Linear Convergence of Adaptively Iterative Thresholding Algorithms for Compressed Sensing

Yu Wang, Jinshan Zeng*, Zhimin Peng, Xiangyu Chang, and Zongben Xu

Abstract—**This paper studies the convergence of the adaptively iterative thresholding (AIT) algorithm for compressed sensing. We first introduce a generalized restricted isometry property (gRIP). Then we prove that the AIT algorithm converges to the original sparse solution at a linear rate under a certain gRIP condition in the noise free case. While in the noisy case, its convergence rate is also linear until attaining a certain error bound. Moreover, as by-products, we also provide some sufficient conditions for the convergence of the AIT algorithm based on the two well-known properties, i.e., the coherence property and the restricted isometry property (RIP), respectively. It should be pointed out that such two properties are special cases of gRIP. The solid improvements on the theoretical results are demonstrated and compared with the known results. Finally, we provide a series of simulations to verify the correctness of the theoretical assertions as well as the effectiveness of the AIT algorithm.**

Index Terms—**restricted isometric property, coherence, iterative hard thresholding, SCAD, compressed sensing, sparse optimization**

I. INTRODUCTION

Let $A \in \mathbf{R}^{m \times n}$, $b \in \mathbf{R}^m$ and $x \in \mathbf{R}^n$. Compressed sensing [1], [2] solves the following constrained ℓ_0-minimization problem

$$\min_{x \in \mathbf{R}^n} \|x\|_0 \quad \text{s.t. } b = Ax + \epsilon, \ \|\epsilon\|_2 \le \sigma \qquad (1)$$

where $\epsilon \in \mathbf{R}^m$ is the measurement noise, $\sigma \in \mathbf{R}$ is the noise variance and $\|x\|_0$ denotes the number of the nonzero components of x. Due to the NP-hardness of problem (1) [3], approximate methods including the greedy method and relaxed method are introduced. The greedy method approaches the sparse solution by successively alternating one or more components that yield the greatest improvement in quality [3]. These algorithms include iterative hard thresholding (IHT) [4], accelerated hard thresholding (AHT) [5], ALPS [6], hard thresholding pursuit (HTP) [7], CLASH [8], OMP [10], [11], StOMP [12], ROMP [13], CoSaMP [14] and SP [15]. The greedy algorithms can be quite efficient and fast in many applications, especially when the signal is very sparse.

Y. Wang, J.S. Zeng and Z.B. Xu are with the Institute for Information and System Sciences, School of Mathematics and Statistics, Xi'an Jiaotong University, Xi'an 710049 (email: shifwang@gmail.com, jsh.zeng@gmail.com,zbxu@mail.xjtu.edu.cn).

Z.M. Peng is with the Department of Mathematics, University of California, Los Angeles (UCLA), Los Angeles, CA 90095, United States (email: zhimin.peng@math.ucla.edu).

X.Y. Chang is with the School of Management, Xi'an Jiaotong University, Xi'an 710049. (email: xiangyuchang@gmail.com).

∗ Corresponding author: Jinshan Zeng (jsh.zeng@gmail.com).

The relaxed method converts the combinatorial ℓ_0-minimization into a more tractable model through replacing the ℓ_0 norm with a nonnegative and continuous function $P(\cdot)$, that is,

$$\min_{x \in \mathbf{R}^n} P(x) \quad \text{s.t. } b = Ax + \epsilon, \ \|\epsilon\|_2 \le \sigma. \qquad (2)$$

One of the most important cases is the ℓ_1-minimization problem (also known as *basis pursuit* (BP)) [16] in the noise free case and *basis pursuit denoising* in the noisy case) with $P(x) = \|x\|_1$, where $\|x\|_1 = \sum_{i=1}^n |x_i|$ is called the ℓ_1 norm. The ℓ_1-minimization problem is a convex optimization problem that can be efficiently solved. Nevertheless, the ℓ_1 norm may not induce further sparsity when applied to certain applications [17], [18], [19], [20]. Therefore, many nonconvex functions were proposed as substitutions of the ℓ_0 norm. Some typical nonconvex examples include the ℓ_q $(0 < q < 1)$ norm [17], [18], [19], smoothly clipped absolute deviation (SCAD) [21] and minimax concave penalty (MCP) [22]. Compared with the ℓ_1-minimization model, the nonconvex relaxed models can often induce better sparsity and reduce the bias, while they are generally more difficult to solve.

The iterative reweighted method and regularization method are two main classes of algorithms to solve (2) when $P(x)$ is nonconvex. The iterative reweighted method includes the iterative reweighted least squares minimization (IRLS) [23], [24], and the iterative reweighted ℓ_1-minimization (IRL1) algorithms [20]. Specifically, the IRLS algorithm solves a sequence of weighted least squares problems, which can be viewed as some approximations to the original optimization problem. Similarly, the IRL1 algorithm solves a sequence of non-smooth weighted ℓ_1-minimization problems, and hence it is the non-smooth counterpart to the IRLS algorithm. However, the iterative reweighted algorithms are slow if the nonconvex penalty cannot be well approximated by the quadratic function or the weighted ℓ_1 norm function. The regularization method transforms problem (2) into the following unconstrained optimization problem

$$\min_{x \in \mathbf{R}^n} \{\|Ax - b\|_2^2 + \lambda P(x)\}, \qquad (3)$$

where $\lambda > 0$ is a regularization parameter. For some special penalties $P(x)$ such as the ℓ_q norms $(q = 0, 1/2, 2/3, 1)$, SCAD and MCP, an optimal solution of the model (3) is a fixed point of the following equation

$$x = H(x - sA^T(Ax - b)),$$

where $H(\cdot)$ is a componentwise thresholding operator which will be defined in detail in the next section and $s > 0$ is

a step size parameter. This yields the corresponding iterative thresholding algorithm ([19], [25], [26], [27], [28], [29])

$$x^{(t+1)} = H(x^{(t)} - sA^T(Ax^{(t)} - b)).$$

Compared to greedy methods and iterative reweighted algorithms, iterative thresholding algorithms have relatively lower computational complexities [30], [31], [32]. So far, most of theoretical guarantees of the iterative thresholding algorithms were developed for the regularization model (3) with fixed λ. However, it is in general difficult to determine an appropriate regularization parameter λ.

Some adaptive strategies for setting the regularization parameters were proposed. One strategy is to set the regularization parameter adaptively so that $\|x^{(t)}\|_0$ remains the same at each iteration. This strategy was first applied to the iterative hard thresholding algorithm (called Hard algorithm for short henceforth) in [33], and later the iterative soft thresholding algorithm[34] (called Soft algorithm for short henceforth) and the iterative half thresholding algorithm [19] (called Half algorithm for short henceforth). The convergence of Hard algorithm was justified when A satisfies the restricted isometry property (RIP) with $\delta_{3k^*} < \frac{1}{\sqrt{32}}$ [33], where k^* is the number of the nonzero components of the truely sparse signal. Later, Maleki [34] investigated the convergence of both Hard and Soft algorithms in terms of the coherence. Recently, Zeng et al. [35] generalized Maleki's results to a wide class of iterative thresholding algorithms. However, most of guarantees in [35] are coherence-based and focus on the noise free case with the step size equal to 1. While it has been observed that in practice, the AIT algorithm can have remarkable performances for noisy cases with a variety of step sizes. In this paper, we develop the theoretical guarantees of the AIT algorithm with different step sizes in both noise free and noisy cases.

A. Main Contributions

The main contributions of this paper are the following.

i) Based on the introduced gRIP, we give a new uniqueness theorem for the sparse signal (see Theorem 1), and then show that the AIT algorithm can converge to the original sparse signal at a linear rate (See Theorem 2). Specifically, in the noise free case, the AIT algorithm converges to the original sparse signal at a linear rate. While in the noisy case, it also converges to the original sparse signal at a linear rate until reaching an error bound.

ii) The tightness of our analyses is further discussed in two specific cases. The coherence based condition for Soft algorithm is the same as those required for both OMP and BP. Moreover, the RIP based condition for Hard algorithm is $\delta_{3k^*+1} < \frac{\sqrt{5}-1}{2} \approx 0.618$, which is better than the results in [7] and [9].

The rest of this paper is organized as follows. In section II, we describe the adaptively iterative thresholding (AIT) algorithm. In section III, we introduce the generalized restricted isometry property, and then provide a new uniqueness theorem. In section IV, we prove the convergence of the AIT algorithm. In section V, we compare the obtained theoretical results with some other known results. In section VI, we implement a series of simulations to verify the correctness of the theoretical results as well as the efficiency of the AIT algorithm. In section VII, we discuss many practical issues on the implementation of the AIT algorithm, and then conclude this paper in section VIII. All the proofs are presented in the Appendices.

Notations. We denote $\mathbf{N}$ and $\mathbf{R}$ as the natural number set and one-dimensional real space, respectively. For any vector $x \in \mathbf{R}^n$, x_i is the i-th component of x for $i = 1, \ldots, n$. For any matrix $A \in \mathbf{R}^{m \times n}$, A_i denotes the i-th column of A. x^T and A^T represent the transpose of vector x and matrix A respectively. For any index set $S \subset \{1, \ldots, n\}$, $|S|$ represents its cardinality. S^c is the complementary set, i.e., $S^c = \{1, \ldots, n\} \setminus S$. For any vector $x \in \mathbf{R}^n$, x_S represents the subvector of x with the components restricted to S. Similarly, A_S represents the submatrix of A with the columns restricted to S. We denote x^* as the original sparse signal with $\|x^*\|_0 = k^*$, and $I^* = \{i : |x_i^*| \neq 0\}$ is the support set of x^*. $\mathbf{I}_r \in \mathbf{R}^{r \times r}$ is the r-dimensional identity matrix. $sgn(\cdot)$ represents the signum function.

II. ADAPTIVELY ITERATIVE THRESHOLDING ALGORITHM

The AIT algorithm for (3) is the following

$$z^{(t+1)} = x^{(t)} - sA^T(Ax^{(t)} - b), \tag{4}$$

$$x^{(t+1)} = H_{\tau^{(t+1)}}(z^{(t+1)}), \tag{5}$$

where $s > 0$ is a step size and

$$H_{\tau^{(t+1)}}(x) = (h_{\tau^{(t+1)}}(x_1), \ldots, h_{\tau^{(t+1)}}(x_n))^T \tag{6}$$

is a componentwise thresholding operator. The thresholding function $h_\tau(u)$ is defined as

$$h_\tau(u) = \begin{cases} f_\tau(u), & |u| > \tau \\ 0, & \text{otherwise} \end{cases}, \tag{7}$$

where $f_\tau(u)$ is the *defining function*. In the following, we give some basic assumptions of the defining function, which were firstly introduced in [35].

Assumption 1. Assume that f_τ satisfies

1) **Odevity**. $f_\tau(u)$ is an odd function of u.
2) **Monotonicity**. $f_\tau(u) < f_\tau(v)$ for any $\tau \leq u < v$.
3) **Boundedness**. There exist two constants $0 \leq c_2 \leq c_1 \leq 1$ such that $u - c_1\tau \leq f_\tau(u) \leq u - c_2\tau$ for $u \geq \tau$.

Note that most of the commonly used thresholding functions satisfy Assumption 1. In Fig. 1, we show some typical thresholding functions including hard [27], soft [25] and half [19] thresholding functions for $\ell_0, \ell_1, \ell_{1/2}$ norms respectively, as well as the thresholding functions for $\ell_{2/3}$ norm [26] and SCAD penalty [21]. The corresponding boundedness parameters are shown in Table I.

TABLE I
BOUNDEDNESS PARAMETERS c FOR DIFFERENT THERSHOLDING FUNCTIONS

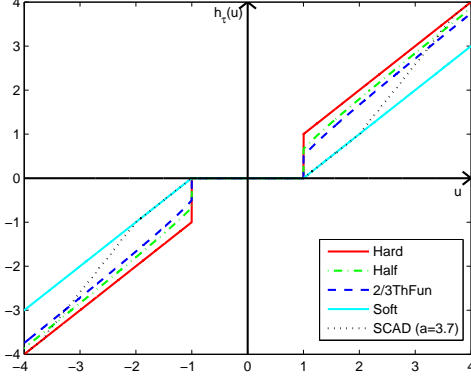|	$f_{\tau,*}$	$f_{\tau,0}$	$f_{\tau,1/2}$	$f_{\tau,2/3}$	$f_{\tau,1}$	$f_{\tau,SCAD}$
c_1		0	$\frac{1}{3}$	$\frac{1}{2}$	1	1
c_2		0	0	0	1	0

Fig. 1. Typical thresholding functions $h_\tau(u)$ with $\tau = 1$.

This paper considers a heuristic way for setting the threshold $\tau^{(t)}$, specifically, we let

$$\tau^{(t)} = |z^{(t)}_{[k+1]}|,$$

where $z^{(t)}_{[k+1]}$ is the $(k+1)$-th largest component of $z^{(t)}$ in magnitude and k is the *specified sparsity level*, $[k+1]$ denotes the index of this component. We formalise the AIT algorithm as in Algorithm 1.

