Journal of Approximation Theory **99**, 95–111 (1999)
Article ID jath.1998.3305, available online at http://www.idealibrary.com on IDEAL®

On the Approximation of Functional Classes Equipped with a Uniform Measure Using Ridge Functions*

Vitaly Maiorov

Department of Mathematics, Technion, Haifa 32000, Israel
E-mail: maiorov@tx.technion.ac.il

and

Ron Meir and Joel Ratsaby

Department of Electrical Engineering, Technion, Haifa 32000, Israel
E-mail: rmeir@ee.technion.ac.il, jer@ee.technion.ac.il

Communicated by Allan Pinkus

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

We introduce a construction of a uniform measure over a functional class $\mathscr{B}^r$ which is similar to a Besov class with smoothness index r. We then consider the problem of approximating $\mathscr{B}^r$ using a manifold M_n which consists of all linear manifolds spanned by n ridge functions, i.e., $M_n = \{\sum_{i=1}^n g_i(a_i \cdot x) \colon a_i \in S^{d-1}, g_i \in L_2([-1, 1])\}$, $x \in B^d$. It is proved that for some subset $A \subset \mathscr{B}^r$ of probabilistic measure $1-\delta$, for *all* $f \in A$ the degree of approximation of M_n behaves asymptotically as $1/n^{r/(d-1)}$. As a direct consequence the probabilistic (n, δ)-width for *nonlinear* approximation denoted as $d_{n,\delta}(\mathscr{B}^r, \mu, M_n)$, where μ is a uniform measure over $\mathscr{B}^r$, is similarly bounded. The lower bound holds also for the specific case of approximation using a manifold of one hidden layer neural networks with n hidden units. © 1999 Academic Press

1. INTRODUCTION

We consider the problem of approximating a functional class $\mathscr{B}^r$ similar to a Besov class using a manifold of ridge functions $M_n = \{\sum_{i=1}^n g_i(a_i \cdot x)\colon a_i \in S^{d-1}, g_i \in L_2([-1, 1])\}$, defined on the unit ball $B^d = \{x \in \mathbb{R}^d \colon \|x\|_2 := (\sum_{i=1}^d x_i^2)^{1/2} \leqslant 1\}$ in the space $\mathbb{R}^d$. Here $S^{d-1} = \{x \in B^d \colon \|x\|_2 = 1\}$ is the

* The work of V. Maiorov was partially supported by the center for Absorption in Science, Ministry of Immigrant Absorption, State of Israel. The work of R. Meir was partially supported by a grant from the Israel Science Foundation. Support of the Ollendorff center at the Department of Electrical Engineering in the Technion is also acknowledged.

0021-9045/99 $30.00

unit sphere in $\mathbb{R}^d$. The degree of approximation of $f \in \mathcal{B}^r$ by M_n in the space L_2 is defined by the expression

$$\operatorname{dist}(f, M_n, L_2) = \inf_{h \in M_n} \|f - h\|_{L_2},$$

where $\|f\|_{L_2}$ denotes the L_2-norm of f on B^d.

Vostrecov and Kreines [27] and Lin and Pinkus [8, 9] studied issues of fundamentality of ridge functions in functional spaces. The specific case in which the ridge functions are sigmoidal, e.g., $g_i(y) = \sigma(y) = 1/1 + e^{-y}$, $1 \leqslant i \leqslant n$, and translations are permitted corresponds to a manifold $\mathcal{H}_n = \{\sum_{i=1}^n c_i \sigma(a_i \cdot x + b_i) : a_i \in \mathbb{R}^d, b_i, c_i \in \mathbb{R}\}$, of one hidden layer neural networks with n hidden units. There have been many investigations concerning the approximation properties of $\mathcal{H}_n$, e.g., Barron [1], Mhaskar [15], Girosi *et al.* [4–6], DeVore *et al.* [2], Petrushev [19], and Maiorov and Meir [13].

Recently a series of results was obtained for estimates of approximation of functions by the ridge-manifold M_n in the two-dimensional case, $d = 2$ (see Oskolkov [18], and Temlyakov [23]). In particular, Oskolkov showed that for $d = 2$ the orders of approximation of radial functions by the ridge-manifold M_n and by the space of algebraic polynomials of degree n coincide. In Maiorov [12], the asymptotic behavior of the distance

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

for the Sobolev class $W_2^{r,d}$, $d \geqslant 2$, was obtained.

In this work we are interested in assessing how massive is the subset of functions in $\mathcal{B}^r$ such that for *all* functions in this subset a certain degree of approximation holds. In order to formalize the statement that a high percentage of the functions in $\mathcal{B}^r$ are approximated by M_n to a certain degree we construct a uniform measure over $\mathcal{B}^r$. The volume of a subset A of the unit ball in $\mathbb{R}^n$ which is equipped with a uniform probability measure is proportional to the probability of A. Similarly if $\mathcal{B}^r$ is equipped with the uniform measure then a subset in $\mathcal{B}^r$ of high probability is interpreted as being massive in the sense of occupying almost all of $\mathcal{B}^r$.

To proceed we first construct a uniform measure μ over $\mathcal{B}^r$. We then calculate lower and upper bounds on the degree of approximation by M_n which holds for all functions in some subset $A \subset \mathcal{B}^r$ of probability $1 - \delta$. Specifically we obtain a degree of approximation such that for some $A \subset \mathcal{B}^r$, with $\mu(A) \geqslant 1 - e^{-\alpha(n)}$ and $\alpha(n) = c_0 n^{d/(d-1)}$, then for all $f \in A$, $c_1/n^{r/(d-1)} \leqslant \operatorname{dist}(f, M_n, L_2) \leqslant c_2/n^{r/(d-1)}$, for some constants c_0, $c_1, c_2 > 0$ depending on r and d, but not on n.

In [12], upper and lower bounds on the distance

$$\operatorname{dist}(W_2^{r,d}, M_n, L_2) = \sup_{f \in W_2^{r,d}} \operatorname{dist}(f, M_n, L_2)$$

for a Sobolev class $W_2^{r,d}$ were obtained. However, this type of result only guarantees the existence of a function in $W_2^{r,d}$ for which the lower bound holds. That is, it is a "worst case" result. In this work we extend that result by obtaining tight lower and upper bounds that hold for all functions in a subset of large measure in $\mathscr{B}^r$.

