INTERDISCIPLINARY MATHEMATICS INSTITUTE

2013:03

Sparse approximation and recovery by greedy algorithms

E. Livshitz and V. Temlyakov

COLLEGE OF ARTS AND SCIENCES

UNIVERSITY OF SOUTH CAROLINA

Sparse approximation and recovery by greedy algorithms*

E. Livshitz†and V. Temlyakov ‡

March 14, 2013

Abstract

We study sparse approximation by greedy algorithms. Our contribution is two-fold. First, we prove exact recovery with high probability of random K-sparse signals within $\lceil K(1 + \epsilon) \rceil$ iterations of the Orthogonal Matching Pursuit (OMP). This result shows that in a probabilistic sense the OMP is almost optimal for exact recovery. Second, we prove the Lebesgue-type inequalities for the Weak Chebyshev Greedy Algorithm, 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. However, even in the case of a Hilbert space our results add some new elements to known results on the Lebesque-type inequalities for the RIP dictionaries. Our technique is a development of the recent technique created by Zhang.

1   Introduction

This paper deals with sparse approximation. Driven by applications in biology, medicine, and engineering approximation problems are formulated in

---

very high dimensions, which bring to the fore new phenomena. One aspect of the high-dimensional context is a focus on sparse signals (functions). The main motivation for the study of sparse signals is that many real world signals can be well approximated by sparse ones. A very important step in solving multivariate problems with large dimension occurred during last 20 years. Researchers began to use sparse representations as a way to model the corresponding function classes. This approach 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 of importance of a Banach space setting in numerical analysis in general and in sparse approximation in particular (see, for instance, [10], Preface, and [5]). An element (function, signal) $f \in X$ is said to be K-sparse with respect to $\mathcal{D}$ if it has a representation $f = \sum_{i=1}^{K} x_i g_i$, $g_i \in \mathcal{D}$, $i = 1, \ldots, K$. 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

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

Here are two fundamental problems of sparse approximation.

P1. Exact recovery. Suppose we know that $f_0 \in \Sigma_K(\mathcal{D})$. How can we recover it?

P2. Approximate recovery. How to design a practical algorithm that builds m-term approximations comparable to best m-term approximations?

It is known that in both of the above problems greedy-type algorithms play a fundamental role. We discuss one of them here. There are two special cases of the above general setting of the sparse approximation problem.

(I). Instead of a Banach space X we consider a Hilbert space H. Approximation is still with respect to a redundant dictionary $\mathcal{D}$.

(II). We approximate in a Banach space X with respect to a basis Ψ instead of a redundant dictionary $\mathcal{D}$.

This section discusses setting (I) and the corresponding generalizations to the Banach space setting. Section 4 addresses setting (II). We begin our discussion with the Orthogonal Greedy Algorithm (OGA) in a Hilbert space. The Orthogonal Greedy Algorithm is called the Orthogonal Matching Pursuit (OMP) in signal processing. We will use the name Orthogonal Matching Pursuit for this algorithm in this paper. It is natural to compare perfor-

2

mance of the OMP with the best m-term approximation with regard to a dictionary $\mathcal{D}$. We recall some notations and definitions from the theory of greedy algorithms. Let H be a real Hilbert space with an inner product $\langle \cdot, \cdot \rangle$ and the norm $\|x\| := \langle x, x \rangle^{1/2}$. We say a set $\mathcal{D}$ of functions (elements) from H is a dictionary if each $g \in \mathcal{D}$ has a unit norm ($\|g\| = 1$) and the closure of span $\mathcal{D}$ is H. Let a sequence $\tau = \{t_k\}_{k=1}^{\infty}$, $0 \leq t_k \leq 1$, be given. The following greedy algorithm was defined in [6] under the name Weak Orthogonal Greedy Algorithm (WOGA).

Weak Orthogonal Matching Pursuit (WOMP). Let f_0 be given. Then for each $m \geq 1$ we inductively define:

(1) $\varphi_m \in \mathcal{D}$ is any element satisfying

$$|\langle f_{m-1}, \varphi_m \rangle| \geq t_m \sup_{g \in \mathcal{D}} |\langle f_{m-1}, g \rangle|.$$

(2) Let $H_m := \operatorname{span}(\varphi_1, \ldots, \varphi_m)$ and let $P_{H_m}(\cdot)$ denote an operator of orthogonal projection onto H_m. Define

$$G_m(f_0, \mathcal{D}) := P_{H_m}(f_0).$$

(3) Define the residual after mth iteration of the algorithm

$$f_m := f_0 - G_m(f_0, \mathcal{D}).$$

In the case $t_k = 1$, $k = 1, 2, \ldots$, WOMP is called the Orthogonal Matching Pursuit (OMP). In this paper we only consider the case $t_k = t$, $k = 1, 2, \ldots$, $t \in (0, 1]$.

The theory of the WOMP is well developed (see [10]). In first results on performance of the WOMP in problems **P1** and **P2** researchers imposed the incoherence assumption on a dictionary $\mathcal{D}$. The reader can find detailed discussion of these results in [10], Section 2.6 and [4]. Recently, exact recovery results and Lebesgue-type inequalities for the WOMP under assumption that $\mathcal{D}$ satisfies Restricted Isometry Property (RIP) introduced in compressed sensing theory (see Definition 2.1 below) have been proved (see [12], [3], [11]). A breakthrough result in this direction was obtained by Zhang [12]. In particular, he proved that if $\delta_{31K}^{RIP}(\mathcal{D}) < 1/3$ then the OMP recovers exactly all K-sparse signals within $30K$ iterations. In other words, $f_{30K} = 0$. It is interesting and difficult problem to improve the constant 30. There are several papers devoted to this problem (see [3] and [11]). In this paper we develop Zhang's technique in two directions: (1) to obtain exact recovery with high

3

probability of random K-sparse signals within $\lceil K(1+\epsilon) \rceil$ iterations of the OMP and (2) to obtain recovery results and the Lebesgue-type inequalities in the Banach space setting.

In Section 2 we prove exact recovery results under RIP conditions on a dictionary combined with assumptions on the sparse signal to be recovered (see Theorem 2.1). We prove that the corresponding assumptions on a sparse signal are satisfied with high probability if it is a random signal. In particular, we prove the following theorem.

Theorem 1.1. *For any $\epsilon > 0$ there exist $\delta = \delta(\epsilon) > 0$ and $K_0 = K_0(\epsilon)$ such that for any dictionary $\mathcal{D}$, $\delta_{2K}^{RIP}(\mathcal{D}) < \delta$, $K \geq K_0$, the following statement holds. Let $f_0 \in \Sigma_K(\mathcal{D})$ and its nonzero coefficients are uniformly distributed on $[-1, 1]$ independent random variables. Then $f_{\lceil K(1+\epsilon) \rceil} = 0$ with probability greater than $1 - \exp(-C(\epsilon)K)$.*

This theorem shows that in a probabilistic sense the OMP is almost optimal for exact recovery.

