

ON THE APPROXIMATION BY WEIGHTED RIDGE FUNCTIONS ¹

Vugar E. Ismailov

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

Abstract. We characterize the best L_2 approximation to a multivariate function by linear combinations of ridge functions multiplied by some fixed weight functions. In the special case when the weight functions are constants, we propose explicit formulas for both the best approximation and approximation error.

2000 Mathematics Subject Classification: 41A30, 41A50, 41A63.

1. Introduction

A function $g(\mathbf{a} \cdot \mathbf{x})$, where $\mathbf{a} \in \mathbb{R}^n \setminus \{\mathbf{0}\}$, $\mathbf{x} \in \mathbb{R}^n$, $\mathbf{a} \cdot \mathbf{x}$ is the inner product and g is a univariate function, is called a *ridge function* (in \mathbf{x}) with the direction \mathbf{a} . These functions and their linear combinations appear naturally in computerized tomography, statistics, partial differential equations (where they are called *plane waves*), neural networks, and approximation theory. Ridge approximation in L_2 was actively studied in the late 90's by K.I. Oskolkov [7], V.E. Maiorov [6], A. Pinkus [9], V.N. Temlyakov [10], P. Petrushev [8] and others.

Let D be the unit disk in \mathbb{R}^2 . In [5], Logan and Shepp along with other results gave a closed-form expression for the best L_2 approximation to a function $f(x_1, x_2) \in L_2(D)$ from the set

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

Their solution requires that the directions $\mathbf{a}^1, \dots, \mathbf{a}^r$ be equally-spaced and involves finite sums of convolutions with explicit kernels. In n dimensional case, the author [3] obtained an expression of simpler form for the best L_2 approximation to square-integrable multivariate functions over some domain, provided that $r = n$ and the directions $\mathbf{a}^1, \dots, \mathbf{a}^r$ are linearly independent.

It should be noted that problems of approximation from the set $\mathcal{R}(\mathbf{a}^1, \dots, \mathbf{a}^r)$ were also considered in the uniform norm. For example, one essential approximation method, its defects and advantages were discussed in [9]. Lin and Pinkus [4] characterized $\mathcal{R}(\mathbf{a}^1, \dots, \mathbf{a}^r)$, 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 $\mathbf{a}^1, \dots, \mathbf{a}^r \in \mathbb{R}^n \setminus \{\mathbf{0}\}$, but unknown continuous g_1, \dots, g_r . Two other characterizations of $\mathcal{R}(\mathbf{a}^1, \dots, \mathbf{a}^r)$ may be found in Diaconis and Shahshahani [2]. Buhmann and Pinkus [1] solved

¹Supported by INTAS Grant YSF-06-1000015-6283

the inverse problem: assume that we are given a function $f \in \mathcal{R}(\mathbf{a}^1, \dots, \mathbf{a}^r)$. How can we identify the functions g_i , $i = 1, \dots, r$?

In this paper, we would like to consider the approximation from the more general set

$$\mathcal{R}(\mathbf{a}^1, \dots, \mathbf{a}^r; w_1, \dots, w_r) = \left\{ \sum_{i=1}^r w_i(\mathbf{x}) g_i(\mathbf{a}^i \cdot \mathbf{x}) : g_i : \mathbb{R} \rightarrow \mathbb{R}, i = 1, \dots, r \right\},$$

where w_1, \dots, w_r are fixed multivariate functions. We are going to characterize the best L_2 approximation in this set (see theorem 2.4) for the case $r \leq n$. Then, in the special case when the weight functions w_1, \dots, w_r are constants, we will prove two theorems on explicit formulas for the best approximation and the error of approximation respectively. Unfortunately, we do not yet know any reasonable answer to these problems in other possible cases of r .

2. Characterization of the best approximation

Let X be a subset of \mathbb{R}^n with a finite Lebesgue measure. Consider the approximation of a function $f(\mathbf{x}) = f(x_1, \dots, x_n)$ in $L_2(X)$ from the manifold $\mathcal{R}(\mathbf{a}^1, \dots, \mathbf{a}^r; w_1, \dots, w_r)$, where $r \leq n$. We suppose that the functions $w_i(\mathbf{x})$ and the products $w_i(\mathbf{x}) \cdot g_i(\mathbf{a}^i \cdot \mathbf{x})$, $i = 1, \dots, r$, belong to the space $L_2(X)$. Besides, we assume that the vectors $\mathbf{a}^1, \dots, \mathbf{a}^r$ are linearly independent. We say that a function $g_w^0 = \sum_{i=1}^r w_i(\mathbf{x}) g_i^0(\mathbf{a}^i \cdot \mathbf{x})$ in $\mathcal{R}(\mathbf{a}^1, \dots, \mathbf{a}^r; w_1, \dots, w_r)$ is the best approximation (or extremal) to f if

