

Article

On the Rate of Convergence of Greedy Algorithms

Vladimir Temlyakov ^{1,2,3,4}¹ Steklov Mathematical Institute of Russian Academy of Sciences, 117312 Moscow, Russia; temlyak@math.sc.edu² Department of Mechanics and Mathematics, Lomonosov Moscow State University, 119991 Moscow, Russia³ Moscow Center of Fundamental and Applied Mathematics, 119333 Moscow, Russia⁴ Department of Mathematics, University of South Carolina, Columbia, SC 29208, USA

Abstract: In this paper, a new criterion for the evaluation of the theoretical efficiency of a greedy algorithm is suggested. Using this criterion, we prove some results on the rate of convergence of greedy algorithms, which provide expansions. We consider both the case of Hilbert spaces and the more general case of Banach spaces. The new component of this paper is that we bound the error of approximation by the product of two norms—the norm of f and the A_1 -norm of f . Typically, only the A_1 -norm of f is used. In particular, we establish that some greedy algorithms (Pure Greedy Algorithm (PGA) and its modifications) are as good as the Orthogonal Greedy Algorithm (OGA) in this new sense of the rate of convergence, while it is known that the PGA is much worse than the OGA in the standard sense. Our new results provide better bounds for the accuracy than known results in the case of small $\|f\|$.

Keywords: greedy approximation; rate of convergence; pure greedy algorithm

MSC: 41A25



Citation: Temlyakov, V. On the Rate of Convergence of Greedy Algorithms. *Mathematics* **2023**, *11*, 2559. <https://doi.org/10.3390/math11112559>

Academic Editor: Kazaros Kazarian and Mikhail Dyachenko

Received: 8 May 2023

Revised: 27 May 2023

Accepted: 31 May 2023

Published: 2 June 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

This paper is devoted to the theoretical study of the efficiency of some greedy algorithms. Greedy algorithms are very useful in applications; in particular, adaptive methods are used in PDE solvers, and sparse approximation is used in image/signal/data processing, as well as in the design of neural networks, and in convex optimization. This fact motivated deep theoretical study of a variety of greedy algorithms. In this paper, we study two the most popular greedy algorithms—the Pure Greedy Algorithm (PGA) and the Orthogonal Greedy Algorithm (OGA) and their natural modifications. The reader can find other important greedy algorithms in the book [1] and in the papers (Regularized Orthogonal Matching Pursuit) [2], (Compressive Sampling Matching Pursuit) [3], (Subspace Pursuit) [4], (Rescaled Pure Greedy Algorithm) [5], and (Biorthogonal Greedy Algorithm) [6]. The reader can find some results on the application of greedy algorithms in convex optimization in the papers [7–14].

There are different criteria for the theoretical efficiency of greedy algorithms. All of them are based on the accuracy of the algorithm (the error after the m th iteration of the algorithm) and take different forms. One of those criteria uses the worst error over a given class of incomes (elements, which we approximate by the algorithm). We discuss a variant of this criterion in this paper. There is another, a more delicate criterion, which is based on the Lebesgue-type inequalities for individual elements. We do not discuss this criterion here. The reader can find a survey of the corresponding results in [15], Ch. 8. We now proceed to a detailed discussion of our results.

Let us begin with a general description of the problem. Let X be a Banach space with the norm $\|\cdot\|_X$ and $Y \subset X$ be a subspace of X with a stronger norm $\|f\|_Y \geq \|f\|_X$,

$f \in Y$. Consider a homogeneous approximation operator (linear or nonlinear) $G : Y \rightarrow X$, $G(af) = aG(f)$, $f \in Y$, $a \in \mathbb{R}$, and the error of approximation:

$$e(B_Y, G)_X := \sup_{f \in B_Y} \|f - G(f)\|_X, \quad B_Y := \{f : \|f\|_Y \leq 1\}.$$

Then, for any $f \in Y$, we have:

$$\|f - G(f)\|_X \leq e(B_Y, G)_X \|f\|_Y. \quad (1)$$

The characteristic $e(B_Y, G)_X$ plays an important role in approximation theory, with many classical examples of spaces X and Y ; for instance, $X = L_p$ and Y is one of the smoothness spaces such as Sobolev, Nikol'skii, or Besov space.

In this paper, we focus on the following version of the inequality (1): find the best $\gamma(\alpha, G, X, Y)$ such that the inequality:

$$\|f - G(f)\|_X \leq \gamma(\alpha, G, X, Y) \|f\|_X^{1-\alpha} \|f\|_Y^\alpha, \quad \alpha \in [0, 1], \quad (2)$$

holds for all $f \in Y$. Clearly, $\gamma(1, G, X, Y) = e(B_Y, G)_X$. Additionally, it is clear that under assumption: $\|f - G(f)\|_X \leq \|f\|_X$, $f \in Y$, we obtain the trivial bound:

$$\gamma(\alpha, G, X, Y) \leq e(B_Y, G)_X^\alpha.$$

In this paper, we discuss greedy approximation with respect to a given dictionary and prove some nontrivial inequalities for $\gamma(\alpha, G, X, Y)$ both in the case of X being a Hilbert space and X being a Banach space. In particular, we establish that some greedy algorithms (Pure Greedy Algorithm (PGA) and its generalizations) are as good as the Orthogonal Greedy Algorithm (OGA) in the sense of inequality (2), while it is known that the PGA is much worse than the OGA in the sense of the inequality (1) (for definitions and precise formulations, see below).

Let H be a real Hilbert space with the inner product $\langle \cdot, \cdot \rangle$ and norm $\|\cdot\|$. We say that a set of elements (functions) \mathcal{D} from H is a dictionary (symmetric dictionary) if each $g \in \mathcal{D}$ has norm one ($\|g\| = 1$) and $\overline{\text{span}} \mathcal{D} = H$. In addition, we assume for convenience the property of symmetry:

$$g \in \mathcal{D} \quad \text{implies} \quad -g \in \mathcal{D}.$$

We define the Pure Greedy Algorithm (PGA). We describe this algorithm for a general dictionary \mathcal{D} . If $f \in H$, we let $g(f) \in \mathcal{D}$ be an element from \mathcal{D} which maximizes $\langle f, g \rangle$. We assume for simplicity that such a maximizer exists; if not, suitable modifications are necessary (see Weak Greedy Algorithm below) in the algorithm that follows. We define:

$$G(f, \mathcal{D}) := \langle f, g(f) \rangle g(f) \quad \text{and} \quad R(f, \mathcal{D}) := f - G(f, \mathcal{D}).$$

Pure Greedy Algorithm (PGA). We define $f_0 := f$ and $G_0(f, \mathcal{D}) := 0$. Then, for each $m \geq 1$, we inductively define

$$G_m(f, \mathcal{D}) := G_{m-1}(f, \mathcal{D}) + G(f_{m-1}, \mathcal{D}),$$

$$f_m := f - G_m(f, \mathcal{D}) = R(f_{m-1}, \mathcal{D}).$$

Note that for a given element f , the sequence $\{G_m(f, \mathcal{D})\}$ may not be unique.

