

# Sparse approximation and recovery by greedy algorithms in Banach spaces\*

V. Temlyakov †

March 28, 2013

## Abstract

We study sparse approximation by greedy algorithms. We prove the Lebesgue-type inequalities for the Weak Chebyshev Greedy Algorithm (WCGA), a generalization of the Weak Orthogonal Matching Pursuit to the case of a Banach space. The main novelty of these results is a Banach space setting instead of a Hilbert space setting. The results are proved for redundant dictionaries satisfying certain conditions. Then we apply these general results to the case of bases. In particular, we prove that the WCGA provides almost optimal sparse approximation for the trigonometric system in  $L_p$ ,  $2 \leq p < \infty$ .

## 1 Introduction

This paper is devoted to theoretical aspects of sparse approximation. The main motivation for the study of sparse approximation is that many real world signals can be well approximated by sparse ones. Sparse approximation automatically implies a need for nonlinear approximation, in particular, for greedy approximation. We give a brief description of a sparse approximation problem. In a general setting we are working in a Banach space  $X$  with a redundant system of elements  $\mathcal{D}$  (dictionary  $\mathcal{D}$ ). There is a solid justification

---

\**Math Subject Classifications. primary: 41A65; secondary: 41A25, 41A46, 46B20.*

†University of South Carolina and Steklov Institute of Mathematics. Research was supported by NSF grant DMS-1160841

of importance of a Banach space setting in numerical analysis in general and in sparse approximation in particular (see, for instance, [11], Preface, and [7]). An element (function, signal)  $f \in X$  is said to be  $K$ -sparse with respect to  $\mathcal{D}$  if it has a representation  $f = \sum_{i=1}^K x_i g_i$ ,  $g_i \in \mathcal{D}$ ,  $i = 1, \dots, K$ . The set of all  $K$ -sparse elements is denoted by  $\Sigma_K(\mathcal{D})$ . For a given element  $f_0$  we introduce the error of best  $m$ -term approximation

$$\sigma_m(f_0, \mathcal{D}) := \inf_{f \in \Sigma_m(\mathcal{D})} \|f_0 - f\|.$$

We are interested in the following fundamental problem of sparse approximation.

**Problem.** How to design a practical algorithm that builds sparse approximations comparable to best  $m$ -term approximations?

We demonstrate in this paper that the Weak Chebyshev Greedy Algorithm (WCGA) which we define momentarily is a solution to the above problem. This paper is devoted to the Banach space setting. Let  $X$  be a real Banach space with norm  $\|\cdot\| := \|\cdot\|_X$ . We say that a set of elements (functions)  $\mathcal{D}$  from  $X$  is a dictionary if each  $g \in \mathcal{D}$  has norm one ( $\|g\| = 1$ ), and the closure of  $\text{span } \mathcal{D}$  is  $X$ . For a nonzero element  $g \in X$  we let  $F_g$  denote a norming (peak) functional for  $g$ :

$$\|F_g\|_{X^*} = 1, \quad F_g(g) = \|g\|_X.$$

The existence of such a functional is guaranteed by the Hahn-Banach theorem.

Let  $\tau := \{t_k\}_{k=1}^\infty$  be a given weakness sequence of nonnegative numbers  $t_k \leq 1$ ,  $k = 1, \dots$ . We define the Weak Chebyshev Greedy Algorithm (WCGA) (see [9]) as a generalization for Banach spaces of the Weak Orthogonal Matching Pursuit (WOMP). In a Hilbert space the WCGA coincides with the WOMP. The WOMP is very popular in signal processing, in particular, in compressed sensing. We study in detail the WCGA in this paper.

**Weak Chebyshev Greedy Algorithm (WCGA).** Let  $f_0$  be given. Then for each  $m \geq 1$  we have the following inductive definition.

- (1)  $\varphi_m := \varphi_m^{c,\tau} \in \mathcal{D}$  is any element satisfying

$$|F_{f_{m-1}}(\varphi_m)| \geq t_m \sup_{g \in \mathcal{D}} |F_{f_{m-1}}(g)|.$$

- (2) Define

$$\Phi_m := \Phi_m^\tau := \text{span}\{\varphi_j\}_{j=1}^m,$$

and define  $G_m := G_m^{c,\tau}$  to be the best approximant to  $f_0$  from  $\Phi_m$ .

(3) Let

$$f_m := f_m^{c,\tau} := f_0 - G_m.$$

In this paper we only consider the case when  $t_k = t \in (0, 1]$ ,  $k = 1, 2, \dots$ .

The trigonometric system is a classical system that is known to be difficult to study. In this paper we study among other problems the problem of nonlinear sparse approximation with respect to it. Let  $\mathcal{RT}$  denote the real trigonometric system  $1, \sin 2\pi x, \cos 2\pi x, \dots$  on  $[0, 1]$  and let  $\mathcal{RT}_p$  to be its version normalized in  $L_p([0, 1])$ . Denote  $\mathcal{RT}_p^d := \mathcal{RT}_p \times \dots \times \mathcal{RT}_p$  the  $d$ -variate trigonometric system. We need to consider the real trigonometric system because the algorithm WCGA is well studied for the real Banach space. In order to illustrate performance of the WCGA we discuss in this section the above mentioned problem for the trigonometric system. There is a natural algorithm, the Thresholding Greedy Algorithm (TGA), that can be considered for the above problem. We give a definition of the TGA for a general basis  $\Psi$ . Let a Banach space  $X$ , with a normalized basis  $\Psi = \{\psi_k\}_{k=1}^\infty$ , be given. We consider the following greedy algorithm. For a given element  $f \in X$  we consider the expansion

$$f = \sum_{k=1}^{\infty} c_k(f) \psi_k. \quad (1.1)$$

For an element  $f \in X$  we say that a permutation  $\rho$  of the positive integers is decreasing if

$$|c_{k_1}(f)| \geq |c_{k_2}(f)| \geq \dots, \quad (1.2)$$

where  $\rho(j) = k_j$ ,  $j = 1, 2, \dots$ , and write  $\rho \in D(f)$ . If the inequalities are strict in (1.2), then  $D(f)$  consists of only one permutation. We define the  $m$ th greedy approximant of  $f$ , with regard to the basis  $\Psi$  corresponding to a permutation  $\rho \in D(f)$ , by the formula

$$G_m(f, \Psi) := G_m(f, \Psi, \rho) := \sum_{j=1}^m c_{k_j}(f) \psi_{k_j}.$$

The following Lebesgue-type inequality was proved in [8].

**Theorem 1.1.** *For each  $f \in L_p([0, 1]^d)$  we have*

$$\|f - G_m(f, \mathcal{RT}_p^d)\|_p \leq C(d) m^{h(p)} \sigma_m(f, \mathcal{RT}_p^d)_p, \quad 1 \leq p \leq \infty,$$

where  $h(p) := |1/2 - 1/p|$ .

It was also proved in [8] that the above inequality is sharp.

**Remark 1.1.** *There is a positive absolute constant  $C$  such that for each  $m$  and  $1 \leq p \leq \infty$  there exists a function  $f \neq 0$  with the property*

$$\|G_m(f, \mathcal{RT}_p)\|_p \geq Cm^{h(p)}\|f\|_p. \quad (1.3)$$

Remark 1.1 shows that the TGA does not work well for the trigonometric system in  $L_p$ ,  $p \neq 2$ . This leads to a natural attempt to consider some other algorithms that may have some advantages over the TGA in the case of the trigonometric system. In this paper we discuss the performance of the Weak Chebyshev Greedy Algorithm (WCGA) with respect to the trigonometric system. We prove here the following Lebesgue-type inequality for the WCGA (see Example 2 in Section 4).

**Theorem 1.2.** *Let  $\mathcal{D}$  be the normalized in  $L_p$ ,  $2 \leq p < \infty$ , real  $d$ -variate trigonometric system. Then for any  $f_0 \in L_p$  the WCGA with weakness parameter  $t$  gives*

$$\|f_{C(t,p,d)m \ln(m+1)}\|_p \leq C\sigma_m(f_0, \mathcal{D})_p. \quad (1.4)$$

The Open Problem 7.1 (p. 91) from [10] asks if (1.4) holds without an extra  $\ln(m+1)$  factor. Theorem 1.2 is the first result on the Lebesgue-type inequalities for the WCGA with respect to the trigonometric system. It provides a progress in solving the above mentioned open problem, but the problem is still open.

Theorem 1.2 shows that the WCGA is very well designed for the trigonometric system. We show in Example 1 of Section 4 that an analog of (1.4) holds for uniformly bounded orthogonal systems. We note that it is known (see [11]) that the TGA is very well designed for bases  $L_p$ -equivalent to the Haar basis,  $1 < p < \infty$ . We discuss performance of the WCGA in more detail in Section 5.

