GREEDY STRATEGIES FOR CONVEX OPTIMIZATION

HAO NGUYEN AND GUERGANA PETROVA

Abstract. We investigate two greedy strategies for finding an approximation to the minimum of a convex function E defined on a Hilbert space H. We prove convergence rates for these algorithms under suitable conditions on the objective function E. These conditions involve the behavior of the modulus of smoothness and the modulus of uniform convexity of E.
Key Words: Greedy Algorithms, Convex Optimization, Rates of Convergence.

1. Introduction

Convex optimization has many application domains such as automatic control systems, signal processing, communications and networks, electronic circuit design, data analysis and modeling, statistical estimation, finance, and combinatorial optimization. A general description for convex optimization is that we are given a Banach space X and a convex function E on X whose minimum we wish to compute. Thus, we are interested in the development and analysis of algorithms for approximating

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

where D is a convex subset of X. E is called the *objective* function and, by the convexity assumption, satisfies the condition

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

The classical results on convex optimization deal with objective functions E defined on subsets in $\mathbb{R}^d$ with moderate values of d, see e.g. [2]. However, several of the applications listed above lead to optimization on Banach spaces of dimension d, where d is quite large or even ∞. The design of algorithms for such high dimensional problems is quite challenging, typical convergent results involve the dimension d and suffer from the curse of dimensionality.

Recently, several researchers (see e.g. [10, 11, 16, 12, 5]) have proposed strategies for solving (1.1), where the curse of dimensionality is overcome by using greedy techniques, similar to those originally developed for the approximation of a given

This research was supported by the Office of Naval Research Contract ONR N00014-11-1-0712, by the NSF Grant DMS 1521067, and by the Bulgarian Science Fund Grant DFNI-T01/0001.

element $x \in X$. The minimum in (1.1) is approximated by $E(x_m)$, $m = 0, 1, \dots$, where each x_m is constructed as a linear combination of m elements (i.e. x_m is m *sparse*) from a given dictionary $\mathcal{D}$. Recall that $\mathcal{D}$ is called a symmetric dictionary if each $\varphi \in \mathcal{D}$ has norm $\|\varphi\| = 1$, if $\varphi \in \mathcal{D}$, then $-\varphi \in \mathcal{D}$, and the closure of $span\,\mathcal{D}$ is X. Some of the a priori convergence results given by the above authors for these greedy algorithms are proven under two assumptions:

(i) An assumption on the smoothness of E.

(ii) An assumption that the minimum in (1.1) is taken at a point $\bar{x}$ which is in the convex hull of the dictionary $\mathcal{D}$.

In this paper, we investigate the special case when $X = H$ is a Hilbert space, the dictionary $\mathcal{D}$ is an orthonormal basis $\{\varphi_j\}_{j=1}^{\infty}$ for H, and $D = H$ (which corresponds to global minimization). We assume that the global minimum is attained at some point $\bar{x} \in H$. It follows then that the minimum is taken on the set

$$\Omega := \{x \in H : \; E(x) \leq E(0)\}.$$

We assume throughout this paper that the set Ω is bounded in H. We impose the following assumptions on the objective function E:

Condition 0: E has a Frechet derivative $E'(x) \in H$ at each point x in Ω and

$$\|E'(x)\| \leq M_0, \quad x \in \Omega,$$

where throughout $\|\cdot\|$ denotes the norm on H.

Uniform Smoothness (US): There are constants $\alpha > 0$, $1 < q \leq 2$, and $M > 0$, such that for all x, x' with $\|x - x'\| \leq M$, $x \in \Omega$,

$$E(x') - E(x) - \langle E'(x), x' - x \rangle \leq \alpha \|x' - x\|^q. \tag{1.2}$$

Uniform Convexity (UC): There are constants $\beta > 0$, $2 \leq p < \infty$, and $M > 0$, such that for all x, x' with $\|x - x'\| \leq M$, $x \in \Omega$,

$$E(x') - E(x) - \langle E'(x), x' - x \rangle \geq \beta \|x' - x\|^p. \tag{1.3}$$

The **US** condition has already been proposed in [14] and the **UC** condition has been discussed in [8]. We show in §2 that the **US** condition is equivalent to a condition on the modulus of smoothness of E, and the **UC** condition is equivalent to a condition on the modulus of uniform convexity of E, as usually defined in convex optimization (see e.g. [15]), and introduced by us in §2.

We study two greedy procedures for solving (1.1). The first is a convex minimization analogue of the Orthogonal Matching Pursuit Algorithm used for approximation (see [7]). We denote this convex minimization algorithm by OMP(co)[1]. The second is the Weak Orthogonal Matching Pursuit (WOMP(co)), as introduced by Temlyakov (see [11]) in the case of general Banach spaces and named there as the Weak Chebyshev Greedy Algorithm. These greedy procedures, which are defined in §3, iteratively generate a sequence x_m, $m = 0, 1, \ldots$, where each x_m is m sparse, and then use $E(x_m)$ as the approximation to the minimum $E(\bar{x})$.

