Sparse Signal Recovery via Tree Search Matching Pursuit

Jaeseok Lee, Jun Won Choi, and Byonghyo Shim

Abstract: **Recently, greedy algorithm has received much attention as a cost-effective means to reconstruct the sparse signals from compressed measurements. Much of previous work has focused on the investigation of a single candidate to identify the support (index set of nonzero elements) of the sparse signals. Well-known drawback of the greedy approach is that the chosen candidate is often not the optimal solution due to the myopic decision in each iteration. In this paper, we propose a tree search based sparse signal recovery algorithm referred to as the tree search matching pursuit (TSMP). Two key ingredients of the proposed TSMP algorithm to control the computational complexity are the *pre-selection* to put a restriction on columns of the sensing matrix to be investigated and the *tree pruning* to eliminate unpromising paths from the search tree. In numerical simulations of Internet of Things (IoT) environments, it is shown that TSMP outperforms conventional schemes by a large margin.**

Index Terms: **Compressive sensing, sparse recovery, greedy algorithm, tree search, tree pruning.**

I. INTRODUCTION

In recent years, compressive sensing (CS) has received much attention as a means to recover sparse signals in underdetermined system [1-16]. Key finding of the CS paradigm is that one can recover signals with far fewer measurements than traditional approaches use as long as the signals to be recovered are sparse and the sensing mechanism roughly preserves the energy of signals of interest.

It is now well known that the problem to recover the sparest signal $\mathbf{x}$ using the measurements $\mathbf{y} = \mathbf{\Phi x}$ is formulated as the ℓ_0-minimization problem

$$\min_{\mathbf{x}} \|\mathbf{x}\|_0 \quad \text{subject to } \mathbf{y} = \mathbf{\Phi x} \tag{1}$$

where $\mathbf{\Phi} \in \mathbb{R}^{\mathbb{M}\times\mathbb{N}}$ is often called sensing matrix. Since solving this problem is combinatoric in nature and known to be NP-hard [1], early works focused on the ℓ_1-relaxation method, such as Basis Pursuit (BP) [1], BP denoising (BPDN) [17] (also known as Lasso [5]), and Dantzig selector [6]. Another line of research receiving much attention in recent years is a greedy approach. In a nutshell, greedy algorithm attempts to find the support (index set of nonzero entries) in an iterative fashion, returning a sequence of estimates. Although the greedy algorithm, such as orthogonal matching pursuit (OMP) [7], is relatively simple to implement and also computationally efficient, performance might not be so appealing, in particular for the noisy scenario.

The aim of this paper is to introduce a tree search based sparse signal recovery algorithm for pursuing an improvement in the sparse signal recovery. Tree search strategy is distinct from the conventional greedy strategy in the sense that it examines a number of candidates to make the best possible decision. Compared with conventional greedy algorithms, the tree search algorithm provides better recovery performance at the expense of an increased computational burden. The proposed algorithm, referred to as the *tree search matching pursuit* (TSMP), significantly reduces the computational burden of the tree search yet achieves excellent recovery performance in both noiseless and noisy scenarios.

Two key ingredients of the TSMP algorithm accomplishing this mission are the *pre-selection* to put a restriction on columns of $\mathbf{\Phi}$ to be investigated and the *tree pruning* to eliminate unpromising paths from the search tree. In the pre-selection stage, we choose a small number of promising columns in the sensing matrix using a conventional greedy algorithm. If we denote the set of column indices obtained in the pre-selection stage as Θ, then we set $N \gg |\Theta| > K$ where K is the sparsity of the input vector ($\|\mathbf{x}\|_0 = \mathbf{K}$). Once the pre-selection is finished, a tree search is performed to find the best estimate of support. In the construction of the tree, we only use elements in the pre-selected set Θ so that the number of paths in the tree is reduced from $\binom{N}{K}$ to $\binom{|\Theta|}{K}$. Further, in order to alleviate computational burden in the main tree search, we introduce an aggressive tree pruning strategy by which unpromising paths are promptly discarded from the tree.

In our analysis, we show that TSMP accurately identifies the support of K-sparse signal and hence reconstructs the original sparse signal accurately in the noiseless setting if the sensing matrix satisfies the property so called restricted isometry property (RIP) (Theorem 11). In the noisy setting, we show that an accurate identification of support is possible if the signal power is sufficiently larger than the noise power (Theorem 20). We confirm by numerical simulations that TSMP performs close to the best estimator, called Oracle estimator [18][1], in the high SNR regime.

The rest of this paper is organized as follows. In Section II, we discuss basics of tree search and then introduce the TSMP

Manuscript received August 16, 2016; approved for publication by Dr. Saewoong Bahk, Co-Editor-in-Chief, August 18, 2016.

This work was partly supported by the ICT R&D program of MSIP/IITP (No.B0126-16-1017, Spectrum Sensing and Future Radio Communication Platforms) and the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIP) (NRF-20161A2B3015576).

J. Lee is with Dept. of Information and Communication Engineering, Daegu Gyeongbuk Institute of Science and Technology, Daegu, Korea, email: jayslee@dgist.ac.kr.

J. Choi is with Dept. of Electrical Engineering, Hanyang University, Seoul, Korea, email: junwchoi@hanyang.ac.kr.

B. Shim is with Institute of New Media and Communications and School of Electrical and Computer Engineering, Seoul National University, Seoul, Korea, email: bshim@snu.ac.kr.

[1]The estimator that has a prior knowledge on the support (which component of the sparse vector is zero or not) is often called *Oracle estimator*.

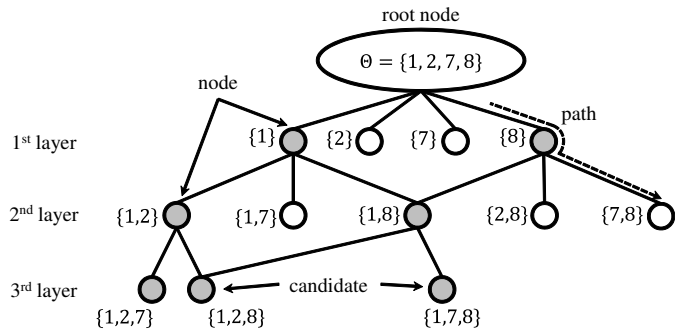


Fig. 1. Illustration of basic tree search for $K = 3$. If $T = \{1, 2, 7\}$, then six paths ($\{1\}$, $\{2\}$, $\{7\}$, $\{1, 2\}$, $\{1, 7\}$ and $\{1, 2, 7\}$) correspond to the true path.

algorithm. In Section III, we analyze the recovery condition of TSMP under which the support is accurately identified in the noiseless and noisy scenarios. In Section IV, we provide the numerical results and conclude the paper in Section V.

II. TREE SEARCH MATCHING PURSUIT (TSMP)

The proposed TSMP algorithm consists of two operational steps: pre-selection and tree search. We first describe the basic definition and operation of the tree search and then discuss the pre-selection operation and the greedy tree search.

A. Tree Search - Basic Definitions

Before we proceed to the details of tree search, we briefly explain the basic definitions and operation of the tree search.

- *Tree* is defined as a collection of nodes where each *node* (depicted as a circle in Fig. 1) represents an index chosen by the tree search. In our problem, the height of the tree is K and the tree search operation starts from the root node and traces down to the bottom of the tree. The *layer* is defined as the (vertical) distance from the root node. Denoting the root node as the 0-th layer, nodes below the root node correspond to the nodes in the first layer and those in the bottom of the tree correspond to the nodes in the K-th layer.
- A sequence of nodes starting from the root node to the given node is called a *path* and denoted as $\hat{S}$. To be specific, path in the i-th layer contains indices from the first to the i-th layer and hence expressed as $\hat{S}^i = \{s_1, s_2, \cdots, s_i\}$ where s_j is the index chosen in the j-th layer. In particular, a path from the root node to the bottom of the tree (i.e., path containing K-indices) is called the *candidate* (see Fig. 1). In the tree search, indices chosen from the pre-selected index set Θ are used exclusively. Also, already chosen indices in a path would not be re-selected. Therefore, the index s_{i+1} in the $i+1$-th layer is chosen from the set $\Theta \backslash \hat{S}^i$.
- We say a path is the *true path* if all elements of the path is from the support (i.e., $\hat{S}^i \in T$).

B. Pre-selection: A First Stage Pruning

The purpose of the pre-selection is to estimate indices that are highly likely to be the elements of support T. Alternatively put, we do our best guess to choose columns of sensing matrix $\mathbf{\Phi}$ that are associated with nonzero elements in the sparse vector. If we denote the set of indices as Θ, then the search set is reduced from $\Omega = \{1, 2, \cdots, N\}$ to Θ, a small subset of Ω. In the construction of the tree, we use elements of the pre-selected set Θ exclusively so that we can limit the number of paths in the tree and eventually reduce the search complexity. In obtaining the pre-selection set Θ, one can basically use any sparse recovery algorithm returning more than K indices. One simple option is to use the OMP algorithm running more than K-iterations [19] or the generalized OMP (gOMP) algorithm [20]. In this work, we use a simple modification of OMP returning $2K - 1$ indices. Specifically, this approach chooses K indices in the first iteration and one index for the rest. While this option does not make a big difference in empirical performance over alternative choices (e.g., gOMP choosing 2 indices per iteration), it simplifies the performance guarantee analysis (see Section III-A).

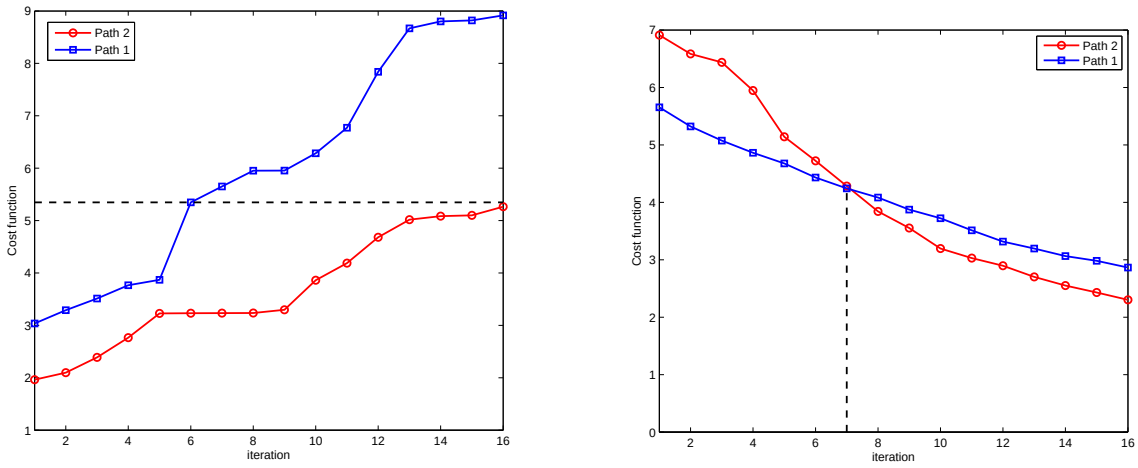
C. Tree Search

In the setting we described, the tree has a maximum depth K, and the goal is to find a candidate S^* having the smallest cost function $J(S^*) = \min_S \|\mathbf{y} - \mathbf{\Phi_S}\hat{\mathbf{x}}_\mathbf{S}\|_\mathbf{2}$ among all possible candidates (paths with cardinality K). Since we choose candidates from the pre-selected set, the number of all possible candidates in the tree search is $\binom{|\Theta|}{K}$. Although this number is better than brute-force enumeration, searching all of these to identify the best candidate is still burdensome.

In order to deal with the problem at hand, we introduce an aggressive pruning strategy discarding part of the tree for which a chance of finding the optimal solution is very small. This pruning decision is done by comparing the cost function of the path against a properly chosen threshold. That is, if the cost function of a path is greater than the threshold, the path has little hope to survive and thus eliminated from the tree immediately. It is worth mentioning that in contrast to typical tree search problems, it is not easy, and in fact not possible, to decide the pruning of a path using only $\hat{S}^i$ (i.e., indices of already visited nodes). We note that in many tree search problems, the cost function of a path *increases monotonically* with the iteration (e.g., Viterbi algorithm for maximum-likelihood detection) [21], [22], [23]. Therefore, if a path whose partial cost function generated by the contribution of already visited nodes exceeds the cost function of the candidate, the path under investigation cannot be the solution of the problem and hence is pruned immediately from the tree (see Fig. 2(a)). This type of strategy, unfortunately, cannot be applied to our tree search problem since the partial cost function $J(\hat{S}^i)$ of a path $\hat{S}^i$, which corresponds to the magnitude of the residual, is a *monotonic decreasing* function of the iteration[2].

