Constr. Approx. (1999) 15: 291–300

CONSTRUCTIVE APPROXIMATION

On the Degree of Approximation by Manifolds of Finite Pseudo-Dimension

V. Maiorov and J. Ratsaby

Abstract. The pseudo-dimension of a real-valued function class is an extension of the VC dimension for set-indicator function classes. A class $\mathcal{H}$ of finite pseudo-dimension possesses a useful statistical smoothness property. In [10] we introduced a nonlinear approximation width $\rho_n(\mathcal{F}, L_q) = \inf_{\mathcal{H}^n} \operatorname{dist}(\mathcal{F}, \mathcal{H}^n, L_q)$ which measures the worst-case approximation error over all functions $f \in \mathcal{F}$ by the best manifold of pseudo-dimension n. In this paper we obtain tight upper and lower bounds on $\rho_n(W_p^{r,d}, L_q)$, both being a constant factor of $n^{-r/d}$, for a Sobolev class $W_p^{r,d}$, $1 \le p, q \le \infty$. As this is also the estimate of the classical Alexandrov nonlinear n-width, our result proves that approximation of $W_p^{r,d}$ by the family of manifolds of pseudo-dimension n is as powerful as approximation by the family of all nonlinear manifolds with continuous selection operators.

1. Introduction

Vapnik and Chervonenkis [12], [13], [14] and later Blumer, Ehrenfeucht, Haussler, and Warmuth [2] studied the classical problem of pattern recognition in which they obtained results concerning uniform strong law convergence for families of indicator, as well as real-valued, functions. As a consequence of their theory a new measure of richness of classes of indicator functions, called the Vapnik–Chervonenkis (VC) dimension, was introduced. Henceforth, let X be an arbitrary space equipped with a probability measure. The VC dimension is defined as follows:

Definition 1 (Vapnik–Chervonenkis Dimension)**.** Given a class $\mathcal{H}$ of indicator functions of sets in X the Vapnik–Chervonenkis dimension of $\mathcal{H}$, denoted as VC($\mathcal{H}$), is defined as the largest integer m such that there exists a sample $x^m = \{x_1, \ldots, x_m\}$, with $x_i \in X$, $1 \le i \le m$, such that the cardinality of the set of sign vectors $S_{x^m}(\mathcal{H}) = \{[h(x_1), \ldots, h(x_m)] : h \in \mathcal{H}\}$ satisfies $|S_{x^m}(\mathcal{H})| = 2^m$. If m is arbitrarily large, then the VC dimension of $\mathcal{H}$ is infinite.

Date received: March 12, 1997. Dates revised: August 26, 1997, October 24, 1997, March 16, 1998, June 15, 1998. Date accepted: June 25, 1998. Communicated by Vladimir N. Temlyakov.
AMS classification: 41A45, 41A46, 42A10, 42A61.
Key words and phrases: Nonlinear widths, Pseudo-Dimension, Sobolev class.

Example. Let $\mathcal{H}$ be the class of indicator functions of interval sets on $X = \mathbf{R}$. With a single point $x_1 \in X$ we have $|\{[h(x_1)] : h \in \mathcal{H}\}| = 2$. For two points $x_1, x_2 \in X$ we have $|\{[h(x_1), h(x_2)] : h \in \mathcal{H}\}| = 4$. When $m = 3$, for any points $x_1, x_2, x_3 \in X$ we have $|\{[h(x_1), h(x_2), h(x_3)] : h \in \mathcal{H}\}| < 2^3$ thus $\mathrm{VC}(\mathcal{H}) = 2$. ■

Pollard [7] and later Haussler [4] have extended the uniform SLN results to classes of real-valued functions. In this case, the complexity measure analogous to the VC dimension is the so-called *pseudo-dimension*, denoted as $\dim_p(\mathcal{H})$ which is defined as follows: Let $\mathrm{sgn}(y)$ be defined as 1 for $y > 0$ and -1 for $y \leq 0$. For a Euclidean vector $v \in \mathbf{R}^m$ denote by $\mathrm{sgn}(v) = [\mathrm{sgn}(v_1), \ldots, \mathrm{sgn}(v_m)]$.

Definition 2 (Pseudo-Dimension)**.** Given a class $\mathcal{H}$ of real-valued functions defined on X. The pseudo-dimension of $\mathcal{H}$, denoted $\dim_p(\mathcal{H})$, is the largest integer m such that there exist $x_1, \ldots, x_m \in X$ and a vector $v \in \mathbf{R}^m$ for which the cardinality of the set of sign vectors satisfies $\{\mathrm{sgn}[h(x_1) + v_1, \ldots, h(x_m) + v_m] : h \in \mathcal{H}\}$ is equal to 2^m. If m can be arbitrarily large, then $\dim_p(\mathcal{H}) = \infty$.

