The Journal of Fourier Analysis and Applications
Volume 7, Issue 4, 2001

Greedy Algorithms and Best m-Term Approximation with Respect to Biorthogonal Systems

Peter Oswald

Communicated by Hans G. Feichtinger

ABSTRACT. *The article extends upon previous work by Temlyakov, Konyagin, and Wojtaszczyk on comparing the error of certain greedy algorithms with that of best m-term approximation with respect to a general biorthogonal system in a Banach space X. We consider both necessary and sufficient conditions which cover most of the special cases previously considered. Some new results concerning the Haar system in* L_1, L_∞, *and BMO are also included.*

1. Introduction

Throughout this article, let X be a real separable Banach space, and $\Phi = \{\phi_k,\ k \in I\}$ a minimal, normalized, dense system in X. We identify I with the set of natural numbers $\mathbb{N} \equiv \{1, 2, \dots\}$ (although all considerations apply to finite-dimensional spaces, too, we will assume $\dim X = \infty$). The normalization condition reads $\|\phi_k\|_X = 1$, $k \in \mathbb{N}$, and the density requirement says that the union of all linear subspaces $V_\Lambda = \mathrm{span}\{\phi_k,\ k \in \Lambda\}$ generated by finite index sets $\Lambda \subset I$ is dense in X. Minimality is equivalent to the existence of a biorthogonal system $\Psi = \{\psi_k,\ k \in I\} \subset X'$ such that

$$\langle \psi_l, \phi_k \rangle_{X' \times X} = \delta_{kl} = \begin{cases} 1 & , \quad k = l \\ 0 & , \quad k \neq l \end{cases} .$$

In order for the following discussion to make sense, we will assume that

$$M_\Psi := \max_{k \in I} \|\psi_k\|_{X'} < \infty . \tag{1.1}$$

As is well known, this condition is equivalent to requiring that the coefficient sequence $\hat{f} := \{\hat{f}_k := \langle \psi_l, f \rangle_{X' \times X}\}$ is a null sequence (i.e., $\hat{f}_k \to 0$ if $k \to \infty$) for any $f \in X$.

Math Subject Classifications. 41A46, 42A10.
Keywords and Phrases. nonlinear m-term approximation, greedy algorithms, Haar system, unconditionality, biorthogonal systems in Banach spaces.

ISSN 1069-5869

For later use, we introduce the following notation. In agreement with the above notation, we set $V_\Lambda := \{f \in X : \hat{f}_k = 0, \ k \notin \Lambda\}$ for arbitrary $\Lambda \subset I$, and denote by g_Λ the generic element of V_Λ. If the index set Λ is finite, i.e., $\#\Lambda < \infty$, then the g_Λ are called *polynomials,* and we introduce the notation

$$\mathbf{1}_\Lambda := \sum_{k \in \Lambda} \phi_k \in V_\Lambda \, .$$

The nonlinear set

$$\Sigma_m := \bigcup_{\#\Lambda \leq m} V_\Lambda$$

contains all *m-term polynomials*, i.e., polynomials with $\leq m$ non-zero coefficients. The following partial sum operators are well defined for any finite $\Lambda \subset I$:

$$S_\Lambda f := \sum_{k \in \Lambda} \hat{f}_k \phi_k \, , \qquad S_{I \setminus \Lambda} f = f - S_\Lambda f \, .$$

For arbitrary null sequences c, we set $|c| := \{|c_k|, \ k \in I\}$, $c \geq c'$ means that $|c_k| \geq |c'_k|$, $k \in I$, and $c^* = \{c^*_k = c_{l(k)}, \ k \in I\}$ denotes the decreasing rearrangement of c which is defined by a one-to-one mapping $l_c : I \mapsto I$ such that $|c^*|$ is monotonically decreasing, i.e.,

$$\left|c_{l_c(1)}\right| \geq \left|c_{l_c(2)}\right| \geq \dots \, .$$

Obviously, l_c is only unique up to index permutations for c_k with equal absolute value. The reader is assured that the results of this article do not depend on the specific choice of the index mapping l_c.

Temlyakov [9, 10, 11] has written a series of papers on the error behavior of greedy algorithms associated with various classical systems (uniformly bounded orthonormal systems, wavelet systems, etc.) in various Banach spaces (L_p-spaces, $\mathcal{F}^r_q$-spaces, etc.). Using the above notation, the simplest *greedy algorithm* is given by

$$f \in X \longmapsto G_m f := \sum_{k \in \Lambda_m} c_k \phi_k \, , \quad \Lambda_m := \left\{ l_{\hat{f}}(k), \ k = 1, \dots, m \right\} \, , \tag{1.2}$$

where the mapping $l_{\hat{f}} : I \to I$ was defined above. The name *greedy* is justified since $G_m f = G_{m-1} f + G_1(f - G_{m-1} f)$, $m \geq 2$, is the result of recursively applying the greedy operator G_1. Theoretically, if $N = \dim X$ is finite, and all $\hat{f}_k$ are computed exactly, then a simple sort is sufficient to define $l_{\hat{f}}$ and, thus, $G_m f$ for all $1 \leq m < N$. We will not discuss implementational issues or other versions of greedy algorithms. (See [11, 3, 4, 12, 13] and the references cited therein; to some of them, the approach below carries over with minor modifications.) Rather, we concentrate on the worst-case comparison of the greedy approximation error $\|f - G_m f\|_X$ with its lower bound given by the *best m-term approximation* with respect to Φ:

$$\sigma_m(f)_X := \inf_{g \in \Sigma_m} \|f - g\|_X \equiv \inf_{\Lambda \subset I : \#\Lambda \leq m} \ \inf_{g_\Lambda \in V_\Lambda} \|f - g_\Lambda\|_X \, . \tag{1.3}$$

More precisely, we are interested in estimates for the quantity

$$\delta_{X,\Phi}(m) := \sup_{f \in X} \frac{\|f - G_m f\|_X}{\sigma_m(f)_X} \, , \qquad m \geq 1 \, . \tag{1.4}$$

Although similar definitions can be considered for other variants of greedy algorithms, we solely deal with the investigation of this quantity which describes the *worst case behavior* of the greedy algorithm G_m for each fixed m with respect to arbitrary $f \in X$.

When analyzing the approach taken in [9, 10, 11] in various partial situations, it is easy to see that it relies on a few basic ingredients, mostly estimates for m-term polynomials. In the preprint version [8] of the present article, we presented a rather technical set of inequalities yielding upper and lower bounds for $\delta_{X,\Phi}(m)$ in the general case. We also considered some simplified conditions based on monotone comparison functions, and applied them to various examples. When presenting our results from [8], V.N. Temlyakov brought to our attention the papers [7] and [15] which contain closely related results. In particular, [7] gives a characterization of all bases Φ in X such that

$$\delta_{X,\Phi}(m) = \mathrm{O}(1) , \qquad m \to \infty , \tag{1.5}$$

in terms of *unconditionality* and inequalities for the polynomials $\mathbf{1}_\Lambda$ (the so-called *democracy condition*). The connection of (1.5) with the weaker notion of *quasi-greedy bases*, i.e., Φ for which the greedy algorithm converges for any $f \in X$, and with various other related properties is also studied in [7]. Similar results have been found in [15] for general biorthogonal systems.

The aim of this article is to present a unified approach to estimating $\delta_{X,\Phi}(m)$, by introducing generalizations of the unconditionality and democracy conditions, and to derive more practical criteria if

$$\nu_1(f) \leq \|f\|_X \leq \nu_2(f) , \qquad f \in X , \tag{1.6}$$

where ν_i , $i = 1, 2$, are monotone comparison functions. These results can be found in Section 2.

