GREEDY STRATEGIES FOR CONVEX MINIMIZATION

A Dissertation

by

HAO THANH NGUYEN

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Guergana Petrova
Committee Members,	Ronald DeVore
	Thomas Schlumprecht
	Pavel Tsvetkov
Head of Department,	Emil Straube

August 2013

Major Subject: Mathematics

ABSTRACT

We have investigated two greedy strategies for finding an approximation to the minimum of a convex function E, defined on a Hilbert space H. We have proved convergence rates for a modification of the orthogonal matching pursuit and its weak version under suitable conditions on the objective function E. These conditions involve the behavior of the moduli of smoothness and the modulus of uniform convexity of E.

DEDICATION

I dedicate this work to my family, for their love and support.

ACKNOWLEDGEMENTS

I would like to thank everyone who has supported and encouraged me during the course of this study and my education at Texas A & M University.

First and foremost, I am grateful to my academic advisor, Professor Guergana Petrova for her constant support, guidance and encouragement. She has professionally inspired me in many ways, particularly to become a better researcher and never give up when facing obstacles. She is the best advisor in the world!

I greatly appreciate my committee, Professors R. DeVore, T. Schlumprecht and P. Tsvetkov, for their critical comments and excellent questions regarding this work. I would like to acknowledge the Department of Mathematics for the outstanding teaching and research environment as well as generous financial support. I thank Ms. Monique Stewart for her help and hard work in uncountable non-academic matters. I also thank the Vietnam Education Foundation (VEF) for sponsoring my PhD program.

I would like to thank Professors Dong-Yen Nguyen (Vietnam Institute of Mathematics) and Thanh-Ha Thi Nguyen (Hanoi National University of Education) for their distant advice. In addition, I thank all my friends from the Math Department and Vietnamese community who have comfortably shared with me many interesting topics and life experiences.

Lastly, I thank my family for their love, patience and support; and in particular, my little son Henry, who has been a great motivation to finish this dissertation.

TABLE OF CONTENTS

Page

ABSTRACT ii

DEDICATION iii

ACKNOWLEDGEMENTS iv

TABLE OF CONTENTS v

1. INTRODUCTION 1

   1.1 Convex Optimization 1
   1.2 Problem Setting 2

2. CONDITIONS ON E 6

   2.1 Convexity and Frechet Differentiability 6
   2.2 Moduli of Smoothness 9
   2.3 Modulus of Uniform Convexity 13

3. GREEDY ALGORITHMS 19

   3.1 Introduction 19
   3.2 **OMP** and **WCGA** for Optimization 20

4. MAIN RESULTS 23

   4.1 Auxiliary Lemmas 23
   4.2 Convergence Results for **OMP** 29
   4.3 Convergence Results for **WCGA** 38

5. SUMMARY 44

REFERENCES 46

1. INTRODUCTION

1.1 Convex Optimization

The main goal in convex optimization is the development and analysis of algorithms for solving the problem

$$\inf_{x\in\Omega} E(x), \tag{1.1}$$

where E is a given convex function, defined on a bounded convex set Ω. E is called the *objective* function and satisfies the convexity condition

$$E(\gamma x + \delta y) \leq \gamma E(x) + \delta E(y), \quad x, y \in \Omega, \quad \gamma, \delta \geq 0, \quad \gamma + \delta = 1.$$

Convex optimization has many application domains such as automatic control systems, estimation and signal processing, communications and networks, electronic circuit design, data analysis and modeling, statistics, finance, combinatorial optimization and others. Some of the new application areas have stimulated renewed interest in the subject. While the classical convex optimization deals with objective functions E defined on subsets in $\mathbb{R}^n$ with moderate values of n, see [4], some of the new applications require that n is quite large or even ∞. The design of algorithms for such cases is quite challenging and typical convergent results involving the dimension n suffer from the curse of dimensionality. Recently, Temlyakov (see [15, 16]) has proposed several greedy strategies for solving (1.1) for some classes of objective functions E defined on Banach spaces, where he overcomes the curse of dimensionality. The approximate solution to the optimization problem is constructed as a linear combination of elements from a given system (dictionary) of elements. In this way, the greedy approximation technique, originally developed in the context of nonlinear approximation theory, has been successfully adjusted for finding a sparse solution to the optimization problem.

1.2 Problem Setting

This dissertation investigates greedy based strategies for solving (1.1) for particular classes of objective functions E defined on a Hilbert space H and for domain $\Omega = \{x \in H : E(x) \le E(0)\}$. The function E satisfies the following 3 conditions:

Condition 0: E is a convex function defined on a Hilbert space H, is Frechet differentiable at each point in Ω and its Frechet derivative is uniformly bounded on Ω by a constant M_0. The set Ω is bounded.

Condition 1: There are constants $\alpha \ge 0$, $1 < q \le 2$ and $M > 0$, such that

$$E(x') - E(x) - \langle E'(x), x' - x\rangle \le \alpha \|x' - x\|^q, \quad \text{for } x, x', \text{ such that } \|x - x'\| \le M, \quad x \in \Omega. \tag{1.2}$$

Condition 2: There are constants $\beta \ge 0$, $2 \le p < \infty$ and $M > 0$, such that

$$E(x') - E(x) - \langle E'(x), x' - x\rangle \ge \beta \|x' - x\|^p, \quad \text{for } x, x', \text{ such that } \|x - x'\| \le M, \quad x \in \Omega. \tag{1.3}$$

In both **Condition 1** and **Condition 2**, $E'(x)$ is the Frechet derivative of E at x.

Temlyakov (see [15], [16]) has studied various greedy strategies for solving problem (1.1). The objective functions E he considers are defined on Banach spaces X with norm $\|\cdot\|$ and the tool he uses for greedy approximation is symmetric dictionaries $\mathcal{D}$. Recall that $\mathcal{D}$ is called a symmetric dictionary if each element $\varphi \in \mathcal{D}$ has norm $\|\varphi\| \le 1$, if $\varphi \in \mathcal{D}$, then $-\varphi \in \mathcal{D}$, and the closure of

$$\text{span } \mathcal{D} := \{\Sigma_{i\in I} c_i \varphi_i, I \text{ finite}, c_i \in \mathbb{R}, \varphi_i \in \mathcal{D}\}$$

is X. Among other results, he has shown that the Orthogonal Matching Pursuit (**OMP**), when applied to a function E satisfying certain suitable assumptions on its modulus of smoothness, described by a parameter q, gives an error at the m-th step

$$E(x_m) - E(\bar{x}) \leq Cm^{1-q}, \tag{1.4}$$

with a constant C, where $\bar{x}$ is the minimizer of E and x_m is the output of the **OMP**.
Our main results are Theorem 4.2.3 and Theorem 4.3.3 from Chapter IV, where we prove an improved convergence rate for both **OMP** and the Weak Chebyshev Greedy Algorithm (**WCGA**) when they are used to find the minimum of a function E that satisfies **Conditions 0, 1** and **2**. For example, we show that if the objective function E satisfies **Condition 0** and **Condition 1**, is strongly convex on H (therefore satisfies **Condition 2** with $p = 2$) and its minimizer $\bar{x}$ is sparse with respect to an orthonormal basis, the error at the m-th step of the **OMP** satisfies the inequality

$$E(x_m) - E(\bar{x}) \leq C_0 m^{1-\frac{q}{2-q}},$$

and

$$\|x_m - \bar{x}\| \leq C_1 m^{\frac{1}{2}-\frac{q}{2(2-q)}},$$

where $C_0 = C_0(q, \bar{x}), C_1 = C_1(q, \bar{x})$. In contrast, the results from Temlyakov could provide only the rate $1 - q$, see (1.4).
In summary, we have shown that imposing more conditions on the convexity of E (like **Condition 2**) results in provable improved convergence rate for both **OMP** and **WCGA**. Namely, we prove the following two theorems.

Theorem 1.2.1. *Let the objective function E satisfy* **Conditions 0,1,2** *with $M_0 < \alpha q M^{q-1}$. Let problem* (1.1) *has a solution $\bar{x} = \sum_i c_i(\bar{x})\varphi_i \in \Omega$ with support $\bar{S} := \{i :$*

$c_i(\bar{x}) \neq 0\} < \infty$, where $\{\varphi_i\}$ is an orthonormal basis. Then, at Step k, the **OMP** *applied to E and $\{\varphi_i\}$ outputs x_k, where either $x_k = \bar{x}$ or:*

- *When $p \neq q \neq 2$, the error $e_k := E(x_k) - E(\bar{x})$, $k \geq 2$ satisfies the inequality*

$$e_k \leq C_0 \left(\frac{|\bar{S}|^{\frac{q}{2(q-1)}}}{k + C_1 |\bar{S}|^{\frac{q}{2(q-1)}} - 1} \right)^{\frac{p(q-1)}{p-q}}.$$

  In addition, the sequence $\{x_k\}_{k=2}^{\infty}$ satisfies

$$\|x_k - \bar{x}\| \leq \left(\frac{C_0}{\beta} \right)^{\frac{1}{p}} \left(\frac{|\bar{S}|^{\frac{q}{2(q-1)}}}{k + C_1 |\bar{S}|^{\frac{q}{2(q-1)}} - 1} \right)^{\frac{(q-1)}{p-q}},$$

  where $C_0 = C_0(p, q, \alpha, \beta)$ and $C_1 = C_1(p, q, \alpha, \beta, E)$.

- *When $p = q = 2$, we have*

$$e_k \leq C_2 \left(1 - \frac{\beta}{\alpha |\bar{S}|} \right)^{k-1}, \quad k \geq 2,$$

$$\|x_k - \bar{x}\| \leq \left(\frac{C_2}{\beta} \right)^{\frac{1}{2}} \left(1 - \frac{\beta}{\alpha |\bar{S}|} \right)^{\frac{k-1}{2}}, \quad k \geq 2,$$

  with $C_2 = C_2(E)$.

Theorem 1.2.2. *Let the objective function E satisfy* **Conditions 0,1,2** *with $M_0 < \alpha q M^{q-1}$. Let problem (1.1) has a solution $\bar{x} = \sum_i c_i(\bar{x})\varphi_i \in \Omega$ with support $\bar{S} := \{i : c_i(\bar{x}) \neq 0\} < \infty$, where $\{\varphi_i\}$ is an orthonormal basis. Then, at Step k, the* **WCGA** *applied to E and $\{\varphi_i\}$ outputs x_k^{w}, where either $x_k^{\mathrm{w}} = \bar{x}$ or:*

- *When $p \neq q \neq 2$, the error $e_k^{\mathrm{w}} := E(x_k^{\mathrm{w}}) - E(\bar{x})$ satisfies the inequality*

$$e_k^{\mathrm{w}} \leq C_0 \left(\frac{|\bar{S}|^{\frac{q}{2(q-1)}}}{\sum_{i=2}^{k} t_i^{\frac{q}{q-1}} + C_1 |\bar{S}|^{\frac{q}{2(q-1)}}} \right)^{\frac{p(q-1)}{p-q}}, \quad k \geq 2.$$

In addition, the sequence $\{x_k^{\mathrm{w}}\}_{k=2}^{\infty}$ *satisfies*

$$\|x_k^{\mathrm{w}} - \bar{x}\| \leq \left(\frac{C_0}{\beta}\right)^{\frac{1}{p}} \left(\frac{|\bar{S}|^{\frac{q}{2(q-1)}}}{\sum_{i=2}^{k} t_i^{\frac{q}{q-1}} + C_1|\bar{S}|^{\frac{q}{2(q-1)}}}\right)^{\frac{(q-1)}{p-q}}, \quad k \geq 2,$$

where $C_0 = C_0(p, q, \alpha, \beta)$ *and* $C_1 = C_1(p, q, \alpha, \beta, E)$*.*

- *When* $p = q = 2$,

$$e_k^{\mathrm{w}} \leq C_2 \prod_{i=2}^{k} \left(1 - \frac{\beta}{\alpha|\bar{S}|} t_i^2\right), \quad k \geq 2,$$

$$\|x_k^{\mathrm{w}} - \bar{x}\| \leq \left(\frac{C_2}{\beta}\right)^{\frac{1}{2}} \prod_{i=2}^{k} \left(1 - \frac{\beta}{\alpha|\bar{S}|} t_i^2\right)^{1/2}, \quad k \geq 2,$$

with $C_2 = C_2(E)$*.*

This dissertation is organized as follows. In Chapter 2, we expand on the conditions imposed on the objective function E. In Chapter 3, we present some of the general results about greedy algorithms and describe the two algorithms (**OMP** and **WCGA**) that we are using for convex optimization. Some auxiliary lemmas needed for our analysis and the main results are presented in Chapter 4.