From the discussion thus far, it is clear that we need to compute the cost function of a *candidate* to make an effective pruning decision, and hence, we need to somehow estimate indices of unvisited nodes. In order to distinguish the index set of unvisited node from that of already visited nodes, in the sequel we call it as *noncausal index set*. This noncausal index set $\{\hat{s}_{i+1}, \hat{s}_{i+2}, \cdots, \hat{s}_K\}$ is temporarily needed for the pruning decision purpose and can be easily obtained by choosing $K - i$

[2]If $\hat{S}^i \subset \hat{S}^{i+1}$, then $\|\mathbf{r_{\hat{S}^i}}\|_\mathbf{2} \geq \|\mathbf{r_{\hat{S}^{i+1}}}\|_\mathbf{2}$.



(a) Conventional tree search. (b) Proposed tree search.

Fig. 2. Comparison of cost function between the conventional tree search and proposed tree search. In the conventional tree search such as Viterbi algorithm, if the cost function of path 1 exceeds the final cost function of path 2, then the path 1 cannot be the optimal solution and hence it is removed immediately. We cannot expect this feature in the proposed tree search since the cost function is a monotonic decreasing function of the iteration.

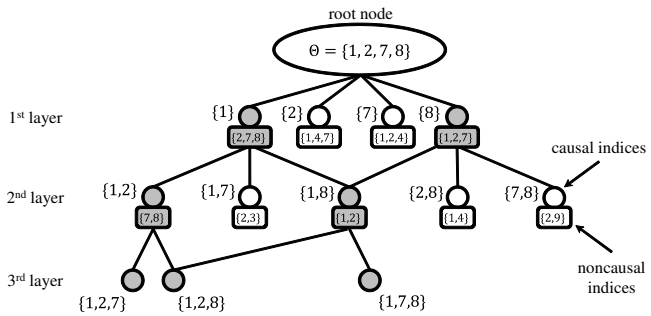


Fig. 3. Detailed operation of the TSMP algorithm. White nodes are pruned from the tree search since they satisfy the pruning criterion.

promising indices of columns in $\Omega \setminus \hat{S}^i$. When obtaining the noncausal index set, we can choose $K-i$ indices which are highly likely to be the support among the elements of $\Theta \setminus \hat{S}^i$. To this end, we choose $K-i$ indices $\{\hat{s}_{i+1}, \cdots, \hat{s}_K\}$ minimizing the magnitude of a residual:

$$(\hat{s}_{i+1}, \cdots, \hat{s}_K) = \arg \min_{\substack{s \subset \Omega \setminus \hat{S}^i, \\ |s|=K-i}} \|\mathbf{r}_{\hat{\mathbf{s}}^\mathbf{i}} - \mathbf{\Phi}_\mathbf{s} \hat{\mathbf{x}}_\mathbf{s}\|_2 \qquad (2)$$

where

$$\begin{aligned} \mathbf{r}_{\hat{\mathbf{s}}^\mathbf{i}} &= \mathbf{y} - \mathbf{\Phi}_{\hat{\mathbf{s}}^\mathbf{i}} \hat{\mathbf{x}}_{\hat{\mathbf{s}}^\mathbf{i}}, \\ \hat{\mathbf{x}}_s &= \mathbf{\Phi}_\mathbf{s}^\dagger \mathbf{r}_{\hat{\mathbf{s}}^\mathbf{i}}. \end{aligned}$$

Once *roughly* estimated candidate $\bar{S} = \hat{S}^i \cup \{\hat{s}_{i+1}, \cdots, \hat{s}_K\}$ is obtained, we compute the residual $\mathbf{r}_{\bar{\mathbf{s}}} = \mathbf{y} - \mathbf{\Phi}_{\bar{\mathbf{s}}} \hat{\mathbf{x}}_{\bar{\mathbf{s}}}$ ($\hat{\mathbf{x}}_{\bar{s}} = \mathbf{\Phi}_{\bar{\mathbf{s}}}^\dagger \mathbf{y}$) to decide the pruning of this path. To be specific, if the ℓ_2-norm of the residual is greater than the threshold ϵ (i.e., $\|\mathbf{r}_{\bar{\mathbf{s}}}\|_\mathbf{2} > \epsilon$), then the path has little hope to survive and is discarded from the tree (see Fig. 3). Initially, the pruning threshold

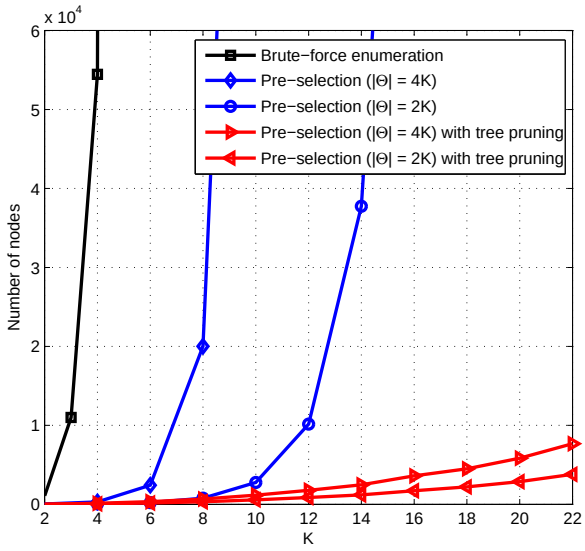


Fig. 4. The number of nodes visited of the tree search as a function of K. In the tree pruning, nodes corresponding to noncausal path are also considered.

ϵ is set to a large number[3] and whenever the search of a layer is finished, updated to the minimum ℓ_2-norm of the residual among all survived paths ($\epsilon = \min \|\mathbf{r}_{\bar{\mathbf{s}}}\|_\mathbf{2}$). We summarize the proposed TSMP algorithm in Table 1.

III. RECOVERY CONDITION ANALYSIS

In this section, we analyze the recovery condition under which TSMP can accurately identify K-sparse signals in noise-

[3]There is no issue in the choice of ϵ in the noiseless scenario since $\|\mathbf{r}_\mathbf{T}\|_\mathbf{2} = \mathbf{0}$ is smaller than any positive value of ϵ. In the noisy scenario, however, it is possible that the proposed algorithm fails if the initial value of ϵ is too small ($\epsilon < \min_{\bar{S}} \|\mathbf{r}_{\bar{\mathbf{S}}}\|_\mathbf{2}$) since in this case no path will be survived.

Table 1. The TSMP algorithm

Input: measurement $\mathbf{y}$, sensing matrix $\mathbf{\Phi}$, sparsity K, initial threshold ϵ_1
Output: Estimated signal $\hat{\mathbf{x}}$
Initialization: $i := 0, W^0 := \emptyset$

$\Theta = f_{\text{preselection}}(\mathbf{y}, \mathbf{\Phi}, \mathbf{p})$ (*preselection*)
while $i < K$ **do**
  $i := i + 1, W^i := \emptyset, \epsilon_{i+1} := \epsilon_i$
  for $l = 1$ **to** $|W^{i-1}|$ **do**
   $\theta := \Theta \setminus \hat{S}^{i-1}(l)$
   for $j = 1$ **to** $|\theta|$ **do**
    $\hat{S}^i := \hat{S}^{i-1}(l) \cup \{s_i(j)\}$ (*update j-th path*)
    if $\hat{S}^i \notin W^i$ **then** (*check duplicated path*)
     $(\hat{s}_{i+1}, \cdots, \hat{s}_K)$
     $= \arg \min_{\substack{s \subset \Omega \setminus \hat{S}^i, \\ |s| = K-i}} \|\mathbf{r}_{\hat{\mathbf{s}}^\mathbf{i}} - \mathbf{\Phi}_\mathbf{s}\hat{\mathbf{x}}_\mathbf{s}\|_\mathbf{2}$ (*support estimation*)
     $\bar{S} = \hat{S}^i \cup \{\hat{s}_{i+1}, \cdots, \hat{s}_K\}$
     $\mathbf{r}_{\bar{\mathbf{s}}} = \mathbf{P}^\perp_{\bar{\mathbf{s}}}\mathbf{y}$
     if $\|\mathbf{r}_{\bar{\mathbf{s}}}\|_\mathbf{2} \le \epsilon_\mathbf{i}$ **then** (*pruning decision*)
      $W^i := W^i \cup \hat{S}^i, I^* := \bar{S}$
      if $\|\mathbf{r}_{\mathbf{I}^*}\|_\mathbf{2} = \mathbf{0}$ **then**
       $j = |\theta| + 1, l = |W^{i-1}| + 1$
       $i = K + 1$
      end if
      if $\|\mathbf{r}_{\mathbf{I}^*}\|_\mathbf{2} \le \epsilon_{\mathbf{i+1}}$ **then**
       $\epsilon_{i+1} := \|\mathbf{r}_{\mathbf{I}^*}\|_2$
      end if
     end if
    end if
   end for
  end for
end while
return $\hat{\mathbf{x}}^* = \mathbf{\Phi}^\dagger_{\mathbf{I}^*}\mathbf{y}$ (*signal reconstruction*)

The $f_{\text{preselection}}(\cdot)$ is a function to choose multiple promising indices (see Section II-B).

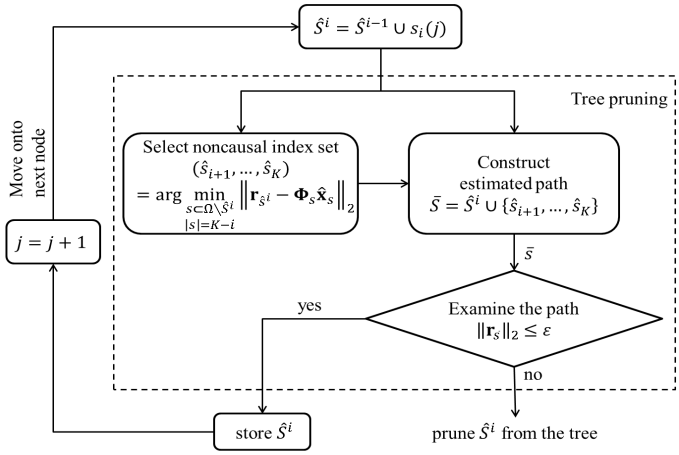


Fig. 5. The pruning operation of TSMP in the i-th layer where j is the index of column in $\Theta \setminus \hat{S}^{i-1}$. Note that TSMP investigates each path $\hat{S}^i$ and performs the pruning of the path if $\|\mathbf{r}_{\bar{\mathbf{s}}}\|_\mathbf{2} > \epsilon$.

less and noisy scenarios. In our analysis, we use the restricted isometry property (RIP) of the sensing matrix.

Definition 1: The sensing matrix $\mathbf{\Phi}$ is said to satisfy the RIP of order K if there exists a constant $\delta(\mathbf{\Phi}) \in (\mathbf{0}, \mathbf{1})$ such that

$$(1 - \delta(\mathbf{\Phi}))\|\mathbf{x}\|_\mathbf{2}^\mathbf{2} \le \|\mathbf{\Phi x}\|_\mathbf{2}^\mathbf{2} \le (\mathbf{1} + \delta(\mathbf{\Phi}))\|\mathbf{x}\|_\mathbf{2}^\mathbf{2}$$

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

In particular, the minimum of all constants $\delta(\mathbf{\Phi})$ satisfying Definition 1 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 brevity.

The following lemmas are useful in our analysis.

Lemma 2: *(Monotonicity of RIP [2]):* If the sensing matrix $\mathbf{\Phi}$ satisfies the RIP of both orders K_1 and K_2, then $\delta_{K_1} < \delta_{K_2}$ for any $K_1 < K_2$.

Lemma 3: *(Consequences of RIP [2], [9]):* If $0 < \delta_{|I|} < 1$ exists for $I \subset \Omega$, then for any vector $\mathbf{x} \in \mathbb{R}^{|\mathbb{I}|}$,

$$\left(1 - \delta_{|I|}\right)\|\mathbf{x}\|_\mathbf{2} \le \|\mathbf{\Phi}'_\mathbf{I}\mathbf{\Phi}_\mathbf{I}\mathbf{x}\|_\mathbf{2} \le \left(\mathbf{1} + \delta_{|\mathbf{I}|}\right)\|\mathbf{x}\|_\mathbf{2},$$
$$\frac{1}{1 + \delta_{|I|}}\|\mathbf{x}\|_\mathbf{2} \le \|\left(\mathbf{\Phi}'_\mathbf{I}\mathbf{\Phi}_\mathbf{I}\right)^{-\mathbf{1}}\mathbf{x}\|_\mathbf{2} \le \frac{\mathbf{1}}{\mathbf{1} - \delta_{|\mathbf{I}|}}\|\mathbf{x}\|_\mathbf{2}.$$

Lemma 4: *(Lemma 2.1 in [4]):* Let $I_1, I_2 \subset \Omega$ and $I_1 \cap I_2 = \emptyset$. If $0 < \delta_{|I_1|+|I_2|} < 1$ exists, then

$$\|\mathbf{\Phi}'_{\mathbf{I_1}}\mathbf{\Phi}_{\mathbf{I_2}}\mathbf{x}\|_\mathbf{2} \le \delta_{|\mathbf{I_1}|+|\mathbf{I_2}|}\|\mathbf{x}\|_\mathbf{2}.$$

A. Exact Recovery from Noiseless Measurements

In this subsection, we analyze a condition ensuring that TSMP recovers the original sparse signal accurately from the noiseless measurements. As mentioned, TSMP consists of *preselection* and *tree search*. In order to guarantee an accurate identification of the support, TSMP should satisfy the following two conditions:

C-1) At least one support index should be selected in the pre-selection process (i.e., $T \cap \Theta \neq \emptyset$).
C-2) At least one true path[4] should be survived in the tree pruning process.