We also give some new results for the univariate Haar system H. In Section 3 we prove the equality

$$\delta_{L_p,H}(m) = 3m + 1 , \qquad m \geq 1 , \quad p = 1, \infty , \tag{1.7}$$

which implies that the trivial estimate

$$\delta_{X,\Phi}(m) \leq 3M_\Psi m + 1 , \qquad m \geq 1 , \tag{1.8}$$

cannot be improved. Finally, the asymptotic behavior of $\delta_{BMO,H}(m)$ and $\delta_{dBMO,H}(m)$ is determined.

2. Abstract Estimates

Fix a system $\Phi \subset X$ satisfying the properties of Section 1. For arbitrary $f \in X$, define

$$\|f\|_{X,\Phi;1} := \inf_{g \in X : \hat{g} \geq \hat{f}} \|g\|_X , \qquad \|f\|_{X,\Phi;2} := \sup_{g \in X : \hat{g} \leq \hat{f}} \|g\|_X . \tag{2.1}$$

In general, these quantities are not norms on X ($\|\cdot\|_{X,\Phi;1}$ does not satisfy the triangle inequality, while $\|\cdot\|_{X,\Phi;2}$ may take the value $+\infty$). It is well known [5, Proof of Theorem I.3.2] that

$$\|f\|_{X,\Phi;2} = \sup_{\epsilon_k = \pm 1} \left\| \sum_{k \in I} \epsilon_k \hat{f}_k \phi_k \right\|_X . \tag{2.2}$$

Obviously,

$$\|f\|_{X,\Phi;1} \leq \|f\|_X \leq \|f\|_{X,\Phi;2} \qquad \forall\, f \in X . \tag{2.3}$$

Let us introduce the quantities

$$A_{X,\Phi}(m) := \sup_{g_\Lambda \in \Sigma_m} \frac{\|g_\Lambda\|_{X,\Phi;2}}{\|g_\Lambda\|_{X,\Phi;1}} , \quad m \geq 1 , \tag{2.4}$$

and

$$B_{X,\Phi}(m) := \sup_{\Lambda' \cap \Lambda'' = \emptyset,\ 1 \le \#\Lambda' = \#\Lambda'' \le m} \frac{\|\mathbf{1}_{\Lambda'}\|_{X,\Phi;2}}{\|\mathbf{1}_{\Lambda''}\|_{X,\Phi;1}}, \quad m \ge 1 . \tag{2.5}$$

The quantities $A_{X,\Phi}(m)$ indicate how close Φ is to being an unconditional basis (indeed, unconditionality is equivalent to

$$A_{X,\Phi}(m) = \mathrm{O}(1), \qquad m \to \infty , \tag{2.6}$$

see [5, Theorem I.3.2]). On the other hand, $\{B_{X,\Phi}(m)\}$ is connected to the notion of democracy resp. superdemocracy introduced in [7]. Roughly speaking,

$$B_{X,\Phi}(m) = \mathrm{O}(1), \qquad m \to \infty , \tag{2.7}$$

implies democracy and superdemocracy (and is equivalent to them if (2.6) is satisfied). Examples showing that $\{A_{X,\Phi}(m)\}$ and $\{B_{X,\Phi}(m)\}$ may behave independently for $m \to \infty$ can be found in [7]. Our first result generalizes Theorem 1 of [7] as well as Theorem 4 of [15].

Theorem 1.

For any system $\Phi \subset X$ satisfying the assumptions of Section 1, we have

$$\delta_{X,\Phi}(m) \le 1 + 2A_{X,\Phi}(m) + B_{X,\Phi}(m) , \tag{2.8}$$

where $m \ge 1$. This upper estimate is asymptotically sharp as we have

$$\delta_{X,\Phi}(m) \asymp \max(A_{X,\Phi}(m), B_{X,\Phi}(m)) , \qquad m \to \infty . \tag{2.9}$$

We start the proof of Theorem 1 with a formula for $A_{X,\Phi}(m)$ which exhibits the relationship with unconditionality more explicitly. Set

$$U_{X,\Phi}(m) := \sup_{g \in X} \sup_{\#\Lambda \le m} \frac{\|S_\Lambda g\|_X}{\|g\|_X}, \quad m \ge 1 . \tag{2.10}$$

Note that the second infimum could have been restricted to all Λ with $\#\Lambda = m$, without changing the value of $U_{X,\Phi}(m)$ (since $\hat{g}_k \to 0$ we can enlarge any Λ to the necessary cardinality while essentially preserving the value of $\|S_\Lambda g\|_X$).

Lemma 1.

We have

$$U_{X,\Phi}(m) \le A_{X,\Phi}(m) = \sup_{g \in X} \sup_{\Lambda' \cap \Lambda'' = \emptyset,\ \#\Lambda' + \#\Lambda'' \le m} \frac{\|S_{\Lambda'} g - S_{\Lambda''} g\|_X}{\|g\|_X} \le 2U_{X,\Phi}(m) \tag{2.11}$$

for all $m \ge 1$.

Proof. By definition of $A_{X,\Phi}(m)$ and by (2.1) and (2.2) it follows that

$$\begin{aligned} A_{X,\Phi}(m) &= \sup_{g \in X} \sup_{1 \le \#\Lambda \le m} \sup_{g_\Lambda \in V_\Lambda : \hat{g}_\Lambda \le \hat{g}} \frac{\|g_\Lambda\|_{X,\Phi;2}}{\|g\|_X} \\ &= \sup_{g \in X} \sup_{1 \le \#\Lambda \le m} \max_{\epsilon_k = \pm 1} \frac{\left\| \sum_{k \in \Lambda} \epsilon_k \hat{g}_k \phi_k \right\|_X}{\|g\|_X} \\ &= \sup_{g \in X} \sup_{\Lambda' \cap \Lambda'' = \emptyset,\ \#\Lambda' + \#\Lambda'' \le m} \frac{\|S_{\Lambda'} g - S_{\Lambda''} g\|_X}{\|g\|_X} . \end{aligned}$$

The remaining inequalities in (2.11) are obvious.

Another preparation is the following, more technical observation. If we define

$$\|\mathbf{1}_\Lambda\|^*_{X,\Phi,1} := \inf_{g_\Lambda \in V_\Lambda,\, g_{I\setminus\Lambda}\in V_{I\setminus\Lambda}:\, \hat{g}_\Lambda \ge \hat{\mathbf{1}}_\Lambda,\, \|\hat{g}_{I\setminus\Lambda}\|_{\ell_\infty}\le 1} \left\| g_\Lambda + g_{I\setminus\Lambda}\right\|_X ,$$

and

$$B^*_{X,\Phi}(m) := \sup_{\Lambda'\cap\Lambda''=\emptyset,\ 1\le\#\Lambda'=\#\Lambda''\le m} \frac{\|\mathbf{1}_{\Lambda'}\|_{X,\Phi;2}}{\|\mathbf{1}_{\Lambda''}\|^*_{X,\Phi;1}} , \quad m \ge 1 , \tag{2.12}$$

then

$$B^*_{X,\Phi}(m) \le B_{X,\Phi}(m) \le B^*_{X,\Phi}(m) + A_{X,\Phi}(m) , \quad m \ge 1 . \tag{2.13}$$

The lower estimate is obvious since $\|\mathbf{1}_{\Lambda''}\|_{X,\Phi;1} \le \|\mathbf{1}_{\Lambda''}\|^*_{X,\Phi;1}$ by definition. To establish the upper bound, let $\varepsilon > 0$ be fixed. According to the definition (2.5), we can find disjoint sets Λ', Λ'' of cardinality $1 \le \#\Lambda' = \#\Lambda'' = m' \le m$, and functions

$$g_{\Lambda'} = \sum_{k\in\Lambda'} \epsilon_k \phi_k \qquad (\epsilon_k = \pm 1) ,$$

and $g = g_{\Lambda''} + g_{I\setminus\Lambda''}$ satisfying $\hat{g}_{\Lambda''} \ge \hat{\mathbf{1}}_{\Lambda''}$, such that

$$(1-\varepsilon) B_{X,\Phi}(m) \le \frac{\|g_{\Lambda'}\|_X}{\|g\|_X} .$$

Now, introduce a partitioning of $\Lambda' = \tilde{\Lambda} \cup \bar{\Lambda}$ into the two disjoint subsets

$$\tilde{\Lambda} = \left\{k \in \Lambda' :\ |\hat{g}_k| > 1\right\} , \quad \bar{\Lambda} = \Lambda'\setminus\tilde{\Lambda} .$$