Example. Consider $X = \{1, \ldots, n\}$. Consider the family of indicator functions $H = \{h(x) = 1_{\{x \in A\}} : A \subset X\}$. Then $\mathrm{VC}(H) = n$. Now consider the family of real-valued functions $G = \{g(x) = h(x) + x : h \in H\}$. Then, since there exists a $v = (-1, -2, \ldots, -n)$ such that $|\{\mathrm{sgn}[g(1) + v_1, \ldots, g(n) + v_n] : g \in G\}| = 2^n$, it follows that $\dim_p(G) = n$. Let $F = \{f(x) = 1_{\{g(x) > 0\}} : g \in G\}$. Then $\mathrm{VC}(F) = 0$. ■

For several useful invariance properties of the pseudo-dimension, see Pollard [8] and Haussler [4, Theorem 5]. We mention two useful properties. The first, appearing as Theorem 4 in Haussler [4], states that for the case of finite-dimensional vector spaces of functions the pseudo-dimension equals its dimension.

Property 1. *Let $\mathcal{H}$ be an n-dimensional vector space of functions from a set X into* $\mathbf{R}$. *Then* $\dim_p(\mathcal{H}) = n$.

The second property is attributed to Vapnik and Chervonenkis [12] and appears as Proposition A2.1(ii) of Blumer, Ehrenfeucht, Haussler, and Warmuth, which we restate for real-valued functions. For nonnegative integers $i > m$ we take $\binom{m}{i} = 0$.

Property 2. *Let $\mathcal{H}^n$ be a class of functions from X into* $\mathbf{R}$ *of pseudo-dimension* $n \geq 1$. *Let $m \geq 0$. Then for any sample $x^m = \{x_1, \ldots, x_m\}$, $x_i \in X$, $1 \leq i \leq m$, and vector $v \in \mathbf{R}^m$, the cardinality of the set* $S_{x^m}(\mathcal{H}^n) = \{\mathrm{sgn}[h(x_1)+v_1, \ldots, h(x_m)+v_m] : h \in \mathcal{H}^n\}$ *satisfies*

$$\left|S_{x^m}(\mathcal{H}^n)\right| \leq \sum_{i=0}^{n} \binom{m}{i} \leq m^n + 1.$$

This property follows from the next argument: denote by $1_{\{x \in D\}}$ the indicator function for the set $D \subset X$, i.e., it equals 1 if $x \in D$ and 0 otherwise. Since $\dim_p(\mathcal{H}^n) \leq n$ then there does not exist a set $x_1, \ldots, x_m \in X$ and $y_1, \ldots, y_m \in \mathbf{R}$, $m > n$, such that

$$A = \left|\{\text{sgn}[h(x_1) + y_1, h(x_2) + y_2, \ldots, h(x_m) + y_m] : h \in \mathcal{H}^n\}\right| = 2^m.$$

Moreover, A equals

$$B = \left|\{[1_{\{h(x_1)+y_1 \geq 0\}}, 1_{\{h(x_2)+y_2 \geq 0\}}, \ldots, 1_{\{h(x_m)+y_m \geq 0\}}] : h \in \mathcal{H}^n\}\right|.$$

We define the function $g_h(z) = g_h(x, y) \equiv h(x) + y$, and we write $z = (x, y)$. We consider the class of indicator functions

$$\mathcal{G} = \{1_{\{g_h(z) \geq 0\}} : h \in \mathcal{H}^n\}.$$

It follows from the above that there does not exist a set $z_1, \ldots, z_m \in X \times \mathbf{R}$ with $m > n$ such that

$$C = \left|\{[1_{\{g_h(z_1) \geq 0\}}, \ldots, 1_{\{g_h(z_m) \geq 0\}}] : h \in \mathcal{H}^n\}\right| = 2^m.$$

Then by definition of the VC dimension for indicator classes it follows that $\text{VC}(\mathcal{G}) \leq n$. Using Proposition A2.1(ii) of Blumer et al., we obtain $C \leq m^n + 1$ which is true for all $m \geq 0$ and $n \geq 1$. Since $A = B = C$ it follows that $A \leq m^n + 1$, which proves Property 2.

Consider some normed space $\mathcal{F}$ consisting of functions $f(x)$, $x \in X$. In Ratsaby and Maiorov [9], [10] we introduced a new nonlinear n width of a subset F of a space $\mathcal{F}$ defined as

(1) $$\rho_n(F, \mathcal{F}) \equiv \inf_{\mathcal{H}^n} \sup_{f \in F} \inf_{h \in \mathcal{H}^n} \|f - h\|_{\mathcal{F}},$$

where $\mathcal{H}^n$ runs over all classes in $\mathcal{F}$ with $\dim_p(\mathcal{H}^n) \leq n$.