2. CONDITIONS ON E

In this chapter, we discuss the conditions imposed on the objective function E and justify that they are a natural choice.

2.1 Convexity and Frechet Differentiability

Let us first recall that a function E is Frechet differentiable at $x \in S \subset H$ if there exists a bounded linear functional denoted by $E'(x)$ such that

$$\lim_{h \to 0} \frac{|E(x+h) - E(x) - \langle E'(x), h \rangle|}{\|h\|} = 0.$$

The next lemma (see [4],[13]) shows a relation between convexity and **Condition 2** with constant $\beta = 0$. The proof of this lemma for convex functions on $I\!R^n$ can be found in [4] but for the readers' convenience, we include here the proof in our current setting in which the convex function E is defined on a Hilbert space H.

Lemma 2.1.1. *Let E be a Frechet differentiable function on H. E is convex on H if and only if for some constant $M > 0$,*

$$E(x') - E(x) - \langle E'(x), x' - x \rangle \geq 0, \quad \text{for all } x, x' \in H, \quad \|x - x'\| \leq M. \tag{2.1}$$

Proof. The inequality is trivial if $x = x'$. Assume first that E is convex and $x \neq x' \in H$. Then, for all $0 < t \leq 1$, $(1-t)x + tx' \in H$ and

$$E((1-t)x + tx') \leq (1-t)E(x) + tE(x').$$

This is equivalent to

$$E(x') - E(x) - \frac{E(x + t(x' - x)) - E(x)}{t} \geq 0, \quad 0 < t \leq 1.$$

Therefore, we have

$$E(x') - E(x) - \lim_{t \to 0^+} \frac{E(x + t(x' - x)) - E(x)}{t} \geq 0. \tag{2.2}$$

On the other hand, it follows from the definition of Frechet derivative of E at x for $h = t(x' - x)$ that

$$\lim_{t \to 0} \frac{|E(x + t(x' - x)) - E(x) - t\langle E'(x), x' - x\rangle|}{|t|} = 0,$$

and therefore

$$\lim_{t \to 0^+} \frac{E(x + t(x' - x) - E(x)}{t} = \langle E'(x), x' - x\rangle. \tag{2.3}$$

Combining (2.2) and (2.3) gives

$$E(x') - E(x) - \langle E'(x), x' - x\rangle \geq 0 \quad x, x' \in H.$$

Conversely, assume that E satisfies inequality (2.1). For any $x, x' \in H$, $\|x - x'\| \leq M$, we denote by $y = tx + (1-t)x'$. Applying (2.1) for x, y (note that $\|x - y\| \leq M$) and for x', y yields

$$E(x) - E(y) - \langle E'(y), x - y\rangle \geq 0, \qquad E(x') - E(y) - \langle E'(y), x' - y\rangle \geq 0.$$

Multiplying the first inequality by t and the second by $1-t$ and adding them results in

$$tE(x) + (1-t)E(x') - E(y) \geq 0.$$

Therefore, we have $E(tx+(1-t)x') \leq tE(x)+(1-t)E(x')$ for $x, x' \in H$, $\|x-x'\| \leq M$, which means that E is locally convex on H. Then, see [18], E is convex on H. The proof is complete. $\square$

Recall that a function E is said to be strongly convex on H, if there is a constant $\beta > 0$, such that

$$E(x') - E(x) - \langle E'(x), x' - x\rangle \geq \beta\|x' - x\|^2, \quad x, x' \in H.$$

The constant β is called the convexity parameter of the function E. Note that a strongly convex function on H satisfies **Condition 2** with a value for p, $p = 2$.

Remark 2.1.2. *If* **Condition 2** *holds for a function E that is convex and a set Ω that is convex and bounded, then* **Condition 2** *holds for all x, $x' \in \Omega$ with possibly smaller β. Further in this dissertation, we assume that* **Condition 2** *holds for all x, $x' \in \Omega$ and all x, x', such that $x \in \Omega$ and $\|x - x'\| \leq M$.*

Proof. Assume that for a convex function E, **Condition 2** holds with some constant β_0. Since Ω is bounded, there is $L > 0$, such that $diam(\Omega) \leq LM$. Let $x, x' \in \Omega$. If $\|x - x'\| \leq M$, **Condition 2** holds for the pair (x, x'). If not, we chose a point x_1,

$$x_1 = \gamma x' + (1-\gamma)x \in \Omega, \quad \gamma := \frac{M}{\|x - x'\|} \geq L^{-1}.$$

Clearly $\|x - x_1\| = M$, and therefore $E(x_1) - E(x) - \langle E'(x), x_1 - x\rangle \geq \beta_0\|x_1 - x\|^p$. Because of the convexity of E, $E(x_1) \leq \gamma E(x') + (1-\gamma)E(x)$. A combination of the last two inequalities and the fact that $x_1 - x = \gamma(x' - x)$ result in

$$E(x') - E(x) - \langle E'(x), x' - x\rangle \geq \beta_0 \gamma^{p-1} \|x' - x\|^p \geq \beta_0 L^{1-p} \|x' - x\|^p.$$

Therefore, **Condition 2** holds for all $x, x' \in \Omega$ and all x, x' such that $x \in \Omega$, $\|x - x'\| \leq M$ with $\beta = \min\{\beta_0, \beta_0 L^{1-p}\}$. □

2.2 Moduli of Smoothness

Next, we would like to point out that the class of functions E considered by Temlyakov (see [15] and [16]) is the same as the class of functions satisfying **Condition 0** and **Condition 1**. To see this, we first need the following definitions.

Definition 2.2.1. *Given a convex function $E : H \to I\!R$ and a set $S \subset H$, the modulus of smoothness of E on S is defined by*

$$\rho(E, u) := \rho(E, u, S) := \frac{1}{2} \sup_{x \in S, \|y\|=1} \{E(x + uy) + E(x - uy) - 2E(x)\}, \quad u > 0. \tag{2.4}$$

Definition 2.2.2. *Given a convex function $E : H \to I\!R$ and a set S, the modulus of uniform smoothness of E on S is defined by*

$$\rho_1(E, u, S) := \sup_{x \in S, \|y\|=1, \lambda \in (0,1)} \left\{ \frac{(1-\lambda)E(x - \lambda u y) + \lambda E(x + (1-\lambda)uy) - E(x)}{\lambda(1-\lambda)} \right\}. \tag{2.5}$$

Definition 2.2.3. *A convex function $E : H \to I\!R$ for which*

$$\lim_{u \to 0} \frac{\rho(E, u, S)}{u} = 0$$

is said to be uniformly smooth on S.

Note that the two moduli of smoothness defined above are equivalent (see [19], page 205). For the readers' convenience, we provide the proof below.

Lemma 2.2.4. *Let E be a convex function defined on H and $\rho(E,\cdot,S)$ and $\rho_1(E,\cdot,S)$ be its modulus of smoothness and modulus of uniform smoothness on $S \subset H$, respectively. Then*

$$4\rho(E,\frac{u}{2},S) \le \rho_1(E,u,S) \le 2\rho(E,u,S). \tag{2.6}$$

Proof. First, we prove the left inequality. It follows from the definition of ρ_1 that for every $y \in H$, $\|y\| = 1$,

$$\rho_1(E,u,S) \ge 4\frac{E(x-\frac{u}{2}y)+E(x+\frac{u}{2}y)-2E(x)}{2}.$$

Taking the suppremum over $x \in S$ and $y \in H$, $\|y\| = 1$ results in the desired inequality. For the second inequality, we use the convexity of E. Observe that

$$E(x-\lambda uy) = E((1-\lambda)x+\lambda(x-uy)) \le (1-\lambda)E(x)+\lambda E(x-uy),$$

and

$$E(x+(1-\lambda)uy) = E(\lambda x+(1-\lambda)(x+uy)) \le \lambda E(x)+(1-\lambda)E(x+uy).$$

We multiply the first inequality by $1-\lambda$, the second one by λ and add them. This results in

$$\begin{aligned} &\frac{(1-\lambda)E(x-\lambda uy)+\lambda E(x+(1-\lambda)uy)-E(x)}{\lambda(1-\lambda)} \\ \le\ & E(x+uy)+E(x-uy)-2E(x) \le 2\rho(E,u,S). \end{aligned}$$

Taking the supremum over $\lambda \in (0,1)$, $x \in S$, $y \in H$, $\|y\| = 1$ completes the proof. $\square$

The following lemma is a particular case of Corollary 3.5.7 from [19].

Lemma 2.2.5. *Let E be a convex function defined on a Hilbert space H and E be Frechet differentiable on $S \subset H$. The following statements are equivalent.*

1. *There exist constants $\alpha \geq 0$, $M > 0$ and $q \in (1,2]$ such that for any $x \in S, x' \in H, \|x - x'\| \leq M$, we have*

$$E(x') - E(x) - \langle E'(x), x' - x\rangle \leq \alpha \|x' - x\|^q. \tag{2.7}$$

2. *There exist constants $\alpha_1 \geq 0$, $M > 0$ and $q \in (1,2]$ such that*

$$\rho_1(E, u, S) \leq \alpha_1 u^q, \quad 0 < u \leq M. \tag{2.8}$$

Proof. Let us assume that statement 1 is true. For any $x \in S, y \in H, \|y\| = 1$ and any $0 < u \leq M$, let $x' = x + uy$ and $x'' = x - uy$. Then, we have $\|x - x'\| = u \leq M, \|x'' - x\| = u \leq M$. We apply (2.7) for x', x and x'', x to obtain

$$E(x + uy) - E(x) - u\langle E'(x), y\rangle \leq \alpha u^q,$$

$$E(x - uy) - E(x) + u\langle E'(x), y\rangle \leq \alpha u^q.$$

Therefore, we have

$$E(x + uy) + E(x - uy) - 2E(x) \leq 2\alpha u^q.$$

We take the supremum over $x \in S, y \in H, \|y\| = 1$ and derive

$$\rho(E, u, S) \le \alpha u^q, \quad 0 < u \le M.$$

The latter inequality and Lemma 2.2.4 results in

$$\rho_1(E, u, S) \le 2\alpha u^q, \quad 0 < u \le M,$$

which is (2.8) with $\alpha_1 = 2\alpha$.

Conversely, suppose we have statement 2. Therefore, for any $\lambda \in (0, 1)$ and any $x \in S, y \in H, \|y\| = 1, 0 < u \le M$,

$$\frac{(1-\lambda)E(x - \lambda uy) + \lambda E(x + (1-\lambda)uy) - E(x)}{\lambda(1-\lambda)} \le \alpha_1 u^q.$$

This is the same as

$$\frac{E(x - \lambda uy) - E(x)}{\lambda} + \frac{E(x + (1-\lambda)uy) - E(x)}{1-\lambda} \le \alpha_1 u^q.$$

We let $\lambda \to 0^+$ and by the continuity of E and the definition of Frechet derivative $E'(x)$ with $h = -\lambda uy$, we obtain

$$\langle E'(x), -uy\rangle + E(x + uy) - E(x) \le \alpha_1 u^q.$$

Now, for any $x \in S, x' \in H, \|x' - x\| \le M$, we let $u = \|x' - x\|, y = \frac{x'-x}{\|x'-x\|}$. The above inequality can be written as

$$E(x') - E(x) - \langle E'(x), x' - x\rangle \le \alpha_1 \|x' - x\|^q, \quad x \in S, \|x - x'\| \le M,$$

which is (2.7) with $\alpha = \alpha_1$. $\square$

2.3 Modulus of Uniform Convexity

In this section, we discuss a concept which is dual to the modulus of uniform smoothness for convex functions, called modulus of uniform convexity (see [3], [19]).

