UDC 517.5

V. E. Maiorov (Technion, Haifa, Israel)

BEST APPROXIMATION BY RIDGE FUNCTIONS IN L_p-SPACES

НАЙКРАЩЕ НАБЛИЖЕННЯ ХРЕБТОВИМИ ФУНКЦІЯМИ В L_p-ПРОСТОРАХ

We study the approximation of function classes by the manifold R_n formed by all possible linear combinations of n ridge functions of the form $r(a\cdot x)$). We prove that for any $1\leq q\leq p\leq\infty$, 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 $n^{-r/(d-1)}$.

Досліджено наближення класів функцій многовидом R_n, що утворений усіма можливими лінійними комбінаціями n хребтових функцій вигляду $r(a\cdot x)$). Доведено, що для будь-яких $1\leq q\leq p\leq\infty$ відхилення класу Соболєва W_p^r від множини R_n хребтових функцій у просторі $L_q(B^d)$ характеризується точним порядком $n^{-r/(d-1)}$.

1. Introduction. In this work we continue the study of approximation of multivariate functions by classes consisting of 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 the parameters $a\in\mathbb{R}^d$, $r\colon\mathbb{R}\to\mathbb{R}$ and $a\cdot x$ is the usual inner product. Let $L_q=L_p(B^d)$, $1\leq q\leq\infty$, be Banach space of all q-integrable functions on the unite ball $B^d=\{|x|\leq 1\}$, 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 $\mathbb{R}^d$. 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. We denote $R=R(\mathbb{S}^{d-1})$. Let n be a natural number. Consider the class of functions

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

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

Approximation by ridge functions has been studied by several authors. In the works [26] and [6] necessary and sufficient conditions are found on a set A in order that the closure of the set $R(A)$ coincides 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 a 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], Temlyakov [21]. In Gordon, Maiorov, Meyer and Reisner [5] the results about best approximation by ridge functions in the Banach space L_p are considered.

In this work we consider the problems of 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, q satisfy $1 \leq q \leq p \leq \infty$. Earlier, the asymptotical estimates of approximation by R_n were studied in [5, 9], but only for $2 \leq q \leq p \leq \infty$.

Let $\rho = (\rho_1, \ldots, \rho_d)$ be a multiindex vector, that is, ρ is the vector with nonnegative integer coordinates, $|\rho| = \rho_1 + \ldots + \rho_d$. Introduce the differential operator $\mathcal{D}^\rho = \partial^{|\rho|}/\partial^{\rho_1} x_1 \ldots \partial^{\rho_d} x_d$.

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

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

For subsets $W, V \subset L_q$ we define the deviation of W from V by

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

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

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

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

We describe briefly the proof of the 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 will be done 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 ridge function $r(a \cdot x)$ 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 for any k the identity holds $\langle r_a, P_k \rangle = u(a)v(r)$, where u and v are some function of a and linear functional of r, respectively. 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 to the polynomial space $\mathcal{P}_s^d$. In Section 4 we prove Theorem 1 using the results of Sections 2, 3. In the Appendix we present well-known results from the theory of orthogonal polynomials on the segment and from the theory of harmonic analysis on the sphere, which we use in the proofs of Theorem 1.

In the rest of this paper notations like c, c', c_0, $c_1, \ldots$, etc. denote positive constants which do not depend on the parameter n and may depend only on r, d, p, or q. For two positive sequences $\{a_n\}$ and $\{b_n\}$, $n = 0, 1, \ldots$ we write $a_n \asymp b_n$ if there exist positive constants c_1 and c_2 such that $c_1 \leq a_n/b_n \leq c_2$ for all $n = 0, 1, \ldots$.

2. Projection of R_n to the polynomial space. In this section we construct special orthonormal systems of polynomials on the unit ball B^d. Orthogonal systems of polynomials on the ball play the important part in problems of approximations of multivariable functions by manifolds of linear combinations of ridge functions (the plane waves) (see [2, 18, 9, 10]). In the works [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 in the construction of orthogonal systems of polynomials on the ball [8]. Note, in particular, that in the case $d = 2$, these Gegenbauer polynomials coincide with Chebyshev polynomials of second kind.

