl. Acad. Sci. Azerb.* **50** (2024), no. 1, 152-161.

[7] A.Kh. Asgarova, Approximation with RBF neural networks using unit smoothing factors and translations, *Adv. Studies: Euro-Tbilisi Math. J.* **18** (2025), no. 3, 79-90.

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

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

[10] 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.

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

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

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

[14] V.E. Ismailov, Approximation error of single hidden layer neural networks with fixed weights, *Inform. Processing Lett.* **185** (2024), Paper No. 106467

[15] A. Ismayilova, M. Ismayilov, On the universal approximation property of radial basis function neural networks, *Ann. Math. Artif. Intell.* **92** (2024), no. 3, 691-701.

[16] J.P. Kahane, *Lectures on Mean Periodic Functions*, Tata Institute of Fundamental Research, Bombay, 1959.

[17] S.Ya.Khavinson, *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.

[18] J. Liu, *Radial Basis Function (RBF) neural network control for mechanical systems: design, analysis and Matlab simulation*, 2013, Springer, 365 pp.

[19] I. Marrero, Relaxed conditions for universal approximation by radial basis function neural networks of Hankel translates, *AIMS Math.* **10** (2025), no. 5, 10852-10865.

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

[21] 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.

[22] G.A. Montazer, D. Giveki, An improved radial basis function neural network for object image retrieval, *Neurocomputing* **168** (2015), 221-233.

[23] H. Muzhou, H. Xuli, The multidimensional function approximation based on constructive wavelet RBF neural network, *Appl. Soft Comput.* **11** (2011), 2173-2177.

[24] J. Park, I.W. Sandberg, Approximation and Radial-Basis-Function Networks, *Neural Comput.* **5** (1993), no.2, 305-316.

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

[26] A. Pinkus, *Ridge functions.* Cambridge Tracts in Mathematics, 205. Cambridge University Press, 2015, 207 pp.

[27] C.A. Rogers, A.J. Kassab, E.A. Divo, Z. Ostrowski, R.A. Bialecki, An inverse POD-RBF network approach to parameter estimation in mechanics, *Inverse Problems in Sci. and Engineer.* **20 (5)** (2012), 749–767.

[28] D.J. Singh, N. K Verma, A.K. Ghosh, J. Sanwale, A. Malagaudanavar, Aerodynamic parameter estimation using two-stage radial basis function neural network, In 2017 International Conference on Sensing, Diagnostics, Prognostics, and Control (SDPC), IEEE, 2017, 461-467.

[29] 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.

[30] S. Yang, T.O. Ting, K.L. Man, S.-U. Guan, Investigation of neural networks for function approximation, *Procedia Computer Sci.* **17** (2013), 586-594.

[31] S.-H. Yoo, S.-K. Oh, W. Pedrycz, Optimized face recognition algorithm using radial basis function neural networks and its practical applications, *Neural Netw.* **69** (2015), 111-125.

[32] H. Yuan, X. Shen, X. Chen, Prediction of fishing ground based on RBF neural network, *Procedia Engineer.* **15** (2011), 3240-3244.

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

Received: August 16, 2025; Revised: December 27, 2025; Accepted: January 30, 2026