Vietnam Journal of Mathematics **28**:4 (2000) 381-385

Vietnam Journal
of
MATHEMATICS

Short Communication

On Entropy Numbers and Non-Linear Approximation by Sets of Finite Pseudo-Dimension

Dinh Dung

Institute of Information Technology
Nghia Do, Cau Giay, Hanoi, Vietnam

Received August 10, 1999

1. Let X be a quasi-normed linear space and W, $M \subset X$. For approximation of elements from W by M, the quantity

$$E(W,M,X) := \sup_{f\in W} \inf_{\varphi\in M} \|f-\varphi\|$$

gives the worst case error of approximation. When M is a linear manifold we have the linear approximation problem. Non-linear approximation is that when the set of the approximation M is a non-linear manifold.

Given a family $\mathcal{M}$ of subsets in X, we are interested in the best approximation by M from $\mathcal{M}$ in terms of the quantity

$$d(W,\mathcal{M},X) := \inf_{M\in\mathcal{M}} E(W,M,X). \tag{1}$$

In linear approximation, if $\mathcal{M}$ is the family of all linear manifolds of dimension $\leq n$, then $d(W,\mathcal{M},X)$ in (1) defines the well-known Kolmogorov n-width $d_n(W,X)$.

We are interested in non-linear approximation in terms of the entropy number $\varepsilon_n(W,X)$ and the non-linear n-width $\rho_n(W,X)$.

The quantity $d(W,\mathcal{M},X)$ is called entropy number $\varepsilon_n(W,X)$ if in (1) $\mathcal{M}$ is the family of all subsets of X such that $|M| \leq 2^n$, where $|M|$ denotes the cardinality of M. It is inverse to the ε-entropy $H_\varepsilon(W,X) := \log N_\varepsilon(W,X)$ where $N_\varepsilon(W,X)$ is the cardinality of the minimal ε-net of W. The ε-entropy $H_\varepsilon(W,X)$ was introduced by Kolmogorov and Tikhomirov [5].

The non-linear n-width $\rho_n(W,X)$ introduced recently by Ratsaby and Maiorov [7, 8] is defined only for a space X of real-valued functions on a set Ω. It is $d(W,\mathcal{M},X)$, if $\mathcal{M}$ in (1) is the family of all subsets of X of pseudo-dimension $\leq n$.

The notion of pseudo-dimension is defined as follows. For a real number t

let sgn(t) be 1 for $t > 0$ and -1 otherwise. For $x \in \mathbb{R}^n$ let $\text{sgn}(x) = \{\text{sgn}(x_1), \text{sgn}(x_2), ..., \text{sgn}(x_n)\}$. Let M be a set of real-valued functions defined on Ω. The pseudo-dimension of M is defined as the largest integer n such that there exist points $a^1, a^2, ..., a^n$ in Ω and $b \in \mathbb{R}^n$ such that the cardinality of the set $\{\text{sgn}(y) : y = \{f(a^1)+b_1, f(a^2)+b_2, ..., f(a^n)+b_n\}, f \in M\}$ is 2^n. If n is arbitrarily large then the pseudo-dimension of M is infinite. We denote the pseudo-dimension of M by $\dim_p(M)$.

The definition of pseudo-dimension of a real-valued functions set is introduced by Pollard [6] and later by Haussler [4] as an extension of the Vapnik–Chervonekis dimension [12] of an indicator function set. The pseudo-dimension defined above and the Vapnik–Chervonekis dimension are measures of capacity of function set. They play an important role in theory of pattern recognition and regression estimation, empirical processes and computational learning theory (see also [7, 8] for details).

If M is a linear manifold of dimension n in X, then $\dim_p(M) = n$ (see [4. 6]). From the definition we can see that $\dim_p(M) \le \log|M|$, and consequently,

$$\rho_n(W, X) \le \varepsilon_n(W, X) \tag{2}$$

for any subset W in the quasi-normed linear space X of real valued functions on Ω.

We establish in our paper asymptotic orders of the non-linear n-width ρ_n and entropy number ε_n in the space $L_q(\mathbb{T}^d)$ of the unit $\mathbf{SW}^r_q$ of Sobolev space and the unit ball of $\mathbf{SB}^r_{p,\theta}$ of the Besov space of functions on $\mathbb{T}^d$ with common mixed smoothness r.