Let us compare this width to the classical Alexandrov nonlinear width, see Tikhomirov [11] and DeVore [3]. Let $\|\cdot\|$ be a norm on $\mathcal{F}$. Let M_n be a mapping from $\mathbf{R}^n$ into the Banach space $\mathcal{F}$ which associates each $a \in \mathbf{R}^n$ with the element $M_n(a) \in \mathcal{F}$. Functions $f \in \mathcal{F}$ are approximated by functions in the manifold $\mathcal{M}_n = \{M_n(a) : a \in \mathbf{R}^n\}$. The measure of approximation of f by $\mathcal{M}_n$ is defined as the distance $\inf_{a \in \mathbf{R}^n} \|f - M_n(a)\|$. The degree of approximation of a subset F of $\mathcal{F}$ by $\mathcal{M}_n$ is defined as $\sup_{f \in F} \inf_{a \in \mathbf{R}^n} \|f - M_n(a)\|$. Denote by r a selection operator which takes an element $f \in F$ to $\mathbf{R}^n$. Given such an operator r then the approximation of f by a manifold $\mathcal{M}_n$ is $M_n(r(f))$. The distance between the set F and the manifold $\mathcal{M}_n$ is then defined as $\sup_{f \in F} \|f - M_n(r(f))\|$. Restricting r to be a continuous operator, the continuous *nonlinear* n width of F is then defined as $a_n(F; \mathcal{F}) = \inf_{r:\text{cont.}, \mathcal{M}_n} \sup_{f \in F} \|f - M_n(r(f))\|$, where the infimum is taken over all continuous selection operators r and all n-dimensional manifolds $\mathcal{M}_n$. This width is determined for various F and $\mathcal{F}$ in [3].

The Alexandrov nonlinear width does not in general reflect the degree of approximation of the more natural selection operator r which chooses the best approximation for an $f \in F$ as its closest element in $\mathcal{M}_n$, i.e., that whose distance from f equals $\inf_{g \in \mathcal{M}_n} \|f - g\|$, the reason being that such an r is not necessarily continuous. In contrast, the ρ_n width imposes no restriction as far as the selection mechanism is concerned, namely, it selects for f an element $h(f)$ where $\|f - h(f)\| = \inf_{h \in \mathcal{H}^n} \|f - h\|$. Many interesting manifolds such as rational functions and splines with free knots are included in the family $\{\mathcal{H}^n\}$ of all n-pseudo-dimensional manifolds.

Introduce definitions. Let $X = [0, 1]^d$ be the unit cube in the space $\mathbf{R}^d$, $d \geq 1$. Consider a normed space $L_q = L_q(X)$, $1 \leq q \leq \infty$, consisting of functions $f(x)$, $x \in X$, such that

$$\|f\|_{L_q} = \left(\int_X |f(x)|^q \, dx \right)^{1/q} < \infty.$$

In [10] we estimated this width in the L_∞-metric for a Sobolev class

$$W_\infty^{r,d} = \left\{ f : \|D^k f\|_{L_\infty([0,1]^d)} \leq 1, |k| \leq r \right\},$$

where for $k = [k_1, \ldots, k_d] \in \mathbf{Z}^d$ the norm $|k| = \sum_{i=1}^d k_i$, and $D^k f = (\partial^{k_1+\cdots+k_d})/(\partial x_1^{k_1} \cdots \partial x_d^{k_d}) f$. In the current work we estimate ρ_n for a general Sobolev class embedded in the space L_q which is defined as follows: Let $r > 0$, $1 \leq p, q \leq \infty$, and if $p \leq q$, and let the condition $r/d > 1/p - 1/q$ be satisfied. Then

$$W_p^{r,d} \equiv \left\{ f : \|D^k f\|_{L_p[0,1]^d} \leq 1, |k| \leq r \right\}.$$

2. Main Results

The main result is the following two theorems. The previously mentioned embedding condition on r, d, p, and q is written more concisely as $r/d > (1/p - 1/q)_+$ where $(y)_+ = 0$ if $y \leq 0$ and $(y)_+ = y$ for $y > 0$.

Theorem 1 (Lower Bound). *For $r > 0$, $1 \leq p, q \leq \infty$, and integers $1 \leq n, d \leq \infty$, we have*

$$\rho_n(W_p^{r,d}, L_q) \geq \frac{c_1}{n^{r/d}}$$

for some finite constant $c_1 > 0$ independent of n.

Theorem 2 (Upper Bound). *For $r > 0$, $1 \leq p, q \leq \infty$ satisfying $r/d > (1/p - 1/q)_+$, and integers $1 \leq n, d \leq \infty$, we have*

$$\rho_n(W_p^{r,d}, L_q) \leq \frac{c_2}{n^{r/d}}$$

for some finite constant $c_2 > 0$ independent of n.

We now proceed with the proofs of the two theorems.