As a consequence, we obtain asymptotically tight lower and upper bounds on the distance between $\mathscr{B}^r$ and M_n measured by a probabilistic (n, δ)-width which is defined as

$$d_{n,\delta}(\mathscr{B}^r, \mu, M_n) = \inf_{\substack{A \subset \mathscr{B}^r \\ \mu(A) = 1-\delta}} \sup_{f \in A} \operatorname{dist}(f, M_n, L_2), \tag{1}$$

where $0 \leqslant \delta \leqslant 1$ and the infimum runs over all subsets A of $\mathscr{B}^r$ with probability $\mu(A) = 1 - \delta$. From the construction of the class $\mathscr{B}^r$ one can see that for any $0 \leqslant \delta \leqslant 1$ there exists a subset $A \in \mathscr{B}^r$ such that $\mu(A) = 1 - \delta$. Quantities similar to (1) were considered in [25, 11, 14] where μ was taken to be a Gaussian or Wiener measure and the approximation was linear.

From (1) the next inverse formulation follows

$$\mu\{f \in \mathscr{B}^r : \operatorname{dist}(f, M_n, L_2) \geqslant d_{n,\delta}\} = \delta,$$

where $d_{n,\delta} = d_{n,\delta}(\mathscr{B}^r, \mu, M_n)$. Indeed, from (1) it follows that there exists the subset A in $\mathscr{B}^r$ such that $\mu(A) = 1 - \delta$ and

$$\begin{aligned} &\mu\{f \in \mathscr{B}^r : \operatorname{dist}(f, M_n, L_2) \geqslant d_{n,\delta}\} \\ &\quad = \mu\{f \in \mathscr{B}^r : \operatorname{dist}(f, M_n, L_2) \geqslant \sup_{h \in A} \operatorname{dist}(h, M_n, L_2)\} \end{aligned}$$

and hence

$$\mu\{f \in \mathscr{B}^r : \operatorname{dist}(f, M_n, L_2) \geqslant d_{n,\delta}\} = \mu\{\mathscr{B}^r \backslash A\} = 1 - \mu(A) = \delta.$$

The main contributions of this paper are threefold: (i) the construction of a uniform measure over a functional class $\mathscr{B}^r$ which is similar to a Besov class. (ii) Proving a lower bound on the degree of approximation by ridge functions which holds for all functions in some subset of $\mathscr{B}^r$ of probability measure $1 - \delta$ with respect to the uniform measure. (iii) Introducing a probabilistic width $d_{n,\delta}$ for *nonlinear* approximation and estimating $d_{n,\delta}(\mathscr{B}^r, \mu, M_n)$ for a uniform measure μ.

2. PRELIMINARIES

We begin by introducing some notation. For an integer $m \geqslant 1$ let $\mathbb{Z}_m = \{1, 2, ..., m\}$. Consider the ball of radius r in $\mathbb{R}^m$ denoted by $B^m(r) = \{x \in \mathbb{R}^m: \|x\|_2 \leqslant r\}$ and set $B^m = B^m(1)$. For a vector $z \in \mathbb{R}^m$, $\operatorname{sgn}(z) = (\operatorname{sgn}(z_1), ..., \operatorname{sgn}(z_m))$, $\operatorname{sgn}(z_i) = 1$ for $z_i \geqslant 0$, $\operatorname{sgn}(z_i) = -1$ for $z_i < 0$. We denote by $\|v\|_{l_p^m}$ or simply $\|v\|_p$, $p \geqslant 1$, the l_p^m Euclidean norm of $v \in \mathbb{R}^m$. For any Euclidean sets A and B in $\mathbb{R}^m$ we use a distance function $\operatorname{dist}(a, B, l_p^m) = \inf_{b \in B} \|a - b\|_{l_p^m}$ for any $a \in A$, and $\operatorname{dist}(A, B, l_p^m) = \sup_{a \in A} \operatorname{dist}(a, B, l_p^m)$.

Define the space of functions

$$L_2 = L_2(B^d) = \left\{ f \colon \|f\|_{L_2} := \left(\int_{B^d} |f(x)|^2 \, dx \right)^{1/2} < \infty \right\}.$$

We write $\int_{B^d} f(x) \, dx$ where $x = (x_1, ..., x_d)$, and $dx = dx_1 \cdots dx_d$.

The notation $a_n \asymp b_n$ in this paper means that there exist constants $c_1, c_2 > 0$ which depend only on the smoothness parameter r of the class $\mathscr{B}^r$ and the dimensionality d of the domain B^d such that for every $n \geqslant 1$, $c_1 \leqslant a_n / b_n \leqslant c_2$.

We define the class of functions $\mathscr{B}^r$ using the classical means of approximation, namely, algebraic polynomials. Consider the space $\mathscr{P}_s = \operatorname{span}\{x_1^{k_1} \cdots x_d^{k_d}: |k| = k_1 + \cdots + k_d \leqslant s\}$, $s = 0, 1, ...,$ consisting of all algebraic polynomials on $\mathbb{R}^d$ of total degree at most s. Let $\mathscr{P}_s^h = \operatorname{span}\{x_1^{k_1} \cdots x_d^{k_d}: |k| = s\}$ be the subspace of $\mathscr{P}_s$ consisting of homogeneous polynomials of degree s. Set $m_s = \dim \mathscr{P}_s^h$. It is known (cf. [22]) that $m_s = \binom{d+s-1}{d-1} \asymp s^{d-1}$.

Let the set of polynomials $Q_s = \{q_l\}_{l=1}^{m_s}$ be a basis in $\mathscr{P}_s^h$. The set of polynomials $\bigcup_{s=0}^{\infty} Q_s$ is a complete system of functions in the space L_2. Using the method of orthogonalization in L_2 we can construct a complete orthogonal system of polynomials in L_2

$$P = \bigcup_{s=0}^{\infty} \{p_{s,1}, ..., p_{s,m_s}\},$$

such that the set $P_s^h = \{p_{s,1}, ..., p_{s,m_s}\}$ is a complete orthonormal system of functions in the subspace $\mathscr{P}_s^h$. Note in particular that in [12] we constructed one specific orthonormal system of algebraic polynomials in L_2.