Algorithm 1: Adaptively Iterative Thresholding Algorithm

Initialization: Normalize A such that $\|A_j\|_2 = 1$ for $j = 1, \ldots, n$.
Given a sparsity level k, a step size $s > 0$ and an initial point $x^{(0)}$.
Let $t := 0$;
Step 1: Calculate $z^{(t+1)} = x^{(t)} - sA^T(Ax^{(t)} - b)$;
Step 2: Set $\tau^{(t+1)} = |z^{(t+1)}_{[k+1]}|$ and I^{t+1} as the index set of
 the largest k components of $z^{(t+1)}$ in magnitude;
Step 3: Update: if $i \in I^{t+1}$, $x^{(t+1)}_i = f_{\tau^{(t+1)}}(z^{(t+1)}_i)$, otherwise
 $x^{(t+1)}_i = 0$;
Step 4: $t = t + 1$ and repeat ***Steps 1-3*** until convergence.

Remark 1. *At the $(t+1)$-th iteration, the AIT algorithm yields a sparse vector $x^{(t+1)}$ with k nonzero components. The sparsity level k is a crucial parameter for the performance of the AIT algorithm. When $k \geq k^*$, the results will get better as k decreases. Once $k < k^*$, the AIT algorithm fails to find the original sparse solution. Thus, k should be specified as an upper bound estimate of k^*.*

Remark 2. *In Algorithm 1, the columns of matrix A are required to be normalized. Such operation is only for a clearer definition of the following introduced generalized restricted isometry property (gRIP) and more importantly, better theoretical analyses. However, as shown in Section VII B, this requirement is generally not necessary for the use of the AIT algorithm in the perspective of the recovery performance. We will conduct a series of experiments in Section VII B for a detailed explanation.*

III. GENERALIZED RESTRICTED ISOMETRY PROPERTY

This section introduces the generalized restricted isometry property (gRIP) and then gives the uniqueness theorem.

Definition 1. *For any matrix $A \in \mathbf{R}^{m \times n}$, and a constant pair (p, q) where $p \in [1, \infty)$, $q \in [1, \infty]$ and $\frac{1}{p} + \frac{1}{q} = 1$, then the (k, p, q)-generalized restricted isometry constant (gRIC) $\beta_{k,p,q}$ of A is defined as*

$$\beta_{k,p,q} = \sup_{S \subset \{1,\ldots,n\}, |S| \leq k} \sup_{x \in \mathbf{R}^{|S|} \setminus \{0\}} \frac{\|(\mathbf{I}_{|S|} - A_S^T A_S)x\|_q}{\|x\|_p}. \tag{8}$$

We will show that the introduced gRIP satisfies the following proposition.

Proposition 1. *For any positive constant pair (p, q) with $\frac{1}{p} + \frac{1}{q} = 1$, the generalized restricted isometric constant $\beta_{k,p,q}$ associated with A and k must satisfy*

$$\frac{1}{3}\beta_{k,p,q} \leq \sup_{z \in \mathbf{R}^n \setminus \{0\}, \|z\|_0 \leq k} \frac{|z^T(A^TA - \mathbf{I}_n)z|}{\|z\|_p^2} \leq \beta_{k,p,q}. \tag{9}$$

The proof of this proposition is presented in Appendix A. It can be noted that the gRIP closely relates to the coherence property and restricted isometry property (RIP), whose definitions are listed in the following.

Definition 2. *For any matrix $A \in \mathbf{R}^{m \times n}$, the coherence of A is defined as*

$$\mu = \max_{i \neq j} \frac{|\langle A_i, A_j \rangle|}{\|A_i\|_2 \cdot \|A_j\|_2}, \tag{10}$$

where A_i denotes the i-th column of A for $i = 1, \ldots, n$.

Definition 3. *For any matrix $A \in \mathbf{R}^{m \times n}$, given $1 \leq k \leq n$, the restricted isometry constant (RIC) of A with respect to k, δ_k, is defined to be the smallest constant δ such that*

$$(1 - \delta)\|z\|_2^2 \leq \|Az\|_2^2 \leq (1 + \delta)\|z\|_2^2, \tag{11}$$

for all k-sparse vector, i.e., $\|z\|_0 \leq k$.

By Definition 3, RIC can also be written as:

$$\delta_k = \sup_{z \in \mathbf{R}^n \setminus \{0\}, \|z\|_0 \leq k} \frac{|z^T(A^TA - \mathbf{I}_n)z|}{\|z\|_2^2}, \tag{12}$$

which is very similar to the middle part of (9). In fact, Proposition 2 shows that coherence and RIP are two special cases of gRIP.

Proposition 2. *For any column-normalized matrix $A \in \mathbf{R}^{m \times n}$, that is, $\|A_j\|_2 = 1$ for $j = 1, \ldots, n$, it holds*

(i) $\beta_{k,1,\infty} = \mu$, *for $2 \leq k \leq n$.*
(ii) $\beta_{k,2,2} = \delta_k$, *for $1 \leq k \leq n$.*

The proof of this proposition is shown in Appendix B.

A. Uniqueness Theorem Characterized via gRIP

We first give a lemma to show the relation between two different norms for a k-sparse vector space.

Lemma 1. *For any vector $x \in \mathbf{R}^n$ with $\|x\|_0 = k \leq n$, and for any $1 \leq q \leq p \leq \infty$, then*

$$\|x\|_p \leq \|x\|_q \leq k^{\frac{1}{q} - \frac{1}{p}}\|x\|_p. \tag{13}$$

This lemma is trivial based on the well-known norm equivalence theorem so the proof is omitted. Note that Lemma 1 is equivalent to

$$\|x\|_p \le k^{\max\{\frac{1}{p}-\frac{1}{q},0\}}\|x\|_q, \forall p,q \in [1,\infty]. \qquad (14)$$

With Lemma 1, the following theorem shows that a k-sparse solution of the equation $Ax = b$ will be the unique sparsest solution if A satisfies a certain gRIP condition.

Theorem 1. *Let x^* be a k-sparse solution of $Ax = b$. If A satisfies $(2k,p,q)$-gRIP with*

$$0 < \beta_{2k,p,q} < (2k)^{\min\{\frac{1}{q}-\frac{1}{p},0\}},$$

then x^* is the unique sparsest solution.

The proof of Theorem 1 is given in Appendix C. According to Proposition 2 and Theorem 1, we can obtain the following uniqueness results characterized via coherence and RIP, respectively.

Corollary 1. *Let x^* be a k-sparse solution of the equation $Ax = b$. If μ satisfies*

$$0 < \mu < \frac{1}{2k},$$

then x^* is the unique sparsest solution.

It was shown in [36] that when $\mu < \frac{1}{2k-1}$, the k-sparse solution should be unique. In another perspective, it can be noted that the condition $\mu < \frac{1}{2k}$ is equivalent to $k < \frac{1}{2\mu}$ while $\mu < \frac{1}{2k-1}$ is equivalent to $k < \frac{1}{2\mu} + \frac{1}{2}$. Since k should be an integer, these two conditions are almost the same.

Corollary 2. *Let x^* be a k-sparse solution of the equation $Ax = b$. If δ_{2k} satisfies*

$$0 < \delta_{2k} < 1,$$

then x^* is the unique sparsest solution.

According to [37], the RIP condition obtained in Corollary 2 is the same as the state-of-the-art result and more importantly, is tight in the sense that once the condition is violated, then we can construct two different signals with the same sparsity.

IV. CONVERGENCE ANALYSIS

In this section, we will study the convergence of the AIT algorithm based on the introduced gRIP.

A. Characterization via gRIP

To describe the convergence of the AIT algorithm, we first define

$$L_1 = 2^{p-1}(k^*)^{\max\{1-\frac{p}{q},0\}} + (2^{p-1} - (c_2)^p + 1)k^*,$$

$$L_2 = 2^p(2k^*)^{\max\{1-\frac{p}{q},0\}} + 2^{p-1}(c_1)^p k^*,$$

and

$$L = \min\{\sqrt[p]{L_1}, \sqrt[p]{L_2}\},$$

where $p \in [1,\infty)$, $q \in [1,\infty]$ and c_1, c_2 are the corresponding boundedness parameters.

Theorem 2. *Let $\{x^{(t)}\}$ be a sequence generated by the AIT algorithm. Assume that A satisfies $(3k^* + 1,p,q)$-gRIP with the constant $\beta_{3k^*+1,p,q} < \frac{1}{L}$, and let*
  (i) $k = k^*$;
  (ii) $\underline{s} < s < \overline{s}$, *where*

$$\underline{s} = \frac{(2k^*)^{\max\{\frac{1}{q}-\frac{1}{p},0\}} - \frac{1}{L}}{(2k^*)^{\max\{\frac{1}{q}-\frac{1}{p},0\}} - \beta_{3k^*+1,p,q}},$$

  and

$$\overline{s} = \frac{(2k^*)^{\max\{\frac{1}{q}-\frac{1}{p},0\}} + \frac{1}{L}}{(2k^*)^{\max\{\frac{1}{q}-\frac{1}{p},0\}} + \beta_{3k^*+1,p,q}}.$$

Then

$$\|x^{(t)} - x^*\|_p \le (\rho_s)^t \|x^* - x^{(0)}\|_p + \frac{sL}{1-\rho_s}\|A^T\epsilon\|_q,$$

where $\rho_s = \gamma_s L < 1$ with

$$\gamma_s = |1-s|(2k^*)^{\max\{\frac{1}{q}-\frac{1}{p},0\}} + s\beta_{3k^*+1,p,q}.$$

Particularly, when $\epsilon = 0$, it holds

$$\|x^{(t)} - x^*\|_p \le (\rho_s)^t \|x^* - x^{(0)}\|_p.$$

The proof of this Theorem is presented in Appendix D. Under the conditions of this theorem, we can verify that $0 < \rho_s < 1$. We first note that $\beta_{3k^*+1,p,q} < \frac{1}{L} < 1 \le (2k^*)^{\max\{\frac{1}{q}-\frac{1}{p},0\}}$, then it holds $\underline{s} < 1 < \overline{s}$. The definition of γ_s gives $\gamma_s =$

$$\begin{cases} (1-s)(2k^*)^{\max\{\frac{1}{q}-\frac{1}{p},0\}} + s\beta_{3k^*+1,p,q}, & if\ \underline{s} < s \le 1 \\ (s-1)(2k^*)^{\max\{\frac{1}{q}-\frac{1}{p},0\}} + s\beta_{3k^*+1,p,q}, & if\ 1 < s < \overline{s} \end{cases}.$$

If $\underline{s} < s \le 1$, it holds

$$\gamma_s < (1-\underline{s})(2k^*)^{\max\{\frac{1}{q}-\frac{1}{p},0\}} + \underline{s}\beta_{3k^*+1,p,q} = \frac{1}{L}.$$

Similarly, if $1 < s \le \overline{s}$

$$\gamma_s < (\overline{s}-1)(2k^*)^{\max\{\frac{1}{q}-\frac{1}{p},0\}} + \overline{s}\beta_{3k^*+1,p,q} = \frac{1}{L}.$$

Therefore, we have $\gamma_s < \frac{1}{L}$ and thus, $\rho_s = \gamma_s L < 1$.

Theorem 2 demonstrates that in the noise free case, the AIT algorithm converges to the original sparse signal at a linear rate, while in the noisy case, it also converges at a linear rate until reaching an error bound. Moreover, it can be noted that the constant ρ_s depends on the step size s. Since $\beta_{3k^*+1,p,q} < \frac{1}{L} < (2k^*)^{\max\{\frac{1}{q}-\frac{1}{p},0\}}$, ρ_s reaches its minimum at $s = 1$. The trend of ρ_s with respect to s is shown in Fig. 2. The optimal convergence rate is obtained when $s = 1$. This observation is consistent with the conclusion drawn in [6].

By Proposition 2, it shows that the coherence and RIP are two special cases of gRIP, thus we can easily obtain some recovery guarantees based on coherence and RIP respectively in the next two subsections.

Remark 3. *From Theorem 2, we can see that the step size should lie in an appropriate interval, which depends on the gRIP constant, which is generally NP-hard to verify. However, we would like to emphasize that the theoretical result obtained in Thoeorem 2 is of importance in theory and it can give some insights and theoretical guarantees of the implementation of*
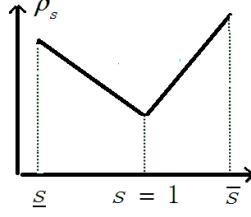
Fig. 2. The trend of ρ_s with respect to s.

the AIT algorithm, though it seems stringent. Empirically, we find that a small interval of the step size, i.e., $[0.9, 1]$ is generally sufficient for the convergence of the AIT algorithm. This is also supported by the numerical experiments conducted in section VI. In [8], it demonstrates that many algorithms perform well with either constant or adaptive step sizes. In section VII C, we will discuss and compare different step-size schemes including the constant and an adaptive step-size strategies on the performance of AIT algorithms.