The proof of Theorem 1.2 uses technique developed for proving the Lebesgue-type inequalities for redundant dictionaries with special properties. We present these results in Sections 2 and 3. These results are an extension of earlier results from [5]. In Section 4 we test the power of general results from Section 2 on specific dictionaries, namely, on bases. Section 4 provides a number of examples, including the trigonometric system, where the technique from Sections 2 and 3 can be successfully applied. In particular, results from

Section 4 demonstrate that the general technique from Sections 2 and 3 provides almost optimal  $m$ -term approximation results for uniformly bounded orthogonal systems (see Example 1). Example 7 shows that an extra assumption that a uniformly bounded orthogonal system  $\Psi$  is a quasi-greedy basis allows us to improve inequality (1.4):

$$\|f_{C(t,p,\Psi)m \ln \ln(m+3)}\|_p \leq C\sigma_m(f_0, \Psi)_p.$$

## 2 Lebesgue-type inequalities. General results.

A very important advantage of the WCGA is its convergence and rate of convergence properties. The WCGA is well defined for all  $m$ . Moreover, it is known (see [9] and [11]) that the WCGA with  $\tau = \{t\}$  converges for all  $f_0$  in all uniformly smooth Banach spaces with respect to any dictionary. That is, when  $X$  is a real Banach space and the modulus of smoothness of  $X$  is defined as follows

$$\rho(u) := \frac{1}{2} \sup_{x,y;\|x\|=\|y\|=1} \left| \|x+uy\| + \|x-uy\| - 2 \right|, \quad (2.1)$$

then the uniformly smooth Banach space is the one with  $\rho(u)/u \rightarrow 0$  when  $u \rightarrow 0$ .

We discuss here the Lebesgue-type inequalities for the WCGA with  $\tau = \{t\}$ ,  $t \in (0, 1]$ . For notational convenience we consider here a countable dictionary  $\mathcal{D} = \{g_i\}_{i=1}^\infty$ . The following assumptions **A1** and **A2** were used in [5]. For a given  $f_0$  let sparse element (signal)

$$f := f^\epsilon = \sum_{i \in T} x_i g_i$$

be such that  $\|f_0 - f^\epsilon\| \leq \epsilon$  and  $|T| = K$ . For  $A \subset T$  denote

$$f_A := f_A^\epsilon := \sum_{i \in A} x_i g_i.$$

**A1.** We say that  $f = \sum_{i \in T} x_i g_i$  satisfies the Nikol'skii-type  $\ell_1 X$  inequality with parameter  $r$  if

$$\sum_{i \in A} |x_i| \leq C_1 |A|^r \|f_A\|, \quad A \subset T, \quad r \geq 1/2. \quad (2.2)$$

We say that a dictionary  $\mathcal{D}$  has the Nikol'skii-type  $\ell_1 X$  property with parameters  $K, r$  if any  $K$ -sparse element satisfies the Nikol'skii-type  $\ell_1 X$  inequality with parameter  $r$ .

**A2.** We say that  $f = \sum_{i \in T} x_i g_i$  has incoherence property with parameters  $D$  and  $U$  if for any  $A \subset T$  and any  $\Lambda$  such that  $A \cap \Lambda = \emptyset$ ,  $|A| + |\Lambda| \leq D$  we have for any  $\{c_i\}$

$$\|f_A - \sum_{i \in \Lambda} c_i g_i\| \geq U^{-1} \|f_A\|. \quad (2.3)$$

We say that a dictionary  $\mathcal{D}$  is  $(K, D)$ -unconditional with a constant  $U$  if for any  $f = \sum_{i \in T} x_i g_i$  with  $|T| \leq K$  inequality (2.3) holds.

The term *unconditional* in **A2** is justified by the following remark. The above definition of  $(K, D)$ -unconditional dictionary is equivalent to the following definition. Let  $\mathcal{D}$  be such that any subsystem of  $D$  distinct elements  $e_1, \dots, e_D$  from  $\mathcal{D}$  is linearly independent and for any  $A \subset [1, D]$  with  $|A| \leq K$  and any coefficients  $\{c_i\}$  we have

$$\left\| \sum_{i \in A} c_i e_i \right\| \leq U \left\| \sum_{i=1}^D c_i e_i \right\|.$$

It is convenient for us to use the following assumption **A3** which is a corollary of assumptions **A1** and **A2**.

**A3.** We say that  $f = \sum_{i \in T} x_i g_i$  has  $\ell_1$  incoherence property with parameters  $D, V$ , and  $r$  if for any  $A \subset T$  and any  $\Lambda$  such that  $A \cap \Lambda = \emptyset$ ,  $|A| + |\Lambda| \leq D$  we have for any  $\{c_i\}$

$$\sum_{i \in A} |x_i| \leq V |A|^r \|f_A - \sum_{i \in \Lambda} c_i g_i\|. \quad (2.4)$$

A dictionary  $\mathcal{D}$  has  $\ell_1$  incoherence property with parameters  $K, D, V$ , and  $r$  if for any  $A \subset B$ ,  $|A| \leq K$ ,  $|B| \leq D$  we have for any  $\{c_i\}_{i \in B}$

$$\sum_{i \in A} |c_i| \leq V |A|^r \left\| \sum_{i \in B} c_i g_i \right\|.$$

It is clear that **A1** and **A2** imply **A3** with  $V = C_1 U$ . Also, **A3** implies **A1** with  $C_1 = V$  and **A2** with  $U = V K^r$ . Obviously, we can restrict ourselves to  $r \leq 1$ .

We now proceed to main results of this paper on the WCGA with respect to redundant dictionaries. The following Theorem 2.1 in the case  $q = 2$  was proved in [5].

**Theorem 2.1.** *Let  $X$  be a Banach space with  $\rho(u) \leq \gamma u^q$ ,  $1 < q \leq 2$ . Suppose  $K$ -sparse  $f^\epsilon$  satisfies **A1**, **A2** and  $\|f_0 - f^\epsilon\| \leq \epsilon$ . Then the WCGA with weakness parameter  $t$  applied to  $f_0$  provides*

$$\|f_{C(t,\gamma,C_1)U^{q'} \ln(U+1)K^{r_{q'}}}\| \leq C\epsilon \quad \text{for} \quad K + C(t,\gamma,C_1)U^{q'} \ln(U+1)K^{r_{q'}} \leq D$$

with an absolute constant  $C$ .

It was pointed out in [5] that Theorem 2.1 provides a corollary for Hilbert spaces that gives sufficient conditions somewhat weaker than the known RIP conditions on  $\mathcal{D}$  for the Lebesgue-type inequality to hold. We formulate the corresponding definitions and results. Let  $\mathcal{D}$  be the Riesz dictionary with depth  $D$  and parameter  $\delta \in (0, 1)$ . This class of dictionaries is a generalization of the class of classical Riesz bases. We give a definition in a general Hilbert space (see [11], p. 306).

**Definition 2.1.** *A dictionary  $\mathcal{D}$  is called the Riesz dictionary with depth  $D$  and parameter  $\delta \in (0, 1)$  if, for any  $D$  distinct elements  $e_1, \dots, e_D$  of the dictionary and any coefficients  $a = (a_1, \dots, a_D)$ , we have*

$$(1 - \delta)\|a\|_2^2 \leq \left\| \sum_{i=1}^D a_i e_i \right\|^2 \leq (1 + \delta)\|a\|_2^2. \quad (2.5)$$

We denote the class of Riesz dictionaries with depth  $D$  and parameter  $\delta \in (0, 1)$  by  $R(D, \delta)$ .

The term Riesz dictionary with depth  $D$  and parameter  $\delta \in (0, 1)$  is another name for a dictionary satisfying the Restricted Isometry Property (RIP) with parameters  $D$  and  $\delta$ . The following simple lemma holds.

**Lemma 2.1.** *Let  $\mathcal{D} \in R(D, \delta)$  and let  $e_j \in \mathcal{D}$ ,  $j = 1, \dots, s$ . For  $f = \sum_{i=1}^s a_i e_i$  and  $A \subset \{1, \dots, s\}$  denote*

$$S_A(f) := \sum_{i \in A} a_i e_i.$$

If  $s \leq D$  then

$$\|S_A(f)\|^2 \leq (1 + \delta)(1 - \delta)^{-1}\|f\|^2.$$

Lemma 2.1 implies that if  $\mathcal{D} \in R(D, \delta)$  then it is  $(D, D)$ -unconditional with a constant  $U = (1 + \delta)^{1/2}(1 - \delta)^{-1/2}$ .

**Theorem 2.2.** *Let  $X$  be a Hilbert space. Suppose  $K$ -sparse  $f^\epsilon$  satisfies **A2** and  $\|f_0 - f^\epsilon\| \leq \epsilon$ . Then the WOMP with weakness parameter  $t$  applied to  $f_0$  provides*

$$\|f_{C(t,U)K}\| \leq C\epsilon \quad \text{for } K + C(t,U)K \leq D$$

*with an absolute constant  $C$ .*

Theorem 2.2 implies the following corollaries.

