



ELSEVIER

Contents lists available at ScienceDirect

Journal of Functional Analysis

www.elsevier.com/locate/jfa

A unified way of analyzing some greedy algorithms

A.V. Dereventsov^{a,*}, V.N. Temlyakov^{a,b,c}^a *Department of Mathematics, University of South Carolina, Columbia, SC 29208, USA*^b *Steklov Institute of Mathematics, Moscow, 119991, Russian Federation*^c *Lomonosov Moscow State University, Moscow, 119991, Russian Federation*

ARTICLE INFO

Article history:

Received 1 February 2018

Accepted 15 July 2019

Available online xxxx

Communicated by T. Schlumprecht

Keywords:

Banach space

Greedy approximation

Greedy algorithm

Weak Biorthogonal Greedy Algorithm

ABSTRACT

In this paper we propose a unified way of analyzing a certain kind of greedy-type algorithms in Banach spaces. We define a class of the Weak Biorthogonal Greedy Algorithms that contains a wide range of greedy algorithms. In particular, we show that the following well-known algorithms — the Weak Chebyshev Greedy Algorithm and the Weak Greedy Algorithm with Free Relaxation — belong to this class. We investigate the properties of convergence, rate of convergence, and numerical stability of the Weak Biorthogonal Greedy Algorithms. Numerical stability is understood in the sense that the steps of the algorithm are allowed to be performed with controlled computational inaccuracies. We carry out a thorough analysis of the connection between the magnitude of those inaccuracies and the convergence properties of the algorithm. To emphasize the advantage of the proposed approach, we introduce here a new greedy algorithm — the Rescaled Weak Relaxed Greedy Algorithm — from the above class, and derive the convergence results without analyzing the algorithm explicitly. Additionally, we explain how the proposed approach can be extended to some other types of greedy algorithms.

© 2019 Elsevier Inc. All rights reserved.

* Corresponding author.

E-mail addresses: dereventsov@gmail.com (A.V. Dereventsov), temlyakovv@gmail.com (V.N. Temlyakov).

<https://doi.org/10.1016/j.jfa.2019.108286>

0022-1236/© 2019 Elsevier Inc. All rights reserved.

1. Introduction

The theory of greedy approximation continues to actively develop, which is driven not only by a theoretical interest but by an applied prospect as well. As a result, many known greedy-type algorithms are being examined and new ones are being invented to fit specific applications. The tools employed in the analysis of these algorithms, however, are mostly identical, which is due to the fact that the behavior of a greedy algorithm is largely dictated by the geometry of the space rather than the particular details in the algorithm's description.

The main goal of this paper is to establish and investigate those underlying geometrical connections. In this effort we present a unified way of analyzing a certain kind of greedy-type algorithms in Banach spaces. Namely, we define a class of Weak Biorthogonal Greedy Algorithms (denoted further as $WBG\mathcal{A}$) and study the convergence properties for algorithms from this class. In particular, well-known algorithms like the Weak Chebyshev Greedy Algorithm and the Weak Greedy Algorithm with Free Relaxation belong to this class.

Motivated by the issue of numerical stability, we also discuss modifications of the class $WBG\mathcal{A}$, which allow for the steps of the algorithms to be performed with imprecise calculations. In accordance with the historical development, we call such algorithms *approximate* greedy algorithms. We show that each of the considered approximate greedy algorithms belongs to a larger class, which we introduce and call the Approximate Weak Biorthogonal Greedy Algorithms (denoted further as $AWBG\mathcal{A}$). We examine the relations between the $AWBG\mathcal{A}$ and various computational errors that can occur in practice, and emphasize the robustness of the algorithms from this class.

To the best of our knowledge, this paper is the first successful attempt to establish a unified way of analyzing a wide class of greedy-type algorithms. The step from the analysis of a specific greedy algorithm to a unified analysis of a rather wide class of algorithms motivated us to introduce here a new greedy-type algorithm — the Rescaled Weak Relaxed Greedy Algorithm — from the above class, and derive the corresponding convergence results without analyzing the algorithm explicitly.

This paper is organized as follows. In Section 2 we describe the problem setting, recall basic definitions and concepts, and introduce and state simple conclusions of our main results. Section 3 contains a discussion on greedy algorithms and historical comments related to the topic. Specifically, in this section we state the algorithms of interest and present a new greedy algorithm. In Section 4 we define the class $WBG\mathcal{A}$ and show that the algorithms of interest belong to that class. Theoretical results related to the convergence properties of algorithms from the $WBG\mathcal{A}$ are stated in Section 5 and proven in Section 6. Section 7 is devoted to the class $AWBG\mathcal{A}$ (approximate versions of algorithms from the $WBG\mathcal{A}$). In Section 8 we state the convergence results for the $AWBG\mathcal{A}$ and in Section 9 we prove the stated results. In Section 10 we comment on another type of greedy algorithms and explain how our findings can be adapted to fit those algorithms.

In Section 11 we explain and discuss our motivation for the study of the above class of algorithms.

2. Overview

The purpose of a greedy algorithm is to iteratively construct an approximation of an element $f \in X$ (the target element) by a linear combination of appropriate elements of a given set $\mathcal{D} \subset X$ (the dictionary). The exact method of selecting the elements $\{\varphi_m\}_{m=1}^\infty$ from the dictionary and constructing approximators $\{G_m\}_{m=1}^\infty$ depends on the algorithm. Generally, an m -th iteration of a greedy algorithm consists of the following stages:

- Select a suitable element φ_m from the dictionary \mathcal{D} ;
- Construct an approximation G_m for the target element f ;
- Update the remainder $f_m := f - G_m$, and proceed to the iteration $m + 1$.

This paper is focused on greedy algorithms in a Banach space setting. Let X be a real Banach space with a norm $\|\cdot\|$. We say that a set of elements (functions) \mathcal{D} from X is a dictionary if each $g \in \mathcal{D}$ has the norm bounded by one ($\|g\| \leq 1$) and $\overline{\text{span}}\mathcal{D} = X$. For convenience we assume that every dictionary \mathcal{D} is symmetric, i.e.

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

By $\mathcal{A}_1(\mathcal{D})$ we denote the closure of the convex hull of \mathcal{D} in X .

In a real Hilbert space it is natural to use the inner product to select the element φ_m from the dictionary \mathcal{D} , e.g.

$$\varphi_m \text{ is such that } \langle f_{m-1}, \varphi_m \rangle = \sup_{g \in \mathcal{D}} \langle f_{m-1}, g \rangle.$$

While in a general Banach space one does not have an inner product, there are alternative tools which can be utilized. One known approach is to select an element φ_m by solving an optimization problem, e.g.

$$\varphi_m \text{ is such that } \inf_{\lambda \in \mathbb{R}} \|f_{m-1} - \lambda \varphi_m\| = \inf_{g \in \mathcal{D}, \lambda \in \mathbb{R}} \|f_{m-1} - \lambda g\|.$$

The algorithms that select elements of the dictionary in such a way are called the X -greedy algorithms.

Another approach for selecting φ_m is to use the norming functionals (defined below) to act as an analogue of an inner product, e.g.

$$\varphi_m \text{ is such that } F_{f_{m-1}}(\varphi_m) = \sup_{g \in \mathcal{D}} F_{f_{m-1}}(g). \quad (2.1)$$

The algorithms that employ norming functionals are called the *dual* greedy algorithms.

In this paper we concentrate on the dual greedy algorithms. In Section 10 we briefly discuss how our findings can be applied to analyze the X-greedy algorithms. For a more detailed discussion on various types of greedy algorithms in the settings of Hilbert and Banach spaces, we refer the reader to the book [23] and the references therein.

For a nonzero element $f \in X$ we let F_f denote a norming (peak) functional for f :

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

The existence of such a functional is provided by the Hahn–Banach theorem, however the uniqueness is generally not guaranteed. A Banach space in which each element $f \in X \setminus \{0\}$ has a unique norming functional is the smooth space, i.e. the space with the Gateaux-differentiable norm. In such a case the value of the norming functional of an element f on an element g is the Gateaux-derivative of the norm $\|\cdot\|$ at f in the direction g , i.e.

$$F_f(g) = \lim_{u \rightarrow 0} \frac{\|f + ug\| - \|f\|}{u}.$$

Note that in many spaces of interest the norming functionals are known explicitly. For instance, in $L_p(\Omega)$, $1 \leq p < \infty$ we have

$$F_f(g) = \int_{\Omega} \frac{\text{sign}(f)|f|^{p-1}g}{\|f\|_p^{p-1}} d\mu.$$

For convenience of presentation, we introduce the norm $\|\cdot\|_{\mathcal{D}}$ in the dual space X' , associated with the dictionary \mathcal{D} and given by the formula

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

For more information on norming functionals and their connection to smoothness and other geometrical aspects of the space we refer the reader to the book [2].

Since formally the selection step (2.1) of a dual greedy algorithm might not always be attainable, it has become standard to work with *weak* greedy algorithms (proposed in [18]), which perform the greedy selection step not precisely but with a controlled relaxation given by a sequence $\tau = \{t_m\}_{m=1}^{\infty}$, $t_m \in [0, 1]$ (the weakness sequence), e.g.

$$\varphi_m \text{ is such that } F_{f_{m-1}}(\varphi_m) \geq t_m \sup_{g \in \mathcal{D}} F_{f_{m-1}}(g) = t_m \|F_{f_{m-1}}\|_{\mathcal{D}}.$$

We now can define the class of Weak Biorthogonal Greedy Algorithms, which contains a large class of weak dual greedy algorithms. We say that an algorithm belongs to the class $\mathcal{WBG\!A}$ if it satisfies the following criteria for every $m \geq 1$, where we set $f_0 := f$ and $G_0 := 0$:

- (1) **Greedy selection.** At the m -th iteration the algorithm selects an element of the dictionary $\varphi_m \in \mathcal{D}$, which satisfies

$$F_{f_{m-1}}(\varphi_m) \geq t_m \|F_{f_{m-1}}\|_{\mathcal{D}};$$

- (2) **Error reduction.** The remainder f_m is changed in such a way that

$$\|f_m\| \leq \inf_{\lambda \geq 0} \|f_{m-1} - \lambda \varphi_m\|;$$

- (3) **Biorthogonality.** The remainder f_m is biorthogonal to the constructed approximant $G_m \in \text{span}\{\varphi_1, \dots, \varphi_m\}$

$$F_{f_m}(G_m) = 0.$$

Note that the $\mathcal{WBG}\mathcal{A}$ is not an algorithm but rather the class (collection) of algorithms that possess the properties (1)–(3). In particular, in Section 3 we state some well-known dual greedy algorithms that belong to this class and introduce a new one. In Section 4 we discuss some aspects of the class $\mathcal{WBG}\mathcal{A}$ and show that the stated algorithms belong to this class.

