Vugar E. ISMAILOV

ON THE APPROXIMATION BY LINEAR COMBINATIONS OF RIDGE FUNCTIONS IN L_2 METRIC

Abstract

The problem considered in the paper is to approximate a multivariate function by linear combinations of ridge functions with fixed directions in L_2 metric. We establish characteristic properties of an extremal element and find its analytic form.

1. Introduction

A *ridge function* is a multivariate function of the form

$$g\left(\mathbf{a}\cdot\mathbf{x}\right)=g\left(a_1x_1+\cdots+a_nx_n\right),$$

where $g:\mathbb{R}\to\mathbb{R}$ and $\mathbf{a}=(a_1,...,a_n)$ is a fixed vector (direction) in $\mathbb{R}^n\setminus\{\mathbf{0}\}$. In other words, it is a multivariate function constant on the parallel hyperplanes $\mathbf{a}\cdot\mathbf{x}=\alpha,\alpha\in\mathbb{R}$. Ridge functions and their combinations arise in various contexts. They arise naturally in problems of computerized tomography (see, e.g., [7, 10]), statistics (see, e.g., [1, 5]), partial differential equations [6], neural networks (see, e.g., [2, 11, 13, 14]), and approximation theory (see, e.g., [2, 3, 4, 8, 9, 11, 12]).

The term ridge function is rather recent. These functions have been considered for a long time under the name of *plane waves* (see, for example, [6]). We refer the reader to Pinkus [12] for an excellent short survey on ridge functions and various motivations for their research.

In some applications, one is interested in the following sets of ridge functions. The first is given by

$$\mathcal{R}\left(\mathbf{a}^1,...,\mathbf{a}^r\right)=\left\{\sum_{i=1}^{r}g_i\left(\mathbf{a}^i\cdot\mathbf{x}\right):g_i:\mathbb{R}\to\mathbb{R},i=1,...,r\right\}.$$

That is, we consider linear combinations of ridge functions with a finite number of fixed directions. It is clear that this is a linear space.

The second set is

$$\mathcal{R}_r=\left\{\sum_{i=1}^{r}g_i\left(\mathbf{a}^i\cdot\mathbf{x}\right):\mathbf{a}^i\in\mathbb{R}^n\setminus\{\mathbf{0}\},g_i:\mathbb{R}\to\mathbb{R},\quad i=1,...,r\right\}.$$

In this case, we fix r and variate both the functions g_i and the directions $\mathbf{a}^i$. Notice that $\mathcal{R}_r$ is not a linear space.

Some problems of approximation by functions from both the set $\mathcal{R}\left(\mathbf{a}^1,...,\mathbf{a}^r\right)$ and the set $\mathcal{R}_r$ were investigated by a number of authors. For example, one

essential method of approximating from $\mathcal{R}(\mathbf{a}^1, ..., \mathbf{a}^r)$, its defects and advantages were discussed in [12]. Lin and Pinkus [8] characterized $\mathcal{R}(\mathbf{a}^1, ..., \mathbf{a}^r)$, i.e. they found means of determining if a continuous function f (defined on $\mathbb{R}^n$) is of the form $\sum_{i=1}^{r} g_i(\mathbf{a}^i \cdot \mathbf{x})$ for some given $\mathbf{a}^1, ..., \mathbf{a}^{\mathbf{r}} \in \mathbb{R}^n \setminus \{\mathbf{0}\}$, but unknown continuous $g_1, ..., g_r$. Two other characterizations of $\mathcal{R}(\mathbf{a}^1, ..., \mathbf{a}^r)$ in the uniform norm may be found in Diaconis and Shahshahani [3]. The efficiency of approximation by functions from $\mathcal{R}_r$ in L_2 metric was investigated by Petrushev [11]. Maiorov [9] established upper and lower bounds for the error of approximation to functions in Sobolev class from $\mathcal{R}_r$ in L_2 metric.

In the following, we are going to consider some essential problems in approximating from the set $\mathcal{R}(\mathbf{a}^1, ..., \mathbf{a}^n)$ in L_2 metric. These problems include both the characterization and the construction of an extremal element. We are also interested in formulas for the approximation error. Unfortunately, our results involve cases in which the space dimension is equal to the number of fixed directions. Nevertheless, we hope that these results will encourage some of readers to seek more general and better ways of approaching the considered here problems. Since for the basis vectors $\mathbf{a}^1, ..., \mathbf{a}^n$ the functions $g_i(\mathbf{a}^i \cdot \mathbf{x})$, $i = 1, ..., n$, are univariate, the results of the paper are also applicable to the approximation of a multivariate function by sums of univariate functions.

