Industrial Mathematics Institute

2001:26

Best basis selection for approximation in L_p

R. DeVore, G. Petrova and V. Temlyakov

IMI

Preprint Series

Department of Mathematics
University of South Carolina

Best basis selection for approximation in L_p*

Ronald DeVore, Guergana Petrova, and Vladimir Temlyakov

January 16, 2002

Abstract

We study the approximation of a function class $\mathcal{F}$ in L_p by choosing first a basis B and then using n-term approximation with the elements of B. Into the competition for best bases we enter all greedy (i.e. democratic and unconditional [20]) bases for L_p. We show that if the function class $\mathcal{F}$ is well oriented with respect to a particular basis B then, in a certain sense, this basis is the best choice for this type of approximation. Our results extend the recent results of Donoho [9] from L_2 to L_p, $p \neq 2$.

AMS subject classification: 42C40, 46B70,26B35 , 42B25.

Key Words: Best basis, n-term approximation, degree of approximation, approximation classes, democratic bases, unconditional bases, greedy bases.

Dedicated with much admiration to Professor M.J.D. Powell on the occasion of his 65-th birthday and retirement.

1 Introduction

Although nonlinear approximation takes many forms, its usual setting is n-term approximation. Let X be a Banach space with norm $\|\cdot\|_X$ and let $B = (b_k)$ be a (Schauder) basis for X. Throughout this paper whenever considering a basis B for X, we shall assume that the elements b_k of the basis are normalized by $\|b_k\|_X = 1$. For each $n \geq 0$, we denote by $\Sigma_n(B)$ the set of all functions

$$S = \sum_{k \in \Lambda} a_k b_k, \tag{1.1}$$

where $\#\Lambda \leq n$. Thus the functions in $\Sigma_n(B)$ can be written as a linear combination of at most n of the basis elements b_k. In the case $n = 0$, Σ_0 is the empty set. If $f \in X$, the error of n-term approximation using the basis B is given by

$$\sigma_n(f)_X := \sigma_n(f, B)_X := \inf_{S \in \Sigma_n(B)} \|f - S\|_X, \quad n \geq 0. \tag{1.2}$$

*This work has been supported in part by the Office of Naval Research Contract N00014-91-J-1076 and the National Science Foundation grants DMS 9970326, DMS 0104112

Note in particular that $\sigma_0(f)_X = \|f\|_X$.

We sometimes wish to understand this type of approximation for a class $\mathcal{F}$ of functions from X. Given such a class, we define

$$\sigma_n(\mathcal{F}_0)_X := \sigma_n(\mathcal{F}_0, B)_X := \sup_{f\in\mathcal{F}_0} \sigma_n(f, B)_X \tag{1.3}$$

with $\mathcal{F}_0$ the unit ball of $\mathcal{F}$. The asymptotic behavior for $\sigma_n(\mathcal{F}_0, B)_{L_p}$ is known for many function classes $\mathcal{F}$ and many bases B. For example, for the trigonometric basis $B = \mathcal{T}^d$ consisting of the complex exponentials $e^{ik\cdot x}$ with $k \in \mathbb{Z}^d$, this asymptotic behavior has been found (see [14]) for all Sobolev and Besov classes and all $1 \le p \le \infty$. Also, much interest has centered around this form of approximation when the basis B consists of wavelets [9, 11, 26] or spline functions [8]. We also note that n-term approximation has found many interesting applications in image/signal processing [10, 12], statistical estimation [16], and numerical methods for PDEs [3, 4].

Given a basis B and a Banach space X, one of the central questions in n-term approximation is to characterize the set of functions which have a common rate of approximation. For $0 < q \le \infty$ and $\alpha > 0$, we let $\mathcal{A}_q^\alpha(B, X)$ denote the set of all functions $f \in X$ for which

$$\|f\|_{\mathcal{A}_q^\alpha(B,X)} := \begin{cases} (\sum_{n\ge0}(n+1)^{\alpha q-1}\sigma_n(f,B)_X^q)^{1/q}, & 0 < q < \infty, \\ \sup_{n\ge0}(n+1)^\alpha\sigma_n(f,B)_X, & q = \infty, \end{cases} \tag{1.4}$$

is finite. The set $\mathcal{A}_q^\alpha(B, X)$ is called an approximation class and (1.4) defines a quasi-norm on this class. Approximation classes have been characterized for wavelet bases [11], [25], splines [24, 8], and more general classes [17],[19],[15]. In addition to the above references, we mention the papers [5] and [19] for good expositions of the essential elements in establishing such characterizations. We mention in this introduction only one result which will serve to orient the reader.

Let $B = (b_k)$ be an orthonormal basis for L_2. Then

$$\mathcal{A}_\infty^\alpha(B, L_2) = \{f = \sum_k a_k b_k : (a_k) \in \ell_{\tau,\infty}\}, \quad 1/\tau = \alpha + 1/2, \tag{1.5}$$

where $w\ell_\tau := \ell_{\tau,\infty}$ is the Lorentz space weak ℓ_τ. In other words, the class $\mathcal{A}_\infty^\alpha$ is the set of all functions f whose representation with respect to the basis B yields coefficients $a_n = a_n(f)$ whose rearrangement $c_n(f)$ into decreasing size (in absolute value) decays like $O(n^{-\alpha-1/2})$. Similar characterizations are known when the approximation takes place in L_p, $p \ne 2$ and the basis B is greedy. These results are explained in detail in §2.

It is natural to try to allow more flexibility into the approximation process by allowing the choice of basis B to depend either on the function f or the class $\mathcal{F}$. We call this type of approximation highly nonlinear because of the extra degree of nonlinearity. We denote by $\mathcal{B}$ a collection of bases B which will enter into the competition for best basis.

The first results for best basis approximation were given by Kashin [18] who showed that for any orthonormal basis B and any $0 < \alpha \le 1$, we have

$$\sigma_n(\text{Lip }\alpha, B)_{L_2} \ge cn^{-\alpha} \tag{1.6}$$

where the constant c depends only on α. It follows from this that any of the standard wavelet or Fourier bases are best for the Lipschitz classes when the approximation is carried out in L_2 and the competition is held over all orthonormal bases. The estimate (1.6) rests on some fundamental estimates for the best basis approximation of finite dimensional hypercubes using orthonormal bases. We shall also make use of these results in our analysis.

Donoho [9] has also studied the problem of best bases for a function class $\mathcal{F}$. He calls a basis B best for $\mathcal{F}$ if

$$\sigma_n(\mathcal{F}_0, B)_X = O(n^{-\alpha}), \quad n \to \infty, \tag{1.7}$$

and no other basis B' from $\mathcal{B}$ satisfies

$$\sigma_n(\mathcal{F}_0, B')_X = O(n^{-\beta}), \quad n \to \infty, \tag{1.8}$$

for a value of $\beta > \alpha$. Donoho has shown that in some cases it is possible to determine a best basis (in the above sense) for the class $\mathcal{F}$ by intrinsic properties of how the class gets represented with respect to the basis. In Donoho's analysis (as was the case for Kashin as well) the space X is L_2 (or equivalently any Hilbert space) and the competition for a best basis takes place over all complete orthonormal systems (i.e. $\mathcal{B}$ consists of all complete orthonormal bases for L_2).

In view of (1.5), the question of how well a function class $\mathcal{F}_0$ can be approximated by elements from B is transferred to learning the structure of the coefficient body $\Theta(\mathcal{F}_0, B) := \{(a_k(f)) : \ f \in \mathcal{F}_0\}$. The main result of [9] is that if $\mathcal{F}$ is a function space and B is an orthonormal basis for X for which $\Theta(\mathcal{F}_0, B)$ is bounded, orthosymmetric, and solid then B is a best basis for $\mathcal{F}$ in the above sense. In particular, if B is an unconditional basis for $\mathcal{F}$, then B is a best basis for this function class, since the body for $\mathcal{F}$ with respect to this basis is inscribed and circumscribed by multiples of an orthosymmetric solid set.

Any extension of the Kashin and Donoho results from L_2 to L_p requires a substitute of the notion of orthonormal bases in L_2. For this purpose, we shall use the notion of greedy basis (introduced in [20]) which in the case of Banach spaces is equivalent to the notion of a democratic, unconditional basis.

We say that a basis B for X is democratic if for any two finite subsets Λ, Λ' of the same cardinality, we have

$$\|\sum_{k\in\Lambda} b_k\|_X \le C \|\sum_{k\in\Lambda'} b_k\|_X \tag{1.9}$$

with C an absolute constant.

Notice, that in a Hilbert space H each orthonormal basis is a democratic and unconditional basis. In this paper, we shall consider the search for a best basis B for $\mathcal{F}$ in the sense of Donoho. We shall enter into the competition for best basis all unconditional democratic bases. Given a function class $\mathcal{F} \subset X$, we shall say that the basis B is *aligned* to $\mathcal{F}$ if whenever $f \in \mathcal{F}$, $f = \sum_k a_k(f) b_k$, and $g \in X$ with $g = \sum_k a_k(g) b_k$ then

$$|a_k(g)| \le |a_k(f)|, \quad k = 1, 2, \dots, \longrightarrow c_0 \|g\|_{\mathcal{F}} \le \|f\|_{\mathcal{F}} \tag{1.10}$$

with $c_0 = c_0(\mathcal{F}) > 0$ an absolute constant. Our main result, given in §3, shows that whenever B is an unconditional democratic basis for L_p and $\mathcal{F}$ is a function class in L_p such that B is aligned to $\mathcal{F}$ then B is a best basis for $\mathcal{F}$ in the sense described above.