We prove convergence and rate of convergence results for the $\mathcal{WBG}\mathcal{A}$, which provide the above mentioned unified way of analyzing greedy-type algorithms from this class. Our results on convergence and rate of convergence are formulated in terms of modulus of smoothness of the space. For a Banach space X the modulus of smoothness ρ is defined as

$$\rho(u) := \rho(u, X) = \sup_{\|x\|=\|y\|=1} \frac{\|x + uy\| + \|x - uy\|}{2} - 1. \quad (2.2)$$

The uniformly smooth Banach space is the one with the property

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

We say that the modulus of smoothness $\rho(u)$ is of power type $1 < q \leq 2$ if $\rho(u) \leq \gamma u^q$ with some $\gamma > 0$. In particular, it is known that L_p -spaces with $p \in [1, \infty)$ have modulus of smoothness of power type $\min\{p, 2\}$ (see e.g. [8, Lemma B.1]).

The main results of the paper — Theorems 5.1 and 5.2 — are stated in Section 5. Below we give a demonstrative corollary of those theorems.

Corollary 2.1. *Let X be a uniformly smooth Banach space with modulus of smoothness $\rho(u)$ of power type $1 < q \leq 2$, that is $\rho(u) \leq \gamma u^q$. Then an algorithm from the $\mathcal{WBG}\mathcal{A}$ with a constant weakness sequence $\tau = \{t\}$, $t \in (0, 1]$ converges for any target element $f \in X$.*

Moreover, if $f \in \mathcal{A}_1(\mathcal{D})$, then we have

$$\|f_m\| \leq C(q, \gamma, t) m^{-1/p},$$

where $C(q, \gamma, t) = 16\gamma^{1/q} t^{-1/p}$ and $p = q/(q-1)$.

Corollary 2.1 addresses two important characteristics of an algorithm — convergence and rate of convergence. We also consider here another essential characteristic — computational stability, which is crucial for practical implementation. A systematic study of the stability of greedy algorithms in Banach spaces was started in [21], where sufficient conditions for convergence and rate of convergence were established. A further development of the theory of stability of greedy algorithms was conducted in [5], where necessary and sufficient conditions for the convergence of a single algorithm were obtained.

In Section 7 we address the stability issue of algorithms from the class $\mathcal{WBG}\mathcal{A}$. We analyze an extended version of the $\mathcal{WBG}\mathcal{A}$, which allows the steps of an algorithm to be performed with some controlled inaccuracies. Algorithms with imprecise step evaluations are well-known in the greedy approximation theory; this type of algorithms are called the *approximate* greedy algorithms (see e.g. [13,10,11]). Permissions of approximate evaluations are natural for the applications as they simplify and speed up the execution of an algorithm.

We begin by defining the variables that represent these inaccuracies (called the error parameters or the error sequences). Let $\{\delta_m\}_{m=0}^\infty$, $\{\eta_m\}_{m=1}^\infty$ be sequences of real numbers from the interval $[0, 1]$, and let $\{\epsilon_m\}_{m=1}^\infty$ be a sequence of non-negative numbers. Denote by $\{F_m\}_{m=0}^\infty$ a sequence of such functionals that for any $m \geq 0$

$$\|F_m\| \leq 1 \quad \text{and} \quad F_m(f_m) \geq (1 - \delta_m)\|f_m\|,$$

where $\{f_m\}_{m=0}^\infty$ is the sequence of remainders produced by the algorithm.

We say that an algorithm belongs to the class $\mathcal{AWBG}\mathcal{A}$ if for every $m \geq 1$ it satisfies the following relaxed versions of conditions (1)–(3):

- (1) **Greedy selection.** At the m -th iteration the algorithm selects an element of the dictionary $\varphi_m \in \mathcal{D}$, which satisfies

$$F_{m-1}(\varphi_m) \geq t_m \|F_{m-1}\|_{\mathcal{D}};$$

- (2) **Error reduction.** The remainder f_m is changed in such a way that

$$\|f_m\| \leq (1 + \eta_m) \inf_{\lambda \geq 0} \|f_{m-1} - \lambda \varphi_m\|;$$

- (3) **Biorthogonality.** The constructed approximant $G_m \in \text{span}\{\varphi_1, \dots, \varphi_m\}$ is such that

$$|F_m(G_m)| \leq \epsilon_m.$$

Similarly to the $WBG\mathcal{A}$, in Section 8 we state convergence and rate of convergence results for the $AWBG\mathcal{A}$ (Theorems 8.1 and 8.2). For illustrative purposes we formulate here a simple corollary of those results.

Corollary 2.2. *Let X be a uniformly smooth Banach space with the modulus of smoothness $\rho(u)$ of power type $1 < q \leq 2$. Then an algorithm from the $AWBG\mathcal{A}$ with a constant weakness sequence $\tau = \{t\}$, $t \in (0, 1]$ and error parameters $\{\delta_m\}_{m=0}^\infty$, $\{\eta_m\}_{m=1}^\infty$, $\{\epsilon_m\}_{m=1}^\infty$ with*

$$\lim_{m \rightarrow \infty} \delta_m = \lim_{m \rightarrow \infty} \eta_m = \lim_{m \rightarrow \infty} \epsilon_m = 0$$

converges for any target element $f \in X$.

Moreover, if $f \in \mathcal{A}_1(\mathcal{D})$ and

$$\delta_m + \epsilon_m / \|f_m\| \leq 1/4, \quad \delta_m + \eta_{m+1} \leq C(q, \gamma, t)^{-p} \|f_m\|^p,$$

then we get

$$\|f_m\| \leq C(q, \gamma, t) m^{-1/p},$$

where $C(q, \gamma, t) = 16\gamma^{1/q} t^{-1/p}$ and $p = q/(q-1)$.

3. Dual greedy algorithms

The goal of this paper is to establish a unified way of analyzing a certain kind of dual greedy algorithms. In this section we briefly overview the most prominent algorithms and explicitly define the ones that are directly related to our new results. By $\tau := \{t_m\}_{m=1}^\infty$ we denote a weakness sequence, i.e. a given sequence of non-negative numbers $t_m \leq 1$, $m = 1, 2, 3, \dots$

We define first the Weak Chebyshev Greedy Algorithm (WCGA) (introduced in [19]) that can be thought of as a generalization of the Orthogonal Greedy Algorithm (defined and studied in [4], also known as the Orthogonal Matching Pursuit [16]) to a Banach space setting.

Weak Chebyshev Greedy Algorithm (WCGA)

Set $f_0^c := f$ and $G_0^c := 0$. Then for each $m \geq 1$

1. Take any $\varphi_m^c \in \mathcal{D}$ satisfying $F_{f_{m-1}^c}(\varphi_m^c) \geq t_m \|F_{f_{m-1}^c}\|_{\mathcal{D}}$;
2. Denote $\Phi_m^c := \text{span}\{\varphi_j^c\}_{j=1}^m$ and find $G_m^c \in \Phi_m^c$ such that

$$G_m^c = \operatorname{argmin}_{G \in \Phi_m^c} \|f - G\|;$$

3. Set $f_m^c := f - G_m^c$.

Note that in order to construct the m -th approximant G_m^c the WCGA has to solve an m -dimensional optimization problem

$$(\lambda_1^m, \dots, \lambda_m^m) = \underset{(\lambda_1, \dots, \lambda_m) \in \mathbb{R}^m}{\operatorname{argmin}} \left\| f - \sum_{k=1}^m \lambda_k \varphi_k^c \right\|, \quad G_m^c = \sum_{k=1}^m \lambda_k^m \varphi_k^c.$$

From a computational perspective, such a problem, even though convex, becomes more expensive with each iteration.

The greedy algorithm that attempts to maximally simplify the approximant construction process is the Weak Dual Greedy Algorithm (WDGA) [7], which can be thought of as a generalization of the Pure Greedy Algorithm ([4], also known as the Matching Pursuit, see e.g. [15]). In the WDGA the approximation G_m is constructed by solving a one-dimensional convex optimization problem

$$\lambda_m = \underset{\lambda \geq 0}{\operatorname{argmin}} \|f_{m-1} - \lambda \varphi_m\|, \quad G_m = G_{m-1} + \lambda_m \varphi_m.$$

Such an approach is much simpler computationally when compared to the full m -dimensional projection performed by the WCGA, however the convergence results for the WDGA are quite scarce (see [7, 12]).

There is a famous technique of constructing the next approximation as a linear combination of the previous approximant and a newly chosen element of the dictionary, i.e.

$$G_m = \omega_m G_{m-1} + \lambda_m \varphi_m$$

with appropriately chosen coefficients ω_m and λ_m . Greedy algorithms that employ such an approach are called the *relaxed* greedy algorithms. While the parameter λ_m is generally selected by the line search, the selection of ω_m is where relaxed greedy algorithms differ.

A particular approach in this direction is to omit the selection of ω_m by taking it to be the convex complement of λ_m , i.e. $\omega_m = 1 - \lambda_m$. Such a technique is in the style of the renown Frank–Wolfe algorithm (see e.g. [9, 14]). The greedy algorithm that utilizes this concept is the Weak Relaxed Greedy Algorithm [19], which constructs an approximant in the following way

$$\lambda_m = \underset{0 \leq \lambda \leq 1}{\operatorname{argmin}} \|f - ((1 - \lambda)G_{m-1} + \lambda \varphi_m)\|, \quad G_m = (1 - \lambda_m)G_{m-1} + \lambda_m \varphi_m.$$

Despite the attractive computational simplicity of this approach, it is naturally limited in the sense that the algorithm converges only for the target elements from the closure of the convex hull of the elements of the dictionary (i.e. the class $\mathcal{A}_1(\mathcal{D})$).

An easy way to mitigate this restriction is by not imposing any restraining connections on coefficients ω_m and λ_m . Perhaps the most straightforward realization of this concept,

which is closer to the style of the WCGA, is to solve the minimization problem over both parameters ω_m, λ_m . This is the idea behind the Weak Greedy Algorithm with Free Relaxation [22].