For any natural N we denote the set of multi-indexes

$$\Delta_N = \{(s, l): s = 2^N + 1, ..., 2^{N+1}, l = 1, ..., m_s\}.$$

Introduce the subspace $\Phi_N = \operatorname{span}\{p_{s,l}: (s,l) \in \Delta_N\}$. Let G_N^r, $r>0$, be the ball with radius 2^{-rN} in the space Φ_N, that is,

$$G_N^r = \left\{ \sum_{(s,l)\in\Delta_N} c_{s,l}\, p_{s,l} \in \Phi_N : \left(\sum_{(s,l)\in\Delta_N} |c_{s,l}|^2 \right)^{1/2} \leqslant 2^{-rN} \right\}.$$

Denote by $\mathscr{B}^r$, the set of all functions $f \in L_2(B^d)$ which can represented as infinite sums of functions from G_N^r, namely

$$\mathscr{B}^r = \left\{ f: f = \sum_{N=0}^{\infty} f_N,\ f_N \in G_N^r,\ N=0, 1, ... \right\}.$$

It is not hard to see that the class $\mathscr{B}^r$ is essentially equivalent to the class H^r, consisting of all functions f for which the best approximation by algebraic polynomials of degree 2^N satisfies the inequality

$$\operatorname{dist}(f, \mathscr{P}_{2^N}, L_2) \leqslant 2^{-rN} \qquad (N=0, 1, ...).$$

From Jackson's Theorem (see [24]), it follows that the Sobolev class $W_2^{r,d}$ belongs to the class H^r and hence also to the class $c\mathscr{B}^r$, for some constant c. Observe also that the latter class (discussed also in [23]) is analogous to the Besov class [26] which is defined using trigonometric polynomials.

As an approximating function class we will use the following nonlinear manifold

$$M_n = \left\{ h(x) = \sum_{l=1}^{n} h_l(a_l \cdot x) : a_l \in S^{d-1},\ h_l \in L_2([-1, 1]) \right\} \qquad (x \in B^d) \qquad (2)$$

which represents the union of all linear manifolds that are spanned by n ridge functions from the space $L_2([-1,1])$ of square-integrable functions on the segment $[-1,1]$.

3. UNIFORM MEASURE CONSTRUCTION

The construction of a uniform measure over a functional class is nontrivial. For example, it is not possible to construct such a measure over a Sobolev or Besov class. For this reason we consider the class $\mathscr{B}^r$ which permits such a construction.

Let $P=\{p_{s,l}\}$ be a complete system of orthonormal polynomials in L_2, as constructed in Section 2. Then we can express the class $\mathscr{B}^r$ as

$$\mathscr{B}^r=\Bigg\{f\in L_2\colon f(x)=\sum_{N=0}^{\infty}\sum_{(s,l)\in\Delta_N}c_{s,l}p_{s,l}(x),$$

$$\Bigg(\sum_{(s,l)\in\Delta_N}|c_{s,l}|^2\Bigg)^{1/2}\leqslant 2^{-rN},\text{ for all } N\geqslant 0\Bigg\}. \tag{3}$$

Consider the subspace $\Phi_N=\operatorname{span}\{p_{s,l}\colon (s,l)\in\Delta_N\}$. We have that Φ_N is orthogonal to $\Phi_{N'}$, for all $N\neq N'$, and $\mathscr{B}^r$ is isomorphic to the set D^r of inifinite sequences of finite dimensional vectors, i.e.,

$$\mathscr{B}^r\simeq D^r:=\prod_{N=0}^{\infty}B^{|\Delta_N|}(2^{-rN})$$

$$:=\{c=(c^0,...,c^N,...)\colon c^N\in B^{|\Delta_N|}(2^{-rN})\}, \tag{4}$$

where $c^N:=(c_{s,l})_{(s,l)\in\Delta_N}$, and $|\Delta_N|$ is the cardinality of Δ_N, $N\geqslant 0$.

Note that the cardinality of Δ_N satisfies the asymptotic

$$|\Delta_N|=\sum_{s=2^N+1}^{2^{N+1}}\dim\mathscr{P}_s^h=\sum_{s=2^N+1}^{2^{N+1}}m_s\asymp\sum_{s=2^N+1}^{2^{N+1}}s^{d-1}\asymp 2^{dN}.$$

Let $b_n\equiv B^{|\Delta_n|}(2^{-rn})$ be the ball of radius 2^{-rn} in $\mathbb{R}^{|\Delta_n|}$, and denote the volume of b_n by $\operatorname{vol}(b_n)$. Let $v_n(dc^n)=dc^n/\operatorname{vol}(b_n)$ be the normed Lebesgue measure on b_n, $v_n(b_n)=1$, and

$$D_N^r=\prod_{n=0}^{N}b_n.$$

For $c=(c^0,...,c^N)\in D_N^r$ define the measure on D_N^r as

$$\lambda_N(dc)=\prod_{n=0}^{N}v_n(dc^n).$$

Now, let $B\subset D_N^r$. We have

$$\lambda_{N+1}(B\times b_{N+1})=\int_{B\times b_{N+1}}\lambda_N(dc)\,v_{N+1}(dc^{N+1})$$

$$=\frac{1}{\prod_{n=0}^{N}\operatorname{vol}(b_n)}\,\frac{1}{\operatorname{vol}(b_{N+1})}\int_{B\times b_{N+1}}dx\,dy$$

$$=\frac{1}{\prod_{n=0}^{N+1}\operatorname{vol}(b_n)}\int_B\int_{b_{N+1}}dy\,dx$$

which equals $\mathrm{vol}_{(B)}/\prod_{n=0}^{N}\mathrm{vol}_{(b_n)}=\lambda_N(B)$. It follows from the Kolmogorov Extension of Measure Theorem (see, for example, Shiryayev [21, Theorem 3, and observation, p. 163]) that there exists a unique probability measure λ on D^r such that for every $B\subset D_N^r$

$$\lambda((c^0, ..., c^N, ...)\in D^r: (c^0, ..., c^N)\in B)=\lambda_N(B).$$

This uniform measure λ on D^r induces a uniform measure μ on $\mathscr{B}^r$, which will now be used to establish our main result.

4. MAIN RESULTS

Let $r>0$ and an integer $d\geqslant 1$ be given. Fix an integer $n\geqslant 1$, and set $\alpha(n)=c_1 n^{d/(d-1)}$, for some constant $c_1>0$ depending only on r and d. Let μ be the uniform measure over $\mathscr{B}^r$ constructed in Section 3.

THEOREM 1.

$$\mu\left\{f\in\mathscr{B}^r:\operatorname{dist}(f, M_n, L_2)\geqslant\frac{c_2}{n^{r/(d-1)}}\right\}\geqslant 1-e^{-\alpha(n)}$$

for some constant $c_2>0$ depending only on r and d.

THEOREM 2. *For all $f\in\mathscr{B}^r$*

$$\operatorname{dist}(f, M_n, L_2)\leqslant\frac{c_3}{n^{r/(d-1)}},$$

where $c_3>0$ is some constant depending only on r and d.

