V.N. Temlyakov

Greedy expansions in convex optimization

V.N. Temlyakov *

May 3, 2012

Abstract

This paper is a follow up to the previous author's paper on convex optimization. In that paper we began the process of adjusting greedy-type algorithms from nonlinear approximation for finding sparse solutions of convex optimization problems. We modified there three the most popular in nonlinear approximation in Banach spaces greedy algorithms – Weak Chebyshev Greedy Algorithm, Weak Greedy Algorithm with Free Relaxation and Weak Relaxed Greedy Algorithm – for solving convex optimization problems. We continue to study sparse approximate solutions to convex optimization problems. It is known that in many engineering applications researchers are interested in an approximate solution of an optimization problem as a linear combination of elements from a given system of elements. There is an increasing interest in building such sparse approximate solutions using different greedy-type algorithms. In this paper we concentrate on greedy algorithms that provide expansions, which means that the approximant at the mth iteration is equal to the sum of the approximant from the previous iteration ($(m-1)$th iteration) and one element from the dictionary with an appropriate coefficient. The problem of greedy expansions of elements of a Banach space is well studied in nonlinear approximation theory. At a first glance the setting of a problem of expansion of a given element and the setting of the problem of expansion in an optimization problem are very different. However, it turns out that the same technique can be used for solving both problems. We show how the technique developed in nonlinear approximation theory, in particular, the greedy expansions technique can be adjusted

for finding a sparse solution of an optimization problem given by an
expansion with respect to a given dictionary.

1 Introduction

This paper is a follow up to the author's paper [13] on convex optimiza-
tion. In [13] we began the process of adjusting greedy-type algorithms from
nonlinear approximation for finding sparse solutions of convex optimization
problems. We modified in [13] three the most popular in nonlinear approx-
imation in Banach spaces greedy algorithms – Weak Chebyshev Greedy Al-
gorithm, Weak Greedy Algorithm with Free Relaxation and Weak Relaxed
Greedy Algorithm – for solving convex optimization problems. We continue
to study sparse approximate solutions to convex optimization problems. We
apply the technique developed in nonlinear approximation known under the
name of *greedy approximation*. A typical problem of convex optimization is
to find an approximate solution to the problem

$$\inf_x E(x) \tag{1.1}$$

under assumption that E is a convex function. Usually, in convex optimiza-
tion function E is defined on a finite dimensional space $\mathbb{R}^n$ (see [2], [6]).
Recent needs of numerical analysis call for consideration of the above opti-
mization problem on an infinite dimensional space, for instance, a space of
continuous functions. Thus, we consider a convex function E defined on a
Banach space X. It is pointed out in [15] that in many engineering applica-
tions researchers are interested in an approximate solution of problem (1.1)
as a linear combination of elements from a given system $\mathcal{D}$ of elements. There
is an increasing interest in building such sparse approximate solutions using
different greedy-type algorithms (see, for instance, [15], [7], [3],[14], and [13]).
The problem of approximation of a given element $f \in X$ by linear combina-
tions of elements from $\mathcal{D}$ is well studied in nonlinear approximation theory
(see, for instance [4], [11], [12]). Many of known greedy-type algorithms pro-
vide such approximation in a form of expansion of a given element into a
series with respect to a given dictionary $\mathcal{D}$. In the paper [13] we showed how
some of the greedy algorithms that provide good approximation, but not
an expansion, can be adjusted for solving an optimization problem. In this
paper we concentrate on greedy algorithms that provide expansions, which
means that the approximant at the mth iteration is equal to the sum of

the approximant from the previous iteration ($(m-1)$th iteration) and one element from the dictionary with an appropriate coefficient.

We point out that at a first glance the setting of a problem of expansion of a given element and the setting of the expansion problem in an optimization are very different. However, it turns out that the same technique can be used for solving both problems. We show how the technique developed in nonlinear approximation theory, in particular, the greedy expansions technique can be adjusted for finding a sparse solution of an optimization problem (1.1) given by an expansion with respect to a given dictionary $\mathcal{D}$.

We begin with a brief description of greedy expansion methods in Banach spaces. Let X be a Banach space with norm $\|\cdot\|$. We say that a set of elements (functions) $\mathcal{D}$ from X is a dictionary, respectively, symmetric dictionary, if each $g \in \mathcal{D}$ has norm bounded by one ($\|g\| \leq 1$),

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

and the closure of span $\mathcal{D}$ is X. In this paper symmetric dictionaries are considered. We denote the closure (in X) of the convex hull of $\mathcal{D}$ by $A_1(\mathcal{D})$. For a nonzero element $f \in X$ we let F_f denote a norming (peak) functional for f:

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

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

We assume that the set

$$D := \{x : E(x) \leq E(0)\}$$

is bounded. For a bounded set S define the modulus of smoothness of E on S as follows

$$\rho(E, u) := \rho(E, S, u) := \frac{1}{2} \sup_{x \in S, \|y\|=1} |E(x + uy) + E(x - uy) - 2E(x)|. \quad (1.2)$$

We assume that E is Fréchet differentiable. Then convexity of E implies that for any x, y

$$E(y) \geq E(x) + \langle E'(x), y - x \rangle \qquad (1.3)$$

or, in other words,

$$E(x) - E(y) \leq \langle E'(x), x - y \rangle = \langle -E'(x), y - x \rangle. \qquad (1.4)$$

We will often use the following simple lemma.

3

Lemma 1.1. *Let E be Fréchet differentiable convex function. Then the following inequality holds for $x \in S$*

$$0 \le E(x + uy) - E(x) - u\langle E'(x), y\rangle \le 2\rho(E, S, u\|y\|). \qquad (1.5)$$

Proof. The left inequality follows directly from (1.3). Next, from the definition of modulus of smoothness it follows that

$$E(x + uy) + E(x - uy) \le 2(E(x) + \rho(E, S, u\|y\|)). \qquad (1.6)$$

Inequality (1.3) gives

$$E(x - uy) \ge E(x) + \langle E'(x), -uy\rangle = E(x) - u\langle E'(x), y\rangle. \qquad (1.7)$$

Combining (1.6) and (1.7), we obtain

$$E(x + uy) \le E(x) + u\langle E'(x), y\rangle + 2\rho(E, S, u\|y\|).$$

This proves the second inequality. $\qquad\qquad\square$

From the definition of a dictionary it follows that any element $f \in X$ can be approximated arbitrarily well by finite linear combinations of the dictionary elements. The primary goal of greedy expansion theory is to study representations of an element $f \in X$ by a series

$$f \sim \sum_{j=1}^{\infty} c_j(f)g_j(f), \quad g_j(f) \in \mathcal{D}, \quad c_j(f) > 0, \quad j = 1, 2, \dots. \qquad (1.8)$$

In building the representation (1.8) we should construct two sequences: $\{g_j(f)\}_{j=1}^{\infty}$ and $\{c_j(f)\}_{j=1}^{\infty}$. In greedy expansion theory the construction of $\{g_j(f)\}_{j=1}^{\infty}$ is based on ideas used in greedy-type nonlinear approximation (greedy-type algorithms). This justifies the use of the term *greedy expansion* for (1.8). The construction of $\{g_j(f)\}_{j=1}^{\infty}$ is, clearly, the most important and difficult part in building the representation (1.8). On the basis of the contemporary theory of nonlinear approximation with respect to redundant dictionaries, we may conclude that the method of using a norming functional in greedy steps of an algorithm is the most productive in approximation in Banach spaces.