Weak Greedy Algorithm with Free Relaxation (WGAFR)

Set $f_0^e := f$ and $G_0^e := 0$. Then for each $m \geq 1$

1. Take any $\varphi_m^e \in \mathcal{D}$ satisfying $F_{f_{m-1}^e}(\varphi_m^e) \geq t_m \|f_{m-1}^e\|_{\mathcal{D}}$;
2. Find $\omega_m \in \mathbb{R}$ and $\lambda_m \geq 0$ such that

$$\|f - ((1 - \omega_m)G_{m-1}^e + \lambda_m \varphi_m^e)\| = \inf_{\lambda \geq 0, \omega \in \mathbb{R}} \|f - ((1 - \omega)G_{m-1}^e + \lambda \varphi_m^e)\|$$

and define $G_m^e := (1 - \omega_m)G_{m-1}^e + \lambda_m \varphi_m^e$;

3. Set $f_m^e := f - G_m^e$.

Note that the second step of the WGAFR is generally much simpler computationally than the second step the WCGA since at the m -th iteration the WGAFR solves a 2-dimensional optimization problem, while the WCGA has to solve an m -dimensional one.

Nevertheless, in some problems it might be preferable to simplify the algorithm even further by decoupling the choice of parameters ω_m, λ_m and hence reducing the dimensionality of the minimization problem. Based on this idea, we propose here a new relaxed dual greedy algorithm that employs two one-dimensional minimizations — the Rescaled Weak Relaxed Greedy Algorithm (RWRGA).

Rescaled Weak Relaxed Greedy Algorithm (RWRGA)

Set $f_0^r := f$ and $G_0^r := 0$. Then for each $m \geq 1$

1. Select any $\varphi_m^r \in \mathcal{D}$ satisfying $F_{f_{m-1}^r}(\varphi_m^r) \geq t_m \|f_{m-1}^r\|_{\mathcal{D}}$;
2. Find $\lambda_m \geq 0$ such that

$$\|f_{m-1}^r - \lambda_m \varphi_m^r\| = \inf_{\lambda \geq 0} \|f_{m-1}^r - \lambda \varphi_m^r\|;$$

3. Find $\mu_m \in \mathbb{R}$ such that

$$\|f - \mu_m(G_{m-1}^r + \lambda_m \varphi_m^r)\| = \inf_{\mu \in \mathbb{R}} \|f - \mu(G_{m-1}^r + \lambda_m \varphi_m^r)\|$$

and define $G_m^r := \mu_m(G_{m-1}^r + \lambda_m \varphi_m^r)$;

4. Set $f_m^r := f - G_m^r$.

Note that the idea of selecting the parameters ω_m and λ_m independently is quite popular and has been considered before in various areas. In particular, the approach of

choosing $\{\omega_m\}_{m=1}^\infty$ as values of a predefined sequence is considered in [1] for a Hilbert space, and in [22] for a general Banach space setting (see also [23, Chapter 6.5]).

Recently, an interesting new idea of building relaxed greedy algorithms was proposed in [17], where the λ_m is specified explicitly in terms of f_{m-1} and of parameters γ and q characterizing modulus of smoothness (2.2) of the Banach space X :

$$\lambda_m := \text{sign}(F_{f_{m-1}}(\varphi_m)) \|f_{m-1}\| \left(\frac{|F_{f_{m-1}}(\varphi_m)|}{2\gamma q} \right)^{1/(q-1)},$$

and the scaling parameter μ_m is chosen via the line search.

Such an algorithm is even simpler computationally than the RWRGA as it involves only one 1-dimensional optimization problem rather than two. However, this kind of further simplification might not always be beneficial as it results in certain drawbacks. For instance, one requires an *a priori* knowledge about the Banach space in terms of the values of γ and q . Moreover, since the value of λ_m directly depends on the evaluation of $F_{f_{m-1}}$, it is hard to predict how the algorithm behaves under the imprecise evaluation of the norming functionals. Surely, this algorithm is beneficial in certain settings when the smoothness of the space is known (e.g. L_p -spaces) and the precision of computations is not an issue; however the detailed comparison of various greedy algorithms is outside of the scope of this paper.

In the next section we show that despite the differences in the approximant construction and computational complexity, all three algorithms (the WCGA, the WGAFR, and the RWRGA) admit the same rate of convergence estimates (at least theoretically) due to the fact that the stages of those algorithms (selection of the element φ_m and construction of the approximant G_m) possess the same geometrical properties (1)–(3).

4. Weak Biorthogonal Greedy Algorithms

While in the previous section we introduced the variety of seemingly different dual greedy algorithms, in this section we show that the convergence analysis of those algorithms is based on underlying geometrical properties which are identical for all algorithms.

This is our motivation to establish those properties by defining the class of the Weak Biorthogonal Greedy Algorithms (denoted further as $\mathcal{WBG}\mathcal{A}$), introduced in Section 2.

We say that an algorithm belongs to the class $\mathcal{WBG}\mathcal{A}$ if sequences of remainders $\{f_m\}_{m=0}^\infty$, approximators $\{G_m\}_{m=0}^\infty$, and selected elements $\{\varphi_m\}_{m=1}^\infty$ satisfy the following conditions at every iteration $m \geq 1$:

- (1) Greedy selection: $F_{f_{m-1}}(\varphi_m) \geq t_m \|F_{f_{m-1}}\|_{\mathcal{D}}$;
- (2) Error reduction: $\|f_m\| \leq \inf_{\lambda \geq 0} \|f_{m-1} - \lambda \varphi_m\|$;
- (3) Biorthogonality: $F_{f_m}(G_m) = 0$.

Remark. Note that it is sufficient to take the infimum over $\lambda \geq 0$ in the error reduction condition (2) since the greedy selection condition (1) implies that

$$\inf_{\lambda \in \mathbb{R}} \|f_{m-1} - \lambda \varphi_m\| = \inf_{\lambda \geq 0} \|f_{m-1} - \lambda \varphi_m\|.$$

Indeed, for any $\lambda < 0$ we have

$$\begin{aligned} \|f_{m-1} - \lambda \varphi_m\| &\geq F_{f_{m-1}}(f_{m-1} - \lambda \varphi_m) \\ &\geq \|f_{m-1}\| - \lambda t_m \|F_{f_{m-1}}\|_{\mathcal{D}} \geq \|f_{m-1}\|. \end{aligned}$$

Remark. Note that the biorthogonality condition (3) guarantees that the algorithm fully utilizes the constructed approximant G_m to reduce the norm of remainder f_m . Indeed, for any $\lambda \in \mathbb{R}$ we have

$$\|f_m - \lambda G_m\| \geq F_{f_m}(f_m - \lambda G_m) = \|f_m\|.$$

The above estimate implies that the remainder f_m is orthogonal to the approximant G_m in the sense of Birkhoff–James orthogonality (see e.g. [3]).

We now prove that the algorithms from Section 3 — the WCGA, the WGAFR, and the RWRGA — belong to the class $\mathcal{WBG}\mathcal{A}$.

Proposition 4.1. *The WCGA, the WGAFR, and the RWRGA belong to the class $\mathcal{WBG}\mathcal{A}$.*

We will use the following well-known result.

Lemma 4.1 ([23, Lemma 6.9]). *Let X be a uniformly smooth Banach space and L be a finite-dimensional subspace of X . For any $f \in X \setminus L$ let f_L denote the best approximant of f from L . Then for any $\phi \in L$ we have*

$$F_{f-f_L}(\phi) = 0.$$

Proof of Proposition 4.1. From the definitions of the corresponding algorithms it is clear that the greedy selection condition (1) holds for all three algorithms.

For the WCGA / WGAFR the error reduction condition (2) follows from the fact that G_m^c / G_m^e is the best approximation to f from the subspace $\text{span}\{\varphi_1^c, \dots, \varphi_m^c\} / \text{span}\{G_{m-1}^e, \varphi_m^e\}$ respectively. Evidently, the biorthogonality condition (3) is guaranteed by the previous observation and Lemma 4.1.

For the RWRGA we write

$$\begin{aligned} \|f_m^r\| &= \inf_{\mu \in \mathbb{R}} \|f - \mu(G_{m-1}^r + \lambda_m \varphi_m^r)\| \\ &\leq \|f - G_{m-1}^r - \lambda_m \varphi_m^r\| = \inf_{\lambda \geq 0} \|f_{m-1} - \lambda \varphi_m^r\|. \end{aligned}$$

Hence condition (2) holds and condition (3) follows from Lemma 4.1 and the observation that G_m^r is the best approximation from $\text{span}\{G_{m-1}^r + \lambda_m \varphi_m^r\}$. \square

We emphasize the advantage of analyzing the class $\mathcal{WBG}\mathcal{A}$ by noting that any result proven for the class holds for each algorithm in the class. For instance, it allows us to derive theoretical estimates on the behavior of the RWRGA, which is a novel algorithm and has not been studied previously. In the next section we claim and prove the statements regarding the convergence and the rate of convergence for the $\mathcal{WBG}\mathcal{A}$ and compare them to the known results for some algorithms.

5. Convergence results for the $\mathcal{WBG}\mathcal{A}$

In the formulation of the convergence results for the $\mathcal{WBG}\mathcal{A}$ we utilize a special sequence which is defined for a given modulus of smoothness $\rho(u)$ and a given weakness sequence $\tau = \{t_m\}_{m=1}^\infty$.

Definition. For a given weakness sequence τ and a number $0 < \theta \leq 1/2$ we define $\xi_m := \xi_m(\rho, \tau, \theta)$ as a root of the equation

$$\rho(u) = \theta t_m u. \quad (5.1)$$

Remark. Note that a modulus of smoothness $\rho(u)$, defined in (2.2), is an even convex function on $(-\infty, \infty)$ and $\rho(2) \geq 1$. Hence the function

$$s(u) := \begin{cases} \rho(u)/u, & u \neq 0 \\ 0, & u = 0 \end{cases}$$

is continuous and increasing on $[0, \infty)$ with $s(2) \geq 1/2$. Thus the equation (5.1) has a unique solution $\xi_m = s^{-1}(\theta t_m)$ such that $0 < \xi_m \leq 2$.

We now state our main convergence result for the $\mathcal{WBG}\mathcal{A}$.

