Proceedings of the Institute of Mathematics and Mechanics,
National Academy of Sciences of Azerbaijan
Volume 41, Number 1, 2015, Pages 146–152

ALTERNATING ALGORITHM FOR THE APPROXIMATION BY SUMS OF TWO COMPOSITIONS AND RIDGE FUNCTIONS

VUGAR E. ISMAILOV

Abstract. In the current paper, we consider the problem of approximation of a continuous multivariate function defined on a convex compact set X by sums of compositions of univariate functions with two fixed multivariate functions, in particular, by sums of ridge functions with two fixed directions. Under some assumptions imposed on X, we prove that the sequence produced by the alternating algorithm converges to the error of approximation.

1. Introduction

Let X be a convex compact subset of the space $\mathbb{R}^d$. Fix two continuous maps $s : X \longrightarrow \mathbb{R}$, $p : X \longrightarrow \mathbb{R}$ and consider the following spaces of compositions of functions:

$$\begin{aligned} S &= \{h \circ s : \ h \in C(\mathbb{R})\}, \\ P &= \{g \circ p : \ g \in C(\mathbb{R})\}, \\ U &= S + P. \end{aligned}$$

We are going to deal with the problem of approximating a continuous function $f : X \to \mathbb{R}$ using functions from the space U. By $s(X)$ and $p(X)$ we will denote the images of X under the mappings s and p respectively. Define the following operators

$$H \ : \ C(X) \to S, \ (Hf)(a) = \frac{1}{2}\left(\max_{\substack{x\in X \\ s(x)=a}} f(x) + \min_{\substack{x\in X \\ s(x)=a}} f(x) \right), \quad \text{for all } a \in s(X)$$

and

$$G \ : \ C(X) \to P, \ (Gf)(b) = \frac{1}{2}\left(\max_{\substack{x\in X \\ p(x)=b}} f(x) + \min_{\substack{x\in X \\ p(x)=b}} f(x) \right), \quad \text{for all } b \in p(X).$$

We are interested in algorithmic methods for computing the distance to a given continuous function $f : X \to \mathbb{R}$ from the manifold U. Historically, there is one procedure called the Diliberto-Straus algorithm [3]. This procedure can be

---

2010 *Mathematics Subject Classification.* 41A30, 41A63, 68W25.
Key words and phrases. Approximation error; Alternating algorithm; Path; Ridge function.

described as follows: Starting with $f_1 = f$ compute $f_2 = f_1 - Hf_1$, $f_3 = f_2 - Gf_2$, $f_4 = f_3 - Hf_3$, and so forth. Clearly, $f - f_n \in U$ and the sequence $\{\|f_n\|\}_{n=1}^{\infty}$ is nonincreasing. The question is if and when $\|f_n\|$ converges to the error of approximation from U?

Many approximation theoretic problems associated with the set U were considered in the relatively recent monograph by Khavinson [11]. In this monograph, the Diliberto-Straus algorithm was given a special attention (see [11, p.112-126]). Khavinson analyzed the algorithm in its simplest case, in which $s(x)$ and $p(x)$ are the coordinate functions and X is a rectangle in $\mathbb{R}^2$ with sides parallel to the coordinate axes. But the question on the convergence of the algorithm for other sets and other approximating functions remained unanswered there.

Note that the space U, in a particular case, turns into the space of sums of two ridge functions with fixed directions. A ridge function is a multivariate function of the form $g(\mathbf{a}\cdot\mathbf{x})$, where $\mathbf{a} \in \mathbb{R}^d \backslash \{\mathbf{0}\}$ is a fixed vector (direction), $\mathbf{x} \in \mathbb{R}^d$ is the variable, $\mathbf{a}\cdot\mathbf{x}$ is the Euclidean inner product and g is a univariate function. The literature abounds with the use of ridge functions and their linear combinations. Ridge functions arise naturally in various fields. They arise in partial differential equations (where they are called *plane waves*), in computerized tomography (the name ridge function was coined by Logan and Shepp [14] in one of the seminal papers on tomography), in statistics (especially, in the theory of projection pursuit and projection regression). Ridge functions are also the underpinnings of many central models in neural networks which has become increasing more popular in computer science, statistics, engineering, physics, etc. (see [17] and a great deal of references therein). A ridge function has a very simple structure. This structure makes ridge functions an interesting and useful approximating tool in multivariate approximation theory (see, e.g., [2, 5, 6, 9, 15, 16] and references therein).

