Ukrainian Mathematical Journal, Vol. 62, No. 3, 2010

BEST APPROXIMATION BY RIDGE FUNCTIONS IN L_p-SPACES

V. E. Maiorov

UDC 517.5

We study the approximation of the classes of functions by the manifold R_n formed by all possible linear combinations of n ridge functions of the form $r(a \cdot x)$). It is proved that, for any $1 \le q \le p \le \infty$, the deviation of the Sobolev class W_p^r from the set R_n of ridge functions in the space $L_q(B^d)$ satisfies the sharp order $n^{-r/(d-1)}$.

1. Introduction

In the present work, we continue the investigation of approximation of multivariate functions by classes formed by linear combinations of ridge functions started in [9, 10, 11, 5]. Ridge functions are defined as functions on $\mathbb{R}^d$ of the form $r(a \cdot x)$ with parameters $a \in \mathbb{R}^d$ and $r\colon \mathbb{R} \to \mathbb{R}$, and $a \cdot x$ is the ordinary inner product. Let $L_q = L_p(B^d)$, $1 \le q \le \infty$, be a Banach space of all q-integrable functions on the unit ball $B^d = \{|x| \le 1\}$, where $|x|^2 = x_1^2 + \ldots + x_d^2$, with the norm

$$\|f\|_q = \left(\int\limits_{B^d} |f(x)|^q \, dx \right)^{1/q}.$$

Let A be a set on the unit sphere $\mathbb{S}^{d-1} = \{|x| = 1\}$ in $\mathbb{R}^d$. We introduce the set of ridge functions

$$R(A) = \Big\{ r_a := r(a \cdot x) \colon\ r \in L_{2,\mathrm{loc}}(\mathbb{R}),\ a \in A \Big\},$$

where r runs over the class $L_{2,\mathrm{loc}}(\mathbb{R})$ of square integrable functions on all compact subsets of $\mathbb{R}$ and a runs over A. Denote $R = R(\mathbb{S}^{d-1})$. Let n be a natural number. Consider a class of functions

$$R_n = R + \ldots + R,$$

formed by all possible linear combinations of n functions from the set R.

Approximation by ridge functions has been studied by several authors. In [26] and [6], necessary and sufficient conditions are established for the closure of the set $R(A)$ on a set A to coincide with the space of continuous functions. In addition, Lin and Pinkus [7] proved that, for any fixed n, the set R_n is not dense in the spaces of continuous functions on compact sets. The approximation properties of ridge manifolds were studied by Barron [1], DeVore, Oskolkov, and Petrushev [2], Maiorov [9], Maiorov and Meir [12]. Makovoz [14], Mhaskar and Micchelli [16], Mhaskar [15], Oskolkov [17], Petrushev [18], Pinkus [19], and Temlyakov [21]. In [5], Gordon,

Technion, Haifa, Israel.

Published in Ukrains'kyi Matematychnyi Zhurnal, Vol. 62, No. 3, pp. 396–408, March, 2010. Original article submitted May 26, 2009.

 0041–5995/10/6203–0452

Maiorov, Meyer, and Reisner considered the results about the best approximation by ridge functions in the Banach space L_p.

In the present work, we study the problems of the best approximation of multivariate functions from the Sobolev classes W_p^r by the class R_n in the space L_q, where the parameters p and q satisfy the inequalities $1 \le q \le p \le \infty$. Earlier, the asymptotic estimates of approximation by R_n were studied in [5, 9] but only for $2 \le q \le p \le \infty$.

Let $\rho = (\rho_1, \dots, \rho_d)$ be a multiindex vector, i.e., ρ is a vector with nonnegative integer coordinates and $|\rho| = \rho_1 + \dots + \rho_d$. We introduce a differential operator $\mathcal{D}^\rho = \partial^{|\rho|}/\partial^{\rho_1} x_1 \dots \partial^{\rho_d} x_d$.

Let r be any natural number. In the space $L_p = L_p(B^d)$, we consider the Sobolev class of functions [23]

$$W_p^r := \left\{ f\colon \|f\|_{W_p^r} := \|f\|_p + \sum_{|\rho|=r} \|\mathcal{D}^\rho f\|_p \le 1 \right\}.$$

For subsets $W, V \subset L_q$, we define the deviation of W from V as follows:

$$e(W, V)_q = \sup_{f \in W} e(f, V)_q,$$

where $e(f, V)_q = \inf_{v \in V} \|f - v\|_q$.

Theorem 1. *Let $d \ge 2$, $r > 0$, and $1 \le p \le q \le \infty$ be arbitrary numbers. Then the following asymptotic inequality holds for the deviation of the Sobolev class W_p^r from the class R_n :*

$$e(W_p^r, R_n)_q \asymp n^{-r/(d-1)}.$$

We now present a brief description of the proof of Theorem 1. In order to obtain the lower bound in Theorem 1, we construct, for any n, a function $f \in W_p^r$ depending on n such that the distance of f from the class R_n is greater than $cn^{-r/(d-1)}$. The construction of the function f is realized in the following way: In Section 2, we introduce an orthonormal system $\{P_k(x)\}_{k=1}^\infty$ of polynomials on the ball B^d and study the Fourier coefficients $\langle r_a, P_k \rangle$ of the ridge function $r(a \cdot x)$ with respect to the system $\{P_k(x)\}$. In particular, we show that the coefficients allow the separation of variables r and a (see [13, 9, 10]), namely, the identity $\langle r_a, P_k \rangle = u(a)v(r)$, where u is a function of a and v is a linear functional of r, holds for any k. In Section 3, we estimate the Vapnik–Chervonenkis dimension of the projection $\mathrm{Pr}_s R_n$ of the class of ridge functions R_n onto the polynomial space $\mathcal{P}_s^d$. In Section 4, we prove Theorem 1 by using the results of Sections 2 and 3. In the Appendix, we present the well-known results from the theory of orthogonal polynomials on the segment and from the theory of harmonic analysis on the sphere used in the proof of Theorem 1.

In what follows, by c, c', c_0, $c_1, \dots$, etc. we denote positive constants independent of the parameter n which may depend only on r, d, p, or q. For two positive sequences $\{a_n\}$ and $\{b_n\}$, $n = 0, 1, \dots$, we write $a_n \asymp b_n$ if there exist positive constants c_1 and c_2 such that $c_1 \le a_n/b_n \le c_2$ for all $n = 0, 1, \dots$.