Theorem 5.1. Let X be a uniformly smooth Banach space with the modulus of smoothness $\rho(u)$. Let $\tau = \{t_m\}_{m=1}^\infty$ be such a sequence that for any $\theta > 0$ we have

$$\sum_{m=1}^{\infty} t_m \xi_m(\rho, \tau, \theta) = \infty. \quad (5.2)$$

Then an algorithm from the $\mathcal{WBG}\mathcal{A}$ with the weakness sequence τ converges for any target element $f \in X$.

Condition (5.2) essentially holds for weakness sequences τ that do not decay rapidly. We demonstrate this idea with the following corollary.

Corollary 5.1. *Let X be a uniformly smooth Banach space with the modulus of smoothness $\rho(u)$ of power type $1 < q \leq 2$. Assume that*

$$\sum_{m=1}^{\infty} t_m^p = \infty, \quad p = \frac{q}{q-1}. \quad (5.3)$$

Then an algorithm from the $\mathcal{WBG}\mathcal{A}$ with the weakness sequence τ converges for any target element $f \in X$.

Another important result of the paper stated below describes the rate of convergence of the $\mathcal{WBG}\mathcal{A}$.

Theorem 5.2. *Let X be a uniformly smooth Banach space with the modulus of smoothness $\rho(u)$ of power type $1 < q \leq 2$. Let $f \in X$ be a target element. Take a number $\epsilon \geq 0$ and an f^ϵ from X such that*

$$\|f - f^\epsilon\| \leq \epsilon, \quad f^\epsilon/A(\epsilon) \in \mathcal{A}_1(\mathcal{D})$$

with some number $A(\epsilon) > 0$. Then for an algorithm from the $\mathcal{WBG}\mathcal{A}$ we have

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

where $C(q, \gamma) = 4(2\gamma)^{1/q}$ and $p = q/(q-1)$.

This result indicates a pragmatic advantage of the class $\mathcal{WBG}\mathcal{A}$: it does not use any *a priori* information about the target element f or the space X and automatically adjusts to its smoothness property. Moreover, the above formulation of Theorem 5.2 covers the case of noisy data since f can be viewed as a noisy version of f^ϵ .

Theorem 5.2 shows that the rate of convergence of an algorithm depends on how close the target element is to the closure of the convex hull of the elements of the dictionary, which is a known fact in the area of greedy approximation. The following corollary gives an estimate on the rate of convergence for a target element from the class $\mathcal{A}_1(\mathcal{D})$.

Corollary 5.2. *Let X be a uniformly smooth Banach space with the modulus of smoothness $\rho(u)$ of power type $1 < q \leq 2$. Then for any target element $f \in \mathcal{A}_1(\mathcal{D})$ an algorithm from the $\mathcal{WBG}\mathcal{A}$ provides*

$$\|f_m\| \leq C(q, \gamma) \left(1 + \sum_{k=1}^m t_k^p \right)^{-1/p},$$

where $C(q, \gamma) = 4(2\gamma)^{1/q}$ and $p = q/(q-1)$.

Remark. Theorems 5.1 and 5.2 are known for the WCGA (see [19]) and the WGAFR (see [22]). For the RWRGA stated results are novel.

6. Proofs of results from Section 5

First, we prove the convergence results for the \mathcal{WBGA} . We start with the well-known lemma.

Lemma 6.1 ([23, Lemma 6.10]). *For any bounded linear functional F and any dictionary \mathcal{D} we have*

$$\|F\|_{\mathcal{D}} = \sup_{g \in \mathcal{D}} F(g) = \sup_{f \in \mathcal{A}_1(\mathcal{D})} F(f).$$

The following lemma is our main tool for analyzing the behavior of algorithms from the \mathcal{WBGA} .

Lemma 6.2 (Error Reduction Lemma). *Let X be a uniformly smooth Banach space with the modulus of smoothness $\rho(u)$. Take a number $\epsilon \geq 0$ and two elements f, f^ϵ from X such that*

$$\|f - f^\epsilon\| \leq \epsilon, \quad f^\epsilon / A(\epsilon) \in \mathcal{A}_1(\mathcal{D})$$

with some number $A(\epsilon) \geq \epsilon$. Then for an algorithm from the \mathcal{WBGA} we have

$$\|f_m\| \leq \|f_{m-1}\| \inf_{\lambda \geq 0} \left(1 - \lambda t_m A(\epsilon)^{-1} \left(1 - \frac{\epsilon}{\|f_{m-1}\|} \right) + 2\rho \left(\frac{\lambda}{\|f_{m-1}\|} \right) \right)$$

for any $m \geq 1$.

Proof of Lemma 6.2. Due to the error reduction property (2) it is sufficient to bound the quantity $\inf_{\lambda \geq 0} \|f_{m-1} - \lambda \varphi_m\|$. From the definition of modulus of smoothness (2.2) we have for any $\lambda \geq 0$

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

Next,

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

By the greedy selection condition (1) and Lemma 6.1 we get

$$\begin{aligned} F_{f_{m-1}}(\varphi_m) &\geq t_m \sup_{g \in \mathcal{D}} F_{f_{m-1}}(g) = t_m \sup_{\phi \in \mathcal{A}_1(\mathcal{D})} F_{f_{m-1}}(\phi) \\ &\geq t_m A(\epsilon)^{-1} F_{f_{m-1}}(f^\epsilon) \geq t_m A(\epsilon)^{-1} (F_{f_{m-1}}(f) - \epsilon). \end{aligned}$$

The biorthogonality property **(3)** provides

$$F_{f_{m-1}}(f) = F_{f_{m-1}}(f_{m-1} + G_{m-1}) = F_{f_{m-1}}(f_{m-1}) = \|f_{m-1}\|.$$

Combining the provided estimates completes the proof. \square

Next, we prove the general convergence result for the $\mathcal{WBG}\mathcal{A}$.

Proof of Theorem 5.1. The Error Reduction Lemma 6.2 implies that $\{\|f_m\|\}_{m=0}^\infty$ is a non-increasing sequence. Therefore, we have

$$\lim_{m \rightarrow \infty} \|f_m\| = \alpha.$$

We prove that $\alpha = 0$ by contradiction. Assume that for any m we have

$$\|f_m\| \geq \alpha > 0.$$

We set $\epsilon = \alpha/2$ and find f^ϵ such that

$$\|f - f^\epsilon\| \leq \epsilon \quad \text{and} \quad f^\epsilon/A(\epsilon) \in \mathcal{A}_1(\mathcal{D})$$

with some $A(\epsilon)$. Then, by the Error Reduction Lemma 6.2 we get

$$\|f_m\| \leq \|f_{m-1}\| \inf_{\lambda \geq 0} (1 - \lambda t_m A(\epsilon)^{-1}/2 + 2\rho(\lambda/\alpha)).$$

We specify $\theta := \alpha/8A(\epsilon)$ and take $\lambda = \alpha\xi_m(\rho, \tau, \theta)$, where ξ_m is the solution of (5.1), and obtain

$$\|f_m\| \leq \|f_{m-1}\|(1 - 2\theta t_m \xi_m).$$

Then the assumption (5.2)

$$\sum_{m=1}^{\infty} t_m \xi_m = \infty$$

implies the contradiction

$$\|f_m\| \rightarrow 0 \quad \text{as} \quad m \rightarrow \infty. \quad \square$$

Proof of Corollary 5.1. Denote $\rho^q(u) = \gamma u^q$. Then

$$\rho(u)/u \leq \rho^q(u)/u$$

and therefore for any $\theta > 0$ we have

$$\xi_m(\rho, \tau, \theta) \geq \xi_m(\rho^q, \tau, \theta).$$

For $\rho^q(u)$ we get from the definition of ξ_m (5.1) that

$$\xi_m(\rho^q, \tau, \theta) = (\theta t_m / \gamma)^{\frac{1}{q-1}}.$$

Thus condition (5.3) implies that

$$\sum_{m=1}^{\infty} t_m \xi_m(\rho, \tau, \theta) \geq \sum_{m=1}^{\infty} t_m \xi_m(\rho^q, \tau, \theta) \asymp \sum_{m=1}^{\infty} t_m^p = \infty.$$

It remains to apply Theorem 5.1. \square

Finally, we prove the rate of convergence for algorithms from the $\mathcal{WBG}\mathcal{A}$.

Proof of Theorem 5.2. Fix an $m \geq 1$. Note that the required estimate holds trivially if $A(\epsilon) \leq \epsilon$ or $\|f_m\| \leq 2\epsilon$. Consider the case $A(\epsilon) > \epsilon$ and $\|f_m\| > 2\epsilon$. Then due to monotonicity of the norms of remainders f_k we get that $\|f_k\| > 2\epsilon$ for any $k \leq m$. Hence the Error Reduction Lemma 6.2 provides

$$\|f_k\| \leq \|f_{k-1}\| \inf_{\lambda \geq 0} \left(1 - \frac{\lambda t_k A(\epsilon)^{-1}}{2} + 2\gamma \left(\frac{\lambda}{\|f_{k-1}\|} \right)^q \right). \quad (6.1)$$

Take λ to be the root of the equation

$$\frac{\lambda t_k}{4A(\epsilon)} = 2\gamma \left(\frac{\lambda}{\|f_{k-1}\|} \right)^q \quad \text{i.e.} \quad \lambda = \|f_{k-1}\|^{\frac{q}{q-1}} \left(\frac{t_k}{8\gamma A(\epsilon)} \right)^{\frac{1}{q-1}}.$$

Denote $A_q := 4(8\gamma)^{1/q-1}$, then we get from (6.1)

$$\|f_k\| \leq \|f_{k-1}\| \left(1 - \frac{1}{4} \frac{\lambda t_k}{A(\epsilon)} \right) = \|f_{k-1}\| \left(1 - \frac{t_k^p \|f_{k-1}\|^p}{A_q A(\epsilon)^p} \right).$$

Raising both sides of this inequality to the power p and taking into account the inequality $x^r \leq x$ for $r \geq 1$, $0 \leq x \leq 1$, we get

$$\|f_k\|^p \leq \|f_{k-1}\|^p \left(1 - \frac{t_k^p \|f_{k-1}\|^p}{A_q A(\epsilon)^p} \right).$$

Then by using the estimates $\|f\| \leq A(\epsilon) + \epsilon$ and $A_q > 1$, we obtain (see e.g. [4, Lemma 3.4])

$$\|f_m\|^p \leq A_q (A(\epsilon) + \epsilon)^p \left(1 + \sum_{k=1}^m t_k^p \right)^{-1},$$

which implies

$$\|f_m\| \leq C(q, \gamma)(A(\epsilon) + \epsilon) \left(1 + \sum_{k=1}^m t_k^p\right)^{-1/p}$$

with $C(q, \gamma) = A_q^{1/p} = 4(2\gamma)^{1/q}$. \square

7. Approximate Weak Biorthogonal Greedy Algorithms

In this section we address the issue of numerical stability of the algorithms from the class $\mathcal{WBG}\mathcal{A}$ by considering a wider class that we call the Approximate Weak Biorthogonal Greedy Algorithms introduced in Section 2.

