Journal of Approximation Theory **99**, 68–94 (1999)
Article ID jath.1998.3304, available online at http://www.idealibrary.com on IDEAL®

On Best Approximation by Ridge Functions

V. E. Maiorov*

Department of Mathematics, Technion, Haifa, Israel

Communicated by Allan Pinkus

Received July 30, 1996; accepted in revised form August 11, 1998

We consider best approximation of some function classes by the manifold M_n consisting of sums of n arbitrary ridge functions. It is proved that the deviation of the Sobolev class $W_2^{r,d}$ from the manifold M_n in the space L_2 behaves asymptotically as $n^{-r/(d-1)}$. © 1999 Academic Press

1. INTRODUCTION

In this work we study approximation of multivariate functions by ridge functions. Ridge functions are defined as functions of the form $h(a \cdot x)$, where $a, x \in \mathbf{R}^d$, $(d \geqslant 2)$, h: $\mathbf{R} \to \mathbf{R}$, and $a \cdot x$ is the usual inner product. In particular, we study approximation by sums of n arbitrary ridge functions.

For a subset $A \subset \mathbf{R}^d$ consider the space of ridge functions given by

$$M(A) = span\{h(a \cdot x) \colon a \in A, h \in L(\mathbf{R})\},$$

where h runs over the space $L(\mathbf{R})$ of functions integrable on any compact subset of $\mathbf{R}$. Fix n, and consider the set

$$M_n = \bigcup \{M(A) \colon card\ A \leqslant n\},$$

which is the union of all sets $M(A)$, where A runs over all subsets in $\mathbf{R}^d$ of cardinality at most n.

Approximation by ridge functions has been studied by several authors. In Vostrecov and Kreines [26] and Lin and Pinkus [10] necessary and sufficient conditions are found on set A in order that the closure of the set $M(A)$ coincides with the space of continuous functions. In addition, Lin and Pinkus [10] proved that for any fixed n the set M_n is not dense in $C(\mathbf{R}^d)$. Additional results in this direction were obtained by Kroo [9].

* The author was partially supported by the center for Absorption in Science, Ministry of Immigrant Absorption, State of Israel.

0021-9045/99 $30.00

Recently a series of results were established estimating the degree of approximation of functions by the ridge-manifold M_n in the two-dimensional case, $d=2$ (see Oskolkov [17] and Temlyakov [21]). In particular, Oskolkov showed that in the case $d=2$ orders of approximation of radial functions by the ridge-manifold M_n and by the space of algebraic polynomials of degree n coincide.

Note also the results connected with approximation by ridge functions of a special form, neural networks (Mhaskar and Micchelli [16], Mhaskar [15], DeVore *et al.* [4], Petrushev [18], Maiorov and Meir [13]).

We first introduce some definitions. Let K be a compact set in the space $\mathbf{R}^d$. Consider the space $L_2(K, \mathbf{R}^d)$ of functions defined on $\mathbf{R}^d$ with support on the set K and norm

$$\|f\|_2=\left(\int_{\mathbf{R}^d} |f(x)|^2\,dx\right)^{1/2}=\left(\int_K |f(x)|^2\,dx\right)^{1/2}.$$

We denote the ball of radius r in $\mathbf{R}^d$ by $B^d(r)=\{x=(x_1, ..., x_d)\colon \sum_{i=1}^d x_i^2 \leqslant r^2\}$. In the sequel we mainly consider the unit ball $B^d(1)$. We simplify the notation somewhat by using $B^d=B^d(1)$ and $L_2=L_2(B^d, \mathbf{R}^d)$. The results obtained here can be immediately extended to general compact domains K by use of standard extension theorems, as in [1].

For any two sets W, $H\subset L_2$ we define the distance of W from H by

$$dist(W, H, L_2)=\sup_{f\in W} dist(f, H, L_2),$$

where $dist(f, H, L_2)=\inf_{h\in H} \|f-h\|_{L^2}$.

Furthermore, for any function $f\in L_p$ we denote by $\mathscr{F}f$ its Fourier transform

$$\mathscr{F}f(u)=(2\pi)^{-d/2}\int_{\mathbf{R}^d} f(x)\,e^{iu\cdot x}\,dx,$$

where $u\in\mathbf{R}^d$ and $u\cdot x$ is the inner product of u and x. The inverse Fourier transform will be denoted by $\mathscr{F}^{-1}$.

In the space L_2 define the derivative of order $\rho\geqslant 0$ as

$$\mathscr{D}^\rho f:=\mathscr{F}^{-1}\{|u|^\rho\,\mathscr{F}f(u)\},$$

where $|u|=\sqrt{u_1^2+\cdots+u_d^2}$. In the space L_2 consider the class of functions

$$W_2^{r,d}=\left\{f\colon \max_{\rho\leqslant r} \|\mathscr{D}^\rho f\|_2\leqslant 1\right\}.$$

When r is an integer, the class $W_2^{r,d}$ is equivalent to the Sobolev class of functions f from L_2, for which all distributional derivatives $D^\nu f$ of order smaller or equal to r, satisfy an inequality $\|D^\nu f\|_2 \leqslant 1$.

Let c, and $c_1, c_2, \ldots$ be positive constants depending solely upon the parameters r and d. 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 $c_1 \leqslant a_n/b_n \leqslant c_2$ for all $n=0, 1, \ldots$.

The main result of our work is the following.

THEOREM 1. *Let $r>0$, $d \geqslant 2$. Then the asymptotic relation*

$$dist(W_2^{r,d}, M_n, L_2) \asymp n^{-r/(d-1)}$$

holds.

We describe briefly the proof of the main theorem. In order to obtain the lower bound, we construct for any n a function $f \in W_2^{r,d}$ depending on n such that the distance of f from the manifold M_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 construct an orthonormal system $\{p_k(x)\}_{k=1}^{\infty}$ of algebraic polynomials on the ball B^d. Further, we construct a set of polynomials of the form $\{\sum_{k=1}^m u_k p_k : (u_1, \ldots, u_m) \in U\}$, where U is some discrete set in $\mathbf{R}^m$, on which a Bernstein inequality is satisfied. In Section 3 we show that the coefficients of decomposition of the ridge functions $g_a = g(a \cdot x)$, i.e., the inner products $\langle g_a, p_k \rangle$ are algebraic polynomials $\pi(a, b)$ of degree k in the variables a and are linear functions in some variables $b_1, \ldots, b_l$ depending only on the function g. On the basis of the results of Sections 2 and 3 the problem of approximation of functions from the class $W_2^{r,d}$ is reduced to the problem of approximation of finite-dimensional sets by a polynomial manifold of a special form. In Section 4 we obtain estimates for approximation of the cube in $\mathbf{R}^m$ by polynomial manifolds. In Section 5 we prove Theorem 1 using the results of Sections 2–4. 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 proof of Theorem 1.

2. ORTHOGONAL SYSTEM OF ALGEBRAIC POLYNOMIALS ON THE BALL

Consider the Hilbert space L_2 of all square integrable functions on the ball B^d with the inner product

$$\langle f, g \rangle = \int_{B^d} f(x)\, \overline{g(x)}\, dx \qquad (f, g \in L_2).$$

In the present section we construct an orthogonal system of algebraic polynomials on the ball B^d which have the form of ridge functions (or ridge polynomials). We make use of results from the theory of orthogonal polynomials on a segment, and from the harmonic analysis on the Euclidean sphere. Orthogonal subspaces consisting of ridge polynomials may be found in the papers [11, 4, 17, 18]. Some properties of the Gegenbauer polynomials given in [18] will be exploited in our work.

We will use the orthogonal system of polynomials on the ball B^d in Section 3 for the decomposition of the initial ridge functions by ridge polynomials.

Let s be a natural number. Consider the space

$$\mathcal{P}_s = span\{x^k \equiv x_1^{k_1} \cdots x_d^{k_d} : |k| = k_1 + \cdots + k_d \leqslant s\},$$

consisting of all algebraic polynomials on $\mathbf{R}^d$ of degree at most s. Denote by $\mathcal{P}_s^h$ the subspace in $\mathcal{P}_s$ consisting of all homogeneous polynomials of degree s, i.e., $\mathcal{P}_s^h = span\{x_1^{k_1} \cdots x_d^{k_d} : |k| = s\}$.

Let $S^{d-1} = \{\xi = (\xi_1, ..., \xi_d) \in \mathbf{R}^d : \xi_1^2 + \cdots + \xi_d^2 = 1\}$ be the unit sphere in the space $\mathbf{R}^d$.

Consider the Hilbert space $L_2(S^{d-1})$ of square integrable functions on the sphere S^{d-1} with the inner product

$$(h_1, h_2) = \int_{S^{d-1}} h_1(\xi)\, \overline{h_2(\xi)}\, d\xi \qquad (h_1, h_2 \in L_2(S^{d-1})), \tag{1}$$

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

Consider (see the Appendix) in the space $L_2(S^{d-1})$ the subspace $\mathcal{H}_l$ of spherical harmonics of degree l. Let $\{h_{lk}\}_{k \in K^l}$ be the orthogonal system (A.7) in the subspace $\mathcal{H}_l$.

Let further s be any even number. Consider the space $H_{2s} = \mathcal{H}_s \oplus \mathcal{H}_{s+2} \oplus \cdots \oplus \mathcal{H}_{2s}$ which is the direct sum of the orthogonal subspace $\mathcal{H}_s$, $\mathcal{H}_{s+2}$, ..., $\mathcal{H}_{2s}$ of the spherical harmonics with even degrees from s to $2s$. Denote by N_s the dimension of space H_{2s}.

