Applications of Orthogonal Matching Pursuit in Compressed Sensing*

Long Jingfan[1], **Wei Xiujie**[2], **Ye Peixin**[2]
[1]Beijing Information Science and Technology University Bejing, China
[2]School of Mathematics and LPMC Nankai University Tianjin 300071, China
yepx@nankai.edu.cn

Abstract In this paper, we study the efficiency of compressed sensing by using Orthogonal Matching Pursuit (OMP). We show that if a Matrix Φ has coherence less than $\frac{1}{20K^{0.8}}$ and satisfies the Restricted Isometry Property (RIP) of order $[CK^{1.2}]$ with constant $\delta = cK^{-0.2}$, then a K-sparse signal x can be recovered from $y = \Phi x$ via Orthogonal Matching Pursuit in at most optimal approximation on the first $[CK^{1.2}]$ iterations.

Keywords: Orthogonal Matching Pursuit (OMP); Restricted Isometry Property (RIP); Compressed Sensing (CS); K-sparse signal.

1. Introduction.

Compressed Sensing is a new paradigm in signal and image processing. It seeks to faithfully capture a signal or image with the fewest number of measurements, cf. [1-9]. Rather than model a signal as a bandlimited function or an image as a pixel array, it models both of these as a sparse vector in some representation system. This model fits well real world signals and images. For example, images are well approximated by a sparse wavelet decomposition. One replaces the bandlimited model of signals by the assumption that the signal is sparse or compressible with respect to some basis or dictionary of waveform and enlarges the concept of sample to include the applications of any linear functional. Given this model, how should we design a sensor to capture the signal with the fewest number of measurements? we will focus on the discrete sensing problem where we are given a vector in R^N with N large and we wish to capture it through n measurements given by inner products with fixed vectors. Such a measurement system can be represented by an $n \times N$ matrix A. The vector $y = Ax$ is the vector of n measurements we make of x. The information that y holds about x is extracted through a decoder Δ . So $\Delta(x)$ should be designed to be a faithful approximation to x. The fact that this may be possible is embedded in some old mathematical results in functional analysis, geometry and approximation, cf. [10-14]. We will discuss what are the best matrices to use in sensing and how to extract the information contained in the sensed vector y. We shall focus on the relation between the number of samples we take of a signal and how well we can approximate the signal.

In this paper, we study the efficiency of compressed sensing via Orthogonal Matching Pursuit (OMP). Let us begin with the demonstration of the use of greedy algorithms in the compressed sensing problem. The emerging theory of compressed sensing (CS) has provided a new framework for signal acquisition [1], [3], [8]. Now we recall some basic concepts of CS. Suppose that $1 \le K \le n \le N$ and $0 < \delta < 1$. A signal $x = (x_j)_{j=1}^N \in \mathbb{R}^N$ is said to be K-sparse if x has at most K nonzero coordinates. An $n \times N$ matrix Φ is said to satisfies Restricted Isometry Property (RIP) ([4]) of order K with isometry constant δ if, for all K-sparse vectors x, we have

$$(1-\delta)\|x\|^2 \le \|\Phi x\| \le (1+\delta)\|x\|^2.$$

Suppose that $\phi_1, \phi_2, ..., \phi_N$ are the columns of a matrix Φ, we assume that $\|\phi_i\| = 1,\ 1 \le i \le N$. The coherence of Φ is defined as

$$\mu(\Phi) := \sup_{\phi_i, \phi_j \in \Phi, i \neq j} |\langle \phi_i, \phi_j \rangle|.$$

Let Φ be a $M \times N$ matrix ($M < N$). The basic problem in CS is to construct a stable and fast algorithm for recovery a signal $x \in \mathbb{R}^d$ (K-sparse) from measurements $y = \Phi x \in \mathbb{R}^M$ and to determine (M, N, K) for which such algorithms exist.

Candes and Tao [4] proved that Basic Pursuit (BP)

$$\widehat{x}(y) = \mathrm{argmin}\{|z|_1 : \Phi z = y\}$$

can provide the exact recovery of arbitrary K-sparse $x \in \mathbb{R}^N$ by $M = O(K \log(N/K))$ measurements.