2. For a nonnegative integer r, the univariate symmetric difference operator Δ^s_h, $h \in \mathbb{T}$, is defined inductively by $\Delta^s_h := \Delta^1_h \Delta^{s-1}_h$, starting from the operator

$$\Delta^1_h f := f(\cdot + h/2) - f(\cdot - h/2).$$

For a natural number s and $e \subset E := \{1, 2, ..., d\}$, we let the multivariable mixed s-th difference operator Δ^s_h, $h \in \mathbb{T}^d$, be defined by

$$\Delta^s_h(e) f := \prod_{j \in e}^{d} \Delta^s_{h_j} f,$$

where the univariate operator $\Delta^s_{h_j}$ is applied to the variable x_j (in particular, $\Delta^s_h(\emptyset) f \equiv f$).

For $r > 0$ and $1 \le p \le \infty$, $0 < \theta \le \infty$, let $\mathbf{B}^r_{p,\theta}$ denote the Besov space of all functions on $\mathbb{T}^d$, for which the quasi-norm

$$\|f\|_{\mathbf{B}^r_{p,\theta}} := \sum_{e \subset E} |f|_{B^{r,e}_{p,\theta}}$$

is finite, where $\|\cdot\|_p$ is the usual p-integral norm in $L_p := L_p(\mathbb{T}^d)$ and

$$|f|_{B^{r,e}_{p,\theta}} := \left(\int_{\mathbb{T}^d} \prod_{j \in e} |h|^{-1-\theta r} \|\Delta^s_h(e) f\|^\theta_p dh \right)^{1/\theta}, \ \theta < \infty$$

(the integral is changed to the supremum for $\theta = \infty$) for some $s > r$. The definition of $\mathbf{B}^r_{p,\theta}$ can be extended for any $r \in \mathbb{R}$ (see, e.g. [3]).

The Sobolev space $\mathbf{W}^r_p$ is defined in the same way as $\mathbf{B}^r_{p,\theta}$ by replacing

$\|f\|_{\mathbf{B}^r_{p,\theta}}$ and $|f|_{\mathbf{B}^{r,e}_{p,\theta}}$ by $\|f\|_{\mathbf{W}^r_p}$ and $|f|_{W^{r,e}_p} := \|(\prod_{j\in e}\partial^r/\partial x_j^r)f\|_p$, respectively, where $\partial^r/\partial x_j^r$ is the fractional partial differential operator of order r in the sense of Weil.

3. We give the main result of our paper. Let

$$\mathbf{SB}^r_{p,\theta} := \{f \in \mathbf{W}^r_p : \|f\|_{\mathbf{W}^r_p} \le 1\}$$

and

$$\mathbf{SW}^r_p := \{f \in \mathbf{B}^r_{p,\theta} : \|f\|_{\mathbf{B}^r_{p,\theta}} \le 1\}$$

be the unit balls in $\mathbf{B}^r_{p,\theta}$ and $\mathbf{W}^r_p$, respectively.

We use the notation $F \asymp F'$ if $F \ll F'$ and $F' \ll F$, and $F \ll F'$ if $F \le CF'$ with C an absolute constant. Denote by γ_n either ε_n or ρ_n and put $a_+ := \max\{a, 0\}$.

Theorem 1. *Let $1 < p,\ q < \infty$, $0 < \theta \le \infty$. Then we have for either $r > 1/p$ or $r > (1/p - 1/q)_+$ and $\theta \ge \min\{q, 2\}$*

$$\gamma_n(\mathbf{SB}^r_{p,\theta}, L_q) \asymp n^{-r}(\log n)^{(d-1)(r+1/2-1/\theta)}, \tag{3}$$

and for $r > (1/p - 1/q)_+$

$$\gamma_n(\mathbf{SW}^r_p, L_q) \asymp \left(n/\log^{d-1} n\right)^{-r}. \tag{4}$$

In addition, we can explicitly construct a subset M in L_q of cardinality $|M| \le 2^n$ and a mapping $S : M \to M$ so that

$$E(W, M, L_q) \le \sup_{f\in W} \|f - S(f)\|_q \ll E(n),$$

where W denotes $\mathbf{SB}^r_{p,\theta}$ of $\mathbf{SW}^r_p$ and $E(n)$ the right side of (3) or (4), respectively.

Theorem 2. *Let $0 < p, q, \theta \le \infty$, $1 \le \tau \le \infty$ and $r > \alpha$. Assume that either $r - \alpha > 1/p$ or $r - \alpha > (1/p - 1/q)_+$ and $\theta \ge \tau$. Then we have*

$$\varepsilon_n(\mathbf{SB}^r_{p,\theta}.\mathbf{B}^\alpha_{q,\tau}) \ll E_{\theta,\tau}(n),$$

where

$$E_{\theta,\tau}(n) = n^{-r+\alpha}(\log n)^{(d-1)(r-\alpha+1/\tau-1/\theta}.$$

In addition, we can explicitly construct a finite subset $\mathbf{V}^*$ in $\mathbf{V}$, a subset M in $\mathbf{M}_n(\mathbf{V}^*)$ of cardinality $|M| \le 2^n$, and a mapping $S : \mathbf{B}^r_{p,\theta} \longrightarrow M$ so that

$$E(\mathbf{SB}^r_{p,\theta}, M, \mathbf{B}^\alpha_{q,\tau}) \le \sup_{f\in \mathbf{SB}^r_{p,\theta}} \|f - S(f)\|_{\mathbf{B}^\alpha_{q,\tau}} \ll E_{\theta,\tau}(n).$$