Our main results are Theorem 4.5 and Theorem 4.6 which establish a priori convergence rates for both OMP(co) and the WOMP(co) when they are used to find the minimum of a function E that satisfies **Condition 0**, the **US** and the **UC** condition. They give a crisp and general analysis of the discussed algorithms under generalized smoothness and uniform convexity assumptions for the objective function. For example, we show that if the objective function E satisfies **Condition 0** and the **US** condition, is strongly convex on H (therefore satisfies the **UC** condition with $p = 2$), and its minimizer $\bar{x}$ is sparse with respect to $\mathcal{D}$, then the error at the m-th step of the OMP(co) satisfies the inequality

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

and

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

where $C_0 = C_0(q, E)$ and $C_1 = C_1(q, E)$. In contrast, the results from [11] and [16] do not impose the **UC** condition and only give the rate $1 - q$. The above theorems generalize also the results from [9] (see Theorem 2.8), where convergence rates for the OMP(co) are proved in the case $p = q = 2$ and $H = \mathbb{R}^d$. In the case $p = q = 2$, our results can be derived from [12] (see Theorem 1.2), where the convergence of the WOMP(co) is discussed in the case of Banach spaces and special dictionaries. In summary, we show that imposing more conditions on the convexity of the objective function E (like the **UC** condition) results in provably improved convergence rates for both OMP(co) and WOMP(co).

2. Conditions on E

In this section, we discuss the compatibility of **Condition 0** , the **US** and the **UC** conditions, imposed on the objective function E, and their relation to the modulus of smoothness and modulus of uniform convexity of E. We recall that a function E is Frechet differentiable at $x \in \Omega$ if there exists a bounded linear functional, denoted

[1]Here and later we will use the abbreviation (co) if an algorithm is used for convex optimization

by $E'(x)$, such that

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

We start with discussing the connection between the **US** condition and the modulus of uniform smoothness of E on Ω.

2.1. **The uniform smoothness condition.**

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

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

and the modulus of uniform smoothness of E on S is defined by

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

These two moduli of smoothness are equivalent (see [15], page 205), as the following lemma states.

Lemma 2.1. *Let E be a convex function defined on H, and let $S\subset H$, then*

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

The next lemma shows the relation between the modulus of uniform smoothness and the **US** condition.

Lemma 2.2. *Let E be a convex function defined on a Hilbert space H and E be Frechet differentiable on a set $S\subset H$. The following statements are equivalent for any $q\in(1,2]$ and $M>0$.*

(i) *There exists $\alpha>0$, such that for all $x\in S, x'\in H, \|x-x'\|\le M$,*

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

(ii) *There exists $\alpha_1>0$, such that*

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

The same result holds with ρ replaced by ρ_1.

Proof. While this is a particular case of Corollary 3.5.7 from [15], for completeness of this paper, we provide a simple proof of the lemma. Note also that the statement (ii) $\to$ (i) follows from Lemma 1.1 in [11]. First, observe that because of Lemma 2.1, statement (ii) for ρ and ρ_1 are equivalent, and so we can use them interchangeably. Assume that the first statement is true. For any $x \in S$, $y \in H$, $\|y\| = 1$ and any $0 < u \leq M$, let $x' := x + uy$, $x'' := x - uy$. Then, we have $\|x - x'\| = u \leq M$ and $\|x'' - x\| = u \leq M$. We apply (2.7) for the pairs (x', x) and (x'', x) to obtain

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

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 $\rho(E, u, S) \leq \alpha u^q$, $0 < u \leq M$, which gives the lemma for ρ.

Conversely, suppose that (ii) holds for ρ_1. Then, for any $\lambda \in (0, 1)$ and any $x \in S$, $y \in H, \|y\| = 1$, $0 < u \leq M$,

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

This is the same as writing

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

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

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

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

$$E(x') - E(x) - \langle E'(x), x' - x\rangle \leq \alpha_1 \|x' - x\|^q,$$

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

2.2. **The uniform convexity condition.**

We first observe the following claim.