If $\tilde{\Lambda} = \emptyset$ then we can find a real number $\lambda \ge 1$ and a set $\tilde{\Lambda}''$, again disjoint from Λ' and of cardinality m' such that

$$|\hat{g}_k| \ge \lambda , \quad k \in \tilde{\Lambda}'' , \qquad |\hat{g}_k| \le \lambda , \quad k \notin \tilde{\Lambda}'' .$$

This shows that

$$\lambda^{-1} \|g_{\Lambda'}\|_X \le \|\mathbf{1}_{\Lambda'}\|_{X,\Phi;2} , \quad \lambda^{-1}\|g\|_X \ge \left\|\mathbf{1}_{\tilde{\Lambda}''}\right\|^*_{X,\Phi;1} ,$$

which by (2.12) implies $(1-\varepsilon)B_{X,\Phi}(m) \le B^*_{X,\Phi}(m)$ in this case.

If $1 \le \tilde{m} := \#\tilde{\Lambda} \le m' \le m$ then we simply write

$$\frac{\|g_{\Lambda'}\|_X}{\|g\|_X} \le \frac{\left\|S_{\tilde{\Lambda}} g_{\Lambda'}\right\|_X}{\|g\|_X} + \frac{\left\|S_{\bar{\Lambda}} g_{\Lambda'}\right\|_X}{\|g\|_X} .$$

The second term can again be bounded by $B^*_{X,\Phi}(m)$ by repeating the above argument with Λ', m', and $g_{\Lambda'}$ replaced by $\bar{\Lambda}$, $\bar{m} = m' - \tilde{m}$, and $S_{\bar{\Lambda}} g_{\Lambda'}$, respectively. For the first term, recall that by definition of $\tilde{\Lambda}$, the coefficient bound for $g_{\Lambda'}$, and Lemma 1 we have

$$\left\|S_{\tilde{\Lambda}} g_{\Lambda'}\right\|_X \le \left\|S_{\tilde{\Lambda}} g\right\|_{X,\Phi;2} \le A_{X,\Phi}(m)\|g\|_X .$$

Thus, altogether we have

$$(1-\varepsilon)B_{X,\Phi}(m) \le A_{X,\Phi}(m) + B^*_{X,\Phi}(m) ,$$

which gives the upper bound in (2.13) if $\varepsilon \to 0$.

We can now prove the upper bound (2.8). For any given $f \in X$, let the index set Λ_m of cardinality $\#\Lambda_m = m$ be defined by $G_m f = S_{\Lambda_m} f$. For any $\varepsilon > 0$, we can find a polynomial $g_\Lambda \in V_\Lambda$ such that $g := f - g_\Lambda$ satisfies $\|g\|_X \le (1-\varepsilon)^{-1}\sigma_m(f)_X$. Without loss of generality, we can assume that $\#\Lambda = m$, too. Now, set

$$\Lambda' = \Lambda \backslash \Lambda_m , \qquad \Lambda'' = \Lambda_m \backslash \Lambda .$$

These sets are disjoint and have equal cardinality $m' = \#\Lambda' = \#\Lambda'' \le m$. Since $S_{I\backslash(\Lambda\cup\Lambda_m)} f = S_{I\backslash(\Lambda\cup\Lambda_m)} g$, we can write

$$f - G_m f = S_{I\backslash(\Lambda\cup\Lambda_m)} g + S_{\Lambda'} f = g - S_\Lambda g - S_{\Lambda''} g + S_{\Lambda'} f .$$

Applying the triangle inequality, we get

$$\begin{aligned}(1-\varepsilon)\frac{\|f - G_m f\|_X}{\sigma_m(f)_X} &\le 1 + \frac{\|S_\Lambda g\|_X}{\|g\|_X} + \frac{\|S_{\Lambda''} g\|_X}{\|g\|_X} + \frac{\|S_{\Lambda'} f\|_X}{\|g\|_X} \\ &\le 1 + 2U_{X,\Phi}(m) + \frac{\|\mathbf{1}_{\Lambda'}\|_{X,\Phi;2}}{\|\mathbf{1}_{\Lambda''}\|_{X,\Phi;1}} \\ &\le 1 + 2A_{X,\Phi}(m) + B_{X,\Phi}(m) .\end{aligned}$$

In the estimation we have used that by definition of $G_m f$, Λ', and Λ'', we have

$$\left|\hat f_k\right| \le \left|\hat g_{k'}\right| = \left|\hat f_{k'}\right| \qquad \forall\, k \in \Lambda' ,\ k' \in \Lambda'' .$$

Letting $\varepsilon \to 0$ and taking the supremum with respect to all $f \in X$, we have (2.8).

To prove (2.9), we need to establish a matching lower bound. By definition of $U_{X,\Phi}(m)$, for any $\varepsilon > 0$, we can find a non-zero $g \in X$ and an index set Λ of cardinality $\#\Lambda = m$ such that

$$\|S_\Lambda g\|_X \ge (1-\varepsilon) U_{X,\Phi}(m) \|g\|_X .$$

Set

$$f = (M_\Psi \|g\|_X + 1)\,\mathbf{1}_\Lambda + g - S_\Lambda g .$$

Since $|\hat g_k| \le M_\Psi \|g\|_X$ by (1.1), we have

$$\|f - G_m f\|_X = \|g - S_\Lambda g\|_X \ge \big((1-\varepsilon)U_{X,\Phi}(m) - 1\big)\, \|g\|_X .$$

On the other hand, f and g differ by a polynomial from V_Λ which gives $\sigma_m(f)_X \le \|g\|_X$. Altogether, for $\varepsilon \to 0$ we obtain

$$\delta_{X,\Phi}(m) \ge U_{X,\Phi}(m) - 1 \ge \frac{A_{X,\Phi}(m)}{2} - 1 , \tag{2.14}$$

where we already have incorporated the result of Lemma 1.

Analogously, from definition (2.12), for any $\varepsilon > 0$ we find disjoint sets Λ', Λ'' with $\#\Lambda' = \#\Lambda'' \le m$, and functions $g_{\Lambda'} \in V_{\Lambda'}$, $g \in X$, such that

$$\hat g_{\Lambda'} \le \hat{\mathbf{1}}_{\Lambda'} , \qquad |\hat g_k| \begin{cases} \ge 1 , & k \in \Lambda'', \\ \le 1 , & k \notin \Lambda'' \end{cases}$$

and

$$\|g_{\Lambda'}\|_X \ge (1-\varepsilon) B^*_{X,\Phi}(m) \|g\|_X .$$

Choose any $\tilde{\Lambda}$ disjoint with Λ' and Λ'' such that the cardinality of $\Lambda = \Lambda' \cup \tilde{\Lambda}$ equals m. Set

$$f = g_{\Lambda'} + S_{I\setminus(\Lambda''\cup\Lambda)}g + (1+\varepsilon)\left(S_{\Lambda''}g + \mathbf{1}_{\tilde{\Lambda}}\right) .$$

Then

$$\begin{aligned}\|f - G_m f\|_X &= \left\|g_{\Lambda'} + S_{I\setminus(\Lambda''\cup\Lambda)}g\right\|_X \\ &\geq \left((1-\varepsilon)B^*_{X,\Phi}(m) - 1 - 2U_{X,\Phi}(m)\right)\|g\|_X ,\end{aligned}$$

and (by subtracting a suitable polynomial from V_Λ)

$$\sigma_m(f)_X \leq \inf_{g_\Lambda \in V_\Lambda} \|f - g_\Lambda\|_X \leq \|g + \varepsilon S_{\Lambda''}g\|_X \leq \left(1 + \varepsilon U_{X,\Phi}(m)\right)\|g\|_X .$$

