INTERDISCIPLINARY
MATHEMATICS
INSTITUTE

2010:01

Orthogonal super greedy algorithm and application in compressed sensing

E.-T. Liu and V. Temlyakov

COLLEGE OF ART AND SCIENCE
UNIVERSITY OF SOUTH CAROLINA

IMI

PREPRINT SERIES

Orthogonal Super Greedy Algorithm and Applications in Compressed Sensing*

Entao Liu [†]and V.N. Temlyakov [‡]

January 26, 2010

Abstract

The general theory of greedy approximation is well developed. Much less is known about how specific features of a dictionary can be used for our advantage. In this paper we discuss incoherent dictionaries. We build a new greedy algorithm which is called the Orthogonal Super Greedy Algorithm (OSGA). OSGA is more efficient than a standard Orthogonal Greedy Algorithm (OGA). We show that the rates of convergence of OSGA and OGA with respect to incoherent dictionaries are the same. Greedy approximation is also a fundamental tool for sparse signal recovery. The performance of Orthogonal Multi Matching Pursuit (OMMP), a counterpart of OSGA in the compressed sensing setting, is also analyzed under RIP conditions.

1 Introduction

We discuss here greedy approximation with regard to a redundant system (dictionary). The general theory of greedy approximation with regard to an arbitrary dictionary is well developed (see [15]). Much less is known about how specific features of a dictionary can be used for our advantage

Math Subject Classifications. primary: 41A65; secondary: 41A25, 41A46, 46B20.

[†]Dept. of Mathematics, 1523 Greene Street, University of South Carolina, Columbia, SC 29208, USA e-mail: liue@mailbox.sc.edu

[‡]Dept. of Mathematics, 1523 Greene Street, University of South Carolina, Columbia, SC 29208, USA e-mail: temlyak@math.sc.edu

– either to improve rate of convergence results for known algorithms or to build more efficient algorithms with the same rate of convergence as known general algorithms. A specific feature of a dictionary – M-coherence in our case – allows us to build a more efficient greedy algorithm – Orthogonal Super Greedy Algorithm – than a known algorithm – Orthogonal Greedy Algorithm – with the same rate of convergence. We study rate of convergence of greedy algorithms for elements of the closure of the convex hull of the symmetrized dictionary, which is standard in the theory of greedy approximation setting. We present these results in Sections 2–3.

It is well known that greedy algorithms are a suitable tool for recovering sparse signals (see, for instance, [12], [16], [17], [9], [10], [11], [13], [5], [4]). Along with ℓ_1-minimization they play a fundamental role in compressed sensing. In Sections 4–6 we illustrate how the Orthogonal Super Greedy Algorithm can be used in the compressed sensing setting. In signal processing the standard name for the Greedy Algorithm is the Matching Pursuit. For instance, the Orthogonal Greedy Algorithm is called the Orthogonal Matching Pursuit. For the reader's convenience we will use the term Matching Pursuit for Greedy Algorithm in Sections 4–6. In particular, the Orthogonal Super Greedy Algorithm will be called the Orthogonal Multi Matching Pursuit. Further discussion and comments will be given in Sections 2, 4, 5, 6.

2 Weak Orthogonal Greedy Algorithm

We begin with a known result on the rate of convergence of the Orthogonal Greedy Algorithm. We recall some notations and definitions from the theory of greedy algorithms. Let H be a real Hilbert space with an inner product $\langle \cdot, \cdot \rangle$ and the norm $\|x\| := \langle x, x\rangle^{1/2}$. We say a set $\mathcal{D}$ of functions (elements) from H is a dictionary if each $g \in \mathcal{D}$ has a unit norm ($\|g\| = 1$) and $\overline{\operatorname{span}}\mathcal{D} = H$. Sometimes it will be convenient for us to consider along with $\mathcal{D}$ the symmetrized dictionary $\mathcal{D}^{\pm} := \{\pm g, g \in \mathcal{D}\}$. Let

$$M(\mathcal{D}) := \sup_{\substack{\varphi \neq \psi \\ \varphi, \psi \in \mathcal{D}}} |\langle \varphi, \psi \rangle|$$

be the coherence parameter of dictionary $\mathcal{D}$. Let a sequence $\tau = \{t_k\}_{k=1}^{\infty}$, $0 \le t_k \le 1$, be given. The following greedy algorithm was defined in [14].

Weak Orthogonal Greedy Algorithm (WOGA). We define $f_0^{o,\tau} := f$. Then for each $m \geq 1$ we inductively define:

(1) $\varphi_m^{o,\tau} \in \mathcal{D}$ is any element satisfying

$$|\langle f_{m-1}^{o,\tau}, \varphi_m^{o,\tau} \rangle| \geq t_m \sup_{g \in \mathcal{D}} |\langle f_{m-1}^{o,\tau}, g \rangle|.$$

(2) Let $H_m^\tau := \operatorname{span}(\varphi_1^{o,\tau}, \dots, \varphi_m^{o,\tau})$ and let $P_{H_m^\tau}(f)$ denote an operator of orthogonal projection onto H_m^τ. Define

$$G_m^{o,\tau}(f, \mathcal{D}) := P_{H_m^\tau}(f).$$

(3) Define the residual after mth iteration of the algorithm

$$f_m^{o,\tau} := f - G_m^{o,\tau}(f, \mathcal{D}).$$

In the case $t_k = 1$, $k = 1, 2, \dots$, WOGA is called the Orthogonal Greedy Algorithm (OGA). Denote by $A_1(\mathcal{D})$ the closure of the convex hull of $\mathcal{D}^\pm$. The following theorem is from [14].

Theorem 2.1. *Let $\mathcal{D}$ be an arbitrary dictionary in H. Then for each $f \in A_1(\mathcal{D})$ we have*

$$\|f - G_m^{o,\tau}(f, \mathcal{D})\| \leq (1 + \sum_{k=1}^m t_k^2)^{-1/2}. \tag{2.1}$$

We note that in a particular case $t_k = t$, $k = 1, 2, \dots$, the right hand side takes form $(1 + mt^2)^{-1/2}$ that is equal to $(1 + m)^{-1/2}$ for $t = 1$.

We now introduce a new algorithm. Let a natural number s and a sequence $\tau := \{t_k\}_{k=1}^\infty$, $t_k \in [0, 1]$, be given. Consider the following Weak Orthogonal Super Greedy Algorithm with parameter s.

