

# Oracle-order Recovery Performance of Greedy Pursuits with Replacement against General Perturbations

Laming Chen and Yuantao Gu\*

Received Dec. 30, 2011; revised Jan. 20, 2013.

## Abstract

Applying the theory of compressive sensing in practice always takes different kinds of perturbations into consideration. In this paper, the recovery performance of greedy pursuits with replacement for sparse recovery is analyzed when both the measurement vector and the sensing matrix are contaminated with additive perturbations. Specifically, greedy pursuits with replacement include three algorithms, compressive sampling matching pursuit (CoSaMP), subspace pursuit (SP), and iterative hard thresholding (IHT), where the support estimation is evaluated and updated in each iteration. Based on restricted isometry property, a unified form of the error bounds of these recovery algorithms is derived under general perturbations for compressible signals. The results reveal that the recovery performance is stable against both perturbations. In addition, these bounds are compared with that of oracle recovery—least squares solution with the locations of some largest entries in magnitude known a priori. The comparison shows that the error bounds of these algorithms only differ in coefficients from the lower bound of oracle recovery for some certain signal and perturbations, as reveals that oracle-order recovery performance of greedy pursuits with replacement is guaranteed. Numerical simulations are performed to verify the conclusions.

**Keywords:** Compressive sensing, sparse recovery, general perturbations, performance analysis, restricted isometry property, greedy pursuits, compressive sampling matching pursuit, subspace pursuit, iterative hard thresholding, oracle recovery.

## 1 Introduction

Compressive sensing, or compressive sampling (CS) [1–3], is a novel signal processing technique proposed to effectively sample and compress sparse signals, i.e., signals that can be represented by few significant coefficients in some basis. Assume that the signal of interest  $\mathbf{x} \in \mathbb{R}^N$  can be represented by  $\mathbf{x} = \mathbf{\Psi}\mathbf{s}$ , where  $\mathbf{\Psi} \in \mathbb{R}^{N \times N}$  is the basis matrix and  $\mathbf{s}$  is

---

\*This work was supported partially by the National Natural Science Foundation of CHINA (60872087 and U0835003) and Agilent Technologies Foundation # 2205. The authors are with Department of Electronic Engineering, Tsinghua University, Beijing 100084, China. The corresponding author of this paper is Yuantao Gu (Email: gyt@tsinghua.edu.cn).

$K$ -sparse, which means only  $K$  out of its  $N$  entries are nonzero. One of the essential issues of CS theory lies in recovering  $\mathbf{x}$  (or equivalently,  $\mathbf{s}$ ) from its linear observations,

$$\mathbf{y} = \Phi\mathbf{x} = \Phi\Psi\mathbf{s}, \quad (1)$$

where  $\Phi \in \mathbb{R}^{M \times N}$  is a sensing matrix with more columns than rows and  $\mathbf{y}$  is the measurement vector. Unfortunately, directly finding the sparsest solution to (1) is NP-hard, which is not practical for sparse recovery. This leads to one of the major aspects of CS theory—designing effective recovery algorithms with low computational complexity and fine recovery performance.

## 1.1 Overview of Sparse Recovery Algorithms

A family of convex relaxation algorithms for sparse recovery [4] had been introduced before the theory of CS was established. Based on linear programming (LP) techniques, it is shown that  $\ell_1$  norm optimization, also known as basis pursuit (BP),

$$\min_{\mathbf{s}} \|\mathbf{s}\|_1 \quad \text{s.t.} \quad \mathbf{y} = \Phi\Psi\mathbf{s} \quad (2)$$

yields the sparse solution as long as  $\mathbf{A} = \Phi\Psi$  satisfies the restricted isometry property (RIP) with a constant parameter [5–7]. Recovery algorithms based on convex optimization include interior-point methods [8, 9] and homotopy methods [10, 11].

In contrast with convex relaxation algorithms, non-convex optimization algorithms solve (1) by minimizing  $\ell_p$  norm with respect to  $0 < p < 1$ , which is not convex. Typical algorithms include focal underdetermined system solver (FOCUSS) [12], iteratively reweighted least squares (IRLS) [13], smoothed  $\ell^0$  (SL0) [14], and zero-point attracting projection (ZAP) [15, 16]. Compared with convex relaxation algorithms, theoretical analysis based on RIP shows that fewer measurements are required for exact recovery by non-convex optimization methods [17].

A family of iterative greedy algorithms has received much attention due to their simple implementation and low computational complexity. The basic idea underlying these algorithms is to iteratively estimate the support set of the unknown sparse signal, i.e., the set of locations of its nonzero entries. In each iteration, one or more indices are added to the support estimation by correlating the columns of  $\mathbf{A}$  with the regularized measurement vector. Typical examples include orthogonal matching pursuit (OMP) [18], regularized OMP (ROMP) [19], and stage-wise OMP (StOMP) [20]. Compared with convex relaxation algorithms, greedy pursuits need more measurements, but they tend to be more computationally efficient.

Recently, several greedy algorithms including compressive sampling matching pursuit (CoSaMP) [21] and subspace pursuit (SP) [22] have been proposed by incorporating the idea of backtracking. In each iteration, SP algorithm refines the  $K$  columns of matrix  $\mathbf{A}$  that span the subspace where the measurement vector  $\mathbf{y}$  lies. Specifically, SP adds  $K$  more

indices to the  $K$  candidates of support estimate, and discards the most unreliable  $K$  ones. Similarly, CoSaMP adds  $2K$  more indices in each iteration, while computes the regularized measurement vector in a different way. By evaluating the reliability of all candidates in each iteration, these algorithms can provide comparable performance to convex relaxation algorithms, and exhibit low computational complexity as matching pursuit algorithms.

Another kind of greedy pursuits, including iterative hard thresholding (IHT) [23] and its normalized variation [24], is proposed with the advantages of low computational complexity and theoretical performance guarantee. In each iteration, the entries of the iterative solution except for the most  $K$  reliable ones are set to zero. Together with CoSaMP and SP, these algorithms can be considered as greedy pursuits with replacement involved in each iteration.

## 1.2 Overview of Perturbation Analysis

To apply the theory of CS in practice, the effect of noise and perturbation must be taken into consideration. The common analysis is the additive noise to the measurements, i.e.,

$$\tilde{\mathbf{y}} = \mathbf{y} + \mathbf{e}, \quad (3)$$

where  $\mathbf{e}$  is termed measurement noise. Most existing algorithms have been proved to be stable in this scenario, including theoretical analysis of BP [6], ROMP [25], CoSaMP [21], SP [22], and IHT [23, 24]. It is shown that the error bounds of the recovered solutions are proportional to the  $\ell_2$  norm of  $\mathbf{e}$ . A certain distribution of the measurement noise can be introduced to achieve better results, such as Gaussian [26–29] or others [30, 31].

Until recently, only a few researches involve the perturbation to the sensing matrix  $\Phi$ , which is also termed system perturbation. Existing works include analysis of BP [32], CoSaMP [33], SP [34],  $\ell_p$  norm minimization with  $p \in (0, 1)$  [35], and OMP [36]. In these works, (3) is extended by introducing a perturbed sensing matrix, i.e.,  $\tilde{\Phi} = \Phi + \Delta$ . It is of great significance to analyze the stability of recovery algorithms against both perturbations when the theory of CS is applied in practice. Other related works include mismatch of sparsity basis [37] and sparsity-cognizant total least-squares [38].

Two practical scenarios are usually considered in CS applications. First, when  $\Phi$  represents a system model,  $\Delta$  denotes the precision error when the system is physically implemented. Thus the whole sensing process is

$$\tilde{\mathbf{y}} = (\Phi + \Delta) \mathbf{x} + \mathbf{e}, \quad (4)$$

and only the nominal sensing matrix and contaminated measurement vector are available for recovery, i.e.,  $\hat{\mathbf{x}} = \mathbf{R}(\Phi, \tilde{\mathbf{y}})$  where  $\mathbf{R}(\cdot)$  denotes a certain recovery algorithm.

In countless other problems,  $\Delta$  is involved due to mis-modeling of the actual system  $\Phi$ . Thus  $\tilde{\Phi} = \Phi + \Delta$  and the sensing process is

$$\tilde{\mathbf{y}} = \tilde{\Phi} \mathbf{x} + \mathbf{e}. \quad (5)$$

Both the sensing matrix and measurement vector are contaminated, and the recovered solution is  $\hat{\mathbf{x}} = \mathbf{R}(\tilde{\Phi}, \tilde{\mathbf{y}})$ .

Existing works [32–34, 36] are all based on the latter scenario, thus it is well considered and fully analyzed in this paper. The first scenario (4) is briefly discussed after that as a remark.

### 1.3 Our Work

This paper mainly considers the recovery performance of greedy pursuits with replacement against general perturbations. Specifically, when both measurement noise and system perturbation exist, the error bounds of the solutions of CoSaMP, SP, and IHT are derived and discussed in detail for not strictly sparse signals, i.e., compressible signals. It is shown that these relative error bounds are linear in the relative perturbations of the measurement vector and the sensing matrix, which indicates that the recovery performance is stable against general perturbations. These error bounds are also compared with that of oracle recovery, and it reveals that the results in this paper are optimal up to the coefficients. Numerical simulations are performed to verify the conclusions in this paper. Previous related works including [29, 33, 34] are compared with this work in Section VI.

The remainder of this paper is organized as follows. Section II gives a brief review of RIP and three greedy pursuits with replacement. Section III presents the main theorems about the recovery performance of greedy pursuits with replacement against general perturbations, and the results are shown to be of the same order as oracle recovery by comparing them in Section IV. Numerical simulations are performed in Section V. Several related works are discussed in Section VI, and the paper is concluded in Section VII. Some detailed descriptions and discussions of greedy pursuits with replacement and proofs are postponed to Appendix.

## 2 Preliminary

In this section, the definition of RIP and descriptions of several greedy pursuits, including CoSaMP, SP, and IHT, as well as their recovery performance against measurement noise, are introduced.

### 2.1 Restricted Isometry Property

The restricted isometry property (RIP) for any matrix  $\mathbf{A}$  describes the degree of orthogonality among its different columns.

**Definition 1** [5] *For positive integer  $K$ , define the restricted isometry constant (RIC)  $\delta_K$  of a matrix  $\mathbf{A}$  as the smallest non-negative number such that*

$$(1 - \delta_K) \|\mathbf{s}\|_2^2 \leq \|\mathbf{A}\mathbf{s}\|_2^2 \leq (1 + \delta_K) \|\mathbf{s}\|_2^2 \quad (6)$$

holds for any  $K$ -sparse vector  $\mathbf{s}$ .

According to Definition 1,  $\delta_K < 1$  implies that every  $K$  columns of  $\mathbf{A}$  are linearly independent, and  $\delta_{2K} \ll 1$  implies that  $\mathbf{A}$  almost maintains the  $\ell_2$  distance between any pair of  $K$ -sparse signals. It is also easy to check that if  $\mathbf{A}$  satisfies the RIP with  $\delta_{K_1}$  and  $\delta_{K_2}$ , and  $K_1 \leq K_2$ , then  $\delta_{K_1} \leq \delta_{K_2}$ .