Together with Lemma 1, (2.13), and after letting $\varepsilon \to 0$, this gives

$$\delta_{X,\Phi}(m) \geq B_{X,\Phi}(m) - 3A_{X,\Phi}(m) - 1 , \quad m \geq 1 . \tag{2.15}$$

Combining (2.14) and (2.15), it is not hard to derive the lower bound in

$$\frac{1}{8}\max\left(A_{X,\Phi}(m), B_{X,\Phi}(m)\right) \leq \delta_{X,\Phi}(m) \leq 4\max\left(A_{X,\Phi}(m), B_{X,\Phi}(m)\right) ,$$

while the upper bound is obvious from (2.8). This proves (2.9), and completes the proof of Theorem 1. Note that the proof shows that $A_{X,\Phi}(m)$ could be replaced by $U_{X,\Phi}(m)$ in both relations (2.8) and (2.9). It is also possible to replace $B_{X,\Phi}(m)$ by $B^*_{X,\Phi}(m)$ in (2.9). □

Although Theorem 1 gives the correct asymptotic behavior for the quantities $\delta_{X,\Phi}(m)$ in the general case, its application to particular systems is tedious, partly due to the complicated, implicit definitions of $A_{X,\Phi}(m)$ and $B_{X,\Phi}(m)$. We will show next that the upper estimates can be simplified if it is possible to introduce suitable comparison functions $\nu_i : X \longmapsto \mathbb{R}_+ \cup \{\infty\}$, $i = 1, 2$, such that

$$\nu_1(f) \leq \|f\|_X , \qquad \|g_\Lambda\|_X \leq \nu_2(g_\Lambda) , \tag{2.16}$$

holds for all $f \in X$ and all polynomials $g_\Lambda \in V_\Lambda$ and any Λ with $\#\Lambda < \infty$ (assumption (2.16) and the considerations below show that ν_2 only needs to be defined for polynomials g_Λ, not necessarily for general $f \in X$).

We call a $\nu : X \longmapsto \mathbb{R}_+ \cup \{\infty\}$

- *monotone* if $\nu(f) \leq \nu(g)$ whenever $\hat{f} \leq \hat{g}$, and
- *weakly rearrangement-invariant* if $\nu(\mathbf{1}_{\Lambda'}) \leq \beta\nu(\mathbf{1}_{\Lambda''})$ with some fixed $1 \leq \beta < \infty$ for all finite disjoint index sets Λ', Λ'' satisfying $\#\Lambda' = \#\Lambda''$.

Clearly, these definitions depend on Φ. If $\nu(f) = \|\hat{f}\|$ is given by a symmetric sequence norm $\|\cdot\|$ such as (a multiple of) an ℓ_τ-norm then it satisfies both these conditions (with $\beta = 1$). Some other examples of practical use include Littlewood–Paley type norms (see, e.g., [10]).

It is easy to see that the comparison functions

$$\nu_{X,\Phi;1}(f) := \|f\|_{X,\Phi;1} , \qquad \nu_{X,\Phi;2}(f) := \|f\|_{X,\Phi;2}$$

are monotone, and satisfy (2.16). Moreover, this choice is *optimal* in the following sense: if two monotone comparison functions ν_1, ν_2 satisfy (2.16) then

$$\nu_1(f) \leq \nu_{X,\Phi;1}(f) , \qquad \nu_{X,\Phi;2}(g_\Lambda) \leq \nu_2(g_\Lambda) \qquad \forall\, f \in X ,\ g_\Lambda \in V_\Lambda . \tag{2.17}$$

With this observation at hand, we have the following obvious lemma.

Lemma 2.

Assume that there are two monotone comparison functions ν_1, ν_2 *such that* (2.16) *holds. Then*

$$U_{X,\Phi}(m) \le A_{X,\Phi}(m) \le A(m) := \sup_{g_\Lambda \in \Sigma_m} \frac{\nu_2(g_\Lambda)}{\nu_1(g_\Lambda)} , \tag{2.18}$$

and

$$B^*_{X,\Phi}(m) \le B_{X,\Phi}(m) \le B(m) := \sup_{\Lambda' \cap \Lambda'' = \emptyset,\, \#\Lambda' = \#\Lambda'' \le m} \frac{\nu_2(\mathbf{1}_{\Lambda'})}{\nu_1(\mathbf{1}_{\Lambda''})} . \tag{2.19}$$

If either ν_1 *or* ν_2 *is weakly rearrangement-invariant then* $B(m) \le \beta A(m)$.

Proof. The inequalities (2.18) and (2.19) immediately follow from the definitions of the quantities and (2.17). The last statement is also trivial: if ν_2 is weakly rearrangement-invariant then $\nu_2(\mathbf{1}_{\Lambda''}) \le \beta \nu_2(\mathbf{1}_{\Lambda'})$ can be used; if ν_1 is weakly rearrangement-invariant then $\nu_1(\mathbf{1}_{\Lambda'}) \ge \beta^{-1}\nu_2(\mathbf{1}_{\Lambda''})$ is appropriate. This concludes the proof of Lemma 2. □

A yet simpler criterion is formulated in the following.

Corollary 1.

Let $\nu = \nu_1$ *be a monotone and weakly rearrangement-invariant comparison function such that the first inequality of* (2.16) *is satisfied. Then,*

$$A_{X,\Phi}(m) \le \hat{A}(m) := \sup_{g_\Lambda \in \Sigma_m} \frac{\|g_\Lambda\|_X}{\nu(g_\Lambda)} , \qquad B_{X,\Phi}(m) \le \beta \hat{A}(m) . \tag{2.20}$$

Consequently, we have

$$\delta_{X,\Phi}(m) \le 1 + (2+\beta)\hat{A}(m) , \qquad m \ge 1 . \tag{2.21}$$

This result suffices for most of the applications to the examples considered in [9, 10, 11], as was demonstrated in Section 3 of [8].

Remark 1. Theorem 1 (in conjunction with Lemma 2) and Corollary 1 yield upper bounds for $\delta_{X,\Phi}(m)$ the optimality of which depends on the proper choice of the comparison functions. Some simplified lower bounds for either $A_{X,\Phi}(m)$ or $B_{X,\Phi}(m)$ in terms of comparison functions have been formulated in [8]. In many situations, using examples based on the polynomials $\mathbf{1}_\Lambda$ will yield matching lower bounds. For example, the quantity $B^*_{X,\Phi}(m)$ can often be estimated from below by constructing disjoint Λ' and Λ'' ($\#\Lambda' = \#\Lambda'' \le m$) such that the ratio $\|\mathbf{1}_{\Lambda'}\|_X / \|\mathbf{1}_{\Lambda''}\|_X$ is large (and comparable to the upper bounds). For examples, we refer to Section 3 and [8].

Remark 2. We conclude with showing the crude estimate (1.8). Obviously, by definition of $\hat{f}$ and M_Ψ we have

$$\nu(f) := M_\Psi^{-1} \left\| \hat{f} \right\|_{\ell_\infty} \le \|f\|_X \qquad \forall\, f \in X ,$$

and

$$\|g_\Lambda\|_X \le \sum_{k \in \Lambda} \left|(\hat{g}_\Lambda)_k\right| \le m \left\|\hat{g}_\Lambda\right\|_{\ell_\infty} \le m M_\Psi \nu(g_\Lambda)$$

for all $g_\Lambda \in \Sigma_m$. Thus, by Corollary 1 we have (1.8).

3. Applications to the Haar System

In this section, we present some applications of the material of the previous section to the Haar system. In most cases, we use Corollary 1. Roughly speaking, the art consists in detecting a suitable monotone and weakly symmetric comparison function ν for which the lower bound in (2.16) holds tightly, and to find the appropriate $\hat{A}(m)$. Matching lower bounds are obtained by using Remark 1. As was mentioned before, all univariate examples considered in [9, 10, 11] are covered by our approach, see [8] for some more details.