**Corollary 2.1.** *Let  $X$  be a Hilbert space. Suppose any  $K$ -sparse  $f$  satisfies **A2**. Then the WOMP with weakness parameter  $t$  applied to  $f_0$  provides*

$$\|f_{C(t,U)K}\| \leq C\sigma_K(f_0, \mathcal{D}) \quad \text{for } K + C(t,U)K \leq D$$

*with an absolute constant  $C$ .*

**Corollary 2.2.** *Let  $X$  be a Hilbert space. Suppose  $\mathcal{D} \in R(D, \delta)$ . Then the WOMP with weakness parameter  $t$  applied to  $f_0$  provides*

$$\|f_{C(t,\delta)K}\| \leq C\sigma_K(f_0, \mathcal{D}) \quad \text{for } K + C(t,\delta)K \leq D$$

*with an absolute constant  $C$ .*

We emphasized in [5] that in Theorem 2.1 we impose our conditions on an individual function  $f^\epsilon$ . It may happen that the dictionary does not have the Nikol'skii  $\ell_1 X$  property and  $(K, D)$ -unconditionality but the given  $f_0$  can be approximated by  $f^\epsilon$  which does satisfy assumptions **A1** and **A2**. Even in the case of a Hilbert space the above results from [5] add something new to the study based on the RIP property of a dictionary. First of all, Theorem 2.2 shows that it is sufficient to impose assumption **A2** on  $f^\epsilon$  in order to obtain exact recovery and the Lebesgue-type inequality results. Second, Corollary 2.1 shows that the condition **A2**, which is weaker than the RIP condition, is sufficient for exact recovery and the Lebesgue-type inequality results. Third, Corollary 2.2 shows that even if we impose our assumptions in terms of RIP we do not need to assume that  $\delta < \delta_0$ . In fact, the result works for all  $\delta < 1$  with parameters depending on  $\delta$ .

Theorem 2.1 follows from the combination of Theorems 2.3 and 2.4. In case  $q = 2$  these theorems were proved in [5].



**Theorem 2.3.** *Let  $X$  be a Banach space with  $\rho(u) \leq \gamma u^q$ ,  $1 < q \leq 2$ . Suppose for a given  $f_0$  we have  $\|f_0 - f^\epsilon\| \leq \epsilon$  with  $K$ -sparse  $f := f^\epsilon$  satisfying **A3**. Then for any  $k \geq 0$  we have for  $K + m \leq D$*

$$\|f_m\| \leq \|f_k\| \exp\left(-\frac{c_1(m-k)}{K^{rq'}}\right) + 2\epsilon, \quad q' := \frac{q}{q-1},$$

where  $c_1 := \frac{t^{q'}}{2(16\gamma)^{\frac{1}{q-1}} V^{q'}}$ .

In all theorems that follow we assume  $rq' \geq 1$ .

**Theorem 2.4.** *Let  $X$  be a Banach space with  $\rho(u) \leq \gamma u^q$ ,  $1 < q \leq 2$ . Suppose  $K$ -sparse  $f^\epsilon$  satisfies **A1**, **A2** and  $\|f_0 - f^\epsilon\| \leq \epsilon$ . Then the WCGA with weakness parameter  $t$  applied to  $f_0$  provides*

$$\|f_{C'U^{q'} \ln(U+1)K^{rq'}}\| \leq CU\epsilon \quad \text{for} \quad K + C'U^{q'} \ln(U+1)K^{rq'} \leq D$$

with an absolute constant  $C$  and  $C' = C_2(q)\gamma^{\frac{1}{q-1}}C_1^{q'}t^{-q'}$ .

We formulate an immediate corollary of Theorem 2.4 with  $\epsilon = 0$ .

**Corollary 2.3.** *Let  $X$  be a Banach space with  $\rho(u) \leq \gamma u^q$ . Suppose  $K$ -sparse  $f$  satisfies **A1**, **A2**. Then the WCGA with weakness parameter  $t$  applied to  $f$  recovers it exactly after  $C'U^{q'} \ln(U+1)K^{rq'}$  iterations under condition  $K + C'U^{q'} \ln(U+1)K^{rq'} \leq D$ .*

We formulate versions of Theorem 2.4 with assumptions **A1**, **A2** replaced by a single assumption **A3** and replaced by two assumptions **A2** and **A3**. The corresponding modifications in the proofs go as in the proof of Theorem 2.3.

**Theorem 2.5.** *Let  $X$  be a Banach space with  $\rho(u) \leq \gamma u^q$ ,  $1 < q \leq 2$ . Suppose  $K$ -sparse  $f^\epsilon$  satisfies **A3** and  $\|f_0 - f^\epsilon\| \leq \epsilon$ . Then the WCGA with weakness parameter  $t$  applied to  $f_0$  provides*

$$\|f_{C(t,\gamma,q)V^{q'} \ln(VK)K^{rq'}}\| \leq CVK^r\epsilon \quad \text{for} \quad K + C(t,\gamma,q)V^{q'} \ln(VK)K^{rq'} \leq D$$

with an absolute constant  $C$  and  $C(t,\gamma,q) = C_2(q)\gamma^{\frac{1}{q-1}}t^{-q'}$ .

**Theorem 2.6.** *Let  $X$  be a Banach space with  $\rho(u) \leq \gamma u^q$ ,  $1 < q \leq 2$ . Suppose  $K$ -sparse  $f^\epsilon$  satisfies **A2**, **A3** and  $\|f_0 - f^\epsilon\| \leq \epsilon$ . Then the WCGA with weakness parameter  $t$  applied to  $f_0$  provides*

$$\|f_{C(t,\gamma,q)V^{q'} \ln(U+1)K^{r_{q'}}}\| \leq CU\epsilon \quad \text{for} \quad K + C(t,\gamma,q)V^{q'} \ln(U+1)K^{r_{q'}} \leq D$$

with an absolute constant  $C$  and  $C(t,\gamma,q) = C_2(q)\gamma^{\frac{1}{q-1}}t^{-q'}$ .

Theorems 2.5 and 2.3 imply the following analog of Theorem 2.1.

**Theorem 2.7.** *Let  $X$  be a Banach space with  $\rho(u) \leq \gamma u^q$ ,  $1 < q \leq 2$ . Suppose  $K$ -sparse  $f^\epsilon$  satisfies **A3** and  $\|f_0 - f^\epsilon\| \leq \epsilon$ . Then the WCGA with weakness parameter  $t$  applied to  $f_0$  provides*

$$\|f_{C(t,\gamma,q)V^{q'} \ln(VK)K^{r_{q'}}}\| \leq C\epsilon \quad \text{for} \quad K + C(t,\gamma,q)V^{q'} \ln(VK)K^{r_{q'}} \leq D$$

with an absolute constant  $C$  and  $C(t,\gamma,q) = C_2(q)\gamma^{\frac{1}{q-1}}t^{-q'}$ .

The following edition of Theorems 2.1 and 2.7 is also useful in applications. It follows from Theorems 2.6 and 2.3.

**Theorem 2.8.** *Let  $X$  be a Banach space with  $\rho(u) \leq \gamma u^q$ ,  $1 < q \leq 2$ . Suppose  $K$ -sparse  $f^\epsilon$  satisfies **A2**, **A3** and  $\|f_0 - f^\epsilon\| \leq \epsilon$ . Then the WCGA with weakness parameter  $t$  applied to  $f_0$  provides*

$$\|f_{C(t,\gamma,q)V^{q'} \ln(U+1)K^{r_{q'}}}\| \leq C\epsilon \quad \text{for} \quad K + C(t,\gamma,q)V^{q'} \ln(U+1)K^{r_{q'}} \leq D$$

with an absolute constant  $C$  and  $C(t,\gamma,q) = C_2(q)\gamma^{\frac{1}{q-1}}t^{-q'}$ .

### 3 Proofs

We begin with a proof of Theorem 2.3.

