Advances in Computational Mathematics 1(1993)197–214

Wavelet compression and nonlinear n-widths*

R.A. DeVore

Department of Mathematics, University of South Carolina, Columbia, SC 29208, USA

G. Kyriazis

Department of Mathematics, Purdue University, West Lafayette, IN 47907, USA

D. Leviatan

Department of Mathematics, Raymond & Beverley Sackler Faculty of Exact Sciences, Tel Aviv University, 69928 Tel Aviv, Israel

and

V.M. Tikhomirov☆

Department of Mathematics, Moscow State University, Moscow, Russia

Received 4 September 1992; revised 8 January 1993

Abstract

It is shown that certain algorithms of compression based on wavelet decompositions are optimal in the sense of nonlinear n-widths.

1. Introduction

The use of wavelet decompositions for image and data compression has recently received much attention. The question arises as to whether wavelets have any significant advantage over other methods of compression such as fractals or those based on Fourier transforms (such as the Discrete Fourier Transform). The purpose of the present paper is to show that methods of wavelet compression based on quantization are optimal in a certain mathematical sense.

Wavelet compression based on quantization strategies can be viewed as a method of nonlinear approximation (see [4]). In order to treat compression as a mathematical problem, we view the image or data to be compressed as a function

*This research was supported by the Office of Naval Research Contract N0014-91-J1343 and a Binational Science Foundation Grant 89-00505.

☆This research was completed while D. Leviatan and V.M. Tikhomirov were visiting scholars at the University of South Carolina.

f defined on a cube (which we take to be the unit cube $\Omega := [0, 1]^d$) in $\mathbb{R}^d$. We shall assume that the error of compression is to be measured in an L_p norm with $1 \le p \le \infty$. With this understanding, we show that wavelet compression performs asymptotically as well as any nonlinear method of approximation for compressing classes of functions described by smoothness conditions. To measure the smoothness of functions, we shall use the Besov space norms since they are robust enough to allow us to measure all orders of smoothness in all of the spaces L_s, $0 < s \le \infty$. Similar results to those stated here can also be obtained for Sobolev spaces or potential spaces (in fact, these follow from our results by using the embeddings of these spaces into corresponding Besov spaces). We note that much justification for the use of classifying data and images in terms of Besov spaces has been given [4]. In particular, these spaces are rich enough to include discontinuous images. Moreover, it is well known that these spaces occur naturally in the characterization of approximation order by most methods of nonlinear approximation.

In order to compare various methods of compression, we must decide which methods of nonlinear approximation will be allowed in our theory. For this, we shall introduce several definitions of nonlinear widths which have been studied over the last several decades. We shall show that these various definitions are asymptotically equivalent. Therefore, in each of these frameworks, wavelet compression is optimal in the sense alluded to above. We note that each of these definitions requires a certain type of continuity in the nonlinear approximation. In numerical terms, this can be viewed as stability of the method of compression.

The various notions of widths are among the most important methods in approximation theory for measuring the size of a compact set $K \subseteq X$, where X is a normed linear space. Perhaps the best known width is the Kolmogorov n-width

$$d_n(K, X) := \inf_{X_n} \sup_{x \in K} d(x, X_n),$$

where the infimum is taken over all n-dimensional subspaces X_n of X. Here, and later, we use the notation $d(x, A)$ to denote the distance from the point x to the set A in the norm under consideration (at present the norm on X). The Kolmogorov width and other standard widths are based on approximation from linear subspaces. In the present paper, however, we are mainly interested in nonlinear approximation, since this occurs more naturally in problems of compression.

There are several definitions of nonlinear widths which have been introduced by different authors. We review these in the following section and show that all of these definitions are asymptotically equivalent. The remainder of the paper is then devoted to proving the following theorem.

THEOREM 1.1

Let $L_s := L_s(\Omega)$, $0 < s \le \infty$, where $\Omega := [0, 1]^d$ is the unit cube in $\mathbb{R}^d$, and let $U(B_q^\alpha(L_s))$, $\alpha > 0$, $0 < q \le \infty$, denote the unit ball of the Besov space $B_q^\alpha(L_s)$. Assume that $\alpha > d(1/s - 1/p)_+$ so that $U(B_q^\alpha(L_s))$ is embedded compactly into L_p. Then, if w_n is any of the nonlinear widths given in section 2, we have

$$w_n(U(B_q^\alpha(L_s)), L_p) \asymp n^{-\alpha/d}. \tag{1.1}$$

The lower bound in (1.1) is proved by DeVore et al. [3] and so we will concentrate here on proving the upper bound. We shall prove the upper estimate using wavelets, thus showing how wavelet compression can provide an asymptotically optimal nonlinear method of approximation.

Theorem 1.1 says that if we wish to compress all functions in the class $U(B_q^\alpha(L_s))$, then wavelet-based compression, with an appropriate method of quantization, will perform as well as any method in the sense of (1.1).

During the preparation of the present paper, a manuscript by Dinh Dung and Vu Quoc Thanh [9] has been brought to our attention which gives the upper estimate in (1.1) for $1 \le s \le \infty$ for functions defined on the d-dimensional torus. The proof uses the Fourier expansion of $f \in L_s$ and the de la Vallée Poussin means, and therefore is not directly applicable to $0 < s < 1$. As is well known, Besov spaces with $0 < s < 1$ play an important role in characterizing the function classes with a given rate of nonlinear approximation. For example, for any p, if α is large enough, then a function in L_p is approximable by the best nonlinear methods at the rate $n^{-\alpha/d}$ whenever $f \in B_q^\alpha(L_s)$ where $1/s = \alpha/d + 1/p$, thus $0 < s < 1$.

2. Nonlinear widths

In this section, we are going to introduce various nonlinear widths and derive comparisons between them. Most of these comparisons are not new, but we bring their proofs here for the sake of completeness. We start with the Alexandroff n-width, which may not be the simplest to define but historically was first. For its definition, we need the notion of a complex in X.

Given a normed linear space X, by a complex $\mathscr{C}$ in X we mean the following. There is a set of points $\{x_1, \dots, x_N\} \subset X$ and a collection $\mathscr{J}$ of subsets of $\mathscr{N} := \{1, \dots, N\}$ with the properties that $J \in \mathscr{J}$ and $J' \subset J$ imply $J' \in \mathscr{J}$, and that $J \in \mathscr{J}$ implies $\{x_j\}_{j \in J}$ are in general position. We denote by $\sigma_J := \operatorname{conv}\{x_j : j \in J\}$ the simplex associated to $J \in \mathscr{J}$. We require that for each $J_1, J_2 \in \mathscr{J}$, there is an $I \in \mathscr{J}$ such that $\sigma_{J_1} \cap \sigma_{J_2} = \sigma_I$. Then $\mathscr{C} := \mathscr{C}(x_1, \dots, x_N; \mathscr{J}) := \cup_{J \in \mathscr{J}} \sigma_J$. We also define $\dim \mathscr{C} := \max\{|J| - 1 : J \in \mathscr{J}\}$.