Denote

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

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 have $r_{\mathcal{D}}(f) = \|F_f\|_{\mathcal{D}}$.

We begin with a description of a general scheme that provides an expansion for a given element f. Later, specifying this general scheme, we will obtain different methods of expansion.

Dual-Based Expansion (DBE). Let $t \in (0, 1]$ and $f \neq 0$. Denote $f_0 := f$. Assume $\{f_j\}_{j=0}^{m-1} \subset X$, $\{\varphi_j\}_{j=1}^{m-1} \subset \mathcal{D}$ and a set of coefficients $\{c_j\}_{j=1}^{m-1}$ of expansion have already been constructed. If $f_{m-1} = 0$ then we stop (set $c_j = 0$, $j = m, m+1, \ldots$ in the expansion) and get $f = \sum_{j=1}^{m-1} c_j \varphi_j$. If $f_{m-1} \neq 0$ then we conduct the following two steps.

(1) Choose $\varphi_m \in \mathcal{D}$ such that

$$\sup_{F_{f_{m-1}}} F_{f_{m-1}}(\varphi_m) \geq t r_{\mathcal{D}}(f_{m-1}).$$

(2) Define

$$f_m := f_{m-1} - c_m \varphi_m,$$

where $c_m > 0$ is a coefficient either prescribed in advance or chosen from a concrete approximation procedure.

We call the series

$$f \sim \sum_{j=1}^{\infty} c_j \varphi_j \tag{1.9}$$

the Dual-Based Expansion of f with coefficients $c_j(f) := c_j$, $j = 1, 2, \ldots$ with respect to $\mathcal{D}$.

Denote

$$S_m(f, \mathcal{D}) := \sum_{j=1}^{m} c_j \varphi_j.$$

Then it is clear that

$$f_m = f - S_m(f, \mathcal{D}).$$

The reader can find some convergence results for the DBE in Sections 6.7.2 and 6.7.3 of [12].

Let $\mathcal{C} := \{c_m\}_{m=1}^{\infty}$ be a fixed sequence of positive numbers. We restrict ourselves to positive numbers because of the symmetry of the dictionary $\mathcal{D}$.

5

X-**Greedy Algorithm with coefficients** $\mathcal{C}$ (**XGA**($\mathcal{C}$)**).** We define $f_0 := f$, $G_0 := 0$. Then, for each $m \geq 1$ we have the following inductive definition.

(1) $\varphi_m \in \mathcal{D}$ is such that (assuming existence)

$$\|f_{m-1} - c_m\varphi_m\|_X = \inf_{g \in \mathcal{D}} \|f_{m-1} - c_m g\|_X.$$

(2) Let

$$f_m := f_{m-1} - c_m\varphi_m, \qquad G_m := G_{m-1} + c_m\varphi_m.$$

Dual Greedy Algorithm with weakness τ **and coefficients** $\mathcal{C}$ (**DGA**($\tau,\mathcal{C}$)**).** Let $\tau := \{t_m\}_{m=1}^{\infty}$, $t_m \in [0,1]$, be a weakness sequence. We define $f_0 := f$, $G_0 := 0$. Then, for each $m \geq 1$ we have the following inductive definition.

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

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

(2) Let

$$f_m := f_{m-1} - c_m\varphi_m, \qquad G_m := G_{m-1} + c_m\varphi_m.$$

In the case $\tau = \{t\}$, $t \in (0,1]$, we write t instead of τ in the notation.

It is easy to see that for any Banach space X its modulus of smoothness $\rho(u)$ is an even convex function satisfying the inequalities

$$\max(0, u - 1) \leq \rho(u) \leq u, \quad u \in (0, \infty).$$

In Section 6.7.3 of [12] we considered a variant of the Dual-Based Expansion with coefficients chosen by a certain simple rule. The rule depends on two numerical parameters, $t \in (0,1]$ (the weakness parameter from the definition of the DBE) and $b \in (0,1)$ (the tuning parameter of the approximation method). The rule also depends on a majorant μ of the modulus of smoothness of the Banach space X.

Let X be a uniformly smooth Banach space with modulus of smoothness $\rho(u)$, and let $\mu(u)$ be a continuous majorant of $\rho(u)$: $\rho(u) \leq \mu(u)$, $u \in [0, \infty)$ such that $\mu(u)/u$ goes to 0 monotonically. It is clear that $\mu(2) \geq 1$.

Dual Greedy Algorithm with parameters (t, b, μ) (**DGA**(t, b, μ)**).** Let X and $\mu(u)$ be as above. For parameters $t \in (0,1]$, $b \in (0,1]$ we define

sequences $\{f_m\}_{m=0}^{\infty}$, $\{\varphi_m\}_{m=1}^{\infty}$, $\{c_m\}_{m=1}^{\infty}$ inductively. Let $f_0 := f$. 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

$$F_{f_{m-1}}(\varphi_m) \geq tr_{\mathcal{D}}(f_{m-1}).$$

(2) Choose $c_m > 0$ from the equation

$$\|f_{m-1}\|\mu(c_m/\|f_{m-1}\|) = \frac{tb}{2}c_m r_{\mathcal{D}}(f_{m-1}).$$

(3) Define

$$f_m := f_{m-1} - c_m\varphi_m.$$

We note that (2) is equivalent to solving the equation

$$\frac{\mu(c_m/\|f_{m-1}\|)}{c_m/\|f_{m-1}\|} = \frac{tb}{2}r_{\mathcal{D}}(f_{m-1}).$$

It follows from the definitions of t, b and $r_{\mathcal{D}}(f_{m-1})$ that the right hand side of the above equation is $\leq 1/2$. Therefore, there always exists a unique solution to this equation and it satisfies the inequality

$$c_m/\|f_{m-1}\| \leq 2.$$

For illustration we present two theorems on convergence and rate of convergence of the $\mathrm{DGA}(\tau, b, \mu)$ (see Section 6.7.3 of [12]).