In our work the system of orthogonal polynomials on the unit ball is obtained, in a 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 describe this construction.

Let $L_2(\mathbb{S}^{d-1})$ be the Hilbert space consisting of all the complex-valued square-integrable functions $h(\xi)$ on the sphere $\mathbb{S}^{d-1}$ with the 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 by $d\xi$ we denote the normalized Lebesgue measure on the sphere S^{d-1}.

Consider (see the Appendix) in the space $L_2(\mathbb{S}^{d-1})$ the subspace $\mathcal{H}$ consisting of the restrictions on $\mathbb{S}^{d-1}$ of the harmonic functions on $\mathbb{R}^d$. Let $\mathcal{H}_s$ be the subspace in $\mathcal{H}$ 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 the 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, using the relation $\dim\ \mathcal{H}_s^{\text{hom}} \asymp s^{d-2}$ (see (A.2)) we obtain

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

We introduce in the space $\mathcal{H}$ the family of functions $\mathcal{B}(\mathbb{S}^{d-1}) := \{h_i\}_{i=0}^{\infty}$ consisting (see (A.2)) of all ordered spherical harmonics, that is, the functions

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

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

Now we consider (see Appendix) the Gegenbauer polynomials $C_n^{d/2}(t)$, $t \in \mathbb{R}$, of degree n associated with $d/2$. We norm the polynomial $C_n^{d/2}$ 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 two arbitrary indices from $\mathbb{Z}_+$. Construct on $\mathbb{R}^d$ the function

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

From (1) we see 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)). Let the set I consists of the couples of nonnegative integers (i, j) such that $\deg h_i \leq j$ and a parity of numbers $\deg h_i$ and j are equal. Let I_s be the subset in I consisting of the couples (i, j) with $\deg h_i \leq j \leq s$. Construct the system of polynomials

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

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

Lemma 1 (see [10])**.** a) *The polynomial set Π_s is 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 any unit vector and let A be an orthogonal matrix such that $a = Ae$, where $e = (1, 0, \ldots, 0)$. Let A^* be the adjoint matrix to A. We denote $a^* = A^* e$.

Lemma 2. *Let $r_a = r(a \cdot x)$ be any 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, that is can be represented as*

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

where

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

is the j-th coefficient of Fourier – Gegenbauer of the function r.

Proof. By the invariance of the measures dx and $d\xi$ relative to the rotation in $\mathbb{R}^d$ and $\mathbb{S}^d$ we have

$$\int\limits_{B^d} r(a \cdot x) P_{ij}(x)\, dx = \nu_j \int\limits_{B^d} r(a \cdot x)\, dx \int\limits_{\mathbb{S}^{d-1}} h_i(\xi) u_j(\xi \cdot x)\, d\xi =$$

$$= \nu_j \int\limits_{B^d} r(e \cdot x)\, dx \int\limits_{\mathbb{S}^{d-1}} h_i(A^*\xi) u_j(\xi \cdot x)\, d\xi =$$

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

Decompose the function r to the Fourier – Gegenbauer serious $r(t) = \sum_{k=0}^{\infty} \hat{r}_k u_k(t)$ in the space $L_2(I, w)$. Using the properties (A.2) and (A.3) from Appendix we obtain

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

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

It follows from the property (A.4) that

$$\nu_j \int\limits_{\mathbb{S}^{d-1}} h_i(A^*\xi)\,d\xi \int\limits_{B^d} r(e\cdot x)u_j(\xi\cdot x)\,d\xi =$$

$$= \frac{\nu_j \hat r_j}{u_j(1)} \int\limits_{\mathbb{S}^d} h_i(A^*\xi)u_j(e\cdot\xi)\,dx = \hat r_j h_i(A^*e).$$

The lemma is proved.

Denote by $m = \dim \mathcal{P}_s^d$ 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 ridge functions class R_n. Consider the projection of the function r to 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 coefficient of Fourier – Gegenbauer of the function g_k.