We shall often use the following important remark:

Remark 2.1

A famous lemma of Pontrjagin–Nöbling (see [1, p. 128]) states that for any complex $\mathscr{C}$ in X, of dimension n, there is a complex $\mathscr{C}(x_1, \dots, x_N; \mathscr{J})$ of dimension n, in $\mathbb{R}^{2n+1}$ and a homeomorphism g that maps $\mathscr{C}$ onto $\mathscr{C}(x_1, \dots, x_N, \mathscr{J})$.

Finally, we need the *topological dimension* of a compact set, sometimes called the *Lebesgue dimension.* Let $K \subset X$ be compact. Then $\dim K = n$ if n is the smallest number such that for any $\varepsilon > 0$ there exists a covering $\mathcal{A} := \{A_i\}_{i \in \mathcal{N}}$, $\mathcal{N} := \{1, \ldots, N\}$, of K by open sets such that

$$\operatorname{diam} A_i < \varepsilon, \quad i \in \mathcal{N},$$

and

$$\bigcap_{j=1}^{n+2} A_{i_j} = \varnothing \quad \text{for all } \{i_1, \ldots, i_{n+2}\} \subset \mathcal{N}.$$

Such a covering is said to be of *order* $n+1$.

It should be noted that our definition of the dimension of a complex coincides with the Lebesgue dimension.

Alexandroff introduced the following n-width of a compact subset K of X:

$$a_n(K, X) := \inf_{\dim \mathcal{C} = n} \inf_{F \in C(K, \mathcal{C})} \sup_{x \in K} \|x - F(x)\|,$$

where the first infimum is taken over all continuous functions from K into $\mathcal{C}$ and the second over all complexes $\mathcal{C} \subset X$ with $\dim \mathcal{C} = n$. We have also the dual notion of n-co-width

$$a^n(K \mid X) := \inf_{\dim \mathcal{C} = n} \inf_{G \in C(K, \mathcal{C})} \sup_{x \in K} \operatorname{diam}(G^{-1}(G(x))), \tag{2.2}$$

where the second infimum is taken over all n-dimensional complexes $\mathcal{C}$ (not necessarily subsets of X). Here as usual, $G^{-1}(A) := \{x : G(x) \in A\}$.

It is easy to see that the co-width a^n is equivalent to the Urysohn width

$$u_n(K, X) := \inf_{\mathcal{A}} \max \operatorname{diam}(A_i),$$

where the infimum is taken over all coverings $\mathcal{A} := \{A_i\}$ of K of order $n+1$.

We first prove that the above two concepts are equivalent in the following sense:

LEMMA 2.1

For any normed linear space X and any compact set $K \subset X$, we have

$$a_n(K, X) \le a^n(K \mid X) \le 2a_n(K, X), \quad n = 0, 1, \ldots . \tag{2.3}$$

Proof

To prove the right inequality, let $\varepsilon > 0$ be given and let F be a continuous mapping of K into a complex $\mathcal{C} \subset X$ of dimension $\le n$ such that

$$\sup_{x \in K} \|x - F(x)\| < a_n + \varepsilon.$$

We take $G := F$. Then, for each $x_1, x_2 \in G^{-1}(G(x))$, we have $F(x_1) = F(x_2) = G(x)$ and so

$$\| x_1 - x_2 \| \le \| x_1 - F(x_1) \| + \| x_2 - F(x_2) \| \le 2a_n + 2\varepsilon.$$

Therefore,

$$\sup_{x \in K} \operatorname{diam}(G^{-1}(G(x))) \le 2a_n + 2\varepsilon .$$

Since $\varepsilon > 0$ is arbitrary, this proves the right inequality in (2.3).

To prove the remaining inequality, we can assume that the linear space X has dimension $> n$, for otherwise there is a simplex X which contains K and it follows that $a_n = 0$. Now, let $\varepsilon > 0$ be given, and let $G : K \to \mathscr{C}$ be such that $\dim \mathscr{C} = n$ and

$$\sup_{x \in K} \operatorname{diam}(G^{-1}(G(x))) \le a_n + \varepsilon .$$

Since the Lebesgue dimension of $G(K)$ is at most n, for every $\eta > 0$, there is a covering $\mathscr{A} := \{A_i\}_{i=1}^{N}$ of $G(K)$ of order n with $\operatorname{diam} A_i < \eta$. We can assume that $G^{-1}(A_i) \neq \varnothing$ for each $i = 1, \ldots, N$, since otherwise we can discard A_i in the covering of $G(K)$. For each i, we choose an $x_i \in G^{-1}(A_i)$. By the continuity of G and the compactness of K, we have

$$\operatorname{diam}(G^{-1}(A_i)) < a^n + 2\varepsilon , \quad i = 1, \ldots , N, \tag{2.4}$$

provided η is chosen sufficiently small. We would like to form the complex $\mathscr{C}(x_1, \ldots, x_N : \mathscr{J})$ where $\mathscr{J}$ is the collection of all sets J for which $\cap_{j \in J} A_j \neq \varnothing$. However, for the definition of a complex we need to make sure that $\{x_j\}_{j \in J}$ are in general position whenever $J \in \mathscr{J}$. We can accomplish this by perturbing the x_i slightly to obtain points $\bar{x}_i$ with $\| x_i - \bar{x}_i \| < \varepsilon$ and the $\{\bar{x}_j\}_{j \in J}$ are in general position, $J \in \mathscr{J}$. Note that the points $\bar{x}_j$ are no longer necessarily in $G^{-1}(A_j)$. We construct the "nerve" $\mathscr{N}_{\mathscr{A}} := \mathscr{C}(\bar{x}_1, \ldots, \bar{x}_N; \mathscr{J})$.