Calculating the exact value of RIC is intractable because it involves all submatrices comprised of  $K$  columns of  $\mathbf{A}$ . Fortunately, random matrices possess small RICs for overwhelming probability. For example, if the entries of  $\mathbf{A} \in \mathbb{R}^{M \times N}$  are independently and identically distributed Gaussian random variables with zero mean and variance  $1/M$ , and  $M$  satisfies

$$M \geq \frac{CK \log(N/K)}{\varepsilon^2},$$

then  $\delta_K \leq \varepsilon$  with at least probability  $1 - e^{-cM}$ , where  $C$  and  $c$  are two constants [5, 21]. Recent theoretical results about the bounds of RIC can be found in [39, 40].

Though RIP is usually used to show the impact of matrix on sparse signals, the following lemma permits us to generalize the result from sparse signals to general signals.

**Lemma 1** (Proposition 3.5 in [21]) *Suppose that the matrix  $\mathbf{A}$  satisfies RIP of level  $K$  with  $\delta_K$ , then for any signal  $\mathbf{s} \in \mathbb{R}^N$ , it holds that*

$$\|\mathbf{A}\mathbf{s}\|_2 \leq \sqrt{1 + \delta_K} \left( \|\mathbf{s}\|_2 + \frac{1}{\sqrt{K}} \|\mathbf{s}\|_1 \right). \quad (7)$$

Lemma 1 is used in the theoretical analysis of this paper for compressible signals.

## 2.2 Greedy Pursuits with Replacement

Greedy pursuits with replacement include CoSaMP, SP, and IHT algorithms, and they are briefly introduced as follows. The pseudo codes for these algorithms and some discussions are postponed to Appendix A.

The CoSaMP algorithm iteratively refines the support of  $K$ -sparse vector  $\mathbf{s}$ . In each iteration,  $2K$  more indices are selected by correlating  $\mathbf{A}$  with the residue vector  $\mathbf{r}$ , then the best  $K$  candidates out of at most  $3K$  ones are kept and the residue vector is updated. The details of CoSaMP can be found in [21].

The SP algorithm is firstly proposed in [41] and further developed in [22]. Unlike CoSaMP, SP adds  $K$  more indices in each iteration, and keeps the best  $K$  candidates. In addition, the residue vector of SP is orthogonal to the subspace spanned by the columns of  $\mathbf{A}$  indexed by the  $K$  candidates, making the new  $K$  indices added in each iteration totally different from the previously identified  $K$  ones. Please refer to [22] for more details.

The IHT algorithm is firstly proposed in [42], and later developed and analyzed in [23]. To improve the convergence performance of the method, a normalized variation NIHT is

Table 1: The Specification of the Constants

	CoSaMP	SP	IHT
$a$	1	1.26	1
$b$	4	3	3
$c$	0.171	0.206	0.353
$C$	$\frac{4\delta_{4K}}{(1-\delta_{4K})^2}$	$\frac{2\delta_{3K} + 2\delta_{3K}^2}{(1-\delta_{3K})^3}$	$\sqrt{8}\delta_{3K}$
$C_1$	$\frac{6 + 2\delta_{4K}}{(1-\delta_{4K})^2}$	$\frac{4(1+\delta_{3K})}{(1-\delta_{3K})^2}$	$2\sqrt{1+\delta_{3K}}$
$D$	$\frac{6 + 2\delta_{4K}}{1 - 6\delta_{4K} + \delta_{4K}^2}$	$\frac{5 - 4\delta_{3K} - 8\delta_{3K}^2 - \delta_{3K}^4}{1 - 6\delta_{3K} + 6\delta_{3K}^2 - 2\delta_{3K}^3 + \delta_{3K}^4}$	$\frac{2\sqrt{1+\delta_{3K}}}{1 - \sqrt{8}\delta_{3K}}$

proposed in [24] which retains theoretical performance guarantee similar to IHT. Without loss of generality, only IHT algorithm is discussed in the following analysis.

The following Theorem 1 reveals the error bounds of the solutions of greedy pursuits with replacement when only measurement noise exists. For two sets  $S_1$  and  $S_2$ ,  $S_1 - S_2$  denotes the set comprised of all elements  $x \in S_1$  and  $x \notin S_2$ .  $\mathbf{s}_S$  denotes the subvector composed of entries of  $\mathbf{s}$  indexed by set  $S$ .

**Theorem 1** *Given a noisy measurement vector  $\tilde{\mathbf{y}} = \mathbf{A}\mathbf{s} + \mathbf{e}$  where  $\mathbf{s}$  is  $K$ -sparse, the estimated solution  $\mathbf{s}^{[l]}$  in the  $l$ -th iteration of CoSaMP and IHT algorithms satisfies*

$$\|\mathbf{s} - \mathbf{s}^{[l]}\|_2 \leq C \|\mathbf{s} - \mathbf{s}^{[l-1]}\|_2 + C_1 \|\mathbf{e}\|_2, \quad (8)$$

and the estimated support set  $S^l$  in the  $l$ -th iteration of SP algorithm satisfies

$$\|\mathbf{s}_{S-S^l}\|_2 \leq C \|\mathbf{s}_{S-S^{l-1}}\|_2 + C_1 \|\mathbf{e}\|_2, \quad (9)$$

where  $S$  denotes the support of  $\mathbf{s}$ .

Furthermore, if the matrix  $\mathbf{A}$  satisfies  $\delta_{bK} \leq c$ , then  $C < 1$ , and it can be derived that

$$\|\mathbf{s} - \mathbf{s}^{[l]}\|_2 \leq aC^l \|\mathbf{s}\|_2 + D \|\mathbf{e}\|_2 \quad (10)$$

holds for greedy pursuits with replacement. The specific values of the constants  $a$ ,  $b$ ,  $c$ ,  $C$ ,  $C_1$ , and  $D$  are illustrated in TABLE 1.

PROOF Theorem 1 is concluded from the results of [21–23]. The proof is postponed to Appendix B. ■

Now we have introduced CoSaMP, SP, and IHT algorithms, where sparsity level  $K$  needs to be known a priori for replacement. According to (10), when only measurement noise exists, the error bounds of the solutions are exponential decay function of iteration number, and the limits of them are proportional to the  $\ell_2$  norm of the noise. Some remarks based on (10) are derived in Appendix A, and the main conclusions in this paper are based on (10) as well. It needs to be emphasized that the analysis is performed as worst-case, and the demands of RICs of  $\mathbf{A}$  are sufficient conditions. Better recovery performance of these three algorithms is normally achieved in practice, thus they are widely applied in various applications.

### 3 Recovery Performance of Greedy Pursuits with Replacement

In this section, the error bounds of greedy pursuits with replacement are derived for general signals when both measurement noise and system perturbation exist.

#### 3.1 Notations and Assumptions

Several notations and assumptions are stated first and they will be used in the following analysis.  $\mathbf{A}_S$  denotes the submatrix composed of the columns of  $\mathbf{A}$  indexed by set  $S$ .  $\mathbf{A}^T$  denotes the transpose of matrix  $\mathbf{A}$ .  $\mathbf{A}^\dagger$  denotes the pseudo-inverse of matrix  $\mathbf{A}$ . The signal  $\mathbf{s}$  is assumed to be a compressible signal in the following context, which means that it can be well approximated by a sparse signal. Vector  $\mathbf{s}_K$  is assumed to be the best  $K$ -term approximation to  $\mathbf{s}$ , and define the approximation error  $\mathbf{s}_K^c = \mathbf{s} - \mathbf{s}_K$ . The approximation error can be quantified as

$$r_K = \frac{\|\mathbf{s}_K^c\|_2}{\|\mathbf{s}\|_2}, \quad s_K = \frac{\|\mathbf{s}_K^c\|_1}{\sqrt{K} \|\mathbf{s}\|_2}. \quad (11)$$

When  $\mathbf{s}$  is  $K$ -sparse, the ratios  $r_K$  and  $s_K$  are both zero. If  $\mathbf{s}$  is compressible, then for reasonable  $K$ , the ratios are expected to be far less than 1.

The symbol  $\|\mathbf{A}\|_2$  denotes the spectral norm of a matrix  $\mathbf{A}$ . The measurement noise  $\mathbf{e}$  and system perturbation  $\mathbf{\Delta}$  can be quantified with the following relative perturbations

$$\varepsilon_{\mathbf{y}} = \frac{\|\mathbf{e}\|_2}{\|\mathbf{y}\|_2}, \quad \varepsilon_{\mathbf{\Phi}} = \frac{\|\mathbf{\Delta}\|_2}{\|\mathbf{\Phi}\|_2}, \quad (12)$$

where  $\|\mathbf{y}\|_2$  and  $\|\mathbf{\Phi}\|_2$  are nonzero.  $\varepsilon_{\mathbf{y}}$  and  $\varepsilon_{\mathbf{\Phi}}$  quantify the relative measurement noise and relative system perturbation, respectively. In practical scenarios, the exact forms of  $\mathbf{e}$  and  $\mathbf{\Delta}$  are not known in advance, thus the relative perturbations are applied instead. In this paper,  $\varepsilon_{\mathbf{y}}$  and  $\varepsilon_{\mathbf{\Phi}}$  are assumed less than 1. Define

$$\tilde{\mathbf{A}} = (\mathbf{\Phi} + \mathbf{\Delta})\mathbf{\Psi} = \mathbf{\Phi}\mathbf{\Psi} + \mathbf{\Delta}\mathbf{\Psi} \triangleq \mathbf{A} + \mathbf{E}. \quad (13)$$

Let  $\|\mathbf{A}\|_2^{(K)}$  denote the largest spectral norm taken over all  $K$ -column submatrices of  $\mathbf{A}$ . Assume

$$\varepsilon_{\mathbf{A}} = \frac{\|\mathbf{E}\|_2}{\|\mathbf{A}\|_2}, \quad \varepsilon_{\mathbf{A}}^{(K)} = \frac{\|\mathbf{E}\|_2^{(K)}}{\|\mathbf{A}\|_2^{(K)}}, \quad (14)$$

which also represent the relative system perturbation. It is easy to derive that

$$\varepsilon_{\mathbf{A}} \leq \kappa_{\Psi} \varepsilon_{\Phi}, \quad \varepsilon_{\mathbf{A}}^{(K)} \leq \frac{\|\Phi\|_2 \|\Psi\|_2}{\sqrt{1 - \delta_K}} \varepsilon_{\mathbf{A}},$$

where  $\kappa_{\Psi} = \|\Psi\|_2 \|\Psi^{-1}\|_2$  is the condition number of  $\Psi$ . Particularly,  $\kappa_{\Psi}$  is equal to one if  $\mathbf{x}$  is sparse in orthogonal basis. In general cases, the relative perturbations (14) are approximately the same as  $\varepsilon_{\Phi}$  if  $\kappa_{\Psi}$  is not very large. The following lemma quantifies the RICs of the matrix  $\tilde{\mathbf{A}}$ .

**Lemma 2** (Theorem 1 in [32]) *Assume the matrix  $\mathbf{A}$  satisfies RIP of level  $K$  with  $\delta_K$ , and the relative perturbation  $\varepsilon_{\mathbf{A}}^{(K)}$  is associated with matrix  $\mathbf{E}$ . Define the constant*

$$\tilde{\delta}_{K,\max} \triangleq (1 + \delta_K) \left(1 + \varepsilon_{\mathbf{A}}^{(K)}\right)^2 - 1, \quad (15)$$

then the RIC  $\tilde{\delta}_K$  of matrix  $\tilde{\mathbf{A}} = \mathbf{A} + \mathbf{E}$  satisfies  $\tilde{\delta}_K \leq \tilde{\delta}_{K,\max}$ .