From Theorems 1 and 2 we have the following corollary which estimates the probabilistic width defined in (1).

COROLLARY 1. *Let $0\leqslant\delta<1-2e^{-\alpha(n)}$. Then*

$$\frac{c_2}{n^{r/(d-1)}}\leqslant d_{n,\delta}(\mathscr{B}^r,\mu, M_n)\leqslant\frac{c_3}{n^{r/(d-1)}}$$

for some constants $c_2, c_3>0$ depending only on r and d.

Indeed let $0\leqslant\delta<1-2e^{-\alpha(n)}$ be any number. Then for any set $A\subset\mathscr{B}^r$ with the measure $\mu(A)=1-\delta$ we have $\mu(A)\geqslant 2e^{-\alpha(n)}$. Therefore from

Theorem 1 it follows that there exists a function $f \in A$ such that $\operatorname{dist}(f, M_n, L_2) \geqslant c_2 n^{-r/(d-1)}$. Hence

$$d_{n,\delta}(\mathscr{B}^r, \mu, M_n) \geqslant \operatorname{dist}(f, M_n, L_2) \geqslant \frac{c_2}{n^{r/(d-1)}}.$$

The upper bound in Corollary 1 follows directly from Theorem 2.

We note that Traub *et al.* [25] consider also the so called average case setting which introduces the notion of an average distance with respect to a measure over a functional space in our case defined for $0 < p < \infty$ as

$$d_n^{avg}(\mathscr{B}^r, \mu, M_n)_p = \left(\int_{f \in \mathscr{B}^r} |\operatorname{dist}(f, M_n, L_2)|^p \, \mu(df) \right)^{1/p}.$$

The following corollary follows easily from Theorems 1 and 2.

Corollary 2. *For any* $0 < p < \infty$,

$$\frac{c_2}{n^{r/(d-1)}} \leqslant d_n^{avg}(\mathscr{B}^r, \mu, M_n)_p \leqslant \frac{c_3}{n^{r/(d-1)}}$$

for some constants $c_2, c_3 > 0$ *depending only on* r, d, *and* p.

We proceed to prove Theorem 1, first stating several auxiliary lemmas. From the definition of the orthonormal system $P = \{p_{s,l}\}$ it follows that an $h \in M_n$ can be expressed as a sum $\sum_{N=0}^{\infty} \sum_{(s,l) \in \varDelta_N} c_{s,l}(h) \, p_{s,l}(x)$ with the coefficients, $c_{s,l}(h) = \langle h, p_{s,l} \rangle = \int_{B^d} h(x) \, p_{s,l}(x) \, dx$. Let $N \in \mathbb{Z}_+$ be some number, and $I \subset \varDelta_N$ be any subset. Consider the set of sign-valued vectors

$$\varGamma_n^I := \{(\operatorname{sgn}(c_{s,l}(h))_{(s,l) \in I} : h \in M_n\}. \tag{5}$$

We will use the next lemma which follows from Lemma 3 of [12].

Lemma 1. *Assume that* N *and* n *are such that* $|\varDelta_N| = [c_5 n^{d/(d-1)}]$, *for some absolute constant* $c_5 > 0$. *Then for any subset* $I \subset \varDelta_N$ *with* $|I| \geqslant |\varDelta_N|/10$ *we have*

$$|\varGamma_n^I| \leqslant 2^{c_4 |I|} \leqslant 2^{c_6 n^{d/(d-1)}},$$

where $c_4 = 0.23$, *and* $c_6 = c_4 c_5$.

The next lemma then follows.

LEMMA 2. *Let* $|\Delta_N| = [c_5 n^{d/(d-1)}]$, *and let* $I \subset \Delta_N$, $|I| \geqslant |\Delta_N|/10$. *Introduce the sets of sign-valued vectors* $E^{|I|} = \{-1, +1\}^{|I|}$, *and* $\hat{E}^{|I|} = \{\varepsilon \in E^{|I|}: \operatorname{dist}(\varepsilon, \Gamma_n^I, l_2^{|I|}) \geqslant 2\sqrt{|I|/3}\}$. *Then*

$$|\hat{E}^{|I|}| \geqslant 2^{|I|} - 2^{c_7 |I|}$$

for some absolute constant $0 < c_7 < 1$.

Proof. Set $k = |I|$. From Lemma 1 it follows that the cardinality $|\Gamma_n^I| \leqslant 2^{c_4 k}$. Fix any $\varepsilon^* \in E^k$. Denote by

$$D_{\varepsilon^*} = \left\{\varepsilon \in E^k \colon \|\varepsilon - \varepsilon^*\|_{l_2^k}^2 \geqslant \frac{4k}{9}\right\}.$$

Now $|D_{\varepsilon^*}|$ is independent of the specific choice of $\varepsilon^* \in E^k$. As such $|D_{\varepsilon^*}| = |\{\varepsilon \in E^k \colon \|\varepsilon - \underline{1}\|_{l_2^k}^2 \geqslant 4k/9\}|$ where $\underline{1} = [1, ..., 1] \in E^k$. The latter equals $\sum_{i \geqslant k/9} \binom{k}{i}$ and is bounded from below by $2^k - 2^{c_8 k}$, $c_8 = 1 - 2(7/18)^2 \log_2 e = 0.55...$, where we used an upper bound on the tails of the binomial distribution (cf. [3]).

Set $\bar{D}_\varepsilon = E^k \backslash D_\varepsilon$. Then $|\bar{D}_{\varepsilon^*}| = |E^k \backslash D_{\varepsilon^*}| \leqslant 2^{c_8 k}$. We also have $\hat{E}^k := \bigcap_{\varepsilon^* \in \Gamma_n^I} D_{\varepsilon^*} = E^k \backslash (\bigcup_{\varepsilon^* \in \Gamma_n^I} \bar{D}_{\varepsilon^*})$. It follows that

$$|\hat{E}^k| \geqslant |E^k| - \left|\bigcup_{\varepsilon^* \in \Gamma_n^I} \bar{D}_{\varepsilon^*}\right| \geqslant |E^k| - |\Gamma_n^I|\, 2^{c_8 k} \geqslant 2^k - 2^{c_4 k} 2^{c_8 k}.$$

Set $c_7 = c_4 + c_8 = 0.78...$. Thus $|\hat{E}^k| \geqslant 2^k - 2^{c_7 k}$, which proves the lemma.