By a theorem of Alexandroff, there is a continuous mapping between $\mathscr{C}$ and the nerve $\mathscr{N}_{\mathscr{A}}$. In fact, we can constructively exhibit such a mapping $g : \mathscr{C} \to \mathscr{N}_{\mathscr{A}}$ by

$$g(y) := \sum_{i=1}^{N} \frac{d(y, \mathscr{C} \setminus A_i)}{\sum_{j=1}^{N} d(y, \mathscr{C} \setminus A_j)} \, \bar{x}_i , \quad y \in \mathscr{C} .$$

Obviously, the mapping $F := g \circ G : K \to \mathscr{N}_{\mathscr{A}}$ is continuous. If $x \in K$ and $y = G(x)$, then $d(y, \mathscr{C} \setminus A_i) \neq 0$ implies that $x \in G^{-1}(A_i)$, so that by virtue of (2.4), $\| x - x_i \| < a^n + 2\varepsilon$ and $\| x - \bar{x}_i \| < a^n + 3\varepsilon$. Since $F(x)$ is a convex combination of these $\bar{x}_i$'s, we have

$$\| x - F(x) \| < a^n + 3\varepsilon , \quad x \in K.$$

Thus we have for arbitrary ε

$$a_n < a^n + 3\varepsilon,$$

and the left inequality in (2.3) follows. □

The continuous n-co-width $c^n(K|X)$ is defined as follows:

$$c^n(K \mid X) := \inf_{F \in C(K,\mathbb{R}^n)} \sup_{x \in K} \operatorname{diam}(F^{-1}(F(x))). \tag{2.5}$$

Since the image $F(K)$ is a compact set, it is contained in a simplex in $\mathbb{R}^n$ and therefore we have

$$a^n(K \mid X) \le c^n(K \mid X). \tag{2.6}$$

Also, the existence of a homeomorphic realization of any n-dimensional complex in $\mathbb{R}^{2n+1}$ (see remark 2.1) immediately implies

$$c^{2n+1}(K \mid X) \le a^n(K \mid X). \tag{2.7}$$

Finally, we define the manifold n-width $\delta_n(K,X)$ as

$$\delta_n(K,X) := \inf_{M,a} \sup_{x \in K} \| x - M(a(x)) \|, \tag{2.8}$$

where the infimum is taken over all continuous mappings $a : K \to \mathbb{R}^n$ and $M : \mathbb{R}^n \to X$.

The following lemma, which was given in [9], shows that δ_n and a_n are asymptotically equivalent.

LEMMA 2.2

For any normed linear space X and any compact subset $K \subset X$, we have

$$\delta_{2n+1}(K,X) \le a_n(K,X) \le \delta_n(K,X), \quad n = 0, 1, \ldots . \tag{2.9}$$

Proof

For completeness of the present paper, we prove (2.9). We first give a proof of the upper inequality in (2.9), which is quite similar to the proof of the lower inequality in lemma 2.1. We can assume that X has linear space dimension $> n$. Let $\varepsilon > 0$. Then there are continuous mappings $M : \mathbb{R}^n \to X$ and $a : K \to \mathbb{R}^n$ such that

$$\| x - M(a(x)) \| \le \delta_n + \varepsilon, \quad x \in K. \tag{2.10}$$

The compact set $a(K) \subset \mathbb{R}^n$ has Lebesgue dimension $\le n$ (see e.g. [1, p. 163]). Therefore, for each $\eta > 0$, there is a finite covering $\mathcal{A} := \{A_j\}_{j=1}^N$ of $a(K)$ by open sets A_j, each of diameter $< \eta$ and any point $\alpha \in a(K)$ appears in at most $n+1$ of these sets. The union of the sets A_j is a bounded subset of $\mathbb{R}^n$ and therefore, by the continuity of M, for η sufficiently small,

$$|M(\alpha) - M(\beta)| < \varepsilon, \quad \alpha, \beta \in A_j, \quad j = 1, \ldots, N.$$

We let $G_j := a^{-1}(A_j)$, $j = 1, \ldots, N$. These sets are open relative to K. We can assume that each G_j is nonempty since otherwise A_j could be discarded in the covering for $a(K)$. For each $j = 1, \ldots, N$, we choose a point $\alpha_j \in A_j$ such that $\alpha_j = a(x)$ for some $x \in G_j$ and we let $x_j := M(\alpha_j) \in X$. Further, we let $\bar{x}_j$ satisfy $\| x_j - \bar{x}_j \| < \varepsilon$ and $\mathscr{C} := \{\bar{x}_1, \ldots, \bar{x}_N; \mathscr{J}\}$, where $\mathscr{J}$ is the collection of all sets J for which $\cap_{j \in J} A_j \neq \varnothing$. By choosing the perturbed points correctly, we guarantee that $\mathscr{C}$ is an n-dimensional complex in X.

To approximate the elements of K, we define the mapping

$$F(x) := \sum_{i=1}^{N} \frac{d(x, K \backslash G_i)\bar{x}_i}{\sum_{j=1}^{N} d(x, K \backslash G_j)}, \quad x \in K.$$

Then F is a continuous mapping of K into $\mathscr{C}$. If $d(x, K \backslash G_i) \neq 0$, that is, $x \in G_i$, then $a(x)$ and α_i are in A_i. Hence, $\|M(a(x)) - x_i\| < \varepsilon$ by our construction. Therefore,

$$\| x - x_i \| \leq \| x - M(a(x)) \| + \| M(a(x)) - x_i \| < \delta_n + 2\varepsilon.$$

It follows that $\| x - \bar{x}_i \| < \delta_n + 3\varepsilon$. Since $F(x)$ is a convex linear combination of these $\bar{x}_i$, we obtain $\|x - F(x)\| < \delta_n + 3\varepsilon$. This proves the upper estimate in (2.9) because $\varepsilon > 0$ was arbitrary.

To prove the lower estimate in (2.9), let $\varepsilon > 0$ be given. Then there is a complex $\mathscr{C} \subset X$ of dimension n and a continuous mapping F from K into $\mathscr{C}$ satisfying $\|x - F(x)\| \leq a_n + \varepsilon$. Let G be a homeomorphism from $\mathscr{C}$ into an n-dimensional complex of $\mathbb{R}^{2n+1}$ (see remark 2.1). We let $a(x) := G(F(x))$, $x \in K$. Then, a is a continuous mapping of K into $\mathbb{R}^{2n+1}$. Now G^{-1} is a continuous function defined on the compact set $G(\mathscr{C})$ which takes values in X. We can extend G^{-1} to a continuous mapping M defined on all of $\mathbb{R}^{2n+1}$. For example, the classical Tietze–Urysohn extension theorem will provide such an extension. Since $M(a(x)) = G^{-1}(G(F(x)) = F(x)$, we have $\|x - M(a(x))\| \leq a_n + \varepsilon$. This proves the lower estimate in (2.9) because ε is arbitrary. □

In summary, we have shown that any two of the nonlinear n-widths introduced in this paper are asymptotically equivalent.