Claim 1. *If the* **UC** *condition holds for a convex function E and a set Ω that is convex and bounded, then the* **UC** *condition holds for all x, $x' \in \Omega$ with β replaced by $\beta_0 > 0$.*

Proof. Since Ω is bounded, there is $L > 0$, such that $diam(\Omega) \leq LM$. Let $x, x' \in \Omega$. If $\|x - x'\| \leq M$, the **UC** condition holds for the pair (x, x') provided $\beta_0 \leq \beta$. If $\|x - x'\| > M$, we chose a point x_1, such that

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

Clearly $\|x - x_1\| = M$, and therefore

$$E(x_1) - E(x) - \langle E'(x), x_1 - x\rangle \geq \beta\|x_1 - x\|^p.$$

Because of the convexity of E,

$$E(x_1) - E(x) \leq \gamma[E(x') - E(x)].$$

A combination of the last two inequalities and the fact that $x_1 - x = \gamma(x' - x)$ result in

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

Therefore, the claim has been proven with $\beta_0 = \min\{\beta, \beta L^{1-p}\}$. □

Note that the **UC** condition is a generalization of the notion of strongly convex functions. Recall that a function E is called strongly convex on H, if there is a constant $\beta > 0$, called the convexity parameter of E, such that

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

Next, we discuss the compatibility between the convexity of E and the **UC** condition.

Lemma 2.3. *Let E be a Frechet differentiable function on H. E is convex on H if and only if*

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

Proof. For convex functions on $I\!R^n$, a proof (without the restriction $\|x' - x\| \leq M$) can be found in [2]. Simple modifications of this proof (which we do not give) result in a proof of the lemma. □

Finally, we present a concept which is dual to the modulus of uniform smoothness for convex functions, called the modulus of uniform convexity (see [1, 15]) and show how it is related to the **UC** condition. Given a convex function $E : H \to I\!R$ and a set $S \subset H$, its modulus of uniform convexity on S is defined by

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

We prove a lemma (see [15]) that shows the equivalence of the **UC** condition and certain behavior of the modulus of uniform convexity δ_1 of E.

Lemma 2.4. *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 for any $p \in [2, \infty)$ and $M > 0$.*

(i) *There exists $\beta > 0$, such that for all $x \in S$, $x' \in H, \|x - x'\| \leq M$,*

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

(ii) *There exists* $\beta_1 > 0$, *such that*

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

Proof. Assume that the first statement is true. For any $x \in S$, $y \in H$, $\|y\| = 1$, $0 < u \leq M$ and $\lambda \in (0,1)$, let $x' := x - \lambda u y$, $x'' := x + (1-\lambda)uy$. 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

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

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

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

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

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

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

We take the infimum over $x \in S$, $y \in H$, $\|y\| = 1$ and $\lambda \in (0,1)$ and obtain that $\delta_1(E,u,S) \geq 2^{2-p}\beta u^p$, $0 < u \leq M$, which is (2.11) with $\beta_1 = 2^{2-p}\beta$.

Conversely, suppose that for some $\beta > 0$ we have $\delta_1(E,u,S) \geq \beta u^p$ for all $0 < u \leq M$. It follows from the definition of δ_1 that for any $\lambda \in (0,1)$, $x \in S$, $y \in H$, $\|y\| = 1$ and $0 < u \leq M$,

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

This is the same as writing

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

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

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

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

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

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

We now summarize that as a result of Lemma 2.2 and Lemma 2.4, we have proven the following lemma.

Lemma 2.5. *Let E be a convex function defined on a Hilbert space H. Let us denote by Ω the set $\Omega = \{x \in H : E(x) \leq E(0)\}$ and let E be Frechet differentiable on Ω. Let $\delta_1(E, \cdot, \Omega)$ and $\rho_1(E, \cdot, \Omega)$ be the modulus of uniform convexity and modulus of uniform smoothness of E on Ω, respectively. The following two statements are equivalent*

(i) *E satisfies the* **US** *and the* **UC** *conditions.*

(ii) *There exist constants $\alpha_1 > 0, \beta_1 > 0$, such that*

$$\beta_1 u^p \leq \delta_1(E, u, \Omega) \leq \rho_1(E, u, \Omega) \leq \alpha_1 u^q, \quad u \in (0, M].$$

3. Greedy algorithms for optimization

In this section, we introduce the two algorithms for convex minimization in a Hilbert space H that we will analyze. As usual, we assume that $\{\varphi_j\}_{j=1}^{\infty}$ is an orthonormal basis for H. We begin with the OMP(co) algorithm.

Orthogonal Matching Pursuit (OMP(co)):

- **Step** 0: Define $x_0 := 0$. If $E'(x_0) = 0$, stop the algorithm and define $x_k := x_0$, $k \geq 1$.
- **Step** m: Assuming x_{m-1} has been defined and $E'(x_{m-1}) \neq 0$, find

$$\varphi_{j_m} := \text{argmax}\{|\langle E'(x_{m-1}), \varphi\rangle|, \varphi \in \mathcal{D}\}, \tag{3.12}$$

and define

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

If $E'(x_m) = 0$, stop the algorithm and define $x_k := x_m$, $k > m$. Otherwise, go to **Step** $m + 1$.

Note that if the algorithm stops at step m, then x_m is the minimizer $\bar{x}$, because of the following well-known lemma.