DEFINITION 1. Let B^m denote the unit ball in $\mathbb{R}^m$. For any set $A \subset B^m$ denote the volume of A as $\operatorname{vol}(A)$. The uniform measure over the ball denoted by ν is defined such that for every $A \subset B^m$, $\nu(A) = \operatorname{vol}(A)/\operatorname{vol}(B^m)$.

Denote by

$$A := \left\{x \in B^m \colon |x_k| > \frac{3}{8\sqrt{m}}, \text{ for at least } \frac{m}{10} \text{ coordinates } k\right\}.$$

We will use the following lemma.

LEMMA 3. *For any* $m \geqslant 1$

$$\nu(A) \geqslant 1 - 3e^{-c_9 m}$$

for some absolute constant $c_9 > 0$.

Proof. We aim at finding a lower bound on $\nu(A)$ by first expressing the measure of the set A under the uniform measure over B^m as the measure of another set under the Gaussian measure over $\mathbb{R}^m$. Introduce the auxiliary set in $\mathbb{R}^m$

$$\hat{A}=\left\{x\in\mathbb{R}^m\colon |x_k|>\frac{3}{4\sqrt{m}}\|x\|_2, \text{ for at least } \frac{m}{10} \text{ coordinates } k\right\}.$$

Denote by $B^m(\alpha,\beta):=B^m(\beta)\backslash B^m(\alpha)$. We have

$$\nu(A)\geqslant\nu(B^m(\tfrac{1}{2},1)\cap\hat{A})\geqslant\nu(\hat{A}\cap B^m)-(\tfrac{1}{2})^m. \tag{6}$$

Let $\chi_A(x)$ denote the indicator function of the set A. Switching to polar coordinates we have, since $x\in\hat{A}$ implies $ax\in\hat{A}$ for all $a\neq 0$

$$\begin{aligned}\nu(\hat{A}\cap B^m)&=\frac{1}{\mathrm{vol}(B^m)}\int_{B^m}\chi_{\hat{A}}(x)\,dx\\ &=\frac{1}{\mathrm{vol}(B^m)}\int_0^1 r^{m-1}\,dr\int_{S^{m-1}}\chi_{\hat{A}}(s)\,ds \qquad (s\in S^{m-1}),\end{aligned} \tag{7}$$

where ds is the Lebesgue measure on S^{m-1}. Assume that m is even (for m odd the proof is analogous). The volume of the unit ball $\mathrm{vol}(B^m)=\pi^{m/2}/(m/2)!$. It is known (cf. [17]) that $\int_0^\infty x^{m-1}e^{-x^2}\,dx=\frac{1}{2}\Gamma(m/2)$. Hence it follows that

$$\frac{1}{\mathrm{vol}(B^m)}\int_0^1 r^{m-1}\,dr=\pi^{-m/2}\int_0^\infty r^{m-1}e^{-r^2}\,dr.$$

Therefore using once more polar coordinates we obtain from (7)

$$\begin{aligned}\nu(\hat{A}\cap B^m)&=\pi^{-m/2}\int_0^\infty r^{m-1}e^{-r^2}\,dr\int_{S^{m-1}}\chi_{\hat{A}}(s)\,ds\\ &=\pi^{-m/2}\int_{\mathbb{R}^m}\chi_{\hat{A}}(x)\,e^{-|x|^2}\,dx.\end{aligned} \tag{8}$$

Define a Gaussian measure over $\mathbb{R}^m$ as $\gamma(G)=\pi^{-m/2}\int_G e^{-|x|^2}\,dx$, $G\subset\mathbb{R}^m$. From (8) it is seen that $\nu(\hat{A}\cap B^m)=\gamma(\hat{A})$. Let

$$D=\left\{x\in\mathbb{R}^m\colon |x_k|\geqslant\frac{3}{2}, \text{ for at least } \frac{m}{10} \text{ coordinates } k\right\}.$$

Then it follows that

$$\nu(\hat{A}\cap B^m)=\gamma(\hat{A})\geqslant\gamma(\hat{A}\cap B^m(2\sqrt{m}))\geqslant\gamma(D\cap B^m(2\sqrt{m})),$$

and therefore

$$\begin{aligned} v(\hat{A} \cap B^m) &\geqslant \gamma(D) + \gamma(B^m(2\sqrt{m})) - \gamma(D \cup B^m(2\sqrt{m})) \\ &\geqslant \gamma(D) + \gamma(B^m(2\sqrt{m})) - 1. \end{aligned} \tag{9}$$

Let $I \subset \mathbb{Z}_m = \{1, 2, ..., m\}$. Consider the subset in D

$$D_I = \{x \in D\colon |x_i| \geqslant \tfrac{3}{2} \text{ for all } i \in I,\ |x_i| < \tfrac{3}{2} \text{ for all } i \in \mathbb{Z}_m \backslash I\}.$$

We have

$$\gamma(D) = \sum_{I \subset \mathbb{Z}_m} \gamma(D_I) = \sum_{l=1}^{m} \sum_{I \subset \mathbb{Z}_m,\ |I| = l} \gamma(D_I) \geqslant \sum_{l=m/10}^{m} \sum_{I \subset \mathbb{Z}_m,\ |I| = l} \gamma(D_I).$$

For $|I| = l$

$$\gamma(D_I) = p^l (1-p)^{m-l},$$

where

$$\frac{1}{\sqrt{\pi}} \int_{|t| \geqslant 3/2} e^{-t^2}\, dt = 0.134 \equiv p.$$

Hence from the definition of the Gaussian measure γ it follows that

$$\gamma(D) \geqslant \sum_{l=m/10}^{m} \binom{m}{l} p^l (1-p)^{m-l} > 1 - e^{-c_{10} m} \tag{10}$$

for some $0 < c_{10} < 1$ where we used a bound on the tail of the binomial distribution [3].

We now estimate $\gamma(B^m(2\sqrt{m}))$. We will show that

$$\gamma(B^m(2\sqrt{m})) \geqslant 1 - e^{-c_{11} m} \tag{11}$$

for some absolute constant $c_{11} > 0$.