Here we only deal with the univariate Haar system $H = \{h_k\}$ which is the prototype of wavelet systems. We use the following notation. Set $\Delta_1 := [0, 1]$, and call

$$\Delta_{2^{j-1}+l} := \left[(l-1)2^{-j+1}, l2^{-j+1}\right], \qquad l = 1, \ldots, 2^{j-1},$$

are the dyadic intervals of level $j \geq 1$. χ_Δ denotes the characteristic function of an interval Δ. With each of these intervals, we associate a Haar function h_k with support in Δ_k by setting $h_1 := \chi_{\Delta_1}$ and $h_{2^{j-1}+l} := \chi_{\Delta_{2^j+2l-1}} - \chi_{\Delta_{2^j+2l}}$ for the remaining $k = 2^{j-1} + l \geq 2$.

We will first consider the Haar system in the Banach spaces $L_p := L_p(0, 1)$, $1 \leq p \leq \infty$. More precisely, H denotes now the L_p-normalized system $\{|\Delta_k|^{-1/p} h_k\}$. Thus, if we talk about Haar coefficients of $f \in L_p$, we have in mind the sequence $\hat{f}^p \equiv \{|\Delta_k|^{1/p} \hat{f}_k\}$, where $f = \sum_{k=1}^{\infty} \hat{f}_k h_k$. Obviously, $\hat{f}^\infty = \hat{f}$.

Let us start with the case $1 < p < \infty$ [10]. For these p, H is an unconditional basis in L_p, and the comparison function of our choice is the Littlewood–Paley norm

$$\nu(f) = C_p^{-1} \left\| \left(\sum_{k=1}^{\infty} \left|\hat{f}_k\right|^2 \chi_{\Delta_k} \right)^{1/2} \right\|_{L_p} .$$

For some choice of the positive constant C_p we have

$$\nu(f) \leq \|f\|_{L_p} \leq C_p^2 \nu(f) \qquad \forall\, f \in L_p(0, 1) ,$$

compare [6, III, Theorem 9]. Due to this norm equivalence, $\hat{A}(m) \leq C_p^2 < \infty$ for all m. Thus, to show the basic result of [10],

$$\delta_{L_p,H}(m) \approx 1 \, , m \geq 1 \, , 1 < p < \infty \, , \tag{3.1}$$

i.e., the asymptotic optimality of the greedy algorithm (up to a constant factor), as a consequence of Corollary 1, we only need to verify that ν is weakly rearrangement-invariant (the monotonicity i) is obvious). The proof of ii) with some $\beta = \beta_p < \infty$ is essentially contained in [10, Lemma 2.1-2] (note differences in notation), we do not repeat it here.

Next we come to the case $p = \infty$ which has been dealt with in [9, Section 6.2], see also [2] for earlier results. To be precise, the L_∞-space we are dealing with is the closed, separable subspace in L_∞ generated by H. A suitable comparison function is defined by the norm

$$\nu(f) := \left\|\hat{f}\right\|_{\ell_\infty} \leq \|f\|_{L_\infty} \cdot$$

which satisfies i) and ii) (with $\beta = 1$). Inequality (2.20) is fulfilled with $\hat{A}(m) = m$ since

$$\|g_\Lambda\|_{L_\infty} \leq \sum_{k \in \Lambda} |\hat{g}_\Lambda| \leq m \cdot \nu(g_\Lambda)$$

for all Haar polynomials $g_\Lambda \in \Sigma_m$. This gives

$$\delta_{L_\infty,H}(m) \le 3m+1\ , \quad m \ge 1\ ,$$

where we have used Corollary 1.

Since $2^m > m$ for all $m \ge 1$, the two m-term Haar polynomials

$$\mathbf{1}_\Lambda = h_1 + \sum_{j=1}^{m-1} h_{2^{j-1}+1}\ , \qquad \mathbf{1}_{\Lambda'} = \sum_{l=1}^{m} h_{2^m+m}\ ,$$

correspond to disjoint Λ and Λ' and satisfy

$$\|\mathbf{1}_\Lambda\|_{L_\infty} = m\ , \qquad \|\mathbf{1}_{\Lambda'}\|_{L_\infty} = 1\ .$$

This gives the lower bound

$$\delta_{L_\infty,H}(m) \ge m\ , \quad m \ge 1\ ,$$

if we consider $f = (1+\epsilon)\mathbf{1}_{\Lambda'} + \mathbf{1}_\Lambda$, and let $\epsilon > 0$ tend to zero.

The following result shows that we can do better. Moreover, it shows that the estimation techniques of Section 2 are essentially sharp.

Theorem 2.

We have the identity

$$\delta_{L_\infty,H}(m) = 3m+1\ , \quad m \ge 1\ . \tag{3.2}$$

Since $M_\Psi = 1$ for the biorthogonal system $\Psi = \{|\Delta_k|^{-1}h_k\}$ of H, this also shows that (1.8) *is the best possible.*

Proof. Only an improved lower bound needs to be established. Let $k \ge 3$, and define

$$g_r = h_{2^{rk-1}+1} + \frac{1}{2}h_{2^{rk-2}+1} + \dots \frac{1}{2^{(k-2)}}h_{2^{(r-1)k+1}+1}\ ,$$

and

$$f_r = h_{2^{(r-1)k}+1} - b_k^{-1} g_r\ , \quad r = 1,\dots,2m\ ,$$

where $b_k \equiv \sum_{s=0}^{k-2} 2^{-s} = 2 - 2^{-(k-2)} > 1$. Obviously,

$$g_r(x) = \begin{cases} b_k\ , & x \in \Delta_{2^{rk}+1}\ , \\ -2^{-(k-2)}\ , & x \in \Delta_{2^{(r-1)k+1}+1} \backslash \Delta_{2^{rk}+1}\ , \\ 0\ , & x \in [0,1] \backslash \Delta_{2^{(r-1)k+1}+1}\ . \end{cases}$$

and

$$f_r(x) = \begin{cases} 0\ , & x \in \Delta_{2^{rk}+1} \cup ([0,1] \backslash \Delta_{2^{(r-1)k}+1})\ , \\ 1 + 2^{-(k-2)} b_k^{-1}\ , & x \in \Delta_{2^{(r-1)k+1}+1} \backslash \Delta_{2^{rk}+1}\ , \\ -1\ , & x \in \Delta_{2^{(r-1)k}+1} \backslash \Delta_{2^{(r-1)k+1}+1}\ . \end{cases}$$

Note that

$$\|f_r\|_{L_\infty} < 1 + 2^{-(k-2)}\ , \qquad \left\|h_{2^{(r-1)k}+1} - g_r\right\|_{L_\infty} = 1 + 2^{-(k-2)}\ . \tag{3.3}$$

Set

$$f = \sum_{r=1}^{m} f_r + (1-\epsilon)\left(\sum_{r=m+1}^{2m-1}\left(f_r - 2h_{2^{(r-1)k}+1}\right) - \left(h_{2^{(2m-1)k}+1} + g_{2m}\right)\right),$$

where $0 < \epsilon < 1$ is arbitrary. It is easy to see that the m largest in absolute value coefficients of f are 1, and associated with the Haar functions $h_{2^{(r-1)k}+1}$, $r = 1, \ldots, m$. Thus,

$$\begin{aligned} \|f - G_m f\|_{L_\infty} &\geq \left|(f - \textstyle\sum_{r=1}^{m} h_{2^{(r-1)k}+1})(0+)\right| \\ &= |-m + (1-\epsilon)(-2(m-1) - 1 - b_k)| \\ &= 3m + 1 - (2m+1)\epsilon - (1-\epsilon)2^{-(k-2)} . \end{aligned} \tag{3.4}$$