2.Main results

Let X be a subset of $\mathbb{R}^n$ whose Lebesgue measure is finite. Consider the approximation of a function $f(\mathbf{x}) = f(x_1, ..., x_n)$ from $L_2(X)$ by functions from the manifold $\mathcal{R}(\mathbf{a}^1, ..., \mathbf{a}^n)$. We suppose that the functions $g_i(\mathbf{a}^i \cdot \mathbf{x}), i = 1, ..., n$, belong to the space $L_2(X)$ and the vectors $\mathbf{a}^1, ..., \mathbf{a}^n$ are linearly independent. We say that a function $g^0 = \sum_{i=1}^{n} g_i^0(\mathbf{a}^i \cdot \mathbf{x})$ from $\mathcal{R}(\mathbf{a}^1, ..., \mathbf{a}^n)$ is an extremal element or a best approximation to f if

$$\|f - g_0\|_{L_2(X)} = \inf_{g \in \mathcal{R}(\mathbf{a}^1, ..., \mathbf{a}^n)} \|f - g\|_{L_2(X)}$$

Consider the mapping $J : X \to \mathbb{R}^n$ given by the formulas

$$y_i = \mathbf{a}^i \cdot \mathbf{x}, \quad i = 1, ..., n. \tag{2.1}$$

Since the vectors $\mathbf{a}^i = (a_1^i, ..., a_n^i)$, $i = 1, ..., n$, are linearly independent, it is an injection. The Jacobian of this mapping is a constant different from zero:

$$\det\left[\frac{\partial y_i}{\partial x_j}\right] = \det\left[a_j^i\right] \neq 0.$$

Solving the system of linear equations (2.1) with respect to x_i, $i = 1, ..., n$, we obtain that

$$x_i = \mathbf{b}^i \cdot \mathbf{y}, \quad i = 1, ..., n,$$

where $\mathbf{y} = (y_1, ..., y_n)$, $\mathbf{b}^i = (b_1^i, ..., b_n^i)$, $i = 1, ..., n$, and $\left[b_j^i\right] = \left[a_j^i\right]^{-1}$. Introduce the notation

$$Y = J(X)$$

and

$$Y_i = \left\{ y_i \in \mathbb{R} : \quad y_i = \mathbf{a}^i \cdot \mathbf{x}, \quad \mathbf{x} \in X \right\}, \; i = 1, ..., n.$$

For any function $u \in L_2(X)$ put

$$u^* = u^*(\mathbf{y}) \stackrel{def}{=} u\left(\mathbf{b}^1 \cdot \mathbf{y}, ..., \mathbf{b}^n \cdot \mathbf{y}\right).$$

It is obvious that $u^* \in L_2(Y)$. Besides,

$$\int_Y u^*(\mathbf{y})\, d\mathbf{y} = \left|\det\left[a_j^i\right]\right| \cdot \int_X u(\mathbf{x})\, d\mathbf{x} \tag{2.2}$$

and

$$\|u^*\|_{L_2(Y)} = \left|\det\left[a_j^i\right]\right|^{1/2} \cdot \|u\|_{L_2(X)}. \tag{2.3}$$

From (2.3) we obtain that the following lemma is valid.

Lemma 2.1. *Let $f(\mathbf{x}) \in L_2(X)$. A function $\sum_{i=1}^{n} g_i^0(\mathbf{a}^i \cdot \mathbf{x})$ is extremal to the function $f(\mathbf{x})$ if and only if $\sum_{i=1}^{n} g_i^0(y_i)$ is an extremal element from the space $L_2(Y_1) + ... + L_2(Y_n)$ to the function $f^*(\mathbf{y})$.*

Lemma 2.2. *Let $f(\mathbf{x}) \in L_2(X)$. A function $\sum_{i=1}^{n} g_i^0(\mathbf{a}^i \cdot \mathbf{x}) \in \mathcal{R}(\mathbf{a}^1, ..., \mathbf{a}^n)$ is extremal to the function $f(\mathbf{x})$ if and only if*

$$\int_X \left(f(\mathbf{x}) - \sum_{i=1}^{n} g_i^0\left(\mathbf{a}^i \cdot \mathbf{x}\right) \right) h\left(\mathbf{a}^j \cdot \mathbf{x}\right) d\mathbf{x} = 0, \tag{2.4}$$

for any ridge function $h(\mathbf{a}^j \cdot \mathbf{x}) \in L_2(X)$ $j = 1, ..., n$.