Definition 2.3.1. *Given a convex function $E : H \to I\!R$ and a set S, its modulus of uniform convexity on S is defined by*

$$\delta_1(E,u,S) := \inf_{x\in S, \|y\|=1, \lambda\in(0,1)} \left\{ \frac{(1-\lambda)E(x-\lambda uy) + \lambda E(x+(1-\lambda)uy) - E(x)}{\lambda(1-\lambda)} \right\}. \tag{2.9}$$

Next, we prove a lemma (see [19]) that shows the equivalence of **Condition 2** and certain behavior of the modulus of uniform convexity δ_1 of E.

Lemma 2.3.2. *Let E be a convex function defined on a Hilbert space H and E be Frechet differentiable on $S \subset H$. The following statements are equivalent.*

1. *There exist constants $\beta \geq 0$, $M > 0$ and $p \in [2,\infty)$ such that for any $x \in S, x' \in H, \|x - x'\| \leq M$,*

$$E(x') - E(x) - \langle E'(x), x' - x\rangle \geq \beta \|x' - x\|^p. \tag{2.10}$$

2. *There exist constants $\beta_1 \geq 0$, $M > 0$ and $p \in [2,\infty)$ such that*

$$\delta_1(E,u,S) \geq \beta_1 u^p, \quad 0 < u \leq M. \tag{2.11}$$

Proof. Let us assume that statement 1 is true. For any $x \in S, y \in H, \|y\| = 1$, $0 < u \leq M$ and $\lambda \in (0,1)$, let $x' = x - \lambda uy$ and $x'' = x + (1-\lambda)uy$. Then, we have

$\|x - x'\| = \lambda u \leq M, \|x'' - x\| = (1-\lambda)u \leq M$. We apply (2.10) for $x \in S, x' \in H$ and $x \in S, x'' \in H$ to derive

$$E(x - \lambda u y) - E(x) + \lambda u \langle E'(x), y\rangle \geq \beta \lambda^p u^p,$$

$$E(x + (1-\lambda)uy) - E(x) - (1-\lambda)u\langle E'(x), y\rangle \geq \beta(1-\lambda)^p u^p.$$

Multiplying the first inequality by $(1-\lambda)$, the second one by λ and adding them yields

$$(1-\lambda)E(x-\lambda u y) + \lambda E(x + (1-\lambda)uy) - E(x) \geq \beta\lambda(1-\lambda)(\lambda^{p-1} + (1-\lambda)^{p-1})u^p.$$

Note that for $\lambda \in (0,1)$, $(\lambda^{p-1} + (1-\lambda)^{p-1}) \geq 2^{2-p}$. Therefore, we have

$$\frac{(1-\lambda)E(x-\lambda u y) + \lambda E(x + (1-\lambda)uy) - E(x)}{\lambda(1-\lambda)} \geq 2^{2-p}\beta u^p.$$

We take the infimum over $x \in S, y \in H, \|y\| = 1$ and $\lambda \in (0,1)$ and obtain

$$\delta_1(E, u, S) \geq 2^{2-p}\beta u^p, \quad 0 < u \leq M,$$

which is (2.11) with $\beta_1 = 2^{2-p}\beta$.

Conversely, suppose we have statement 2 *i.e* for some $\beta_1 > 0$,

$$\delta_1(E, u, S) \geq \beta_1 u^p, \quad 0 < u \leq M.$$

It follows from the definition of δ_1 that for any $\lambda \in (0,1)$, $x \in S, y \in H, \|y\| = 1$ and

$0 < u \le M$,

$$\frac{(1-\lambda)E(x-\lambda uy)+\lambda E(x+(1-\lambda)uy)-E(x)}{\lambda(1-\lambda)} \ge \beta_1 u^p.$$

This is the same as

$$\frac{E(x-\lambda uy)-E(x)}{\lambda}+\frac{E(x+(1-\lambda)uy)-E(x)}{1-\lambda} \ge \beta_1 u^p.$$

We let $\lambda \to 0^+$ and by the continuity of E and the definition of Frechet derivative $E'(x)$ at x for $h = -\lambda uy$, we obtain

$$\langle E'(x), -uy\rangle + E(x+uy) - E(x) \ge \beta_1 u^p.$$

Now, for any $x \in S, x' \in H, \|x'-x\| \le M$, we let $u = \|x'-x\|, y = \frac{x'-x}{\|x'-x\|}$ and derive

$$E(x') - E(x) - \langle E'(x), x'-x\rangle \ge \beta_1 \|x'-x\|^p,$$

which is (2.10) with $\beta = \beta_1$. □

We combine Lemma 2.2.5 and Lemma 2.3.2 to state conditions on E that are equivalent to **Condition 1** and **Condition 2**.

Lemma 2.3.3. *Let E be a convex function defined on a Hilbert space H, $\Omega = \{x \in H : E(x) \le E(0)\}$ and $\delta_1(E,\cdot,\Omega)$, $\rho_1(E,\cdot,\Omega)$ be its moduli of uniform convexity and uniform smoothness on Ω, respectively. Let E be Frechet differentiable on Ω. The following two statements are equivalent*

- *E satisfies* **Condition 1** *and* **Condition 2**.
- *There exist constants $\alpha_1 \ge 0, \beta_1 \ge 0$, $M > 0$, $p \in [2,\infty)$ and $q \in (1,2]$, such*

that

$$\beta_1 u^p \le \delta_1(E,u,\Omega) \le \rho_1(E,u,\Omega) \le \alpha_1 u^q, \quad u \in (0,M]. \tag{2.12}$$

Remark 2.3.4. *All our proofs go through if instead of* **Condition 0**, **Condition 1** *and* **Condition 2**, *E satisfies the following three conditions:*

Condition $0'$**:** *E is a function defined on a Hilbert space H, is Frechet differentiable at each point in H and its Frechet derivative is uniformly bounded on Ω by a constant M_0.*

Condition $1'$**:** *There are constants $\alpha \ge 0$, $1 < q \le 2$ and $M > 0$, such that*

$$E(x') - E(x) - \langle E'(x), x'-x\rangle \le \alpha\|x'-x\|^q, \quad \text{for } x, x' \text{ such that } \|x-x'\| \le M. \tag{2.13}$$

Condition $2'$**:** *There are constants $\beta \ge 0$, $2 \le p < \infty$ and $M > 0$, such that*

$$E(x') - E(x) - \langle E'(x), x'-x\rangle \ge \beta\|x'-x\|^p, \quad \text{for } x, x' \text{ such that } \|x-x'\| \le M. \tag{2.14}$$

In both **Condition** $1'$ *and* **Condition** $2'$, *E' is the Frechet derivative of E.*

Note that **Condition** $2'$ implies that E is locally convex on H (see Lemma 2.1.1), that is

$$E(\gamma x + \delta y) \le \gamma E(x) + \delta E(y), \quad x, y \in H, \quad \|x-y\| \le M, \quad \gamma, \delta \ge 0, \quad \gamma + \delta = 1,$$

and therefore (see [18]) E is convex on H. In this case Ω is a level set of a convex function and hence Ω is convex.

Conditions similar to **Condition 1** and **Condition 2** have been considered by Zhang in [20], where he solves a sparse optimization problem in $\mathbb{R}^n$, using greedy based strategies. More precisely, the class of functions he considers is the set of all convex functions E for which there are constants $\alpha(s), \beta(s) > 0$ such that

$$\beta(s)\|x' - x\|_2^2 \leq E(x') - E(x) - \langle E'(x), x' - x\rangle \leq \alpha(s)\|x' - x\|_2^2, \tag{2.15}$$

for any $x, x' \in I\!\!R^n$, where $x - x'$ has $\leq s$ nonzero coordinates. In particular, when $E(x) = \|Ax - b\|_2^2$, where $\|x\|_2^2 = x_1^2 + x_2^2 + \cdots + x_n^2$ is the Euclidean norm of $x \in I\!\!R^n$ and A is a given $k \times n$ matrix with $k << n$, $E'(x)$ can be computed explicitly. In this case, we have that the linear functional $E'(x)$ is given by the formula

$$\langle E'(x), \cdot\rangle = 2\langle A^T(Ax - b), \cdot\rangle,$$

and we can compute

$$\begin{aligned}
& E(x') - E(x) - \langle E'(x), x' - x\rangle \\
= \quad & \|Ax' - b\|_2^2 - \|Ax - b\|_2^2 - 2\langle A^T(Ax - b), x' - x\rangle \\
= \quad & \|Ax' - b\|_2^2 - \|Ax - b\|_2^2 - 2\langle Ax - b, Ax' - Ax\rangle \\
= \quad & \|Ax'\|_2^2 - 2\langle Ax', b\rangle + \|b\|_2^2 - (\|Ax\|_2^2 - 2\langle Ax, b\rangle + \|b\|_2^2) \\
& -2(\langle Ax, Ax'\rangle - \langle b, Ax'\rangle - \|Ax\|^2 + \langle b, Ax\rangle) \\
= \quad & \|Ax'\|_2^2 - 2\langle Ax', Ax\rangle + \|Ax\|_2^2 \\
= \quad & \|Ax' - Ax\|_2^2 = \|A(x - x')\|_2^2.
\end{aligned}$$

Let us denote by $z = x' - x$. Then, condition (2.15) becomes

$$\beta(s)\|z\|_2^2 \leq \|Az\|_2^2 \leq \alpha(s)\|z\|_2^2. \tag{2.16}$$

Condition (2.16) is known as the *Restricted Isometry Property* (**RIP**) and is widely used in Compressed Sensing. The **RIP** (see, for example, [7],[2],[11]) was first introduced by Candes and Tao (see [5], [6], [10]) with $\beta(s) = 1 - \delta_s, \alpha(s) = 1 + \delta_s$, $\delta_s \in (0, 1)$ and vectors z that have at most s non-zero coordinates. In our case, the constants $\alpha(s), \beta(s)$ only need to be uniformly bounded as in [20].

3. GREEDY ALGORITHMS

3.1 Introduction

Greedy algorithms were first introduced in non-linear approximation as a tool to find the best approximant to a function using the elements of a symmetric dictionary (see [1],[9],[12],[16]). In [9], DeVore and Temlyakov studied three greedy algorithms (pure greedy, orthogonal greedy and relaxed greedy) for approximating functions f from a Hilbert space H using elements from a symmetric dictionary $\mathcal{D}$.

Here, we describe the orthogonal greedy algorithm from [9] for approximating f in a Hilbert space H with inner product $\langle \cdot, \cdot \rangle$. A modification of this algorithm for the purposes of solving problem (1.1) was investigated by Temlyakov in [16].
In what follows below, we denote by $g(h) := \text{argmax}\{\langle h, \varphi \rangle, \varphi \in \mathcal{D}\}$ for any element $h \in H$.

Orthogonal Greedy Algorithm (OGA)

- **Step** 0:

  Start with initial guess $G_0(f) := 0$ and set $R_0(f) := f$.

- **Step** k, $k = 1, \dots, m$:

  Find the space $H_k := H_k(f) := \text{span}\{g(R_0(f)), \dots, g(R_{k-1}(f))\}$.

  Find $G_k(f) := G_k(f, \mathcal{D}) := P_{H_k} f$, where $P_{H_k} f$ is the orthogonal projection of f onto H_k.

  Find the residual $R_k(f) := R_k(f, \mathcal{D}) := f - G_k(f)$.