2. Projection of R_n onto the Polynomial Space

In the present section, we construct special orthonormal systems of polynomials on the unit ball B^d. The orthogonal systems of polynomials on the ball play an important role in the problems of approximation of multivariable functions by the manifolds of linear combinations of ridge functions (plane waves) (see [2, 18, 9, 10]).

In [2, 18], these methods were developed for the construction of orthogonal projections on polynomial subspaces and approximation by ridge functions. The system of Gegenbauer orthogonal polynomials is the main tool used for the construction of orthogonal systems of polynomials on the ball [8]. Note that, in a special case $d = 2$, the Gegenbauer polynomials coincide with the Chebyshev polynomials of second kind.

In the present work, the system of orthogonal polynomials on the unit ball is obtained, in a certain sense, by the convolution of two orthogonal systems. These are the system of Gegenbauer polynomials on the segment $[-1, 1]$ and the system of spherical harmonics on the unit sphere $\mathbb{S}^{d-1}$. We now describe this construction.

Let $L_2(\mathbb{S}^{d-1})$ be the Hilbert space formed by all complex-valued square-integrable functions $h(\xi)$ on the sphere $\mathbb{S}^{d-1}$ with inner product

$$(h_1, h_2) = \int\limits_{\mathbb{S}^{d-1}} h_1(\xi)\bar{h}_2(\xi)d\xi, \qquad h_1,\ h_2 \in L_2(S^{d-1}),$$

where $d\xi$ denotes the normalized Lebesgue measure on the sphere S^{d-1}.

In the space $L_2(\mathbb{S}^{d-1})$, we consider (see the Appendix) a subspace $\mathcal{H}$ of the restrictions of harmonic functions defined on $\mathbb{R}^d$ to $\mathbb{S}^{d-1}$. Let $\mathcal{H}_s$ be the subspace of $\mathcal{H}$ generated by all spherical harmonics of degree at most s, i.e., all harmonic polynomials of degree at most s. Let $\mathcal{H}_s^{\text{hom}}$ be the subspace of $\mathcal{H}_s$ formed by all homogeneous spherical harmonics of degree s. The functions $\{h_{sk}\}_{k\in K^s}$ (see Appendix) generate a basis in the space $\mathcal{H}_s^{\text{hom}}$.

The space $\mathcal{H}_s = \mathcal{H}_0^{\text{hom}} \oplus \mathcal{H}_1^{\text{hom}} \oplus \ldots \oplus \mathcal{H}_s^{\text{hom}}$ is the direct sum of orthogonal subspaces of the spherical harmonics of degrees $0, 1, \ldots, s$. Denote by N_s the dimension of the space $\mathcal{H}_s$. We have $N_s \asymp s^{d-1}$. Indeed, by using the relation $\dim\ \mathcal{H}_s^{\text{hom}} \asymp s^{d-2}$ (see (A.2)) we obtain

$$N_s = \dim\ \mathcal{H}_s = \dim\ \mathcal{H}_0^{\text{hom}} + \dim\ \mathcal{H}_1^{\text{hom}} + \ldots + \dim\ \mathcal{H}_s^{\text{hom}} \asymp s^{d-1}.$$

In the space $\mathcal{H}$, we introduce a family of functions $\mathcal{B}(\mathbb{S}^{d-1}) := \{h_i\}_{i=0}^{\infty}$ formed (see (A.2)) by all ordered spherical harmonics, i.e., the functions

$$\bigcup_{s=0}^{\infty} \{h_{s,k}\}_{k\in K^s},$$

where K^s is defined in the Appendix. The set $\mathcal{B}(\mathbb{S}^{d-1})$ is an orthonormal basis in the space $\mathcal{H}$, i.e., for the subscripts $i \neq i'$, we have $(h_i, h_{i'}) = \delta_{ii'}$, where $\delta_{ii'} = 0$ for $i \neq i'$, and $\delta_{ii} = 1$.

We now consider (see the Appendix) the Gegenbauer polynomials $C_n^{d/2}(t)$, $t \in \mathbb{R}$, of degree n associated with $d/2$. Every polynomial $C_n^{d/2}$ is normed by a factor, i.e., we set

$$u_n(t) = v_n^{-1/2} C_n^{d/2}(t), \qquad v_n = \frac{\pi^{1/2}(d)_n \Gamma\left((d+1)/2\right)}{(n + d/2)n!\Gamma(d/2)},$$

where $(a)_0 = 1$, and $(a)_n = a(a+1)\ldots(a+n-1)$.

Let i and j be any two indices from $\mathbb{Z}_+$. In $\mathbb{R}^d$, we construct a function

$$P_{ij}(x) = \nu_j \int\limits_{S^{d-1}} h_i(\xi) u_j(x\cdot\xi)d\xi, \qquad \nu_j = \left(\frac{(j+1)_{d-1}}{2(2\pi)^{d-1}}\right)^{1/2}. \tag{1}$$

It follows from (1) that, for any $i, j \in \mathbb{Z}_+$ the function P_{ij} is a polynomial on $\mathbb{R}^d$ of degree j. Note that if the indices i and j are such that the degrees of the polynomials h_i and u_j satisfy the inequality $\deg h_i > \deg u_j = j$, then $P_{ij}(x) \equiv 0$ (see (A.8)). Assume that the set I consists of the couples of nonnegative integers (i, j) such that $\deg h_i \le j$ and the numbers $\deg h_i$ and j have the same parity. Let I_s be a subset of I formed by the couples (i, j) with $\deg h_i \le j \le s$. We construct a system of polynomials

$$\Pi := \Pi(B^d) := \{P_{ij}\}_{(i,j)\in I} \tag{2}$$

and consider a finite subsystem $\Pi_s = \{P_{ij}\}_{(i,j)\in I_s}$ of the system Π formed by the polynomials of degree at most s.

Lemma 1 (see [10]).

(a) The set of polynomials Π_s forms an orthonormal basis in the space of polynomials $\mathcal{P}_s^d$.

(b) The set of polynomials $\Pi(B^d)$ is a complete orthonormal system of functions in the space $L_2(B^d)$.