Before we proceed, we provide some useful definitions in our analysis. Let κ be the largest correlation in magnitude between the observation $\mathbf{y}$ and the columns associated with *correct* indices. That is,

$$\kappa = \max_{j \in T} |\phi'_j \mathbf{y}|.$$

Further, let ζ be the K-th largest correlation in magnitude between $\mathbf{y}$ and the columns corresponding to *incorrect* indices. That is,

$$\zeta = \min_{j \in I_K} |\phi'_j \mathbf{y}|.$$

where $I_K = \arg \max_{|I|=K, I \subset T^c} \|\mathbf{\Phi}'_{\mathbf{I_K}}\mathbf{y}\|_\mathbf{2}$.

The following lemmas provide the lower and the upper bounds of κ and ζ, respectively.

[4]If $\hat{S}^i$ is a true path, it contains indices only in T ($\hat{S}^i \subset T$).

Lemma 5: κ satisfies

$$\kappa \geq \frac{1-\delta_K}{\sqrt{K}} \left\|\mathbf{x_T}\right\|_2. \tag{3}$$

Proof: See Appendix A.

Lemma 6: ζ satisfies

$$\zeta \leq \frac{\delta_{2K}}{\sqrt{K}} \left\|\mathbf{x_T}\right\|_2. \tag{4}$$

Proof: See Appendix B.

Recall that the modified OMP algorithm used for the pre-selection chooses K indices in the first iteration. Thus, if at least one correct index is chosen in the first iteration, then more than one correct index will be chosen in the pre-selection process. In order to to choose at least one correct index in the first iteration, we should have $\kappa > \zeta$.

Theorem 7 (Sufficient condition for C-1)) At least one element in the support is found in the first iteration of the pre-selection stage under

$$\delta_{2K} < \frac{1}{2}. \tag{5}$$

Proof: From definitions of κ and ζ, we observe that C-1) condition is satisfied if

$$\kappa > \zeta. \tag{6}$$

Also, using Lemma 5 and 6, we see that (6) is guaranteed under

$$\frac{1-\delta_K}{\sqrt{K}} \left\|\mathbf{x_T}\right\|_2 > \frac{\delta_{2K}}{\sqrt{K}} \left\|\mathbf{x_T}\right\|_2 \tag{7}$$

and hence $\delta_{2K} + \delta_K < 1$. Using Lemma 2 (monotonicity of RIP), we obtain the desired result.

We next analyze the condition guaranteeing that the candidate $\bar{S}$ of the tree search equals the support T. This condition is satisfied if at least one true path is survived in each layer and further a true index is added to this path. Note that for a given true path $\hat{S}^i \subset T$, the problem to choose the noncausal index set is equivalent to the problem of reconstructing $(K-i)$-sparse signal from the measurement $\mathbf{r_{\hat{S}^i}}$. Before we proceed, we provide definitions useful in our analysis. Let Ψ_l be the combination of $\hat{S}^i$ and l indices chosen in noncausal index set search. That is, $\Psi_l = \hat{S}^i \cup \{\hat{s}_{i+1}, \cdots, \hat{s}_{i+l}\}$. Next, let λ^l be the largest correlation in magnitude between the residual $\mathbf{r_{\Psi_1}}$ and columns associated with *correct* indices:

$$\lambda^l = \max_{j \in T \setminus \Psi_l} |\phi_j' \mathbf{r_{\Psi_1}}|.$$

Further, let γ^l be the $(K-i)$-th largest correlation in magnitude between $\mathbf{r_{\Psi_1}}$ and columns associated with *incorrect* indices:

$$\gamma^l = \min_{j \in D} |\phi_j' \mathbf{r_{\Psi_1}}|$$

where $D = \arg \max_{|I|=K-i, I \subset \Omega \setminus T} \|\mathbf{\Phi_I'} \mathbf{r_{\Psi_1}}\|_\mathbf{2}$.

In the following lemmas, we provide a lower bound of λ^l and an upper bound of γ^l.

Lemma 8: Suppose a path Ψ_l is contained in T (i.e., $\Psi_l \subset T$), then

$$\lambda^l \geq \frac{1-2\delta_K}{\sqrt{K-1}(1-\delta_K)} \|\mathbf{x_{T \setminus \Psi_1}}\|_\mathbf{2}. \tag{8}$$

Proof: See Appendix C.

Lemma 9: Suppose a path Ψ_l is contained in T, then

$$\gamma^l \leq \frac{\delta_{2K-1}}{\sqrt{K-1}(1-\delta_K)} \|\mathbf{x_{T \setminus \Psi_1}}\|_\mathbf{2}. \tag{9}$$

Proof: See Appendix D.

As mentioned, in order to recover the original sparse signals accurately, at least one true path should be survived in each layer. In other words, when a path $\hat{S}^i$ is contained in T ($\hat{S}^i \subset T$), then the noncausal index set should also be contained in T (i.e., $\{\hat{s}_{i+1}, \cdots, \hat{s}_K\} = T \setminus \hat{S}^i$) and further this candidate should not be pruned. That is,

$$\|\mathbf{r_{\hat{S}^i \cup \{\hat{s}_{i+1}, \cdots, \hat{s}_K\}}}\| = \|\mathbf{r_T}\| < \epsilon. \tag{10}$$

Since $\|\mathbf{r_T}\|_\mathbf{2} = \mathbf{0}$ for the noiseless scenario, the condition (10) always holds for any positive ϵ. Thus, what we essentially need is a condition ensuring that the noncausal index set is contained in T (i.e., $\{\hat{s}_{i+1}, \cdots, \hat{s}_K\} \subset T$).

Theorem 10: If a path $\hat{S}^i$ contains correct indices only, then the noncausal index set of TSMP consist only of correct ones under

$$\delta_{2K-1} < \frac{1}{3}. \tag{11}$$

for any $1 \leq i \leq K-1$. In other words, $\bar{S} = \hat{S}^i \cup \{\hat{s}_{i+1}, \cdots, \hat{s}_K\} = T$ under $\delta_{2K-1} < \frac{1}{3}$. *Proof*: To satisfy the hypothesis, among $K-i$ indices of columns maximally correlated with $\mathbf{r_{\Psi_1}}$, at least one should be an element of the support. From the definitions of λ^l and γ^l, it is clear that this is satisfied if

$$\lambda^l > \gamma^l. \tag{12}$$

Using Lemma 8 and 9, we can easily show that (12) is guaranteed under $2\delta_K + \delta_{2K-1} < 1$. Finally, using Lemma 2, we obtain the desired result.

If Theorem 7 and 10 are jointly satisfied, $\{\hat{s}_{i+1}, \cdots, \hat{s}_K\} \subset T$ for any true path $\hat{S}^i$ ($\hat{S}^i \subset T$) and $\bar{S} = T$. Therefore, $\hat{S}^i$ will not be pruned from the tree (we recall that $\|\mathbf{r_{\bar{S}}}\|_\mathbf{2} = \mathbf{0} < \epsilon$ for any positive ϵ). Overall recovery condition of TSMP for the noiseless scenario can be obtained by combining Theorem 7 and 10.

Theorem 11 (Recovery condition of TSMP) The TSMP algorithm identifies the support of any K-sparse signal from $\mathbf{y} = \mathbf{\Phi x}$ accurately if the sensing matrix $\mathbf{\Phi}$ satisfies the RIP with

$$\delta_{2K} < \frac{1}{3}. \tag{13}$$

Proof: The condition (13) is obtained by choosing stricter condition between Theorem 7 and 10.

When compared to the conditions of other greedy approaches such as the OMP algorithm ($\delta_{K+1} < \frac{1}{\sqrt{K}+1}$) [7], [8], the CoSaMP algorithm ($\delta_{4K} < 0.1$) [9], the SP algorithm ($\delta_{3K} < 0.165$) [10], the IHT algorithm ($\delta_{3K} < 0.175$) [11], and the gOMP algorithm ($\delta_{LK} < \frac{\sqrt{L}}{\sqrt{L}+2\sqrt{K}}$) [20], we observe that TSMP provides more relaxed recovery condition.

B. Reconstruction from Noisy Measurements

Now, we turn to the noisy scenario and analyze the condition of TSMP to accurately identify the support in the presence of noise. In this scenario, the measurement $\mathbf{y}$ is expressed as $\mathbf{y} = \mathbf{\Phi x} + \mathbf{v}$ where $\mathbf{v}$ is an additive noise vector. Even though details are a bit complicated, main architecture of the proof is reminiscent of the argument in the noiseless scenario. In fact, two requirements of TSMP to identify the support are 1) at least one support element should be chosen in the pre-selection process (i.e., $T \cap \Omega \neq \emptyset$), and 2) true path ($\bar{S} = T$) should be survived in the pruning process. Before we proceed, we provide useful definitions in our analysis. First, let ρ be the largest correlation (in magnitude) between the observation $\mathbf{y}$ and the columns associated with true indices. That is,

$$\rho = \max_{j \in T} |\phi_j' \mathbf{y}|.$$

Next, let η be the K-th largest correlation (in magnitude) between the observation $\mathbf{y}$ and the columns associated with incorrect indices. Then η is expressed as

$$\eta = \min_{j \in I_K} |\phi_j' \mathbf{y}|.$$

where $I_K = \arg \max_{|I|=K, I \subset T^c} \|\mathbf{\Phi_I' r_\Lambda}\|_\mathbf{2}$.

In the following lemmas, we provide the lower bound of ρ and the upper bound of η.

Lemma 12: ρ satisfies

$$\rho \geq \frac{1}{\sqrt{K}} \left[(1 - \delta_K) \|\mathbf{x_T}\|_\mathbf{2} - \sqrt{\mathbf{1} + \delta_\mathbf{K}} \|\mathbf{v}\|_\mathbf{2} \right] \tag{14}$$

Proof: See Appendix E.

Lemma 13: η satisfies

$$\eta \leq \frac{1}{\sqrt{K}} \left[\delta_{2K} \|\mathbf{x_T}\|_\mathbf{2} + \sqrt{\mathbf{1} + \delta_\mathbf{K}} \|\mathbf{v}\|_\mathbf{2} \right]. \tag{15}$$

Proof: See Appendix F.

The following theorem provides the condition ensuring that at least one support element is identified by the pre-selection stage.

Theorem 14 (Pre-selection condition) At least one element in the support is found in the pre-selection stage if the nonzero coefficients of the original sparse signal $\mathbf{x}$ satisfy

$$\min_{j \in T} |x_j| > \frac{2\sqrt{1 + \delta_K}}{1 - \delta_K - \delta_{2K}} \|\mathbf{v}\|_\mathbf{2}. \tag{16}$$

Proof: From definitions, it is clear that at least one true index is chosen in the first iteration if

$$\rho > \eta. \tag{17}$$

Using the lower bound of ρ and the upper bound of η, we obtain the sufficient condition of (17) as

$$\begin{aligned} &\frac{1}{\sqrt{K}} \left[(1 - \delta_K) \|\mathbf{x_T}\|_\mathbf{2} - \sqrt{\mathbf{1} + \delta_\mathbf{K}} \|\mathbf{v}\|_\mathbf{2} \right] \\ &\quad > \frac{1}{\sqrt{K}} \left[\delta_{2K} \|\mathbf{x_T}\|_\mathbf{2} + \sqrt{\mathbf{1} + \delta_\mathbf{K}} \|\mathbf{v}\|_\mathbf{2} \right]. \end{aligned} \tag{18}$$

After some manipulations, we have

$$\|\mathbf{x_T}\|_\mathbf{2} > \frac{\mathbf{2}\sqrt{\mathbf{1} + \delta_\mathbf{K}}}{\mathbf{1} - \delta_\mathbf{K} - \delta_\mathbf{2K}} \|\mathbf{v}\|_\mathbf{2}. \tag{19}$$

Since $\|\mathbf{x_T}\|_\mathbf{2} \geq \min_{\mathbf{j} \in \mathbf{T}} |\mathbf{x_j}|$, (19) is guaranteed under

$$\min_{j \in T} |x_j| > \frac{2\sqrt{1 + \delta_K}}{1 - \delta_K - \delta_{2K}} \|\mathbf{v}\|_\mathbf{2}, \tag{20}$$

which completes the proof.