This algorithm is well studied from the point of view of convergence and rate of convergence. The reader can find the corresponding results and historical comments in [1], Ch. 2. In this paper, we focus on the rate of convergence. Typically, in approximation theory, we define the rate of convergence for specific classes. In classical approximation theory, these are smoothness classes. Clearly, in the general setting with arbitrary H and \mathcal{D} , we do not have a concept of smoothness similar to the classical smoothness of functions. It turns out that the geometrically defined class, namely, the closure of the convex hull of

\mathcal{D} , which we denote by $A_1(\mathcal{D})$, is a very natural class. For each $f \in H$, we associate the following norm:

$$\|f\|_{A_1(\mathcal{D})} := \inf\{M > 0 : f/M \in A_1(\mathcal{D})\}.$$

Clearly, $\|f\| \leq \|f\|_{A_1(\mathcal{D})}$. Then, the problem of the rate of convergence of the PGA can be formulated as follows (see [1], p. 95). Find the order of decay of the sequence:

$$\gamma_m(H) := \sup_{\{G_m(f, \mathcal{D})\}, f, \mathcal{D}} \frac{\|f - G_m(f, \mathcal{D})\|}{\|f\|_{A_1(\mathcal{D})}},$$

where the supremum is taken over all possible choices of $\{G_m(f, \mathcal{D})\}$, over all elements $f \in H$, $f \neq 0$, $\|f\|_{A_1(\mathcal{D})} < \infty$, and over all dictionaries \mathcal{D} . This problem is a central theoretical problem in greedy approximation in Hilbert spaces and it is still open. We mention some of the known results here and refer the reader, for the detailed history of the problem, to [1], Ch. 2. It is clear that for any $f \in H$, such that $\|f\|_{A_1(\mathcal{D})} < \infty$ we have:

$$\|f - G_m(f, \mathcal{D})\| \leq \gamma_m(H) \|f\|_{A_1(\mathcal{D})}.$$

In this paper, we discuss the following extension of the asymptotic characteristic $\gamma_m(H)$: for $\alpha \in (0, 1]$ define:

$$\gamma_m(\alpha, H) := \sup_{\{G_m(f, \mathcal{D})\}, f, \mathcal{D}} \frac{\|f - G_m(f, \mathcal{D})\|}{\|f\|^{1-\alpha} \|f\|_{A_1(\mathcal{D})}^\alpha}.$$

Clearly,

$$\gamma_m(1, H) = \gamma_m(H), \quad \gamma_m(\alpha, H) \geq \gamma_m(\beta, H) \quad \text{if } \alpha \leq \beta. \quad (3)$$

The first upper bound on $\gamma_m(H)$ was obtained in [16]:

$$\gamma_m(H) \leq m^{-1/6}.$$

Actually, the proof in [16] (see also [1], pp. 92–93) gives:

$$\gamma_m(1/3, H) \leq m^{-1/6}.$$

We establish here the following bounds:

$$\frac{1}{2} m^{-\alpha/2} \leq \gamma_m(\alpha, H) \leq m^{-\alpha/2}, \quad \alpha \leq 1/3. \quad (4)$$

Additionally, in Section 2, we find the right behavior of the asymptotic characteristic similar to $\gamma_m(\alpha, H)$ for a more general algorithm than the PGA, namely, for the Weak Greedy Algorithm with parameter b .

It is interesting to compare the rates of convergence of the PGA and the Orthogonal Greedy Algorithm (OGA). We now give a brief definition of the OGA. We define $f_0^o := f$, $G_0^o(f, \mathcal{D}) := 0$ and for $m \geq 1$, we inductively define $G_m^o(f, \mathcal{D})$ to be the orthogonal projection of f onto the span of $g(f_0^o), \dots, g(f_{m-1}^o)$ and set $f_m^o := f - G_m^o(f, \mathcal{D})$. The analogs of the characteristics $\gamma_m(H)$ and $\gamma_m(\alpha, H)$ for the OGA are denoted by $\gamma_m^o(H)$ and $\gamma_m^o(\alpha, H)$. The following bound is proved in [16] (see also [1], p. 93):

$$\gamma_m^o(H) \leq m^{-1/2}. \quad (5)$$

It is known (see [17]) that $\gamma_m(H)$ decays slower than $m^{-0.1898}$. Therefore, from the point of view of the characteristics $\gamma_m(H)$ and $\gamma_m^o(H)$, the OGA is much better than the PGA. We establish here the following bounds:

$$\frac{1}{2}m^{-\alpha/2} \leq \gamma_m^o(\alpha, H) \leq m^{-\alpha/2}, \quad \alpha \leq 1. \quad (6)$$

This means that from the point of view of the characteristics $\gamma_m(\alpha, H)$ and $\gamma_m^o(\alpha, H)$, the OGA is the same (in the sense of order) as PGA for $\alpha \leq 1/3$. This is a very surprising fact.

We do not know if the upper bound in (4) holds for $\alpha > 1/3$. However, the inequality in (3) and the lower bound for the $\gamma_m(H)$ show that:

$$\gamma_m(\alpha, H) \geq \gamma_m(H) \geq cm^{-0.1898}.$$

Therefore, the upper bound in (4) cannot be extended beyond $\alpha_0 := 0.3796$.

Section 3 deals with the case of a Banach space X . The results for the Banach space case are similar to those for Hilbert spaces, but are not as sharp as their counterparts.

Novelty. In this paper, we suggest a new criterion for the evaluation of the theoretical efficiency of a greedy algorithm. The classical criterion uses the worst error (for instance, $\gamma_m(H)$) of approximation of elements from the class $A_1(\mathcal{D})$ by our algorithm (for instance, the PGA). In other words, this criterion uses the norm $\|f\|_{A_1(\mathcal{D})}$ for estimating the error of approximation of f . Our new criterion uses two norms, $\|f\|_{A_1(\mathcal{D})}$ and $\|f\|$; more precisely, the weighted product of these norms $\|f\|^{1-\alpha}\|f\|_{A_1(\mathcal{D})}^\alpha$, $\alpha \in [0, 1]$. The most important qualitative discovery of this paper is that the PGA and its natural modifications have the same theoretical efficiency for some α as the OGA and its modifications. It is known that in accordance with the old criterion, the PGA is much worse than the OGA.

Method. The standard way of analyzing the accuracy of a greedy algorithm is based on estimating from below the difference $\|f_{m-1}\| - \|f_m\|$ and then solving (estimating) the corresponding recurrent inequalities. For instance, this method works very well for the OGA. In the paper [16], this standard method was modified in such a way that it allowed us to obtain some new nontrivial upper bounds for $\gamma_m(H)$. The method from [16] simultaneously analyzes two sequences, the $\{\|f_m\|\}_{m=1}^\infty$ and the sequence of sums of absolute values of the coefficients of the $G_m(f, \mathcal{D})$. In this paper, we further develop the method from [16].

The results of this paper (see Theorem 2 and lower bounds in Section 2) show that our method, which is a development of the method from [16], is optimal for proving the upper bounds for $\gamma_m^{t,b}(\alpha, H)$ in the case $\alpha \leq \frac{(2-b)t}{(2-b)t+2}$. It is known that the bound $\gamma_m(H) \leq m^{-1/6}$, which was obtained in [16], is not optimal. The bound $\gamma_m(H) \leq 4m^{-11/62}$ was proved in [18] by a method distinct from the one in [16]. The method from [18] was further developed in [19,20]. It would be interesting to understand if the method from [18] and its further developments allow us to prove an analog of Theorem 2 for $\alpha > \frac{(2-b)t}{(2-b)t+2}$.

