Applied Mathematics E-Notes, 7(2007), 71-76 © ISSN 1607-2510
Available free at mirror sites of http://www.math.nthu.edu.tw/∼amen/

A Note On The Best L_2 Approximation By Ridge Functions*

Vugar E. Ismailov†

Received 2 February 9 2006

Abstract

We propose an explicit formula for the best L_2 approximation to a multivariate function by linear combinations of ridge functions over some set in $\mathbb{R}^n$.

1 Introduction

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

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

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, 12]), statistics (see, e.g., [1, 5]), partial differential equations [6], neural networks (see, e.g., [2, 13, 15, 16]), and approximation theory (see, e.g., [2, 3, 4, 8, 9, 11, 13, 14, 15]).

The term "ridge function" is rather recent. It was coined by Logan and Shepp [10] in one of the seminal papers on computerized tomography. However, these functions have been considered for a long time under the name of *plane waves* (see, for example, [6]).

In some applications (especially in tomography), one is interested in the set

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

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.

Some problems of approximation from the set $\mathcal{R}\left(\mathbf{a}^1, ..., \mathbf{a}^r\right)$ were investigated by a number of authors. For example, one essential approximation method, its defects and advantages were discussed in [14]. Lin and Pinkus [9] characterized $\mathcal{R}\left(\mathbf{a}^1, ..., \mathbf{a}^r\right)$,

*Mathematics Subject Classifications: 41A30, 41A50, 41A63.

†Mathematics and Mechanics Institute, Azerbaijan National Academy of Sciences, Az-1141, Baku, Azerbaijan

i.e. they found means of determining if a continuous function f (defined on $\mathbb{R}^n$) is of the form $\sum_{i=1}^{r} g_i\left(\mathbf{a}^i \cdot \mathbf{x}\right)$ 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 $\mathcal{R}\left(\mathbf{a}^1, ..., \mathbf{a}^r\right)$ in the uniform norm may be found in Diaconis and Shahshahani [3].

Let D be the unit disk in $\mathbb{R}^2$. Logan and Shepp [10] gave a complicated, but explicitly determined expression for the best L_2 approximation to a function $f\left(x_1, x_2\right) \in L_2\left(D\right)$ from $\mathcal{R}\left(\mathbf{a}^1, ..., \mathbf{a}^r\right)$ with equally-spaced directions $\mathbf{a}^1, ..., \mathbf{a}^{\mathbf{r}}$. We are not aware of any result of this kind in n dimensional case. Our purpose is to consider the problem of the best L_2 approximation to a multivariate function $f\left(x_1, ..., x_n\right)$ from $\mathcal{R}\left(\mathbf{a}^1, ..., \mathbf{a}^r\right)$ over some set in $\mathbb{R}^n$. We use only the basic facts from the theory of Hilbert spaces to characterize and then construct the best approximation. Unfortunately, the results of the paper involve only cases in which the space dimension is equal to the number of fixed directions. However, our formula for the best approximation is rather simple.

2 Construction of the Best Approximation

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

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

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

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

Since the vectors $\mathbf{a}^i = \left(a_1^i, ..., a_n^i\right), i = 1, ..., n$, 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 (1) with respect to $x_i, \ \ i = 1, ..., n$, we obtain that

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

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

$$Y = J\left(X\right)$$

and

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

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

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

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

$$\int_Y u^*(\mathbf{y})\, d\mathbf{y} = \left|\det\left[a_j^i\right]\right| \cdot \int_X u(\mathbf{x})\, d\mathbf{x} \tag{2}$$

and

$$\|u^*\|_{L_2(Y)} = \left|\det\left[a_j^i\right]\right|^{1/2} \cdot \|u\|_{L_2(X)}. \tag{3}$$

From (3) we obtain that the following lemma is valid.

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

The following two lemmas are observations well known from 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 (lemma 2.2) and the sum of squares of norms of orthogonal vectors is equal to the square of the norm of their sum (lemma 2.3).