We shall also prove in §4 several variants and improvements of this result. For example, we enlarge the search for a best basis to all unconditional bases and we treat approximation in a general Banach space X. We are also able to give a finer description of the asymptotic decay of the best basis error. For example, in Theorem 4.2 we show that whenever a function class $\mathcal{F}$ is aligned to a greedy basis B for X and satisfies

$$\limsup_{m\to\infty} m^\alpha \sigma_m(\mathcal{F}_0, B)_X > 0, \tag{1.11}$$

then for any unconditional basis B' we have

$$\limsup_{m\to\infty} m^\alpha (\log_2 m)^\alpha \sigma_m(\mathcal{F}_0, B')_X > 0. \tag{1.12}$$

The purpose of this paper is to introduce techniques based on metric entropy and encoding for determining optimal bases. The results of this paper can be applied in a variety of settings. In a subsequent paper, we shall employ these results (and some modifications as well) to determine best basis selection for Besov and Triebel-Lizorkin classes when the approximation takes place in L_p.

2 Democratic and unconditional bases

In this section, we state and prove some elementary results about unconditional and democratic bases which are preparatory to our main results in the sections that follow. While we could work in more generality, we shall restrict our attention to the case $X = L_p$, $1 < p < \infty$, since these are the main examples of the theory we put forward.

The main result of this section is the characterization of the approximation classes for n-term approximation using a greedy basis. It is possible to obtain these characterizations directly from the properties of greedy bases. In [25], this was done for bases that are L_p-equivalent to the univariate Haar basis. Recently, it has been independently shown by Gribonval and Nielsen [17] and Kerkyacharian and Picard [19] that the method from [25] works for arbitrary greedy basis. To keep this paper as self contained as possible, we shall derive the characterization of approximation classes in this section. Rather than use the method of [25], we shall use the approach of proving Jackson and Bernstein inequalities for the approximation process. This is a standard vehicle for characterizing approximation classes (see for example [5]).

Let $B := (b_k)$ be an unconditional basis for L_p with $\|b_k\|_{L_p} = 1$. This means that whenever $|\alpha_k| \le |\beta_k|$, $k = 1, 2, \ldots$, then

$$\|\sum_{k=1}^\infty \alpha_k b_k\|_{L_p} \le C \|\sum_{k=1}^\infty \beta_k b_k\|_{L_p}, \tag{2.1}$$

with C an absolute constant.

As we have noted in the introduction, the basis B is said to be *democratic* if for any finite sets Λ and Λ' of the same cardinality, we have

$$\|\sum_{k\in\Lambda} b_k\|_{L_p} \le C \|\sum_{k\in\Lambda'} b_k\|_{L_p}, \tag{2.2}$$

with C an absolute constant independent of the cardinality. There is a result in functional analysis [22] that says that for any unconditional basis $B = (b_k)$, normalized so that

$\|b_k\|_{L_p} = 1$, there is a subsequence k_j, $j = 1, 2, \ldots$, such that (b_{k_j}) satisfies

$$\|\sum_{j=1}^{\infty} \alpha_{k_j} b_{k_j}\|_{L_p}^p \asymp \sum_{j=1}^{\infty} |\alpha_{k_j}|^p.$$

It follows that for any democratic and unconditional basis B for L_p, we have

$$\|\sum_{k\in\Lambda} b_k\|_{L_p} \asymp (\#\Lambda)^{1/p} \tag{2.3}$$

with the constants of equivalency depending at most on B and p.

For an unconditional, democratic basis B in L_p, then the above results combine to show that

$$C_1 \min_{k\in\Lambda} |a_k|\,(\#\Lambda)^{1/p} \le \|\sum_{k\in\Lambda} a_k b_k\|_{L_p} \le C_2 \max_{k\in\Lambda} |a_k|\,(\#\Lambda)^{1/p} \tag{2.4}$$

for any finite set Λ with $C_1, C_2 > 0$ absolute constants[1].

The concept of democratic basis was first introduced by Konyagin and Temlyakov [20] in their study of greedy algorithms. Given a basis B for X and a positive integer n, the greedy algorithm assigns to $f \in X$ the n-term greedy approximant $G_n(f) := \sum_{j\in\Lambda_n(f)} a_j(f) b_j$, where $\Lambda_n(f)$ is the set of the n indices of the largest coefficients $a_j(f)$ in absolute value. The basis B is said to be *greedy* if there is an absolute constant C such that

$$\|f - G_n(f)\|_X \le C\sigma_n(f, B)_X \tag{2.5}$$

holds for all f and n. The main result in [20] is that a basis B is greedy for X if and only if it is unconditional and democratic for X. In other words, for each such basis, the greedy strategy of picking the n largest coefficients of f is equivalent (up to absolute constants) to finding the best n-term approximation to f.

Consider next the approximation classes $\mathcal{A}_q^\alpha(B, L_p)$ which were defined in the introduction (see (1.4)). It will be convenient in what follows to use the following equivalent quasi-norm

$$\|f\|_{\mathcal{A}_q^\alpha(B,L_p)} := \begin{cases} \left(\sum_{j=-1}^{\infty} \left[2^{j\alpha}\sigma_{2^j}(f,B)_{L_p}\right]^q\right)^{1/q}, & 0 < q < \infty, \\ \sup_{j\ge -1} 2^{j\alpha}\sigma_{2^j}(f,B)_{L_p}, & q = \infty, \end{cases} \tag{2.6}$$

where by definition $\sigma_{1/2}(f,B)_{L_p} := \sigma_0(f,B)_{L_p}$.

It has been observed before [25],[17],[19] that for greedy bases, it is easy to characterize these approximation classes in terms of the basis coefficients. We shall derive these results in the next subsections for the completeness of this paper using the approach of [25, 5] which was employed for wavelet bases. Our characterization of these spaces will be in terms of the size of the basis coefficients as measured by their membership in certain Lorentz spaces.

[1] We shall use the following convention concerning constants in this paper. The letter C will be used for generic constants which may vary at each occurrence. Constants that are important in further considerations will be denoted by $C_1, C_2, \ldots$

Let us recall the definition of the Lorentz sequence spaces $\ell_{\tau,q}$, $0 < \tau < \infty$, $0 < q \le \infty$. Given a sequence (c_n), we denote by c_n^*, $n = 1, 2, \ldots$, its decreasing rearrangement. In other words, c_n^* is the n-th largest of the numbers $|c_k|$. The sequence (c_n) is said to be in $\ell_{\tau,q}$ if

$$\|(c_n)\|_{\ell_{\tau,q}} := \|(n^{1/\tau} c_n^*)\|_{\ell_q(w)} \tag{2.7}$$

is finite where $\ell_q(w)$ is the ℓ_q sequence space with the Haar measure $w(n) := 1/n$. The space $w\ell_\tau := \ell_{\tau,\infty}$ is called weak ℓ_τ and is equivalently described by

$$\#\{k : |c_k| > \epsilon\} \le M^\tau \epsilon^{-\tau}, \tag{2.8}$$

for all $\epsilon > 0$. The norm $\|(c_k)\|_{\omega\ell_\tau}$ is the smallest value of M such that (2.8) holds.

We shall use the following result concerning the interpolation of Lorentz spaces (see [1]). If $0 < \tau_1, \tau_2 < \infty$ and $0 < q_1, q_2 \le \infty$, then for any $0 < \theta < 1$ and $0 < q \le \infty$, we have

$$(\ell_{\tau_1,q_1}, \ell_{\tau_2,q_2})_{\theta,q} = \ell_{\tau,q} \tag{2.9}$$

where $\frac{1}{\tau} = \frac{1-\theta}{\tau_1} + \frac{\theta}{\tau_2}$. Here, $(X,Y)_{\theta,q}$ denotes the interpolation spaces generated by the real method of interpolation (K-functionals).

2.1 Direct theorem

Let $B = (b_k)$ be a greedy basis for L_p, i.e. B is unconditional and democratic. In this subsection, we shall prove that if $f = \sum_k a_k(f) b_k$ with $(a_k(f))_{k>0}$ in $w\ell_\tau$, $1/\tau = s + 1/p$, $s > 0$, then,

$$\sigma_n(f, B)_{L_p} \le C n^{-s} \|(a_k(f))\|_{w\ell_\tau}, \quad n = 1, 2, \ldots, \tag{2.10}$$

with the constant C depending only on p and s. To prove this, we let $\Lambda_j := \{k : 2^{-j} < |a_k(f)| \le 2^{-j+1}\}$. Then, for each $k = 1, 2, \ldots$, we have

$$\sum_{j=-\infty}^{k} \#\Lambda_j \le M^\tau 2^{k\tau}.$$

Let $S_j := \sum_{k \in \Lambda_j} a_k(f) b_k$ and $T_k := \sum_{j=-\infty}^{k} S_j$. Then $T_k \in \Sigma_n$ with $n = M^\tau 2^{k\tau}$. We have

$$\|f - T_k\|_{L_p} \le \sum_{j=k+1}^{\infty} \|S_j\|_{L_p}.$$

We fix $j > k$ and estimate $\|S_j\|_{L_p}$. Since $|a_k(f)| \le 2^{-j+1}$ for all $k \in \Lambda_j$, we have from (2.4),

$$\|S_j\|_{L_p} \le C 2^{-j} (\#\Lambda_j)^{1/p} \le C M^{\tau/p} 2^{j(\tau/p - 1)}.$$

We therefore conclude that

$$\|f - T_k\|_{L_p} \le C M^{\tau/p} \sum_{j=k+1}^{\infty} 2^{j(\tau/p-1)} \le C M (M 2^k)^{\tau/p - 1},$$

because $\tau/p - 1 = -s\tau < 0$. In other words, for $n = M^\tau 2^{k\tau}$, we have

$$\sigma_n(f, B)_{L_p} \le C M n^{1/p - 1/\tau} = C M n^{-s}.$$

From the monotonicity of σ_n it follows that the last inequality holds for all $n \ge 1$.

