INTERDISCIPLINARY
MATHEMATICS
INSTITUTE

2012:01

Lebesgue-type inequalities for greedy approximation in Banach spaces

D. Savu and V. Temlyakov

IMI

PREPRINT SERIES

COLLEGE OF ARTS AND SCIENCES
UNIVERSITY OF SOUTH CAROLINA

Lebesgue-type inequalities for greedy approximation in Banach spaces*

D. Savu †and V.N. Temlyakov ‡

January 19, 2012

Abstract

We study sparse representations and sparse approximations with respect to incoherent dictionaries. We address the problem of designing and analyzing greedy methods of approximation. A key question in this regard is: How to measure efficiency of a specific algorithm? Answering this question we prove the Lebesgue-type inequalities for algorithms under consideration. A very important new ingredient of the paper is that we perform our analysis in a Banach space instead of a Hilbert space. It is known that in many numerical problems users are satisfied with a Hilbert space setting and do not consider a more general setting in a Banach space. There are known arguments that justify interest in Banach spaces. In this paper we give one more argument in favor of consideration of greedy approximation in Banach spaces. We introduce a concept of M-coherent dictionary in a Banach space which is a generalization of the corresponding concept in a Hilbert space. We analyze the Quasi-Orthogonal Greedy Algorithm (QOGA), which is a generalization of the Orthogonal Greedy Algorithm (Orthogonal Matching Pursuit) for Banach spaces. It is known that the QOGA recovers exactly S-sparse signals after S iterations

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

† Dept. of Mathematics, 1523 Greene Street, University of South Carolina, Columbia, SC 29208, USA e-mail:

‡ Dept. of Mathematics, 1523 Greene Street, University of South Carolina, Columbia, SC 29208, USA e-mail: temlyakovv@gmail.com

provided $S < (1+1/M)/2$. This result is well known for the Orthogonal Greedy Algorithm in Hilbert spaces. The following question is of great importance: Are there dictionaries in $\mathbb{R}^n$ such that their coherence in ℓ_p^n is less than their coherence in ℓ_2^n for some $p \in (1,\infty)$? We show that the answer to the above question is "yes". Thus, for such dictionaries, replacing the Hilbert space ℓ_2^n by a Banach space ℓ_p^n we improve an upper bound for sparsity that guarantees an exact recovery of a signal.

1 Introduction

We study sparse representations and sparse approximations with respect to incoherent dictionaries. Sparse representations of a function are not only a powerful analytic tool but they are utilized in many application areas such as image/signal processing and numerical computation. The backbone of finding sparse representations is the concept of m-term approximation of the target function by the elements of a given system of functions (dictionary). Since the elements of the dictionary used in the m-term approximation may be adapted to the function being approximated, this type of approximation is very efficient when approximants can be found. We address the problem of designing and analyzing greedy methods of approximation. A key question in this regard is: How to measure efficiency of a specific algorithm? Answering this question we follow the pattern introduced in [1] and prove the Lebesgue-type inequalities for algorithms under consideration. A very important new ingredient of the paper is that we perform our analysis in a Banach space instead of a Hilbert space. It is known that in many numerical problems users are satisfied with a Hilbert space setting and do not consider a more general setting in a Banach space. There are known arguments (see [11], p. xiii) that justify interest in Banach spaces. The first argument is an *a priori* argument that the spaces L_p are very natural and should be studied along with the L_2 space. The second argument is an *a posteriori* argument. The study of greedy approximation in Banach spaces has discovered that the characteristic of a Banach space X that governs the behavior of greedy approximation is the *modulus of smoothness* $\rho(u)$ of X. It is known that the spaces L_p, $2 \le p < \infty$ have modulo of smoothness of the same order: u^2. Thus, many results that are known for the Hilbert space L_2 and proved using some special structure of a Hilbert space can be generalized to Banach spaces L_p, $2 \le p < \infty$. The new proofs use only the geometry of the unit sphere of the space expressed in the form $\rho(u) \le \gamma u^2$ (see [11]).

In this paper we give one more argument in favor of consideration of greedy approximation in Banach spaces. We introduce a concept of M-coherent dictionary in a Banach space which is a generalization of the corresponding concept in a Hilbert space. We analyze the Quasi-Orthogonal Greedy Algorithm (QOGA), which is a generalization of the Orthogonal Greedy Algorithm (Orthogonal Matching Pursuit) for Banach spaces. It is known (see [10]) that the QOGA recovers exactly S-sparse signals after S iterations provided $S < (1 + 1/M)/2$. This result is well known for the Orthogonal Greedy Algorithm in Hilbert spaces. The following question is of great importance: Are there dictionaries in $\mathbb{R}^n$ such that their coherence in ℓ_p^n is less than their coherence in ℓ_2^n for some $p \in (1, \infty)$? In Section 3 we show that the answer to the above question is "yes". Thus, for such dictionaries, replacing the Hilbert space ℓ_2^n by a Banach space ℓ_p^n we improve an upper bound for sparsity that guarantees an exact recovery of a signal. We note that the computational complexity of the QOGA in the case of $\ell_p^n, 1 < p < \infty$, is close to that of the OGA in the case of ℓ_2^n because we can write the formula for the norming functional explicitly.

We now proceed to a detailed introduction and begin with an explanation of the concept of *Lebesgue-type inequality.*

Lebesgue [4] proved the following inequality: for any 2π-periodic continuous function f we have

$$\|f - S_n(f)\|_\infty \le (4 + \frac{4}{\pi^2} \ln n) E_n(f)_\infty, \tag{1.1}$$

where $S_n(f)$ is the nth partial sum of the Fourier series of f and $E_n(f)_\infty$ is the error of the best approximation of f by the trigonometric polynomials of order n in the uniform norm $\|\cdot\|_\infty$. The inequality (1.1) relates the error of a particular method (S_n) of approximation by the trigonometric polynomials of order n to the best-possible error $E_n(f)_\infty$ of approximation by the trigonometric polynomials of order n. By the Lebesgue-type inequality we mean an inequality that provides an upper estimate for the error of a particular method of approximation of f by elements of a special form, say, form $\mathcal{A}$, by the best-possible approximation of f by elements of the form $\mathcal{A}$. In the case of approximation with regard to bases (or minimal systems), the Lebesgue-type inequalities are known both in linear and in nonlinear settings (see surveys [3], [8] and [9]).