In this article we study signal recovery via Orthogonal Matching Pursuit (OMP). Although theoretical results for OMP are essentially worse than for BP, its computational simplicity allows OMP to achieve very good result in practise [19]. We give the definition of the Orthogonal Matching Pursuit in terms of the theory of transmission of signals.

* This work is supported by National Nature Science Foundation under Grant 10971251, 11271199 , 11101220 and supported by the Program for new century excellent talents in University of China (NCET-10-0513). Corresponding author Ye Peixin, E-mail: yepx@nankai.edu.cn

Algorithm: Orthogonal Matching Pursuit

Input: Φ, y.
Initiation: $r^0 := y,\ x^0 := 0,\ \Lambda^0 := \emptyset,\ l = 0$.
Iteration: Define $\Lambda^{l+1} := \Lambda^l \cup \operatorname{argmax}_i |\langle r^l, \phi_i\rangle|$,
$x^{l+1} := \operatorname{argmin}_{z:\ \operatorname{supp}(z)\in\Lambda^{l+1}} \|y - \Phi z\|$
$r^{l+1} := y - \Phi x^{l+1}$.
If $r^{l+1} = 0$, stop. Otherwise, we set $l := l+1$ and begin a new iteration.
Output: If the algorithm stops at the l-th iteration, output is $\widehat{x} = x_l$.

Now let us recall some results on recovery of sparse signals by the OMP, it is well known that if

$$K < \frac{1}{2}(\mu(\Phi)^{-1} + 1) \tag{1}$$

then OMP will recover arbitrary K-sparse signal x from $y = \Phi x$ in exactly K iterations. Temlyakov and Zheltov [18] showed that the strict inequality $<$ in (1) cannot be replaced by the wide inequality $\leq$. The stability of recovery via OMP in the term of coherence of has been studied in [11], [19], [9], [10], [18], [15]. Recently M. Davenport and M. wakin [6], and E. Liu and V.N. Temlyakov [14] showed that if Φ satisfies RIP of order $K+1$ with isometry constant

$$\delta = \frac{1}{3K^{1/2}}(see\ [6]),\ \ \delta = \frac{1}{(1+2^{1/2})K^{1/2}}(see\ [14]),$$

then OMP recovers arbitrary K-sparse signal $x \in \mathbb{R}^N$ from $y = \Phi x$ in exactly K iterations.

To compare these results we recall estimates on coherence and RIP for normalized random Bernoulli matrices Φ (each entry is $M^{-1/2}$ with probability 1/2). For rather big c_μ we have with high probability that

$$\mu(\Phi) \leq c_\mu M^{-1/2} \log^{1/2} N. \tag{2}$$

R. Baraniuk, M. Davenport R. Devore and M. Wakin [2] (see also earlier B.S. Kashins work [13]) showed that random Bernoulli matrix Φ with high probability satisfy RIP of order K with isometry constant $\delta > 0$ with

$$K \asymp \frac{\delta^2 M}{\log(N/M)}.$$

Thus both results require $M = O(K^2)$ $(M \lesssim K^2 \log N)$ measurements for recovery of K-sparse signal. The aim of this article to show that OMP can recover sparse signals by essentially less number of measurements. This result is a improvement of recent results of Eugene Livshitz in [16].

Theorem 1. *There exist absolute constants* $C = 1.6 \times 10^4 > 0$ *and* $c = 10^{-5} > 0$ *such that if* Φ *satisfies the RIP of order* $[CK^{1.2}]$ with isometry *constant* $\delta = cK^{-0.2}$ *and has coherence* $\mu(\Phi) \leq 1/(20K^{0.8})$, *then for any* $K-$*sparse* $x \in \mathbb{R}^N$, *OMP will recover* x *exactly from* $y = \Phi x$ *in at most* $[CK^{1.2}]$ *iterations.*

Theorem 1 together with (1) and (2), imply the estimate: given a fixed random $M \times N$ Bernoulli matrix Φ, the recovery by OMP will be exact (with high probability) for all K-sparse $x \in \mathbb{R}^N$ where

$$K \lesssim (\frac{M}{\log N})^{5/8};$$

in other words, to recovery a K-sparse signals by the OMP it will suffice to perform M measurements, where

$$M \lesssim K^{1.6} \log N.$$

2. Auxiliary lemmas.

We use several standard lemmas to prove Theorem 1. First we use the following two results on the convergence rate of the OMP.