Theorem 1.1. *Let X be a uniformly smooth Banach space with the modulus of smoothness $\rho(u)$ and let $\mu(u)$ be a continuous majorant of $\rho(u)$ with the property $\mu(u)/u \downarrow 0$ as $u \to +0$. Then, for any $t \in (0,1]$ and $b \in (0,1)$ the $\mathrm{DGA}(t, b, \mu)$ converges for each dictionary $\mathcal{D}$ and all $f \in X$.*

Theorem 1.2. *Assume X has a modulus of smoothness $\rho(u) \leq \gamma u^q$, $q \in (1,2]$ and $b \in (0,1)$. 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 $\mathrm{DGA}(t, b, \mu)$ is given by*

$$\|f_m\| \leq C(t, b, \gamma, q)m^{-\frac{t(1-b)}{p(1+t(1-b))}}, \quad p := \frac{q}{q-1}.$$

We now formulate the corresponding generalizations of the above algorithms to the case of smooth convex function E. Denote

$$E_{\mathcal{D}}(x) := \sup_{g \in \mathcal{D}} \langle -E'(x), g \rangle.$$

Gradient Based Expansion. Let $t \in (0, 1]$. Denote $G_0 := 0$. Assume $\{G_j\}_{j=0}^{m-1} \subset X$, $\{\varphi_j\}_{j=1}^{m-1} \subset \mathcal{D}$ and a set of coefficients $\{c_j\}_{j=1}^{m-1}$ of expansion have already been constructed. If $E'(G_{m-1}) = 0$ then we stop (set $c_j = 0$, $j = m, m+1, \dots$ in the expansion). If $E'(G_{m-1}) \neq 0$ then we conduct the following two steps.

(1) Choose $\varphi_m \in \mathcal{D}$ such that

$$\langle -E'(G_{m-1}), \varphi_m \rangle \geq t E_{\mathcal{D}}(G_{m-1}).$$

(2) Define

$$G_m := G_{m-1} + c_m \varphi_m,$$

where $c_m > 0$ is a coefficient either prescribed in advance or chosen from a concrete approximation procedure.

We call the series

$$\sum_{j=1}^{\infty} c_j \varphi_j \tag{1.10}$$

the Gradient Based Expansion with coefficients c_j, $j = 1, 2, \dots$ with respect to $\mathcal{D}$.

Let $\mathcal{C} := \{c_m\}_{m=1}^{\infty}$ be a fixed sequence of positive numbers. We restrict ourselves to positive numbers because of the symmetry of the dictionary $\mathcal{D}$.

E-Greedy Algorithm with coefficients $\mathcal{C}$ (EGA($\mathcal{C}$)). We define $G_0 := 0$. Then, for each $m \geq 1$ we have the following inductive definition.

(1) $\varphi_m \in \mathcal{D}$ is such that (assuming existence)

$$E(G_{m-1} + c_m \varphi_m) = \inf_{g \in \mathcal{D}} E(G_{m-1} + c_m g).$$

(2) Let

$$G_m := G_{m-1} + c_m \varphi_m.$$

Gradient Greedy Algorithm with weakness τ and coefficients $\mathcal{C}$ (GGA($\tau, \mathcal{C}$)). Let $\tau := \{t_m\}_{m=1}^{\infty}$, $t_m \in [0, 1]$, be a weakness sequence. We define $G_0 := 0$. Then, for each $m \geq 1$ we have the following inductive definition.

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

$$\langle -E'(G_{m-1}), \varphi_m \rangle \geq t_m E_{\mathcal{D}}(G_{m-1}).$$

(2) Let

$$G_m := G_{m-1} + c_m \varphi_m.$$

In the case $\tau = \{t\}$, $t \in (0, 1]$, we write t instead of τ in the notation.

Let E be a uniformly smooth convex function with modulus of smoothness $\rho(E, D, u)$, and let $\mu(u)$ be a continuous majorant of $\rho(E, D, u)$: $\rho(E, D, u) \leq \mu(u)$, $u \in [0, \infty)$ such that $\mu(u)/u$ goes to 0 monotonically.

Gradient Greedy Algorithm with parameters (τ, b, μ) (GGA(τ, b, μ)). Let E and $\mu(u)$ be as above. For parameters $\tau = \{t_k\}$, $t_k \in (0, 1]$, $b \in (0, 1]$ we define sequences $\{G_m\}_{m=0}^{\infty}$, $\{\varphi_m\}_{m=1}^{\infty}$, $\{c_m\}_{m=1}^{\infty}$ inductively. Let $G_0 := 0$. If for $m \geq 1$ $E'(G_{m-1}) = 0$ then we stop. If $E'(G_{m-1}) \neq 0$ then we conduct the following three steps.

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

$$\langle -E'(G_{m-1}), \varphi_m \rangle \geq t_m E_{\mathcal{D}}(G_{m-1}). \tag{1.11}$$

(2) Choose $c_m > 0$ from the equation

$$\mu(c_m) = \frac{t_m b}{2} c_m E_{\mathcal{D}}(G_{m-1}) \tag{1.12}$$

provided it has a solution $c_m > 0$ and set $c_m = 1$ otherwise.

(3) Define

$$G_m := G_{m-1} + c_m \varphi_m. \tag{1.13}$$

We note that equation (1.12) is equivalent to the equation

$$\frac{\mu(c_m)}{c_m} = \frac{t_m b}{2} E_{\mathcal{D}}(G_{m-1}).$$

Our assumption $E'(G_{m-1}) \neq 0$ implies that $E_{\mathcal{D}}(G_{m-1}) > 0$. Therefore, the above equation either has a solution $c_m > 0$ or $\mu(u)/u \leq \frac{t_m b}{2} E_{\mathcal{D}}(G_{m-1})$ for all u.

The greedy step (1) in the above algorithm is a standard greedy step which is based on $E'(G_{m-1})$. The choice of the coefficient c_m from equation (1.12) requires knowledge of both $E_{\mathcal{D}}(G_{m-1})$ and $\mu(u)$. The quantity $E_{\mathcal{D}}(G_{m-1})$ can be computed (in case X is finite dimensional and $\mathcal{D}$ is finite). The function

9

$\mu(u)$ comes from our assumption on E and may be the one which does not describe smoothness of E in the best way. Here is a modification of the GGA(τ, b, μ) which does not use μ.

Gradient E-Greedy Algorithm with parameters (τ) (GEGA(τ)). Let E be uniformly smooth convex function. For parameters $\tau = \{t_k\}$, $t_k \in (0, 1]$ we define sequences $\{G_m\}_{m=0}^{\infty}$, $\{\varphi_m\}_{m=1}^{\infty}$, $\{c_m\}_{m=1}^{\infty}$ inductively. Let $G_0 := 0$. If for $m \geq 1$ $E'(G_{m-1}) = 0$ then we stop. If $E'(G_{m-1}) \neq 0$ then we conduct the following three steps.