Indeed using polar coordinates we have

$$\begin{aligned} \gamma(B^m(2\sqrt{m})) &= \pi^{-m/2} \int_{B^m(2\sqrt{m})} e^{-|x|^2}\, dx \\ &= 1 - \pi^{-m/2} \int_{\mathbb{R}^m \backslash B^m(2\sqrt{m})} e^{-|x|^2}\, dx \\ &= 1 - \pi^{-m/2} d(S^{d-1}) \int_{2\sqrt{m}}^{\infty} r^{m-1} e^{-r^2}\, dr, \end{aligned}$$

where $d(S^{d-1})$ is the Lebesgue measure of the sphere S^{d-1}. Using the substitution $r=\sqrt{mt/2}$, and the estimate $\int_8^\infty t^{k-1}e^{-kt}\,dr \leqslant (1/7k)\,e^{-8k}8^{k+1/2}$, $k \geqslant 1$ (see [10, p. 471, form. (6.5)]), we obtain

$$\int_{2\sqrt{m}}^{\infty} r^{m-1}e^{-r^2}\,dr = \tfrac{1}{2}\,(m/2)^{m/2}\int_8^\infty t^{m/2-1}e^{-(m/2)\,t}\,dt \leqslant \tfrac{1}{2}(m/2)^{m/2}\,e^{-c'_{11}m},$$

where $c'_{11}=4-\frac{3}{2}\ln 2$. Since $d(S^{d-1})=\mathrm{vol}(B^m)/m=\pi^{m/2}/m\Gamma(m/2)\asymp \pi^{m/2} \times e^{m/2}/(m(m/2)^{m/2}\sqrt{2\pi m})$, then

$$\gamma(B^m(2\sqrt{m})) \geqslant 1-\pi^{-m/2}\,d(S^{d-1})\,\tfrac{1}{2}(m/2)^{m/2}\,e^{-c'_{11}m} \geqslant 1-e^{-c_{11}m},$$

where $c_{11}=\frac{3}{2}(1-\ln 2)$.

Using (6), (9), (10), and (11) we obtain that

$$\nu(A) \geqslant 1-e^{-c_{10}m}-e^{-c_{11}m}-2^{-m} \geqslant 1-3e^{-c_9 m},$$

for absolute constant $c_9=\min\{c_{10}, c_{11}, \ln 2\}$. ∎

We now proceed with finding a lower bound on the measure stated in Theorem 1.

4.1. *Proof of Theorem* 1

The proof of Theorem 1 is based on the following observation. Let $m=|\Delta_N|$. In the space $\mathbb{R}^m$, consider the set $E^m=\{-1,+1\}^m$ endowed with a uniform discrete measure α, and let $\Gamma_n^{\Delta_N}$ be the subset in E^m defined in (5). From Lemma 2 it follows that the measure of elements in E^m which are "badly" approximated by the manifold $\Gamma_n^{\Delta_N}$, i.e., the α measure of set $G=\{\varepsilon\in E^m\colon \mathrm{dist}(\varepsilon, \Gamma_n^{\Delta_N}, l_2^m)\geqslant 2\sqrt{m/3}\}$ satisfies the inequality

$$\alpha(G) \geqslant 1-2^{-cm}$$

for $c>0$. This implies that almost all elements from E^m, in the sense of the induced probabilistic measure over E^m, are "badly" approximated by $\Gamma_n^{\Delta_N}$. The statement of the theorem follows upon making use of the isomorphism (4).

We proceed with the detailed proof. Let $N^*>0$ be some integer which will be taken later to be sufficiently large. Since

$$\begin{aligned}\mathrm{dist}(f, M_n, L_2)^2 &= \inf_{h\in M_n}\sum_{N=0}^{\infty}\sum_{(s,l)\in\Delta_N}|c_{s,l}(f)-c_{s,l}(h)|^2\\ &\geqslant \inf_{h\in M_n}\sum_{(s,l)\in\Delta_{N^*}}|c_{s,l}(f)-c_{s,l}(h)|^2,\end{aligned}$$

then for an arbitrary $\varepsilon > 0$

$$\mu\{f \in \mathscr{B}^r : \operatorname{dist}(f, M_n, L_2) > \varepsilon\}$$
$$\geqslant \mu\left\{f \in \mathscr{B}^r : \inf_{h \in M_n} \sum_{(s,l) \in \Delta_{N^*}} |c_{s,l}(f) - c_{s,l}(h)|^2 > \varepsilon^2\right\}. \tag{12}$$

Let m, N^* and n be such that $m = |\Delta_{N^*}| = c_5 n^{d/(d-1)}$. To any $h \in M_n$ there corresponds a vector $\hat{h} \in \mathbb{R}^m$ defined as

$$\hat{h} = (c_{s,l}(h))_{(s,l) \in \Delta_{N^*}}. \tag{13}$$

Denote by

$$\hat{M}_n = \{\hat{h} = (\hat{h}_1, ..., \hat{h}_m) \in \mathbb{R}^m : h \in M_n\}.$$

Due to the isomorphism statement of (4) the approximation problem is now reduced to approximation in an m-dimensional Euclidean space.

Let $\varepsilon = 2^{rN^*}/4$ in (12). We have

$$\Sigma := \mu\left\{f \in \mathscr{B}^r : \inf_{h \in M_n} \sum_{k \in \Delta_{N^*}} |c_k(f) - c_k(h)|^2 > \varepsilon^2\right\}$$
$$= \nu\left\{y \in B^m : \inf_{\hat{h} \in \hat{M}_n} \sum_{i=1}^{m} |y_i - \hat{h}_i|^2 > \tfrac{1}{4}\right\}.$$

Let $I \subseteq \mathbb{Z}_m$. Define the set

$$Q_I = \left\{x \in B^m : |x_i| \geqslant \frac{3}{8\sqrt{m}}, \text{ for all } i \in I, |x_i| \leqslant \frac{3}{8\sqrt{m}} \text{ for all } i \in \mathbb{Z}_m \backslash I\right\}.$$

From the definition of Q_I we have $\bigcup_{I \in \mathbb{Z}_m} Q_I = B^m$. Thus

$$\Sigma = \sum_{I \in \mathbb{Z}_m} \nu\left\{y \in Q_I : \inf_{\hat{h} \in \hat{M}_n} \sum_{i=1}^{m} |y_i - \hat{h}_i|^2 > \tfrac{1}{4}\right\}.$$

For all $I \subset \mathbb{Z}_m$, $|I| \geqslant m/10$, and $y \in Q_I$ we have

$$\sum_{i=1}^{m} |y_i - \hat{h}_i|^2 \geqslant \sum_{i \in I} |y_i - \hat{h}_i|^2 \geqslant \frac{9}{64m} \sum_{i \in I} \left|\frac{y_i}{|y_i|} - \frac{\hat{h}_i}{|y_i|}\right|^2.$$