3. Besov spaces

If $f \in L_s(\Omega)$, $0 < s \leq \infty$, we let $\omega_r(f, t)_s$, $t > 0$, denote the modulus of smoothness of order r of f, i.e.

$$\omega_r(f,t)_s := \sup_{|h|\le t} \|\Delta_h^r(f,\cdot)\|_{L_s(\Omega(rh))}, \tag{3.1}$$

where $|h|$ is the Euclidean length of the vector h; Δ_h^r is the rth order difference with step h; and the norm in (3.1) is the L_s (quasi) norm on $\Omega(rh) := \{x : [x, x+rh] \subset \Omega\}$. Since for $0 < s < 1$ we do not have a norm, it is useful to note that for any $\mu \le \min(1, s)$ and any sequence $(f_n) \subset L_s$ we have

$$\left\|\sum f_n\right\|_{L_s} \le \left(\sum \|f_n\|_{L_s}^{\mu}\right)^{1/\mu}. \tag{3.2}$$

If α, s, $q > 0$, and r is an integer larger than α, we say that f belongs to the Besov space $B_q^\alpha(L_s)$ provided

$$|f|_{B_q^\alpha(L_s)} := \left(\int_0^\infty (t^{-\alpha}\omega_r(f,t)_s)^q \frac{\mathrm{d}t}{t}\right)^{1/q} \tag{3.3}$$

is finite. When $q = \infty$, the integral in (3.3) is replaced by sup. The (quasi) norm for $B_q^\alpha(L_s)$ is then defined by

$$\|f\|_{B_q^\alpha(L_s)} := \|f\|_{L_s} + |f|_{B_q^\alpha(L_s)}. \tag{3.4}$$

Different values of r result in norms (3.4) which are equivalent.

Besov spaces give a way of measuring smoothness analogous to the better known Sobolev spaces. The space $B_q^\alpha(L_s)$ consists of functions with smoothness of order α in L_s. The index $q > 0$ is secondary and gives subtle differences in smoothness.

4. Wavelet decomposition

One of the main points of the present paper is to show that wavelet compression provides an optimal method of nonlinear approximation. In this section, we shall introduce wavelet decompositions and discuss how they characterize the Besov spaces. We begin by discussing the B-spline wavelets whose properties were developed in [6].

We fix $\alpha > 0$ and $0 < s, q \le \infty$, and let $\Omega := [0, 1]^d$. We shall consider the Besov spaces $B_q^\alpha(L_s)$ with $\alpha > d(1/s - 1)_+$. This guarantees that $B_q^\alpha(L_s)$ is compactly embedded in L_1. Here and later, all spaces are defined on the cube Ω.

Let ψ be the tensor product B-spline $\psi(x) := N(x_1) \ldots N(x_d)$, $x = (x_1, \ldots, x_d) \in \mathbb{R}^d$, with $N(t) := N(t; 0, \ldots, r)$ the univariate B-spline of order r with knots at the points $0, \ldots, r$. We fix the integer r so that $r > \alpha$. From ψ, we can form the translated dilates $\psi_{j,k} := \psi(2^k \cdot j)$, $j \in \mathbb{Z}^d$, $k \in \mathbb{Z}$. It is well known and proved in [6] that each $f \in L_s$ has a wavelet decomposition

$$f = P(f) + \sum_{k=1}^{\infty} \sum_{j \in \Lambda_k} a_{j,k}(f)\psi_{j,k} \tag{4.1}$$

in the sense of convergence in L_s. Here,

$$P(f) = \sum_{j \in [0, r-1]^d} a_j(f) x^j$$

is a polynomial of coordinate degree $< r$ and Λ_k is the set of those indices $j \in \mathbb{Z}^d$ such that $\psi(2^k \cdot - j)$ does not vanish identically on Ω. It follows that $|\Lambda_k| \le C 2^{kd}$.

One of the important facts about wavelet decompositions is that their coefficients characterize the Besov spaces. We shall discuss this only under our restriction $\alpha/d > (1/s - 1)_+$, although similar results hold for all α, q, s. The Besov space $B_q^\alpha(L_s)$ is characterized by the condition

$$|f|^*_{B_q^\alpha(L_s)} := \left(\sum_{k=1}^{\infty} 2^{k\alpha q} \left(\sum_{j \in \Lambda_k} |a_{j,k}(f)|^s \, 2^{-kd} \right)^{q/s} \right)^{1/q} < \infty. \tag{4.2}$$

The expression $|\cdot|^*_{B_q^\alpha(L_s)}$ is a quasi-seminorm and $\|\cdot\|_{L_s} + |\cdot|^*_{B_q^\alpha(L_s)}$ is equivalent to the usual norm on $B_q^\alpha(L_s)$ given in (3.4).

There are in fact many decompositions (4.1) (see [6]) for which the characterization (4.2) holds. That is, there are many ways to choose the coefficient functionals a_j and $a_{j,k}$. It will be important for us to see that in some cases these coefficients can be chosen to depend continuously on $f \in B_q^\alpha(L_s)$. For this, we shall concentrate on one particular decomposition (4.1) based on quasi-interpolants, which we shall now describe.

Let $\mathcal{S}^k$ denote the linear span of the B-splines $\psi_{j,k} := \psi(2^k \cdot - j)$, $j \in \Lambda_k$. The quasi-interpolant operator Q_k (see e.g. [6] for its definition) is a bounded linear projector from $L_1(\Omega)$ into $\mathcal{S}^k$ which can be represented by

$$Q_k(f) = \sum_{j \in \Lambda_k} q_{j,k}(f)\psi_{j,k}, \tag{4.3}$$

with the coefficient, linear functionals given by

$$q_{j,k}(f) = \int_{I_{j,k}} f \, \mu_{j,k} \, \mathrm{d}x. \tag{4.4}$$

Here, $I_{j,k}$ is a cube of side length 2^{-k} which is associated to the points $j2^{-k}$. To be specific, we can take $I_{j,k}$ to be the cube of side length 2^{-k} and smallest vertex $j2^{-k}$ provided that this cube is contained in Ω. Otherwise, we take $I_{j,k}$ as any cube of this type which is contained in $\Omega \cap \operatorname{supp} \psi_{j,k}$: The functions $\mu_{j,k}$ are in L_∞ with $\|\mu_{j,k}\|_{L_\infty} \le C 2^{kd}$, $j \in \Lambda_k$.

The quasi-interpolant operator Q_k can generally not be applied to a function in L_s when $s < 1$. This is the reason that a more elaborate operator (which is defined on L_s) is used in [6]. However, for treating only the Besov spaces $B_q^\alpha(L_s)$ with $\alpha > d(1/s - 1)_+$, we shall see that Q_k in and of itself is sufficient. For this we develop some properties of Q_k.