*Proof.* Let

$$f := f^\epsilon = \sum_{i \in T} x_i g_i, \quad |T| = K, \quad g_i \in \mathcal{D}.$$

Denote by  $T^m$  the set of indices of  $g_j \in \mathcal{D}$  picked by the WCGA after  $m$  iterations,  $\Gamma^m := T \setminus T^m$ . Denote by  $A_1(\mathcal{D})$  the closure in  $X$  of the convex

hull of the symmetrized dictionary  $\mathcal{D}^\pm := \{\pm g, g \in D\}$ . We will bound  $\|f_m\|$  from above. Assume  $\|f_{m-1}\| \geq \epsilon$ . Let  $m > k$ . We bound from below

$$S_m := \sup_{\phi \in A_1(\mathcal{D})} |F_{f_{m-1}}(\phi)|.$$

Denote  $A_m := \Gamma^{m-1}$ . Then

$$S_m \geq F_{f_{m-1}}(f_{A_m}/\|f_{A_m}\|_1),$$

where  $\|f_A\|_1 := \sum_{i \in A} |x_i|$ . Next, by Lemma 6.9, p. 342, from [11] we obtain

$$F_{f_{m-1}}(f_{A_m}) = F_{f_{m-1}}(f^\epsilon) \geq \|f_{m-1}\| - \epsilon.$$

Thus

$$S_m \geq \|f_{A_m}\|_1^{-1} (\|f_{m-1}\| - \epsilon). \quad (3.1)$$

From the definition of the modulus of smoothness we have for any  $\lambda$

$$\|f_{m-1} - \lambda\varphi_m\| + \|f_{m-1} + \lambda\varphi_m\| \leq 2\|f_{m-1}\| \left(1 + \rho\left(\frac{\lambda}{\|f_{m-1}\|}\right)\right) \quad (3.2)$$

and by (1) from the definition of the WCGA and Lemma 6.10 from [11], p. 343, we get

$$\begin{aligned} |F_{f_{m-1}}(\varphi_m)| &\geq t \sup_{g \in \mathcal{D}} |F_{f_{m-1}}(g)| = \\ &t \sup_{\phi \in A_1(\mathcal{D})} |F_{f_{m-1}}(\phi)| = tS_m. \end{aligned}$$

Then either  $F_{f_{m-1}}(\varphi_m) \geq tS_m$  or  $F_{f_{m-1}}(-\varphi_m) \geq tS_m$ . Both cases are treated in the same way. We demonstrate the case  $F_{f_{m-1}}(\varphi_m) \geq tS_m$ . We have for  $\lambda \geq 0$

$$\|f_{m-1} + \lambda\varphi_m\| \geq F_{f_{m-1}}(f_{m-1} + \lambda\varphi_m) \geq \|f_{m-1}\| + \lambda tS_m.$$

From here and from (3.2) we obtain

$$\|f_m\| \leq \|f_{m-1} - \lambda\varphi_m\| \leq \|f_{m-1}\| + \inf_{\lambda \geq 0} (-\lambda tS_m + 2\|f_{m-1}\| \rho(\lambda/\|f_{m-1}\|)).$$

We discuss here the case  $\rho(u) \leq \gamma u^q$ . Using (3.1) we get

$$\|f_m\| \leq \|f_{m-1}\| \left(1 - \frac{\lambda t}{\|f_{A_m}\|_1} + 2\gamma \frac{\lambda^q}{\|f_{m-1}\|^q}\right) + \frac{\epsilon \lambda t}{\|f_{A_m}\|_1}.$$

Let  $\lambda_1$  be a solution of

$$\frac{\lambda t}{2\|f_{A_m}\|_1} = 2\gamma \frac{\lambda^q}{\|f_{m-1}\|^q}, \quad \lambda_1 = \left( \frac{t\|f_{m-1}\|^q}{4\gamma\|f_{A_m}\|_1} \right)^{\frac{1}{q-1}}.$$

Our assumption (2.4) gives

$$\begin{aligned} \|f_{A_m}\|_1 &= \|(f^\epsilon - G_{m-1})_{A_m}\|_1 \leq VK^r \|f^\epsilon - G_{m-1}\| \\ &\leq VK^r (\|f_0 - G_{m-1}\| + \|f_0 - f^\epsilon\|) \leq VK^r (\|f_{m-1}\| + \epsilon). \end{aligned}$$

Specify

$$\lambda = \left( \frac{t\|f_{A_m}\|_1^{q-1}}{16\gamma(VK^r)^q} \right)^{\frac{1}{q-1}}.$$

Then, using  $\|f_{m-1}\| \geq \epsilon$  we get

$$\left( \frac{\lambda}{\lambda_1} \right)^{q-1} = \frac{\|f_{A_m}\|_1^q}{4\|f_{m-1}\|^q (VK^r)^q} \leq 1$$

and obtain

$$\|f_m\| \leq \|f_{m-1}\| \left( 1 - \frac{t^{q'}}{2(16\gamma)^{\frac{1}{q-1}} (VK^r)^{q'}} \right) + \frac{\epsilon t^{q'}}{(16\gamma)^{\frac{1}{q-1}} (VK^r)^{q'}}.$$

Denote  $c_1 := \frac{t^{q'}}{2(16\gamma)^{\frac{1}{q-1}} V^{q'}}$ . Then

$$\|f_m\| \leq \|f_k\| \exp\left(-\frac{c_1(m-k)}{K^{rq'}}\right) + 2\epsilon.$$

□

We proceed to a proof of Theorem 2.4. Modifications of this proof which are in a style of the above proof of Theorem 2.3 give Theorems 2.5 and 2.6.

*Proof.* We use the above notations  $T^m$  and  $\Gamma^m := T \setminus T^m$ . Let  $k \geq 0$  be fixed. Suppose

$$2^{n-1} < |\Gamma^k| \leq 2^n.$$

For  $j = 1, 2, \dots, n, n+1$  consider the following pairs of sets  $A_j, B_j$ :  $A_{n+1} = \Gamma^k$ ,  $B_{n+1} = \emptyset$ ; for  $j \leq n$ ,  $A_j := \Gamma^k \setminus B_j$  with  $B_j \subset \Gamma^k$  is such that  $|B_j| \geq |\Gamma^k| - 2^{j-1}$  and for any set  $J \subset \Gamma^k$  with  $|J| \geq |\Gamma^k| - 2^{j-1}$  we have

$$\|f_{B_j}\| \leq \|f_J\|.$$

We note that this implies that if for some  $Q \subset \Gamma^k$  we have

$$\|f_Q\| < \|f_{B_j}\| \quad \text{then} \quad |Q| < |\Gamma^k| - 2^{j-1}. \quad (3.3)$$

For a given  $b > 1$ , to be specified later, denote by  $L$  the index such that  $(B_0 := \Gamma^k)$

$$\begin{aligned} \|f_{B_0}\| &< b\|f_{B_1}\|, \\ \|f_{B_1}\| &< b\|f_{B_2}\|, \\ &\dots \\ \|f_{B_{L-2}}\| &< b\|f_{B_{L-1}}\|, \\ \|f_{B_{L-1}}\| &\geq b\|f_{B_L}\|. \end{aligned}$$

Then

$$\|f_{B_j}\| \leq b^{L-1-j} \|f_{B_{L-1}}\|, \quad j = 1, 2, \dots, L. \quad (3.4)$$

We now proceed to a general step. Let  $m > k$  and let  $A, B \subset \Gamma^k$  be such that  $A = \Gamma^k \setminus B$ . As above we bound  $S_m$  from below. It is clear that  $S_m \geq 0$ . Denote  $A_m := A \cap \Gamma^{m-1}$ . Then

$$S_m \geq F_{f_{m-1}}(f_{A_m} / \|f_{A_m}\|_1).$$

Next,

$$F_{f_{m-1}}(f_{A_m}) = F_{f_{m-1}}(f_{A_m} + f_B - f_B).$$

Then  $f_{A_m} + f_B = f^\epsilon - f_\Lambda$  with  $F_{f_{m-1}}(f_\Lambda) = 0$ . Moreover, it is easy to see that  $F_{f_{m-1}}(f^\epsilon) \geq \|f_{m-1}\| - \epsilon$ . Therefore,

$$F_{f_{m-1}}(f_{A_m} + f_B - f_B) \geq \|f_{m-1}\| - \epsilon - \|f_B\|.$$

Thus

$$S_m \geq \|f_{A_m}\|_1^{-1} \max(0, \|f_{m-1}\| - \epsilon - \|f_B\|).$$

By (2.2) we get

$$\|f_{A_m}\|_1 \leq C_1 |A_m|^r \|f_{A_m}\| \leq C_1 |A|^r \|f_{A_m}\|.$$

Then

$$S_m \geq \frac{\|f_{m-1}\| - \|f_B\| - \epsilon}{C_1 |A|^r \|f_{A_m}\|}. \quad (3.5)$$

From the definition of the modulus of smoothness we have for any  $\lambda$

$$\|f_{m-1} - \lambda\varphi_m\| + \|f_{m-1} + \lambda\varphi_m\| \leq 2\|f_{m-1}\|(1 + \rho(\frac{\lambda}{\|f_{m-1}\|}))$$

and by (1) from the definition of the WCGA and Lemma 6.10 from [11], p. 343, we get

$$\begin{aligned} |F_{f_{m-1}}(\varphi_m)| &\geq t \sup_{g \in \mathcal{D}} |F_{f_{m-1}}(g)| = \\ &t \sup_{\phi \in A_1(\mathcal{D})} |F_{f_{m-1}}(\phi)|. \end{aligned}$$

From here we obtain

$$\|f_m\| \leq \|f_{m-1}\| + \inf_{\lambda \geq 0} (-\lambda t S_m + 2\|f_{m-1}\|\rho(\lambda/\|f_{m-1}\|)).$$

We discuss here the case  $\rho(u) \leq \gamma u^q$ . Using (3.5) we get

$$\|f_m\| \leq \|f_{m-1}\| \left( 1 - \frac{\lambda t}{C_1 |A|^r \|f_{A_m}\|} + 2\gamma \frac{\lambda^q}{\|f_{m-1}\|^q} \right) + \frac{\lambda t (\|f_B\| + \epsilon)}{C_1 |A|^r \|f_{A_m}\|}.$$