WOSGA(s, τ). Initially, $f_0 := f$. Then for each $m \geq 1$ we inductively define:

(1) $\varphi_{(m-1)s+1}, \dots, \varphi_{ms} \in \mathcal{D}$ are elements of the dictionary $\mathcal{D}$ satisfying the following inequality. Denote $I_m := [(m-1)s + 1, ms]$ and assume that

$$\min_{i \in I_m} |\langle f_{m-1}, \varphi_i \rangle| \geq t_m \sup_{g \in \mathcal{D}, g \neq \varphi_i, i \in I_m} |\langle f_{m-1}, g \rangle|.$$

2) Let $H_m := H_m(f) := \operatorname{span}(\varphi_1, \dots, \varphi_{ms})$ and let P_{H_m} denote an operator of orthogonal projection onto H_m. Define

$$G_m(f) := G_m(f, \mathcal{D}) := G_m^s(f, \mathcal{D}) := P_{H_m}(f).$$

3) Define the residual after mth iteration of the algorithm

$$f_m := f_m^s := f - G_m(f, \mathcal{D}).$$

In this paper we study rate of convergence of WOSGA(s, τ) in the case $t_k = t$, $k = 1, 2, \dots$. In this case we write t instead of τ in the notations. We assume that the dictionary $\mathcal{D}$ is M-coherent and that $f \in A_1(\mathcal{D})$. We begin with the case $t = 1$. In this case we impose an additional assumption that the φ_i from the first step exist. Clearly, it is the case if $\mathcal{D}$ is finite. We call the algorithm WOSGA$(s, 1)$ the Orthogonal Super Greedy Algorithm with parameter s (OSGA(s)). In Section 3 we prove the following error bound for the OSGA(s).

Theorem 2.2. *Let $\mathcal{D}$ be a dictionary with coherence parameter $M := M(\mathcal{D})$. Then for $s \le (2M)^{-1}$ OSGA(s) provides, after m iterations, an approximation of $f \in A_1(\mathcal{D})$ with the following upper bound on the error:*

$$\|f_m\|^2 \le 40.5(sm)^{-1}, \quad m = 1, 2, \dots \quad .$$

We note that OSGA(s) adds s new elements of the dictionary at each iteration and makes one orthogonal projection at each iteration. For comparison, OGA adds one new element of the dictionary at each iteration and makes one orthogonal projection at each iteration. After m iterations of OSGA(s) and after ms iterations of OGA both algorithms provide ms-term approximants with a guaranteed error bound for $f \in A_1(\mathcal{D})$ of the same order: $O((ms)^{-1/2})$. Both algorithms use the same number ms of elements of the dictionary. However, OSGA(s) makes m orthogonal projections and OGA makes ms (s times more) orthogonal projections. Thus, in the sense of number of orthogonal projections OSGA(s) is s times simpler (more efficient) than OGA. We gain this simplicity of OSGA(s) under an extra assumption of $\mathcal{D}$ being M-coherent and $s \le (2M)^{-1}$. Therefore, if our dictionary $\mathcal{D}$ is M-coherent then OSGA(s) with small enough s approximates with error whose guaranteed upper bound for $f \in A_1(\mathcal{D})$ is of the same order as that for OGA.

3 Rate of convergence

Proof of Theorem 2.2. Denote

$$F_m := \operatorname{span}(\varphi_i, i \in I_m).$$

Then H_m is a direct sum of H_{m-1} and F_m. Therefore,

$$\begin{aligned} f_m = f - P_{H_m}(f) &= f_{m-1} + G_{m-1}(f) - P_{H_m}(f_{m-1} + G_{m-1}(f)) \\ &= f_{m-1} - P_{H_m}(f_{m-1}). \end{aligned}$$

It is clear that the inclusion $F_m \subset H_m$ implies

$$\|f_m\| \leq \|f_{m-1} - P_{F_m}(f_{m-1})\|. \tag{3.1}$$

Using the notation $p_m := P_{F_m}(f_{m-1})$, we continue

$$\|f_{m-1}\|^2 = \|f_{m-1} - p_m\|^2 + \|p_m\|^2$$

and by (3.1)

$$\|f_m\|^2 \leq \|f_{m-1}\|^2 - \|p_m\|^2. \tag{3.2}$$

To estimate $\|p_m\|^2$ from below for $f \in A_1(\mathcal{D})$, we first make some auxiliary observations. Let

$$f = \sum_{j=1}^{\infty} c_j g_j, \quad g_j \in \mathcal{D}, \quad \sum_{j=1}^{\infty} |c_j| \leq 1, \quad |c_1| \geq |c_2| \geq \dots \quad . \tag{3.3}$$

Every element of $A_1(\mathcal{D})$ can be approximated arbitrarily well by elements of the form (3.3). It will be clear from the below argument that it is sufficient to consider elements f of the form (3.3). Suppose ν is such that $|c_\nu| \geq 2/s \geq |c_{\nu+1}|$. Then the above assumption on the sequence $\{c_j\}$ implies that $\nu \leq s/2$ and $|c_{s+1}| < 1/s$. We claim that elements $g_1, \dots, g_\nu$ will be chosen among $\varphi_1, \dots, \varphi_s$ at the first iteration. Indeed, for $j \in [1, \nu]$ we have

$$|\langle f, g_j \rangle| \geq |c_j| - M \sum_{k \neq j}^{\infty} |c_k| \geq 2/s - M(1 - 2/s) > 2/s - M.$$

For all g distinct from $g_1, \dots, g_s$ we have

$$|\langle f, g \rangle| \leq M + 1/s.$$

Our assumption $s \leq 1/(2M)$ implies that $M + 1/s \leq 2/s - M$. Thus, we do not pick any of $g \in \mathcal{D}$ distinct from $g_1, \dots, g_s$ until we have chosen all $g_1, \dots, g_\nu$.

Denote

$$f' := f - \sum_{j=1}^{\nu} c_j g_j = \sum_{j=\nu+1}^{\infty} c_j g_j.$$

It is clear from the above argument that

$$f_1 = f - P_{H_1(f)}(f) = f' - P_{H_1(f)}(f');$$

$$f_m = f - P_{H_m(f)}(f) = f' - P_{H_m(f)}(f').$$

We now estimate $\|p_m\|^2$. For f_{m-1} consider the following quantity

$$q_s := q_s(f_{m-1}) := \sup_{\substack{h_i \in \mathcal{D} \\ i \in [1,s]}} \|P_{H(s)}(f_{m-1})\|,$$

where $H(s) := \operatorname{span}(h_1, \dots, h_s)$. Then

$$\|P_{H(s)}(f_{m-1})\| = \max_{\psi \in H(s), \|\psi\| \le 1} |\langle f_{m-1}, \psi \rangle|.$$

Let $\psi = \sum_{i=1}^{s} a_i h_i$. Then by Lemma 2.1 from [11] we bound

$$(1 - Ms) \sum_{i=1}^{s} a_i^2 \le \|\psi\|^2 \le (1 + Ms) \sum_{i=1}^{s} a_i^2. \tag{3.4}$$