As can be seen from Lemma 2, the upper bound of the RIC of perturbed matrix  $\tilde{\mathbf{A}}$  is slightly bigger than that of  $\mathbf{A}$  if the relative perturbation  $\varepsilon_{\mathbf{A}}^{(K)}$  is small. Notice that  $\tilde{\delta}_{K,\max}$  represents a worst-case of  $\tilde{\delta}_K$ , thus better RICs of  $\tilde{\mathbf{A}}$  are normally achieved in practice.

### 3.2 Error Bounds of Greedy Pursuits with Replacement

The relative error bounds of greedy pursuits with replacement under general perturbations for compressible signals are given in this subsection. The following Theorem 2 summarizes the main result.

**Theorem 2** *Suppose  $\mathbf{y} = \Phi \mathbf{x}$ , where  $\mathbf{x} = \Psi \mathbf{s}$  and  $\mathbf{s}$  is a compressible vector. The available information for recovering  $\mathbf{x}$  is  $\tilde{\mathbf{y}} = \mathbf{y} + \mathbf{e}$ ,  $\tilde{\Phi} = \Phi + \Delta$ , and basis matrix  $\Psi$ . If the available perturbed matrix  $\tilde{\mathbf{A}} = \tilde{\Phi} \Psi$  satisfies RIP with*

$$\tilde{\delta}_{bK} \leq c, \quad (16)$$

and the non-perturbed matrix  $\mathbf{A}$  satisfies RIP of level  $K$  with  $\delta_K$ , then in the  $l$ -th iteration, the relative error of the solution  $\mathbf{x}^{[l]} = \Psi \mathbf{s}^{[l]}$  of greedy pursuits with replacement satisfies

$$\frac{\|\mathbf{x} - \mathbf{x}^{[l]}\|_2}{\|\mathbf{x}\|_2} \leq \kappa_{\Psi} \left( a \tilde{C}^l + r_K + \tilde{D} \sqrt{1 + \delta_K} \left( \varepsilon_{\mathbf{y}} + \varepsilon_{\mathbf{A}}^{(K)} + (1 + \varepsilon_{\mathbf{y}}) (r_K + s_K) \right) \right), \quad (17)$$

where the specific values of the constants  $a$ ,  $b$ ,  $c$ ,  $\tilde{C}$ , and  $\tilde{D}$  are illustrated in TABLE 1 and TABLE 2.



Table 2: The Specification of the Constants

	CoSaMP	SP	IHT
$\tilde{C}$	$\frac{4\tilde{\delta}_{4K}}{(1 - \tilde{\delta}_{4K})^2}$	$\frac{2\tilde{\delta}_{3K} + 2\tilde{\delta}_{3K}^2}{(1 - \tilde{\delta}_{3K})^3}$	$\sqrt{8\tilde{\delta}_{3K}}$
$\tilde{D}$	$\frac{6 + 2\tilde{\delta}_{4K}}{1 - 6\tilde{\delta}_{4K} + \tilde{\delta}_{4K}^2}$	$\frac{5 - 4\tilde{\delta}_{3K} - 8\tilde{\delta}_{3K}^2 - \tilde{\delta}_{3K}^4}{1 - 6\tilde{\delta}_{3K} + 6\tilde{\delta}_{3K}^2 - 2\tilde{\delta}_{3K}^3 + \tilde{\delta}_{3K}^4}$	$\frac{2\sqrt{1 + \tilde{\delta}_{3K}}}{1 - \sqrt{8\tilde{\delta}_{3K}}}$

Furthermore, after at most

$$l = \left\lceil \log_{\tilde{C}} \left( \frac{\varepsilon_{\mathbf{y}} + \varepsilon_{\mathbf{A}}^{(K)} + s_K}{a} \right) \right\rceil \quad (18)$$

iterations, these algorithms estimate  $\mathbf{x}$  with accuracy

$$\frac{\|\mathbf{x} - \mathbf{x}^{[l]}\|_2}{\|\mathbf{x}\|_2} \leq \kappa_{\Psi} \left( \tilde{D} \sqrt{1 + \delta_K} + 1 \right) \left( \varepsilon_{\mathbf{y}} + \varepsilon_{\mathbf{A}}^{(K)} + (1 + \varepsilon_{\mathbf{y}}) (r_K + s_K) \right). \quad (19)$$

For exact sparse signals, a better and more intuitive result can be derived from Theorem 2 by setting  $r_K$  and  $s_K$  to zero. The result is stated as Theorem 3.

**Theorem 3** Suppose  $\mathbf{y} = \Phi \mathbf{x}$ , where  $\mathbf{x} = \Psi \mathbf{s}$  and  $\mathbf{s}$  is  $K$ -sparse. The available information for recovering  $\mathbf{x}$  is  $\tilde{\mathbf{y}} = \mathbf{y} + \mathbf{e}$ ,  $\tilde{\Phi} = \Phi + \Delta$ , and basis matrix  $\Psi$ . If the available perturbed matrix  $\tilde{\mathbf{A}} = \tilde{\Phi} \Psi$  satisfies RIP with

$$\tilde{\delta}_{bK} \leq c, \quad (20)$$

and the non-perturbed matrix  $\mathbf{A}$  satisfies RIP of level  $K$  with  $\delta_K$ , then in the  $l$ -th iteration, the relative error of the solution  $\mathbf{x}^{[l]} = \Psi \mathbf{s}^{[l]}$  of greedy pursuits with replacement satisfies

$$\frac{\|\mathbf{x} - \mathbf{x}^{[l]}\|_2}{\|\mathbf{x}\|_2} \leq \kappa_{\Psi} \left( a \tilde{C}^l + \tilde{D} \sqrt{1 + \delta_K} \left( \varepsilon_{\mathbf{y}} + \varepsilon_{\mathbf{A}}^{(K)} \right) \right), \quad (21)$$

where the specific values of the constants  $a$ ,  $b$ ,  $c$ ,  $\tilde{C}$ , and  $\tilde{D}$  are illustrated in TABLE 1 and TABLE 2.

Furthermore, after at most

$$l = \left\lceil \log_{\tilde{C}} \left( \frac{\varepsilon_{\mathbf{y}} + \varepsilon_{\mathbf{A}}^{(K)}}{a} \right) \right\rceil \quad (22)$$

iterations, these algorithms estimate  $\mathbf{x}$  with accuracy

$$\frac{\|\mathbf{x} - \mathbf{x}^{[l]}\|_2}{\|\mathbf{x}\|_2} \leq \kappa_{\Psi} \left( \tilde{D} \sqrt{1 + \delta_K} + 1 \right) \left( \varepsilon_{\mathbf{y}} + \varepsilon_{\mathbf{A}}^{(K)} \right). \quad (23)$$

### 3.3 Proof of Theorem 2

PROOF Recalling the sensing process (5) with general perturbations, it is equivalent to

$$\begin{aligned}\tilde{\mathbf{y}} &= \mathbf{A}\mathbf{s} + \mathbf{e} = \left(\tilde{\mathbf{A}} - \mathbf{E}\right)\mathbf{s}_K + \mathbf{A}\mathbf{s}_K^c + \mathbf{e} \\ &= \tilde{\mathbf{A}}\mathbf{s}_K + (\mathbf{e} - \mathbf{E}\mathbf{s}_K + \mathbf{A}\mathbf{s}_K^c).\end{aligned}\quad (24)$$

Define  $\tilde{\mathbf{e}} = \mathbf{e} - \mathbf{E}\mathbf{s}_K + \mathbf{A}\mathbf{s}_K^c$  as the error term. According to Theorem 1, under condition (16) about the RICs of  $\tilde{\mathbf{A}}$ , the solution  $\mathbf{s}^{[l]}$  in the  $l$ -th iteration satisfies

$$\|\mathbf{s}_K - \mathbf{s}^{[l]}\|_2 \leq a\tilde{C}^l \|\mathbf{s}_K\|_2 + \tilde{D} \|\tilde{\mathbf{e}}\|_2, \quad (25)$$

where  $\tilde{C} < 1$  and  $\tilde{D}$  are constants specified in TABLE 2. From the triangle inequality, it can be derived that

$$\|\mathbf{s} - \mathbf{s}^{[l]}\|_2 \leq \|\mathbf{s} - \mathbf{s}_K\|_2 + \|\mathbf{s}_K - \mathbf{s}^{[l]}\|_2. \quad (26)$$

Substituting (25) into (26), and from the fact that

$$\|\mathbf{s}_K\|_2 \leq \|\mathbf{s}\|_2, \quad (27)$$

the error bound in the  $l$ -th iteration satisfies

$$\|\mathbf{s} - \mathbf{s}^{[l]}\|_2 \leq a\tilde{C}^l \|\mathbf{s}\|_2 + \|\mathbf{s}_K^c\|_2 + \tilde{D} \|\tilde{\mathbf{e}}\|_2. \quad (28)$$

To estimate  $\|\tilde{\mathbf{e}}\|_2$ , the triangle inequality, Lemma 1, and definitions (11) imply

$$\begin{aligned}\|\tilde{\mathbf{e}}\|_2 &\leq \|\mathbf{e}\|_2 + \|\mathbf{E}\mathbf{s}_K\|_2 + \|\mathbf{A}\mathbf{s}_K^c\|_2 \\ &\leq \|\mathbf{e}\|_2 + \|\mathbf{E}\|_2^{(K)} \|\mathbf{s}_K\|_2 + \|\mathbf{A}\|_2^{(K)} \|\mathbf{s}\|_2 (r_K + s_K).\end{aligned}\quad (29)$$

It can be derived that

$$\begin{aligned}\|\mathbf{e}\|_2 &= \varepsilon_{\mathbf{y}} \|\mathbf{A}\mathbf{s}\|_2 \leq \varepsilon_{\mathbf{y}} (\|\mathbf{A}\mathbf{s}_K\|_2 + \|\mathbf{A}\mathbf{s}_K^c\|_2) \\ &\leq \varepsilon_{\mathbf{y}} \|\mathbf{A}\|_2^{(K)} \|\mathbf{s}\|_2 (1 + r_K + s_K),\end{aligned}\quad (30)$$

which indicates

$$\|\tilde{\mathbf{e}}\|_2 \leq \|\mathbf{A}\|_2^{(K)} \|\mathbf{s}\|_2 \left( \varepsilon_{\mathbf{y}} + \varepsilon_{\mathbf{A}}^{(K)} + (1 + \varepsilon_{\mathbf{y}}) (r_K + s_K) \right). \quad (31)$$

Since

$$\|\mathbf{x} - \mathbf{x}^{[l]}\|_2 \leq \|\Psi\|_2 \|\mathbf{s} - \mathbf{s}^{[l]}\|_2, \quad (32)$$

$$\|\mathbf{x}\|_2 \geq \|\Psi^{-1}\|_2^{-1} \|\mathbf{s}\|_2, \quad (33)$$

together with (28) and (31), inequality (17) can be obtained.