We begin our discussion with the Orthogonal Greedy Algorithm (OGA) in a Hilbert space. It is natural to compare performance of the OGA with the best m-term approximation with regard to a dictionary $\mathcal{D}$. We let $\Sigma_m(\mathcal{D})$ denote the collection of all functions (elements) in H which can be expressed as a linear combination of at most m elements of $\mathcal{D}$. Thus, each function $s \in \Sigma_m(\mathcal{D})$ can be written in the form

$$s = \sum_{g \in \Lambda} c_g g, \quad \Lambda \subset \mathcal{D}, \quad \#\Lambda \le m,$$

where the c_g are real or complex numbers. In some cases, it may be possible to write an element from $\Sigma_m(\mathcal{D})$ in this form in more than one way. The space $\Sigma_m(\mathcal{D})$ is not linear: the sum of two functions from $\Sigma_m(\mathcal{D})$ is generally not in $\Sigma_m(\mathcal{D})$.

For a function $f \in H$ we define its best m-term approximation error

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

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 $\operatorname{span} \mathcal{D}$ is H. Sometimes it will be convenient for us to consider along with $\mathcal{D}$ the symmetrized dictionary $\mathcal{D}^{\pm} := \{\pm g, g \in \mathcal{D}\}$. Let

$$M(\mathcal{D}) := \sup_{\substack{\varphi \neq \psi \\ \varphi, \psi \in \mathcal{D}}} |\langle \varphi, \psi \rangle|$$

be the coherence parameter of dictionary $\mathcal{D}$. Let a sequence $\tau = \{t_k\}_{k=1}^{\infty}$, $0 \le t_k \le 1$, be given. The following greedy algorithm was defined in [7].

Weak Orthogonal Greedy Algorithm (WOGA). We define $f_0^{o,\tau} := f$. Then for each $m \ge 1$ we inductively define:

(1) $\varphi_m^{o,\tau} \in \mathcal{D}$ is any element satisfying

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

(2) Let $H_m^{\tau} := \operatorname{span}(\varphi_1^{o,\tau}, \dots, \varphi_m^{o,\tau})$ and let $P_{H_m^{\tau}}(f)$ denote an operator of orthogonal projection onto H_m^{τ}. Define

$$G_m^{o,\tau}(f, \mathcal{D}) := P_{H_m^{\tau}}(f).$$

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

$$f_m^{o,\tau} := f - G_m^{o,\tau}(f, \mathcal{D}).$$

In the case $t_k = 1$, $k = 1, 2, \dots$, WOGA is called the Orthogonal Greedy Algorithm (OGA). Denote by $A_1(\mathcal{D})$ the closure of the convex hull of $\mathcal{D}^{\pm}$.

The first general Lebesgue-type inequality for the OGA for the M-coherent dictionary was obtained in [2]. They proved that

$$\|f_m^o\| \le 8m^{1/2}\sigma_m(f) \quad \text{for} \quad m < 1/(32M).$$

The constants in this inequality were improved in [12]:

$$\|f_m^o\| \le (1 + 6m)^{1/2}\sigma_m(f) \quad \text{for} \quad m < 1/(3M). \tag{1.2}$$

The following inequalities were proved in [1].

Theorem 1.1. *Let a dictionary $\mathcal{D}$ have the mutual coherence $M = M(\mathcal{D})$. Assume $m \le 0.05M^{-2/3}$. Then, for $l \ge 1$ satisfying $2^l \le \log m$, we have*

$$\|f_{m(2^l-1)}^o\| \le 6m^{2^{-l}}\sigma_m(f).$$

Corollary 1.1. *Let a dictionary $\mathcal{D}$ have the mutual coherence $M = M(\mathcal{D})$. Assume $m \le 0.05M^{-2/3}$. Then we have*

$$\|f_{[m\log m]}^o\| \le 24\sigma_m(f).$$

It was pointed out in [1] that the inequality $\|f_{[m\log m]}^o\| \le 24\sigma_m(f)$ from the above Corollary is almost (up to a $\log m$ factor) perfect Lebesgue-type inequality. However, we are paying a big price for it in the sense of a strong assumption on m. It was mentioned in [1] that it was not known if the assumption $m \le 0.05M^{-2/3}$ can be substantially weakened. It was shown in [13] that the assumption $m \le 0.05M^{-2/3}$ can be substantially weakened.

Theorem 1.2. *Let a dictionary $\mathcal{D}$ have the mutual coherence $M = M(\mathcal{D})$. For any $\delta \in (0, 1/4]$ set $L(\delta) := [1/\delta] + 1$. Assume m is such that $20Mm^{1+\delta}2^{L(\delta)} \le 1$. Then we have*

$$\|f_{m(2^{L(\delta)+1}-1)}^o\| \le \sqrt{3}\sigma_m(f).$$

Very recently E. Livshitz [5] improved the above Lebesgue-type inequality. He proved that

$$\|f^o_{2m}\| \le 3\sigma_m(f)$$

for $m \le (20M)^{-1}$. The proof in [5] is different from the proof of Theorem 1.2 from [13]. It is more technically involved.

In this paper we prove Lebesgue-type inequalities for greedy approximation in Banach spaces. 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 $\operatorname{span} \mathcal{D}$ is X. In this paper we study greedy algorithms with regard to $\mathcal{D}$. 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 Hahn-Banach theorem. 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 present here a generalization of the concept of M-coherent dictionary to the case of Banach spaces.

Let $\mathcal{D}$ be a dictionary in a Banach space X. We define the coherence parameter of this dictionary in the following way

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

We note that, 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 [10] the following greedy algorithm which generalizes the Weak Orthogonal Greedy Algorithm to a Banach space setting.

Weak Quasi-Orthogonal Greedy Algorithm (WQOGA). Let $t \in (0,1]$. Denote $f_0 := f_0^{q,t} := f$ (here and below index q stands for *quasi-orthogonal*) and find $\varphi_1 := \varphi_1^{q,t} \in \mathcal{D}$ 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 $f_0, f_1, \dots, f_{m-1}$ and dictionary elements $\varphi_1, \dots, \varphi_{m-1}$. 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 $c_1^m, \dots, c_m^m$ satisfying

$$F_{\varphi_j}(f - \sum_{i=1}^{m} c_i^m \varphi_i) = 0, \quad j = 1, \dots, m. \tag{1.3}$$

If there is no solution to (1.3) 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 $c_1^m, \dots, c_m^m$ satisfying (1.3).