Therefore,

$$(1 + Ms)^{-1} \sum_{i=1}^{s} \langle f_{m-1}, h_i \rangle^2 \le \|P_{H(s)}(f_{m-1})\|^2 \le (1 - Ms)^{-1} \sum_{i=1}^{s} \langle f_{m-1}, h_i \rangle^2.$$

Thus

$$\|p_m\|^2 \ge \frac{1 - Ms}{1 + Ms} q_s^2. \tag{3.5}$$

Using the notation $J_l := [(l-1)s + \nu + 1, ls + \nu]$ we write for $m \ge 2$

$$\begin{aligned} \|f_{m-1}\|^2 &= \langle f_{m-1}, f' \rangle = \sum_{l=1}^{\infty} \langle f_{m-1}, \sum_{j \in J_l} c_j g_j \rangle \\ &\le q_s (1 + Ms)^{1/2} \sum_{l=1}^{\infty} (\sum_{j \in J_l} c_j^2)^{1/2}. \end{aligned} \tag{3.6}$$

Since the sequence $\{c_j\}$ has the property

$$|c_{\nu+1}| \geq |c_{\nu+2}| \geq \dots, \quad \sum_{j=\nu+1}^{\infty} |c_j| \leq 1, \quad |c_{\nu+1}| \leq 2/s \tag{3.7}$$

we may apply the simple inequality,

$$(\sum_{j\in J_l} c_j^2)^{1/2} \leq s^{1/2}|c_{(l-1)s+\nu+1}|,$$

so that we bound the sum in the right side of (3.6)

$$\begin{aligned} \sum_{l=1}^{\infty}(\sum_{j\in J_l} c_j^2)^{1/2} &\leq s^{1/2}\sum_{l=1}^{\infty}|c_{(l-1)s+\nu+1}| \\ &\leq s^{1/2}(2/s + \sum_{l=2}^{\infty} s^{-1}\sum_{j\in J_{l-1}}|c_j|) \leq 3s^{-1/2}. \end{aligned} \tag{3.8}$$

Inequalities (3.6) and (3.8) imply

$$q_s \geq (s^{1/2}/3)(1+Ms)^{-1/2}\|f_{m-1}\|^2$$

By (3.5) we have

$$\|p_m\|^2 \geq \frac{s(1-Ms)}{9(1+Ms)^2}\|f_{m-1}\|^4. \tag{3.9}$$

Our assumption $Ms \leq 1/2$ implies

$$\frac{1-Ms}{(1+Ms)^2} \geq 2/9$$

and, therefore, (3.9) gives

$$\|p_m\|^2 \geq (s/A)\|f_{m-1}\|^4, \quad A := 40.5.$$

Thus, by (3.2) we get

$$\|f_m\|^2 \leq \|f_{m-1}\|^2(1-(s/A)\|f_{m-1}\|^2). \tag{3.10}$$

Using (3.7) we get for $\|f'\|$

$$\|f'\| \leq \sum_{l=1}^{\infty} \|\sum_{j\in J_l} c_j g_j\| \leq \sum_{l=1}^{\infty}(1+Ms)^{1/2}(\sum_{j\in J_l} c_j^2)^{1/2} \leq (1+Ms)^{1/2} 3s^{-1/2},$$

and

$$\|f_1\|^2 \leq \|f'\|^2 \leq 27/(2s) \leq A/s.$$

We need the following known lemma (see, for example, [7]).

Lemma 3.1. *Let $\{a_m\}_{m=1}^{\infty}$ be a sequence of non-negative numbers satisfying the inequalities*

$$a_1 \leq D, \quad a_{m+1} \leq a_m(1-a_m/D), \quad m=1,2,\dots.$$

Then we have for each m

$$a_m \leq D/m.$$

By Lemma 3.1 with $a_m := \|f_m\|^2$, $D := A/s$, we obtain

$$\|f_m\|^2 \leq A(sm)^{-1}, \quad m=1,2,\dots \quad .$$

This completes the proof of Theorem 2.2.

We now proceed to the case of WOSGA(s,t) with $t\in(0,1)$.

Theorem 3.1. *Let $\mathcal{D}$ be a dictionary with coherence parameter $M := M(\mathcal{D})$. Then for $s \leq (2M)^{-1}$ WOSGA(s,t) provides, after m iterations, an approximation of $f \in A_1(\mathcal{D})$ with the following upper bound on the error:*

$$\|f_m\|^2 \leq A(t)(sm)^{-1}, \quad m=1,2,\dots \quad A(t) := (81/8)(1+t)^2 t^{-4}.$$

Proof. Proof of this theorem mimics the proof of Theorem 2.2, except that in the auxiliary observations we choose a threshold B/s with $B := (3+t)/(2t)$ instead of $2/s$: $|c_\nu| \geq B/s \geq |c_{\nu+1}|$, so that our assumption $Ms \leq 1/2$ implies that $M + 1/s \leq t(B/s - M)$. This, in turn, implies that all $g_1,\dots,g_\nu$ will be chosen at the first iteration. As a result, the sequence $\{c_j\}$ satisfies the following conditions

$$|c_{\nu+1}| \geq |c_{\nu+2}| \geq \dots, \quad \sum_{j=\nu+1}^{\infty} |c_j| \leq 1, \quad |c_{\nu+1}| \leq B/s. \tag{3.11}$$

To find an analog of inequality (3.5), we begin with the fact that

$$q_s^2 \le \sup_{\substack{h_i \in \mathcal{D} \\ i \in [1,s]}} (1-Ms)^{-1} \sum_{i=1}^{s} \langle f_{m-1}, h_i \rangle^2.$$

Now, in order to relate q_s^2 to $\|p_m\|^2$, consider an arbitrary set $\{h_i\}_{i=1}^s$ of distinct elements of the dictionary $\mathcal{D}$. Let V be a set of all indices $i \in [1,s]$ such that $h_i = \varphi_{k(i)}$, $k(i) \in I_m$. Denote $V' := \{k(i), i \in V\}$. Then

$$\sum_{i=1}^{s} \langle f_{m-1}, h_i \rangle^2 = \sum_{i \in V} \langle f_{m-1}, h_i \rangle^2 + \sum_{i \in [1,s] \setminus V} \langle f_{m-1}, h_i \rangle^2. \tag{3.12}$$

From the definition of $\{\varphi_k\}_{k \in I_m}$ we get

$$\max_{i \in [1,s] \setminus V} |\langle f_{m-1}, h_i \rangle| \le t^{-1} \min_{k \in I_m \setminus V'} |\langle f_{m-1}, \varphi_k \rangle|. \tag{3.13}$$