We first want to establish the approximation properties of Q_k. Since these properties are well known when $s \geq 1$, we shall assume that $s < 1$ and estimate how well $Q_k(f)$ approximates $f \in B_s^\beta(L_s)$ with $\beta := d(1/s - 1)$. We recall that this Besov space is embedded in L_1 (see [6]) and there is a polynomial P_J of degree $< r$ (for example, any polynomial of best $L_s(J)$ approximation will do) such that

$$\int_J |f - P_J| \, \mathrm{d}x \leq C |f|_{B_s^\beta(L_s(J))} \tag{4.5}$$

for each cube $J \subset \Omega$ with the Besov semi-norm taken with respect to J.

THEOREM 4.1

For each $f \in B_s^\beta(L_s)$, $\beta := d(1/s - 1)$, $0 < s < 1$, we have

$$\| f - Q_k(f) \|_{L_s} \leq C 2^{-k\beta} |f|_{B_s^\beta(L_s)}, \quad k = 1, 2, \ldots, \tag{4.6}$$

with C depending only on r, d, s.

Proof

We fix a cube $I = I_{i,k} \subseteq \Omega$ and estimate $\|f - Q_k(f)\|_{L_s(I)}$. Let Λ_I denote the indices $j \in \Lambda_k$ for which $\psi_{j,k}$ does not vanish identically on I. Further, let P be a polynomial of best $L_s(\tilde{I})$ approximation to f, with $\tilde{I}$ the smallest cube which contains $\cup_{j \in \Lambda_I}(\Omega \cap \operatorname{supp} \psi_{j,k})$.

From (4.4), we have for each $j \in \Lambda_I$,

$$|q_{j,k}(f - P)| < C 2^{kd} \int_{I_{j,k}} |f - P| \, \mathrm{d}x. \tag{4.7}$$

It follows by (3.2) that

$$\begin{aligned} \| Q_k(f - P) \|_{L_s(I)}^s &\leq C 2^{kds} \sum_{j \in \Lambda_I} \left(\int_{I_{j,k}} |f - P| \, \mathrm{d}x \right)^s \int_\Omega |\psi_{j,k}|^s \, \mathrm{d}x \\ &\leq C 2^{kds} 2^{-kd} \left(\int_{\tilde{I}} |f - P| \, \mathrm{d}x \right)^s. \end{aligned} \tag{4.8}$$

Now, $Q_k(P) = P$, because Q_k is a projector. Also, by using Hölder's inequality, we find that $\|f - P\|_{L_s(I)}^s$ also does not exceed the right-hand side of (4.6). Therefore, using (4.5), we have

$$\|f - Q_k(f)\|_{L_s(I)}^s \le C\{\|f - P\|_{L_s(I)}^s + \|Q_k(f - P)\|_{L_s(I)}^s\}$$

$$\le C2^{kds}2^{-kd}\left(\int_{\tilde{I}} |f - P|\,\mathrm{d}x\right)^s \le C2^{kds-kd}|f|_{B_s^\beta(L_s(\tilde{I}))}^s. \tag{4.9}$$

Now, $d/s - d = \beta$, and a point $x \in \Omega$ falls in at most C of the cubes $\tilde{I}$, $I = I_{i,k}$, $i \in \Lambda_k$, with C depending only on d. Therefore, we can add the estimates (4.9) over all the disjoint I and use the set subadditivity of $|f|_{B_s^\beta(L_s(J))}^s$ (see [6]) to obtain (4.6). □

More generally, for any $\gamma \ge d(1/s - 1)$ we have

$$\|f - Q_k(f)\|_{L_s} \le C2^{-k\gamma}|f|_{B_s^\gamma(L_s)}. \tag{4.10}$$

Indeed, it is well known (see e.g. [11]) that there is an $S \in \mathcal{S}^k$ which satisfies

$$|f - S|_{B_s^\beta(L_s)} \le C2^{k(\beta-\gamma)}|f|_{B_s^\gamma(L_s)}. \tag{4.11}$$

Since $Q_k(S) = S$,

$$\|f - Q_k(f)\|_{L_s} = \|f - S - Q_k(f - S)\|_{L_s},$$

so that (4.10) follows from (4.11) and (4.6).

We fix γ with $\alpha < \gamma < r$ and let $K(f, t) := K(f, t;\, B_s^\beta(L_s), B_s^\gamma(L_s))$ denote the K-functional for the pair of Besov spaces $(B_s^\beta(L_s), B_s^\gamma(L_s))$. If $f \in B_s^\beta(L_s)$ and $g \in B_s^\gamma(L_s)$ are properly chosen, then

$$\|f - Q_k(f)\|_{L_s} \le C\{\|f - g - Q_k(f - g)\|_{L_s} + \|g - Q_k(g)\|_{L_s}\}$$

$$\le C2^{-k\beta}\left\{|f - g|_{B_s^\beta(L_s)} + 2^{-k(\gamma-\beta)}|g|_{B_s^\gamma(L_s)}\right\}$$

$$\le C2^{-k\beta}K(f, 2^{-k(\gamma-\beta)}). \tag{4.12}$$

If we apply the real method of interpolation to the pair $B_s^\beta(L_s), B_s^\gamma(L_s))$ for $\theta := (\alpha - \beta)/(\gamma - \beta)$ and $0 < q \le \infty$, we obtain the Besov space $B_q^\alpha(L_s)$ (see, for example, corollary 6.2 of [6]). It therefore follows that $f \in B_q^\alpha(L_s)$ is equivalent to $f \in L_s$ and

$$\left(\sum_{k=1}^{\infty}[2^{k(\alpha-\beta)}K(f,2^{-k(\gamma-\beta)})]^q\right)^{1/q} < \infty, \tag{4.13}$$

and the left-hand side of (4.13) is an equivalent semi-norm for $B_q^\alpha(L_s)$.

From inequality (4.12), we conclude that $f\in B_q^\alpha(L_s)$ implies

$$\left(\sum_{k=1}^{\infty}[2^{k\alpha}\|f-Q_k(f)\|_{L_s}]^q\right)^{1/q} \le C\left(\sum_{k=1}^{\infty}[2^{k(\alpha-\beta)}K(f,2^{-k(\gamma-\beta)})]^q\right)^{1/q}. \tag{4.14}$$

We can now explain how to obtain a wavelet decomposition (4.1) from the Q_k. For each $f\in B_q^\alpha(L_s)$, we can write $f=Q_0(f)+\sum_{k=1}^{\infty}(Q_k(f))$ in the sense of L_s convergence. We then rewrite

$$Q_k(f)-Q_{k-1}(f)=\sum_{j\in\Lambda_k}a_{j,k}(f)\,\psi_{j,k} \tag{4.15}$$

to obtain the wavelet decomposition (4.1). The coefficient functionals depend continuously on $f\in L_1$ and hence on $f\in B_q^\alpha(L_s)$.