Conclusion. The PGA at the m th iteration searches for an element $g(f_{m-1})$, which maximizes the inner product $\langle f_{m-1}, g \rangle$ over all $g \in \mathcal{D}$. Then, the update is very easy: $f_m = f_{m-1} - \langle f_{m-1}, g(f_{m-1}) \rangle g(f_{m-1})$. The OGA, like the PGA at the m th iteration, searches for an element $g(f_{m-1}^o)$, which maximizes the inner product $\langle f_{m-1}^o, g \rangle$ over all $g \in \mathcal{D}$. However, the second steps of the PGA and OGA at the m th iteration are different—the OGA makes an orthogonal projection onto the span of $g(f_0^o), \dots, g(f_{m-1}^o)$. Clearly, this step of the OGA is more difficult than the corresponding step of the PGA. Moreover, it is clear from the definition of the PGA that it provides an expansion of f into a series with respect to \mathcal{D} . The OGA does not provide an expansion. Thus, the advantage of the PGA over the OGA is that it is simpler and provides an expansion. The advantage of the OGA over the PGA is that in accordance with the old criterion, we can guarantee better accuracy for an $f \in A_1(\mathcal{D})$. The results of this paper show that in accordance with the new criterion, the OGA does not have an advantage in the sense of accuracy for some parameters α .

Similar results are obtained for the modifications of the PGA—the Weak Greedy Algorithm with parameter b in Hilbert spaces (see Section 2) and the Dual Greedy Algorithm with parameters (t, b, μ) in Banach spaces (see Section 3).

2. Hilbert Space: The Weak Greedy Algorithm with Parameter b

Let a sequence $\tau = \{t_k\}_{k=1}^{\infty}$, $0 \leq t_k \leq 1$, and a parameter $b \in (0, 1]$ be given. We define the Weak Greedy Algorithm with parameter b .

Weak Greedy Algorithm with parameter b (WGA(τ, b)). We define $f_0 := f_0^{\tau, b} := f$. Then, for each $m \geq 1$, we inductively define:

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

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

(2)

$$f_m := f_m^{\tau, b} := f_{m-1} - b \langle f_{m-1}, \varphi_m \rangle \varphi_m;$$

(3)

$$G_m(f, \mathcal{D}) := G_m^{\tau, b}(f, \mathcal{D}) := b \sum_{j=1}^m \langle f_{j-1}, \varphi_j \rangle \varphi_j.$$

In the case $t_k = t$, $k = 1, 2, \dots$, we write t in the notation instead of τ .

We proceed to the rate of convergence. The following Theorem 1 was proved in [21].

Theorem 1. Let \mathcal{D} be an arbitrary dictionary in H . Assume $\tau := \{t_k\}_{k=1}^{\infty}$ is a nonincreasing sequence and $b \in (0, 1]$. Then, for $f \in A_1(\mathcal{D})$, we have:

$$\|f - G_m^{\tau, b}(f, \mathcal{D})\| \leq e_m(\tau, b), \quad (7)$$

where:

$$e_m(\tau, b) := \left(1 + b(2-b) \sum_{k=1}^m t_k^2 \right)^{-\frac{(2-b)t_m}{2(2+(2-b)t_m)}}. \quad (8)$$

Theorem 1 implies the following inequality for any f and any \mathcal{D} :

$$\frac{\|f - G_m^{\tau, b}(f, \mathcal{D})\|}{\|f\|_{A_1(\mathcal{D})}} \leq e_m(\tau, b). \quad (9)$$

We now extend Theorem 1 to provide a bound for:

$$\gamma_m^{t, b}(\alpha, H) := \sup_{\mathcal{D}} \sup_{f \in A_1(\mathcal{D}), f \neq 0} \sup_{G_m^{t, b}(f, \mathcal{D})} \frac{\|f - G_m^{t, b}(f, \mathcal{D})\|}{\|f\|^{1-\alpha} \|f\|_{A_1(\mathcal{D})}^{\alpha}}. \quad (10)$$

We prove the following Theorem 2 in the case $t_k = t$, $k = 1, 2, \dots$.

Theorem 2. For any Hilbert space H , we have:

$$\gamma_m^{t, b}(\alpha, H) \leq (1 + mb(2-b)t^2)^{-\alpha/2}, \quad (11)$$

provided $\alpha \leq \frac{(2-b)t}{(2-b)t+2}$.

Proof. The proof of this theorem goes along the lines of the proof of Theorem 1 in [21]. Let \mathcal{D} be a dictionary in H . We introduce some notations:

$$f_k := f_k^{t, b}, \quad \varphi_k := \varphi_k^{t, b}, \quad k = 0, 1, \dots,$$

$$a_m := \|f_m\|^2, \quad y_m := \langle f_{m-1}, \varphi_m \rangle, \quad m = 1, 2, \dots,$$

and consider the sequence $\{B_n\}$ defined as follows:

$$B_0 := \|f\|_{A_1(\mathcal{D})}, \quad B_m := B_{m-1} + by_m, \quad m = 1, 2, \dots$$

It is clear that $\|f_n\|_{A_1(\mathcal{D})} \leq B_n$, $n = 0, 1, \dots$. By Lemma 3.5 from [16] (see also [1], p. 91, Lemma 2.17), we get:

$$\sup_{g \in \mathcal{D}} \langle f_{m-1}, g \rangle \geq \|f_{m-1}\|^2 / B_{m-1}. \quad (12)$$

From here and from the equality:

$$\|f_m\|^2 = \|f_{m-1}\|^2 - b(2-b)\langle f_{m-1}, \varphi_m \rangle^2$$

we obtain the following relations:

$$a_m = a_{m-1} - b(2-b)y_m^2, \quad (13)$$

$$B_m = B_{m-1} + by_m, \quad (14)$$

$$y_m \geq ta_{m-1} / B_{m-1}. \quad (15)$$

From (13) and (15), we obtain:

$$a_m \leq a_{m-1}(1 - b(2-b)t^2a_{m-1}B_{m-1}^{-2}).$$

Using that $B_{m-1} \leq B_m$, we derive from here:

$$a_mB_m^{-2} \leq a_{m-1}B_{m-1}^{-2}(1 - b(2-b)t^2a_{m-1}B_{m-1}^{-2}). \quad (16)$$

We shall need the following simple known Lemma (see, for example, [1], p. 91, in case $C_1 = C_2$).

