A note on the equioscillation theorem for best ridge function approximation

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

Abstract. We consider the approximation of a continuous function, defined on a compact set of the $d$-dimensional Euclidean space, by sums of two ridge functions. We obtain a necessary and sufficient condition for such a sum to be a best approximation. The result resembles the classical Chebyshev equioscillation theorem for polynomial approximation.

Mathematics Subject Classifications: 41A30, 41A50, 46B50, 46E15

Keywords: ridge function; Chebyshev equioscillation theorem; a best approximation; path; weak* convergence

1. Introduction

Let $Q$ be compact set in the $d$-dimensional Euclidean space and $C(Q)$ be the space of continuous real-valued functions on $Q$. Consider the approximation of a function $f \in C(Q)$ by sums of the form $g_1(a_1 \cdot x) + g_2(a_2 \cdot x)$, where $a_i$ are fixed vectors (directions) in $\mathbb{R}^d \setminus \{0\}$ and $g_i$ are continuous univariate functions. We are interested in characterization of a best approximation. Note that functions of the form $g(a \cdot x)$ are called ridge functions. These functions and their linear combinations arise naturally in problems of computerized tomography (see, e.g., [26]), statistics (see, e.g., [9, 11]), partial differential equations [19] (where they are called plane waves), neural networks (see, e.g., [35] and references therein), and approximation theory (see, e.g., [12, 27, 31, 32]). In the past few years, problems of ridge function representation have gained special attention among researchers (see e.g. [1, 24, 25, 34]). For more on ridge functions and application areas see a recently published monograph by Pinkus [33].

Characterization theorems for best approximating elements are essential in approximation theory. The classical and most striking example of such a theorem are the Chebyshev equioscillation theorem. This theorem characterizes the unique best uniform approximation to a continuous real-valued function $F(t)$ by polynomials $P(t)$ of degree at most $n$, by the oscillating nature of the difference $F(t) - P(t)$. The result says that if such polynomial has the property that for some particular $n + 2$ points $t_i$ in $[0, 1]$

$$F(t_i) - P(t_i) = (-1)^i \max_{x \in [0,1]} |F(x) - P(x)|, \ i = 1, \ldots, n + 2,$$

then $P$ is the best approximation to $F$ on $[0, 1]$. The monograph of Natanson [30] contains a very rich commentary on this theorem. Some general alternation type theorems applying
functions by sums of two ridge functions. To be more precise, let $Q$ be a compact subset
of the space $\mathbb{R}^d$. Fix two directions $\mathbf{a}_1$ and $\mathbf{a}_2$ in $\mathbb{R}^d$ and consider the following space

$$
\mathcal{R} = \mathcal{R}(\mathbf{a}_1, \mathbf{a}_2) = \{ g_1(\mathbf{a}_1 \cdot \mathbf{x}) + g_2(\mathbf{a}_2 \cdot \mathbf{x}) : g_1, g_2 \in C(\mathbb{R}) \}.
$$

Note that the space $\mathcal{R}$ is a linear space. Assume a function $f \in C(Q)$ is given. We ask
and answer the following question: which geometrical conditions imposed on $G_0 \in \mathcal{R}$ is
necessary and sufficient for the equality

$$
\| f - G_0 \| = \inf_{G \in \mathcal{R}} \| f - G \|?
$$

Here $\| \cdot \|$ denotes the standard uniform norm in $C(Q)$. Recall that functions $G_0$ satisfying
(1.1) are called best approximations or extremal elements.

It should be remarked that in the special case when $Q \subset \mathbb{R}^2$ and $\mathbf{a}_1$ and $\mathbf{a}_2$ coincide
with the coordinate directions, the above question was answered by Khavinson [21]. In
[21], he obtained an equioscillation theorem for a best approximating sum $\varphi(x) + \psi(y)$. In
our papers [12, 16], Chebyshev type theorems were proven for ridge functions under
additional assumption that $Q$ is convex. For a more recent and detailed discussion of an
equioscillation theorem in ridge function approximation see Pinkus [33].

2. Equioscillation theorem for ridge functions

We start with a definition of paths with respect to two directions. These objects will
play an essential role in our further analysis.