Remark 1.1. *We note that (1.3) has a unique solution if* $\det(F_{\varphi_j}(\varphi_i))_{i,j=1}^m \neq 0$. *We apply the WQOGA in the case of a dictionary with the coherence parameter $M := M(\mathcal{D})$. Then, by a simple well known argument on the linear independence of the rows of the matrix $(F_{\varphi_j}(\varphi_i))_{i,j=1}^m$, we conclude that (1.3) has a unique solution for any $m < 1 + 1/M$. In particular, this follows from a simple Lemma 2.1 that we will also need later on. 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 [10] (see also [11], p. 382) that the WQOGA is as good as the WOGA in the sense of exact recovery of sparse signals with respect to incoherent dictionaries. The following result was obtained in [10].

Theorem 1.3. *Let $t \in (0,1]$. Assume that $\mathcal{D}$ has coherence parameter M. Let $S < \frac{t}{1+t}(1+1/M)$. Then for any f of the form*

$$f = \sum_{i=1}^{S} a_i \psi_i, \tag{1.4}$$

where ψ_i are distinct elements of $\mathcal{D}$, the WQOGA recovers it exactly after S iterations. In other words we have that $f_S^{q,t} = 0$.

It is known (see [13] and [11], pp. 303–305) that the bound $S < \frac{1}{2}(1 + 1/M)$ is sharp for exact recovery by the OGA.

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 this paper the norm Y will be either the norm X of our Banach space or the norm $\|\cdot\|_{\mathcal{D}}$ defined above. In Section 2 we prove the following two Lebesgue-type inequalities.

Theorem 1.4. *Assume that $\mathcal{D}$ is an M-coherent dictionary. Then for $m \le 1/(3M)$ we have for the QOGA*

$$\|f_m\|_{\mathcal{D}} \le 13.5\sigma_m(f)_{\mathcal{D}}. \tag{1.5}$$

Theorem 1.5. *Assume that $\mathcal{D}$ is an M-coherent dictionary in a Banach space X. There exists an absolute constant C such that for $m \le 1/(3M)$ we have for the QOGA*

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

Corollary 1.2. *Using the inequality $\|g\|_{\mathcal{D}} \le \|g\|_X$ we obtain from Theorem 1.5*

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

Inequality (1.5) is a perfect (up to a constant 13.5) Lebesgue-type inequality. It indicates that the norm $\|\cdot\|_{\mathcal{D}}$ used in this paper is a suitable norm for analyzing performance of the QOGA. Corollary 1.2 shows that the Lebesgue-type inequality (1.5) in the norm $\|\cdot\|_{\mathcal{D}}$ implies the Lebesgue-type inequality in the norm $\|\cdot\|_X$.

The paper is a follow up to results of the dissertation of the first author [6].

2 Lebesgue-type inequalities for the WQOGA

In this section we analyze the Weak Quasi-Orthogonal Greedy Algorithm. It is clear from the definition of this algorithm (see (1.3)) that we need to solve linear systems in running this algorithm. We begin this section with a simple lemma about linear systems that we will use later on.

Lemma 2.1. *Let $P = (p_{i,j})_{i,j=1}^{k+1}$ be a $(k+1) \times (k+1)$ matrix such that $p_{i,i} = 1$, $i = 1, \dots, k+1$, and $|p_{i,j}| \le M < 1/k$, $i \ne j$. Then the inverse matrix P^{-1} exists and for the solution of the equation $Pa = b$ we have*

$$\|a\|_1 \le (1 - kM)^{-1} \|b\|_1.$$

Proof. Let $a = (a_1, \dots, a_{k+1})$ and $b = (b_1, \dots, b_{k+1})$. Then we have

$$\sum_{j=1}^{k+1} p_{i,j} a_j = b_i, \qquad i = 1, \dots, k+1.$$

Multiplying these equations by $\operatorname{sign}(a_i)$ and summing up, we obtain

$$\sum_{i=1}^{k+1} (1 - kM)|a_i| \le \sum_{i=1}^{k+1} \sum_{j=1}^{k+1} p_{i,j} a_j \operatorname{sign}(a_i) \le \|b\|_1$$

which gives the required inequality. $\square$

We now proceed to analysis of the Quasi-Orthogonal Greedy Algorithm. First, we address the issue of existence of an element φ_1 at the first step of the QOGA. We note that the argument from [10] proves Proposition 2.1.

Proposition 2.1. *Let $\mathcal{D}$ be an M-coherent dictionary. Then for any f of the form (1.4) with $S < \frac{1}{2}(1 + M^{-1})$ there exists $\varphi \in \mathcal{D}$ such that*

$$|F_\varphi(f)| = \sup_{g \in \mathcal{D}} |F_g(f)| = \sup_{1 \le j \le S} |F_{\psi_j}(f)|.$$

Second, we address the following issue. In the case of a finite dictionary $\mathcal{D}$ it is clear that the conditions $\sigma_m(f)_\mathcal{D} = 0$ and $f \in \Sigma_m(\mathcal{D})$ are equivalent. We extend this property of finite dictionaries to infinite incoherent dictionaries.