B. Characterization via Coherence

Let $p = 1, q = \infty$. In this case, $L_1 = (3 - c_2)k^*$, $L_2 = (4 + c_1)k^*$, and $L = (3 - c_2)k^*$. According to Theorem 2 and Proposition 2, assume that $\mu < \frac{1}{(3-c_2)k^*}$, then the AIT algorithm converges linearly with the convergence rate constant

$$\rho_s = \gamma_s L = (|1 - s| + s\mu)L < 1$$

if we take $k = k^*$ and $\frac{1 - \frac{1}{L}}{1 - \mu} < s < \frac{1 + \frac{1}{L}}{1 + \mu}$. In the following, we show that the constant γ_s and thus ρ_s can be further improved when $p = 1$ and $q = \infty$.

Theorem 3. *Let $\{x^{(t)}\}$ be a sequence generated by the AIT algorithm for $b = Ax + \epsilon$. Assume that A satisfies $0 < \mu < \frac{1}{(3-c_2)k^*}$, and if we take*
  (i) $k = k^*$;
  (ii) $1 - \frac{1}{L} < s < \min\{\frac{1}{L\mu}, 1 + \frac{1}{L}\}$,
then it holds

$$\|x^{(t)} - x^*\|_1 \le (\rho_s)^t \|x^* - x^{(0)}\|_1 + \frac{sL}{1 - \rho_s}\|A^T\epsilon\|_\infty,$$

where $\rho_s = \gamma_s L < 1$ with

$$\gamma_s = \max\{|1 - s|, s\mu\}.$$

Particularly, when $\epsilon = 0$, it holds

$$\|x^{(t)} - x^*\|_1 \le (\rho_s)^t \|x^* - x^{(0)}\|_1.$$

The proof of this Theorem is given in Appendix E. As shown in Theorem 3, the constant γ_s can be improved from $|1 - s| + s\mu$ to $\max\{|1 - s|, s\mu\}$, and also the feasible range of the step size parameter s gets larger from $\left(\frac{1 - \frac{1}{L}}{1 - \mu}, \frac{1 + \frac{1}{L}}{1 + \mu}\right)$ to $\left(1 - \frac{1}{L}, \min\{\frac{1}{L\mu}, 1 + \frac{1}{L}\}\right)$. We list the coherence-based convergence conditions of several typical AIT algorithms in Table II. As shown in Table II, it can be observed that the recovery condition for Soft algorithm is the same as those of OMP [38] and BP [39].

TABLE II
COHERENCE BASED CONDITIONS FOR DIFFERENT AIT ALGORITHMS

AIT	Hard	Half	Soft	SCAD
c_2	0	0	1	0
μ	$\frac{1}{3k^*-1}$	$\frac{1}{3k^*-1}$	$\frac{1}{2k^*-1}$	$\frac{1}{3k^*-1}$

C. Characterization via RIP

Let $p = 2, q = 2$. In this case, $L_1 = 2 + (3 - c_2^2)k^*$, $L_2 = 4 + 2c_1^2 k^*$, and thus

$$L = \min\{\sqrt{4 + 2c_1^2 k^*}, \sqrt{2 + (3 - c_2^2)k^*}\}.$$

According to Theorem 2, and by Proposition 2, we can directly claim the following corollary.

Corollary 3. *Let $\{x^{(t)}\}$ be a sequence generated by the AIT algorithm for $b = Ax + \epsilon$. Assume that A satisfies $\delta_{3k^*+1} < \frac{1}{L}$, and if we take*
  (i) $k = k^*$;
  (ii) $\underline{s} \le s \le \overline{s}$, where $\underline{s} = \frac{1 - \frac{1}{L}}{1 - \delta_{3k^*+1}}$ and $\overline{s} = \frac{1 + \frac{1}{L}}{1 + \delta_{3k^*+1}}$.
Then

$$\|x^{(t)} - x^*\|_2 \le (\rho_s)^t \|x^* - x^{(0)}\|_2 + \frac{sL}{1 - \rho_s}\|A^T\epsilon\|_2,$$

where $\rho_s = \gamma_s L < 1$ with $\gamma_s = |1 - s| + s\delta_{3k^*+1}$. Particularly, when $\epsilon = 0$, it holds

$$\|x^{(t)} - x^*\|_2 \le (\rho_s)^t \|x^* - x^{(0)}\|_2.$$

According to Corollary 3, the RIP based sufficient conditions for some typical AIT algorithms are listed in Table III.

TABLE III
RIP BASED CONDITIONS FOR DIFFERENT AIT ALGORITHMS

AIT	Hard	Half	Soft	SCAD
c_1	0	1/3	1	1
δ_{3k^*+1}	$\frac{1}{2}$	$\frac{3}{\sqrt{36+2k^*}}$	$\frac{1}{\sqrt{2+2k^*}}$	$\frac{1}{\sqrt{4+2k^*}}$

Moreover, we note that the condition in Corollary 3 for Hard algorithm can be further improved via using the specific expression of the hard thresholding operator. This can be shown as the following theorem.

Theorem 4. *Let $\{x^{(t)}\}$ be a sequence generated by Hard algorithm for $b = Ax + \epsilon$. Assume that A satisfies $\delta_{3k^*+1} < \frac{\sqrt{5}-1}{2}$, and if we take $k = k^*$ and $s = 1$, then*

$$\|x^{(t)} - x^*\|_2 \le \rho^t \|x^* - x^{(0)}\|_2 + \frac{\sqrt{5}+1}{2 - 2\rho}\|A^T\epsilon\|_2,$$

where $\rho = \frac{\sqrt{5}+1}{2}\delta_{3k^*+1} < 1$. Particularly, when $\epsilon = 0$, it holds

$$\|x^{(t)} - x^*\|_2 \le \rho^t \|x^* - x^{(0)}\|_2.$$

The proof of Theorem 4 is presented in Appendix F.

V. COMPARISON WITH PREVIOUS WORKS

This section discusses some related works of the AIT algorithm, and then compares its computational complexity and sufficient conditions for convergence with other algorithms.

1) On related works of the AIT algorithm: In [34], Maleki provided some similar results for two special AIT algorithms, i.e., Hard and Soft algorithms with $k = k^*$ and $s = 1$ for the noiseless case. The sufficient conditions for convergence are $\mu < \frac{1}{3.1k^*}$ and $\mu < \frac{1}{4.1k^*}$ for Hard and Soft algorithms, respectively. In [35], Zeng et al. improved and extended Maleki's results to a wide class of the AIT algorithm with step size $s = 1$. The sufficient condition based on coherence was improved to $\mu < \frac{1}{(3+c_1)k^*}$, where the boudedness parameter c_1 can be found in Table I. Compared with these two tightly related works, several significant improvements are made in this paper.

(i) **Weaker convergence conditions.** The conditions obtained in this paper is weaker than those in both [34] and [35]. More specifically, we give a unified convergence condition based on the introduced gRIP. Particularly, as shown in Theorem 3, the coherence based conditions for convergence are $\mu < \frac{1}{(3-c_2)k^*-1}$, which is much better than the condition $\mu < \frac{1}{(3+c_1)k^*}$ obtained in [35]. Moreover, except Hard algorithm, we firstly show the convergence of the other AIT algorithms based on RIP.

(ii) **Better convergence rate.** The asymptotic linear convergence rate was justified in both [34] and [35]. However, in this paper, we show the global linear convergence rate of the AIT algorithm, which means it converges at a linear rate from the first iteration.

(iii) **More general model.** In this paper, besides the noiseless model $b = Ax$, we also consider the performance of the AIT algorithm for the noisy model $b = Ax + \epsilon$, which is very crucial since the noise is almost inevitable in practice.

(iv) **More general algorithmic framework.** In both [34] and [35], the AIT algorithm was only considered with unit step size ($s = 1$). While in this paper, we show that the AIT algorithm converges when s is in an appropriate range.

Among these AIT algorithms, Hard algorithm has been widely studied. In [36], it was demonstrated that if A has unit-norm columns and coherence μ, then A has the (r, δ_r)-RIP with

$$\delta_r \leq (r-1)\mu. \tag{15}$$

In terms of RIP, Blumensath and Davies [33] justified the performance of Hard algorithm when applied to signal recovery problem. It was shown that if A satisfies a certain RIP with $\delta_{3k^*} < \frac{1}{\sqrt{32}}$, then Hard algorithm has global convergence guarantee. Later, Foucart improved this condition to $\delta_{3k^*} < \frac{1}{2}$ or $\delta_{2k^*} < \frac{1}{4}$ [4] and further improved it to $\delta_{3k^*} < \frac{1}{\sqrt{3}} \approx 0.5773$ (Theorem 6.18, [9]). Now we can improve this condition to $\delta_{3k^*+1} < \frac{\sqrt{5}-1}{2} \approx 0.618$ as shown by Theorem 4.

2) On comparison with other algorithms: For better comparison, we list the state-of-the-art results on sufficient conditions of some typical algorithms including BP, OMP, CoSaMP, Hard, Soft, Half and general AIT algorithms in Table IV.

TABLE IV
SUFFICIENT CONDITIONS FOR DIFFERENT ALGORITHMS

Algorithm	μ	(r, δ_r)
BP	$\frac{1}{2k^*-1}$ ([39])	$(2k^*, 0.707)$ ([41])
OMP	$\frac{1}{2k^*-1}$ ([38])	$(13k^*, \frac{1}{6})$ (Thm. 6.25,[9])
CoSaMP	$\frac{0.384}{4k^*-1}$ ⋆	$(4k^*, 0.384)$ ([14])
Hard	$\frac{1}{3k^*-1}$ (Thm. 3)	$(3k^*+1, 0.618)$ (Thm. 4)
Soft	$\frac{1}{2k^*-1}$ (Thm. 3)	$(3k^*+1, \frac{1}{\sqrt{2+2k^*}})$ (Coro. 3)
Half	$\frac{1}{3k^*-1}$ (Thm. 3)	$(3k^*+1, \frac{3}{\sqrt{36+2k^*}})$ (Coro. 3)
General AIT	$\frac{1}{(3-c_2)k^*-1}$ (Thm. 3)	$(3k^*+1, \frac{1}{\sqrt{4+2c_1^2k^*}})$ (Coro. 3)

⋆: a coherence based sufficient condition for CoSaMP derived by the fact that $\delta_{4k^*} < 0.384$ and $\delta_r \leq (r-1)\mu$.

From Table IV, in the perspective of coherence, the sufficient conditions of AIT algorithms are slightly stricter than those of BP and OMP algorithms except Soft algorithm. However, AIT algorithms are generally faster than both BP and OMP algorithms with lower computational complexities, especially for large scale applications due to their linear convergence rates. As shown in the next section, the number of iterations required for the convergence of the AIT algorithm is empirically of the same order of the original sparsity level k^*, that is, $\mathcal{O}(k^*)$. At each iteration of the AIT algorithm, only some simple matrix-vector multiplications and a projection on the vector need to be done, and thus the computational complexity per iteration is $\mathcal{O}(mn)$. Therefore, the total computational complexity of the AIT algorithm is $\mathcal{O}(k^*mn)$. While the total computational complexities of BP and OMP algorithms are generally $\mathcal{O}(m^2n)$ and $\max\{\mathcal{O}(k^*mn), \mathcal{O}(\frac{(k^*)^2(k^*+1)^2}{4})\}$, respectively. It should be pointed out that the computational complexity of OMP algorithm is related to the commonly used halting rule of OMP algorithm, that is, the number of maximal iterations is set to be the true sparsity level k^*.

Another important greedy algorithm, CoSaMP algorithm, identifies multicomponents (commonly $2k^*$) at each iteration. From Table IV, the RIP based sufficient condition of CoSaMP is $\delta_{4k^*} < 0.384$ and a deduced coherence based sufficient condition is $\mu < \frac{0.384}{4k^*-1}$. In the perspective of coherence, our conditions for AIT algorithms are better than CoSaMP, though this comparison is not very reasonable. On the other hand, our conditions for AIT algorithms except Hard algorithm are generally worse than that of CoSaMP in the perspective of RIP. However, when the true signal is very sparse, the conditions of AIT algorithms may be better than that of CoSaMP. At each iteration of CoSaMP algorithm, some simple matrix-vector multiplications and a least squares problem should be considered. Thus, the computational complexity per iteration of CoSaMP algorithm is generally $\max\{\mathcal{O}(mn), \mathcal{O}((3k^*)^3)\}$, which is higher than those of AIT algorithms, especially when k^* is relatively large.

Besides BP and greedy algorithms, another class of tightly related algorithms is the reweighted techniques that have

been also widely used for solutions to ℓ_q regularization with $q \in (0,1)$. Two well-known examples of such reweighted techniques are the iterative reweighted least squares (IRLS) method [23] and the reweighted l_1 minimization (IRL1) method [20]. The convergence analysis conducted in [24] shows that the IRLS method converges with an asymptotically superlinear convergence rate under the assumptions that A possesses a certain null-space property (NSP). However, from Theorem 2, the rates of convergence of AIT algorithms are globally linear. Furthermore, Lai et al. [42] applied the IRLS method to the unconstrained l_q minimization problem and also extended the corresponding convergence results to the matrix case. It was shown also in [43] that the IRL1 algorithm can converge to a stationary point and the asymptotic convergence speed is approximately linear when applied to the unconstrained l_q minimization problem. Both in [42] and [43], the authors focused on the unconstrained l_q minimization problem with a fixed regularization parameter λ, while in this paper, we focus on a different model with an adaptive regularization parameter.