Let $a \in \mathbb{S}^{d-1}$ be an arbitrary vector and let A be an orthogonal matrix such that $a = Ae$, where $e = (1, 0, \dots, 0)$. Also let A^* be the matrix adjoint to A. We denote $a^* = A^*e$.

Lemma 2. *Let $r_a = r(a \cdot x)$ be an arbitrary ridge function from the class R. Then any Fourier coefficient of the function r_a with respect to the orthonormal system $\Pi_s = \{P_{ij}\}$ admits separation of the variables a and r, i.e., can be represented in the form*

$$\langle r_a, P_{ij}\rangle = h_i(a^*)\,\hat{r}_j,$$

where

$$\hat{r}_j = \int\limits_I r(t)u_j(t)w_{d/2}(t)\,dt$$

is the jth Fourier–Gegenbauer coefficient of the function r.

Proof. In view of the invariance of the measures dx and $d\xi$ under rotations in $\mathbb{R}^d$ and $\mathbb{S}^d$, we conclude that

$$\begin{aligned}
\int\limits_{B^d} r(a\cdot x)P_{ij}(x)\,dx &= \nu_j \int\limits_{B^d} r(a\cdot x)\,dx \int\limits_{\mathbb{S}^{d-1}} h_i(\xi)u_j(\xi\cdot x)\,d\xi \\
&= \nu_j \int\limits_{B^d} r(e\cdot x)\,dx \int\limits_{\mathbb{S}^{d-1}} h_i(A^*\xi)u_j(\xi\cdot x)\,d\xi \\
&= \nu_j \int\limits_{\mathbb{S}^{d-1}} h_i(A^*\xi)\,d\xi \int\limits_{B^d} r(e\cdot x)u_j(\xi\cdot x)\,dx.
\end{aligned}$$

Further, we decompose the function r in the Fourier–Gegenbauer serious

$$r(t) = \sum_{k=0}^{\infty} \hat{r}_k u_k(t)$$

in the space $L_2(I, w)$. By using properties (A.2) and (A.3) established in the Appendix, we find

$$\int_{B^d} r(e\cdot x)u_j(\xi\cdot x)\,dx = \sum_{k=0}^{\infty} \hat{r}_k \int_{B^d} u_k(e\cdot x)u_j(\xi\cdot x)\,dx$$

$$= \hat{r}_j \int_{B^d} u_j(e\cdot x)u_j(\xi\cdot x)\,dx = \hat{r}_j \frac{u_j(e\cdot\xi)}{u_j(1)}.$$

It follows from property (A.4) that

$$\nu_j \int_{\mathbb{S}^{d-1}} h_i(A^*\xi)\,d\xi \int_{B^d} r(e\cdot x)u_j(\xi\cdot x)\,d\xi = \frac{\nu_j\hat{r}_j}{u_j(1)} \int_{\mathbb{S}^d} h_i(A^*\xi)u_j(e\cdot\xi)\,dx = \hat{r}_j h_i(A^*e).$$

The lemma is proved.

By $m = \dim \mathcal{P}_s^d$ we denote the dimension of the space $\mathcal{P}_s^d$. Let $r(x) = \sum_{k=1}^n r_k(a_k\cdot x)$ be any function from the class of ridge functions R_n. Consider the projection of the function r onto the space $\mathcal{P}_s^d$

$$\mathrm{Pr}_s r(x) := \sum_{P_{ij}\in\Pi_s} \langle r, P_{ij}\rangle P_{ij}(x). \tag{3}$$

According to Lemma 2, we can write

$$\mathrm{Pr}_s r(x) = \sum_{k=1}^{n} \sum_{P_{ij}\in\Pi_s} h_i(a_k^*)\,\hat{r}_{kj}\,P_{ij}(x), \tag{4}$$

where $\hat{r}_{kj}$ is the j th Fourier–Gegenbauer coefficient of the function g_k.

3. Vapnik–Chervonenkis Dimension of the Class $\mathrm{Pr}_s R_n$

We now recall the notion of Vapnik–Chervonenkis dimension (for details, see [24]). Consider a function $\operatorname{sgn} a = 1$ for $a \ge 0$ and $\operatorname{sgn} a = -1$ for $a < 0$. For a vector $h = (h_1, \ldots, h_n)$ in $\mathbb{R}^n$, by $\operatorname{sgn} h$ we denote the vector $(\operatorname{sgn} h_1, \ldots, \operatorname{sgn} h_n)$. Let $H = \{h\}$ be a set of real-valued functions defined on $\mathbb{R}^d$. By $\operatorname{sgn} H$ we denote the set of all vectors $\{\operatorname{sgn} h\}$, $h \in H$.

Definition. *The Vapnik–Chervonenkis dimension* $\dim_{VC} H$ *of a set of functions* $H = \{h\}$ *is defined as the maximum natural number* m *for which there exists a collection* $\{\xi_1, \ldots, \xi_m\}$ *in* $\mathbb{R}^d$ *such that the cardinality of the set of "sgn" vectors*

$$S = \{(\operatorname{sgn} h(\xi_1), \ldots, \operatorname{sgn} h(\xi_m)) : h \in H\}$$

is equal to 2^m. *Thus, the set* S *coincides with the set of all vertices of the unit cube in the space* $\mathbb{R}^m$.

Let $\{\xi_1, \dots, \xi_m\}$ be any collection of points in $\mathbb{R}^d$. Consider a set of vectors in $\mathbb{R}^d$

$$\Pi_{m,s,n} = \Big\{(P(\xi_1 + t), \dots, P(\xi_m + t)):\ P \in \mathrm{Pr_s}\, R_n,\ t \in \mathbb{R}^d\Big\}.$$

It is necessary to estimate the cardinality $|\operatorname{sgn} \Pi_{m,s,n}|$ of the set of "sgn" vectors $\operatorname{sgn} \Pi_{m,s,n}$. To this end, we use the following result:

Lemma 3 ([9], Lemma 3)**.** *Let m, s, l, and q be any natural numbers such that $l + q \le m/2$. Also let $\pi_{\alpha\beta}(\sigma)$, $\alpha = 1, \dots, m$, $\beta = 1, \dots, q$ be any fixed polynomials with real coefficients in the variables $\sigma \in \mathbb{R}^l$, each of degree $2s$. Consider m polynomials in the $l + q$ variables $b \in \mathbb{R}^q$ and $\sigma \in \mathbb{R}^l$*

$$\pi_\alpha(b, \sigma) = \sum_{\beta=1}^{q} b_\beta \pi_{\alpha\beta}(\sigma), \qquad \alpha = 1, \dots, m, \tag{5}$$

and a polynomial manifold in $\mathbb{R}^m$

$$\Pi^*_{m,s,l,q} = \Big\{(\pi_1(b, \sigma), \dots, \pi_m(b, \sigma)):\ (b, \sigma) \in \mathbb{R}^q \times \mathbb{R}^l\Big\}.$$

Then the following estimate holds for the cardinality of the set $\operatorname{sgn} \Pi^*_{m,s,p,q}$:

$$\left|\operatorname{sgn} \Pi^*_{m,s,l,q}\right| \le (4s)^l (l + q + 1)^{l+2} \left(\frac{2em}{l + q}\right)^{l+q}.$$

Lemma 4. *There exist absolute constants c_0, c_1, and c_2 such that*

$$c_0 n \le s^{d-1} \le 2c_0 n, \qquad c_1 s^d \le m \le c_2 s^d,$$

and the cardinality of the set $\operatorname{sgn} \Pi_{m,s,n}$ satisfies the inequality

$$|\operatorname{sgn} \Pi_{m,s,n}| \le 2^{cm},$$

where $c \le 1/4$ is an absolute constant.

Proof. Consider a polynomial space $\mathcal{P}_s^d$ with orthonormal basis $\Pi_s = \{P_{i,j}\}_{(i,j)\in I_s}$. Let $P \in \mathrm{Pr_s}\, R_n$ be an arbitrary polynomial. Then

$$P(x) = \sum_{(i,j)\in I_s} \langle r, P_{ij}\rangle P_{ij}(x), \tag{6}$$

where the function $r(x) = \sum_{k=1}^{n} r_k(a_k \cdot x)$ belongs to the manifold R_n. We show that, for any point $\xi \in \mathbb{R}^d$, the polynomial $P(\xi + t)$ can be represented as a linear combination of polynomials in the variables $a_1^*, \dots, a_n^*$, and t. It follows from identity (4) that

$$P(x) = \sum_{k=1}^{n} \sum_{(i,j)\in I_s} h_i(a_k^*) \hat{r}_{kj} P_{ij}(x). \tag{7}$$

Recall that the set I_s consists of the couples (i, j) from the set I satisfying the inequality $\deg h_i \le j \le s$. For any j, we introduce a set I_s^j formed by all numbers i such that $\deg h_i = j$. Then the polynomial $P(x)$ can be rewritten as

$$P(x) = \sum_{k=1}^{n} \sum_{j=1}^{s} \hat{r}_{kj} \sum_{i \in I_s^j} h_i(a_k^*) P_{ij}(x). \tag{8}$$

Since $\{P_{i,j}\}_{(i,j)\in I_s}$ is an orthonormal basis in the space $\mathcal{P}_s^d$, for any t, there exists a nondegenerate matrix

$$\Gamma(t) = \{\gamma_{ij}^{i'j'}(t)\}_{(i,j),(i',j')\in I_s},$$

where (i, j) and (i', j') are, respectively, the column and row subscripts of the matrix $\Gamma(t)$, such that

$$P_{ij}(\xi + t) = \sum_{(i',j')\in I_s}^{m} \gamma_{ij}^{i'j'}(t) P_{i'j'}(\xi). \tag{9}$$

Note that all functions $\gamma_{ij}^{i'j'}(t)$ are polynomials from the space $\mathcal{P}_s^d$. Hence, it follows from (8) and (9) that

$$P(\xi + t) = \sum_{k=1}^{n} \sum_{j=1}^{s} \hat{r}_{kj} \sum_{i \in I_s^j} \sum_{(i',j')\in I_s} \gamma_{ij}^{i'j'}(t) h_i(a_k^*) P_{i'j'}(\xi). \tag{10}$$

We enumerate the set $\{(k, j)\colon 1 \le k \le n,\ 1 \le j \le s\}$ in $\{\beta = 1, \dots, q\}$, where $q = ns$ and put $b_\beta = \hat{r}_{kj}$. For any point ξ_α, $\alpha = 1, \dots, m$, and subscript $\beta = 1, \dots, ns$ we define a function on $\mathbb{R}^{(d+1)n}$ as follows:

$$\pi_{\alpha\beta}(a_1^*, \dots, a_n^*, t) = \sum_{i \in I_s^j} \sum_{(i',j')\in I_s} \gamma_{ij}^{i'j'}(t) h_i(a_k^*) P_{i'j'}(\xi_\alpha).$$

Thus, identity (10) can be rewritten as

$$P(\xi + t) = \sum_{\beta=1}^{q} b_\beta \pi_{\alpha\beta}(a_1^*, \dots, a_n^*, t), \qquad \alpha = 1, \dots, m.$$

We introduce variables $\sigma = (a_1^*, \dots, a_n^*, t)$ from the polynomial space $\mathcal{P}_{2s}^l$ with $l = (d + 1)n$. Thus, the vector set $\Pi_{m,s,n}$ belongs to the set $\Pi_{m,2s,l,q}^*$ with $q = sn$. This and Lemma 3 imply that

$$|\operatorname{sgn} \Pi_{m,2s,n}| \le (8s)^l (l + q + 1)^{l+2} \left(\frac{2em}{l + q}\right)^{l+q}. \tag{11}$$

Let c_0 and $c_1 < c_2$ be positive numbers (they are chosen in what follows). Assume that the numbers m, s, and l satisfy the conditions

$$c_0 n \le s^{d-1} \le 2c_0 n, \qquad c_1 s^d \le m \le c_2 s^d, \qquad \text{and} \qquad l = (d + 1)n. \tag{12}$$

Substituting these conditions in inequality (11), we complete the proof of Lemma 4.