We have $N_s \asymp s^{d-1}$. Indeed, using (see (A.9)) the relation $\dim \mathcal{H}_s \asymp s^{d-2}$ we obtain

$$N_s = \dim H_{2s} = \dim \mathcal{H}_s + \dim \mathcal{H}_{s+2} + \cdots + \dim \mathcal{H}_{2s} \asymp s^{d-1}.$$

Consider in the space H_{2s} the family $\mathcal{B}_s = \{h_i\}_{i=1}^{N_s}$ consisting (see the Appendix) of the functions

$$\{h_{s,k}\}_{k \in K^s} \cup \{h_{s+2,k}\}_{k \in K^{s+2}} \cup \cdots \cup \{h_{2s,k}\}_{k \in K^{2s}}.$$

The set $\mathscr{B}_s$ is an orthonormal basis in the space H_{2s}, i.e., for any indices $1 \leqslant i,\ i' \leqslant N_s$ we have $(h_i, h_{i'}) = \delta_{ii'}$, where $\delta_{ii'} = 0$ for $i \neq i'$, and $\delta_{ii} = 1$.

Now consider (see the Appendix) the Gegenbauer polynomials $C_n^{d/2}(t)$, $t \in \mathbf{R}$, of degree n associated with $d/2$. Set

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

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

Introduce the set of pairs of indices

$$\Phi_s = \left\{ (i, j) : i = 1, 2, ..., N_s; j = \frac{s}{2}, \frac{s}{2} + 1, ..., s \right\}.$$

Construct for any pair $(i, j) \in \Phi_s$ the function on $\mathbf{R}^d$

$$P_{ij}(x) = v_{2j}^{1/2} \int_{S^{d-1}} h_i(\xi)\, u_{2j}(x \cdot \xi)\, d\xi \qquad \left(v_{2j} = \frac{(2j+1)_{d-1}}{2(2\pi)^{d-1}} \right), \tag{2}$$

which are algebraic polynomials of degree $2j$. Here we denote by $x \cdot \xi$ the inner product of the vectors x and ξ.

Consider in Φ_s a subset

$$\Psi_s = \{ (i, j) \in \Phi_s : \deg h_i \leqslant 2j \}. \tag{3}$$

Let us estimate the asymptotic behavior of the cardinality of the set Ψ_s. From (3) we have $|\Psi_s| = \sum_{j=s/2}^{s} (\dim \mathscr{H}_s + \dim \mathscr{H}_{s+2} + \cdots + \dim \mathscr{H}_{2j})$. Using (A.9) we obtain

$$|\Psi_s| \asymp \sum_{j=s/2}^{s} [s^{d-2} + (s+2)^{d-2} + \cdots + (2j)^{d-2}] \asymp s^d. \tag{4}$$

Consider the set $\Pi_s = \{ P_{ij}(x) : (i, j) \in \Psi_s \}$. We will prove that Π_s is an orthonormal system of polynomials on B^d.

LEMMA 1. (1) *For any two multi-indices* (ij), $(i'j') \in \Psi_s$ *the identity*

$$\langle P_{ij}, P_{i'j'} \rangle = \delta_{ii'} \delta_{jj'}$$

holds.

(2) *If* $(i, j) \in \Phi_s \backslash \Psi_s$, *then* $P_{ij}(x) \equiv 0$.

Proof. From the definition (2) of the polynomials P_{ij}, and the properties (A.4) (for $j \neq j'$) and (A.5) (for $j = j'$) of the Gegenbauer polynomials we have

$$\int_{B^d} P_{ij}(x)\,\overline{P_{i'j'}(x)}\,dx = (v_{2j}v_{2j'})^{1/2} \int_{S^{d-1}\times S^{d-1}} h_i(\xi)\,\overline{h_{i'}(\eta)}\,d\xi\,d\eta$$
$$\times \int_{B^d} u_{2j}(x\cdot\xi)\,u_{2j'}(x\cdot\eta)\,dx$$
$$= \delta_{jj'} \frac{v_{2j}}{u_{2j}(1)} \int_{S^{d-1}} h_i(\xi)\,d\xi \int_{S^{d-1}} \overline{h_{i'}(\eta)}\,u_{2j}(\xi\cdot\eta)\,d\eta.$$

Let $j = j'$. Since $(i', j') \in \Psi_s$ then $\deg h_{i'} \leqslant 2j' = 2j$. Therefore we obtain from (A.6)

$$\frac{v_{2j}}{u_{2j}(1)} \int_{S^{d-1}} \overline{h_{i'}(\eta)}\,u_{2j}(\xi\cdot\eta)\,d\eta = \overline{h_{i'}(\xi)}.$$

Hence

$$\int_{B^d} P_{ij}(x)\,\overline{P_{i'j'}(x)}\,dx = \delta_{jj'} \int_{S^{d-1}} h_i(\xi)\,\overline{h_{i'}(\xi)}\,d\xi = \delta_{jj'}\delta_{ii'}.$$

In the final equality we have used the property of orthonormality of the system $\{h_i\}$. The first statement of the lemma is proved. The second statement of the lemma follows from property (A.10). ∎

Let r be any positive number, and r' be the smallest even number such that $r' \geqslant r$. Set $\mu := \mu_r := 2r' - 1$. Let j be any index from set $I_s = \{1, ..., s\}$. Denote by α_j and β_j integers such that $s - j = (\mu + 1)\,\alpha_j + \beta_j$, where $\alpha_j \in \mathbf{Z}$, and $\beta_j \in \{0, ..., \mu\}$.

Set $\gamma_j = \sqrt{v_{2j}/v_{2j}}$. It is easy to verify the asymptotic relation $\gamma_j \asymp j^{-1/2}$. Consider the function a from Ψ_s to $\mathbf{R}$ defined by

$$a(i, j) = a_{ij} = (-1)^{\mu - \beta_j} \binom{\mu}{\beta_j} \gamma_j \varepsilon_{i\alpha_j},$$

where $\varepsilon_{i\alpha_j}$ is some number equal -1 or 1. The set of functions a corresponding to all possible selections of the $\varepsilon_{i\alpha_j} = \pm 1$, $(i, j) \in \Psi_s$ will be denoted A_s^r.

Consider the set of polynomials on $\mathbf{R}^d$

$$\mathscr{P}(A_s^r) = \left\{ \sum_{(i,j)\in\Psi_s} a_{ij} P_{ij}(x) \colon a \in A_s^r \right\}. \tag{5}$$

For any polynomial $P_a \in \mathcal{P}(A_s^r)$ the asymptotic relation $\|P_a\|_2 \asymp s^{(d-1)/2}$ holds. Indeed, from Lemma 1 and (4) we have

$$\|P_a\|_2^2 = \left\| \sum_{(i,j) \in \Psi_s} a_{ij} P_{ij} \right\|_2^2 = \sum_{(i,j) \in \Psi_s} |a_{ij}|^2$$
$$= \sum_{(i,j) \in \Psi_s} \left(\frac{\mu}{\beta_j}\right)^2 \gamma_j^2 \asymp s^{-1} |\Psi_s| \asymp s^{d-1}. \tag{6}$$

In Theorem 2 we will prove that on the class of polynomials $\mathcal{P}(A_s^r)$ a Bernstein inequality holds.

THEOREM 2. *Let $r > 0$ be any number, and s be any even positive integer. Then for any polynomial $P_a \in \mathcal{P}(A_s^r)$ the inequality*

$$\|\mathcal{D}^r P_a\|_2 \leqslant c s^{r+(d-1)/2}$$

holds.

We will first prove this inequality for even r.

LEMMA 2. *Let r, s be any even positive integers with $s \geqslant r/2$. Then for any polynomial $P_a \in \mathcal{P}(A_s^r)$ the inequality*

$$\|\mathcal{D}^r P_a\|_2 \leqslant c s^{r+(d-1)/2}$$

holds.

Proof. From the definition of the operator $\mathcal{D}^r$ it follows that for any polynomial $u = u(t)$, $t \in \mathbf{R}$, and for any unit vector $\xi \in S^{d-1}$

$$\mathcal{D}^r u(x \cdot \xi) = (-\Delta)^{r/2} u(x \cdot \xi) = (-1)^{r/2} \left(\frac{d^r u}{dt^r}\right)_{t = x \cdot \xi},$$

where Δ is the Laplace operator. From definition (2) we have

$$\mathcal{D}^r P_{ij}(x) = v_{2j}^{1/2} \int_{S^{d-1}} h_i(\xi)\, \mathcal{D}^r u_{2j}(x \cdot \xi)\, d\xi = v_{2j}^{1/2} \int_{S^{d-1}} h_i(\xi) \left(\frac{d^r u_{2j}}{dt^r}\right)_{t = x \cdot \xi} d\xi.$$

Set $\lambda = d/2$. Taking into consideration that $\Phi_s = \{(i,j): i = 1, 2, ..., N_s; j = s/2, (s/2)+1, ..., s\}$, and $u_{2j}(t) = v_{2j}^{-1/2} C_{2j}^{\lambda}(t)$, we obtain

$$\mathcal{D}^r P_a(x) = \sum_{(i,j) \in \Phi_s} a_{ij} \mathcal{D}^r P_{ij}(x) = \sum_{i=1}^{N_s} \int_{S^{d-1}} h_i(\xi) \sum_{j=s/2}^{s} \gamma_j^{-1} a_{ij} \left(\frac{d^r C_{2j}^{\lambda}}{dt^r}\right)_{t = x \cdot \xi} d\xi. \tag{7}$$