Lemma 1. Let $\{x_m\}_{m=0}^\infty$ be a sequence of non-negative numbers satisfying the inequalities:

$$x_0 \leq C_1, \quad x_{m+1} \leq x_m(1 - x_mC_2), \quad m = 0, 1, 2, \dots, \quad C_1, C_2 > 0.$$

Then, we have for each m :

$$x_m \leq (C_1^{-1} + C_2m)^{-1}.$$

Proof. The proof is by induction on m . For $m = 0$, the statement is true by assumption. We assume $x_m \leq (C_1^{-1} + C_2m)^{-1}$ and prove that $x_{m+1} \leq (C_1^{-1} + C_2(m+1))^{-1}$. If $x_{m+1} = 0$, this statement is obvious. Assume, therefore, that $x_{m+1} > 0$. Then, we have:

$$x_{m+1}^{-1} \geq x_m^{-1}(1 - x_mC_2)^{-1} \geq x_m^{-1}(1 + x_mC_2) = x_m^{-1} + C_2 \geq C_1^{-1} + (m+1)C_2,$$

which implies $x_{m+1} \leq (C_1^{-1} + C_2(m+1))^{-1}$. \square

We apply Lemma 1 with $x_m := a_mB_m^{-2}$. Then, the inequality $\|f\| \leq \|f\|_{A_1(\mathcal{D})}$ implies that we can take $C_1 = 1$. We set $C_2 = b(2-b)t^2$ and obtain from (16) and Lemma 1:

$$a_mB_m^{-2} \leq (1 + mb(2-b)t^2)^{-1}. \quad (17)$$

Relations (13) and (15) imply:

$$a_m \leq a_{m-1} - b(2-b)y_mta_{m-1}/B_{m-1} = a_{m-1}(1 - b(2-b)ty_m/B_{m-1}). \quad (18)$$

We now need the following simple inequality: for any $x < 1$ and any $a > 0$, we have:

$$(1-x)(1+x/a)^a \leq 1. \quad (19)$$

Rewriting (14) in the form:

$$B_m = B_{m-1}(1 + by_m/B_{m-1}) \quad (20)$$

and using the inequality (19) with $x = b(2-b)ty_m/B_{m-1}$ and $a = (2-b)t$, we get from (18) and (20) that:

$$a_m B_m^{(2-b)t} \leq a_{m-1} B_{m-1}^{(2-b)t} \leq \dots \leq \|f\|^2 \|f\|_{A_1(\mathcal{D})}^{(2-b)t}. \quad (21)$$

Combining (17) and (21), we obtain:

$$a_m^{(2-b)t+2} \leq \|f\|^4 \|f\|_{A_1(\mathcal{D})}^{2(2-b)t} (1 + mb(2-b)t^2)^{-(2-b)t},$$

which completes the proof of Theorem 2 with $\alpha_0 := \frac{(2-b)t}{(2-b)t+2}$. The case $\alpha \leq \alpha_0$ follows from Lemma 2 below. \square

Lower bounds. Let H be an infinite dimensional Hilbert space and $\{e_k\}_{k=1}^\infty$ be an orthonormal system in H . Suppose that our symmetric dictionary \mathcal{D} consists of $\pm e_k$, $k = 1, 2, \dots$, and other elements $g \in \mathcal{D}$ have the property $\langle g, e_k \rangle = 0$, $k = 1, 2, \dots$. We present an example in the case $t = 1$. Let $b \in (0, 1]$ and m be given. We consider two cases (I) $b \leq 1/4$ and (II) $b \in (1/4, 1]$.

(I). Set $m' := [2bm] + 1$ and:

$$f = \sum_{k=1}^{m'} e_k.$$

Then, at each iteration, the WGA(1, b) will pick one of the e_k , $k \in [1, m']$ with the largest coefficient. After m iterations, we will get:

$$f_m = \sum_{k=1}^{m'} c_k e_k, \quad c_k \geq 1 - (m/m' + 1)b \geq 1/4.$$

Therefore, we obtain:

$$\|f_m\| \geq (m')^{1/2}/4, \quad \|f\| = (m')^{1/2}, \quad \|f\|_{A_1(\mathcal{D})} \leq m'.$$

Thus, for any $\alpha \in [0, 1]$, we find:

$$\frac{\|f_m\|}{\|f\|^{1-\alpha} \|f\|_{A_1(\mathcal{D})}^\alpha} \geq (m')^{-\alpha/2}/4. \quad (22)$$

(II). Set:

$$f = \sum_{k=1}^{2m} e_k.$$

Then,

$$f_m = \sum_{k=1}^{2m} c_k e_k, \quad c_k = 1, k \in G, \quad c_k = 1 - b, k \notin G, \quad |G| = m.$$

Therefore, we obtain:

$$\|f_m\| \geq (m)^{1/2}, \quad \|f\| = (2m)^{1/2}, \quad \|f\|_{A_1(\mathcal{D})} \leq 2m.$$

Thus, for any $\alpha \in [0, 1]$, we find:

$$\frac{\|f_m\|}{\|f\|^{1-\alpha} \|f\|_{A_1(\mathcal{D})}^\alpha} \geq 2^{-1/2} (2m)^{-\alpha/2}. \quad (23)$$

Bounds (22) and (23) show that in the case $t = 1$, inequality (11) is sharp in the sense of dependence on m and b , when m goes to ∞ and b goes to 0.

Proof of (6). The lower bound in (6) follows from (23) because in the case of an orthonormal system, the PGA and OGA algorithms coincide. We now prove the upper bound. \square

Lemma 2. Let $g_m(\alpha, H)$ denote either $\gamma_m(\alpha, H)$ or $\gamma_m^o(\alpha, H)$. Suppose that for some $\beta \in (0, 1]$, we have:

$$g_m(\beta, H) \leq C\varphi(m)^{-\beta/2}.$$

Then, for any $\alpha \in (0, \beta)$, we have:

$$g_m(\alpha, H) \leq C^{\alpha/\beta} \varphi(m)^{-\alpha/2}. \quad (24)$$

Proof. By the definition of the PGA and OGA definitions, we have $\|f_m\| \leq \|f\|$ and $\|f_m^o\| \leq \|f\|$. For both PGA and OGA algorithms, the proof is identical. We will carry it out for the PGA. Our assumption gives for any f , any dictionary \mathcal{D} , and any realization of $G_m(f, \mathcal{D})$:

$$\|f_m\| = \|f - G_m(f, \mathcal{D})\| \leq \|f\|^{1-\beta} \|f\|_{A_1(\mathcal{D})}^\beta C\varphi(m)^{-\beta/2}.$$

Therefore, for any $a \in [0, 1]$, we have:

$$\|f_m\| \leq \|f\|^{1-a} (\|f\|^{1-\beta} \|f\|_{A_1(\mathcal{D})}^\beta C\varphi(m)^{-\beta/2})^a. \quad (25)$$

Choosing $a = \alpha/\beta$, we obtain (24) from (25). \square

The upper bound in (6) follows from (5) and Lemma 2.