We next analyze the condition under which the true path is survived by the tree pruning stage. In order to meet this requirement, 1) under the condition that the path under consideration is true ($\hat{S}^i \subset T$), corresponding noncausal index set should also be true ($\{\hat{s}_{i+1}, \cdots, \hat{s}_K\} \subset T$) and further 2) this true path should not be removed by the tree pruning (i.e., if $\bar{S} = T$, then $\|\mathbf{r_{\bar{S}}}\|_\mathbf{2} < \epsilon$). Before we proceed, we introduce two useful definitions in our analysis. Similar to the analysis in the noiseless setting, we denote the combination of path $\hat{S}^i$ and l indices chosen in noncausal index set search as Ψ_l (i.e., $\Psi_l = \hat{S}^i \cup \{\hat{s}_{i+1}, \cdots, \hat{s}_{i+l}\}$). Also, we let β^l be the largest correlation in magnitude between ϕ_j ($j \in T \setminus \Psi_l$) and $\mathbf{r_{\Psi_l}}$:

$$\beta^l = \arg \max_{j \in T \setminus \Psi_l} |\phi_j' \mathbf{r_{\Psi_l}}|,$$

and let α^l be the $(K - i)$-th largest correlation in magnitude between $\mathbf{r_{\Psi_l}}$ and columns associated with incorrect indices:

$$\alpha^l = \arg \min_{j \in D} |\phi_j' \mathbf{r_{\Psi_l}}|$$

where $D = \arg \max_{|I|=K-i, I \subset \Omega \setminus T} \|\mathbf{\Phi_I' r_{\Psi_l}}\|_\mathbf{2}$. The following two lemmas provide the lower and upper bounds of β^l and α^l, respectively.

Lemma 15: If Ψ_l contains true indices exclusively, then β^l satisfies

$$\beta^l > \frac{1}{\sqrt{K - 1}} \left[\frac{1 - 2\delta_K}{1 - \delta_K} \|\mathbf{x_{T \setminus \Psi_l}}\|_\mathbf{2} - \sqrt{\mathbf{1} + \delta_\mathbf{K-1}} \|\mathbf{v}\|_\mathbf{2} \right]. \tag{21}$$

Proof: See Appendix G.

Lemma 16: If Ψ_l contains true indices exclusively, then α^l satisfies

$$\alpha^l \leq \frac{1}{\sqrt{K - 1}} \left[\frac{\delta_{2K-1}}{1 - \delta_K} \|\mathbf{x_{T \setminus \Psi_l}}\|_\mathbf{2} + \sqrt{\mathbf{1} + \delta_\mathbf{K-1}} \|\mathbf{v}\|_\mathbf{2} \right]. \tag{22}$$

Proof: See Appendix H.

Using these lemmas, we can identify the condition guaranteeing that the noncausal index set generated from the true path ($\hat{S}^i \subset T$) is also true.

Theorem 17: Suppose a path $\hat{S}^i$ consists of true indices exclusively (i.e., $\hat{S}^i \subset T$), then the noncausal index set also contains the true ones ($\{\hat{s}_{i+1}, \cdots, \hat{s}_K\} = T \setminus \hat{S}^i$) under

$$\min_{j \in T} |x_j| > \frac{2(1 - \delta_K)\sqrt{1 + \delta_{K-1}}}{1 - 2\delta_K - \delta_{2K-1}} \|\mathbf{v}\|_\mathbf{2}. \tag{23}$$

Proof: Similar to Theorem 10, one can observe that the non-causal index set of given true path $\hat{S}^i$ contains only true indices, that is, $\{\hat{s}_{i+1}, \cdots, \hat{s}_K\} = T \setminus \hat{S}^i$ if

$$\beta^l > \alpha^l. \tag{24}$$

Using Lemma 15 and 16, we obtain the sufficient condition of (24) as

$$\frac{1}{\sqrt{K-1}}\left[\frac{1-2\delta_K}{1-\delta_K}\|\mathbf{x_{T\setminus\Psi_l}}\|_\mathbf{2} - \sqrt{\mathbf{1+\delta_{K-1}}}\|\mathbf{v}\|_\mathbf{2}\right] > \frac{1}{\sqrt{K-1}}\left[\frac{\delta_{2K-1}}{1-\delta_K}\|\mathbf{x_{T\setminus\Psi_l}}\|_\mathbf{2} + \sqrt{\mathbf{1+\delta_{K-1}}}\|\mathbf{v}\|_\mathbf{2}\right] \tag{25}$$

where $\Psi_l \subset T$. After some manipulations, we have

$$\|\mathbf{x_{T\setminus\Psi_l}}\|_\mathbf{2} > \frac{\mathbf{2(1-\delta_K)}\sqrt{\mathbf{1+\delta_{K-1}}}}{\mathbf{1-2\delta_K-\delta_{2K-1}}}\|\mathbf{v}\|_\mathbf{2}. \tag{26}$$

Since $\|\mathbf{x_{T\setminus\hat{S}^i}}\|_\mathbf{2} \geq \min_{\mathbf{j}\in\mathbf{T}}|\mathbf{x_j}|$, we get the desired result.

Next, we turn to the analysis of the condition under which the magnitude of $\mathbf{r_T}$ becomes the minimum among all combinations of K indices.

Theorem 18: The candidate whose residual is minimum (in magnitude) becomes the support if

$$\min_{j\in T}|x_j| > \frac{2(1-\delta_K)}{1-3\delta_{2K}}\|\mathbf{v}\|_\mathbf{2}. \tag{27}$$

In other words, $\|\mathbf{r_T}\|_\mathbf{2} < \|\mathbf{r_{\bar{S}}}\|_\mathbf{2}$ for any $\bar{S} \neq T$ under (27).

Proof: One can notice that the hypothesis is satisfied if the upper bound of $\|\mathbf{r_T}\|_\mathbf{2}$ is smaller than the lower bound of $\|\mathbf{r_{\bar{S}}}\|_\mathbf{2}$. First, we obtain the upper bound of $\|\mathbf{r_T}\|_\mathbf{2}$ as

$$\begin{aligned}\|\mathbf{r_T}\|_\mathbf{2} &= \|\mathbf{P_T^\perp y}\|_\mathbf{2} \\ &= \|\mathbf{P_T^\perp}\left(\mathbf{\Phi_T x_T + v}\right)\|_\mathbf{2} \\ &= \|\mathbf{P_T^\perp v}\|_\mathbf{2} \\ &\leq \|\mathbf{v}\|_\mathbf{2}\end{aligned} \tag{28}$$

where $\mathbf{P_T^\perp} = \mathbf{I} - \mathbf{\Phi_T}\left(\mathbf{\Phi_T'\Phi_T}\right)^{-\mathbf{1}}\mathbf{\Phi_T'}$ is the projection onto the orthogonal complement of T and (28) is because $\mathbf{P_T^\perp \Phi_T x_T} = \mathbf{0}$.

Next, we obtain the lower bound of $\|\mathbf{r_{\bar{S}}}\|_\mathbf{2}$. For any $\bar{S} \neq T$, we have

$$\begin{aligned}\|\mathbf{r_{\bar{S}}}\|_\mathbf{2} &= \|\mathbf{P_{\bar{S}}^\perp y}\|_\mathbf{2} \\ &= \|\mathbf{P_{\bar{S}}^\perp}(\mathbf{\Phi x + v})\|_\mathbf{2} \\ &= \|\mathbf{P_{\bar{S}}^\perp}(\mathbf{\Phi_{\bar{S}} x_{\bar{S}} + \Phi_{T\setminus\bar{S}} x_{T\setminus\bar{S}} + v})\|_\mathbf{2} \\ &= \|\mathbf{P_{\bar{S}}^\perp}(\mathbf{\Phi_{T\setminus\bar{S}} x_{T\setminus\bar{S}} + v})\|_\mathbf{2} \qquad (29) \\ &\geq \|\mathbf{P_{\bar{S}}^\perp \Phi_{T\setminus\bar{S}} x_{T\setminus\bar{S}}}\|_\mathbf{2} - \|\mathbf{P_{\bar{S}}^\perp v}\|_\mathbf{2} \qquad (30)\end{aligned}$$

where (29) is because $\mathbf{P_{\bar{S}}^\perp \Phi_{\bar{S}} x_{\bar{S}}} = \mathbf{0}$ and (30) is from the triangle inequality. The first term in the right-hand side of (30) is lower bounded as

$$\begin{aligned}\|\mathbf{P_{\bar{S}}^\perp \Phi_{T\setminus\bar{S}} x_{T\setminus\bar{S}}}\|_\mathbf{2} &= \|(\mathbf{I} - \mathbf{\Phi_{\bar{S}}}(\mathbf{\Phi_{\bar{S}}'\Phi_{\bar{S}}})^{-\mathbf{1}}\mathbf{\Phi_{\bar{S}}'})\mathbf{\Phi_{T\setminus\bar{S}} x_{T\setminus\bar{S}}}\|_\mathbf{2} \\ &\geq \|\mathbf{\Phi_{T\setminus\bar{S}} x_{T\setminus\bar{S}}}\|_\mathbf{2} \\ &\quad -\|\mathbf{\Phi_{\bar{S}}}(\mathbf{\Phi_{\bar{S}}'\Phi_{\bar{S}}})^{-\mathbf{1}}\mathbf{\Phi_{\bar{S}}'\Phi_{T\setminus\bar{S}} x_{T\setminus\bar{S}}}\|_\mathbf{2} \qquad (31) \\ &\geq \sqrt{1-\delta_{|T\setminus\bar{s}|}}\|\mathbf{x_{T\setminus\bar{s}}}\|_\mathbf{2} \\ &\quad -\sqrt{1+\delta_{|\bar{s}|}} \\ &\quad \times\|(\mathbf{\Phi_{\bar{S}}'\Phi_{\bar{S}}})^{-\mathbf{1}}\mathbf{\Phi_{\bar{S}}'\Phi_{T\setminus\bar{S}} x_{T\setminus\bar{S}}}\|_\mathbf{2} \qquad (32) \\ &\geq \sqrt{1-\delta_{|T\setminus\bar{s}|}}\|\mathbf{x_{T\setminus\bar{s}}}\|_\mathbf{2} \\ &\quad -\frac{\sqrt{1+\delta_{|\bar{s}|}}}{1-\delta_{|\bar{s}|}}\|\mathbf{\Phi_{\bar{S}}'\Phi_{T\setminus\bar{S}} x_{T\setminus\bar{S}}}\|_\mathbf{2} \qquad (33)\end{aligned}$$

$$\begin{aligned}&\geq \sqrt{1-\delta_{|T\setminus\bar{s}|}}\|\mathbf{x_{T\setminus\bar{s}}}\|_\mathbf{2} \\ &\quad -\frac{\sqrt{1+\delta_{|\bar{s}|}}\delta_{K+|T\setminus\bar{s}|}}{1-\delta_{|\bar{s}|}}\|\mathbf{x_{T\setminus\bar{s}}}\|_\mathbf{2} \qquad (34) \\ &> \sqrt{1-\delta_{2K}}\|\mathbf{x_{T\setminus\bar{s}}}\|_\mathbf{2} \\ &\quad -\frac{\sqrt{1+\delta_K}\delta_{2K}}{1-\delta_K}\|\mathbf{x_{T\setminus\bar{s}}}\|_\mathbf{2} \qquad (35)\end{aligned}$$

where (32) is from Definition 1, (33) is from from Lemma 3, and (34) and (35) are from Lemma 4 and 2, respectively. Using this together with $\|\mathbf{P_{\bar{S}}^\perp v}\|_\mathbf{2} \leq \|\mathbf{v}\|_\mathbf{2}$, we have

$$\|\mathbf{r_{\bar{S}}}\|_\mathbf{2} > \sqrt{\mathbf{1-\delta_{2K}}}\|\mathbf{x_{T\setminus\bar{S}}}\|_\mathbf{2} - \frac{\sqrt{\mathbf{1+\delta_K}}\boldsymbol{\delta}_\mathbf{2K}}{\mathbf{1-\delta_K}}\|\mathbf{x_{T\setminus\bar{S}}}\|_\mathbf{2} - \|\mathbf{v}\|_\mathbf{2}. \tag{36}$$

for any $\bar{S} \neq T$. Since $\|\mathbf{r_T}\|_\mathbf{2} < \|\mathbf{r_{\bar{S}}}\|_\mathbf{2}$ always holds if the upper bound of $\|\mathbf{r_T}\|_\mathbf{2}$ is smaller than the lower bound of $\|\mathbf{r_{\bar{S}}}\|_\mathbf{2}$, it is clear from (28) and (36) that the hypothesis is satisfied under

$$\|\mathbf{x_{T\setminus\bar{S}}}\|_\mathbf{2} > \frac{\mathbf{2(1-\delta_K)}}{\mathbf{1-3\delta_{2K}}}\|\mathbf{v}\|_\mathbf{2}. \tag{37}$$

Noting that $\|\mathbf{x_{T\setminus\bar{S}}}\|_\mathbf{2} \geq \min_{\mathbf{j}\in\mathbf{T}}|\mathbf{x_j}|$, we get the desired result.