3. Vapnik – Chervonenkis dimension of the class $\mathrm{Pr_s} R_n$. We recall the notions of Vapnik – Chervonenkis dimension (in details see [24]). Consider the function $\operatorname{sgn} a = 1$, if $a \ge 0$, and $\operatorname{sgn} a = -1$, if $a < 0$. For a vector $h = (h_1,\ldots,h_n)$ in $\mathbb{R}^n$, we denote by $\operatorname{sgn} h$ 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 functions set* $H = \{h\}$ *is defined as the maximal natural number* m *such that there exists a collection* $\{\xi_1,\ldots,\xi_m\}$ *in* $\mathbb{R}^d$ *such that the cardinality of the sgn vectors set* $S = \{(\operatorname{sgn} h(\xi_1),\ldots,\operatorname{sgn} h(\xi_m))\colon h\in H\}$ *is equal to* 2^m. *That is, the set* S *coincides with the set of all vertices of the unit cube in the space* $\mathbb{R}^m$.

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

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

We will need to estimate the cardinality $|\operatorname{sgn}\Pi_{m,s,n}|$ of the sign vectors set $\operatorname{sgn}\Pi_{m,s,n}$. 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* $l+q\le m/2$. *Let* $\pi_{\alpha\beta}(\sigma)$, $\alpha=1,\ldots,m$, $\beta=1,\ldots,q$ *be any fixed polynomials with real coefficients in the variables* $\sigma\in\mathbb{R}^l$, *each of degree* $2s$. *Construct m polynomials in the* $l+q$ *variables* $b\in\mathbb{R}^q$ *and* $\sigma\in\mathbb{R}^l$

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

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

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

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

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

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

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

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

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

where $c \leq 1/4$ *is some absolute constant.*

Proof. Consider the polynomial space $\mathcal{P}^d_s$ with the orthonormal basis $\Pi_s = \{P_{i,j}\}_{(i,j)\in I_s}$ Let $P \in \operatorname{Pr_s} R_n$ be any 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 will show that for every point $\xi \in \mathbb{R}^d$ the polynomial $P(\xi + t)$ can be represented as a linear combination of polynomials on the variables $a^*_1, \ldots, a^*_n$ and t. It follows from the identity (4)

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

Recall that the set I_s consists of the couples (i,j) from the set I satisfying $\deg h_i \leq j \leq s$. For every j introduce the set I^j_s consisting of all numbers i such that $\deg h_i = j$. Then the polynomial $P(x)$ can be written as

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

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

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

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

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

Note that all the functions $\gamma^{i'j'}_{ij}(t)$ are polynomials from the space $\mathcal{P}^d_s$. Hence, from (8) and (9) we obtain

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

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

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

Then the identity (10) can be written as

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

Introduce the variables $\sigma = (a_1^*,\dots,a_n^*,t)$ which belong to 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 $q = sn$. From here and Lemma 3 we obtain

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

Let c_0 and $c_1 < c_2$ be a positive numbers with we will choose below. We assume the numbers m, s, l satisfy the conditions

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

Substituting these conditions to the inequality (11) we complete Lemma 4.

From Lemma 4 we directly obtain the following consequence.

Consequence 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 the cube which belongs to the unit ball B^d. We define the 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 it on the space $\mathbb{R}^d$ such 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 the lattice subset in the cube Ω consisting of m points

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

Let $\xi_1,\dots,\xi_m$ be the points of the set Ξ^m. Introduce the set

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

of sign vectors in $\mathbb{R}^m$. Consider the collection of function

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

Obviously, every function f_ε from $\mathcal{F}^m$ belongs to the Sobolev class W^r_∞. Denote by g_ε the polynomial of best approximation of the function f_ε by the space $\mathcal{P}^d_s$ in L_∞-norm, that is satisfying

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

We know [22] that the error of the best approximation of any function $f \in W^r_\infty$ from the polynomial space $\mathcal{P}^d_s$ in the L_∞-norm is bounded above as follows,

$$\|f_\varepsilon - g_\varepsilon\|_\infty \leq cs^{-r}.$$

Hence we have the next result.

