Recovery of Sparse Signals via Generalized Orthogonal Matching Pursuit: A New Analysis

Jian Wang, *Student Member, IEEE*, Suhyuk Kwon, *Student Member, IEEE*, Ping Li, and Byonghyo Shim, *Senior Member, IEEE*

***Abstract*—As an extension of orthogonal matching pursuit (OMP) for improving the recovery performance of sparse signals, generalized OMP (gOMP) has recently been studied in the literature. In this paper, we present a new analysis of the gOMP algorithm using the restricted isometry property (RIP). We show that if a measurement matrix $\mathbf{\Phi} \in \mathcal{R}^{m \times n}$ satisfies the RIP with isometry constant $\delta_{\max\{9,S+1\}K} \leq \frac{1}{8}$, then gOMP performs stable reconstruction of all K-sparse signals $\mathbf{x} \in \mathcal{R}^n$ from the noisy measurements $\mathbf{y} = \mathbf{\Phi x} + \mathbf{v}$, within $\max\left\{K, \left\lfloor \frac{8K}{S} \right\rfloor\right\}$ iterations, where $\mathbf{v}$ is the noise vector and S is the number of indices chosen in each iteration of the gOMP algorithm. For Gaussian random measurements, our result indicates that the number of required measurements is essentially $m = \mathcal{O}\left(K \log \frac{n}{K}\right)$, which is a significant improvement over the existing result $m = \mathcal{O}\left(K^2 \log \frac{n}{K}\right)$, especially for large K.**

***Index Terms*—Compressed Sensing (CS), Generalized Orthogonal Matching Pursuit (gOMP), Mean Square Error (MSE), Restricted Isometry Property (RIP), sparse recovery, stability.**

I. Introduction

ORTHOGONAL MATCHING PURSUIT (OMP) is a greedy algorithm widely used for the recovery of sparse signals [1]–[9]. The goal of OMP is to recover a K-sparse signal vector $\mathbf{x} \in \mathcal{R}^n$ ($\|\mathbf{x}\|_0 \leq K$) from its linear measurements

$$\mathbf{y} = \mathbf{\Phi x}, \tag{1}$$

where $\mathbf{\Phi} \in \mathcal{R}^{m \times n}$ is called the measurement matrix. For each iteration, OMP estimates the support (positions of non-zero elements) of $\mathbf{x}$ by adding an index of the column in $\mathbf{\Phi}$ that is most correlated with the current residual. The vestiges of columns in the estimated support are then eliminated from the measurements $\mathbf{y}$, yielding an updated residual for the next iteration. See [1], [3] for details on the OMP algorithm.

While the number of iterations of the OMP algorithm is typically set to the sparsity level K of the underlying signal to be recovered, there have been recent efforts to relax this constraint with the aim of enhancing recovery performance. In one direction, an approach allowing more iterations than the sparsity has been suggested [10]–[12]. In another direction, algorithms identifying multiple indices for each iteration have been proposed. Well-known examples include stagewise OMP (StOMP) [13], regularized OMP (ROMP) [14], CoSaMP [15], and subspace pursuit (SP) [16]. The key feature of these algorithms is to introduce special operations in the identification step to select multiple promising indices. Specifically, StOMP selects indices whose magnitudes of correlation exceed a deliberately designed threshold. ROMP chooses a set of K indices and then reduces the number of candidates using a predefined regularization rule. CoSaMP and SP add multiple indices and then prune a large portion of the chosen indices to refine the identification step. In contrast to these algorithms performing deliberate refinement of the identification step, a recently proposed extension of OMP, referred to as generalized OMP (gOMP) [17] (also known as OSGA or OMMP [18]–[20]), simply chooses S columns that are most correlated with the residual. A detailed description of the gOMP algorithm is given in Table I.

The main motivation of gOMP is to reduce computational complexity. Since the OMP algorithm chooses one index at a time, computational complexity depends heavily on the sparsity K. When K is large, therefore, computational cost of OMP might be problematic. In the gOMP algorithm, however, more than one "correct" index can be chosen in each iteration due to the identification of multiple indices at a time. Therefore, the number of iterations needed to complete the algorithm is usually much smaller than that of the OMP algorithm. In fact, it has been shown that the computational complexity of gOMP is $2smn + (2S^2 + S)s^2m$, where s is the number of actually performed iterations [17], while that of OMP is $2Kmn + 3K^2m$.

Manuscript received February 18, 2014; revised March 31, 2015 and September 06, 2015; accepted October 24, 2015. Date of publication November 05, 2015; date of current version January 20, 2016. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Ruixin Niu. The work of J. Wang and P. Li were supported in part by NSF-III-1360971, NSF-Bigdata-1419210, ONRN00014-13-1-0764, and AFOSR-FA9550-13-1-0137. The work of J. Wang was also supported in part by Grant NSFC 61532009 and Grant 15KJA520001 of Jiangsu Province. The work of B. Shim was supported in part by ICT R&D program of MSIP/IITP, B0126-15-1017, Spectrum Sensing and Future Radio Communication Platforms and the National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIP) (2014R1A5A1011478).

J. Wang was with Department of Statistics and Biostatistics, Department of Computer Science, Rutgers University, Piscataway, NJ 08901 USA. He is now with B-DAT Lab, School of information and Control, Nanjing University of Information Science and Technology, Nanjing 210044, China (e-mail: wangjianeee@gmail.com).

S. Kwon was with Institute of New Media and Communications, Seoul National University, Seoul 151-742, Korea. He is now with Samsung Display, Asan 465, Korea (e-mail: shkwon@islab.snu.ac.kr).

P. Li is with Department of Statistics and Biostatistics, Department of Computer Science, Rutgers University, Piscataway, NJ 08901 USA (e-mail: pingli@stat.rutgers.edu).

B. Shim is with Institute of New Media and Communications and School of Electrical and Computer Engineering, Seoul National University, Seoul 151-742, Korea (e-mail: bshim@snu.ac.kr).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSP.2015.2498132

TABLE I
THE gOMP ALGORITHM

Input:	measurement matrix $\mathbf{\Phi} \in \mathcal{R}^{m\times n}$, measurements $\mathbf{y} \in \mathcal{R}^m$, sparsity level K, number of indices for each selection $S \leq K$.
Initialize:	iteration count $k = 0$, estimated list $T^0 = \emptyset$, residual vector $\mathbf{r}^0 = \mathbf{y}$.
While	$\|\mathbf{r}^k\|_2 > \epsilon$ and $k < \frac{m}{S}$ **do**
	$k = k + 1$.
	$\Lambda^k = \arg\max_{\Lambda:\|\Lambda\|=S} \|(\mathbf{\Phi}'\mathbf{r}^{k-1})_\Lambda\|_1$. **(Identification)**
	$T^k = T^{k-1} \cup \Lambda^k$. **(Augmentation)**
	$\hat{\mathbf{x}}^k = \arg\min_{supp(\mathbf{u})=T^k} \|\mathbf{y} - \mathbf{\Phi}\mathbf{u}\|_2$. **(Estimation)**
	$\mathbf{r}^k = \mathbf{y} - \mathbf{\Phi}\hat{\mathbf{x}}^k$. **(Residual Update)**
End	
Output	the estimated support $\hat{T} = \underset{\mathcal{A}:\|\mathcal{A}\|=K}{\arg\min} \|\hat{\mathbf{x}}^k - \hat{\mathbf{x}}^k_{\mathcal{A}}\|_2$ and signal $\hat{\mathbf{x}}$ satisfying $\hat{\mathbf{x}}_{\{1,\cdots,n\}\setminus\hat{T}} = \mathbf{0}$ and $\hat{\mathbf{x}}_{\hat{T}} = \mathbf{\Phi}^\dagger_{\hat{T}}\mathbf{y}$.

In many situations where $K < m \ll n$, the computational complexity is dominated by the first term for both approaches. Since s is generally much smaller than K, gOMP has lower computational complexity than OMP.

In analyzing the theoretical performance of gOMP, the restricted isometry property (RIP) has been frequently used [17]–[25]. A measurement matrix $\mathbf{\Phi}$ is said to satisfy the RIP of order K if there exists a constant $\delta(\mathbf{\Phi}) \in [0, 1)$, such that [26]

$$(1 - \delta(\mathbf{\Phi}))\,\|\mathbf{x}\|_2^2 \leq \|\mathbf{\Phi}\mathbf{x}\|_2^2 \leq (1 + \delta(\mathbf{\Phi}))\,\|\mathbf{x}\|_2^2, \tag{2}$$

for any K-sparse vector $\mathbf{x}$. In particular, the minimum of all constants $\delta(\mathbf{\Phi})$ satisfying (2) is called the restricted isometry constant (RIC) and denoted by $\delta_K(\mathbf{\Phi})$. In the sequel, we use δ_K instead of $\delta_K(\mathbf{\Phi})$ for notational simplicity. In [17], it has been shown that gOMP ensures perfect recovery of any K-sparse signal $\mathbf{x}$ from $\mathbf{y} = \mathbf{\Phi}\mathbf{x}$ within K iterations under

$$\delta_{SK} < \frac{\sqrt{S}}{\sqrt{K} + 3\sqrt{S}}, \tag{3}$$

where S is the number of indices chosen in each iteration. In the noisy scenario where the measurements are corrupted by a noise vector $\mathbf{v}$ (i.e., $\mathbf{y} = \mathbf{\Phi}\mathbf{x} + \mathbf{v}$), it has been shown that if conditions (3) and $\delta_{SK+S} < 1$ are satisfied, the output $\hat{\mathbf{x}}$ of gOMP after K iterations satisfies [17]

$$\|\hat{\mathbf{x}} - \mathbf{x}\|_2 \leq C\sqrt{K}\|\mathbf{v}\|_2, \tag{4}$$

where C is a constant.

While the empirical recovery performance of gOMP is promising, theoretical results to date are relatively weak when compared to state-of-the-art recovery algorithms. For example, performance guarantees of basis pursuit (BP) [27] and CoSaMP are given by $\delta_{2K} < \sqrt{2} - 1$ [28] and $\delta_{4K} < 0.1$ [15], but conditions for gOMP require that the RIC should be inversely proportional to $\sqrt{K}$ [17]–[25]. Another weakness for existing theoretical results of gOMP lies in the lack of stability guarantees for the noisy scenario. For example, it can be seen from (4) that the ℓ_2-norm of the recovery error of gOMP is upper bounded by $C\sqrt{K}\|\mathbf{v}\|_2$. This implies that even for a small $\|\mathbf{v}\|_2$, the ℓ_2-norm of recovery error can be unduly large when the sparsity K approaches infinity. In contrast, recovery error bound of BP denoising (BPDN) and CoSaMP is directly proportional to $\|\mathbf{v}\|_2$, and hence these are stable under measurement noise [15], [27].

The main purpose of this paper is to provide an improved performance analysis of the gOMP algorithm. Specifically, we show that if the measurement matrix $\mathbf{\Phi}$ satisfies the RIP with isometry constant

$$\delta_{\max\{9,S+1\}K} \leq \frac{1}{8}, \tag{5}$$

gOMP achieves stable recovery of any K-sparse signal $\mathbf{x}$ from the noisy measurements $\mathbf{y} = \mathbf{\Phi}\mathbf{x} + \mathbf{v}$ within $\max\left\{K, \left\lfloor \frac{8K}{S} \right\rfloor\right\}$ iterations. That is, the ℓ_2-norm of recovery error satisfies

$$\|\hat{\mathbf{x}} - \mathbf{x}\|_2 \leq C\|\mathbf{v}\|_2, \tag{6}$$

where C is a constant. In the special case where $\|\mathbf{v}\|_2 = 0$ (i.e., the noise-free case), we show that gOMP accurately recovers all K-sparse signals in $\max\left\{K, \left\lfloor \frac{8K}{S} \right\rfloor\right\}$ iterations under

$$\delta_{7K} \leq \frac{1}{8}. \tag{7}$$

When compared to previous results [17]–[25], there are two important aspects to these new results.

i) Our results show that the gOMP algorithm can recover sparse signals under a similar RIP condition required by the state-of-the-art sparse recovery algorithms (e.g., BP and CoSaMP). For Gaussian random measurement matrices, this implies that the number of measurements required for gOMP is essentially $m = \mathcal{O}\left(K \log \frac{n}{K}\right)$ [26], [29], which is significantly smaller than the previous result, $m = \mathcal{O}\left(K^2 \log \frac{n}{K}\right)$.

ii) In [17], it was shown that the ℓ_2-norm of the recovery error in the noisy scenario depends linearly on $\sqrt{K}\|\mathbf{v}\|_2$. However, our new result suggests that the recovery distortion of gOMP is upper bounded by a constant times $\|\mathbf{v}\|_2$, which strictly ensures the stability of gOMP under measurement noise.

We briefly summarize notations used in this paper. For a vector $\mathbf{x} \in \mathcal{R}^n$, $T = supp(\mathbf{x}) = \{i | x_i \neq 0\}$ represents the set of its non-zero positions. $\Omega = \{1, \cdots, n\}$. For a set $A \subseteq \Omega$, $|A|$ denotes the cardinality of A. $T \setminus A$ is the set of all elements contained in T but not in A. $\mathbf{\Phi}_A \in \mathcal{R}^{m\times|A|}$ is the submatrix of $\mathbf{\Phi}$ that only contains columns indexed by A. $\mathbf{\Phi}'_A$ means the transpose of the matrix $\mathbf{\Phi}_A$. $\mathbf{x}_A \in \mathcal{R}^{|A|}$ is the vector, which equals $\mathbf{x}$ for elements indexed by A. If $\mathbf{\Phi}_A$ is full column rank, then $\mathbf{\Phi}^\dagger_A = \left(\mathbf{\Phi}'_A\mathbf{\Phi}_A\right)^{-1}\mathbf{\Phi}'_A$ is the pseudoinverse of $\mathbf{\Phi}_A$. By $span(\mathbf{\Phi}_A)$ we mean the span of columns in $\mathbf{\Phi}_A$. $\mathcal{P}_A = \mathbf{\Phi}_A\mathbf{\Phi}^\dagger_A$ is the projection onto $span(\mathbf{\Phi}_A)$. $\mathcal{P}^\perp_A = \mathbf{I} - \mathcal{P}_A$ is the projection onto the orthogonal complement of $span(\mathbf{\Phi}_A)$. At the kth iteration of gOMP, we use T^k, $\Gamma^k = T \setminus T^k$, $\hat{\mathbf{x}}^k$, and $\mathbf{r}^k$ to denote the estimated support, the remaining support set, the estimated sparse signal, and the residual vector, respectively.

The remainder of the paper is organized as follows. In Section II, we provide theoretical and empirical results of gOMP. In Section III, we present the proof of theoretical results, and in Section IV we conclude the paper.

II. Sparse Recovery With gOMP

A. Main Results

In this section, we provide the performance guarantees of gOMP in recovering sparse signals in the presence of noise. Since the noise-free scenario can be considered as a special case of the noisy scenario, extension to the noise-free scenario is straightforward. In the noisy scenario, perfect reconstruction of sparse signals is not possible, and hence we use the ℓ_2-norm of the recovery error as a performance measure. We first show that the ℓ_2-norm of the residual, after a specified number of iterations, is upper bounded by a quantity depending only on $\|\mathbf{v}\|_2$.

Theorem 1: Let $\mathbf{x} \in \mathcal{R}^n$ be any K-sparse vector, $\mathbf{\Phi} \in \mathcal{R}^{m\times n}$ be a measurement matrix, and $\mathbf{y} = \mathbf{\Phi}\mathbf{x} + \mathbf{v}$ be the noisy measurements, where $\mathbf{v}$ is a noise vector. Then, under

$$\delta_{\max\{Sk+7|\Gamma^k|, Sk+S+|\Gamma^k|\}} \leq \frac{1}{8}, \tag{8}$$

the residual of gOMP satisfies

$$\left\| \mathbf{r}^{k+\max\left\{|\Gamma^k|, \left\lfloor \frac{8|\Gamma^k|}{S} \right\rfloor\right\}} \right\|_2 \leq \mu_k \|\mathbf{v}\|_2, \tag{9}$$

where μ_K is a constant depending only on $\delta_{\max\{Sk+7|\Gamma^k|, Sk+S+|\Gamma^k|\}}$.

The proof will be given in Section III. One can observe from Theorem 1 that if gOMP already performs k iterations, then it requires at most $\max\left\{|\Gamma^k|, \left\lfloor \frac{8|\Gamma^k|}{S} \right\rfloor\right\}$ additional iterations to ensure that the ℓ_2-norm of residual falls below $\mu_K\|\mathbf{v}\|_2$. In particular, when $k = 0$ (i.e., at the beginning of the iterations), $|\Gamma^k| = K$ and the RIC in (8) is simplified to δ_{7K}, and hence we have a simple interpretation of Theorem 1.