Let  $\lambda_1$  be a solution of

$$\frac{\lambda t}{2C_1 |A|^r \|f_{A_m}\|} = 2\gamma \frac{\lambda^q}{\|f_{m-1}\|^q}, \quad \lambda_1 = \left( \frac{t \|f_{m-1}\|^q}{4\gamma C_1 |A|^r \|f_{A_m}\|} \right)^{\frac{1}{q-1}}.$$

Our assumption (2.3) gives

$$\|f_{A_m}\| \leq U(\|f_{m-1}\| + \epsilon).$$

Specify

$$\lambda = \left( \frac{t \|f_{A_m}\|^{q-1}}{16\gamma C_1 |A|^r U^q} \right)^{\frac{1}{q-1}}.$$

Then  $\lambda \leq \lambda_1$  and we obtain

$$\|f_m\| \leq \|f_{m-1}\| \left( 1 - \frac{t^{q'}}{2(16\gamma)^{\frac{1}{q-1}} (C_1 U |A|^r)^{q'}} \right) + \frac{t^{q'} (\|f_B\| + \epsilon)}{(16\gamma)^{\frac{1}{q-1}} (C_1 |A|^r U)^{q'}}. \quad (3.6)$$

Denote  $c_1 := \frac{t^{q'}}{2(16\gamma)^{\frac{1}{q-1}} (C_1 U)^{q'}}$ . This implies for  $m_2 > m_1 \geq k$

$$\|f_{m_2}\| \leq \|f_{m_1}\| (1 - c_1/|A|^{rq'})^{m_2 - m_1} + \frac{2c_1(m_2 - m_1)}{|A|^{2r}} (\|f_B\| + \epsilon). \quad (3.7)$$

Define  $m_0 := k$  and, inductively,

$$m_j = m_{j-1} + \beta |A_j|^{r_{q'}}, \quad j = 1, \dots, n.$$

At iterations from  $m_{j-1} + 1$  to  $m_j$  we use  $A = A_j$  and obtain from (3.6)

$$\|f_{m_j}\| \leq \|f_{m_{j-1}}\| e^{-c_1 \beta} + 2(\|f_{B_j}\| + \epsilon).$$

We continue it up to  $j = L$ . Denote  $\eta := e^{-c_1 \beta}$ . Then

$$\|f_{m_L}\| \leq \|f_k\| \eta^L + 2 \sum_{j=1}^L (\|f_{B_j}\| + \epsilon) \eta^{L-j}.$$

We bound the  $\|f_k\|$ . It follows from the definition of  $f_k$  that  $\|f_k\|$  is the error of best approximation of  $f_0$  by the subspace  $\Phi_k$ . Representing  $f_0 = f + f_0 - f$  we see that  $\|f_k\|$  is not greater than the error of best approximation of  $f$  by the subspace  $\Phi_k$  plus  $\|f_0 - f\|$ . This implies  $\|f_k\| \leq \|f_{B_0}\| + \epsilon$ . Therefore we continue

$$\begin{aligned} &\leq (\|f_{B_0}\| + \epsilon) \eta^L + 2 \sum_{j=1}^L (\|f_{B_{L-1}}\| (\eta b)^{L-j} b^{-1} + \epsilon \eta^{L-j}) \\ &\leq b^{-1} \|f_{B_{L-1}}\| \left( (\eta b)^L + 2 \sum_{j=1}^L (\eta b)^{L-j} \right) + \frac{2\epsilon}{1-\eta}. \end{aligned}$$

We will specify  $\beta$  later. However, we note that it will be chosen in such a way that guarantees  $\eta < 1/2$ . Choose  $b = \frac{1}{2\eta}$ . Then

$$\|f_{m_L}\| \leq \|f_{B_{L-1}}\| 8e^{-c_1 \beta} + 4\epsilon. \quad (3.8)$$

By (2.3) we get

$$\|f_{\Gamma^{m_L}}\| \leq U(\|f_{m_L}\| + \epsilon) \leq U(\|f_{B_{L-1}}\| 8e^{-c_1 \beta} + 5\epsilon).$$

We note that in the proof of Theorem 2.5 we use the above inequality with  $U = VK^r \leq VK$ . If  $\|f_{B_{L-1}}\| \leq 10U\epsilon$  then by (3.8)

$$\|f_{m_L}\| \leq CU\epsilon.$$

If  $\|f_{B_{L-1}}\| \geq 10U\epsilon$  then making  $\beta$  sufficiently large to satisfy  $16Ue^{-c_1\beta} < 1$  so that  $\beta = \frac{C_3 \ln(U+1)}{c_1}$ , we get

$$U(\|f_{B_{L-1}}\|8e^{-c_1\beta} + 5\epsilon) < \|f_{B_{L-1}}\|$$

and therefore

$$\|f_{\Gamma^{m_L}}\| < \|f_{B_{L-1}}\|.$$

This implies

$$|\Gamma^{m_L}| < |\Gamma^k| - 2^{L-2}.$$

We begin with  $f_0$  and apply the above argument (with  $k = 0$ ). As a result we either get the required inequality or we reduce the cardinality of support of  $f$  from  $|T| = K$  to  $|\Gamma^{m_{L_1}}| < |T| - 2^{L_1-2}$ ,  $m_{L_1} \leq \beta 2^{aL_1}$ ,  $a := rq'$ . We continue the process and build a sequence  $m_{L_j}$  such that  $m_{L_j} \leq \beta 2^{aL_j}$  and after  $m_{L_j}$  iterations we reduce the support by at least  $2^{L_j-2}$ . We also note that  $m_{L_j} \leq \beta 2^a K^a$ . We continue this process till the following inequality is satisfied for the first time

$$m_{L_1} + \dots + m_{L_n} \geq 2^{2a} \beta K^a. \quad (3.9)$$

Then, clearly,

$$m_{L_1} + \dots + m_{L_n} \leq 2^{2a+1} \beta K^a.$$

Using the inequality

$$(a_1 + \dots + a_n)^\theta \leq a_1^\theta + \dots + a_n^\theta, \quad a_j \geq 0, \quad \theta \in (0, 1]$$

we derive from (3.9)

$$\begin{aligned} 2^{L_1-2} + \dots + 2^{L_n-2} &\geq (2^{a(L_1-2)} + \dots + 2^{a(L_n-2)})^{\frac{1}{a}} \\ &\geq 2^{-2} (2^{aL_1} + \dots + 2^{aL_n})^{\frac{1}{a}} \\ &\geq 2^{-2} ((\beta)^{-1}(m_{L_1} + \dots + m_{L_n}))^{\frac{1}{a}} \geq K. \end{aligned}$$

Thus, after not more than  $N := 2^{2a+1} \beta K^a$  iterations we recover  $f$  exactly and then  $\|f_N\| \leq \|f_0 - f\| \leq \epsilon$ .

□



## 4 Examples

In this section we discuss applications of Theorems from Section 2 for specific dictionaries  $\mathcal{D}$ . Mostly,  $\mathcal{D}$  will be a basis  $\Psi$  for  $X$ . Because of that we use  $m$  instead of  $K$  in the notation of sparse approximation. In some of our examples we take  $X = L_p$ ,  $2 \leq p < \infty$ . Then it is known that  $\rho(u) \leq \gamma u^2$  with  $\gamma = (p-1)/2$ . In some other examples we take  $X = L_p$ ,  $1 < p \leq 2$ . Then it is known that  $\rho(u) \leq \gamma u^p$ , with  $\gamma = 1/p$ .

**Example 1.** Let  $\Psi$  be a uniformly bounded orthogonal system normalized in  $L_p(\Omega)$ ,  $2 \leq p < \infty$ ,  $\Omega$  is a bounded domain. Then we have

$$C_1(\Omega, p) \|\psi_j\|_2 \leq \|\psi_j\|_p \leq C_2(\Omega, p) \|\psi_j\|_2, \quad j = 1, 2, \dots$$

Next, for  $f = \sum_i c_i(f) \psi_i$

$$\begin{aligned} \sum_{i \in A} |c_i(f)| &= \left\langle f, \sum_{i \in A} (\text{sign } c_i(f)) \psi_i \|\psi_i\|_2^{-2} \right\rangle \\ &\leq \|f\|_2 \left\| \sum_{i \in A} (\text{sign } c_i(f)) \psi_i \|\psi_i\|_2^{-2} \right\|_2 \leq C_3(\Omega, p) |A|^{1/2} \|f\|_p. \end{aligned}$$

Therefore  $\Psi$  satisfies **A3** with  $D = \infty$ ,  $V = C_3(\Omega, p)$ ,  $r = 1/2$ . Theorem 2.7 gives

$$\|f_{C(t,p,D)m \ln(m+1)}\|_p \leq C \sigma_m(f_0, \Psi)_p. \quad (4.1)$$

**Example 1q.** Let  $\Psi$  be a uniformly bounded orthogonal system normalized in  $L_p(\Omega)$ ,  $1 < p \leq 2$ ,  $\Omega$  is a bounded domain. Then we have

$$C_1(\Omega, p) \|\psi_j\|_2 \leq \|\psi_j\|_p \leq C_2(\Omega, p) \|\psi_j\|_2, \quad j = 1, 2, \dots$$