It is also easy to see that (4.2) holds for this decomposition. For this, we need the stability of the B-spline basis $\psi_{j,k}$: there are constants C_1, $C_2>0$ depending only on s and r such that

$$C_1\left(\sum_{j\in\Lambda_k}|c_j|^s\,2^{-kd}\right)^{1/s} \le \Big\|\sum_{j\in\Lambda_k}c_j\psi_{j,k}\Big\|_{L_s} \le C_2\left(\sum_{j\in\Lambda_k}|c_j|^s\,2^{-kd}\right)^{1/s}. \tag{4.16}$$

This shows that

$$\|Q_k(f)-Q_{k-1}(f)\|_{L_s}\approx\left(\sum_{j\in\Lambda_k}|a_{j,k}(f)|^s\,2^{-kd})\right)^{1/s}, \tag{4.17}$$

with constants of equivalency depending only on s and r. Using the triangle inequality, we can replace $\|f-Q_k(f)\|_{L_s}$ in (4.14) by $\|Q_k(f)-Q_{k-1}(f)\|_{L_s}$ and then replace this by the right-hand side of (4.17). This shows that the right-hand side of (4.2) does not exceed the Besov space semi-norm. We can prove the reverse of the inequality in (4.2) by using a Bernstein-type inequality for the elements in $\mathcal{S}^k$. Since the details are exactly the same as those already given in section 4 of [6], we do not discuss them here.

There are more general wavelet decompositions which also provide the characterization (4.2) with the coefficients depending continuously on $f\in B_q^\alpha(L_s)$. We shall mention in some detail only the orthogonal wavelets with compact support constructed by Daubechies [2].

In Daubechies' construction, one begins with a univariate function ϕ of compact support which has orthogonal shifts, and from it one obtains a function ψ whose normalized shifted dilates $2^{k/2}\psi(2^k\cdot -j)$ are an orthonormal basis for $L_2(\mathbb{R})$. If we define $\psi_0 := \phi$ and $\psi_1 := \psi$, then the family Ψ of functions $\psi_v(x) := \psi_{v_1}(x_1)\ldots\psi_{v_d}(x_d)$, $x\in\mathbb{R}^d$, with $v=(v_1,\ldots,v_d)$ a nonzero vertex of the cube Ω, will form by shifts and dilations an orthogonal basis for $L_2(\mathbb{R}^d)$.

The functions in $B_q^\alpha(L_s(\Omega))$ can be extended (in many different ways) to all of $\mathbb{R}^d$ (see, for example, [8]) so as to be in $B_q^\alpha(L_s(\mathbb{R}^d))$. If we take the wavelet decomposition of this extended function in terms of the functions $\psi(2^k\cdot -j)$, $j\in\mathbb{Z}^d$, $k\in\mathbb{Z}$, $\psi\in\Psi$, then only a finite number of terms at each dyadic level k contribute to the description of f on Ω. This gives a wavelet decomposition

$$f=\sum_{k\in\mathbb{Z}}\sum_{\psi\in\Psi}\sum_{j\in\Lambda_k(\psi)}a_{j,k,\psi}(f)2^{kd/2}\psi(2^k\cdot -j), \tag{4.18}$$

with $\Lambda_k(\psi)$ the indices $j\in\mathbb{Z}^d$ for which $\psi(2^k\cdot -j)$ does not vanish identically on Ω.

In an analogous fashion to (4.2) (see [10]), the coefficients in this decomposition determine whether f is in a Besov space $B_q^\alpha(L_s(\Omega))$ provided $\alpha/d>(1/s-1)_+$. It is also easy to see that the coefficients can be taken to depend continuously on $f\in L_1$. Indeed, these coefficients are a multiple of the inner product of $\psi_{j,k}$ with the extension of f. Therefore, these coefficients will depend continuously on $f\in B_q^\alpha(L_s)$ whenever the extension operator is chosen to be continuous on $B_q^\alpha(L_s)$.

5. Wavelet compression

In this section, we shall use wavelet compression to estimate the nonlinear n-width of the unit ball $U(B_q^\alpha(L_s))$ in L_p. In order that this ball be compactly embedded in L_p, we assume that $\alpha>(d/s-d/p)_+$. We continue to use the notation w_n to denote any of the nonlinear widths δ_n, a_n, a^n, u_n, and c^n. We have mentioned in the introduction that

$$w_n(U(B_q^\alpha(L_s)),L_p)\geq Cn^{-\alpha/d}$$

follows from [3].

In this section, we shall reverse this estimate and thereby determine $w_n(U(B_q^\alpha(L_s)),L_p)$ asymptotically. Since we have shown that all of these definitions of nonlinear widths are asymptotically equivalent, we need only prove that for each $n=0,1,\ldots$, there is a positive integer $m:=m(n)$ with $m(n)\leq Cn$ and

$$a^m(U(B_q^\alpha(L_s))\,|\,L_p)\leq Cn^{-\alpha/d}, \tag{5.1}$$

with C independent of n. We shall accomplish this upper estimate with wavelet compression. In this way, we show that wavelet compression provides an asymptotically optimal nonlinear approximation method. Since

$$U(B_q^\alpha(L_s)) \subset U(B_\infty^\alpha(L_s)) =: U(B) =: U,$$

we need only consider from here on the unit ball U of the space $B := B_\infty^\alpha(L_s)$. Moreover, we can assume that $n = 2^{dN}$ for some integer N since the general case follows easily from this and the monotonicity of a^m.

We shall take the sum in (4.2) as the defining semi-norm for B. It follows that functions $f \in U$ satisfy

$$\sum_{j\in\Lambda_k} |a_{j,k}(f)|^s \, 2^{-kd} 2^{k\alpha s} \le 1, \quad k = 1, 2, \ldots . \tag{5.2}$$

We shall restrict our discussion to the spline wavelet ψ given in the previous section. However, we remark at the outset that the following discussion applies with small modification to any wavelet or wavelet set for which (4.2) holds and for which the wavelet coefficients depend continuously on $f \in B_q^\alpha(L_s)$.