Theorem 3. *Let $0 < p, q, \theta, \tau \le \infty$ and $r > \alpha$. Then we have*

$$\rho_n(\mathbf{SB}^r_{p,\theta}, \mathbf{B}^\alpha_{q,\tau}) \gg n^{-r+\alpha}(\log n)^{(d-1)(r-\alpha+1/\tau-1/\theta)}.$$

The asymptotic order of $\varepsilon_n(\mathbf{SW}^r_p, L_q)$ was proved by Smolyak [9] for $p = q = 2$, by Dinh Dung for $1 < p = q < \infty$ [2] and by Temlyakov for $1 < p \ne q < \infty$ [10] and $r > 1$, and Belinsky [1] for $1 < p \le q \le \infty$ and $1/p - 1/q < r \le 1$. The asymptotic order of $\varepsilon_n(\mathbf{SB}^r_{p,\infty}, L_q)$ was proved by Temlyakov [10] for $1 < p, q < \infty$ and $r > 1$, and Belinsky [1] for $1 < p \le q \le \infty$ and $1/p - 1/q < r \le 1$. We are restricted to consider the case $1 < p, q < \infty$ of $\varepsilon_n(\mathbf{SW}^r_p, L_q)$ and $\varepsilon_n(\mathbf{SB}^r_{p,\theta}, L_q)$. See [1, 11] for details of recent results on the cases $p, q = 1, \infty$ and $\theta = \infty$. The asymptotic order of ρ_n of the unit ball of the multivariate classical Sobolev space was obtained by Ratsaby and Maiorov [8].

4. Theorem 1 is easily proved from Theorems 2 and 3 using the inequality (2), the well-known Littlewood–Paley theorem and the following

Lemma 1. *Let the linear space L be equipped with two quasi-norms $\|\cdot\|_X$ and $\|\cdot\|_Y$, and W be a subset of L. If $\varepsilon_m(W, X) > 0$, we have*

$$\varepsilon_{n+m}(W, Y) \le \varepsilon_n(SX, Y)\varepsilon_m(W, X).$$

In order to prove Theorems 2 and 3, we used non-linear n-term approximations with regard to the family formed from the integer translates of the mixed dyadic scales of the tensor product multivariate de la Vallée Poussin kernel, discretization methods, a lemma of Haussler on upper bound for the ε-packing number (see [8, Lemma 1]) and the following estimates of ε_n for sets in finite-dimensional spaces.

For $0 < p \le \infty$ denote by l^m_p the space of all sequences $x = \{x_k\}^m_{k=1}$ of (complex) numbers, equipped with the quasi-norm $\|x\|_{l^m_p} := \left(\sum_{k=1}^m |x_k|^p\right)^{1/p}$ with the change to the max norm when $p = \infty$.

Let $0 < p, \theta \le \infty$, $\mathbf{N} = \{N_k\}_{k\in Q}$ be a sequence of natural numbers with Q a finite set of indices. Denote by $\mathbf{b}^{\mathbf{N}}_{p,\theta}$ the space of all such sequences $x = \{x^k\}_{k\in Q} = \{\{x^k_j\}^{N_k}_{j=1}\}_{k\in Q}$, for which the mixed quasi-norm

$$\|x\|_{\mathbf{b}^{\mathbf{N}}_{p,\theta}} := \left(\sum_{k\in Q} \|x^k\|^\theta_{X^k}\right)^{1/\theta}, \ \theta < \infty,$$

is finite (the sum is changed to supremum for $\theta = \infty$), where $X^k := l^{N_k}_p$. Let $\mathbf{S}^{\mathbf{N}}_{p,\theta}$ be the unit ball in $\mathbf{b}^{\mathbf{N}}_{p,\theta}$.