- **Step** $m + 1$:

  Output $G_m(f)$.

If we denote by $\mathcal{A}_1(\mathcal{D}, P)$ the closure in H of the set

$$\{f \in H : f = \Sigma_{k\in I} c_k \varphi_k,\ \varphi_k \in \mathcal{D}, I \text{ finite},\ \Sigma_{k\in I}|c_k| \leq P\},$$

the following theorem from [9] shows the approximation properties of **OGA**.

Theorem 3.1.1 (see [9], Theorem 3.7)**.** *Let $\mathcal{D}$ be a symmetric dictionary in H. Then for each $f \in \mathcal{A}_1(\mathcal{D}, P)$, we have*

$$\|f - G_m(f)\| \leq P m^{-\frac{1}{2}}.$$

3.2 **OMP** and **WCGA** for Optimization

In this section, we describe the analogues of **OGA** that are used for convex optimization. The algorithm (**OMP**) and its weak version (**WCGA**) are introduced in [16], where the approximation properties of these methods are also proved.

Orthogonal Matching Pursuit (OMP)

- **Step** 0:

  Start with initial guess $x_0 = 0$.

  If $E'(x_0) = 0$, $x_m := x_0 = 0$. Go to **Step** $m + 1$.

- **Step** $k, k = 1, 2, \ldots m$:

  Find $\varphi_{j_k} := \text{argmax}\{|\langle E'(x_{k-1}), \varphi\rangle|, \varphi \in \mathcal{D}\}$.

  Find $x_k := \text{argmin}_{x\in span\{\varphi_{j_1},\varphi_{j_2},\ldots,\varphi_{j_k}\}} E(x)$.

  If $E'(x_k) = 0$, $x_m := x_k$. Go to **Step** $m + 1$.

- **Step** $m + 1$:

  Output x_m.

In order to describe the main convergence results for this algorithm, proved in [16], we need some additional notation. We denote the closure (in H) of the convex hull of $\mathcal{D}$ by $\mathcal{A}_1(\mathcal{D})$.

Theorem 3.2.1 ([16] Theorem 2.2)**.** *Let E be a uniformly smooth convex function defined on a Banach space X and let the set $\Omega := \{x : E(x) \leq E(0)\}$ be bounded. Let the modulus of smoothness of E, $\rho(E, u, \Omega) \leq \gamma u^q$, where $1 < q \leq 2$. Take $\epsilon > 0$ and an element $\varphi^\epsilon \in \mathcal{D}$, such that*

$$E(\varphi^\epsilon) \leq \inf_{x\in\Omega} E(x) + \epsilon, \quad \varphi^\epsilon/A(\epsilon) \in \mathcal{A}_1(\mathcal{D}),$$

for some constant $A(\epsilon) \geq 1$. Then, the output x_m of the **OMP** *satisfies the inequality*

$$E(x_m) - \inf_{x\in\Omega} E(x) \leq \max\{2\epsilon, C_1 A(\epsilon)^q \big(C_2 + m\big)^{1-q}\},$$

with constants $C_1 = C_1(q, \gamma)$ and $C_2 = C_2(E, q, \gamma)$.

Next, we describe the weak version of the **OMP**, called Weak Chebysev Greedy Algorithm (**WCGA**). We call the sequence $\{t_k\}_{k=1}^{\infty}$, $t_k \in (0, 1]$, a weakness sequence. It is used to weaken the condition on the choice of φ_{j_k} (see [14], [17]). Notice that when all $t_k = 1$, the **WCGA** is actually the **OMP**.

Weak Chebysev Greedy Algorithm (WCGA)

- **Step** 0:

  Start with initial guess $x_0^{\mathrm{w}} = 0$.

  If $E'(x_0^{\mathrm{w}}) = 0$, $x_m^{\mathrm{w}} := x_0^{\mathrm{w}} = 0$. Go to **Step** $m + 1$.

- **Step** $k, k = 1, 2, \ldots m$:

Find φ_{j_k} such that $|\langle E'(x^{\mathrm{w}}_{k-1}), \varphi_{j_k}\rangle| \geq t_k \sup_{\varphi\in\mathcal{D}} |\langle E'(x^{\mathrm{w}}_{k-1}), \varphi\rangle|$.

Find $x^{\mathrm{w}}_k := \mathrm{argmin}_{x\in span\{\varphi_{j_1},\varphi_{j_2},\ldots,\varphi_{j_k}\}} E(x)$.

If $E'(x^{\mathrm{w}}_k)$=0, $x^{\mathrm{w}}_m := x^{\mathrm{w}}_k$. Go to **Step** $m+1$.

- **Step** $m+1$:

  Output x^{w}_m.

For the weak version, the following result is proved by Temlyakov.

Theorem 3.2.2 ([16] Theorem 2.2)**.** *Let E be a uniformly smooth convex function defined on a Banach space X and let the set $\Omega := \{x : E(x) \leq E(0)\}$ be bounded. Let the modulus of smoothness of E, $\rho(E, u, \Omega) \leq \gamma u^q$, where $1 < q \leq 2$. Take $\epsilon > 0$ and an element $\varphi^\epsilon \in \mathcal{D}$, such that*

$$E(\varphi^\epsilon) \leq \inf_{x\in\Omega} E(x) + \epsilon, \quad \varphi^\epsilon / A(\epsilon) \in \mathcal{A}_1(\mathcal{D}),$$

for some constant $A(\epsilon) \geq 1$. Then, the output x^{w}_m of the **WCGA** *satisfies the inequality*

$$E(x^{\mathrm{w}}_m) - \inf_{x\in\Omega} E(x) \leq \max\{2\epsilon, C_1 A(\epsilon)^q \big(C_2 + \Sigma_{k=1}^m t_k^{q/(q-1)}\big)^{1-q}\},$$

with constants $C_1 = C_1(q, \gamma)$ and $C_2 = C_2(E, q, \gamma)$.

4. MAIN RESULTS

4.1 Auxiliary Lemmas

In this section, we will prove several auxiliary lemmas that we will need for the analysis of the above greedy algorithms. We start with the statement of the following lemma.

Lemma 4.1.1. *Let E be a Frechet differentiable convex function, defined on a convex set Ω. Then E has a global minimum at $x_0 \in \Omega$ if and only if $E'(x_0) = 0$.*

Proof. Clearly, if E has a global minimum at x_0, by Fermat's theorem, see [4], we have that $E'(x_0) = 0$. This result holds not only for a convex function E, but for any function E. Now, suppose that E is convex on Ω and $E'(x_0) = 0$. For any $x \in \Omega$, $t \in (0, 1]$, $(1 - t)x_0 + tx \in \Omega$, and

$$E((1 - t)x_0 + tx) \leq (1 - t)E(x_0) + tE(x).$$

This is equivalent to

$$E(x) - E(x_0) - \frac{E((1 - t)x_0 + tx) - E(x_0)}{t} \geq 0, \quad 0 < t \leq 1.$$

Therefore, we have

$$E(x) - E(x_0) - \lim_{t \to 0^+} \frac{E(x_0 + t(x - x_0)) - E(x_0)}{t} \geq 0.$$

Since E is Frechet differentiable at x_0 and $E'(x_0) = 0$, we have

$$E(x) \geq E(x_0), \quad x \in \Omega.$$

The proof is completed. $\square$

Lemma 4.1.2. *For every convex function E defined on the whole Hilbert space H, the set $\Omega := \{x \in H : E(x) \leq E(0)\}$ is convex and all $\{x_k\}_{k=1}^{\infty}$ generated from the* **OMP** *and all $\{x_k^{\mathrm{w}}\}_{k=1}^{\infty}$ generated from* **WCGA** *are in Ω.*

Proof. The first statement is straight forward by the property of convex function. It is usually said as every level set of a convex function is also convex. Indeed, for any $x, y \in \Omega$ and any $t \in [0, 1]$,

$$E((1-t)x + ty) \leq (1-t)E(x) + tE(y) \leq (1-t)E(0) + tE(0) = E(0),$$

which means that $(1-t)x + ty \in \Omega$.
Now, we will prove that all $\{x_k\}_{k=1}^{\infty}$ generated from **OMP** belong to Ω. The proof for the **WCGA** sequence is similar. By the definition of the algorithm,

$$x_k := \mathrm{argmin}_{x \in \mathrm{span}\{\varphi_{j_1}, \varphi_{j_2}, \ldots, \varphi_{j_k}\}} E(x),$$

$$x_{k-1} := \mathrm{argmin}_{x \in \mathrm{span}\{\varphi_{j_1}, \varphi_{j_2}, \ldots, \varphi_{j_{k-1}}\}} E(x),$$

and therefore we have $E(x_k) \leq E(x_{k-1})$. Thus the sequences $\{E(x_k)\}_{k=1}^{\infty}$ is decreasing and hence $E(x_k) \leq E(x_0) = E(0)$ for every $k = 1, 2, \ldots$. In other words, the sequence $\{x_k\}_{k=1}^{\infty} \subset \Omega$. $\square$

Lemma 4.1.3. *Let $F : H \to I\!R$ be a Frechet differentiable function. Let $V_k :=$ span$\{\varphi_{j_1}, \ldots, \varphi_{j_k}\} \subset H$ and $x_k = \mathrm{argmin}\{F(x) : x \in V_k\}$. Then*

$$\langle F'(x_k), \varphi \rangle = 0 \quad \textit{for every} \quad \varphi \in V_k.$$

Proof. F is Frechet differentiable at x_k, and hence

$$\lim_{y\to 0}\frac{|F(x_k+y)-F(x_k)-\langle F'(x_k),y\rangle|}{\|y\|}=0.$$

We fix $\ell\in\{1,\dots,k\}$ and let $y=t\varphi_{j_\ell}$, where $t\in\mathbb{R}$. Without loss of generality we can assume that $\|\varphi_{j_\ell}\|=1$. We have

$$\lim_{t\to 0}\left|\frac{F(x_k+t\varphi_{j_\ell})-F(x_k)}{t}-\langle F'(x_k),\varphi_{j_\ell}\rangle\right|=0,$$

and therefore

$$\langle F'(x_k),\varphi_{j_\ell}\rangle=\lim_{t\to 0}\frac{F(x_k+t\varphi_{j_\ell})-F(x_k)}{t}.$$

From the definition of x_k it follows that $F(x_k+t\varphi_{j_\ell})-F(x_k)\geq 0$ for every $t\in\mathbb{R}$. Then, we have

$$\langle F'(x_k),\varphi_{j_\ell}\rangle=\lim_{t\to 0^+}\frac{F(x_k+t\varphi_{j_\ell})-F(x_k)}{t}\geq 0,$$

and

$$\langle F'(x_k),\varphi_{j_\ell}\rangle=\lim_{t\to 0^-}\frac{F(x_k+t\varphi_{j_\ell})-F(x_k)}{t}\leq 0.$$

This results in $\langle F'(x_k),\varphi_{j_\ell}\rangle=0$. □

Our next lemma can be viewed as a generalized version of Lemma 2.16 in [17]. Lemmas of this type are well known in approximation theory (see [8]).

Lemma 4.1.4. *Let $\ell>0$, $r>0$, $B>0$, and $\{a_m\}_{m=1}^{\infty}$ and $\{r_m\}_{m=2}^{\infty}$ be sequences of non-negative numbers satisfying the inequalities*

$$a_1\leq B,\quad a_{m+1}\leq a_m\left(1-\frac{r_{m+1}}{r}a_m^{\ell}\right),\quad m=1,2,\dots.$$

Then, we have for $m = 2, 3, \dots$

$$a_m \le \begin{cases} \dfrac{r^{1/\ell}}{(rB^{-\ell} + \Sigma_{k=2}^{m} r_k)^{1/\ell}}, & \textit{if} \;\; \ell \ge 1, \\[2ex] \dfrac{r^{1/\ell}}{(rB^{-\ell} + \ell\Sigma_{k=2}^{m} r_k)^{1/\ell}}, & \textit{if} \;\; 0 < \ell \le 1. \end{cases} \tag{4.1}$$

Proof. Let us first notice that since all a_m's are non-negative, it follows from the recursive relation that

$$0 \le 1 - \frac{r_{m+1}}{r} a_m^\ell \le 1, \quad m = 1, 2, \dots. \tag{4.2}$$

We prove the lemma by induction.