Theorem 2.1. *Let $\mathcal{D}$ be an M-coherent dictionary. If, for $m < \frac{1}{2}(1 + M^{-1})$, $\sigma_m(f)_\mathcal{D} = 0$ then $f \in \Sigma_m(\mathcal{D})$.*

Proof. The proof is by induction. First, consider the case $m = 1$. Let $\sigma_1(f)_{\mathcal{D}} = 0$. If $f = 0$ then $f \in \Sigma_1(\mathcal{D})$. Suppose that $f \neq 0$. Take $\epsilon > 0$ and find $\varphi_1 \in \mathcal{D}$ and a coefficient b_1 such that

$$\|f - b_1\varphi_1\|_{\mathcal{D}} \le \epsilon.$$

We prove that for sufficiently small ϵ the inequality

$$\|f - b_2\varphi_2\|_{\mathcal{D}} \le \epsilon, \quad \varphi_2 \in \mathcal{D},$$

implies that $\varphi_2 = \varphi_1$. Indeed, assuming the contrary $\varphi_2 \neq \varphi_1$ we obtain that

$$\|b_1\varphi_1 - b_2\varphi_2\|_{\mathcal{D}} \le 2\epsilon. \tag{2.1}$$

Therefore

$$|b_1 - b_2 F_{\varphi_1}(\varphi_2)| \le 2\epsilon, \quad |b_1 F_{\varphi_2}(\varphi_1) - b_2| \le 2\epsilon,$$

which implies by Lemma 2.1 that $b_1 \to 0$ and $b_2 \to 0$ with $\epsilon \to 0$. On the other hand

$$|b_1| \ge \|f\|_{\mathcal{D}} - \epsilon, \quad |b_2| \ge \|f\|_{\mathcal{D}} - \epsilon. \tag{2.2}$$

It is clear that (2.1) and (2.2) contradict each other if ϵ is small enough compared to $\|f\|_{\mathcal{D}}$. Thus, for sufficiently small ϵ only one dictionary element φ_1 provides good approximation. This and the assumption $\sigma_1(f)_{\mathcal{D}} = 0$ imply $f = b\varphi_1$.

Consider the case $m > 1$. Following the induction argument assume that if $\sigma_{m-1}(f)_{\mathcal{D}} = 0$ then $f \in \Sigma_{m-1}(\mathcal{D})$. We now have $\sigma_m(f)_{\mathcal{D}} = 0$. If $\sigma_{m-1}(f)_{\mathcal{D}} = 0$ then by the induction assumption $f \in \Sigma_{m-1}$. So, assume that $\sigma_{m-1}(f)_{\mathcal{D}} > 0$. Let

$$\|f - \sum_{i=1}^{m} b_i\varphi_i\|_{\mathcal{D}} \le \epsilon, \quad \varphi_i \in \mathcal{D}, \quad i = 1, \dots, m. \tag{2.3}$$

As in the case $m = 1$ suppose that there are $\psi_1, \dots, \psi_m$ from the dictionary such that at least one of φ_i, say, φ_m is distinct from them and

$$\|f - \sum_{i=1}^{m} c_j\psi_j\|_{\mathcal{D}} \le \epsilon, \quad \psi_i \in \mathcal{D}, \quad i = 1, \dots, m. \tag{2.4}$$

Inequalities (2.3) and (2.4) imply

$$\|\sum_{i=1}^{m} b_i\varphi_i - \sum_{i=1}^{m} c_j\psi_j\|_{\mathcal{D}} \le 2\epsilon.$$

As above by Lemma 2.1 we obtain from here

$$b_m \to 0 \quad \text{as} \quad \epsilon \to 0. \tag{2.5}$$

Therefore, for small enough ϵ we get

$$\|f - \sum_{i=1}^{m-1} b_i\varphi_i\|_{\mathcal{D}} \le \epsilon + |b_m| < \sigma_{m-1}(f)_{\mathcal{D}}.$$

The obtained contradiction implies that for small enough ϵ inequalities (2.3) and (2.4) imply that $\{\psi_1,\dots,\psi_m\} = \{\varphi_1,\dots,\varphi_m\}$. This reduces the problem to a finite dimensional case, where existence of best approximation is well known. $\square$

We first present analysis of the QOGA. Analysis of the WQOGA goes exactly the same lines as analysis of the QOGA with constants depending on t. We will not repeat this analysis and will only formulate the corresponding results for the WQOGA. Let $\Psi_m := \{\psi_j\}_{j=1}^m$ be a set of m distinct elements of the dictionary $\mathcal{D}$. For an $f \in X$ define a quasi-orthogonal projection $P^q_{\Psi_m}(f)$ of f onto $\operatorname{span}(\psi_1,\dots,\psi_m)$ as an element

$$P^q_{\Psi_m}(f) = \sum_{i=1}^m c_i^m \psi_i$$

satisfying

$$F_{\psi_j}(f - \sum_{i=1}^m c_i^m\psi_i) = 0, \quad j = 1,\dots,m. \tag{2.6}$$

Lemma 2.1 guarantees that for an M-coherent dictionary the quasi-orthogonal projection $P^q_{\Psi_m}(f)$ exists for all f and Ψ_m provided $m \le 1/M$. We define projective best m-term approximation in the norm Y as follows

$$\sigma^q_m(f)_Y := \inf_{\Psi_m} \|f - P^q_{\Psi_m}(f)\|_Y.$$

In this paper the norm Y will be either the norm X of our Banach space or the norm $\|\cdot\|_{\mathcal{D}}$ defined above.

We begin with the key lemma that guarantees that under certain assumptions the QOGA picks good elements from the dictionary.

Lemma 2.2. *Assume that* $\sigma_m^q(f)_{\mathcal{D}} > 0$. *Suppose* $m \leq 1/(3M)$ *and*

$$\|f - P^q_{\Phi_m}(f)\|_{\mathcal{D}} \leq (1+\epsilon)\sigma_m^q(f)_{\mathcal{D}},$$
$$\Phi_m = \{\varphi_i\}_{i=1}^m, \quad \varphi_i \in \mathcal{D}, \quad i = 1, \dots, m, \tag{2.7}$$

with some fixed $\epsilon > 0$. *If*

$$\|f\|_{\mathcal{D}} > 3(1+\epsilon)\sigma_m^q(f)_{\mathcal{D}}$$

then the QOGA picks one of the φ_i, $i = 1, \dots, m$, *at the first iteration. If the QOGA has picked elements from* $\{\varphi_1, \dots, \varphi_m\}$ *at the first* $k < m$ *iterations and*

$$\|f_k\|_{\mathcal{D}} > 3(1+\epsilon)\sigma_m^q(f)_{\mathcal{D}} \tag{2.8}$$

then the QOGA picks one of the φ_i, $i = 1, \dots, m$, *at the* $(k+1)$*th iteration.*