Thus far, we investigated the condition under which the non-causal index set is true when the causal path is true (Theorem 17) and the condition ensuring that the true path has the minimum residual (in magnitude) and hence survives during the tree pruning (Theorem 18). Recalling that the pruning threshold is updated by the minimum value of the residual (in magnitude) in each layer ($\epsilon = \min\|\mathbf{r_{\bar{S}}}\|_\mathbf{2}$) and a path whose residual magnitude is larger than ϵ is pruned, it is clear that the support T will never be pruned if the conditions of Theorem 17 and 18 are jointly satisfied. Formal description of our findings is as follows.

Theorem 19: The true path $\hat{S}^i \subset T$ survives in the pruning process for any i under

$$\min_{j\in T}|x_j| > \max(\mu, \omega)\|\mathbf{v}\|_\mathbf{2} \tag{38}$$

where $\mu = \frac{2(1-\delta_K)}{1-3\delta_{2K}}$ and $\omega = \frac{2(1-\delta_K)\sqrt{1+\delta_{K-1}}}{1-2\delta_K-\delta_{2K-1}}\|\mathbf{v}\|_\mathbf{2}$. *Proof*: Immediate from Theorem 17 and 18.

By combining the results of pre-selection (Theorem 14) and tree search (Theorem 19), we obtain the main result for the noisy setting.

Theorem 20: The TSMP algorithm accurately identifies the support from the noisy measurement $\mathbf{y} = \mathbf{\Phi x} + \mathbf{v}$ under

$$\min_{j \in T} |x_j| > \gamma \|\mathbf{v}\|_2 \tag{39}$$

where $\gamma = \max(\nu, \mu, \omega)$ and $\mu = \frac{2(1-\delta_K)}{1-3\delta_{2K}}$, $\omega = \frac{2(1-\delta_K)\sqrt{1+\delta_{K-1}}}{1-2\delta_K-\delta_{2K-1}}\|\mathbf{v}\|_2$, and $\nu = \frac{2\sqrt{1+\delta_K}}{1-\delta_K-\delta_{2K}}\|\mathbf{v}\|_2$.

Proof: Immediate from Theorem 14 and 19.

It is worth noting that under (39), which essentially corresponds to the high signal-to-noise ratio (SNR) regime, we can identify the accurate support information so that we can simply remove all non-support elements (zero entries in $\mathbf{x}$) and columns associated with these from the system model. In doing so, we can obtain the *overdetermined* system $\mathbf{y} = \mathbf{\Phi_T x_T} + \mathbf{v}$ and the reconstructed signal becomes equivalent to the output of the best possible estimator referred to as Oracle estimator $\hat{\mathbf{x}} = \mathbf{\Phi_T^\dagger y}$.

C. Stable Reconstruction of TSMP

We next investigate the stability of the TSMP algorithm. By stability, we mean that the ℓ_2-norm of the estimation error $\|\mathbf{x} - \hat{\mathbf{x}}_{\bar{\mathbf{s}}}\|_2 = \|\mathbf{x} - \mathbf{\Phi}_{\bar{\mathbf{s}}}^\dagger \mathbf{y}\|_2$ is upper bounded by the constant multiple of the noise power and thus stable reconstruction of the sparse signal is guaranteed. Following theorem indicates that TSMP guarantees the stable reconstruction of K-sparse signal $\mathbf{x}$ regardless of the accurate identification of the support.

Theorem 21: The output $\hat{\mathbf{x}}_{\bar{s}}$ of the TSMP algorithm satisfies

$$\|\mathbf{x} - \hat{\mathbf{x}}_{\bar{\mathbf{s}}}\|_2 < \tau \|\mathbf{v}\|_2. \tag{40}$$

where $\tau = \frac{(\gamma+1)(1-\delta_K)+2\gamma\delta_{2K}}{(1-\delta_K)\sqrt{1-\delta_{2K}}}$.

Proof: From Definition 1, it is clear that

$$\|\mathbf{x} - \hat{\mathbf{x}}_{\bar{\mathbf{s}}}\|_2 \leq \frac{\|\mathbf{\Phi}(\mathbf{x} - \hat{\mathbf{x}}_{\bar{\mathbf{s}}})\|_2}{\sqrt{1-\delta_{|\mathbf{T}\cup\bar{\mathbf{s}}|}}}. \tag{41}$$

Since $\mathbf{x} - \hat{\mathbf{x}}$ is at most $2K$-sparse, we further have

$$\begin{aligned}
\|\mathbf{x} - \hat{\mathbf{x}}_{\bar{\mathbf{s}}}\|_2 &\leq \frac{\|\mathbf{\Phi}(\mathbf{x} - \hat{\mathbf{x}}_{\bar{\mathbf{s}}})\|_2}{\sqrt{1-\delta_{|T\cup\bar{s}|}}} \\
&\leq \frac{\|\mathbf{\Phi}(\mathbf{x} - (\mathbf{\Phi}'_{\bar{\mathbf{s}}}\mathbf{\Phi}_{\bar{\mathbf{s}}})^{-1}\mathbf{\Phi}'_{\bar{\mathbf{s}}}\mathbf{y})\|_2}{\sqrt{1-\delta_{2K}}} \\
&= \frac{\|\mathbf{\Phi x} - \mathbf{\Phi}_{\bar{\mathbf{s}}}(\mathbf{\Phi}'_{\bar{\mathbf{s}}}\mathbf{\Phi}_{\bar{\mathbf{s}}})^{-1}\mathbf{\Phi}'_{\bar{\mathbf{s}}}(\mathbf{\Phi x} + \mathbf{v})\|_2}{\sqrt{1-\delta_{2K}}} \\
&= \frac{\|\mathbf{P}^\perp_{\bar{\mathbf{s}}}\mathbf{\Phi}_{\mathbf{T}\setminus\bar{\mathbf{s}}}\mathbf{x}_{\mathbf{T}\setminus\bar{\mathbf{s}}} - \mathbf{P}_{\bar{\mathbf{s}}}\mathbf{v}\|_2}{\sqrt{1-\delta_{2K}}} \\
&\leq \frac{\|\mathbf{P}^\perp_{\bar{\mathbf{s}}}\mathbf{\Phi}_{\mathbf{T}\setminus\bar{\mathbf{s}}}\mathbf{x}_{\mathbf{T}\setminus\bar{\mathbf{s}}}\|_2 + \|\mathbf{P}_{\bar{\mathbf{s}}}\mathbf{v}\|_2}{\sqrt{1-\delta_{2K}}} \\
&\leq \frac{\|\mathbf{P}^\perp_{\bar{\mathbf{s}}}\mathbf{\Phi}_{\mathbf{T}\setminus\bar{\mathbf{s}}}\mathbf{x}_{\mathbf{T}\setminus\bar{\mathbf{s}}}\|_2 + \|\mathbf{v}\|_2}{\sqrt{1-\delta_{2K}}}
\end{aligned} \tag{42}$$

where $\mathbf{P}_{\bar{\mathbf{s}}} = \mathbf{\Phi}_{\bar{\mathbf{s}}}(\mathbf{\Phi}'_{\bar{\mathbf{s}}}\mathbf{\Phi}_{\bar{\mathbf{s}}})^{-1}\mathbf{\Phi}'_{\bar{\mathbf{s}}}$. Also,

$$\begin{aligned}
\|\mathbf{P}^\perp_{\bar{\mathbf{s}}}\mathbf{\Phi}_{\mathbf{T}\setminus\bar{\mathbf{s}}}\mathbf{x}_{\mathbf{T}\setminus\bar{\mathbf{s}}}\|_2 &= \|(\mathbf{I} - \mathbf{\Phi}_{\bar{\mathbf{s}}}(\mathbf{\Phi}'_{\bar{\mathbf{s}}}\mathbf{\Phi}_{\bar{\mathbf{s}}})^{-1}\mathbf{\Phi}'_{\bar{\mathbf{s}}})\mathbf{\Phi}_{\mathbf{T}\setminus\bar{\mathbf{s}}}\mathbf{x}_{\mathbf{T}\setminus\bar{\mathbf{s}}}\|_2 \\
&\leq \|\mathbf{\Phi}_{\mathbf{T}\setminus\bar{\mathbf{s}}}\mathbf{x}_{\mathbf{T}\setminus\bar{\mathbf{s}}}\|_2 \\
&\quad + \|\mathbf{\Phi}_{\bar{\mathbf{s}}}(\mathbf{\Phi}'_{\bar{\mathbf{s}}}\mathbf{\Phi}_{\bar{\mathbf{s}}})^{-1}\mathbf{\Phi}'_{\bar{\mathbf{s}}}\mathbf{\Phi}_{\mathbf{T}\setminus\bar{\mathbf{s}}}\mathbf{x}_{\mathbf{T}\setminus\bar{\mathbf{s}}}\|_2
\end{aligned} \tag{43}$$

where (43) is from the triangle inequality. From (43), we have

$$\begin{aligned}
\|\mathbf{P}^\perp_{\bar{\mathbf{s}}}\mathbf{\Phi}_{\mathbf{T}\setminus\bar{\mathbf{s}}}\mathbf{x}_{\mathbf{T}\setminus\bar{\mathbf{s}}}\|_2 &\leq \sqrt{1+\delta_{|T\setminus\bar{s}|}}\|\mathbf{x}_{\mathbf{T}\setminus\bar{\mathbf{s}}}\|_2 \\
&\quad + \sqrt{1+\delta_{|\bar{s}|}} \\
&\quad \times \|(\mathbf{\Phi}'_{\bar{\mathbf{s}}}\mathbf{\Phi}_{\bar{\mathbf{s}}})^{-1}\mathbf{\Phi}'_{\bar{\mathbf{s}}}\mathbf{\Phi}_{\mathbf{T}\setminus\bar{\mathbf{s}}}\mathbf{x}_{\mathbf{T}\setminus\bar{\mathbf{s}}}\|_2 && (44) \\
&\leq \sqrt{1+\delta_{|T\setminus\bar{s}|}}\|\mathbf{x}_{\mathbf{T}\setminus\bar{\mathbf{s}}}\|_2 \\
&\quad + \frac{\sqrt{1+\delta_{|\bar{s}|}}}{1-\delta_{|\bar{s}|}}\|\mathbf{\Phi}'_{\bar{\mathbf{s}}}\mathbf{\Phi}_{\mathbf{T}\setminus\bar{\mathbf{s}}}\mathbf{x}_{\mathbf{T}\setminus\bar{\mathbf{s}}}\|_2 && (45) \\
&\leq \sqrt{1+\delta_{|T\setminus\bar{s}|}}\|\mathbf{x}_{\mathbf{T}\setminus\bar{\mathbf{s}}}\|_2 \\
&\quad + \frac{\sqrt{1+\delta_{|\bar{s}|}}\delta_{K+|T\setminus\bar{s}|}}{1-\delta_{|\bar{s}|}}\|\mathbf{x}_{\mathbf{T}\setminus\bar{\mathbf{s}}}\|_2 && (46) \\
&\leq \sqrt{1+\delta_K}\|\mathbf{x}_{\mathbf{T}\setminus\bar{\mathbf{s}}}\|_2 \\
&\quad + \frac{\sqrt{1+\delta_K}\delta_{2K}}{1-\delta_K}\|\mathbf{x}_{\mathbf{T}\setminus\bar{\mathbf{s}}}\|_2 && (47) \\
&< \sqrt{1+\delta_{2K}}\|\mathbf{x}_{\mathbf{T}\setminus\bar{\mathbf{s}}}\|_2 \\
&\quad + \frac{\sqrt{1+\delta_K}\delta_{2K}}{1-\delta_K}\|\mathbf{x}_{\mathbf{T}\setminus\bar{\mathbf{s}}}\|_2 && (48) \\
&< \frac{1-\delta_K+2\delta_{2K}}{1-\delta_K}\|\mathbf{x}_{\mathbf{T}\setminus\bar{\mathbf{s}}}\|_2 && (49)
\end{aligned}$$

where (44) is from Definition 1, and (45) and (46) are from Lemma 3 and 4, respectively. Plugging (49) into (42), we have

$$\|\mathbf{x} - \hat{\mathbf{x}}_{\bar{\mathbf{s}}}\|_2 < \frac{(\mathbf{1}-\delta_{\mathbf{K}}+\mathbf{2}\delta_{\mathbf{2K}})\|\mathbf{x}_{\mathbf{T}\setminus\bar{\mathbf{s}}}\|_2}{(\mathbf{1}-\delta_{\mathbf{K}})\sqrt{\mathbf{1}-\delta_{\mathbf{2K}}}} + \frac{\|\mathbf{v}\|_2}{\sqrt{\mathbf{1}-\delta_{\mathbf{2K}}}}. \tag{50}$$

Note that when the support is chosen accurately, $\bar{S} = T$ and thus

$$\|\mathbf{x}_{\mathbf{T}\setminus\bar{\mathbf{s}}}\|_2 = \mathbf{0}. \tag{51}$$

Whereas, if $\bar{S} \neq T$, then by the contraposition of Theorem 20[5], we have

$$\|\mathbf{x}_{\mathbf{T}\setminus\bar{\mathbf{s}}}\|_2 \leq \gamma \|\mathbf{v}\|_2. \tag{52}$$

By combining (51) and (52), we obtain the desired result.