Definition 2.1 (see [16]). A finite or infinite ordered set $p = (p_1, p_2, \ldots) \subset Q$
with $p_i \neq p_{i+1}$, and either $\mathbf{a}_1 \cdot p_1 = \mathbf{a}_1 \cdot p_2, \mathbf{a}_2 \cdot p_2 = \mathbf{a}_2 \cdot p_3, \mathbf{a}_1 \cdot p_3 = \mathbf{a}_1 \cdot p_4, \ldots$ or
$\mathbf{a}_2 \cdot p_1 = \mathbf{a}_2 \cdot p_2, \mathbf{a}_1 \cdot p_2 = \mathbf{a}_1 \cdot p_3, \mathbf{a}_2 \cdot p_3 = \mathbf{a}_2 \cdot p_4, \ldots$ is called a path with respect to the
directions $\mathbf{a}_1$ and $\mathbf{a}_2$.

In the sequel, we will simply use the term “path” instead of the expression “path with
respect to the directions $\mathbf{a}_1$ and $\mathbf{a}_2$”. If in a finite path $(p_1, \ldots, p_n, p_{n+1})$, $p_{n+1} = p_1$ and $n$
is an even number, then the path $(p_1, \ldots, p_n)$ is said to be closed. Note that for a closed
path $(p_1, \ldots, p_{2n})$ and any function $G \in \mathcal{R}$, $G(p_1) - G(p_2) + \cdots - G(p_{2n}) = 0$.

Paths, in the special case when $Q \subset \mathbb{R}^2$, $\mathbf{a}_1$ and $\mathbf{a}_2$ coincide with the coordinate
directions, are geometrically explicit objects. In this case, a path is a finite ordered set
$(p_1, \ldots, p_n)$ in $\mathbb{R}^2$ with the line segments $[p_i, p_{i+1}], i = 1, \ldots, n$, alternatively perpendicular
to the $x$ and $y$ axes (see, e.g., [2], [8] [10], [17], [18], [20], [28]). These objects were first
introduced by Diliberto and Straus [7] (in [7], they are called “permissible lines”). They
appeared further in a number of papers with several different names such as “bolts” (see,
On the other hand, for any function $G$, there exists an infinite path ($\exists \infty$) case, $f$ valid. The following linear functional $p$ exists a closed path ($\exists C$) approximation to a function $h$ e.g., \([2, 20, 28]\), “trips” (see \([29]\)), “links” (see, e.g., \([3, 22, 23]\)), etc. Paths with respect to two directions $a_1$ and $a_2$ were exploited in some papers devoted to ridge function interpolation (see, e.g., \([3, 13]\)). In \([14, 15]\), paths were generalized to those with respect to a finite set of functions. The last objects turned out to be very useful in problems of interpolation (see, e.g., \([3, 13]\)).

In the sequel, we need the concept of an “extremal path”, which is defined as follows.

**Definition 2.2 (see \([16]\)).** A finite or infinite path $(p_1, p_2, \ldots)$ is said to be extremal for a function $h \in C(Q)$ if $h(p_i) = (-1)^i \|h\|, i = 1, 2, \ldots$ or $h(p_i) = (-1)^{i+1} \|h\|, i = 1, 2, \ldots$

The purpose of this note is to prove the following theorem.

**Theorem 2.1.** Assume $Q$ is a compact subset of $\mathbb{R}^d$. A function $G_0 \in \mathcal{R}$ is a best approximation to a function $f \in C(Q)$ if and only if there exists a closed or infinite path $p = (p_1, p_2, \ldots)$ extremal for the function $f - G_0$.