Theorem A. ([7]) *Let* $y = \Phi x$. *Then, for any* $l > 1$, *we have*

$$\|r^l\| \leq \|x\|_1 l^{-1/2}.$$

Theorem B. ([17]) *For any* l, $1 \leq l \leq \frac{1}{20\mu(\Phi)}$, $\forall \varepsilon > 0$, *we have*

$$\|r^{2l}\| \leq 2.47(1+\varepsilon)\sigma_l(y, \Phi)$$

For $l \geq 0$, we set

$$z^l := x - x^l.$$

Then, by the definition of the Orthogonal Matching Pursuit,

$$r^l = y - \Phi x^l = \Phi x - \Phi x^l = \Phi z^l, l > 0. \tag{3}$$

Suppose that

$$x = (x_1, \ldots, x_N),\ z^l = (z_1^l, \ldots, z_N^l),\ l \geq 0.$$

We set

$$V_0 = \operatorname{supp} x,\ \sharp V_0 \leq K. \tag{4}$$

By $x|_V,\ V \subset V_0$, we define $(\widetilde{x_1}, \ldots, \widetilde{x_N})$ of $\mathbb{R}^N$ with $\widetilde{x}_i = x_i,\ i \in V$, and $\widetilde{x}_i = 0,\ i \notin V$. For any $V \subset V_0$, we define

$$R(V) = \sum_{i\in V} x_i^2.$$

Lemma 1. *Assume that* $l + K \leq CK^{1.2}$. *Then we have*

$$\sum_{i\in\Lambda^l} (z_i^l)^2 \leq \frac{2\delta}{1-\delta} R(V_0 \backslash \Lambda^l), \tag{5}$$

$$R(V_0\backslash\Lambda^l) \leq \frac{1}{1-\delta}\|r^l\|^2. \tag{6}$$

Proof: We clearly have $|z^l|_0 \leq |x|_0 + |x^l|_0 \leq K + l \leq CK^{1.2}$, so from the RIP and (3), we get

$$(1-\delta)\sum_{i=1}^N (z_i^l)^2 \leq \|\Phi z^l\|^2 = \|r^l\|^2 \leq (1+\delta)\sum_{i=1}^N (z_i^l)^2. \tag{7}$$

On the other hand, using the definition of $R(\cdot)$ and the RIP for $x|_{V_0\setminus\Lambda^l}$,

$$\|x|_{V_0\setminus\Lambda^l}\|^2 = R(V_0\setminus\Lambda^l)$$

$$(1-\delta)R(V_0\setminus\Lambda^l) \le \|\Phi(x|_{V_0\setminus\Lambda^l})\|^2 \le (1+\delta)R(V_0\setminus\Lambda^l). \quad (8)$$

From the definition of the Orthogonal Matching Pursuit, we have

$$\|\Phi z^l\|^2 = \|r^l\|^2 \le \|\Phi(x|_{V_0\setminus\Lambda^l})\|^2.$$

Therefore using (7) and (8) we have

$$(1-\delta)\sum_{i=1}^{N}(z_i^l)^2 \le \|r^l\|^2 \le \|\Phi(x|_{V_0\setminus\Lambda^l})\|^2 \le (1+\delta)\sum_{i=1}^{N}(z_i^l)^2.$$

$$(1-\delta)\big(\sum_{i\in\Lambda^l}(z_i^l)^2 + \sum_{i\in V_0\setminus\Lambda^l}(z_i^l)^2\big) \le (1+\delta)R(V_0\setminus\Lambda^l).$$

$$\sum_{i\in\Lambda^l}(z_i^l)^2 + R(V_0\setminus\Lambda^l) \le \frac{1+\delta}{1-\delta}R(V_0\setminus\Lambda^l).$$

$$\sum_{i\in\Lambda^l}(z_i^l)^2 \le \big(\frac{1+\delta}{1-\delta}-1\big)R(x|_{V_0\setminus\Lambda^l}) = \frac{2\delta}{1-\delta}R(V_0\setminus\Lambda^l).$$