Set $a'_{ij}=\gamma_j^{-1}a_{ij}$. Define $C_k^{\alpha}(t)\equiv 0$ for negative k and all α. From property (A.2) we have

$$\frac{d^r C_{2j}^{\lambda}}{dt^r}(t)=2^r(\lambda)_r\, C_{2j-r}^{\lambda+r}(t).$$

From this and (7) it follows that

$$\mathscr{D}^r P_a(x)=2^r(\lambda)_r \sum_{i=1}^{N_s} \int_{S^{d-1}} h_i(\xi) \sum_{j=s/2}^{s} a'_{ij} C_{2j-r}^{\lambda+r}(x\cdot\xi)\, d\xi. \tag{8}$$

We express the interior sum in (8) as a linear combination of the polynomials C_k^{λ}, in the following way. From identity (A.3) we have for any natural numbers q and $\alpha\geqslant 2$

$$C_{2q}^{\alpha}(t)=\sum_{k=0}^{q}\frac{2k+\alpha-1}{\alpha-1}\, C_{2k}^{\alpha-1}(t).$$

Since $\lambda+1=d/2+1\geqslant 2$, then using this formula r times we obtain

$$C_{2q}^{\lambda+r}(t)=\sum_{k_1=0}^{q}\sum_{k_2=0}^{k_1}\sum_{k_3=0}^{k_2}\cdots\sum_{k_r=0}^{k_{r-1}}\left(\prod_{i=1}^{r}\frac{2k_i+\lambda+r-i}{\lambda+r-i}\right) C_{2k_r}^{\lambda}(t).$$

Change the order of summation to obtain

$$C_{2q}^{\lambda+r}(t)=\sum_{k_r=0}^{q}\;\sum_{k_{r-1}=k_r}^{q}\;\sum_{k_{r-2}=k_{r-1}}^{q}\cdots\sum_{k_1=k_2}^{q}\left(\prod_{i=1}^{r}\frac{2k_i+\lambda+r-i}{\lambda+r-i}\right) C_{2k_r}^{\lambda}(t). \tag{9}$$

Set $k=k_r$ and consider the function in the variable q

$$Q_k(q)=\sum_{k_{r-1}=k}^{q}\;\sum_{k_{r-2}=k_{r-1}}^{q}\cdots\sum_{k_1=k_2}^{q}\prod_{i=1}^{r-1}(2k_i+\lambda+r-i). \tag{10}$$

The expression (9) may be rewritten as

$$C_{2q}^{\lambda+r}(t)=\frac{1}{(\lambda)_r}\sum_{k=0}^{q}(2k+\lambda)\, Q_k(q)\, C_{2k}^{\lambda}(t). \tag{11}$$

Consider the interior sum in (8) and set $S_i(t)=\sum_{j=s/2}^{s} a'_{ij} C_{2j-r}^{\lambda+r}(t)$. From identity (11) it follows that

$$S_i(t)=\frac{1}{(\lambda)_r}\sum_{j=s/2}^{s} a'_{ij}\sum_{k=0}^{j-r/2}(2k+\lambda)\, Q_k(j-r/2)\, C_{2k}^{\lambda}(t).$$

Let $j_k = \min\{s/2, k+(r/2)\}$, and in the last expression change the order of summation so that

$$S_i(t) = \frac{1}{(\lambda)_r} \sum_{k=0}^{s-r/2} (2k+\lambda)\, C_{2k}^{\lambda}(t) \sum_{j=j_k}^{s} a'_{ij} Q_k(j-r/2).$$

Substitute this expression into (8) and take into consideration that $\gamma_k^{-1} \int_{S^{d-1}} h_i(\xi)\, C_{2k}^{\lambda}(x\cdot\xi)\, d\xi = P_{ik}(x)$, and $P_{ik}(x) \equiv 0$ for all $(i,k) \notin \Psi_s$. Thus

$$\begin{aligned}
\mathscr{D}^r P_a(x) &= 2^r \sum_{i=1}^{N_s} \int_{S^{d-1}} \left[h_i(\xi) \sum_{k=0}^{s-r/2} (2k+\lambda)\, C_{2k}^{\lambda}(x\cdot\xi) \sum_{j=j_k}^{s} a'_{ij} Q_k(j-r/2) \right] d\xi \\
&= 2^r \sum_{i=1}^{N_s} \sum_{k=0}^{s-r/2} (2k+\lambda) \sum_{j=j_k}^{s} a'_{ij} Q_k(j-r/2) \int_{S^{d-1}} h_i(\xi)\, C_{2k}^{\lambda}(x\cdot\xi)\, d\xi \\
&= 2^r \sum_{(i,k)\in\Psi_s,\, k\leqslant s-r/2} (2k+\lambda)\, \gamma_k \sum_{j=j_k}^{s} a'_{ij} Q_k(j-r/2)\, P_{ik}(x).
\end{aligned}$$

Applying Lemma 1 and Parseval's identity we obtain

$$\|\mathscr{D}^r P_a\|_2^2 = 2^{2r} \sum_{(i,k)\in\Psi_s,\, k\leqslant s-r/2} \left| (2k+\lambda)\, \gamma_k \sum_{j=j_k}^{s} a'_{ij} Q_k(j-r/2) \right|^2. \tag{12}$$

In order to estimate this last sum we will need the following

PROPOSITION 1. *For any index pair* $(i,k)\in\Psi_s$ *such that* $k \leqslant s-r/2$ *the inequality*

$$\left| \sum_{j=j_k}^{s} a'_{ij} Q_k(j-r/2) \right| \leqslant c s^{r-1}$$

holds.

Proof. Since $a'_{ij} = \gamma_j^{-1} a_{ij} = (-1)^{\mu-\beta_j} \binom{\mu}{\beta_j} \varepsilon_{i\alpha_j}$ by (5), we have

$$S \equiv \sum_{j=j_k}^{s} a'_{ij} Q_k(j-r/2) = \sum_{j=j_k}^{s} (-1)^{\mu-\beta_j} \binom{\mu}{\beta_j} \varepsilon_{i\alpha_j} Q_k(j-r/2).$$

Define the integer l to satisfy

$$s-(\mu+1)(l-1) \leqslant j_k < s-(\mu+1)\,l, \tag{13}$$

and write the sum S as

$$S = \left(\sum_{j=j_k}^{s-(\mu+1)l} + \sum_{j=s-(\mu+1)l+1}^{s} \right) (-1)^{\mu-\beta_j} \binom{\mu}{\beta_j} \varepsilon_{i\alpha_j} Q_k(j-r/2) =: S_1 + S_2. \tag{14}$$

We first estimate S_1. From definition (10) of the function $Q_k(q)$ we have

$$Q_k(j-r/2)=\sum_{k_{r-1}=k}^{j-r/2}\ \sum_{k_{r-2}=k_{r-1}}^{j-r/2}\cdots\sum_{k_1=k_2}^{j-r/2}\ \prod_{i=1}^{r-1}(2k_i+\lambda+r-i)$$
$$\leqslant(j-r/2-k+1)^{r-1}\prod_{i=1}^{r-1}(2j+\lambda-i).$$

From here taking into consideration (13) we obtain for all $j\in\{k+r/2, ..., s-(\mu+1)l\}$

$$Q_k(j-r/2)\leqslant(\mu+2)^{r-1}\prod_{i=1}^{r-1}(2s+\lambda-i)\leqslant(\mu+2)^{r-1}(2s+\lambda-1)^{r-1}$$
$$\leqslant c_1 s^{r-1}.$$

Applying (13) once more we obtain

$$|S_1|=\left|\sum_{j=j_k}^{s-(\mu+1)l}(-1)^{\mu-\beta_j}\binom{\mu}{\beta_j}\varepsilon_{i\alpha_j}Q_k(j-r/2)\right|$$
$$\leqslant c_1 s^{r-1}\sum_{\beta=0}^{\mu}\binom{\mu}{\beta}=c_2 s^{r-1}. \tag{15}$$

We now show that $S_2=0$. Indeed, using the relation $s-j=(\mu+1)\alpha_j+\beta_j$, where $\alpha_j\in\mathbf{Z}$ and $\beta_j\in\{0, ..., \mu\}$ we have

$$S_2=\sum_{j=s-(\mu+1)l+1}^{s}(-1)^{\mu-\beta_j}\binom{\mu}{\beta_j}\varepsilon_{i\alpha_j}Q_k(j-r/2)$$
$$=\sum_{\alpha}\varepsilon_{i\alpha}\sum_{\beta=0}^{\mu}(-1)^{\mu-\beta}\binom{\mu}{\beta}Q_k(s-(\mu+1)\alpha-\beta-r/2), \tag{16}$$

where α run over some set of integers. From (10) it is seen that for any k the function $Q_k(j)$ in the variable j is a polynomial of degree $2(r-1)<2(r'-1)<\mu$. At the same time, the inner sum in (16) is a finite difference of the polynomial $Q_k(j)$ of order μ. Hence

$$\sum_{\beta=0}^{\mu}(-1)^{\mu-\beta}\binom{\mu}{\beta}Q_k(s-(\mu+1)\alpha+\beta-r/2)=0$$

for all α. Therefore from (16) we have $S_2=0$. From (14) and (15) we obtain Proposition 1. ▌

Proof of Lemma 2 (*continued*). From identity (12) and Proposition 1 it follows that

$$\|\mathscr{D}^r P_a\|_2^2 \leqslant 2^{2r} \sum_{(i,k)\in \Psi_s,\, k\leqslant s-r/2} [(2k+\lambda)\,\gamma_k c s^{r-1}]^2$$
$$\leqslant [c(2s+\lambda)\,\gamma_{s/2} s^{r-1}]^2\,|\Psi_s| \leqslant c s^{2r-1}\,|\Psi_s|.$$

From this and (4) follows the statement of Lemma 2.