2. Convergence of the algorithm

We start with the definition of the following objects called paths.

Definition 2.1 (see [8]). An ordered set $l = (x_1, x_2, ..., x_n) \subset X$, where $x_i \neq x_{i+1}$, with either $s(x_1) = s(x_2), p(x_2) = p(x_3), s(x_3) = s(x_4), ...$ or $p(x_1) = p(x_2), s(x_2) = s(x_3), p(x_3) = p(x_4), ...$ is called a path with respect to the functions s and p.

If in a path $(x_1, ..., x_n, x_{n+1})$, $x_{n+1} = x_1$ and n is an even number, then the path $l = (x_1, ..., x_n)$ is called to be closed. If $s(x)$ and $p(x)$ are the coordinate functions on $\mathbb{R}^2$, then definition 2.1 defines an ordinary path (or a bolt of lightning in a number of papers, see, e.g., [1, 11]). It is well known that the idea of ordinary paths, first introduced by Diliberto and Straus [3], played significant role in many problems of the approximation and interpolation of bivariate functions by sums of univariate functions (see, e.g., [4, 10, 11, 12, 13]). Paths with respect to two directions $\mathbf{a}$ and $\mathbf{b}$ (that is, with respect to the functions $\mathbf{a}\cdot\mathbf{x}$ and $\mathbf{b}\cdot\mathbf{x}$) were exploited in some papers devoted to ridge functions (see, e.g., [2, 5, 9]). In [7], 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 representation by linear superpositions.

With each path $l = (x_1, ..., x_n)$, we associate the following path functional

$$r_l(f) = \sum_{i=1}^{n} (-1)^{n+1} f(x_i).$$

It is an exercise to check that r_l is a linear bounded functional on $C(X)$ with the norm $\|r_l\| \leq 1$ and $\|r_l\| = 1$ if and only if the set of points x_i with odd indices i does not intersect with that of even indices. Besides, if l is closed, then the closed path functional $r_l \in U^{\perp}$, where $U^{\perp}$ is the annihilator of the subspace $U \subset C(X)$.

In the sequel, we assume that max and min functions in the definition of the operators H and G are continuous. Note that in the case of ridge functions, that is, when $s(x)$ and $p(x)$ are scalar product functions, the mentioned max and min functions are always continuous (see [9]).

The following theorem is valid.

Theorem 2.1. *Let $X \subset \mathbb{R}^d$ be a convex compact set with the property: for any path $l = (x_1, ..., x_n) \subset X$ there exist points $x_{n+1}, ..., x_{n+m} \in X$ such that $(x_1, ..., x_{n+m})$ is a closed path and m is not more than some positive integer N independent of l. Then $\|f_n\|$ converges to the error of approximation $dist(f, U)$.*

Proof. Let us write the above iteration (see Introduction) in the following form

$$\begin{aligned} f_1 &= f, \ f_{n+1} = f_n - q_n, \text{ where} \\ q_n &= Hf_n, \text{ if } n \text{ is odd;} \\ q_n &= Gf_n, \text{ if } n \text{ is even.} \end{aligned}$$

By the above assumption, all the functions q_n are continuous. Introduce the functions

$$\begin{aligned} u_n &= q_1 + \cdots + q_{2n-1}; \\ v_n &= q_2 + \cdots + q_{2n}. \end{aligned}$$

Clearly, $u_n \in S$ and $v_n \in P$. Besides, $f_{2n} = f - u_n - v_{n-1}$ and $f_{2n+1} = f - u_n - v_n$.