Theorem 2 (Upper Bound of Residual): Let $\mathbf{x} \in \mathcal{R}^n$ be any K-sparse vector, $\mathbf{\Phi} \in \mathcal{R}^{m\times n}$ be the measurement matrix, and $\mathbf{y} = \mathbf{\Phi}\mathbf{x} + \mathbf{v}$ be the noisy measurements, where $\mathbf{v}$ is the noise vector. Then under $\delta_{7K} \leq \frac{1}{8}$, the residual of gOMP satisfies

$$\left\| \mathbf{r}^{\max\left\{K, \left\lfloor \frac{8K}{S} \right\rfloor\right\}} \right\|_2 \leq \mu_0 \|\mathbf{v}\|_2, \tag{10}$$

where μ_0 is a constant depending only on δ_{7K}.

From Theorem 2, we also obtain the exact recovery condition of gOMP in the noise-free scenario. In fact, in the absence of noise, Theorem 2 suggests that $\left\| \mathbf{r}^{\max\left\{K, \left\lfloor \frac{8K}{S} \right\rfloor\right\}} \right\|_2 = 0$ under $\delta_{7K} \leq \frac{1}{8}$. Therefore, gOMP recovers any K-sparse signal accurately within $\max\left\{K, \left\lceil \frac{8K}{S} \right\rceil\right\}$ iterations under $\delta_{7K} \leq \frac{1}{8}$.

We next show that the ℓ_2-norm of the recovery error is also upper bounded by the product of a constant and $\|\mathbf{v}\|_2$.

Theorem 3 (Stability Under Measurement Perturbations): Let $\mathbf{x} \in \mathcal{R}^n$ be any K-sparse vector, $\mathbf{\Phi} \in \mathcal{R}^{m\times n}$ be the measurement matrix, and $\mathbf{y} = \mathbf{\Phi}\mathbf{x} + \mathbf{v}$ be the noisy measurements, where $\mathbf{v}$ is the noise vector. Then under $\delta_{\max\{9,S+1\}K} \leq \frac{1}{8}$, gOMP satisfies

$$\left\| \hat{\mathbf{x}}^{\max\left\{K, \left\lfloor \frac{8K}{S} \right\rfloor\right\}} - \mathbf{x} \right\|_2 \leq \mu \|\mathbf{v}\|_2 \tag{11}$$

and

$$\|\hat{\mathbf{x}} - \mathbf{x}\|_2 \leq C\|\mathbf{v}\|_2, \tag{12}$$

where μ and C are constants depending on $\delta_{\max\{9,S+1\}K}$.

Proof: See Appendix A. □

Remark 1 (Comparison With Previous Results): From Theorem 2 and 3, we observe that gOMP is far more effective than seen in previous results. Indeed, the upper bounds in Theorem 2 and 3 are absolute constants and independent of the sparsity K, while those in previous works are inversely proportional to $\sqrt{K}$ (e.g., $\delta_{SK} < \frac{\sqrt{S}}{(2+\sqrt{2})\sqrt{K}}$ [18], $\delta_{SK} < \frac{\sqrt{S}}{\sqrt{K}+3\sqrt{S}}$ [17], and $\delta_{SK} < \frac{\sqrt{S}}{\sqrt{K}+2\sqrt{S}}$ [22]). Clearly these upper bounds will vanish when K is large.

Remark 2 (Number of Measurements): It is well known that a random measurement matrix $\mathbf{\Phi} \in \mathcal{R}^{m\times n}$, which has independent and identically distributed (i.i.d.) entries with Gaussian distribution $\mathcal{N}\left(0, \frac{1}{m}\right)$, obeys the RIP with $\delta_K \leq \varepsilon$ with overwhelming probability if $m = \mathcal{O}\left(\frac{K \log \frac{n}{K}}{\varepsilon^2}\right)$ [26], [29]. When the recovery conditions in [17]–[25] are used, the number of required measurements is expressed as $m = \mathcal{O}\left(K^2 \log \frac{n}{K}\right)$. Whereas, our new conditions require $m = \mathcal{O}\left(K \log \frac{n}{K}\right)$, which is significantly smaller than the previous result, in particular for large K.

Remark 3 (Comparison With Information-Theoretic Results): It is worth comparing our result with information-theoretic results in [30]–[32]. Those results, which are obtained by a single-letter characterization in a large system limit, provide a performance limit of maximum a posteriori (MAP) and minimum mean square error (MMSE) estimation. For Gaussian random measurements, it is known that the MMSE estimation achieves a scaling of $m = \mathcal{O}(K)$ when the signal-to-noise ratio (SNR) is sufficiently large [30], [31]. In contrast, the gOMP algorithm requires $m = \mathcal{O}\left(K \log \frac{n}{K}\right)$, which is slightly larger than the MMSE estimation due to the term $\log \frac{n}{K}$.

Remark 4 (Recovery Error): The constants μ_0, μ and C in Theorem 2 and 3 can be estimated from the RIC. For example, when $\delta_{7K} \leq \delta_{\max\{9,S+1\}K} \leq 0.05$, we have $\mu_0 \leq 49$, $\mu \leq 52$, and $C \leq 110$. It is interesting to compare the constant C of gOMP with MMSE results [30]–[32]. Consider the scenario where $\mathbf{\Phi}$ is a random matrix having i.i.d. elements of zero mean and $\frac{1}{m}$ variance, $\mathbf{x}$ is a sparse vector with each non-zero element taking the value ± 1 with equal probability, and $\mathbf{v}$ is the noise vector with i.i.d. Gaussian elements. Consider $n = 10,000$ and $m = 500$ and suppose $\mathbf{x}$ has a sparsity rate $p = 0.001$ (so that the sparsity level K is 10 on average). Then, for an SNR of 0 dB, the required bound of MMSE is 8.6×10^{-6} (per dimension) [31], which amounts to $\|\mathbf{x} - \hat{\mathbf{x}}\|_2 \leq 0.027\|\mathbf{v}\|_2$ so that this constant 0.027 is smaller than the constant C of gOMP.[1] In fact, the constant C obtained in Theorem 3 is in general loose, as we will see in the simulations. This is mainly because 1) our analysis is based on the RIP framework so that the analysis is in essence the worst-case-analysis, and 2) many inequalities used in our analysis are not tight.

Remark 5 (Comparison With OMP): When $S = 1$, gOMP returns to the OMP algorithm and Theorem 3 suggests that OMP performs stable recovery of all K-sparse signals in

[1]SNR = 0 dB implies that $\frac{\|\mathbf{\Phi}\mathbf{x}\|_2^2}{\|\mathbf{v}\|_2^2} = 1$. Since each element in $\mathbf{\Phi}$ has power $\frac{1}{m}$, we have $\mathbb{E}[(\mathbf{\Phi}\mathbf{x})_j^2] = \frac{pn}{m} = \frac{1}{50}$, which implies that $\mathbb{E}[v_j^2] = \frac{1}{50}$, and hence $\mathbb{E}[\|\mathbf{v}\|_2] = \sqrt{10}$.

$\max\left\{K, \lfloor \frac{8K}{S} \rfloor\right\} = 8K$ iterations under $\delta_{9K} \le \frac{1}{8}$. In a recent work of Zhang [11, Theorem 2.1], it has been shown that OMP can achieve stable recovery of K-sparse signals in $16.6K$ iterations under $\delta_{17.6K} \le \frac{1}{8}$. Clearly, our new result indicates that OMP has a better (less restrictive) RIP condition and also requires a smaller number of iterations.

It is worth mentioning that even though the input signal is not strictly sparse, in many cases, it can be well approximated by a sparse signal. Our result can be readily extended to this scenario.

Corollary 1 (Recovery of Non-Sparse Signals): Let $\mathbf{x}_K \in \mathcal{R}^n$ be the vector that keeps K largest elements of the input vector $\mathbf{x}$ and sets all other entries to zero. Let $\mathbf{\Phi} \in \mathcal{R}^{m\times n}$ be the measurement matrix satisfying $\delta_{\max\{18,2S+2\}K} \le \frac{1}{8}$ and $\mathbf{y} = \mathbf{\Phi x} + \mathbf{v}$ be the noisy measurements, where $\mathbf{v}$ is the noise vector. Then, gOMP produces an estimate $\hat{\mathbf{x}}$ of $\mathbf{x}$ in $\max\left\{2K, \lfloor \frac{16K}{S} \rfloor\right\}$ iterations, such that

$$\|\hat{\mathbf{x}} - \mathbf{x}\|_2 \le D\left(\frac{\|\mathbf{x}-\mathbf{x}_K\|_1}{\sqrt{K}} + \|\mathbf{v}\|_2\right), \tag{13}$$

where D is a constant depending on $\delta_{\max\{18,2S+2\}K}$.

Since Corollary 1 is a straightforward extension of Theorem 3, we omit the proof for brevity (see [14]–[16], [33], [34] for noise-free results). Note that the key idea is to partition the noisy measurements $\mathbf{y}$ of a non-sparse signal into two parts and then apply Theorem 3. The two parts consist of 1) measurements associated with dominant elements of the signal ($\mathbf{y}_1 = \mathbf{\Phi x}_K$), and 2) measurements associated with insignificant elements and the noise vector ($\mathbf{y}_2 = \mathbf{\Phi}(\mathbf{x}-\mathbf{x}_K) + \mathbf{v}$). Therefore, we have

$$\mathbf{y} = \mathbf{y}_1 + \mathbf{y}_2 = \mathbf{\Phi x}_K + \mathbf{\Phi}(\mathbf{x}-\mathbf{x}_K) + \mathbf{v}. \tag{14}$$

B. Empirical Results

We evaluate the recovery performance of the gOMP algorithms through numerical experiments. Our simulations are focused on the noisy scenario (readers are referred to [17], [20], [21] for simulation results in the noise-free scenario). In our simulations, we consider random matrices $\mathbf{\Phi}$ of size 100×200, whose entries are drawn i.i.d. from Gaussian distribution $\mathcal{N}\left(0, \frac{1}{m}\right)$. We generate K-sparse signals $\mathbf{x}$, whose components are i.i.d. and follow a Gaussian-Bernoulli distribution:

$$x_j \sim \begin{cases} 0 & \text{with probability } 1-p, \\ \mathcal{N}(0,1) & \text{with probability } p, \end{cases} \tag{15}$$

where p is the sparsity rate that represents the average fraction of non-zero components in $\mathbf{x}$. As a metric to evaluate the recovery performance in the noisy scenario, we employ the mean square error (MSE). The MSE is defined as

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \hat{x}_i)^2, \tag{16}$$

where $\hat{x}_i$ is the estimate of x_i. In our simulation, the following recovery algorithms are considered:

1) OMP and gOMP ($S = 3, 5$).[2]
2) CoSaMP: We set the maximal iteration number to 50 to avoid repeated iterations (http://www.cmc.edu/pages/faculty/DNeedell).
3) StOMP: We use the false alarm rate control (FAR) strategy, as it works better than the false discovery rate control (FDR) strategy (http://sparselab.stanford.edu/).
4) BPDN (http://cvxr.com/cvx/).
5) Generalized approximate message passing (GAMP) [32], [38], [39]: (http://gampmatlab.wikia.com/).
6) Linear MMSE estimator.

In obtaining the performance result for each simulation point of the recovery method, we perform 2,000 independent trials.

In Fig. 1, we plot the MSE performance for each recovery method as a function of the signal-to-noise ratio (SNR), where the SNR (in dB) is defined as

$$\text{SNR} = 10\log_{10}\frac{\|\mathbf{\Phi x}\|_2^2}{\|\mathbf{v}\|_2^2}. \tag{17}$$

In this case, the system model is expressed as $\mathbf{y} = \mathbf{\Phi x} + \mathbf{v}$, where $\mathbf{v}$ is the noise vector whose elements are generated from Gaussian distribution $\mathcal{N}\left(0, \frac{pn}{m}10^{-\frac{\mathbf{SNR}}{10}}\right)$.[3] The benchmark performance of the Oracle least squares (Oracle-LS) estimator (i.e., the best possible estimation having prior knowledge on the support of input signals) is plotted as well. In general, we observe that the MSE performance improves with the SNR for all methods. While GAMP has the lowest MSE when prior knowledge on the signal and noise distribution is available, it does not perform well when this prior information is incorrect.[4] For the whole SNR region under test, the MSE performance of gOMP is comparable to OMP and also outperforms CoSaMP and BPDN. An interesting point is that the actual recovery error of gOMP is much smaller than that provided in Theorem 3. For example, when $\text{SNR} = 10$ dB, $p = 0.05$, and $v_j \sim \mathcal{N}\left(0, \frac{pn}{m}10^{-\frac{\mathbf{SNR}}{10}}\right)$, we have $\mathbb{E}\|\mathbf{v}\|_2 = \left(pn10^{-\frac{\mathbf{SNR}}{10}}\right)^{1/2} = 1$. Using this together with (A.7), the upper bound for $\|\mathbf{x} - \hat{\mathbf{x}}\|_2$ in Theorem 3 is around 63. In contrast, when $\text{SNR} = 10$ dB, the ℓ_2-norm of the actual recovery error of gOMP is $\|\mathbf{x}-\hat{\mathbf{x}}\|_2 = (n\cdot\text{MSE})^{1/2} \approx 1$ (Fig. 1(a)), which is much smaller than the upper bound indicated in Theorem 3.

Fig. 2 displays the running time of each recovery method as a function of the sparsity rate p. The running time is measured using the MATLAB program on a personal computer with an Intel Core i7 processor and Microsoft Windows 7 environment. Overall, we observe that the running time of OMP, gOMP, and StOMP is smaller than that of CoSaMP, GAMP, and BPDN. In particular, the running time of BPDN is more than one order of magnitude higher than the rest of algorithms require. This is because the complexity of BPDN is a quadratic function of the number of measurements ($\mathcal{O}(m^2n^{3/2})$) [40], while that of the gOMP algorithm is $\mathcal{O}(Kmn)$ [17]. Since gOMP can chooses

[2]We set the residual tolerance to be the noise level.

[3]Since the components of $\mathbf{\Phi}$ have power $\frac{1}{m}$ and the signal $\mathbf{x}$ has sparsity rate p, $\mathbb{E}|(\mathbf{\Phi x})_i|^2 = \frac{pn}{m}$. From the definition of SNR, we have $\mathbb{E}|v_i|^2 = \mathbb{E}|(\mathbf{\Phi x})_i|^2 \cdot 10^{-\frac{\mathbf{SNR}}{10}} = \frac{pn}{m}10^{-\frac{\mathbf{SNR}}{10}}$.

[4]In order to test the mismatch scenario, we use Bernoulli distribution ($x_j \sim \mathcal{B}(1,p)$).

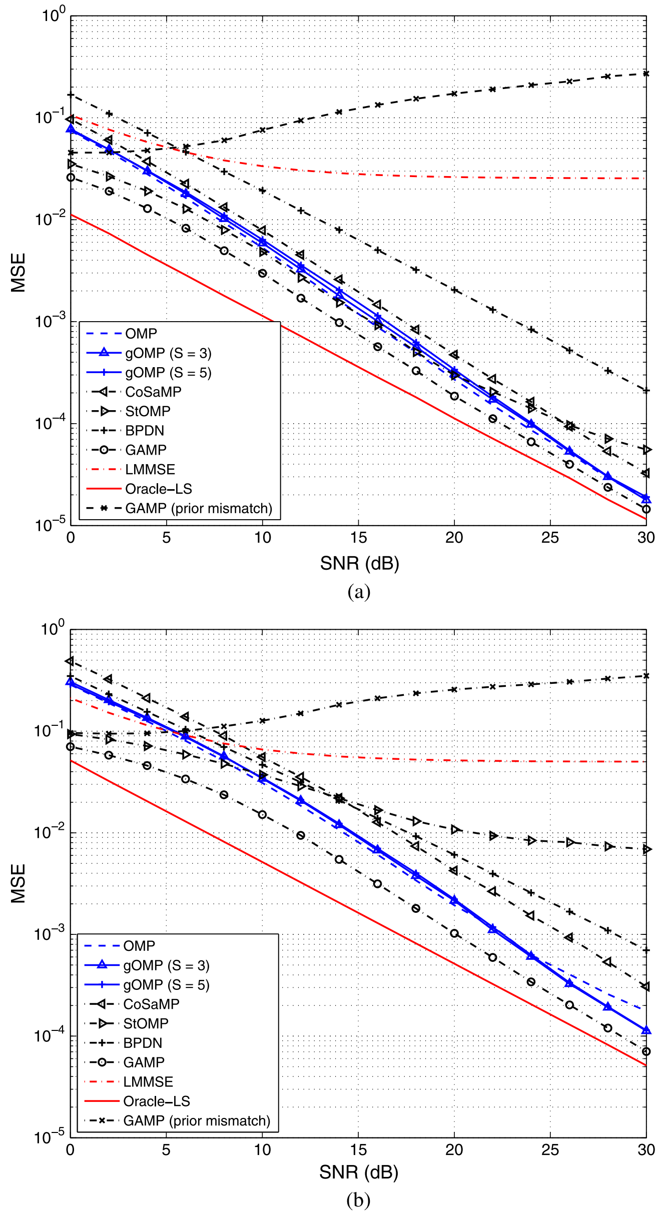


Fig. 1. MSE performance of recovery algorithms as a function of SNR. (a) $p = 0.05$. (b) $p = 0.1$.