Denote by $\varepsilon_i(y) = y_i/|y_i| = \mathrm{sgn}(y_i)$. Then using the fact that for any $a \in \mathbb{R}$ and $\delta \in \{-1, +1\}$ the inequality $|\delta - a| \geqslant \frac{1}{2} |\delta - \mathrm{sgn}(a)|$ holds we have

$$\sum_{i=1}^{m} |y_i - \hat{h}_i|^2 \geqslant \frac{9}{256m} \sum_{i \in I} |\varepsilon_i(y) - \mathrm{sgn}(\hat{h}_i)|^2.$$

We then have for $b = 64/9$

$$\Sigma \geqslant \sum_{I \in \mathbb{Z}_m:\, |I| \geqslant m/10} v\left\{ y \in Q_I:\ \inf_{\hat{h} \in \hat{M}_n} \sum_{i \in I} |\varepsilon_i(y) - \mathrm{sgn}(\hat{h}_i)|^2 > bm \right\}$$
$$= \sum_{j=0}^{9m/10} \sum_{I \in \mathbb{Z}_m:\, |I| = m/10 + j} v\left\{ y \in Q_I:\ \inf_{\hat{h} \in \hat{M}_n} \sum_{i \in I} |\varepsilon_i(y) - \mathrm{sgn}(\hat{h}_i)|^2 > bm \right\}.$$

For $I \in \mathbb{Z}_m$ let $E^{|I|} = \{-1, +1\}^{|I|}$. Define

$$\Gamma_n^{|I|} = \{ (\mathrm{sgn}(\hat{h}_i))_{i \in I}:\, h \in M_n \}.$$

Denote by $\|y\|_{l_2^{|I|}} = (\sum_{i \in I} |y_i|^2)^{1/2}$. Let

$$\hat{E}^{|I|} = \{ \varepsilon \in E^{|I|}:\ \min_{\delta \in \Gamma_n^{|I|}} \|\varepsilon - \delta\|^2_{l_2^{|I|}} \geqslant bm \}. \tag{14}$$

For any $\varepsilon = (\varepsilon_i)_{i \in I} \in E^{|I|}$ define the set

$$Q_{I, \varepsilon} = \{ y \in Q_I:\ \mathrm{sgn}(y_i) = \varepsilon_i, \text{ for all } i \in I \}.$$

Then continuing from above we have

$$\Sigma \geqslant \sum_{j=0}^{9m/10} \sum_{I \in \mathbb{Z}_m:\, |I| = m/10 + j} v\{ y \in Q_I:\ \min_{\delta \in \Gamma_n^{|I|}} \|\varepsilon(y) - \delta\|^2_{l_2^{|I|}} > bm \}$$
$$= \sum_{j=0}^{9m/10} \sum_{I \in \mathbb{Z}_m:\, |I| = m/10 + j} \sum_{\varepsilon \in E^{|I|}} v\{ y \in Q_{I,\varepsilon}:\ \min_{\delta \in \Gamma_n^{|I|}} \|\varepsilon(y) - \delta\|^2_{l_2^{|I|}} > bm \}$$

and since $\hat{E}^{|I|} \subset E^{|I|}$ then

$$\Sigma \geqslant \sum_{j=0}^{9m/10} \sum_{I \in \mathbb{Z}_m:\, |I| = m/10 + j} \sum_{\varepsilon \in \hat{E}^{|I|}} v\{ y \in Q_{I,\varepsilon}:\ \min_{\delta \in \Gamma_n^{|I|}} \|\varepsilon(y) - \delta\|^2_{l_2^{|I|}} > bm$$

Now from (14) for all $\varepsilon \in \hat{E}^{|I|}$ the condition $\min_{\delta \in \Gamma^I} \|\varepsilon - \delta\|^2_{l_2^{|I|}} > bm$ is satisfied. We therefore have

$$\Sigma \geqslant \sum_{j=0}^{9m/10} \sum_{I \in \mathbb{Z}_m:\, |I| = m/10 + j} \sum_{\varepsilon \in \hat{E}^{|I|}} v\{ y \in Q_{I,\varepsilon} \}.$$

Note that $\nu(y \in Q_{I,\varepsilon})$ does not depend on ε. Denote by $a_I := \nu\{y \in Q_{I,\varepsilon}\}$. Thus the latter becomes

$$\sum_{j=0}^{9m/10} \sum_{I \in \mathbb{Z}_m:\, |I| = m/10+j} \sum_{\varepsilon \in \hat{E}^{|I|}} a_I = \sum_{j=0}^{9m/10} \sum_{I \in \mathbb{Z}_m:\, |I| = m/10+j} a_I\, |\hat{E}^{|I|}|.$$

From Lemma 2 it follows that for I such that $|I| = m/10 + j$ the cardinality $|\hat{E}^{|I|}| \geqslant 2^{m/10+j} - 2^{c_7(m/10+j)}$ for some constant $0 < c_7 < 1$. We therefore have

$$\begin{aligned} \Sigma &\geqslant \sum_{j=0}^{9m/10} \sum_{I \in \mathbb{Z}_m:\, |I| = m/10+j} a_I 2^{m/10+j}(1 - 2^{-(1-c_7)(m/10+j)}) \\ &\geqslant (1 - 2^{-(1-c_7)\, m/10}) \sum_{j=0}^{9m/10} \sum_{I \in \mathbb{Z}_m:\, |I| = m/10+j} a_I 2^{m/10+j}. \end{aligned}$$

Since $a_I 2^{m/10+j} = |E^{|I|}|\, a_I = \nu(Q_I)$ then

$$\begin{aligned} \Sigma &\geqslant (1 - 2^{-(1-c_7)\, m/10}) \\ &\quad \times \nu \left\{ y \in B^m \colon |y_k| > \frac{3}{8\sqrt{m}}, \text{ for at least } \frac{m}{10} \text{ coordinates } k \right\}. \end{aligned}$$

Using Lemma 3 we have

$$\Sigma \geqslant (1 - 2^{-(1-c_7)\, m/10})(1 - 3e^{-c_9 m}) \geqslant 1 - e^{-c_{12} m}$$

for some absolute constants c_9, $c_{12} > 0$. Finally, from before, $m = 2^{dN^*}$ and $\rho = 2^{-rN^*}$ then $\rho \asymp m^{-r/d}$. Also, the condition of Lemma 1 has $m = c_5 n^{d/(d-1)}$ thus $\rho = c_{13}/n^{r/(d-1)}$ and therefore

$$\mu \left\{ f \in \mathscr{B}^r \colon \operatorname{dist}(f, M_n, L_2) > \frac{c_{13}}{4n^{r/(d-1)}} \right\} \geqslant 1 - e^{-\alpha(n)},$$

where $\alpha(n) = c_5 c_{11} n^{d/(d-1)}$. This completes the proof of Theorem 1. ∎