The algorithms from the $\mathcal{AWBG}\mathcal{A}$ are allowed to perform the steps not exactly but with some computational inaccuracies. Such numerical inaccuracies are controlled by the weakness sequence $\{t_m\}_{m=1}^\infty$, and by the error sequences $\{\delta_m\}_{m=0}^\infty$, $\{\eta_m\}_{m=1}^\infty$, and $\{\epsilon_m\}_{m=1}^\infty$, where the first three sequences consist of numbers from the interval $[0, 1]$ and the last one is of non-negative numbers. By $\{F_m\}_{m=0}^\infty$ we denote a sequence of functionals that satisfy for any $m \geq 0$ the conditions

$$\|F_m\| \leq 1 \quad \text{and} \quad F_m(f_m) \geq (1 - \delta_m)\|f_m\|,$$

where $\{f_m\}_{m=0}^\infty$ is the sequence of remainders produced by the algorithm.

We say that an algorithm belongs to the class $\mathcal{AWBG}\mathcal{A}$ if sequences of remainders $\{f_m\}_{m=0}^\infty$, approximators $\{G_m\}_{m=0}^\infty$, and selected elements $\{\varphi_m\}_{m=1}^\infty$ satisfy the following conditions at every iteration $m \geq 1$:

- (1) Greedy selection: $F_{m-1}(\varphi_m) \geq t_m\|F_{m-1}\|_{\mathcal{D}}$;
- (2) Error reduction: $\|f_m\| \leq (1 + \eta_m) \inf_{\lambda \geq 0} \|f_{m-1} - \lambda\varphi_m\|$;
- (3) Biorthogonality: $|F_m(G_m)| \leq \epsilon_m$.

We now state the approximate versions of the algorithms from Section 3 (namely, the WCGA, the WGAFR, and the RWRGA) and show that they belong to the class $\mathcal{AWBG}\mathcal{A}$.

We start with an approximate version of the WCGA, which was studied in [21] and [5]. A more general version with both relative and absolute inaccuracies was considered in [6].

Approximate Weak Chebyshev Greedy Algorithm (AWCGA)

Set $f_0^c := f$ and $G_0^c := 0$. Then for each $m \geq 1$ perform the following steps

1. Take any $\varphi_m^c \in \mathcal{D}$ satisfying $F_{m-1}(\varphi_m^c) \geq t_m\|F_{m-1}\|_{\mathcal{D}}$;

2. Denote $\Phi_m^c := \text{span}\{\varphi_k^c\}_{k=1}^m$ and find $G_m^c \in \Phi_m^c$ such that

$$\|f - G_m^c\| \leq (1 + \eta_m) \inf_{G \in \Phi_m^c} \|f - G\|;$$

3. Set $f_m^c := f - G_m^c$.

Next, we introduce an approximate version of the WGAFR.

Approximate Weak Greedy Algorithm with Free Relaxation (AWGAFR)

Set $f_0^e := f$ and $G_0^e := 0$. Then for each $m \geq 1$

1. Take any $\varphi_m^e \in \mathcal{D}$ satisfying $F_{m-1}(\varphi_m^e) \geq t_m \|F_{m-1}\|_{\mathcal{D}}$;
2. Find $\omega_m \in \mathbb{R}$ and $\lambda_m \geq 0$ such that

$$\|f - ((1 - \omega_m)G_{m-1}^e + \lambda_m \varphi_m^e)\| \leq (1 + \eta_m) \inf_{\lambda \geq 0, \omega \in \mathbb{R}} \|f - ((1 - \omega)G_{m-1}^e + \lambda \varphi_m^e)\|$$

and define $G_m^e := (1 - \omega_m)G_{m-1}^e + \lambda_m \varphi_m^e$;

3. Set $f_m^e := f - G_m^e$.

Lastly, we define an approximate version of the RWRGA.

Approximate Rescaled Weak Relaxed Greedy Algorithm (ARWRGA)

Set $f_0^r := f$ and $G_0^r := 0$. Then for each $m \geq 1$

1. Select any $\varphi_m^r \in \mathcal{D}$ satisfying $F_{m-1}(\varphi_m^r) \geq t_m \|F_{m-1}\|_{\mathcal{D}}$;
2. Find $\lambda_m \geq 0$ such that

$$\|f_{m-1}^r - \lambda_m \varphi_m^r\| \leq \left(1 + \frac{\eta_m}{3}\right) \inf_{\lambda \geq 0} \|f_{m-1}^r - \lambda \varphi_m^r\|;$$

3. Find $\mu_m \in \mathbb{R}$ such that

$$\|f - \mu_m(G_{m-1}^r + \lambda_m \varphi_m^r)\| \leq \left(1 + \frac{\eta_m}{3}\right) \inf_{\mu \in \mathbb{R}} \|f - \mu(G_{m-1}^r + \lambda_m \varphi_m^r)\|$$

and define $G_m^r := \mu_m(G_{m-1}^r + \lambda_m \varphi_m^r)$;

4. Set $f_m^r := f - G_m^r$.

We note that steps 2 and 3 of the ARWRGA can be defined with two different error sequences and the corresponding results will hold with the appropriate minor modifications; however, for simplicity of presentation, we state those steps with the same error sequence $\{\eta_m\}_{m=1}^\infty$ with a factor of $1/3$.

Next, we show that the stated algorithms are Approximate Biorthogonal Weak Greedy Algorithms.

Proposition 7.1. *The AWCGA, the AWGAFR, and the ARWRGA belong to the class AWBGA with*

$$\epsilon_m = \inf_{\lambda > 0} \frac{\delta_m + \eta_m + 2\rho(\lambda\|G_m\|)}{\lambda}.$$

Proof of Proposition 7.1. Throughout the proof we will only use the superscripts c, e, r in the estimates that are specific for one of the algorithms, and omit superscripts when an estimate is applicable to any of the algorithms.

From the definitions of the corresponding algorithms it is clear that we only need to establish the sequence $\{\epsilon_m\}_{m=1}^\infty$ for the biorthogonality condition (3). Indeed, by the definition of the modulus of smoothness (2.2) we have for any $\lambda > 0$

$$\|f_m - \lambda G_m\| + \|f_m + \lambda G_m\| \leq 2\|f_m\| \left(1 + \rho \left(\frac{\lambda\|G_m\|}{\|f_m\|} \right) \right).$$

Assume that $F_m(G_m) \geq 0$ (case $F_m(G_m) < 0$ is handled similarly). Then

$$\|f_m + \lambda G_m\| \geq F_m(f_m + \lambda G_m) \geq (1 - \delta_m)\|f_m\| + \lambda F_m(G_m),$$

and thus

$$\|f_m - \lambda G_m\| \leq \|f_m\| \left(1 + \delta_m + 2\rho \left(\frac{\lambda\|G_m\|}{\|f_m\|} \right) \right) - \lambda F_m(G_m).$$

We now estimate $\|f_m - \lambda G_m\|$ for each of the algorithms. From the corresponding definitions we obtain

$$\begin{aligned} \|f_m^c - \lambda G_m^c\| &\geq \inf_{G \in \Phi_m^c} \|f - G\| \geq (1 + \eta_m)^{-1} \|f_m^c\| \geq (1 - \eta_m) \|f_m^c\|, \\ \|f_m^e - \lambda G_m^e\| &\geq \inf_{\mu \geq 0} \|f - \mu G_m^e\| \geq \inf_{\lambda \geq 0, w \in \mathbb{R}} \|f - ((1 - w)G_{m-1}^e + \lambda \varphi_m^e)\| \\ &\geq (1 + \eta_m)^{-1} \|f_m^e\| \geq (1 - \eta_m) \|f_m^e\|, \\ \|f_m^r - \lambda G_m^r\| &\geq \inf_{\mu \geq 0} \|f - \mu G_m^r\| = \inf_{\mu \geq 0} \|f - \mu(G_{m-1}^r + \lambda_m^r \varphi_m^r)\| \\ &\geq (1 + \eta_m)^{-1} \|f_m^r\| \geq (1 - \eta_m) \|f_m^r\|. \end{aligned}$$

Therefore

$$\lambda F_m(G_m) \leq \|f_m\| \left(\delta_m + \eta_m + 2\rho \left(\frac{\lambda\|G_m\|}{\|f_m\|} \right) \right)$$

and, since the inequality holds for any $\lambda > 0$,

$$F_m(G_m) \leq \epsilon_m := \inf_{\lambda > 0} \frac{1}{\lambda} (\delta_m + \eta_m + 2\rho(\lambda\|G_m\|)). \quad \square$$

8. Convergence results for the $\mathcal{AWBG}\mathcal{A}$

In this section we investigate how the weakness and error parameters affect the convergence properties of the class $\mathcal{AWBG}\mathcal{A}$.

First, we state the sufficient conditions on the weakness sequences that guarantee convergence of algorithms from the $\mathcal{AWBG}\mathcal{A}$.

Theorem 8.1. *Let X be a uniformly smooth Banach space with the modulus of smoothness $\rho(u)$ of power type $1 < q \leq 2$. Let sequences $\{t_m\}_{m=1}^\infty$, $\{\delta_m\}_{m=0}^\infty$, $\{\eta_m\}_{m=1}^\infty$, $\{\epsilon_m\}_{m=1}^\infty$ be such that*

$$\sum_{k=1}^{\infty} t_k^p = \infty, \quad (8.1)$$

$$\delta_{m-1} + \eta_m = o(t_m^p), \quad \epsilon_m = o(1). \quad (8.2)$$

Then an algorithm from the $\mathcal{AWBG}\mathcal{A}$ with the weakness sequence τ converges for any target element $f \in X$.