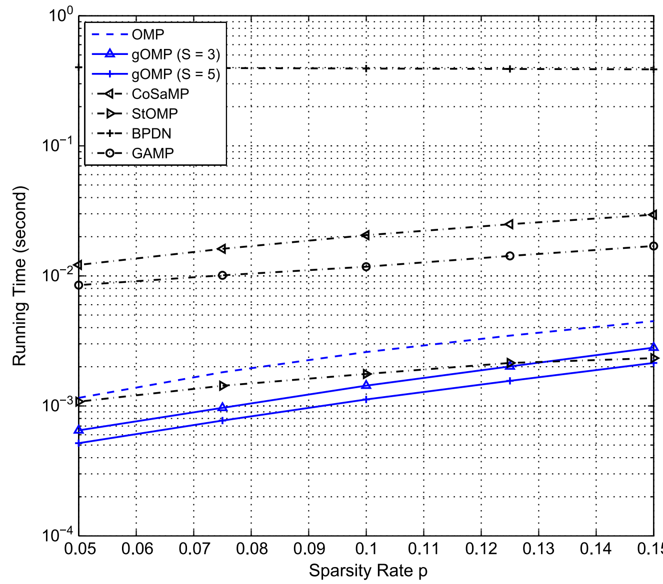


Fig. 2. Running time as a function of sparsity rate p.

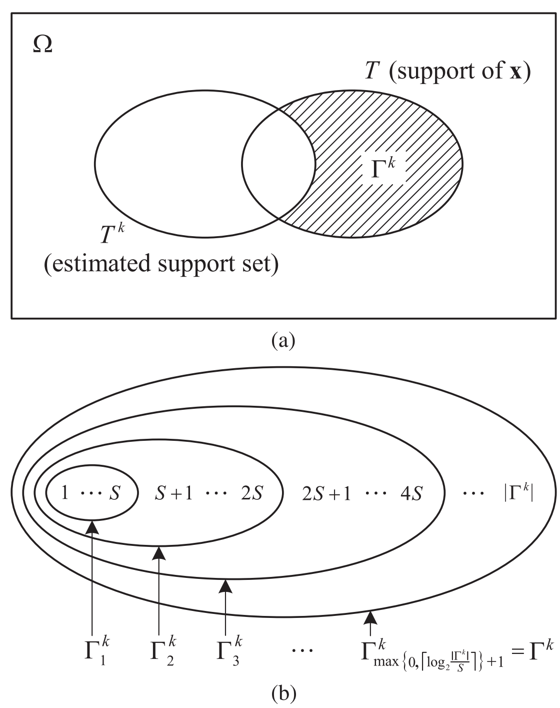


Fig. 3. Illustration of sets T, T^k, and Γ^k. (a) Set diagram of T, T^k, and Γ^k. (b) Illustration of indices in Γ^k_τ.

more than one support index at a time, we also observe that gOMP runs faster than the OMP algorithm.

III. Proof of Theorem 1

A. Preliminaries

Before we proceed to the proof of Theorem 1, we present definitions used in our analysis. Recall that $\Gamma^k = T \setminus T^k$ is the set of remaining support elements after k iterations of gOMP. In what follows, we assume without loss of generality that

$$\Gamma^k = \left\{1, \cdots, |\Gamma^k|\right\}.$$

Then it is clear that $0 \le |\Gamma^k| \le K$. For example, if $k = 0$, then $T^k = \emptyset$ and $|\Gamma^k| = |T| = K$. Whereas, if $T^k \supseteq T$, then $\Gamma^k = \emptyset$ and $|\Gamma^k| = 0$. Also, for notational convenience we assume that $\{x_i\}$ is arranged in descending order of their magnitudes, i.e., $|x_1| \ge |x_2| \ge \cdots \ge |x_{|\Gamma^k|}|$. Now, we define the subset Γ^k_τ of Γ^k as (see Fig. 3(b)):

$$\Gamma^k_\tau = \begin{cases} \emptyset & \tau = 0, \\ \{1, \cdots, 2^{\tau-1}S\} & \tau = 1, \cdots, \max\left\{0, \left\lceil \log_2 \frac{|\Gamma^k|}{S} \right\rceil\right\}, \\ \Gamma^k & \tau = \max\left\{0, \left\lceil \log_2 \frac{|\Gamma^k|}{S} \right\rceil\right\} + 1. \end{cases} \tag{18}$$

Note that the last set $\Gamma^k_{\max\left\{0, \left\lceil \log_2 \frac{|\Gamma^k|}{S} \right\rceil\right\}+1}$ $(= \Gamma^k)$ does not necessarily have $2^{\max\left\{0, \left\lceil \log_2 \frac{|\Gamma^k|}{S} \right\rceil\right\}} S$ elements.

For given set Γ^k and constant $\sigma \geq 2$, let $L \in \left\{1, 2, \cdots, \max\left\{0, \left\lceil \log_2 \frac{|\Gamma^k|}{S} \right\rceil\right\} + 1\right\}$ be a positive integer satisfying[5]

$$\left\|\mathbf{x}_{\Gamma^k \setminus \Gamma_0^k}\right\|_2^2 < \sigma \left\|\mathbf{x}_{\Gamma^k \setminus \Gamma_1^k}\right\|_2^2, \tag{19a}$$

$$\left\|\mathbf{x}_{\Gamma^k \setminus \Gamma_1^k}\right\|_2^2 < \sigma \left\|\mathbf{x}_{\Gamma^k \setminus \Gamma_2^k}\right\|_2^2, \tag{19b}$$

$$\vdots$$

$$\left\|\mathbf{x}_{\Gamma^k \setminus \Gamma_{L-2}^k}\right\|_2^2 < \sigma \left\|\mathbf{x}_{\Gamma^k \setminus \Gamma_{L-1}^k}\right\|_2^2, \tag{19c}$$

$$\left\|\mathbf{x}_{\Gamma^k \setminus \Gamma_{L-1}^k}\right\|_2^2 \geq \sigma \left\|\mathbf{x}_{\Gamma^k \setminus \Gamma_L^k}\right\|_2^2. \tag{19d}$$

If (19d) holds true for all $L \geq 1$, then we ignore (19a)–(19c) and simply take $L = 1$. Note that L always exists because $\|\mathbf{x}^k_{\Gamma^k \setminus \Gamma_{\max\{0, \lceil \log_2 |\Gamma^k|/S \rceil\}+1}}\|_2^2 = 0$ so that (19d) holds true at least for $L = \max\{0, \lceil \log_2 |\Gamma^k|/S \rceil\} + 1$. From (19a)–(19d), we have

$$\left\|\mathbf{x}_{\Gamma^k \setminus \Gamma_\tau^k}\right\|_2^2 \leq \sigma^{L-1-\tau} \left\|\mathbf{x}_{\Gamma^k \setminus \Gamma_{L-1}^k}\right\|_2^2, \ \tau = 0, 1, \cdots, L. \tag{20}$$

Moreover, if $L \geq 2$, we have a lower bound for $|\Gamma^k|$ as (see Appendix B):

$$|\Gamma^k| > \left(\frac{2\sigma - 1}{2\sigma - 2}\right) 2^{L-2} S. \tag{21}$$

Equations (20) and (21) will be used in the proof of Theorem 1 and we will fix $\sigma = \frac{1}{2} \exp\left(\frac{14}{9}\right)$ in the proof.

We now provide two propositions useful in the proof of Theorem 1. The first one offers an upper bound for $\|\mathbf{r}^k\|_2^2$ and a lower bound for $\|\mathbf{r}^l\|_2^2 - \|\mathbf{r}^{l+1}\|_2^2$ $(l \geq k)$.

Proposition 1: For given Γ^k and any integer $l \geq k$, the residual of gOMP satisfies

$$\|\mathbf{r}^k\|_2^2 \leq \left\|\mathbf{\Phi}_\Gamma^k \mathbf{x}_\Gamma^k + \mathbf{v}\right\|_2^2, \tag{22}$$

$$\begin{aligned}&\|\mathbf{r}^l\|_2^2 - \|\mathbf{r}^{l+1}\|_2^2 \\ &\geq \frac{1 - \delta_{|\Gamma_\tau^k \cup T^l|}}{(1 + \delta_S)\left\lceil \frac{|\Gamma_\tau^k|}{S} \right\rceil} \left(\|\mathbf{r}^l\|_2^2 - \left\|\mathbf{\Phi}_{\Gamma^k \setminus \Gamma_\tau^k} \mathbf{x}_{\Gamma^k \setminus \Gamma_\tau^k} + \mathbf{v}\right\|_2^2\right),\end{aligned} \tag{23}$$

where $\tau = 1, 2, \cdots, \max\left\{0, \left\lceil \log_2 \frac{|\Gamma^k|}{S} \right\rceil\right\} + 1$.

Proof: See Appendix C. □

The second proposition is essentially an extension of (23). It characterizes the relationship between residuals of gOMP in different number of iterations.

Proposition 2: For any integer $l \geq k$, $\Delta l > 0$, and $\tau \in \left\{1, \cdots, \max\left\{0, \left\lceil \log_2 \frac{|\Gamma^k|}{S} \right\rceil\right\} + 1\right\}$, the residual $\mathbf{r}^{l+\Delta l}$ of gOMP satisfies

$$\begin{aligned}&\|\mathbf{r}^{l+\Delta l}\|_2^2 - \left\|\mathbf{\Phi}_{\Gamma^k \setminus \Gamma_\tau^k} \mathbf{x}_{\Gamma^k \setminus \Gamma_\tau^k} + \mathbf{v}\right\|_2^2 \\ &\quad \leq C_{\tau, l, \Delta l}\left(\|\mathbf{r}^l\|_2^2 - \left\|\mathbf{\Phi}_{\Gamma^k \setminus \Gamma_\tau^k} \mathbf{x}_{\Gamma^k \setminus \Gamma_\tau^k} + \mathbf{v}\right\|_2^2\right),\end{aligned} \tag{24}$$

where

$$C_{\tau, l, \Delta l} = \exp\left(-\frac{\Delta l \left(1 - \delta_{|\Gamma_\tau^k \cup T^{l+\Delta l-1}|}\right)}{\left\lceil \frac{|\Gamma_\tau^k|}{S} \right\rceil (1 + \delta_S)}\right). \tag{25}$$

Proof: See Appendix E. □

[5]We note that L is a function of k.

B. Outline of Proof

The proof of Theorem 1 is based on mathematical induction in $|\Gamma^k|$, the number of remaining indices after k iterations of gOMP. We first consider the case when $|\Gamma^k| = 0$. This case is trivial since all support indices are already selected $(T \subseteq T^k)$ and hence

$$\begin{aligned}\|\mathbf{r}^k\|_2 &= \|\mathbf{y} - \mathbf{\Phi}\hat{\mathbf{x}}^k\|_2 \\ &= \min_{supp(\mathbf{u}) = T^k} \|\mathbf{y} - \mathbf{\Phi}\mathbf{u}\|_2 \\ &\leq \|\mathbf{y} - \mathbf{\Phi}\mathbf{x}\|_2 \\ &= \|\mathbf{v}\|_2 \\ &\leq \mu_k \|\mathbf{v}\|_2.\end{aligned} \tag{26}$$

Next, we assume that the argument holds up to an integer $\gamma - 1$. Under this inductive assumption, we will prove that it also holds true for $|\Gamma^k| = \gamma$. In other words, we will show that when $|\Gamma^k| = \gamma$,

$$\left\|\mathbf{r}^{k + \max\left\{\gamma, \left\lfloor \frac{8\gamma}{S} \right\rfloor\right\}}\right\|_2 \leq \mu_k \|\mathbf{v}\|_2 \tag{27}$$

holds true under

$$\delta_{\max\{Sk + 7\gamma, Sk + S + \gamma\}} \leq \frac{1}{8}. \tag{28}$$

Although the details of the proof in the induction step are somewhat cumbersome, the main idea is rather simple. First, we show that a decent amount of support indices in Γ^k can be selected within a specified number of additional iterations so that the number of remaining support indices is upper bounded. More precisely,

i) If $L = 1$, the number of remaining support indices after $(k+1)$ iterations is upper bounded as
$$|\Gamma^{k+1}| < \gamma. \tag{29}$$

ii) If $L \geq 2$, the number of remaining support indices after $(k + k_L)$ iterations satisfies
$$|\Gamma^{k + k_L}| < \left|\Gamma^k \setminus \Gamma_{L-1}^k\right|, \tag{30}$$
where
$$k_i = 2\sum_{\tau=0}^{i} \left\lceil \frac{|\Gamma_\tau^k|}{S} \right\rceil, \ i = 0, \cdots, L. \tag{31}$$

Second, since (29) and (30) imply that the number of remaining support indices is no more than $\gamma - 1$, from the induction hypothesis we have

$$\left\|\mathbf{r}^{k+1+\max\left\{\gamma, \left\lfloor \frac{8}{S}|\Gamma^{k+1}| \right\rfloor\right\}}\right\|_2 \leq \mu_k \|\mathbf{v}\|_2, \ L = 1, \tag{32}$$

$$\left\|\mathbf{r}^{k+k_L+\max\left\{\gamma, \left\lfloor \frac{8}{S}|\Gamma^{k+k_L}| \right\rfloor\right\}}\right\|_2 \leq \mu_k \|\mathbf{v}\|_2, \ L \geq 2. \tag{33}$$

Further, by using the upper bound of $k + 1 + \max\left\{\gamma, \left\lfloor \frac{8}{S}|\Gamma^{k+1}| \right\rfloor\right\}$ in (32) and $k + k_L + \max\left\{\gamma, \left\lfloor \frac{8}{S}|\Gamma^{k+k_L}| \right\rfloor\right\}$ in (33), we establish the induction step. Specifically,

i) $L = 1$ case: We obtain from (29) that
$$\begin{aligned}&k + 1 + \max\left\{\gamma, \left\lfloor \frac{8}{S}|\Gamma^{k+1}| \right\rfloor\right\} \\ &\quad \leq k + 1 + \max\left\{\gamma, \left\lfloor \frac{8}{S}(\gamma - 1) \right\rfloor\right\} \\ &\quad \leq k + \max\left\{\gamma, \left\lfloor \frac{8\gamma}{S} \right\rfloor\right\}.\end{aligned} \tag{34}$$

By noting that the residual power of gOMP is non-increasing ($\|\mathbf{r}^i\|_2 \le \|\mathbf{r}^j\|_2$ for $i \ge j$), we have

$$\left\|\mathbf{r}^{k+\max\left\{\gamma, \left\lfloor\frac{8\gamma}{S}\right\rfloor\right\}}\right\|_2 \le \left\|\mathbf{r}^{k+1+\max\left\{\gamma, \left\lfloor\frac{8}{S}|\Gamma^{k+1}|\right\rfloor\right\}}\right\|_2 \le \mu_k\|\mathbf{v}\|_2. \quad (35)$$

ii) $L \ge 2$ case: We observe from (31) that

$$\begin{aligned} k_L &= 2\sum_{\tau=0}^{L}\left\lceil\frac{|\Gamma_\tau^k|}{S}\right\rceil \\ &= 2\sum_{\tau=1}^{L}\left\lceil\frac{|\Gamma_\tau^k|}{S}\right\rceil \\ &\le 2\sum_{\tau=1}^{L}2^{\tau-1} \\ &= 2(2^L-1), \end{aligned} \quad (36)$$

which together with (30) implies that

$$\begin{aligned} & k+k_L+\max\left\{\gamma, \left\lfloor\frac{8}{S}|\Gamma^{k+k_L}|\right\rfloor\right\} \\ &\le k+2(2^L-1)+\max\left\{\gamma, \left\lfloor\frac{8}{S}\left|\Gamma^k\setminus\Gamma_{L-1}^k\right|\right\rfloor\right\} \\ &= k+2(2^L-1)+\max\left\{\gamma, \left\lfloor\frac{8}{S}(\gamma-2^{L-2}S)\right\rfloor\right\} \\ &\le k+\max\left\{\gamma, \left\lfloor\frac{8\gamma}{S}\right\rfloor\right\}, \end{aligned} \quad (37)$$

Hence, we obtain from (33) and (37) that

$$\left\|\mathbf{r}^{k+\max\left\{\gamma, \left\lfloor\frac{8\gamma}{S}\right\rfloor\right\}}\right\|_2 \le \left\|\mathbf{r}^{k+k_L+\max\left\{\gamma, \left\lfloor\frac{8}{S}|\Gamma^{k+k_L}|\right\rfloor\right\}}\right\|_2 \le \mu_k\|\mathbf{v}\|_2. \quad (38)$$

In summary, what remains now is the proofs of (29) and (30).