2.2 Inverse estimate

In this subsection, we prove the following "inverse inequality" for n-term approximation in L_p by elements from a greedy basis $B = (b_k)$. Let $s > 0$ and $1/\tau = s + 1/p$. For each $n \geq 1$, we let $\Sigma_n(B)$ be the set of all $S = \sum_{k\in\Lambda} a_k(S)b_k$ with $\#\Lambda \leq n$. Then, we have

$$\|(a_k(S))_{k\in\Lambda}\|_{w\ell_\tau} \leq Cn^s\|S\|_{L_p}, \quad S \in \Sigma_n(B), \quad n = 1, 2, \ldots. \tag{2.11}$$

For the proof of (2.11), let $\epsilon > 0$ and let $\Lambda_\epsilon := \{k \in \Lambda : |a_k(S)| > \epsilon\}$ and $S_0 := \sum_{k\in\Lambda_\epsilon} a_k(S)b_k$. From (2.1) and (2.4), we know that

$$\epsilon(\#\Lambda_\epsilon)^{1/p} \leq C\|S_0\|_{L_p} \leq C\|S\|_{L_p}. \tag{2.12}$$

Hence,

$$\epsilon(\#\Lambda_\epsilon)^{1/\tau} \leq C(\#\Lambda_\epsilon)^{1/\tau-1/p}\|S\|_{L_p} \leq Cn^s\|S\|_{L_p}, \tag{2.13}$$

because $1/\tau - 1/p = s$ and $\#\Lambda_\epsilon \leq n$. Taking a supremum over all $\epsilon > 0$, the left side is $\|(a_k(S))\|_{w\ell_\tau}$ and we obtain (2.11).

2.3 Characterization of the approximation classes

Let B be a greedy basis for L_p. For any $s > 0$, let X_s be the space of all $f \in L_p$ such that $f = \sum_k a_k(f)b_k$ with $(a_k(f)) \in w\ell_\tau$ where $1/\tau := s + 1/p$. We define the norm of f in X_s by

$$\|f\|_{X_s} := \|(a_k(f))\|_{w\ell_\tau}.$$

From §2.1- §2.2, we have that X_s satisfies the Jackson and Bernstein inequalities for n-term approximation,

$$\sigma_n(f, B)_{L_p} \leq Cn^{-s}\|f\|_{X_s}, \quad n = 1, 2, \ldots, \quad f \in X_s, \tag{2.14}$$

and

$$\|S\|_{X_s} \leq Cn^s\|S\|_{L_p}, \quad S \in \Sigma_n(B). \tag{2.15}$$

From the general theory of approximation spaces in [12], we find

$$\mathcal{A}_q^\alpha(B, L_p) = (L_p, X_s)_{\alpha/s,q} \tag{2.16}$$

with the latter the interpolation spaces for the pair (L_p, X_s). If the unconditional basis $B = (b_k)$ is democratic, then the space $Y_0 := \{f = \sum_k a_k(f)b_k : (a_k(f)) \in \ell_{p,1}\}$ is embedded in L_p and L_p is embedded in the space $Y_1 := \{f = \sum_k a_k(f)b_k : (a_k(f)) \in \ell_{p,\infty}\}$. Using these embeddings and the interpolation result for Lorentz spaces (2.9), we obtain that

$$\mathcal{A}_q^\alpha(B, L_p) = (L_p, X_s)_{\alpha/s,q} = \{f = \sum_k a_k(f)b_k : (a_k) \in \ell_{\tau_\alpha,q}\}, \tag{2.17}$$

with $1/\tau_\alpha = \alpha + 1/p$.

3 Best basis

The next sections of this paper will consider the problem of finding a best basis B to use in conjunction with n-term approximation for approximating a given function class $\mathcal{F}$. The present section will give a coarse version of our theory which uses as its main vehicle the concept of Kolmogorov entropy and encoding. Later sections will give finer results and treat variants of the best basis problem.

We fix $1 < p < \infty$ and let $\mathcal{B}$ denote the class of all democratic, unconditional bases for L_p. A *function class* $\mathcal{F}$ is a collection of functions f equipped with a norm

$$\|f\|_{\mathcal{F}}. \tag{3.1}$$

We denote by $\mathcal{F}_0$ the unit ball of $\mathcal{F}$. We shall consider function classes $\mathcal{F}$ which are embedded in L_p which means

$$\|f\|_{L_p} \leq C\|f\|_{\mathcal{F}}, \tag{3.2}$$

with C a constant depending only on p and $\mathcal{F}$. Other conditions on $\mathcal{F}$ will be imposed subsequently. The typical function classes are smoothness spaces such as the Besov or Triebel-Lizorkin spaces.

We denote by U_τ the unit ball of $w\ell_\tau$ and by aU_τ the ball of radius a centered at the origin. Given any basis $B = (b_k) \in \mathcal{B}$, we have that each $f \in L_p$ has a unique expansion in the basis B: $f = \sum_k a_k(f)b_k$. From the results of §2.3, we know that for any given $\alpha > 0$, a function $f \in L_p$ satisfies $\sigma_n(f, B)_{L_p} = O(n^{-\alpha})$ if and only if $(a_k(f)) \in w\ell_\tau$ with $1/\tau = \alpha + 1/p$.

Given a function class $\mathcal{F}$ with unit ball $\mathcal{F}_0$, we denote by $\mathcal{A}(\mathcal{F}, B) := \{(a_k(f)) : f \in \mathcal{F}_0\}$. We consider $T(\mathcal{F}, B) := \{\tau : \mathcal{A}(\mathcal{F}, B) \subset aU_\tau, \text{for some } a > 0\}$. We know from (2.10) that for each $\tau \in T(\mathcal{F}, B)$, we have

$$\sigma_n(\mathcal{F}_0, B)_{L_p} \leq Cn^{-\alpha}, \quad \alpha := 1/\tau - 1/p, \tag{3.3}$$

with C independent of n.

Now define $\tau(\mathcal{F}, B) := \inf\{\tau : \tau \in T(\mathcal{F}, B)\}$ and $\alpha(\mathcal{F}, B) := \frac{1}{\tau(\mathcal{F},B)} - \frac{1}{p}$. The fact that $\tau(\mathcal{F}, B)$ need not be attained will cause ϵ's to appear in the discussion that follows. The reader could neglect these for a simpler reading. The basis B provides n-term approximation order $O(n^{-\alpha(\mathcal{F},B)+\epsilon})$ for $\mathcal{F}_0$ whenever $\epsilon > 0$. Moreover, if $\alpha > \alpha(\mathcal{F}, B)$, then one can find functions $f \in \mathcal{F}_0$ such that $\sigma_n(f, B)_{L_p} \neq O(n^{-\alpha})$, $n \to \infty$ (we do not give the simple arguments to do this derivation).

We want to enter into a competition all bases $B \in \mathcal{B}$ to find out which basis provides the best approximation order of the form $O(n^{-\alpha})$. Later sections of this paper will consider finer descriptions of approximation order. The best approximation order in the above sense is attained by taking

$$\tau^* := \tau^*(\mathcal{F}) := \inf_{B\in\mathcal{B}} \tau(\mathcal{F}, B) \tag{3.4}$$

and defining

$$\alpha^* := \alpha^*(\mathcal{F}) := \frac{1}{\tau^*} - \frac{1}{p}. \tag{3.5}$$

We shall say that a basis $B^* \in \mathcal{B}$ is *best in coarse order* for $\mathcal{F}$ if $\tau(\mathcal{F}, B^*) = \tau^*(\mathcal{F})$.

The theorem which follows will show how to find best bases for function classes $\mathcal{F}$ which are aligned with respect to a fixed basis B (see (1.10) for the definition of aligned). We shall give several examples of such function classes and bases in the last section of this paper.

The main theorem of this section is the following.

Theorem 3.1 *If $\mathcal{F}$ is a function class in L_p and $\bar{B}$ is a democratic, unconditional basis to which $\mathcal{F}$ is aligned then $\bar{B}$ is a best coarse order basis for $\mathcal{F}$ in the sense described above, i.e. $\bar{B} = B^*$.*

The remainder of this section will be devoted to a proof of this theorem. Let $\bar{B}$ be the basis of the theorem and let $\bar{\tau} := \tau(\mathcal{F}, \bar{B})$ and $\bar{\alpha} := \frac{1}{\bar{\tau}} - \frac{1}{p}$. It is sufficient to show that for any $\beta > \bar{\alpha}$ and any basis $B \in \mathcal{B}$, we have

$$\limsup_{n \to \infty} n^\beta \sigma_n(\mathcal{F}_0, B)_{L_p} > 0. \tag{3.6}$$

Indeed, using the fact that B is greedy, the results of §2.3 combined with (3.6) then give the theorem.