Next, we formulate the following analog of Theorem 5.2, which states such bounds for the weakness and error sequences that the convergence rate of the $\mathcal{AWBG}\mathcal{A}$ is the same as the one of the $\mathcal{WBG}\mathcal{A}$.

Theorem 8.2. *Let X be a uniformly smooth Banach space with the modulus of smoothness $\rho(u)$ of power type $1 < q \leq 2$. Let $f \in X$ be a target element. Take a number $\epsilon \geq 0$ and an f^ϵ from X such that*

$$\|f - f^\epsilon\| \leq \epsilon, \quad f^\epsilon / A(\epsilon) \in \mathcal{A}_1(\mathcal{D})$$

with some number $A(\epsilon) > 0$. Then an algorithm from the $\mathcal{AWBG}\mathcal{A}$ with the weakness and error parameters $\{t_m\}_{m=1}^\infty$, $\{\delta_m\}_{m=0}^\infty$, $\{\eta_m\}_{m=1}^\infty$, $\{\epsilon_m\}_{m=1}^\infty$, satisfying

$$\delta_m + \epsilon_m / \|f_m\| \leq 1/4, \quad (8.3)$$

$$\delta_m + \eta_{m+1} \leq \frac{1}{2} C(q, \gamma)^{-p} A(\epsilon)^{-p} t_{m+1}^p \|f_m\|^p \quad (8.4)$$

for any $m \geq 0$, provides

$$\|f_m\| \leq \max \left\{ 4\epsilon, C(q, \gamma)(A(\epsilon) + \epsilon) \left(1 + \sum_{k=1}^m t_k^p \right)^{-1/p} \right\},$$

where $C(q, \gamma) = 4q(2\gamma)^q(2/q-1)^{1/p}$ and $p = q/(q-1)$.

Corollary 8.1. *Let X be a uniformly smooth Banach space with the modulus of smoothness $\rho(u)$ of power type $1 < q \leq 2$. Then for any $f \in \mathcal{A}_1(\mathcal{D})$ an algorithm from the AWBG \mathcal{A} with the weakness and error parameters $\{t_m\}_{m=1}^\infty$, $\{\delta_m\}_{m=0}^\infty$, $\{\eta_m\}_{m=1}^\infty$, $\{\epsilon_m\}_{m=1}^\infty$, satisfying*

$$\begin{aligned}\delta_m + \epsilon_m / \|f_m\| &\leq 1/4, \\ \delta_m + \eta_{m+1} &\leq \frac{1}{2} C(q, \gamma)^{-p} t_{m+1}^p \|f_m\|^p\end{aligned}$$

for any $m \geq 0$, provides

$$\|f_m\| \leq C(q, \gamma) \left(1 + \sum_{k=1}^m t_k^p\right)^{-1/p},$$

where $C(q, \gamma) = 4q(2\gamma)^q(2/q-1)^{1/p}$ and $p = q/(q-1)$.

Since Theorems 8.1 and 8.2 involve generally unknown weakness parameters $\{\epsilon_m\}_{m=1}^\infty$ and thus are not as straightforward as their analogues from Section 5 (i.e. Theorems 5.1 and 5.2), we derive the following result for the algorithms from Section 7 — the AWCGA, the AWGAFR, and the ARWRGA.

Remark. Theorems 8.1 and 8.2 are known for the AWCGA (see [21] and [5]). For the AWGAFR and the ARWRGA stated results are novel.

Proposition 8.1. *Let X be a uniformly smooth Banach space with the modulus of smoothness $\rho(u)$ of power type $1 < q \leq 2$. Let sequences $\{t_m\}_{m=1}^\infty$, $\{\delta_m\}_{m=0}^\infty$, $\{\eta_m\}_{m=1}^\infty$ be such that*

$$\sum_{k=1}^{\infty} t_k^p = \infty, \quad \delta_{m-1} + \eta_m = o(t_m^p).$$

Then the AWCGA, the AWGAFR, and the ARWRGA converge for any dictionary \mathcal{D} and any target element $f \in X$.

Moreover, if $f \in \mathcal{A}_1(\mathcal{D})$ and the sequences $\{\delta_m\}_{m=0}^\infty$ and $\{\eta_m\}_{m=1}^\infty$ satisfy

$$\begin{aligned}\delta_m &\leq 64^{-p} \gamma^{1-p} \|f_m\|^p t_{m+1}^p, \\ \eta_m &\leq 64^{-p} \gamma^{1-p} \|f_m\|^p t_m^p,\end{aligned}\tag{8.5}$$

then the AWCGA, the AWGAFR, and the ARWRGA provide

$$\|f_m\| \leq C(q, \gamma) \left(1 + \sum_{k=1}^m t_k^p\right)^{-1/p},$$

where $C(q, \gamma) = 4q(2\gamma)^q(2/q-1)^{1/p}$ and $p = q/(q-1)$.

9. Proofs of results from Section 8

First, we need the following analogue of Lemma 6.2.

Lemma 9.1 (*Error Reduction Lemma*). *Let X be a uniformly smooth Banach space with the modulus of smoothness $\rho(u)$. Take a number $\epsilon \geq 0$ and two elements f, f^ϵ from X such that*

$$\|f - f^\epsilon\| \leq \epsilon, \quad f^\epsilon / A(\epsilon) \in \mathcal{A}_1(\mathcal{D})$$

with some number $A(\epsilon) \geq \epsilon$. Then for an algorithm from the AWBG \mathcal{A} we have for any $m \geq 0$ and $\lambda \geq 0$

$$\begin{aligned} \|f_{m+1}\| \leq & \|f_m\| (1 + \eta_{m+1}) \left(1 + \delta_m + 2\rho\left(\frac{\lambda}{\|f_m\|}\right) \right. \\ & \left. - \lambda t_{m+1} A^{-1}(\epsilon) \left(1 - \delta_m - \frac{\epsilon_m + \epsilon}{\|f_m\|} \right) \right). \end{aligned}$$

Proof of Lemma 9.1. The error reduction property (2) implies for any $\lambda \geq 0$

$$\|f_{m+1}\| \leq (1 + \eta_{m+1}) \|f_m - \lambda \phi_{m+1}\|.$$

From the definition of modulus of smoothness (2.2) we have

$$\|f_m - \lambda \phi_{m+1}\| \leq 2\|f_m\| (1 + \rho(\lambda/\|f_m\|)) - \|f_m + \lambda \phi_{m+1}\|$$

and the greedy selection condition (1) implies with Lemma 6.1

$$\begin{aligned} \|f_m + \lambda \phi_{m+1}\| & \geq F_m(f_m + \lambda \phi_{m+1}) = F_m(f_m) + \lambda F_m(\phi_{m+1}) \\ & \geq (1 - \delta_m) \|f_m\| + \lambda t_{m+1} \sup_{\phi \in \mathcal{A}_1(\mathcal{D})} F_m(\phi) \\ & \geq (1 - \delta_m) \|f_m\| + \lambda t_{m+1} A^{-1}(\epsilon) F_m(f^\epsilon). \end{aligned}$$

Assumption on f and f^ϵ and the biorthogonality property (3) provide

$$\begin{aligned} F_m(f^\epsilon) & \geq F_m(f) - \epsilon = F_m(f_m + G_m) - \epsilon \\ & \geq (1 - \delta_m) \|f_m\| - \epsilon_m - \epsilon. \end{aligned}$$

Combining the above estimates we get

$$\begin{aligned} \|f_{m+1}\| \leq & (1 + \eta_{m+1}) \left(2\|f_m\| (1 + \rho(\lambda/\|f_m\|)) - (1 - \delta_m) \|f_m\| \right. \\ & \left. - \lambda t_{m+1} A^{-1}(\epsilon) ((1 - \delta_m) \|f_m\| - \epsilon_m - \epsilon) \right), \end{aligned}$$

which completes the proof. \square

Now we can prove the convergence result for the $\mathcal{AWBG}\mathcal{A}$.

Proof of Theorem 8.1. Assume that for some $f \in \mathcal{D}$ an algorithm from the $\mathcal{AWBG}\mathcal{A}$ does not converge, i.e. there exists $\alpha > 0$ and a sequence $\{m_\nu\}_{\nu=0}^\infty$ such that for any $\nu \geq 0$

$$\|f_{m_\nu}\| > \alpha. \quad (9.1)$$

We will obtain a contradiction by showing that this assumption does not hold for any sufficiently big ν .

Fix $\epsilon = \alpha/4$, take corresponding f^ϵ and $A = A(\epsilon)$, and denote

$$\varkappa := \frac{\alpha^p}{2(16A)^p(2\gamma)^{p-1}}, \quad p = \frac{q}{q-1}.$$

Find such K that for any $k \geq K$ the following estimates hold

$$\delta_k + 2\epsilon_k/\alpha \leq 1/4, \quad (9.2)$$

$$\delta_k + \eta_{k+1} \leq \varkappa t_{k+1}^p. \quad (9.3)$$

Note that existence of such K is guaranteed by conditions (8.2).

Take any $k \geq K$. If $\|f_k\| \leq \alpha/2$, then by the error reduction condition we have

$$\|f_{k+1}\| \leq (1 + \eta_{k+1})\|f_k\| \leq 2\|f_k\| \leq \alpha. \quad (9.4)$$

If $\|f_k\| > \alpha/2$, then Lemma 9.1 and condition (9.2) provide

$$\begin{aligned} \|f_{k+1}\| &\leq \|f_k\|(1 + \eta_{k+1}) \left(1 + \delta_k + 2\rho \left(\frac{\lambda}{\|f_k\|} \right) - \frac{\lambda t_{k+1}}{A} \left(1 - \delta_k - \frac{\epsilon_k + \alpha/4}{\|f_k\|} \right) \right) \\ &\leq \|f_k\|(1 + \eta_{k+1}) \left(1 + \delta_k + 2\gamma \left(\frac{2\lambda}{\alpha} \right)^q - \lambda \frac{t_{k+1}}{4A} \right). \end{aligned}$$

Let λ_k be the positive root of the equation

$$\lambda \frac{t_{k+1}}{8A} = 2\gamma \left(\frac{2\lambda}{\alpha} \right)^q,$$

which implies that

$$\lambda_k = \left(\frac{\alpha^q t_{k+1}}{16 \cdot 2^q A \gamma} \right)^{\frac{1}{q-1}} \quad \text{and} \quad \lambda_k \frac{t_{k+1}}{8A} = \frac{\alpha^p t_{k+1}^p}{(16A)^p (2\gamma)^{p-1}} = 2\varkappa t_{k+1}^p.$$

The choice $\lambda = \lambda_k$ in the previous estimate guarantees

$$\|f_{k+1}\| \leq \|f_k\|(1 + \eta_{k+1})(1 + \delta_k - 2\kappa t_{k+1}^p),$$

and condition (9.3) provides

$$\|f_{k+1}\| \leq \|f_k\|(1 - \kappa t_{k+1}^p). \quad (9.5)$$

Note that estimates (9.4) and (9.5) guarantee that if $\|f_k\| \leq \alpha$ then $\|f_{k+1}\| \leq \alpha$. By condition (8.1) there exists such $N > K$ that

$$\|f_K\| \prod_{k=K+1}^N (1 - \kappa t_k^p) \leq \alpha.$$

Therefore for any ν such that $m_\nu \geq N$ we have

$$\|f_{m_\nu}\| \leq \alpha,$$

which contradicts the assumption (9.1) and proves Theorem 8.1. \square

We proceed with the proof of the estimate on the rate of convergence of the $\mathcal{AWBG}\mathcal{A}$.