Proof. Necessity. Let a function $\sum_{i=1}^{n} g_i^0(\mathbf{a}^i \cdot \mathbf{x})$ be extremal to the function $f(\mathbf{x})$. Suppose that there is a positive integer $j \in [1, n]$ and a ridge function $h(\mathbf{a}^j \cdot \mathbf{x}) \in L_2(X)$ such that

$$\int_X \left(f(\mathbf{x}) - \sum_{i=1}^{n} g_i^0\left(\mathbf{a}^i \cdot \mathbf{x}\right) \right) h\left(\mathbf{a}^j \cdot \mathbf{x}\right) d\mathbf{x} \neq 0. \tag{2.5}$$

Without loss of generality we may assume that $j = 1$ and the integral in (2.5) is negative. Then it follows from (2.2) that

$$\int_Y \left(f^*(\mathbf{y}) - \sum_{i=1}^{n} g_i^0(y_i) \right) h(y_1)\, d\mathbf{y} < 0. \tag{2.6}$$

For sufficiently small positive real number k we obtain from (2.6) that

$$\left\| f^*(\mathbf{y}) - \left(g_1^0(y_1) - kh(y_1)\right) - \sum_{i=2}^{n} g_i^0(y_i) \right\|_{L_2(Y)}^2 = \left\| f^*(\mathbf{y}) - \sum_{i=1}^{n} g_i^0(y_i) \right\|_{L_2(Y)}^2$$

$$+2k\int\limits_{Y}\left(f^{*}(\mathbf{y})-\sum_{i=1}^{n}g_{i}^{0}(y_{i})\right)h(y_{1})\,d\mathbf{y}+k^{2}\left\|h(y_{1})\right\|_{L_{2}(Y)}^{2}$$

$$<\left\|f^{*}(\mathbf{y})-\sum_{i=1}^{n}g_{i}^{0}(y_{i})\right\|_{L_{2}(Y)}^{2}.$$

The last strong inequality means that the function $\sum\limits_{i=1}^{n}g_{i}^{0}(y_{i})$ is not extremal to the function $f^{*}(\mathbf{y})$. Then by lemma 2.1, the function $\sum\limits_{i=1}^{n}g_{i}^{0}(\mathbf{a}^{i}\cdot\mathbf{x})$ is not extremal to $f(\mathbf{x})$. This contradiction shows that our supposition in (2.5) is not true.

Sufficiency. Let (2.4) holds for some element $\sum\limits_{i=1}^{n}g_{i}^{0}(\mathbf{a}^{i}\cdot\mathbf{x})$ from $\mathcal{R}(\mathbf{a}^{1},...,\mathbf{a}^{n})$. Then due to (2.2), we have

$$\int\limits_{Y}\left(f^{*}(\mathbf{y})-\sum_{i=1}^{n}g_{i}^{0}(y_{i})\right)\cdot h_{j}(y_{j})\,d\mathbf{y}=0, \tag{2.7}$$

for any function $h_{j}(y_{j})$ from $L_{2}(Y_{j}),\ j=1,...,n$. From (2.7) we obtain that

$$\left\|f^{*}(\mathbf{y})-\sum_{i=1}^{n}h_{i}(y_{i})\right\|_{L_{2}(Y)}^{2}=\left\|f^{*}(\mathbf{y})-\sum_{i=1}^{n}g_{i}^{0}(y_{i})+\sum_{i=1}^{n}\left(g_{i}^{0}(y_{i})-h_{i}(y_{i})\right)\right\|_{L_{2}(Y)}^{2}$$

$$=\left\|f^{*}(\mathbf{y})-\sum_{i=1}^{n}g_{i}^{0}(y_{i})\right\|_{L_{2}(Y)}^{2}+\left\|\sum_{i=1}^{n}\left(g_{i}^{0}(y_{i})-h_{i}(y_{i})\right)\right\|_{L_{2}(Y)}^{2}$$

$$\geq\left\|f^{*}(\mathbf{y})-\sum_{i=1}^{n}g_{i}^{0}(y_{i})\right\|_{L_{2}(Y)}^{2}.$$

The last inequality means that the function $\sum\limits_{i=1}^{n}g_{i}^{0}(y_{i})$ is extremal to the function $f^{*}(\mathbf{y})$. Then by lemma 2.1, the function $\sum\limits_{i=1}^{n}g_{i}^{0}(\mathbf{a}^{i}\cdot\mathbf{x})$ is extremal to $f(\mathbf{x})$.