$$\|f - g_w^0\|_{L_2(X)} = \inf_{g \in \mathcal{R}(\mathbf{a}^1, \dots, \mathbf{a}^r; w_1, \dots, w_r)} \|f - g\|_{L_2(X)}.$$

Let the system of vectors $\{\mathbf{a}^1, \dots, \mathbf{a}^r, \mathbf{a}^{r+1}, \dots, \mathbf{a}^n\}$ be a completion of the system $\{\mathbf{a}^1, \dots, \mathbf{a}^r\}$ to a basis in \mathbb{R}^n . Let $J : X \rightarrow \mathbb{R}^n$ be the linear transformation given by the formulas

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

Since the vectors \mathbf{a}^i , $i = 1, \dots, n$, are linearly independent, it is an injection. The Jacobian $\det J$ of this transformation is a constant different from zero.

Let the formulas

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

stand for the solution of linear equations (2.1) with respect to x_i , $i = 1, \dots, n$.

Introduce the notation

$$Y = J(X)$$

and

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

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

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

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

$$\int_Y u^*(\mathbf{y}) d\mathbf{y} = |\det J| \cdot \int_X u(\mathbf{x}) d\mathbf{x} \quad (2.2)$$

and

$$\|u^*\|_{L_2(Y)} = |\det J|^{1/2} \cdot \|u\|_{L_2(X)}. \quad (2.3)$$

Set

$$L_2^i = \{w_i^*(\mathbf{y})g(y_i) \in L_2(Y)\}, \quad i = 1, \dots, r.$$

We need the following auxiliary lemmas.

Lemma 2.1. *Let $f(\mathbf{x}) \in L_2(X)$. A function $\sum_{i=1}^r w_i(\mathbf{x})g_i^0(\mathbf{a}^i \cdot \mathbf{x})$ is extremal to the function $f(\mathbf{x})$ if and only if $\sum_{i=1}^r w_i^*(\mathbf{y})g_i^0(y_i)$ is extremal from the space $L_2^1 \oplus \dots \oplus L_2^r$ to the function $f^*(\mathbf{y})$.*

Due to (2.3) the proof of this lemma is obvious.

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

$$\int_X \left(f(\mathbf{x}) - \sum_{i=1}^r w_i(\mathbf{x})g_i^0(\mathbf{a}^i \cdot \mathbf{x}) \right) w_j(\mathbf{x})h(\mathbf{a}^j \cdot \mathbf{x}) d\mathbf{x} = 0$$

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

Lemma 2.3. *The following formula is valid for the error of approximation to a function $f(\mathbf{x})$ in $L_2(X)$ from $\mathcal{R}(\mathbf{a}^1, \dots, \mathbf{a}^r; w_1, \dots, w_r)$:*

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

where $\sum_{i=1}^r w_i(\mathbf{x})g_i^0(\mathbf{a}^i \cdot \mathbf{x})$ is the best approximation to $f(\mathbf{x})$.

Lemmas 2.2 and 2.3 follow from the well-known facts of functional analysis that the best approximation of an element x in a Hilbert space H from a linear subspace Z of H must be the image of x via the orthogonal projection onto Z and the sum of squares of norms of orthogonal vectors is equal to the square of the norm of their sum.

We say that Y is an r -set if it can be represented as $Y_1 \times \dots \times Y_r \times Y_0$, where Y_0 is some set from the space \mathbb{R}^{n-r} . In special case, Y_0 may be equal to $Y_{r+1} \times \dots \times Y_n$, but it is not necessary.

By $Y^{(i)}$, we denote the Cartesian product of the sets Y_1, \dots, Y_r, Y_0 except for Y_i , $i = 1, \dots, r$. That is, $Y^{(i)} = Y_1 \times \dots \times Y_{i-1} \times Y_{i+1} \times \dots \times Y_r \times Y_0$, $i = 1, \dots, r$.