To prove (3.6) we need a technique of estimating from below the quantities $\sigma_n(\mathcal{F}_0, B)_{L_p}$. Let us make some historical remarks relevant to this problem. We already mentioned in the introduction Kashin's [18] and Donoho's [9] results. Those results are based on the method of inscribing a cube in the function class $\mathcal{F}_0$ and then studying lower estimates for the inscribed cube. We will also employ this method. Another ingredient of our technique is based on the following heuristic well known in approximation theory and functional analysis: "the smallest asymptotic characteristics of compact classes are the entropy numbers". Thus, we shall estimate from below the entropy numbers of the given class $\mathcal{F}_0$.

To define the entropy numbers, we need the concept of Kolmogorov entropy which we now define. If K is a compact subset of L_p and $\epsilon > 0$, the covering number $N_\epsilon(K)$ is by definition the smallest integer N for which there is a covering of K by balls in L_p of radius $\leq \epsilon$. Then,

$$H_\epsilon(K) := \log_2 N_\epsilon(K) \tag{3.7}$$

is called the Kolmogorov ϵ- entropy of K. The Kolmogorov entropy allows one to encode the elements of K by a bitstream. Namely, we label each of the balls in an optimal covering of K in a one to one fashion by an integer between 0 and $N_\epsilon(K) - 1$ and then represent each such integer by its binary representation. Thus each ball is identified with a bitstream consisting of at most $\lceil H_\epsilon(K) \rceil$ bits. This in turn gives an encoding of the elements of K. If f is in K, we identify one of the ϵ balls which contains f and associate to f the bitstream associated to this ball. In this way, each f is associated to a bitstream containing at most $\lceil H_\epsilon(K) \rceil$ bits. We can recover f to error at most ϵ from its bitstream by taking the center of the ball with the same bitstream as that of f. Such an encoder is optimal for K, although not very practical.

It is also useful to reverse the roles of $H_\epsilon(K)$ and ϵ. Namely, given an integer N, we can ask for the infimum of all ϵ for which we can cover K by at most 2^N balls of radius ϵ:

$$\epsilon_N(K) := \epsilon_N(K, L_p) := \inf\{\epsilon : H_\epsilon(K) \leq N\}. \tag{3.8}$$

The numbers $\epsilon_N(K)$, $N = 1, 2, \ldots$, are called the *entropy numbers* for K. In other words, $\epsilon_N(K)$ is the smallest error we can obtain using an encoder/decoder pair which uses at most N bits for each $f \in K$.

Entropy numbers are also closely related to n-widths. The nonlinear Kolmogorov (N, m)-width for a centrally symmetric compact set K in Banach space X is defined by

$$d_m(K, X, N) := \inf_{\ell_N, \#\ell_N \leq N} \sup_{f \in K} \inf_{L \in \ell_N} \inf_{g \in L} \|f - g\|_X, \tag{3.9}$$

where ℓ_N is a set of at most N m-dimensional subspaces L. In the case $N = 1$, $d_m(K, X, N)$ coincides with the classical Kolmogorov width $d_m(K, X)$. Let us mention three inequalities which show the relation between n-widths and entropy numbers. Carl [2] has proven that

$$\max_{1 \leq k \leq n} k^r \epsilon_k(K, X) \leq C(r) \max_{1 \leq m \leq n} m^r d_{m-1}(K, X). \tag{3.10}$$

The following two inequalities (see [27]) are also useful in estimating $\sigma_m(\mathcal{F}_0, B)_X$ from below. For any positive constant a we have

$$\max_{1 \leq k \leq n} k^r \epsilon_k(K, X) \leq C(r, a) \max_{1 \leq m \leq n} m^r d_{m-1}(K, X, (an/m)^m). \tag{3.11}$$

For any positive a and r we have

$$\max_{1 \leq k \leq n} k^r \epsilon_{(a+r)k \log_2 k}(K, X) \leq C(r, a) \max_{1 \leq m \leq n} m^r d_{m-1}(K, X, m^{am}). \tag{3.12}$$

We note, in particular, that Theorem 4.1 of the next section can be easily derived from the last inequality.

Now let us turn to proving (3.6). Let $\bar{B} = (\bar{b}_k)$ be an unconditional democratic basis for which (1.10) holds and let $\alpha > \bar{\alpha}$ be arbitrary and let τ be defined by the relation $1/\tau = \alpha + 1/p$. From the definition of $\bar{\tau}$, there is a function $f \in \mathcal{F}_0$ such that its coefficients $a_k := a_k(f)$ with respect to $\bar{B}$, $f = \sum_k a_k \bar{b}_k$, are not in weak ℓ_τ. By rearranging the basis elements in $\bar{B}$, we can assume that the absolute values of the a_k are nonincreasing. This means that for infinitely many of n, we have

$$|a_n| \geq c_0^{-1} n^{-\frac{1}{\tau}} \tag{3.13}$$

with c_0 the constant of (1.10). In going further we shall only consider the set $\mathcal{N}$ of such n. For each $n \in \mathcal{N}$, we consider the compact set

$$K_n(\alpha) := \{g = \sum_{k=1}^{n} a'_k \bar{b}_k : \ |a'_k| \leq n^{-\frac{1}{\tau}}, \ k = 1, 2, \ldots, n\}.$$

Since $f \in \mathcal{F}_0$ and $c_0^{-1}|a'_k| \leq c_0^{-1} n^{-\frac{1}{\tau}} \leq |a_n| \leq |a_k|$, $k = 1, \ldots, n$, it follows from the fact that $\mathcal{F}$ is aligned with respect to $\bar{B}$ (see (1.10)) that

$$K_n(\alpha) \subset \mathcal{F}_0, \quad n \in \mathcal{N}. \tag{3.14}$$

The lemma that follows gives a lower bound for the entropy numbers of $K_n(\alpha)$.

Lemma 3.2 *For any $\alpha > \bar{\alpha}$, and any of the sets $K_n(\alpha)$, $n \in \mathcal{N}$, we have*

$$\epsilon_k(K_n(\alpha)) \geq cn^{-\alpha}, \quad k = \lfloor n \log_2(4/3) \rfloor, \tag{3.15}$$

with c an absolute constant.

Proof: Let E denote the set of vertices of the unit cube in $\mathbb{R}^n$. If $e = (e_1, \ldots, e_n)$ is in E, then the function

$$f_e := n^{-\frac{1}{\tau}} \sum_{k=1}^{n} e_k \bar{b}_k \tag{3.16}$$

is in $K_n(\alpha)$. We can choose (see Lemma 2.2, p. 489, [21]) a subset $E_0 \subset E$ consisting of at least $(4/3)^n$ vectors in E such that any two elements $e, e' \in E_0$ differ in at least $[n/8]$ positions and therefore with an absolute constant c_1, we have

$$\|f_e - f_{e'}\|_{L_p} \geq c_1 n^{-\frac{1}{\tau}} n^{\frac{1}{p}} = c_1 n^{-\alpha}, \tag{3.17}$$

where we have used condition (2.4) for the basis $\bar{B}$. It follows that with any collection of $(4/3)^n$ balls of radius $cn^{-\alpha}$, with $c = c_1/2$, we cannot cover $K_n(\alpha)$. This gives (3.15) and proves the lemma. $\square$

Now, let us suppose that there is a greedy basis $B \in \mathcal{B}$ which satisfies

$$\sigma_m(\mathcal{F}_0, B)_{L_p} \leq C_0 m^{-\beta}, \quad m = 1, 2, \ldots, \tag{3.18}$$

for some $\beta > \bar{\alpha}$ and an absolute constant C_0. For $\alpha := (\bar{\alpha} + \beta)/2$, we will draw a contradiction to (3.15).

We fix any of the special values of $n \in \mathcal{N}$ and we draw a contradiction to (3.15) as follows. Each of the functions $f_k := n^{-\frac{1}{\tau}} \bar{b}_k$, $k = 1, \ldots, n$, $1/\tau = \alpha + 1/p$, is in $\mathcal{F}_0$ because $f_k \in K_n(\alpha) \subset \mathcal{F}_0$. Hence, from (3.18)

$$\sigma_m(f_k, B)_{L_p} \leq C_0 m^{-\beta}, \quad m = 1, 2, \ldots. \tag{3.19}$$

We fix an integer $r > 0$ such that $r\beta > \beta + 1$. We choose the sets $\Lambda(f_k)$, $k = 1, \ldots, n$, of at most n^r indices of the largest coefficients $\mu_{j,k} := \mu_j(f_k)$ in absolute value in the representation of f_k with respect to the basis B. From (3.19) and the fact that B is a greedy basis, we have

$$\|f_k - \sum_{j \in \Lambda(f_k)} \mu_{j,k} b_j\|_{L_p} \leq Cn^{-r\beta}, \quad k = 1, 2, \ldots, n. \tag{3.20}$$

If we reorder the basis elements of B we can assume that $\Lambda(f_k) \subset \{1, \ldots, n^{r+1}\} =: \Lambda^*$, $k = 1, \ldots, n$.