Lemma 3.1. *Let E be a Frechet differentiable convex function, defined on a convex domain Ω. Then E has a global minimum at $\bar{x} \in \Omega$ if and only if $E'(\bar{x}) = 0$.*

Weak Orthogonal Matching Pursuit (WOMP(co)): The description of the WOMP(co) is the same as the OMP(co), with the only difference that a sequence $\{t_k\}_{k=1}^{\infty}$, $t_k \in (0, 1]$ is used to weaken the condition on the choices of the φ_{j_k}'s. Namely, φ_{j_m} is now chosen to satisfy the inequality

$$|\langle E'(x_{m-1}^{\text{w}}), \varphi_{j_m}\rangle| \geq t_m \sup_{\varphi \in \mathcal{D}} |\langle E'(x_{m-1}^{\text{w}}), \varphi\rangle|,$$

where x_{m-1}^{w} is the output of the WOMP(co) at Step $m - 1$. When $t_k = 1$, $k \geq 1$, the WOMP(co) becomes the OMP(co).

Let us remark that neither of these two algorithms generates a unique sequence x_m, $m \geq 0$. The analysis that follows applies to any sequence generated by the corresponding algorithm.
For a comparison with the results we prove in this paper, we recall a result of Temlyakov. Let $A_1(\mathcal{D})$ denote the closure (in X) of the convex hull of $\mathcal{D}$. The following theorem was proved in [11] in the more general setting of a Banach space X and general symmetric dictionary $\mathcal{D}$.

Theorem 3.2 ([11] 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 on Ω satisfy $\rho(E, u, \Omega) \leq \gamma u^q$, $u > 0$, where $1 < q \leq 2$. If for a given $\epsilon > 0$, there is an element $\varphi^\epsilon \in \mathcal{D}$, such that*

$$E(\varphi^\epsilon) \leq \inf_{x\in\Omega} E(x) + \epsilon, \quad \varphi^\epsilon/A(\epsilon) \in A_1(\mathcal{D}),$$

for some constant $A(\epsilon) \geq 1$, then the output x_m^{w} of the WOMP(co) satisfies the inequality

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

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

4. Main results

In this section, we present our main results and the auxiliary lemmas, needed for their proof. First, note that the set $\Omega := \{x \in H : E(x) \leq E(0)\}$ is convex since it is the level set of a convex function. Also, all outputs $\{x_k\}_{k=1}^\infty$ generated by the OMP(co) (or the WOMP(co)) are in Ω, since the sequence $\{E(x_k)\}_{k=1}^\infty$ is decreasing and $E(x_1) \leq E(0)$.

4.1. **Auxiliary lemmas.**

We begin with some lemmas that we use to derive our main results. The next lemma is well-known.

Lemma 4.1. *Let F be a Frechet differentiable function. Let $V_k :=$ span$\{\varphi_{j_1}, \dots, \varphi_{j_k}\}$ and $x_k :=$ argmin$\{F(x) : x \in V_k\}$. Then, we have that $\langle F'(x_k), \varphi\rangle = 0$ for every $\varphi \in V_k$.*

Our next lemma can be viewed as a generalization of Lemma 2.16 from [13].

Lemma 4.2. *Let $\ell > 0$, $r > 0$, $B > 0$, $\{a_m\}_{m=1}^\infty$ and $\{r_m\}_{m=2}^\infty$ be sequences of non-negative numbers satisfying the inequalities*

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

Then, we have

$$a_m \le \max\{1, \ell^{-1/\ell}\} r^{1/\ell}(rB^{-\ell} + \Sigma_{k=2}^m r_k)^{-1/\ell}, \quad m = 2, 3, \dots. \tag{4.13}$$

Proof. Let us first notice that from the recursive relation and the fact that all a_m's are non-negative, we have

$$0 \le 1 - \frac{r_{m+1}}{r} a_m^\ell \le 1, \quad m = 1, 2, \dots. \tag{4.14}$$

We will show that for $m = 2, 3, \dots$

$$a_m^\ell \le \begin{cases} \dfrac{r}{(rB^{-\ell} + \Sigma_{k=2}^m r_k)}, & \text{if } \ \ell \ge 1, \\ \dfrac{r}{(rB^{-\ell} + \ell \Sigma_{k=2}^m r_k)}, & \text{if } \ 0 < \ell \le 1, \end{cases} \tag{4.15}$$

from which inequality (4.13) easily follows.
We prove (4.15) by induction.
Case 1: $\ell \ge 1$.

If $a_2 = 0$, then all $a_m = 0$, $m = 3, 4, \dots$, and the lemma is true. Let us assume that $a_2 > 0$, and therefore $a_1 > 0$. It follows from the recursive relation and (4.14) that for $\ell \ge 1$

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

This gives (4.15) for $m = 2$.