Proof. Define

$$a_m := P^q_{\Phi_m}(f) = \sum_{i=1}^m c_i\varphi_i, \quad G_k := G_k(f, \mathcal{D}) = \sum_{i \in \Lambda_k} b_i \varphi_i,$$
$$\Lambda_k \subset \{1, \dots, m\}, \quad |\Lambda_k| = k.$$

We write

$$a_m - G_k = \sum_{i=1}^m d_i \varphi_i$$

and define $A := \max_{1 \leq i \leq m} |d_i|$. Then

$$F_{\varphi_i}(f_k) = F_{\varphi_i}(f - G_k) = F_{\varphi_i}(f - a_m) + F_{\varphi_i}(a_m - G_k) = F_{\varphi_i}(a_m - G_k)$$

and

$$\begin{aligned} y := \max_{1 \leq i \leq m} |F_{\varphi_i}(f_k)| &= \max_{1 \leq i \leq m} |F_{\varphi_i}(a_m - G_k)| \\ &\geq A(1 - M(m-1)) > A(1 - Mm). \end{aligned} \tag{2.9}$$

For any $g \in \mathcal{D}$ distinct from $\{\varphi_1, \dots, \varphi_m\}$ we obtain

$$|F_g(f_k)| = |F_g(f - a_m + a_m - G_k)| \leq (1+\epsilon)\sigma_m^q(f)_{\mathcal{D}} + AMm.$$

In order to prove our claim it is sufficient to have

$$((1+\epsilon)\sigma_m^q(f)_{\mathcal{D}} + AMm)/y < 1 \tag{2.10}$$

By inequality (2.9) inequality (2.10) follows from

$$y > (1+\epsilon)\sigma_m^q(f)_{\mathcal{D}}(1-Mm)(1-2Mm)^{-1}. \tag{2.11}$$

Proposition 2.1 gives

$$\begin{aligned} y &= \|a_m - G_k\|_{\mathcal{D}} = \|a_m - f + f - G_k\|_{\mathcal{D}} \\ &\geq \|f_k\|_{\mathcal{D}} - \|f - a_m\|_{\mathcal{D}} > 2(1+\epsilon)\sigma_m^q(f)_{\mathcal{D}}. \end{aligned} \tag{2.12}$$

Observing that under our assumption $Mm \leq 1/3$ we have

$$(1-Mm)(1-2Mm)^{-1} \leq 2,$$

we deduce (2.10) from (2.12).

□

Theorem 2.2. *Assume that $\mathcal{D}$ is an M-coherent dictionary. Then for $m \leq 1/(3M)$ we have for the QOGA*

$$\|f_m\|_{\mathcal{D}} \leq 4.5\sigma_m^q(f)_{\mathcal{D}}. \tag{2.13}$$

Proof. If $\sigma_m^q(f)_{\mathcal{D}} = 0$ then $\sigma_m(f)_{\mathcal{D}} = 0$ and by Lemma 2.2 we obtain that $f \in \Sigma_m(\mathcal{D})$. Then Theorem 1.3 guarantees that $f_m^q = 0$. This proves Theorem 2.2 in the case $\sigma_m^q(f)_{\mathcal{D}} = 0$.

Assume now that $\sigma_m^q(f)_{\mathcal{D}} > 0$. Then for any $\epsilon > 0$ there exists a collection

$$\Phi_m = \{\varphi_i\}_{i=1}^m, \quad \varphi_i \in \mathcal{D}, \quad i = 1, \dots, m,$$

such that

$$\|f - P_{\Phi_m}^q(f)\|_{\mathcal{D}} \leq (1+\epsilon)\sigma_m^q(f)_{\mathcal{D}}. \tag{2.14}$$

If

$$\|f_k\|_{\mathcal{D}} > 3(1+\epsilon)\sigma_m^q(f)\mathcal{D} \tag{2.15}$$

for all $k < m$ then by Lemma 2.2 the QOGA picks all the elements $\{\varphi_1, \dots, \varphi_m\}$ after m iterations. Then $f_m = f - P_{\Phi_m}^q(f)$ and (2.14) implies (2.13). Assume now that (2.15) does not hold for some $k < m$:

$$\|f_k\|_{\mathcal{D}} \leq 3(1+\epsilon)\sigma_m^q(f)\mathcal{D}. \tag{2.16}$$

We now need the following lemma.

Lemma 2.3. *Assume that $\mathcal{D}$ is an M-coherent dictionary. Then for $k+1 < 1/M$ we have for the WQOGA*

$$\|f_{k+1}\|_{\mathcal{D}} \leq \frac{1-(k-1)M}{1-kM}\|f_k\|_{\mathcal{D}}. \tag{2.17}$$

First, we use this lemma to complete the proof of Theorem 2.2. Inequalities (2.16) and (2.17) imply

$$\|f_m\|_{\mathcal{D}} \leq \frac{1-(k-1)M}{1-mM}\|f_k\|_{\mathcal{D}} \leq 4.5(1+\epsilon)\sigma_m^q(f)\mathcal{D}.$$

□

Second, we prove Lemma 2.3.

Proof. Consider

$$g_k := f_{k+1} - f_k = \sum_{i=1}^{k+1} a_i\psi_i, \quad \psi_i \in \mathcal{D}.$$

It follows from the definition of the WQOGA that

$$F_{\psi_i}(g_k) = 0, \quad i = 1, \dots, k;$$

$$|F_{\psi_{k+1}}(g_k)| = |F_{\psi_{k+1}}(f_k)| \leq \|f_k\|_{\mathcal{D}}.$$

Therefore, by Lemma 2.1

$$\|a\|_1 \leq (1-kM)^{-1}\|f_k^q\|_{\mathcal{D}}. \tag{2.18}$$

For any ψ_i, $i = 1, \dots, k+1$, we have

$$F_{\psi_i}(f_{k+1}) = 0$$

and for any $\psi \in \mathcal{D}$ distinct from ψ_i, $i = 1, \dots, k+1$, we get from (2.18)

$$|F_\psi(f_{k+1})| \leq |F_\psi(f_k)| + |F_\psi(g_k)| \leq \|f_k\|_{\mathcal{D}} + M\|a\|_1 \leq \frac{1-(k-1)M}{1-kM}\|f_k\|_{\mathcal{D}}.$$

□

We now formulate the corresponding results for the WQOGA. Lemma 2.2 take the following form.