Sections 3 is devoted to the Banach space setting. Let X be a Banach space with norm $\| \cdot \| := \| \cdot \|_X$. As in the case of Hilbert spaces 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 span $\mathcal{D}$ is X. For a nonzero element $g \in X$ we let F_g denote a norming (peak) functional for g:

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

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 $t_k \leq 1$, $k = 1, \ldots$. We define the Weak Chebyshev Greedy Algorithm (WCGA) (see [7]) as a generalization for Banach spaces of the Weak Orthogonal Matching Pursuit. 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

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

4

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

(3) Let
$$f_m := f_m^{c,\tau} := f_0 - G_m.$$

In Section 3 we prove the Lebesgue-type inequalities for the WCGA. 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 [7] and [10]) 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

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

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

For notational convenience we consider here a countable dictionary $\mathcal{D} = \{g_i\}_{i=1}^{\infty}$. For a given f_0, let the sparse element (signal)

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

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

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

We use the following two assumptions.

A1. Nikol'skii-type inequality. The sparse element $f = \sum_{i \in T} x_i g_i$ satisfies Nikol'skii-type $\ell_1 X$ inequality with parameter r if

$$\sum_{i \in A} |x_i| \le C_1 |A|^r \|f_A\|, \quad A \subset T, \quad r \ge 1/2.$$

A2. Incoherence property. The sparse element $f = \sum_{i \in T} x_i g_i$ has incoherence property with parameters D and U if for any $A \subset T$ and any Λ, such that $A \cap \Lambda = \emptyset$ and $|A| + |\Lambda| \le D$, we have for any $\{c_i\}$

$$\left\| f_A - \sum_{i \in \Lambda} c_i g_i \right\| \ge U^{-1} \|f_A\|.$$

The main result of Section 3 is the following.

Theorem 1.2. *Let X be a Banach space with $\rho(u) \leq \gamma u^2$. Suppose K-sparse f^ϵ satisfies $\boldsymbol{A1}$, $\boldsymbol{A2}$ and $\|f_0 - f^\epsilon\| \leq \epsilon$. Then the WCGA with weakness parameter t applied to f_0 provides*

$$\|f_{C(t,\gamma,C_1)U^2 \ln(U+1)K^{2r}}\| \leq C\epsilon \quad for \quad K + C(t,\gamma,C_1)U^2 \ln(U+1)K^{2r} \leq D$$

with an absolute constant C.

Theorem 1.2 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 it as a theorem.

Theorem 1.3. *Let X be a Hilbert space. Suppose K-sparse f^ϵ satisfies $\boldsymbol{A2}$ and $\|f_0 - f^\epsilon\| \leq \epsilon$. Then the WOMP with weakness parameter t applied to f_0 provides*

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

with an absolute constant C.

Theorem 1.3 implies the following corollary.

Corollary 1.1. *Let X be a Hilbert space. Suppose any K-sparse f satisfies $\boldsymbol{A2}$. Then the WOMP with weakness parameter t applied to f_0 provides*

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

with an absolute constant C.

We show in Sections 3 that the RIP condition with parameters D and δ implies the (D, D) unconditionality with $U = (1+\delta)^{1/2}(1-\delta)^{-1/2}$. Therefore, Corollary 1.1 reads as follows in this case.

Corollary 1.2. *Let X be a Hilbert space. Suppose $\mathcal{D}$ satisfies RIP condition with parameters D and δ. Then the WOMP with weakness parameter t applied to f_0 provides*

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

with an absolute constant C.

We emphasize that in Theorem 1.2 we impose our conditions on an individual function f^ϵ. It may happen that the dictionary does not satisfy assumptions of $\ell_1 X$ inequality and (K, D)-unconditionality (see Section 3) but the given f_0 can be approximated by f^ϵ which does satisfy assumptions **A1** and **A2**. Even in the case of a Hilbert space our approach adds something new to the study based on the RIP. First of all, Theorem 1.3 shows that it is sufficient to impose assumption **A2** on an individual f^ϵ in order to obtain exact recovery and the Lebesgue-type inequality results. Second, Corollary 1.1 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 1.2 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 δ.

2 Almost optimality of the OMP

We prove Theorem 1.1 in this section. For the readers convenience we use notations which are standard in signal processing. Let $\mathcal{D} = \{\phi_i\}_{i=1}^N$ be a dictionary in $\mathbb{R}^M$, $M < N$. By Φ denote an $M \times N$ matrix, consisting of elements of $\mathcal{D}$ ($\phi_i \in \mathbb{R}^M$ is the i-th column of Φ). We say that $\mathbf{x} \in \mathbb{R}^N$ is S-sparse if $\mathbf{x}$ has at most S nonzero coordinates.

Definition 2.1. *A matrix Φ satisfies $RIP(S, \delta)$ if the inequality*

$$(1 - \delta)\|\mathbf{x}\|^2 \leq \|\Phi\mathbf{x}\|^2 \leq (1 + \delta)\|\mathbf{x}\|^2 \tag{2.1}$$

holds for all S-sparse $\mathbf{x} \in \mathbb{R}^N$. The minimum of all constants δ, satisfying (2.1), is called the isometric constant $\delta_S(\Phi) = \delta_S(\mathcal{D}) = \delta_S^{RIP}(\mathcal{D})$.

In this section we study the OMP and use the "compressed sensing notation" for the residual of the OMP. Set

$$\mathbf{r}^m := f_m, \; m \geq 0.$$

Consider the set

$$\Omega = \{1, \ldots, N\}.$$

Since $f_0 \in \Sigma_K(\mathcal{D})$, there exists an $\mathbf{x} = (x_1, x_2, \ldots, x_N)$, $\mathrm{supp}\,\mathbf{x} = T$, $T \subset \Omega$, $|T| = K$ such that

$$\mathbf{r}^0 = f_0 = \Phi\mathbf{x}.$$

7

Denote by T^m the set of indices of ϕ_i picked by the OMP after m iterations. According to the definition of the OMP for every $m \geq 0$ we choose $\mathbf{x}^m \in \mathbb{R}^N$, satisfying the following relations

$$\operatorname{supp} \mathbf{x}^m \subset T^m, \quad |T^m| = m, \text{ while } \mathbf{r}^m \neq 0, \tag{2.2}$$

$$G_m(f, \mathcal{D}) = \Phi \mathbf{x}^m,$$

$$\mathbf{r}^m = \Phi \mathbf{x} - \Phi \mathbf{x}^m. \tag{2.3}$$

Let $N(\mathbf{x}, \nu)$ be the minimal integer such that

$$\|\mathbf{x}_\Lambda\|^2 > \nu, \text{ for all } \Lambda \subset T, \ |\Lambda| \geq N(\mathbf{x}, \nu) + 1. \tag{2.4}$$

Theorem 2.1. *There exists an absolute constant $\widehat{C}$ such that for any δ, $0 < \delta < 0.001$, an integer $K \geq K_0 = K_0(\delta)$, and a dictionary $\mathcal{D}$, $\delta_{2K}^{RIP}(\mathcal{D}) < \delta$ the following statement holds. The OMP recovers exactly every K-sparse signal $\mathbf{x}$ within $K + 6N(\mathbf{x}, \widehat{C}\delta^{1/2}K)$ iterations, in other words, $\mathbf{r}^{K+6N(\mathbf{x}, \widehat{C}\delta^{1/2}K)} = 0$.*

Here is a direct corollary of Theorem 2.1.