It is easy to see that the following inequalities hold

$$\|f_1\| \geq \|f_2\| \geq \|f_3\| \geq \cdots \geq dist(f, U). \tag{2.1}$$

From (2.1) it follows that there exists the limit

$$M = \lim_{n \to \infty} \|f_n\| \geq dist(f, U).$$

It is a consequence of the Hahn-Banach extension theorem that

$$dist(f, U) = \sup_{\substack{r \in U^{\perp} \\ \|r\| \leq 1}} |r(f)|,$$

where sup is attained by some functional. To complete the proof it is enough for arbitrary positive number ε to find a functional r such that $r \in U^{\perp}$, $\|r\| \leq 1$ and

$$|r(f)| \geq M - \varepsilon. \tag{2.2}$$

Let ε be arbitrarily small positive real number. Choose an integer k such that

$$\frac{(2k+2)M}{2k+2+N} > M - \frac{\varepsilon}{2} \text{ and } \frac{N\,\|f\|}{2k+2+N} < \frac{\varepsilon}{4}. \tag{2.3}$$

Set now $\alpha = \frac{\varepsilon}{2^{2k+2}}$. There exists a number n_k such that for all $n \geq n_k$

$$\|f_n\| \leq M + \alpha$$

Now we construct a path $l = [x_1, ..., x_{2k+2}]$ with the property that $|r_l(f_{n_k})| \geq M - \varepsilon/4$. Without loss of generality we may assume that n_k is even. Then $n_k + 2k$ is also even and since

$$\max_{\substack{x \in X \\ s(x)=a}} f_{n_k+2k}(x) = -\min_{\substack{x \in X \\ s(x)=a}} f_{n_k+2k}(x), \quad \text{for all } a \in s(X),$$

there exists points x_1 and x_2 such that $f_{n_k+2k}(x_1) = \|f_{n_k+2k}\|$, $f_{n_k+2k}(x_2) = -\|f_{n_k+2k}\|$ and $s(x_1) = s(x_2)$. This can be written in the form

$$f_{n_k+2k-1}(x_1) - Hf_{n_k+2k-1}(s(x_1)) = \|f_{n_k+2k}\|\,; \tag{2.4}$$

$$f_{n_k+2k-1}(x_2) - Hf_{n_k+2k-1}(s(x_2)) = -\|f_{n_k+2k}\|\,. \tag{2.5}$$

Therefore,

$$\begin{aligned} Hf_{n_k+2k-1}(s(x_1)) = Hf_{n_k+2k-1}(s(x_2)) = \\ = f_{n_k+2k-1}(x_1) - \|f_{n_k+2k}\| \leq \|f_{n_k+2k-1}\| - M \leq \alpha \end{aligned} \tag{2.6}$$

and

$$\begin{aligned} Hf_{n_k+2k-1}(s(x_1)) = Hf_{n_k+2k-1}(s(x_2)) = \\ = f_{n_k+2k-1}(x_2) + \|f_{n_k+2k}\| \geq -(\|f_{n_k+2k-1}\| - M) \geq -\alpha. \end{aligned} \tag{2.7}$$

From (2.4)-(2.7) we also can obtain that

$$f_{n_k+2k-1}(x_1) = Hf_{n_k+2k-1}(s(x_1)) + \|f_{n_k+2k}\| \geq M - \alpha; \tag{2.8}$$

$$f_{n_k+2k-1}(x_2) = Hf_{n_k+2k-1}(s(x_1)) - \|f_{n_k+2k}\| \leq -M + \alpha. \tag{2.9}$$