Proof of Theorem 2. Let r' be the smallest even number such that $r' \geqslant r$. Applying the multiplicative inequality for derivatives (see [2]) we have

$$\|\mathscr{D}^r P_a\|_2 \leqslant \|P_a\|_2^{1-(r/r')}\,\|\mathscr{D}^{r'} P_a\|_2^{r/r'}.$$

Using Lemma 2 and inequality (7) we obtain

$$\|\mathscr{D}^r P_a\|_2 \leqslant (c s^{(d-1)/2})^{1-r/r'}\,(c s^{r'+(d-1)/2})^{r/r'} = c s^{r+(d-1)/2},$$

thus establishing Theorem 2. ∎

3. DECOMPOSITION OF RIDGE FUNCTIONS BY THE ORTHOGONAL SYSTEM OF RIDGE POLYNOMIALS

Consider the orthogonal system of polynomials $\Pi_s = \{P_{ij} = P_{ij}(x)\colon (i,j)\in \Psi_s\}$, constructed in Section 2. In this section we will show that for any polynomial $P \in \Pi_s$ and for any ridge function $g_\omega = g(\omega\cdot x)$, where $\omega = (\omega_1, ..., \omega_d) \in \mathbf{R}^d$ and $g \in L(\mathbf{R})$, the coefficient of the decomposition by the system Π_s corresponding to P, i.e.,

$$\langle g_\omega, P\rangle = \int_{B^d} g(\omega\cdot x)\,\overline{P(x)}\,dx$$

is a function $h(\omega, b)$ that is an algebraic polynomial in the variables $\omega_1, ..., \omega_d$, and a linear function of some variables $b_1, ..., b_m$ depending only on the function g.

Let $g(\omega\cdot x)$ be an arbitrary ridge function, where $\omega \in \mathbf{R}^d$, and $g \in L(\mathbf{R})$. Without loss of generality we can assume that $\omega \in S^{d-1}$.

Consider the group $SO(d)$ of all orthogonal matrices of order d with determinant equal to 1. For any vector $\omega \in S^{d-1}$ there exists a matrix $\sigma = \sigma_\omega \in SO(d)$ such that $\omega = \sigma e$, where $e = (1, 0, ..., 0)$. Elements of the matrix are denoted by $(\sigma_{ij})_{i,j=1}^d$.

Let $P(x)$ be any polynomial from the system Π_s. Using the orthogonality of the matrix σ and the invariance of the measure dx with respect to the orthogonal mapping we have

$$\langle g_\omega, P\rangle = \int_{B^d} g(\omega \cdot x)\, \overline{P(x)}\, dx = \int_{B^d} g(\sigma e \cdot x)\, \overline{P(x)}\, dx$$

$$= \int_{B^d} g(x_1)\, \overline{P(\sigma x)}\, dx. \tag{17}$$

Since $P(x)$ is a polynomial of degree $\leqslant 2s$, the integral (17) can be presented as

$$\int_{B^d} g(x_1)\, \overline{P(\sigma x)}\, dx = \sum_{|k| \leqslant 2s} p_k(\sigma; P) \int_{B^d} g(x_1)\, x^k\, dx, \tag{18}$$

where the functions $\sigma \to p_k(\sigma; P)$ are some polynomials of degree $\leqslant 2s$ in the d^2 variables σ_{ij}, $i, j = 1, ..., d$, and $k = (k_1, ..., k_d)$ is a multi-index, $|k| = k_1 + \cdots + k_d$, $x^k = x_1^{k_1} \cdots x_d^{k_d}$.

Fix the multi-index k. Consider the integral

$$\int_{B^d} g(x_1)\, x^k\, dx = \int_{B^d} g(x_1)\, x_1^{k_1} x_2^{k_2} \cdots x_d^{k_d}\, dx_1 \cdots dx_d$$

$$= \int_{-1}^{1} g(x_1)\, x_1^{k_1} dx_1 \int_{B_{x_1}} x_2^{k_2} \cdots x_d^{k_d}\, dx_2 \cdots dx_d, \tag{19}$$

where we denote $B_{x_1} = \{(x_2, ..., x_d) \in \mathbf{R}^{d-1} : x_2^2 + \cdots + x_d^2 \leqslant 1 - x_1^2\}$.

Transforming to polar coordinates in the inner integral, we obtain

$$\int_{B_{x_1}} x_2^{k_2} \cdots x_d^{k_d}\, dx_2 \cdots dx_d = \int_0^{\sqrt{1-x_1^2}} r^{d-2}\, dr \int_{S^{d-2}} (r\xi_2)^{k_2} \cdots (r\xi_d)^{k_d}\, d\xi$$

$$= \int_0^{\sqrt{1-x_1^2}} r^{k_2 + \cdots + k_d + d - 2}\, dr \int_{S^{d-2}} \xi_2^{k_2} \cdots \xi_d^{k_d}\, d\xi$$

$$= q_{k,d} (1 - x_1^2)^{(k_2 + \cdots + k_d + d - 1)/2},$$

where

$$q_{k,d} = \frac{1}{k_2 + \cdots + k_d + d - 1} \int_{S^{d-2}} \xi_2^{k_2} \cdots \xi_d^{k_d}\, d\xi$$

is a constant depending only on the vector k and d. Therefore from (19), applying the change of variable $x_1 = \cos t$, we obtain

$$\int_{B^d} g(x_1)\, x^k\, dx = q_{k,d} \int_{-1}^{1} g(x_1)\, x_1^{k_1} (1 - x_1^2)^{(k_2 + \cdots + k_d + d - 1)/2}\, dx_1$$

$$= q_{k,d} \int_0^{\pi} g(\cos t)(\cos t)^{k_1}\, (\sin t)^{k_2 + \cdots + k_d + d}\, dt. \tag{20}$$

Introduce the notation

$$b_l(g) = \begin{cases} \displaystyle\int_0^{\pi} g(\cos t) \cos(lt)\, dt, & l \text{ even} \\ \displaystyle\int_0^{\pi} g(\cos t) \sin(lt)\, dt, & l \text{ odd.} \end{cases}$$

Then the intral (20) can be rewritten as

$$\int_{B^d} g(x_1)\, x^k\, dx = \sum_{l=0}^{|k|+d} c_{kl} b_l(g),$$

where the coefficients c_{kl} depend only on k and l. Substituting this final expression into (18) and changing the order of summation we obtain

$$\int_{B^d} g(x_1)\, P(\sigma x)\, dx = \sum_{|k| \leqslant 2s} p_k(\sigma; P) \sum_{l=0}^{|k|+d} c_{kl} b_l(g)$$

$$= \sum_{l=0}^{2s+d} b_l(g) \sum_{k:\, l-d \leqslant |k| \leqslant 2s} c_{kl}\, p_k(\sigma; P).$$

From this and identity (17) we conclude

THEOREM 3. *Let* $\omega = \sigma e \in S^{d-1}$, $g \in L(\mathbf{R})$, *let* s *be a natural number, and* $P \in \Pi_s$. *Then the inner product of the functions* $g_\omega = g(\omega \cdot x)$ *and* $P = P(x)$ *may be expressed as*

$$\langle g_\omega, P \rangle = \sum_{l=0}^{2s+d} b_l(g)\, \rho_l(\sigma; P),$$

where

$$\rho_l(\sigma; P) = \sum_{k:\, l-d \leqslant |k| \leqslant 2s} c_{kl}\, p_k(\sigma; P),$$

$p_k(\sigma; P)$ *are some polynomials of degree* $\leqslant 2s$ *in the variables* σ_{ij}, $i, j = 1, ..., d$, *and* c_{kl} *depend only on* k *and* l.

4. SOME ESTIMATES IN THE FINITE-DIMENSIONAL SPACE

In this section we obtain estimates for the approximation of sets in a finite-dimensional space by certain polynomials manifolds. To obtain the lower bounds of approximation, we use estimates of the number of connected components of polynomial manifolds. These estimates are related to the problem of calculation of the pseudo-dimension of manifolds, which is widely studied within the neural network community (see [25, 27, 8, 12, 23]).

We start with some notations. Let m be a fixed natural number. Consider the m-dimensional Hilbert space l_2^m consisting of vectors $a=(a_1, ..., a_m)\in \mathbf{R}^m$ with the norm $\|a\|_{l_2^m}=(\sum_{i=1}^m |a_i|^2)^{1/2}$. Let H be some set in l_2^m. Define the distance of a point a from the set H as $dist(a, H, l_2^m)=\inf_{h\in H}\|a-h\|_{l_2^m}$. Introduce the vector set E^m consisting of all vectors $\varepsilon=(\varepsilon_1, ..., \varepsilon_m)$ with coordinates $\varepsilon_i\in\{1, -1\}$ for all $i=1, ..., m$.