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

$$\langle -E'(G_{m-1}), \varphi_m \rangle \geq t_m E_{\mathcal{D}}(G_{m-1}). \tag{1.14}$$

(2) Choose c_m from the equation

$$E(G_{m-1} + c_m \varphi_m) = \min_c E(G_{m-1} + c\varphi_m). \tag{1.15}$$

(3) Define

$$G_m := G_{m-1} + c_m \varphi_m. \tag{1.16}$$

Our main interest in this paper is in analysis of greedy constructions of sparse approximants for convex optimization problems with respect to an arbitrary dictionary $\mathcal{D}$. We now give a comment that relates the above algorithms to classical gradient-type algorithms and thus justifies the use of the term *gradient* in the names of these algorithms. We specify our dictionary $\mathcal{D}$ to be the unit sphere $\mathcal{S} := \{g \in X : \|g\| = 1\}$ of the space X. Then

$$E_{\mathcal{D}}(x) = \|E'(x)\|_{X^*}.$$

Therefore, the greedy step from the Gradient Based Expansion, the GGA$(\tau, \mathcal{C})$, and the GGA(τ, b, μ) takes the form: choose $\varphi_m \in \mathcal{D}$ such that

$$\langle -E'(G_{m-1}), \varphi_m \rangle \geq t_m \|E'(G_{m-1})\|_{X^*}.$$

In particular, when $X = \mathbb{R}^n$ equipped with Euclidean norm and $t_m = 1$ we obtain

$$\varphi_m = -E'(G_{m-1})/\|E'(G_{m-1})\|_2$$

is a unit vector in the direction opposite to the gradient $E'(G_{m-1})$ direction. In this case the GGA$(\{1\}, b, \mu)$ with $\mu(u) = \gamma u^2$ chooses the step size c_m from the equation

$$\gamma c_m^2 = \frac{b}{2} c_m \|E'(G_{m-1})\|_2 \quad \Rightarrow \quad c_m = \frac{b}{2\gamma} \|E'(G_{m-1})\|_2.$$

Thus
$$G_m = G_{m-1} + c_m \varphi_m = G_{m-1} - \frac{b}{2\gamma} E'(G_{m-1}),$$

which describes the classical Gradient Method.

2 Convergence of the Gradient Based Expansion

In this section we assume that the sets
$$D_C := \{x : E(x) \le E(0) + C\}$$

are bounded for all finite C and that for any bounded set Ω we have
$$\sup_{x \in \Omega} \|E'(x)\|_{X^*} < \infty. \tag{2.1}$$

We begin with the following lemma

Lemma 2.1. *Let E be Fréchet differentiable convex function satisfying the above assumptions. Assume that the approximants $\{G_j\}_{j=0}^\infty$ and coefficients $\{c_j\}_{j=1}^\infty$ from the Gradient Based Expansion satisfy the following two conditions*
$$\sum_{j=1}^\infty c_j E_{\mathcal{D}}(G_j) < \infty, \tag{2.2}$$

$$\sum_{j=1}^\infty c_j = \infty. \tag{2.3}$$

Then
$$\liminf_{m \to \infty} E(G_m) = \inf_{x \in D} E(x). \tag{2.4}$$

Proof. By (1.4)
$$E(G_m) - E(G_{m-1}) \le \langle E'(G_m), G_m - G_{m-1} \rangle = c_m \langle E'(G_m), \varphi_m \rangle.$$

This implies
$$E(G_m) \le E(G_{m-1}) + c_m E_{\mathcal{D}}(G_m).$$

11

Using our assumption (2.2) we obtain

$$E(G_m) \le E(0) + \sum_{j=1}^{m} c_j E_\mathcal{D}(G_j) \le E(0) + C_1.$$

By (2.1) we get from here for all m

$$\|E'(G_m)\|_{X^*} \le C_2.$$

Denote $s_n := \sum_{j=1}^{n} c_j$. Then (2.3) implies (see [1], p. 904) that

$$\sum_{n=1}^{\infty} \frac{c_n}{s_n} = \infty. \tag{2.5}$$

Using (2.2)) we get

$$\sum_{n=1}^{\infty} s_n E_\mathcal{D}(G_n) \frac{c_n}{s_n} = \sum_{n=1}^{\infty} c_n E_\mathcal{D}(G_n) < \infty.$$

Thus, by (2.5)

$$\liminf_{n\to\infty} s_n E_\mathcal{D}(G_n) = 0.$$

Let

$$\lim_{k\to\infty} s_{n_k} E_\mathcal{D}(G_{n_k}) = 0. \tag{2.6}$$

Consider $\{E'(G_{n_k})\}$. A closed bounded set in the dual X^* is weakly* compact (see [5], p. 45). Let $\{F_i\}_{i=1}^{\infty}$, $F_i := -E'(G_{n_{k_i}})$ be a w^*-convergent subsequence. Denote

$$F := w^*\text{-}\lim_{i\to\infty} F_i.$$

We complete the proof of Lemma 2.1 by contradiction. We assume that (2.4) does not hold, that is, there exist $\alpha > 0$ and $N \in \mathbb{N}$ such that

$$E(G_m) - \inf_{x\in D} E(x) \ge 2\alpha, \quad m \ge N, \tag{2.7}$$

and then derive a contradiction.

We begin by deducing from (2.7) that $F \ne 0$. Indeed, by (2.7) there exists $f \in D$ such that

$$E(G_m) - E(f) \ge \alpha, \quad m \ge N. \tag{2.8}$$

By (1.4) we obtain

$$\langle -E'(G_m), f - G_m \rangle \geq E(G_m) - E(f) \geq \alpha. \tag{2.9}$$

Next, we have

$$\langle F, f \rangle = \lim_{i \to \infty} \langle F_i, f \rangle \tag{2.10}$$

and

$$|\langle F_i, G_{n_{k_i}} \rangle| = |\langle F_i, \sum_{j=1}^{n_{k_i}} c_j \varphi_j \rangle|$$

$$= |\sum_{j=1}^{n_{k_i}} c_j \langle F_i, \varphi_j \rangle| \leq s_{n_{k_i}} E_{\mathcal{D}}(G_{n_{k_i}}) \to 0 \tag{2.11}$$

for $i \to \infty$. Relations (2.10), (2.11) and (2.9) imply that $\langle F, f \rangle \geq \alpha$, and hence $F \neq 0$. This implies that there exists $g \in \mathcal{D}$ for which $\langle F, g \rangle > 0$. However,

$$\langle F, g \rangle = \lim_{i \to \infty} \langle F_i, g \rangle \leq \lim_{i \to \infty} E_{\mathcal{D}}(G_{n_{k_i}}) = 0.$$

We have a contradiction, which completes the proof of Lemma 2.1. $\qquad\square$

3   Convergence of GGA($\tau, \mathcal{C}$) and EGA($\mathcal{C}$)

We begin with a simple lemma.

Lemma 3.1. *Let f, $A > 0$, be such that*

$$f/A \in A_1(\mathcal{D}).$$

Then for

$$G_k := \sum_{j=1}^{k} c_j \varphi_j, \quad \varphi_j \in \mathcal{D}, \quad j = 1, \ldots, k,$$

we have

$$E_{\mathcal{D}}(G_k) \geq (E(G_k) - E(f))/(A + A_k), \quad A_k := \sum_{j=1}^{k} |c_j|.$$

13

Proof. We have by (1.4)

$$\langle -E'(G_k), f - G_k \rangle \geq E(G_k) - E(f). \tag{3.1}$$

Next,

$$|\langle -E'(G_k), f \rangle| \leq A E_{\mathcal{D}}(G_k), \tag{3.2}$$

$$|\langle -E'(G_k), G_k \rangle| \leq E_{\mathcal{D}}(G_k) \sum_{j=1}^{k} |c_j|. \tag{3.3}$$

Inequalities (3.1)–(3.3) imply the statement of Lemma 3.1.

$\square$

We now proceed to a convergence result for general uniformly smooth convex function E.