For the second part of the theorem, noticing that when (18) holds, it is obvious that

$$\begin{aligned}a\tilde{C}^l + r_K &\leq \varepsilon_{\mathbf{y}} + \varepsilon_{\mathbf{A}}^{(K)} + r_K + s_K \\ &\leq \varepsilon_{\mathbf{y}} + \varepsilon_{\mathbf{A}}^{(K)} + (1 + \varepsilon_{\mathbf{y}}) (r_K + s_K)\end{aligned}\quad (34)$$

and (19) follows immediately, which completes the proof of Theorem 2.  $\blacksquare$

### 3.4 Discussion

In this subsection, several remarks are drawn to have a deep insight into the results of this paper.

**Remark 1** As can be seen from (19), the relative error of greedy pursuits with replacement is linear in both kinds of perturbations, and almost linear in the approximation error of  $\mathbf{s}$  to a  $K$ -sparse vector. If  $\mathbf{s}$  is  $K$ -sparse and the sensing process is non-perturbed, according to (21) with  $\varepsilon_{\mathbf{y}} = \varepsilon_{\mathbf{A}}^{(K)} = 0$ , the relative error is bounded by an exponential decay function of iteration number, which indicates that the recovered solution can approach the original signal at any given precision. If  $\mathbf{s}$  is  $K$ -sparse and  $\mathbf{x}$  is sensed under general perturbations, (23) demonstrates that the recovery accuracy is bounded by the size of both perturbations, and the recovery performance is stable in this scenario. If the sensing process is non-perturbed for general signal  $\mathbf{x}$ , according to (19) with  $\varepsilon_{\mathbf{y}} = \varepsilon_{\mathbf{A}}^{(K)} = 0$ , the recovery accuracy is determined by how well  $\mathbf{s}$  can be approximated by a  $K$ -sparse vector.

**Remark 2** Under the assumption (16), the base  $\tilde{C}$  of the exponential function in (17) is less than one, which guarantees the convergence of the recovery error. As  $\tilde{\delta}_{bK} \rightarrow c$ , the base  $\tilde{C}$  approaches one and the constant  $\tilde{D}$  approaches infinity. Also, due to the monotonicity of RICs, for less sparse signals with larger  $K$ , the constant  $\tilde{D}$  gets larger.

For these three greedy pursuits with replacement,  $K$  is quite a significant parameter and needs to be known a priori. An improper selection of  $K$  may result in a great loss of recovery performance. If  $K$  is selected too large, the condition (16) can be hardly satisfied, and the parameter  $\tilde{D}$  may be quite large. On the other hand, too small  $K$  indicates large  $r_K$  and  $s_K$ , which also leads to poor recovery performance.

As a typical compressible signal, strong-decaying signal  $\mathbf{s}$  is a vector whose ordered coefficients in certain basis satisfy

$$|\mathbf{s}|_{(l)} \geq p |\mathbf{s}|_{(l+1)}, \quad 1 \leq l \leq N - 1, \quad (35)$$

where  $|\mathbf{s}|_{(l)}$  is the  $l$ -th largest magnitude among the elements of  $\mathbf{s}$ , and  $p > 1$  is a constant controls the speed of the decay: the larger  $p$  is, the faster  $\mathbf{s}$  decays. Simple calculation implies that

$$r_K \leq p^{-K}, \quad s_K \leq \frac{\sqrt{p+1}}{\sqrt{K(p-1)}} p^{-K}. \quad (36)$$

For compressible signal satisfying a power law, i.e.,

$$|\mathbf{s}|_{(l)} \approx R \cdot l^{-p}, \quad 1 \leq l \leq N, \quad (37)$$

where  $R$  is a positive constant denoting the radius of weak- $\ell_{1/p}$  ball and  $p > 1$  controls the speed of the decay, it can also be calculated that

$$r_K \approx K^{1/2-p}, \quad s_K \approx \frac{\sqrt{2p-1}}{p-1} K^{1/2-p}. \quad (38)$$

In practice, the parameter  $K$  should be adjusted according to  $\varepsilon_{\mathbf{y}}$ ,  $\varepsilon_{\mathbf{A}}^{(K)}$ , and the decay parameter  $p$ . In the scenario with relatively large perturbations, according to (19),  $K$  can be selected a small value such that  $\tilde{\delta}_{bK}$  and  $\tilde{D}$  are small, and  $r_K$  and  $s_K$  are comparable to  $\varepsilon_{\mathbf{y}}$  and  $\varepsilon_{\mathbf{A}}^{(K)}$ . In the scenario with relatively small perturbations, the recovery error is mainly determined by  $r_K$  and  $s_K$ , and  $K$  is a trade-off parameter between these ratios and  $\tilde{D}$ .

**Remark 3** Considering the sensing process (4) with  $\tilde{\Phi} = \Phi + \Delta$  and  $\tilde{\mathbf{A}} = \tilde{\Phi}\Psi = \mathbf{A} + \mathbf{E}$ ,

$$\tilde{\mathbf{y}} = (\mathbf{A} + \mathbf{E})\mathbf{s} + \mathbf{e} = \mathbf{A}\mathbf{s}_K + \left(\mathbf{e} + \mathbf{E}\mathbf{s}_K + \tilde{\mathbf{A}}\mathbf{s}_K^c\right). \quad (39)$$

Define  $\tilde{\mathbf{e}} = \mathbf{e} + \mathbf{E}\mathbf{s}_K + \tilde{\mathbf{A}}\mathbf{s}_K^c$  as the error term. Following the steps of proof of Theorem 2, it can be derived that if the matrix  $\mathbf{A}$  satisfies RIP with  $\delta_{bK} \leq c$ , then

$$\|\mathbf{s} - \mathbf{s}^{[l]}\|_2 \leq aC^l \|\mathbf{s}\|_2 + \|\mathbf{s}_K^c\|_2 + D \|\tilde{\mathbf{e}}\|_2. \quad (40)$$

Since

$$\|\tilde{\mathbf{e}}\|_2 \leq \|\mathbf{e}\|_2 + \|\mathbf{E}\|_2^{(K)} \|\mathbf{s}\|_2 + \|\tilde{\mathbf{A}}\|_2^{(K)} \|\mathbf{s}\|_2 (r_K + s_K)$$

and the fact that

$$\|\tilde{\mathbf{A}}\|_2^{(K)} \leq \|\mathbf{A}\|_2^{(K)} + \|\mathbf{E}\|_2^{(K)} \leq \left(1 + \varepsilon_{\mathbf{A}}^{(K)}\right) \|\mathbf{A}\|_2^{(K)}, \quad (41)$$

it can be derived that the relative error of the solution in the  $l$ -th iteration obeys

$$\frac{\|\mathbf{x} - \mathbf{x}^{[l]}\|_2}{\|\mathbf{x}\|_2} \leq \kappa_{\Psi} \left( aC^l + r_K + D\sqrt{1 + \delta_K} \left( \varepsilon_{\mathbf{y}} + \varepsilon_{\mathbf{A}}^{(K)} + (1 + \varepsilon_{\mathbf{y}} + \varepsilon_{\mathbf{A}}^{(K)}) (r_K + s_K) \right) \right). \quad (42)$$

Notice that in this scenario, the relative error bound of greedy pursuits with replacement is also linear in the relative perturbations of both measurement vector and sensing matrix, which indicates that the recovery performance is stable against general perturbations. Also, it should be stressed that for both scenarios, the demands of RIP are all for the available matrix  $\tilde{\mathbf{A}}$  or  $\mathbf{A}$  in the process of recovery.

## 4 Comparison with Oracle Recovery

In this section, the upper and lower bounds of oracle recovery are derived, and are compared with those of greedy pursuits with replacement. It reveals that the results in this paper are optimal up to the coefficients.

### 4.1 Error Bound of Oracle Recovery

Consider the oracle recovery where the locations of the  $K$  largest entries in magnitude of the vector  $\mathbf{s}$  are known a priori. Assume  $S$  is the set of the  $K$  locations, i.e.,  $S = \text{supp}(\mathbf{s}_K)$ .

Recall the sensing process (5) where both perturbations exist. Through least squares (LS) method, the estimated solution  $\hat{\mathbf{x}} = \mathbf{\Psi}\hat{\mathbf{s}}$  of oracle recovery is obtained by

$$\hat{\mathbf{s}}_S = \tilde{\mathbf{A}}_S^\dagger \tilde{\mathbf{y}}, \quad \hat{\mathbf{s}}_{S^c} = \mathbf{0}. \quad (43)$$

It is easy to check that

$$\hat{\mathbf{s}}_S = \tilde{\mathbf{A}}_S^\dagger \tilde{\mathbf{y}} = \tilde{\mathbf{A}}_S^\dagger (\tilde{\mathbf{A}}_S \mathbf{s}_K + \tilde{\mathbf{e}}) = \mathbf{s}_S + \tilde{\mathbf{A}}_S^\dagger \tilde{\mathbf{e}}. \quad (44)$$

Thus the estimation error of  $\hat{\mathbf{s}}$  obeys

$$\begin{aligned} \|\mathbf{s} - \hat{\mathbf{s}}\|_2^2 &= \|\mathbf{s}_S - \hat{\mathbf{s}}_S\|_2^2 + \|\mathbf{s}_{S^c} - \hat{\mathbf{s}}_{S^c}\|_2^2 \\ &= \|\tilde{\mathbf{A}}_S^\dagger \tilde{\mathbf{e}}\|_2^2 + \|\mathbf{s}_K^c\|_2^2. \end{aligned} \quad (45)$$

According to (45), the estimation error obeys

$$\begin{aligned} \|\mathbf{s} - \hat{\mathbf{s}}\|_2 &\leq \|\tilde{\mathbf{A}}_S^\dagger \tilde{\mathbf{e}}\|_2 + \|\mathbf{s}_K^c\|_2 \\ &\leq \|\tilde{\mathbf{A}}_S^\dagger\|_2 \|\tilde{\mathbf{e}}\|_2 + \|\mathbf{s}_K^c\|_2. \end{aligned} \quad (46)$$

Substituting (31) into (46), and together with (32), (33), and the fact that

$$\|\tilde{\mathbf{A}}_S^\dagger\|_2 \leq \frac{1}{\sqrt{1 - \tilde{\delta}_K}} \triangleq \hat{D}, \quad (47)$$

the relative error of LS solution is derived following the steps in the proof of Theorem 2

$$\frac{\|\mathbf{x} - \hat{\mathbf{x}}\|_2}{\|\mathbf{x}\|_2} \leq \kappa_{\mathbf{\Psi}} \left( \hat{D} \sqrt{1 + \delta_K} + 1 \right) \left( \varepsilon_{\mathbf{y}} + \varepsilon_{\mathbf{A}}^{(K)} + (1 + \varepsilon_{\mathbf{y}}) (r_K + s_K) \right). \quad (48)$$

Comparing (48) with (19), it can be derived that after finite iterations, the error bounds of greedy pursuits with replacement and the error bound of oracle recovery only differ in coefficients, and they are all of the same order on the noise level and approximation error

$$O \left( \varepsilon_{\mathbf{y}} + \varepsilon_{\mathbf{A}}^{(K)} + (1 + \varepsilon_{\mathbf{y}}) (r_K + s_K) \right) \quad (49)$$

under general perturbations.