Let $Y^{(i)}$ be the Cartesian product of sets $Y_{1},...,Y_{n}$ except for Y_{i}, $i=1,...,n$. That is $Y^{(i)}=Y_{1}\times...\times Y_{i-1}\times Y_{i+1}\times...\times Y_{n}$, $i=1,...,n$.

Theorem 2.3. *Let $f(\mathbf{x})\in L_{2}(X)$ and $Y=Y_{1}\times...\times Y_{n}$. An element $\sum\limits_{i=1}^{n}g_{i}^{0}(\mathbf{a}^{i}\cdot\mathbf{x})$ from $\mathcal{R}(\mathbf{a}^{1},...,\mathbf{a}^{n})$ is extremal to the function $f(\mathbf{x})$ if and only if*

$$g_{j}^{0}(y_{j})=\frac{1}{\left|Y^{(j)}\right|}\int\limits_{Y^{(j)}}\left(f^{*}(\mathbf{y})-\sum_{\substack{i=1\\i\neq j}}^{n}g_{i}^{0}(y_{i})\right)d\mathbf{y}^{(j)},\quad j=1,...,n, \tag{2.8}$$

where $\left|Y^{(j)}\right|$ *is the Lebesgue measure of* $Y^{(j)}$.

Proof. Necessity. Let a function $\sum_{i=1}^{n} g_i^0\left(\mathbf{a}^i \cdot \mathbf{x}\right)$ is extemal to f. Then by lemma 2.1, the function $\sum_{i=1}^{n} g_i^0\left(y_i\right)$ from $L_2\left(Y_1\right)+...+L_2\left(Y_n\right)$ is extremal to f^*. By lemma 2.2 and equality (2.2),

$$\int_Y f^*(\mathbf{y}) h\left(y_j\right) d\mathbf{y}=\int_Y \sum_{i=1}^{n} g_i^0\left(y_i\right) \cdot h\left(y_j\right) d\mathbf{y}, \tag{2.9}$$

for any function $h\left(y_j\right) \in L_2\left(Y_j\right), \quad j=1,...,n$. Applying Fubini's theorem to the integrals in (2.9), we obtain that

$$\int_{Y_j} h\left(y_j\right)\left[\int_{Y^{(j)}} f^*(\mathbf{y}) d\mathbf{y}^{(j)}\right] dy_j=\int_{Y_j} h\left(y_j\right)\left[\int_{Y^{(j)}} \sum_{i=1}^{n} g_i^0\left(y_i\right) d\mathbf{y}^{(j)}\right] dy_j.$$

Since $h\left(y_j\right)$ is an arbitrary function from $L_2\left(Y_j\right)$,

$$\int_{Y^{(j)}} f^*(\mathbf{y}) d\mathbf{y}^{(j)}=\int_{Y^{(j)}} \sum_{i=1}^{n} g_i^0\left(y_i\right) d\mathbf{y}^{(j)}, \quad j=1,...,n.$$

Therefore,

$$\int_{Y^{(j)}} g_j^0\left(y_j\right) d\mathbf{y}^{(j)}=\int_{Y^{(j)}}\left(f^*(\mathbf{y})-\sum_{\substack{i=1 \\ i \neq j}}^{n} g_i^0\left(y_i\right)\right) d\mathbf{y}^{(j)}, \quad j=1,...,n.$$

Now, since $y_j \notin Y^{(j)}$, we obtain (2.8).

Sufficiency. The proof of the sufficiency is not difficult if note that all equalities in the proof of the necessity can be obtained in the reverse order. That is, (2.9) can be obtained from (2.8). Then by (2.2) and lemma 2.2, we finally conclude that the element $\sum_{i=1}^{n} g_i^0\left(\mathbf{a}^i \cdot \mathbf{x}\right)$ is exremal to the function $f(\mathbf{x})$.

For the sake of completeness and further applications (see theorem 2.5), we give the proof for the following theorem which is known from functional analysis.

Theorem 2.4. *The following formula is valid for the error of approximation to a function* $f(\mathbf{x})$ *in* $L_2(X)$ *from* $\mathcal{R}\left(\mathbf{a}^1,...,\mathbf{a}^n\right)$*:*

$$E(f)=\left(\|f(\mathbf{x})\|_{L_2(X)}-\left\|\sum_{i=1}^{n} g_i^0\left(\mathbf{a}^i \cdot \mathbf{x}\right)\right\|_{L_2(X)}\right)^{\frac{1}{2}},$$

where $\sum_{i=1}^{n} g_i^0 (\mathbf{a}^i \cdot \mathbf{x})$ *is an extremal element to* $f(\mathbf{x})$.