3. Greedy Expansions in Banach Spaces

In this section, we extend the results from Section 2 to the case of a Banach space instead of a Hilbert space. We begin with some definitions. Let X be a real Banach space with norm $\|\cdot\|$. As above, we say that a set of elements (functions) \mathcal{D} from X is a dictionary (symmetric dictionary) if each $g \in \mathcal{D}$ has norm one ($\|g\| = 1$) and $\overline{\text{span}}\mathcal{D} = X$. In addition, we assume for convenience that the dictionary is symmetric:

$$g \in \mathcal{D} \text{ implies } -g \in \mathcal{D}.$$

In this paper, we study greedy algorithms with regard to \mathcal{D} that provide greedy expansions. For a nonzero element $f \in X$, we denote by F_f a norming (peak) functional for f :

$$\|F_f\| = 1, \quad F_f(f) = \|f\|.$$

The existence of such a functional is guaranteed by Hahn–Banach theorem. Denote

$$r_{\mathcal{D}}(f) := \sup_{F_f} \sup_{g \in \mathcal{D}} F_f(g).$$

We note that, in general, a norming functional F_f is not unique. This is why we take \sup_{F_f} over all norming functionals of f in the definition of $r_{\mathcal{D}}(f)$. It is known that in the case of uniformly smooth Banach spaces (our primary object here), the norming functional F_f is unique. In such a case, we do not need \sup_{F_f} in the definition of $r_{\mathcal{D}}(f)$.

We consider here approximation in uniformly smooth Banach spaces. For a Banach space X , we define the modulus of smoothness:

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

A uniformly smooth Banach space is one with the property:

$$\lim_{u \rightarrow 0} \rho(u)/u = 0.$$

It is well known (see, for instance, [22], Lemma B.1) that in the case $X = L_p$, $1 \leq p < \infty$, we have:

$$\rho(u, L_p) \leq \begin{cases} u^p/p & \text{if } 1 \leq p \leq 2, \\ (p-1)u^2/2 & \text{if } 2 \leq p < \infty. \end{cases} \quad (26)$$

We now give a definition of the $\text{DGA}(\tau, b, \mu)$, $\tau = \{t_k\}_{k=1}^\infty$, $t_k \in (0, 1]$ introduced in [21] (see also [1], Ch. 6).

Dual Greedy Algorithm with parameters (τ, b, μ) ($\text{DGA}(\tau, b, \mu)$). Let X be a uniformly smooth Banach space with the modulus of smoothness $\rho(u)$ and let $\mu(u)$ be a majorant of $\rho(u)$: $\rho(u) \leq \mu(u)$, $u \in [0, \infty)$. For a sequence $\tau = \{t_k\}_{k=1}^\infty$, $t_k \in (0, 1]$ and a parameter $b \in (0, 1]$, we define sequences $\{f_m\}_{m=0}^\infty$, $\{\varphi_m\}_{m=1}^\infty$, $\{c_m\}_{m=1}^\infty$, and $\{G_m\}_{m=0}^\infty$ inductively. Let $f_0 := f$ and $G_0 := 0$. If for $m \geq 1$ $f_{m-1} = 0$, then we set $f_j = 0$ for $j \geq m$ and stop. If $f_{m-1} \neq 0$, then we conduct the following three steps:

(1) take any $\varphi_m \in \mathcal{D}$ such that:

$$F_{f_{m-1}}(\varphi_m) \geq t_m r_{\mathcal{D}}(f_{m-1}); \quad (27)$$

(2) choose $c_m > 0$ from the equation:

$$\|f_{m-1}\| \mu(c_m / \|f_{m-1}\|) = \frac{t_m b}{2} c_m r_{\mathcal{D}}(f_{m-1}); \quad (28)$$

(3) define:

$$f_m := f_{m-1} - c_m \varphi_m, \quad G_m := G_m^{\tau, b, \mu} := G_{m-1} + c_m \varphi_m. \quad (29)$$

Along with the algorithm $\text{DGA}(\tau, b, \mu)$, we consider a slight modification of it; when at step (2), we find c_m from the equation (see [21], Remark 3.1):

$$\|f_{m-1}\| \mu(c_m / \|f_{m-1}\|) = \frac{b}{2} c_m F_{f_{m-1}}(\varphi_m). \quad (30)$$

We denote this modification by $\text{DGA}(\tau, b, \mu)^*$.

We proceed to studying the rate of convergence of the $\text{DGA}(\tau, b, \mu)$ in the uniformly smooth Banach spaces with the power type majorant of the modulus of smoothness: $\rho(u) \leq \mu(u) = \gamma u^q$, $1 < q \leq 2$. The following Theorem 3 is from [21] (see also [1], p. 372).

Theorem 3. Let $\tau := \{t_k\}_{k=1}^\infty$ be a nonincreasing sequence $1 \geq t_1 \geq t_2 \geq \dots > 0$ and $b \in (0, 1)$. Assume that X has a modulus of smoothness $\rho(u) \leq \gamma u^q$, $q \in (1, 2]$. Denote $\mu(u) := \gamma u^q$. Then, for any dictionary \mathcal{D} and any $f \in A_1(\mathcal{D})$, the rate of convergence of the $\text{DGA}(\tau, b, \mu)$ is given by:

$$\|f_m\| \leq C(b, \gamma, q) \left(1 + \sum_{k=1}^m t_k^p \right)^{-\frac{t_m(1-b)}{p(1+t_m(1-b))}}, \quad p := \frac{q}{q-1}.$$

Remark 1. It is pointed out in [21], Remark 3.2, that Theorem 3 holds for the algorithm $\text{DGA}(\tau, b, \mu)^*$ as well.

Theorem 3 is an analog of Theorem 1. We now prove an analog of Theorem 2. We extend Theorem 3 to provide a bound for:

$$\gamma_m^{t,b,\mu}(\alpha, X) := \sup_{\mathcal{D}} \sup_{f \in A_1(\mathcal{D}), f \neq 0} \sup_{G_m^{t,b,\mu}(f, \mathcal{D})} \frac{\|f - G_m^{t,b,\mu}(f, \mathcal{D})\|}{\|f\|^{1-\alpha} \|f\|_{A_1(\mathcal{D})}^\alpha}. \quad (31)$$

The corresponding characteristic for the algorithm $\text{DGA}(\tau, b, \mu)^*$ is denoted by $\gamma_m^{t,b,\mu}(\alpha, X)^*$. We prove the following Theorem 4 in the case $t_k = t, k = 1, 2, \dots$

Theorem 4. For any Banach space X with modulus of smoothness $\rho(u, X) \leq \gamma u^q, 1 < q \leq 2, p := \frac{q}{q-1}$, we have:

$$\gamma_m^{t,b,\mu}(\alpha, X) \leq (1 + mct^p)^{-\alpha/p}, \quad c := (1 - b) \left(\frac{b}{2\gamma} \right)^{\frac{1}{q-1}}, \quad (32)$$

provided $\alpha \leq \frac{t(1-b)}{1+t(1-b)}$. The same inequality holds for the $\gamma_m^{t,b,\mu}(\alpha, X)^*$.