3. Proofs

We start with some notation. Let $c_3, c_4, \ldots$ denote absolute constants. For the distance between two function classes $\mathcal{A}, \mathcal{B} \subset L_q$ we use $\operatorname{dist}(\mathcal{A}, \mathcal{B}, L_q) = \sup_{\{a \in \mathcal{A}\}} \inf_{\{b \in \mathcal{B}\}} \|a - b\|_{L_q}$. The l_p^m-norm of a vector $v \in \mathbf{R}^m$ is denoted by $\|v\|_{l_p^m} = \left(\sum_{i=1}^m v_i^p\right)^{1/p}$, $1 \leq p \leq \infty$. For a set $G \subset \mathbf{R}^m$ write $\operatorname{sgn}(G) = \{\operatorname{sgn}(z) : z \in G\}$.

3.1. *Proof of Theorem* 1

Denote by $E_m = \{[z_1, \ldots, z_m] : z_i \in \{-1, +1\}, 1 \le i \le m\}$. The next claim follows immediately from Lorentz, Yolitschek, and Makovoz [6, p. 489].

Claim 1. *There exists a set $G \subset E_m$ of cardinality greater than or equal to $2^{m/16}$ such that for any $v, v' \in G$, where $v \neq v'$, the distance $\|v - v'\|_{l_1^m} \ge m/2$.*

In order to find a lower bound on $\rho_n(W_p^{r,d}, L_q)$ it suffices to bound $\rho_n(F_m, L_1)$ from below, where $F_m \subset W_p^{r,d}$ is a function class which is constructed next. For $y \in \mathbf{R}$, let $\varphi(y)$ be a nonnegative function in $W_p^{r,1}$ which satisfies: $|\varphi(y)| \le 1$, $\varphi(y) = 0$ for $y \notin (0, 1)$, and $\varphi(y) = 1$ for $y \in [\frac{1}{4}, \frac{3}{4}]$.

Let m be a fixed positive integer such that $m = \tilde{m}^d$ for some arbitrary integer $\tilde{m}$ which will be chosen below. Let $D = \{0, 1, \ldots, \tilde{m} - 1\}^d$. For $x \in \mathbf{R}^d$, let $\bar{i} = [i_1, i_2, \ldots, i_d] \in D$ define the function $\varphi_{\bar{i}}(x) = \prod_{j=1}^d \varphi_{i_j}(x_j)$, where for $y \in \mathbf{R}$, $\varphi_{i_j}(y) = \varphi(\tilde{m}y - i_j)$, $0 \le i_j \le \tilde{m} - 1$, $1 \le j \le d$.

Consider the function class

$$F_m = \left\{ f_a(x) = \frac{1}{m^{r/d}} \sum_{\bar{i} \in D} a_{\bar{i}} \varphi_{\bar{i}}(x) : a \in E_m \right\}, \tag{2}$$

where again $E_m = \{a = [a_1, \ldots, a_m] : a_i \in \{-1, +1\}, 1 \le i \le m\}$ and we take the liberty in using a scalar as well as a vector index for a coordinate of the vector a.

Claim 2. *We have $F_m \subset W_p^{r,d}$.*

We now prove the claim. For a multi-integer $\alpha \in \mathbf{Z}_+^d$, satisfying $|\alpha| = \sum_{i=1}^d \alpha_i \le r$, denote by $f^{(\alpha)}$ the partial derivative of order α. Let $\Delta = [0, 1/\tilde{m}]^d$. All integrals below are d-dimensional. We have

$$\| f_a^{(\alpha)} \|_{L_p}^p = \frac{1}{m^{rp/d}} \int_{[0,1]^d} \left| \sum_{\bar{i} \in D} a_{\bar{i}} (\varphi_{\bar{i}}(x))^{(\alpha)} \right|^p dx \tag{3}$$

$$= \frac{1}{m^{rp/d}} \sum_{\bar{i} \in D} \int_{\Delta} \left| a_{\bar{i}} (\varphi(\tilde{m}x))^{(\alpha)} \right|^p dx. \tag{4}$$

Using the fact that $|\alpha| \le r$ and letting $y_i = \tilde{m} x_i$, $1 \le i \le d$, then (4) is bounded from above by

$$\frac{1}{m^{rp/d+1}} \sum_{\bar{i} \in D} \tilde{m}^{rp} \left| a_{\bar{i}} \right|^p \int_{[0,1]^d} \left| \prod_{j=1}^d \varphi^{(\alpha_j)}(y_j) \right|^p dy. \tag{5}$$

The above integral is less than or equal to 1 since $\varphi \in W_p^{r,1}$ thus (5) reduces to

$$\frac{1}{m^{rp/d+1}} m^{rp/d} \sum_{\bar{i} \in D} \left| a_{\bar{i}} \right|^p = 1,$$

since $a_{\bar{i}} \in E_m$. This proves that $f_a \in W_p^{r,d}$. ∎

We now have the following claim:

Claim 3. *Let $G \subset E_m$ be a subset as defined in Claim* 1. *Denote by $F_m(G) = \{f_a(x) : a \in G\}$. Then for any $f \neq f' \in F_m(G)$*

$$\|f - f'\|_{L_1} \geq \frac{c_3}{m^{r/d}},$$

where $c_3 = 2^{-d-1}$.