Theorem 2.4. *Let Y be an r -set. A function $\sum_{i=1}^r w_i(\mathbf{x})g_i^0(\mathbf{a}^i \cdot \mathbf{x})$ is the best approximation to $f(\mathbf{x})$ if and only if*

$$g_j^0(y_j) = \frac{1}{\int_{Y^{(j)}} w_j^{*2}(\mathbf{y}) d\mathbf{y}^{(j)}} \int_{Y^{(j)}} \left(f^*(\mathbf{y}) - \sum_{\substack{i=1 \\ i \neq j}}^r w_i^*(\mathbf{y})g_i^0(y_i) \right) w_j^*(\mathbf{y}) d\mathbf{y}^{(j)}, \quad j = 1, \dots, r. \quad (2.4)$$

Proof. Necessity. Let a function $\sum_{i=1}^r w_i(\mathbf{x})g_i^0(\mathbf{a}^i \cdot \mathbf{x})$ be extremal to f . Then by lemma 2.1, the function $\sum_{i=1}^r w_i^*(\mathbf{y})g_i^0(y_i)$ in $L_2^1 \oplus \dots \oplus L_2^r$ is extremal to f^* . By lemma 2.2 and equality (2.2),

$$\int_Y f^*(\mathbf{y}) w_j^*(\mathbf{y}) h(y_j) d\mathbf{y} = \int_Y w_j^*(\mathbf{y}) h(y_j) \sum_{i=1}^r w_i^*(\mathbf{y}) g_i^0(y_i) d\mathbf{y} \quad (2.5)$$

for any product $w_j^*(\mathbf{y})h(y_j)$ in L_2^j , $j = 1, \dots, r$. Applying Fubini's theorem to the integrals in (2.5), we obtain that

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

Since $h(y_j)$ is an arbitrary function such that $w_j^*(\mathbf{y})h(y_j) \in L_2^j$,

$$\int_{Y^{(j)}} f^*(\mathbf{y}) w_j^*(\mathbf{y}) d\mathbf{y}^{(j)} = \int_{Y^{(j)}} w_j^*(\mathbf{y}) \sum_{i=1}^r w_i^*(\mathbf{y}) g_i^0(y_i) d\mathbf{y}^{(j)}, \quad j = 1, \dots, r.$$

Therefore,

$$\int_{Y^{(j)}} w_j^{*2}(\mathbf{y}) g_j^0(y_j) d\mathbf{y}^{(j)} = \int_{Y^{(j)}} \left(f^*(\mathbf{y}) - \sum_{\substack{i=1 \\ i \neq j}}^r w_i^*(\mathbf{y}) g_i^0(y_i) \right) w_j^*(\mathbf{y}) d\mathbf{y}^{(j)}, \quad j = 1, \dots, r.$$

Now, since $y_j \notin Y^{(j)}$, we obtain (2.4).

Sufficiency. Note that all the equalities in the proof of the necessity can be obtained in the reverse order. Thus, (2.5) can be obtained from (2.4). Then by (2.2) and lemma 2.2, we finally conclude that the function $\sum_{i=1}^r w_i(\mathbf{x})g_i^0(\mathbf{a}^i \cdot \mathbf{x})$ is extremal to $f(\mathbf{x})$. \square

In the following, $|Q|$ will denote the Lebesgue measure of a measurable set Q . The following corollary is obvious.

Corollary 2.5. *Let Y be an r -set. A function $\sum_{i=1}^r g_i^0(\mathbf{a}^i \cdot \mathbf{x})$ in $\mathcal{R}(\mathbf{a}^1, \dots, \mathbf{a}^r)$ is the best approximation to $f(\mathbf{x})$ if and only if*

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

In [3], this corollary was proven for the case $r = n$.

3. Calculation of the approximation error

In this Section, we are going to establish explicit formulas for both the best approximation and approximation error, provided that the weight functions are constants. In this case, since we vary over g_i , the set $\mathcal{R}(\mathbf{a}^1, \dots, \mathbf{a}^r; w_1, \dots, w_r)$ coincide with $\mathcal{R}(\mathbf{a}^1, \dots, \mathbf{a}^r)$. Thus, without loss of generality, we may assume that $w_i(\mathbf{x}) = 1$ for $i = 1, \dots, r$.