Lemma 2.4. *Assume that $\sigma^q_m(f)_{\mathcal{D}} > 0$. Suppose $m \le \frac{2}{3}\frac{t}{1+t}\frac{1}{M}$ and*

$$\|f - P^q_{\Phi_m}(f)\|_{\mathcal{D}} \le (1+\epsilon)\sigma^q_m(f)_{\mathcal{D}},$$

$$\Phi_m = \{\varphi_i\}_{i=1}^m, \quad \varphi_i \in \mathcal{D}, \quad i = 1, \dots, m, \tag{2.19}$$

with some fixed $\epsilon > 0$. If

$$\|f\|_{\mathcal{D}} > C_t(1+\epsilon)\sigma^q_m(f)_{\mathcal{D}}, \quad C_t := 1 + \frac{3+t}{t(1+t)},$$

then the WQOGA picks one of the φ_i, $i = 1, \dots, m$, at the first iteration. If the WQOGA has picked elements from $\{\varphi_1, \dots, \varphi_m\}$ at the first $k < m$ iterations and

$$\|f_k\|_{\mathcal{D}} > C_t(1+\epsilon)\sigma^q_m(f)_{\mathcal{D}} \tag{2.20}$$

then the WQOGA picks one of the φ_i, $i = 1, \dots, m$, at the $(k+1)$th iteration.

Theorem 2.2 is transformed into the following theorem.

Theorem 2.3. *Assume that $\mathcal{D}$ is an M-coherent dictionary. Then for $m \le \frac{2}{3}\frac{t}{1+t}\frac{1}{M}$ we have for the WQOGA*

$$\|f_m\|_{\mathcal{D}} \le C_1(t)\sigma^q_m(f)_{\mathcal{D}}, \quad C_1(t) := \frac{3}{t} + \frac{1+t}{1+t/3}. \tag{2.21}$$

We show that the quantity $\sigma^q_m(f)_{\mathcal{D}}$ appearing in Theorems 2.2 and 2.3 can be replaced by $\sigma_m(f)_{\mathcal{D}}$. This follows from Lemma 2.5.

Lemma 2.5. *Assume that $\mathcal{D}$ is an M-coherent dictionary and $m \le 1/(3M)$. Then for any $f \in X$, any collection $\Phi_m = \{\varphi_1, \dots, \varphi_m\}$ of distinct elements of $\mathcal{D}$ and any $g \in \operatorname{span}\Phi_m$ we have*

$$\|f - P^q_{\Phi_m}(f)\|_{\mathcal{D}} \le 3\|f - g\|_{\mathcal{D}} \tag{2.22}$$

and, therefore,

$$\sigma^q_m(f)_{\mathcal{D}} \le 3\sigma_m(f)_{\mathcal{D}}. \tag{2.23}$$

Proof. We prove this lemma by contradiction. Assume that $g \in \operatorname{span}\Phi_m$ is such that

$$3\|f - g\|_{\mathcal{D}} < \|f - P^q_{\Phi_m}(f)\|_{\mathcal{D}} =: p.$$

Then

$$\|g - P^q_{\Phi_m}(f)\|_{\mathcal{D}} < p(1 + 1/3). \tag{2.24}$$

Let

$$g - P^q_{\Phi_m}(f) = \sum_{i=1}^{m} a_i \varphi_i.$$

Lemma 2.1 and (2.24) give

$$\|a\|_1 < 2pm. \tag{2.25}$$

For $\varphi \in \Phi_m$ we have

$$F_\varphi(f - P^q_{\Phi_m}(f)) = 0.$$

Thus it is sufficient to check $\varphi \notin \Phi_m$. We have for $\varphi \notin \Phi_m$

$$|F_\varphi(f - P^q_{\Phi_m}(f))| \le |F_\varphi(f-g)| + |F_\varphi(g - P^q_{\Phi_m}(f))| \le \|f-g\|_{\mathcal{D}} + \|a\|_1 M.$$

This implies by (2.25)

$$p \le \|f-g\|_{\mathcal{D}} + \|a\|_1 M < \|f-g\|_{\mathcal{D}} + 2pmM \le \|f-g\|_{\mathcal{D}} + 2p/3$$

and

$$p < 3\|f-g\|_{\mathcal{D}}.$$

The obtained contradiction proves Lemma 2.5. □

Theorem 1.4 from the Introduction follows from Theorem 2.2 and Lemma 2.5. Theorem 2.3 and Lemma 2.5 imply the following theorem for the WQOGA.

Theorem 2.4. *Assume that $\mathcal{D}$ is an M-coherent dictionary. Then for $m \le \frac{2}{3}\frac{t}{1+t}\frac{1}{M}$ we have for the WQOGA*

$$\|f_m\|_{\mathcal{D}} \le C_2(t)\sigma_m(f)_{\mathcal{D}}, \quad C_2(t) := 3\left(\frac{3}{t} + \frac{1+t}{1+t/3}\right). \tag{2.26}$$

We now proceed to a corollary of Theorem 2.2 for Lebesgue-type inequalities in the norm X. We begin with a simple lemma that relates different norms of n-term polynomials.

Lemma 2.6. *Assume that $\mathcal{D}$ is an M-coherent dictionary in a Banach space X. Then for f of the form*

$$f = \sum_{i=1}^{n} a_i \varphi_i, \quad n \le 1/M, \quad \varphi_i \in \mathcal{D}, \quad i = 1, \dots, n,$$

we have

$$\|f\|_X \le n(1-(n-1)M)^{-1}\|f\|_{\mathcal{D}}.$$

Proof. Clearly, $\|f\|_X \le \|a\|_1$, where $a = (a_1, \dots, a_n)$. We use a trivial inequality $\|a\|_1 \le n\|a\|_\infty$ and bound $\|a\|_\infty$. Suppose $\|a\|_\infty = |a_p|$ with some $p \in \{1, 2, \dots, n\}$. Then

$$|a_p|(1-(n-1)M) \le |F_{\varphi_p}(f)| \le \|f\|_{\mathcal{D}}.$$

Thus

$$\|a\|_\infty \le (1-(n-1)M)^{-1}\|f\|_{\mathcal{D}}$$

which completes the proof of Lemma 2.6. □

Remark 2.1. *It is known (see [1]) that in a Hilbert space under assumptions of Lemma 2.6 we have*

$$\|f\|_H \le (1+nM)^{1/2}\|a\|_2$$

and, therefore, as in Lemma 2.6 we get

$$\|f\|_H \le (1+nM)^{1/2}(1-(n-1)M)^{-1}\|f\|_{\mathcal{D}}.$$

Theorem 2.5. *Assume that $\mathcal{D}$ is an M-coherent dictionary in a Banach space X. For any $t \in (0,1]$ there exists a constant $C_3(t)$ that may depend only on t such that for $m \le \frac{2}{3}\frac{t}{1+t}\frac{1}{M}$ we have for the WQOGA*

$$\|f_m\|_X \le C_3(t) \inf_{g \in \Sigma_m(\mathcal{D})} (\|f-g\|_X + m\|f-g\|_{\mathcal{D}}).$$