The proof follows: Let $f \neq f'$ be such that $f = f_a(x)$, $f' = f_{a'}(x)$. Then from the definition of $\varphi_{\bar{i}}(x)$ and from Claim 1 we have

$$\begin{aligned}\|f - f'\|_{L_1} &= \frac{1}{m^{r/d}} \int_{[0,1]^d} \left| \sum_{\bar{i} \in D} (a_{\bar{i}} - a'_{\bar{i}}) \varphi_{\bar{i}}(x) \right| dx \\ &= \left(\frac{1}{m^{r/d}} \int_{\Delta} \left| \prod_{j=1}^{d} \varphi(\tilde{m} x_j) \right| dx \right) \sum_{\bar{i} \in D} |a_{\bar{i}} - a'_{\bar{i}}| \\ &\geq \frac{1}{m^{r/d+1}} \left(\int_0^1 |\varphi(y)|\, dy \right)^d \frac{m}{2} \\ &\geq \frac{c_3}{m^{r/d}}. \qquad \blacksquare\end{aligned}$$

For a set of functions $\mathcal{F}$ in L_1 denote by

$$\mathcal{M}_\varepsilon(\mathcal{F}) = \max\{s : \exists\, f_1, \ldots, f_s \in L_1,\ \|f_i - f_j\|_{L_1} \geq \varepsilon,\ \forall\, i \neq j\}$$

the ε-packing number for $\mathcal{F}$ in the L_1-norm.

The next lemma follows directly from Haussler [5, Corollary 3].

Lemma 1. *Let $\mathcal{H}^n = \{h\}$ be a set of Lebesgue-measurable functions on $[0,1]^d$ such that $\|h\|_{L_\infty} \leq \beta$ and $\dim_p(\mathcal{H}^n) \leq n < \infty$. Then for any $\varepsilon > 0$ the following upper bound on the ε-packing number holds*:

$$\mathcal{M}_\varepsilon(\mathcal{H}^n) \leq e(n+1) \left(\frac{4e\beta}{\varepsilon} \right)^n.$$

We now proceed with the proof of Theorem 1.

Proof. Let $\mathcal{H}^n$ be any set of Lebesgue-measurable functions with $\dim_p(\mathcal{H}^n) \leq n$.

For the set $G \subset E_m$ defined in Claim 1 consider the set

$$F_m(G) = \{f_a(x) \in F_m : a \in G\}.$$

Let $\varepsilon > 0$ be any positive real number. Denote

$$\delta = \sup_{f \in F_m(G)} \inf_{h \in \mathcal{H}^n} \|f - h\|_{L_1} + \varepsilon = \operatorname{dist}(F_m(G), \mathcal{H}^n, L_1) + \varepsilon.$$

Define the projection operator $P : F_m(G) \to \mathcal{H}^n$, as follows: For any $f \in F_m(G)$ let

Pf be any element in $\mathcal{H}^n$ such that

$$\|f - Pf\|_{L_1} \le \delta.$$

Set $\beta = m^{-r/d}$. Introduce the cut operator

$$Cf := Cf(x) = \begin{cases} -\beta, & f(x) < -\beta, \\ f(x), & -\beta \le f(x) \le \beta, \\ \beta, & f(x) > \beta. \end{cases}$$

Consider the set of functions $\mathcal{S} := CP(F_m(G)) := \{CPf : f \in F_m(G)\}$. Let $f \neq f' \in F_m(G)$. Then

$$\begin{aligned}\|CPf - CPf'\|_{L_1} &= \|(CPf - f) + (f' - CPf') + (f - f')\|_{L_1} \\ &\ge \|f - f'\|_{L_1} - \|f' - CPf'\|_{L_1} - \|f - CPf\|_{L_1}.\end{aligned}$$

If $|h(x)| \le \beta$ for all $x \in X$, then $\|h - CPh\|_{L_1} \le \|h - Ph\|_{L_1}$. Therefore, using Claim 3 we have

$$\|CPf - CPf'\|_{L_1} \ge \|f - f'\|_{L_1} - \|f' - Pf'\|_{L_1} - \|f - Pf\|_{L_1} \ge \frac{c_3}{m^{r/d}} - 2\delta.$$