C. Proof of (30)

We consider the proof of (30) for the case of $L \ge 2$. Instead of directly proving (30), we show that a sufficient condition for (30) is true. To be specific, since $\mathbf{x}_{\Gamma^k\setminus\Gamma_{L-1}^k}$ consists of $|\Gamma^k\setminus\Gamma_{L-1}^k|$ smallest non-zero elements (in magnitude) in $\mathbf{x}_{\Gamma^k}$, a sufficient condition for (30) is

$$\|\mathbf{x}_{\Gamma^{k+k_L}}\|_2^2 < \left\|\mathbf{x}_{\Gamma^k\setminus\Gamma_{L-1}^k}\right\|_2^2. \quad (39)$$

In this subsection, we show that (39) is true under

$$\delta_{Sk+7\gamma} \le \frac{1}{8}. \quad (40)$$

To this end, we first construct lower and upper bounds for $\|\mathbf{r}^{k+k_L}\|_2$ and then use these bounds to derive a condition guaranteeing (39).

1) **Lower bound for $\|\mathbf{r}^{k+k_L}\|_2^2$:**

$$\begin{aligned} \|\mathbf{r}^{k+k_L}\|_2 &= \|\mathbf{y}-\mathbf{\Phi}\hat{\mathbf{x}}^{k+k_L}\|_2 \\ &= \left\|\mathbf{\Phi}(\mathbf{x}-\hat{\mathbf{x}}^{k+k_L})+\mathbf{v}\right\|_2 \\ &\ge \left\|\mathbf{\Phi}(\mathbf{x}-\hat{\mathbf{x}}^{k+k_L})\right\|_2-\|\mathbf{v}\|_2 \\ &\overset{(a)}{\ge} \left(1-\delta_{|T\cup T^{k+k_L}|}\right)^{\frac{1}{2}}\|\mathbf{x}-\hat{\mathbf{x}}^{k+k_L}\|_2-\|\mathbf{v}\|_2 \\ &\ge \left(1-\delta_{|T\cup T^{k+k_L}|}\right)^{\frac{1}{2}}\|\mathbf{x}_{\Gamma^{k+k_L}}\|_2-\|\mathbf{v}\|_2, \end{aligned} \quad (41)$$

where (a) is from the RIP (note that $\mathbf{x}-\hat{\mathbf{x}}^{k+k_L}$ is supported on $T\cup T^{k+k_L}$).

2) **Upper bound for $\|\mathbf{r}^{k+k_L}\|_2^2$:**
First, by applying Proposition 2, we have

$$\begin{aligned} &\|\mathbf{r}^{k+k_1}\|_2^2-\|\mathbf{\Phi}_{\Gamma^k\setminus\Gamma_1^k}\mathbf{x}_{\Gamma^k\setminus\Gamma_1^k}+\mathbf{v}\|_2^2 \\ &\le C_{1,k,k_1}\times\left(\|\mathbf{r}^k\|_2^2-\|\mathbf{\Phi}_{\Gamma^k\setminus\Gamma_1^k}\mathbf{x}_{\Gamma^k\setminus\Gamma_1^k}+\mathbf{v}\|_2^2\right), \end{aligned} \quad (42)$$

$$\begin{aligned} &\|\mathbf{r}^{k+k_2}\|_2^2-\|\mathbf{\Phi}_{\Gamma^k\setminus\Gamma_2^k}\mathbf{x}_{\Gamma^k\setminus\Gamma_2^k}+\mathbf{v}\|_2^2 \\ &\le C_{2,k+k_1,k_2-k_1}\times\left(\|\mathbf{r}^{k+k_1}\|_2^2-\|\mathbf{\Phi}_{\Gamma^k\setminus\Gamma_2^k}\mathbf{x}_{\Gamma^k\setminus\Gamma_2^k}+\mathbf{v}\|_2^2\right), \end{aligned} \quad (43)$$

$$\vdots$$

$$\begin{aligned} &\|\mathbf{r}^{k+k_L}\|_2^2-\|\mathbf{\Phi}_{\Gamma^k\setminus\Gamma_L^k}\mathbf{x}_{\Gamma^k\setminus\Gamma_L^k}+\mathbf{v}\|_2^2 \\ &\le C_{L,k+k_{L-1},k_L-k_{L-1}} \\ &\quad\times\left(\|\mathbf{r}^{k+k_{L-1}}\|_2^2-\|\mathbf{\Phi}_{\Gamma^k\setminus\Gamma_L^k}\mathbf{x}_{\Gamma^k\setminus\Gamma_L^k}+\mathbf{v}\|_2^2\right). \end{aligned} \quad (44)$$

From (31) and monotonicity of the RIC, we have

$$\begin{aligned} C_{i,k+k_{i-1},k_i-k_{i-1}} &= \exp\left(-2\cdot\frac{1-\delta_{|\Gamma_i^k\cup T^{k+k_i-1}|}}{1+\delta_S}\right) \\ &\overset{(a)}{\le} \exp\left(-2\cdot\frac{1-\delta_{Sk+7\gamma}}{1+\delta_{Sk+7\gamma}}\right) \\ &\overset{(b)}{\le} \exp\left(-\frac{14}{9}\right), \end{aligned} \quad (45)$$

for $i=1,2,\cdots,L$, where (a) is due to monotonicity of the RIC and (b) is from (28). Notice that (a) is because

$$\begin{aligned} \left|\Gamma_i^k\cup T^{k+k_i-1}\right| &\le |T\cup T^{k+k_L}| \\ &= |T^{k+k_L}|+|\Gamma^{k+k_L}| \\ &\le S(k+k_L)+|\Gamma^k| \\ &\overset{(c)}{\le} Sk+2(2^L-1)S+\gamma \\ &\overset{(d)}{<} Sk+8\left(\frac{2\sigma-2}{2\sigma-1}\right)\gamma+\gamma-2S \\ &\overset{(e)}{<} Sk+7\gamma, \end{aligned} \quad (46)$$

where (c) follows from (36), (d) is from (21), and (e) is due to $\sigma=\frac{1}{2}\exp\left(\frac{14}{9}\right)$.

For notational simplicity, we let $\eta=\exp\left(-\frac{14}{9}\right)$. Then (42)–(44) can be rewritten as

$$\begin{aligned} \|\mathbf{r}^{k+k_1}\|_2^2 &\le \eta\|\mathbf{r}^k\|_2^2+(1-\eta)\|\mathbf{\Phi}_{\Gamma^k\setminus\Gamma_1^k}\mathbf{x}_{\Gamma^k\setminus\Gamma_1^k}+\mathbf{v}\|_2^2, \\ \|\mathbf{r}^{k+k_2}\|_2^2 &\le \eta\|\mathbf{r}^{k+k_1}\|_2^2+(1-\eta)\|\mathbf{\Phi}_{\Gamma^k\setminus\Gamma_2^k}\mathbf{x}_{\Gamma^k\setminus\Gamma_2^k}+\mathbf{v}\|_2^2, \\ &\vdots \\ \|\mathbf{r}^{k+k_L}\|_2^2 &\le \eta\|\mathbf{r}^{k+k_{L-1}}\|_2^2+(1-\eta)\|\mathbf{\Phi}_{\Gamma^k\setminus\Gamma_L^k}\mathbf{x}_{\Gamma^k\setminus\Gamma_L^k}+\mathbf{v}\|_2^2. \end{aligned}$$

Further, some additional manipulations yield the following result.

$$
\begin{aligned}
&\|\mathbf{r}^{k+k_L}\|_2^2 \\
&\le \eta^L\|\mathbf{r}^k\|_2^2 + (1-\eta)\sum_{\tau=1}^{L}\eta^{L-\tau}\|\mathbf{\Phi}_{\Gamma^k\setminus\Gamma^k_\tau}\mathbf{x}_{\Gamma^k\setminus\Gamma^k_\tau}+\mathbf{v}\|_2^2 \\
&\overset{(a)}{\le} \eta^L\left\|\mathbf{\Phi}^k_\Gamma\mathbf{x}^k_\Gamma+\mathbf{v}\right\|_2^2 \\
&\quad + (1-\eta)\sum_{\tau=1}^{L}\eta^{L-\tau}\|\mathbf{\Phi}_{\Gamma^k\setminus\Gamma^k_\tau}\mathbf{x}_{\Gamma^k\setminus\Gamma^k_\tau}+\mathbf{v}\|_2^2 \\
&\overset{(b)}{\le} \eta^L\left((1+t)\left\|\mathbf{\Phi}^k_\Gamma\mathbf{x}^k_\Gamma\right\|_2^2+(1+t^{-1})\|\mathbf{v}\|_2^2\right)+(1-\eta) \\
&\quad \times\sum_{\tau=1}^{L}\eta^{L-\tau}\left((1+t)\|\mathbf{\Phi}_{\Gamma^k\setminus\Gamma^k_\tau}\mathbf{x}_{\Gamma^k\setminus\Gamma^k_\tau}\|_2^2+(1+t^{-1})\|\mathbf{v}\|_2^2\right), \\
&\overset{(c)}{\le} \left(\eta^L\|\mathbf{x}_{\Gamma^k\setminus\Gamma^k_0}\|_2^2+(1-\eta)\sum_{\tau=1}^{L}\eta^{L-\tau}\|\mathbf{x}_{\Gamma^k\setminus\Gamma^k_\tau}\|_2^2\right)(1+t) \\
&\quad \times(1+\delta_\gamma)+(1+t^{-1})\left(\eta^L+(1-\eta)\sum_{\tau=1}^{L}\eta^{L-\tau}\right)\|\mathbf{v}\|_2^2,
\end{aligned} \tag{47}
$$

where (a) is from Proposition 1, (b) uses the fact that

$$
\|\mathbf{u}+\mathbf{v}\|_2^2 \le (1+t)\|\mathbf{u}\|_2^2+(1+t^{-1})\|\mathbf{v}\|_2^2 \tag{48}
$$

for $t > 0$ (we will specify t later), and (c) is due to the RIP. (Note that $|\Gamma^k\setminus\Gamma^k_\tau| \le |\Gamma^k| = \gamma$ for $\tau = 1,\cdots,L$.)

By applying (20) to (47), we further have

$$
\begin{aligned}
&\|\mathbf{r}^{k+k_L}\|_2^2 \\
&\le \left(\sigma^{L-1}\eta^L+(1-\eta)\sum_{\tau=1}^{L}\sigma^{L-1-\tau}\eta^{L-\tau}\right)(1+t)(1+\delta_\gamma) \\
&\quad \times\|\mathbf{x}_{\Gamma^k\setminus\Gamma^k_{L-1}}\|_2^2+(1+t^{-1})\left(\eta^L+(1-\eta)\sum_{\tau=1}^{L}\eta^{L-\tau}\right)\|\mathbf{v}\|_2^2 \\
&= \left((\sigma\eta)^L+(1-\eta)\sum_{\tau=0}^{L-1}(\sigma\eta)^\tau\right)\sigma^{-1}(1+\delta_\gamma)\|\mathbf{x}_{\Gamma^k\setminus\Gamma^k_{L-1}}\|_2^2 \\
&\quad \times(1+t)+(1+t^{-1})\left(\eta^L+(1-\eta)\sum_{\tau=0}^{L-1}\eta^\tau\right)\|\mathbf{v}\|_2^2 \\
&\overset{(a)}{<} \left(\sum_{\tau=L}^{\infty}(\sigma\eta)^\tau+\sum_{\tau=0}^{L-1}(\sigma\eta)^\tau\right)\sigma^{-1}(1-\eta)(1+\delta_\gamma)\|\mathbf{x}_{\Gamma^k\setminus\Gamma^k_{L-1}}\|_2^2 \\
&\quad \times(1+t)+(1+t^{-1})(1-\eta)\left(\sum_{\tau=L}^{\infty}\eta^\tau+\sum_{\tau=0}^{L-1}\eta^\tau\right)\|\mathbf{v}\|_2^2 \\
&\overset{(b)}{=} 4\eta(1-\eta)(1+\delta_\gamma)(1-t)\|\mathbf{x}_{\Gamma^k\setminus\Gamma^k_{L-1}}\|_2^2+(1+t^{-1})\|\mathbf{v}\|_2^2,
\end{aligned} \tag{49}
$$

where (a) is because $\sigma \ge 2$, $\sigma\eta < 1$, and $\eta < 1$. Hence

$$
(\sigma\eta)^L < \left(\frac{1-\eta}{1-\sigma\eta}\right)(\sigma\eta)^L = (1-\eta)\sum_{\tau=L}^{\infty}(\sigma\eta)^\tau,
$$

$$
\eta^L = (1-\eta)\left(\frac{\eta^L}{1-\eta}\right) = (1-\eta)\sum_{\tau=L}^{\infty}\eta^\tau,
$$

and (b) uses the fact that $\sigma\eta = \frac{1}{2}$.

Thus far, we have obtained a lower bound for $\|\mathbf{r}^{k+k_L}\|_2$ in (41) and an upper bound for $\|\mathbf{r}^{k+k_L}\|_2$ in (49), respectively. Next, we will use these bounds to prove that (39) holds true under $\delta_{Sk+7\gamma} \le \frac{1}{8}$.

By relating (41) and (49), we have

$$
\|\mathbf{x}_{\Gamma^{k+k_L}}\|_2 \le \alpha\|\mathbf{x}_{\Gamma^k\setminus\Gamma^k_{L-1}}\|_2+\beta\|\mathbf{v}\|_2 \tag{50}
$$

where

$$
\alpha = 2\left(\frac{\eta(1-\eta)(1+\delta_\gamma)(1+t)}{1-\delta_{|T\cup T^{k+k_L}|}}\right)^{\frac{1}{2}} \tag{51}
$$

aan

$$
\beta = \left((1+t^{-1})^{\frac{1}{2}}+1\right)\left(1-\delta_{|T\cup T^{k+k_L}|}\right)^{-1/2}. \tag{52}
$$

Since $\delta_{|T\cup T^{k+k_L}|} \le \delta_{Sk+7\gamma}$ by monotonicity of the RIC,

$$
\alpha \le 2\left(\frac{(1+\delta_{Sk+7\gamma})(1+t)\left(1-\exp\left(-\frac{14}{9}\right)\right)}{(1-\delta_{Sk+7\gamma})\exp\left(\frac{14}{9}\right)}\right)^{\frac{1}{2}}. \tag{53}
$$

By choosing $t = \frac{1}{6}$ in (53), we have

$$
\alpha < 1 \tag{54}
$$

under $\delta_{Sk+7\gamma} \le \frac{1}{8}$.

Now, we consider two cases: 1) $\beta\|\mathbf{v}\|_2 < (1-\alpha)\|\mathbf{x}_{\Gamma^k\setminus\Gamma^k_{L-1}}\|_2$ and 2) $\beta\|\mathbf{v}\|_2 \ge (1-\alpha)\|\mathbf{x}_{\Gamma^k\setminus\Gamma^k_{L-1}}\|_2$. First, if $\beta\|\mathbf{v}\|_2 < (1-\alpha)\|\mathbf{x}_{\Gamma^k\setminus\Gamma^k_{L-1}}\|_2$, (50) implies (39) (i.e., $\|\mathbf{x}_{\Gamma^{k+k_L}}\|_2^2 < \|\mathbf{x}_{\Gamma^k\setminus\Gamma^k_{L-1}}\|_2^2$) so that (30) holds true.