Given any $f \in K_n(\alpha) \subset \mathcal{F}_0$ and any integer m, the m-term greedy approximation $G_m(f)$ to f using the greedy basis B satisfies (see (3.18))

$$\|f - G_m(f)\|_{L_p} \leq Cm^{-\beta}. \tag{3.21}$$

We want next to observe that we can restrict the indices of the basis elements used in $G_m(f)$ to come from Λ^*. For any $f \in K_n(\alpha)$, we have $f = \sum_{k=1}^n \gamma_k f_k$ with $|\gamma_k| \le 1$. Hence, from (3.20) and the fact that the basis B is unconditional,

$$\|f - P_{\Lambda^*} f\|_{L_p} \le \sum_{k=1}^{n} \|f_k - P_{\Lambda^*} f_k\|_{L_p} \le C n^{-r\beta+1} \le C n^{-\beta}, \tag{3.22}$$

where $P_{\Lambda^*} f := \sum\limits_{k\in\Lambda^*} a_k(f) b_k$. This means that

$$\begin{aligned} \|f - P_{\Lambda^*}(G_m(f))\|_{L_p} &\le \|f - P_{\Lambda^*} f\|_{L_p} + \|P_{\Lambda^*}(f - G_m(f))\|_{L_p} \\ &\le Cn^{-\beta} + Cm^{-\beta} \le Cm^{-\beta}, \quad m = 1, \ldots, n, \end{aligned} \tag{3.23}$$

where the first term was estimated by (3.22) and the second by (3.21), and we have used the fact that the basis is unconditional.

We will now describe a rather brutal encoding of the elements of $K_n(\alpha)$ which will violate the lower bound (3.15). Given $f \in K_n(\alpha)$, let $c_k := c_k(f)$, be the coefficients of f with respect to the basis B. From our assumption (3.18), we know that

$$\mathcal{A}(\mathcal{F}, B) \subset aU_\mu, \quad \mu = (\beta + 1/p)^{-1}, \tag{3.24}$$

for some $a > 0$. This means that the coefficients $c_k(f)$, $f \in \mathcal{F}_0$, are all uniformly bounded:

$$|c_k(f)| \le 2^A \tag{3.25}$$

with A dependent only on $\mathcal{F}$.

Let $k \ge 0$ be any integer with $2^k \le n$ and $n \in \mathcal{N}$. We shall fix the value of k in a moment. Given $f \in K_n(\alpha)$, we let $\Lambda_k(f) \subset \Lambda^*$ be such that $Q_k(f) := \sum_{j\in\Lambda_k(f)} c_j(f) b_j$ is the projection of the 2^k-term greedy approximation to f onto Λ^*. We know from (3.23) that

$$\|f - Q_k(f)\|_{L_p} \le C2^{-k\beta} \tag{3.26}$$

with C an absolute constant. Let $\bar{c}_j(f)$ be the quantization of $c_j(f)$ obtained by retaining the first $A + \lceil k(\beta+1)\rceil$ terms of the binary expansion of $c_j(f)$ starting with 2^A. Then

$$|c_j(f) - \bar{c}_j(f)| \le 2^{-k(\beta+1)}. \tag{3.27}$$

We now encode f as follows. For each $j \in \Lambda_k(f)$, we use $\lceil (r+1)\log_2 n\rceil$ bits to give the value of j. This will use a total of at most $2^k\lceil (r+1)\log_2 n\rceil$ bits to specify the indices of the elements in $\Lambda_k(f)$. Next, for each $j \in \Lambda_k(f)$, we use the $A + \lceil k(\beta+1)\rceil$ bits which determine $\bar{c}_j(f)$. This will use a total of at most $2^k(A + \lceil k(\beta+1)\rceil)$ bits. Hence the entire bitstream will consist of at most

$$N = 2^k(A + \lceil k(\beta+1)\rceil + \lceil (r+1)\log_2 n\rceil) \tag{3.28}$$

bits in total. We can partially recover f from this bitstream by

$$\bar{Q}_k(f) := \sum_{k\in\Lambda_k(f)} \bar{c}_k(f) b_k. \tag{3.29}$$

Because of (2.4) for the basis B, we get

$$\|f - \bar{Q}_k(f)\|_{L_p} \le \|f - Q_k(f)\|_{L_p} + \|Q_k(f) - \bar{Q}_k(f)\|_{L_p} \le C2^{-k\beta} + C2^{-k\beta - k}2^{k/p} \le C2^{-k\beta}. \tag{3.30}$$

In other words with an investment of N bits with N given in (3.28), we can encode $f \in K_n(\alpha)$ with accuracy $C2^{-k\beta}$. We now choose k as large as possible so that $N \le \lfloor n \log_2(4/3) \rfloor$. For example, if γ is sufficiently large (independent of n) then $2^k \ge \frac{n}{(\log_2 n)^\gamma}$. We can therefore deduce that with an investment of at most $\lfloor n \log_2(4/3) \rfloor$ bits, we obtain an encoding error of at most

$$C(\log_2 n)^{\gamma\beta} n^{-\beta} \tag{3.31}$$

with C independent of $f \in K_n(\alpha)$. Since $\beta > \alpha$, this contradicts Lemma 3.2 and completes the proof of Theorem 3.1.

4 Finer results for best bases

In this section, we shall give refinements of the results of the previous section. The main new feature will be that we widen the search for best basis from the collection of all greedy bases to the collection of all unconditional bases. The techniques developed in this section also allow us to replace the coarse order results with a finer analysis of approximation orders in best basis selection. We shall also work in slightly more generality than the L_p spaces.

Let X be a Banach space with norm $\|\cdot\|_X$ and let $B = (b_k)$ be a Schauder basis for X normalized so that $\|b_k\|_X = 1$. Each $f \in X$ has a representation $f = \sum_k a_k(f) b_k$ where the a_k are continuous linear functionals on X which are uniformly bounded. The partial sums of this series converge to f in $\|\cdot\|_X$. In other words, if we denote by $P_n(f) := P_n(f, B) = \sum_{k=1}^n a_k(f) b_k$, then

$$\|f - P_n(f, B)\|_X \to 0, \quad n \to \infty. \tag{4.1}$$

By the uniform boundedness principle, the operators P_n are uniformly bounded:

$$\|P_n\|_{X \to X} \le C_0, \quad n = 1, 2, \dots, \tag{4.2}$$

with $C_0 > 0$ an absolute constant.

Let us denote by $E_n(f, B)_X$ the error in approximating f by the elements in the finite dimensional space X_n spanned by $b_1, \dots, b_n$:

$$E_n(f, B)_X := \inf_{g \in X_n} \|f - g\|_X. \tag{4.3}$$

For a class K of functions, we define

$$E_n(K, B)_X := \sup_{f \in K} E_n(f, B)_X. \tag{4.4}$$

It follows from (4.2) that

$$\|f - P_n(f)\|_X \le (C_0 + 1) E_n(f, B)_X. \tag{4.5}$$

We begin with the following theorem in which the function Φ is a positive, monotone decreasing function which satisfies $\Phi(x) \leq C_1\Phi(2x)$ for an absolute constant $C_1 > 0$ and also satisfies $\Phi(x) \geq c_1 x^{-\alpha}$, $x \geq 1$, for absolute constants $c_1, \alpha > 0$.

Theorem 4.1 *Suppose that K is a compact set in X and B is a basis for X such that*

$$\sigma_m(K, B)_X \leq \Phi(m), \quad m = 1, 2, \ldots \tag{4.6}$$

and

$$E_m(K, B)_X \leq C_2 m^{-\beta}, \quad m = 1, 2, \ldots, \tag{4.7}$$

for some fixed constants $C_2, \beta > 0$. Then, the entropy numbers of K satisfy

$$\epsilon_{N \log_2 N}(K) \leq C\Phi(N), \quad N = 1, 2, \ldots. \tag{4.8}$$

Proof: The proof will be similar to the results of the previous section. For each $f \in X$, let $S_m(f)$ denote a best m-term approximation to f using the basis elements B (when such best approximations do not exist the following argument is modified trivially). We shall consider the case $N = 2^n$ since the general case follows from this and the properties of Φ.

Let $\alpha > 0$ be from the properties of Φ and let $a > \alpha/\beta$ be an integer. We consider the set $\Lambda := \{1, 2, \ldots, M\}$, where $M := N^a$. For each $f \in K$, we have

$$\|f - P_M(S_N(f))\|_X \leq \|f - P_M(f)\|_X + \|P_M(f - S_N(f))\|_X \leq CM^{-\beta} + C\Phi(N) \leq C\Phi(N) \tag{4.9}$$

where we have used (4.5) and (4.7) on the first term, (4.2) and (4.6) on the second term and the definition of M and the properties of Φ in the last inequality.

The function $P_M(S_N(f))$ is a linear combination of at most N terms from B with indices in Λ. We denote this set of indices by $\Lambda(f)$ and the coefficients in this linear combination by $c_j(f)$. We have

$$|c_j(f)| \leq \|P_j - P_{j-1}\| \|S_N(f)\|_X \leq C\|f\|_X \leq C_3, \quad f \in K, \tag{4.10}$$

with C_3 an absolute constant (because K is a bounded set in X). We let A be an integer such that $C_3 \leq 2^A$. We define $\bar{c}_k(f)$ to be the approximation to $c_k(f)$ given by the first $\lceil(\alpha + 1)n\rceil + A$ terms of the binary expansion of $c_k(f)$. Then

$$|c_k(f) - \bar{c}_k(f)| \leq N^{-\alpha-1}. \tag{4.11}$$

Thus if $\bar{S}(f) := \sum_{k \in \Lambda(f)} \bar{c}_k(f) b_k$, then

$$\|P_M(S_N(f)) - \bar{S}(f)\|_X \leq N^{-\alpha} \leq C\Phi(N). \tag{4.12}$$

It follows that

$$\|f - \bar{S}(f)\|_X \leq \|f - P_M(S_N(f))\|_X + \|P_M(S_N(f)) - \bar{S}(f)\|_X \leq C\Phi(N). \tag{4.13}$$

We now show we can encode each of the $\bar{S}(f)$, $f \in K$, with at most $CN \log_2 N$ bits. We can identify each of the integers $k \in \Lambda$ with its binary expansion which has at most

$\lceil \log_2 M \rceil$ bits. We use $N\lceil \log_2 M \rceil \le CN \log_2 N$ bits to identify the N indices in $\Lambda(f)$ and then in natural order we use $A + \lceil (\alpha+1)\log_2 N \rceil$ bits to identify each of the $\bar{c}_k(f)$. This latter step uses $\le CN \log_2 N$ bits in all. Hence with at most $CN \log_2 N$ bits we can identify $\bar{S}(f)$. In view of (4.13), this shows that $\epsilon_{CN\log_2 N}(K) \le C\Phi(N)$. □

The next theorem gives a refinement of Theorem 3.1 in that it shows that a greedy aligned basis B is not only best in the sense of the coarse order α but also up to logarithmic factors. Also, it allows the larger competition over all unconditional bases. We also formulate the theorem for more general Banach spaces X than L_p.