Proof of Theorem 8.2. Take any $k \geq 0$. If $\|f_k\| \leq 2\epsilon$, then by the error reduction condition we have

$$\|f_{k+1}\| \leq (1 + \eta_{k+1})\|f_k\| \leq 2\|f_k\| \leq 4\epsilon. \quad (9.6)$$

If $\|f_k\| > 2\epsilon$ then by Lemma 9.1 and condition (8.3) we get for any $\lambda > 0$

$$\begin{aligned} \|f_{k+1}\| &\leq \|f_k\|(1 + \eta_{k+1}) \left(1 + \delta_k + 2\rho \left(\frac{\lambda}{\|f_k\|} \right) - \frac{\lambda t_{k+1}}{A} \left(1 - \delta_k - \frac{\epsilon_k + \epsilon}{\|f_k\|} \right) \right) \\ &\leq \|f_k\|(1 + \eta_{k+1}) \left(1 + \delta_k + 2\gamma \lambda^q \|f_k\|^{-q} - \frac{\lambda t_{k+1}}{4A} \right). \end{aligned}$$

Hence, by taking infimum over all $\lambda > 0$ and using condition (8.4), we obtain

$$\begin{aligned} \|f_{k+1}\| &\leq \inf_{\lambda > 0} \|f_k\|(1 + \eta_{k+1}) \left(1 + \delta_k + 2\gamma \lambda^q \|f_k\|^{-q} - \frac{\lambda t_{k+1}}{4A} \right) \\ &= \|f_k\|(1 + \eta_{k+1})(1 + \delta_k - (q-1)(4qA)^{-p}(2\gamma)^{1-p}t_{k+1}^p\|f_k\|^p) \\ &\leq \|f_k\| \left(1 - \frac{1}{2}(q-1)(4qA)^{-p}(2\gamma)^{1-p}t_{k+1}^p\|f_k\|^p \right) \\ &= \|f_k\|(1 - C^{-p}A^{-p}t_{k+1}^p\|f_k\|^p), \end{aligned} \quad (9.7)$$

where

$$C = C(q, \gamma) := 4q(2\gamma)^q \left(\frac{2}{q-1} \right)^{1/p}.$$

Therefore estimates (9.6) and (9.7) guarantee that as long as conditions (8.3) and (8.4) are satisfied, $\|f_k\| \leq 4\epsilon$ implies $\|f_{k+1}\| \leq 4\epsilon$. Thus for any $m \geq 0$ either $\|f_m\| \leq 4\epsilon$ or $\|f_k\| > 4\epsilon$ for any $k \leq m$. In the latter case, by using estimate (9.7) and taking into account that $\|f\| \leq \|f^\epsilon\| + \epsilon \leq A(\epsilon) + \epsilon$ and $C > 1$ (due to the estimate $\gamma \geq 2^{-q}$ which follows from $u-1 \leq \rho(u) \leq \gamma u^q$), we complete the proof in the same way as in the proof of Theorem 5.2 in Section 6. \square

Finally, we prove the convergence and the rate of convergence of the AWCGA, the AWGAFR, and the ARWRGA.

Proof of Proposition 8.1. It is enough to verify that ϵ_m satisfies conditions of Theorems 8.1 and 8.2. Indeed, from Proposition 7.1 we get

$$\begin{aligned} \epsilon_m &\leq \inf_{\lambda>0} \frac{1}{\lambda} (\delta_m + \eta_m + 2\gamma \|G_m\|^q \lambda^q) \\ &= q(q-1)^{-1/p} (\delta_m + \eta_m)^{1/p} (2\gamma)^{1/q} \|G_m\| \\ &\leq 6(2\gamma)^{1/q} (\delta_m + \eta_m)^{1/p}, \end{aligned}$$

since for all considered algorithms we have $\|G_m\| \leq \|f_m\| + \|f\| \leq 3\|f\| \leq 3$. Therefore $\epsilon_m \rightarrow 0$ as $m \rightarrow \infty$ and Theorem 8.1 guarantees convergence.

Assume now that conditions (8.5) are satisfied. We will show that estimates (8.3) and (8.4) hold as well. Indeed, using the estimates $\gamma \geq 2^{-q}$ and $\|f_m\| \leq 2\|f\| \leq 2$ we obtain

$$\delta_m + \epsilon_m / \|f_m\| \leq 64^{-p} + 6(2\gamma)^{1/q} (2 \cdot 64^{-p} \gamma^{1-p})^{1/p} = 64^{-p} + \frac{3}{16} < \frac{1}{4}.$$

Similarly, since $\|f_{m+1}\| \leq 2\|f_m\|$, we get

$$\begin{aligned} \delta_m + \eta_{m+1} &\leq (2^{-p} + 1) 32^{-p} \gamma^{1-p} \|f_m\|^p t_{m+1}^p \\ &\leq 16^{-p} (2\gamma)^{1-p} t_{m+1}^p \|f_m\|^p \\ &\leq \frac{q-1}{2} (4q)^{-p} (2\gamma)^{1-p} t_{m+1}^p \|f_m\|^p. \end{aligned}$$

Hence Theorem 8.2 guarantees the stated rate of convergence. \square

10. X-greedy algorithms

Algorithms from the class $\mathcal{WBG}\mathcal{A}$ perform the greedy selection step based on the norming functional F_{f_m} . For this reason, this type of algorithms is called *dual* greedy

algorithms. As mentioned in Section 2, there exists another natural class of greedy algorithms — the X -greedy type algorithms — that solve a simple optimization problem rather than compute the norming functional. Thus, in certain applications, such algorithms might be more suitable for implementation than the dual greedy algorithms.

Generally one can obtain an X -greedy algorithm from a given dual algorithm by replacing the greedy selection step

$$\varphi_m = \operatorname{argmin}_{g \in \mathcal{D}} F_{f_{m-1}}(g)$$

with a minimization over all possible choices of a newly added element from the dictionary

$$(\varphi_m, \lambda_m) = \operatorname{argmin}_{g \in \mathcal{D}, \lambda \in \mathbb{R}} \|f_{m-1} - \lambda g\|.$$

It is also known that convergence results for such X -greedy companions can be derived from the proofs of the corresponding results for the dual greedy algorithms. While the exact connection between dual and X -greedy algorithms and the detailed analysis of the latter is outside of the scope of this paper, we illustrate the concept by defining and deriving convergence results for the Rescaled Relaxed X -Greedy Algorithm (RRXGA) — the X -greedy companion for the RWRGA. For a more detailed discussion on X -greedy algorithms and related theoretical results see [20] and [23].

Rescaled Relaxed X -Greedy Algorithm (RRXGA)

Set $f_0 := f$ and $G_0 := 0$. Then for each $m \geq 1$

1. Select any $\varphi_m \in \mathcal{D}$ and $\lambda_m \in \mathbb{R}$ (we assume existence) satisfying

$$\|f_{m-1} - \lambda_m \varphi_m\| = \inf_{g \in \mathcal{D}, \lambda \in \mathbb{R}} \|f_{m-1} - \lambda g\|;$$

2. Find $\mu_m \in \mathbb{R}$ such that

$$\|f - \mu_m(G_{m-1} + \lambda_m \varphi_m)\| = \inf_{\mu \in \mathbb{R}} \|f - \mu(G_{m-1} + \lambda_m \varphi_m)\|$$

and define $G_m := \mu_m(G_{m-1} + \lambda_m \varphi_m)$;

3. Set $f_m := f - G_m$.

Next, we state one of the corresponding convergence results.

Theorem 10.1. *Let X be a uniformly smooth Banach space with the modulus of smoothness $\rho(u)$ of power type $1 < q \leq 2$. Let $f \in X$ be a target element. Take a number $\epsilon \geq 0$ and an f^ϵ from X such that*

$$\|f - f^\epsilon\| \leq \epsilon, \quad f^\epsilon/A(\epsilon) \in \mathcal{A}_1(\mathcal{D})$$

with some number $A(\epsilon) > 0$. Then the RRXGA provides

$$\|f_m\| \leq \max \left\{ 2\epsilon, C(q, \gamma)(A(\epsilon) + \epsilon)(1 + m)^{-1/p} \right\}, \quad p = q/(q - 1),$$

where $C(q, \gamma) = 4(2\gamma)^{1/q}$.

In order to prove the stated theorem, we need the following lemma, which is a generalization of the Error Reduction Lemma 6.2 and can be largely derived from its proof.