Using (3.13) we continue (3.12)

$$\le \sum_{k \in V'} \langle f_{m-1}, \varphi_k \rangle^2 + t^{-2} \sum_{k \in I_m \setminus V'} \langle f_{m-1}, \varphi_k \rangle^2 \le t^{-2} \sum_{k \in I_m} \langle f_{m-1}, \varphi_k \rangle^2.$$

Therefore,

$$q_s^2 \le (1-Ms)^{-1} t^{-2} \sum_{k \in I_m} \langle f_{m-1}, \varphi_k \rangle^2 \le \frac{1+Ms}{t^2(1-Ms)} \|p_m\|^2.$$

This results in the following analog of (3.5)

$$\|p_m\|^2 \ge \frac{t^2(1-Ms)}{1+Ms} q_s^2. \tag{3.14}$$

The use of (3.11) instead of (3.7) gives us the following version of (3.8)

$$\sum_{l=1}^{\infty} (\sum_{j \in J_l} c_j^2)^{1/2} \le (B+1) s^{-1/2}.$$

The rest of the proof repeats the corresponding part of the proof of Theorem 2.2 with $A := \frac{9(B+1)^2}{2t^2} = (81/8)(1+t)^2 t^{-4}$.

4 Introduction of Compressed Sensing

We now proceed to the compressed sensing (CS) setting. In this case $H = \mathbb{R}^m$ equipped with the Euclidean norm $\|x\| := \langle x, x\rangle^{1/2}$ and $\mathcal{D} = \{\varphi_i\}_{i=1}^N$ is a finite set of elements (column vectors) of $\mathbb{R}^m$. Then the dictionary $\mathcal{D}$ is associated with a $m \times N$ matrix $\Phi = [\varphi_1 \dots \varphi_N]$. The condition $y \in A_1(\mathcal{D})$ is equivalent to existence of $x \in \mathbb{R}^N$ such that $y = \Phi x$ and

$$\|x\|_1 := |x_1| + \cdots + |x_N| \leq 1. \tag{4.1}$$

As a direct corollary of Theorem 2.1, we get for any $y \in A_1(\mathcal{D})$ that the Orthogonal Greedy Algorithm guarantees the following upper bound for the error

$$\|y - G_n^o(y, \mathcal{D})\| \leq (n+1)^{-1/2}. \tag{4.2}$$

The bound (4.2) holds for any $\mathcal{D}$ (any Φ).

In compressed sensing the relation $y = \Phi x$ has the following interpretation. Let $\psi_1, \dots, \psi_m$ be the rows of the matrix Φ. Then the corresponding column vectors ψ_i^T belong to $\mathbb{R}^N$. The relation $y = \Phi x$ is equivalent to $y_i = \langle \psi_i^T, x\rangle$, $i = 1, \dots, m$. The number $y_i = \langle \psi_i^T, x\rangle$ is understood as a linear measurement of an unknown vector x. The goal is to recover (or to approximately recover) the unknown vector x from its measurements y.

The following error bound is one of fundamental results of CS (see [2], [8] and also [15] for further results and a discussion). Denote by

$$A_\Phi(y) := \operatorname*{argmin}_{v:\Phi v = y} \|v\|_1 \tag{P1}$$

the result of application of the ℓ_1-minimization algorithm A_Φ to the data vector y. Then under some conditions on the matrix Φ (RIP$(2n, \delta)$ with small enough δ, which will be discussed momentarily) one has for x satisfying (4.1)

$$\|x - A_\Phi(\Phi x)\| \leq Cn^{-1/2}. \tag{4.3}$$

The inequalities (4.2) and (4.3) look alike. However, they provide the error bounds in different spaces: (4.2) in $\mathbb{R}^m$ (the data space) and (4.3) in $\mathbb{R}^N$ (the coefficients space).

To discuss the performance of greedy approximation in CS, let us begin with the Restricted Isometry Property of Φ introduced by Candes and Tao in [3]. RIP is useful in analysis of performance of recovery algorithms. A vector $x \in \mathbb{R}^N$ has the support $T = \mathrm{supp}(x) := \{i \in \mathbb{N} : 1 \leq i \leq N$ and

$x_i \neq 0\}$. For a set of indices T, denote by $|T|$ the cardinality of the set. If $|T| = K$, x is called a K-sparse signal.

Definition 4.1. (Restricted Isometry Property) A $m \times N$ matrix Φ satisfies the Restricted Isometry Property with parameters (K, δ) (we say Φ satisfies RIP(K, δ) for simplicity) if there exist $\delta \in (0, 1)$ such that

$$(1 - \delta)\|x\|^2 \leq \|\Phi x\|^2 \leq (1 + \delta)\|x\|^2 \tag{4.4}$$

holds for every K-sparse x. Moreover, define $\delta_K := \inf\{\delta : (4.4) \text{ holds for any } K\text{-sparse } x\}$.

To avoid confusions, let us clarify the notations that we use in the rest of this paper. For $x \in \mathbb{R}^N$, $x_\Gamma \in \mathbb{R}^{|\Gamma|}$ is a vector whose entries are the entries of x with indices in Γ. For $m \times N$ matrix Φ, Φ_Γ is a $m \times |\Gamma|$ submatrix of Φ with columns indexed in Γ. Given $y \in \mathbb{R}^m$, the orthogonal projection of y onto $\text{span}(\Phi_\Gamma) := \text{span}\{\varphi_i : i \in \Gamma\}$ is denoted by

$$P_\Gamma(y) := \underset{y' : y' \in \text{span}(\Phi_\Gamma)}{\text{argmin}} \|y - y'\|.$$

It is known and easy to check that if $\Phi_\Gamma^* \Phi_\Gamma$ is invertible then $P_\Gamma(y) = \Phi_\Gamma \Phi_\Gamma^\dagger y$, where $\Phi_\Gamma^\dagger := (\Phi_\Gamma^* \Phi_\Gamma)^{-1} \Phi_\Gamma^*$ is the Moore-Penrose pseudoinverse of Φ_Γ and Φ^* is the transpose of Φ.

We complete this section by presenting one observation and two inequalities that will be used in the proofs later. They are all consequences of RIP. The observation, derived directly from the definition of RIP, is that if Φ satisfies RIP of both orders K and K', with $K < K'$, then $\delta_K \leq \delta_{K'}$.

The two inequalities (see [13] and [5]) are used frequently in this paper.