Next, for  $f = \sum_i c_i(f) \psi_i$

$$\begin{aligned} \sum_{i \in A} |c_i(f)| &= \left\langle f, \sum_{i \in A} (\text{sign } c_i(f)) \psi_i \|\psi_i\|_2^{-2} \right\rangle \\ &\leq \|f\|_p \left\| \sum_{i \in A} (\text{sign } c_i(f)) \psi_i \|\psi_i\|_2^{-2} \right\|_{p'} \leq C_4(\Omega, p) |A|^{1-1/p'} \|f\|_p. \end{aligned}$$

Therefore  $\Psi$  satisfies **A3** with  $D = \infty$ ,  $V = C_4(\Omega, p)$ ,  $r = 1 - 1/p'$ . Theorem 2.7 gives

$$\|f_{C(t,p,D)m^{p'-1} \ln(m+1)}\|_p \leq C \sigma_m(f_0, \Psi)_p. \quad (4.2)$$

**Example 2.** Let  $\Psi$  be the normalized in  $L_p$ ,  $2 \leq p < \infty$ , real  $d$ -variate trigonometric system. Then Example 1 applies and gives for any  $f_0 \in L_p$

$$\|f_{C(t,p,d)m \ln(m+1)}\|_p \leq C\sigma_m(f_0, \Psi)_p. \quad (4.3)$$

We note that (4.3) provides some progress in Open Problem 7.1 (p. 91) from [10].

**Example 2q.** Let  $\Psi$  be the normalized in  $L_p$ ,  $1 < p \leq 2$ , real  $d$ -variate trigonometric system. Then Example 1q applies and gives for any  $f_0 \in L_p$

$$\|f_{C(t,p,d)m^{p'-1} \ln(m+1)}\|_p \leq C\sigma_m(f_0, \Psi)_p. \quad (4.4)$$

We need the concept of cotype of a Banach space  $X$ . We say that  $X$  has cotype  $s$  if for any finite number of elements  $u_i \in X$  we have the inequality

$$\left( \text{Average}_{\pm} \left\| \sum_i \pm u_i \right\|^s \right)^{1/s} \geq C_s \left( \sum_i \|u_i\|^s \right)^{1/s}.$$

It is known that the  $L_p$  spaces with  $2 \leq p < \infty$  have cotype  $s = p$  and  $L_p$  spaces with  $1 < p \leq 2$  have cotype 2.

**Remark 4.1.** Suppose  $\mathcal{D}$  is  $(K, K)$ -unconditional with a constant  $U$ . Assume that  $X$  is of cotype  $s$  with a constant  $C_s$ . Then  $\mathcal{D}$  has the Nikol'skii-type  $\ell_1 X$  property with parameters  $K, 1 - 1/s$  and  $C_1 = 2UC_s^{-1}$ .

*Proof.* Our assumption about  $(K, K)$ -unconditionality implies: for any  $A$ ,  $|A| \leq K$ , we have

$$\left\| \sum_{i \in A} \pm x_i g_i \right\| \leq 2U \left\| \sum_{i \in A} x_i g_i \right\|.$$

Therefore, by  $s$ -cotype assumption

$$\left\| \sum_{i \in A} x_i g_i \right\|^s \geq (2U)^{-s} C_s^s \sum_{i \in A} |x_i|^s.$$

This implies

$$\sum_{i \in A} |x_i| \leq |A|^{1-1/s} \left( \sum_{i \in A} |x_i|^s \right)^{1/s} \leq 2UC_s^{-1} |A|^{1-1/s} \left\| \sum_{i \in A} x_i g_i \right\|.$$

□

**Example 3.** Let  $X$  be a Banach space with  $\rho(u) \leq \gamma u^q$ ,  $1 < q \leq 2$  and with cotype  $s$ . Let  $\Psi$  be a normalized in  $X$  unconditional basis for  $X$ . Then  $U \leq C(X, \Psi)$  and  $\Psi$  satisfies **A2** with  $D = \infty$  and any  $K$ .

By Remark 4.1  $\Psi$  satisfies **A1** with  $r = 1 - \frac{1}{s}$ . Theorem 2.4 gives

$$\|f_{C(t,X,\Psi)m^{(1-1/s)q'}}\| \leq C\sigma_m(f_0, \Psi). \quad (4.5)$$

**Example 4.** Let  $\Psi$  be the normalized in  $L_p$ ,  $2 \leq p < \infty$ , multivariate Haar basis  $\mathcal{H}_p^d = \mathcal{H}_p \times \cdots \times \mathcal{H}_p$ . It is an unconditional basis. Also it is known that  $L_p$  space with  $2 \leq p < \infty$  has cotype  $s = 1/p$ . Therefore, Example 3 applies in this case. We give a direct argument here. It is an unconditional basis and therefore  $U \leq C(p, d)$ . Next, for any  $A$

$$\left\| \sum_{i \in A} x_i H_{i,p} \right\|_p \geq C(p, d) \left( \sum_{i \in A} |x_i|^p \right)^{1/p} \geq C(p, d) |A|^{\frac{1}{p}-1} \sum_{i \in A} |x_i|.$$

Therefore, we can take  $r = \frac{1}{p}$ . Theorem 2.4 gives

$$\|f_{C(t,p,d)m^{2/p'}}\|_p \leq C\sigma_m(f_0, \mathcal{H}_p^d). \quad (4.6)$$

Inequality (4.6) provides some progress in Open Problem 7.2 (p. 91) from [10] in the case  $2 < p < \infty$ .

**Example 4q.** Let  $\Psi$  be the normalized in  $L_p$ ,  $1 < p \leq 2$ , univariate Haar basis  $\mathcal{H}_p = \{H_{I,p}\}_I$ , where  $H_{I,p}$  the Haar functions indexed by dyadic intervals of support of  $H_{I,p}$  (we index function 1 by  $[0, 1]$  and the first Haar function by  $(0, 1]$ ). Then for any finite set  $A$  of dyadic intervals we have for  $f = \sum_I c_I(f) H_{I,p}$

$$\sum_{I \in A} |c_I| = \langle f, f_A^* \rangle, \quad f_A^* := \sum_{I \in A} (\text{sign } c_I(f)) H_{I,p} \|H_{I,p}\|_2^{-2}.$$

Therefore,

$$\sum_{I \in A} |c_I| \leq \|f\|_p \|f_A^*\|_{p'}.$$

It is easy to check that

$$\|H_{I,p}\|_{p'} \|H_{I,p}\|_2^{-2} = |I|^{-1/p} |I|^{1/p'} |I|^{-(1-2/p)} = 1.$$

By Lemma 1.23, p. 28, from [11] we get

$$\|f_A^*\|_{p'} \leq C(p) |A|^{1/p'}.$$

Thus

$$\sum_{I \in A} |c_I| \leq C(p) |A|^{1/p'} \|f\|_p.$$

This means that  $\mathcal{H}_p$  satisfies **A3** with  $V = C(p)$  and  $r = 1/p'$ . Also it is an unconditional basis and therefore satisfies **A2** with  $U = C(p)$ . It is known that  $L_p$  space with  $1 < p \leq 2$  has modulus of smoothness  $\rho(u) \leq \gamma u^p$ . Therefore, Theorem 2.8 applies in this case and gives

$$\|f_{C(t,p)m}\|_p \leq C\sigma_m(f_0, \mathcal{H}_p)_p. \quad (4.7)$$

Inequality (4.7) solves the Open Problem 7.2 (p. 91) from [10] in the case  $1 < p \leq 2$ .