Corollary 2.1. *Let K-sparse $\mathbf{x}$ be such that $|x_i| = 1$, $i \in T$, $|T| = K$. Then under assumptions of Theorem 2.1 the OMP recovers $\mathbf{x}$ exactly within $(1 + 6\widehat{C}\delta^{1/2})K$ iterations.*

Proof. Set
$$\Gamma^m := T \setminus T^m.$$
We fix
$$a := \delta^{1/2}. \tag{2.5}$$

Consider $m \in \mathbb{Z}_+$ such that

$$|T^m| + [aK] = m + [aK] \leq K. \tag{2.6}$$

Assume that $K \geq K_0 = K_0(a) \geq 1/a$. Let z^m be the maximal number, satisfying the following inequality

$$|\{i \in \Gamma^m : |x_i| \geq z^m\}| \geq [aK], \ |\{i \in \Gamma^m : |x_i| \leq z^m\}| \geq |\Gamma^m| - [aK]. \tag{2.7}$$

In other words z^m is the $[aK]$th largest element out of $\{|x_i|\}_{i \in \Gamma^m}$. We use the following lemma.

8

Lemma 2.1. *Under (2.6) the following inequality is valid:*

$$\|\mathbf{r}^m\|^2 - \|\mathbf{r}^{m+1}\|^2 \geq (z^m)^2(1 - C_1 a).$$

Proof. According to (2.7), we can choose sets

$$\Gamma_+^m \subset \Gamma^m, \quad |\Gamma_+^m| = [aK] \tag{2.8}$$

and

$$\Gamma_-^m := \Gamma \setminus \Gamma_+^m \tag{2.9}$$

with the following property

$$\min_{i \in T_+^m} |x_i| \geq z^m \geq \max_{i \in \Gamma_-^m} |x_i|. \tag{2.10}$$

Consider $\mathbf{w} \in \mathbb{R}^N$ such that

$$\mathbf{w}_{T \cap T^m \cup \Gamma_+^m} = \mathbf{x}_{T \cap T^m \cup \Gamma_+^m}, \quad \mathbf{w}_{\Omega \setminus (T \cap T^m \cup \Gamma_+^m)} = 0. \tag{2.11}$$

We use several well-known properties of the OMP:

$$\|\mathbf{r}^m\|^2 - \|\mathbf{r}^{m+1}\|^2 \geq \sup_{\phi \in \mathcal{D}} \langle \mathbf{r}^m, \phi \rangle^2, \tag{2.12}$$

$$\sup_{\phi \in \mathcal{D}} |\langle \mathbf{r}^m, \phi \rangle| \geq \frac{|\langle r^m, \Phi \mathbf{u} \rangle|}{\|\mathbf{u}\|_1}, \quad \mathbf{u} \in \mathbb{R}^{\mathbb{N}}, \tag{2.13}$$

$$\langle \mathbf{r}^m, \Phi \mathbf{u} \rangle = 0, \text{ if } \operatorname{supp} \mathbf{u} \subset T^m. \tag{2.14}$$

In particular

$$\langle \mathbf{r}^m, \Phi \mathbf{x}^m \rangle = 0. \tag{2.15}$$

Using (2.13) for $\mathbf{u} = \mathbf{w}_{\Omega \setminus T^m}$, we can estimate

$$
\begin{aligned}
|\sup_{\phi \in \mathcal{D}} \langle \mathbf{r}^m, \phi \rangle| \quad &\geq \quad \frac{|\langle \mathbf{r}^m, \Phi \mathbf{w}_{\Omega \setminus T^m} \rangle|}{\|\mathbf{w}_{\Omega \setminus T^m}\|_1} \overset{(2.14)}{=} \frac{|\langle \mathbf{r}^m, \Phi \mathbf{w} \rangle|}{\|\mathbf{w}_{\Omega \setminus T^m}\|_1} \\
&\overset{(2.15)}{=} \frac{|\langle \mathbf{r}^m, \Phi(\mathbf{w} - \mathbf{x}^m) \rangle|}{\|\mathbf{w}_{\Omega \setminus T^m}\|_1} \geq \frac{|\langle \mathbf{r}^m, \Phi(\mathbf{w} - x^m) \rangle|}{\|\mathbf{w}_{\Omega \setminus T^m}\|_0^{1/2} \|\mathbf{w}_{\Omega \setminus T^m}\|_2}.
\end{aligned}
$$

Applying (2.8) and (2.11), we obtain from the above inequality

$$|\sup_{\phi \in \mathcal{D}} \langle \mathbf{r}^m, \phi \rangle| \geq \frac{|\langle \mathbf{r}^m, \Phi(\mathbf{w} - x^m) \rangle|}{(|\Gamma_+^m|)^{1/2} \|\mathbf{w}_{\Omega \setminus T^m}\|_2} \geq \frac{|\langle \mathbf{r}^m, \Phi(\mathbf{w} - x^m) \rangle|}{(aK)^{1/2} \|\mathbf{w}_{\Omega \setminus T^m}\|_2}. \tag{2.16}$$

9

We estimate

$$\|\mathbf{r}^m\|^2 \;=\; \|\Phi(\mathbf{x} - \mathbf{x}^m)\|^2 \overset{RIP}{\geq} (1-\delta)\|\mathbf{x} - \mathbf{x}^m\|^2 \geq (1-\delta)\|(\mathbf{x} - \mathbf{x}^m)_{\Gamma^m}\|^2$$

$$= \; (1-\delta)\|\mathbf{x}_{\Gamma^m}\|^2 = (1-\delta)\|\mathbf{x}_{\Gamma^m_+ \cup \Gamma^m_-}\|^2$$

$$= \; (1-\delta)(\|\mathbf{x}_{\Gamma^m_+}\|^2 + \|\mathbf{x}_{\Gamma^m_-}\|^2) \overset{(2.10),(2.8)}{\geq} (1-\delta)((z^m)^2[aK] + \|\mathbf{x}_{\Gamma^m_-}\|^2),$$

and

$$\|\Phi(\mathbf{w} - \mathbf{x})\|^2 \overset{(2.11)}{=} \|\Phi\mathbf{x}_{\Gamma^m_-}\|^2 \overset{RIP}{\leq} (1+\delta)\|\mathbf{x}_{\Gamma^m_-}\|^2.$$

Combining two last inequalities, we obtain, for sufficiently large $K_0 = K_0(a) = K_0(\epsilon)$,

$$
\begin{aligned}
\|\mathbf{r}^m\|^2 - \|\Phi(\mathbf{w} - \mathbf{x})\|^2 \;\;&\geq\;\; (1-\delta)(z^m)^2[aK] - 2\delta\|\mathbf{x}_{\Gamma^m_-}\|^2 \\
&\overset{(2.10)}{\geq} (1-\delta)(z^m)^2[aK] - 2\delta(z^m)^2|\Gamma^m_-| \\
&\overset{(2.9)}{\geq} (1-\delta)(z^m)^2[aK] - 2\delta(z^m)^2 K \\
&\geq (1-2\delta)(z^m)^2(aK) - 2\delta(z^m)^2 K \\
&= (z^m)^2 aK\left((1-2\delta) - \frac{2\delta}{a}\right) \\
&\overset{(2.5)}{\geq} (z^m)^2 aK(1-4a). \quad\quad (2.17)
\end{aligned}
$$