Theorem 3.1. *Let E be a uniformly smooth convex function satisfying*

$$E(x + uy) - E(x) - u\langle E'(x), y \rangle \leq 2\mu(u), \tag{3.4}$$

for $x \in D_2$, $\|y\| = 1$, $|u| \leq 1$ with $\mu(u) = o(u)$ as $u \to 0$. Assume that the coefficients sequence $\mathcal{C} := \{c_j\}$, $c_j \in [0, 1]$ satisfies the conditions

$$\sum_{k=1}^{\infty} c_k = \infty, \tag{3.5}$$

$$\sum_{k=1}^{\infty} \mu(c_k) \leq 1. \tag{3.6}$$

Then for the GGA(t,$\mathcal{C}$) and for the EGA($\mathcal{C}$) we have for each dictionary $\mathcal{D}$

$$\lim_{m \to \infty} E(G_m) = \inf_{x \in D} E(x).$$

Proof. We give here a proof that works for both algorithms from Theorem 3.1. Let G_{m-1} be an approximate solution after $m - 1$ iterations of either the GGA(t,$\mathcal{C}$) or the EGA($\mathcal{C}$). Let φ_m be such that

$$\langle -E'(G_{m-1}), \varphi_m \rangle \geq t E_{\mathcal{D}}(G_{m-1}). \tag{3.7}$$

Then

$$\inf_{g \in \mathcal{D}} E(G_{m-1} + c_m g) \leq E(G_{m-1} + c_m \varphi_m).$$

Thus, in both cases $(\text{GGA}(t,\mathcal{C})$ and $\text{EGA}(\mathcal{C}))$ it is sufficient to estimate $E(G_{m-1} + c_m\varphi_m)$ with φ_m satisfying (3.7). By (3.4) under assumption that $G_{m-1} \in D_2$ we get

$$E(G_{m-1} + c_m\varphi_m) \leq E(G_{m-1}) + c_m\langle E'(G_{m-1}), \varphi_m\rangle + 2\mu(c_m).$$

Using definition of φ_m we continue

$$\leq E(G_{m-1}) - c_m t E_{\mathcal{D}}(G_{m-1}) + 2\mu(c_m). \tag{3.8}$$

We now prove by induction that $G_m \in D_2$ for all m. Indeed, clearly $G_0 \in D_2$. Suppose that $G_k \in D_2$, $k = 0, 1, \ldots, m-1$, then (3.8) holds for all $k = 1, \ldots, m$ instead of m and, therefore,

$$E(G_m) \leq E(0) + 2\sum_{k=1}^{m}\mu(c_k) \leq E(0) + 2$$

which implies that $G_m \in D_2$.

Let f^ϵ, $\epsilon \geq 0$, $A(\epsilon) > 0$, be such that

$$E(f^\epsilon) - b \leq \epsilon, \quad f^\epsilon/A(\epsilon) \in A_1(\mathcal{D}), \quad b := \inf_{x \in D} E(x).$$

Applying Lemma 3.1 we obtain from (3.8) (with $A_k := \sum_{j=1}^{k} c_j$)

$$E(G_{m-1} + c_m\varphi_m) \leq E(G_{m-1}) - \frac{tc_m(E(G_{m-1}) - b - \epsilon)}{A(\epsilon) + A_{m-1}} + 2\mu(c_m). \tag{3.9}$$

Denote

$$a_n := E(G_n) - b - \epsilon.$$

By (3.9) we obtain

$$a_m \leq a_{m-1}(1 - \theta_m) + 2\mu(c_m). \tag{3.10}$$

with

$$\theta_m := \frac{tc_m}{A(\epsilon) + A_{m-1}}.$$

We note that our assumption (3.5) implies that

$$\sum_{m=1}^{\infty}\theta_m = \infty. \tag{3.11}$$

15

Without loss of generality we can assume that $A(\epsilon) \geq 1$. Then $\theta_m \leq 1$ and we get from (3.10)

$$a_m \leq a_0 \prod_{j=1}^m (1 - \theta_j) + 2\mu(c_1) \prod_{j=2}^m (1 - \theta_j) + \cdots + 2\mu(c_{m-1})(1 - \theta_m) + 2\mu(c_m).$$

(3.12)

The properties (3.11) and $\sum_m \mu(c_m) < \infty$ imply that

$$\limsup_{m \to \infty} a_m \leq 0.$$

This completes the proof of Theorem 3.1. $\qquad\qquad\qquad\qquad\qquad\qquad\square$

4   Rate of convergence of $\mathbf{GGA}(\tau, \mathcal{C})$ and $\mathbf{EGA}(\mathcal{C})$

In this section we consider the $\mathrm{GGA}(t, \mathcal{C})$ and the $\mathrm{EGA}(\mathcal{C})$ with a specific sequence $\mathcal{C}$. For a special $\mathcal{C}$ we prove the rate of convergence results for the uniformly smooth convex functions with modulus of smoothness $\rho(E, u) \leq \gamma u^q$, $q \in (1, 2]$.

Theorem 4.1. *Let E be a uniformly smooth convex function with modulus of smoothness $\rho(E, u) \leq \gamma u^q$, $q \in (1, 2]$ on D_2. We set $s := \frac{t+1}{t+q}$ and $\mathcal{C}_s := \{ck^{-s}\}_{k=1}^\infty$ with c chosen in such a way that $\gamma c^q \sum_{k=1}^\infty k^{-sq} \leq 1$. Then the $GGA(t, \mathcal{C}_s)$ and $EGA(\mathcal{C}_s)$ (for this algorithm $t = 1$) converge with the following rate: for any $r \in (0, t(1-s))$*

$$E(G_m) - \inf_{\frac{x}{M} \in A_1(\mathcal{D})} E(x) \leq C(r, t, q, \gamma, M) m^{-r}.$$

Proof. In the same way as in the proof of Theorem 3.1 we prove that $G_m \in D_2$ for all m. Then we use inequality (3.9) proved in Section 3. Let f^ϵ, $\epsilon \geq 0$, $M > 0$, be such that

$$E(f^\epsilon) - b \leq \epsilon, \quad f^\epsilon / M \in A_1(\mathcal{D}), \quad b := \inf_{\frac{x}{M} \in A_1(\mathcal{D})} E(x).$$

Using the assumption $f^\epsilon / M \in A_1(\mathcal{D})$, we write (3.9) with $A(\epsilon) = M$

$$E(G_{m-1} + c_m \varphi_m) \leq E(G_{m-1}) - \frac{t c_m (E(G_{m-1}) - b - \epsilon)}{M + A_{m-1}} + 2\gamma c_m^q. \quad (4.1)$$

16

We have

$$A_{m-1} = c \sum_{k=1}^{m-1} k^{-s} \leq c(1 + \int_1^m x^{-s} dx) = c(1 + (1-s)^{-1}(m^{1-s} - 1)))$$

and

$$M + A_{m-1} \leq M + c(1-s)^{-1} m^{1-s}.$$

Therefore, for $m \geq N$ we have with $v := (r + t(1-s))/2$

$$\frac{t c_m}{M + A_{m-1}} \geq \frac{v + t(1-s)}{2m}. \tag{4.2}$$

We need the following technical lemma. This lemma is a more general version of Lemma 2.1 from [8] (see also Remark 5.1 in [10] and Lemma 2.37 on p. 106 of [12]).