For brevity of the further exposition, introduce the notation

$$A = \int_Y f^*(\mathbf{y}) d\mathbf{y} \quad \text{and} \quad f_i^* = f_i^*(y_i) = \int_{Y^{(i)}} f^*(\mathbf{y}) d\mathbf{y}^{(i)}, \quad i = 1, \dots, r.$$

The following theorem is a generalization of the main result of [3] from the case $r = n$ to the cases $r < n$.

Theorem 3.1. *Let Y be an r -set. Set the functions*

$$g_1^0(y_1) = \frac{1}{|Y^{(1)}|} f_1^* - (r-1) \frac{A}{|Y|}$$

and

$$g_j^0(y_j) = \frac{1}{|Y^{(j)}|} f_j^*, \quad j = 2, \dots, r.$$

Then the function $\sum_{i=1}^r g_i^0(\mathbf{a}^i \cdot \mathbf{x})$ is the best approximation from $\mathcal{R}(\mathbf{a}^1, \dots, \mathbf{a}^r)$ to $f(\mathbf{x})$.

The proof is the same as in [3]. It is sufficient to verify that the functions $g_j^0(y_j)$, $j = 1, \dots, r$, satisfy the conditions of corollary 2.5. This becomes obvious if note that

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

for $j = 1, \dots, r$.

Theorem 3.2. *Let Y be an r -set. Then the error of approximation to a function $f(x)$ from the set $\mathcal{R}(\mathbf{a}^1, \dots, \mathbf{a}^r)$ can be calculated by the formula*

$$E(f) = |\det J|^{-1/2} \left(\|f^*\|_{L_2(Y)}^2 - \sum_{i=1}^r \frac{1}{|Y^{(i)}|^2} \|f_i^*\|_{L_2(Y)}^2 + (r-1) \frac{A^2}{|Y|} \right)^{1/2}.$$

Proof. From Eq. (2.3), lemma 2.3 and theorem 3.1, it follows that

$$E(f) = |\det J|^{-1/2} \left(\|f^*\|_{L_2(Y)}^2 - I \right)^{1/2}, \quad (3.1)$$

where

$$I = \left\| \sum_{i=1}^r \frac{1}{|Y^{(i)}|} f_i^* - (r-1) \frac{A}{|Y|} \right\|_{L_2(Y)}^2.$$

The integral I can be written as a sum of the following four integrals:

$$\begin{aligned} I_1 &= \sum_{i=1}^r \frac{1}{|Y^{(i)}|^2} \|f_i^*\|_{L_2(Y)}^2, \quad I_2 = \sum_{i=1}^r \sum_{\substack{j=1 \\ j \neq i}}^r \frac{1}{|Y^{(i)}|} \frac{1}{|Y^{(j)}|} \int_Y f_i^* f_j^* d\mathbf{y}, \\ I_3 &= -2(r-1) \frac{1}{|Y|} A \sum_{i=1}^r \frac{1}{|Y^{(i)}|} \int_Y f_i^* d\mathbf{y}, \quad I_4 = (r-1)^2 \frac{A^2}{|Y|}. \end{aligned}$$

It is not difficult to verify that

$$\int_Y f_i^* f_j^* d\mathbf{y} = \left| Y_0 \times \prod_{\substack{k=1 \\ k \neq i, j}}^r Y_k \right| A^2, \text{ for } i, j = 1, \dots, r, \quad i \neq j, \quad (3.2)$$

and

$$\int_Y f_i^* d\mathbf{y} = \left| Y_0 \times \prod_{\substack{k=1 \\ k \neq i}}^r Y_k \right| A, \text{ for } i = 1, \dots, r. \quad (3.3)$$

Considering (3.2) and (3.3) in the expressions of I_2 and I_3 respectively, we obtain that

$$I_2 = r(r-1) \frac{A^2}{|Y|} \text{ and } I_3 = -2r(r-1) \frac{A^2}{|Y|}.$$

Therefore,

$$I = I_1 + I_2 + I_3 + I_4 = \sum_{i=1}^r \frac{1}{|Y^{(i)}|^2} \|f_i^*\|_{L_2(Y)}^2 - (r-1) \frac{A^2}{|Y|}.$$

Now the last equality with (3.1) complete the proof. \square

Example. Consider the following set

$$X = \{\mathbf{x} \in \mathbb{R}^4 : y_i = y_i(\mathbf{x}) \in [0; 1], \ i = 1, \dots, 4\},$$

where

$$\begin{cases} y_1 = x_1 + x_2 + x_3 - x_4 \\ y_2 = x_1 + x_2 - x_3 + x_4 \\ y_3 = x_1 - x_2 + x_3 + x_4 \\ y_4 = -x_1 + x_2 + x_3 + x_4 \end{cases} \quad (3.4)$$