Let m, s, p and q be natural numbers. Let $\pi_{ij}(\sigma)$, $i=1, ..., m$; $j=1, ..., q$ be any algebraic polynomials with real coefficients in the variables $\sigma=(\sigma_1, ..., \sigma_p)\in\mathbf{R}^p$, each of degree s. Construct the polynomials in the $p+q$ variables $b=(b_1, ..., b_q)\in\mathbf{R}^q$ and $\sigma=(\sigma_1, ..., \sigma_p)\in\mathbf{R}^p$

$$\pi_i(b, \sigma)=\sum_{j=1}^{q} b_j\pi_{ij}(\sigma), \qquad i=1, ..., m. \tag{21}$$

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

$$\Pi_{m, s, p, q}=\{\pi(b, \sigma)=(\pi_1(b, \sigma), ..., \pi_m(b, \sigma)): (b, \sigma)\in\mathbf{R}^q\times\mathbf{R}^p\}.$$

THEOREM 4. *Let m, s, p, q be integers such that $p+q\leqslant m/2$ and*

$$p\log_2(4s)+(p+1)\log_2(p+q+2)+(p+q)\log_2\left(\frac{2em}{p+q}\right)\leqslant m/4. \tag{22}$$

Then there exists a vector ε in E^m an some absolute constant $c>0$ such that

$$dist(\varepsilon, \Pi_{m, s, p, q}, l_2^m)\geqslant c\sqrt{m}.$$

We first prove auxiliary statements. For a given vector $x=(x_1, ..., x_m)\in\mathbf{R}^m$ we define the vector $sgn\, x=(sgn\, x_1, ..., sgn\, x_m)\in E^m$, where $sgn\, a=1$ for $a\geqslant 0$, and $sgn\, a=-1$ for $a<0$. For a set $K\subset\mathbf{R}^m$, denote by $sgn\, K$ a set of vectors $\{sgn\, x: x\in K\}$.

LEMMA 3. *Let m, s, p, q be integers such that $p+q \leqslant m/2$. Then for the cardinality of the set $sgn\, \Pi_{m,s,p,q}$ the following estimate holds*

$$|sgn\, \Pi_{m,s,p,q}| \leqslant (4s)^p\, (p+q+1)^{p+2} \left(\frac{2em}{p+q}\right)^{p+q}.$$

Assume for the moment that we have proved Lemma 3.

Proof of Theorem 4. Let $a<1$ be the absolute constant satisfying the equation $1-\frac{1}{2}(1-a)^2 \log_2 e = \frac{47}{67}$ (i.e., $a = 0.39...$).

Denote $H = sgn\, \Pi_{m,s,p,q}$. Let $h=(h_1, ..., h_m)$ be any vector from H. Define the subset of E^m

$$E_h = \left\{ \varepsilon \in E^m : \sum_{i=1}^{m} |\varepsilon_i - h_i|^2 \leqslant 2am \right\}.$$

Since $h_i = 1$ or -1, we have the estimate for the cardinality of E_h,

$$\begin{aligned} |E_h| &= \left| \left\{ \varepsilon \in E^m : \sum_{i=1}^{m} (\varepsilon_i(-h_i)+1)^2 \leqslant 2am \right\} \right| \\ &= \left| \left\{ \varepsilon \in E^m : \sum_{i=1}^{m} (\varepsilon_i+1)^2 \leqslant 2am \right\} \right| \\ &= \left| \left\{ \varepsilon : \sum_{i:\varepsilon_i=1} 1 \leqslant am/2 \right\} \right| = \sum_{i=0}^{[am/2]} \binom{m}{i}. \end{aligned}$$

From the well-known estimate (see, for example, [5, Chap. 8]) we have

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

where $\beta = m^{-1}[am/2]$, and $b = 1-\frac{1}{2}(1-a)^2 \log_2 e = \frac{47}{64}$. Hence $|E_h| \leqslant 2^{47m/64}$.

Consider in E^m the subset $E' = \bigcap_{h \in H} (E^m \backslash E_h)$. We estimate the cardinality of E' via

$$\begin{aligned} |E'| &= \left| E^m \Big\backslash \bigcup_{h \in H} E_h \right| \geqslant 2^m - |H| \max_{h \in H} |E_h| \\ &\geqslant 2^m - |H|\, 2^{(47/64)m}. \end{aligned} \tag{23}$$

By Lemma 3 and the conditions of Theorem 4 we have

$$|H| \leqslant (4s)^p\, (p+q+1)^{p+2} \left(\frac{2em}{p+q}\right)^{p+q} \leqslant 2^{m/4}.$$

From this and (23) we obtain $|E'| \geqslant 2^m - 2^{(63/64)m} > 0$. Therefore there exist a vector $\varepsilon \in E^m$ such that for every vector $h \in H$ the following inequality holds

$$\|\varepsilon - h\|_{l_2^m} \geqslant \sqrt{2am}.$$

Taking into consideration the fact that every vector h has the form $h = sgn\, \pi(b, \sigma)$, where $\pi(b, \sigma) \in \Pi_{m, s, p, q}$ we obtain the inequality

$$\|\varepsilon - \pi(b, \sigma)\|_{l_2^m} \geqslant \frac{1}{2} \|\varepsilon - sgn\, \pi(b, \sigma)\|_{l_2^m} \geqslant \sqrt{\frac{am}{2}}.$$

Theorem 4 is proved. ∎

We now prove Lemma 3. Set $v = p + q$. The set of points $(b, \sigma) \in \mathbf{R}^v$ on which the polynomial $\pi(b, \sigma)$ vanishes will be denoted by $Z(\pi)$.

Consider the domains $D = \bigcup_{i=1}^m Z(\pi_i)$ in $\mathbf{R}^v$, and $D' = \mathbf{R}^v \backslash D$. We will need an estimate on the number of connected components of the set D'. We denote the number of connected components of any $G \subset \mathbf{R}^v$ by $N(G)$. The following lemma is a direct consequence of a result of Warren [27, Theorem 1, 2].

Lemma 4 (Warren). *For any polynomials $\pi_1(b, \sigma), ..., \pi_m(b, \sigma)$ of the form* (21) *there exist positive numbers $\delta_1, ..., \delta_m$ such that the number of connected components of the set D' satisfies the inequality*

$$N(D') \leqslant \sum_{k=1}^{v} \sum_{i_1 < \cdots < i_k} \sum_{\varepsilon_1, ..., \varepsilon_k = \pm 1} N\left(\bigcap_{l=1}^{k} Z(\pi_{i_l} + \varepsilon_l \delta_{i_l}) \right),$$

where the indices $i_1, ..., i_k$ run over all different integers between 1 *and m, and $\{\varepsilon_1, ..., \varepsilon_k\}$ assume all values in $\{1, -1\}^k$.*

Let $g_1(\sigma), ..., g_n(\sigma)$ be any polynomials in the variable $\sigma \in \mathbf{R}^p$. Let $deg\, g_i$ be the degree of the polynomial g_i. Set $d = deg\, g_i + \cdots + deg\, g_n$.

Consider in $\mathbf{R}^p$ the manifold $S = \{\sigma: g_1(\sigma) \geqslant 0, ..., g_n(\sigma) \geqslant 0\}$. Let $N(S)$ be the number of connected components of the manifold S. From the work of Milnor [14, Theorem 3] it directly follows[1]

[1] It is known (see, for example, [7]), that the number of connected components, $N(S)$, is at most, $rank\, H^*(S)$, the rank of full cohomology group $H^*(S)$. Milnor obtained the estimate $rank\, H^*(S) \leqslant \frac{1}{2}(2 + d)(1 + d)^{p-1}$.

LEMMA 5 (Milnor). *If $g_1, ..., g_n$ are any polynomials in variable $\sigma \in \mathbf{R}^p$, then*

$$N(S) \leqslant \tfrac{1}{2}(2+d)(1+d)^{p-1}.$$

Using Lemma 5 we will also estimate the numbers $N(\bigcap_{l=1}^k Z(\pi_{i_l}+\varepsilon_l \delta_{i_l}))$. Fix k, $i_1, ..., i_k$ and $\varepsilon_1, ..., \varepsilon_k$. Without loss of generality we may take $i_l = l$, and $\varepsilon_l = -1$ for all $l = 1, ..., k$.

LEMMA 6. *For any $1 \leqslant k \leqslant v$ and any positive numbers $\delta_1, ..., \delta_k$, the following inequality holds*

$$N\left(\bigcap_{l=1}^{k} Z(\pi_l - \delta_l)\right) \leqslant (4s)^p\,(k+1)^{p+1}.$$

Proof. Consider the system of k linear equations

$$\begin{cases} \pi_{11}(\sigma)\, b_1 + \cdots + \pi_{1q}(\sigma)\, b_q = \delta_1 \\ \qquad\vdots \\ \pi_{k1}(\sigma)\, b_1 + \cdots + \pi_{kq}(\sigma)\, b_q = \delta_k \end{cases} \tag{24}$$

in the variables $b_1, ..., b_q$, with coefficients $\pi_{ij}(\sigma)$ and constants δ_i.