Recall that $\psi_{j,k} := \psi(2^k \cdot -j)$. We shall approximate $f \in U$ by spline functions

$$S = S_n(f) = P(f) + \sum \alpha_{j,k}^*(f)\, \psi_{j,k} \tag{5.3}$$

with the sum finite and consisting of at most Cn terms. A study of such approximation was made in previous work, most notably in [6], [5], and [7]. Our analysis given here is on the one hand simpler than this previous work because U is compactly embedded in L_p, and on the other hand somewhat different than previous work because we shall need to ensure that the coefficients $\alpha_{j,k}^*(f)$ are selected to depend continuously on $f \in U$. For this, we shall use an idea of Stessin [13].

To begin with, we shall reduce our problem to a finite dimensional one by using the following lemma. We recall that for a fixed dyadic level k, the B-splines $\psi_{j,k}$, $j \in \mathbb{Z}^d$, are L_p stable for each $0 < p \le \infty$.

LEMMA 5.1

For each $n \ge 1$, there is an integer $K := K(n) > 0$ such that

$$\| f - P(f) - \sum_{k=1}^{K} \sum_{j\in\Lambda_k} a_{j,k}(f)\, \psi_{j,k} \|_{L_p} \le n^{-\alpha/d}. \tag{5.4}$$

Proof

We need to estimate the L_p norm of $\sum_{k>K} T_k$ with $T_k := \sum_{j\in\Lambda_k} a_{j,k}(f)\, \psi_{j,k}$. According to the stability of the B-spline basis (see (4.16)),

$$\|T_k\|_{L_p} \le C\left(\sum_{j\in\Lambda_k} |a_{j,k}(f)|^p \, 2^{-kd} \right)^{1/p} \le C\left(\sum_{j\in\Lambda_k} |a_{j,k}(f)|^s \, 2^{-kd} \right)^{1/s} 2^{kd(1/s-1/p)_+}$$

$$\le C 2^{-k(\alpha-(d/s-d/p)_+)}.$$

Here, we have used Hölder's inequality when $s > p$ and the norm comparison $\|\cdot\|_{l_p} \le \|\cdot\|_{l_s}$ when $s \le p$. Also recall that $|\Lambda_k| \le C2^{kd}$. Since $\alpha > (d/s - d/p)_+$, we have that the left-hand side of (5.4) does not exceed

$$C \sum_{k>K} \|T_k\|_p \le C \sum_{k>K} 2^{-k(\alpha-(d/s-d/p)_+)} \le C2^{-K(\alpha-(d/s-d/p)_+)} \le n^{-\alpha/d}$$

provided K is large enough. □

We shall next show how to select coefficients from each dyadic level $k = 1, \ldots, K$ in such a way that the resulting sum is a good approximation to f. We first dispose of the case $s \ge p$ which is trivial and can be handled by linear methods. Namely, in this case it is enough to choose all coefficients at each dyadic level $k \le N$ and no coefficients at higher levels. This results in the approximant $S := S_n(f) := Q_N(f)$ of (4.3). It is well known (see [6]) that

$$\|f - S_n(f)\|_{L_p} = \|f - Q_N(f)\|_{L_p} \le \|f - Q_N(f)\|_{L_s} \le C2^{-N\alpha} = Cn^{-\alpha/d}. \quad (5.5)$$

There are at most $C2^{Nd} \le Cn$ nonzero coefficients in the representation (5.3) of $S_n(f)$. This is our desired estimate in the case $s \ge p$. We now consider the case $s < p$.

The selection of coefficients in the case $s < p$ is a nonlinear process similar to the quantization used in compression. Let $\delta := (\alpha - (d/s - d/p))/2(1/s - 1/p)$. For each $k \in 1, \ldots, K$, we let

$$\varepsilon_k := \begin{cases} 0, & 1 \le k \le N, \\ n^{-1/s}\, 2^{kd/s}\, 2^{-k\alpha}\, 2^{(k-N)\delta/s}, & k > N. \end{cases} \quad (5.6)$$

With this, we let Λ_k^* denote the set of indices $j \in \Lambda_k$ such that

$$|a_{j,k}(f)| \ge \varepsilon_k.$$

We note that

$$|\Lambda_k^*| \le \begin{cases} C2^{kd}, & k \le N, \\ 2^{-(k-N)\delta} n, & k > N. \end{cases} \quad (5.7)$$

Indeed, for $k \le N$, this follows from $|\Lambda_k^*| = |\Lambda_k| \le C2^{kd}$, while for $k > N$, we have

$$2^{(k-N)\delta} n^{-1} |\Lambda_k^*| \le \sum_{j \in \Lambda_k^*} |a_{j,k}|^s\, 2^{-kd}\, 2^{k\alpha s} \le 1$$

because of (5.2).

A typical method of compression would be to choose $a_{j,k}^*(f) = a_{j,k}(f)$ for $j \in \Lambda_k^*$ and $a_{j,k}^*(f) = 0$ otherwise. However, this selection does not depend continuously on f. Therefore, we make the following slight modification of this idea. Let $\lambda_k(x) := x$, $x \in \mathbb{R}$, when $k \leq N$, and let λ_k be the continuous piecewise linear *odd* function which takes the value 0 for $0 \leq x \leq \varepsilon_k$, the value x for $x \geq 2\varepsilon_k$ and is linear on $(\varepsilon_k, 2\varepsilon_k)$. We define

$$a^*_{j,k}(f) := \lambda_k(a_{j,k}(f)), \quad j \in \Lambda_k, \quad k = 1, \ldots, K. \tag{5.8}$$

Since both λ_k and $a_{j,k}$ are continuous, the functionals $a^*_{j,k}(f)$ depend continuously on $f \in U$. We define

$$S_n(f) := P(f) + \sum_{k=1}^{K} \sum_{j \in \Lambda_k} a^*_{j,k}(f)\,\psi_{j,k} \tag{5.9}$$

and prove the following.

THEOREM 5.2

There is a constant C with the following properties: (i) for each $n = 1, 2, \ldots$, the number of nonzero coefficients appearing in the sum (5.9) which defines $S_n(f)$ does not exceed $m(n) \leq Cn$, (ii) for each $f \in U$, we have

$$\|f - S_n(f)\|_p \leq Cn^{-\alpha/d}, \quad n = 1, 2, \ldots . \tag{5.10}$$

Proof

(i) We have $a^*_{j,k}(f) \neq 0$ only if $j \in \Lambda^*_k$. Hence, (i) follows by summing the estimates for $|\Lambda^*_k|$ given in (5.7).