We now assume that (4.15) is true for m and prove it's validity for $m + 1$. As in 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 (4.14), we derive

$$a_{m+1}^{-\ell} \ge a_m^{-\ell}(1 - \frac{r_{m+1}}{r} a_m^\ell)^{-\ell} \ge a_m^{-\ell}(1 + \frac{r_{m+1}}{r} a_m^\ell) = a_m^{-\ell} + \frac{r_{m+1}}{r}. \tag{4.16}$$

Now, from the induction hypothesis we have that

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

which combined with (4.16) proves the lemma in the case $\ell \ge 1$.
Case 2: $0 < \ell < 1$.

Again, we need 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

$$(1 - t)^\ell \le 1 - \ell t, \quad 0 \le t \le 1. \tag{4.17}$$

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

$$\begin{aligned} a_2^{-\ell} &\geq a_1^{-\ell}(1-\frac{r_2}{r}a_1^\ell)^{-\ell} \geq a_1^{-\ell}(1-\ell\frac{r_2}{r}a_1^\ell)^{-1} \geq a_1^{-\ell}(1+\ell\frac{r_2}{r}a_1^\ell) \\ &= a_1^{-\ell} + \ell\frac{r_2}{r} \geq B^{-\ell} + \ell\frac{r_2}{r}, \end{aligned}$$

which gives (4.15) for $m = 2$. Next , we assume that (4.15) is true for m and prove it for $m+1$. We can assume $a_{m+1} > 0$ and therefore $a_m > 0$. From the recursive relation and (4.17) with $t = \frac{r_{m+1}}{r} a_m^\ell \in [0,1]$, we have

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

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 complete. □

4.2. **Convergence rates for OMP(co).**

In this section, we analyze the performance of the OMP(co) algorithm when applied to the minimization problem (1.1) with $D = H$. We assume that the dictionary $\mathcal{D}$ is an orthonormal system $\{\varphi_i\}_{i=1}^\infty$ for H and E attains its global minimum $\bar{x}$. This means that this global minimum is assumed over $\Omega := \{x : \ E(x) \leq E(0)\}$. Let us denote by e_m the error of the algorithm at Step m, namely,

$$e_m := E(x_m) - E(\bar{x}).$$

The next lemma provides a recursive relation for the sequence $\{e_m\}_{m=0}^\infty$.

Lemma 4.3. *Let the objective function E satisfy* **Condition 0**, *the* **US**, *the* **UC** *condition, and μ be a constant such that $\mu > \max\{1, M_0\alpha^{-1}M^{1-q}\}$. 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 error of the OMP(co) applied to E and $\{\varphi_i\}$ satisfies the following recursive inequalities*

$$e_1 \leq E(0) - E(\bar{x}), \tag{4.18}$$

and

$$e_m \leq e_{m-1} - \frac{(\mu-1)\mu^{-q/(q-1)}}{r} e_{k-1}^{\frac{(p-1)q}{(q-1)p}}, \quad m \geq 2, \tag{4.19}$$

where the constant r is given by

$$r = |\bar{S}|^{\frac{q}{2(q-1)}} \alpha^{\frac{1}{q-1}} \left(p\beta_0^{1/p}(p-1)^{(1-p)/p} \right)^{-q/(q-1)}.$$

Proof. Clearly, we have $e_1 = E(x_1) - E(\bar{x}) \le E(0) - E(\bar{x})$, since

$$x_1 := \mathrm{argmin}_{x \in span\{\varphi_{j_1}\}} E(x).$$

Next, we consider Step m, $m = 2, 3, \dots$ of the algorithm. Observe that if at Step $m-1$ we have that $\bar{S} \subseteq \{j_1, \dots, j_{m-1}\}$, then $x_{m-1} = \bar{x}$, $E'(x_{m-1}) = 0$ and the OMP(co) would have stopped with output $x_k = x_{m-1} = \bar{x}$, $k > m-1$. If the algorithm has not stopped, then it generates the next output x_m and φ_{j_m}. Since x_m is the point of minimum of E over $span\{\varphi_{j_1}, \varphi_{j_2}, \dots, \varphi_{j_m}\}$, we have for any $|t| \le M$,

$$E(x_m) \le E(x_{m-1} + t\varphi_{j_m}) \le E(x_{m-1}) + t\, \langle E'(x_{m-1}), \varphi_{j_m} \rangle + \alpha |t|^q, \tag{4.20}$$

where the last inequality invoked the **US** condition. We take

$$t = -\,(\alpha\mu)^{-\frac{1}{q-1}}\, \mathrm{sign}(\langle E'(x_{m-1}), \varphi_{j_m} \rangle) |\langle E'(x_{m-1}), \varphi_{j_m} \rangle|^{\frac{1}{q-1}}.$$

Because of the definition of μ in the statement of the theorem, we have $|t| \le M$. Therefore, we have

$$E(x_m) \le E(x_{m-1}) - \frac{\mu - 1}{\mu}\, (\alpha\mu)^{-\frac{1}{q-1}}\, |\langle E'(x_{m-1}), \varphi_{j_m} \rangle|^{q/(q-1)}. \tag{4.21}$$