Second, if $\beta\|\mathbf{v}\|_2 \ge (1-\alpha)\|\mathbf{x}_{\Gamma^k\setminus\Gamma^k_{L-1}}\|_2$, then (27) directly holds true because

$$
\begin{aligned}
&\left\|\mathbf{r}^{k+\max\{\gamma,\lfloor\frac{8\gamma}{S}\rfloor\}}\right\|_2 \\
&\overset{(a)}{\le} \|\mathbf{r}^{k+k_L}\|_2 \\
&\overset{(b)}{\le} 2\left(\eta(1-\eta)(1+\delta_\gamma)(1+t)\right)^{\frac{1}{2}}\|\mathbf{x}_{\Gamma^k\setminus\Gamma^k_{L-1}}\|_2 \\
&\quad +(1+t^{-1})^{\frac{1}{2}}\|\mathbf{v}\|_2 \\
&\overset{(c)}{=} \alpha\left(1-\delta_{|T\cup T^{k+k_L}|}\right)^{\frac{1}{2}}\|\mathbf{x}_{\Gamma^k\setminus\Gamma^k_{L-1}}\|_2+(1+t^{-1})^{\frac{1}{2}}\|\mathbf{v}\|_2 \\
&\le \left(\frac{\alpha\beta\left(1-\delta_{|T\cup T^{k+k_L}|}\right)^{\frac{1}{2}}}{1-\alpha}+(1+t^{-1})^{\frac{1}{2}}\right)\|\mathbf{v}\|_2 \\
&= \left(\frac{(1+t^{-1})^{\frac{1}{2}}+1}{1-\alpha}-1\right)\|\mathbf{v}\|_2 \\
&\le \mu_k\|\mathbf{v}\|_2,
\end{aligned} \tag{55}
$$

where

$$
\mu_k = \left(1-2\left(\frac{7(1+\delta)\left(1-\exp\left(-\frac{14}{9}\right)\right)}{6(1-\delta)\exp\left(\frac{14}{9}\right)}\right)^{\frac{1}{2}}\right)^{-1}(\sqrt{7}+1)-1 \tag{56}
$$

where $\delta = \delta_{\max\{Sk+7\gamma, Sk+S+\gamma\}}$, (a) is from (37) and the fact that the residual power of gOMP is always non-increasing, (b) is due to (49), and (c) is from (51).

D. Proof of (29)

The proof of (29) is similar to the proof of (30). Instead of directly proving (29), we will show that a sufficient condition for (29) is true. More precisely, we will prove that

$$\|\mathbf{x}_{\Gamma^{k+1}}\|_2^2 < \left\|\mathbf{x}_{\Gamma}^k\right\|_2^2 \tag{57}$$

holds true under

$$\delta_{S(k+2)+\gamma} \leq \frac{1}{8}. \tag{58}$$

We first construct lower and upper bounds for $\|\mathbf{r}^{k+1}\|_2$ and then use these bounds to derive a condition guaranteeing (57).

1) **Lower bound for $\|\mathbf{r}^{k+1}\|_2^2$:**

$$\begin{aligned}
&\|\mathbf{r}^{k+1}\|_2 \\
&= \|\mathbf{y} - \mathbf{\Phi}\hat{\mathbf{x}}^{k+1}\|_2 \\
&= \left\|\mathbf{\Phi}(\mathbf{x} - \hat{\mathbf{x}}^{k+1}) + \mathbf{v}\right\|_2 \\
&\geq \left\|\mathbf{\Phi}(\mathbf{x} - \hat{\mathbf{x}}^{k+1})\right\|_2 - \|\mathbf{v}\|_2 \\
&\overset{(a)}{\geq} \left(1 - \delta_{|T\cup T^{k+1}|}\right)^{\frac{1}{2}} \|\mathbf{x} - \hat{\mathbf{x}}^{k+1}\|_2 - \|\mathbf{v}\|_2 \\
&\geq \left(1 - \delta_{|T\cup T^{k+1}|}\right)^{\frac{1}{2}} \|\mathbf{x}_{\Gamma^{k+1}}\|_2 - \|\mathbf{v}\|_2,
\end{aligned} \tag{59}$$

where (a) is because $\mathbf{x} - \hat{\mathbf{x}}^{k+1}$ is supported on $T \cup T^{k+1}$.

2) **Upper bound for $\|\mathbf{r}^{k+1}\|_2^2$:**
By applying Proposition 1 with $l = k$ and $\tau = 1$, we have

$$\begin{aligned}
&\left\|\mathbf{r}^k\right\|_2^2 - \left\|\mathbf{r}^{k+1}\right\|_2^2 \\
&\geq \frac{1 - \delta_{|\Gamma_1^k \cup T^k|}}{(1+\delta_S)\left\lceil \frac{|\Gamma_1^k|}{S} \right\rceil} \left(\left\|\mathbf{r}^k\right\|_2^2 - \|\mathbf{\Phi}_{\Gamma^k \setminus \Gamma_1^k}\mathbf{x}_{\Gamma^k \setminus \Gamma_1^k} + \mathbf{v}\|_2^2\right) \\
&\overset{(a)}{=} \frac{1 - \delta_{|\Gamma_1^k \cup T^k|}}{1+\delta_S} \left(\left\|\mathbf{r}^k\right\|_2^2 - \|\mathbf{\Phi}_{\Gamma^k \setminus \Gamma_1^k}\mathbf{x}_{\Gamma^k \setminus \Gamma_1^k} + \mathbf{v}\|_2^2\right),
\end{aligned} \tag{60}$$

where (a) is because $|\Gamma_1^k| \leq S$ (see (18)) and hence $\left\lceil \frac{|\Gamma_1^k|}{S} \right\rceil = 1$. Rearranging the terms yields

$$\begin{aligned}
\|\mathbf{r}^{k+1}\|_2^2 \leq &\left(1 - \frac{1 - \delta_{|\Gamma_1^k \cup T^k|}}{1+\delta_S}\right) \|\mathbf{r}^k\|_2^2 \\
&+ \frac{1 - \delta_{|\Gamma_1^k \cup T^k|}}{1+\delta_S} \left\|\mathbf{\Phi}_{\Gamma^k \setminus \Gamma_1^k}\mathbf{x}_{\Gamma^k \setminus \Gamma_1^k} + \mathbf{v}\right\|_2^2.
\end{aligned} \tag{61}$$

From Proposition 1,

$$\begin{aligned}
\|\mathbf{r}^k\|_2^2 &\leq \left\|\mathbf{\Phi}_{\Gamma}^k\mathbf{x}_{\Gamma}^k + \mathbf{v}\right\|_2^2 \\
&\overset{(a)}{\leq} (1+t)\left\|\mathbf{\Phi}_{\Gamma}^k\mathbf{x}_{\Gamma}^k\right\|_2^2 + (1+t^{-1})\|\mathbf{v}\|_2^2 \\
&\overset{(b)}{\leq} (1+t)(1+\delta_\gamma)\|\mathbf{x}_{\Gamma^k}\|_2^2 + (1+t^{-1})\|\mathbf{v}\|_2^2,
\end{aligned} \tag{62}$$

where (a) is from (48) and (b) is due to the RIP. Moreover,

$$\begin{aligned}
&\|\mathbf{\Phi}_{\Gamma^k \setminus \Gamma_1^k}\mathbf{x}_{\Gamma^k \setminus \Gamma_1^k} + \mathbf{v}\|_2^2 \\
&\overset{(a)}{\leq} (1+t)\|\mathbf{\Phi}_{\Gamma^k \setminus \Gamma_1^k}\mathbf{x}_{\Gamma^k \setminus \Gamma_1^k}\|_2^2 + (1+t^{-1})\|\mathbf{v}\|_2^2 \\
&\overset{(b)}{\leq} (1+t)(1+\delta_\gamma)\|\mathbf{x}_{\Gamma^k \setminus \Gamma_1^k}\|_2^2 + (1+t^{-1})\|\mathbf{v}\|_2^2 \\
&\overset{(b)}{\leq} (1+t)(1+\delta_\gamma)\sigma^{-1}\|\mathbf{x}_{\Gamma^k}\|_2^2 + (1+t^{-1})\|\mathbf{v}\|_2^2,
\end{aligned} \tag{63}$$

where (a) is from (48), (b) is due to the RIP, and (c) is from (19d).

Using (61), (62), and (63), we have

$$\begin{aligned}
&\|\mathbf{r}^{k+1}\|_2^2 \leq (1+t)(1+\delta_\gamma) \\
&\times\left(1 - \frac{(1-\sigma^{-1})\left(1 - \delta_{|\Gamma_1^k \cup T^k|}\right)}{1+\delta_S}\right)\|\mathbf{x}_{\Gamma^k}\|_2^2 + (1+t^{-1})\|\mathbf{v}\|_2^2,
\end{aligned}$$

from which we obtain an upper bound for $\|\mathbf{r}^{k+1}\|_2$ as

$$\begin{aligned}
\|\mathbf{r}^{k+1}\|_2 \leq &(1+t)^{\frac{1}{2}}\left(1 - \frac{(1-\sigma^{-1})\left(1 - \delta_{|\Gamma_1^k \cup T^k|}\right)}{1+\delta_S}\right)^{\frac{1}{2}} \\
&\times(1+\delta_\gamma)^{\frac{1}{2}}\|\mathbf{x}_{\Gamma^k}\|_2 + (1+t^{-1})^{\frac{1}{2}}\|\mathbf{v}\|_2.
\end{aligned} \tag{64}$$

Thus far, we have established a lower bound for $\|\mathbf{r}^{k+1}\|_2$ in (59) and an upper bound for $\|\mathbf{r}^{k+1}\|_2$ in (64). Now we combine (59) and (64) to obtain

$$\|\mathbf{x}_{\Gamma^{k+1}}\|_2 \leq \alpha'\|\mathbf{x}_{\Gamma^k}\|_2 + \beta'\|\mathbf{v}\|_2, \tag{65}$$

where

$$\alpha' = \left(\frac{(1+t)(1+\delta_\gamma)}{1 - \delta_{|T\cup T^{k+1}|}}\left(1 - \frac{(1-\sigma^{-1})\left(1 - \delta_{|\Gamma_1^k \cup T^k|}\right)}{1+\delta_S}\right)\right)^{\frac{1}{2}} \tag{66}$$

and

$$\beta' = \left((1+t^{-1})^{\frac{1}{2}} + 1\right)\left(1 - \delta_{|T\cup T^{k+1}|}\right)^{-1/2}. \tag{67}$$

Recalling that $t = \frac{1}{6}$ and $\sigma = \frac{1}{2}\exp\left(\frac{14}{9}\right)$ and also noting that $\delta_{|\Gamma_1^k \cup T^k|} \leq \delta_{|T\cup T^{k+1}|} = \delta_{|T^{k+1}|+|\Gamma^{k+1}|} \leq \delta_{Sk+S+\gamma}$ and $\delta_\gamma \leq \delta_{Sk+S+\gamma}$, one can show from (66) that

$$\alpha' < 1 \tag{68}$$

under $\delta_{Sk+S+\gamma} \leq \frac{1}{8}$.

Now, we consider two cases: 1) $\beta'\|\mathbf{v}\|_2 < (1-\alpha')\|\mathbf{x}_{\Gamma^k}\|_2$ and 2) $\beta'\|\mathbf{v}\|_2 \geq (1-\alpha')\|\mathbf{x}_{\Gamma^k}\|_2$.

First, if $\beta'\|\mathbf{v}\|_2 < (1-\alpha')\|\mathbf{x}_{\Gamma^k}\|_2$, (65) implies (57) (i.e., $\|\mathbf{x}_{\Gamma^{k+1}}\|_2^2 < \|\mathbf{x}_{\Gamma^k}\|_2^2$) so that (29) holds true.

Second, if $\beta'\|\mathbf{v}\|_2 \geq (1-\alpha')\|\mathbf{x}_{\Gamma^k}\|_2$, then (27) directly holds true because

$$\begin{aligned}
&\left\|\mathbf{r}^{k+\max\left\{\gamma, \left\lfloor \frac{8\gamma}{S} \right\rfloor\right\}}\right\|_2 \\
&\overset{(a)}{\leq} \|\mathbf{r}^{k+1}\|_2 \\
&\overset{(c)}{=} \alpha'(1-\delta_{Sk+S+\gamma})^{\frac{1}{2}}\|\mathbf{x}_{\Gamma^k}\|_2 + (1+t^{-1})^{\frac{1}{2}}\|\mathbf{v}\|_2 \\
&\leq \left(\frac{\alpha'\beta'(1-\delta_{Sk+S+\gamma})^{\frac{1}{2}}}{1-\alpha'} + (1+t^{-1})^{\frac{1}{2}}\right)\|\mathbf{v}\|_2 \\
&= \left(\frac{(1+t^{-1})^{\frac{1}{2}} + 1}{1-\alpha'} - 1\right)\|\mathbf{v}\|_2 \\
&\leq \mu_k\|\mathbf{v}\|_2,
\end{aligned} \tag{69}$$

where (a) is due to (34) and the fact that the residual power of gOMP is always non-increasing and (b) is from (64) and (66). This completes the proof of (29).

IV. Conclusion

As a method to enhance the recovery performance of orthogonal matching pursuit (OMP), generalized OMP (gOMP) has received attention in recent years [17]–[25]. While empirical evidence has shown that gOMP is effective in reconstructing sparse signals, theoretical results to date are relatively weak. In this paper, we have presented an improved recovery guarantee of gOMP by showing that the gOMP algorithm can perform stable recovery of all sparse signals from the noisy measurements under the restricted isometry property (RIP) with $\delta_{\max\{9,S+1\}K} \leq \frac{1}{8}$. The presented proof strategy might be useful for obtaining improved results for other greedy algorithms derived from the OMP algorithm.

Appendix A
Proof of Theorem 2

Proof: We first give the proof of (11). Observe that

$$\begin{aligned}
&\left\|\mathbf{r}^{\max\left\{K,\left\lfloor\frac{8K}{S}\right\rfloor\right\}}\right\|_2 \\
&= \left\|\mathbf{y} - \mathbf{\Phi}\hat{\mathbf{x}}^{\max\left\{K,\left\lfloor\frac{8K}{S}\right\rfloor\right\}}\right\|_2 \\
&= \left\|\mathbf{\Phi}\left(\mathbf{x} - \hat{\mathbf{x}}^{\max\left\{K,\left\lfloor\frac{8K}{S}\right\rfloor\right\}}\right) + \mathbf{v}\right\|_2 \\
&\geq \left\|\mathbf{\Phi}\left(\mathbf{x} - \hat{\mathbf{x}}^{\max\left\{K,\left\lfloor\frac{8K}{S}\right\rfloor\right\}}\right)\right\|_2 - \|\mathbf{v}\|_2 \\
&\overset{(a)}{\geq} \left(1 - \delta_{\left|T\cup T^{\max\left\{K,\left\lfloor\frac{8K}{S}\right\rfloor\right\}}\right|}\right)^{\frac{1}{2}} \left\|\mathbf{x} - \hat{\mathbf{x}}^{\max\left\{K,\left\lfloor\frac{8K}{S}\right\rfloor\right\}}\right\|_2 \\
&\quad - \|\mathbf{v}\|_2 \\
&\overset{(b)}{\geq} \left(1 - \delta_{\max\{9,S+1\}K}\right)^{\frac{1}{2}} \left\|\mathbf{x} - \hat{\mathbf{x}}^{\max\left\{K,\left\lfloor\frac{8K}{S}\right\rfloor\right\}}\right\|_2 - \|\mathbf{v}\|_2,
\end{aligned} \tag{A.1}$$

where (a) is from the RIP and (b) is because

$$\begin{aligned}
\left|T \cup T^{\max\left\{K,\left\lfloor\frac{8K}{S}\right\rfloor\right\}}\right| &\leq |T| + \left|T^{\max\left\{K,\left\lfloor\frac{8K}{S}\right\rfloor\right\}}\right| \\
&\leq K + \max\left\{K, \left\lfloor\frac{8K}{S}\right\rfloor\right\} S \\
&\leq \max\{9, S+1\}K.
\end{aligned}$$

Using (27) and (A.1), we have

$$\begin{aligned}
&\left\|\mathbf{x} - \hat{\mathbf{x}}^{\max\left\{K,\left\lfloor\frac{8K}{S}\right\rfloor\right\}}\right\|_2 \\
&\leq \left(1 - \delta_{\max\{9,S+1\}K}\right)^{-1/2} \left(\left\|\mathbf{r}^{\max\left\{K,\left\lfloor\frac{8K}{S}\right\rfloor\right\}}\right\|_2 + \|\mathbf{v}\|_2\right) \\
&\overset{(a)}{\leq} \left(1 - \delta_{\max\{9,S+1\}K}\right)^{-1/2} (\mu_0 + 1)\|\mathbf{v}\|_2 \\
&= \mu\|\mathbf{v}\|_2,
\end{aligned} \tag{A.2}$$

where

$$\mu = \left(1 - 2\left(\frac{7(1+\delta)\left(1 - \exp\left(-\frac{14}{9}\right)\right)}{6(1-\delta)\exp\left(\frac{14}{9}\right)}\right)^{\frac{1}{2}}\right)^{-1} \frac{\sqrt{7}+1}{(1-\delta)^{-1/2}} \tag{A.3}$$

where $\delta = \delta_{\max\{9,S+1\}K}$ and (a) is because $\delta_{7K} \leq \delta_{\max\{9,S+1\}K} \leq \frac{1}{8}$ (see Theorem 2).