Following the technique from [11] we have

$$
\begin{aligned}
|\langle \mathbf{r}^m, \Phi(\mathbf{w} - \mathbf{x}^m)\rangle| \;\;&=\;\; \frac{1}{2}\left|\|\Phi(\mathbf{w} - \mathbf{x}^m)\|_2^2 + \|\mathbf{r}^m\|^2 - \|\Phi(\mathbf{w} - \mathbf{x}^m) - r^m\|^2\right| \\
&\overset{(2.3)}{=} \frac{1}{2}\left|\|\Phi(\mathbf{w} - \mathbf{x}^m)\|_2^2 + (\|\mathbf{r}^m\|^2 - \|\Phi(\mathbf{w} - \mathbf{x})\|^2)\right| \\
&\overset{(2.17)}{\geq} \left(\|\Phi(\mathbf{w} - \mathbf{x}^m)\|_2^2 (z^m)^2 aK(1-4a)\right)^{1/2} \\
&= \|\Phi(\mathbf{w} - \mathbf{x}^m)\|_2 z^m (aK(1-4a))^{1/2} \\
&\overset{RIP}{\geq} (1-\delta)^{1/2}\|\mathbf{w} - \mathbf{x}^m\| z^m (aK(1-4a))^{1/2} \\
&\geq \|\mathbf{w} - \mathbf{x}^m\| z^m (aK)^{1/2}(1 - c_1 a) \\
&\geq \|(\mathbf{w} - \mathbf{x}^m)_{\Omega \setminus T^m}\| z^m (aK)^{1/2}(1 - c_1 a) \overset{(2.2)}{=} \\
&= \|\mathbf{w}_{\Omega \setminus T^m}\| z^m (aK)^{1/2}(1 - c_1 a). \quad\quad (2.18)
\end{aligned}
$$

10

Substituting (2.18) in (2.16) and (2.12), we finally get

$$
\begin{aligned}
\|\mathbf{r}^m\|^2 - \|\mathbf{r}^{m+1}\|^2 \;&\geq\; \frac{\langle \mathbf{r}^m, \Phi(\mathbf{w} - \mathbf{x}^m)\rangle^2}{aK\|\mathbf{w}_{\Omega\setminus T^m}\|_2^2} \;\geq\; \frac{\|\mathbf{w}_{\Omega\setminus T^m}\|^2 (z^m)^2 aK(1 - c_1 a)^2}{aK\|\mathbf{w}_{\Omega\setminus T^m}\|_2^2} \\
&\geq\; (z^m)^2(1 - C_1 a).
\end{aligned}
$$

$\square$

We continue to prove Theorem 2.1. Without loss of generality we may assume that $T = \{1, \ldots, K\}$ and that the sequence $\{|x_i|\}_{i=1}^K$ decreases. Then using the inequality $|T \cap T^m| \leq m$ and the definition (2.7), we have

$$
z^m \geq |x_{m+[aK]}| \geq |x_{m+1+[aK]}|.
$$

Applying Lemma 2.1, we have for $m \geq 1$, $m + [aK] \leq K$,

$$
\|\mathbf{r}^{m-1}\|^2 - \|\mathbf{r}^m\|^2 \geq (z^{m-1})^2(1 - C_1 a) \geq x_{m+[aK]}^2(1 - C_1 a). \tag{2.19}
$$

First we bound $\|\mathbf{r}^K\|$ from above

$$
\begin{aligned}
\|\mathbf{r}^K\|^2 \;&=\; \|\mathbf{r}^0\|^2 - \sum_{m=1}^{K}\left(\|\mathbf{r}^{m-1}\|^2 - \|\mathbf{r}^m\|^2\right) \\
&\overset{(\mathbf{r}^0 = \Phi\mathbf{x})}{=}\; \|\Phi\mathbf{x}\|^2 - \sum_{m=1}^{K}\left(\|\mathbf{r}^{m-1}\|^2 - \|\mathbf{r}^m\|^2\right) \\
&\overset{RIP}{\leq}\; (1+\delta)\sum_{i=1}^{K} x_i^2 - \sum_{m=1}^{K-[aK]}\left(\|\mathbf{r}^{m-1}\|^2 - \|\mathbf{r}^m\|^2\right) \\
&\overset{(2.19)}{\leq}\; (1+\delta)\sum_{i=1}^{K} x_i^2 - \sum_{m=1}^{K-[aK]} x_{m+[aK]}^2(1 - C_1 a) \\
&\leq\; (1+\delta)\sum_{i=1}^{K} x_i^2 - \sum_{i=1+[aK]}^{K} x_i^2(1 - C_1 a) \\
&\leq\; (\delta + C_1 a)\sum_{i=1}^{K} x_i^2 + \sum_{i=1}^{[aK]} x_i^2 \\
&\overset{|x_i|\leq 1}{\leq}\; (\delta + C_1 a)K + [aK] \leq KC_2 a \overset{(2.5)}{=} KC_2\delta^{1/2}. \tag{2.20}
\end{aligned}
$$

11

Then using RIP, we can estimate $\|\mathbf{r}^K\|$ from below

$$\|\mathbf{r}^K\|^2 \geq (1 - \delta) \sum_{i \in T \setminus T^K} x_i^2. \tag{2.21}$$

Set

$$\widehat{C} := \frac{C_2}{1 - \delta}.$$

Combining this definition with (2.20) and (2.21), we obtain

$$\sum_{i \in T \setminus T^K} x_i^2 \leq \widehat{C} K \delta^{1/2}.$$

Thus, using (2.4), we conclude that

$$|T \setminus T^K| \leq N(\mathbf{x}, \widehat{C} K \delta^{1/2}). \tag{2.22}$$

It is known that (see Lemma 1.2 from [1] and Lemma 1 from [2])

$$\delta_{2S}(\Phi) \leq 3\delta_S(\Phi).$$

Then the condition $\delta < 0.001$ implies that

$$\delta_{10K}(\Phi) \leq \delta_{16K}(\Phi) \leq 27\delta_{2K}(\Phi) \leq 27\delta \leq 0.03. \tag{2.23}$$

Now we can apply the improvement of Zhang's theorem obtained by Wang and Shim ([11], Theorem 3.1). It claims that under (2.23) we have

$$\mathbf{r}^{K+6|T \setminus T^K|} = 0,$$

Therefore, taking into account (2.22), we finally get

$$\mathbf{r}^{K+6N(\mathbf{x}, \widehat{C} K \delta^{1/2})} = 0.$$

$\square$

As corollaries of Theorem 2.1 we obtain Theorem 1.1 and the following result.

Theorem 2.2. *For any $\epsilon_1, \epsilon_2 > 0$ there exist $\delta = \delta(\epsilon_1, \epsilon_2) > 0$ and $K_0 = K_0(\epsilon_1, \epsilon_2)$ such that for any dictionary $\mathcal{D}$, $\delta_{2K}^{RIP}(\mathcal{D}) < \delta$, $K \geq K_0$, the following statement holds. If $\mathbf{r}^0 = f_0 \in \Sigma_K(\mathcal{D})$ and its nonzero coefficients belong to $[-1, 1] \setminus (-\epsilon_1, \epsilon_1)$, then $\mathbf{r}^{\lceil K(1+\epsilon_2) \rceil} = 0$.*

Proof. It is clear that for any $\Lambda \subset T$ we have

$$\|\mathbf{x}_\Lambda\|^2 \geq \epsilon_1^2 |\Lambda|.$$

Hence according to (2.4) we get

$$N(\mathbf{x}, \nu) \leq \frac{\nu}{\epsilon_1^2}.$$

Then

$$6N(\mathbf{x}, \widehat{C} K \delta^{1/2}) \leq K 6 \frac{\widehat{C} \delta^{1/2}}{\epsilon_1^2}.$$

Thus, to complete the proof it remains to choose $\delta = \delta(\epsilon_1, \epsilon_2)$ such that

$$6 \frac{\widehat{C} \delta^{1/2}}{\epsilon_1^2} \leq \epsilon_2.$$

$\square$

Lemma 2.2. *Assume that $p < 1$ and numbers x_i, $1 \leq i \leq K$, $K \geq K_0(p)$ are uniformly distributed on $[-1, 1]$ independent random variables. Then*

$$|\{i : |x_i| < p\}| \leq 2pK$$

with probability greater than $1 - \exp(-C(p)K)$.