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

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

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

LEMMA 2.3. The following formula is valid for the error of approximation to a function $f(\mathbf{x})$ in $L_2(X)$ from $\mathcal{R}\left(\mathbf{a}^1, ..., \mathbf{a}^n\right)$:

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

where $\sum\limits_{i=1}^n g_i^0\left(\mathbf{a}^i \cdot \mathbf{x}\right)$ is the best approximation to $f(\mathbf{x})$.

By $Y^{(i)}$, we denote the Cartesian product of the 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 2.4. Let Y be represented as the Cartesian product $Y_1 \times ... \times Y_n$. A function $\sum_{i=1}^n g_i^0\left(\mathbf{a}^i \cdot \mathbf{x}\right)$ in $\mathcal{R}\left(\mathbf{a}^1, ..., \mathbf{a}^n\right)$ is the best approximation to $f(\mathbf{x})$ if and only if

$$g_j^0(y_j) = \frac{1}{\left|Y^{(j)}\right|} \int_{Y^{(j)}} \left(f^*(\mathbf{y}) - \sum_{i=1,...,n;\, i \neq j} g_i^0(y_i)\right) d\mathbf{y}^{(j)}, \quad j = 1, ..., n, \tag{4}$$

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 extremal to f. Then by lemma 2.1, the function $\sum\limits_{i=1}^{n} g_i^0\left(y_i\right)$ in $L_2\left(Y_1\right)+...+L_2\left(Y_n\right)$ is extremal to f^*. By lemma 2.2 and equality (2),

$$\int\limits_{Y} f^*\left(\mathbf{y}\right) h\left(y_j\right) d\mathbf{y} = \int\limits_{Y} \sum_{i=1}^{n} g_i^0\left(y_i\right) \cdot h\left(y_j\right) d\mathbf{y} \tag{5}$$

for any function $h\left(y_j\right) \in L_2\left(Y_j\right),\ j=1,...,n$. Applying Fubini's theorem to the integrals in (5), we obtain that

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

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

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

Therefore,

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

Now, since $y_j \notin Y^{(j)}$, we obtain (4).

Sufficiency. The proof of the sufficiency is not difficult if note that all the equalities in the proof of the necessity can be obtained in the reverse order. That is, (5) can be obtained from (4). Then by (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 extremal to $f\left(\mathbf{x}\right)$.

Our main result is the following theorem.

THEOREM 2.5. Let Y be represented as the Cartesian product $Y_1 \times ... \times Y_n$. Set the functions

$$g_1^0\left(y_1\right) = \frac{1}{\left|Y^{(1)}\right|}\int\limits_{Y^{(1)}} f^*\left(\mathbf{y}\right) d\mathbf{y}^{(1)} - (n-1)\frac{1}{|Y|}\int\limits_{Y} f^*\left(\mathbf{y}\right) d\mathbf{y}$$

and

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

Then the function $\sum\limits_{i=1}^{n} g_i^0\left(\mathbf{a}^i \cdot \mathbf{x}\right)$ is the best approximation to $f\left(\mathbf{x}\right)$.

PROOF. It is sufficient to verify that the functions $g_j^0(y_j)$, $j = 1, ..., n$, satisfy the conditions (4) of Theorem 2.4. This becomes obvious if note that

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

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

REMARK. Using lemma 2.3 and theorem 2.5, one can obtain an efficient formula for the error in approximating from the set $\mathcal{R}\left(\mathbf{a}^1, ..., \mathbf{a}^n\right)$.

Acknowledgment. This research is supported by INTAS, through grant YSF-06-1000015-6283.

References

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

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

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

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

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

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

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

[8] A. Kroo, On approximation by ridge functions, Constr. Approx., 13(1997), 447–460.

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

[10] B. F. Logan and L. A. Shepp, Optimal reconstruction of a function from its projections, Duke Math. J., 42(1975), 645–659.

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

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

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

[14] 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.

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

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