Case 1: $\ell \ge 1$. If $a_2 = 0$, all $a_m = 0$, $m = 3, 4, \dots$, and the lemma is true. Let us assume that $a_2 > 0$. This also means that $a_1 > 0$ and it follows from the recursive relation, (4.2) and the fact that $\ell \ge 1$ that

$$a_2^{-\ell} \ge a_1^{-\ell}(1 - \frac{r_2}{r} a_1^\ell)^{-\ell} \ge a_1^{-\ell}(1 - \frac{r_2}{r} a_1^\ell)^{-1} \ge a_1^{-\ell}(1 + \frac{r_2}{r} a_1^\ell) = a_1^{-\ell} + \frac{r_2}{r} \ge B^{-\ell} + \frac{r_2}{r}.$$

This gives

$$\frac{r}{rB^{-\ell} + r_2} \ge a_2^\ell,$$

and we have (4.1) for $m = 2$.

We now assume that (4.1) is true for m and will prove it for $m + 1$. Similarly to the case $m = 2$, we may assume that $a_{m+1} > 0$. Because of the recursive relation, this also means that $a_m > 0$ and using (4.2) and that $\ell \geq 1$, we derive

$$\begin{aligned} a_{m+1}^{-\ell} &\geq a_m^{-\ell}(1 - \frac{r_{m+1}}{r}a_m^{\ell})^{-\ell} \geq a_m^{-\ell}(1 - \frac{r_{m+1}}{r}a_m^{\ell})^{-1} \\ &\geq a_m^{-\ell}(1 + \frac{r_{m+1}}{r}a_m^{\ell}) = a_m^{-\ell} + \frac{r_{m+1}}{r}. \end{aligned} \tag{4.3}$$

Now, from the induction hypothesis we have that

$$a_m^{-\ell} \geq \frac{rB^{-\ell} + \Sigma_{k=2}^{m} r_k}{r},$$

which combined with (4.3) proves the lemma in the case $\ell \geq 1$.

Case 2: $0 < \ell < 1$**.** Again, we only consider the case when $a_2 > 0$. We will use the fact that for $0 < \ell < 1$, the function $(1-t)^{\ell}$ is concave. Therefore, we have

$$(1-t)^{\ell} \leq 1 - \ell t, \quad 0 \leq t \leq 1. \tag{4.4}$$

We apply this inequality with $t = \frac{r_2}{r}a_1^{\ell} \in [0,1]$ and obtain

$$a_2^{-\ell} \geq a_1^{-\ell}(1 - \frac{r_2}{r}a_1^{\ell})^{-\ell} \geq a_1^{-\ell}(1 - \ell\frac{r_2}{r}a_1^{\ell})^{-1} \geq a_1^{-\ell}(1 + \ell\frac{r_2}{r}a_1^{\ell}) = a_1^{-\ell} + \ell\frac{r_2}{r} \geq B^{-\ell} + \ell\frac{r_2}{r}.$$

This gives

$$\frac{r}{rB^{-\ell} + \ell r_2} \geq a_2^{\ell},$$

and we have (4.1) for $m = 2$. Next , we assume that (4.1) is true for m and prove it for $m+1$. We consider only the case $a_{m+1} > 0$ (the lemma is true if $a_{m+1} = 0$), and therefore $a_m > 0$. From the recursive relation and (4.4) with $t = \frac{r_{m+1}}{r}a_m^{\ell} \in [0,1]$, we have

$$a_{m+1}^{-\ell} \geq a_m^{-\ell}(1-\frac{r_{m+1}}{r}a_m^{\ell})^{-\ell} \geq a_m^{-\ell}(1-\ell\frac{r_{m+1}}{r}a_m^{\ell})^{-1} \geq a_m^{-\ell}(1+\ell\frac{r_{m+1}}{r}a_m^{\ell}) = a_m^{-\ell}+\ell\frac{r_{m+1}}{r}.$$

This inequality, combined with the induction hypothesis gives that

$$a_{m+1}^{-\ell} \geq \frac{rB^{-\ell} + \ell\Sigma_{k=2}^{m+1} r_k}{r},$$

and the proof is completed. □

The previous lemma can be stated in the following way.

Lemma 4.1.5. *Let $\ell > 0$, $r > 0$, $B > 0$, and $\{a_m\}_{m=1}^{\infty}$ and $\{r_m\}_{m=2}^{\infty}$ be sequences of non-negative numbers satisfying the inequalities*

$$a_1 \leq B, \quad a_{m+1} \leq a_m(1 - \frac{r_{m+1}}{r}a_m^{\ell}), \quad m = 1, 2, \ldots.$$

Then, we have for $m = 2, 3, \ldots$

$$a_m \leq \max\{1, \ell^{-1/\ell}\} r^{1/\ell}(rB^{-\ell} + \Sigma_{k=2}^{m} r_k)^{-1/\ell}. \tag{4.5}$$

Proof. The inequality (4.5) follows from Lemm 4.1.4 and the fact that for $0 < \ell \leq 1$

$$\frac{r^{1/\ell}}{(rB^{-\ell} + \ell\Sigma_{k=2}^{m} r_k)^{1/\ell}} \leq \ell^{-1/\ell}\frac{r^{1/\ell}}{(rB^{-\ell} + \Sigma_{k=2}^{m} r_k)^{1/\ell}}.$$

□

4.2 Convergence Results for **OMP**

In this section, we analyze the performance of the **OMP** algorithm when applied to the minimization problem (1.1) in the case that the dictionary $\mathcal{D}$ is an orthonormal system $\{\varphi_i\}_{i=1}^{\infty}$. Let us denote by e_k the error of the algorithm at Step k, namely

$$e_k := E(x_k) - E(\bar{x}).$$

Lemma 4.2.1. *Let the objective function E satisfy* **Conditions 0,1** *and* **2** *with $M_0 < \alpha q M^{q-1}$. Let problem* (1.1) *have a solution $\bar{x} = \sum_i c_i(\bar{x})\varphi_i \in \Omega$ with support*

$$\bar{S} := \{i : \; c_i(\bar{x}) \neq 0\} < \infty,$$

where $\{\varphi_i\}$ is an orthonormal basis. Then the **OMP** *applied to E and $\{\varphi_i\}$ satisfies the following inequality*

$$e_1 \leq E(0) - E(\bar{x}), \tag{4.6}$$

and

$$e_k \leq e_{k-1} - \frac{1}{A|\bar{S}|^{\frac{q}{2(q-1)}}} e_{k-1}^{\frac{(p-1)q}{(q-1)p}}, \quad k \geq 2, \tag{4.7}$$

where

$$A = \frac{(p-1)^{\frac{q(p-1)}{p(q-1)}}}{(q-1)} \left(\frac{\alpha q^q}{\beta^{q/p} p^q}\right)^{\frac{1}{q-1}}.$$

Proof. Clearly $e_1 = E(x_1) - E(\bar{x}) \leq E(0) - E(\bar{x})$ since $x_1 := \operatorname{argmin}_{x \in span\{\varphi_{j_1}\}} E(x)$. Note that by Lemma 4.1.2, all $\{x_k\}$ generated from the algorithm **OMP** are in Ω. Next, we consider Step k, $k = 2, 3, \ldots$ of the algorithm. Observe that if at Step (k-1) we have that $\bar{S} \subset \{j_1, \ldots, j_{k-1}\}$, then $x_{k-1} = \bar{x}$, $E'(x_{k-1}) = 0$ and the **OMP** would have stopped with output $x_{k-1} = \bar{x}$. If the algorithm has not stopped, then we have found x_k and φ_{j_k}. Since $x_k := \text{argmin}_{x \in span\{\varphi_{j_1}, \varphi_{j_2}, \ldots, \varphi_{j_k}\}} E(x)$, we have for any $0 < t < M$

$$E(x_k) \leq E(x_{k-1} - t \cdot \text{sgn}\{\langle E'(x_{k-1}), \varphi_{j_k} \rangle\} \varphi_{j_k}). \tag{4.8}$$

Now, we use **Condition 1** with $x' = x_{k-1} - t \cdot \text{sgn}\{\langle E'(x_{k-1}), \varphi_{j_k} \rangle\} \varphi_{j_k}$ and $x = x_{k-1}$ to obtain

$$E(x_{k-1} - t \cdot \text{sgn}\langle E'(x_{k-1}), \varphi_{j_k} \rangle \varphi_{j_k}) \leq E(x_{k-1}) - t |\langle E'(x_{k-1}), \varphi_{j_k} \rangle| + \alpha t^q, \tag{4.9}$$

since $\|\varphi_{j_k}\| = 1$. It follows from (4.8) and (4.9) that for any $0 < t < M$,

$$E(x_k) \leq E(x_{k-1}) - t |\langle E'(x_{k-1}), \varphi_{j_k} \rangle| + \alpha t^q =: \Phi(t).$$

The function Φ achieves a minimum value

$$\Phi(t^*) = E(x_{k-1}) - \frac{q-1}{q} (\alpha q)^{-\frac{1}{q-1}} |\langle E'(x_{k-1}), \varphi_{j_k} \rangle|^{q/(q-1)}$$

at

$$t = t^* = (\alpha q)^{-\frac{1}{q-1}} |\langle E'(x_{k-1}), \varphi_{j_k} \rangle|^{\frac{1}{q-1}}.$$

Notice that $t^* \leq \left[\frac{M_0}{\alpha q}\right]^{1/(q-1)} < M$. Therefore, we have

$$E(x_k) \leq \Phi(t^*) = E(x_{k-1}) - \frac{q-1}{q} (\alpha q)^{-\frac{1}{q-1}} |\langle E'(x_{k-1}), \varphi_{j_k} \rangle|^{q/(q-1)}. \tag{4.10}$$

Now, we will find a lower bound for $|\langle E'(x_{k-1}), \varphi_{j_k} \rangle|$. First, we use **Condition 2** with $x' = \bar{x}$ and $x = x_{k-1}$ to obtain

$$\langle E'(x_{k-1}), x_{k-1} - \bar{x} \rangle \geq E(x_{k-1}) - E(\bar{x}) + \beta \|\bar{x} - x_{k-1}\|^p. \tag{4.11}$$

Let us recall the generalized Cauchy inequality

$$\frac{p_1}{p_1 + p_2} a + \frac{p_2}{p_1 + p_2} b \geq a^{\frac{p_1}{p_1+p_2}} b^{\frac{p_2}{p_1+p_2}}, \quad a, b, p_1, p_2 > 0,$$

and apply it for $p_1 = p - 1$, $p_2 = 1$, $a = \frac{E(x_{k-1}) - E(\bar{x})}{p-1}$, $b = \beta \|\bar{x} - x_{k-1}\|^p$. We have

$$E(x_{k-1}) - E(\bar{x}) + \beta \|\bar{x} - x_{k-1}\|^p = p \left(\frac{(p-1)}{p} \frac{E(x_{k-1}) - E(\bar{x})}{p-1} + \frac{1}{p} \beta \|\bar{x} - x_{k-1}\|^p \right),$$

and therefore

$$E(x_{k-1}) - E(\bar{x}) + \beta \|\bar{x} - x_{k-1}\|^p \geq C \|\bar{x} - x_{k-1}\| \left(E(x_{k-1}) - E(\bar{x}) \right)^{(p-1)/p},$$

with $C = p\beta^{1/p}(p-1)^{(1-p)/p}$. We combine this inequality with (4.11) to obtain

$$\langle E'(x_{k-1}), x_{k-1} - \bar{x} \rangle \geq C \|\bar{x} - x_{k-1}\| \left(E(x_{k-1}) - E(\bar{x}) \right)^{(p-1)/p}. \tag{4.12}$$