Lemma 4 directly yields the following corollary:

Corollary 1. *The Vapnik–Chervonenkis dimension of the polynomial class* $\mathrm{Pr_s} R_n$ *satisfies the estimate*

$$\dim_{VC} \mathrm{Pr_s} R_n \le l \log_2(4s) + (l+2)\log_2(l+q+1) + (l+q)\log_2\left(\frac{2em}{l+q}\right).$$

4. Approximation of the Class $\mathcal{P}_s^d$ by Ridge Functions

Let

$$\Omega = \left[-\frac{1}{\sqrt{d}}, \frac{1}{\sqrt{d}}\right]^d$$

be a cube lying in the unit ball B^d. We define a function on $\mathbb{R}^d$

$$\omega(x) = \begin{cases} 1, & x \in \frac{1}{2}\Omega, \\ 0, & x \in \mathbb{R}^d \setminus \Omega, \end{cases}$$

and continue this functions to the space $\mathbb{R}^d$ so that ω belongs to the class $W_\infty^r(\mathbb{R}^d)$ and $0 \le \omega(x) \le 1$ for all $x \in \mathbb{R}^d$. Let λ and m be any natural numbers such that $m^{1/d} \le \lambda \le 2m^{1/d}$. Consider a lattice subset from the cube Ω formed by m points as follows:

$$\Xi^m = \left\{\left(\frac{i_1 + 1/2}{\sqrt{d}\,\lambda}, \ldots, \frac{i_d + 1/2}{\sqrt{d}\,\lambda}\right) \colon i_1, \ldots, i_d = -\lambda, \ldots, \lambda - 1\right\}.$$

Let $\xi_1, \ldots, \xi_m$ be the points of the set Ξ^m. We introduce a set

$$E^m = \left\{\varepsilon = (\varepsilon_1, \ldots, \varepsilon_m) \colon \varepsilon_i = \pm 1,\ i - 1, \ldots, m\right\}$$

of sign vectors in $\mathbb{R}^m$. Consider a collection of functions

$$\mathcal{F}^m = \left\{ f_\varepsilon(x) := (2\lambda)^{-r} \sum_{i=1}^m \varepsilon_i \omega\left(2\lambda(x - \xi_i)\right) \colon \varepsilon \in E^m \right\}.$$

Clearly, every function f_ε from $\mathcal{F}^m$ belongs to the Sobolev class W_∞^r. Denote by g_ε the polynomial of the best approximation of the function f_ε in the L_∞-norm of the space $\mathcal{P}_s^d$, i.e., such that

$$\|f_\varepsilon - g_\varepsilon\|_\infty = \min_{g \in \mathcal{P}_s^d} \|f_\varepsilon - g\|_\infty.$$

It is known [22] that the error of the best approximation of any function $f \in W_\infty^r$ from the polynomial space $\mathcal{P}_s^d$ in the L_∞-norm is bounded above as follows,

$$\|f_\varepsilon - g_\varepsilon\|_\infty \le c s^{-r}.$$

Hence, we arrive at the following result:

Proposition 1. *Consider a set of polynomials*

$$G_s^m = \{g_\varepsilon : \varepsilon \in E^m\}.$$

Then the deviation of the set $\mathcal{F}^m$ from the space G_s^m satisfies the inequality

$$e(\mathcal{F}^m, G_s^m)_\infty \le cs^{-r}.$$

Let Q be a set of functions from the space $L(B^d)$. By $\mathrm{Pr_s} Q = \{\mathrm{Pr_s} q : q \in Q\}$ we denote the projection of the set Q onto the subspace $\mathcal{P}_s^d$.

Lemma 5. *Let $1 \le q \le \infty$ be an arbitrary number and let $P \in \mathcal{P}_s^d$ be an arbitrary polynomial. Then*

$$e(P, R_n)_q \ge e(P, \mathrm{Pr_s} R_n)_q.$$

Proof. We have

$$e(P, R_n)_q = \inf_{r_i \in L_{2,\mathrm{loc}}(\mathbb{R}),\, a_i \in \mathbb{R}^d} \left\| P(x) - \sum_{i=1}^n r_i(a_i \cdot x) \right\|_q . \tag{13}$$

Further, we fix a set of vectors $a = \{a_1, \dots, a_n\}$ and consider a linear subspace of functions

$$U_n(a) := \left\{ u = \sum_{i=1}^n u_i(a_i \cdot x) \colon u_i \in L_{2,\mathrm{loc}}(\mathbb{R}) \right\}.$$

Let $U_n(a)^\perp = \{v \in L_q : \langle v, u\rangle = 0 \text{ for all } u \in U_n(a)\}$ be the annihilator subspace in L_q for the subspace $U_n(a)$. We define a number q' such that $1/q + 1/q' = 1$. In view of the duality in the space L_q, we get

$$\inf_{u \in U_n(a)} \|P - u\|_q = \sup_{v \in U_n(a)^\perp,\, \|v\|_{q'} \le 1} \langle P, v\rangle \ge \sup_{v \in U_n(a)^\perp \cap \mathcal{P}_s^d,\, \|v\|_{q'} \le 1} \langle P, v\rangle.$$

Since

$$U_n(a)^\perp \cap \mathcal{P}_s^d = \{v \in \mathcal{P}_s^d : \langle v, U_n(a)\rangle = 0\} = \{v : \langle v, \mathrm{Pr_s}\, U_n(a))\rangle = 0\},$$

by using the duality in the space $\mathcal{P}_s^d$ once again, we obtain

$$e(P, U_n(a))_q \ge \sup_{v \in \mathrm{Pr_s} U_n(a)^\perp \cap \mathcal{P}_s^d,\, \|v\|_{q'} \le 1} \langle P, v\rangle = \inf_{h \in \mathrm{Pr_s}\, U_n(a)} \|P - h\|_q. \tag{14}$$

It follows from (13) and (14) that

$$e(P, R_n)_q = \inf_{a_1,\dots,a_n} e(P, U_n(a))_q \ge \inf_{a_1,\dots,a_n} \inf_{h \in \mathrm{Pr_s} U_n(a)} \|P - h\|_q = e(P, \mathrm{Pr_s} R_n)_q.$$

Lemma 5 is proved.

5. Proof of Theorem 1

Consider a space l_1^m formed by vectors $x \in \mathbb{R}^m$ and equipped with a norm $\|x\|_{l_1^m} = |x_1| + \ldots + |x_m|$. In the space l_1^m we consider a subset $E^m = \{\varepsilon : \varepsilon_1, \ldots, \varepsilon_m = \pm 1\}$. The following lemma is proved in [9]. For the sake of completeness, we present its proof in what follows.