Lemma 4.1. *Suppose Φ satisfies RIP of order s. Then for any set of indices Γ such that $|\Gamma| \leq s$ and any $x \in \mathbb{R}^N$ and $y \in \mathbb{R}^{|\Gamma|}$, we bound*

$$(1 - \delta_{|\Gamma|})\|x_\Gamma\| \leq \|\Phi_\Gamma^* \Phi_\Gamma x_\Gamma\| \leq (1 + \delta_{|\Gamma|})\|x_\Gamma\| \tag{4.5}$$

and

$$\|\Phi_\Gamma^* y\| \leq (1 + \delta_{|\Gamma|})^{1/2}\|y\|. \tag{4.6}$$

Lemma 4.2. *Assume Γ and Λ are two disjoint sets of indices. If Φ satisfies RIP of order $|\Gamma \cup \Lambda|$ with constant $\delta_{|\Gamma \cup \Lambda|}$, then for any vector $x \in \mathbb{R}^{|\Lambda|}$*

$$\|\Phi_\Gamma^* \Phi_\Lambda x\| \leq \delta_{|\Gamma \cup \Lambda|}\|x\|. \tag{4.7}$$

5 Orthogonal Matching Pursuit

Orthogonal Matching Pursuit (OMP) in the CS setting is the counterpart of OGA in the general setting. It is computationally simple and performs well. Considering its applications in signal processing, let us present it in an algorithmic way.

Algorithm: Orthogonal Matching Pursuit (OMP)

Input: Φ and y
Initialization: $r^0 := y$, $x^0 := 0$, $\Lambda^0 := \emptyset$, $\ell := 0$
Iterations: Define $\Lambda^{\ell+1} := \Lambda^\ell \cup \{\operatorname{argmax}_i |\langle r^\ell, \varphi_i\rangle|\}$.
Then $x^{\ell+1} := \operatorname{argmin}_z \|y - \Phi_{\Lambda^{\ell+1}} z\|$ and
$r^{\ell+1} := y - P_{\Lambda^{\ell+1}}(y) = y - \Phi_{\Lambda^{\ell+1}} x^{\ell+1}$.
If $r^{\ell+1} = 0$, stop. Otherwise let $\ell := \ell + 1$ and begin a new iteration.
Output: If algorithm stops at kth iteration, output $\hat{x}$ is such that
$\hat{x}_{\Lambda^k} = x^k$ and $\hat{x}_{(\Lambda^k)^C} = 0$.

It is known that if the columns of the matrix Φ form a dictionary $\mathcal{D}$ that is M-coherent then OMP recovers exactly any K-sparse signal with $K < (1 + M^{-1})/2$. The behavior of OMP with respect to incoherent dictionaries is well studied (see, for instance, [12], [16], [9], [10], [11]). In this section we study performance of OMP assuming RIP and prove a theorem that guarantees the exact recovery of any K-sparse signal.

Let us take a close look at the first iteration of OMP. OMP chooses the index that corresponds to the largest (in magnitude) inner product. When can we conclude that this index is indeed in the support of x? A corollary of the following more general lemma gives the answer.

From now on we always assume $y = \Phi x$, where $x \in \mathbb{R}^N$ is K-sparse.

Lemma 5.1. *Assume $s \in \mathbb{N}$ and $s \leq K$ and Φ satisfies RIP of order $K + s$ with constant $\delta := \delta_{K+s} < \frac{\sqrt{s}}{(1+\sqrt{2})\sqrt{K}}$. Define $\Lambda := \{i_1, \ldots, i_s\}$ such that*

$$|\langle y, \varphi_{i_1}\rangle| \geq \ldots \geq |\langle y, \varphi_{i_s}\rangle| \geq \sup_{\varphi \in \Phi, \varphi \neq \varphi_i, i \in \Lambda} |\langle y, \varphi\rangle|. \tag{5.1}$$

Then $\Lambda \cap T \neq \emptyset$, where T is the support of x.

Proof. We proceed by contradiction. Assume $\Lambda \cap T = \emptyset$. By Lemma 4.2

$$\|\Phi_\Lambda^* y\| = \|\Phi_\Lambda^* \Phi_T x\| \leq \delta \|x\|. \tag{5.2}$$

Let $T' := \{n_1, \ldots, n_s\}$ be a subset of T which contains the s largest (in magnitude) entries of x. Then

$$|x_{n_1}| \geq \ldots \geq |x_{n_s}| \geq \sup_{n\in[1,N], n\notin T'} |x_n|.$$

From the definition of Λ it follows that

$$\begin{aligned} \|\Phi_\Lambda^* y\| &\geq \|\Phi_{T'}^* y\| = \|\Phi_{T'}^*\Phi_T x\| \\ &= \|\Phi_{T'}^*\Phi_{T'} x_{T'} + \Phi_{T'}^*\Phi_{T\setminus T'} x_{T\setminus T'}\| \\ &\geq \|\Phi_{T'}^*\Phi_{T'} x_{T'}\| - \|\Phi_{T'}^*\Phi_{T\setminus T'} x_{T\setminus T'}\|. \end{aligned} \tag{5.3}$$

Applying (4.5) and Lemma 4.2 to the last two terms in (5.3), we obtain

$$\|\Phi_\Lambda^* y\| \geq (1-\delta_s)\|x_{T'}\| - \delta_K \|x_{T\setminus T'}\|. \tag{5.4}$$

Using (5.4) and (5.2) we get

$$\delta\|x\| \geq (1-\delta_s)\|x_{T'}\| - \delta_K \|x_{T\setminus T'}\|.$$

Replacing in the above inequality δ_s and δ_K by δ we obtain

$$\delta\|x\| \geq (1-\delta)\|x_{T'}\| - \delta \|x_{T\setminus T'}\|. \tag{5.5}$$

Next, using the definition of T' we get

$$\|x_{T'}\| \geq \Big(\frac{s}{K}\Big)^{\frac{1}{2}}\|x\|, \tag{5.6}$$

and

$$\|x_{T\setminus T'}\| \leq \Big(1-\frac{s}{K}\Big)^{\frac{1}{2}}\|x\|. \tag{5.7}$$

Combining (5.6), (5.7) and (5.5) gives

$$\delta\|x\| \geq (1-\delta)\Big(\frac{s}{K}\Big)^{\frac{1}{2}}\|x\| - \delta\Big(1-\frac{s}{K}\Big)^{\frac{1}{2}}\|x\|.$$

Thus

$$\Big(\frac{s}{K}\Big)^{\frac{1}{2}} \leq \delta + \delta\Big(\frac{s}{K}\Big)^{\frac{1}{2}} + \delta\Big(1-\frac{s}{K}\Big)^{\frac{1}{2}}. \tag{5.8}$$

Using the inequality

$$\Big(\frac{s}{K}\Big)^{1/2} + \Big(1-\frac{s}{K}\Big)^{1/2} \leq \sqrt{2},$$

we continue (5.8)

$$\leq (1+\sqrt{2})\delta.$$

Clearly, we have $(s/K)^{\frac{1}{2}} > (1+\sqrt{2})\delta$, if $\delta < \frac{\sqrt{s}}{(1+\sqrt{2})\sqrt{K}}$. This contradiction completes the proof.