IV. SIMULATION AND DISCUSSIONS

A. Simulation Setup

In this section, we observe the performance of sparse recovery algorithms including TSMP through empirical simulations. In our simulations, we use the sensing matrix $\mathbf{\Phi}$ of size 100×256 whose entries are from the independent zero-mean Gaussian distribution with variance $\frac{1}{M}$. We generate K-sparse vector $\mathbf{x}$ whose nonzero locations are randomly chosen and nonzero coefficients are drawn from i.i.d. zero-mean Gaussian random variable with variance 1. In each point of the individual recovery algorithm, we perform at least $n = 10,000$ independent trials. In the noiseless setting, we use the exact recovery ratio (ERR)

[5]Here, we need to use slightly modified version of Theorem 20, which says that if $\|\mathbf{x}_{\bar{\mathbf{s}}}\|_2 > \gamma\|\mathbf{v}\|_2$, then $\bar{S} = T$.

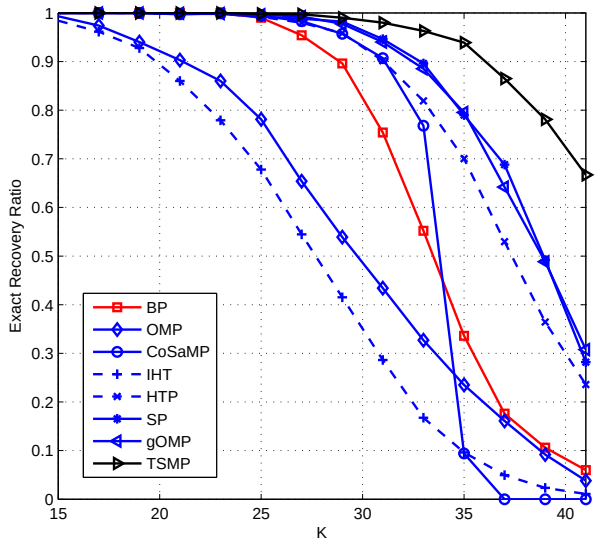


Fig. 6. ERR performance as a function of the sparsity K in the noiseless setting.

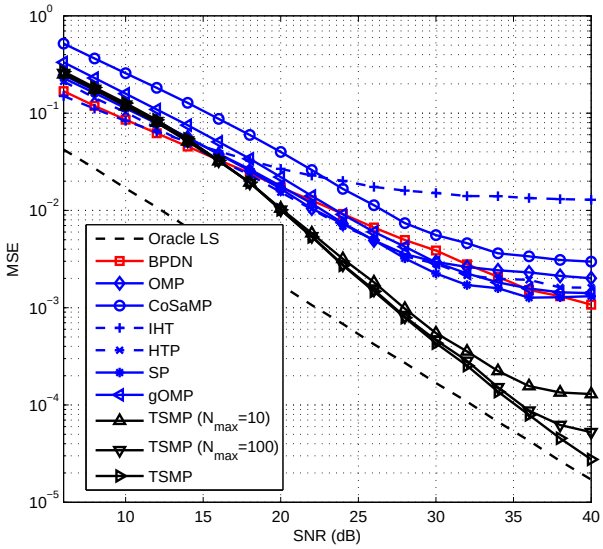


Fig. 8. MSE performance of sparse recovery algorithms in the noisy setting ($K = 30$).

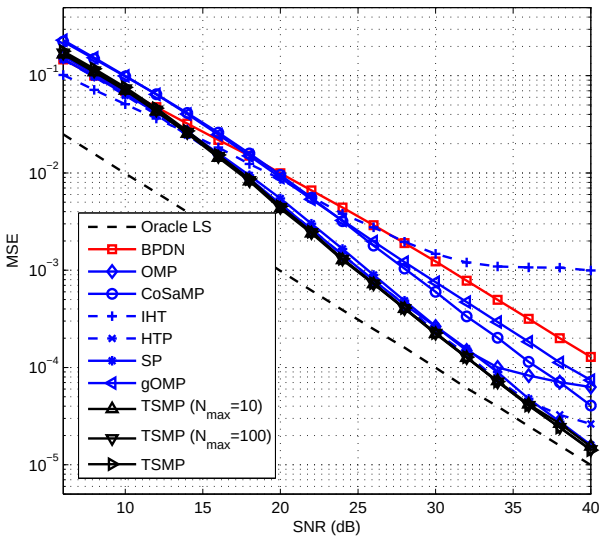


Fig. 7. MSE performance of sparse recovery algorithms ($K = 20$) in 100×256 system.

as a performance measure. In the noisy setting, we use the mean squared error (MSE) of the recovery algorithms which is defined as

$$MSE = \frac{1}{n}\sum_{\ell=1}^{n} \frac{\|\mathbf{x}(\ell) - \hat{\mathbf{x}}(\ell)\|^2}{N}$$

where $\hat{\mathbf{x}}(\ell)$ is the estimate of the original sparse signal $\mathbf{x}(\ell)$. In our simulations, we consider the following algorithms:

1. BP algorithm [5]: we use BP in noiseless setting and basis pursuit denoising (BPDN) in noisy setting.
2. OMP algorithm [7]
3. Iterative hard thresholding (IHT) [11]
4. Hard thresholding pursuit (HTP) [13]
5. Subspace pursuit (SP) [10]
6. gOMP algorithm [20]: we set $L = 2$.
7. Proposed TSMP algorithm
8. TSMP with limited branching: we set the maximum number of child nodes in each layer ($N_{\max} = 10$ and 100).

B. Simulation Results

We first compare the ERR performance of sparse recovery algorithms in the noiseless setting. In Fig. 6, we plot the ERR performance of sparse signal recovery algorithms as a function of the sparsity K. Overall, we observe that overall behavior of TSMP is better than existing algorithms, in particular for large K. In Fig. 7, we plot the MSE performance of the sparse recovery algorithms as a function of signal-to-noise ratio (SNR) in the noisy setting. Note that the decibel (dB) scale of SNR is defined as $\text{SNR} = 10\log_{10}\frac{\|\mathbf{\Phi x}\|^2}{\|\mathbf{v}\|^2}$. In this test, we set the sparsity to $K = 20$ so that 8% of the input vector elements are nonzero. Overall, we see that the performance of sparse recovery algorithm improves with SNR. While the performance gap between conventional sparse recovery algorithms and performance lower bound (obtained by Oracle LS estimator) is maintained across the board, the performance gap between TSMP and Oracle LS estimator decreases as SNR increases. When the sparsity increases ($K = 30$), we clearly see this trend as shown in Fig. 8. In fact, the gain of TSMP over conventional algorithms at $MSE = 10^{-2}$ is around 2 dB but the gain at $MSE = 10^{-3}$ increases to 10 dB. Also, as it can be seen from the figure and also in accordance with Theorem 20, the performance of TSMP is asymptotically optimal in high SNR regime in the sense that it approaches the performance bound of Oracle LS estimator.

Fig. 9 shows the running time complexity of the sparse recovery algorithms as a function of the sparsity level K. All algorithms under test are coded by MATLAB software package and run by a personal computer with Intel Core i5 processor

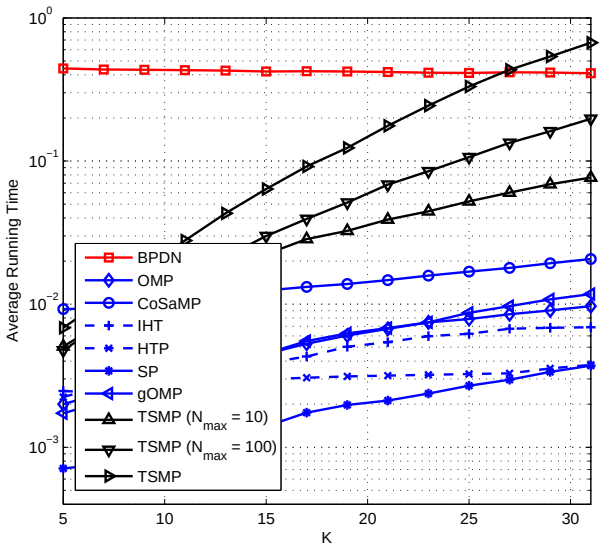


Fig. 9. Average running time of sparse signal recovery algorithms in 100×256 system.

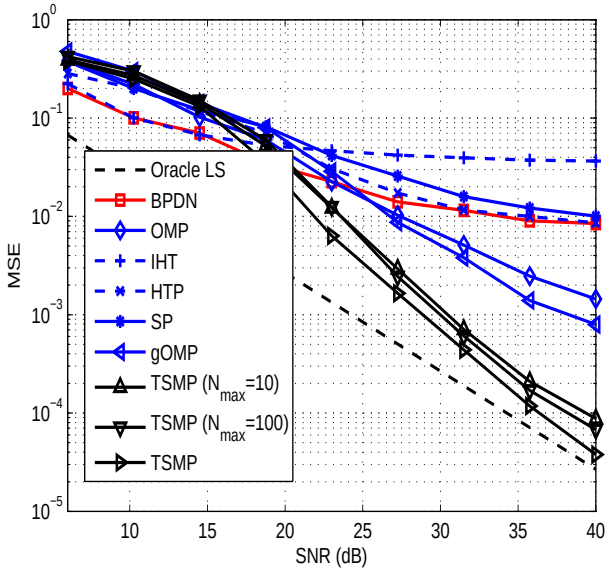


Fig. 10. MSE performance of sparse recovery algorithms when the dimension of the sensing matrix is $1,000 \times 2,500$. We set $K = 400$.

and Microsoft Windows 7 environment. As seen in the figure, among greedy algorithms under test, OMP exhibits the smallest running time. Since TSMP performs the tree search, the running time complexity of TSMP is higher than the conventional greedy algorithms. However, by limiting the number of child nodes in the tree search, computational burden of TSMP can be reduced significantly. In particular, if $N_{\max} = 10$, TSMP achieves substantial reduction in complexity over the original TSMP algorithm with a marginal loss in performance.

In order to see the behavior of TSMP when the dimension increases, we investigate the performance and the complexity of sparse recovery algorithms when the dimension of sensing matrix is $1,000 \times 2,500$. Fig. 10 shows the MSE performance of this system as a function of SNR when $K = 400$ (16% sparsity). Overall behavior is more or less similar to the scenario with 100×256 sensing matrix, but the gain of TSMP over conventional algorithms increases and also TSMP performs close to the Oracle estimator as SNR increases.

In order to show the effectiveness of TSMP in a realistic scenario, we evaluate the performance for internet of things (IoT) network [24]. IoT network, designed to provide network connectivity of almost all things, has many applications including healthcare, automatic metering, environmental monitoring, and surveillance. Common feature of the IoT networks is that the node density is in general higher than the cellular network, yet the packet size is very small and not every devices transmits information at a given time. In view of this, the transmit vector $\mathbf{x}$ can be readily modeled as a sparse vector whose nonzero entries are from the set of finite alphabets. Fig. 11 shows the symbol error rate (SER) performance as a function of SNR, where the transmission is based on 24×48 system with sparsity $K = 10$ (20% sensor node activity). We observe that the performance of conventional algorithms is in general not appealing even for high SNR regime but TSMP is effective and its performance improves with SNR.

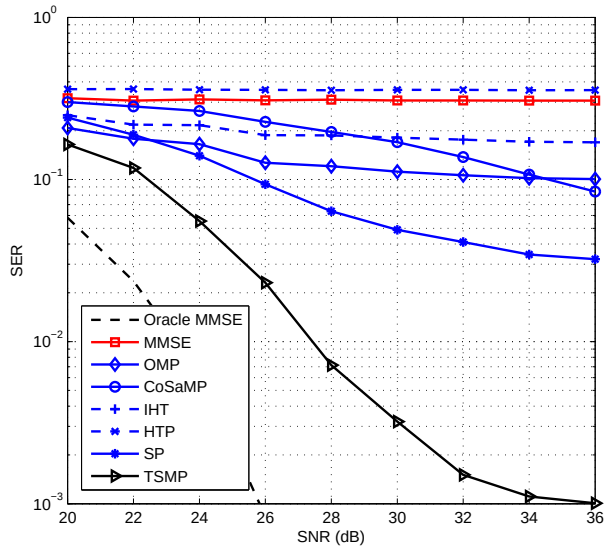


Fig. 11. SER performance of sparse recovery for wireless sensor network detection. Nonzero entries (actual transmit signal) are chosen from 16QAM constellation.