Lemma 6. *Assume that all conditions of Lemma 4 are satisfied. Then there is a vector $\varepsilon^* \in E^m$ such that*

$$e(\varepsilon^*, \Pi_{m,s,n})_{l_1^m} := \inf_{x \in \Pi_{m,s,n}} \|\varepsilon^* - x\|_{l_1^m} \geq am,$$

where a is an absolute strictly positive constant.

Proof. Let $a < 1$ be an absolute constant satisfying the equation

$$1 - \frac{1}{2}(1 - 2a)^2 \log_2 e = \frac{47}{64}$$

(i.e., $a = 0.19\ldots$). We set $\Pi = \operatorname{sgn} \Pi_{m,s,n}$. Let π be any vector from Π. Consider a subset of E^m

$$E_\pi = \left\{ \varepsilon \in E^m : \sum_{i=1}^{m} |\varepsilon_i - \pi_i| \leq 2am \right\}.$$

Since $\pi_i = \pm 1$, we have the following estimate for the cardinality of the set E_π :

$$|E_\pi| = \left| \left\{ \varepsilon \in E^m : \sum_{i=1}^{m} (\varepsilon_i + 1) \leq 2am \right\} \right| = \left| \left\{ \varepsilon : \sum_{i : \varepsilon_i = 1} 1 \leq am \right\} \right| = \sum_{i=0}^{[am]} \binom{m}{i}.$$

Further, by using the well-known estimate (see, e.g., [3], Chapter 8), we find

$$\sum_{i=0}^{[am]} \binom{m}{i} \leq 2^m e^{-2m(1/2-\beta)^2} \leq 2^{bm},$$

where

$$\beta = m^{-1}[am] \qquad \text{and} \qquad b = 1 - \frac{1}{2}(1-2a)^2 \log_2 e = \frac{47}{64}.$$

Hence, $|E_\pi| \leq 2^{47m/64}$. In E^m, we now consider a subset $E' = \bigcap_{\pi \in \Pi} (E^m \setminus E_\pi)$ and estimate the cardinality of E' as follows:

$$|E'| = \left| E^m \setminus \bigcup_{\pi \in H} E_\pi \right| \geq 2^m - |\Pi| \max_{\pi \in \Pi} |E_\pi| \geq 2^m - |\Pi| 2^{(47/64)m}. \tag{15}$$

By Lemma 4, we conclude that $|\Pi| \leq 2^{m/4}$. This and (15) imply that $|E'| \geq 2^m - 2^{(63/64)m} > 0$. Therefore, there exists a vector ε^* such that the following inequality holds for any vector $\pi \in \Pi$:

$$\|\varepsilon^* - \pi\|_{l_1^m} \geq 2am.$$

This yields the inequality

$$e(\varepsilon^*, \Pi_{m,s,n})_{l_1^m} \geq \frac{1}{2} e(\varepsilon^*, \operatorname{sgn} \Pi_{m,s,n})_{l_1^m} \geq am.$$

Lemma 6 is proved.

Lemma 7. *Assume the natural numbers m, s, and n satisfy the conditions of Lemma 4. Then there exists $f_{\varepsilon^*} \in \mathcal{F}^m$ such that*

$$e(f_{\varepsilon^*}, \mathrm{Pr}_s R_n)_1 \geq c_3 n^{-r/(d-1)},$$

where c_3 is an absolute strictly positive constant.

Proof. Let f_ε and P be any functions from the sets $\mathcal{F}^m$ and $\mathrm{Pr}_s R_n$, respectively. We have

$$\|f_\varepsilon - P\|_1 \geq \int_{\Omega} |f_\varepsilon(x) - P(x)| dx = \int_{\Omega/(2\lambda)} \sum_{i=1}^{m} \big| f_\varepsilon(\xi_i + t) - P(\xi_i + t) \big| dt.$$

We now define a function $\bar{\omega}(t) = (2\lambda)^{-r} \omega(2\lambda t)$. Since $f_\varepsilon(\xi_i + t) = \bar{\omega}(t)\varepsilon_i$, for every vector ε, we conclude that, for any t from the cube $\Omega/(2\lambda)$,

$$\sum_{i=1}^{m} \big| f_\varepsilon(\xi_i + t) - P(\xi_i + t) \big| \geq \inf_{P \in \mathrm{Pr}_s R_n,\, \tau \in \mathbb{R}^d} \sum_{i=1}^{m} \big| \bar{\omega}(t)\varepsilon_i - P(\xi_i + \tau) \big|$$

$$= \inf_{P \in \mathrm{Pr}_s R_n,\, \tau \in \mathbb{R}^d} \bar{\omega}(t) \sum_{i=1}^{m} \big| \varepsilon_i - P(\xi_i + \tau) \big|.$$

Thus, we get

$$\|f_\varepsilon - P\|_1 \geq \frac{1}{m} \inf_{t \in \Omega/(2\lambda)} \big| \bar{\omega}(t) \big| \inf_{P \in \mathrm{Pr}_s R_n,\, \tau \in \mathbb{R}^d} \sum_{i=1}^{m} \big| \varepsilon_i - P(\xi_i + \tau) \big| \geq \frac{c_3}{(2\lambda)^r m} e(\varepsilon, \Pi_{m,s,n}, l_1^m).$$

Recall (see (12)) that the numbers m, s, and n satisfy the conditions $c_0 n \leq s^{d-1} \leq 2c_0 n$, $c_1 s^d \leq m \leq c_2 s^d$, and $m^{1/d} \leq \lambda \leq 2m^{1/d}$. Applying Lemma 6, we conclude that there is a function $f_{\varepsilon^*} \in \mathcal{F}^m$ satisfying the inequality

$$\|f_{\varepsilon^*} - P\|_1 \geq \frac{c_3 a}{(2\lambda)^r} \geq c_4 n^{-r/(d-1)}, \qquad c_4 = \frac{c_3 a}{c_2^{r/d} (2c_0)^r},$$

for any polynomial $P \in \mathrm{Pr}_s R_n$.

Lemma 7 is proved.