VI. DISCUSSION

In this section, we numerically discuss some practical issues on the implementation of AIT algorithms, especially, the effects of several algorithmic factors including the estimated sparsity level parameter, the column-normalization operation, different step-size strategies as well as the formats of different thresholding operators on the performance of AIT algorithms. Moreover, we will further demonstrate the performance of several typical AIT algorihtms including Hard, Half and SCAD via comparing with many state-of-the-art algorithms such as CGIHT [50], CoSaMP [14], 0-ALPS(4) [6] in the perspective of the 50% phase transition curves [47], [49].

A. Robustness of the estimated sparsity level

In the preceding proposed algorithms, the specified sparsity level parameter k is taken exactly as the true sparsity level k^*, which is generally unknown in practice. Instead, we can often obtain a rough estimate of the true sparsity level. Therefore, in this experiment, we will explore the performance of the AIT algorithm with a variety of specified sparsity levels. We varied k from 1 to 150 while kept $k^* = 15$. The experiment setup is the same with Section VI. A.

From Fig. 3, we can observe that these AIT algorithms are efficient for a wide range of k. Interestingly, the point $k = k^*$ is a break point of the performance of all these AIT algorithms. When $k < k^*$, all AIT algorithms fail to recover the original sparse signal, while when $k \geq k^*$, a wide interval of k is allowed for small recovery errors, as shown in Fig. 3 (b) and (d). In the noise free case, if $\|x^{(t)} - x^*\|_2 < 10^{-10}$, the feasible intervals of the specified sparsity level k are $[15, 109]$ for SCAD and Soft, $[15, 81]$ for Half and $[15, 65]$ for Hard, respectively. This observation is very important for real applications of AIT algorithms because k^* is usually unknown. In the noisy case, if $\|x^{(t)} - x^*\|_2 < 10^{-2}$, the feasible intervals of sparsity level k are $[15, 105]$ for SCAD, $[15, 40]$ for Soft, $[15, 37]$ for Half and $[15, 26]$ for Hard, respectively.
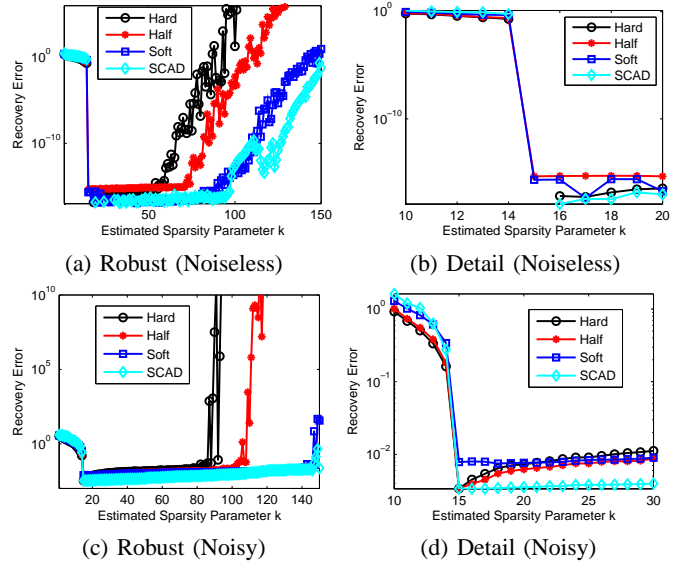

Fig. 3. On robustness of the specified sparsity level. (a) The trends of the recovery precision with different estimated sparsity levels in noiseless case. (b) The detailed trends of the recovery precision with different estimated sparsity levels in noiseless case. (c) The trends of the recovery precision with different estimated sparsity levels in noiseless case. (d) The detailed trends of the recovery precision with different estimated sparsity levels in noisy case.

TABLE V
THE RECOVERY PRECISION OF DIFFERENT AIT ALGORITHMS WITH OR WITHOUT COLUMN-NORMALIZATION (NOISELESS CASE)

Algorithm	Hard	Soft	Half	SCAD
no normalization	5.719e-6	1.425e-8	5.062e-6	9.330e-9
normalization	5.703e-5	1.437e-8	5.935e-5	8.505e-9

B. With vs Without Normalization

As shown in Algorithm 1, the column-normalization on the measurement matrix A is required in consideration of a clearer definition of the introduced gRIP and more importantly, better theoretical analyses. However, in this subsection, we will conduct a series of simulations to show that such requirement is generally not necessary in practice. The experiment setup is similar to Section VI.A. More specifically, we set $m = 250$, $n = 400$ and $k^* = 15$. The nonzero components of x^* were generated randomly according to the standard Gaussian distribution. The matrix A was generated from i.i.d Gaussian distribution $\mathcal{N}(0, 1/250)$ without normalization. In order to adopt Algorithm 1, we let Λ be the corresponding column-normalized factor matrix of A (i.e., Λ is a diagonal matrix and its diagonal element is the l_2-norm of the corresponding column of A), and $\hat{A} = A\Lambda^{-1}$ be the corresponding column-normalized measurement matrix. Assume that $\hat{x}$ is a recovery via Algorithm 1 corresponding to $\hat{A}$, then $\bar{x} = \Lambda^{-1}\hat{x}$ is the corresponding recovery of x^*. For each algorithm, we conducted 10 times experiments independently in both noiseless and noise (signal-to-noise ratio (SNR): 60dB) cases, and recorded the average recovery precision. The recovery precision is defined as $\frac{\|\tilde{x} - x^*\|_2}{\|x^*\|_2}$, where $\tilde{x}$ and x^* represent the recovery and original signal, respectively. The experiment results are shown in Table V and VI.

TABLE VI
THE RECOVERY PRECISION OF DIFFERENT AIT ALGORITHMS WITH OR
WITHOUT COLUMN-NORMALIZATION (WITH 60dB NOISE)

Algorithm	Hard	Soft	Half	SCAD
no normalization	1.217e-3	5.739e-3	1.206e-3	1.282e-3
normalization	1.214e-3	5.498e-3	1.205e-3	1.264e-3

From Table V and VI, we can see that the column-normalization operator has almost no effect on the performance of the AIT algorithm in both noiseless and noise cases. Therefore, in the following experiments, we will adopt the more practical AIT algorithm without the column-normalization for better comparison with the other algorithms.

C. Constant vs Adaptive Step Size

From Algorithm 1, we only consider the constant step-size. However, according to many previous and empirical studies [6], [46], we have known that certain adaptive step-size strategies may improve the performance of AIT algorithms. In this subsection, we will compare the performance of two different step-size schemes, i.e., the constant step-size strategy and an adaptive step-size strategy introduced in [46] via the so-called 50% phase transition curve [49]. More specifically, the adaptive step-size scheme can be described as follows. Assume that $x^{(t)}$ is the t-th iteration, then at $(t+1)$-th iteration, the step size $s^{(t+1)}$ is set as

$$s^{(t+1)} = \frac{\|(A^T(b - Ax^{(t)}))_{I^t}\|^2}{\|(AA^T(b - Ax^{(t)}))_{I^t}\|^2}, \qquad (16)$$

where I^t is the support set of $x^{(t)}$, A is the measurement matrix and b is the measurement vector. Similar to [46], we will call the AIT algorithm with such adaptive step-size strategy the normalised AIT (NAIT) algorithm, and correspondingly, several typical AIT algorithms such as Hard, Soft, Half and SCAD algorithms with such adaptive step-size strategy NHard, NSoft, NHalf and NSCAD for short, respectively. Note that NHard algorithm studied here is actually the same with the normalised iterative hard thresholding (NIHT) algorithm proposed in [46].

50% phase transition curve was first introduced in [48] and has been widely used to compare the recovery ability for different algorithms in compressed sensing [47], [49]. For a fixed n, any given problem setting (k, m, n) can depict a point in the space $(m/n, k/m) \in (0, 1]^2$. For any algorithm, its 50% phase transition curve is actually a function f on the $(k/m, m/n)$ space. More specifically, if the point $(m/n, k/m)$ lies below the curve of the algorithm, i.e. $k/m < f(m/n)$, then it means the algorithm could recover the sparse signal from the given (k, m, n)-problem with high probability, otherwise the successful recovery probability is very low [48]. Moreover, the 50% phase transition curve usually depends on the prior distribution of x^* as depicted in many researches [27], [47], [49].

In these experiments, we consider two common distributions of x^*, the first one is the standard Gaussian distribution, and the second one is a binary distribution, which takes

-1 or 1 with an equal probability. For any given (k, m, n), the measurement matrix $A \in \mathbf{R}^{m \times n}$ is generated from the Gaussian distribution $\mathcal{N}(0, \frac{1}{m})$, and the nonzero components of the original k-sparse signal x^* are generated independently and identically distribution (i.i.d.) according to the Gaussian or binary distributions. For any experiment, we consider it as a successful recovery if

$$\frac{\|\tilde{x} - x^*\|_\infty}{\|x^*\|_\infty} \le 10^{-3},$$

where x^* is the original sparse signal and $\tilde{x}$ is the corresponding recovery signal. We set $n = 512$, $m = 50, 100, ..., 500$. To determine $f(m/n)$, we exploit a bisection search scheme as the same as the experiment setting in [47]. We compare the 50% phase transition curves of Hard, Soft, Half and SCAD algorithms with their adaptive step-size versions, i.e., NHard, NSoft, NHalf, NSCAD in Fig. 4.

From Fig. 4 (a) and (c), we can see that the performances of all AIT algorithms except Soft algorithm adopting the adaptive step-size strategy (16) are significantly better than those of the corresponding AIT algorithms with a constant step size in the Guassian case. In this case, NSCAD has the best performance, then NHalf and NHard, while NSoft is the worst. The performance of NSCAD is slightly better than those of NHalf and NHard, and much better than NSoft. While for the binary case, as shown in Fig. 4 (b) and (d), NSCAD breaks down with the curve fluctuating around 0.1 while NHalf and NHard still perform well. In the binary case, Soft as well as NSoft perform the worst. In addition, we can see that the performances of Soft and NSoft are almost the same in all cases, which means that such adaptive step-size strategy (16) may not bring the improvement on the performance of Soft algorithm. Moreover, some interesting phenomena can also be observed in Fig. 4, that is, the performance of the AIT algorithm depends to some extent on the choice of the thresholding operator, and for different prior distributions of the original sparse signal, the AIT algorithm may perform very different. For these phenomena, we will study in the future work.

D. Comparison with the State-of-the-art Algorithms

We also compare the performance of several AIT algorithms including NHard, NSCAD and NHalf with some typical state-of-the-art algorithms such as conjugate gradient iterative hard thresholding (CGIHT) [50], CoSaMP [14], 0-ALPS(4) [6] in terms of their 50% phase transition curves. For more other algorithms like MP [3], HTP [7], OMP [10], CSMPSP [51], CompMP [52], OLS [53] etc., their 50% phase transition curves can be found in [49], and we omit them here. For all considered algorithms, the estimated sparsity level parameters are set to be the true sparsity level of x^*. The result is shown in Fig 5.

From Fig. 5, we can see that almost all algorithms have better performances for the Gaussian distribution case than for the binary distribution case, especially NSCAD algorithm. More specifically, as shown in Fig. 5(a), for the Gaussian distribution, NSCAD has the best performance among all these
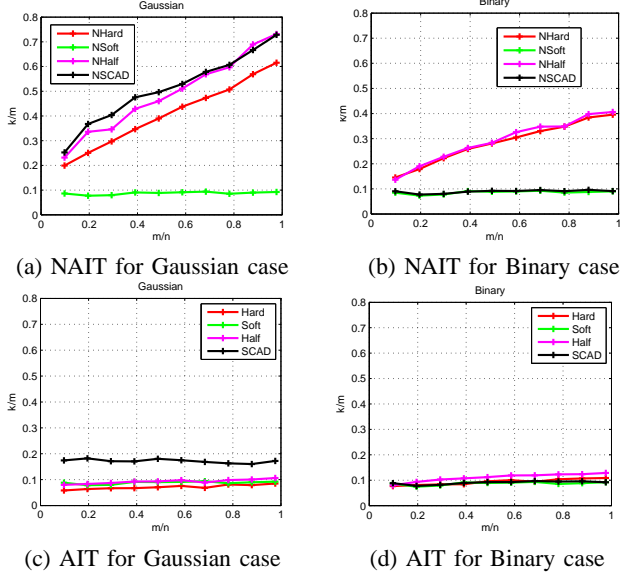
Fig. 4.   50% phase transition curves of different AIT algorithms with two different step-size schemes. (a) AIT algorithms with an adaptive step size for Gaussian case. (a) AIT algorithms with an adaptive step size for Binary case. (c) AIT algorithms with a constant step size for Gaussian case. (d) AIT algorithms with a constant step size for Binary case.
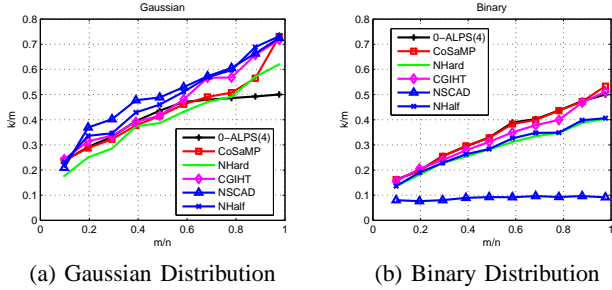

Fig. 5.   50% phase transition curves of different algorithms. (a) Gaussian distribution case. (b) Binary distribution case.

algorithms, and NHalf is slightly worse than NSCAD and better than the other algorithms. While in the binary case, it can be seen from Fig. 5(b), all AIT algorithms perform worse than the other algorithms like CGIHT, CoSaMP, 0-ALPS(4), especially, NSCAD algorithm is much worse than the other algorithms. These experiments demonstrate that AIT algorithms are more appropriate for the recovery problems that the original sparse signals obey the Gaussian distribution.