Now, we will find a lower bound for $|\langle E'(x_{m-1}), \varphi_{j_m} \rangle|$. First, note that from the **UC** condition and Claim 1, applied to $x' = \bar{x}$ and $x = x_{m-1}$ (both are in Ω), we obtain

$$\langle E'(x_{m-1}), x_{m-1} - \bar{x} \rangle \ge E(x_{m-1}) - E(\bar{x}) + \beta_0 \|\bar{x} - x_{m-1}\|^p. \tag{4.22}$$

Let us recall the weighted arithmetic mean -geometric mean inequality

$$\frac{p_1}{p_1 + p_2} a + \frac{p_2}{p_1 + p_2} b \ge a^{\frac{p_1}{p_1+p_2}} b^{\frac{p_2}{p_1+p_2}}, \quad \text{where} \quad a, b \ge 0, \quad p_1, p_2 > 0,$$

and apply it for $p_1 = p - 1$, $p_2 = 1$, $a = \frac{1}{p-1}(E(x_{m-1}) - E(\bar{x}))$, $b = \beta_0 \|\bar{x} - x_{m-1}\|^p$. We have

$$E(x_{m-1}) - E(\bar{x}) + \beta_0 \|\bar{x} - x_{m-1}\|^p = p \left(\frac{(p-1)}{p} \frac{(E(x_{m-1}) - E(\bar{x}))}{p-1} + \frac{1}{p} \beta_0 \|\bar{x} - x_{m-1}\|^p \right),$$

and therefore

$$E(x_{m-1}) - E(\bar{x}) + \beta_0 \|\bar{x} - x_{m-1}\|^p \ge C \|\bar{x} - x_{m-1}\| \left(E(x_{m-1}) - E(\bar{x})\right)^{(p-1)/p},$$

with $C = p\beta_0^{1/p}(p-1)^{(1-p)/p}$. We combine this inequality with (4.22) to obtain

$$\langle E'(x_{m-1}), x_{m-1} - \bar{x} \rangle \ge C \|\bar{x} - x_{m-1}\| \left(E(x_{m-1}) - E(\bar{x})\right)^{(p-1)/p}. \tag{4.23}$$

From the definition of x_{m-1} and Lemma 4.1, it follows that

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

Therefore, if we write

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

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

$$\begin{aligned}\langle E'(x_{m-1}),x_{m-1}-\bar{x}\rangle &= \sum_{i\in\bar{S}\setminus\{j_1,\dots,j_{m-1}\}} c_i(x_{m-1}-\bar{x})\langle E'(x_{m-1}),\varphi_i\rangle,\\ &\le \sum_{i\in\bar{S}\setminus\{j_1,\dots,j_{m-1}\}} |c_i(x_{m-1}-\bar{x})||\langle E'(x_{m-1}),\varphi_{j_m}\rangle| \qquad (4.24)\\ &\le |\langle E'(x_{m-1}),\varphi_{j_m}\rangle||\bar{S}|^{1/2}\|x_{m-1}-\bar{x}\|.\end{aligned}$$

We combine this inequality with (4.23) and derive that

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

Therefore we have the desired lower bound

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

The latter result and (4.21) gives the estimate

$$E(x_m)\le E(x_{m-1})-\frac{(\mu-1)C^{q/(q-1)}}{\mu^{q/(q-1)}\alpha^{1/(q-1)}|\bar{S}|^{\frac{q}{2(q-1)}}}\left(E(x_{m-1})-E(\bar{x})\right)^{\frac{(p-1)q}{(q-1)p}}.$$

Subtracting $E(\bar{x})$ from both sides of this inequality results in (4.19) and the proof is completed. □

Note that Lemma 4.3 holds for any value of μ, provided $\mu>\max\{1,M_0\alpha^{-1}M^{1-q}\}$. We can choose a specific value for μ that provides the best error estimate in (4.19) as the following remark states.

Remark 4.4. *Let the objective function E satisfy* **Condition 0** *, the* **US***, and the* **UC** *condition. 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 error of the OMP(co) applied to E and $\{\varphi_i\}$ satisfies the following recursive inequalities:*

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

and

$$e_m\le e_{m-1}-\frac{C_3}{r}e_{m-1}^{\frac{(p-1)q}{(q-1)p}}=e_{m-1}[1-\frac{C_3}{r}e_{m-1}^{\frac{p-q}{(q-1)p}}],\quad k\ge 2, \qquad (4.25)$$

where r is the constant from Lemma 4.3 and $C_3=C_3(M_0,M,\alpha,q)$ is

$$C_3=\begin{cases}(q-1)q^{-q/(q-1)}, & \text{if } M_0M^{1-q}\alpha^{-1}<q,\\ (M_0M^{1-q}\alpha^{-1}-1)M_0^{-q/(q-1)}M^{-q}\alpha^{q/(q-1)}, & \text{if } M_0M^{1-q}\alpha^{-1}\ge q.\end{cases} \qquad (4.26)$$

Proof. The estimate follows from Lemma 4.3 and the fact that the function

$$g(\mu) = (\mu - 1)\mu^{-q/(q-1)}$$

is increasing on $(1, q)$ and decreasing on (q, ∞) with global maximum at $\mu = q$. □

The next theorem is our main result about the OMP(co) algorithm.