It follows from Lemma 4.2.3 that

$$\langle E'(x_{k-1}), \varphi_i \rangle = 0, \quad i = j_1, \ldots, j_{k-1}.$$

Therefore, if we write

$$x_{k-1} - \bar{x} = \sum_i c_i(x_{k-1} - \bar{x}) \varphi_i,$$

since the support of x_{k-1} is $\{j_1, \ldots, j_{k-1}\}$, we obtain

$$
\begin{aligned}
\langle E'(x_{k-1}), x_{k-1} - \bar{x}\rangle &= \sum_{i \in \bar{S}\setminus\{j_1,\ldots,j_{k-1}\}} c_i(x_{k-1} - \bar{x})\langle E'(x_{k-1}), \varphi_i\rangle, \\
&\leq \sum_{i \in \bar{S}\setminus\{j_1,\ldots,j_{k-1}\}} |c_i(x_{k-1} - \bar{x})||\langle E'(x_{k-1}), \varphi_{j_k}\rangle| \\
&\leq |\langle E'(x_{k-1}), \varphi_{j_k}\rangle||\bar{S}|^{1/2}\|x_{k-1} - \bar{x}\|. \qquad (4.13)
\end{aligned}
$$

We combine this with (4.12) to obtain

$$
|\langle E'(x_{k-1}), \varphi_{j_k}\rangle|\|\bar{x} - x_{k-1}\||\bar{S}|^{1/2} \geq C\|\bar{x} - x_{k-1}\| \left(E(x_{k-1}) - E(\bar{x})\right)^{(p-1)/p},
$$

and therefore we have the desired lower bound

$$
|\langle E'(x_{k-1}), \varphi_{j_k}\rangle| \geq C|\bar{S}|^{-1/2} \left(E(x_{k-1}) - E(\bar{x})\right)^{(p-1)/p}.
$$

We combine this result with (4.10) to obtain the estimate

$$
E(x_k) \leq E(x_{k-1}) - \frac{1}{A|\bar{S}|^{\frac{q}{2(q-1)}}} \left(E(x_{k-1}) - E(\bar{x})\right)^{\frac{(p-1)q}{(q-1)p}},
$$

where

$$
A = \frac{(p-1)^{\frac{q(p-1)}{p(q-1)}}}{(q-1)} \left(\frac{\alpha q^q}{\beta^{q/p} p^q}\right)^{\frac{1}{q-1}}.
$$

Subtracting $E(\bar{x})$ from both sides of the inequality results in (4.7) and the proof is completed. □

Remark 4.2.2. *Note that in the proof of Lemma 4.2.1* **Condition 1** *needs to hold only for vectors x and x' such that $(x - x')$ is 1-sparse.*

The next theorem is the main result about **OMP**.

Theorem 4.2.3. *Let the objective function E satisfy* **Conditions 0,1** *and* **2** *with $M_0 < \alpha q M^{q-1}$. Let problem (1.1) have a solution $\bar{x} = \sum_i c_i(\bar{x})\varphi_i \in \Omega$ with support $\bar{S} := \{i : c_i(\bar{x}) \neq 0\} < \infty$, where $\{\varphi_i\}$ is an orthonormal basis. Then, at Step k, the* **OMP** *applied to E and $\{\varphi_i\}$ outputs x_k, where either $x_k = \bar{x}$ or:*

- *When $p \neq q \neq 2$,*

$$e_k \leq C_0 \left(\frac{|\bar{S}|^{\frac{q}{2(q-1)}}}{k + C_1 |\bar{S}|^{\frac{q}{2(q-1)}} - 1} \right)^{\frac{p(q-1)}{p-q}}, \quad k \geq 2.$$

  In addition, the sequence $\{x_k\}_{k=2}^{\infty}$ satisfies

$$\|x_k - \bar{x}\| \leq \left(\frac{C_0}{\beta} \right)^{\frac{1}{p}} \left(\frac{|\bar{S}|^{\frac{q}{2(q-1)}}}{k + C_1 |\bar{S}|^{\frac{q}{2(q-1)}} - 1} \right)^{\frac{(q-1)}{p-q}},$$

  where $C_0 = C_0(p, q, \alpha, \beta)$ and $C_1 = C_1(p, q, \alpha, \beta, E)$.

- *When $p = q = 2$, we have*

$$e_k \leq C_2 \left(1 - \frac{\beta}{\alpha |\bar{S}|} \right)^{k-1}, \quad k \geq 2,$$

$$\|x_k - \bar{x}\| \leq \left(\frac{C_2}{\beta} \right)^{\frac{1}{2}} \left(1 - \frac{\beta}{\alpha |\bar{S}|} \right)^{\frac{k-1}{2}}, \quad k \geq 2,$$

  with $C_2 = C_2(E)$.

Proof. In the case $p \neq q \neq 2$, we define the sequence of non-negative numbers

$$r_k = 1, \quad a_k = E(x_k) - E(\bar{x}), \quad k = 1, 2, \ldots,$$

and

$$\ell = \frac{p-q}{p(q-1)} > 0, \quad r = \frac{(p-1)^{\frac{q(p-1)}{p(q-1)}}}{(q-1)} \left(\frac{\alpha q^q}{\beta^{q/p} p^q} \right)^{\frac{1}{q-1}} |\bar{S}|^{\frac{q}{2(q-1)}} > 0, \quad B = E(0) - E(\bar{x}).$$

It follows from Lemma 4.2.1 that the above defined sequences satisfy the conditions of Lemma 4.1.5, and therefore we have

$$E(x_k) - E(\bar{x}) \leq C_0 \left(\frac{|\bar{S}|^{\frac{q}{2(q-1)}}}{k + C_1 |\bar{S}|^{\frac{q}{2(q-1)}} - 1} \right)^{\frac{p(q-1)}{p-q}},$$

where

$$C_0 = C_0(p,q,\alpha,\beta) = \frac{(p-1)^{\frac{q(p-1)}{p-q}}}{(q-1)^{\frac{p(q-1)}{p-q}}} \left(\frac{\alpha q^q}{\beta^{q/p} p^q} \right)^{\frac{p}{p-q}} \cdot \max \left\{ 1, \left(\frac{p(q-1)}{p-q} \right)^{\frac{p(q-1)}{p-q}} \right\},$$

$$C_1 = C_1(p,q,\alpha,\beta,E) = \frac{(p-1)^{\frac{q(p-1)}{p(q-1)}}}{(q-1)} \left(\frac{\alpha q^q}{\beta^{q/p} p^q} \right)^{\frac{1}{q-1}} \cdot (E(0) - E(\bar{x}))^{\frac{q-p}{p(q-1)}},$$

Next, we apply **Condition 2** with $x' = x_k$ and $x = \bar{x}$ and use Lemma 4.1.1 to derive the estimate for $\|x_k - \bar{x}\|$.

In the case $p = q = 2$, Lemma 4.2.1 gives that $E(x_1) - E(\bar{x}) \leq E(0) - E(\bar{x})$ and

$$E(x_k) - E(\bar{x}) \leq \left(1 - \frac{\beta}{\alpha |\bar{S}|} \right) (E(x_{k-1}) - E(\bar{x})), \quad k = 2, 3, \ldots.$$

It follows then that

$$E(x_k) - E(\bar{x}) \leq (E(0) - E(\bar{x})) \left(1 - \frac{\beta}{\alpha |\bar{S}|} \right)^{k-1}, \quad k = 2, 3, \ldots.$$

As in the previous case, we use **Condition 2** with $x' = x_k$ and $x = \bar{x}$ and Lemma 4.1.1 to derive the estimate for $\|x_k - \bar{x}\|$. □

Remark 4.2.4. *The constants C_0, C_1 and C_3 from Theorem 4.2.3 can be explicitly found. They are*

$$C_0 = C_0(p,q,\alpha,\beta) = \frac{(p-1)^{\frac{q(p-1)}{p-q}}}{(q-1)^{\frac{p(q-1)}{p-q}}} \left(\frac{\alpha q^q}{\beta^{q/p} p^q}\right)^{\frac{p}{p-q}} \cdot \max\left\{1, \left(\frac{p(q-1)}{p-q}\right)^{\frac{p(q-1)}{p-q}}\right\},$$

$$C_1 = C_1(p,q,\alpha,\beta,E) = \frac{(p-1)^{\frac{q(p-1)}{p(q-1)}}}{(q-1)} \left(\frac{\alpha q^q}{\beta^{q/p} p^q}\right)^{\frac{1}{q-1}} \cdot (E(0)-E(\bar{x}))^{\frac{q-p}{p(q-1)}},$$

and

$$C_2 = C_2(E) := E(0) - E(\bar{x}).$$

Remark 4.2.5. *The same statement as Theorem 4.2.3 holds if instead of* **Conditions 0,1,2** *we have that E satisfies* **Conditions** $0', 1', 2'$.

Next, we show that with our analysis we can recover the convergence rate for **OMP** from [15]. In what follows, $\mathcal{D}$ is a symmetric dictionary and

$$\|\bar{x}\|_1 := \inf\{\sum_{\varphi} |c_\varphi(\bar{x})| : \ \bar{x} = \sum_{\varphi} c_\varphi(\bar{x})\varphi\}.$$

We prove the following lemma.

Lemma 4.2.6. *Let the objective function E satisfy* **Condition 0** *and* **Condition 1** *with $M_0 < \alpha q M^{q-1}$. Let problem* (1.1) *have a solution $\bar{x} = \sum_\varphi c_\varphi(\bar{x})\varphi \in \Omega$, where $\{\varphi\} = \mathcal{D}$ is a symmetric dictionary. Then* **OMP** *applied to E and $\mathcal{D}$ satisfies the inequality*

$$e_1 \le E(0) - E(\bar{x}), \tag{4.14}$$

and

$$e_k \le e_{k-1} - D e_{k-1}^{\frac{q}{q-1}}, \quad k \ge 2, \quad D := \frac{q-1}{q} (\alpha q)^{-\frac{1}{q-1}} \|\bar{x}\|_1^{-q/(q-1)}. \tag{4.15}$$

Proof. As in Lemma 4.2.1, one shows that

$$E(x_k) \leq E(x_{k-1}) - \frac{q-1}{q} (\alpha q)^{-\frac{1}{q-1}} |\langle E'(x_{k-1}), \varphi_{j_k} \rangle|^{q/(q-1)}. \tag{4.16}$$

Now, instead of (4.11), we use Lemma 2.1.1 from Chapter II to derive

$$\langle E'(x_{k-1}), x_{k-1} - \bar{x} \rangle \geq E(x_{k-1}) - E(\bar{x}). \tag{4.17}$$

Lemma 4.2.3 gives that $\langle E'(x_{k-1}), \varphi_i \rangle = 0$, $i = j_1, \dots, j_{k-1}$. For any $\epsilon > 0$, we choose a representation

$$\bar{x} = \sum_{\varphi} c_{\varphi}^{\epsilon}(\bar{x}) \varphi, \quad \text{with} \ \sum_{\varphi} |c_{\varphi}^{\epsilon}(\bar{x})| > \epsilon + \|\bar{x}\|_1.$$

Since the support of x_{k-1} is $\{j_1, \dots, j_{k-1}\}$, we obtain

$$\langle E'(x_{k-1}), x_{k-1} - \bar{x} \rangle = - \sum_{\varphi \neq \varphi_{j_1}, \dots, \varphi_{j_{k-1}}} c_{\varphi}^{\epsilon}(\bar{x}) \langle E'(x_{k-1}), \varphi \rangle \leq |\langle E'(x_{k-1}), \varphi_{j_k} \rangle| (\|\bar{x}\|_1 + \epsilon).$$