## 4.2 Lower Bound Analysis of Oracle Recovery

According to (28), (29), and (32), after

$$l = \left\lceil \log_{\tilde{C}} \left( \frac{\|\mathbf{e}\|_2 + \|\mathbf{E}\|_2^{(K)} \|\mathbf{s}_K\|_2}{a \|\mathbf{A}\|_2^{(K)} \|\mathbf{s}\|_2} \right) \right\rceil$$

iterations, the recovery error of greedy pursuits with replacement satisfies

$$\|\mathbf{x} - \mathbf{x}^{[l]}\|_2 \leq \|\mathbf{\Psi}\|_2 \left( \tilde{D} + \frac{1}{\sqrt{1 - \delta_K}} \right) \left( \|\mathbf{e}\|_2 + \|\mathbf{E}\|_2^{(K)} \|\mathbf{s}_K\|_2 + \|\mathbf{A}\|_2^{(K)} \|\mathbf{s}\|_2 (r_K + s_K) \right). \quad (50)$$

To show the optimality of the results in this paper, we prove that for some certain  $\mathbf{s}$ ,  $\mathbf{e}$ , and  $\mathbf{E}$ , the lower bound of the estimation error of oracle recovery is also linear in the following three terms

$$\|\mathbf{e}\|_2, \quad \|\mathbf{E}\|_2^{(K)} \|\mathbf{s}_K\|_2, \quad \|\mathbf{s}\|_2 (r_K + s_K), \quad (51)$$

which quantify measurement noise, system perturbation, and approximation error, respectively.

According to (45), the estimation error of  $\hat{\mathbf{s}}$  obeys

$$\|\mathbf{s} - \hat{\mathbf{s}}\|_2 \geq \frac{1}{\sqrt{2}} \left( \|\tilde{\mathbf{A}}_S^\dagger \tilde{\mathbf{e}}\|_2 + \|\mathbf{s}_K^c\|_2 \right). \quad (52)$$

The system perturbation  $\mathbf{E}$  can be chosen as  $\mathbf{E} = \varepsilon_{\mathbf{A}}^{(K)} \mathbf{A}$ , and the vector  $\mathbf{s}_K$  is selected such that

$$\|\mathbf{E}\mathbf{s}_K\|_2 = \|\mathbf{E}\|_2^{(K)} \|\mathbf{s}_K\|_2. \quad (53)$$

Meanwhile, the vector  $\mathbf{s}_K^c$  is also assumed  $K$ -sparse, and it can be assumed that

$$(-\mathbf{E}\mathbf{s}_K)^\top \mathbf{A}_S \mathbf{A}_S^\top (\mathbf{A}\mathbf{s}_K^c) \geq 0, \quad (54)$$

otherwise  $(-\mathbf{s}_K^c)$  is applied. Furthermore, the measurement noise  $\mathbf{e}$  is chosen in the column space of  $\mathbf{A}_S$ , and satisfies

$$\mathbf{e}^\top \mathbf{A}_S \mathbf{A}_S^\top (-\mathbf{E}\mathbf{s}_K + \mathbf{A}\mathbf{s}_K^c) \geq 0. \quad (55)$$

Due to the above assumptions and the fact that

$$\tilde{\mathbf{A}} = \mathbf{A} + \mathbf{E} = \left(1 + \varepsilon_{\mathbf{A}}^{(K)}\right) \mathbf{A},$$

it can be derived that

$$\|\tilde{\mathbf{A}}_S^\dagger \tilde{\mathbf{e}}\|_2 = \left\| (\tilde{\mathbf{A}}_S^\top \tilde{\mathbf{A}}_S)^{-1} \tilde{\mathbf{A}}_S^\top \tilde{\mathbf{e}} \right\|_2 \geq \frac{1}{1 + \delta_K} \|\mathbf{A}_S^\top \tilde{\mathbf{e}}\|_2. \quad (56)$$

According to the definition of  $\tilde{\mathbf{e}}$ , (54), and (55), it can be calculated that

$$\begin{aligned} \|\mathbf{A}_S^\top \tilde{\mathbf{e}}\|_2 &\geq \frac{1}{\sqrt{2}} \left( \|\mathbf{A}_S^\top \mathbf{e}\|_2 + \|\mathbf{A}_S^\top (-\mathbf{E}\mathbf{s}_K + \mathbf{A}\mathbf{s}_K^c)\|_2 \right) \\ &\geq \frac{1}{\sqrt{2}} \left( \|\mathbf{A}_S^\top \mathbf{e}\|_2 + \|\mathbf{A}_S^\top \mathbf{E}\mathbf{s}_K\|_2 \right). \end{aligned} \quad (57)$$

Since  $\mathbf{e}$  and  $\mathbf{E}\mathbf{s}_K$  both belong to the column space of  $\mathbf{A}_S$ , (57) and (53) further imply

$$\begin{aligned} \|\mathbf{A}_S^\top \tilde{\mathbf{e}}\|_2 &\geq \frac{\sqrt{1 - \delta_K}}{\sqrt{2}} (\|\mathbf{e}\|_2 + \|\mathbf{E}\mathbf{s}_K\|_2) \\ &= \frac{\sqrt{1 - \delta_K}}{\sqrt{2}} \left( \|\mathbf{e}\|_2 + \|\mathbf{E}\|_2^{(K)} \|\mathbf{s}_K\|_2 \right). \end{aligned} \quad (58)$$

In addition, Cauchy-Schwartz inequality implies

$$\|\mathbf{s}_K^c\|_2 \geq \frac{1}{2} \left( \|\mathbf{s}_K^c\|_2 + \frac{\|\mathbf{s}_K^c\|_1}{\sqrt{K}} \right) = \frac{1}{2} \|\mathbf{s}\|_2 (r_K + s_K). \quad (59)$$

According to (52), (56), (58), (59), and the fact that

$$\|\mathbf{x} - \hat{\mathbf{x}}\|_2 \geq \|\Psi^{-1}\|_2^{-1} \|\mathbf{s} - \hat{\mathbf{s}}\|_2,$$

the lower bound of the estimation error of oracle recovery satisfies

$$\|\mathbf{x} - \hat{\mathbf{x}}\|_2 \geq \frac{\|\Psi^{-1}\|_2^{-1}}{2} \left( \frac{\sqrt{1 - \delta_K}}{1 + \tilde{\delta}_K} \left( \|\mathbf{e}\|_2 + \|\mathbf{E}\|_2^{(K)} \|\mathbf{s}_K\|_2 \right) + \frac{1}{\sqrt{2}} \|\mathbf{s}\|_2 (r_K + s_K) \right), \quad (60)$$

which is also linear in the three terms (51).

According to the above analysis, it can be concluded that there exist  $\mathbf{s}$ ,  $\mathbf{e}$ , and  $\mathbf{E}$  with which the results in this paper are essentially tight up to the coefficients, i.e., the performances of these algorithms and oracle recovery are of the same order. In general, no recovery algorithms can do better than the oracle least squares method. Thus, the greedy pursuits with replacement, CoSaMP, SP, and IHT, can provide oracle-order recovery performance against general perturbations.

## 5 Numerical Simulations

Three numerical simulations are performed in MATLAB and demonstrated as follows. The first simulation compares the relative error of greedy pursuits with replacement and oracle LS method for different sparsity levels. The second simulation compares the relative error with different relative perturbations. The final simulation considers compressible signals and presents the influence of parameter  $K$  on the relative recovery error.

In each trial the matrix  $\Phi$  of size  $512 \times 2048$  is randomly generated with independent Gaussian distributed entries with variance  $\sigma^2 = 1/512$  so that the expectation of the  $\ell_2$  norm of each column is normalized. The basis matrix  $\Psi$  is the identity matrix, i.e.,  $\mathbf{x} = \mathbf{s}$  and  $\mathbf{A} = \Phi$ . The nonzero locations of vector  $\mathbf{s}$  are randomly selected among all possible choices, and the nonzero entries are generated independently from normal distribution. Then, for each pair of relative perturbations,  $\tilde{\mathbf{y}}$  and  $\tilde{\Phi}$  are generated according to (5) and (12). Notice that only  $\varepsilon_{\mathbf{A}}$  is used in the simulations since calculating  $\varepsilon_{\mathbf{A}}^{(K)}$  is NP-hard. However, it is reasonable because  $\varepsilon_{\mathbf{A}} \approx \varepsilon_{\mathbf{A}}^{(K)}$  holds for all  $K$  with high probability when both  $\mathbf{A}$  and  $\mathbf{E}$  are random Gaussian matrices [32].

In the first experiment, the relative perturbations are set to  $\varepsilon_{\mathbf{A}} = 0.05$  and  $\varepsilon_{\mathbf{y}} = 0.05$ . The sparsity level  $K$  varies from 1 to 20, and the simulation is repeated for 1000 times. As can be seen from Fig. 1, when  $K < 20$ , the relative error increases almost linearly as  $K$  increases. When the sparsity level  $K$  continues to increase, the relative error of IHT cannot be bounded any more, which indicates that the RIP condition is violated.

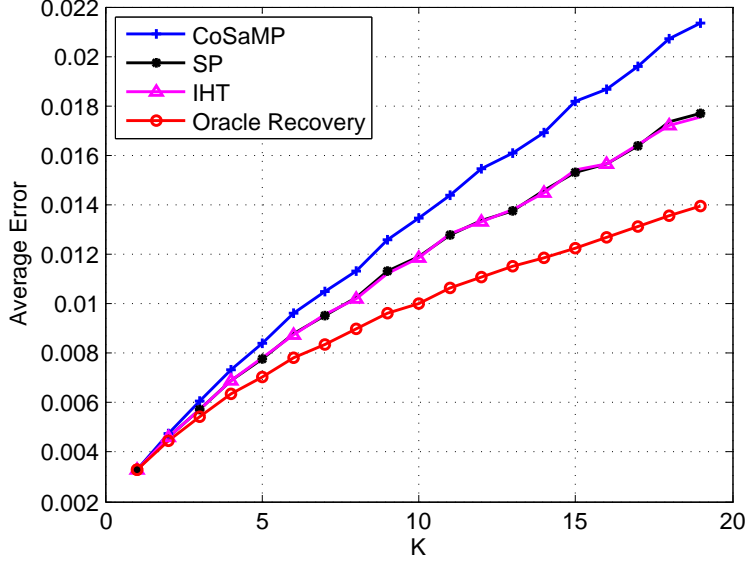


Figure 1: Average relative error (1000 trials) of greedy pursuits with replacement and oracle LS recovery versus sparsity level  $K$  when  $\varepsilon_{\mathbf{A}} = 0.05$  and  $\varepsilon_{\mathbf{y}} = 0.05$ .