**Proof.** Sufficiency. There are two possible cases. The first case happens when there exists a closed path $(p_1, \ldots, p_{2n})$ extremal for the function $f - G_0$. Let us check that in this case, $f - G_0$ is a best approximation. Indeed, on the one hand, the following equalities are valid.

$$
\left| \sum_{i=1}^{2n} (-1)^i f(p_i) \right| = \sum_{i=1}^{2n} (-1)^i \left[ f - G_0 \right] (p_i) = 2n \| f - G_0 \|.
$$

On the other hand, for any function $G \in \mathcal{R}$, we have

$$
\left| \sum_{i=1}^{2n} (-1)^i f(p_i) \right| = \sum_{i=1}^{2n} (-1)^i \left[ f - G \right] (p_i) \leq 2n \| f - G \|.
$$

Therefore, $\| f - G_0 \| \leq \| f - G \|$ for any $G \in \mathcal{R}$. That is, $G_0$ is a best approximation.

The second case happens when we do not have closed paths extremal for $f - G_0$, but there exists an infinite path $(p_1, p_2, \ldots)$ extremal for $f - G_0$. To analyze this case, consider the following linear functional

$$
l_q : C(Q) \to \mathbb{R}, \quad l_q(F) = \frac{1}{n} \sum_{i=1}^{n} (-1)^i F(q_i),
$$

where $q = \{q_1, \ldots, q_n\}$ is a finite path in $Q$. It is easy to see that the norm $\|l_q\| \leq 1$ and $\|l_q\| = 1$ if and only if the set of points of $q$ with odd indices $O = \{q_i \in q : i \text{ is an odd number}\}$ do not intersect with the set of points of $q$ with even indices $E = \{q_i \in q : i \text{ is an even number}\}$. Indeed, from the definition of $l_q$ it follows that $|l_q(F)| \leq \|F\|$ for all functions $F \in C(Q)$, whence $\|l_q\| \leq 1$. If $O \cap E = \emptyset$, then for a function $F_0$ with the property $F_0(q_i) = -1$ if $i$ is odd, $F_0(q_i) = 1$ if $i$ is even and $-1 < F_0(x) < 1$ elsewhere on $Q$, we have $|l_q(F_0)| = \|F_0\|$. Hence, $\|l_q\| = 1$. Recall that such a function $F_0$ exists on the basis of Urysohn’s great lemma.
Note that if \( q \) is a closed path, then \( l_q \) annihilates all members of the class \( \mathcal{R} \). But in general, when \( q \) is not closed, we do not have the equality \( l_q(G) = 0 \), for all members \( G \in \mathcal{R} \). Nonetheless, this functional has the important property that

\[
|l_q(g_1 + g_2)| \leq \frac{2}{n}(\|g_1\| + \|g_2\|),
\]

where \( g_1 \) and \( g_2 \) are ridge functions with the directions \( a_1 \) and \( a_2 \), respectively, that is, \( g_1 = g_1(a_1 \cdot x) \) and \( g_2 = g_2(a_2 \cdot x) \). This property is important in the sense that if \( n \) is sufficiently large, then the functional \( l_q \) is close to an annihilating functional. To prove (2.1), note that \( |l_q(g_1)| \leq \frac{2}{n} \|g_1\| \) and \( |l_q(g_2)| \leq \frac{2}{n} \|g_2\| \). These estimates become obvious if consider the chain of equalities \( g_1(a_1 \cdot x_1) = g_1(a_1 \cdot x_2), g_1(a_1 \cdot x_3) = g_1(a_1 \cdot x_4), \ldots \) (or \( g_1(a_1 \cdot x_2) = g_1(a_1 \cdot x_3), g_1(a_1 \cdot x_4) = g_1(a_1 \cdot x_5), \ldots \)) for \( g_1(a_1 \cdot x) \) and the corresponding chain of equalities for \( g_2(a_2 \cdot x) \)

Now consider the infinite path \( p = (p_1, p_2, \ldots) \) and form the finite paths \( p_k = (p_1, \ldots, p_k) \), \( k = 1, 2, \ldots \). For ease of notation, let us set \( l_k = l_{p_k} \). The sequence \( \{l_k\}_{k=1}^{\infty} \) is a subset of the unit ball of the conjugate space \( C^*(Q) \). By the Banach-Alaoglu theorem, the unit ball is weak* compact in the weak* topology of \( C^*(Q) \) (see, e.g., Rudin [36, p. 66]). From this theorem we derive that the sequence \( \{l_k\}_{k=1}^{\infty} \) must have weak* cluster points. Suppose \( l^* \) denotes one of them. Without loss of generality we may assume that \( l_k \xrightarrow{\text{weak}^*} l^* \), as \( k \to \infty \). From (2.1) it follows that \( l^*(g_1 + g_2) = 0 \). That is, \( l^* \in \mathcal{R}^\perp \), where the symbol \( \mathcal{R}^\perp \) stands for the annihilator of \( \mathcal{R} \). Since in addition \( \|l^*\| \leq 1 \), we can write that