Proof. For i, $1 \leq i \leq K$, we set $\xi_i = 0$, if $|x_i| \geq p$, and $\xi_i = 1$, otherwise. So ξ_i has Bernoulli distribution with

$$\mathsf{P}\{\xi_i = 1\} = p, \quad \mathsf{P}\{\xi_i = 0\} = 1 - p, \quad \mathsf{E}\xi_i = p.$$

By Hoeffding's inequality (see, for instance, [10], p. 198, inequality (4.22)) we obtain

$$\mathsf{P}\left\{\frac{1}{K} \sum_{i=1}^{K} \xi_i - p \geq p\right\} \leq \exp(-Kp^2/2).$$

Clearly,

$$|\{i : |x_i| < p\}| = \sum_{i=1}^{K} \xi_i.$$

Therefore,

$$\mathsf{P}\{|\{i : |x_i| < p\}| \leq 2p\} = \mathsf{P}\left\{\frac{1}{K} \sum_{i=1}^{K} \xi_i \leq 2p\right\} \geq 1 - \exp(-Kp^2/2).$$

$\square$

We now give a proof of Theorem 1.1 from the Introduction.

Proof. Let
$$\varkappa := \varkappa(\delta) := \widehat{C}\delta^{1/2}.$$

According to Lemma 2.1 with probability greater than $1 - \exp(-C(\delta)K)$ we have
$$\left|\{i : |x_i| < \varkappa^{1/3}\}\right| \leq 2\varkappa^{1/3}K \tag{2.24}$$

To prove the theorem we need to estimate $N(\mathbf{x}, \varkappa K)$. Consider $\Lambda \subset T$ such that
$$\|\mathbf{x}_\Lambda\|^2 \leq \varkappa K.$$

Then we estimate

$$
\begin{aligned}
|\Lambda| &= |\{i \in \Lambda : |x_i| < \varkappa^{1/3}\}| + |\{i \in \Lambda : |x_i| \geq \varkappa^{1/3}\}| \\
&\leq |\{i \in \Lambda : |x_i| < \varkappa^{1/3}\}| + \frac{\varkappa K}{(\varkappa^{1/3})^2} \overset{(2.24)}{\leq} 2\varkappa^{1/3}K + \varkappa^{1/3}K = 3\varkappa^{1/3}K.
\end{aligned}
$$

Therefore, by definition (2.4) we have

$$N(\mathbf{x}, \varkappa K) \leq 3\varkappa^{1/3}K = 3(\widehat{C})^{1/3}\delta^{1/6}K.$$

To complete the proof it remains to apply Theorem 2.1 for $\delta < 0.001$ providing
$$N(\mathbf{x}, \varkappa K) < \epsilon K/6.$$

$\square$

3   Lebesgue-type inequalities

We discuss here the Lebesgue-type inequalities for the WCGA with $\tau = \{t\}$, $t \in (0, 1]$. We repeat the above assumptions **A1** and **A2** with remarks on the corresponding properties of dictionaries. 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
$$f_A := f_A^\epsilon := \sum_{i \in A} x_i g_i.$$

Here are two assumptions that we will use.

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

$$\sum_{i \in A} |x_i| \leq C_1 |A|^r \|f_A\|, \quad A \subset T, \quad r \geq 1/2. \tag{3.1}$$

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 Λ such that $A \cap \Lambda = \emptyset$, $|A| + |\Lambda| \leq D$ we have for any $\{c_i\}$

$$\|f_A - \sum_{i \in \Lambda} c_i g_i\| \geq U^{-1} \|f_A\|. \tag{3.2}$$

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 (3.2) 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 $e_1, \ldots, e_D$ from $\mathcal{D}$ is linearly independent and for any A with $|A| \leq K$ and any coefficients $\{c_i\}$ we have

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

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 [10], p. 306).

Definition 3.1. *A dictionary $\mathcal{D}$ is called the Riesz dictionary with depth D and parameter $\delta \in (0, 1)$ if, for any D distinct elements $e_1, \ldots, e_D$ of the dictionary and any coefficients $a = (a_1, \ldots, a_D)$, we have*

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

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

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

Lemma 3.1. *Let $\mathcal{D} \in R(D, \delta)$ and let $e_j \in \mathcal{D}$, $j = 1, \ldots, s$. For $f = \sum_{i=1}^{s} a_i e_i$ and $A \subset \{1, \ldots, s\}$ denote*

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

If $s \le D$ then
$$\|S_A(f)\|^2 \le (1 + \delta)(1 - \delta)^{-1}\|f\|^2.$$

Lemma 3.1 implies that if $\mathcal{D} \in R(D, \delta)$ then it is (D, D)-unconditional with a constant $U = (1 + \delta)^{1/2}(1 - \delta)^{-1/2}$.

We need the concept of cotype of a Banach space X. We say that X has cotype $q \ge 2$ if for any finite number of elements $u_i \in X$ we have the inequality

$$\left(\text{Average}_{\pm} \| \sum_i \pm u_i \|^q \right)^{1/q} \ge C_q \left(\sum_i \|u_i\|^q \right)^{1/q}.$$

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

Remark 3.1. *Suppose $\mathcal{D}$ is (K, K)-unconditional with a constant U. Assume that X is of cotype q with a constant C_q. Then $\mathcal{D}$ has the Nikol'skii-type $\ell_1 X$ property with parameters $K, 1 - 1/q$ and $C_1 = 2UC_q^{-1}$.*

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

$$\| \sum_{i \in A} \pm x_i g_i \| = \| \sum_{i \in A_+} x_i g_i - \sum_{i \in A_-} x_i g_i \| \le 2U \| \sum_{i \in A} x_i g_i \|.$$

Therefore, by q-cotype assumption

$$\| \sum_{i \in A} x_i g_i \|^q \ge (2U)^{-q} C_q^q \sum_{i \in A} |x_i|^q.$$

16

This implies

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

$\square$

The above proof also gives the following individual function version of Remark 3.1.

Remark 3.2. *Suppose $f = \sum_{i \in T} x_i g_i$ has incoherence property with parameters D and U. Assume that X has cotype q with a constant C_q. Then f satisfies the Nikol'skii-type $\ell_1 X$ inequality with parameter $r = 1 - 1/q$ and $C_1 = 2UC_q^{-1}$.*

It is known that a Hilbert space has cotype 2. Therefore, Remark 3.2 shows that assumption **A2** implies assumption **A1** with $r = 1/2$. This explains how Theorem 1.3 is derived from Theorem 1.2.

We note that the (K, CK)-unconditionality assumption on the dictionary $\mathcal{D}$ in a Hilbert space H is somewhat weaker than the assumption $\mathcal{D} \in R(CK, \delta)$. Also, our theorems do not assume that the dictionary satisfies assumptions **A1** and **A2**; we only assume that the individual function f, a K-sparse approximation of a given f_0, satisfies **A1** and **A2**.

In assumption (3.2) we always have $U \geq 1$. In the extreme case $U = 1$ assumption (3.2) is a strong assumption that leads to strong results.