On the other hand, an upper estimate for $\sigma_m(f)_{L_\infty}$ can be obtained from (3.3) as follows:

$$\begin{aligned} \sigma_m(f) &\leq \left\| f + 2(1-\epsilon)\sum_{r=m+1}^{2m} h_{2^{(r-1)k}+1} \right\|_{L_\infty} \\ &\leq \max\left\{\|f_r\|_{L_\infty},\ r = 1, \ldots, 2m-1,\ \left\|h_{2^{(2m-1)k}+1} - g_{2m}\right\|_{L_\infty}\right\} \\ &\leq 1 + 2^{-(k-2)} . \end{aligned}$$

Here, we have used that the supports of the f_r, $r = 1, \ldots, 2m-1$, and $h_{2^{(2m-1)k}+1} - g_{2m}$ are pairwise disjoint by construction. Together with (3.4), letting $\epsilon \to 0$, and then $k \to \infty$, we arrive at the equality for $\delta_{L_\infty,H}(m)$ in (3.2). □

In the final case $p = 1$, we could not find a suitable ν satisfying all the assumptions of Corollary 1. We will therefore use Theorem 1 in conjunction with Lemma 2. Set

$$\nu_1(f) := \sup\left\{\left|\hat{f}_1^1\right|, \sum_{l=1}^{2^{j-1}}\left|\hat{f}_{2^{j-1}+l}^1\right|,\ j \geq 1\right\}, \qquad \nu_2(f) := \left\|\hat{f}^1\right\|_{\ell_1},$$

where both comparison functions are monotone and satisfy (2.16). Moreover, ν_2 (but not ν_1) is weakly rearrangement-invariant. Recall that $\hat{f}_k^1 = |\Delta_k|\hat{f}_k$.

Inequality (2.20) holds with $\hat{A}(m) = m$, $\beta = 1$. Examples for the lower bounds can be constructed as in the case $p = \infty$ from the polynomials

$$\mathbf{1}_\Lambda = \sum_{l=1}^{m} 2^m h_{2^m+l}, \qquad \mathbf{1}_{\Lambda'} = h_1 + \sum_{j=1}^{m-1} 2^{j-1} h_{2^{j-1}+1},$$

which satisfy $\|\mathbf{1}_\Lambda\|_{L_1} = m$, $\|\mathbf{1}_{\Lambda'}\|_{L_1} = 1$. This gives

$$m \leq \delta_{L_1,H}(m) \leq 3m+1,$$

which is the known result from [9, Section 6.1]. As should be expected, an improvement as in Theorem 2 holds.

Theorem 3.

We have the equality

$$\delta_{L_1,H}(m) = 3m + 1, \quad m \geq 1. \tag{3.5}$$

Proof. The following example shows the equality in (3.5). For fixed large k and small $\epsilon > 0$, set

$$f = g + \epsilon \sum_{l=0}^{m-1} 2^{2lk} h_{2^{2lk}+1} - 2 \sum_{l=0}^{m-1} 2^{(2l+1)k} h_{2^{(2l+1)k}+1} ,$$

where

$$g = h_1 + \sum_{j=1}^{2km} 2^{j-1} h_{2^{j-1}+1} = 2^{2km} \chi_{[0,2^{-2km}]} .$$

Obviously,

$$\sigma_m(f)_{L_1} \le \|g\|_{L_1} + \epsilon \left\| \sum_{l=0}^{m-1} 2^{2lk} h_{2^{2lk}+1} \right\|_{L_1} \le 1 + \epsilon m . \qquad (3.6)$$

On the other hand, $G_m(f) = (1+\epsilon) \sum_{l=0}^{m-1} 2^{2lk} h_{2^{2lk}+1}$. Thus,

$$f - G_m(f) = g - \sum_{l=0}^{m-1} 2^{2lk} h_{2^{2lk}+1} - 2 \sum_{l=0}^{m-1} 2^{(2l+1)k} h_{2^{(2l+1)k}+1} .$$

From this formula, we see that

$$\begin{aligned} &|(f - G_m(f))(x)| \\ &\ge \begin{cases} 2^{2lk} - 2\sum_{j=0}^{2l-1} 2^{jk} \ge 2^{2lk}\left(1 - 4 \cdot 2^{-k}\right) & , \quad x \in \Delta'_{2l} , \\ 2^{(2l+1)k+1} - 2\sum_{j=0}^{2l} 2^{jk} \ge 2^{(2l+1)k+1}\left(1 - 2 \cdot 2^{-k}\right) & , \quad x \in \Delta'_{2l+1} , \\ 2^{2mk} - 2\sum_{j=0}^{2m-1} 2^{jk} \ge 2^{2mk}\left(1 - 4 \cdot 2^{-k}\right) & , \quad x \in \Delta'_{2m} , \end{cases} \end{aligned}$$

where $l = 0, 1, \dots, m-1$, and the intervals Δ'_r are defined by

$$\Delta'_r = \Delta_{2^{rk}+1} \backslash \Delta_{2^{(r+1)k}+1} , \quad r = 0, \dots, 2m-1 , \qquad \Delta'_{2m} = \Delta_{2^{2mk}+1} .$$

This implies

$$\|f - G_m(f)\|_{L_1} \ge (3m+1)\left(1 - 2^{-k}\right)\left(1 - 2^{-(k-2)}\right) ,$$

and together with (3.6) the lower bound in (3.5) if $k \to \infty$ and $\epsilon \to 0$. □

Remark 3. The examples for the lower bounds in Theorem 2 and 3 also show that no algorithm based on nonlinear partial sum operators

$$f \longmapsto S_{\Lambda(f)}(f)$$

with respect to the Haar system or based on more general scaled versions

$$f \longmapsto S^{\omega(f)}_{\Lambda(f)}(f) \equiv \sum_{k \in \Lambda(f)} \omega(f)_k \hat{f}_k h_k ,$$

where $|\Lambda(f)| \le m$ and $\omega(f)_k \ge 0$ for all $k \in \Lambda(f)$ can perform better than within a factor of $\approx m$ compared to the best m-term approximation in the L_∞- resp. L_1-norm. This remark also applies to the tree optimization algorithms discussed in [2] and [1]. It seems to be an open question whether better methods of low complexity can be found in these cases.

Let us investigate the changes if we replace L_∞ by spaces of functions of bounded mean oscillation. We consider two slightly different situations. The space BMO is the subspace of $L_1(0,1)$ defined as the closure of H under the norm

$$\|f\|_{BMO} = \left|f|_{[0,1]}\right| + \sup_{\Delta\subset[0,1]} \frac{1}{|\Delta|} \|f - f|_\Delta\|_{L_1(\Delta)} , \tag{3.7}$$

where $f|_\Delta = |\Delta|^{-1}\int_\Delta f(x)\,dx$ is the average value of f with respect to Δ, and the supremum in (3.7) is taken with respect to all intervals Δ in $[0,1]$ (for some details, see [6, Section V.3]). Analogously, $dBMO$ (the dyadic BMO-space) is defined with the weaker norm

$$\|f\|_{dBMO} = \left|f|_{[0,1]}\right| + \sup_{k\ge1} \frac{1}{|\Delta_k|} \left\|f - f|_{\Delta_k}\right\|_{L_1(\Delta_k)} \;(\le \|f\|_{BMO}) . \tag{3.8}$$

These spaces traditionally serve as replacements for L_∞ resp. C in questions of Fourier analysis (together with versions of the Hardy space H_1, they form 'better' endpoints for the scale of L_p-spaces, $1 < p < \infty$). Note that L_∞ is continuously imbedded into BMO and $dBMO$, more precisely, we have

$$\|f\|_{BMO} \le |f|_{[0,1]}| + \|f\|_{L_\infty} \le 2\|f\|_{L_\infty} \qquad \forall\, f \in L_\infty(0,1) . \tag{3.9}$$

Let us consider the behavior of greedy algorithms with respect to the Haar system (note that the Haar functions h_k have unit norm in both BMO and $dBMO$). Obviously, the comparison function $\nu(f) := \|\hat{f}\|_{\ell_\infty}$ is defined by a symmetric sequence norm and satisfies

$$\nu(f) \le \left\|\hat{f}\right\|^*_{\ell_\infty} = \left|\hat{f}_1\right| + \sup_{k\ge2}\left|\hat{f}_k\right| \le \|f\|_{dBMO}\,(\le \|f\|_{BMO}) \tag{3.10}$$

for all $f \in dBMO$ (resp. $f \in BMO$). In order to use Corollary 1, we need estimates for the BMO-norm of arbitrary m-term polynomials.