This completes the proof of (5). From (7) it follows that

$$R(V_0\setminus\Lambda^l) = \sum_{i\in V_0\setminus\Lambda^l}(z_i^l)^2 \le \sum_{i=1}^{N}(z_i^l)^2 \le \frac{1}{1-\delta}\|r^l\|^2.$$

□

Given an increasing sequence $0 = l_0 < l_1 < \cdots < l_s,\ s \ge 1$, we denote

$$V_k := V_0\setminus\Lambda^{l_k},\ R_k := R(V_k),\ 0 \le k \le s. \quad (9)$$

Lemma 2. *Suppose that* $l_k + K \le CK^{1.2},\ 1 \le k \le s$. *Then for arbitrary* $p \in \mathbb{N}$, *we have*

$$\|r^{l_k+p}\|^2 \le \frac{R_k}{p}\big(\frac{4\delta}{1-\delta}CK^{1.2} + 2K\big).$$

Proof: Since $r^{l_k} = \Phi z^{l_k}$, it follows by Theorem A that

$$\|r^{l_k+p}\|^2 \le \frac{\|z^{l_k}\|_1^2}{p}. \quad (10)$$

So, in order to prove Lemma 2 it suffices to estimate

$$\|z^{l_k}\|^2 = \big(\sum_{i=1}^{N}|z_i^{l_k}|\big)^2 = \big(\sum_{i\in V_0\cup\Lambda^{l_k}}|z_i^{l_k}|\big)^2$$

$$\le 2\Big(\big(\sum_{i\in V_0\setminus\Lambda^{l_k}}|z_i^{l_k}|\big)^2 + \big(\sum_{i\in\Lambda^{l_k}}|z_i^{l_k}|\big)^2\Big).$$

Applying (9) and (4), we obtain

$$\begin{aligned}\big(\sum_{i\in V_0\cup\Lambda^{l_k}}|z_i^{l_k}|\big)^2 &= \big(\sum_{i\in V_0\cup\Lambda^{l_k}}|x_i^l|\big)^2 = \big(\sum_{i\in V_k}|x_i^l|\big)^2\\ &\le \sharp V_k\big(\sum_{i\in V_k}|x_i^l|\big)^2 \le \sharp V_0 R_k\\ &\le R_k K. \end{aligned} \quad (11)$$

Using (5) from Lemma 1 we get

$$\big(\sum_{i\in\Lambda^{l_k}}|z_i^{l_k}|\big)^2\big) \le \sharp\Lambda^{l_k}\sum_{i\in\Lambda^{l_k}}(z_i^{l_k})^2 = l_k\sum_{i\in\Lambda^{l_k}}(z_i^{l_k})^2$$

$$\le CK^{1.2}\big(\frac{2\delta}{1-\delta}\big)R(V_0\setminus\Lambda^{l_k}) = CK^{1.2}\big(\frac{2\delta}{1-\delta}\big)R_k.$$

Combining with (11) we obtain the desirable inequality

$$\|z^{l_k}\|^2 \le R_k\big(\frac{4\delta}{1-\delta}CK^{1.2} + 2K\big).$$

This and inequality (10) concludes the proof of Lemma 2. □

Lemma 3. *Let* $1 \le p \le K^{0.8}$ *and* $l_k + 2p \le CK^{1.2},\ 1 \le k \le s$. *Then for an arbitrary* $W \subset V_k$ *such that* $\sharp W = p$ *we have*

$$R(V_k\setminus\Lambda^{l_k+2p}) \le 6.16(R(V_k\setminus W) + \frac{2\delta}{1-\delta}R_k).$$

Proof: According to RIP, (3), (5) and (9), it is found that

$$\begin{aligned}\big(\sigma_p(r^{l_k})\big)^2 &\le \|r^{l_k} - \Phi(x|_W)\|^2 = \|\Phi(z^{l_k}) - \Phi(z^{l_k}|_W)\|^2\\ &= \|\Phi(z^{l_k} - z^{l_k}|_W)\|^2 \le (1+\delta)\sum_{1\le i\le N,\ i\notin W}(z^{l_k})^2\\ &\le (1+\delta)\big(\sum_{i\in V_k\setminus W}(z^{l_k})^2 + \sum_{1\le i\le N,\ i\notin V_k}(z^{l_k})^2\big)\\ &\le (1+\delta)\big(\sum_{i\in V_k\setminus W}(x_i)^2 + \sum_{i\in\Lambda^{l_k}}(z^{l_k})^2\big)\\ &\le (1+\delta)\big(R(V_k\setminus W) + \frac{2\delta}{1-\delta}R(V_0\setminus\Lambda^{l_k})\big)\\ &\le (1+\delta)\big(R(V_k\setminus W) + \frac{2\delta}{1-\delta}R_k\big).\end{aligned}$$