For any fixed σ, denote by $D_l(\sigma)$ the sum of squares of all minors of order l of the matrix$(\pi_{ij}(\sigma))_{i=1}^{k}{}_{j=1}^{q}$, and by $\overline{D_{l+1}}(\sigma)$ the sum of squares of all minors of order $l+1$ of the extended matrix

$$\begin{pmatrix} \pi_{11}(\sigma) & \cdots & \pi_{1q}(\sigma) & \delta_1 \\ \vdots & \ddots & \vdots & \vdots \\ \pi_{k1}(\sigma) & \cdots & \pi_{kq}(\sigma) & \delta_k \end{pmatrix}.$$

It follows from a theorem of Kronecker–Capelli that the set $\mathscr{S}$ of all vectors $\sigma \in \mathbf{R}^p$ for which there exists a solution to the system (24) may be expressed as

$$\mathscr{S} = \bigcup_{l=1}^{k-1} \mathscr{S}_l, \qquad \mathscr{S}_l = \{\sigma \in \mathbf{R}^p : D_l(\sigma) > 0, \overline{D_{l+1}}(\sigma) = 0\}. \tag{25}$$

Since the $\pi_{ij}(\sigma)$ are polynomials of degree s then $D_l(\sigma)$ and $\overline{D_{l+1}}(\sigma)$ are polynomials of degree $2sl$ and $2s(l+1)$, respectively. From the continuous dependence of the solutions of (24) on the coefficients $\pi_{ij}(\sigma)$, where σ runs over some connected component of $\mathscr{S}_l$, it follows that $N(\bigcap_{l=1}^k Z(\pi_l - \delta_l)) = N(\mathscr{S})$. From (25) we have

$$N\left(\bigcap_{l=1}^{k} Z(\pi_l - \delta_l)\right) = N(\mathscr{S}) \leqslant \sum_{l=1}^{k-1} N(\mathscr{S}_l). \tag{26}$$

Note that for any l a set $\mathcal{S}_l$ may be presented as the set of solutions of the system of inequalities

$$D_l(\sigma) > 0, \qquad -\overline{D_{l+1}}(\sigma) \geqslant 0. \tag{27}$$

We claim that $N(\mathcal{S}_l) \leqslant (4s(k+1))^p$. Indeed, assume that $N(\mathcal{S}_l) \geqslant m+1$, where $m = (4s(k+1))^p$. Then there exist $m+1$ disjoint components $Q_1, ..., Q_{m+1}$ of the set $\mathcal{S}_l$. For each $1 \leqslant i \leqslant m$, choose a point σ_i in Q_i. Put $\alpha = \min_i D_l(\sigma_i)$, and note that $\alpha > 0$.

Consider the set in $\mathbf{R}^p$

$$\mathcal{S}'_l = \{\sigma \in \mathbf{R}^p : D_l(\sigma) \geqslant \alpha, \ -\overline{D_{l+1}}(\sigma) \geqslant 0\}.$$

From Lemma 5 it follows that the number of connected components of the set $\mathcal{S}'$ satisfies the inequality $N(\mathcal{S}'_l) \leqslant \frac{1}{2}(2+d)(1+d)^{p-1}$, where $d = \deg D_l + \deg \overline{D_{l+1}}$. Since $d \leqslant 2sl + 2s(l+1) \leqslant 4s(k+1)$ we have

$$N(\mathcal{S}'_l) \leqslant \tfrac{1}{2}(2+4s(k+1))(1+4s(k+1))^{p-1} \leqslant (4s(k+1))^p = m.$$

On the other hand, since $\mathcal{S}'_l \cap Q_i \neq \varnothing$ for all $i = 1, ..., m+1$, and $Q_1, ..., Q_{m+1}$ do not intersect, then

$$N(\mathcal{S}'_l) \geqslant N(\mathcal{S}_l) \geqslant m+1$$

yielding a contradiction. Hence $N(\mathcal{S}_l) \leqslant (4s(k+1))^p$. From here and (26) we obtain

$$N\left(\bigcap_{l=1}^{k} Z(\pi_l - \delta_l)\right) \leqslant \sum_{l=1}^{k-1} N(\mathcal{S}_l) \leqslant \sum_{l=1}^{k-1} (4s(k+1))^p \leqslant (4s)^p\,(k+1)^{p+1}.$$

Lemma 6 is proved. ▮

From Lemmas 4 and 6 one obtains the following estimate for the number of connected components of the set $D' = \mathbf{R}^\nu \setminus \bigcup_{i=1}^m Z(\pi_i)$

$$\begin{aligned} N(D') &\leqslant \sum_{k=1}^{\nu} \sum_{i_1 < \cdots < i_k} \sum_{\varepsilon_1, ..., \varepsilon_k = \pm 1} (4s)^p\,(k+1)^{p+1} \\ &\leqslant (4s)^p\,(p+q+1)^{p+1} \sum_{k=1}^{\nu} \binom{m}{k} 2^k \\ &\leqslant 2^\nu (4s)^p\,(p+q+1)^{p+1} \sum_{k=1}^{\nu} \frac{m^k}{k!} \leqslant 2^\nu (4s)^p\,(p+q+1)^{p+1}\,\nu \left(\frac{em}{\nu}\right)^\nu, \end{aligned}$$

where we make use of the relation $\nu = p+q \leqslant m/2$. Hence we obtain

LEMMA 7. *Let m, s, p, q be natural numbers, $p+q \leqslant m/2$. Then for any polynomials $\pi_1(b, \sigma), ..., \pi_m(b, \sigma)$ of the form* (21) *in the variables $(b, \sigma) \in \mathbf{R}^{q+p}$ the following estimate for the number of connected components of the set $D' = \mathbf{R}^{q+p} \setminus \bigcup_{i=1}^m Z(\pi_i)$*

$$N(D') \leqslant (4s)^p \, (p+q+1)^{p+2} \left(\frac{2em}{p+q} \right)^{p+q}$$

holds.

Proof of Lemma 3. The statement of Lemma 3 follows directly from Lemma 7. Indeed, note that a vector function $sgn\, \pi(y)$ is constant on any connected component of the set D'. Therefore the cardinality of the vector set $sgn\, \Pi_{m, s, p, q}$ is at most $N(D')$. From this and Lemma 7 we obtain Lemma 3. ▮

5. PROOF OF THEOREM 1

Proof of the Lower Bound in Theorem 1. Let n be any natural number and s be any even number such that $\tau n \leqslant s^{d-1} \leqslant 2\tau n$, where τ is some positive constant depending only on d, which we define below. We have the asymptotice $n \asymp s^{d-1}$. Consider the set of polynomials

$$\mathscr{P}(A_s^r) = \left\{ \sum_{(i,j) \in \Psi_s} a_{ij} P_{ij}(x) \colon a = (a_{ij})_{(i,j) \in \Psi_s} \in A_s^r \right\}, \qquad (28)$$

introduced in Section 2.

Estimate the distance of the set $\mathscr{P}(A_s^r)$ from the manifold M_n

$$dist(\mathscr{P}(A_s^r), M_n, L_2) = \max_{a \in A_s^r} \inf_{g \in M_n} \|P_a(x) - g(x)\|_2.$$

Let $P_a(x) = \sum_{(i,j) \in \Psi_s} a_{ij} P_{ij}(x)$ be an arbitrary function from $\mathscr{P}(A_s^r)$ and let

$$g(x) = \sum_{t=1}^{n} g_t(\omega_t \cdot x), \qquad (\omega_t \in S^{d-1}, g_t \in L(\mathbf{R}) \text{ for all } t)$$

be an arbitrary function from the manifold M_n.

By Lemma 1, the system of polynomials $\{P_{ij}\}_{(i,j)\in\Psi_s}$ is orthonormal. Therefore

$$\begin{aligned}\|P_a(x)-g(x)\|_2^2 &= \Big\| \sum_{(i,j)\in\Psi_s} a_{ij}P_{ij}(x)-g(x)\Big\|_2^2 \\ &\geqslant \Big\| \sum_{(i,j)\in\Psi_s} a_{ij}P_{ij}(x)-\sum_{(i,j)\in\Psi_s}\langle g, P_{ij}\rangle P_{ij}(x)\Big\|_2^2 \\ &= \Big\| \sum_{(i,j)\in\Psi_s} (a_{ij}-\langle g, P_{ij}\rangle)\, P_{ij}(x)\Big\|_2^2 \\ &= \sum_{(i,j)\in\Psi_s} |a_{ij}-\langle g, P_{ij}\rangle|^2. \qquad (29)\end{aligned}$$

Fix indices i, j, and consider the inner product $\langle g, P_{ij}\rangle$. According to Theorem 3 we have

$$\langle g, P_{ij}\rangle = \sum_{t=1}^{n} \langle g_t(\omega_t\cdot\bullet), P_{ij}\rangle = \sum_{t=1}^{n}\sum_{l=0}^{2s+d} b_l(g_t)\, \rho_l(\sigma^t; P_{ij}), \qquad (30)$$

where $\sigma^t=(\sigma^t_{ij})^d_{i,j=1}$ are orthogonal matrices from the group $SO(d)$ for which $\omega_t=\sigma^t e$, $e=(1, 0, ..., 0)$, and $\rho_l(\sigma^t; P_{ij})$ are some polynomials of degree not greater than $2s$ in the variables σ^t_{ij}, $i, j=1, ..., d$. From (30) it follows that

$$\begin{aligned}&\inf_{g\in M_n} \sum_{(i,j)\in\Psi_s} |a_{ij}-\langle g, P_{ij}\rangle|^2 \\ &\quad= \inf_{\{\sigma^t\},\{g_t\}} \sum_{(i,j)\in\Psi_s} \Big| a_{ij}-\sum_{t=1}^{n}\sum_{l=0}^{2s+d} b_l(g_t)\,\rho_l(\sigma^t; P_{ij})\Big|^2, \qquad (31)\end{aligned}$$

where the infimum is calculated over all collections of matrices $\sigma^1, ..., \sigma^n \in SO(d)$ and functions $g_1, ..., g_n \in L(\mathbf{R})$.

Set $p=nd^2$, and $q=n(2s+d+1)$. Enumerate arbitrarily all elements of the matrices $\sigma^1, ..., \sigma^n$ to form the vector $\sigma=(\sigma_1, ..., \sigma_p)$. For fixed i, j we also enumerate the set of polynomials to form $\{\rho_l(\sigma^t; P_{ij})\colon t=1, ..., n;\ l=0, ..., 2s+d\}$, and denote them by $\{\rho_k(\sigma; P_{ij})\colon k=1, ..., q\}$. Then from (31) it follows that

$$\inf_{g\in M_n} \sum_{(i,j)\in\Psi_s} |a_{ij}-\langle g, P_{ij}\rangle|^2 \geqslant \inf_{(b,\sigma)\in\mathbf{R}^q\times\mathbf{R}^p} \sum_{(i,j)\in\Psi_s} \Big| a_{ij}-\sum_{k=1}^{q} b_k\, \rho_k(\sigma; P_{ij})\Big|^2, \qquad (32)$$

where $b=(b_1, ..., b_q)$ and $\sigma=(\sigma_1, ..., \sigma_p)$ run over $\mathbf{R}^q$ and $\mathbf{R}^p$, respectively.