Taking $\epsilon \to 0$, we get

$$\langle E'(x_{k-1}), x_{k-1} - \bar{x} \rangle \leq |\langle E'(x_{k-1}), \varphi_{j_k} \rangle| \|\bar{x}\|_1. \tag{4.18}$$

Next, we substitute (4.18) in (4.17) and (4.16) to derive that

$$E(x_k) \leq E(x_{k-1}) - \frac{q-1}{q} (\alpha q)^{-\frac{1}{q-1}} \|\bar{x}\|_1^{-q/(q-1)} (E(x_{k-1}) - E(\bar{x}))^{q/(q-1)}.$$

We subtract $E(\bar{x})$ from both sides of this inequality and the proof is completed. □

The next theorem is the same as Theorem 2.2 from [15], see Theorem 3.2.2 from

Theorem 4.2.7. *Let the objective function E satisfy* **Condition 0** *and* **Condition 1** *with $M_0 < \alpha q M^{q-1}$. Let problem (1.1) have a solution $\bar{x} = \sum_{\varphi} c_{\varphi}(\bar{x})\varphi$, where $\{\varphi\} = \mathcal{D}$ is a symmetric dictionary. Then, at Step k, the* **OMP** *applied to E and $\mathcal{D}$ outputs x_k, where*

$$e_k = E(x_k) - E(\bar{x}) \leq C_0(C_1 + k)^{1-q}, \quad C_0 = C_0(q, \alpha, \bar{x}), \quad C_1 = C_1(q, \alpha, \bar{x}).$$

Proof. We define the sequence of non-negative numbers

$$r_k = 1, \quad a_k = E(x_k) - E(\bar{x}), \quad k = 1, 2, \dots,$$

and

$$\ell = \frac{1}{q-1} > 0, \quad r = D^{-1} > 0, \quad B = E(0) - E(\bar{x}).$$

It follows from Lemma 4.2.6 that the above defined sequences satisfy the conditions of Lemma 4.1.5, and therefore we have

$$e_k = E(x_k) - E(\bar{x}) \leq C_0 \left(k + C_1\right)^{1-q},$$

where

$$C_0 = \alpha(q-1)^{1-q} q^q \|\bar{x}\|_1^q, \quad C_1 = (q-1)^{-1} \left(\frac{\alpha q^q \|\bar{x}\|_1^q}{E(0) - E(\bar{x})} \right)^{\frac{1}{q-1}} - 1,$$

□

4.3 Convergence Results for **WCGA**

In this section we analyze the convergence of the **WCGA** when applied to problem (1.1) in the case of $\mathcal{D}$ being an orthonormal basis. Let

$$e_k^{\mathrm{w}} := E(x_k^{\mathrm{w}}) - E(\bar{x})$$

be the error of the algorithm. Similar results as Lemma 4.2.1 hold here.

Lemma 4.3.1. *Let the objective function E satisfy* **Conditions 0,1**, *and* **2** *with $M_0 < \alpha q M^{q-1}$. Let problem* (1.1) *have a solution $\bar{x} = \sum_i c_i(\bar{x})\varphi_i \in \Omega$ with support*

$$\bar{S} := \{i : c_i(\bar{x}) \neq 0\} < \infty,$$

where $\{\varphi_i\}$ is an orthonormal basis. Then, at Step k, the **WCGA** *applied to E and $\{\varphi_i\}$ outputs x_k^{w}, where either $x_k^{\mathrm{w}} = \bar{x}$ or*

$$e_1^{\mathrm{w}} \leq E(0) - E(\bar{x}), \tag{4.19}$$

$$e_k^{\mathrm{w}} \leq e_{k-1}^{\mathrm{w}} - \frac{t_k^{\frac{q}{q-1}}}{A|\bar{S}|^{\frac{q}{2(q-1)}}} \left[e_{k-1}^{\mathrm{w}}\right]^{\frac{(p-1)q}{(q-1)p}}, \quad k = 2, 3, \ldots, \tag{4.20}$$

where

$$A = \frac{(p-1)^{\frac{q(p-1)}{p(q-1)}}}{(q-1)} \left(\frac{\alpha q^q}{\beta^{q/p} p^q}\right)^{\frac{1}{q-1}}.$$

Proof. The proof follows the lines of the one of Lemma 4.2.1. The derivation of (4.19) is the same as in Lemma 4.2.1.

Next, we consider Step k, $k = 2, 3, \ldots$. As in Lemma 4.2.1, if at Step (k-1) we have that $\bar{S} \subseteq \{j_1, \ldots, j_{k-1}\}$, then $x_{k-1}^{\mathrm{w}} = \bar{x}$, $E'(x_{k-1}^{\mathrm{w}}) = 0$ and the **WCGA** would have stopped with output $x_{k-1}^{\mathrm{w}} = \bar{x}$. If the algorithm have not stopped, then we have found x_k^{w} and φ_{j_k} and obtained, as in Lemma 4.2.1, the estimates

$$E(x_k^{\mathrm{w}}) \leq \Phi(t^*) = E(x_{k-1}^{\mathrm{w}}) - \frac{q-1}{q}\,(\alpha q)^{-\frac{1}{q-1}}\,|\langle E'(x_{k-1}^{\mathrm{w}}), \varphi_{j_k}\rangle|^{q/(q-1)}. \tag{4.21}$$

and

$$\begin{aligned} \langle E'(x_{k-1}^{\mathrm{w}}), x_{k-1}^{\mathrm{w}} - \bar{x}\rangle &\geq E(x_{k-1}^{\mathrm{w}}) - E(\bar{x}) + \beta\|\bar{x} - x_{k-1}^{\mathrm{w}}\|^p \\ &\geq C|\bar{S}|^{-1/2}\left(E(x_{k-1}^{\mathrm{w}}) - E(\bar{x})\right)^{(p-1)/p}, \end{aligned} \tag{4.22}$$

with $C = p\beta^{1/p}(p-1)^{(1-p)/p}$. Like in Lemma 4.2.1, we have

$$\begin{aligned} \langle E'(x_{k-1}^{\mathrm{w}}), x_{k-1}^{\mathrm{w}} - \bar{x}\rangle &= \Sigma_{i\in\bar{S}\setminus j_1,\ldots,j_{k-1}} c_i(x_{k-1}^{\mathrm{w}} - \bar{x})\langle E'(x_{k-1}^{\mathrm{w}}), \varphi_i\rangle \\ &\leq \Sigma_{i\in\bar{S}\setminus j_1,\ldots,j_{k-1}} |c_i(x_{k-1}^{\mathrm{w}} - \bar{x})||\langle E'(x_{k-1}^{\mathrm{w}}), \varphi_i\rangle| \\ &\leq t_k^{-1}|\langle E'(x_{k-1}^{\mathrm{w}}), \varphi_{j_k}\rangle|\Sigma_{i\in\bar{S}}|c_i(x_{k-1}^{\mathrm{w}} - \bar{x})| \\ &\leq t_k^{-1}|\langle E'(x_{k-1}^{\mathrm{w}}), \varphi_{j_k}\rangle|\|\bar{x} - x_{k-1}^{\mathrm{w}}\||\bar{S}|^{1/2}, \end{aligned} \tag{4.23}$$

where in the second inequality we have used the choice of φ_{j_k}.

Using together (4.22) and (4.23) results in

$$t_k^{-1}|\langle E'(x_{k-1}^{\mathrm{w}}), \varphi_{j_k}\rangle|\|\bar{x} - x_{k-1}^{\mathrm{w}}\||\bar{S}|^{1/2} \geq C|\bar{S}|^{-1/2}\left(E(x_{k-1}^{\mathrm{w}}) - E(\bar{x})\right)^{(p-1)/p},$$

and therefore

$$|\langle E'(x_{k-1}^{\mathrm{w}}), \varphi_{j_k}\rangle| \geq C t_k|\bar{S}|^{-1/2}\left(E(x_{k-1}^{\mathrm{w}}) - E(\bar{x})\right)^{(p-1)/p}.$$

We combine this result with (4.21) to obtain the desired estimate

$$E(x_k^{\mathrm{w}}) \leq E(x_{k-1}^{\mathrm{w}}) - \frac{t_k^{\frac{q}{q-1}}}{A|\bar{S}|^{\frac{q}{2(q-1)}}} \left(E(x_{k-1}^{\mathrm{w}}) - E(\bar{x})\right)^{\frac{(p-1)q}{(q-1)p}},$$

with constant $A = \frac{(p-1)^{\frac{q(p-1)}{p(q-1)}}}{(q-1)} \left(\frac{\alpha q^q}{\beta^{q/p} p^q}\right)^{\frac{1}{q-1}}$.

Subtracting $E(\bar{x})$ from both sides of the inequality results in (4.7) and the proof is completed. □

Remark 4.3.2. *Note that in the proof of Lemma 4.2.1* **Condition 1** *needs to hold only for vectors x and x' such that $(x - x')$ are 1-sparse.*

The next theorem is the main result about the **WCGA** algorithm.

Theorem 4.3.3. *Let the objective function E satisfy* **Conditions 0,1,2** *with $M_0 < \alpha q M^{q-1}$. Let problem* (1.1) *have a solution $\bar{x} = \sum_i c_i(\bar{x})\varphi_i \in \Omega$ with support $\bar{S} := \{i : c_i(\bar{x}) \neq 0\} < \infty$, where $\{\varphi_i\}$ is an orthonormal basis. Then, at Step k, the* **WCGA** *applied to E and $\{\varphi_i\}$ outputs x_k^{w}, where either $x_k^{\mathrm{w}} = \bar{x}$ or*

- *When $p \neq q \neq 2$,*

$$e_k^{\mathrm{w}} \leq C_0 \left(\frac{|\bar{S}|^{\frac{q}{2(q-1)}}}{\sum_{i=2}^k t_i^{\frac{q}{q-1}} + C_1|\bar{S}|^{\frac{q}{2(q-1)}}}\right)^{\frac{p(q-1)}{p-q}}, \quad k \geq 2.$$

In addition, the sequence $\{x_k^{\mathrm{w}}\}_{k=2}^\infty$ satisfies

$$\|x_k^{\mathrm{w}} - \bar{x}\| \leq \left(\frac{C_0}{\beta}\right)^{\frac{1}{p}} \left(\frac{|\bar{S}|^{\frac{q}{2(q-1)}}}{\sum_{i=2}^k t_i^{\frac{q}{q-1}} + C_1|\bar{S}|^{\frac{q}{2(q-1)}}}\right)^{\frac{(q-1)}{p-q}}, \quad k \geq 2,$$

where $C_0 = C_0(p, q, \alpha, \beta)$ and $C_1 = C_1(p, q, \alpha, \beta, E)$.

- *When* $p = q = 2$,
$$e_k^{\mathrm{w}} \le C_2 \prod_{i=2}^{k} \left(1 - \frac{\beta}{\alpha|\bar{S}|} t_i^2\right),$$
$$\|x_k^{\mathrm{w}} - \bar{x}\| \le \left(\frac{C_2}{\beta}\right)^{\frac{1}{2}} \prod_{i=2}^{k} \left(1 - \frac{\beta}{\alpha|\bar{S}|} t_i^2\right)^{1/2}, \quad k = 2, 3, \dots,$$
with $C_2 = C_2(E)$.