VII. Conclusion

We have conducted a study of a wide class of AIT algorithms for compressed sensing. It should be pointed out that almost all of the existing iterative thresholding algorithms like Hard, Soft, Half and SCAD are included in such class of algorithms. The main contribution of this paper is the establishment of the convergence analyses of the AIT algorithm. In summary, we have shown when the measurement matrix satisfies a certain gRIP condition, the AIT algorithm can converge to the original sparse signal at a linear rate in the noiseless case, and approach to the original sparse signal at a linear rate until achieving an error bound in the noisy

case. As two special cases of gRIP, the coherence and RIP based conditions can be directly derived for the AIT algorithm. Moreover, the tightness of our analyses can be demonstrated by two specific cases, that is, the coherence-based condition for Soft algorithm is the same as those of OMP and BP, and the RIP based condition for Hard algorithm is better than the recent result $\delta_{3k^*} < \frac{1}{\sqrt{3}} \approx 0.5773$ obtained in Theorem 6.18 in [9]. Furthermore, the efficiency of the algorithm and the correctness of the theoretical results are also verified via a series of numerical experiments.

In section VII, we have numerically discussed many practical issues on the implementation of AIT algorithms, including the specified sparsity level parameter k, the column-normalization requirement as well as different step-size setting schemes. We can observe the following several interesting phenomena:

(i)   The AIT algorithm is robust to the specified sparsity level parameter k, that is, the parameter k can be specified in a large range to guarantee the well performance of the AIT algorithm.

(ii)   The column-normalization of the measurement matrix A is not necessary for the use of AIT algorithms in the perspective of the recovery performance.

(iii)   Some adaptive step-size strategies may significantly improve the performance of AIT algorithms.

(iv)   The performance of AIT algorithm depends to some extent on the prior distribution of the original sparse signal. Compared with the binary distribution, AIT algorithms are more appropriate for the recovery of the sparse signal generated by the Gaussian distribution.

(v)   The performance of the AIT algorithm depends on the specific thresholding operator.

All of these phenomena are of interest, and we will study them in our future work.

Appendix A: Proof of Proposition 1

Proof: For any index set $S \subset \{1, \ldots, n\}$ with $|S| \leq k$ and a vector $x \in \mathbf{R}^{|S|}$, since $\frac{1}{p} + \frac{1}{q} = 1$, then ℓ_p and ℓ_q norms are dual to each other, which implies that

$$\|(\mathbf{I}_{|S|} - A_S^T A_S)x\|_q = \sup_{y \in \mathbf{R}^{|S|} \setminus \{0\}} \frac{|y^T(\mathbf{I}_{|S|} - A_S^T A_S)x|}{\|y\|_p}.$$
(17)

By Definition 1, then

$$\beta_{k,p,q} = \sup_{|S|\leq k} \sup_{x,y\in\mathbf{R}^{|S|}\backslash\{0\}} \frac{\left|y^T(\mathbf{I}_{|S|} - A_S^T A_S)x\right|}{\|x\|_p\|y\|_p}. \quad (18)$$

It is obvious that

$$\beta_{k,p,q} \geq \sup_{|S|\leq k} \sup_{x\in\mathbf{R}^{|S|}\backslash\{0\}} \frac{\left|x^T(\mathbf{I}_{|S|} - A_S^T A_S)x\right|}{\|x\|_p^2}$$
$$= \sup_{z\in\mathbf{R}^n\backslash\{0\},\|z\|_0\leq k} \frac{\left|z^T(\mathbf{I}_n - A^T A)z\right|}{\|z\|_p^2},$$

which implies the right-hand side of (9).

On the other hand, by (18), we can also observe that

$$\beta_{k,p,q} = \sup_{|S|\leq k} \sup_{\|x\|_p,\|y\|_p\leq 1} |y^T x - y^T A_S^T A_S x|, \quad (19)$$

and for any $x,y\in\mathbf{R}^{|S|}$,

$$\begin{aligned}
&|y^T x - y^T A_S^T A_S x|\\
&= \left|\frac{1}{2}(\|x\|_2^2 + \|y\|_2^2 - \|x-y\|_2^2)\right.\\
&\quad \left.-\frac{1}{2}(\|A_S x\|_2^2 + \|A_S y\|_2^2 - \|A_S x - A_S y\|_2^2)\right|\\
&\leq \frac{1}{2}\left|\|x\|_2^2 - \|A_S x\|_2^2\right| + \frac{1}{2}\left|\|y\|_2^2 - \|A_S y\|_2^2\right|\\
&\quad + \frac{1}{2}\left|\|x-y\|_2^2 - \|A_S(x-y)\|_2^2\right|. \quad (20)
\end{aligned}$$

Furthermore, it can be noted that

$$\begin{aligned}
&\sup_{u,v\in\mathbf{R}^{|S|},\|u\|_p,\|v\|_p\leq 1} \left|\|u-v\|_2^2 - \|A_S(u-v)\|_2^2\right|\\
&\leq 4 \sup_{w\in\mathbf{R}^{|S|},\|w\|_p\leq 1} \left|\|w\|_2^2 - \|A_S w\|_2^2\right|, \quad (21)
\end{aligned}$$

since $\|u-v\|_p \leq 2$ for $\|u\|_p \leq 1$ and $\|v\|_p \leq 1$. Plugging (20) and (21) into (19), it yields

$$\begin{aligned}
\beta_{k,p,q}^A &\leq 3 \sup_{|S|\leq k} \sup_{\|x\|_p\leq 1} \left|\|x\|_2^2 - \|A_S x\|_2^2\right|\\
&= 3 \sup_{\|z\|_0\leq k,\|z\|_p\leq 1} \left|\|z\|_2^2 - \|A z\|_2^2\right|\\
&= 3 \sup_{z\in\mathbf{R}^n\backslash\{0\},\|z\|_0\leq k} \frac{\left|z^T(A^T A - \mathbf{I}_n)z\right|}{\|z\|_p^2},
\end{aligned}$$

which implies the left-hand side of (9). Therefore, the proof of this proposition is completed. ∎

APPENDIX B: PROOF OF PROPOSITION 2

Proof: (i) The definition of gRIP induces $\beta_{k,1,\infty} \geq \beta_{2,1,\infty}$ for all $k \geq 2$. Therefore, if we can claim the following two facts: (a) $\beta_{2,1,\infty} \geq \mu$, and (b) $\beta_{k,1,\infty} \leq \mu$ for all $k \geq 2$, then Proposition 2 (i) follows.

We first justify the fact (a). Suppose the maximal element of $\mathbf{I}_n - A^T A$ in magnitude appears at the i_0-th row and the j_0-th column. Because for any j, the j-th diagonal elements of $\mathbf{I}_n - A^T A$ equals to $1 - \|A_j\|^2 = 0$, we know $i_0 \neq j_0$. Without loss of generality, we assume that $i_0 < j_0$. Let A_{i_0} and A_{j_0} be the i_0-th and j_0-th column vector of A, respectively, then Definition 2 gives

$$\mu = |A_{i_0}^T A_{j_0}|.$$

Let $S = \{i_0, j_0\}$ and $\mathbf{e} = (0,1)^T$. Then

$$\begin{aligned}
\beta_{2,1,\infty} &\geq \|(\mathbf{I}_2 - A_S^T A_S)\mathbf{e}\|_\infty\\
&= \|\mathbf{e} - A_S^T A_{j_0}\|_\infty\\
&= \mu. \quad (22)
\end{aligned}$$

Then we prove the fact (b). For any vector $x \in \mathbf{R}^k$ and a subset $S \subset \{1,2,\ldots,n\}$ with $|S| = k$, let $B = \mathbf{I}_k - A_S^T A_S$ and $z = Bx$. Then

$$|z_i| = \left|\sum_{j=1}^k B_{ij} x_j\right| \leq \sum_{j=1}^k |B_{ij} x_j| \leq \mu\|x\|_1,$$

for any $i = 1, \ldots, k$. It implies that

$$\|Bx\|_\infty \leq \mu\|x\|_1.$$

By the definition of $\beta_{k,1,\infty}$, it implies

$$\beta_{k,1,\infty} \leq \mu. \quad (23)$$

According to (22) and (23), for all $2 \leq k \leq n$, it holds

$$\beta_{k,1,\infty} = \mu.$$

(ii) From the inequality (9) and the equality (12), we know

$$\delta_k \leq \beta_{k,p,q}. \quad (24)$$

To prove

$$\delta_k \geq \beta_{k,p,q}, \quad (25)$$

note that equality (19) leads to

$$\beta_{k,p,q} \leq \sup_{S\subset\{1,\ldots,n\},|S|\leq k} \|\mathbf{I}_{|S|} - A_S^T A_S\|_2, \quad (26)$$

and further

$$\begin{aligned}
&\sup_{S\subset\{1,\ldots,n\},|S|\leq k} \|\mathbf{I}_{|S|} - A_S^T A_S\|_2 \quad (27)\\
&= \sup_{|S|\leq k} \sup_{x\in\mathbf{R}^{|S|}\backslash\{0\}} \frac{|x^T(\mathbf{I}_{|S|} - A_S^T A_S)x|}{\|x\|_2^2}\\
&= \sup_{z\in\mathbf{R}^n\backslash\{0\},\|z\|_0\leq k} \frac{|z^T(\mathbf{I}_n - A^T A)z|}{\|z\|_2^2}\\
&= \delta_k,
\end{aligned}$$

where the last equality holds by the equivalent definition of RIP (this can be also referred to Definition 1 in [4]). From (24)-(27), we can conclude that

$$\delta_k = \beta_{k,p,q}.$$

∎

APPENDIX C: PROOF OF THEOREM 1

Proof: We prove this theorem by contradiction. Assume x^{**} satisfies $Ax^{**} = b$ and $\|x^{**}\|_0 \leq k$. Then

$$A(x^* - x^{**}) = 0,$$

which implies

$$(\mathbf{I}_n - A^T A)(x^* - x^{**}) = x^* - x^{**}.$$

Let $x = x^* - x^{**}$, S be the support of x and x_S be a subvector of x with the components restricted to S. It follows

$$(\mathbf{I}_{|S|} - A_S^T A_S)x_S = x_S,$$

and further

$$\|(\mathbf{I}_{|S|} - A_S^T A_S)x_S\|_q = \|x_S\|_q, \tag{28}$$

for any $q \in [1, \infty)$. Since $\|x^*\|_0 \leq k$ and $\|x^{**}\|_0 \leq k$, then $|S| \leq 2k$. For any $p \in [1, \infty)$, and by the definition of gRIP, we have

$$\|(\mathbf{I}_{|S|} - A_S^T A_S)x_S\|_q \leq \beta_{2k,p,q}\|x_S\|_p.$$

By Lemma 1, there holds

$$\|x_S\|_p \leq (2k)^{\max\{\frac{1}{p}-\frac{1}{q},0\}}\|x_S\|_q.$$

By the assumption of this theorem, then

$$\|(\mathbf{I}_{|S|} - A_S^T A_S)x_S\|_q \leq \beta_{2k,p,q}(2k)^{\max\{\frac{1}{p}-\frac{1}{q},0\}}\|x_S\|_q$$
$$< \|x_S\|_q,$$

which contradicts with (28). Therefore, x^* is the unique sparsest solution. ∎

APPENDIX D: PROOF OF THEOREM 2

Before justifying the convergence of the AIT algorithm based on gRIP, we first introduce two lemmas.

Lemma 2. *For any $x, y \in \mathbf{R}^n$, and $p \in [1, \infty)$, then*

$$\|x + y\|_p^p \leq 2^{p-1}(\|x\|_p^p + \|y\|_p^p). \tag{29}$$

Moreover, if $x_i \cdot y_i \geq 0$ for $i = 1, \ldots, n$, then

$$\|x + y\|_p^p \geq \|x\|_p^p + \|y\|_p^p. \tag{30}$$

The proof of Lemma 2 is obvious since $f(z) = z^p$ is convex for $p \geq 1$ and any $z \geq 0$, and $\|x\|_p \leq \|x\|_1$ for any $x \in \mathbf{R}^n$. We will omit it due to the limitation of the length of the paper.

Lemma 3. *For any $t \geq 1$ and $q \in [1, \infty]$, if $k \geq k^*$, the following inequality holds for the AIT algorithm:*

$$\tau^{(t)} \leq (\sum_{i \in I_+^t} |z_i^{(t)} - x_i^*|^q)^{1/q} = \|z_{I_+^t}^{(t)} - x_{I_+^t}^*\|_q, \tag{31}$$

where I_+^t is the index set of the largest $k + 1$ components of $z^{(t)}$ in magnitude.