In the second experiment, for fixed sparsity level  $K = 5, 10,$  and  $15$ , the relative perturbations  $\varepsilon_{\mathbf{A}}$  and  $\varepsilon_{\mathbf{y}}$  both vary from 0 to 0.1 with step size 0.005. The simulation is conducted 1000 trials to obtain the average relative error. The results are demonstrated in Fig. 2, where they are represented for the same  $K$  in the same row, and for the same recovery algorithm in the same column, with labels in the left and on the top, respectively. As can be seen from them, for the same  $K$ , the surface of oracle LS method lies below the others, while the surface of CoSaMP is above on the top. The surfaces of SP and IHT are almost the same in the middle. It verifies that the relative errors of CoSaMP, SP, and IHT are linear in the measurement perturbation as well as system perturbation, and these greedy pursuits with replacement can provide oracle-order recovery performance. It needs to be pointed out that for the same recovery algorithm, the gradients of these surfaces increase as the increase of  $K$ , which confirms that the coefficients of these error bounds are related to  $K$  as discussed in Section III and IV. To display the results more intuitively, the average relative error is plotted versus  $\varepsilon_{\mathbf{A}}$  for different  $K$  when  $\varepsilon_{\mathbf{y}} = 0$ , as in Fig. 3. As can be seen, the relative error scales almost linearly as  $\varepsilon_{\mathbf{A}}$ , and the slopes of these curves increase as  $K$  for the same recovery algorithm.

In the final experiment, we consider compressible signal satisfying power law (37) with  $R = 1$  and  $p = 2$ . The relative perturbations are set to  $\varepsilon_{\mathbf{A}} = 0.01$  and  $\varepsilon_{\mathbf{y}} = 0.01$ . The parameter  $K$  in the recovery algorithms varies from 5 to 100, and the results are plotted in Fig. 4. Since IHT fails when  $K \geq 20$ , the recovery error of it is not shown for  $K \geq 20$  in the figure. As can be seen from Fig. 4, the optimal  $K$  is around 22 for CoSaMP and SP, 19 for IHT, and around 40 for oracle recovery. Too small  $K$  or large  $K$  will affect the



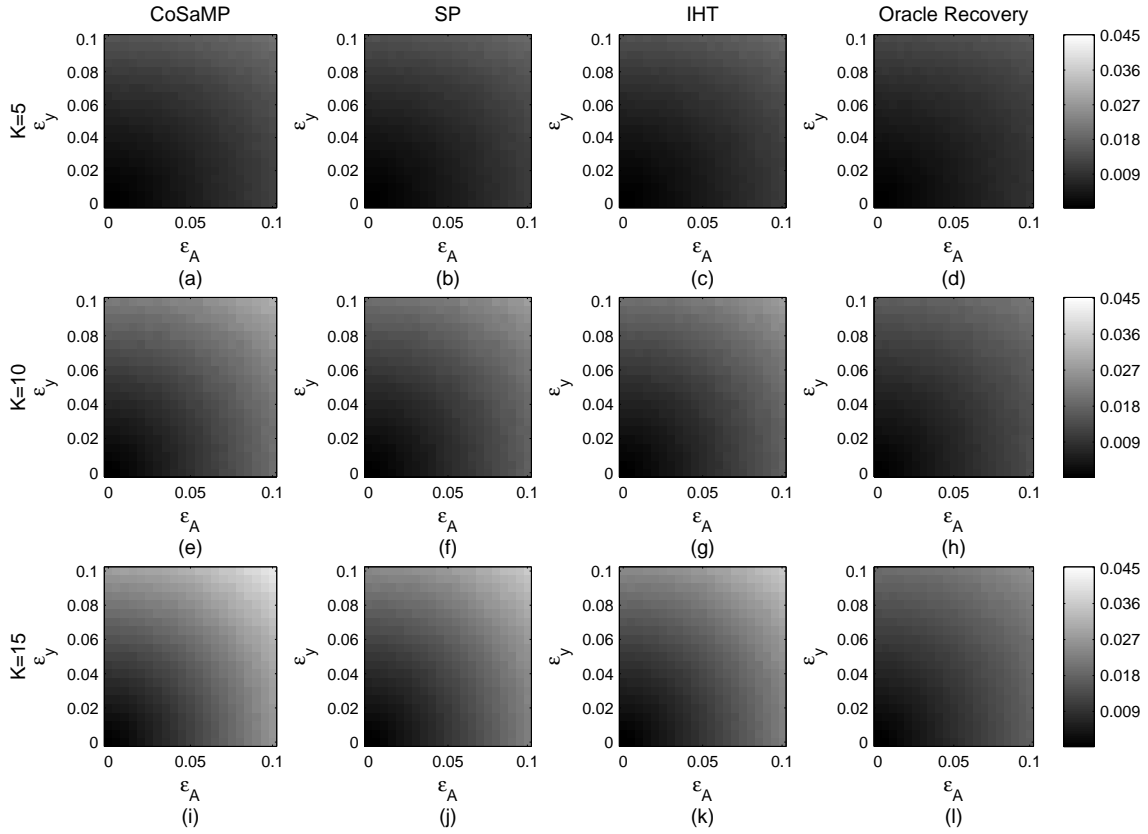


Figure 2: Average relative error (1000 trials) of greedy pursuits with replacement and oracle LS recovery versus different  $\epsilon_A$  and  $\epsilon_y$ . The results are represented for the same  $K$  in the same row, and for the same recovery algorithm in the same column, with labels in the left and on the top, respectively.

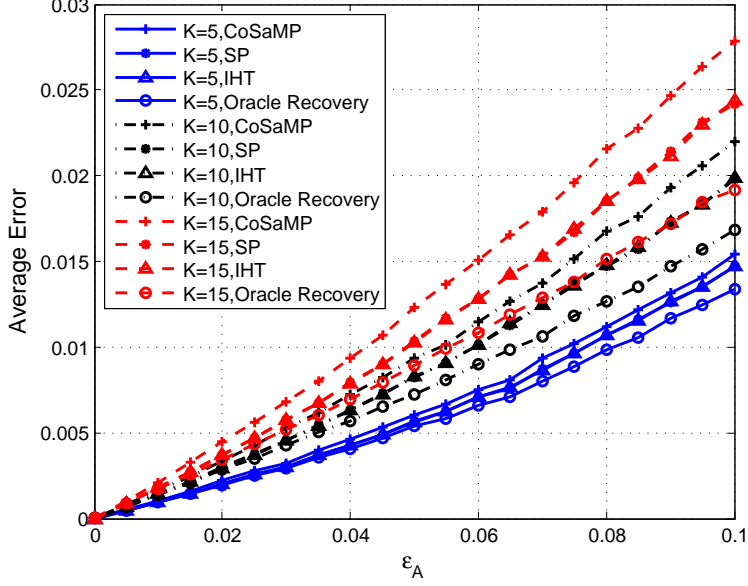


Figure 3: Average relative error (1000 trials) of greedy pursuits with replacement and oracle LS recovery versus  $\varepsilon_{\mathbf{A}}$  for different sparsity level  $K$  when  $\varepsilon_{\mathbf{y}} = 0$ .

recovery performance for compressible signals, as is discussed in Remark 2. According to (38), for  $K = 22$  and  $p = 2$ , the approximation error  $r_K$  is about 0.010 and  $s_K$  is about 0.017, which are comparable to  $\varepsilon_{\mathbf{A}}$  and  $\varepsilon_{\mathbf{y}}$ . This further confirms the analysis in Remark 2 that the parameter  $K$  should be adjusted in accordance with  $\varepsilon_{\mathbf{A}}^{(K)}$  and  $\varepsilon_{\mathbf{y}}$ .

## 6 Related Works

A similar concept of this paper is presented in [29], where the authors also establish an oracle-order performance guarantee for CoSaMP, SP, and IHT algorithms. Based on RIP, the analysis considers the recovery of a  $K$ -sparse signal with the assumption that the measurement vector is corrupted by additive random white Gaussian noise. The main result of [29] is stated as follows, where some notations are replaced for the sake of consistency to our work.

Assume that the white Gaussian noise vector is with covariance matrix  $\sigma^2 \mathbf{I}$ , and that the columns of matrix  $\mathbf{A}$  are normalized. Under certain conditions of RICs and with probability exceeding  $1 - (\sqrt{\pi(1+a)} \log N \cdot N^a)^{-1}$ , it holds that

$$\|\mathbf{x} - \hat{\mathbf{x}}\|_2^2 \leq 2C^2(1+a) \log N \cdot K\sigma^2, \quad (61)$$

where  $\hat{\mathbf{x}}$  is the recovered signal, and  $C$  is a constant related to specific recovery algorithms, sensing matrix, and sparsity level. The result is similar to those for the dantzig-selector and the basis pursuit, but with different constants. The  $\log N$  factor in (61) is proven to

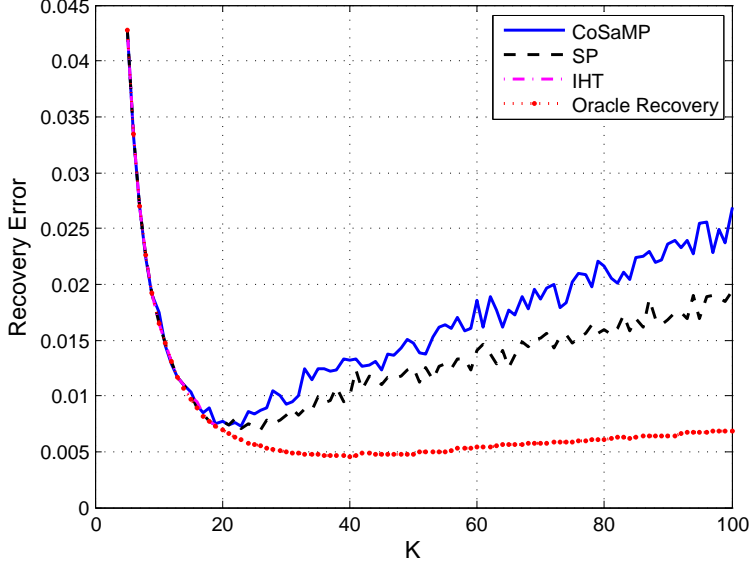


Figure 4: Relative error of greedy pursuits with replacement and oracle LS recovery for compressible signal versus different parameter  $K$  when  $\varepsilon_{\mathbf{A}} = 0.01$  and  $\varepsilon_{\mathbf{y}} = 0.01$ . IHT algorithm becomes invalid when  $K \geq 20$ , thus it is not plotted for the wider range.

be unavoidable in [43], therefore this bound is optimal up to a constant factor. The result is also extended to the nearly-sparse case in [29].