When $s = 1$ in Lemma 5.1, we have the following corollary.

Corollary 5.1. *Suppose Φ satisfies RIP($K+1, \delta_{K+1}$) with $\delta_{K+1} < \frac{1}{(1+\sqrt{2})\sqrt{K}}$. If $i \in [1, N]$ such that*

$$|\langle y, \varphi_i \rangle| \geq \sup_{\varphi\in\Phi, \varphi\neq\varphi_i} |\langle y, \varphi\rangle|.$$

Then $i \in T$, where T is the support of x.

This Corollary tells us that under the above conditions on Φ OMP chooses the right index at each step.

We now proceed to the main result of this section.

Theorem 5.2. *Assume Φ satisfies RIP($K+1, \delta$) with $\delta := \delta_{K+1} < \frac{1}{(1+\sqrt{2})\sqrt{K}}$. Then OMP recovers any K-sparse $x \in \mathbb{R}^N$ exactly in K iterations.*

Theorem 5.2 is a follow up to the corresponding result of Davenport and Wakin, [6]. They proved an analog of Theorem 5.2 under assumption $\delta < \frac{1}{3\sqrt{K}}$. Thus, our contribution is in improving the constant from 1/3 to $1/(1+\sqrt{2})$. It is conjectured in [5] that there exist K-sparse x and Φ satisfying RIP of order $K+1$ with $\delta_{K+1} \leq \frac{1}{\sqrt{K}}$ such that OMP can not recover x in K iterations. If this conjecture is true, then we cannot increase the constant $1/(1+\sqrt{2})$ to 1. The question of best constant in Theorem 5.2 is an interesting open question.

Proof of Theorem 5.2. Take any $\ell < K$. From the definition of OMP, we know that Λ^ℓ contains the indices selected at the first ℓ iterations. All indices contained in Λ^ℓ are distinct because we make an orthogonal projection at each iteration. Assume $\Lambda^\ell \subseteq T$, which is equivalent to the statement that the OMP has selected ℓ correct indices (all selected indices are contained in T) after ℓ iterations. As a consequence of this assumption we get that $r^\ell = y - P_{\Lambda^\ell}(y)$ is at most K-sparse and supported on T. If i is selected by the $(\ell+1)$th iteration, it satisfies

$$|\langle r^\ell, \varphi_i \rangle| \geq \sup_{\varphi\in\Phi, \varphi\neq\varphi_i} |\langle r^\ell, \varphi\rangle|.$$

By Corollary 5.1, if Φ satisfies RIP of order $K+1$ with constant $\delta := \delta_{K+1} < \frac{1}{(1+\sqrt{2})\sqrt{K}}$, we know that $i \in T$. Since $r^\ell = y - P_{\Lambda^\ell}(y)$, the inner product of φ_i and any column indexed in Λ^ℓ is zero. This implies $i \notin \Lambda^\ell$. Therefore, $i \in T \backslash \Lambda^\ell$. If Φ satisfies RIP of order $K+1$ with constant $\delta < \frac{1}{(1+\sqrt{2})\sqrt{K}}$, the above argument works for any $\ell \in [0, K-1]$.

By induction from $\ell = 0$ to $\ell = K-1$, we get the conclusion that $\Lambda^K = T$, $r^K = y - P_T(y) = y - \Phi_T x_T = 0$, and OMP stops after K iterations.

The output $\hat{x}$ satisfies

$$\begin{aligned} \hat{x}_{\Lambda^K} &= x^K = \operatorname*{argmin}_z \|y - \Phi_{\Lambda^K} z\| \\ &= \operatorname*{argmin}_z \|y - \Phi_T z\| = x_T, \end{aligned}$$

and

$$\hat{x}_{(\Lambda^K)^C} = x_{T^C} = 0.$$

Hence, $\hat{x} = x$. This completes the proof.

6 Orthogonal Multi Matching Pursuit

Theorem 5.2 requires RIP condition with $\delta_{K+1} < \frac{1}{(1+\sqrt{2})\sqrt{K}}$ so that OMP recovers a K-sparse signal exactly in K iterations. For doing this OMP must select a correct index at each iteration. In this section we study an algorithm that may pick wrong indices in its iterations yet finally recovers the signal exactly. We study here the Orthogonal Multi Matching Pursuit with parameter s (OMMP(s)), which is the Orthogonal Super Greedy Algorithm with parameter s adjusted to the compressed sensing setting. Next, let us establish the algorithm.

Algorithm: Orthogonal Multi Matching Pursuit (OMMP)

Input: Φ, y and s
Initialization: $r^0 := y$, $x^0 := 0$, $\Lambda^0 := \emptyset$, $\ell := 0$
Iterations: Define $\Lambda^{\ell+1} := \Lambda^\ell \cup \{i_1, \ldots, i_s\}$ such that
$|\langle r^\ell, \varphi_{i_1}\rangle| \geq \ldots \geq |\langle r^\ell, \varphi_{i_s}\rangle| \geq \sup\limits_{\substack{\varphi\in\Phi \\ \varphi\neq\varphi_{i_k}, k=1,\ldots,s}} |\langle r^\ell, \varphi\rangle|$.
Then $x^{\ell+1} := \mathrm{argmin}_z \|y - \Phi_{\Lambda^{\ell+1}} z\|$ and
$r^{\ell+1} := y - \Phi_{\Lambda^{\ell+1}} x^{\ell+1}$.
If $r^{\ell+1} = 0$, stop. Otherwise let $\ell := \ell + 1$ and begin a new iteration.
Output: If algorithm stops at kth iteration, output $\hat{x}$ such that
$\hat{x}_{\Lambda^k} = x^k$ and $\hat{x}_{(\Lambda^k)^C} = 0$.

Theorem 6.1. *For $s \in \mathbb{N}$ and $s \leq K$, assume Φ satisfies RIP of order sK with a constant $\delta := \delta_{sK} < \frac{\sqrt{s}}{(2+\sqrt{2})\sqrt{K}}$. Then for any K-sparse $x \in \mathbb{R}^N$, OMMP(s) recovers x exactly within at most K iterations.*

Proof. We will prove that OMMP(s) picks at least one correct index at each iteration.

After running OMMP for ℓ iterations, we obtain a set of indices Λ^ℓ. We prove that at the $(\ell+1)$th iteration OMMP selects at least one correct index. We carry out this proof without assumption that OMMP selected correct indices at previous iterations. Denote $T_1 := \Lambda^\ell \cap T$, $K_1 := |T_1|$ and $T_2 := T \backslash \Lambda^\ell$, $K_2 := |T_2|$. We continue our proof by contradiction. Assume that at the next iteration the OMMP(s) does not select any correct indices. This means $\Lambda' \cap T = \emptyset$, where $\Lambda' := (\Lambda^{\ell+1} \backslash \Lambda^\ell)$. From the definition of OMMP, we easily derive that $|\Lambda'| = s$.