Lemma 3.

For any m-term Haar polynomial $g_\Lambda \in \Sigma_m$ we have

$$\|g_\Lambda\|_{BMO} \le m \left\|\hat{g}_\Lambda\right\|_{\ell_\infty} \tag{3.11}$$

and

$$\|g_\Lambda\|_{dBMO} \le C\left(1 + \sqrt{\log_2 m}\right) \left\|\hat{g}_\Lambda\right\|_{\ell_\infty} . \tag{3.12}$$

Both estimates are asymptotically sharp (up to constant factors).

Proof. The first inequality is trivial since

$$\|g_\Lambda\|_{BMO} \le \sum_{k\in\Lambda} \left|(\hat{g}_\Lambda)_k\right| \cdot \|h_k\|_{BMO} = \sum_{k\in\Lambda} \left|(\hat{g}_\Lambda)_k\right| \le m \left\|\hat{g}_\Lambda\right\|_{\ell_\infty} .$$

The function

$$g_\Lambda(x) = g(2x+1) , \quad g(x) = \begin{cases} \tilde{g}(x) = \sum_{j=1}^{[m/2]} h_{2^{j-1}+1} , & x \in [0,1] \\ -\tilde{g}(-x) , & x \in [-1,0) \end{cases} , \tag{3.13}$$

is a Haar polynomial for some index set Λ with $\#\Lambda = 2[m/2] \le m$. Thus, $g_\Lambda \in \Sigma_m$ and, obviously, $\|\hat{g}_\Lambda\|_{\ell_\infty} = 1$. Consider as Δ the interval of length $2^{-[m/2]}$ with midpoint at $x = 1/2$. By construction, $g_\Lambda(x) = \pm[m/2]$, $x \in \Delta$, depending on whether $x < 1/2$ or $x > 1/2$. This gives $g_\Lambda|_\Delta = 0$ and

$$\|g_\Lambda\|_{BMO} \ge |\Delta|^{-1} \|g_\Lambda\|_{L_1(\Delta)} = [m/2] .$$

Altogether, this gives the sharpness assertion for (3.11). Since Δ is *not* a dyadic interval, this reasoning does not lead to a lower bound for the $dBMO$-norm.

The inequality (3.12) follows from a coefficient norm equivalence for the Franklin system in BMO first established by Wojtaszczyk [14] (see also [6, Section VI.5, Theorem 11]) and a well-known connection between the Franklin series in BMO and Haar series in $dBMO$ (which follows by duality from the corresponding statements for Hardy spaces). We give an elementary alternative proof which is based on a neat extremal property of Haar polynomials in the $dBMO$ norm (see Lemma 4 below). Let any g_Λ be given, where $\|\hat{g}_\Lambda\|_{\ell_\infty} \le 1$ and $\#\Lambda \le m$. Let us observe the following. The function

$$g_k(x) \equiv \left(g_\Lambda - g_\Lambda|_{\Delta_k}\right)(x\,|\Delta_k| + x_k) = \left(\sum_{l\in\Lambda\,:\,\Delta_l\subset\Delta_k} \left(\hat{g}_\Lambda\right)_l h_l\right)(x\,|\Delta_k| + x_k) \ , \quad x \in [0,1] ,$$

where x_k is the left endpoint of Δ_k, coincides with a certain $g_{\Lambda'}$ with $\Lambda' \subset \mathbb{N}\backslash\{1\}$ satisfying $\#\Lambda' \le m$, $\|\hat{g}_{\Lambda'}\|_{\ell_\infty} \le 1$, and

$$|\Delta_k|^{-1} \left\| g_\Lambda - g_\Lambda|_{\Delta_k} \right\|_{L_1(\Delta_k)} = \|g_k\|_{L_1} = \|g_{\Lambda'}\|_{L_1} \ .$$

Thus,

$$\max_{g_\Lambda\,:\,\|\hat{g}_\Lambda\|_{\ell_\infty}\le 1,\,\#\Lambda\le m} \|g_\Lambda\|_{dBMO} \le 1 + \max_{g_\Lambda\,:\,\|\hat{g}_\Lambda\|_{\ell_\infty}\le 1,\ \Lambda\subset\mathbb{N}\backslash\{1\},\,\#\Lambda\le m} \|g_\Lambda\|_{L_1} , \tag{3.14}$$

i.e., estimates for the $dBMO$-norm reduce to estimates for the more convenient L_1-norm, and can be obtained easily.

Indeed, take any $\Lambda \subset \mathbb{N}\backslash\{1\}$, $\#\Lambda \le m$, and set $\Lambda_j = \Lambda \cap \{2^{j-1}+1, \dots, 2^j\}$, $j \ge 1$. Let $n = [\log_2 m]$ and $\tilde{\Lambda}_n = \cup_{j\le n} \Lambda_j$. Then for any g_Λ with $\|\hat{g}_\Lambda\|_{\ell_\infty} \le 1$, we have

$$\|g_\Lambda\|_{L_1} \le \left\| \sum_{k\in\tilde{\Lambda}_n} \left(\hat{g}_\Lambda\right)_k h_k \right\|_{L_1} + \sum_{j=n+1}^{\infty} \left\| \sum_{k\in\Lambda_j} \left(\hat{g}_\Lambda\right)_k h_k \right\|_{L_1} .$$

But

$$\sum_{j=n+1}^{\infty} \left\| \sum_{k\in\Lambda_j} \left(\hat{g}_\Lambda\right)_k h_k \right\|_{L_1} \le \sum_{j=n+1}^{\infty} 2^{-j+1} \cdot \#\Lambda_j \le 2^{-n} m \le 1 ,$$

and, by Lemma 4 below,

$$\left\| \sum_{k\in\tilde{\Lambda}_n} \left(\hat{g}_\Lambda\right)_k h_k \right\|_{L_1} \le A_n \le C\sqrt{n} .$$

According to (3.14), this gives (3.12). The sharpness of this inequality also follows from Lemma 4. The polynomial

$$\mathbf{1}_\Lambda := \sum_{k=2}^{2^n} h_k \in \Sigma_m$$

will do. This concludes the proof of Lemma 3. □

We have postponed the proof of the following.

Lemma 4.
We have

$$\left\|\sum_{k=2}^{2^n} c_k h_k\right\|_{L_1} \le \left\|\sum_{k=2}^{2^n} c_k h_k\right\|_{dBMO} \le A_n \|c\|_{\ell_\infty}, \tag{3.15}$$

where

$$A_n = \sum_{k=0}^{n} |n-2k| \binom{n}{k} = \left\{ \begin{array}{ll} 2m \cdot 2^{-2m} \binom{2m}{m}, & n = 2m \\ (2m+1) \cdot 2^{-2m} \binom{2m}{m}, & n = 2m+1 \end{array} \right\} \sim \sqrt{\frac{2n}{\pi}}.$$

Equality in (3.15) *is achieved for* $c_k = 1$, $k = 2, \dots, 2^n$.