Let the function

$$f = 8x_1x_2x_3x_4 - \sum_{i=1}^4 x_i^4 + 2 \sum_{i=1}^3 \sum_{j=i+1}^4 x_i^2 x_j^2$$

be given over X . Consider the approximation of this function from the set $\mathcal{R}(\mathbf{a}^1, \mathbf{a}^2, \mathbf{a}^3)$, where $\mathbf{a}^1 = (1; 1; 1; -1)$, $\mathbf{a}^2 = (1; 1; -1; 1)$, $\mathbf{a}^3 = (1; -1; 1; 1)$. Putting $\mathbf{a}^4 = (-1; 1; 1; 1)$, we complete the system of vectors $\mathbf{a}^1, \mathbf{a}^2, \mathbf{a}^3$ to the basis $\{\mathbf{a}^1, \mathbf{a}^2, \mathbf{a}^3, \mathbf{a}^4\}$ in \mathbb{R}^4 . The linear transformation J defined by (3.4) maps the set X onto the set $Y = [0; 1]^4$. The inverse transformation is given by the formulas

$$\begin{cases} x_1 = \frac{1}{4}y_1 + \frac{1}{4}y_2 + \frac{1}{4}y_3 - \frac{1}{4}y_4 \\ x_2 = \frac{1}{4}y_1 + \frac{1}{4}y_2 - \frac{1}{4}y_3 + \frac{1}{4}y_4 \\ x_3 = \frac{1}{4}y_1 - \frac{1}{4}y_2 + \frac{1}{4}y_3 + \frac{1}{4}y_4 \\ x_4 = -\frac{1}{4}y_1 + \frac{1}{4}y_2 + \frac{1}{4}y_3 + \frac{1}{4}y_4 \end{cases}$$

It can be easily verified that $f^* = y_1y_2y_3y_4$ and Y is a 3-set with $Y_i = [0; 1]$, $i = 1, 2, 3$. Besides, $Y_0 = [0; 1]$. After easy calculations we obtain that $A = \frac{1}{16}$; $f_i^* = \frac{1}{8}y_i$ for $i = 1, 2, 3$; $\det J = -16$; $\|f^*\|_{L_2(Y)}^2 = \frac{1}{81}$; $\|f_i^*\|_{L_2(Y)}^2 = \frac{1}{192}$, $i = 1, 2, 3$. Now from theorems 3.1 and 3.2 it follows that the function $\frac{1}{8} \sum_{i=1}^3 (\mathbf{a}^i \cdot \mathbf{x}) - \frac{1}{8}$ is a best approximant from $\mathcal{R}(\mathbf{a}^1, \mathbf{a}^2, \mathbf{a}^3)$ to f and $E(f) = \frac{1}{576} \sqrt{2} \sqrt{47}$.

References

- [1] M.D.Buhmann and A.Pinkus, Identifying linear combinations of ridge functions, *Advances in Applied Math.* **22** (1999), 103-118.
- [2] P.Diaconis and M.Shahshahani, On nonlinear functions of linear combinations, *SIAM J. Sci. Stat. Comput.* **5** (1984), 175-191.
- [3] V.E.Ismailov, A note on the best L_2 approximation by ridge functions, *Appl. Math. E-Notes* **7** (2007), 71-76.

- [4] V.Ya Lin and A.Pinkus, Fundamentality of ridge functions, *J.Approx. Theory* **75** (1993), 295-311.
- [5] B.F.Logan and L.A.Shepp, Optimal reconstruction of a function from its projections, *Duke Math.J.* **42** (1975), 645-659.
- [6] V.E.Maiorov, On best approximation by ridge functions, *J.Approx. Theory* **99** (1999), 68-94.
- [7] K.I. Oskolkov, Ridge approximation, Chebyshev-Fourier analysis and optimal quadrature formulas, *Proc. Steklov Inst. Math.*, **219** (1997), 265-280.
- [8] P.P.Petrushev, Approximation by ridge functions and neural networks, *SIAM J.Math. Anal.* **30** (1998), 155-189.
- [9] 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.
- [10] V.N.Temlyakov, On approximation by ridge functions, Preprint. Department of Mathematics, University of South Carolina, 1996.