In other relevant works [33, 34], the error bounds of CoSaMP and SP are also derived under general perturbations. The analysis is performed following the steps of [32] for basis pursuit. Define

$$\alpha_K = \frac{\|\mathbf{x} - \mathbf{x}_K\|_2}{\|\mathbf{x}_K\|_2}, \quad \beta_K = \frac{\|\mathbf{x} - \mathbf{x}_K\|_1}{\sqrt{K}\|\mathbf{x}_K\|_2}, \quad \kappa_{\mathbf{A}}^{(K)} = \frac{\sqrt{1 + \delta_K}}{\sqrt{1 - \delta_K}}.$$

The main result in [33] states that under the conditions that

$$\delta_{4K} \leq \frac{1.1}{(1 + \varepsilon_{\mathbf{A}}^{(4K)})^2} - 1 \quad (62)$$

and

$$\alpha_K + \beta_K \leq \frac{1}{2\kappa_{\mathbf{A}}^{(K)}}, \quad (63)$$

the recovered solution of CoSaMP satisfies

$$\|\mathbf{x} - \hat{\mathbf{x}}\|_2 \leq C \cdot \left( \|\mathbf{x} - \mathbf{x}_K\|_2 + \frac{1}{\sqrt{K}} \|\mathbf{x} - \mathbf{x}_K\|_1 + \|\mathbf{e}\|_2 + (\|\mathbf{E}\|_2 \alpha_K + \|\mathbf{E}\|_2^{(K)}) \|\mathbf{y}\|_2 \right). \quad (64)$$

Further define

$$\varepsilon_{\mathbf{A}, K, \mathbf{y}} \triangleq \left( \frac{\varepsilon_{\mathbf{A}}^{(K)} \kappa_{\mathbf{A}}^{(K)} + \varepsilon_{\mathbf{A}} \gamma_{\mathbf{A}} \alpha_K}{1 - \kappa_{\mathbf{A}}^{(K)} (\alpha_K + \beta_K)} + \varepsilon_{\mathbf{y}} \right) \|\mathbf{y}\|_2, \quad (65)$$

where

$$\gamma_{\mathbf{A}} = \frac{\|\mathbf{A}\|_2}{\sqrt{1 - \delta_K}}. \quad (66)$$

The main result in [34] states that under the conditions that

$$\delta_{3K} \leq \frac{1.083}{(1 + \varepsilon_{\mathbf{A}}^{(3K)})^2} - 1 \quad (67)$$

and that

$$\alpha_K + \beta_K < \frac{1}{\kappa_{\mathbf{A}}^{(K)}}, \quad (68)$$

the recovered solution of SP satisfies

$$\|\mathbf{x} - \hat{\mathbf{x}}\|_2 \leq (C_K'' + 1)\|\mathbf{x} - \mathbf{x}_K\|_2 + C_K'' \frac{\|\mathbf{x} - \mathbf{x}_K\|_1}{\sqrt{K}} + C_K' \varepsilon_{\mathbf{A}, K, \mathbf{y}}. \quad (69)$$

Our analysis, on the other hand, derives a unified form of the relative error bounds of CoSaMP, SP, and IHT under general perturbations. Specified in Theorem 2, for compressible signals, when measurement vector and sensing matrix are perturbed, under the condition of (16), the relative error bound of recovered solution in each iteration is derived as (17). It is also proved that after finite iterations, the error bound is linear in both perturbations and almost linear in the approximation error, as in (19). The result is compared with oracle recovery, where the locations of  $K$  largest entries in magnitude are known a priori. It is also proved that for some certain  $\mathbf{x}$ ,  $\mathbf{e}$ , and  $\mathbf{E}$ , the lower bound of oracle recovery only differs in coefficients from the error bounds of greedy pursuits with replacement. Therefore, the oracle-order recovery performance of them is guaranteed in this work.

The main difference between [29] and our work is that we consider a more general completely perturbed scenario, and the optimality of the recovery performance is also in this sense. Compared with [33, 34], the demand of RICs in our work is for the perturbed matrix  $\tilde{\mathbf{A}}$  other than  $\mathbf{A}$ , which is due to the fact that only  $\tilde{\mathbf{A}}$  is available for recovery. Also, the demand of RICs is with a constant parameter here. In addition, the condition such as (63) or (68) is not required in our assumption. Our results are compared with oracle recovery and shown to be optimal up to the coefficients, and they are verified by plentiful numerical simulations.

## 7 Conclusion

In this paper, the relative error bounds of three greedy pursuits with replacement under general perturbations are derived. It is shown that these bounds are linear in both perturbations, which leads to the stability of these algorithms against general perturbations. Furthermore, these error bounds are compared with that of oracle LS solution with the locations of  $K$  largest entries in magnitude known a priori. Since the error bounds of CoSaMP,

Table 3: The Pseudo Code of CoSaMP

---

**Initialization**

1.  $l = 0$ ;
2.  $\mathbf{s}^{[0]} = \mathbf{0}$ ,  $\mathbf{r} = \mathbf{y}$ ;

---

**Repeat**

3.  $l = l + 1$ ;
4.  $\mathbf{u} = \mathbf{A}^T \mathbf{r}$ ,  $\Omega = \text{supp}(\mathbf{u}_{2K})$ ;
5.  $\tilde{\mathcal{S}} = \Omega \cup \text{supp}(\mathbf{s}^{[l-1]})$ ;
6.  $\mathbf{b}_{\tilde{\mathcal{S}}} = \mathbf{A}_{\tilde{\mathcal{S}}}^\dagger \mathbf{y}$ ,  $\mathbf{b}_{\tilde{\mathcal{S}}^c} = \mathbf{0}$ ;
7.  $\mathbf{s}^{[l]} = \mathbf{b}_K$ ,  $\mathbf{r} = \mathbf{y} - \mathbf{A} \mathbf{s}^{[l]}$ ;

**Until** stopping criterion is satisfied

---

SP, and IHT algorithms are of the same order as the lower bound of oracle solution for some certain signal and perturbations, it can be concluded that these greedy pursuits with replacement can provide oracle-order recovery performance against general perturbations. Numerical simulations verify that the relative recovery errors of these algorithms are linear in both perturbations, and they exhibit oracle-order recovery performance. Discussions and simulations also reveal how moderate parameter  $K$  achieves better recovery performance for compressible signals.

## Appendix A Discussions of Algorithms

Greedy pursuits with replacement are briefly introduced in Section II-B. The pseudo codes of CoSaMP, SP, and IHT are demonstrated in Table 3, Table 4, and Table 5, respectively. The definitions of the symbols are presented in Section III-A, and  $\text{supp}(\mathbf{u})$  denotes the support set of  $\mathbf{u}$ .

Based on Theorem 1, two remarks are derived as follows.

**Remark 4** In the noiseless scenario, after finite iterations, the recovered solutions of CoSaMP and SP are guaranteed to be identical to the sparse signal  $\mathbf{s}$ . The result can be verified through the following statement.

Suppose  $s_{\min}$  is the smallest magnitude of the nonzero entries of  $\mathbf{s}$ . Then after

$$l = \left\lceil \log_C \left( \frac{s_{\min}}{a \|\mathbf{s}\|_2} \right) \right\rceil + 1$$

iterations, the recovered solution  $\mathbf{s}^{[l]}$  obeys

$$\|\mathbf{s} - \mathbf{s}^{[l]}\|_2 < s_{\min}. \quad (70)$$

Table 4: The Pseudo Code of SP

---

**Initialization**

1.  $l = 0$ ;
2.  $\mathbf{u} = \mathbf{A}^T \mathbf{y}$ ,  $S^0 = \text{supp}(\mathbf{u}_K)$ ;
3.  $\mathbf{s}_{S^0}^{[0]} = \mathbf{A}_{S^0}^\dagger \mathbf{y}$ ,  $\mathbf{s}_{(S^0)^c}^{[0]} = \mathbf{0}$ ;
4.  $\mathbf{y}_r^0 = \mathbf{y} - \mathbf{A} \mathbf{s}^{[0]}$ ;

---

**Repeat**

5.  $l = l + 1$ ;
6.  $\mathbf{u} = \mathbf{A}^T \mathbf{y}_r^{l-1}$ ,  $\Omega = \text{supp}(\mathbf{u}_K)$ ;
7.  $\tilde{S}^l = \Omega \cup S^{l-1}$ ;
8.  $\mathbf{s}_p = \mathbf{A}_{\tilde{S}^l}^\dagger \mathbf{y}$ ,  $S^l = \text{supp}((\mathbf{s}_p)_K)$ ;
9.  $\mathbf{s}_{S^l}^{[l]} = \mathbf{A}_{S^l}^\dagger \mathbf{y}$ ,  $\mathbf{s}_{(S^l)^c}^{[l]} = \mathbf{0}$ ;
10.  $\mathbf{y}_r^l = \mathbf{y} - \mathbf{A} \mathbf{s}^{[l]}$ ;

**Until** stopping criterion is satisfied

---

If the support of  $\mathbf{s}$  is perfectly recovered, then for CoSaMP and SP, the solution is already identical to  $\mathbf{s}$ . Otherwise, at least one nonzero entry is not detected, thus the recovery error is no less than  $s_{\min}$ , which contradicts (70). Notice that the solution of IHT does not possess the above property, since exact support recovery for IHT does not imply exact signal recovery.

**Remark 5** According to (10), after

$$l = \left\lceil \log_C \left( \frac{\|\mathbf{e}\|_2}{a \|\mathbf{s}\|_2} \right) \right\rceil$$

iterations, the error bound of the recovered solution satisfies

$$\|\mathbf{s} - \mathbf{s}^{[l]}\|_2 \leq (D + 1) \|\mathbf{e}\|_2, \quad (71)$$

which means that the error bound is proportional to the  $\ell_2$  norm of the noise, and the recovery performance of greedy pursuits with replacement is stable in this scenario.

## Appendix B Proof of Theorem 1

**PROOF** For CoSaMP algorithm, the inequality (8) can be obtained by following the steps of the proof of Theorem 4.1 in [21], while preserving RICs during the derivation. As for the second part of the theorem, it is easy to derive that if  $\delta_{4K} < 0.171$ , then  $C < 1$ . By

Table 5: The Pseudo Code of IHT

---

**Initialization**

1.  $l = 0$ ;
  2.  $\mathbf{s}^{[0]} = \mathbf{0}$ ;
- 

**Repeat**

3.  $l = l + 1$ ;
4.  $\mathbf{u} = \mathbf{A}^T (\mathbf{y} - \mathbf{A}\mathbf{s}^{[l-1]})$ ;
5.  $\mathbf{s}^{[l]} = (\mathbf{s}^{[l-1]} + \mathbf{u})_K$ ;

**Until** stopping criterion is satisfied

---

recursion, it can be proved from (8) that

$$\|\mathbf{s} - \mathbf{s}^{[l]}\|_2 \leq C^l \|\mathbf{s} - \mathbf{s}^{[0]}\|_2 + \frac{C_1}{1-C} \|\mathbf{e}\|_2, \quad (72)$$

which arrives (10).

For IHT algorithm, the inequality (8) is directly derived in the proof of Theorem 5 in [23]. Similar to CoSaMP, the second part of the theorem can also be proved from (8).

For SP algorithm, the result cannot be directly found in [22]. However, it can be proved based on several facts about SP in [22], and they are stated here without proofs.