Since we have

$$p \le K^{0.8} \le 1/(20\mu(\Phi)),$$

we can apply Theorem B and get

$$\|r^{l_k+2p}\| \le 2.47(1+\varepsilon)\sigma_p(r^{l_k}),\ \forall\varepsilon > 0.$$

Using (6) from lemma 1 we obtain

$$\begin{aligned}R(V_k\setminus\Lambda^{l_k+2p}) = R(V_0\setminus\Lambda^{l_k+2p}) &\le \frac{1}{1-\delta}\|r^{l_k+2p}\|^2\\ &\le (2.47(1+\varepsilon))^2(\sigma_p(r^{l_k}))^2\\ &\le \frac{1+\delta}{1-\delta}(2.47(1+\varepsilon))^2\big(R(V_k\setminus W) + \frac{2\delta}{1-\delta}R_k\big)\\ &\le 6.16\big(R(V_k\setminus W) + \frac{2\delta}{1-\delta}R_k\big).\end{aligned}$$

□

References

[1] R.G. Baraniuk, Compressive sensing,*IEEE Signal Processing Magazine* 24 (2007), 118-121.

[2] R. Baraniuk, M. Davenport, R. DeVore and M. Wakin, A simple proof of the restricted isometry property for random matrices, *Constr. Approx.* 28 (2008), 253-263.

[3] E.J. Candes Compressive sampling, *Int. Congress of Mathematics.* vol. 3 (2006), 1433-1452

[4] E.J. Candes and T. Tao, Decoding by linear programming, *IEEE Trans. Inform. Theory* 51 (2005), 4203-4215.

[5] W. Dai and O. Milenkovic, Subspace pursuit for compressive sensing signal reconstruction, *IEEE Trans. Inform. Theory* 55 (2009), 2230-2249.

[6] M. A. Davenport and M. B. Wakin, Analysis of orthogonal matching pursuit using the restricted isometry property, *IEEE Trans. Inform. Theory* 56 (2010), 4395-4401.

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

[8] D.L. Donoho, Compressed sensing, *IEEE Trans. Inform. Theory* 52 (2006), 1289-1306.

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

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

[11] A.C. Gilbert, M. Muthukrishnan, J. Strauss, Approximation of functions over redundant dictionaries using coherence, *in: Proc. 14th Annu. ACM-SIAM Symp. Discrete Algorithms.* (2003) 234-252.

[12] R. Gilbonval, M. Nielsen, On the strong uniqueness of highly sparse expansions from redundant dictionaries, *in: Proc. Int Conf. Independent Component Anal.* (ICA'04), (2004).

[13] B. Kashin, The widths of certain finite dimensional sets and classes of smooth functions,. *Izvestia* 41 (1977), 334-351.

[14] E. Liu and V.N. Temlyakov, The Orthogonal super greedy algorithm and applications in compressed sensing, *IEEE Trans. Inform Theory* 58 (2012) 2040-2047.

[15] E.D. Livshitz, On the optimality of Orthogonal Greedy Algorithm for μ-coherent dictionaries. *J. Approx. Theory* 164 (2012) 668-681.

[16] E.D. Livshits, On the efficiency of the Orthogonal Matching Pursuit in compressed sensing,*Mat. Sb.* 203 (2012) 33-44.

[17] Y. Peixin, W. Xiujie, Lebesgue-type inequality for Orthogonal Matching Pursuit for μ-coherent dictionaries, *TELKOMNIKA Indonesian Journal of Electrical Engineering,* in press.

[18] V.N. Temlyakov, P. Zheltov, On performance of greedy algorithms, *J. Approx. Theory* 163 (2011) 1134-1145.

[19] J.A. Tropp, Greedy is good: algorithmic results for sparse approximation, *IEEE Trans. Inform. Theroy* 50 (2004) 2231-2242.