Proof

It is obvious that

$$\left\| f - \sum_{i=1}^{n} g_i^0 \right\|_{L_2(X)}^2 = \|f\|_{L_2(X)}^2 -$$

$$-2 \int_X f(\mathbf{x}) \sum_{i=1}^{n} g_i^0 (\mathbf{a}^i \cdot \mathbf{x}) \, d\mathbf{x} + \left\| \sum_{i=1}^{n} g_i^0 \right\|_{L_2(X)}^2. \tag{2.10}$$

On the other hand by lemma 2.2,

$$\left\| f - \sum_{i=1}^{n} g_i^0 \right\|_{L_2(X)}^2 = \|f\|_{L_2(X)}^2 - \int_X f(\mathbf{x}) \sum_{i=1}^{n} g_i^0 (\mathbf{a}^i \cdot \mathbf{x}) \, d\mathbf{x}. \tag{2.11}$$

From (2.10) and (2.11) it follows that

$$\int_X f(\mathbf{x}) \sum_{i=1}^{n} g_i^0 (\mathbf{a}^i \cdot \mathbf{x}) \, d\mathbf{x} = \left\| \sum_{i=1}^{n} g_i^0 \right\|_{L_2(X)}^2. \tag{2.12}$$

Considering (2.12) in (2.11), we finally conclude that

$$E(f)^2 = \|f\|_{L_2(X)}^2 - \left\| \sum_{i=1}^{n} g_i^0 \right\|_{L_2(X)}^2.$$

Using theorem 2.4, it is possible to calculate the approximation error if we a priori can determine an extremal element. The following theorem shows how one can construct an extremal element from integrals of the approximated function.

Theorem 2.5. *Let a function* $f(\mathbf{x})$ *be in* $L_2(X)$ *and* $Y = Y_1 \times ... \times Y_n$. *Set the functions*

$$g_1^0(y_1) = \frac{1}{|Y^{(1)}|} \int_{Y^{(1)}} f^*(\mathbf{y}) \, d\mathbf{y}^{(1)} - (n-1) \frac{1}{|Y|} \int_Y f^*(\mathbf{y}) \, d\mathbf{y}$$

and

$$g_j^0(y_j) = \frac{1}{|Y^{(j)}|} \int_{Y^{(j)}} f^*(\mathbf{y}) \, d\mathbf{y}^{(j)} \quad j = 2, ..., n.$$

Then the function $\sum_{i=1}^{n} g_i^0 (\mathbf{a}^i \cdot \mathbf{x})$ *is extremal to* $f(\mathbf{x})$.

Proof. It is sufficient to verify that the functions $g_j^0(y_j)$, $j = 1, ..., n$, satisfy the conditions (2.8) of theorem 2.3. This becomes obvious if note that

$$\sum_{\substack{i=1 \\ i \neq j}}^{n} \frac{1}{|Y^{(j)}|} \frac{1}{|Y^{(i)}|} \int_{Y^{(j)}} \left[\int_{Y^{(i)}} f^*(\mathbf{y})\, d\mathbf{y}^{(i)} \right] d\mathbf{y}^{(j)} = (n-1) \frac{1}{|Y|} \int_{Y} f^*(\mathbf{y})\, d\mathbf{y}$$

for $j = 1, ..., n$.

Using theorems 2.4 and 2.5, one can obtain an efficient formula for the error in approximating from the set $\mathcal{R}(\mathbf{a}^1, ..., \mathbf{a}^n)$.

References

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

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

[3]. Diaconis P. and Shahshahani M., On nonlinear functions of linear combinations, *SIAM J. Sci. Stat. Comput.* **5** (1984), 175-191.

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

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

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

[7]. Kazantsev I.G., Tomographic reconstruction from arbitrary directions using ridge functions, *Inverse Problems* **14** (1998), 635-645.

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

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

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

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

[12]. Pinkus A., 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.

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

[14]. Wu W., Feng G., Li X., Training multilayer perceptrons via minimization of sum of ridge functions, *Adv., Comput. Math.* **17** (2002), 331-347.

Vugar E. Ismailov
Institute of Mathematics and Mechanics of NAS of Azerbaijan
9, F. Agayev str., AZ1141, Baku, Azerbaijan
Tel.: (99412) 439 47 20 (off.)
Received December 28, 2005; Revised February 27, 2006.
Translated by author.