Lemma 4.1. *Let four positive numbers $\alpha < \beta \leq 1$, A, $U \in \mathbb{N}$ be given and let a sequence $\{a_n\}_{n=1}^\infty$ have the following properties: $a_1 < A$ and we have for all $n \geq 2$*

$$a_n \leq a_{n-1} + A(n-1)^{-\alpha}; \tag{4.3}$$

if for some $\nu \geq U$ we have

$$a_\nu \geq A\nu^{-\alpha}$$

then

$$a_{\nu+1} \leq a_\nu(1 - \beta/\nu). \tag{4.4}$$

Then there exists a constant $C = C(\alpha, \beta, A, U)$ such that for all $n = 1, 2, \ldots$ we have

$$a_n \leq Cn^{-\alpha}.$$

We apply this lemma with $a_n := E(G_n) - b - \epsilon$, $\alpha := r$, $\beta := v := (r + t(1-s))/2$, $U = N$ and A specified later. Let us check the conditions (4.3) and (4.4) of Lemma 4.1. By the inequality

$$E(G_m) \leq E(G_{m-1}) + 2\rho(E, c_m) \leq E(G_{m-1}) + 2\gamma c^q m^{-sq}$$

the condition (4.3) holds for $A \geq 2\gamma c^q$. Assume that $a_m \geq Am^{-r}$. Then using $sq \geq 1 + r$ we get

$$c_m^q = c^q m^{-sq} \leq c^q m^{-1-r}. \tag{4.5}$$

17

Setting A to be big enough to satisfy

$$2\gamma c_m^q \le \frac{A(t(1-s)-\beta)}{2m^{1+r}}$$

we obtain from (4.1), (4.2), and (4.5)

$$a_{m+1} \le a_m(1-\beta/m)$$

provided $a_m \ge Am^{-r}$. Thus (4.4) holds. Applying Lemma 4.1 we get

$$a_m \le C(r,t,q,\gamma,M)m^{-r}.$$

$\square$

We note that in the special case when $\mathcal{D}$ is the unit sphere $\mathcal{S}$ of X the rate of convergence in Theorem 4.1 can be improved.

Theorem 4.2. *Let E be a uniformly smooth convex function with modulus of smoothness $\rho(E,u) \le \gamma u^q$, $q \in (1,2]$ on D_2 which we assume to be bounded. For a $\delta \in (0,1)$ we set $s := 1-\delta$ and $\mathcal{C}_s := \{ck^{-s}\}_{k=1}^{\infty}$ with $c := c(\delta)$ chosen in such a way that $\gamma c^q \sum_{k=1}^{\infty} k^{-sq} \le 1$. Suppose $\mathcal{D} = \mathcal{S}$. Then the $GGA(t,\mathcal{C}_s)$ and $EGA(\mathcal{C}_s)$ (for this algorithm $t = 1$) converge with the following rate:*

$$E(G_m) - \inf_{x \in D} E(x) \le C(E,\delta,q,\gamma,t)m^{-s(q-1)}.$$

Proof. As we already mentioned in the Introduction in the case $\mathcal{D} = \mathcal{S}$ we have

$$E_{\mathcal{D}}(x) = \|E'(x)\|_{X^*}.$$

By (3.4) under assumption that $G_{m-1} \in D_2$ we get

$$E(G_{m-1} + c_m\varphi_m) \le E(G_{m-1}) + c_m\langle E'(G_{m-1}), \varphi_m \rangle + 2\gamma(c_m)^q.$$

Using definition of φ_m we continue

$$\le E(G_{m-1}) - c_m t\|E'(G_{m-1})\|_{X^*} + 2\gamma(c_m)^q. \tag{4.6}$$

As in the proof of Theorem 3.1 we derive from here that $G_m \in D_2$ for all m. Using notation $a_m := E(G_m) - \inf_{x \in D} E(x)$ we obtain

$$a_{m-1} = \sup_{f \in D}(E(G_{m-1}) - E(f)) \le \|E'(G_{m-1})\|_{X^*} \sup_{f \in D} \|G_{m-1} - f\|. \tag{4.7}$$

Inequality (4.7) and our assumption that D_2 is bounded imply

$$\|E'(G_{m-1})\|_{X^*} \geq a_{m-1}/C_1.$$

Substituting this bound into (4.6) we get

$$a_m \leq a_{m-1}\left(1 - tc_mC_1^{-1}\right) + 2\gamma(c_m)^q. \tag{4.8}$$

As in the proof of Theorem 4.1 we use Lemma 4.1. It is clear that for a_{m-1} satisfying

$$a_{m-1} \geq Am^{-s(q-1)}$$

with large enough A we have

$$a_{m-1}tc_mC_1^{-1} \geq 4\gamma(c_m)^q.$$

Therefore, (4.8) gives in this case

$$a_m \leq a_{m-1}\left(1 - \frac{tc_m}{2C_1}\right). \tag{4.9}$$

It follows from the definition of c_m that

$$\frac{tc_m}{2C_1} \geq \frac{s(q-1)+1}{m-1} \quad \text{for} \quad m \geq U.$$

Thus by Lemma 4.1 we obtain

$$a_m \leq C(E, \delta, q, \gamma, t)m^{-s(q-1)}$$

which proves Theorem 4.2. $\qquad\square$

5 Convergence and rate of convergence of the $\mathbf{GGA}(\tau, b, \mu)$

We begin with a convergence result.

Theorem 5.1. *Let E be a uniformly smooth convex function with the modulus of smoothness $\rho(E, D, u)$ and let $\mu(u)$ be a continuous majorant of $\rho(E, D, u)$ with the property $\mu(u)/u \downarrow 0$ as $u \to +0$. Assume that for $x \in D$*

$$\|E'(x)\|_{X^*} \leq C_D.$$

Then, for any $t \in (0, 1]$ and $b \in (0, 1)$ we have for the $GGA(t, b, \mu)$

$$\lim_{m \to \infty} E(G_m) = \inf_{x \in D} E(x). \tag{5.1}$$

19

Proof. In this case $\tau = \{t\}$, $t \in (0,1]$. We have by (3.4) under assumption that $G_{m-1} \in D$

$$E(G_{m-1} + c_m \varphi_m) \leq E(G_{m-1}) + c_m \langle E'(G_{m-1}), \varphi_m \rangle + 2\mu(c_m).$$

Using definition of φ_m we continue

$$\leq E(G_{m-1}) - c_m t E_\mathcal{D}(G_{m-1}) + 2\mu(c_m). \tag{5.2}$$

Using the choice of c_m we find

$$E(G_m) \leq E(G_{m-1}) - t(1-b)c_m E_\mathcal{D}(G_{m-1}). \tag{5.3}$$

In particular, (5.3) implies that $\{E(G_m)\}$ is a monotone decreasing sequence and therefore our assumption that $G_{m-1} \in D$ implies that $G_m \in D$. Clearly, $G_0 \in D$. Thus we obtain that $G_m \in D$ for all m. Also, (5.3) implies that

$$t(1-b)c_m E_\mathcal{D}(G_{m-1}) \leq E(G_{m-1}) - E(G_m).$$

Thus

$$\sum_{m=1}^{\infty} c_m E_\mathcal{D}(G_{m-1}) < \infty. \tag{5.4}$$

We have the following two cases:

$$(I) \quad \sum_{m=1}^{\infty} c_m = \infty, \qquad (II) \quad \sum_{m=1}^{\infty} c_m < \infty.$$