Proof. Take an arbitrary $g_m \in \Sigma_m(\mathcal{D})$. Then

$$\|f_m\|_X = \|f-G_m\|_X \le \|f-g_m\|_X + \|g_m - G_m\|_X. \tag{2.27}$$

Further $g_m - G_m \in \Sigma_{2m}(\mathcal{D})$ and by Lemma 2.6

$$\|g_m - G_m\|_X \le C_4(t)m\|g_m - G_m\|_{\mathcal{D}}. \tag{2.28}$$

Next

$$\|g_m - G_m\|_{\mathcal{D}} \le \|f-g_m\|_{\mathcal{D}} + \|f-G_m\|_{\mathcal{D}}. \tag{2.29}$$

By Theorem 2.4 we get

$$\|f-G_m\|_{\mathcal{D}} \le C_2(t)\sigma_m(f)_{\mathcal{D}} \le C_2(t)\|f-g_m\|_{\mathcal{D}}. \tag{2.30}$$

Combining (2.27)–(2.30) we obtain

$$\|f_m\|_X \le \|f-g_m\|_X + C_5(t)m\|f-g_m\|_{\mathcal{D}}.$$

□

Corollary 2.1. *Using the inequality* $\|g\|_{\mathcal{D}} \leq \|g\|_X$ *we obtain from Theorem 2.5*

$$\|f_m\|_X \leq C_3(t)(1+m)\sigma_m(f)_X.$$

The proof of Theorem 2.5 and Remark 2.1 give the following variant of Theorem 2.5 in a Hilbert space.

Theorem 2.6. *Assume that* $\mathcal{D}$ *is an* M*-coherent dictionary in a Hilbert space* H*. Then for* $m \leq \frac{2}{3}\frac{t}{1+t}\frac{1}{M}$ *we have*

$$\|f_m\|_H \leq C_6(t) \inf_{g \in \Sigma_m(\mathcal{D})} (\|f-g\|_H + m^{1/2}\|f-g\|_{\mathcal{D}}).$$

3 Examples

The purpose of the following examples is to illustrate a possible use of the Quasi-Orthogonal Algorithms in approximations, in place of their Orthogonal correspondents. We will consider recovery in the finite dimensional space $\mathbb{R}^d$ and the associate Hilbert space $\ell_2(\mathbb{R}^d)$ or a Banach space $\ell_p(\mathbb{R}^d)$, where $1 \leq p \leq \infty,\ p \neq 2$.

We propose to replace the recovery with respect to a dictionary in ℓ_2 that we will denote D_2, with the recovery in ℓ_p, where the dictionary is D_p, the re-normed version of D_2.

While we have proved that the Quasi-Orthogonal Algorithms have the same power of exact recovery as their Orthogonal counterparts, the benefit will come from an improved geometry in the new space considered, namely the coherence parameter M_p of the dictionary D_p would be smaller than the coherence M_2 of the original dictionary.

It is possible, then, that we can guarantee either exact recovery of more terms in the re-normed dictionary, or a better approximation than in the original setting.

Example 3.1. *For a very simple example of this geometry change, let us consider the basis* $D_2 = \{(1/2, \sqrt{3}/2), (\sqrt{3}/2, 1/2)\}$ *in* $\mathbb{R}^2$*. The* 4*-norm of its elements is* $\sqrt[4]{10}/2$*, so*

$$D_4 = \left\{ \left(\frac{1}{\sqrt[4]{10}}, \frac{\sqrt{3}}{\sqrt[4]{10}} \right), \left(\frac{\sqrt{3}}{\sqrt[4]{10}}, \frac{1}{\sqrt[4]{10}} \right) \right\}.$$

The norming functionals for $p = 4$ are unique and they form a conjugate (dual) dictionary:

$$D_4^* = \left\{ \left(\left(\frac{1}{10}\right)^{\frac{3}{4}}, \left(\frac{9}{10}\right)^{\frac{3}{4}} \right), \left(\left(\frac{9}{10}\right)^{\frac{3}{4}}, \left(\frac{1}{10}\right)^{\frac{3}{4}} \right) \right\}.$$

The new geometry is characterized by the coherence parameter

$$M_4 = \frac{2\sqrt{3}}{5} \approx 0.692 < 0.865 \approx \frac{\sqrt{3}}{2} = M_2,$$

which shows an improvement over the original coherence, see Figure 1.

Adding two other vectors, each orthogonal on one of the original vectors of this basis produces a redundant dictionary with $N = 4$ elements that shows exactly the same improvement of coherence.

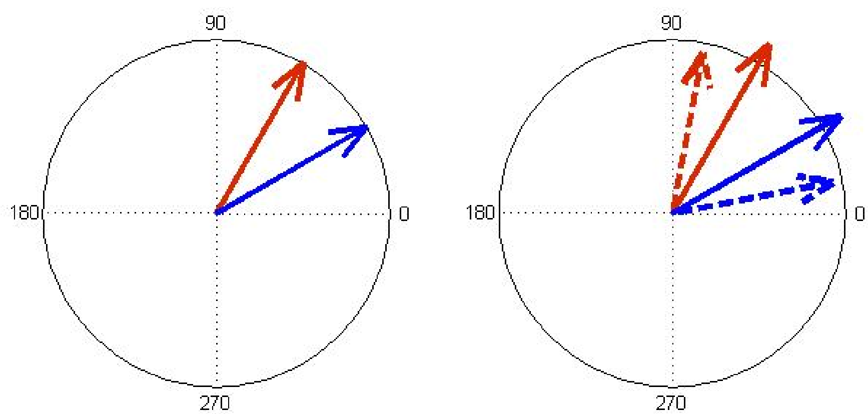

Figure 1: The vectors of the original dictionary in ℓ_2 with coherence $M_2 \approx 0.865$ are shown on the left side, and their re-normed versions in ℓ_4 with improved coherence $M_4 \approx 0.692$ are shown on the right side. The elements of the conjugate dictionary $\ell_{4/3}$ are represented with interrupted lines.