Now, we turn to the proof of (12). Using the best K-term approximation $\left(\hat{\mathbf{x}}^{\max\left\{K,\left\lfloor\frac{8K}{S}\right\rfloor\right\}}\right)_K$ of $\hat{\mathbf{x}}^{\max\left\{K,\left\lfloor\frac{8K}{S}\right\rfloor\right\}}$, we have

$$\begin{aligned}
&\left\|\left(\hat{\mathbf{x}}^{\max\left\{K,\left\lfloor\frac{8K}{S}\right\rfloor\right\}}\right)_K - \mathbf{x}\right\|_2 \\
&= \left\|\left(\hat{\mathbf{x}}^{\max\left\{K,\left\lfloor\frac{8K}{S}\right\rfloor\right\}}\right)_K - \hat{\mathbf{x}}^{\max\left\{K,\left\lfloor\frac{8K}{S}\right\rfloor\right\}}\right. \\
&\quad \left. + \hat{\mathbf{x}}^{\max\left\{K,\left\lfloor\frac{8K}{S}\right\rfloor\right\}} - \mathbf{x}\right\|_2 \\
&\overset{(a)}{\leq} \left\|\left(\hat{\mathbf{x}}^{\max\left\{K,\left\lfloor\frac{8K}{S}\right\rfloor\right\}}\right)_K - \hat{\mathbf{x}}^{\max\left\{K,\left\lfloor\frac{8K}{S}\right\rfloor\right\}}\right\|_2 \\
&\quad + \left\|\hat{\mathbf{x}}^{\max\left\{K,\left\lfloor\frac{8K}{S}\right\rfloor\right\}} - \mathbf{x}\right\|_2 \\
&\overset{(b)}{\leq} 2\left\|\hat{\mathbf{x}}^{\max\left\{K,\left\lfloor\frac{8K}{S}\right\rfloor\right\}} - \mathbf{x}\right\|_2 \\
&\leq 2\mu\|\mathbf{v}\|_2,
\end{aligned} \tag{A.4}$$

where (a) is from the triangle inequality and (b) is because $\left(\hat{\mathbf{x}}^{\max\left\{K,\left\lfloor\frac{8K}{S}\right\rfloor\right\}}\right)_K$ is the best K-term approximation to $\hat{\mathbf{x}}^{\max\left\{K,\left\lfloor\frac{8K}{S}\right\rfloor\right\}}$, and hence is a better approximation than $\mathbf{x}$ (note that both $\left(\hat{\mathbf{x}}^{\max\left\{K,\left\lfloor\frac{8K}{S}\right\rfloor\right\}}\right)_K$ and $\mathbf{x}$ are K-sparse).

On the other hand,

$$\begin{aligned}
&\left\|\left(\hat{\mathbf{x}}^{\max\left\{K,\left\lfloor\frac{8K}{S}\right\rfloor\right\}}\right)_K - \mathbf{x}\right\|_2 \\
&\overset{(a)}{\geq} (1-\delta_{2K})^{-1/2}\left\|\mathbf{\Phi}\left(\left(\hat{\mathbf{x}}^{\max\left\{K,\left\lfloor\frac{8K}{S}\right\rfloor\right\}}\right)_K - \mathbf{x}\right)\right\|_2 \\
&= (1-\delta_{2K})^{-1/2}\left\|\mathbf{\Phi}\left(\hat{\mathbf{x}}^{\max\left\{K,\left\lfloor\frac{8K}{S}\right\rfloor\right\}}\right)_K - \mathbf{y} + \mathbf{v}\right\|_2 \\
&\overset{(b)}{\geq} (1-\delta_{2K})^{-1/2}\left(\left\|\mathbf{\Phi}\left(\hat{\mathbf{x}}^{\max\left\{K,\left\lfloor\frac{8K}{S}\right\rfloor\right\}}\right)_K - \mathbf{y}\right\|_2 - \|\mathbf{v}\|_2\right) \\
&\overset{(c)}{\geq} (1-\delta_{2K})^{-1/2}\left(\|\mathbf{\Phi}\hat{\mathbf{x}} - \mathbf{y}\|_2 - \|\mathbf{v}\|_2\right) \\
&= (1-\delta_{2K})^{-1/2}\left(\|\mathbf{\Phi}(\hat{\mathbf{x}} - \mathbf{x}) - \mathbf{v}\|_2 - \|\mathbf{v}\|_2\right) \\
&\overset{(d)}{\geq} (1-\delta_{2K})^{-1/2}\left(\|\mathbf{\Phi}(\hat{\mathbf{x}} - \mathbf{x})\|_2 - 2\|\mathbf{v}\|_2\right) \\
&\overset{(e)}{\geq} (1-\delta_{2K})^{-1/2}\left((1+\delta_{2K})^{\frac{1}{2}}\|\hat{\mathbf{x}} - \mathbf{x}\|_2 - 2\|\mathbf{v}\|_2\right) \\
&\geq \left(1-\delta_{\max\{9,S+1\}K}\right)^{-1/2} \\
&\quad \times \left(\left(1+\delta_{\max\{9,S+1\}K}\right)^{\frac{1}{2}}\|\hat{\mathbf{x}} - \mathbf{x}\|_2 - 2\|\mathbf{v}\|_2\right),
\end{aligned} \tag{A.5}$$

where (a) is from the RIP, (b) and (d) are from the triangle inequality, (c) is because $\left(\hat{\mathbf{x}}^{\max\left\{K,\left\lfloor\frac{8K}{S}\right\rfloor\right\}}\right)_K$ is supported on $\hat{T}$ and

$$\hat{\mathbf{x}}_{\hat{T}} = \mathbf{\Phi}_{\hat{T}}^{\dagger}\mathbf{y} = \arg\min_{\mathbf{u}} \|\mathbf{y} - \mathbf{\Phi}_{\hat{T}}\mathbf{u}\|_2,$$

and (e) follows from the RIP.

Combining (A.4) and (A.5) yields

$$\begin{aligned}
&\left(1-\delta_{\max\{9,S+1\}K}\right)^{-1/2} \\
&\quad \times \left(\left(1+\delta_{\max\{9,S+1\}K}\right)^{\frac{1}{2}}\|\hat{\mathbf{x}} - \mathbf{x}\|_2 - 2\|\mathbf{v}\|_2\right) \leq 2\mu\|\mathbf{v}\|_2.
\end{aligned}$$

That is,

$$\|\hat{\mathbf{x}} - \mathbf{x}\|_2 \leq C\|\mathbf{v}\|_2, \tag{A.6}$$

where

$$C = 2\left(\frac{1+\delta_{\max\{9,S+1\}K}}{1-\delta_{\max\{9,S+1\}K}}\right)^{\frac{1}{2}}\mu + 2\left(1-\delta_{\max\{9,S+1\}K}\right)^{-1/2}$$
$$= \frac{2(1+\delta)^{\frac{1}{2}}}{1-\delta}\left(1-2\left(\frac{7(1+\delta)\left(1-\exp\left(-\frac{14}{9}\right)\right)}{6(1-\delta)\exp\left(\frac{14}{9}\right)}\right)^{\frac{1}{2}}\right)^{-1}$$
$$\times(\sqrt{7}+1)+2(1-\delta)^{-1/2}, \quad \text{(A.7)}$$

where $\delta = \delta_{\max\{9,S+1\}K}$, which completes the proof. □

Appendix B
Proof of (20)

Proof: Recall from (19c) that

$$\left\|\mathbf{x}_{\Gamma^k\setminus\Gamma^k_{L-2}}\right\|_2^2 < \sigma\left\|\mathbf{x}_{\Gamma^k\setminus\Gamma^k_{L-1}}\right\|_2^2. \quad \text{(B.1)}$$

Subtracting both sides by $\|\mathbf{x}_{\Gamma^k\setminus\Gamma^k_{L-1}}\|_2^2$, we have

$$\left\|\mathbf{x}_{\Gamma^k_{L-1}\setminus\Gamma^k_{L-2}}\right\|_2^2 < (\sigma-1)\left\|\mathbf{x}_{\Gamma^k\setminus\Gamma^k_{L-1}}\right\|_2^2. \quad \text{(B.2)}$$

Since $|x_1| \geq |x_2| \geq \cdots \geq |x_{|\Gamma^k|}|$, and also noting that $\Gamma^k \setminus \Gamma^k_{L-1} = \{2^{L-2}+1, \cdots, |\Gamma^k|\}$ (see (18)), the elements of $\mathbf{x}_{\Gamma^k\setminus\Gamma^k_{L-1}}$ are the smallest $|\Gamma^k| - 2^{L-2}S$ elements (in magnitude) of the vector $\mathbf{x}_{\Gamma^k}$. Furthermore, since $\sigma - 1 \geq 1$, (B.2) is equivalent to

$$\left|\Gamma^k_{L-1}\setminus\Gamma^k_{L-2}\right| < (\sigma-1)\left(|\Gamma^k| - 2^{L-2}S\right). \quad \text{(B.3)}$$

Now we consider two cases. First, when $L = 2$, one can rewrite (B.3) as

$$\left|\Gamma^k_{L-1}\right| < (\sigma-1)\left(|\Gamma^k| - S\right), \quad \text{(B.4)}$$

and hence

$$|\Gamma^k| > \left(\frac{\sigma}{\sigma-1}\right)S. \quad \text{(B.5)}$$

Second, when $L \geq 3$, (B.3) becomes

$$2^{L-3}S < (\sigma-1)\left(|\Gamma^k| - 2^{L-2}S\right). \quad \text{(B.6)}$$

Equivalently,

$$|\Gamma^k| > \left(\frac{2\sigma-1}{2\sigma-2}\right)2^{L-2}S. \quad \text{(B.7)}$$

Combining these two cases yields the desired result. □

Appendix C
Proof of Proposition 1

Proof: We first consider the proof of (22). $(T^k \cap T) \subseteq T^k$ implies that

$$\|\mathbf{r}^k\|_2^2 = \left\|\mathcal{P}^{\perp}_{T^k}\mathbf{y}\right\|_2^2 \leq \left\|\mathcal{P}^{\perp}_{T^k\cap T}\mathbf{y}\right\|_2^2. \quad \text{(C.1)}$$

Also, noting that $\mathcal{P}^{\perp}_{T^k\cap T}\mathbf{y}$ is the projection of $\mathbf{y}$ onto the orthogonal complement of $span(\mathbf{\Phi}_{T^k\cap T})$, we have

$$\left\|\mathcal{P}^{\perp}_{T^k\cap T}\mathbf{y}\right\|_2^2 = \min_{supp(\mathbf{z})=T^k\cap T}\|\mathbf{y}-\mathbf{\Phi}\mathbf{z}\|_2^2. \quad \text{(C.2)}$$

From (C.1) and (C.2), we have

$$\begin{aligned}\|\mathbf{r}^k\|_2^2 &\leq \|\mathbf{y} - \mathbf{\Phi}_{T^k\cap T}\mathbf{x}_{T^k\cap T}\|_2^2 \\ &= \|\mathbf{\Phi}_T\mathbf{x}_T + \mathbf{v} - \mathbf{\Phi}_{T^k\cap T}\mathbf{x}_{T^k\cap T}\|_2^2 \\ &= \|\mathbf{\Phi}_{\Gamma^k}\mathbf{x}_{\Gamma^k} + \mathbf{v}\|_2^2, \quad \text{(C.3)}\end{aligned}$$

where (C.3) is from $T \setminus (T^k \cap T) = T \setminus T^k = \Gamma^k$.

Now, we turn to the proof of (23). The proof consists of two steps. First, we will show that the residual power difference of the gOMP satisfies

$$\|\mathbf{r}^l\|_2^2 - \|\mathbf{r}^{l+1}\|_2^2 \geq \frac{1}{1+\delta_S}\left\|\mathbf{\Phi}'_{\Lambda^{k+1}}\mathbf{r}^l\right\|_2^2. \quad \text{(C.4)}$$

Second, we will show that

$$\left\|\mathbf{\Phi}'_{\Lambda^{l+1}}\mathbf{r}^l\right\|_2^2 \geq \frac{1-\delta_{|\Gamma^k_\tau\cup T^l|}}{\left\lceil\frac{|\Gamma^k_\tau|}{S}\right\rceil}\left(\|\mathbf{r}^l\|_2^2 - \left\|\mathbf{\Phi}_{\Gamma^k\setminus\Gamma^k_\tau}\mathbf{x}_{\Gamma^k\setminus\Gamma^k_\tau}+\mathbf{v}\right\|_2^2\right). \quad \text{(C.5)}$$

Finally, (23) is established by combining (C.4) and (C.5).

- **Proof of** (C.4):
  Recall that the gOMP algorithm orthogonalizes the measurements $\mathbf{y}$ against previously chosen columns of $\mathbf{\Phi}$, yielding the updated residual in each iteration. That is,

$$\mathbf{r}^{l+1} = \mathcal{P}^{\perp}_{T^{l+1}}\mathbf{y}. \quad \text{(C.6)}$$

  Since $\mathbf{r}^l = \mathbf{y} - \mathbf{\Phi}\hat{\mathbf{x}}^l$, we have

$$\mathbf{r}^{l+1} = \mathcal{P}^{\perp}_{T^{k+1}}(\mathbf{r}^l + \mathbf{\Phi}\hat{\mathbf{x}}^l) = \mathcal{P}^{\perp}_{T^{l+1}}\mathbf{r}^l, \quad \text{(C.7)}$$

  where (C.7) is because $\mathbf{\Phi}\hat{\mathbf{x}}^l \in span(\mathbf{\Phi}_{T^l})$ and $T^l \subset T^{l+1}$, and hence $\mathcal{P}^{\perp}_{T^{l+1}}\mathbf{\Phi}\hat{\mathbf{x}}^l = \mathbf{0}$. As a result,

$$\mathbf{r}^l - \mathbf{r}^{l+1} = \mathbf{r}^l - \mathcal{P}^{\perp}_{T^{l+1}}\mathbf{r}^l = \mathcal{P}_{T^{l+1}}\mathbf{r}^l. \quad \text{(C.8)}$$

  Noting that $\Lambda^{l+1} \subseteq T^{l+1}$, we have

$$\|\mathbf{r}^l - \mathbf{r}^{l+1}\|_2 = \|\mathcal{P}_{T^{l+1}}\mathbf{r}^l\|_2 \geq \|\mathcal{P}_{\Lambda^{l+1}}\mathbf{r}^l\|_2. \quad \text{(C.9)}$$

  Since $\mathcal{P}_{\Lambda^{l+1}} = \mathcal{P}'_{\Lambda^{l+1}} = (\mathbf{\Phi}^{\dagger}_{\Lambda^{l+1}})'\mathbf{\Phi}'_{\Lambda^{l+1}}$, we further have

$$\begin{aligned}\|\mathbf{r}^l - \mathbf{r}^{l+1}\|_2 &\geq \left\|\left(\mathbf{\Phi}^{\dagger}_{\Lambda^{l+1}}\right)'\mathbf{\Phi}'_{\Lambda^{l+1}}\mathbf{r}^l\right\|_2 \\ &\geq (1+\delta_S)^{-1/2}\left\|\mathbf{\Phi}'_{\Lambda^{l+1}}\mathbf{r}^l\right\|_2, \quad \text{(C.10)}\end{aligned}$$

  where (C.10) is because the singular values of $\mathbf{\Phi}_{\Lambda^{l+1}}$ lie between $(1-\delta_S)^{1/2}$ and $(1+\delta_S)^{1/2}$, and hence the smallest singular value of $\mathbf{\Phi}^{\dagger}_{\Lambda^{l+1}}$ is lower bounded by $(1+\delta_S)^{-1/2}$.[6]
- **Proof of** (C.5):
  We first introduce a lemma useful in our proof.