Proof: When $q = \infty$, we need to show

$$\tau^{(t)} \leq \max_{i \in I_+^t} |z_i^{(t)} - x_i^*|, \tag{32}$$

then Lemma 1 shows that (31) holds for all $q \in [1, \infty]$.

Let I^t be the index set of the largest k components of $z^{(t)}$ in magnitude, then $I_+^t = I^t \cup \{[k + 1]\}$, where $[k + 1]$ represents the index of the $(k + 1)$-th largest component of $z^{(t)}$ in magnitude. We will prove (32) in the following two cases.

Case (i). If $I^* = I^t$, then

$$\tau^{(t)} = |z_{[k+1]}^{(t)}| = |z_{[k+1]}^{(t)} - x_{[k+1]}^*|$$
$$\leq \max_{i \in I_+^t} |z_i^{(t)} - x_i^*|. \tag{33}$$

Case (ii). If $I^* \neq I^t$, then there exists $i_0 \in I^t$ such that

$$i_0 \notin I^*.$$

Otherwise $I^t \subset I^*$ and $I^t \neq I^*$ which contradicts with $|I^t| \geq k^*$ and $|I^*| = k^*$. Thus, $x_{i_0}^* = 0$ and

$$\tau^{(t)} = |z_{[k+1]}^{(t)}| \leq |z_{i_0}^{(t)}| = |z_{i_0}^{(t)} - x_{i_0}^*|$$
$$\leq \max_{i \in I_+^t} |z_i^{(t)} - x_i^*|. \tag{34}$$

Combining (33) and (34) gives (32). ∎

Proof of Theorem 2: In order to prove this theorem, we only need to justify the following two inequalities, i.e., for any $t \in \mathbf{N}$,

$$\|z_{S^t}^{(t+1)} - x_{S^t}^*\|_q \leq \gamma_s\|x^{(t)} - x^*\|_p + s\|A^T\epsilon\|_q, \tag{35}$$

and for any $t \geq 1$,

$$\|x^{(t)} - x^*\|_p \leq L\|z_{S^{t-1}}^{(t)} - x_{S^{t-1}}^*\|_q. \tag{36}$$

Then combining (35) and (36), it holds

$$\|x^{(t+1)} - x^*\|_p \leq L\|z_{S^t}^{(t+1)} - x_{S^t}^*\|_q$$
$$\leq \rho_s\|x^{(t)} - x^*\|_p + sL\|A^T\epsilon\|_q.$$

Since $0 < \rho_s < 1$ under the assumption of this theorem, then by induction for any $t \geq 1$, we have

$$\|x^{(t)} - x^*\|_p \leq (\rho_s)^t\|x^* - x^{(0)}\|_p + \frac{sL}{1-\rho_s}\|A^T\epsilon\|_q.$$

First, we turn to prove the inequality (35). By the **Step 1** of Algorithm 1, for any $t \in \mathbf{N}$,

$$z^{(t+1)} = x^{(t)} - sA^T(Ax^{(t)} - b),$$

and we note that $b = Ax^* + \epsilon$, then

$$z^{(t+1)} - x^* = (\mathbf{I}_n - sA^T A)(x^{(t)} - x^*) + sA^T\epsilon$$
$$= (1 - s)(x^{(t)} - x^*) + s(\mathbf{I}_n - A^T A)(x^{(t)} - x^*) + sA^T\epsilon.$$

For any $t \in \mathbf{N}$ and $q \in [1, \infty]$, let $S^t = I_+^{t+1} \bigcup I^t \bigcup I^*$. Noting that $I^t, I^* \subset S^t$, it follows

$$A(x^{(t)} - x^*) = A_{S^t}(x_{S^t}^{(t)} - x_{S^t}^*).$$

Then we have

$$z_{S^t}^{(t+1)} - x_{S^t}^* = (1 - s)(x_{S^t}^{(t)} - x_{S^t}^*)$$
$$+ s(\mathbf{I}_{|S^t|} - A_{S^t}^T A_{S^t})(x_{S^t}^{(t)} - x_{S^t}^*) + sA_{S^t}^T\epsilon.$$

Therefore,

$$\|z_{S^t}^{(t+1)} - x_{S^t}^*\|_q \leq |1 - s| \cdot \|x_{S^t}^{(t)} - x_{S^t}^*\|_q$$
$$+ s\|(\mathbf{I}_{|S^t|} - A_{S^t}^T A_{S^t})(x_{S^t}^{(t)} - x_{S^t}^*)\|_q + s\|A_{S^t}^T\epsilon\|_q. \tag{37}$$

Since $\|x^{(t)}\|_0 \leq k = k^*$ and $\|x^*\|_0 = k^*$ then

$$|I^t| \leq k^*, |I_+^{t+1}| \leq k^* + 1, |I^*| = k^*,$$

and hence $|S^t| \le 3k^* + 1$. For any $p \in [1, \infty)$, by (14) and the definition of gRIP (8), it holds

$$\|x^{(t)} - x^*\|_q \le (2k^*)^{\max\{\frac{1}{q} - \frac{1}{p}, 0\}} \|x^{(t)} - x^*\|_p, \qquad (38)$$

and

$$\|(\mathbf{I}_{|S^t|} - A_{S^t}^T A_{S^t})(x_{S^t}^{(t)} - x_{S^t}^*)\|_q$$
$$\le \beta_{3k^*+1, p, q} \|x_{S^t}^{(t)} - x_{S^t}^*\|_p = \beta_{3k^*+1, p, q} \|x^{(t)} - x^*\|_p. (39)$$

Plugging (38) and (39) into (37), then

$$\|z_{S^t}^{(t+1)} - x_{S^t}^*\|_q$$
$$\le \left(|1 - s|(2k^*)^{\max\{\frac{1}{q} - \frac{1}{p}, 0\}} + s\beta_{3k^*+1, p, q} \right) \|x^{(t)} - x^*\|_p$$
$$+ s\|A_{S^t}^T \epsilon\|_q$$
$$\le \gamma_s \|x^{(t)} - x^*\|_p + s\|A^T \epsilon\|_q.$$

Thus, we have obtained the inequality (35).

Then we turn to the proof of (36). We will prove it in two steps.

Step a): For any $p \in [1, \infty)$,

$$\|x^{(t)} - x^*\|_p^p = \|x_{I^*}^{(t)} - x_{I^*}^*\|_p^p + \|x_{I^t \setminus I^*}^{(t)}\|_p^p. \qquad (40)$$

By Lemma 2,

$$\|x_{I^*}^{(t)} - x_{I^*}^*\|_p^p = \|x_{I^*}^{(t)} - z_{I^*}^{(t)} + z_{I^*}^{(t)} - x_{I^*}^*\|_p^p$$
$$\le 2^{p-1}\|z_{I^*}^{(t)} - x_{I^*}^{(t)}\|_p^p + 2^{p-1}\|z_{I^*}^{(t)} - x_{I^*}^*\|_p^p. \qquad (41)$$

Moreover, by the *Step 3* of Algorithm 1 and Assumption 1, for any $i \in I^t$:

$$sgn(x_i^{(t)}) = sgn(z_i^{(t)}) \text{ and } |x_i^{(t)}| \le |z_i^{(t)}|.$$

Thus, for any $i \in I^t \setminus I^*$, it holds

$$x_i^{(t)} \cdot (z_i^{(t)} - x_i^{(t)}) \ge 0. \qquad (42)$$

With (42) and by Lemma 2, we have

$$\|z_{I^t \setminus I^*}^{(t)}\|_p^p = \|x_{I^t \setminus I^*}^{(t)} + (z_{I^t \setminus I^*}^{(t)} - x_{I^t \setminus I^*}^{(t)})\|_p^p.$$
$$\ge \|x_{I^t \setminus I^*}^{(t)}\|_p^p + \|z_{I^t \setminus I^*}^{(t)} - x_{I^t \setminus I^*}^{(t)}\|_p^p. \quad (43)$$

Plugging (41) and (43) into (40), it becomes

$$\|x^{(t)} - x^*\|_p^p \le 2^{p-1}(\|z_{I^*}^{(t)} - x_{I^*}^{(t)}\|_p^p + \|z_{I^*}^{(t)} - x_{I^*}^*\|_p^p)$$
$$+ \|z_{I^t \setminus I^*}^{(t)}\|_p^p - \|z_{I^t \setminus I^*}^{(t)} - x_{I^t \setminus I^*}^{(t)}\|_p^p. \quad (44)$$

Furthermore, by the *Step 2* of Algorithm 1, Assumption 1 and Lemma 3, for any $t \ge 1$, we have:

(a) if $i \in I^t$, $c_2 \tau^{(t)} \le |z_i^{(t)} - x_i^{(t)}| \le c_1 \tau^{(t)} \le \tau^{(t)}$;
(b) if $i \notin I^t$, $|z_i^{(t)} - x_i^{(t)}| = |z_i^{(t)}| \le \tau^{(t)}$;
(c) $\tau^{(t)} \le \|z_{I_+^t}^{(t)} - x_{I_+^t}^*\|_q$.

By the above facts (a)-(c), it holds

$$\|z_{I^*}^{(t)} - x_{I^*}^{(t)}\|_p^p \le k^* \max_{i \in I^*} |z_i^{(t)} - x_i^{(t)}|^p \le k^* |\tau^{(t)}|^p, \quad (45)$$

and

$$\|z_{I^t \setminus I^*}^{(t)} - x_{I^t \setminus I^*}^{(t)}\|_p^p \ge |I^t \setminus I^*| \min_{i \in I^t \setminus I^*} |z_i^{(t)} - x_i^{(t)}|^p$$
$$\ge |I^t \setminus I^*|(c_2)^p |\tau^{(t)}|^p, \qquad (46)$$

where $|I^t \setminus I^*|$ represents the cardinality of the index set $I^t \setminus I^*$. Plugging (45), (46) into (44), it follows

$$\|x^{(t)} - x^*\|_p^p \le 2^{p-1}\|z_{I^*}^{(t)} - x_{I^*}^*\|_p^p + \|z_{I^t \setminus I^*}^{(t)}\|_p^p$$
$$+ (2^{p-1}k^* - (c_2)^p |I^t \setminus I^*|)|\tau^{(t)}|^p. \quad (47)$$

Furthermore, we note that

$$\|z_{I^t \setminus I^*}^{(t)}\|_p^p = \|z_{I^t \setminus I^*}^{(t)} - x_{I^t \setminus I^*}^*\|_p^p$$
$$\le |I^t \setminus I^*| \max_{i \in I^t \setminus I^*} |z_i^{(t)} - x_i^*|^p$$
$$= |I^t \setminus I^*| \cdot \|z_{I^t \setminus I^*}^{(t)} - x_{I^t \setminus I^*}^*\|_\infty^p$$
$$\le |I^t \setminus I^*| \cdot \|z_{I^t \setminus I^*}^{(t)} - x_{I^t \setminus I^*}^*\|_q^p,$$

where the first equality holds because $x_{I^t \setminus I^*}^* = 0$, and the second inequality holds because of Lemma 1. Therefore, (47) becomes

$$\|x^{(t)} - x^*\|_p^p$$
$$\le 2^{p-1}\|z_{I^*}^{(t)} - x_{I^*}^*\|_p^p + |I^t \setminus I^*| \cdot \|z_{I^t \setminus I^*}^{(t)} - x_{I^t \setminus I^*}^*\|_q^p$$
$$+ \left(2^{p-1}k^* - (c_2)^p |I^t \setminus I^*|\right)|\tau^{(t)}|^p$$
$$\le 2^{p-1}\|z_{I^*}^{(t)} - x_{I^*}^*\|_p^p$$
$$+ (2^{p-1}k^* - (c_2)^p |I^t \setminus I^*| + |I^t \setminus I^*|)\|z_{I_+^t}^{(t)} - x_{I_+^t}^*\|_q^p$$
$$\le 2^{p-1}(k^*)^{\max\{1 - \frac{p}{q}, 0\}}\|z_{I^*}^{(t)} - x_{I^*}^*\|_q^p$$
$$+ (2^{p-1} - (c_2)^p + 1)k^* \|z_{I_+^t}^{(t)} - x_{I_+^t}^*\|_q^p$$
$$\le L_1 \|z_{S^{t-1}}^{(t)} - x_{S^{t-1}}^*\|_q^p, \qquad (48)$$

where the second inequality holds by the fact (c), i.e., $\tau^{(t)} \le \|z_{I_+^t}^{(t)} - x_{I_+^t}^*\|_q$, the third inequality holds by Lemma 1 and $|I^t \setminus I^*| \le k^*$ and the last inequality holds because $S^{t-1} = I_+^t \cup I^{t-1} \cup I^*$. Thus, it implies

$$\|x^{(t)} - x^*\|_p \le \sqrt[p]{L_1}\|z_{S^{t-1}}^{(t)} - x_{S^{t-1}}^*\|_q. \qquad (49)$$