Now we assume that $\delta \le c_3/4m^{r/d}$. From the above inequality and Claim 1 it follows that for any $g, g' \in \mathcal{S}$, $g \neq g'$,

$$\|g - g'\|_{L_1} \ge \frac{c_3}{2m^{r/d}},$$

and the cardinality $|\mathcal{S}| = 2^{m/16}$. Fix $\alpha = c_3/2m^{r/d}$. Then

$$\mathcal{M}_\alpha(\mathcal{S}) \ge 2^{m/16}. \tag{6}$$

From the other hand, we have $\|g\|_{L_\infty} \le \beta$ for any $g \in \mathcal{S}$. From Definition 2 of pseudo-dimension it directly follows that $\dim_p(CP(F_m(G))) \le \dim_p(P(F_m(G)))$. Since $P(F_m(G)) \subset \mathcal{H}^n$, then $\dim_p(P(F_m(G))) \le \dim_p(\mathcal{H}^n) \le n$. Hence $\dim_p(\mathcal{S}) = \dim_p(CP(F_m(G))) \le n$. According to Lemma 1 we have

$$\mathcal{M}_\alpha(\mathcal{S}) \le e(n+1)\left(\frac{4e\beta}{\alpha}\right)^n. \tag{7}$$

Recall that $\beta = m^{-r/d}$, $c_3 = 2^{-d-1}$, and $\alpha = c_3 m^{-r/d}/2 = m^{-r/d}/2^{d+1}$. From (6) and (7) we obtain the inequality

$$2^{m/16} \le e(n+1)\left(\frac{4e\beta}{\alpha}\right)^n = e(n+1)(2^{d+4}e)^n.$$

Let $\gamma_n = [32\log_2(2^{d+4}e)]n$. Recall that $m = \tilde{m}^d$. Choose the integer $\tilde{m}$ such that $\gamma_n^{1/d} \le \tilde{m} \le 2\gamma_n^{1/d}$, which is possible since $\gamma_n^{1/d} > 1$. Then the last inequality implies that

$$2\log_2(2^{d+4}e) \le \frac{\log_2(e(n+1))}{n} + \log_2(2^{d+4}e)$$

which is false for all $n \ge 1$. It follows that the assumption of $\delta \le c_3/4m^{r/d}$ is contradicted for any $n \ge 1$. Hence $\delta > c_3/4m^{r/d} \equiv c_1/n^{r/d}$. According to the definition of δ, ε is any positive number, so it follows that $\mathrm{dist}(F_m(G), \mathcal{H}^n, L_1) \ge c_1/n^{r/d}$. Using Claim 2 we obtain

$$\mathrm{dist}(W_p^{r,d}, \mathcal{H}^n, L_q) \ge \mathrm{dist}(W_p^{r,d}, \mathcal{H}^n, L_1) \ge \mathrm{dist}(F_m, \mathcal{H}^n, L_1) \ge \frac{c_1}{n^{r/d}}$$

which proves the theorem. ■

3.2. *Proof of Theorem* 2

To establish an upper bound it suffices to consider a specific manifold of pseudo-dimension n and use the L_∞-metric for approximation. For a positive integer n consider the family Ξ_n of possible partitions of the domain $X = [0,1]^d$ attained by the constructive spline algorithm of Birman and Solomjak [1]. We will not describe the algorithm here but only mention the particular properties of the family of partitions obtained by the algorithm. Let Π be a partition of X into a finite number of half-open d-dimensional cubes $\Delta_k = \{x \in \mathbf{R}^d : a_{k,i} \le x_i < b_{k,i}, 1 \le i \le d\}$, where $a_{k,i}, b_{k,i} \in [0,1]$. Let the cardinality of Π, denoted $|\Pi|$, be the number of cubes in the partition Π. A partition Π' which is obtained from Π by dividing certain cubes Δ_k into 2^d cubes is called an elementary extension of Π. The class Ξ_n consists of all partitions of cardinality n which can be obtained from the trivial partition $\Pi_0 = X = [0,1]^d$ by a finite number of elementary extensions.

We consider the approximating manifold $\mathcal{G}^n$ to be comprised of all functions g which are piecewise polynomials of degree $r-1$ over the n cubes of any partition Π_n in the above family Ξ_n. Specifically, denote by $1_{\Delta_k}(x)$ the characteristic function over Δ_k and let $P(\Delta_k)$ be the space of all functions on X of the form $p(x)1_{\Delta_k}(x)$ where $p(x)$ is an algebraic polynomial of x of degree at most $r-1$. Associated with a partition Π_n define the class $P(\Pi_n)$ consisting of all functions $g(x) = \sum_{k=1}^n p_k(x)$ where $p_k \in P(\Delta_k)$, $\Delta_k \in \Pi_n$, $1 \le k \le n$.