Proof. Like in the proof of Theorem 4.2.3, we apply Lemma 4.1.5 and Lemma 4.2.1 to the sequences of non-negative numbers

$$r_k = t_k^{\frac{q}{q-1}}, \quad a_k = E(x_k^{\mathrm{w}}) - E(\bar{x}), \quad k = 1, 2, \dots,$$

and

$$\ell = \frac{p-q}{p(q-1)} > 0, \quad r = \frac{(p-1)^{\frac{q(p-1)}{p(q-1)}}}{(q-1)} \left(\frac{\alpha q^q}{\beta^{q/p} p^q}\right)^{\frac{1}{q-1}} |\bar{S}|^{\frac{q}{2(q-1)}} > 0, \quad B = E(0) - E(\bar{x}).$$

It follows from Lemma 4.2.1 that the above defined sequences satisfy the conditions of Lemma 4.1.5, and therefore we have

$$E(x_k^{\mathrm{w}}) - E(\bar{x}) \le C_0 \left(\frac{|\bar{S}|^{\frac{q}{2(p-1)}}}{\sum_{k=i}^{k} t_i^{\frac{q}{q-1}} + C_1 |\bar{S}|^{\frac{q}{2(p-1)}}}\right)^{\frac{p(q-1)}{p-q}},$$

where

$$C_0 = C_0(p, q, \alpha, \beta) = \frac{(p-1)^{\frac{q(p-1)}{p-q}}}{(q-1)^{\frac{p(q-1)}{p-q}}} \left(\frac{\alpha q^q}{\beta^{q/p} p^q}\right)^{\frac{p}{p-q}} \cdot \max\left\{1, \left(\frac{p(q-1)}{p-q}\right)^{\frac{p(q-1)}{p-q}}\right\},$$

and

$$C_1 = C_1(p, q, \alpha, \beta, E) = \frac{(p-1)^{\frac{q(p-1)}{p(q-1)}}}{(q-1)} \left(\frac{\alpha q^q}{\beta^{q/p} p^q}\right)^{\frac{1}{q-1}} \cdot (E(0) - E(\bar{x}))^{\frac{q-p}{p(q-1)}},$$

Next, we apply **Condition 2** with $x' = x_k^{\rm w}$ and $x = \bar{x}$ and use Lemma 4.1.1 to derive the estimate for $\|x_k^{\rm w} - \bar{x}\|$.

In the case $p = q = 2$, Lemma 4.3.1 gives that $E(x_1^{\rm w}) - E(\bar{x}) \le E(0) - E(\bar{x})$ and

$$E(x_k^{\rm w}) - E(\bar{x}) \le \left(1 - \frac{\beta}{\alpha|\bar{S}|} t_k^2\right) (E(x_{k-1}^{\rm w}) - E(\bar{x})), \quad k = 2, 3, \dots.$$

It follows then that

$$E(x_k^{\rm w}) - E(\bar{x}) \le (E(0) - E(\bar{x})) \prod_{i=2}^{k} \left(1 - \frac{\beta}{\alpha|\bar{S}|} t_i^2\right), \quad k = 2, 3, \dots.$$

As in the previous case, we use **Condition 2** with $x' = x_k^{\rm w}$ and $x = \bar{x}$ and Lemma 4.1.1 to derive the estimate for $\|x_k^{\rm w} - \bar{x}\|$. □

Remark 4.3.4. *The constants C_0, C_1, C_2 and C_3 can be computed explicitly. They are*

$$C_0 = C_0(p, q, \alpha, \beta) = \frac{(p-1)^{\frac{q(p-1)}{p-q}}}{(q-1)^{\frac{p(q-1)}{p-q}}} \left(\frac{\alpha q^q}{\beta^{q/p} p^q}\right)^{\frac{p}{p-q}} \cdot \max\left\{1, \left(\frac{p(q-1)}{p-q}\right)^{\frac{p(q-1)}{p-q}}\right\},$$

$$C_1 = C_1(p, q, \alpha, \beta, E) = \frac{(p-1)^{\frac{q(p-1)}{p(q-1)}}}{(q-1)} \left(\frac{\alpha q^q}{\beta^{q/p} p^q}\right)^{\frac{1}{q-1}} \cdot (E(0) - E(\bar{x}))^{\frac{q-p}{p(q-1)}},$$

and

$$C_2 = C_2(E) := E(0) - E(\bar{x}).$$

Remark 4.3.5. *The same theorem as Theorem 4.3.3 holds if instead of* **Conditions 0,1,2** *we impose* **Conditions** $0', 1', 2'$ *on E.*

The next theorem is the weak version of Theorem 4.2.6 and it is the same as Theorem 2.2 from [16].

Theorem 4.3.6. *Let the objective function E satisfy* **Condition 0** *and* **Condition 1** *with $M_0 < \alpha q M^{q-1}$. Let problem (1.1) have a solution $\bar{x} = \sum_{\varphi} c_{\varphi}(\bar{x})\varphi$, where $\{\varphi\} = \mathcal{D}$ is a symmetric dictionary. Then, at Step k, the* **WCGA** *applied to E and $\mathcal{D}$ outputs x_k^{w}, where*

$$e_k^{\mathrm{w}} = E(x_k^{\mathrm{w}}) - E(\bar{x}) \leq C_0(C_1 + \Sigma_{i=2}^k t_i^{\frac{q}{q-1}})^{1-q}, \quad C_0 = C_0(q, \alpha, \bar{x}), \quad C_1 = C_1(q, \alpha, \bar{x}, E).$$

Proof. We define the sequence of non-negative numbers

$$r_k = t_k^{\frac{q}{q-1}}, \quad a_k = E(x_k^{\mathrm{w}}) - E(\bar{x}), \quad k = 1, 2, \dots,$$

and

$$\ell = \frac{1}{q-1}, \quad r = D^{-1}, \quad B = E(0) - E(\bar{x}).$$

It follows from Lemma 4.2.6 that the above defined sequences satisfy the conditions of Lemma 4.1.5, and therefore we have

$$e_k^{\mathrm{w}} = E(x_k^{\mathrm{w}}) - E(\bar{x}) \leq C_0 \left(\Sigma_{i=2}^k t_i^{\frac{q}{q-1}} + C_1 \right)^{1-q},$$

where

$$C_0 = \alpha(q-1)^{1-q} q^q \|\bar{x}\|_1^q, \quad C_1 = (q-1)^{-1} \left(\frac{\alpha q^q \|\bar{x}\|_1^q}{E(0) - E(\bar{x})} \right)^{\frac{1}{q-1}}.$$

□

5. SUMMARY

In this dissertation, we have proved that if a convex function E satisfies certain assumption on its modulus of smoothness and modulus of uniform convexity and its minimizer is sparse with respect to an orthonormall basis in a Hilbert space H, then the orthogonal matching pursuit and the weak Chebysev greedy algorithm have rates of convergence depending on the order of smoothness and the order of convexity.

In our analysis, we have used the fact that the minimizer $\bar{x}$ is sparse with respect to an orthonormal basis $\mathcal{D}$ for the Hilbert space H. My future research plans are to investigate the convergence rates of two other greedy algorithms: the Weak Relaxed Greedy Algorithm (**WRGA**) and the Weak Greedy Algorithm with Free Relaxation (**WGAFR**). These greedy algorithms (see [15, 16]) have less computational cost and are described below.

Weak Relaxed Greedy Algorithm (WCGA):

- **Step** 0:

  Start with initial $x_0 = 0$.

  If $E'(x_0) = 0$, $x_m := x_0 = 0$. Go to **Step** $m + 1$.

- **Step** $k, k = 1, 2, \ldots m$:

  Find φ_{j_k} such that $|\langle E'(x_{k-1}), \varphi_{j_k}\rangle| \geq t_k \sup_{\varphi\in\mathcal{D}} |\langle E'(x_{k-1}), \varphi\rangle|$.

  Find $\lambda_k := \text{argmin}_{\lambda\in[0,1]}\{E((1-\lambda)x_{k-1} + \lambda\varphi_{j_k})\}$.

  Set $x_k := (1-\lambda_k)x_{k-1} + \lambda_k\varphi_{j_k}$.

  If $E'(x_k)$=0, $x_m := x_k$. Go to **Step** $m + 1$.

- **Step** $m + 1$:

Output x_m.

Weak Greedy Algorithm with Free Relaxation (WGAFR):

- **Step** 0:

  Start with initial $x_0 = 0$.

  If $E'(x_0) = 0$, $x_m := x_0$. Go to **Step** $m + 1$.

- **Step** $k, k = 1, 2, \ldots m$:

  Find φ_{j_k} such that $|\langle E'(x_{k-1}), \varphi_{j_k}\rangle| \geq t_k| \sup_{\varphi \in \mathcal{D}} \langle E'(x_{k-1}), \varphi\rangle|$.

  Find $(\lambda_k, \mu_k) := \text{argmin}_{\{\lambda,\mu\}}\{E(\lambda x_{k-1} + \mu\varphi_k)\}$.

  Set $x_k := \lambda_k x_{k-1} + \mu_k\varphi_k$.

  If $E'(x_k)$=0, $x_m := x_k$. Go to **Step** $m + 1$.

- **Step** $m + 1$:

  Output x_m.

REFERENCES

[1] A. Barron, A. Cohen, W. Dahmen, and R. DeVore, "Approximation and learning by greedy algorithms", *The Annals of Statistics*, vol. 36, iss. 1, pp. 64-94, 2008.

[2] R. Berinde, A. Gilbert, P. Indyk, H. Karloff, and M. Strauss, "Combining geometry and combinatorics: A unified approach to sparse signal recovery“, *Communication, Comtrol, and Computing, 46th Annual Conference*, Allerton, September 2008.

[3] J. Borwein, A. Guiro, P. Hajek, and J. Vanderwerff, "Uniformly convex functions on Banach spaces", *Proc. Amer. Math. Soc.* vol. 137, pp. 1081-1091, 2009.

[4] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, New York, NY, 2004.

[5] E. Candes and T. Tao, "Decoding by linear programming", *IEEE Trans. Inf. Theory*, vol. 51, iss. 12, pp. 4203-4215, 2005.

[6] E. Candes, "Compressive Sampling", *Int. Congress of Mathematics*, vol. 3, pp. 1433-1452, Madrid, Spain, 2006.

[7] A. Cohen, W. Dahmen, and R. DeVore, "Compressed sensing and best k-terms approximation", *J. Amer. Math. Soc.* vol. 22, pp. 211-231, 2009.

[8] R. DeVore and G. Lorentz, *Constructive Approximation*, Springer-Verlag, Berlin-Heidelberg, 1993.

[9] R. DeVore and V. Temlyakov, "Some remarks on greedy algorithms", *Advances in Computational Mathematics*, vol. 5, pp. 173-187, 1996.

[10] D. Donoho, "Compressed sensing", *IEEE. Trans. Inf. Theory*, vol. 52, iss. 4, pp. 1289-1306, 2006.

[11] S. Foucart, “Hard Thresholding Pursuit: An algorithm for compressive sensing”, *SIAM Journal on Numerical Analysis*, vol. 49, iss. 6, pp. 2543-2563, 2011.

[12] S. Mallat and Z. Zhang, “Matching Pursuit in a time-frequency dictionary”, *IEEE Trans. Signal Proc*, vol. 41, pp. 3397-3415, 1993.

[13] R. Phelps, *Convex Functions, Monotone Operators and Differentiability*, 2nd edition, Lecture Notes in Mathematics , Springer Verlag, Berlin-Heidelberg, 1993.

[14] V. Temlyakov, “Weak greedy algorithms”, *Avances in Computational Mathematics*, vol. 12, iss. 2, pp. 213-227, 2011.

[15] V. Temlyakov, “Greedy expansions in convex optimization”, *arXiv*: 1206.0393, 2012.

[16] V. Temlyakov, “Greedy approximation in convex optimization”, *arXiv*: 1206.0392, 2012.

[17] V. Temlyakov, *Greedy Approximation*, Cambridge monographs on Applied and Computational Mathematics, Cambridge University Press, New York, NY 2011.

[18] M. Yang, “ Extension of convex functions”, *arXiv*: 1207.0944, 2013.

[19] C. Zalinescu, *Convex Analysis in General Vector Spaces*, World Scientific Publishing Co. Inc., River Edge, NJ, 2002 .

[20] T. Zhang, “Sparse recovery with orthogonal matching pursuit under RIP”, *IEEE Transaction Information Theory* , vol. 57, iss. 9, pp. 6215-6221, 2010.