Using relations (29), (31), and (32) we obtain

$$dist(\mathscr{P}(A_s^r), M_n, L_2)^2 \geqslant \max_{a \in A_s^r} \inf_{g \in M_n} \sum_{(i,j) \in \Psi_s} |a_{ij} - \langle g, P_{ij} \rangle|^2$$

$$\geqslant \max_{a \in A_s^r} \inf_{(b,\sigma) \in \mathbf{R}^q \times \mathbf{R}^p} \sum_{(i,j) \in \Psi_s} \left| a_{ij} - \sum_{k=1}^{q} b_k \rho_k(\sigma; P_{ij}) \right|^2. \tag{33}$$

Set $\pi_{ij}(b, \sigma) = \sum_{k=1}^{q} b_k \rho_k(\sigma; P_{ij})$, and consider the sum

$$I(a, b, \sigma) := \sum_{(i,j) \in \Psi_s} |a_{ij} - \pi_{ij}(b, \sigma)|^2. \tag{34}$$

Recall that $a_{ij} = (-1)^{\mu - \beta_j} \binom{\mu}{\beta_j} \gamma_j \varepsilon_{i, \alpha_j}$, where μ is some integer depending only on r, and the numbers α_j and β_j are defined for a given j from the relation $s - j = (\mu+1)\alpha_j + \beta_j$, where $\alpha_j \in \mathbf{Z}$, $\beta_j \in \{0, ..., \mu\}$, while the numbers $\varepsilon_{i, \alpha_j}$ equal $+1$ or -1.

Consider a subset $\Psi_s^0 = \{(i,j) \in \Psi_s : \beta_j = 0\}$ in Ψ_s. Estimate the quantity (34)

$$I(a, b, \sigma) = \sum_{(i,j) \in \Psi_s} \left| (-1)^{\mu - \beta_j} \binom{\mu}{\beta_j} \gamma_j \varepsilon_{i, \alpha_j} - \pi_{ij}(b, \sigma) \right|^2$$

$$= \sum_{(i,j) \in \Psi_s} \left| (-1)^{\mu - \beta_j} \binom{\mu}{\beta_j} \gamma_j \varepsilon_{i, \alpha_j} - \pi_{i, s-(\mu+1)\alpha_j - \beta_j}(b, \sigma) \right|^2$$

$$\geqslant \sum_{(i,j) \in \Psi_s^0} |(-1)^{\mu} \gamma_j \varepsilon_{i, \alpha_j} - \pi_{i, s-(\mu+1)\alpha_j}(b, \sigma)|^2.$$

Since $\gamma_j^2 \asymp j^{-1} \geqslant 2s^{-1}$ for all $(i,j) \in \Psi_s$, we have

$$I(a, b, \sigma) \geqslant 2s^{-1} \sum_{(i,j) \in \Psi_s^0} |\varepsilon_{i, \alpha_j} - (-1)^{\mu} \gamma_j^{-1} \pi_{i, s-(\mu+1)\alpha_j}(b, \sigma)|^2. \tag{35}$$

Set $m = |\Psi_s^0|$. From the definition of the numbers β_j and (4) we have the asymptotics $m \asymp |\Psi_s|/(\mu+1) \asymp s^d$.

Since on the set of indices $(i,j) \in \Psi_s^0$ we have $j = s - (\mu+1)\alpha_j$, i.e. the relation between j and α_j is one-to-one, the set of vectors $\{(\varepsilon_{i,\alpha_j})_{(i,j) \in \Psi_s^0} : \varepsilon_{i,\alpha_j} \in \{-1, +1\}$ *for all* $(i,j) \in \Psi_s^0\}$ coincides with the set

$$\{(\varepsilon_{i,j})_{(i,j) \in \Psi_s^0} : \varepsilon_{i,j} \in \{-1, +1\} \text{ for all } (i,j) \in \Psi_s^0\}$$

$$= \{\varepsilon = (\varepsilon_1, ..., \varepsilon_m) : \varepsilon_1, ..., \varepsilon_m \in \{-1, +1\}\} =: E^m.$$

Denote the elements of the set

$$\{(-1)^{\mu}\gamma_j^{-1}\pi_{i,s-(\mu+1)\alpha_j}(b,\sigma):(i,j)\in\Psi_s^0\}$$

by $(\pi_1(b,\sigma),...,\pi_m(b,\sigma))$. Then from the inequality (35) we have

$$\max_{a\in A_s^{\mu}}\inf_{(b,\sigma)\in\mathbf{R}^q\times\mathbf{R}^p} I(a,b,\sigma)\geqslant 2s^{-1}\max_{\varepsilon\in E^m}\inf_{(b,\sigma)\in\mathbf{R}^q\times\mathbf{R}^p}\sum_{i=1}^{m}|\varepsilon_i-\pi_i(b,\sigma)|^2. \tag{36}$$

From relations (33), (34), and (36) we have

$$\begin{aligned} dist(\mathscr{P}(A_s^r),M_n,L_2)^2 &\geqslant \max_{a\in A_s^{\mu}}\inf_{(b,\sigma)\in\mathbf{R}^q\times\mathbf{R}^p} I(a,b,\sigma)\\ &\geqslant 2s^{-1}\max_{\varepsilon\in E^m}\inf_{(b,\sigma)\in\mathbf{R}^q\times\mathbf{R}^p}\sum_{i=1}^{m}|\varepsilon_i-\pi_i(b,\sigma)|^2. \end{aligned}$$

Recall that

$$\tau n\leqslant s^{d-1}\leqslant 2\tau n,\qquad p=nd^2,\qquad q=n(2s+d+1),\qquad c_1s^d\leqslant m\leqslant c_2s^d.$$

Set $\tau=200d^5$. It is easy to verify that the inequality (22) is satisfied. Thus from Theorem 4 we obtain

$$\max_{\varepsilon\in E^m}\inf_{(b,\sigma)\in\mathbf{R}^q\times\mathbf{R}^p}\sum_{i=1}^{m}|\varepsilon_i-\pi_i(b,\sigma)|^2\geqslant cm\asymp s^d.$$

Hence there exists a polynomial $P_a\in\mathscr{P}(A_s^{\mu})$ such that

$$dist(P_a,M_n,L_2)\geqslant cs^{(d-1)/2}.$$

According to Theorem 2 the function $\hat{P}_a(x)=(c/s^{r+(d-1)/2})\,P_a(x)$ belongs to class $W_2^{r,d}$. Therefore taking into consideration that $s\asymp n^{1/(d-1)}$, we obtain

$$dist(W_2^{r,d},M_n,L_2)\geqslant dist(\hat{P}_a,M_n,L_2)\geqslant c\,\frac{1}{s^{r+(d-1)/2}}\,s^{(d-1)/2}\asymp\frac{1}{n^{r/(d-1)}}.$$

Proof of the Upper Bound in Theorem 1. Now we prove the upper bound in Theorem 1. Let s be any natural number. Consider the space $\mathscr{P}_s$ of all polynomials of degree not greater than s, and the space $\mathscr{P}_s^h$ of homogeneous polynomials of degree s.

PROPOSITION 2. *If* $n=\dim\mathscr{P}_s^h$ *then* $\mathscr{P}_s\subset M_n$.

Proof. Let $a_1, ..., a_n$ be n vectors in $\mathbf{R}^d$ for which the $\{(a_i \cdot x)^s\}_{i=1}^n$ are linearly independent. Since each $(a_i \cdot x)^s$ is in $\mathcal{P}_s^h$, and $n = \dim \mathcal{P}_s^h$, these functions span $\mathcal{P}_s^h$.

For each $0 \leqslant k \leqslant s$ the $\{(a_i \cdot x)^k\}_{i=0}^s$ span $\mathcal{P}_k$. This is immediate and may be proved as follows. Every polynomial in $\mathcal{P}_k^h$ may be obtained as an appropriate derivative of order $s-k$ of a polynomial in $\mathcal{P}_s^h$. Every derivative of order $s-k$ of $(a_i \cdot x)^s$ is a constant times $(a_i \cdot x)^k$. Thus the set $\{(a_i \cdot x)^k\}$ spans $\mathcal{P}_k^h$ for every $0 \leqslant k < m$.

This implies that

$$\mathcal{P}_s = \left\{ \sum_{i=1}^n p_i(a_i \cdot x) \colon p_i \text{ is polynomial of degree } \leqslant s \right\}.$$

The space on the right hand side is a subspace of M_n. Thus $\mathcal{P}_s \subset M_n$.

Now we prove the upper bound in Theorem 1. First we formulate a well-known result (see [22]): the error in the best approximation of any function $f \in W_2^{r,d}$ from the polynomial space $\mathcal{P}_s$ in the L_2-norm is bounded above as follows,

$$dist(f, \mathcal{P}_s, L_2) \leqslant cs^{-r}.$$

Since $n = \dim \mathcal{P}_s^h$ then $n \asymp s^{d-1}$. Hence from Proposition 2 it follows

$$dist(W_2^{r,d}, M_n, L_2) \leqslant dist(W_2^{r,d}, \mathcal{P}_s, L_2) \leqslant cs^{-r} \asymp n^{-r/(d-1)}.$$

The upper bound is established, and Theorem 1 is proved. ∎

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 $(a)_0 := 0$, $(a)_n := a(a+1)\cdots(a+N-1)$.