Now, according to Theorem 3.1, p. 305, of Birman and Solomjak [1], for any $f \in W_p^{r,d}$ there exists a partition $\Pi_{n,f} \in \Xi_n$ of X and an associated class $P(\Pi_{n,f})$, both dependent on f, such that

$$\text{dist}(f, P(\Pi_{n,f}), L_\infty) = \inf_{g \in P(\Pi_{n,f})} \|f - g\|_{L_\infty} \le \frac{c}{n^{r/d}} \tag{8}$$

for some constant $c > 0$ independent of n. As this is true for any $f \in W_p^{r,d}$ then it holds also for the worst function $f \in W_p^{r,d}$, i.e.,

$$\sup_{f \in W_p^{r,d}} \inf_{g \in P(\Pi_{n,f})} \|f - g\|_{L_\infty} \le \frac{c}{n^{r/d}}.$$

Consider the manifold

$$\mathcal{G}_n = \bigcup_{\Pi_n \in \Xi_n} P(\Pi_n).$$

We now find the upper bound for the pseudo-dimension of $\mathcal{G}_n$. Let m be a positive integer. Let $x_1, \ldots, x_m \in X$ be any set of m points in X and let v be any vector in $\mathbf{R}^m$. Denote by $\overline{g} = [g(x_1)+v_1, \ldots, g(x_m)+v_m]$ and $\text{sgn}(\overline{g}) = [\text{sgn}(g(x_1)+v_1), \ldots, \text{sgn}(g(x_m)+v_m)]$. We have,

$$\begin{aligned}\left|\{\text{sgn}(\overline{g}) : g \in \mathcal{G}^n\}\right| &= \left|\bigcup_{\Pi_n \in \Xi_n} \{\text{sgn}(\overline{g}) : g \in P(\Pi_n)\}\right| \\ &\le \sum_{\Pi_n \in \Xi_n} \left|\{\text{sgn}(\overline{g}) : g \in P(\Pi_n)\}\right|.\end{aligned}$$

As all cubes in Π_n are mutually disjoint the last expression equals

$$\sum_{\Pi_n \in \Xi_n} \prod_{\Delta_k \in \Pi_n} \left|\{\text{sgn}(\overline{p}_k) : p \in P(\Delta_k)\}\right|, \tag{9}$$

where $\overline{p}_k = [p(x_{i_1})+v_{i_1}, \ldots, p(x_{i_{m_k}})+v_{i_{m_k}}]$, $\{x_{i_1}, \ldots, x_{i_{m_k}}\}$ is the subset of $\{x_1, \ldots, x_m\}$ which is contained in Δ_k, and $\{y_{i_1}, \ldots, y_{i_{m_k}}\}$ are the corresponding y values.

The class $P(\Delta_k)$ is a vector space of functions of the form $p(x) = \sum_{i:|i|\le r-1} a_i x_1^{i_1} \cdots x_d^{i_d}$, $a_i \in \mathbf{R}$ for a multi-integer $i \in \mathbf{Z}_+^d$, where $|i| = \sum_{j=1}^d i_j$. The number of such terms is bounded from above by $\alpha \equiv c_4 2^{rd}$ for some absolute constant $c_4 > 0$. Therefore the dimension of the linear manifold $P(\Delta_k)$ is bounded from above by α which is independent of k and n. By Property 1, it follows that the pseudo-dimension of the class $P(\Delta_k)$ is bounded from above by α, for any $1 \le k \le n$.

For a fixed partition Π_n, it follows from Property 2 that $|\{\mathrm{sgn}(\overline{p}_k) : p \in P(\Delta_k)\}| \le m_k^\alpha + 1$, for each $1 \le k \le n$. Let C_n denote the number of partitions Π_n in Ξ_n. Then (9) is bounded from above by

$$
\begin{aligned}
C_n \prod_{k=1}^n (m_k^\alpha + 1) &\le C_n \left(\prod_{k=1}^n (m_k + 1)\right)^\alpha \\
&\le C_n \left(\frac{1}{n}\sum_{k=1}^n (m_k + 1)\right)^{n\alpha} \\
&= C_n \left(1 + \frac{m}{n}\right)^{n\alpha},
\end{aligned}
$$

where we used a well-known inequality relating the arithmetic and geometric means. We now show that $C_n \le 2^{c_5 n}$ for some constant $c_5 > 0$ independent of n.

As noted earlier, the family Ξ_n is defined in such a way that every partition $\Pi_n \in \Xi_n$ is the result of a finite sequence of elementary extensions starting from the trivial partition. Consider such a sequence of partitions $\{\Pi_{n_i}\}_{i=0}^k$ which starts from the trivial partition Π_{n_0} of cardinality $n_0 = 1$ and ends in a partition $\Pi_{n_k} \in \Xi_n$ of cardinality $n_k = n$. From p. 302 of [1] it follows that there are no more than 2^n possible cardinality sequences $n_0, \ldots, n_k$, corresponding to possible partition sequences. For a fixed cardinality sequence there are no more than $2^{\sum_{i=0}^k n_i}$ possible sequences of partitions $\{\Pi_{n_i}\}_{i=0}^k$ for which $|\Pi_{n_i}| = n_i$ and all elements n_i, $1 \le i \le k$, satisfy the inequality above equation (2.20) and Lemma 2.3, p. 301, in [1]. Using (8), it then follows that $\sum_{i=0}^k n_i \le c_6 n$ for some constant $c_6 > 0$ independent of n. Therefore the total number C_n of possible partition sequences $\{\Pi_{n_i}\}_{i=0}^k$ is no more than $2^n 2^{c_6 n} = 2^{c_5 n}$. Thus (10) is bounded from above by $2^{c_5 n}(1 + m/n)^{n\alpha}$. Solving for the m such that this last expression is strictly less than 2^m yields $m \le c_7 n$, for some constant $c_7 > 0$ independent of n. Thus we have proved that for $m \ge c_7 n$, for any set of points $x_1, \ldots, x_m \in X$ and $v_1, \ldots, v_m \in \mathbf{R}$, $|\{\mathrm{sgn}(\overline{g}) : g \in \mathcal{G}^n\}| < 2^m$ which proves that $\dim_p(\mathcal{G}^n) \le c_7 n$.