First, we consider case (I). Our argument here is as in Lemma 2.1. Denote $s_n := \sum_{j=1}^{n} c_j$. Then our assumption implies (see [1], p. 904) that

$$\sum_{n=1}^{\infty} \frac{c_n}{s_n} = \infty. \tag{5.5}$$

Using (5.4)) we get

$$\sum_{n=1}^{\infty} s_n E_\mathcal{D}(G_{n-1}) \frac{c_n}{s_n} = \sum_{n=1}^{\infty} c_n E_\mathcal{D}(G_{n-1}) < \infty.$$

Thus, by (5.5)

$$\liminf_{n \to \infty} s_n E_\mathcal{D}(G_{n-1}) = 0.$$

20

Clearly, the above relation implies

$$\liminf_{n \to \infty} s_n E_{\mathcal{D}}(G_n) = 0.$$

The rest of the proof in this case repeats the corresponding part from the proof of Lemma 2.1. As a result we obtain

$$\liminf_{m \to 0} E(G_m) = \inf_{x \in D} E(x).$$

Monotonicity of $\{E(G_m)\}$ implies that we can replace liminf by lim in the above relation.

Second, we consider the case (II). Our assumption implies that $c_m \to 0$ as $m \to \infty$. From the definition (1.12) of c_m we obtain

$$E_{\mathcal{D}}(G_{m-1}) = \frac{2}{tb} \mu(c_m)/c_m \to 0, \quad m \to \infty. \tag{5.6}$$

We show that relation (5.6) implies the following two properties (5.7) and (5.8)

$$\lim_{m \to 0} \langle E'(G_m), G_m \rangle = 0, \tag{5.7}$$

$$\lim_{m \to 0} \langle E'(G_m), f \rangle = 0. \tag{5.8}$$

Indeed, for (5.7) we have

$$|\langle E'(G_m), G_m \rangle| = |\sum_{j=1}^{m} \langle E'(G_m), \varphi_j \rangle c_j| \le E_{\mathcal{D}}(G_m) \sum_{j=1}^{m} c_j \to 0.$$

We now prove (5.8). For arbitrary $\epsilon > 0$ find f^ϵ such that

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

with some $A(\epsilon)$. Then

$$|\langle E'(G_m), f \rangle| = |\langle E'(G_m), f^\epsilon \rangle + \langle E'(G_m), f - f^\epsilon \rangle| \le E_{\mathcal{D}}(G_m)A(\epsilon) + C_D\epsilon.$$

We complete the proof of case (II) by contradiction. We assume that (5.1) does not hold, that is, there exist $\alpha > 0$ and $N \in \mathbb{N}$ such that

$$E(G_m) - \inf_{x \in D} E(x) \ge 2\alpha, \quad m \ge N, \tag{5.9}$$

21

and then derive a contradiction. By (5.9) there exists $f \in D$ such that

$$E(G_m) - E(f) \geq \alpha, \quad m \geq N. \tag{5.10}$$

By (1.4) we obtain

$$\langle -E'(G_m), f - G_m \rangle \geq E(G_m) - E(f) \geq \alpha. \tag{5.11}$$

This contradicts to (5.7) and (5.8).

$\square$

Theorem 5.2. *Let E be a uniformly smooth convex function. Assume that for $x \in D$*

$$\|E'(x)\|_{X^*} \leq C_D.$$

Then, for any $t \in (0, 1]$ we have for the GEGA($\{t\}$)

$$\lim_{m \to \infty} E(G_m) = \inf_{x \in D} E(x). \tag{5.12}$$

Proof. Let E be a uniformly smooth convex function with the modulus of smoothness $\rho(E, D, u)$ and let $\mu(u)$ be a continuous majorant of $\rho(E, D, u)$ with the property $\mu(u)/u \downarrow 0$ as $u \to +0$. As in (5.3) we obtain

$$E(G_m) \leq E(G_{m-1} + c'_m \varphi_m) \leq E(G_{m-1}) - t(1 - b)c'_m E_{\mathcal{D}}(G_{m-1}) \tag{5.13}$$

with c'_m chosen from the equation

$$\mu(c'_m) = \frac{tb}{2} c'_m E_{\mathcal{D}}(G_{m-1}) \tag{5.14}$$

with some fixed $b \in (0, 1)$.

The proof of Theorem 5.1 used only assumptions on E and analogs of relations (5.13) and (5.14). Therefore the same proof gives (5.12). $\square$

We proceed to study the rate of convergence of the GGA(τ, b, μ) for the uniformly smooth convex function with the power-type majorant of modulus of smoothness: $\rho(E, D, u) \leq \mu(u) = \gamma u^q$, $1 < q \leq 2$.

Theorem 5.3. *Let $\tau := \{t_k\}_{k=1}^{\infty}$ be a nonincreasing sequence $1 \geq t_1 \geq t_2 \cdots > 0$ and $b \in (0, 1)$. Assume that uniformly smooth convex function E has a modulus of smoothness $\rho(E, D, u) \leq \gamma u^q$, $q \in (1, 2]$. Denote $\mu(u) = \gamma u^q$. Then the rate of convergence of the GGA(τ, b, μ) is given by*

$$E(G_m) - \inf_{x \in A_1(\mathcal{D})} E(x) \leq C(b, \gamma, q)(1 + \sum_{k=1}^{m} t_k^p)^{-\frac{t_m(1-b)(q-1)}{q+t_m(1-b)}}, \quad p := \frac{q}{q-1}.$$

Proof. As in (5.3), we get

$$E(G_m) \le E(G_{m-1}) - t_m(1-b)c_m E_{\mathcal{D}}(G_{m-1}). \tag{5.15}$$

Thus we need to estimate $c_m E_{\mathcal{D}}(G_{m-1})$ from below. Denote $b_n := 1 + \sum_{j=1}^{n} c_j$. By Lemma 3.1 we obtain

$$E_{\mathcal{D}}(G_{m-1}) \ge (E(G_{m-1}) - w)/b_{m-1}, \quad w := \inf_{x \in A_1(\mathcal{D})} E(x). \tag{5.16}$$