**Example 5.** Let  $X$  be a Banach space with  $\rho(u) \leq \gamma u^2$ . Assume that  $\Psi$  is a normalized Schauder basis for  $X$ . Then for any  $f = \sum_i c_i(f)\psi_i$

$$\sum_{i \in A} |c_i(f)| \leq C(\Psi) |A| \|f\|.$$

This implies that  $\Psi$  satisfies **A3** with  $D = \infty$ ,  $V = C(\Psi)$ ,  $r = 1$  and any  $T$ . Theorem 2.7 gives

$$\|f_{C(t,X,\Psi)m^2 \ln m}\| \leq C\sigma_m(f_0, \Psi). \quad (4.8)$$

We note that the above simple argument still works if we replace the assumption that  $\Psi$  is a Schauder basis by the assumption that a dictionary  $D$  is  $(1, D)$ -unconditional with constant  $U$ . Then we obtain

$$\|f_{C(t,\gamma,U)K^2 \ln K}\| \leq C\sigma_K(f_0, \Psi), \quad \text{for } K + C(t, \gamma, U)K^2 \ln K \leq D.$$

**Example 5q.** Let  $X$  be a Banach space with  $\rho(u) \leq \gamma u^q$ ,  $1 < q \leq 2$ . Assume that  $\Psi$  is a normalized Schauder basis for  $X$ . Then for any  $f = \sum_i c_i(f)\psi_i$

$$\sum_{i \in A} |c_i(f)| \leq C(\Psi) |A| \|f\|.$$

This implies that  $\Psi$  satisfies **A3** with  $D = \infty$ ,  $V = C(\Psi)$ ,  $r = 1$  and any  $T$ . Theorem 2.7 gives

$$\|f_{C(t,X,\Psi)m^{q'} \ln m}\| \leq C\sigma_m(f_0, \Psi). \quad (4.9)$$

We note that the above simple argument still works if we replace the assumption that  $\Psi$  is a Schauder basis by the assumption that a dictionary  $\mathcal{D}$  is  $(1, D)$ -unconditional with constant  $U$ . Then we obtain

$$\|f_{C(t, \gamma, q, U)K^{q'} \ln K}\| \leq C\sigma_K(f_0, \mathcal{D}), \quad \text{for } K + C(t, \gamma, q, U)K^{q'} \ln K \leq D.$$

We now discuss application of Theorem 2.1 to quasi-greedy bases. We begin with a brief introduction to the theory of quasi-greedy bases. Let  $X$  be an infinite-dimensional separable Banach space with a norm  $\|\cdot\| := \|\cdot\|_X$  and let  $\Psi := \{\psi_m\}_{m=1}^\infty$  be a normalized basis for  $X$ . The concept of quasi-greedy basis was introduced in [4].

**Definition 4.1.** *The basis  $\Psi$  is called quasi-greedy if there exists some constant  $C$  such that*

$$\sup_m \|G_m(f, \Psi)\| \leq C\|f\|.$$

Subsequently, Wojtaszczyk [13] proved that these are precisely the bases for which the TGA merely converges, i.e.,

$$\lim_{n \rightarrow \infty} G_n(f) = f.$$

The following lemma is from [1] (see also [3]).

**Lemma 4.1.** *Let  $\Psi$  be a quasi-greedy basis of  $X$ . Then for any finite set of indices  $\Lambda$  we have for all  $f \in X$*

$$\|S_\Lambda(f, \Psi)\| \leq C \ln(|\Lambda| + 1)\|f\|.$$

We now formulate a result about quasi-greedy bases in  $L_p$  spaces. The following theorem is from [12]. We note that in the case  $p = 2$  Theorem 4.1 was proved in [13]. Some notations first. For a given element  $f \in X$  we consider the expansion

$$f = \sum_{k=1}^{\infty} c_k(f)\psi_k$$

and the decreasing rearrangement of its coefficients

$$|c_{k_1}(f)| \geq |c_{k_2}(f)| \geq \dots$$

Denote

$$a_n(f) := |c_{k_n}(f)|.$$

**Theorem 4.1.** Let  $\Psi = \{\psi_m\}_{m=1}^\infty$  be a quasi-greedy basis of the  $L_p$  space,  $1 < p < \infty$ . Then for each  $f \in X$  we have

$$C_1(p) \sup_n n^{1/p} a_n(f) \leq \|f\|_p \leq C_2(p) \sum_{n=1}^\infty n^{-1/2} a_n(f), \quad 2 \leq p < \infty;$$

$$C_3(p) \sup_n n^{1/2} a_n(f) \leq \|f\|_p \leq C_4(p) \sum_{n=1}^\infty n^{1/p-1} a_n(f), \quad 1 < p \leq 2.$$

**Example 6.** Let  $\Psi$  be a normalized quasi-greedy basis for  $L_p$ ,  $2 \leq p < \infty$ . Theorem 4.1 implies for any  $f = \sum_i c_i(f) \psi_i$

$$\sum_{i \in A} |c_i(f)| \leq \sum_{n=1}^{|A|} a_n(f) \leq C_1(p)^{-1} \sum_{n=1}^{|A|} n^{-1/p} \|f\|_p \leq C(p) |A|^{1-1/p} \|f\|_p.$$

This means that  $\Psi$  satisfies **A3** with  $D = \infty$ ,  $V = C(p)$ ,  $r = 1 - \frac{1}{p}$ . Theorem 2.7 gives

$$\|f_{C(t,p)m^{2(1-1/p)} \ln(m+1)}\| \leq C \sigma_m(f_0, \Psi). \quad (4.10)$$

**Example 6q.** Let  $\Psi$  be a normalized quasi-greedy basis for  $L_p$ ,  $1 < p \leq 2$ . Theorem 4.1 implies for any  $f = \sum_i c_i(f) \psi_i$

$$\sum_{i \in A} |c_i(f)| \leq \sum_{n=1}^{|A|} a_n(f) \leq C_3(p)^{-1} \sum_{n=1}^{|A|} n^{-1/2} \|f\|_p \leq C(p) |A|^{1/2} \|f\|_p.$$

This means that  $\Psi$  satisfies **A3** with  $D = \infty$ ,  $V = C(p)$ ,  $r = 1/2$ . Theorem 2.7 gives

$$\|f_{C(t,p)m^{p'/2} \ln(m+1)}\| \leq C \sigma_m(f_0, \Psi). \quad (4.11)$$

**Example 7.** Let  $\Psi$  be a normalized uniformly bounded orthogonal quasi-greedy basis for  $L_p$ ,  $2 \leq p < \infty$ . For existence of such bases see [6]. Then orthogonality implies that we can take  $r = 1/2$ . We obtain from Lemma 4.1 that  $\Psi$  is  $(K, \infty)$  unconditional with  $U \leq C \ln(K+1)$ . Theorem 2.8 gives

$$\|f_{C(t,p,\Psi)m \ln \ln(m+3)}\|_p \leq C \sigma_m(f_0, \Psi)_p. \quad (4.12)$$

**Example 7q.** Let  $\Psi$  be a normalized uniformly bounded orthogonal quasi-greedy basis for  $L_p$ ,  $1 < p \leq 2$ . For existence of such bases see [6]. Then orthogonality implies that we can take  $r = 1/2$ . We obtain from Lemma 4.1 that  $\Psi$  is  $(K, \infty)$  unconditional with  $U \leq C \ln(K+1)$ . Theorem 2.8 gives

$$\|f_{C(t,p,\Psi)m^{p'/2} \ln \ln(m+3)}\|_p \leq C \sigma_m(f_0, \Psi)_p. \quad (4.13)$$

## 5 Discussion

We study sparse approximation. In a general setting we study an algorithm (approximation method)  $\mathcal{A} = \{A_m(\cdot, \mathcal{D})\}_{m=1}^{\infty}$  with respect to a given dictionary  $\mathcal{D}$ . The sequence of mappings  $A_m(\cdot, \mathcal{D})$  defined on  $X$  satisfies the condition: for any  $f \in X$ ,  $A_m(f, \mathcal{D}) \in \Sigma_m(\mathcal{D})$ . In other words,  $A_m$  provides an  $m$ -term approximant with respect to  $\mathcal{D}$ . It is clear that for any  $f \in X$  and any  $m$  we have

$$\|f - A_m(f, \mathcal{D})\| \geq \sigma_m(f, \mathcal{D}).$$

We are interested in such pairs  $(\mathcal{D}, \mathcal{A})$  for which the algorithm  $\mathcal{A}$  provides approximation close to best  $m$ -term approximation. We introduce the corresponding definitions.

**Definition 5.1.** *We say that  $\mathcal{D}$  is a greedy dictionary with respect to  $\mathcal{A}$  if there exists a constant  $C_0$  such that for any  $f \in X$  we have*

$$\|f - A_m(f, \mathcal{D})\| \leq C_0 \sigma_m(f, \mathcal{D}). \quad (5.1)$$

If  $\mathcal{D}$  is a greedy dictionary with respect to  $\mathcal{A}$  then  $\mathcal{A}$  provides ideal (up to a constant  $C_0$ )  $m$ -term approximations for every  $f \in X$ .

**Definition 5.2.** *We say that  $\mathcal{D}$  is an almost greedy dictionary with respect to  $\mathcal{A}$  if there exist two constant  $C_1$  and  $C_2$  such that for any  $f \in X$  we have*

$$\|f - A_{C_1 m}(f, \mathcal{D})\| \leq C_2 \sigma_m(f, \mathcal{D}). \quad (5.2)$$

If  $\mathcal{D}$  is an almost greedy dictionary with respect to  $\mathcal{A}$  then  $\mathcal{A}$  provides almost ideal sparse approximation. It provides  $C_1 m$ -term approximant as good (up to a constant  $C_2$ ) as ideal  $m$ -term approximant for every  $f \in X$ . We also need a more general definition. Let  $\phi(u)$  be a function such that  $\phi(u) \geq 1$ .