**Lemma 3** (Lemma 3 in [22]) *Let  $\mathbf{s} \in \mathbb{R}^N$  be a  $K$ -sparse vector, and  $\tilde{\mathbf{y}} = \mathbf{A}\mathbf{s} + \mathbf{e}$  is a noisy measurement vector where  $\mathbf{A} \in \mathbb{R}^{M \times N}$  satisfies the RIP with parameter  $\delta_{3K}$ . For an arbitrary set  $\hat{S} \subset \{1, \dots, N\}$  such that  $|\hat{S}| \leq K$ , define  $\hat{\mathbf{s}}$  as*

$$\hat{\mathbf{s}}_{\hat{S}} = \mathbf{A}_{\hat{S}}^\dagger \tilde{\mathbf{y}}, \quad \hat{\mathbf{s}}_{\hat{S}^c} = \mathbf{0}. \quad (73)$$

Then

$$\|\mathbf{s} - \hat{\mathbf{s}}\|_2 \leq \frac{1}{1 - \delta_{3K}} \|\mathbf{s}_{S - \hat{S}}\|_2 + \frac{1 + \delta_{3K}}{1 - \delta_{3K}} \|\mathbf{e}\|_2. \quad (74)$$

**Lemma 4** (Theorem 10 in [22]) *It holds that*

$$\begin{aligned} \|\mathbf{s}_{S - \hat{S}^t}\|_2 &\leq \frac{2\delta_{3K}}{(1 - \delta_{3K})^2} \|\mathbf{s}_{S - S^{t-1}}\|_2 + \frac{2(1 + \delta_{3K})}{1 - \delta_{3K}} \|\mathbf{e}\|_2, \\ \|\mathbf{s}_{S - S^t}\|_2 &\leq \frac{1 + \delta_{3K}}{1 - \delta_{3K}} \|\mathbf{s}_{S - \hat{S}^t}\|_2 + \frac{2}{1 - \delta_{3K}} \|\mathbf{e}\|_2. \end{aligned}$$

Therefore

$$\|\mathbf{s}_{S - S^t}\|_2 \leq \frac{2\delta_{3K} + 2\delta_{3K}^2}{(1 - \delta_{3K})^3} \|\mathbf{s}_{S - S^{t-1}}\|_2 + \frac{4(1 + \delta_{3K})}{(1 - \delta_{3K})^2} \|\mathbf{e}\|_2.$$

Lemma 4 directly implies inequality (9). It's easy to check that if  $\delta_{3K} \leq 0.206$ , then  $C < 1$ , and

$$\begin{aligned} \|\mathbf{s}_S - \mathbf{s}^l\|_2 &\leq C^l \|\mathbf{s}_S - \mathbf{s}^0\|_2 + \frac{C_1}{1-C} \|\mathbf{e}\|_2 \\ &\leq C^l \|\mathbf{s}\|_2 + \frac{C_1}{1-C} \|\mathbf{e}\|_2. \end{aligned} \tag{75}$$

Applying Lemma 3 to (75) and with the fact that  $\delta_{3K} \leq 0.206$ , inequality (10) can be derived. ■

## References

- [1] E. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Information Theory*, vol. 52, no. 2, pp. 489-509, Feb. 2006.
- [2] D. Donoho, "Compressed sensing," *IEEE Trans. Information Theory*, vol. 52, no. 4, pp. 1289-1306, Apr. 2006.
- [3] E. Candès and T. Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies?" *IEEE Trans. Information Theory*, vol. 52, no. 12, pp. 5406-5425, Dec. 2006.
- [4] S. Chen, D. Donoho, and M. Saunders, "Atomic decomposition by basis pursuit," *SIAM Journal on Scientific Computing*, vol. 20, no. 1, pp. 33-61, Aug. 1998.
- [5] E. Candès and T. Tao, "Decoding by linear programming," *IEEE Trans. Information Theory*, vol. 51, no. 12, pp. 4203-4215, Dec. 2005.
- [6] E. Candès, "The restricted isometry property and its implications for compressed sensing," *Comptes Rendus Mathématique*, vol. 346, no. 9-10, pp. 589-592, May 2008.
- [7] T. Cai, L. Wang, and G. Xu, "New bounds for restricted isometry constants," *IEEE Trans. Information Theory*, vol. 56, no. 9, pp. 4388-4394, Sept. 2010.
- [8] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge University Press, 2004.
- [9] S. Kim, K. Koh, M. Lusig, S. Boyd, and D. Gorinevsky, "An interior-point method for large-scale  $\ell_1$ -regularized least squares," *IEEE Journal of Selected Topics in Signal Process.*, vol. 1, no. 4, pp. 606-617, Dec. 2007.
- [10] M. Osborne, B. Presnell, and B. Turlach, "A new approach to variable selection in least squares problems," *IMA Journal Numerical Analysis*, vol. 20, no. 3, pp. 389-403, July 2000.



- [11] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani, "Least angle regression," *Annals of Statistics*, vol. 32, no. 2, pp. 407-499, Apr. 2004.
- [12] I. Gorodnitsky and B. Rao, "Sparse signal reconstruction from limited data using FOCUSS: A re-weighted minimum norm algorithm," *IEEE Trans. Signal Process.*, vol. 45, no. 3, pp. 600-616, Mar. 1997.
- [13] R. Chartrand and W. Yin, "Iteratively reweighted algorithms for compressive sensing," *Proc. IEEE ICASSP*, pp. 3869-3872, Mar. 2008.
- [14] H. Mohimani, M. Babaie-Zadeh, and C. Jutten, "A fast approach for overcomplete sparse decomposition based on smoothed  $\ell^0$  norm," *IEEE Trans. Signal Process.*, vol. 57, no. 1, pp. 289-301, Jan. 2009.
- [15] J. Jin, Y. Gu, and S. Mei, "A Stochastic gradient approach on compressive sensing signal reconstruction based on adaptive filtering framework," *IEEE Journal of Selected Topics in Signal Process.*, vol. 4, no. 2, pp. 409-420, Apr. 2010.
- [16] X. Wang, Y. Gu, and L. Chen, "Proof of convergence and performance analysis for sparse recovery via zero-Point attracting projection," *IEEE Trans. Signal Processing*, vol. 60, no. 8, pp. 4081-4093, Aug. 2012.
- [17] R. Chartrand, "Exact reconstruction of sparse signals via nonconvex minimization," *IEEE Signal Process. Letters*, vol. 14, no. 10, pp. 707-710, Oct. 2007.
- [18] J. Tropp and A. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Information Theory*, vol. 53, no. 12, pp. 4655-4666, Dec. 2007.
- [19] D. Needell and R. Vershynin, "Uniform uncertainty principle and signal recovery via regularized orthogonal matching pursuit," *Foundations of Computational Mathematics*, vol. 9, no. 3, pp. 317-334, June 2009.
- [20] D. Donoho, Y. Tsaig, I. Drori, and J. Starck, "Sparse solution of underdetermined systems of linear equations by stagewise orthogonal matching pursuit," *IEEE Trans. Information Theory*, vol. 58, no. 2, pp. 1094-1121, Feb. 2012.
- [21] D. Needell and J. Tropp, "CoSaMP: Iterative signal recovery from incomplete and inaccurate samples," *Appl. Comput. Harmon. Anal.*, vol. 26, no. 3, pp. 301-321, May 2009.
- [22] W. Dai and O. Milenkovic, "Subspace pursuit for compressive sensing signal reconstruction," *IEEE Trans. Information Theory*, vol. 55, no. 5, pp. 2230-2249, May 2009.
- [23] T. Blumensath and M. Davies, "Iterative hard thresholding for compressed sensing," *Appl. Comput. Harmon. Anal.*, vol. 27, no. 3, pp. 265-274, Nov. 2009.

- [24] T. Blumensath and M. Davies, “Normalized iterative hard thresholding: Guaranteed stability and performance,” *IEEE Journal of Selected Topics in Signal Process.*, vol. 4, no. 2, pp. 298-309, Apr. 2010.
- [25] D. Needell and R. Vershynin, “Signal recovery from incomplete and inaccurate measurements via regularized orthogonal matching pursuit,” *IEEE Journal of Selected Topics in Signal Process.*, vol. 4, no. 2, pp. 310-316, Apr. 2010.
- [26] E. Candès and T. Tao, “The dantzig selector: Statistical estimation when  $p$  is much larger than  $n$ ,” *Annals of Statistics*, vol. 35, no. 6, pp. 2313-2351, Dec. 2007.
- [27] P. Bickel, Y. Ritov, and A. Tsybakov, “Simultaneous analysis of lasso and dantzig selector,” *Annals of Statistics*, vol. 37, no. 4, pp. 1705-1732, Aug. 2009.
- [28] Z. Ben-Haim, Y. Eldar, and M. Elad, “Coherence-based performance guarantees for estimating a sparse vector under random noise,” *IEEE Trans. Signal Process.*, vol. 58, no. 10, pp. 5030-5043, Oct. 2010.
- [29] R. Giryes and M. Elad, “RIP-based near-oracle performance guarantees for SP, CoSaMP, and IHT,” *IEEE Trans. Signal Process.*, vol. 60, no. 3, pp. 1465-1468, Mar. 2012.
- [30] I. Rish and G. Grabarnik, “Sparse signal recovery with exponential-family noise,” *Allerton Conference on Communication, Control, and Computing*, pp. 60-66, Sept. 2009.
- [31] M. Raginsky, R. Willett, Z. Harmany, and R. Marcia, “Compressed sensing performance bounds under poisson noise,” *IEEE Trans. Signal Process.*, vol. 58, no. 8, pp. 3990-4002, Aug. 2010.
- [32] M. Herman and T. Strohmer, “General deviants: An analysis of perturbations in compressed sensing,” *IEEE Journal of Selected Topics in Signal Process.*, vol. 4, no. 2, pp. 342-349, Apr. 2010.
- [33] M. Herman and D. Needell, “Mixed operators in compressed sensing,” *Conference on Information Sciences and Systems*, pp. 1-6, Mar. 2010.
- [34] Q. Wang and Z. Liu, “Sampling matrix perturbation analysis of subspace pursuit for compressive sensing,” *Communications in Computer and Information Science*, vol. 86, pp. 581-588, 2011.
- [35] A. Aldroubi, X. Chen, and A. Powell, “Perturbations of measurement matrices and dictionaries in compressed sensing,” *Appl. Comput. Harmon. Anal.*, vol. 33, no. 2, pp. 282-291, Sept. 2012.
- [36] J. Ding, L. Chen, and Y. Gu, “Perturbation analysis of orthogonal matching pursuit,” *IEEE Trans. Signal Process.*, vol. 61, no. 2, pp. 398-410, Jan. 2013.

- [37] Y. Chi, L. Scharf, A. Pezeshki, and A. Calderbank, "Sensitivity to basis mismatch in compressed sensing," *IEEE Trans. Signal Process.*, vol. 59, no. 5, pp. 2182-2195, May 2011.
- [38] H. Zhu, G. Leus, and G. Giannakis, "Sparsity-cognizant total least-squares for perturbed compressive sampling," *IEEE Trans. Signal Process.*, vol. 59, no. 5, pp. 2002-2016, May 2011.
- [39] B. Bah and J. Tanner, "Improved bounds on restricted isometry constants for Gaussian matrices," *SIAM Journal on Matrix Analysis and Applications*, vol. 31, no. 5, pp. 2882-2898, July 2010.
- [40] J. Blanchard, C. Cartis, and J. Tanner, "Compressed sensing: How sharp is the restricted isometry property?" vol. 53, no. 1, pp. 105-125, Feb. 2011.
- [41] W. Dai and O. Milenkovic, "Subspace pursuit for compressive sensing: Closing the gap between performance and complexity," available online: <http://citeseer.ist.psu.edu/viewdoc/summary?doi=10.1.1.154.8384>.
- [42] T. Blumensath and M. Davies, "Iterative thresholding for sparse approximations," *Journal of Fourier Analysis and Applications*, vol. 14, no. 5-6, pp. 629-654, Dec. 2008.
- [43] E. Candès, "Modern statistical estimation via oracle inequalities," *Acta Numerica*, vol. 15, pp. 257-325, May 2006.