Lemma 10.2 (*Generalized Error Reduction Lemma*). Let X be a uniformly smooth Banach space with the modulus of smoothness $\rho(u)$. Take a number $\epsilon \geq 0$ and two elements f, f^ϵ from X such that

$$\|f - f^\epsilon\| \leq \epsilon, \quad f^\epsilon/A(\epsilon) \in \mathcal{A}_1(\mathcal{D})$$

with some number $A(\epsilon) \geq \epsilon$. Suppose that f is represented as $f = f' + G'$ in such a way that $F_{f'}(G') = 0$ and an element $\varphi' \in \mathcal{D}$ is chosen to satisfy

$$F_{f'}(\varphi') \geq \theta \|F_{f'}\|_{\mathcal{D}}, \quad \theta \in [0, 1].$$

Then we have

$$\inf_{\lambda \geq 0} \|f' - \lambda \varphi'\| \leq \|f'\| \inf_{\lambda \geq 0} \left(1 - \lambda \theta A(\epsilon)^{-1} \left(1 - \frac{\epsilon}{\|f'\|} \right) + 2\rho \left(\frac{\lambda}{\|f'\|} \right) \right).$$

Proof of Theorem 10.1. It is sufficient to prove an analog of the Error Reduction Lemma 6.2 for the RRXGA. Let G_m and f_m be the approximant and the residual of the RRXGA at the m -th iteration. Using Lemma 10.2 with $G' = G_m$ and $f' = f_m$ find such $\varphi' \in \mathcal{D}$ that the generalized error reduction holds. Hence it follows from the definition of the RRXGA that for the residual f_{m+1} we have

$$\begin{aligned} \|f_{m+1}\| &\leq \inf_{\lambda \geq 0} \|f_m - \lambda \varphi_{m+1}\| \leq \inf_{\lambda \geq 0} \|f' - \lambda \varphi'\| \\ &\leq \|f_m\| \inf_{\lambda \geq 0} \left(1 - \lambda \theta A(\epsilon)^{-1} \left(1 - \frac{\epsilon}{\|f_m\|} \right) + 2\rho \left(\frac{\lambda}{\|f_m\|} \right) \right). \end{aligned}$$

Taking supremum over all $\theta \leq 1$ gives the Error Reduction Lemma for the RRXGA. We complete the proof of Theorem 10.1 in the same way as we derived Theorem 5.2 from Lemma 6.2. \square

11. Discussion

This paper is devoted to a theoretical study of fundamental techniques used in sparse representation of data, which can be structured or unstructured. The contemporary

challenge is unstructured data, which comes from different sources such as medical, engineering, and networks, among others. In order to apply a sparsity based method one needs to use a dictionary (a representation system), which provides sparse representation for the data at hand. In some cases the data structure itself gives us an idea of which dictionary to use. For instance, in signal processing, say, in music processing, it is natural to use the trigonometric system or the Gabor system as a dictionary for sparse representation. In other cases, especially in the case of unstructured data, we need to learn a dictionary providing sparse representation for the given data. Dictionary learning is an important and rapidly developing area of numerical mathematics. Clearly, we cannot expect that a dictionary learnt from given unstructured data will have some special structure. Therefore, the theory broadly applicable to Big Data problems must address the problem of sparse representation with respect to an arbitrary (structured and unstructured) dictionary. This motivates our setting with greedy approximation with respect to an arbitrary dictionary.

In order to address the contemporary needs of data managing, a very general model of approximation with regard to a redundant system (dictionary) has been considered in many recent papers. As such a model, we choose a Banach space X with elements as target functions and an arbitrary system \mathcal{D} of elements of this space such that the closure of $\text{span } \mathcal{D}$ coincides with X as a representation system. We would like to have an algorithm of constructing m -term approximants that adds at each step only one new element from \mathcal{D} and keeps elements of \mathcal{D} obtained at the previous steps. This requirement is an analogue of *on-line* computation that is desirable in practical algorithms. Clearly, we are looking for good algorithms which converge for each target function. It is not obvious that such an algorithm exists in a setting at the above level of generality (X, \mathcal{D} are arbitrary).

An important argument that motivates us to study this problem in the infinite dimensional setting is that in many contemporary data management applications the dimensionality d of an ambient space \mathbb{R}^d is quite large, which negatively affects the rate of convergence of many native algorithms (the so-called *curse of dimensionality*). As greedy algorithms are naturally designed for use in infinitely-dimensional spaces, our results provide bounds on the convergence rate that are independent of the dimensionality of the space.

While in many applications users are satisfied with a Hilbert space setting, we comment on our interest in a more general Banach space setting. An *a-priori* argument for a Banach space approach is that the spaces L_p are very natural in many real-life applications and hence they should be studied along with the L_2 space. An *a-posteriori* argument is that the study of greedy approximation in Banach spaces has discovered that the behavior of greedy algorithms is governed by the smoothness of the space that is generally described by the modulus of smoothness $\rho(u)$. It is known that the spaces L_p for $2 \leq p < \infty$ have modulus of smoothness of the power type 2, i.e. the same as a Hilbert space. 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 ,

$2 \leq p < \infty$ as they only rely on the geometry of the unit sphere of the space expressed in the form $\rho(u) \leq \gamma u^2$. Also, we note that in the case L_p -space the implementation of a dual greedy algorithm is not substantially more difficult than in L_2 since the norming functionals have the following simple explicit form

$$F_f(g) = \int_{\Omega} \frac{\text{sign}(f)|f|^{p-1}g}{\|f\|_p^{p-1}} d\mu.$$

Thus at the greedy selection step of the algorithm one has to find an approximate solution to the optimization problem

$$\text{maximize} \int_{\Omega} \text{sign}(f_m)|f_m|^{p-1}g d\mu \quad \text{s.t.} \quad g \in \mathcal{D},$$

which is not much more difficult in the case $p \neq 2$ than in the case $p = 2$.

Finally, we comment on the issue of numerical stability. Since the selection step of a greedy algorithm is usually the most expensive (as we need to search over the whole dictionary), there is a natural desire to simplify it by relaxing the search criteria, which results in a *weak* greedy algorithm, where instead of finding $\sup_{g \in \mathcal{D}} F(g)$ we are satisfied with an element $\varphi \in \mathcal{D}$ such that

$$F(\varphi) \geq t_m \sup_{g \in \mathcal{D}} F(g).$$

Theorem 5.2 demonstrates a major (and surprising) advantage of the Weak Biorthogonal Greedy Algorithms — the weak version of an algorithm from the $\mathcal{WBG}\mathcal{A}$ with the weakness sequence $\tau = \{t\}$, $t \in (0, 1]$ has the same convergence properties as the strong version of the algorithm with $t = 1$. Furthermore, Theorem 8.1 guarantees in this case the stable convergence under very mild conditions on the relative errors — they must go to zero with $m \rightarrow \infty$.

We have carried over a thorough study of three fundamental properties of the Weak Biorthogonal Greedy Algorithms — convergence, rate of convergence, and computational stability. The above discussion suggests that algorithms from the $\mathcal{WBG}\mathcal{A}$ are easily implementable in many Banach spaces of interest (e.g. L_p) and possess the necessary theoretical properties that are essential for realization in real-life applications.

Acknowledgments

The work was supported by the Russian Federation Government Grant No. 14.W03.31.0031.

The paper contains results obtained in frames of the program “Center for the storage and analysis of big data”, supported by the Ministry of Science and High Education of Russian Federation (contract 11.12.2018N°13/1251/2018 between the Lomonosov

Moscow State University and the Fond of support of the National technological initiative projects).

We are grateful to the reviewer for useful comments which resulted in an improved structure of the paper.

References

- [1] A.R. Barron, A. Cohen, W. Dahmen, R.A. DeVore, Approximation and learning by greedy algorithms, *Ann. Appl. Stat.* 36 (2008) 64–94.
- [2] B. Beauzamy, *Introduction to Banach Spaces and Their Geometry*, North-Holland, 1982.
- [3] G. Birkhoff, et al., Orthogonality in linear metric spaces, *Duke Math. J.* 1 (1935) 169–172.
- [4] R.A. DeVore, V.N. Temlyakov, Some remarks on greedy algorithms, *Adv. Comput. Math.* 5 (1996) 173–187.
- [5] A. Dereventsov, On the approximate weak Chebyshev greedy algorithm in uniformly smooth Banach spaces, *J. Math. Anal. Appl.* 436 (2016) 288–304.
- [6] A. Dereventsov, On the generalized approximate weak Chebyshev greedy algorithm, *Studia Math.* 237 (2017) 153–175.
- [7] S. Dilworth, D. Kutzarova, V.N. Temlyakov, Convergence of some greedy algorithms in Banach spaces, *J. Fourier Anal. Appl.* 8 (2002) 489–506.
- [8] M.J. Donahue, C. Darken, L. Gurvits, E. Sontag, Rates of convex approximation in non-Hilbert spaces, *Constr. Approx.* 13 (1997) 187–220.
- [9] M. Frank, P. Wolfe, An algorithm for quadratic programming, *Naval Res. Logist. Quart.* 3 (1956) 95–110.
- [10] V.V. Galatenko, E.D. Livshits, On the convergence of approximate weak greedy algorithms, *East J. Approx.* 9 (2003) 43–50.
- [11] V.V. Galatenko, E.D. Livshits, Generalized approximate weak greedy algorithms, *Math. Notes* 78 (2005) 170–184.
- [12] M. Ganichev, N.J. Kalton, Convergence of the weak dual greedy algorithm in L_p -spaces, *J. Approx. Theory* 124 (2003) 89–95.
- [13] R. Gribonval, M. Nielsen, Approximate weak greedy algorithms, *Adv. Comput. Math.* 14 (2001) 361–378.
- [14] M. Jaggi, Revisiting Frank–Wolfe: projection-free sparse convex optimization, in: *ICML* (1), 2013, pp. 427–435.
- [15] S.G. Mallat, Z. Zhang, Matching pursuits with time-frequency dictionaries, *IEEE Trans. Signal Process.* 41 (1993) 3397–3415.
- [16] Y.C. Pati, R. Rezaifar, P.S. Krishnaprasad, Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition, in: *Proceedings of 27th Asilomar Conference on Signals, Systems and Computers*, IEEE, 1993, pp. 40–44.
- [17] G. Petrova, Rescaled pure greedy algorithm for Hilbert and Banach spaces, *Appl. Comput. Harmon. Anal.* 41 (2016) 852–866.
- [18] V.N. Temlyakov, Weak greedy algorithms, *Adv. Comput. Math.* 12 (2000) 213–227.
- [19] V.N. Temlyakov, Greedy algorithms in Banach spaces, *Adv. Comput. Math.* 14 (2001) 277–292.
- [20] V.N. Temlyakov, Nonlinear methods of approximation, *Found. Comput. Math.* 3 (2003).
- [21] V.N. Temlyakov, Greedy-type approximation in Banach spaces and applications, *Constr. Approx.* 21 (2005) 257–292.
- [22] V.N. Temlyakov, Relaxation in greedy approximation, *Constr. Approx.* 28 (2008) 1–25.
- [23] V.N. Temlyakov, *Greedy Approximation*, Cambridge University Press, 2011.