Proposition 1. *Consider the set of polynomials*

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

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

$$e(\mathcal{F}^m, G^m_s)_\infty \leq cs^{-r}.$$

Let Q be a set of functions in the space $L(B^d)$. Denote by $\mathrm{Pr}_\mathrm{s} Q = \{\mathrm{Pr}_\mathrm{s} q : q \in Q\}$ the projection of a set Q to the subspace $\mathcal{P}^d_s$.

Lemma 5. *Let $1 \leq q \leq \infty$ be any number and $P \in \mathcal{P}^d_s$ be any polynomial. Then*

$$e(P, R_n)_q \geq e(P, \mathrm{Pr}_\mathrm{s} R_n)_q.$$

Proof. We have

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

We fix the set of vectors $a = \{a_1, \ldots, a_n\}$ and consider the linear subspace of functions

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

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

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

Since

$$U_n(a)^\perp \cap \mathcal{P}^d_s = \big\{ v \in \mathcal{P}^d_s : \langle v, U_n(a) \rangle = 0 \big\} = \big\{ v : \langle v, \mathrm{Pr}_\mathrm{s}\, U_n(a)) \rangle = 0 \big\},$$

then using once more the duality in the space $\mathcal{P}^d_s$, 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

$$e(P,R_n)_q = \inf_{a_1,\dots,a_n} e(P,U_n(a))_q \ge$$

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

Lemma 5 is proved.

5. Proof of Theorem 1. Consider the space l_1^m consisting of vectors $x\in\mathbb{R}^m$ equipped with the norm $\|x\|_{l_1^m} = |x_1| + \ldots + |x_m|$. In the space l_1^m we consider the subset $E^m = \{\varepsilon\colon \varepsilon_1,\dots,\varepsilon_m = \pm 1\}$. The following lemma is proved in [9]. For completeness we cite its proof.

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

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

where a *is an absolute and strictly positive constant.*

Proof. Let $a<1$ be the absolute constant satisfying the equation $1-\frac{1}{2}(1-2a)^2\log_2 e = \frac{47}{64}$ (i.e., $a = 0.19\ldots$). Set $\Pi = \operatorname{sgn}\Pi_{m,s,n}$. Let π be any vector from Π. Consider the subset in E^m

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

Since $\pi_i=\pm1$ we have the estimate for cardinality of the set E_π

$$|E_\pi| = \left|\left\{\varepsilon\in E^m\colon\ \sum_{i=1}^m (\varepsilon_i+1)\le 2am\right\}\right| =$$

$$= \left|\left\{\varepsilon\colon\ \sum_{i\colon\,\varepsilon_i=1} 1 \le am\right\}\right| = \sum_{i=0}^{[am]}\binom{m}{i}.$$

From the well-known estimate (see, for example, [3], Chapter 8) we have

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

where $\beta = m^{-1}[am]$, and $b = 1-\frac{1}{2}(1-2a)^2\log_2 e = \frac{47}{64}$. Hence $|E_\pi|\le 2^{47m/64}$. Consider in E^m the subset $E' = \bigcap_{\pi\in\Pi}(E^m\setminus E_\pi)$. We estimate the cardinality of E' via

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

By Lemma 4 we have $|\Pi|\le 2^{m/4}$. From this and (15) we obtain $|E'|\ge 2^m - 2^{(63/64)m} > 0$. Therefore there exists a vector ε^* such that for every vector $\pi\in\Pi$ the following inequality holds:

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

From here we obtain 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 is a $f_{\varepsilon^*} \in \mathcal{F}^m$ such that*

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

where c_3 is an absolute and 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

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

Define the 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, then for any t from the cube $\Omega/(2\lambda)$ we have