(2) There is an identity which relates the Gegenbauer polynomials to its derivatives (see [6])

$$\frac{d^m C_n^\lambda}{dt^m}(t) = 2^m(\lambda)_m\, C_{n-m}^{\lambda+m}(t) \qquad (m = 1, 2, ..., n). \tag{A.2}$$

(3) The following relation between contiguous Gegenbauer polynomials holds (see [6])

$$(m+\lambda)\, C_{m+1}^{\lambda-1} = (\lambda-1)(C_{m+1}^\lambda - C_{m-1}^\lambda) \qquad (\lambda > 1),$$

and $C_0^\lambda(t) = 1$. Set $m = 2n-1$. This identity readily implies

$$C_{2n}^\lambda = \frac{2n+\lambda-1}{\lambda-1}\, C_{2n}^{\lambda-1} + C_{2n-2}^\lambda,$$

from which we obtain by induction

$$C_{2n}^\lambda = \sum_{k=0}^{n} \frac{2k+\lambda-1}{\lambda-1}\, C_{2k}^{\lambda-1}. \tag{A.3}$$

(4) Let $\mathscr{P}_n$ 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 $v_n = \pi^{1/2}(d)_n\, \Gamma((d+1)/2)/(n+d/2)\, n!\, \Gamma(d/2)$. The polynomials $u_n(\xi \cdot x)$, $\xi \in S^{d-1}$, are in $\mathscr{P}_n$ and the $u_n(\xi \cdot x)$ are orthogonal to $\mathscr{P}_{n-1}$ in $L_2(B^d)$ (see [18]):

$$\int_{B^d} u_n(\xi \cdot x)\, p(x)\, dx = 0 \qquad \forall \xi \in S^{d-1} \quad \text{and} \quad \forall p \in \mathscr{P}_{n-1}. \tag{A.4}$$

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

$$\int_{B^d} u_n(\xi \cdot x)\, u_n(\eta \cdot x)\, dx = \frac{u_n(\xi \cdot \eta)}{u_n(1)}. \tag{A.5}$$

(6) 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])

$$\int_{S^{d-1}} h(\xi)\, u_n(\xi\cdot\eta)\, d\xi=\frac{u_n(1)}{v_n}\, h(\eta), \qquad \text{where} \quad v_n=\frac{(n+1)_{d-1}}{2(2\pi)^{d-1}}. \tag{A.6}$$

An Orthogonal System of Polynomials on the Sphere

We state some facts (see [6, 24]) 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 $x_1, ..., x_d$. Any polynomial from $\mathcal{H}_s$ is decomposable by a linear combination of polynomials of the form

$$h_{sk}(x)=A_{sk}\prod_{j=0}^{d-2} r_{d-j}^{k_j-k_{j-1}+1}\, C_{k_j-k_{j+1}}^{(d-j-2)/2+k_{j+1}}\left(\frac{x_{d-j}}{r_{d-j}}\right)(x_2\pm ix_1)^{k_{d-2}}, \tag{A.7}$$

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

$$K^s=\{k=(k_0, k_1, ..., k_{d-3}, \varepsilon k_{d-2})\colon 0\leqslant k_{d-2}\leqslant\cdots\leqslant k_1\leqslant k_0=s,\ \varepsilon=\pm 1\},$$

and A_{sk} is the normalization factor

$$A_{sk}=\frac{1}{\Gamma(d/2)} \times\prod_{j=0}^{d-3}\frac{2^{2k_{j+1}+d-j-4}(k_j-k_{j+1})(d-j+2k_j-2)\,\Gamma^2((d-j-2)/2+k_{j+2})}{\sqrt{\pi}\,\Gamma(k_j+k_{j+1}+d-j-2)}.$$

It is known that the dimension of the space $\mathcal{H}_s$ is given by

$$\dim\mathcal{H}_s=|K^s|=\binom{s+d-1}{s}-\binom{s+d-3}{s-2}, \tag{A.8}$$

if $s\geqslant 2$, 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

$$\dim\mathcal{H}_s=\left(2+\frac{2}{(d-2)!}+c(s, d)\right) s(s+1)\cdots(s+d-3)\asymp s^{d-2}, \tag{A.9}$$

where $0\leqslant c(s, d)\leqslant 1$ 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 multi-indices $k, k' \in K^s$, the following holds

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

Note that the spaces $\mathscr{H}_s$ and $\mathscr{H}_{s'}$ for $s \neq s'$ are orthogonal with respect to the inner product (1). 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 $\{p: p \in \mathscr{P}_n\}$ of degree $\leqslant n$ belongs to the space $\mathscr{H}_0 \oplus \mathscr{H}_1 \oplus \cdots \oplus \mathscr{H}_n$, which is the direct sum of the orthogonal subspaces $\mathscr{H}_0, \mathscr{H}_1, ..., \mathscr{H}_n$. From the above it follows that for any polynomial $p \in \mathscr{P}_n$ and for any function $h \in \mathscr{H}_{n+1} \oplus \mathscr{H}_{n+2} \oplus \cdots$ the equality

$$\int_{S^{d-1}} p(\xi)\, \overline{h(\xi)}\, d\xi = 0. \tag{A.10}$$

holds.

ACKNOWLEDGMENTS

The author is extremely grateful to Allan Pinkus and Ron Meir for helpful discussions. The author is also grateful to Konstantin Oskolkov, Vladimir Temlyakov, and Pencho Petrushev for cooperation and their providing preprints of papers [17, 21, 18].

REFERENCES

1. R. A. Adams, "Sobolev Spaces," Academic Press, New York, 1975.
2. C. Bennett and R. Shapley, "Interpolation of Operators," Academic Press, New York, 1988.
3. R. A. DeVore, R. Howard, and C. A. Micchelli, Optimal nonlinear approximation, *Manuscripta Math.* **63** (1989), 469–478.
4. R. A. DeVore, K. Oskolkov, and P. Petrushev, Approximation by feed-forward neural networks, *Ann. Numer. Math.* **4** (1997), 261–287.
5. L. Devroye, L. Györfy, and G. Lugosi, "A Probabilistic Theory of Pattern Recognition," Springer Verlag, New York, 1996.
6. A. Erdelyi (Ed.), "Higher Transcendental Functions," Vol. 2, Bateman Manuscript Project, McGraw–Hill, New York, 1953.
7. P. J. Hilton and S. Wylie, "Homology Theory," Cambridge Univ. Press, Cambridge, UK, 1960.
8. M. Karpinski and A. J. Macintyre, Polynomial bounds for VC dimension of sigmoidal neural networks, *in* "Proc. 27th ACM Symp. Thy. of Computing, 1995," pp. 200–208.
9. A. Kroo, On approximation by ridge functions, *Constr. Approx.* **13** (1997), 447–460.
10. V. Ya. Lin and A. Pinkus, Fundamentality of ridge functions, *J. Approx. Theory* **75** (1993), 295–311.

11. B. Logan and L. Shepp, Optimal reconstruction of functions from its projections, *Duke Math. J.* **42** (1975), 645–659.
12. V. Maiorov and J. Ratsaby, On the degree of approximation using manifolds of finite pseudo-dimension, *Construct. Approxim.* **15** (1999), 291–300.
13. V. Maiorov and R. Meir, On the near optimality of the stochastic approximation of smooth functions by neural networks, *Adv. Comp. Math.*, to appear.
14. J. Milnor, On the Betti numbers of real varieties, *Proc. Amer. Math. Soc.* **15** (1964), 275–280.
15. H. N. Mhaskar, Neural networks for optimal approximation of smooth and analytic functions, *Neural Comput.* **8** (1996), 164–177.
16. H. N. Mhaskar and C. A. Micchelli, Dimension independent bounds on the degree of approximation by neural networks, *IBM J. Res. Develop.* **38** (1994), 277–284.
17. K. I. Oskolkov, Ridge approximation, Chebyshev–Fourier analysis and optimal quadrature formulas, *Proc. Steklov Inst. Math.* **219** (1997), 265–280.
18. P. P. Petrushev, Approximation by ridge functions and neural networks, *SIAM. J. on Math. Analysis* **30** (1998), 155–189.
19. A. Pinkus, "*n*-Widths in Approximation Theory," Springer-Verlag, Berlin, 1985.
20. E. M. Stein and G. Weiss, "Introduction to Fourier Analysis on Euclidean Spaces," Princeton, NJ, 1971.
21. V. Temlyakov, On approximation by ridge functions, preprint, Dept. of Math., University of South Carolina, 1996, 12 pp.
22. A. F. Timan, "Theory of Approximation of Function of the Real Variable," Macmillan Co., New York, 1963.
23. M. Vidyasagar, "A Theory of Learning and Generalization," Springer-Verlag, London, 1997.
24. N. Ya Vilenkin, "Special Functions and a Theory of Representation of Group," Fizmatgiz, Moscow, 1965.
25. A. G. Vitushkin, "Estimation of the Complexity of the Tabulation Problem," Fizmatgiz, Moscow, 1959.
26. B. A. Vostrecov and M. A. Kreines, Approximation of continuous functions by superpositions of plane waves, *Dokl. Akad. Nauk SSSR* **140** (1961), 1237–1240; *Soviet Math. Dokl.* **2** (1961), 1320–1329.
27. H. E. Warren, Lower bounds for approximation by nonlinear manifold, *Trans. Amer. Math. Soc.* **133** (1968), 167–178.