Proof. The proof is identical for both characteristics $\gamma_m^{t,b,\mu}(\alpha, X)$ and $\gamma_m^{t,b,\mu}(\alpha, X)^*$. We carry it out for the $\gamma_m^{t,b,\mu}(\alpha, X)$. From the definition of the modulus of smoothness, we have:

$$\|f_{n-1} - c_n \varphi_n\| + \|f_{n-1} + c_n \varphi_n\| \leq 2\|f_{n-1}\|(1 + \rho(c_n/\|f_{n-1}\|)). \quad (33)$$

Using the definition of φ_n :

$$F_{f_{n-1}}(\varphi_n) \geq \text{tr}_{\mathcal{D}}(f_{n-1}), \quad (34)$$

we get:

$$\begin{aligned} \|f_{n-1} + c_n \varphi_n\| &\geq F_{f_{n-1}}(f_{n-1} + c_n \varphi_n) \\ &= \|f_{n-1}\| + c_n F_{f_{n-1}}(\varphi_n) \geq \|f_{n-1}\| + c_n \text{tr}_{\mathcal{D}}(f_{n-1}). \end{aligned} \quad (35)$$

Combining (33) and (35), we get:

$$\|f_n\| = \|f_{n-1} - c_n \varphi_n\| \leq \|f_{n-1}\|(1 + 2\rho(c_n/\|f_{n-1}\|)) - c_n \text{tr}_{\mathcal{D}}(f_{n-1}). \quad (36)$$

Using the choice of c_m , we get from here:

$$\|f_m\| \leq \|f_{m-1}\| - t(1 - b)c_m r_{\mathcal{D}}(f_{m-1}). \quad (37)$$

Thus, we need to estimate from below $c_m r_{\mathcal{D}}(f_{m-1})$. It is clear that:

$$\|f_{m-1}\|_{A_1(\mathcal{D})} = \|f - \sum_{j=1}^{m-1} c_j \varphi_j\|_{A_1(\mathcal{D})} \leq \|f\|_{A_1(\mathcal{D})} + \sum_{j=1}^{m-1} c_j. \quad (38)$$

Denote $B_n := \|f\|_{A_1(\mathcal{D})} + \sum_{j=1}^n c_j$. Then, by (38), we have:

$$\|f_{m-1}\|_{A_1(\mathcal{D})} \leq B_{m-1}.$$

Next, by Lemma 6.10 from [1], p. 343, we obtain:

$$\begin{aligned} r_{\mathcal{D}}(f_{m-1}) &= \sup_{g \in \mathcal{D}} F_{f_{m-1}}(g) = \sup_{\varphi \in A_1(\mathcal{D})} F_{f_{m-1}}(\varphi) \\ &\geq \|f_{m-1}\|_{A_1(\mathcal{D})}^{-1} F_{f_{m-1}}(f_{m-1}) \geq \|f_{m-1}\| / B_{m-1}. \end{aligned} \quad (39)$$

Substituting (39) into (37), we get:

$$\|f_m\| \leq \|f_{m-1}\|(1 - t(1 - b)c_m/B_{m-1}). \quad (40)$$

From the definition of B_m , we find:

$$B_m = B_{m-1} + c_m = B_{m-1}(1 + c_m/B_{m-1}).$$

Using the inequality:

$$(1 + x)^\alpha \leq 1 + \alpha x, \quad 0 \leq \alpha \leq 1, \quad x \geq 0,$$

we obtain:

$$B_m^{t(1-b)} \leq B_{m-1}^{t(1-b)}(1 + t(1 - b)c_m/B_{m-1}). \quad (41)$$

Multiplying (40) and (41), we get:

$$\|f_m\|B_m^{t(1-b)} \leq \|f_{m-1}\|B_{m-1}^{t(1-b)} \leq \dots \leq \|f\|\|f\|_{A_1(\mathcal{D})}^{t(1-b)}. \quad (42)$$

The function $\mu(u)/u = \gamma u^{q-1}$ is increasing on $[0, \infty)$. Therefore, the c_m from (28) is greater than or equal to c'_m from the following equation (43) (see (39)):

$$\gamma\|f_{m-1}\|(c'_m/\|f_{m-1}\|)^q = \frac{tb}{2}c'_m\|f_{m-1}\|/B_{m-1}, \quad (43)$$

$$c'_m = \left(\frac{tb}{2\gamma}\right)^{\frac{1}{q-1}} \frac{\|f_{m-1}\|^{\frac{q}{q-1}}}{B_{m-1}^{\frac{1}{q-1}}}. \quad (44)$$

Using notations:

$$p := \frac{q}{q-1}, \quad c := (1 - b)\left(\frac{b}{2\gamma}\right)^{\frac{1}{q-1}},$$

we get from (37), (39), (44):

$$\|f_m\| \leq \|f_{m-1}\| \left(1 - ct^p \frac{\|f_{m-1}\|^p}{B_{m-1}^p}\right). \quad (45)$$

Noting that $B_m \geq B_{m-1}$, we derive from (45) that:

$$(\|f_m\|/B_m)^p \leq (\|f_{m-1}\|/B_{m-1})^p (1 - ct^p (\|f_{m-1}\|/B_{m-1})^p). \quad (46)$$

Taking into account that $\|f\| \leq \|f\|_{A_1(\mathcal{D})}$, we obtain from (46) by Lemma 1 with $C_1 = 1, C_2 = ct^p$:

$$(\|f_m\|/B_m)^p \leq (1 + mct^p)^{-1}. \quad (47)$$

Combining (42) and (47), we get:

$$\|f_m\| \leq \|f\|^{1-\alpha_0} \|f\|_{A_1(\mathcal{D})}^{\alpha_0} (1 + mct^p)^{-\alpha_0/p}, \quad p := \frac{q}{q-1}, \quad \alpha_0 := \frac{t(1-b)}{1+t(1-b)}.$$

This completes the proof of Theorem 4 for $\alpha = \alpha_0$. The case $\alpha < \alpha_0$ follows from the case $\alpha = \alpha_0$ and the corresponding analog of Lemma 2. \square

Let us discuss an application of Theorem 4 in the case of a Hilbert space. It is well known and easy to check that for a Hilbert space H , one has:

$$\rho(u) \leq (1 + u^2)^{1/2} - 1 \leq u^2/2.$$

Let us figure out how the $\text{DGA}(t, b, u^2/2)$ works in a Hilbert space. Consider the m th step of it. Let $\varphi_m \in \mathcal{D}$ be from (27) with $t_m = t$ (we assume existence in case $t = 1$). Then, it is clear that for φ_m , we have:

$$\langle f_{m-1}, \varphi_m \rangle \geq t \|f_{m-1}\| r_{\mathcal{D}}(f_{m-1}) = t \sup_{g \in \mathcal{D}} \langle f_{m-1}, g \rangle.$$

The $\text{WGA}(t, 1)$ would use φ_m with the coefficient $\langle f_{m-1}, \varphi_m \rangle$ at this step. The $\text{DGA}(t, b, u^2/2)^*$, like $\text{WGA}(t, b)$, uses the same φ_m and only a fraction of $\langle f_{m-1}, \varphi_m \rangle$:

$$c_m = b \|f_{m-1}\| F_{f_{m-1}}(\varphi_m) = b \langle f_{m-1}, \varphi_m \rangle. \quad (48)$$

Thus, the choice $b = 1$ in (48) corresponds to the WGA. However, it is clear from the above considerations that our technique, designed for general Banach spaces, does not work in the case $b = 1$. By Theorem 4 with $\mu(u) = u^2/2$, the $\text{DGA}(t, b, \mu)$ and $\text{DGA}(t, b, \mu)^*$ provide the following error estimate:

$$\|f_m\| \leq \|f\|^{1-\alpha} \|f\|_{A_1(\mathcal{D})}^{\alpha} (1 + mct^2)^{-\alpha/2}, \quad \alpha \leq \alpha_0 := \frac{t(1-b)}{1+t(1-b)}. \quad (49)$$

Note that the inequality (49) is similar to the corresponding inequality, which follows from Theorem 2, for $\alpha \leq \alpha_1 := \frac{t(2-b)}{2+t(2-b)}$. It is easy to check that $\alpha_0 < \alpha_1$, which means that Theorem 2 gives a stronger result than the corresponding corollary of Theorem 4.