The residual after ℓ iterations can be written in the following form

$$\begin{aligned} r^\ell &= y - P_{\Lambda^\ell}(y) \\ &= \Phi_{T_1} x_{T_1} + \Phi_{T_2} x_{T_2} - P_{\Lambda^\ell}(\Phi_{T_1} x_{T_1} + \Phi_{T_2} x_{T_2}) \\ &= \Phi_{T_2} x_{T_2} - P_{\Lambda^\ell}(\Phi_{T_2} x_{T_2}). \end{aligned} \tag{6.1}$$

Let $x_p \in \mathbb{R}^{|\Lambda^\ell|}$ give the coefficients of projection $P_{\Lambda^\ell}(\Phi_{T_2} x_{T_2})$. Then we continue (6.1)

$$= \Phi_{T_2} x_{T_2} - \Phi_{\Lambda^\ell} x_p. \tag{6.2}$$

We now focus on the second term of (6.2). Since Φ satisfies RIP of order sK, it also satisfies RIP of order $|\Lambda^\ell|$. By (4.5) $\Phi^*_{\Lambda^\ell}\Phi_{\Lambda^\ell}$ is invertible. By the property of the Moore-Penrose pseudoinverse we get

$$x_p = (\Phi^*_{\Lambda^\ell}\Phi_{\Lambda^\ell})^{-1}\Phi^*_{\Lambda^\ell}\Phi_{T_2}x_{T_2}.$$

Since $\Lambda^\ell \cap T_2 = \emptyset$ we apply (4.5) and Lemma 4.2

$$\begin{aligned}
\|x_p\| &= \|(\Phi^*_{\Lambda^\ell}\Phi_{\Lambda^\ell})^{-1}\Phi^*_{\Lambda^\ell}\Phi_{T_2}x_{T_2}\| \\
&\le \frac{\delta_{s\ell+K_2}}{1-\delta_{s\ell}}\|x_{T_2}\| \\
&\le \frac{\delta}{1-\delta}\|x_{T_2}\|. \qquad (6.3)
\end{aligned}$$

By (6.2) we have

$$\begin{aligned}
\|\Phi^*_{\Lambda'}r^\ell\| &= \|\Phi^*_{\Lambda'}(\Phi_{T_2}x_{T_2} - \Phi_{\Lambda^\ell}x_p)\| \\
&\le \|\Phi^*_{\Lambda'}\Phi_{T_2}x_{T_2}\| + \|\Phi^*_{\Lambda'}\Phi_{\Lambda^\ell}x_p\|. \qquad (6.4)
\end{aligned}$$

By our assumption $\Lambda' \cap T = \emptyset$ and, therefore, $\Lambda' \cap T_2 = \emptyset$. Moreover, the definition of Λ' indicates that Λ' and Λ^ℓ are also disjoint. Using Lemma 4.2 we continue (6.4)

$$\begin{aligned}
&\le \delta_{s+K_2}\|x_{T_2}\| + \delta_{s(\ell+1)}\|x_p\| \\
&\le \delta\|x_{T_2}\| + \delta\|x_p\| \\
&\le \delta\|x_{T_2}\| + \frac{\delta^2}{1-\delta}\|x_{T_2}\|. \qquad (6.5)
\end{aligned}$$

In the last inequality we used (6.3).

The remainder of the proof consists of two cases depending on the size of T_2.

First, we consider the case $K_2 > s$. Denote $T_3 := \{n_1, \ldots, n_s\}$ such that $T_3 \subseteq T_2$ and

$$|x_{n_1}| \ge \ldots \ge |x_{n_s}| \ge \sup_{n\in T_2, n\notin T_3} |x_n|.$$

The definition of the Λ' implies

$$\begin{aligned}
\|\Phi^*_{\Lambda'}r^\ell\| &\ge \|\Phi^*_{T_3}r^\ell\| = \|\Phi^*_{T_3}(\Phi_{T_2}x_{T_2} - \Phi_{\Lambda^\ell}x_p)\| \\
&\ge \|\Phi^*_{T_3}(\Phi_{T_3}x_{T_3} + \Phi_{T_2\setminus T_3}x_{T_2\setminus T_3} - \Phi_{\Lambda^\ell}x_p)\| \\
&\ge \|\Phi^*_{T_3}\Phi_{T_3}x_{T_3}\| - \|\Phi^*_{T_3}\Phi_{T_2\setminus T_3}x_{T_2\setminus T_3}\| - \|\Phi^*_{T_3}\Phi_{\Lambda^\ell}x_p\|. \qquad (6.6)
\end{aligned}$$

By the definition of T_3

$$\|x_{T_3}\| \geq \Big(\frac{s}{K_2}\Big)^{\frac{1}{2}} \|x_{T_2}\|, \tag{6.7}$$

and

$$\|x_{T_2 \setminus T_3}\| \leq \Big(1 - \frac{s}{K_2}\Big)^{\frac{1}{2}} \|x_{T_2}\|. \tag{6.8}$$

We estimate the last three terms in (6.6). Applying (4.5) and Lemma 4.2 we obtain

$$\begin{aligned} \|\Phi_{T_3}^* \Phi_{T_3} x_{T_3}\| &\geq (1-\delta_s)\|x_{T_3}\| \\ &\geq (1-\delta)\|x_{T_3}\| \\ &\geq (1-\delta)\Big(\frac{s}{K_2}\Big)^{\frac{1}{2}} \|x_{T_2}\|, \end{aligned} \tag{6.9}$$

$$\begin{aligned} \|\Phi_{T_3}^* \Phi_{T_2 \setminus T_3} x_{T_2 \setminus T_3}\| &\leq \delta_{K_2} \|x_{T_2 \setminus T_3}\| \\ &\leq \delta \|x_{T_2 \setminus T_3}\| \\ &\leq \delta \Big(1 - \frac{s}{K_2}\Big)^{\frac{1}{2}} \|x_{T_2}\|, \end{aligned} \tag{6.10}$$

and

$$\|\Phi_{T_3}^* \Phi_{\Lambda^\ell} x_p\| \leq \delta_{(\ell+1)s} \|x_p\| \leq \frac{\delta^2}{1-\delta} \|x_{T_2}\|. \tag{6.11}$$