Theorem 4.5. *Let the objective function E satisfy* **Condition 0***, the* **US***, and the* **UC** *conditions. Let problem (1.1) with $D = \Omega := \{x : E(x) \leq E(0)\}$ 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 for H. Then, the OMP(co) applied to E and $\{\varphi_i\}$ outputs a sequence $\{x_k\}_{k=0}^{\infty}$ for which*
(i) *When $p \neq q$, for $k = 2, 3, \ldots$,*

$$e_k \leq C|\bar{S}|^{\frac{pq}{2(p-q)}} k^{-\frac{p(q-1)}{p-q}},$$
$$\|x_k - \bar{x}\| \leq C'|\bar{S}|^{\frac{q}{2(p-q)}} k^{-\frac{q-1}{p-q}}$$

where C and C' depend only on p, q, α, β, E.
(ii) *When $p = q = 2$, we have the exponential decay*

$$e_k \leq C_2\gamma^{k-1},$$
$$\|x_k - \bar{x}\| \leq C_2^{\frac{1}{2}}\beta_0^{-\frac{1}{2}}\gamma^{(k-1)/2}, \quad k = 2, 3, \ldots,$$

where $\gamma := 1 - \frac{\tilde{C}_3}{|\bar{S}|}$ is in $(0, 1)$, $C_2 = E(0) - E(\bar{x})$, and $\tilde{C}_3$ is a constant that depends on α, β, and E.

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

$$r_k = C_3, \quad e_k = E(x_k) - E(\bar{x}), \quad k = 1, 2, \ldots,$$

and the numbers

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

It follows from Remark 4.4 that we can apply Lemma 4.2 to the above defined sequences and obtain

$$e_k = E(x_k) - E(\bar{x}) \leq C_0 \left(\frac{|\bar{S}|^{\frac{q}{2(q-1)}}}{C_1|\bar{S}|^{\frac{q}{2(q-1)}} + C_3(k-1)} \right)^{\frac{p(q-1)}{p-q}}, \tag{4.27}$$

where

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

and

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

One easily derives the estimate for e_k in (i) from (4.27). The estimate for $\|x_k - \bar{x}\|$ in (i) now follows from the **UC** condition with $x' = x_k$, $x = \bar{x}$ and Lemma 3.1 .

In the case $p = q = 2$, as before $E(x_1) - E(\bar{x}) \le E(0) - E(\bar{x})$, and Lemma 4.3 and Remark 4.4 give that

$$e_k = E(x_k) - E(\bar{x}) \le \left(1 - \frac{\tilde{C}_3}{|\bar{S}|}\right)(E(x_{k-1}) - E(\bar{x})), \quad k = 2,3,\dots,$$

where

$$\tilde{C}_3 = \begin{cases} \dfrac{\beta_0}{\alpha}, & \text{if } M_0M^{-1}\alpha^{-1} < 2, \\ 4\beta_0(M_0M^{-1}\alpha^{-1}-1)M_0^{-2}M^{-2}\alpha, & \text{if } M_0M^{-1}\alpha^{-1} \ge 2. \end{cases} \tag{4.28}$$

It follows that

$$e_k = E(x_k) - E(\bar{x}) \le (E(0) - E(\bar{x}))\left(1 - \frac{\tilde{C}_3}{|\bar{S}|}\right)^{k-1}, \quad k = 2,3,\dots.$$

As in the previous case, we use the **UC** condition with $x' = x_k$, $x = \bar{x}$ and Lemma 3.1 to derive the estimate for $\|x_k - \bar{x}\|$. □

4.3. Main results for WOMP(co).

The convergence analysis of the WOMP(co) is similar to the one for the OMP(co). We omit the details here and just state the estimates for the error

$$e_k^{\mathrm{w}} := E(x_k^{\mathrm{w}}) - E(\bar{x}), \quad k = 1,2,\dots,$$

for the output sequence $\{x_k^{\mathrm{w}}\}$ of the WOMP(co), pointing out the main differences in the proof.