\[
|l^*(f)| = |l^*(f - G)| = \|f - G\|,
\]

for all functions \( G \in \mathcal{R} \). On the other hand, since the infinite bolt \( p \) is extremal for \( f - G_0 \)

\[
|l_k(f - G_0)| = \|f - G_0\|, \quad k = 1, 2, \ldots
\]

Therefore,

\[
|l^*(f)| = |l^*(f - G_0)| = \|f - G_0\|.
\]

From (2.2) and (2.3) we conclude that

\[
\|f - G_0\| \leq \|f - G\|,
\]

for all \( G \in \mathcal{R} \). In other words, \( G_0 \) is a best approximation to \( f \). We proved the sufficiency of the theorem.

**Necessity.** The proof of this part is mainly based on the following theorem of Singer.

**Theorem 2.2 (see Singer [37]).** Let \( X \) be a compact space, \( U \) be a linear subspace of \( C(X) \), \( f \in C(X) \setminus U \) and \( u_0 \in U \). Then \( u_0 \) is a best approximation to \( f \) if and only if there exists a regular Borel measure \( \mu \) on \( X \) such that

1. The total variation \( \|\mu\| = 1 \);
2. \( \mu \) is orthogonal to the subspace \( U \), that is, \( \int_X ud\mu = 0 \) for all \( u \in U \);
(3) For the Jordan decomposition \( \mu = \mu^+ - \mu^- \),
\[
f(x) - u_0(x) = \begin{cases} 
\|f - u_0\| & \text{for } x \in S^+, \\
-\|f - u_0\| & \text{for } x \in S^-,
\end{cases}
\]
where \( S^+ \) and \( S^- \) are closed supports of the positive measures \( \mu^+ \) and \( \mu^- \), respectively.

Let us show how we use this theorem in the proof of necessity part of our theorem. Assume \( G_0 \in \mathcal{R} \) is a best approximation. For the subspace \( \mathcal{R} \), the existence of a measure \( \mu \) satisfying the conditions (1)-(3) is a direct consequence of Theorem 2.2. Let \( x_0 \) be any point in \( S^+ \). Consider the point \( y_0 = a_1 \cdot x_0 \) and a \( \delta \)-neighborhood of \( y_0 \). That is, choose an arbitrary \( \delta > 0 \) and consider the set \( I_\delta = (y_0 - \delta, y_0 + \delta) \cap a_1 \cdot Q \). Here, \( a_1 \cdot Q = \{a_1 \cdot x : x \in Q\} \). For any subset \( E \subset \mathbb{R} \), put
\[
E^i = \{x \in Q : a_i \cdot x \in E\}, \ i = 1, 2.
\]

Clearly, for some sets \( E \), one or both the sets \( E^i \) may be empty. Since \( I_\delta^1 \cap S^+ \) is not empty (note that \( x_0 \in I_\delta^1 \)), it follows that \( \mu^+(I_\delta^1) > 0 \). At the same time \( \mu(I_\delta^1) = 0 \), since \( \mu \) is orthogonal to all functions \( g_1(a_1 \cdot x) \). Therefore, \( \mu^-(I_\delta^1) > 0 \). We conclude that \( I_\delta^1 \cap S^- \) is not empty. Denote this intersection by \( A_\delta \). Tending \( \delta \) to 0, we obtain a set \( A \) which is a subset of \( S^- \) and has the property that for each \( x \in A \), we have \( a_1 \cdot x = a_1 \cdot x_0 \). Fix any point \( x_1 \in A \). Changing \( a_1 \), \( \mu^+ \), \( S^+ \) to \( a_2 \), \( \mu^- \) and \( S^- \) correspondingly, repeat the above process with the point \( y_1 = a_2 \cdot x_1 \) and a \( \delta \)-neighborhood of \( y_1 \). Then we obtain a point \( x_2 \in S^+ \) such that \( a_2 \cdot x_2 = a_2 \cdot x_1 \). Continuing this process, one can construct points \( x_3 \), \( x_4 \), and so on. Note that the set of all constructed points \( x_i \), \( i = 0, 1, \ldots \), forms a path. By Theorem 2.2, this path is extremal for the function \( f - G_0 \). We have proved the necessity and hence Theorem 2.1.