Hence the manifold $\mathcal{G}^{n/c_7}$ has pseudo-dimension n and we conclude by

$$
\begin{aligned}
\rho_n(W_p^{r,d}, L_q) &= \inf_{\mathcal{H}^n} \sup_{f \in W_p^{r,d}} \inf_{h \in \mathcal{H}^n} \|f - h\|_{L_q} \\
&\le \sup_{f \in W_p^{r,d}} \inf_{g \in \mathcal{G}^{n/c_7}} \|f - g\|_{L_\infty} \\
&\le \sup_{f \in W_p^{r,d}} \inf_{g \in P(\Pi_{n/c_7}, f)} \|f - g\|_{L_\infty},
\end{aligned}
$$

the latter being bounded from above by $c_2/n^{r/d}$ using (8) for some constant $c_2 > 0$ independent of n. This proves the theorem. ■

Acknowledgments. V. Maiorov acknowledges the support of the Ministry of Absorption of the state of Israel. J. Ratsaby acknowledges the support of the Ministry of Arts and Science of Israel and of the Ollendorff center of the Faculty of Electrical Engineering at the Technion. The authors are extremely grateful to Allan Pinkus for helpful discussions.

References

1. M. S. Birman, M. Z. Solomjak (1967): *Piecewise-polynomial approximations of functions of the classes* W_p^α. Math. USSR-Sb., **2**(3):295–317.
2. A. Blumer, A. Ehrenfeucht, D. Haussler, M. Warmuth (1989): *Learnability and the Vapnik–Chervonenkis dimension*. J. ACM, **36**(4):929–965.
3. R. A. DeVore (to appear): *Degree of nonlinear approximation*. In: Approximation Theory VI, Vol. 1 (C. K. Chui, L. L. Schumaker, J. D. Ward, eds.), pp. 175–201.
4. D. Haussler (1992): *Decision theoretic generalizations of the PAC model for neural net and other learning applications*. Inform. Comput., **100**(1):78–150.
5. D. Haussler (1995): *Sphere packing numbers for subsets of the Boolean n-cube with bounded Vapnik–Chervonenkis dimension*. J. Combin. Theory, Ser. A, **69**:217–232.
6. G. G. Lorentz, M. v. Golitschek, Y. Makovoz (1996): Constructive Approximation, Advanced Problems. New York: Springer-Verlag.
7. D. Pollard (1984): Convergence of Stochastic Processes. Springer Series in Statistics. New York: Springer-Verlag.
8. D. Pollard (1989): Empirical Processes: Theory and Applications. NSF-CAMS Regional Conference Series in Probability and Statistics, Vol. 2. Inst. of Math., Stat. and Am. Stat. Assoc.
9. J. Ratsaby, V. Maiorov (1997): *Generalization of the PAC-Model for learning with partial information*. In: Proceedings of the Third European Conference on Computational Learning Theory, EuroCOLT 97. New York: Springer-Verlag.
10. J. Ratsaby, V. Maiorov (1997): *On the value of partial information for learning from examples*. J. Complexity, **13**:509–544.
11. V. M. Tikhomirov (1976): Some Problems in Approximation Theory. Moscow: Moscow State University. (In Russian.)
12. V. N. Vapnik, A. Ya. Chervonenkis (1971): *On the uniform convergence of relative frequencies of events to their probabilities*. Theoret. Prob. Its Appl., **16**(2):264–280.
13. V. N. Vapnik, A. Ya. Chervonenkis (1981): *Necessary and sufficient conditions for the uniform convergence of means to their expectations*. Theoret. Prob. Its Appl., **26**(3):532–553.
14. V. N. Vapnik (1982): Estimation of Dependences Based on Empirical Data. Berlin: Springer-Verlag.

V. Maiorov
Department of Mathematics
Technion
Haifa 32000
Israel
maiorov@tx.technion.ac.il

J. Ratsaby
Department of Electrical Engineering
Technion
Haifa 32000
Israel
jer@ee.technion.ac.il