A remark on lower bounds. In Section 2, we obtained the lower bounds, which are sharp in both parameters m and b . Clearly, the most important parameter is m . Here, we obtain the lower bounds in m , which apply to any algorithm providing m -term approximation after m iterations. Recall the definition of the concept of m -term approximation with respect to a given dictionary \mathcal{D} . Given an integer $m \in \mathbb{N}$, we denote by $\Sigma_m(\mathcal{D})$ the set of all m -term approximants with respect to \mathcal{D} :

$$\Sigma_m(\mathcal{D}) := \left\{ h \in X : h = \sum_{i=1}^m c_i g_i, \quad g_i \in \mathcal{D}, c_i \in \mathbb{R}, i = 1, \dots, m \right\}.$$

Define for a Banach space X :

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

to be the best m -term approximation of $f \in X$ in the X -norm with respect to \mathcal{D} .

Let $1 < q \leq 2$. Consider $X = \ell_q$. It is known ([23], p. 67) that ℓ_q , $1 < q \leq 2$, is a uniformly smooth Banach space with a modulus of smoothness $\rho(u)$ of power type q : $\rho(u) \leq \gamma u^q$. Choose $\mathcal{D} := \mathcal{E}$ as a symmetrized standard basis $\{\pm e_j\}_{j=1}^{\infty}$, $e_j := (0, \dots, 0, 1, 0, \dots)$, for ℓ_q . For a given $m \in \mathbb{N}$, set:

$$f := \sum_{i=1}^{2m} e_i.$$

Then, the following relations are obvious:

$$\|f\|_{\ell_q} = (2m)^{1/q}, \quad \|f\|_{A_1(\mathcal{E})} = 2m, \quad \sigma_m(f, \mathcal{E})_{\ell_q} = m^{1/q}.$$

Therefore, for any $\alpha \in [0, 1]$:

$$\frac{\sigma_m(f, \mathcal{E})_{\ell_q}}{\|f\|_{\ell_q}^{1-\alpha} \|f\|_{A_1(\mathcal{E})}^\alpha} \geq \frac{1}{2} m^{-\alpha/p}, \quad p := \frac{q}{q-1}. \quad (50)$$

This means that the upper bounds provided by Theorem 4 are sharp in m for any fixed parameters t and b . More precisely, for every $q \in (1, 2]$, there exists a Banach space X with $\rho(u, X) \leq \gamma u^q$, a dictionary $\mathcal{D} \subset X$, and $f \in X$ such that the following inequality holds:

$$\frac{\sigma_m(f, \mathcal{D})_X}{\|f\|_X^{1-\alpha} \|f\|_{A_1(\mathcal{D})}^\alpha} \geq \frac{1}{2} m^{-\alpha/p}, \quad p := \frac{q}{q-1}. \quad (51)$$

Inequality (51) follows directly from (50) with $X = \ell_q$ and $\mathcal{D} = \mathcal{E}$.

Remark 2. Inequality (51) gives the lower bound for the best m -term approximation. It is known (see [1], Ch. 6) that there are greedy-type algorithms; for instance, the Weak Chebyshev Greedy Algorithm and the Weak Greedy Algorithm with Free Relaxation with the weakness parameter $t \in (0, 1]$, which provide the following rate of convergence for $f \in X$ with $\rho(u, X) \leq \gamma u^q$, $1 < q \leq 2$,

$$\|f_m\|_X \leq C(q, \gamma)(1 + mt^p)^{-1/p} \|f\|_{A_1(\mathcal{D})}, \quad p := \frac{q}{q-1}.$$

This means that the corresponding upper bound in (51) (in the sense of order) can be realized by a greedy-type algorithm.

Note that for specific X and \mathcal{D} , the inequality (51) may be improved. We illustrate it on the example of $X = \ell_q$ with $q \in (2, \infty)$ and $\mathcal{D} = \mathcal{E}$. Without loss of generality, we can assume that:

$$f = \sum_{i=1}^{\infty} c_i e_i, \quad c_1 \geq c_2 \geq \dots \geq 0.$$

Then,

$$\|f\|_{\ell_q} = \left(\sum_{i=1}^{\infty} c_i^q \right)^{1/q}, \quad \|f\|_{A_1(\mathcal{E})} = \|f\|_{\ell_1} = \sum_{i=1}^{\infty} c_i. \quad (52)$$

We now estimate from above:

$$\sigma_m(f, \mathcal{E})_{\ell_q} = \left(\sum_{i=m+1}^{\infty} c_i^q \right)^{1/q}.$$

Our monotonicity assumption on $\{c_i\}$ implies:

$$c_m \leq m^{-1/q} \|f\|_{\ell_q}, \quad c_m \leq m^{-1} \|f\|_{\ell_1}$$

and, therefore, for any $\beta \in [0, 1]$:

$$c_m \leq m^{-(1/q)(1-\beta)-\beta} \|f\|_{\ell_q}^{1-\beta} \|f\|_{\ell_1}^\beta. \quad (53)$$

Setting $\alpha := \beta(1 - 1/q) + 1/q$, we obtain from here with $p := \frac{q}{q-1}$:

$$\sigma_m(f, \mathcal{E})_{\ell_q} \leq c_m^{1-1/q} \|f\|_{\ell_1}^{1/q} \leq m^{-\alpha/p} \|f\|_{\ell_q}^{1-\alpha} \|f\|_{\ell_1}^\alpha. \quad (54)$$

Therefore, for any $\alpha \in [0, 1]$, we have for all $f \in \ell_q$:

$$\frac{\sigma_m(f, \mathcal{E})_{\ell_q}}{\|f\|_{\ell_q}^{1-\alpha} \|f\|_{A_1(\mathcal{E})}^\alpha} \leq m^{-\alpha/p}, \quad p := \frac{q}{q-1}. \quad (55)$$

Note that it is known (see (26) above) that the space ℓ_q with $q \in [2, \infty)$ has the modulus of smoothness of the power type 2. Thus, Theorem 4 gives an analog of (55) with a weaker rate of decay $m^{-\alpha/2}$ than $m^{-\alpha/p}$ in (55).