Remark. Theorem 2.1 was proven by Ismailov [12] and in a more general form by Pinkus [33] under additional assumption that \( Q \) is convex. Convexity assumption was made to guarantee continuity of the following functions
\[
g_{1,i}(t) = \max_{x \in Q} F(x) \quad \text{and} \quad g_{2,i}(t) = \min_{x \in Q} F(x), \ i = 1, 2,
\]
where \( F \) is an arbitrary continuous function on \( Q \). Note that in the proof given above we need not continuity of these functions.

References

[1] R. A. Aliev and V. E. Ismailov, On a smoothness problem in ridge function representation, Adv. in Appl. Math. 73 (2016), 154-169.

[2] V. I. Arnold, On functions of three variables (Russian) Dokl. Akad. Nauk SSSR 114 (1957), 679-681; English transl. in Amer. Math. Soc. Transl. 28 (1963), 51-54.
[3] D. Braess and A. Pinkus, Interpolation by ridge functions, *J. Approx. Theory* **73** (1993), 218-236.

[4] B. Brosowski and A. R. da Silva, A general alternation theorem, Approximation theory (Memphis, TN, 1991), 137–150, Lecture Notes in Pure and Appl. Math., 138, Dekker, New York, 1992.

[5] R. C. Buck, Alternation theorems for functions of several variables, *J. Approx. Theory* **1** (1968), 325–334.

[6] R. C. Cowsik, A. Klopotowski, M. G. Nadkarni, When is $f(x, y) = u(x) + v(y)$ ?, *Proc. Indian Acad. Sci. Math. Sci.* **109** (1999), 57–64.

[7] 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.

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

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

[10] 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.

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

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

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

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

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

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

[17] 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.
[18] 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.

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

[20] 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.

[21] S. Ya. Khavinson (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 transl. Math. USSR Izv. 3 (1969), 617-632.

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

[23] A. Klopotowski, M. G. Nadkarni, Shift invariant measures and simple spectrum, Colloq. Math. 84/85 (2000), 385-394.

[24] S. V. Konyagin, A. A. Kuleshov, On the continuity of finite sums of ridge functions (Russian), Mat. Zametki 98 (2015), 308-309; English transl. in Math. Notes 98 (2015), 336-338.

[25] S. V. Konyagin, A. A. Kuleshov, On some properties of finite sums of ridge functions defined on convex subsets of $\mathbb{R}^d$ (Russian), Trudy Matem. Inst. imeni Steklova 293 (2016), 193-200; English transl. in Proc. Steklov Inst. Math. 293 (2016), 186–193.

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

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

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

[30] I. P. Natanson, Constructive function theory, Vol. I. Uniform approximation, Translated from the Russian by Alexis N. Obolensky Frederick Ungar Publishing Co., New York 1964, 232 pp.
[31] B. Pelletier, Approximation by ridge function fields over compact sets, *J. Approx. Theory* **129** (2004), 230-239.

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

[33] A. Pinkus, *Ridge Functions*, Cambridge Tracts in Mathematics, 205. Cambridge University Press, Cambridge, 2015, 218 pp.

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

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

[36] W. Rudin, *Functional analysis*, McGraw-Hill Series in Higher Mathematics. McGraw-Hill Book Co., 1973, 397 pp.

[37] I. Singer, The theory of best approximation and functional analysis. Conference Board of the Mathematical Sciences Regional Conference Series in Applied Mathematics, No. 13. *Society for Industrial and Applied Mathematics, Philadelphia, Pa.*, 1974, 95 pp.