V. Conclusions

In this paper, we proposed a tree search based sparse signal recovery algorithm referred to as the tree search matching pursuit (TSMP). In order to overcome the shortcoming of greedy algorithm in choosing *short-sighted* candidates, the TSMP algorithm performs the tree search and choose the best one among multiple promising candidates. The computational complexity caused by the tree search is controlled by deliberately designed preprocessing and tree pruning operation. From the RIP based analysis and

also numerical simulations, we observed that TSMP shows excellent recovery performance in both noiseless and noisy scenarios. The tree search strategy developed in this paper is an interesting and useful framework to solve the sparse signal recovery problem and many important works remain to be answered. Our future work will address more complicated scenarios. For example, exploration of Bayesian tree search algorithm when the *a prior information* is available and modification of tree search for the multiple measurement vector (MMV) scenario would be interesting direction to be investigated.

APPENDICES

I. Proof of Lemma 5

From the definition of κ, we have

$$\begin{aligned}\kappa &= \max_{j\in T}|\phi_j'\mathbf{y}| \\ &= \|\mathbf{\Phi y}\|_\infty \geq \frac{\mathbf{1}}{\sqrt{\mathbf{K}}}\|\mathbf{\Phi y}\|_\mathbf{2} \end{aligned} \quad (53)$$

where (53) is from the inequality $\|\mathbf{u}\|_\infty \geq \frac{\mathbf{1}}{\sqrt{\|\mathbf{u}\|_\mathbf{0}}}\|\mathbf{u}\|_\mathbf{2}$ for any vector $\mathbf{u}$. Using Definition 1, we further have

$$\begin{aligned}\kappa &\geq \frac{1}{\sqrt{K}}\|\mathbf{\Phi}'\mathbf{y}\|_\mathbf{2} = \frac{\mathbf{1}}{\sqrt{\mathbf{K}}}\|\mathbf{\Phi}_\mathbf{T}'\mathbf{\Phi}_\mathbf{T}\mathbf{x}_\mathbf{T}\|_\mathbf{2} && (54)\\ &\geq \frac{1-\delta_K}{\sqrt{K}}\|\mathbf{x}_\mathbf{T}\|_\mathbf{2} && (55)\end{aligned}$$

where (55) is from Lemma 3.

II. Proof of Lemma 6

From the definition of ζ, we have

$$\sqrt{K}\zeta \leq \sqrt{\sum_{j\in I_K}|\phi_j'\mathbf{y}|^\mathbf{2}} = \|\mathbf{\Phi}_{\mathbf{I_K}}'\mathbf{y}\|_\mathbf{2} \quad (56)$$

where $I_K = \arg\max_{|I|=K, I\subset T^c}\|\mathbf{\Phi}_\mathbf{I}'\mathbf{y}\|_\mathbf{2}$. Since I_K and T are disjoint ($I_K \subset T^c$), we have

$$\|\mathbf{\Phi}_{\mathbf{I_K}}'\mathbf{y}\|_\mathbf{2} = \|\mathbf{\Phi}_{\mathbf{I_K}}'\mathbf{\Phi}_\mathbf{T}\mathbf{x}_\mathbf{T}\|_\mathbf{2} \leq \delta_\mathbf{2K}\|\mathbf{x}_\mathbf{T}\|_\mathbf{2} \quad (57)$$

where (57) is from Lemma 4. From (56) and (57), we get the desired result.

III. Proof of Lemma 8

Let $\Psi_l = \hat{S}^i \cup \{\hat{s}_{i+1}, \cdots, \hat{s}_{i+l}\} \subset T$ and $\lambda^l = \max_{j\in T\setminus\Psi_l}|\phi_j'\mathbf{r}_{\mathbf{\Psi_1}}|$, then we have

$$\begin{aligned}\max_{j\in T\setminus\Psi_l}|\phi_j'\mathbf{r}_{\mathbf{\Psi_1}}| &= \|\mathbf{\Phi}_{\mathbf{T}\setminus\mathbf{\Psi_1}}'\mathbf{r}_{\mathbf{\Psi_1}}\|_\infty && (58)\\ &\geq \frac{1}{\sqrt{|T\setminus\Psi_l|}}\|\mathbf{\Phi}_{\mathbf{T}\setminus\mathbf{\Psi_1}}'\mathbf{P}_{\mathbf{\Psi_1}}^\perp\mathbf{y}\|_\mathbf{2} && (59)\\ &= \frac{1}{\sqrt{|T\setminus\Psi_l|}}\|\mathbf{\Phi}_{\mathbf{T}\setminus\mathbf{\Psi_1}}'\mathbf{P}_{\mathbf{\Psi_1}}^\perp\mathbf{\Phi}_{\mathbf{T}\setminus\mathbf{\Psi_1}}\mathbf{x}_{\mathbf{T}\setminus\mathbf{\Psi_1}}\|_\mathbf{2} && (60)\\ &\geq \frac{1}{\sqrt{K-i-l}}[\|\mathbf{\Phi}_{\mathbf{T}\setminus\mathbf{\Psi_1}}'\mathbf{\Phi}_{\mathbf{T}\setminus\mathbf{\Psi_1}}\mathbf{x}_{\mathbf{T}\setminus\mathbf{\Psi_1}}\|_\mathbf{2} \\ &\quad -\|\mathbf{\Phi}_{\mathbf{T}\setminus\mathbf{\Psi_1}}'\mathbf{P}_{\mathbf{\Psi_1}}\mathbf{\Phi}_{\mathbf{T}\setminus\mathbf{\Psi_1}}\mathbf{x}_{\mathbf{T}\setminus\mathbf{\Psi_1}}\|_\mathbf{2}] && (61)\end{aligned}$$

where $\mathbf{P}_{\mathbf{\Psi_1}}^\perp = \mathbf{I} - \mathbf{\Phi}_{\lambda^\mathbf{l}}(\mathbf{\Phi}_{\mathbf{\Psi_1}}'\mathbf{\Phi}_{\mathbf{\Psi_1}})^{-\mathbf{1}}\mathbf{\Phi}_{\mathbf{\Psi_1}}'$ and (59) is from the inequality $\|\mathbf{u}\|_\infty \geq \frac{\mathbf{1}}{\sqrt{\|\mathbf{u}\|_\mathbf{0}}}\|\mathbf{u}\|_\mathbf{2}$ for any vector $\mathbf{u}$, and (61) is from the triangle inequality. From Lemma 3, we have

$$\|\mathbf{\Phi}_{\mathbf{T}\setminus\mathbf{\Psi_1}}'\mathbf{\Phi}_{\mathbf{T}\setminus\mathbf{\Psi_1}}\mathbf{x}_{\mathbf{T}\setminus\mathbf{\Psi_1}}\|_\mathbf{2} \geq (1-\delta_{|T\setminus\Psi_l|})\|\mathbf{x}_{\mathbf{T}\setminus\mathbf{\Psi_1}}\|_\mathbf{2} \quad (62)$$

Also,

$$\begin{aligned}\|\mathbf{\Phi}_{\mathbf{T}\setminus\mathbf{\Psi_1}}'\mathbf{P}_{\mathbf{\Psi_1}}\mathbf{\Phi}_{\mathbf{T}\setminus\mathbf{\Psi_1}}\mathbf{x}_{\mathbf{T}\setminus\mathbf{\Psi_1}}\|_\mathbf{2} &= \|\mathbf{\Phi}_{\mathbf{T}\setminus\mathbf{\Psi_1}}'\mathbf{\Phi}_{\mathbf{\Psi_1}}(\mathbf{\Phi}_{\mathbf{\Psi_1}}'\mathbf{\Phi}_{\mathbf{\Psi_1}})^{-\mathbf{1}} \\ &\quad \times\mathbf{\Phi}_{\mathbf{\Psi_1}}'\mathbf{\Phi}_{\mathbf{T}\setminus\mathbf{\Psi_1}}\mathbf{x}_{\mathbf{T}\setminus\mathbf{\Psi_1}}\|_\mathbf{2} && (63)\\ &\leq \delta_{|T|}\|(\mathbf{\Phi}_{\mathbf{\Psi_1}}'\mathbf{\Phi}_{\mathbf{\Psi_1}})^{-\mathbf{1}} \\ &\quad \times\mathbf{\Phi}_{\mathbf{\Psi_1}}'\mathbf{\Phi}_{\mathbf{T}\setminus\mathbf{\Psi_1}}\mathbf{x}_{\mathbf{T}\setminus\mathbf{\Psi_1}}\|_\mathbf{2} && (64)\\ &\leq \frac{\delta_{|T|}}{1-\delta_{|\Psi_l|}} \\ &\quad \times\|\mathbf{\Phi}_{\mathbf{\Psi_1}}'\mathbf{\Phi}_{\mathbf{T}\setminus\mathbf{\Psi_1}}\mathbf{x}_{\mathbf{T}\setminus\mathbf{\Psi_1}}\|_\mathbf{2} && (65)\\ &\leq \frac{\delta_{|T|}^2}{1-\delta_{|\Psi_l|}}\|\mathbf{x}_{\mathbf{T}\setminus\mathbf{\Psi_1}}\|_\mathbf{2} && (66)\end{aligned}$$

where (64) and (66) are from Lemma 4, and (65) is from Lemma 3. Using (61), (62), (66), and Lemma 2, we have

$$\begin{aligned}\lambda^l &\geq \frac{1}{\sqrt{K-i-l}}[\|\mathbf{\Phi}_{\mathbf{T}\setminus\mathbf{\Psi_1}}'\mathbf{\Phi}_{\mathbf{T}\setminus\mathbf{\Psi_1}}\mathbf{x}_{\mathbf{T}\setminus\mathbf{\Psi_1}}\|_\mathbf{2} \\ &\quad -\|\mathbf{\Phi}_{\mathbf{T}\setminus\mathbf{\Psi_1}}'\mathbf{P}_{\mathbf{\Psi_1}}\mathbf{\Phi}_{\mathbf{T}\setminus\mathbf{\Psi_1}}\mathbf{x}_{\mathbf{T}\setminus\mathbf{\Psi_1}}\|_\mathbf{2}] \\ &\geq \frac{1}{\sqrt{K-i-l}}\left[(1-\delta_{|T\setminus\Psi_l|})\|\mathbf{x}_{\mathbf{T}\setminus\mathbf{\Psi_1}}\|_\mathbf{2} - \frac{\delta_{|\mathbf{T}|}^\mathbf{2}}{\mathbf{1}-\delta_{|\mathbf{\Psi_1}|}}\|\mathbf{x}_{\mathbf{T}\setminus\mathbf{\Psi_1}}\|_\mathbf{2}\right] \\ &\geq \frac{1}{\sqrt{K-i-1}}\left[(1-\delta_{K-i-l})\|\mathbf{x}_{\mathbf{T}\setminus\mathbf{\Psi_1}}\|_\mathbf{2} - \frac{\delta_\mathbf{K}^\mathbf{2}}{\mathbf{1}-\delta_\mathbf{K}}\|\mathbf{x}_{\mathbf{T}\setminus\mathbf{\Psi_1}}\|_\mathbf{2}\right] \\ &> \frac{1}{\sqrt{K-i}}\left[(1-\delta_K)\|\mathbf{x}_{\mathbf{T}\setminus\mathbf{\Psi_1}}\|_\mathbf{2} - \frac{\delta_\mathbf{K}^\mathbf{2}}{\mathbf{1}-\delta_\mathbf{K}}\|\mathbf{x}_{\mathbf{T}\setminus\mathbf{\Psi_1}}\|_\mathbf{2}\right] \\ &= \frac{1-2\delta_K}{\sqrt{K-i}(1-\delta_K)}\|\mathbf{x}_{\mathbf{T}\setminus\mathbf{\Psi_1}}\|_\mathbf{2}, && (67)\end{aligned}$$

and since $1 \leq i \leq K-1$, we obtain the desired result.