Now since

$$\max_{\substack{x \in X \\ p(x)=b}} f_{n_k+2k-1}(x) = -\min_{\substack{x \in X \\ p(x)=b}} f_{n_k+2k-1}(x), \quad \text{for all } b \in p(X),$$

and (2.9) is valid, there exist points x_3 and $x_3^{'}$ such that

$$f_{n_k+2k-1}(x_3) \geq M - \alpha, \quad p(x_3) = p(x_2) \tag{2.10}$$

and

$$f_{n_k+2k-1}(x_3^{'}) \leq -M + \alpha, \quad p(x_3^{'}) = p(x_1) \tag{2.11}$$

In the inequalities (2.8)-(2.11), replace f_{n_k+2k-1} by $f_{n_k+2k-2} - Gf_{n_k+2k-2}$. Then we will have

$$f_{n_k+2k-2}(x_1) - Gf_{n_k+2k-2}(p(x_1)) \geq M - \alpha; \tag{2.12}$$

$$f_{n_k+2k-2}(x_2) - Gf_{n_k+2k-2}(p(x_2)) \leq -M + \alpha; \tag{2.13}$$

$$f_{n_k+2k-2}(x_3) - Gf_{n_k+2k-2}(p(x_3)) \geq M - \alpha; \tag{2.14}$$

$$f_{n_k+2k-2}(x_3') - Gf_{n_k+2k-2}(p(x_3')) \leq -M + \alpha. \tag{2.15}$$

If take into account the inequalities

$$\begin{aligned} f_{n_k+2k-2}(x_1) &\leq \|f_{n_k+2k-2}\| \leq M + \alpha, \\ f_{n_k+2k-2}(x_3) &\leq \|f_{n_k+2k-2}\| \leq M + \alpha, \\ f_{n_k+2k-2}(x_2) &\geq -\|f_{n_k+2k-2}\| \geq -M - \alpha, \\ f_{n_k+2k-2}(x_3') &\geq -\|f_{n_k+2k-2}\| \geq -M - \alpha, \end{aligned}$$

in the above inequalities (2.12)-(2.15), we can write that

$$-2\alpha \leq Gf_{n_k+2k-2}(p(x_1)) = Gf_{n_k+2k-2}(p(x_3')) \leq 2\alpha; \tag{2.16}$$

$$-2\alpha \leq Gf_{n_k+2k-2}(p(x_2)) = Gf_{n_k+2k-2}(p(x_3)) \leq 2\alpha. \tag{2.17}$$

Considering these inequalities in (2.12)-(2.15), we obtain that

$$\begin{aligned} f_{n_k+2k-2}(x_1) &\geq M - 3\alpha; \\ f_{n_k+2k-2}(x_2) &\leq -M + 3\alpha; \end{aligned}$$

$$f_{n_k+2k-2}(x_3) \geq M - 3\alpha; \tag{2.18}$$

Now since

$$\max_{\substack{x \in X \\ s(x)=a}} f_{n_k+2k-2}(x) = -\min_{\substack{x \in X \\ s(x)=a}} f_{n_k+2k-2}(x), \quad \text{for all } a \in s(X),$$

and (2.18) is valid, there exists a point x_4 such that $s(x_4) = s(x_3)$ and

$$f_{n_k+2k-2}(x_4) \leq -M + 3\alpha.$$

Repeating the above process for the function

$$f_{n_k+2k-2}(x) = f_{n_k+2k-3}(x) - Hf_{n_k+2k-3}(s(x))$$

we obtain that

$$-4\alpha \leq Hf_{n_k+2k-3}(s(x_i)) \leq 4\alpha, \ i = 1, 2, 3, 4, \tag{2.19}$$

and

$$\begin{aligned} f_{n_k+2k-3}(x_1) &\geq M - 7\alpha; \\ f_{n_k+2k-3}(x_2) &\leq -M + 7\alpha; \\ f_{n_k+2k-3}(x_3) &\geq M - 7\alpha; \\ f_{n_k+2k-3}(x_4') &\leq -M + 7\alpha. \end{aligned}$$