Proof. According to our above considerations, the best constant A_n in (3.15) is given by

$$A_n = \max_{|c_k| \le 1} \left\|\sum_{k=2}^{2^n} c_k h_k\right\|_{L_1}.$$

We will show that this maximum is attained for $c_k = 1$, $k \ge 2$. Let $c = \{c_k,\ 2 \le k \le 2^n\}$ be a maximizer. Let l be the largest coefficient such that $c_l \ne 1$ (consequently, $-1 \le c_l < 1$). Without loss of generality, let $l = 2^{j-1} + r$ for some $1 \le j \le n$ and $r = 1, \dots, 2^{j-1}$. Changing the coefficient c_l will only influence the values of the polynomial on Δ_l. We can write

$$\left\|\sum_{k=2}^{2^n} c_k h_k\right\|_{L_1(\Delta_l)} = \left\| \underbrace{\sum_{k=2}^{2^{j-1}} c_k h_k}_{=g_1} + c_l h_l + \underbrace{\sum_{k=2^j+1}^{2^n} h_k}_{=g_2} \right\|_{L_1(\Delta_l)},$$

where we have used that $c_k = 1$ for $k > l$. Observe that g_1 is constant on all of Δ_l while $g_2(x + 2^j) = g_2(x)$ for all $x \in \Delta_l^+ = \Delta_{2^j+2r-1}$, the left half of Δ_l. Thus, by the elementary identity $|a+b| + |a-b| = 2\max(|a|, |b|)$, we have

$$\begin{aligned} \left\|\sum_{k=2}^{2^n} c_k h_k\right\|_{L_1(\Delta_l)} &= \|g_1 + g_2 + c_l\|_{L_1(\Delta_l^+)} + \|g_1 + g_2 - c_l\|_{L_1(\Delta_l^+)} \\ &= 2\,\|\max(|g_1 + g_2|, |c_l|)\|_{L_1(\Delta_l^+)} \le 2\,\|\max(|g_1 + g_2|, 1)\|_{L_1(\Delta_l^+)} \\ &= \|g_1 + g_2 + 1\|_{L_1(\Delta_l^+)} + \|g_1 + g_2 - 1\|_{L_1(\Delta_l^+)} = \left\|\sum_{k=2}^{2^n} c'_k h_k\right\|_{L_1(\Delta_l)}, \end{aligned}$$

where $c'_k = c_k$ for $k \ne l$, and $c'_l = 1$. Thus, c' is also a maximizer in the above expression for A_n. Repeating this reasoning, we arrive at the statement. □

It remains to compute A_n. Obviously,

$$A_n = \left\|\sum_{k=2}^{2^n} h_k\right\|_{L_1} = 2^{-n} \sum_{\delta = (\delta_1, \dots, \delta_n) \in \{1, -1\}^n} |\delta_1 + \cdots + \delta_n| = 2^{-n} \sum_{k=0}^{n} |n-2k| \binom{n}{k}.$$

If $n = 2m$ is even, then we continue

$$A_{2m} = \frac{4}{2^{2m}} \sum_{k=0}^{m-1} (m-k) \binom{2m}{k} = \frac{2m}{2^{2m}} \left(2^{2m} - \binom{2m}{m} \right) - \frac{4}{2^{2m}} \sum_{k=1}^{m-1} k \binom{2m}{k}$$

$$= \frac{2m}{2^{2m}} \left(2^{2m} - \binom{2m}{m} \right) - \frac{8m}{2^{2m}} \sum_{k=0}^{m-2} \binom{2m-1}{k}$$

$$= \frac{2m}{2^{2m}} \left(2^{2m} - \binom{2m}{m} \right) - \frac{4m}{2^{2m}} \left(2^{2m-1} - 2 \binom{2m-1}{m-1} \right) = \frac{2m}{2^{2m}} \binom{2m}{m} .$$

The case $n = 2m+1$ is treated analogously. Note that $A_{2m+1} = (1 + (2m)^{-1}) A_{2m}$. From Stirling's formula we find that

$$A_{2m} \sim \frac{2m}{2^{2m}} \left(\frac{2m}{e} \right)^{2m} \left(\frac{m}{e} \right)^{-2m} \frac{\sqrt{4\pi m}}{2\pi m} = \sqrt{\frac{4m}{\pi}} .$$

Altogether, this establishes Lemma 4.

After these preparations, we can formulate the following.

Theorem 4.
We have

$$\delta_{BMO,H}(m) \asymp m , \tag{3.16}$$

and

$$\delta_{dBMO,H}(m) \asymp \sqrt{\log_2 m} \tag{3.17}$$

as $m \to \infty$.

Proof. The upper bounds follow from Corollary 1 and the above Lemma 3. The lower bounds can easily be derived from the examples already mentioned (note that, as a happy coincidence, the asymptotic sharpness results for the inequalities in Lemma 3 are of the type $\mathbf{1}_\Lambda$ as required in our scheme). For example, to get the lower bound in (3.16), set

$$f = g_\Lambda + (1+\epsilon) \sum_{l=1}^{m} h_{2^m+l}$$

in the *BMO*-case, where g_Λ is defined in (3.13), and set

$$f = \sum_{k=2}^{2^n} h_k + (1+\epsilon) \sum_{l=1}^{m} h_{2^m+l}$$

in the *dBMO*-case ($n = [\log_2 m]$, $\epsilon > 0$). In both cases, $G_m f = (1+\epsilon) \sum_{l=1}^{m} h_{2^m+l}$, thus, the error of the greedy approximation can be recovered from the above estimates. On the other hand, the best m-term approximations for these two examples are certainly bounded from above by $\|\mathbf{1}_{\Lambda'}\|_{L_\infty} \le 1$ for some $\Lambda' \subset \{2^m+1, \dots, 2^m+m\}$. This completes the proof. □

Acknowledgment

The author thanks V.N. Temlyakov and P. Wojtaszczyk for useful comments on the earlier version [8] of this article, and for making accessible their papers [7, 15].

References

[1] Cohen, A., Dahmen, W., Daubechies, I., and DeVore, R. (1999). Tree approximation and optimal encoding, IMI Research Report 99:09, University of South Carolina.

[2] DeVore, R., Petrushev, P., and Yu, X.M. (1992). Wavelet approximation in the space C, in *Progress in Approximation Theory,* Gonchar, A.A. and Saff, E.B., Eds., Springer-Verlag, New York, 261–283.

[3] DeVore, R. and Temlyakov, V.N. (1996). Some remarks on greedy algorithms, *Advances in Computational Mathematics,* **5**, 173–187.

[4] DeVore, R. and Temlyakov, V.N. (1997). Nonlinear approximation in finite-dimensional spaces, *J. Complexity,* **13**, 489–508.

[5] Guerre-Delabrière, S. (1992). *Classical Sequences in Banach Spaces,* Marcel-Dekker, New York.

[6] Kashin, B.S. and Saakyan, A.A. (1989). Orthogonal series, *Transl. Math. Monographs,* **75**, AMS, Providence.

[7] Konyagin, S.V. and Temlyakov, V.N. (1999). A remark on Greedy approximation in Banach spaces, *East J. on Approx.,* **5**, 1–19.

[8] Oswald, P. (1999). Greedy algorithms and best m-term approximation with respect to minimal systems, Tech. Rep. BL99.01597, Bell Laboratories, Lucent Technologies, September, 20 p.

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

[10] Temlyakov, V.N. (1998). The best m-term approximation and greedy algorithms, *Advances in Computational Mathematics,* **8**, 249–265.

[11] Temlyakov, V.N. (1998). Greedy algorithm and m-term trigonometric approximation, *Constructive Approximation,* **14**, 569–587.

[12] Temlyakov, V.N. (1999). Greedy algorithms and m-term approximation with regard to redundant dictionaries, *J. Approx. Theory,* **98**, 117–145.

[13] Temlyakov, V.N.(2000). Weak greedy algorithms, *Advances in Comp. Math.* **12**, 213–227.

[14] Wojtaszczyk, P. (1982). The Franklin system is an unconditional basis in H_1, *Arkiv Mat.,* **20**, 293–300.

[15] Wojtaszczyk, P. (2000). Greedy algorithms for general systems, *J. Approx. Th.,* **107**, 293–314.

Received February 17, 2000

Bell Laboratories, Lucent Technologies Room 2C403 600 Mountain Av. Murray Hill, NJ 07974
e-mail: poswald@research.bell-labs.com