Substituting (5.16) into (5.15) and using notation $a_m := E(G_m) - w$ we get

$$a_m \le a_{m-1}(1 - t_m(1-b)c_m/b_{m-1}). \tag{5.17}$$

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

$$(1+x)^\alpha \le 1 + \alpha x, \quad 0 \le \alpha \le 1, \quad x \ge 0,$$

we obtain

$$b_m^{t_m(1-b)} \le b_{m-1}^{t_m(1-b)}(1 + t_m(1-b)c_m/b_{m-1}). \tag{5.18}$$

Multiplying (5.17) and (5.18), and using that $t_m \le t_{m-1}$, we get

$$a_m b_m^{t_m(1-b)} \le a_{m-1} b_{m-1}^{t_{m-1}(1-b)} \le a_0. \tag{5.19}$$

The function $\mu(u)/u = \gamma u^{q-1}$ is increasing on $[0, \infty)$. Therefore the c_m is greater than or equal to c_m' from (see (5.16))

$$\gamma(c_m')^q = \frac{t_m b}{2} c_m' a_{m-1}/b_{m-1}, \tag{5.20}$$

$$c_m' = \left(\frac{t_m b}{2\gamma}\right)^{\frac{1}{q-1}} \left(\frac{a_{m-1}}{b_{m-1}}\right)^{\frac{1}{q-1}}. \tag{5.21}$$

Using notations

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

23

we obtain

$$a_m \le a_{m-1}\left(1 - \frac{t_m^p}{A}\frac{a_{m-1}^{\frac{1}{q-1}}}{b_{m-1}^p}\right), \tag{5.22}$$

from (5.17) and (5.21). Noting that $b_m \ge b_{m-1}$, we infer from (5.22) that

$$\frac{a_m^{\frac{1}{q-1}}}{b_m^p} \le \frac{a_{m-1}^{\frac{1}{q-1}}}{b_{m-1}^p}\left(1 - \frac{t_m^p}{A}\frac{a_{m-1}^{\frac{1}{q-1}}}{b_{m-1}^p}\right). \tag{5.23}$$

We obtain from (5.23) by an analog of Lemma 2.16 from Chapter 2 of [12] (see [9], Lemma 3.1)

$$\frac{a_m^{\frac{1}{q-1}}}{b_m^p} \le C(E,b,\gamma)\left(1 + \sum_{k=1}^m t_k^p\right)^{-1}. \tag{5.24}$$

Combining (5.19) and (5.24), we get

$$a_m \le C(E,b,\gamma,q)(1 + \sum_{k=1}^m t_k^p)^{-\frac{t_m(1-b)(q-1)}{q+t_m(1-b)}}, \quad p := \frac{q}{q-1}.$$

This completes the proof of Theorem 5.3. $\qquad\square$

We note that in the special case when $\mathcal{D}$ is the unit sphere $\mathcal{S}$ of X the rate of convergence in Theorem 5.3 can be improved.

Theorem 5.4. *Let $\tau := \{t_k\}_{k=1}^\infty$ be a weakness sequence $t_k \in [0,1]$ and $b \in (0,1)$. Assume that uniformly smooth convex function E has a modulus of smoothness $\rho(E,D,u) \le \gamma u^q$, $q \in (1,2]$. Denote $\mu(u) = \gamma u^q$. Suppose that $\mathcal{D} = \mathcal{S}$. Then the rate of convergence of the $GGA(\tau,b,\mu)$ is given by*

$$E(G_m) - \inf_{x \in D} E(x) \le C(E,b,\gamma,q)\left(1 + \sum_{k=1}^m t_k^p\right)^{1-q}, \quad p := \frac{q}{q-1}. \tag{5.25}$$

Proof. As we already mentioned in the Introduction in the case $\mathcal{D} = \mathcal{S}$ we have

$$E_\mathcal{D}(x) = \|E'(x)\|_{X^*}.$$

24

For $\mu(u) = \gamma u^q$ we obtain for the c_m

$$\gamma c_m^q = \frac{t_m b}{2} c_m \|E'(G_{m-1})\|_{X^*} \quad \Rightarrow \quad c_m = \left(\frac{t_m b}{2\gamma} \|E'(G_{m-1})\|_{X^*}\right)^{\frac{1}{q-1}}.$$

Therefore, by (5.15) we get

$$E(G_m) \leq E(G_{m-1}) - t_m(1-b)\left(\frac{t_m b}{2\gamma} \|E'(G_{m-1})\|_{X^*}\right)^{\frac{1}{q-1}} \|E'(G_{m-1})\|_{X^*}. \tag{5.26}$$

Equation (5.26) implies that $G_m \in D$ for all m. Using the notation $a_m := E(G_m) - w$, $w := \inf_{x \in D} E(x)$ we obtain

$$a_{m-1} = \sup_{f \in D}(E(G_{m-1}) - E(f)) \leq \|E'(G_{m-1})\|_{X^*} \sup_{f \in D} \|G_{m-1} - f\|. \tag{5.27}$$

Inequality (5.27) and our assumption that D is bounded imply

$$\|E'(G_{m-1})\|_{X^*} \geq a_{m-1}/C_1.$$

Substituting this bound into (5.26) we get

$$a_m \leq a_{m-1}\left(1 - t_m^p a_{m-1}^{\frac{1}{q-1}} C_2^{-1}\right). \tag{5.28}$$

Inequality (5.28) is similar to (5.22). We derive (5.25) from (5.28) in the same way as (5.24) was derived from (5.22). $\qquad\square$

References

[1] N. K. Bary, Trigonometric series, Nauka, Moscow, 1961 (in Russian); English transl. in Pergamon Press, Oxford, 1964.

[2] J.M. Borwein and A.S. Lewis, Convex Analysis and Nonlinear Optimization. Theory and Examples, Canadian Mathematical Society, Springer, 2006.

[3] V. Chandrasekaran, B. Recht, P.A. Parrilo, and A.S. Willsky, The convex geometry of linear inverse problems, Proceedings of the 48th Annual Allerton Conference on Communication, Control and Computing, 2010, 699–703.

[4] R.A. DeVore, Nonlinear approximation, Acta Numerica, **7** (1998), 51–150.

[5] P. Habala, P. Hájek and V. Zizler, Introduction to Banach spaces [I], Matfyzpress, Univerzity Karlovy, 1996.

[6] Yu. Nesterov, Introductory Lectures on Convex Optimization: A Basic Course, Kluwer Academic Publishers, Boston, 2004.

[7] S. Shalev-Shwartz, N. Srebro, and T. Zhang, Trading accuracy for sparsity in optimization problems with sparsity constrains, SIAM Journal on Optimization, **20(6)** (2010), 2807–2832.

[8] V.N. Temlyakov, Greedy Algorithms and m-term Approximation With Regard to Redundant Dictionaries, J. Approx. Theory **98** (1999), 117–145.

[9] V.N. Temlyakov, Weak Greedy Algorithms, Adv. Comput. Math., **12** (2000), 213–227.

[10] V.N. Temlyakov, Greedy-Type Approximation in Banach Spaces and Applications, Constr. Approx., **21** (2005), 257–292.

[11] V.N. Temlyakov, Greedy approximation, Acta Numerica, **17** (2008), 235–409.

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

[13] V.N. Temlyakov, Greedy approximation in convex optimization, Manuscript, 2012, 1–25.

[14] A. Tewari, P. Ravikumar, and I.S. Dhillon, Greedy Algorithms for Structurally Constrained High Dimensional Problems, prerint, (2012), 1–10.

[15] T. Zhang, Sequential greedy approximation for certain convex optimization problems, IEEE Transactions on Information Theory, **49(3)** (2003), 682–691.