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

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

Thus we obtain

$$\|f_\varepsilon - P\|_1 \geq \frac{1}{m} \inf_{t \in \Omega/(2\lambda)} \left| \bar{\omega}(t) \right| \inf_{P \in \mathrm{Pr_s} R_n,\, \tau \in \mathbb{R}^d} \sum_{i=1}^{m} \left| \varepsilon_i - P(\xi_i + \tau) \right| \geq$$

$$\geq \frac{c_3}{(2\lambda)^r m}\, e(\varepsilon, \Pi_{m,s,n}, l_1^m).$$

Recall (see (12)) that the numbers m, s, 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 obtain that there is a function $f_{\varepsilon^*} \in \mathcal{F}^m$ satisfying

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

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*

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

Proof. Let $f \in \mathcal{F}^m$ be any function. According to Proposition 1 there is a polynomial $g \in G_s^m$ that

$$\|f - g\|_\infty \leq c s^{-r}. \tag{16}$$

Using Lemma 5 and twice the inequality (16) we obtain

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

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

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

Lemma 8 is proved.

Now we prove Theorem 1. We know that the collection of functions $\mathcal{F}^m$ belongs to the class W_∞^r. Therefore, using Hölder's inequality for $1 \leq q \leq p \leq \infty$ and Lemma 8, we obtain

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

The upper bound

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

was proved in the paper [9].

Theorem 1 is completely proved.

6. Appendix. We discuss some well-known results connected with orthogonal polynomials, which we use in this present work.

The 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 and are termed 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 the weighted space $L_2(I, w)$, $I = [-1, 1]$, $w(t) := w_\lambda(t) := (1 - t^2)^{\lambda - 1/2}$, and

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

where we use the usual notation $(a)_0 := 0$, $(a)_N := a(a+1)\ldots(a+N-1)$.

2. Let $\mathcal{P}_n$ denote the set of all algebraic polynomials of total degree n in d real variables. Set $u_n(t) = v_n^{-1/2} C_n^{d/2}(t)$, where $v_n = \dfrac{\pi^{1/2}(d)_n \Gamma\left((d+1)/2\right)}{(n + d/2) n! \Gamma(d/2)}$. The polynomials $u_n(\xi \cdot x)$, $\xi \in S^{d-1}$, are in $\mathcal{P}_n$ and the $u_n(\xi \cdot x)$ are orthogonal to $\mathcal{P}_{n-1}$ in $L_2(B^d)$ (see [18]):

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

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

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

4. For each 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])

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

An orthogonal system of polynomials on the sphere. We state some facts (see [4, 25, 20]) from the theory of harmonic analysis on the sphere. Let s be any positive integer. Consider a space $\mathcal{H}_s$ consisting of the homogeneous harmonic polynomials of degree s in the d variables $x_1, \ldots, x_d$. Any polynomial from $\mathcal{H}_s$ is decomposable by 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 $r_{d-j}^2 = x_1^2 + \ldots + x_{d-j}^2$. The vector k with integer coordinates belongs to the set

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

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

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

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

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

where $0 \le c(s,d) \le 1$ is some 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., for any multiindices $k, k' \in K^s$, the following holds:

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

Note that the spaces $\mathcal{H}_s$ and $\mathcal{H}_{s'}$ for $s \ne s'$ 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 on the sphere $\{p \colon p \in \mathcal{P}_n\}$ of degree $\le n$ is contained in the space $\mathcal{H}_0 \oplus \mathcal{H}_1 \oplus \ldots \oplus \mathcal{H}_n$, which is the direct sum of the orthogonal subspaces $\mathcal{H}_0, \mathcal{H}_1, \ldots, \mathcal{H}_n$. From the above it follows that for any polynomial $p \in \mathcal{P}_n$ and for any function $h \in \mathcal{H}_{n+1} \oplus \mathcal{H}_{n+2} \oplus \ldots$ the equality

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

holds.