4.2. *Proof of Theorem* 2

Let $\mathscr{P}_s$ and $\mathscr{P}_s^h$ be as defined in Section 2. Choose n such that $n = \dim(\mathscr{P}_s^h)$. Then from Proposition 2 of [12] it follows that $\mathscr{P}_s \subset M_n$. Let N' be the integer such that $2^{N'-1} \leqslant s \leqslant 2^{N'}$. Considering the definition of $\mathscr{B}^r$ we have for all $f \in \mathscr{B}^r$, $f(x) = \sum_{N=0}^{\infty} \sum_{(s,l) \in \Delta_N} c_{s,l} p_{s,l}(x)$ and

$$\operatorname{dist}(f, M_n, L_2)^2 \leqslant \operatorname{dist}(f, \mathscr{P}_s, L^2)^2 \leqslant \left\| \sum_{N \geqslant N'} \sum_{(s,l) \in \Delta_N} c_{s,l} p_{s,l} \right\|_{L_2}^2.$$

Therefore from the Parseval equality and the definition of class $\mathscr{B}^r$ we obtain

$$\operatorname{dist}(f, M_n, L_2)^2 \leqslant \sum_{N \geqslant N'} \Big\| \sum_{(s,l) \in \Delta_N} c_{s,l} p_{s,l} \Big\|_{L_2}^2$$
$$\leqslant \sum_{N \geqslant N'} 2^{-2rN} \leqslant c_{14} 2^{-2rN'} = c_{14} s^{-2r},$$

for some constant $c_{14} > 0$. It is known (cf. [22]) that $\dim(\mathscr{P}_s^h) = \binom{d+s-1}{d-1} \asymp s^{d-1}$. Since $n \asymp s^{d-1}$ then $s \asymp n^{1/(d-1)}$. Thus

$$\operatorname{dist}(f, M_n, L_2) \leqslant c_{15} n^{-r/(d-1)},$$

which proves the theorem.

ACKNOWLEDGMENTS

The authors are extremely grateful to Allan Pinkus for helpful discussions.

REFERENCES

1. A. R. Barron, Approximation and estimation bounds for artificial neural networks, *Mach. Learning* **14** (1994), 115–133.
2. R. A. DeVore, K. Oskolkov, and P. Petrushev, Approximation by feed-forward neural networks, *Ann. Numer. Math.* **4** (1997), 261–287.
3. W. Hoeffding, Probability inequalities for sums of bounded random variables, *J. Amer. Statist. Assoc.* **58** (1963), 13–30.
4. F. Girosi and T. Poggio, Networks and the best approximation property, *Biolog. Cybernetics* **63** (1990), 169–176.
5. F. Girosi and G. Anzelloti, Rates of convergence for radial basis functions and neural networks, *in* "Artificial Neural Networks for Speech and Vission," Chapman & Hall, London, 1993.
6. F. Girosi, Approximation error bounds that use VC-bounds, *in* "Proceedings of the International Conference on Neural Networks, Paris, October 1995," pp. 295–302.
7. P. Levy, "Problems Concrets D'Analyse Fontionnelle," Gauthier–Villars, Paris, 1951.
8. V. Ya. Lin and A. Pinkus, Fundamentality of ridge functions, *J. Approx. Theory* **75** (1993), 295–311.
9. V. Ya. Lin and A. Pinkus, Approximation of multivariate functions, *in* "Advances in Computational Mathematics, New Delhi, 1993," World Scientific, Singapore, 1994.
10. G. G. Lorentz, M. Golitschek, and Y. Makovoz, "Constructive Approximation, Advanced Problems," Springer-Verlag, New York/Berlin, 1996.
11. V. Maiorov, Linear widths of function spaces equipped with the Gaussian measure, *J. Approx. Theory* **77** (1994), 74–88.
12. V. Maiorov, On best approximation by ridge functions, *J. Approx. Theory* **99** (1999), 68–94.

13. V. Maiorov and R. Meir, On the near optimality of the stochastic approximation of smooth functions by neural networks, *Adv. Comput. Math.*, to appear.
14. V. E. Maiorov and G. W. Wasilkowski, Probabilistic and average linear widths in L_∞-norm with respect to r-fold Wiener measure, *J. Approx. Theory* **84** (1996), 31–40.
15. H. N. Mhaskar, Neural networks for optimal approximation of smooth and analytic functions, *Neural Computa.* **8** (1996), 164–177.
16. S. M. Nikolski, "Approximation of the Many Variables Functions and Theorems of Embedding," Nauka, Moscow, 1969.
17. I. S. Gradshteyn and I. M. Ryzhik, "Table of Integrals, Series, and Products," 1980.
18. K. I. Oskolkov, Ridge approximation, Chebyshev–Fourier analysis and optimal quadrature formulas, *Proc. Steklov Inst. Math.* **219** (1997), 265–280.
19. P. P. Petrushev, Approximation by ridge functions and neural networks, *SIAM*, *J. on Math. Analysis*, **30** (1998), 155–189.
20. A. Pinkus, "*n*-Widths in Approximation Theory," Springer-Verlag, Berlin, 1985.
21. A. N. Shiryayev, "Probability," Springer-Verlag, Berlin, 1984.
22. E. M. Stein and G. Weis, "Introduction to Fourier Analysis on Euclidean Spaces," Princeton Univ. Press, Princeton, NJ, 1971.
23. V. Temlyakov, On approximation by ridge functions, preprint, Dept. of Math., University of South Carolina, 1996, 12 pp.
24. A. F. Timan, "Theory of Approximation of Functions of a Real Variable," MacMillan, New York, 1963.
25. J. F. Traub, G. W. Wasilkowski, and H. Wozniakowski, "Information-Based Complexity," Academic Press, San Diego, 1988.
26. H. Triebel, "Theory of Function Spaces," Birkhäuser, Basel, 1983.
27. 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.* **4** (1993), 1588–1559.
28. H. E. Warren, Lower bounds for approximation by non-linear manifolds, *Trans. Amer. Math. Soc.* **133** (1968), 167–178.
29. H. Weyl, Harmonics of homogeneous manifolds, *Ann. of Math.* **35** (1934), 486–499.