(ii) We write

$$f - S_n(f) = \left[f - P(f) - \sum_{k=1}^{K} \sum_{j \in \Lambda_k} a_{j,k}(f)\,\psi_{j,k} \right] + \left[\sum_{k=1}^{K} T^*_k \right] = \Sigma_1 + \Sigma_2, \tag{5.11}$$

with

$$T^*_k := \sum_{j \in \Lambda_k} (a_{j,k}(f) - a^*_{j,k}(f))\,\psi_{j,k}. \tag{5.12}$$

According to lemma 5.1, $\|\Sigma_1\|_p \leq n^{-\alpha/d}$ and it is therefore sufficient to show that $\|\Sigma_2\|_p \leq Cn^{-\alpha/d}$. For this, we estimate $\|T^*_k\|_p$. By definition $T^*_k := 0$, $k = 1, \ldots, N$. Also, since $a^*_{j,k}(f) = a_{j,k}(f)$ whenever $|a_{j,k}(f)| \geq 2\varepsilon_k$ and $|\lambda_k(x) - x| \leq |x|$, we have

$$|a_{j,k}(f) - a^*_{j,k}(f)|^p \leq |a_{j,k}(f)|^p \leq (2\varepsilon_k)^{p-s} |a_{j,k}(f)|^s.$$

Therefore, from (4.16),

$$\begin{aligned}\|T_k^*\|_p &\le \left(\sum_{j\in\Lambda_k} |a_{j,k}(f) - a_{j,k}^*(f)|^p\, 2^{-kd} \right)^{1/p} \\ &\le C\varepsilon_k^{1-s/p} \left(\sum_{j\in\Lambda_k} |a_{j,k}(f)|^s\, 2^{-kd} \right)^{1/p} \\ &\le C\varepsilon_k^{1-s/p}\, 2^{-k\alpha s/p} \\ &\le Cn^{-1/s+1/p}\, 2^{k(d/s-d/p-\alpha)}\, 2^{(k-N)(\delta/s-\delta/p)}. \end{aligned} \tag{5.13}$$

Since $\alpha - d/s + d/p > \delta/s - \delta/p$, we have

$$\|\Sigma_2\|_p \le \sum_{k=N+1}^{K} \|T_k^*\|_p \le Cn^{-1/s+1/p}\, 2^{N(d/s-d/p)}\, 2^{-N\alpha} \le Cn^{-\alpha/d} \tag{5.14}$$

as desired. □

Proof of theorem 1.1

We have already noted that the lower estimate in (1.1) is valid. We also mentioned that to prove the upper estimate in (1.1) it is enough to prove (5.1). We fix n and let $\mu := r^d + \sum_{k=1}^{K} |\Lambda_k|$. To each $f \in U$, we associate a vector $G(f) \in \mathbb{R}^\mu$ whose first r^d coordinates are the coefficients of the polynomial $P(f)$ and whose remaining coordinates are the coefficients $a_{j,k}^*(f)$ of $S_n(f)$, with $S_n(f)$ given by (5.5) in the case $s \ge p$ and given by theorem 5.2 in the case $s < p$. The coefficients $a_{j,k}^*(f)$ are to appear in $G(f)$ in lexicographic order. Each vector $G(f)$ has at most $m := m(n)$ non-zero coordinates with $m(n)$ depending only on n and $m(n) \le Cn$. For example, for $s < p$, this follows from theorem 5.2.

Next, we note that the coefficient $a_{j,k}^*(f)$ and those of the polynomial $P(f)$ do not exceed an absolute constant Γ. Indeed, this follows from the continuity of these coefficient functionals with respect to the L_1 norm on the compact subset U of L_1.

We define $x_i := \Gamma^* e_i$, where e_i are the coordinate vectors in $\mathbb{R}^\mu$ and Γ^* is a constant which will be chosen in a moment. We let $\mathcal{J}$ denote the collection of all sets $J \subset \{1, \ldots, \mu\}$ with $|J| \le m$. Then $\mathcal{C} := \mathcal{C}(x_1, \ldots, x_m; \mathcal{J})$ is a complex in $\mathbb{R}^\mu$ of dimension $m-1$. If we choose Γ^* large enough, each of the simplices $\sigma_J := \operatorname{conv}\{x_j\}_{j\in J}$ will contain the $|J|-1$ dimensional cube with vertices $\Gamma e_j, j \in J$. Hence, $G(f)$ is a point in $\mathcal{C}$. Since $\|f - S_n(f)\|_{L_p} \le Cn^{-\alpha/d}$ for each $f \in U$, it follows that if $S_n(f) = S_n(g)$ for $f, g \in U$, then

$$\|f - g\| \le 2Cn^{-\alpha/d}.$$

Therefore, $\operatorname{diam} G^{-1}(G(f)) \le 2Cn^{-\alpha/d}$. Since G is a continuous mapping, we have proved (5.1). □

References

[1] P.S. Aleksandrov, *Combinatorial Topology*, Vol. 1 (Graylock Press, Rochester, NY, 1956).

[2] I. Daubechies, Orthonormal basis of compactly supported wavelets, Commun. Pure Appl. Math. 41(1988)909–996.

[3] R.A. DeVore, R. Howard and C.A. Micchelli, Optimal nonlinear approximation, Manus. Math. 63(1989)469–478.

[4] R.A. DeVore, B. Jawerth and B.J. Lucier, Image compression through transform coding, IEEE Trans. Inf. Theory IT-38(1992)719–746.

[5] R.A. DeVore, B. Jawerth and V.A. Popov, Compression of wavelet decompositions, Amer. J. Math. 114(1992)737–785.

[6] R.A. DeVore and V.A. Popov, Interpolation in Besov spaces, Trans. Amer. Math. Soc. 305(1988) 397–414.

[7] R.A. DeVore, P. Petrushev and X.M. Yu, Wavelet approximation in the space C, in: *Progress in Approximation Theory*, ed. A.A. Gonchar and E. Saff (Springer, New York, 1992) pp. 261–283.

[8] R.A. DeVore and R. Sharpley, Besov space in $L_p(\Omega)$, $0 < p < 1$, Trans. Amer. Math. Soc., to appear.

[9] Dinh Dung and Vu Quoc Thanh, On nonlinear n-widths, Preprint.

[10] G. Kyriazis, Wavelet decompositions and spaces of functions, Ph.D. Thesis, University of South Carolina (1992).

[11] P. Oswald, On the degree of nonlinear spline approximation in Besov–Sobolev spaces, J. Approx. Theory 61(1990)131–157.

[12] E.M. Stein, *Singular Integrals and Differentiability Properties of Functions* (Princeton University Press, Princeton, NJ, 1970).

[13] M.I. Stessin, Alexandroff widths of finite dimensional set and of classes of smooth functions, Dokl. Acad. Nauk USSR 220(1975)1278–1281.