1. *Barron A. R.* Universal approximation bounds for superposition of a sigmoidal function // IEEE Trans. Inform. Theory. – 1993. – **39**. – P. 930 – 945.
2. *DeVore R. A., Oskolkov K., Petrushev P.* Approximation by feed-forward neural networks // Ann. Numer. Math. – 1997. – **4**. – P. 261 – 287.
3. *Devroye L., Györfy L., Lugosi G.* A probabilistic theory of pattern recognition. – New York: Springer Verlag, 1996.
4. *Erdelyi A., ed.* Higher transcendental functions. Vol. 2. Bateman manuscript project. – New York, N. Y.: McGraw Hill, 1953.
5. *Gordon Y., Maiorov V., Meyer M., Reisner S.* On best approximation by ridge functions in the uniform norm // Constr. Approxim. – 2002. – **18**. – P. 61 – 85.
6. *Lin V. Ya., Pinkus A.* Fundamentality of ridge functions // J. Approxim. Theory. – 1993. – **75**. – P. 295 – 311.
7. *Lin V. Ya., Pinkus A.* Approximation of multivariate functions // Adv. Comput. Math. – World Sci. (Singapore), 1994. – P. 257 – 265.
8. *Logan B., Shepp L.* Optimal reconstruction of functions from its projections // Duke Math. J. – 1975. – **42**. – P. 645 – 659.
9. *Maiorov V.* On best approximation by ridge functions // J. Approxim. Theory. – 1999. – **99**. – P. 68 – 94.
10. *Maiorov V.* On best approximation of classes by radial functions // Ibid. – 2003. – **120**. – P. 36 – 70.
11. *Maiorov V., Meir R., Ratsaby J.* On the approximation of functional classes equipped with a uniform measure using ridge functions // Ibid. – 1999. – **99**. – P. 95 – 111.
12. *Maiorov V., Meir R.* On the near optimality of the stochastic approximation of smooth functions by neural networks // Adv. Comput. Math. – 2000. – **13**. – P. 79 – 103.
13. *Maiorov V., Oskolkov K. I., Temlyakov V. N.* Gridge approximation and Radon compass // Approxim. Theory (a Volume dedicated to Blagovest Sendov / Ed. B. Bojanov. – Sofia: DARBA, 2002. – P. 284 – 309.
14. *Makovoz Y.* Random approximation and neural networks // J. Approxim. Theory. – 1996. – **85**. – P. 98 – 109.
15. *Mhaskar H. N.* Neural networks for optimal approximation of smooth and analytic functions // Neural Comput. – 1996. – **8**. – P. 164 – 177.
16. *Mhaskar H. N., Micchelli C. A.* Dimension independent bounds on the degree of approximation by neural networks // IBM J. Research and Development. – 1994. – **38**. – P. 277 – 284.
17. *Oskolkov K. I.* Ridge approximation, Chebyshev – Fourier analysis and optimal quadrature formulas // Proc. Steklov Inst. Math. – 1997. – **219**. – P. 265 – 280.
18. *Petrushev P. P.* Approximation by ridge functions and neural networks // SIAM J. Math. Anal. – 1998. – **30**. – P. 291 – 300.
19. *Pinkus A.* Approximation by ridge functions, some density problems from neutral networks // Surface Fitting and Multiresolution Method. – 1997. – **2**. – P. 279 – 292.
20. *Stein E. M., Weiss G.* Introduction to Fourier analysis on Euclidean spaces. – Princeton, New Jersey, 1971.
21. *Temlyakov V.* On approximation by ridge functions. – Preprint.
22. *Timan A. F.* Theory of approximation of function of the real variable. – New York: Macmillan Co., 1963.
23. *Tribel H.* Interpolation theory function spaces, differential operators. – Berlin: Veb Deutscher Verlag Wissenschaften, 1978.
24. *Vapnik V., Chervonkis A.* Necessary and sufficient conditions for the uniform convergence of empirical means to their expectations // Theory Probab. Appl. – 1981. – **3**. – P. 532 – 553.
25. *Vilenkin N. Ya.* Special functions and a theory of representation of group. – Moscow: Fizmatgiz, 1965.
26. *Vostrecov B. A., Kreines M. A.* Approximation of continuous functions by superpositions of plane waves // Dokl. Akad. Nauk SSSR. – 1961. – **140**. – P. 1237 – 1240 (Soviet Math. Dokl. – 1961. – **2**. – P. 1320 – 1329).

Received 26.05.09