Theorem 4.6. *Let the objective function E satisfy* **Condition 0***, the* **US***, and the* **UC** *conditions. Let problem* (1.1) *with $D = \Omega = \{x : E(x) \le E(0)\}$ have a solution $\bar{x} = \sum_i c_i(\bar{x})\varphi_i \in \Omega$ with support $\bar{S} := \{i : c_i(\bar{x}) \ne 0\} < \infty$, where $\{\varphi_i\}$ is an orthonormal basis. Then, the WOMP applied to E and $\{\varphi_i\}$ outputs a sequence $\{x_k^{\mathrm{w}}\}_{k=0}^{\infty}$, for which*

(i) *When $p \ne q$, for each $k = 2,3,\dots$, we have*

$$e_k^{\mathrm{w}} \le \tilde{C}|\bar{S}|^{\frac{pq}{2(p-q)}} \left(\sum_{j=2}^{k} t_j^{\frac{q}{q-1}}\right)^{\frac{p(q-1)}{p-q}},$$

$$\|x_k^{\mathrm{w}} - \bar{x}\| \le \tilde{C}'|\bar{S}|^{\frac{q}{2(p-q)}} \left(\sum_{j=2}^{k} t_j^{\frac{q}{q-1}}\right)^{\frac{(q-1)}{p-q}},$$

where $\tilde{C}$ *and* $\tilde{C}'$ *depend only on* p, q, α, β, E.

(ii) *When* $p = q = 2$, *we have for each* $k = 2, 3, \ldots$,

$$e_k^{\mathrm{w}} \le C_2 \prod_{j=2}^{k} \left(1 - \frac{\tilde{C}_3}{|\bar{S}|} t_j^2\right),$$

$$\|x_k^{\mathrm{w}} - \bar{x}\| \le C_2^{\frac{1}{2}} \beta^{-\frac{1}{2}} \prod_{j=2}^{k} \left(1 - \frac{\tilde{C}_3}{|\bar{S}|} t_j^2\right)^{1/2},$$

with $C_2 = E(0) - E(\bar{x})$ *and* $\tilde{C}_3$ *depends on* α, β, *and* E.

Proof. The proof follows the lines of that of Theorem 4.5 and the corresponding lemmas. The difference is that instead of estimate (4.24), we have

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

and that we use Lemma 4.2 with $r_k = C_3 t_k^{\frac{q}{q-1}}$, $k = 2, 3, \ldots$. □

References

[1] J. Borwein, A. Guiro, P. Hajek, and J. Vanderwerff, *Uniformly convex functions on Banach Spaces*, Proc. Amer. Math. Soc., **137**, 1081–1091, 2009.

[2] S. Boyd, L. Vandenberghe, Convex optimization, Cambridge University Press, 2009.

[3] E. Candes and T. Tao, *Decoding by linear programming*, IEEE Trans. Inf. Theory, **51**(12), 4203–4215, 2005.

[4] R. DeVore, V. Temlyakov, *Some remarks on greedy algorithms*, Advances in Computational Math., **5**, 173–187, 1996.

[5] R. DeVore, V. Temlyakov, *Convex optimization on Banach spaces*, Foundations of Computational Mathematics, accepted.

[6] D. Donoho, *Compressed Sensing*, IEEE Trans. Inf. Theory, **52**(4), 1289–1306, 2006.

[7] S. Mallat, Z. Zhang, *Matching Pursuits with Time-Frequency Dictionaries*, IEEE Transactions on Signal Processing, **41**(12), 3397–3415, 1993.

[8] A. Juditsky, Y. Nesterov, *Deterministic and stochastic primal-dual subgradient algorithms for uniformly convex minimization*, Stochastic Systems, **4**, 1–37, 2014.

[9] S. Shalev-Shwartz, N. Srebro, T. Zhang, *Trading Accuracy for Sparsity in Optimization Problems with Sparsity Constraints*, SIAM Journal on Optimization, **20**, 2807–2832, 2010.

[10] V. Temlyakov, *Greedy expansions in convex optimization*, Proceedings of the Steklov Institute of Mathematics, **284**(1), 244–262, 2014.

[11] V. Temlyakov, *Greedy approximation in convex optimization*, Constr. Approx., **41**(2), 269–296, 2015.

[12] V. Temlyakov, *Chebyshev Greedy Algorithm in convex optimization*, preprint, arXiv: 1312.1244v1.

[13] V. Temlyakov, Greedy approximation, Cambridge monographs on Applied and Computational Mathematics, Cambridge University Press, 2011.

[14] A. Tewari, P. Ravikumar, I. Dhillon, *Greedy algorithms for structurally constrained high dimensional problems*, Advances in Neural Information Processing Systems (NIPS), **24**, 2011.

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

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

[17] T. Zhang, *Sparse recovery with Orthogonal Matching Pursuit under RIP*, IEEE Trans. Info. Th, **57**, 5215–6221, 2011.

Hao Nguyen
WorldQuant LLC,
Hanoi, Vietnam
haospt@gmail.com

Guergana Petrova
Department of Mathematics, Texas A&M University,
College Station, TX 77843, USA
gpetrova@math.tamu.edu