Theorem 4.2 *Let B be a normalized unconditional basis for X with the property*

$$\| \sum_{j\in\Lambda} b_j \|_X \asymp (\#\Lambda)^{\mu}, \tag{4.14}$$

for some $\mu > 0$. Assume that the function class $\mathcal{F}$ is aligned with B and for some $\alpha > 0$, $\gamma \in I\!R$, we have

$$\limsup_{m\to\infty} m^{\alpha} (\log_2 m)^{\gamma} \sigma_m(\mathcal{F}_0, B)_X > 0. \tag{4.15}$$

Then for any unconditional basis B' we have

$$\limsup_{m\to\infty} m^{\alpha} (\log_2 m)^{\alpha+\gamma} \sigma_m(\mathcal{F}_0, B')_X > 0. \tag{4.16}$$

Proof: The proof is quite similar to that of Theorem 3.1. We shall first construct a compact set $K \subset \mathcal{F}$ for which

$$\epsilon_{2^n}(K) \ge c 2^{-n\alpha} n^{-\gamma} \tag{4.17}$$

holds for infinitely many values of n with an absolute constant $c > 0$.

For each $f \in X$, we let $f = \sum_k a_k(f) b_k$ denote its expansion in the basis B. We denote by $\lambda_n(f)$ the 2^n-th largest of the $|a_k(f)|$. Given $f \in X$, we have $f = \sum_{k=0}^{\infty} [S_k(f) - S_{k-1}(f)]$, $S_{-1}(f) := 0$, where $S_k = \sum_{j\in\Lambda_k(f)} a_j(f) b_j$ and $\Lambda_k(f)$ is the set of indices of the 2^k largest coefficients of f in absolute value (ties in the size of coefficients can be handled in an arbitrary way). Since by our assumption the basis B is unconditional, it follows from (4.14) that

$$\sigma_{2^m}(f, B)_X \le C \sum_{k>m} \|S_k(f) - S_{k-1}(f)\|_X \le C \sum_{k\ge m} \lambda_k(f) 2^{k\mu}. \tag{4.18}$$

It follows from (4.18) and our assumption (4.15) that there is a constant $c_2 > 0$ and a set $\mathcal{N}$ of infinitely many values of n such that for each $n \in \mathcal{N}$ there is a function $f_n \in \mathcal{F}_0$ such that

$$\lambda_n(f_n) \ge c_2 2^{-n(\alpha+\mu)} n^{-\gamma}. \tag{4.19}$$

In going further, we shall use the abbreviated notation $\Lambda_n := \Lambda_n(f_n)$ and $\delta_n := c_2 2^{-n(\alpha+\mu)} n^{-\gamma}$ with c_2 from (4.19). For each $n \in \mathcal{N}$, we define

$$K_n := \{f : f = \sum_{j\in\Lambda_n} a_j(f) b_j,\ |a_j(f)| \le \delta_n\}. \tag{4.20}$$

The alignment property of $\mathcal{F}$ gives that $K_n \subset a\mathcal{F}_0$ for an absolute constant $a > 0$ (independent of n). The same proof as in Lemma 3.2 shows that

$$\epsilon_k(K_n) \geq c2^{-n\alpha}n^{-\gamma}, \quad k = \lfloor 2^n \log_2(4/3) \rfloor. \tag{4.21}$$

We define $K = \cup_{n\in\mathcal{N}} K_n$, so that $K \subset a\mathcal{F}_0$. The decay of $2^{n\mu}\delta_n$ to zero guarantees that K is compact. We also have

$$\epsilon_k(K) \geq \epsilon_k(K_n) \geq c2^{-n\alpha}n^{-\gamma}, \quad k = \lfloor 2^n \log_2(4/3) \rfloor, \quad n \in \mathcal{N}. \tag{4.22}$$

Now let B' be any unconditional basis for X and assume that

$$\sigma_m(\mathcal{F}_0, B')_X = o(m^{-\alpha}(\log_2 m)^{-\alpha-\gamma}), \quad m \to \infty. \tag{4.23}$$

We will find a rearrangement B^* of B' which satisfies (4.7) for a value of $\beta > 0$ which will be specified below. From the fact that $K_n \subset a\mathcal{F}_0$, we have that for each $j \in \Lambda_n$, we have $\delta_n b_j \in a\mathcal{F}_0$, and therefore from (4.23)

$$\sigma_{2^m}(b_j, B')_X \leq C\delta_n^{-1}2^{-m\alpha}m^{-\alpha-\gamma}, \quad m = 1, 2, \ldots. \tag{4.24}$$

Let us denote by $\Lambda_m^*(b_j)$ any set of 2^m indices such that a linear combination of the basis functions $b_k' \in B'$, $k \in \Lambda_m^*(b_j)$, yields a 2^m-term approximation whose error in approximating b_j is bounded by the right side of (4.24).

We shall define the rearranged basis B^* inductively as the limit of finite ordered sets B_s^*, i.e. $B^* = \cup_{s=1}^\infty B_s^*$. The set B_1^* consists of the one element b_1'. Assuming that B_{s-1}^* has been defined for some s, we construct B_s^* by adjoining the following basis elements which are to follow in order the elements of B_{s-1}^*. The first element we add is b_s' if it does not already appear in B_{s-1}^*. Let $n \in \mathcal{N}$ with $n \leq s$. For all $j \in \Lambda_n$, we add all elements b_k' for which $k \in \Lambda_{m_s}^*(b_j)$ with $m_s := \lfloor (1 + 1/\alpha) \rfloor s$. These elements can be added in any order but they must follow the elements in B_{s-1}^* in their order. If one of these elements already is in B_{s-1}^* then of course we do not add it again. Thus, to form B_s^* from B_{s-1}^*, we have added at most

$$1 + s2^s2^{m_s} \tag{4.25}$$

new basis elements. This gives the bound

$$N_s := \#B_s^* \leq Cs2^s2^{m_s} \leq 2^{2s+m_s} \tag{4.26}$$

for the number of elements in B_s^*.

It is now easy to check (4.7). If $f \in K$, we shall bound $E_N(f, B^*)_X$ for $N = N_s$ and $s \in \{1, 2, \ldots\}$. If $f \in K_n$, $n \leq s$, then f is a linear combination of 2^n basis functions b_j with coefficients $\leq \delta_n$. In view of our construction (see (4.24)) of B_s^*, each of these basis functions satisfy

$$E_N(b_j, B^*)_X \leq \sigma_{2^{m_s}}(b_j, B')_X \leq C\delta_n^{-1}2^{-m_s\alpha}m_s^{-\alpha-\gamma}. \tag{4.27}$$

It follows that

$$E_N(f, B^*)_X \leq C2^n2^{-m_s\alpha}m_s^{-\alpha-\gamma} \leq C2^s2^{-m_s\alpha}m_s^{-\alpha-\gamma} \leq C2^{-\alpha s/2}. \tag{4.28}$$

On the other hand, when $f \in K_n$, with $n > s$, we have from (4.20) and (4.14) that

$$E_N(f, B^*)_X \le \|f\|_X \le C\delta_n 2^{n\mu} = C2^{-n\alpha} n^{-\gamma} \le C2^{-\alpha s/2}. \tag{4.29}$$

In view of (4.26), the right sides of (4.28) and (4.29) can both be bounded by $CN^{-\beta}$ with $\beta := \frac{\alpha}{6+2/\alpha}$. This establishes (4.7) for this value of β and for $N = N_s$, $s = 1, 2, \ldots$. For other values of N this follows from the monotonicity of E_N.

We have shown for the rearranged basis B^* that condition (4.7) of Theorem 4.1 is satisfied. We also have from (4.23) that

$$\sigma_m(K, B^*)_X = o(m^{-\alpha}(\log_2 m)^{-\alpha-\gamma}), \tag{4.30}$$

because $K \subset a\mathcal{F}_0$. We can apply Theorem 4.1 with $\Phi(m) = o(m^{-\alpha}(\log_2 m)^{-\alpha-\gamma})$ and obtain

$$\epsilon_N(K) = o(N^{-\alpha}(\log_2 N)^{-\gamma}), \quad N = 1, 2, \ldots. \tag{4.31}$$

This is a contradiction to (4.22). Thus, we can never have (4.23) and therefore we have (4.16). $\square$

Let us close this section by making some remarks about the special case when $X = H$ is a Hilbert space and all bases under consideration are orthonormal bases. In this case, the above results can be improved by removing the $(\log_2 m)^\alpha$ factor that appeared in (4.16). This rests on a deeper analysis of m-term approximation given by Kashin [18].