Step b): By Lemma 2,

$$\|x^{(t)} - x^*\|_p^p = \|x_{I^t}^{(t)} - x_{I^t}^*\|_p^p + \|x_{I^* \setminus I^t}^*\|_p^p$$
$$\le 2^{p-1}\|z_{I^t}^{(t)} - x_{I^t}^*\|_p^p + 2^{p-1}\|z_{I^t}^{(t)} - x_{I^t}^{(t)}\|_p^p$$
$$+ 2^{p-1}\|z_{I^* \setminus I^t}^{(t)} - x_{I^* \setminus I^t}^*\|_p^p + 2^{p-1}\|z_{I^* \setminus I^t}^{(t)}\|_p^p$$
$$= 2^{p-1}\|z_{I^t \cup I^*}^{(t)} - x_{I^t \cup I^*}^*\|_p^p$$
$$+ 2^{p-1}(\|z_{I^t}^{(t)} - x_{I^t}^{(t)}\|_p^p + \|z_{I^* \setminus I^t}^{(t)}\|_p^p). \quad (50)$$

Moreover, by Lemma 1, it holds

$$\|z_{I^t \cup I^*}^{(t)} - x_{I^t \cup I^*}^*\|_p^p$$
$$\le (|I^t \cup I^*|)^{\max\{1 - \frac{p}{q}, 0\}}\|z_{I^t \cup I^*}^{(t)} - x_{I^t \cup I^*}^*\|_q^p$$
$$\le (2k^*)^{\max\{1 - \frac{p}{q}, 0\}}\|z_{I^t \cup I^*}^{(t)} - x_{I^t \cup I^*}^*\|_q^p, \quad (51)$$

where the last inequality holds for $|I^t \cup I^*| \le 2k^*$. We also have

$$\|z_{I^t}^{(t)} - x_{I^t}^{(t)}\|_p^p \le k^* \max_{i \in I^t} |z_i^{(t)} - x_i^{(t)}|^p$$
$$\le k^*(c_1 \tau^{(t)})^p \le k^*(c_1)^p \|z_{I_+^t}^{(t)} - x_{I_+^t}^*\|_q^p. \quad (52)$$

Since $|I^t| = |I^*| = k^*$, then

$$|I^* \setminus I^t| = |I^t \setminus I^*|.$$

Thus, it holds

$$
\begin{aligned}
\|z_{I^* \setminus I^t}^{(t)}\|_p^p &\leq |I^* \setminus I^t| \max_{i \in I^* \setminus I^t} |z_i^{(t)}|^p \leq |I^* \setminus I^t| \cdot |\tau^{(t)}|^p \\
&= |I^t \setminus I^*| \cdot |\tau^{(t)}|^p \leq |I^t \setminus I^*| \min_{i \in I^t \setminus I^*} |z_i^{(t)}|^p \\
&\leq \|z_{I^t \setminus I^*}^{(t)}\|_p^p = \|z_{I^t \setminus I^*}^{(t)} - x_{I^t \setminus I^*}^*\|_p^p \\
&\leq (k^*)^{\max\{1 - \frac{p}{q}, 0\}} \|z_{I^t \setminus I^*}^{(t)} - x_{I^t \setminus I^*}^*\|_q^p. \quad (53)
\end{aligned}
$$

Plugging (51), (52) and (53) into (50), and further since $S^{t-1} = I_+^t \cup I^{t-1} \cup I^*$, and thus $I_+^t \subset S^{t-1}, I^t \subset I_+^t \subset S^{t-1}, I^t \cup I^* \subset S^{t-1}, I^t \setminus I^* \subset S^{t-1}$, it becomes

$$
\begin{aligned}
&\|x^{(t)} - x^*\|_p^p \\
&\leq (2^p (2k^*)^{\max\{1 - \frac{p}{q}, 0\}} + 2^{p-1} (c_1)^p k^*) \|z_{S^{t-1}}^{(t)} - x_{S^{t-1}}^*\|_q^p \\
&= L_2 \|z_{S^{t-1}}^{(t)} - x_{S^{t-1}}^*\|_q^p. \quad (54)
\end{aligned}
$$

Thus, we have

$$\|x^{(t)} - x^*\|_p \leq \sqrt[p]{L_2} \|z^{(t)} - x^*\|_q. \quad (55)$$

From (49) and (55), for any $t \geq 1$, it holds

$$
\begin{aligned}
\|x^{(t)} - x^*\|_p &\leq \min\{\sqrt[p]{L_1}, \sqrt[p]{L_2}\} \|z^{(t)} - x^*\|_q \\
&= L \|z^{(t)} - x^*\|_q \\
&= L \|z_{S^{t-1}}^{(t)} - x_{S^{t-1}}^*\|_q, \quad (56)
\end{aligned}
$$

where the last equality holds for $S^{t-1} = I_+^t \cup I^{t-1} \cup I^*$. Thus, we have obtained (36).

Therefore, we end the proof of this theorem. ∎

APPENDIX E: PROOF OF THEOREM 3

Proof: The proof is similar to that of Theorem 2. According to the proof of Theorem 2, we have known that (37)-(39) hold for all pairs of (p, q) with $\frac{1}{p} + \frac{1}{q} = 1$, and thus obviously hold for $p = 1$ and $q = \infty$. In the following, instead of the inequality (35), we will derive a tighter upper bound of $\|z_{S^t}^{(t+1)} - x_{S^t}^*\|_\infty$, that is,

$$
\begin{aligned}
&\|z_{S^t}^{(t+1)} - x_{S^t}^*\|_\infty \\
&\leq \max\{\mu s, |1 - s|\} \|x^{(t)} - x^*\|_1 + s\|A^T \epsilon\|_\infty. \quad (57)
\end{aligned}
$$

Now we turn to prove the inequality (57). According to (4), it can be observed that

$$
\begin{aligned}
&\|z^{(t+1)} - x^*\|_\infty \leq \\
&\| \left((1 - s)\mathbf{I}_n + s(\mathbf{I}_n - A^T A)\right) (x^{(t)} - x^*)\|_\infty + s\|A^T \epsilon\|_\infty.
\end{aligned}
$$

Let $B = (1 - s)\mathbf{I}_n + s(\mathbf{I}_n - A^T A)$ and B_{ij} be the (i, j)-th element of B. Since $\|A_j\|_2 = 1$ for all $j = 1, \ldots, n$, then

$$B_{ii} = 1 - s,$$

for all $i = 1, \ldots, n$. Moreover, by the definition of the coherence μ, the absolutes of all the off-diagonal elements of $\mathbf{I}_n - A^T A$ are no bigger than μ. Thus,

$$|B_{ij}| \leq s\mu,$$

for any $i \neq j$. As a consequence, it holds

$$\max_{i,j \in \{1, \ldots, n\}} |B_{ij}| \leq \max\{|1 - s|, s\mu\} = \gamma_s.$$

Furthermore, for any $i = 1, \ldots, n$,

$$
\begin{aligned}
|z_i^{(t+1)} - x_i^*| &= \left| \sum_{j=1}^n B_{ij}(x_j^{(t)} - x_j^*) + sA_i^T \epsilon \right| \\
&\leq \left| \sum_{j=1}^n B_{ij}(x_j^{(t)} - x_j^*) \right| + s\|A^T \epsilon\|_\infty \\
&\leq \gamma_s \|x^{(t)} - x^*\|_1 + s\|A^T \epsilon\|_\infty.
\end{aligned}
$$

This implies

$$\|z_{S^t}^{(t+1)} - x_{S^t}^*\|_\infty \leq \gamma_s \|x^{(t)} - x^*\|_1 + s\|A^T \epsilon\|_\infty.$$

Therefore, we obtain the (57). According to the proof of Theorem 2, we have that the inequality (36) still holds when $p = 1$ and $q = \infty$, that is,

$$\|x^{(t)} - x^*\|_1 \leq L \|z_{S^{t-1}}^{(t)} - x_{S^{t-1}}^*\|_\infty. \quad (58)$$

Similar to the rest of the proof of Theorem 2, combining (57) and (58), we can conclude the proof of this theorem. ∎

APPENDIX F: PROOF OF THEOREM 4

Proof: The proof of this theorem is also very similar to that of Theorem 2. According to the proof of Theorem 2, we have known that (35) holds for all pairs of (p, q) with $\frac{1}{p} + \frac{1}{q} = 1$, and thus obviously holds for $p = 2$ and $q = 2$, that is,

$$\|z_{S^t}^{(t+1)} - x_{S^t}^*\|_2 \leq \delta_{3k^*+1} \|x^{(t)} - x^*\|_2 + \|A^T \epsilon\|_2, \quad (59)$$

where $S^t = I_+^{t+1} \cup I^t \cup I^*$, I_+^{t+1} is the index set of the largest $k + 1$ components of $z^{(t+1)}$, I^t and I^* represent the support sets of $x^{(t)}$ and x^*, respectively. In the following, instead of the inequality (36), we will derive a tighter upper bound of $\|x^{(t)} - x^*\|_2$, that is,

$$\|x^{(t)} - x^*\|_2 \leq \frac{\sqrt{5} + 1}{2} \|z_{S^{t-1}}^{(t)} - x_{S^{t-1}}^*\|_2. \quad (60)$$

Now we turn to prove the inequality (60). It can be noted that

$$\|x^{(t)} - x^*\|_2^2 = \|x_{I^t}^{(t)} - x_{I^t}^*\|_2^2 + \|x_{I^* \setminus I^t}^{(t)} - x_{I^* \setminus I^t}^*\|_2^2. \quad (61)$$

On one hand, since $x_i^{(t)} = z_i^{(t)}$ for any $i \in I^t$, then

$$\|x_{I^t}^{(t)} - x_{I^t}^*\|_2^2 = \|z_{I^t}^{(t)} - x_{I^t}^*\|_2^2. \quad (62)$$

On the other hand, we can also observe that $x_i^{(t)} = 0$ for any

$i \in I^* \setminus I^t$, and thus

$$\|x_{I^* \setminus I^t}^{(t)} - x_{I^* \setminus I^t}^*\|_2^2 = \|x_{I^* \setminus I^t}^*\|_2^2 = \sum_{i \in I^* \setminus I^t} (x_i^* - z_i^{(t)} + z_i^{(t)})^2$$

$$\leq \sum_{i \in I^* \setminus I^t} \left(\frac{\sqrt{5}+3}{2}(x_i^* - z_i^{(t)})^2 + \frac{\sqrt{5}+1}{2}(z_i^{(t)})^2 \right)$$

$$\leq \sum_{i \in I^* \setminus I^t} \frac{\sqrt{5}+3}{2}(x_i^* - z_i^{(t)})^2 + \sum_{i \in I^t \setminus I^*} \frac{\sqrt{5}+1}{2}(z_i^{(t)})^2$$

$$= \frac{\sqrt{5}+3}{2}\|z_{I^* \setminus I^t}^{(t)} - x_{I^* \setminus I^t}^*\|_2^2 + \frac{\sqrt{5}+1}{2}\|z_{I^t \setminus I^*}^{(t)}\|_2^2$$

$$= \frac{\sqrt{5}+3}{2}\|z_{I^* \setminus I^t}^{(t)} - x_{I^* \setminus I^t}^*\|_2^2 + \frac{\sqrt{5}+1}{2}\|z_{I^t \setminus I^*}^{(t)} - x_{I^t \setminus I^*}^*\|_2^2. \tag{63}$$

The first inequality holds by the following relation

$$(a+b)^2 = a^2 + b^2 + 2ab \leq (1 + \frac{\sqrt{5}+1}{2})a^2 + (1 + \frac{\sqrt{5}-1}{2})b^2$$

for any $a, b \in \mathbf{R}$. The second inequality holds due to the following facts:

  (a) for any $i \in I^* \setminus I^t$, $|z_i^t| \leq \tau^{(t)}$,
  (b) for any $i \in I^t \setminus I^*$, $|z_i^t| \geq \tau^{(t)}$,
  (c) $|I^* \setminus I^t| = |I^t \setminus I^*|$,
and hence
$$\max_{i \in I^* \setminus I^t} |z_i^{(t)}| \leq \min_{i \in I^t \setminus I^*} |z_i^{(t)}|.$$

The last equality holds for $x_i^* = 0, \forall i \in I^t \setminus I^*$. Plugging (62) and (63) into (61), we have

$$\|x^{(t)} - x^*\|_2^2 \leq \|z_{I^t}^{(t)} - x_{I^t}^*\|_2^2 + \frac{\sqrt{5}+1}{2}\|z_{I^t \setminus I^*}^{(t)} - x_{I^t \setminus I^*}^*\|_2^2$$

$$+ \frac{\sqrt{5}+3}{2}\|z_{I^* \setminus I^t}^{(t)} - x_{I^* \setminus I^t}^*\|_2^2$$

$$= \|z_{I^t \cap I^*}^{(t)} - x_{I^t \cap I^*}^*\|_2^2 + \frac{\sqrt{5}+3}{2}\|z_{I^t \setminus I^*}^{(t)} - x_{I^t \setminus I^*}^*\|_2^2$$

$$+ \frac{\sqrt{5}+3}{2}\|z_{I^* \setminus I^t}^{(t)} - x_{I^* \setminus I^t}^*\|_2^2$$

$$\leq \frac{\sqrt{5}+3}{2}\|z_{S^{t-1}}^{(t)} - x_{S^{t-1}}^*\|_2^2,$$

where $S^{t-1} = I_+^t \cup I^{t-1} \cup I^*$. The last inequality holds because the sets $I^t \cap I^*$, $I^t \setminus I^*$ and $I^* \setminus I^t$ do not intersect with each other and

$$(I^t \cap I^*) \cup (I^t \setminus I^*) \cup (I^* \setminus I^t) = (I^t \cup I^*) \subset (I_+^t \cup I^*) \subset S^{t-1},$$

and $\frac{\sqrt{5}+3}{2} > 1$. Therefore, the above inequality implies (60). Similar to the rest of the proof of Theorem 2, combining (59) and (60), we can conclude the proof of this theorem. ∎

REFERENCES

[1] D.L. Donoho, Compressed sensing, IEEE Transactions on Information Theory, 52 (4): 1289-1306, 2006.