Proposition 3.1. *Let X be a uniformly smooth Banach space. Assume that $f = \sum_{i \in T} x_i g_i$, $|T| = K$, and the set of indices T has the following property. For any $g \in \mathcal{D}$ distinct from g_i, $i \in T$, and any c_i, c we have*

$$\| \sum_{i \in T} c_i g_i - cg\| \geq \| \sum_{i \in T} c_i g_i\|. \tag{3.3}$$

Then the WCGA with $t_k \neq 0$, $k = 1, 2, \ldots$, recovers f exactly after K iterations.

Proof. It is known (see, for instance, [10], Lemma 6.9, p. 342) that (3.3) implies

$$F_f(g) = 0, \qquad g \in \mathcal{D} \setminus \{g_i\}_{i \in T}.$$

Thus, at the first iteration the WCGA picks $\varphi_1 \in \{g_i\}_{i \in T}$. Then f_1 has the form $\sum_{i \in T} c_i g_i$ and we repeat the above argument. Then $\varphi_2 \in \{g_i\}_{i \in T} \setminus \{\varphi_1\}$. After K iterations all g_i, $i \in T$, will be taken and therefore we will have $f_K = 0$. $\qquad\square$

Proposition 3.1 can be applied in the following situation. Assume that $\Psi = \{\psi_i\}_{i=1}^{\infty}$ is a monotone basis for a uniformly smooth Banach space X. Then any $f = \sum_{i=1}^{K} x_i \psi_i$ will be recovered by the WCGA after K iterations. In particular, this applies to the Haar basis in L_p, $1 < p < \infty$.

We now proceed to main results of this section.

Theorem 3.1. *Let X be a Banach space with $\rho(u) \leq \gamma u^2$. Suppose for a given f_0 we have $\|f_0 - f^\epsilon\| \leq \epsilon$ with K-sparse $f := f^\epsilon$ satisfying **A1** and **A2**. Then for any $k \geq 0$ we have for $K + m \leq D$*

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

where $c_1 := \frac{t^2}{32\gamma C_1^2 U^2}$.

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_i 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 [10] we obtain

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

Thus

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

18

By (3.1) we get

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

Then

$$S_m \geq \frac{\|f_{m-1}\| - \epsilon}{C_1 K^r \|f_{A_m}\|}. \tag{3.4}$$

From the definition of the modulus of smoothness we have for any λ

$$\|f_{m-1} - \lambda \varphi_m\| + \|f_{m-1} + \lambda \varphi_m\| \leq 2\|f_{m-1}\| \left(1 + \rho\left(\frac{\lambda}{\|f_{m-1}\|}\right)\right) \tag{3.5}$$

and by (1) from the definition of the WCGA and Lemma 6.10 from [10], p. 343, we get

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

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

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

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

From here and from (3.5) we obtain

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

We discuss here the case $\rho(u) \leq \gamma u^2$. Using (3.4) we get

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

Let λ_1 be a solution of

$$\frac{\lambda t}{2 C_1 K^r \|f_{A_m}\|} = 2\gamma \frac{\lambda^2}{\|f_{m-1}\|^2}, \quad \lambda_1 = \frac{t \|f_{m-1}\|^2}{4\gamma C_1 K^r \|f_{A_m}\|}.$$

Our assumption (3.2) gives

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

19

Specify
$$\lambda = \frac{t\|f_{A_m}\|}{16\gamma C_1 K^r U^2}.$$

Then, using $\|f_{m-1}\| \geq \epsilon$ we get
$$\frac{\lambda}{\lambda_1} = \frac{\|f_{A_m}\|^2}{4\|f_{m-1}\|^2} \leq 1$$

and obtain
$$\|f_m\| \leq \|f_{m-1}\| \left(1 - \frac{t^2}{32\gamma C_1^2 U^2 K^{2r}}\right) + \frac{\epsilon t^2}{16\gamma C_1^2 U^2 K^{2r}}.$$

Denote $c_1 := \frac{t^2}{32\gamma C_1^2 U^2}$. Then
$$\|f_m\| \leq \|f_k\| \exp\left(-\frac{c_1(m-k)}{K^{2r}}\right) + 2\epsilon.$$

$\square$

Theorem 3.2. *Let X be a Banach space with $\rho(u) \leq \gamma u^2$. Suppose K-sparse f^ϵ satisfies **A1**, **A2** and $\|f_0 - f^\epsilon\| \leq \epsilon$. Then the WCGA with weakness parameter t applied to f_0 provides*

$$\|f_{C(t,\gamma,C_1)U^2 \ln(U+1)K^{2r}}\| \leq CU\epsilon \quad for \quad K + C(t,\gamma,C_1)U^2 \ln(U+1)K^{2r} \leq D$$

with an absolute constant C and $C(t,\gamma,C_1) = C_2\gamma C_1^2 t^{-2}$.

We formulate an immediate corollary of Theorem 3.2 with $\epsilon = 0$.

Corollary 3.1. *Let X be a Banach space with $\rho(u) \leq \gamma u^2$. Suppose K-sparse f satisfies **A1**, **A2**. Then the WCGA with weakness parameter t applied to f recovers it exactly after $C(t,\gamma,C_1)U^2 \ln(U+1)K^{2r}$ iterations under condition $K + C(t,\gamma,C_1)U^2 \ln(U+1)K^{2r} \leq D$.*

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 $j = 1, 2, \ldots, n, n+1$ 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

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

20

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

$$\|f_Q\| < \|f_{B_j}\| \quad \text{then} \quad |Q| < |\Gamma^k| - 2^{j-1}. \tag{3.6}$$

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

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

Then

$$\|f_{B_j}\| \leq b^{L-1-j}\|f_{B_{L-1}}\|, \quad j = 1, 2, \dots, L. \tag{3.7}$$

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 (3.1) we get

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

Then

$$S_m \geq \frac{\|f_{m-1}\| - \|f_B\| - \epsilon}{C_1|A|^r\|f_{A_m}\|}. \tag{3.8}$$

From the definition of the modulus of smoothness we have for any λ

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

and by (1) from the definition of the WCGA and Lemma 6.10 from [10], p. 343, we get

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

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

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 $\rho(u) \leq \gamma u^2$. Using (3.8) we get

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

Let λ_1 be a solution of

$$\frac{\lambda t}{2C_1|A|^r\|f_{A_m}\|} = 2\gamma\frac{\lambda^2}{\|f_{m-1}\|^2}, \qquad \lambda_1 = \frac{t\|f_{m-1}\|^2}{4\gamma C_1|A|^r\|f_{A_m}\|}.$$

Our assumption (3.2) gives

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

Specify

$$\lambda = \frac{t\|f_{A_m}\|}{16\gamma C_1|A|^r U^2}.$$

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

$$\|f_m\| \leq \|f_{m-1}\| \left(1 - \frac{t^2}{32\gamma C_1^2 U^2|A|^{2r}}\right) + \frac{t^2(\|f_B\| + \epsilon)}{16\gamma C_1^2|A|^{2r}U^2}. \qquad (3.9)$$

Denote $c_1 := \frac{t^2}{32\gamma C_1^2 U^2}$ and $c_2 := \frac{t^2}{16\gamma C_1^2 U^2}$. This implies for $m_2 > m_1 \geq k$

$$\|f_{m_2}\| \leq \|f_{m_1}\|(1 - c_1/|A|^{2r})^{m_2-m_1} + \frac{c_2(m_2 - m_1)}{|A|^{2r}}(\|f_B\| + \epsilon).$$