Lemma 4.3 *Let $\Phi := \{\varphi_j\}_{j=1}^N$ be an orthonormal system in H, and define*

$$K_N(\Phi) := \{f : f = \sum_{j=1}^N a_j\varphi_j, |a_j| \le 1, j = 1, \ldots, N\}.$$

Then, there exists an absolute constant $c_0 > 0$ such that for any orthonormal basis B we have

$$\sigma_m(K_N(\Phi), B)_H \ge \frac{3}{4} N^{1/2}, \quad \textit{for} \quad m \le c_0 N. \tag{4.32}$$

From this result, one can easily derive

Corollary 4.4 *Let Φ be an orthonormal basis for H and $\mathcal{F}$ be a function class aligned with Φ such that for some $\alpha > 0$, $\gamma \in I\!R$, we have*

$$\limsup_{m\to\infty} m^\alpha (\log_2 m)^\gamma \sigma_m(\mathcal{F}_0, \Phi)_H > 0. \tag{4.33}$$

Then, for any orthonormal basis B we have

$$\limsup_{m\to\infty} m^\alpha (\log_2 m)^\gamma \sigma_m(\mathcal{F}_0, B)_H > 0. \tag{4.34}$$

Proof: The proof follows the same lines as the proof of Theorem 4.2 with $\mu = 1/2$. From the assumption (4.33), one finds for infinitely many n a function $f_n \in \mathcal{F}_0$ with $c_n(f_n) \ge cn^{-\alpha-1/2}(\log_2 n)^{-\gamma}$, for some $c > 0$, independent on n. From this, one derives that the cube

$$K_n := \{f : f = \sum_{j=1}^n a_j\varphi_j, |a_j| \le cn^{-\alpha-1/2}(\log_2 n)^{-\gamma}, j = 1, \ldots, n\}$$

is contained in $a\mathcal{F}_0$. An application of Lemma 4.3 then gives (4.34). $\square$

5 n-term approximation with varying bases.

In this section, we shall consider variants of the best basis problem considered above. Up until this point, we have analyzed the problem where the basis B is chosen and then fixed for approximating the function class $\mathcal{F}$. It is not allowed to vary with m. In this section, we shall consider the more general problem in which the basis B can be chosen from a library of bases $\mathcal{B}$ and B is allowed to vary for each value of m and each f. For a given function class $\mathcal{F}$ and a given library of bases $\mathcal{B}$, we define

$$\sigma_m(\mathcal{F}_0, \mathcal{B})_X := \sup_{f \in \mathcal{F}_0} \inf_{B \in \mathcal{B}} \sigma_m(f, B)_X. \tag{5.1}$$

To obtain meaningful results, it is necessary to restrict the size of $\mathcal{B}$ since otherwise, we could have that each $f \in \mathcal{F}_0$ is in one of the bases B, in which case the right side of (5.1) is zero. Such restrictions are also necessary for any numerical algorithm. We shall show that if the library $\mathcal{B}$ is not too large and satisfies certain other minimal conditions then we can bound $\sigma_m(\mathcal{F}_0, \mathcal{B})_X$ from below by the entropy numbers of $\mathcal{F}_0$.

Let a function class $\mathcal{F} \subset X$ be given. For constants $\beta > 0$, $c_1, c_2 > 0$, we denote by $\mathcal{B}(\mathcal{F}, \beta, c_1, c_2)$ the set of all bases B for X that satisfy the following two conditions

$$E_n(\mathcal{F}_0, B)_X \leq C_1 n^{-\beta}, \quad n = 1, 2, \ldots, \tag{5.2}$$

$$\|P_n(f, B)\|_X \leq C_2 \|f\|_X, \quad f \in X, \quad n = 1, 2, \ldots, \tag{5.3}$$

where as before $P_n(\cdot, B)$ is the projector from X onto the span of the first n elements of B. Note that condition (5.2) implies that $\mathcal{F}_0$ is a compact subset of X.

Theorem 5.1 *Let $\mathcal{F}$ be a function class in X such that for some $\alpha > 0$, $\gamma \in I\!R$, we have*

$$\limsup_{m \to \infty} m^\alpha (\log_2 m)^\gamma \epsilon_m(\mathcal{F}_0) > 0. \tag{5.4}$$

Then for every fixed $a > 0$ and any subset $\mathcal{B}_{M_m}$ of $\mathcal{B}(\mathcal{F}, \beta, c_1, c_2)$ of cardinality $M_m \leq m^{am}$ we have

$$\limsup_{m \to \infty} m^\alpha (\log_2 m)^{\alpha+\gamma} \sigma_m(\mathcal{F}, \mathcal{B}_{M_m})_X > 0. \tag{5.5}$$

Proof: The proof (which we only sketch) is very similar to that of Theorem 4.2. One assumes that (5.5) does not hold and uses a generalization of Theorem 4.1 in which the single basis is replaced by the family $\mathcal{B}_{M_m}$. In the encoder used to obtain the analogue of (4.8) it is necessary to encode the basis selection from $\mathcal{B}_{M_m}$. However, this will use at most $C \log_2 M_m = Cm \log_2 m$ bits. □

Our next theorem will give a lower bound for the approximation of certain function classes from a library of bases consisting of orthonormal bases. In the proof of this theorem, we shall need certain results on finite dimensional geometry which we now develop.

Let

$$V_N := \{z = (z_1, \ldots, z_N) : \ |z_j| = 1, \ j = 1, \ldots, N\} \tag{5.6}$$

be the set of vertices in the boundary of the unit cube $[-1, 1]^N$ in $I\!R^N$. The following theorem has been proved by Kashin [18].

Theorem 5.2 *There is an absolute constant* $C_0 > 0$ *such that for every sequence* $(x_i)_{i=1}^{\infty} \subset \mathbb{R}^N$ *satisfying the conditions*

$$\sum_{i=1}^{\infty} \|x_i\|^2_{\ell_2(\mathbb{R}^N)} = 1, \tag{5.7}$$

$$\max_i \|x_i\|_{\ell_2(\mathbb{R}^N)} \le \rho, \tag{5.8}$$

the following inequality holds for all $n\rho^2 \le 1$ *and* $y \ge 0$,

$$\#\{z \in V_N : \sup_{\#\Omega = n} \sum_{i \in \Omega} |\langle z, x_i\rangle|^2 \ge y n^{1/2}\rho\} \le C_0 2^N (n\rho^2)^{-1} \exp\left(-y^2/8\right). \tag{5.9}$$

Remark: The statement in [18] of the above theorem has the condition $\max_i \|x_i\|_{\ell_2(\mathbb{R}^N)} = \rho$ in place of (5.8) but the same proof gives the above statement.

We use Theorem 5.2 to prove the following result.

Lemma 5.3 *There is an absolute constant* $C_1 > 0$ *with the following property. For any* M *sequences* $x^\ell := (x_i^\ell)_{i=1}^{\infty} \subset \mathbb{R}^N$, $\ell = 1, \dots, M$, *satisfying the conditions*

$$\sum_{i=1}^{\infty} \|x_i^\ell\|^2_{\ell_2(\mathbb{R}^N)} = 1, \quad \ell = 1, \dots, M, \tag{5.10}$$

$$\max_{1 \le \ell \le M} \max_i \|x_i^\ell\|_{\ell_2(\mathbb{R}^N)} \le \rho, \tag{5.11}$$

and for every m *satisfying*

$$m\rho^2 \le \frac{C_1}{1 + \log_2 M}, \tag{5.12}$$

we have

$$\min_{z \in V_N} \max_{1 \le \ell \le M} \sup_{\#\Omega = m} \sum_{i \in \Omega} |\langle z, x_i^\ell\rangle|^2 \le 1/4. \tag{5.13}$$

Proof: It is enough to consider the case $m = \lfloor \frac{C_1}{(1+\log_2 M)\rho^2} \rfloor$. We apply Theorem 5.2 with $y = \frac{1}{4\rho\sqrt{m}}$ and obtain for each $1 \le \ell \le M$,

$$\begin{aligned} \#V_N^\ell := \#\{z \in V_N : \sup_{\#\Omega = m} \sum_{i \in \Omega} |\langle z, x_i^\ell\rangle|^2 > 1/4\} &\le C_0 2^N (m\rho^2)^{-1} e^{-(128 m\rho^2)^{-1}} \\ &\le C_0 2^N C_2 \frac{1 + \log_2 M}{C_1} e^{-\frac{1+\log_2 M}{128 C_1}}, \end{aligned} \tag{5.14}$$

with C_2 an absolute constant. Here and later, we use the fact that xe^{-x} is decreasing for $x \ge 1$. The right side of (5.14) is less than $2^{N-1}/M$ if C_1 is sufficiently small. For this C_1, it follows that $V_N \setminus \cup_{\ell=1}^M V_N^\ell$ is nonempty and therefore (5.13) holds. $\square$

We shall now use these results to study approximation from a library of bases in a Hilbert space H. Let $\Phi_N := \{\phi_1, \dots, \phi_N\}$ be an orthonormal system in H and let, as before,

$$K_N(\Phi_N) := \{\sum_{j=1}^N a_j \phi_j : |a_j| \le 1,\ j = 1, \dots, N\} \tag{5.15}$$

be the N-dimensional cube associated to Φ_N. We have the following lemma.

Lemma 5.4 *If $B^\ell = (b_i^\ell)_{i=1}^\infty$, $\ell = 1, \dots, M$, are orthonormal bases for H, and $\mathcal{B} := \{B^\ell\}_{\ell=1}^M$ then we have*

$$\sigma_m(K_N(\Phi_N), \mathcal{B})_H \geq \frac{\sqrt{3}\sqrt{N}}{2}, \tag{5.16}$$

for $m \leq \frac{C_1 N}{1+\log_2 M}$ with C_1 the constant of Lemma 5.3.