Example 3.2. *Another dictionary for $d = 2$, namely*

$$D_2 = \{(\cos(\pi/8), \sin(\pi/8)), (\cos(5\pi/8), \sin(5\pi/8)\}, (1,0), (0,1)\}$$

re-normed for $p = 3/2$ has a coherence improvement given by:

$$M_{3/2} \approx 0.886 < 0.930 \approx M_2,$$

see Figure 2.

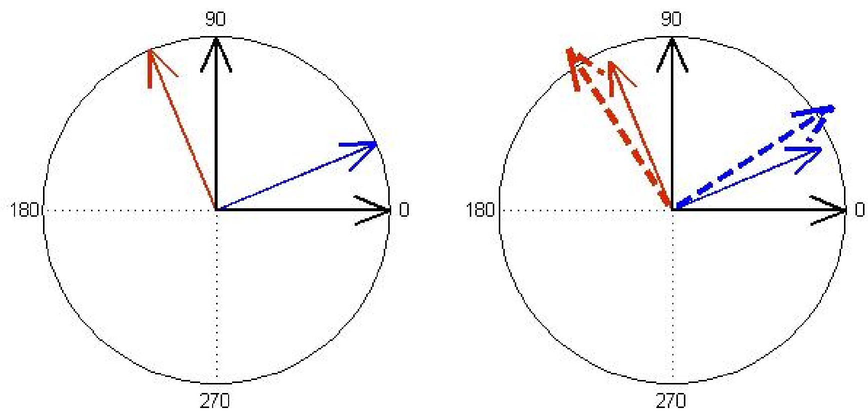

Figure 2: The vectors of the original dictionary in ℓ_2 with coherence $M_2 \approx 0.930$ are shown on the left side, and their re-normed versions in $\ell_{3/2}$ with improved coherence $M_{3/2} \approx 0.886$ are shown on the right side. The corresponding dual vectors in ℓ_3 are represented with interrupted lines.

We see that we can use either a $p > 2$ or a $p < 2$ to improve on the coherence parameter of a dictionary.

Example 3.3. *For a higher-dimension example, say $d = 16$, we consider a dictionary with $N = 2d$ vectors in which the first d come from the Discrete Cosine Transform (DCT), which is popular in designing filters in Signal Processing. Thus, our first d vectors are the column vectors of the square-matrix*

$$\left(\cos\left(\frac{\pi}{d}(k-1)(n-1/2)\right)\right)_{k=1,\dots,d}^{n=1,\dots,d}.$$

We obtain the other d vectors by Orthogonal Greedy Approximation with $m = \frac{5}{8}d = 10$ terms from the above DCT, of the elements of the canonical *basis e_i, for $i = 1, \ldots, d$, where $e_i(j) = \delta_{ij}$ (Kronecker delta). All the $2d$ vectors are normalized in ℓ_2 to obtain D_2.*

The improvement obtained by re-norming to D_3, when $d = 16$ is given by the coherence:

$$M_3 \approx 0.327 < 0.380 \approx M_2.$$

Both the OGA and QGA guarantee recovery of vectors with sparsity $S \leq \frac{1}{2}\left(\frac{1}{M} + 1\right)$. In this example, QGA would guarantee the recovery of vectors with sparsity $S = 2$, one more term than the classic OGA.

Example 3.4. *The same construction for $d = 64$, where the first d vectors in D_2 are the normalized vectors in ℓ_2 from DCT and the other d vectors are the $m = 40$-term Orthogonal Greedy approximants from DCT of the elements of the corresponding canonical basis, holds the following improvement of the geometry to D_3:*

$$M_3 \approx 0.221 < 0.312 \approx M_2.$$

These examples show the potential improvement of the geometry of dictionaries by switching the orthogonal recovery in the Hilbert setting to a quasi-orthogonal recovery in a Banach counterpart. This improvement would result in a greater range of sparsities to be recovered or possibly better m-term approximants using essentially the same vectors in the dictionary.

References

[1] Donoho, D.L., Elad, M., Temlyakov, V.N.: On Lebesgue-type inequalities for greedy approximation. J. Approx. Theory. **147**(2), 185–195 (2007)

[2] Gilbert, A.C., Muthukrishnan, S., Strauss, M.J.: Approximation of functions over redundant dictionaries using coherence. In Proc. 14th Annual ACM-SIAM Symposium on Discrete Algorithms. pp. 243–252 (2002)

[3] S.V. Konyagin and V.N. Temlyakov, Greedy Approximation with Regard to bases and General Minimal Systems, Serdica Math. J., **28**, 305–328 (2002).

[4] H. Lebesgue, Sur les intégrales singuliéres, Ann. Fac. Sci. Univ. Toulouse (3), **1**, 25–117 (1909).

[5] E.D. Livshitz, On the optimality of Orthogonal Greedy Algorithm for M-coherent dictionaries, Preprint, arXiv:1003.5349v1, 1–14 (2010).

[6] D. Savu, Sparse Approximation in Banach Spaces, Ph. D. Dissertation, University of South Carolina (2009).

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

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

[9] Temlyakov, V.N. Greedy approximation. Acta Numerica. pp. 235–409 (2008)

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

[11] V.N. Temlyakov, Greedy Approximation, Cambridge University Press, (2011).

[12] Tropp, J.: Greedy is good: Algorithmic results for sparse approximation. IEEE Trans. Inform. Theory. **50**(10), 2231–2242 (2004)

[13] V.N. Temlyakov and P. Zheltov, On performance of greedy algorithms, IMI-Preprint, **10:02**, 1–13 (2010).