Lemma 8. *The deviation of the set $\mathcal{F}^m$ from the class R_n satisfies the inequality*

$$e(\mathcal{F}^m, R_n)_1 \geq c_4 n^{-r/(d-1)}.$$

Proof. Let $f \in \mathcal{F}^m$ be an arbitrary function. According to Proposition 1, one can find a polynomial $g \in G_s^m$ such that

$$\|f - g\|_\infty \le cs^{-r}. \tag{16}$$

By using Lemma 5 and (twice) inequality (16), we obtain

$$e(\mathcal{F}^m, R_n)_1 \ge e(G_s^m, R_n, L_1) - cs^{-r}$$
$$\ge e(G_s^m, \mathrm{Pr_s} R_n)_1 - cs^{-r} \ge e(\mathcal{F}^m, \mathrm{Pr_s} R_n)_1 - 2cs^{-r}.$$

We choose c_2 and c_0 such that $c_4 > 2c$. Then it follows from Lemma 7 that

$$e(\mathcal{F}^m, R_n)_1 \ge c_4 n^{-r/(d-1)} - 2cs^{-r} \asymp n^{-r/(d-1)}.$$

Lemma 8 is proved.

We now prove Theorem 1. It is known that the collection of functions $\mathcal{F}^m$ belongs to the class W_∞^r. Therefore, by using Hölder's inequality for $1 \le q \le p \le \infty$ and Lemma 8, we get

$$e(W_p^r, R_n)_q \ge e(W_\infty^r, R_n)_1 \ge e(\mathcal{F}^m, R_n)_1 \ge cn^{-r/(d-1)}.$$

The upper bound

$$e(W_p^r, R_n)_q \le cn^{-r/(d-1)}$$

was established in [9].

The proof of Theorem 1 is completed.

6. Appendix

We now discuss some well-known results connected with the orthogonal polynomials used in the present work.

Gegenbauer Polynomials. The Gegenbauer polynomials are usually defined via the generating function:

$$(1 - 2tz + z^2)^{-\lambda} = \sum_{k=0}^{\infty} C_k^\lambda(t) z^k,$$

where $|z| < 1$, $|t| < 1$, and $\lambda > 0$. The coefficients $C_k^\lambda(t)$ are algebraic polynomials of degree k. They are called the Gegenbauer polynomials associated with λ.

The Gegenbauer polynomials possess the following properties:

1. The family of polynomials $\{C_k^\lambda\}$ is a complete orthogonal system for a weighted space $L_2(I, w)$, where $I = [-1, 1]$, $w(t) := w_\lambda(t) := (1 - t^2)^{\lambda - 1/2}$, and

$$\int_I C_m^\lambda(t) C_n^\lambda(t) w(t)\,dt = \begin{cases} 0, & m \ne n, \\ v_{n,\lambda}, & m = n, \end{cases} \qquad \text{with} \qquad v_{n,\lambda} := \frac{\pi^{1/2}(2\lambda)_n \Gamma(\lambda + 1/2)}{(n+\lambda) n! \Gamma(\lambda)}. \tag{A.1}$$

Here, we use the ordinary notation: $(a)_0 := 0$ and $(a)_N := a(a+1)\dots(a+N-1)$.

2. Let $\mathcal{P}_n$ be the set of all algebraic polynomials of total degree n in d real variables. We set $u_n(t) = v_n^{-1/2} C_n^{d/2}(t)$, where

$$v_n = \frac{\pi^{1/2}(d)_n \Gamma\left((d+1)/2\right)}{(n+d/2)n!\Gamma(d/2)}.$$

The polynomials $u_n(\xi \cdot x)$, $\xi \in S^{d-1}$, are in $\mathcal{P}_n$ and the polynomials $u_n(\xi \cdot x)$ are orthogonal to $\mathcal{P}_{n-1}$ in $L_2(B^d)$ (see [18]):

$$\int_{B^d} u_n(\xi \cdot x) p(x) dx = 0 \qquad \forall \xi \in S^{d-1} \quad \text{and} \quad \forall p \in \mathcal{P}_{n-1}. \tag{A.2}$$

3. For any $\xi, \eta \in S^{d-1}$ we have (see [18])

$$\int_{B^d} u_n(\xi \cdot x) u_n(\eta \cdot x) dx = \frac{u_n(\xi \cdot \eta)}{u_n(1)}. \tag{A.3}$$

4. For any polynomial $h(x) \in \mathcal{P}_n$ such that $h(x) = (-1)^n h(-x)$ for all $x \in \mathbf{R}^d$, we have (see [18])

$$\int_{S^{d-1}} h(\xi) u_n(\xi \cdot \eta) d\xi = \frac{u_n(1)}{\nu_n} h(\eta), \qquad \text{where} \quad \nu_n = \frac{(n+1)_{d-1}}{2(2\pi)^{d-1}}. \tag{A.4}$$

An Orthogonal System of Polynomials on the Sphere. We now present some facts (see [4, 25, 20]) from the theory of harmonic analysis on the sphere. Let s be an arbitrary positive integer. Consider a space $\mathcal{H}_s$ formed by homogeneous harmonic polynomials of degree s in d variables $x_1, \dots, x_d$. Any polynomial from $\mathcal{H}_s$ is decomposable in a linear combination of polynomials of the form

$$h_{sk}(x) = A_{sk} \prod_{j=o}^{d-2} r_{d-j}^{k_j - k_{j-1} + 1} C_{k_j - k_{j+1}}^{\frac{d-j-2}{2} + k_{j+1}} \left(\frac{x_{d-j}}{r_{d-j}} \right) (x_2 \pm i x_1)^{k_{d-2}}, \tag{A.5}$$

where $r_{d-j}^2 = x_1^2 + \dots + x_{d-j}^2$. The vector k with integer coordinates belongs to the set

$$K^s = \left\{ k = (k_0, k_1, \dots, k_{d-3}, \varepsilon k_{d-2}) \colon 0 \le k_{d-2} \le \dots \le k_1 \le k_0 = s, \;\; \varepsilon = \pm 1 \right\},$$

and A_{sk} is the normalization factor. It is known that the dimension of the space $\mathcal{H}_s$ is given by the formulas

$$\dim \mathcal{H}_s = |K^s| = \binom{s+d-1}{s} - \binom{s+d-3}{s-2}, \tag{A.6}$$

for $s \geq 2$, $\dim \mathcal{H}_0 = 1$, and $\dim \mathcal{H}_1 = d$. It is easy to see that the dimension of $\mathcal{H}_s$ is asymptotically given by

$$\dim \mathcal{H}_s = \left(2 + \frac{2}{(d-2)!} + c(s,d)\right) s(s+1)\dots(s+d-3) \asymp s^{d-2}, \qquad (A.7)$$

where $0 \leq c(s,d) \leq 1$ is a function depending only on s and d.