**Definition 5.3.** *We say that  $\mathcal{D}$  is a  $\phi$ -greedy dictionary with respect to  $\mathcal{A}$  if there exists a constant  $C_3$  such that for any  $f \in X$  we have*

$$\|f - A_{\phi(m)m}(f, \mathcal{D})\| \leq C_3 \sigma_m(f, \mathcal{D}). \quad (5.3)$$

If  $\mathcal{D} = \Psi$  is a basis then in the above definitions we replace dictionary by basis. In the case  $\mathcal{A} = \{G_m(\cdot, \Psi)\}_{m=1}^\infty$  is the TGA the theory of greedy and almost greedy bases is well developed (see [11]). We present two results on characterization of these bases. A basis  $\Psi$  in a Banach space  $X$  is called *democratic* if there is a constant  $C(\Psi)$  such that

$$\left\| \sum_{k \in A} \psi_k \right\| \leq C(\Psi) \left\| \sum_{k \in B} \psi_k \right\| \quad (5.4)$$

if  $|A| = |B|$ . This concept was introduced in [4]. In [2] we defined a democratic basis as the one satisfying (5.4) if  $|A| \leq |B|$ . It is known that for quasi-greedy bases the above two definitions are equivalent. It was proved in [4] (see Theorem 1.15, p. 18, [11]) that a basis is greedy with respect to TGA if and only if it is unconditional and democratic. It was proved in [2] (see Theorem 1.37, p. 38, [11]) that a basis is almost greedy with respect to TGA if and only if it is quasi-greedy and democratic.

Example 4q is the first result about almost greedy bases with respect to WCGA in Banach spaces. It shows that the univariate Haar basis is an almost greedy basis with respect to the WCGA in the  $L_p$  spaces for  $1 < p \leq 2$ . Example 1 shows that uniformly bounded orthogonal bases are  $\phi$ -greedy bases with respect to WCGA with  $\phi(u) = C(t, p, D) \ln(u+1)$  in the  $L_p$  spaces for  $2 \leq p < \infty$ . We do not know if these bases are almost greedy with respect to WCGA. They are good candidates for that.

It is known (see [11], p. 17) that the univariate Haar basis is a greedy basis with respect to TGA for all  $L_p$ ,  $1 < p < \infty$ . Example 4 only shows that it is a  $\phi$ -greedy basis with respect to WCGA with  $\phi(u) = C(t, p)u^{1-2/p}$  in the  $L_p$  spaces for  $2 \leq p < \infty$ . It is much weaker than the corresponding results for the  $\mathcal{H}_p$ ,  $1 < p \leq 2$ , and for the trigonometric system,  $2 \leq p < \infty$  (see Example 2). We do not know if this result on the Haar basis can be substantially improved. At the level of our today's technique we can observe that the Haar basis is ideal (greedy basis) for the TGA in  $L_p$ ,  $1 < p < \infty$ , almost ideal (almost greedy basis) for the WCGA in  $L_p$ ,  $1 < p \leq 2$ , and that the trigonometric system is very good for the WCGA in  $L_p$ ,  $2 \leq p < \infty$ .

Example 2q shows that our results for the trigonometric system in  $L_p$ ,  $1 < p < 2$ , are not as strong as for  $2 \leq p < \infty$ . We do not know if it is a lack of appropriate technique or it reflects the nature of the WCGA with respect to the trigonometric system.

We note that properties of a given basis with respect to TGA and WCGA could be very different. For instance, the class of quasi-greedy bases (with



respect to TGA) is a rather narrow subset of all bases. It is close in a certain sense to the set of unconditional bases. The situation is absolutely different for the WCGA. If  $X$  is uniformly smooth then WCGA converges for each  $f \in X$  with respect to any dictionary in  $X$ . Moreover, Example 5q shows that if  $X$  is a Banach space with  $\rho(u) \leq \gamma u^q$  then any basis  $\Psi$  is  $\phi$ -greedy with respect to WCGA with  $\phi(u) = C(t, X, \Psi)u^{q'-1} \ln(u+1)$ .

It is interesting to compare Theorem 2.3 with the following known result. The following theorem provides rate of convergence (see [11], p. 347). As above we denote by  $A_1(\mathcal{D})$  the closure in  $X$  of the convex hull of the symmetrized dictionary  $\mathcal{D}^\pm := \{\pm g : g \in \mathcal{D}\}$ .

**Theorem 5.1.** *Let  $X$  be a uniformly smooth Banach space with modulus of smoothness  $\rho(u) \leq \gamma u^q$ ,  $1 < q \leq 2$ . Take a number  $\epsilon \geq 0$  and two elements  $f_0, f^\epsilon$  from  $X$  such that*

$$\|f_0 - f^\epsilon\| \leq \epsilon, \quad f^\epsilon/A(\epsilon) \in A_1(\mathcal{D}),$$

with some number  $A(\epsilon) > 0$ . Then, for the WCGA we have

$$\|f_m^{e,\tau}\| \leq \max \left( 2\epsilon, C(q, \gamma)(A(\epsilon) + \epsilon) \left(1 + \sum_{k=1}^m t_k^{q'}\right)^{-1/q'} \right).$$

Both Theorem 5.1 and Theorem 2.3 provide stability of the WCGA with respect to noise. In order to apply them for noisy data we interpret  $f_0$  as a noisy version of a signal and  $f^\epsilon$  as a noiseless version of a signal. Then, assumption  $f^\epsilon/A(\epsilon) \in A_1(\mathcal{D})$  describes our smoothness assumption on the noiseless signal and assumption  $f^\epsilon \in \Sigma_K(\mathcal{D})$  describes our structural assumption on the noiseless signal. In fact, Theorem 5.1 simultaneously takes care of two issues: noisy data and approximation in an interpolation space. Theorem 5.1 can be applied for approximation of  $f_0$  under assumption that  $f_0$  belongs to one of interpolation spaces between  $X$  and the space generated by the  $A_1(\mathcal{D})$ -norm (atomic norm).

Concluding, we briefly describe the contribution of this paper. First, we present a study of the Lebesgue-type inequalities with respect to the WCGA in Banach spaces with  $\rho(u) \leq \gamma u^q$ ,  $1 < q \leq 2$ , under conditions **A1** and **A2**. In the case  $q = 2$  it has been done in [5]. The case  $1 < q < 2$  uses the same ideas as in [5]. Second, we introduce a new condition **A3** and study the WCGA with respect to dictionaries satisfying either **A3** or **A2** and **A3**. Condition **A3** and a combination of **A2** and **A3** turn out to be

more powerful in applications than **A1** combined with **A2**. Third, we apply the general theory developed in Sections 2 and 3 for bases. Surprisingly, this technique works very well for very different bases. It provides first results on the Lebesgue-type inequalities for the WCGA with respect to bases in Banach spaces. Some of these results (for the  $\mathcal{H}_p$ ,  $1 < p \leq 2$ , and for the  $\mathcal{RT}_p$ ,  $2 \leq p < \infty$ ) are strong. This demonstrates that the technique used is an appropriate and powerful method.

## References

- [1] S.J. Dilworth, N.J. Kalton, and Denka Kutzarova, On the existence of almost greedy bases in Banach spaces, *Studia Math.*, **158** (2003), 67–101.
- [2] S.J. Dilworth, N.J. Kalton, Denka Kutzarova, and V.N. Temlyakov, The Thresholding Greedy Algorithm, Greedy Bases, and Duality, *Constr. Approx.*, **19** (2003), 575–597.
- [3] S.J. Dilworth, M. Soto-Bajo and V.N. Temlyakov, Quasi-greedy bases and Lebesgue-type inequalities, *Sud. Math.*, **211** (2012), 41–69.
- [4] S.V. Konyagin and V.N. Temlyakov, A remark on greedy approximation in Banach spaces, *East. J. Approx.* **5** (1999), 365–379.
- [5] E.D. Livshitz and V.N. Temlyakov, Sparse approximation and recovery by greedy algorithms, arXiv:1303.3595v1 [math.NA] 14 Mar 2013.
- [6] Morten Nielsen, An example of an almost greedy uniformly bounded orthonormal basis for  $L_p(0, 1)$ , *J. Approx. Theory*, **149** (2007), 188–192.
- [7] D. Savu and V.N. Temlyakov, Lebesgue-Type Inequalities for Greedy Approximation in Banach Spaces, *IEEE Transactions on Information Theory*, **58** (2013), 1098–1106.
- [8] V.N. Temlyakov, Greedy Algorithm and  $m$ -Term Trigonometric Approximation, *Constr. Approx.*, **14** (1998), 569–587.
- [9] V.N. Temlyakov, Greedy algorithms in Banach spaces, *Adv. Comput. Math.*, **14** (2001), 277–292.

- [10] V.N. Temlyakov, Nonlinear Methods of Approximation, Found. Comput. Math., **3** (2003), 33–107.
- [11] V.N. Temlyakov, Greedy approximation, Cambridge University Press, 2011.
- [12] V.N. Temlyakov, Mingrui Yang and Peixin Ye, Greedy approximation with regard to non-greedy bases, Adv. Comput. Math. **34** (2011), 319–337.
- [13] P. Wojtaszczyk, Greedy Algorithm for General Biorthogonal Systems, J. Approx. Theory **107** (2000), 293-314.