Substituting (6.9),(6.10) and (6.11) in (6.6) we obtain

$$\|\Phi_{\Lambda'}^* r^\ell\| \geq (1-\delta)\Big(\frac{s}{K_2}\Big)^{\frac{1}{2}} \|x_{T_2}\| - \delta\Big(1 - \frac{s}{K_2}\Big)^{\frac{1}{2}} \|x_{T_2}\| - \frac{\delta^2}{1-\delta}\|x_{T_2}\|. \tag{6.12}$$

Comparing (6.5) and (6.12) we get

$$\delta\Big(\Big(\frac{s}{K_2}\Big)^{\frac{1}{2}} + \Big(1 - \frac{s}{K_2}\Big)^{\frac{1}{2}}\Big) + \delta + \frac{2\delta^2}{1-\delta} \geq \Big(\frac{s}{K_2}\Big)^{\frac{1}{2}},$$

which implies

$$(1+\sqrt{2})\delta + \frac{2\delta^2}{1-\delta} \geq \Big(\frac{s}{K_2}\Big)^{\frac{1}{2}}.$$

If $\delta < \frac{\sqrt{s}}{(2+\sqrt{2})\sqrt{K}} < \frac{1}{3}$, then

$$(1+\sqrt{2})\delta + \frac{2\delta^2}{1-\delta} < (2+\sqrt{2})\delta < \Big(\frac{s}{K}\Big)^{\frac{1}{2}} < \Big(\frac{s}{K_2}\Big)^{\frac{1}{2}}. \tag{6.13}$$

This contradiction implies $\Lambda' \cap T \neq \emptyset$.

Second, we consider the case $K_2 \leq s$. In this case $|T_2| \leq |\Lambda'|$. Using this and the definition of Λ' we get

$$\begin{aligned}
\|\Phi_{\Lambda'}^* r^\ell\| &\geq \|\Phi_{T_2}^* r^\ell\| = \|\Phi_{T_2}^*(\Phi_{T_2} x_{T_2} - \Phi_{\Lambda^\ell} x_p)\| \\
&\geq \|\Phi_{T_2}^* \Phi_{T_2} x_{T_2}\| - \|\Phi_{T_2}^* \Phi_{\Lambda^\ell} x_p\| \\
&\geq (1 - \delta_{K_2})\|x_{T_2}\| - \delta_{s+K_2}\|x_p\| \\
&\geq (1 - \delta)\|x_{T_2}\| - \frac{\delta^2}{1-\delta}\|x_{T_2}\|. \qquad (6.14)
\end{aligned}$$

Comparing (6.5) and (6.14), we obtain

$$\delta\|x_{T_2}\| + \frac{\delta^2}{1-\delta}\|x_{T_2}\| > (1-\delta)\|x_{T_2}\| - \frac{\delta^2}{1-\delta}\|x_{T_2}\|$$

or equivalently

$$2\delta + \frac{2\delta^2}{1-\delta} > 1, \qquad (6.15)$$

The above inequality does not hold for $\delta \leq 1/3$. This contradiction implies that in the second case we also have $\Lambda' \cap T \neq \emptyset$.

These two cases show that if $\delta < \frac{\sqrt{s}}{(2+\sqrt{2})\sqrt{K}}$ then $\Lambda' \cap T \neq \emptyset$. Therefore, OMMP picks at least one correct index at every iteration. This means $T \subseteq \Lambda^K$. Thus

$$\hat{x}_{\Lambda^K} = x^K = \operatorname*{argmin}_z \|y - \Phi_{\Lambda^K} z\| = x_{\Lambda^K}$$

and $\hat{x}_{(\Lambda^K)^C} = x_{(\Lambda^K)^C} = 0$. Then we conclude that $\hat{x} = x$.

Acknowledgements. We are grateful to J.L. Nelson for help in preparation of the paper. This research was supported by the National Science Foundation Grant DMS-0906260.

References

[1] Candés, E., Romberg, J., Tao, T.: Robust Uncertainty Principles: Exact Signal Reconstruction From Highly Incomplete Frequency Information. IEEE Trans. Inform. Theory. **52**(2), 489–509 (2006)

[2] Candés, E., Romberg, J., Tao, T.: Stable signal recovery from incomplete and inaccurate measurements. Commun. Pure Appl. Math. **59**(8), 1207–1223 (2006)

[3] Candés, E., Tao, T.: Decoding by Linear Programming. IEEE Trans. Inform. Theory. **51**(12), 4203–4215 (2005)

[4] Cohen, A., Dahmen, W., DeVore, R.A.: Compressed sensing and k-term approximation. J. Amer. Math. Soc. **22**, 211–231(2009)

[5] Dai, W., Milenkovic, O.: Subspace pursuit for compressive sensing: Closing the gap between performance and complexity. IEEE Trans. Inform. Theory. **55**(5), 2230–2249 (2009)

[6] Davenport, M., Wakin, M.: Analysis of Orthogonal Matching Pursuit using the Restricted Isometry Property. Preprint (2009)

[7] DeVore, R.A., Temlyakov, V.N.: Some remarks on greedy algorithms. Adv. Comput. Math. **5**, 173–187 (1996)

[8] Donoho, D.L.: Compressed Sensing. IEEE Trans. Inform. Theory. **52**(4), 1289–1306 (2006)

[9] Donoho, D.L., Elad, M., Temlyakov, V.N.: Stable recovery of sparse overcomplete representations in the presence of noise. IMI Preprint (2004)

[10] Donoho, D.L., Elad, M., Temlyakov, V.N.: Stable recovery of sparse overcomplete representations in the presence of noise. IEEE Trans. Inform. Theory. **52**, 6–18 (2006)

[11] Donoho, D.L., Elad, M., Temlyakov, V.N.: On Lebesgue-type inequalities for greedy approximation. J. Approx. Theory. **147**(2), 185–195 (2007)

[12] Gilbert, A.C., Muthukrishnan, S., Strauss, M.J.: Approximation of functions over redundant dictionaries using coherence. In Proc. 14th Annual ACM-SIAM Symposium on Discrete Algorithms. pp. 243–252 (2002)

[13] Needell, D, Tropp, J.: COSAMP: Iterative signal recovery from incomplete and inaccurate samples. Appl. Comput. Harmonic Analysis. **26**(3), 301–321 (2009)

[14] Temlyakov, V.N.: Weak greedy algorithms. Adv. Comput. Math. **12**, 213–227 (2000)

[15] Temlyakov, V.N. Greedy approximation. Acta Numerica. pp. 235–409 (2008)

[16] Tropp, J.: Greedy is good: Algorithmic results for sparse approximation. IEEE Trans. Inform. Theory. **50**(10), 2231–2242 (2004)

[17] Tropp, J., Gilbert, A.: Signal Recovery From Random Measurements Via Orthogonal Matching Pursuit. IEEE Trans. Inform. Theory. **53**(12), 4655–4666 (2007)