[2] E.J. Candes, J. Romberg, and T. Tao, Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information, IEEE Transactions on Information Theory, 52 (2): 489-509, 2006.

[3] S. Mallat and Z. Zhang, Matching pursuits with time-frequency dictionaries, IEEE Transactions on Signal Processing, 41 (12): 3397-3415, 1993.

[4] S. Foucart, Sparse recovery algorithms: Sufficient conditions in terms of restricted isometry constants, in Proceedings of the 13th International Conference on Approximation Theory, M. Neantu and L. Schumaker, eds., San Antonio, TX, 2010, Springer.

[5] V. Cevher, On accelerated hard thresholding methods for sparse approximation, Technical report, http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.205.725, 2011.

[6] A. Kyrillidis and V. Cevher, Recipes on hard thresholding methods, in Proceedings of the 4th IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP), pp. 353-356, 2011.

[7] S. Foucart, Hard thresholding pursuit: an algorithm for compressive sensing, SIAM Journal on Numerical Analysis, 49(6), 2543-2563, 2011.

[8] A. Kyrillidis and V. Cevher, Combinatorial selection and least absolute shrinkage via the CLASH algorithm, 2012 IEEE International Symposium on Information Theory Proceedings (ISIT), pp. 2216–2220, 2012.

[9] S. Foucart and H. Rauhut, A mathematical introduction to compressive sensing, Berlin: Springer, 2013.

[10] Y. Pati, R. Rezaifar and P. Krishnaprasad, Orthogonal matching pursuit: recursive function approximatin with applications to wavelet decomposition, In Asilomar Conf. Signals, Syst., Comput., Pacific Grove, CA, 1993.

[11] J.A. Tropp and A. Gilbert, Signal recovery from random measurements via orthogonal mathching pursuit, IEEE Transactions on Information Theory, 2007, 53: 4655-4666.

[12] D.L. Donoho, Y. Tsaig, O. Drori and J.-L. Starck, Sparse solution of underdetermined systems of linear equations by stagewise orthogonal matching pursuit, IEEE Transactions on Information Theory, 58 (2): 1094 - 1121, 2012.

[13] D. Needell and R. Vershynin, Signal recovery from incomplete and inaccurate measurements via Regularized Orthogonal Matching Pursuit, IEEE Journal of Selected Topics in Signal Processing, 4: 310-316, 2010.

[14] D. Needell and J.A. Tropp, CoSaMP: Iterative signal recovery from incomplete and inaccurate samples, Applied and Computational Harmonic Analysis, 26 (3): 301-321, 2008.

[15] W. Dai and O. Milenkovic, Subspace pursuit for compressive sensing signal recontruction, IEEE Transactions on Information Theory, 55 (5): 2230-2249, 2009.

[16] S.S. Chen, D.L. Donoho, and M. A. Saunders, Atomic decomposition by basis pursuit, SIAM Journal on Scientific Computing, 20: 33-61, 1998.

[17] R. Chartrand, Exact reconstruction of sparse signals via nonconvex minimization. IEEE Signal Processing Letters, 14 (10): 707-710, 2007.

[18] R. Chartrand and V. Staneva, Restricted isometry properties and nonconvex compressive sensing, Inverse Problems, 24: 1-14, 2008.

[19] Z.B. Xu, X.Y. Chang, F.M. Xu and H. Zhang, $L_{1/2}$ regularization: a thresholding representation theory and a fast solver, IEEE Transactions on Neural Networks and Learning Systems, 23: 1013-1027, 2012.

[20] E.J. Candes, M.B. Wakin and S.P. Boyd, Enhancing sparsity by reweighted ℓ_1 minimization, Journal of Fourier Analysis and Applications, 14 (5): 877-905, 2008.

[21] J. Fan, and R. Li, Variable selection via nonconcave penalized likelihood and its oracle properties, Journal of the American Statistical Association, 96: 1348-1360, 2001.

[22] C.H. Zhang, Nearly unbiased variable selection under minimax concave penalty, The Annals of Statistics, 38 (2): 894-942, 2010.

[23] I.F. Gorodnitsky and B.D. Rao, Sparse signal reconstruction from limited data using FOCUSS: a re-weighted minimum norm algorithm, IEEE Transactions on Signal Processing, 45 (3): 600-616, 1997.

[24] I. Daubechies, R. Devore, M. Fornasier and C.S. Gunturk, Iteratively reweighted least squares minimization for sparse recovery, Communications on Pure and Applied Mathematics, 63: 1-38, 2010.

[25] I. Daubechies, M. Defries and C. De Mol, An iterative thresholding algorithm for linear inverse problems with a sparisity constraint, Communications on Pure and Applied Mathematics, 57: 1413-1457, 2004.

[26] W.F. Cao, J. Sun and Z.B. Xu, Fast image deconvolution using closed-form thresholding formulas of L_q ($q = 1/2, 2/3$) regularization, Journal of Visual Communication and Image Representation, 24: 31-41, 2013.

[27] T. Blumensath and M.E. Davies, Iterative thresholding for sparse approximation, Journal of Fourier Analysis and Application, 14 (5): 629-654, 2008.

[28] J.S. Zeng, S.B. Lin, Y. Wang and Z.B. Xu, $L_{1/2}$ regularization: convergence of iterative Half thresholding algorithm, IEEE Transactions on Signal Processing, 62(9): 2317-2329, 2014.

[29] J.S. Zeng, S.B. Lin and Z.B. Xu, Sparse regularization: convergence of iterative jumping thresholding algorithm, http://arxiv.org/abs/1402.5744, 2014.

[30] Y.T. Qian, S. Jia, J. Zhou and A. Robles-Kelly, Hyperspectral unmixing via $L_{1/2}$ sparsity-constrained nonnegative matrix factorization, IEEE Transactions on Geoscience and Remote Sensing, 49 (11): 4282-4297, 2011.

[31] J.S. Zeng, J. Fang, Z. B. Xu, Sparse SAR imaging based on $L_{1/2}$ regularization, Science China Information Sciences, 55: 1755-1775, 2012.

[32] J.S. Zeng, Z. B. Xu, B.C. Zhang, W. Hong, Y.R. Wu. Accelerated $L_{1/2}$ regularization based SAR imaging via BCR and reduced Newton skills, Signal Processing, 93: 1831-1844, 2013.

[33] T. Blumensath and M.E. Davies, Iterative hard thresholding for compressed sensing, Applied and Computational Harmonic Analysis, 27: 265-274, 2008.

[34] A. Maleki, Coherence analysis of iteative thresholding algorithms, in Forty-Seventh Annual Allerton Conference, Allerton House, UIUC, Illinois, USA, 2009.

[35] J.S. Zeng, S.B. Lin and Z.B. Xu, Sparse solution of underdetermined linear equations via adaptively iterative thresholding, Signal Processing, 97: 152-161, 2014.

[36] T.T. Cai, G. Xu and J. Zhang, On recovery of sparse signals via ℓ_1 minimization, IEEE Transactions on Information Theory, 55 (7): 3388-3397, 2009.

[37] E.M. Candes, The restricted isometry property and its implications for compressed sensing, Comptes Rendus Mathematique, 346(9):589-592.

[38] J.A. Tropp, Greed is Good: Algorithmic Results for Sparse Approximation, IEEE Transactions on Information Theory, 50 (10): 2231-2242, 2004.

[39] D.L. Donoho and M. Elad, Optimally sparse representation in general (nonorthogonal) dictionaries via ℓ_1 minimization, Proceedings of the National Academy of Sciences, 100 (5): 2197-2202, 2003.

[40] M.B. Wakin and M.A. Davenport, Analysis of orthogonal matching pursuit using the restricted isometry property, IEEE Transactions on Information Theory, 56 (9): 4395-4401, 2010.

[41] T.T. Cai, and A. R. Zhang, Sparse representation of a polytope and recovery of a sparse signals and low-rank matrices, IEEE Transactions on Information Theory, 60(1): 122 - 132, 2014.

[42] M.J. Lai, Y.Y. Xu, and W.T. Yin, Improved iteratively reweighted least squares for unconstrained smoothed ℓ_q minimization, SIAM J. Numer. Anal., 51: 927-957, 2013.

[43] X. Chen and W. Zhou, Convergence of the reweighted l_1 minimization algorithm for l_2-l_p minimization, Comput. Optim. Appl., 59: 47-61, 2014.

[44] M. Rudelson and R. Vershynin, On sparse reconstruction from Fourier and Gaussian measurements. Comm. Pure Appl. Math., 61: 1025-1045, 2008.

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

[46] T. Blumensath, M.E. Davis, Normalised Iterative Hard Thresholding; guaranteed stability and performance, IEEE Journal of Selected Topics in Signal Processing, 4(2): 298-309, 2010.

[47] J.D. Blanchard, and J. Tanner, Performance comparisons of greedy algorithms in compressed sensing, http://people.maths.ox.ac.uk/tanner/papers/PCGACS.pdf, 2013.

[48] D.L. Donoho, High-dimensional centrally symmetric polytopes with neighborliness proportional to dimension, Discrete & Computational Geometry, 35(4): 617-652, 2006.

[49] B.L. Sturm, M.G. Christensen, R. Gribonval, Cyclic pure greedy algorithms for recovering compressively sampled sparse signals, 2011 Conference Record of the Forty Fifth Asilomar Conference on Signals, Systems and Computers (ASILOMAR), pp.1143-1147, Nov. 2011.

[50] J.D. Blanchard, J. Tanner, and K. Wei, CGIHT: conjugate gradient iterative hard thresholding for compressed sensing and matrix completion, IEEE Transactions on Signal Processing, 63(2): 528-537, 2015.

[51] A. Maleki, D.L. Donoho, Optimally tuned iterative reconstruction algorithms for compressed sensing. IEEE Journal of Selected Topics in Signal Processing, 4(2):330-341, 2010

[52] G. Rath and C. Guillemot, A complementary matching pursuit algorithm for sparse approximation, in Proceedings of European Signal Process, Conference, Lausanne, Switzerland, Aug. 2008.

[53] L. Rebollo-Neira and D. Lowe, Optimized orthogonal matching pursuit approach, IEEE Signal Processing Letters, 9(4):137-140, 2002.

Yu Wang received the B.Sc. degree in Applied Mathematics in 2013 in Xi'an Jiaotong University, Shaanxi, China. From 2007 to 2009, he was a member of the Special Class of the Gifted Young in Xi'an Jiaotong University. From 2009 to 2013, he was in the Science Topnotch Program in China. He is currently pursuing M.Sc. degree in Applied Mathematics in Xi'an Jiaotong University.

Jinshan Zeng received the B.S. degree in Information and Computing Sciencises from Xi'an Jiaotong University, Xi'an, in 2008. He is currently pursuing the Ph.D. degree with the School of Mathematics and Statistics, Xi'an Jiaotong University. He has been with the Department of Mathematics, University of California, Los Angeles, as a Visiting Scholar since Nov. 2013. His current research interests include sparse optimization, signal processing and synthetic aperture radar imaging.

Zhimin Peng received his B.S. in computational mathematics from Xi'an Jiaotong University in 2011, and then M.A. in applied math from Rice University in 2013. Now he is a Ph.D. student in the Department of Mathematics at UCLA. His research interests include developing efficient algorithms for anomaly detection and solving large scale convex optimization problems.

Xiangyu Chang received the Ph.D. degree in applied mathematics from Xi'an Jiaotong University, China, in 2014. He is currently an assistant professor at the school of management in Xi'an Jiaotong University, China. His current research interests include statistical machine learning, high-dimensional statistics, and social network analysis.

Zongben Xu received his Ph.D. degree in mathematics from Xi'an Jiaotong University, China, in 1987. He now serves as the Chief Scientist of National Basic Research Program of China (973 Project), and Director of the Institute for Information and System Sciences of the university. He is owner of the National Natural Science Award of China in 2007,and winner of CSIAM Su Buchin Applied Mathematics Prize in 2008. He delivered a 45 minute talk on the International Congress of Mathematicians 2010. He was elected as member of Chinese Academy of Science in 2011. His current research interests include intelligent information processing and applied mathematics.

Gaussian

0−ALPS(4)	
CoSaMP	
NIHT	
CGIHT	
NSCAD	
NHalf	

Binary

0−ALPS(4)	
CoSaMP	
NIHT	
CGIHT	
NSCAD	
NHalf	

Gaussian

Binary