The family of functions $\{h_{sk}\}_{k \in K^s}$ is an orthonormal system in the space $L_2(S^{d-1})$, i.e., the following relation holds for any multiindices $k, k' \in K^s$:

$$(h_{sk}, h_{sk'}) = \int_{S^{d-1}} h_{sk}(\xi)\overline{h_{sk'}(\xi)} d\xi = \delta_{kk'}. \qquad (A.8)$$

Note that, for $s \neq s'$, the spaces $\mathcal{H}_s$ and $\mathcal{H}_{s'}$ are orthogonal with respect to the inner product (A.8). The family of functions $\bigcup_{s=0}^{\infty}\{h_{sk}\}_{k \in K^s}$ is a complete orthonormal system in the space $L_2(S^{d-1})$.

The set of polynomials of degree $\leq n$ on the sphere $\{p\colon p \in \mathcal{P}_n\}$ is contained in the space $\mathcal{H}_0 \oplus \mathcal{H}_1 \oplus \dots \oplus \mathcal{H}_n$, which is the direct sum of the orthogonal subspaces $\mathcal{H}_0, \mathcal{H}_1, \dots, \mathcal{H}_n$. This implies that the following equality holds for any polynomial $p \in \mathcal{P}_n$ and any function $h \in \mathcal{H}_{n+1} \oplus \mathcal{H}_{n+2} \oplus \dots$:

$$\int_{S^{d-1}} p(\xi)\overline{h(\xi)} d\xi = 0.$$

REFERENCES

1. A. R. Barron, " Universal approximation bounds for superposition of a sigmoidal function," *IEEE Trans. Inform. Theory*, **39**, 930–945 (1993).
2. R. A. DeVore, K. Oskolkov, and P. Petrushev, "Approximation by feed-forward neural networks," *Ann. Numer. Math.*, **4**, 261–287 (1997).
3. L. Devroye, L. Györfy, and G. Lugosi, *A Probabilistic Theory of Pattern Recognition*, Springer, New York (1996).
4. A. Erdelyi, ed., *Higher Transcendental Functions. Vol. 2. Bateman Manuscript Project*, McGraw Hill, New York (1953).
5. Y. Gordon, V. Maiorov, M. Meyer, and S. Reisner, "On best approximation by ridge functions in the uniform norm," *Constr. Approxim.*, **18**, 61–85 (2002).
6. V. Ya. Lin and A. Pinkus, "Fundamentality of ridge functions," *J. Approxim. Theory*, **75**, 295–311 (1993).
7. V. Ya. Lin and A. Pinkus, "Approximation of multivariate functions," *Adv. Comput. Math.*, World Sci. (Singapore), 257–265 (1994).
8. B. Logan and L. Shepp, " Optimal reconstruction of functions from their projections," *Duke Math. J.*, **42**, 645–659 (1975).
9. V. Maiorov, " On best approximation by ridge functions," *J. Approxim. Theory*, **99**, 68–94 (1999).
10. V. Maiorov, "On best approximation of classes by radial functions," *J. Approxim. Theory*, **120**, 36–70 (2003).
11. V. Maiorov, R. Meir, and J. Ratsaby, "On the approximation of functional classes equipped with a uniform measure using ridge functions," *J. Approxim. Theory*, **99**, 95–111 (1999).
12. V. Maiorov and R. Meir, "On the near optimality of the stochastic approximation of smooth functions by neural networks," *Adv. Comput. Math.*, **13**, 79–103 (2000).
13. V. Maiorov, K. I. Oskolkov, and V. N. Temlyakov, "Gridge approximation and Radon compass," *Approxim. Theory (a Volume dedicated to Blagovest Sendov)*, Ed. B. Bojanov, DARBA, Sofia, 284–309 (2002).
14. Makovoz Y., "Random approximation and neural networks," *J. Approxim. Theory*, **85**, 98–109 (1996).
15. H. N. Mhaskar, "Neural networks for optimal approximation of smooth and analytic functions," *Neural Comput.*, **8**, 164–177 (1996).
16. H. N. Mhaskar and C. A. Micchelli, "Dimension independent bounds on the degree of approximation by neural networks," *IBM J. Res. Develop.*, **38**, 277–284 (1994).
17. K. I. Oskolkov, "Ridge approximation, Chebyshev–Fourier analysis and optimal quadrature formulas," *Proc. Steklov Inst. Math.*, **219**, 265–280 (1997).
18. P. P. Petrushev, "Approximation by ridge functions and neural networks," *SIAM J. Math. Anal.*, **30**, 291–300 (1998).
19. A. Pinkus, "Approximation by ridge functions, some density problems from neutral networks," *Surface Fitting and Multiresolution Method*, **2**, 279–292 (1997).
20. E. M. Stein and G. Weiss, *Introduction to Fourier Analysis on Euclidean Spaces*, New Jersey, Princeton (1971).
21. V. Temlyakov, *On Approximation by Ridge Functions*, Preprint.
22. A. F. Timan, *Theory of Approximation of Functions of the Real Variable*, Macmillan Co., New York (1963).

23. H. Tribel, *Interpolation Theory, Function Spaces, Differential Operators*, VEB Deutscher Verlag Wissenschaften, Berlin (1978).
24. V. Vapnik and A. Chervonkis, "Necessary and sufficient conditions for the uniform convergence of empirical means to their expectations," *Theory Probab. Appl.*, **3**, 532–553 (1981).
25. N. Ya. Vilenkin, *Special Functions and the Theory of Representations of Groups* [in Russian], Fizmatgiz, Moscow (1965).
26. B. A. Vostretsov and M. A. Kreines, "Approximation of continuous functions by superpositions of plane waves," *Sov. Math. Dokl.*, **2**, 1320–1329 (1961).