22

Define $m_0 := k$ and, inductively,

$$m_j = m_{j-1} + \beta|A_j|^{2r}, \quad j = 1, \ldots, n.$$

At iterations from $m_{j-1} + 1$ to m_j we use $A = A_j$ and obtain from (3.9) that

$$\|f_m\| \leq \|f_{m-1}\|(1-u) + 2u(\|f_B\| + \epsilon), \quad u := c_1|A|^{-2r}.$$

Using $1 - u \leq e^{-u}$ and $\sum_{k=0}^{\infty}(1-u)^k = 1/u$ we derive from here

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

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 Φ_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 Φ_k plus $\|f_0 - f\|$. This implies $\|f_k\| \leq \|f_{B_0}\| + \epsilon$. Therefore using (3.7) we continue

$$\leq (\|f_{B_0}\| + \epsilon)\eta^L + 2\sum_{j=1}^{L}(\|f_{B_{L-1}}\|(\eta b)^{L-j}b^{-1} + \epsilon\eta^{L-j})$$

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

We will specify β 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

$$\|f_{m_L}\| \leq \|f_{B_{L-1}}\|8e^{-c_1\beta} + 4\epsilon. \tag{3.10}$$

By (3.2) we get

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

If $\|f_{B_{L-1}}\| \leq 10U\epsilon$ then by (3.10)

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

23

If $\|f_{B_{L-1}}\| \geq 10U\epsilon$ then making β 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 (see (3.6))

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

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 $|\Gamma^{m_{L_1}}| < |T| - 2^{L_1-2}$, $m_{L_1} \leq \beta 2^{2rL_1}$. We continue the process and build a sequence m_{L_j} such that $m_{L_j} \leq \beta 2^{2rL_j}$ and after m_{L_j} iterations we reduce the support by at least 2^{L_j-2}. We also note that $m_{L_j} \leq \beta 2^{2r} K^{2r}$. We continue this process till the following inequality is satisfied for the first time

$$m_{L_1} + \cdots + m_{L_n} \geq 2^{4r} \beta K^{2r}. \tag{3.11}$$

Then, clearly,

$$m_{L_1} + \cdots + m_{L_n} \leq 2^{4r+1} \beta K^{2r}.$$

Using the inequality

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

we derive from (3.11)

$$2^{L_1-2} + \cdots + 2^{L_n-2} \geq \left(2^{2r(L_1-2)} + \cdots + 2^{2r(L_n-2)}\right)^{\frac{1}{2r}}$$

$$\geq 2^{-2} \left(2^{2rL_1} + \cdots + 2^{2rL_n}\right)^{\frac{1}{2r}}$$

$$\geq 2^{-2} \left((\beta)^{-1}(m_{L_1} + \cdots + m_{L_n})\right)^{\frac{1}{2r}} \geq K.$$

Thus, after not more than $N := 2^{4r+1} \beta K^{2r}$ iterations we recover f exactly and then $\|f_N\| \leq \|f_0 - f\| \leq \epsilon$.

□

Theorem 1.2 from the Introduction follows from Theorems 3.2 and 3.1.

4 Discussion

We begin with presenting some known results about exact recovery and the Lebesgue-type inequalities for incoherent dictionaries. In this case we use another natural generalization of the WOMP. This generalization of the WOMP was introduced in [9]. In the paper [5] we proved Lebesgue-type inequalities for that algorithm. We now formulate the corresponding results. We recall a generalization of the concept of M-coherent dictionary to the case of Banach spaces (see, for instance, [10]).

Let $\mathcal{D}$ be a dictionary in a Banach space X. The coherence parameter of this dictionary is defined as

$$M(\mathcal{D}) := \sup_{g \neq h; g,h \in \mathcal{D}} \sup_{F_g} |F_g(h)|.$$

In general, a norming functional F_g is not unique. This is why we take $\sup_{F_g}$ over all norming functionals of g in the definition of $M(\mathcal{D})$. We do not need $\sup_{F_g}$ in the definition of $M(\mathcal{D})$ if for each $g \in \mathcal{D}$ there is a unique norming functional $F_g \in X^*$. Then we define $\mathcal{D}^* := \{F_g, g \in \mathcal{D}\}$ and call $\mathcal{D}^*$ a *dual dictionary* to a dictionary $\mathcal{D}$. It is known that the uniqueness of the norming functional F_g is equivalent to the property that g is a point of Gateaux smoothness:

$$\lim_{u \to 0}(\|g + uy\| + \|g - uy\| - 2\|g\|)/u = 0$$

for any $y \in X$. In particular, if X is uniformly smooth then F_f is unique for any $f \neq 0$. We considered in [9] the following greedy algorithm which generalizes the Weak Orthogonal Greedy Algorithm to a Banach space setting.

Weak Quasi-Orthogonal Greedy Algorithm (WQOGA). Let $t \in (0,1]$ and f_0 be given. Find $\varphi_1 := \varphi_1^{q,t} \in \mathcal{D}$ (here and below index q stands for *quasi-orthogonal*) such that

$$|F_{\varphi_1}(f_0)| \geq t \sup_{g \in \mathcal{D}} |F_g(f_0)|.$$

Next, we find c_1 satisfying

$$F_{\varphi_1}(f - c_1\varphi_1) = 0.$$

Denote $f_1 := f_1^{q,t} := f - c_1\varphi_1$.

We continue this construction in an inductive way. Assume that we have already constructed residuals $f_0, f_1, \ldots, f_{m-1}$ and dictionary elements $\varphi_1, \ldots, \varphi_{m-1}$. Now, we pick an element $\varphi_m := \varphi_m^{q,t} \in \mathcal{D}$ such that

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

Next, we look for $c_1^m, \ldots, c_m^m$ satisfying

$$F_{\varphi_j}\left(f - \sum_{i=1}^m c_i^m \varphi_i\right) = 0, \quad j = 1, \ldots, m. \tag{4.1}$$

If there is no solution to (4.1) then we stop, otherwise we denote $G_m := G_m^{q,t} := \sum_{i=1}^m c_i^m \varphi_i$ and $f_m := f_m^{q,t} := f - G_m$ with $c_1^m, \ldots, c_m^m$ satisfying (4.1).

Remark 4.1. *Note that (4.1) has a unique solution if $\det(F_{\varphi_j}(\varphi_i))_{i,j=1}^m \neq 0$. Applying the WQOGA in the case of a dictionary with the coherence parameter $M := M(\mathcal{D})$ gives, by a simple well known argument on the linear independence of the rows of the matrix $(F_{\varphi_j}(\varphi_i))_{i,j=1}^m$, the conclusion that (4.1) has a unique solution for any $m < 1 + 1/M$. Thus, in the case of an M-coherent dictionary $\mathcal{D}$, we can run the WQOGA for at least $[1/M]$ iterations.*

In the case $t = 1$ we call the WQOGA the Quasi-Orthogonal Greedy Algorithm (QOGA). In the case of QOGA we need to make an extra assumption that the corresponding maximizer $\varphi_m \in \mathcal{D}$ exists. Clearly, it is the case when $\mathcal{D}$ is finite.

It was proved in [9] (see also [10], p. 382) that the WQOGA is as good as the WOMP in the sense of exact recovery of sparse signals with respect to incoherent dictionaries. The following result was obtained in [9].