Lemma 1: Let $\mathbf{u}, \mathbf{z} \in \mathcal{R}^n$ be two distinct vectors and let $W = supp(\mathbf{u}) \cap supp(\mathbf{z})$. Also, let U be the set of S indices corresponding to S most significant elements in $\mathbf{u}$. Then, for any integer $S \geq 1$,

$$\langle\mathbf{u},\mathbf{z}\rangle \leq \left(\left\lceil\frac{|W|}{S}\right\rceil\right)^{\frac{1}{2}}\|\mathbf{u}_U\|_2\|\mathbf{z}_W\|_2. \quad \text{(C.11)}$$

Proof: See Appendix D. □

Now we are ready to prove (C.5). Let $\mathbf{u} = \mathbf{\Phi}'\mathbf{r}^l$ and let $\mathbf{z} \in \mathcal{R}^n$ be the vector satisfying $\mathbf{z}_{T\cap T^k\cup\Gamma^k_\tau} = \mathbf{x}_{T\cap T^k\cup\Gamma^k_\tau}$ and $\mathbf{z}_{\Omega\setminus(T\cap T^k\cup\Gamma^k_\tau)} = \mathbf{0}$. Since $supp(\mathbf{u}) = \Omega\setminus T^l$ and $supp(\mathbf{z}) = T \cap T^k \cup \Gamma^k_\tau$, and also noting that $T^k \subseteq T^l$, we have $W =$

[6]Suppose the matrix $\mathbf{\Phi}_{\Lambda^{l+1}}$ has singular value decomposition $\mathbf{\Phi}_{\Lambda^{l+1}} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}'$, then $\mathbf{\Phi}^{\dagger}_{\Lambda^{l+1}} = \mathbf{V}\mathbf{\Sigma}^{\dagger}\mathbf{U}'$, where $\mathbf{\Sigma}^{\dagger}$ is the pseudoinverse of $\mathbf{\Sigma}$, which is formed by replacing every non-zero diagonal entry with its reciprocal and transposing the resulting matrix.

$supp(\mathbf{u}) \cap supp(\mathbf{z}) = \Gamma_\tau^k \setminus T^l$. Moreover, since Λ^{l+1} contains the indices corresponding to S most significant elements in $\mathbf{u} = \mathbf{\Phi}'\mathbf{r}^l$, we have $U = \Lambda^{l+1}$. Using Lemma 1,

$$\begin{aligned}\langle \mathbf{\Phi}'\mathbf{r}^l, \mathbf{z}\rangle &\leq \left(\left\lceil \frac{|\Gamma_\tau^k \setminus T^l|}{S}\right\rceil\right)^{\frac{1}{2}} \left\|\mathbf{\Phi}'_{\Lambda^{l+1}}\mathbf{r}^l\right\|_2 \left\|\mathbf{z}_{\Gamma_\tau^k \setminus T^l}\right\|_2 \\ &\leq \left(\left\lceil \frac{|\Gamma_\tau^k|}{S}\right\rceil\right)^{\frac{1}{2}} \left\|\mathbf{\Phi}'_{\Lambda^{l+1}}\mathbf{r}^l\right\|_2 \left\|\mathbf{z}_{\Gamma_\tau^k \setminus T^l}\right\|_2 \\ &\leq \left(\left\lceil \frac{|\Gamma_\tau^k|}{S}\right\rceil\right)^{\frac{1}{2}} \left\|\mathbf{\Phi}'_{\Lambda^{l+1}}\mathbf{r}^l\right\|_2 \left\|\mathbf{z}_{\Omega \setminus T^l}\right\|_2. \end{aligned} \tag{C.12}$$

On the other hand,

$$\begin{aligned}&\langle \mathbf{\Phi}'\mathbf{r}^l, \mathbf{z}\rangle \\ &= \langle \mathbf{\Phi}'\mathbf{r}^l, \mathbf{z} - \hat{\mathbf{x}}^l\rangle + \langle \mathbf{\Phi}'\mathbf{r}^l, \hat{\mathbf{x}}^l\rangle \\ &\overset{(a)}{=} \langle \mathbf{\Phi}'\mathbf{r}^l, \mathbf{z} - \hat{\mathbf{x}}^l\rangle \\ &= \left\langle \mathbf{\Phi}(\mathbf{z} - \hat{\mathbf{x}}^l), \mathbf{r}^l\right\rangle \\ &\overset{(b)}{=} \frac{1}{2}\left(\left\|\mathbf{\Phi}(\mathbf{z} - \hat{\mathbf{x}}^l)\right\|_2^2 + \|\mathbf{r}^l\|_2^2 - \left\|\mathbf{r}^l - \mathbf{\Phi}(\mathbf{z} - \hat{\mathbf{x}}^l)\right\|_2^2\right) \\ &\overset{(c)}{=} \frac{1}{2}\left(\left\|\mathbf{\Phi}(\mathbf{z} - \hat{\mathbf{x}}^l)\right\|_2^2 + \|\mathbf{r}^l\|_2^2 - \|\mathbf{\Phi}(\mathbf{x} - \mathbf{z}) + \mathbf{v}\|_2^2\right) \\ &= \frac{1}{2}\left(\left\|\mathbf{\Phi}(\mathbf{z} - \hat{\mathbf{x}}^l)\right\|_2^2 + \|\mathbf{r}^l\|_2^2 - \left\|\mathbf{\Phi}_{\Gamma^k \setminus \Gamma_\tau^k}\mathbf{x}_{\Gamma^k \setminus \Gamma_\tau^k} + \mathbf{v}\right\|_2^2\right) \\ &\overset{(d)}{\geq} \left\|\mathbf{\Phi}(\mathbf{z} - \hat{\mathbf{x}}^l)\right\|_2 \left(\|\mathbf{r}^l\|_2^2 - \left\|\mathbf{\Phi}_{\Gamma^k \setminus \Gamma_\tau^k}\mathbf{x}_{\Gamma^k \setminus \Gamma_\tau^k} + \mathbf{v}\right\|_2^2\right)^{\frac{1}{2}} \\ &\overset{(e)}{\geq} \left(1 - \delta_{|\Gamma_\tau^k \cup T^l|}\right)^{\frac{1}{2}} \|\mathbf{z} - \hat{\mathbf{x}}^l\|_2 \\ &\quad \times \left(\|\mathbf{r}^l\|_2^2 - \left\|\mathbf{\Phi}_{\Gamma^k \setminus \Gamma_\tau^k}\mathbf{x}_{\Gamma^k \setminus \Gamma_\tau^k} + \mathbf{v}\right\|_2^2\right)^{\frac{1}{2}} \\ &\geq \left(1 - \delta_{|\Gamma_\tau^k \cup T^l|}\right)^{\frac{1}{2}} \left\|(\mathbf{z} - \hat{\mathbf{x}}^l)_{\Omega \setminus T^l}\right\|_2 \\ &\quad \times \left(\|\mathbf{r}^l\|_2^2 - \left\|\mathbf{\Phi}_{\Gamma^k \setminus \Gamma_\tau^k}\mathbf{x}_{\Gamma^k \setminus \Gamma_\tau^k} + \mathbf{v}\right\|_2^2\right)^{\frac{1}{2}} \\ &\overset{(f)}{\geq} \left(1 - \delta_{|\Gamma_\tau^k \cup T^l|}\right)^{\frac{1}{2}} \|\mathbf{z}_{\Omega \setminus T^l}\|_2 \\ &\quad \times \left(\|\mathbf{r}^l\|_2^2 - \left\|\mathbf{\Phi}_{\Gamma^k \setminus \Gamma_\tau^k}\mathbf{x}_{\Gamma^k \setminus \Gamma_\tau^k} + \mathbf{v}\right\|_2^2\right)^{\frac{1}{2}}, \end{aligned} \tag{C.13}$$

where (a) is because $supp(\hat{\mathbf{x}}^l) = T^l$ and $supp(\mathbf{\Phi}'\mathbf{r}^l) = \Omega \setminus T^l$ and hence $\langle \mathbf{\Phi}'\mathbf{r}^l, \hat{\mathbf{x}}^l\rangle = \mathbf{0}$, (b) uses the fact that $\langle \mathbf{u}, \mathbf{v}\rangle = \frac{1}{2}(\|\mathbf{u}\|_2^2 + \|\mathbf{v}\|_2^2 - \|\mathbf{u} - \mathbf{v}\|_2^2)$, (c) is from $\mathbf{r}^l + \mathbf{\Phi}\hat{\mathbf{x}}^l = \mathbf{y} = \mathbf{\Phi}\mathbf{x} + \mathbf{v}$, (d) uses the inequality $\frac{1}{2}(a+b) \geq \sqrt{ab}$ (with $a = \|\mathbf{\Phi}(\mathbf{z} - \hat{\mathbf{x}}^l)\|_2^2$ and $b = \|\mathbf{r}^l\|_2^2 - \|\mathbf{\Phi}_{\Gamma^k \setminus \Gamma_\tau^k}\mathbf{x}_{\Gamma^k \setminus \Gamma_\tau^k} + \mathbf{v}\|_2^2$),[7] (e) is from the RIP ($\|\mathbf{z} - \hat{\mathbf{x}}^l\|_0 = |\Gamma_\tau^k \cup T^l|$), and (e) is due to $(\hat{\mathbf{x}}^l)_{\Omega \setminus T^l} = \mathbf{0}$. Finally, using (C.12) and (C.13), we have

$$\left\|\mathbf{\Phi}'_{\Lambda^{l+1}}\mathbf{r}^l\right\|_2 \geq \left(\frac{1 - \delta_{|\Gamma_\tau^k \cup T^l|}}{\left\lceil \frac{|\Gamma_\tau^k|}{S}\right\rceil}\left(\|\mathbf{r}^l\|_2^2 - \left\|\mathbf{\Phi}_{\Gamma^k \setminus \Gamma_\tau^k}\mathbf{x}_{\Gamma^k \setminus \Gamma_\tau^k} + \mathbf{v}\right\|_2^2\right)\right)^{\frac{1}{2}},$$

which is the desired result. □

[7]Note that we only need to consider the case $\|\mathbf{r}^l\|_2^2 - \|\mathbf{\Phi}_{\Gamma^k \setminus \Gamma_\tau^k}\mathbf{x}_{\Gamma^k \setminus \Gamma_\tau^k} + \mathbf{v}\|_2^2 \geq 0$. For the alternative case where $\|\mathbf{r}^l\|_2^2 - \|\mathbf{\Phi}_{\Gamma^k \setminus \Gamma_\tau^k}\mathbf{x}_{\Gamma^k \setminus \Gamma_\tau^k} + \mathbf{v}\|_2^2 < 0$, (C.5) directly holds true, since $\|\mathbf{\Phi}'_{\Lambda^{l+1}}\mathbf{r}^l\|_2^2 \geq 0$.

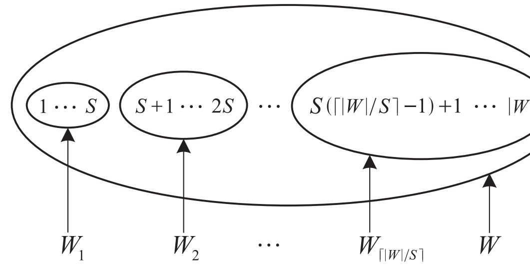


Fig. 4. Illustration of indices in W_i.

APPENDIX D
PROOF OF LEMMA 1

Proof: We consider two cases: 1) $1 \leq S \leq |W|$ and 2) $S > |W|$.

We first consider the case $1 \leq S \leq |W|$. Without loss of generality, we assume that $W = \{1, 2, \cdots, |W|\}$ and that the elements of $\mathbf{u}_W$ are arranged in descending order of their magnitudes. We define the subset W_i of W as

$$W_i = \begin{cases} \{S(i-1)+1, \cdots, Si\} & i = 1, \cdots, \left\lceil \frac{|W|}{S}\right\rceil - 1, \\ \left\{S\left(\left\lceil \frac{|W|}{S}\right\rceil - 1\right)+1, \cdots, |W|\right\} & i = \lceil \frac{|W|}{S}\rceil. \end{cases} \tag{D.1}$$

See Fig. 4 for the illustration of indices in W_i. Note that when $\left\lceil \frac{|W|}{S}\right\rceil > \frac{|W|}{S}$, the last set $W_{\left\lceil \frac{|W|}{S}\right\rceil}$ has less than S elements.

Observe that

$$\langle \mathbf{u}, \mathbf{z}\rangle = \langle \mathbf{u}_W, \mathbf{z}_W\rangle \leq \sum_i |\langle \mathbf{u}_{W_i}, \mathbf{z}_{W_i}\rangle| \leq \sum_i \|\mathbf{u}_{W_i}\|_2 \|\mathbf{z}_{W_i}\|_2, \tag{D.2}$$

where the second inequality is due to the Hölder's inequality. Using the definition of U, we have $\|\mathbf{u}_U\|_2 \geq \|\mathbf{u}_{W_1}\|_2 = \max_i \|\mathbf{u}_{W_i}\|_2$, and hence

$$\langle \mathbf{u}, \mathbf{z}\rangle \leq \|\mathbf{u}_U\|_2 \sum_i \|\mathbf{z}_{W_i}\|_2 \tag{D.3}$$

$$\leq \|\mathbf{u}_U\|_2 \left(\left\lceil \frac{|W|}{S}\right\rceil \sum_i \|\mathbf{z}_{W_i}\|_2^2\right)^{\frac{1}{2}} \tag{D.4}$$

$$= \left(\left\lceil \frac{|W|}{S}\right\rceil\right)^{\frac{1}{2}} \|\mathbf{u}_U\|_2 \|\mathbf{z}_W\|_2, \tag{D.5}$$

where (D.4) follows from the fact that $\sum_{i=1}^d a_i \leq \left(d\sum_{i=1}^d a_i^2\right)^{1/2}$ with $a_i = \|\mathbf{z}_{W_i}\|_2$ and $d = \left\lceil \frac{|W|}{S}\right\rceil$.

Now, we consider the alternative case ($S > |W|$). In this case, it is clear that $\left(\left\lceil \frac{|W|}{S}\right\rceil\right)^{1/2} = 1$ and $\|\mathbf{u}_U\|_2 \geq \|\mathbf{u}_W\|_2$, and hence

$$\begin{aligned}\left(\left\lceil \frac{|W|}{S}\right\rceil\right)^{\frac{1}{2}} \|\mathbf{u}_U\|_2 \|\mathbf{z}_W\|_2 &= \|\mathbf{u}_U\|_2 \|\mathbf{z}_W\|_2 \\ &\geq \|\mathbf{u}_W\|_2 \|\mathbf{z}_W\|_2 \\ &\geq \langle \mathbf{u}_W, \mathbf{z}_W\rangle \\ &= \langle \mathbf{u}, \mathbf{z}\rangle, \end{aligned} \tag{D.6}$$

which completes the proof. □

APPENDIX E
PROOF OF PROPOSITION 2

Proof: Recall from Proposition 1 that for given Γ^k and any integer $l \geq k$, the residual of gOMP satisfies

$$\|\mathbf{r}^l\|_2^2 - \|\mathbf{r}^{l+1}\|_2^2 \geq \frac{1-\delta_{|\Gamma_\tau^k \cup T^l|}}{(1+\delta_S)\left\lceil \frac{|\Gamma_\tau^k|}{S} \right\rceil} \times \left(\|\mathbf{r}^l\|_2^2 - \|\mathbf{\Phi}_{\Gamma^k \setminus \Gamma_\tau^k}\mathbf{x}_{\Gamma^k \setminus \Gamma_\tau^k} + \mathbf{v}\|_2^2\right), \quad \text{(E.1)}$$

where $\tau = 1, 2, \cdots, \max\left\{0, \left\lceil \log_2 \frac{|\Gamma^k|}{S} \right\rceil\right\} + 1$. Since $a > 1 - \exp(-a)$ for $a > 0$ and

$$\frac{1-\delta_{|\Gamma_\tau^k \cup T^l|}}{\left\lceil \frac{|\Gamma_\tau^k|}{S} \right\rceil (1+\delta_S)} > 0,$$

we have

$$\frac{1-\delta_{|\Gamma_\tau^k \cup T^l|}}{\left\lceil \frac{|\Gamma_\tau^k|}{S} \right\rceil (1+\delta_S)} \geq 1 - \exp\left(-\frac{1-\delta_{|\Gamma_\tau^k \cup T^l|}}{\left\lceil \frac{|\Gamma_\tau^k|}{S} \right\rceil (1+\delta_S)}\right). \quad \text{(E.2)}$$

Using (E.1) and (E.2),

$$\|\mathbf{r}^l\|_2^2 - \|\mathbf{r}^{l+1}\|_2^2 \geq \left(1 - \exp\left(-\frac{1-\delta_{|\Gamma_\tau^k \cup T^l|}}{\left\lceil \frac{|\Gamma_\tau^k|}{S} \right\rceil (1+\delta_S)}\right)\right) \times \left(\|\mathbf{r}^l\|_2^2 - \|\mathbf{\Phi}_{\Gamma^k \setminus \Gamma_\tau^k}\mathbf{x}_{\Gamma^k \setminus \Gamma_\tau^k} + \mathbf{v}\|_2^2\right). \quad \text{(E.3)}$$