Lemma 2. *Let $0 < p \le 1$. Then for any positive integer n we can explicitly construct a subset $M \subset l^m_\infty$ of cardinality $|M| \le 2^n$, and a mapping $S : l^m_p \to M$ so that*

$$\varepsilon_n(B^m_p, l^m_\infty) \le \sup_{x\in B^m_p} \|x - S(x)\|_{l^m_\infty} \le C(p)A_p(m, n),$$

where

$$A_p(m,n) = \begin{cases} m^{-1/p}2^{-n/m}, & \text{for } n \ge m \\ n^{-1/p}\log^{1/p}(m/n), & \text{for } n < m. \end{cases}$$

Lemma 3. *Let $0 < p, \theta, \tau \le \infty$ and $p \le \theta$. Then for any positive integer $n < m = \sum_{k\in Q} N_k$, we can explicitly construct a subset $M \subset \mathbf{b}^{\mathbf{N}}_{\infty,\tau}$ of cardinality $|M| \le 2^n\binom{m}{n}$, and a mapping $S : \mathbf{b}^{\mathbf{N}}_{p,\theta} \to M$ so that*

$$\varepsilon_n(\mathbf{S}^{\mathbf{N}}_{p,\theta}, \mathbf{b}^{\mathbf{N}}_{\infty,\tau}) \le \sup_{x\in \mathbf{S}^{\mathbf{N}}_{p,\theta}} \|x - S(x)\|_{\mathbf{b}^{\mathbf{N}}_{\infty,\tau}} \le Cn^{-1/p}|Q|^{1/\tau+1/p-1/\theta}.$$

Remark. Lemma 2 was proved by Maiorov for the case $p = 1$ (see, e.g., [10]) by a method which is not suitable for the case $p < 1$.

References

1. E. S. Belinsky, Estimates of entropy numbers and Gaussian measure for classes of functions with bounded mixed derivative, *J. Approx. Theory* **93** (1998) 114-127.
2. Dinh Dung, *Approximation of Smooth Functions of Several Variables by Means of Harmonic Analysis*, Doctoral Dissertation, Moscow State Univ., Moscow, 1985. (Russian).
3. Dinh Dung, Continuous algorithms in n-term approximation and non-linear widths, *J. Approx. Theory* **102** (2000) 217-242.
4. D. Haussler, Decision theoretic generalization of the PAC model for neural net and other learning applications, *Inform. Comput.* **100** (1) (1982) 78-150.
5. A. N. Kolmogorov and V. M. Tikhomirov, ε-entropy and ε-capacity of sets in function space, *Uspekhi Mat. Nauk* **14** (1959) 3-86; English transl. in *Amer. Math. Soc. Transl.* **17** (2) (1961).
6. D. Pollard, Empirical processes: theory and applications, *NSF-CBMS Regional Conference Series in Probability and Statistics*, Vol. 2, Inst. Math., Stat. and Ann. Stat. Assoc., Providence, RI, 1989.
7. J. Ratsaby and V. Maiorov, The degree of approximation of sets in Euclidean space using sets with bounded Vapnik–Chervonekis dimension, *Discrete Applied Math.* **86** (1998) 81-93.
8. J. Ratsaby and V. Maiorov, On the degree of approximation by manifolds of finite pseudo-dimension, *Constr. Approx.* **15**(1999) 291-300.
9. S. A. Smolyak, The ε-entropy of the classes $E_s^{\alpha k}$ and $W_s^{\alpha}(B)$ in the L_2-metric, *Dokl. Akad. Nauk SSSR* **131** (1960) 30-33; English transl. in *Soviet Math. Dokl.* **1** (1960).
10. V. N. Temlyakov, Estimates of the asymptotic characteristics of classes of functions with bounded mixed derivative or difference, *Trudy Mat. Inst. Steklov* **189** (1989) 138-167; English transl. in *Proc. Steklov Inst. Math.* **12** (1992).
11. V. N. Temlyakov, An inequality for trigonometric polynomials and its application for estimating the entropy numbers *J. Complexity* **11** (1995) 293-307.
12. V. N. Vapnik and A. Ya. Chervonekis, Necessary and sufficient conditions for the uniform convergence of means to their expectations, *Theory Probab. Appl.* **26** (1981) 264-280.

Corrigendum

The author would like to make a correction to the paper: Dinh Dung, On asymptotic orders of n-term approximations and nonlinear n-widths, *Vietnam J. Math.* **27** (4) (1999) 363-367. The assumption $r > 1/p$ in the main results (Theorem) should be changed to the assumption: either $r > 1/p$ or $r > (1/p - 1/q)_+$ and $\theta \geq \min\{q, 2\}$. Therefore, the power of log in the asymptotic order $n^{-r}(\log n)^{(d-1)(r+1/2-1/\theta)}$ of $\sigma_n(\mathbf{SB}^r_{p,\theta}, \mathbf{V}, L_q)$ and $\gamma_n(\mathbf{SB}^r_{p,\theta}, L_q)$ should be positive and the corresponding comments should be corrected. He apologizes to the reader for this error.