Theorem 4.1. *Let $t \in (0, 1]$. Assume that $\mathcal{D}$ has coherence parameter M. Let $K < \frac{t}{1+t}(1 + 1/M)$. Then for any f_0 of the form*

$$f_0 = \sum_{i=1}^K a_i g_i,$$

where g_i are distinct elements of $\mathcal{D}$, the WQOGA recovers it exactly after K iterations. In other words, $f_K^{q,t} = 0$.

It is known (see [10], pp. 303–305) that the bound $K < \frac{1}{2}(1 + 1/M)$ is sharp for exact recovery by the OGA.

We introduce a new norm, associated with a dictionary $\mathcal{D}$, by the formula

$$\|f\|_{\mathcal{D}} := \sup_{g \in \mathcal{D}} |F_g(f)|, \quad f \in X.$$

We define best m-term approximation in the norm Y as follows

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

In [5] the norm Y was either the norm X of our Banach space or the norm $\|\cdot\|_{\mathcal{D}}$ defined above. The following two Lebesgue-type inequalities were proved in [5].

Theorem 4.2. *Assume that $\mathcal{D}$ is an M-coherent dictionary. Then for $m \leq \frac{1}{3M}$ we have for the QOGA*

$$\|f_m\|_{\mathcal{D}} \leq 13.5\sigma_m(f)_{\mathcal{D}}. \tag{4.2}$$

Theorem 4.3. *Assume that $\mathcal{D}$ is an M-coherent dictionary in a Banach space X. There exists an absolute constant C such that, for $m \leq 1/(3M)$, we have for the QOGA*

$$\|f_m\|_X \leq C \inf_{g \in \Sigma_m(\mathcal{D})} (\|f - g\|_X + m\|f - g\|_{\mathcal{D}}).$$

Corollary 4.1. *Using the inequality $\|g\|_{\mathcal{D}} \leq \|g\|_X$, Theorem 4.3 obtains*

$$\|f_m\|_X \leq C(1 + m)\sigma_m(f)_X.$$

Inequality (4.2) is a perfect (up to a constant 13.5) Lebesgue-type inequality. It indicates that the norm $\|\cdot\|_{\mathcal{D}}$ used in [5] is a suitable norm for analyzing performance of the QOGA. Corollary 4.1 shows that the Lebesgue-type inequality (4.2) in the norm $\|\cdot\|_{\mathcal{D}}$ implies the Lebesgue-type inequality in the norm $\|\cdot\|_X$.

Thus, results of this paper complement the above discussed results from [9] and [5]. Results from [9] and [5] deal with incoherent dictionaries and use the QOGA for exact recovery and the Lebesgue-type inequalities. Results of this paper deal with dictionaries which satisfy assumptions **A1** and **A2** and we analyze the WCGA here. In the case of a Hilbert space, assumptions **A1**

and **A2** are satisfied if $\mathcal{D}$ has RIP. It is well known that the RIP condition is much weaker than the incoherence condition in the case of a Hilbert space. It is interesting to note that we do not know how the coherence parameter $M(\mathcal{D})$ is related to properties **A1** and **A2** in the case of a Banach space.

We now give a few applications of Theorem 1.2 for specific dictionaries $\mathcal{D}$. We begin with the case when $\mathcal{D}$ is a basis Ψ for X. In some of our examples we take $X = L_p$, $2 \le p < \infty$. Then it is known that $\rho(u) \le \gamma u^2$ with $\gamma = (p-1)/2$.

Example 1. Let X be a Banach space with $\rho(u) \le \gamma u^2$ and with cotype q. Let Ψ be a normalized in X unconditional basis for X. Then $U \le C(X, \Psi)$. By Remark 3.1 Ψ satisfies **A1** with $r = 1 - \frac{1}{q}$. Theorem 1.2 gives

$$\|f_{C(t,X,\Psi)K^{2-2/q}}\| \le C\sigma_K(f_0, \Psi). \tag{4.3}$$

We note that (4.3) provides some progress in Open Problem 7.1 (p. 91) from [8].

Example 2. Let Ψ be a uniformly bounded orthogonal system normalized in $L_p(\Omega)$, $2 \le p < \infty$, Ω is a bounded domain. Then we can take $r = 1/2$. The inequality

$$\|g\|_p \le CK^{1/2-1/p}\|g\|_2$$

for K-sparse g implies that

$$\|S_A(f)\|_p \le CK^{1/2-1/p}\|S_A(f)\|_2 \le CK^{1/2-1/p}\|f\|_2 \le CK^{1/2-1/p}\|f\|_p.$$

Therefore $U \le CK^{1/2-1/p}$. Theorem 1.2 gives

$$\|f_{C(t,p,D)K^{2/p'}\ln K}\|_p \le C\sigma_K(f_0, \Psi)_p. \tag{4.4}$$

Inequality (4.4) provides some progress in Open Problem 7.2 (p. 91) from [8].

Theorem 1.2 can also be applied for quasi-greedy bases and other greedy-type bases (see [10]). We plan to discuss these applications in detail in our future work.

In this paper we limit ourselves to the case of Banach spaces satisfying the condition $\rho(u) \le \gamma u^2$. In particular, as we mentioned above the L_p spaces with $2 \le p < \infty$ satisfy this condition. Clearly, the L_p spaces with $1 < p \le 2$ are also of interest. For the clarity of presentation we do not discuss the case $\rho(u) \le \gamma u^q$ in this paper. The technique from Section 3 works in this case too and we will present the corresponding results in our future work.

References

[1] E.J. Candès and T. Tao, Decoding by linear programming, IEEE Transactions on Information Theory, **51:12** (2005), 4203–4215.

[2] T.T. Cai, L.Wang, and G. Xu, Shifting inequality and recovery for sparse signals, IEEE Transactions on Information Theory, **57:7** (2011), 4689–4708.

[3] S. Foucart, Sparse recovery algorithms: sufficient conditions in terms of restricted isometry constants, Approximation Theory XIII: San Antonio, 2010, 65–77, 2012.

[4] E.D. Livshitz, On the optimality of the Orthogonal Greedy Algorithm for μ-coherent dictionaries, J. Approx. Theory, **164:5** (2012), 668–681.

[5] D. Savu and V.N. Temlyakov, Lebesgue-Type Inequalities for Greedy Approximation in Banach Spaces, IEEE Transactions on Information Theory, **58** (2013), 1098–1106.

[6] V.N. Temlyakov, Weak greedy algorithms. Adv. Comput. Math. **12**, 213–227 (2000).

[7] V.N. Temlyakov, Greedy algorithms in Banach spaces, Adv. Comput. Math., **14** (2001), 277–292.

[8] V.N. Temlyakov, Nonlinear Methods of Approximation, Found. Comput. Math., **3** (2003), 33–107.

[9] V.N. Temlyakov, Greedy Approximations, Foundations of Computational Mathematics, Santander 2005, London Mathematical Society Lecture Notes Series, **331**, Cambridge University Press, 371–394 (2006).

[10] V.N. Temlyakov, Greedy approximation, Cambridge University Press, 2011.

[11] J. Wang and B. Shim, Improved Recovery Bounds of Orthogonal Matching Pursuit using Restricted Isometry Property, arXiv:1211.4293v1 [cs.IT] 19 Nov 2012.

[12] T. Zhang, Sparse Recovery with Orthogonal Matching Pursuit under RIP, IEEE Transactions on Information Theory, **57** (2011), 6215–6221.