We briefly discuss another example of the L_q -type spaces with $q \in (2, \infty)$. Consider the space $L_q(0, 2\pi)$ of real 2π -periodic functions and take $\mathcal{D} = \mathcal{T}$ to be the real trigonometric system. For a given $m \in \mathbb{N}$, set:

$$f := \sum_{k=1}^{2m} \cos(2^k x).$$

Then, it is well known that:

$$\|f\|_{L_q} \leq C(q)(2m)^{1/2}, \quad \|f\|_{A_1(\mathcal{T})} = 2m.$$

Moreover,

$$\sigma_m(f, \mathcal{T})_{L_q} \geq \sigma_m(f, \mathcal{T})_{L_2} \geq Cm^{1/2}.$$

Therefore, for any $\alpha \in [0, 1]$, we obtain:

$$\frac{\sigma_m(f, \mathcal{T})_{L_q}}{\|f\|_{L_q}^{1-\alpha} \|f\|_{A_1(\mathcal{T})}^\alpha} \geq C'(q)m^{-\alpha/2}. \quad (56)$$

4. Discussion

In this paper, we propose and study a new criterion for the evaluation of efficiency of a greedy algorithm. This criterion takes into account two characteristics of an element f , which we approximate: its norm $\|f\|$ (either in a Hilbert or in a Banach space) and the $\|f\|_{A_1(\mathcal{D})}$. We test this new criterion on two standard classes of algorithms. The first class consists of algorithms which provide for each element f an expansion with respect to a given dictionary \mathcal{D} . In this paper, we discuss two such algorithms: the PGA and the WGA(t, b) for Hilbert spaces, and the DGA(t, b, μ) for Banach spaces. Using our new criterion, we compare the efficiency of these algorithms with algorithms from the second class, which include the OGA (Hilbert spaces) and its generalization for Banach spaces—the Weak Chebyshev Greedy Algorithm. Algorithms from the second class are known for providing optimal in the sense of order rate of m -term approximation for the class $A_1(\mathcal{D})$. These algorithms do not provide expansions. In this paper, we show that from the point of view of our new criterion algorithms, PGA, WGA(t, b), and DGA(t, b, μ) are optimal in the sense of order. We illustrate this fact on the example of WGA(1, b). Theorem 2 with $t = 1$ gives the following upper bound for $\alpha \leq \frac{2-b}{4-b}$:

$$\gamma_m^{1,b}(\alpha, H) \leq (1 + mb)^{-\alpha/2}. \quad (57)$$

Inequality (22) shows that this bound cannot be improved in the sense of order. Moreover, inequality (51) with $p = 2$, which corresponds to a Hilbert space, shows that (57) cannot be improved in the sense of order with respect to m even if instead of the algorithm WGA(1, b), we use the best m -term approximations.

We point out that our new results provide better bounds for the accuracy than known results in the case of small $\|f\|$. For simplicity, we illustrate that on the example of the PGA. Suppose that $f \in A_1(\mathcal{D})$. Then, Ref [16] gives the bound $\|f_m\| \leq m^{-1/6}$, which does not include the norm $\|f\|$. The upper bound in (4) with $\alpha = 1/3$ gives:

$$\|f_m\| \leq \|f\|^{2/3} m^{-1/6},$$

which is better than $\|f_m\| \leq m^{-1/6}$, when $\|f\|$ is small.

In this paper, we only applied our new criterion to a few important greedy algorithms and established a new qualitative effect on optimality (in the sense of order) of some of the algorithms, which provide expansions. It will be interesting to apply this criterion to other greedy algorithms.

Funding: This research was funded by Russian Science Foundation, grant number 23-71-30001.

Data Availability Statement: Not applicable.

Acknowledgments: The author is grateful to the referees for their useful comments and suggestions. This research was supported by the Russian Science Foundation (project No. 23-71-30001) at the Lomonosov Moscow State University.

Conflicts of Interest: The author declares no conflict of interest.

References

1. Temlyakov, V. *Greedy Approximation*; Cambridge University Press: New York, USA, 2011.
2. Needell, D.; Vershynin, R. Uniform uncertainty principle and signal recovery via orthogonal matching pursuit. *Found. Comput. Math.* **2009**, *9*, 317–334.
3. Needell, D.; Tropp, J.A. CoSaMP: Iterative signal recovery from incomplete and inaccurate samples. *Appl. Comput. Harmon. Anal.* **2009**, *26*, 301–321.
4. Dai, W.; Milenkovic, O. Subspace pursuit for compressive sensing signal reconstruction. *IEEE Trans. Inf. Theory* **2009**, *55*, 2230–2249.
5. Petrova, G. Rescaled Pure Greedy Algorithm for Hilbert and Banach spaces. *Appl. Comput. Harmon. Anal.* **2016**, *41*, 852–866.
6. Dereventsov, A.V.; Temlyakov, V.N. A unified way of analyzing some greedy algorithms. *J. Funct. Anal.* **2019**, *277*, 108286.
7. DeVore, R.A.; Temlyakov, V.N. Convex optimization on Banach spaces. *Found. Comput. Math.* **2016**, *16*, 369–394.
8. Chandrasekaran, V.; Recht, B.; Parrilo, P.; Willsky, A.A. The convex geometry of linear inverse problems. *Found. Comput. Math.* **2012**, *12*, 805–849.
9. Dereventsov, A.V.; Temlyakov, V.N. Biorthogonal greedy algorithms in convex optimization. *Appl. Comput. Harmon. Anal.* **2022**, *60*, 489–511.
10. Figueiredo, M.A.; Nowak, R.D.; Wright, S.J. Gradient projection for Sparse Reconstruction: Application to Compressed Sensing and Other Inverse Problems. *IEEE Sel. Top. Signal Process.* **2007**, *1*, 586–597.
11. Jaggi, M. Revisiting Frank-Wolfe: Projection-free sparse convex optimization. *ICML* **2013**, *1*, 427–435.
12. Shalev-Shwartz, S.; Srebro, N.; Zhang, T. Trading accuracy for sparsity in optimization problems with sparsity constraints. *SIAM J. Optim.* **2010**, *20*, 2807–2832.
13. Tewari, A.P.; Ravikumar, K.; Dhillon, I.S. Greedy algorithms for structurally constrained high dimensional problems. *Adv. Neural Inf. Process. Syst.* **2011**, *24*, 882–890.
14. Zhang, T. Sequential greedy approximation for certain convex optimization problems. *IEEE Trans. Inf. Theory* **2003**, *49*, 682–691.
15. Temlyakov, V. *Multivariate Approximation*; Cambridge University Press: New York, NY, USA, 2018.
16. DeVore, R.A.; Temlyakov, V.N. Some remarks on Greedy Algorithms. *Adv. Comp. Math.* **1996**, *5*, 173–187.
17. Livshitz, E.D. On lower estimates of rate of convergence of greedy algorithms. *Izv. RAN, Ser. Matem.* **2009**, *73*, 125–144.
18. Konyagin, S.V.; Temlyakov, V.N. Rate of convergence of Pure Greedy Algorithm. *East J. Approx.* **1999**, *5*, 493–499.
19. Sil'nichenko, A.V. Rate of convergence of greedy algorithms. *Matem. Zametki*, **2004**, *76*, 628–632.
20. Nelson, J.L.; Temlyakov, V.N. Greedy expansions in Hilbert Spaces. *Trudy MIAN* **2013**, *280*, 234–246.
21. Temlyakov, V.N. Greedy Expansions in Banach Spaces. *Adv. Comput. Math.* **2007**, *26*, 431–449.
22. Donahue, M.; Gurvits, L.; Darken, C.; Sontag, E. Rate of convex approximation in non-Hilbert spaces. *Constr. Approx.* **1997**, *13*, 187–220.
23. Lindenstrauss, J.; Tzafriri, L. *Classical Banach Spaces I*; Springer: Berlin/Heidelberg, Germany, 1977.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.