Subtracting both sides of (E.3) by $\|\mathbf{r}^l\|_2^2 - \|\mathbf{\Phi}_{\Gamma^k \setminus \Gamma_\tau^k}\mathbf{x}_{\Gamma^k \setminus \Gamma_\tau^k} + \mathbf{v}\|_2^2$, we have

$$\begin{aligned}\|\mathbf{r}^{l+1}\|_2^2 &- \|\mathbf{\Phi}_{\Gamma^k \setminus \Gamma_\tau^k}\mathbf{x}_{\Gamma^k \setminus \Gamma_\tau^k} + \mathbf{v}\|_2^2 \\ &\leq \exp\left(-\frac{1-\delta_{|\Gamma_\tau^k \cup T^l|}}{\left\lceil \frac{|\Gamma_\tau^k|}{S} \right\rceil (1+\delta_S)}\right) \\ &\quad \times \left(\|\mathbf{r}^l\|_2^2 - \|\mathbf{\Phi}_{\Gamma^k \setminus \Gamma_\tau^k}\mathbf{x}_{\Gamma^k \setminus \Gamma_\tau^k} + \mathbf{v}\|_2^2\right),\end{aligned}$$

and also

$$\begin{aligned}\|\mathbf{r}^{l+2}\|_2^2 &- \|\mathbf{\Phi}_{\Gamma^k \setminus \Gamma_\tau^k}\mathbf{x}_{\Gamma^k \setminus \Gamma_\tau^k} + \mathbf{v}\|_2^2 \\ &\leq \exp\left(-\frac{1-\delta_{|\Gamma_\tau^k \cup T^{l+1}|}}{\left\lceil \frac{|\Gamma_\tau^k|}{S} \right\rceil (1+\delta_S)}\right) \\ &\quad \times \left(\|\mathbf{r}^{l+1}\|_2^2 - \|\mathbf{\Phi}_{\Gamma^k \setminus \Gamma_\tau^k}\mathbf{x}_{\Gamma^k \setminus \Gamma_\tau^k} + \mathbf{v}\|_2^2\right), \\ &\vdots \\ \|\mathbf{r}^{l+\Delta l}\|_2^2 &- \|\mathbf{\Phi}_{\Gamma^k \setminus \Gamma_\tau^k}\mathbf{x}_{\Gamma^k \setminus \Gamma_\tau^k} + \mathbf{v}\|_2^2 \\ &\leq \exp\left(-\frac{1-\delta_{|\Gamma_\tau^k \cup T^{l+\Delta l-1}|}}{\left\lceil \frac{|\Gamma_\tau^k|}{S} \right\rceil (1+\delta_S)}\right) \\ &\quad \times \left(\|\mathbf{r}^{l+\Delta l-1}\|_2^2 - \left\|\mathbf{\Phi}_{\Gamma^k \setminus \Gamma_\tau^k}\mathbf{x}_{\Gamma^k \setminus \Gamma_\tau^k} + \mathbf{v}\right\|_2^2\right).\end{aligned}$$

Some additional manipulations yield the following result:

$$\begin{aligned}\|\mathbf{r}^{l+\Delta l}\|_2^2 &- \|\mathbf{\Phi}_{\Gamma^k \setminus \Gamma_\tau^k}\mathbf{x}_{\Gamma^k \setminus \Gamma_\tau^k} + \mathbf{v}\|_2^2 \\ &\leq \prod_{i=l}^{l+\Delta l-1} \exp\left(-\frac{1-\delta_{|\Gamma_\tau^k \cup T^i|}}{\left\lceil \frac{|\Gamma_\tau^k|}{S} \right\rceil (1+\delta_S)}\right) \\ &\quad \times \left(\|\mathbf{r}^l\|_2^2 - \|\mathbf{\Phi}_{\Gamma^k \setminus \Gamma_\tau^k}\mathbf{x}_{\Gamma^k \setminus \Gamma_\tau^k} + \mathbf{v}\|_2^2\right).\end{aligned}$$

Since $\delta_{|\Gamma_\tau^k \cup T^l|} \leq \delta_{|\Gamma_\tau^k \cup T^{l+1}|} \leq \cdots \leq \delta_{|\Gamma_\tau^k \cup T^{l+\Delta l-1}|}$, we further have

$$\begin{aligned}\|\mathbf{r}^{l+\Delta l}\|_2^2 &- \|\mathbf{\Phi}_{\Gamma^k \setminus \Gamma_\tau^k}\mathbf{x}_{\Gamma^k \setminus \Gamma_\tau^k} + \mathbf{v}\|_2^2 \\ &\leq C_{\tau,l,\Delta l}\left(\|\mathbf{r}^l\|_2^2 + \|\mathbf{\Phi}_{\Gamma^k \setminus \Gamma_\tau^k}\mathbf{x}_{\Gamma^k \setminus \Gamma_\tau^k} + \mathbf{v}\|_2^2\right),\end{aligned}$$

where

$$C_{\tau,l,\Delta l} = \exp\left(-\frac{\Delta l\left(1-\delta_{|\Gamma_\tau^k \cup T^{l+\Delta l-1}|}\right)}{\left\lceil \frac{|\Gamma_\tau^k|}{S} \right\rceil (1+\delta_S)}\right), \quad \text{(E.4)}$$

which completes the proof. □

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for their valuable comments and suggestions that improved the quality of the paper.

REFERENCES

[1] Y. C. Pati, R. Rezaiifar, and P. S. Krishnaprasad, "Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition," in *Proc. 27th Annu. Asilomar Conf. Signals, Systems, Computers*, Pacific Grove, CA, Nov. 1993, vol. 1, pp. 40–44.

[2] S. G. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Trans. Signal Process.*, vol. 41, no. 12, pp. 3397–3415, Dec. 1993.

[3] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4655–4666, Dec. 2007.

[4] M. A. Davenport and M. B. Wakin, "Analysis of orthogonal matching pursuit using the restricted isometry property," *IEEE Trans. Inf. Theory*, vol. 56, no. 9, pp. 4395–4401, Sep. 2010.

[5] T. T. Cai and L. Wang, "Orthogonal matching pursuit for sparse signal recovery with noise," *IEEE Trans. Inf. Theory*, vol. 57, no. 7, pp. 4680–4688, Jul. 2011.

[6] J. Wang and B. Shim, "On the recovery limit of sparse signals using orthogonal matching pursuit," *IEEE Trans. Signal Process.*, vol. 60, no. 9, pp. 4973–4976, Sep. 2012.

[7] 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.

[8] S. Kwon, J. Wang, and B. Shim, "Multipath matching pursuit," *IEEE Trans. Inf. Theory*, vol. 60, no. 5, pp. 2986–3001, May 2014.

[9] J. Wang, "Support recovery with orthogonal matching pursuit in the presence of noise," *IEEE Trans. Signal Process.*, vol. 63, no. 21, pp. 5868–5877, Nov. 2015.

[10] E. D. Livshits, "On the efficiency of the orthogonal matching pursuit in compressed sensing," *Sbornik: Mathematics*, vol. 203, no. 2, p. 183, 2012.

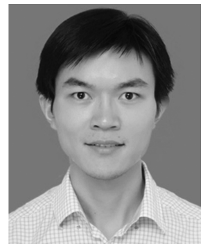
[11] T. Zhang, "Sparse recovery with orthogonal matching pursuit under RIP," *IEEE Trans. Inf. Theory*, vol. 57, no. 9, pp. 6215–6221, Sep. 2011.

[12] S. Foucart, "Stability and robustness of weak orthogonal matching pursuits," in *Proc. Recent Adv. Harmon. Anal. Appl.*, 2013, pp. 395–405.

[13] D. L. Donoho, I. Drori, Y. Tsaig, and J. L. Starck, "Sparse solution of underdetermined linear equations by stagewise orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol. 58, no. 2, pp. 1094–1121, Feb. 2012.

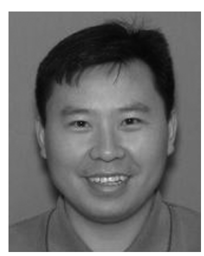
[14] D. Needell and R. Vershynin, "Signal recovery from incomplete and inaccurate measurements via regularized orthogonal matching pursuit," *IEEE J. Sel. Topics Signal Process.*, vol. 4, no. 2, pp. 310–316, Apr. 2010.

[15] D. Needell and J. A. Tropp, "CoSaMP: Iterative signal recovery from incomplete and inaccurate samples," *Appl. Computational Harmonic Analysis*, vol. 26, no. 3, pp. 301–321, Mar. 2009.
[16] W. Dai and O. Milenkovic, "Subspace pursuit for compressive sensing signal reconstruction," *IEEE Trans. Inf. Theory*, vol. 55, no. 5, pp. 2230–2249, May 2009.
[17] J. Wang, S. Kwon, and B. Shim, "Generalized orthogonal matching pursuit," *IEEE Trans. Signal Process.*, vol. 60, no. 12, pp. 6202–6216, Dec. 2012.
[18] E. Liu and V. N. Temlyakov, "The orthogonal super greedy algorithm and applications in compressed sensing," *IEEE Trans. Inf. Theory*, vol. 58, no. 4, pp. 2040–2047, Apr. 2012.
[19] E. Liu and V. N. Temlyakov, "Super greedy type algorithms," *Adv. Comput. Math.*, vol. 37, no. 4, pp. 493–504, 2012.
[20] S. Huang and J. Zhu, "Recovery of sparse signals using OMP and its variants: Convergence analysis based on RIP," *Inverse Problems*, vol. 27, no. 3, p. 035003, 2011.
[21] R. Maleh, "Improved rip analysis of orthogonal matching pursuit," arXiv preprint arXiv:1102.4311, 2011.
[22] S. Satpathi, R. L. Das, and M. Chakraborty, "Improving the bound on the RIP constant in generalized orthogonal matching pursuit," *IEEE Signal Process. Lett.*, vol. 20, no. 11, pp. 1074–1077, Nov. 2013.
[23] Y. Shen, B. Li, W. Pan, and J. Li, "Analysis of generalised orthogonal matching pursuit using restricted isometry constant," *Electron. Lett.*, vol. 50, no. 14, pp. 1020–1022, Jul. 2014.
[24] W. Dan, "Analysis of orthogonal multi-matching pursuit under restricted isometry property," *Science China Mathematics*, vol. 57, no. 10, pp. 2179–2188, Oct. 2014.
[25] B. Li, Y. Shen, Z. Wu, and J. Li, "Sufficient conditions for generalized orthogonal matching pursuit in noisy case," *Signal Process.*, vol. 108, pp. 111–123, Mar. 2015.
[26] E. J. Candès and T. Tao, "Decoding by linear programming," *IEEE Trans. Inf. Theory*, vol. 51, no. 12, pp. 4203–4215, Dec. 2005.
[27] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM J. Sci. Comput.*, vol. 20, no. 1, pp. 33–61, 1998.
[28] E. J. Candès, "The restricted isometry property and its implications for compressed sensing," *Comptes Rendus Mathematique, Ser. I*, vol. 346, no. 9–10, pp. 589–592, May 2008.
[29] R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin, "A simple proof of the restricted isometry property for random matrices," *Construct. Approx.*, vol. 28, no. 3, pp. 253–263, Dec. 2008.
[30] S. Rangan, V. Goyal, and A. K. Fletcher, "Asymptotic analysis of map estimation via the replica method and compressed sensing," in *Proc. Adv. Neural Inf. Process. Syst.*, 2009, pp. 1545–1553.
[31] D. Guo, D. Baron, and S. Shamai, "A single-letter characterization of optimal noisy compressed sensing," in *Proc. IEEE 47th Ann. Allerton Conf. Commun., Contr. Comp.*, Monticello, IL, Oct. 2009, pp. 52–59.
[32] D. L. Donoho, A. Maleki, and A. Montanari, "Message-passing algorithms for compressed sensing," *Proc. Nat. Acad. Sci.*, vol. 106, no. 45, pp. 18914–18919, Nov. 2009.
[33] T. Blumensath and M. E. Davies, "Iterative hard thresholding for compressed sensing," *Appl. Comput. Harmon. Anal.*, vol. 27, no. 3, pp. 265–274, Dec. 2009.
[34] S. Foucart, "Hard thresholding pursuit: An algorithm for compressive sensing," *SIAM J. Numer. Anal.*, vol. 49, no. 6, pp. 2543–2563, 2011.
[35] X. Yuan and S. Yan, "Forward basis selection for sparse approximation over dictionary," in *Proc. AISTATS*, 2012, pp. 1377–1388.
[36] X. Yuan and T. Zhang, "Truncated power method for sparse eigenvalue problems," *J. Mach. Learn. Res.*, vol. 14, no. 1, pp. 899–925, Apr. 2013.
[37] Y. Dong, P. Fan, and K. B. Letaief, "High-speed railway wireless communications: Efficiency versus fairness," *IEEE Trans. Veh. Technol.*, vol. 63, no. 2, pp. 925–930, Feb. 2014.
[38] S. Rangan, "Generalized approximate message passing for estimation with random linear mixing," in *Proc. IEEE Int. Symp. Inf. Theory*, Aug. 2011, pp. 2168–2172.
[39] J. Vila and P. Schniter, "Expectation-maximization bernoulli-gaussian approximate message passing," in *Proc. 45th Annu. Asilomar Conf. Signals, Systems, Computers*, Pacific Grove, CA, USA, Nov. 2011, pp. 799–803.
[40] Y. Nesterov and A. Nemirovskii, *Interior-Point Polynomial Algorithms in Convex Programming*. Philadelphia, PA, USA: SIAM, 1994.



Jian Wang (S'11) received the B.S. degree in Material Engineering and the M.S. degree in Information and Communication Engineering from Harbin Institute of Technology, China, and the Ph.D. degree in Electrical and Computer Engineering from Korea University, Korea, in 2006, 2009, and 2013, respectively. From 2013 to 2015, he held positions as Postdoctoral Research Associate at Department of Statistics & Biostatistics, Department of Computer Science, Rutgers University, Piscataway, NJ, and Department of Electrical & Computer Engineering, Duke University, Durham, NC, USA. He is currently a professor in Nanjing University of Information Science & Technology, Nanjing, China, and also an assistant research professor in Seoul National University, Seoul, Korea. His research interests include compressed sensing, sparse signal and matrix recovery, signal processing for wireless communications, and statistical learning.

Suhyuk Kwon (S'11) received the B.S., M.S., and Ph.D. degrees in School of Information and Communication from Korea University, Seoul, Korea, in 2008, 2010, and 2014, respectively. From 2014 to 2015, he was a postdoctoral associate at Department of Electrical and Computer Engineering, Seoul National University, Seoul, Korea. He is now with Samsung Display, Asan, Korea. His research interests include compressive sensing, signal processing, and information theory.



Ping Li received his Ph.D. in Statistics from Stanford University, where he also earned two masters degrees in Computer Science and Electric Engineering. He is a recipient of the AFOSR (Air Force Office of Scientific Research) Young Investigator Award (AFOSR-YIP) and a receipt of the ONR (Office of Naval Research) Young Investigator Award (ONR-YIP). Ping Li (with co-authors) won the Best Paper Award in NIPS 2014, the Best Paper Award in ASONAM 2014, and the Best Student Paper Award in KDD 2006.

Byonghyo Shim (SM'09) received the B.S. and M.S. degrees in control and instrumentation engineering from Seoul National University, Korea, in 1995 and 1997, respectively. He received the M.S. degree in mathematics and the Ph.D. degree in electrical and computer engineering from the University of Illinois at Urbana-Champaign (UIUC), USA, in 2004 and 2005, respectively.

From 1997 and 2000, he was with the Department of Electronics Engineering, Korean Air Force Academy as an Officer (First Lieutenant) and an Academic Full-time Instructor. From 2005 to 2007, he was with Qualcomm Inc., San Diego, CA, USA, as a Staff Engineer. From 2007 to 2014, he was with the School of Information and Communication, Korea University, Seoul, as an Associate Professor. Since September 2014, he has been with the Department of Electrical and Computer Engineering, Seoul National University, where he is presently an Associate Professor. His research interests include wireless communications, statistical signal processing, estimation and detection, compressive sensing, and information theory.

Dr. Shim was the recipient of the 2005 M. E. Van Valkenburg Research Award from the Electrical and Computer Engineering Department of the University of Illinois and 2010 Hadong Young Engineer Award from IEIE. He is currently an Associate Editor of the IEEE Wireless Communications Letters, Journal of Communications and Networks, and a Guest Editor of the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS (JSAC).