By the same way as above, we can find a point x_5 such that $p(x_5) = p(x_4)$ and

$$f_{n_k+2k-3}(x_5) \geq M - 7\alpha.$$

Continuing this process sequentially backwards until the function f_{n_k}, we obtain the points $x_1, x_2, ..., x_{2k+2}$ with the property that $s(x_1) = s(x_2)$, $p(x_2) = p(x_3), ..., s(x_{2k+1}) = s(x_{2k+2})$. In other words, these points form a path, which we denote by l. Note that in the above process, we also deal with the points like $x_3^{'}$, but these points play only auxiliary role: they are needed in obtaining inequalities like (2.16), (2.17), (2.19), etc. At the points of the path l the function f_{n_k}will obey the inequalities:

$$f_{n_k}(x_i) \geq M - (2^{2k} - 1)\alpha \geq M - \varepsilon/4, \text{ for } i = 1, 3, ..., 2k+1$$

and

$$f_{n_k}(x_j) \leq -M + (2^{2k} - 1)\alpha \leq -M + \varepsilon/4, \text{ for } j = 2, 4, ..., 2k+2.$$

Using these inequalities, we can estimate the value of the functional r_l at f_{n_k} :

$$|r_l(f_{n_k})| \geq \frac{(k+1)(M - \varepsilon/4) - (k+1)(-M + \varepsilon/4)}{2k+2} = M - \varepsilon/4. \tag{2.20}$$

By the hypothesis of the theorem we can close the path l by adding to it not more than N points. Without loss of generality we may assume that a closed path $t \subset X$ is obtained from l by adding precisely N points. Then we can write that

$$|r_t(f)| = |r_t(f_{n_k})| > \frac{(2k+2)\,|r_l(f_{n_k})|}{2k+2+N} - \frac{N\,\|f_{n_k}\|}{2k+2+N} \tag{2.21}$$

Now since $\|f_{n_k}\| \leq \|f\|$ it follows from (2.3), (2.20) and (2.21) that

$$|r_t(f)| > M - \varepsilon. \tag{2.22}$$

Since the functional $r_t \in U^{\perp}$, $\|r_t\| \leq 1$, the inequality (2.22) together with (2.2) completes the proof. □

Acknowledgements

This research was supported by the Science Development Foundation under the President of the Republic of Azerbaijan - Grant EIF-2013-9(15)-46/11/1.

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

[3] 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.

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

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

[6] V. E. Ismailov, A review of some results on ridge function approximation, *Azerb. J. Math.* **3** (2013), no.1, 3-51.
[7] V. E. Ismailov, A note on the representation of continuous functions by linear superpositions, *Expositiones Mathematicae* **30** (2012), 96-101.
[8] V. E. Ismailov, On the approximation by compositions of fixed multivariate functions with univariate functions, *Studia Mathematica* **183** (2007), 117-126.
[9] V. E. Ismailov, Characterization of an extremal sum of ridge functions. *J. Comput. Appl. Math.* **205** (2007), 105–115.
[10] V. E. Ismailov, On methods for computing the exact value of the best approximation by sums of functions of one variable (Russian), *Sibirsk. Mat. Zh.* **47** (2006), 1076–1082; translation in *Siberian Math. J.* **47** (2006), no. 5, 883–888.
[11] 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.
[12] W. A. Light, E. W. Cheney, On the approximation of a bivariate function by the sum of univariate functions. *J. Approx. Theory* **29** (1980), 305–322.
[13] W. A. Light, E. W. Cheney, *Approximation theory in tensor product spaces,* Lecture Notes in Mathematics, 1169. Springer-Verlag, Berlin, 1985, 157 pp.
[14] B. F. Logan and L. A. Shepp, Optimal reconstruction of a function from its projections, *Duke Math.J.* **42** (1975), 645-659.
[15] V. E. Maiorov, On best approximation by ridge functions, *J.Approx. Theory* **99** (1999), 68-94.
[16] 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.
[17] A. Pinkus, Approximation theory of the MLP model in neural networks, *Acta Numerica* **8** (1999), 143-195.

Vugar E. Ismailov
Institute of Mathematics and Mechanics of NAS of Azerbaijan, 9 B. Vahabzadeh str., AZ1141, Baku, Azerbaijan.
E-mail address: vugaris@mail.ru

Received: December 8, 2014; Accepted: June 5, 2015