Proof: Let $\xi_i^\ell = N^{-1/2} P_N b_i^\ell$, where P_N is the orthogonal projector onto $\operatorname{span}\{\phi_1, \dots, \phi_N\}$, $\ell = 1, \dots, M$, and $i = 1, 2, \dots$. Then, (5.10) and (5.11) of Lemma 5.3 are satisfied for $x_i^\ell = \xi_i^\ell$ with $\rho = N^{-1/2}$. Applying this lemma, we see that there is an $f_0 = \sum_{j=1}^N a_j \phi_j$ with $|a_j| = 1$, $j = 1, \dots, N$, such that for all $1 \leq \ell \leq M$, we have

$$\sup_{\#\Omega=m} \sum_{i\in\Omega} \langle f_0, \xi_i^\ell \rangle^2 \leq 1/4, \quad m \leq \frac{C_1 N}{1+\log_2 M}. \tag{5.17}$$

Since $\langle f_0, \xi_i^\ell \rangle = \langle N^{-1/2} f_0, b_i^\ell \rangle$, for all i, this implies that for each $\ell = 1, \dots, M$,

$$\sigma_m(N^{-1/2} f_0, B^\ell)_H^2 = \|N^{-1/2} f_0\|_H^2 - \sup_{\#\Omega=m} \sum_{i\in\Omega} \langle f_0, \xi_i^\ell \rangle^2 \geq 3/4, \tag{5.18}$$

where the last inequality follows from (5.17). Since $f_0 \in K_N(\Phi_N)$ we have proved the lemma. $\square$

We now give an example of how this last lemma can be utilized. For the Hilbert space H and the orthonormal system Φ, let $H^r(\Phi)$ be the class of functions $f \in H$ such that

$$f = \sum_{j=1}^\infty a_j(f)\phi_j, \quad \sum_{j=2^{s-1}}^{2^s-1} a_j(f)^2 \leq 2^{-2rs}, \quad s = 1, 2, \dots. \tag{5.19}$$

Theorem 5.5 *Let Φ be an orthonormal basis and let $\mathcal{B} := \{B^\ell\}_{\ell=1}^M$ be a collection of any M orthonormal bases in H. Then*

$$\sigma_m(H^r(\Phi), \mathcal{B})_H \geq c(r)(m(1+\log_2 M))^{-r}. \tag{5.20}$$

Proof: Given m, let $N = 2^s$ with s the smallest integer such that $m \leq \frac{C_1 N}{1+\log_2 M}$ with C_1 the constant of Lemma 5.3. We apply Lemma 5.4 and obtain coefficients a_j, $j = 2^s, \dots, 2^{s+1}-1$ with $|a_j| = 1$, for all j, such that the function $f_s := \sum_{j=2^s}^{2^{s+1}-1} a_j \phi_j$ satisfies

$$\min_{1\leq\ell\leq M} \sigma_m(f_s, B^\ell)_H \geq \frac{\sqrt{3}\sqrt{N}}{2}. \tag{5.21}$$

The function $g_s := N^{-r-1/2} f_s$ is in $H^r(\Phi)$ and satisfies

$$\min_{1\leq\ell\leq M} \sigma_m(g_s, B^\ell)_H \geq \frac{\sqrt{3}}{2} N^{-r} \tag{5.22}$$

and (5.20) follows. $\square$

As an example of this theorem consider the case that $H = L_2(\mathbb{T})$ where $\mathbb{T}$ is the unit circle and let Φ be the Fourier basis for H. Then the class $H^r(\Phi)$ is identical (after renormalization) with the unit ball of the Besov space $B^r_\infty(L_2(\mathbb{T}))$. It is well known and easy to see that the basis consisting of the first m terms of the Fourier basis provides approximation of order m^{-r} for functions in this unit ball. Hence the gain in this order of approximation by allowing m-term approximation using M bases is at most the factor $(\log_2 M)^{-r}$.

In the case of function spaces one can prove a generalization of Kashin's result ($M = 1$) for Hölder classes H^r_∞:

$$\inf_{\{B^\ell\}} \sup_{f\in H^r_\infty} \min_{1\le\ell\le M} \sigma_m(f, B^\ell)_{L_2} \ge C(r)(m(1+\log_2 M))^{-r}. \tag{5.23}$$

We note that the $\log_2 M$ factor in (5.23) can not be replaced by slower growing function of M. Indeed, for $m = 1$, $M = 2^n$ one has

$$\inf_{\{B^\ell\}} \sup_{f\in H^r_\infty} \min_{1\le\ell\le M} \sigma_1(f, B^\ell)_{L_2} \le \epsilon_n(H^r_\infty, L_2) \ll n^{-r}.$$

References

[1] J. Bergh, J. Löfström, *Interpolation Spaces*, Springer-Verlag, Berlin-New York, 1976.

[2] B. Carl, *Entropy numbers, s-numbers, and eigenvalue problems*, J. Funct. Anal., **41**(1981), 290–306.

[3] A. Cohen, W. Dahmen, R. DeVore, *Adaptive wavelet methods for elliptic operator - convergence rates equations*, Math. Comp. **70**(2001), 27–75.

[4] A. Cohen, W. Dahmen, R. DeVore, *Adaptive wavelet Methods II-beyond the elliptic case*, preprint.

[5] A. Cohen, R. DeVore, R. Hochmuth, *Restricted nonlinear approximation*, Constr. Approx., **16**(2000), 85–113.

[6] I. Daubechies, *Ten Lectures on Wavelets*, CBMS-NSF Regional Conference Series in Applied Mathematics, SIAM, Philadelphia, 1992.

[7] R. DeVore, *Nonlinear approximation*, Acta Numer., **7**(1998), 51–150.

[8] R. DeVore, V. Popov, *Interpolation of Besov spaces*, Trans, Amer. Math. Soc., **305** 1(1988), 397–414.

[9] D. Donoho, *Unconditional bases are optimal bases for data compression and for statistical estimation*, Appl. Comput. Harmon. Anal., **1**(1993), 100–115.

[10] R. DeVore, B. Jawerth, B. Lucier, *Image compression through transform coding*, IEEE Proc. Inform. Theory, **38**(1992), 719–746.

[11] R. DeVore, B. Jawerth, V. Popov, *Compression of wavelet decompositions*, Amer. J. Math, (1992), 737–785.

[12] R. DeVore, G. Lorentz, *Conctructive Approximation*, Springer–Verlag, Berlin-New York, 1993.

[13] R. DeVore, R. Sharpley, *Besov spaces on domains in* $I\!R^d$, TAMS, **335**(1993), 843–864.

[14] R. DeVore, V. Temlyakov, *Nonlinear approximation by trigonometric sums*, J. Fourier Anal. Appl., **2**(1995), 29–48.

[15] S. Dilworth, N. Kalton, D. Kutzarova, V. Temlyakov, *The Thresholding Greedy Algorithm, Greedy Bases, and Duality*, IMI Preprint No. 23, 2001.

[16] D. Donoho, I. Johnstone, *Ideal spatial adaptation via wavelet shrinkage*, Biometrika, **81**(1994), 425–455.

[17] R. Gribonval, M. Nielsen, *Some remarks on non-linear approximation with Schauder bases*, East J. Approx., **7** 3(2001), 267–285.

[18] B. Kashin, *Approximation properties of complete orthonormal systems*, (Russian) Studies in the theory of functions of several real variables and the approximation of functions. Trudy Mat. Inst. Steklov., **172**(1985), 187–191.

[19] G. Kerkyacharian, D. Picard, *Entropy, universal coding, approximation and bases properties*, Universities Paris 6 & 7 Preprint No. 663, 2001, 1–32.

[20] S. Konyagin, V. Temlyakov, *A remark on greedy approximation in Banach spaces*, East J. Approx., **5** 3(1999), 365–379.

[21] G. Lorentz, M. Golitschek, Y. Makovoz, *Constructive Approximation. Advanced Problems*, Springer–Verlag, Berlin, 1996.

[22] J. Lindenstrauss, L. Tzafriri, *Classical Banach spaces*, Lecture Notes in Mathematics, Vol. 338. Springer-Verlag, Berlin-New York, 1973.

[23] Y. Meyer, *Ondelettes et Opérateurs*, Hemann, Paris, 1990.

[24] J. Peetre, *A Theory of Interpolation Spaces*, Notes, Universidade de Brasilia, 1963.

[25] V. Temlyakov, *Non-linear m-term approximation with regard to the multivariate Haar system*, East J. Approx., **4** 1(1998), 87–106.

[26] V. Temlyakov, *The best m-term approximation and greedy algorithms*, Adv. Comput. Math., **8** 3(1998), 249–265.

[27] V. Temlyakov, *Nonlinear Kolmogorov widths*, (Russian) Mat. Zametki, **63** 6(1998), 891–902, translation in Math. Notes, **63** 5-6(1998), 785–795.

Ronald A. DeVore, Dept. of Mathematics, University of South Carolina, Columbia, SC 29208. email: devore@math.sc.edu

Guergana Petrova, Dept. of Mathematics, Texas A& M University, College Station, TX 77843. email:petrova@math.tamu.edu
Vladimir Temlyakov, Dept. of Mathematics, University of South Carolina, Columbia, SC 29208. email: temlyak@math.sc.edu