IV. Proof of Lemma 9

From the definition of $\gamma^l = \min_{j\in D}|\phi_j'\mathbf{r}_{\mathbf{\Psi_1}}|$, we have

$$\sqrt{|D|}\gamma^l \leq \sqrt{\sum_{j\in D}|\phi_j'\mathbf{r}_{\mathbf{\Psi_1}}|^\mathbf{2}} = \|\mathbf{\Phi}_\mathbf{D}'\mathbf{r}_{\mathbf{\Psi_1}}\|_\mathbf{2} \quad (68)$$

where $D = \arg\max_{|I|=K-i, I\subset\Omega\setminus T}\|\mathbf{\Phi}_\mathbf{I}'\mathbf{r}_{\mathbf{\Psi_1}}\|_\mathbf{2}$ and $\Psi_l \subset T$. Using the triangle inequality, we have

$$\begin{aligned}\|\mathbf{\Phi}_\mathbf{D}'\mathbf{r}_{\mathbf{\Psi_1}}\|_\mathbf{2} &= \|\mathbf{\Phi}_\mathbf{D}'\mathbf{P}_{\mathbf{\Psi_1}}^\perp\mathbf{y}\|_\mathbf{2} && (69)\\ &= \|\mathbf{\Phi}_\mathbf{D}'\mathbf{P}_{\lambda^\mathbf{l}}^\perp\mathbf{\Phi}_{\mathbf{T}\setminus\lambda^\mathbf{l}}\mathbf{x}_{\mathbf{T}'\setminus\lambda^\mathbf{l}}\|_\mathbf{2} && (70)\\ &\leq \|\mathbf{\Phi}_\mathbf{D}'\mathbf{\Phi}_{\mathbf{T}\setminus\mathbf{\Psi_1}}\mathbf{x}_{\mathbf{T}\setminus\mathbf{\Psi_1}}\|_\mathbf{2} \\ &\quad +\|\mathbf{\Phi}_\mathbf{D}'\mathbf{P}_{\mathbf{\Psi_1}}\mathbf{\Phi}_{\mathbf{T}\setminus\mathbf{\Psi_1}}\mathbf{x}_{\mathbf{T}\setminus\mathbf{\Psi_1}}\|_\mathbf{2}. && (71)\end{aligned}$$

Since $D \subset \Omega\setminus T$, we have

$$\|\mathbf{\Phi}_\mathbf{D}'\mathbf{\Phi}_{\mathbf{T}\setminus\mathbf{\Psi_1}}\mathbf{x}_{\mathbf{T}\setminus\mathbf{\Psi_1}}\|_\mathbf{2} \leq \delta_{|\mathbf{D}|+|\mathbf{T}\setminus\mathbf{\Psi_1}|}\|\mathbf{x}_{\mathbf{T}\setminus\mathbf{\Psi_1}}\|_\mathbf{2} \quad (72)$$

where (72) is from Lemma 4. Also,

$$\begin{aligned}\|\mathbf{\Phi}'_{\mathbf{D}}\mathbf{P}_{\mathbf{\Psi}_l}\mathbf{\Phi}_{\mathbf{T}\setminus\mathbf{\Psi}_l}\mathbf{x}_{\mathbf{T}\setminus\mathbf{\Psi}_l}\|_{\mathbf{2}} &= \|\mathbf{\Phi}'_{\mathbf{D}}\mathbf{\Phi}_{\mathbf{\Psi}_l}(\mathbf{\Phi}'_{\mathbf{\Psi}_l}\mathbf{\Phi}_{\mathbf{\Psi}_l})^{-\mathbf{1}} \\ &\quad\times\mathbf{\Phi}'_{\mathbf{\Psi}_l}\mathbf{\Phi}_{\mathbf{T}'\setminus\mathbf{\Psi}_l}\mathbf{x}_{\mathbf{T}'\setminus\mathbf{\Psi}_l}\|_{\mathbf{2}} \quad (73)\\ &\le \delta_{|D|+|\Psi_l|}\|(\mathbf{\Phi}'_{\mathbf{\Psi}_l}\mathbf{\Phi}_{\mathbf{\Psi}_l})^{-\mathbf{1}} \\ &\quad\times\mathbf{\Phi}'_{\mathbf{\Psi}_l}\mathbf{\Phi}_{\mathbf{T}\setminus\mathbf{\Psi}_l}\mathbf{x}_{\mathbf{T}\setminus\mathbf{\Psi}_l}\|_{\mathbf{2}} \quad (74)\\ &\le \frac{\delta_{|D|+|\Psi_l|}}{1-\delta_{|\Psi_l|}} \\ &\quad\times\|\mathbf{\Phi}'_{\mathbf{\Psi}_l}\mathbf{\Phi}_{\mathbf{T}\setminus\mathbf{\Psi}_l}\mathbf{x}_{\mathbf{T}\setminus\mathbf{\Psi}_l}\|_{\mathbf{2}} \quad (75)\\ &\le \frac{\delta_{|D|+|\Psi_l|}\delta_K}{1-\delta_{|\Psi_l|}}\|\mathbf{x}_{\mathbf{T}\setminus\mathbf{\Psi}_l}\|_{\mathbf{2}} \quad (76)\end{aligned}$$

where (74) and (76) are from Lemma 4 and (75) is from Lemma 3. Using (71), (72), and (76), we have

$$\begin{aligned}\|\mathbf{\Phi}'_{\mathbf{D}}\mathbf{r}_{\lambda^l}\|_{\mathbf{2}} &\le \|\mathbf{\Phi}'_{\mathbf{D}}\mathbf{r}_{\mathbf{\Psi}_l}\|_{\mathbf{2}} + \|\mathbf{\Phi}'_{\mathbf{D}}\mathbf{P}_{\mathbf{\Psi}_l}\mathbf{\Phi}_{\mathbf{T}\setminus\mathbf{\Psi}_l}\mathbf{x}_{\mathbf{T}\setminus\mathbf{\Psi}_l}\|_{\mathbf{2}} \quad (77)\\ &\le \left(\delta_{|D|+|T\setminus\Psi_l|} + \frac{\delta_{|D|+|\Psi_l|}\delta_K}{1-\delta_{|\Psi_l|}}\right)\|\mathbf{x}_{\mathbf{T}\setminus\mathbf{\Psi}_l}\|_{\mathbf{2}} \quad (78)\\ &\le \left(\delta_{2K-2i-l} + \frac{\delta_{K+l}\delta_K}{1-\delta_{i+l}}\right)\|\mathbf{x}_{\mathbf{T}'\setminus\mathbf{\Psi}_l}\|_{\mathbf{2}} \quad (79)\\ &< \left(\delta_{2K-i} + \frac{\delta_{2K-i}\delta_K}{1-\delta_K}\right)\|\mathbf{x}_{\mathbf{T}'\setminus\mathbf{\Psi}_l}\|_{\mathbf{2}} \quad (80)\\ &= \frac{\delta_{2K-i}}{1-\delta_K}\|\mathbf{x}_{\mathbf{T}'\setminus\mathbf{\Psi}_l}\|_{\mathbf{2}}, \quad (81)\end{aligned}$$

where (79) is because $|L| = K - i$ and $1 \le i \le K-1$, and (80) is from Lemma 2. From (68) and (81), we have the desired result as

$$\begin{aligned}\gamma^l &\le \frac{\delta_{2K-i}}{\sqrt{|D|}(1-\delta_K)}\|\mathbf{x}_{\mathbf{T}\setminus\mathbf{\Psi}_l}\|_{\mathbf{2}} \quad (82)\\ &= \frac{\delta_{2K-i}}{\sqrt{K-i}(1-\delta_K)}\|\mathbf{x}_{\mathbf{T}'\setminus\mathbf{\Psi}_l}\|_{\mathbf{2}}. \quad (83)\end{aligned}$$

Since $1 \le i \le K-1$, we obtain the desired result.

V. Proof of Lemma 12

From the definition of ρ in (14), we have

$$\begin{aligned}\rho &= \max_{j\in T}|\phi'_j\mathbf{y}| \quad (84)\\ &= \|\mathbf{\Phi}'_{\mathbf{T}}\mathbf{y}\|_{\infty} \quad (85)\\ &\ge \frac{1}{\sqrt{|T|}}\|\mathbf{\Phi}'_{\mathbf{T}}\mathbf{y}\|_{\mathbf{2}} \quad (86)\\ &= \frac{1}{\sqrt{K}}\|\mathbf{\Phi}'_{\mathbf{T}}(\mathbf{\Phi}_{\mathbf{T}}\mathbf{x}_{\mathbf{T}}+\mathbf{v})\|_{\mathbf{2}} \quad (87)\\ &\ge \frac{1}{\sqrt{K}}\left(\|\mathbf{\Phi}'_{\mathbf{T}}\mathbf{\Phi}_{\mathbf{T}}\mathbf{x}_{\mathbf{T}}\|_{\mathbf{2}} - \|\mathbf{\Phi}'_{\mathbf{T}}\mathbf{v}\|_{\mathbf{2}}\right) \quad (88)\end{aligned}$$

where (86) is from the inequality $\|\mathbf{u}\|_{\infty} \ge \frac{\mathbf{1}}{\sqrt{\|\mathbf{u}\|_{\mathbf{0}}}}\|\mathbf{u}\|_{\mathbf{2}}$ for any vector $\mathbf{u}$. Note that

$$\|\mathbf{\Phi}'_{\mathbf{T}}\mathbf{\Phi}_{\mathbf{T}}\mathbf{x}_{\mathbf{T}}\|_{\mathbf{2}} \ge (\mathbf{1}-\delta_{\mathbf{K}})\|\mathbf{x}_{\mathbf{T}}\|_{\mathbf{2}} \quad (89)$$

and

$$\|\mathbf{\Phi}'_{\mathbf{T}}\mathbf{v}\|_{\mathbf{2}} \le \sqrt{\mathbf{1}+\delta_{\mathbf{K}}}\|\mathbf{v}\|_{\mathbf{2}}. \quad (90)$$

Using (89) and (90), ρ is lower bounded as

$$\rho \ge \frac{1}{\sqrt{K}}\left[(1-\delta_K)\|\mathbf{x}_{\mathbf{T}}\|_{\mathbf{2}} - \sqrt{\mathbf{1}+\delta_{\mathbf{K}}}\|\mathbf{v}\|_{\mathbf{2}}\right], \quad (91)$$

which is the desired result.

VI. Proof of Lemma 13

From the definition of η in (14), we have

$$\sqrt{K}\eta \le \sqrt{\sum_{j\in I_K}|\phi'_j\mathbf{y}|^{\mathbf{2}}} = \|\mathbf{\Phi}'_{\mathbf{I_K}}\mathbf{y}\|_{\mathbf{2}} \quad (92)$$

where $I_K = \arg\max_{|I|=K, I\subset T^c}\|\mathbf{\Phi}'_{\mathbf{I}}\mathbf{y}\|_{\mathbf{2}}$. Using the triangle inequality, we have

$$\begin{aligned}\|\mathbf{\Phi}'_{\mathbf{I_K}}\mathbf{y}\|_{\mathbf{2}} &= \|\mathbf{\Phi}'_{\mathbf{I_K}}(\mathbf{\Phi}_{\mathbf{T}}\mathbf{x}_{\mathbf{T}}+\mathbf{v})\|_{\mathbf{2}} \quad (93)\\ &\le \|\mathbf{\Phi}'_{\mathbf{I_K}}\mathbf{\Phi}_{\mathbf{T}}\mathbf{x}_{\mathbf{T}}\|_{\mathbf{2}} + \|\mathbf{\Phi}'_{\mathbf{I_K}}\mathbf{v}\|_{\mathbf{2}}. \quad (94)\end{aligned}$$

Since I_K and T are disjoint ($I_K \subset T^c$), we further have

$$\|\mathbf{\Phi}'_{\mathbf{I_K}}\mathbf{\Phi}_{\mathbf{T}}\mathbf{x}_{\mathbf{T}}\|_{\mathbf{2}} \le \delta_{\mathbf{2K}}\|\mathbf{x}_{\mathbf{T}}\|_{\mathbf{2}} \quad (95)$$

and also,

$$\|\mathbf{\Phi}'_{\mathbf{I_K}}\mathbf{v}\|_{\mathbf{2}} \le \sqrt{1+\delta_K}\|\mathbf{v}\|_{\mathbf{2}}. \quad (96)$$

Using (95) and (96), we have

$$\|\mathbf{\Phi}'_{\mathbf{I_K}}\mathbf{y}\|_{\mathbf{2}} \le \delta_{\mathbf{2K}}\|\mathbf{x}_{\mathbf{T}}\|_{\mathbf{2}} + \sqrt{\mathbf{1}+\delta_{\mathbf{K}}}\|\mathbf{v}\|_{\mathbf{2}} \quad (97)$$

and since $\|\mathbf{\Phi}'_{\mathbf{I_K}}\mathbf{y}\|_{\mathbf{2}} \ge \sqrt{\mathbf{K}}\eta$, we have

$$\eta \le \frac{1}{\sqrt{K}}\left[\delta_{2K}\|\mathbf{x}_{\mathbf{T}}\|_{\mathbf{2}} + \sqrt{\mathbf{1}+\delta_{\mathbf{K}}}\|\mathbf{v}\|_{\mathbf{2}}\right], \quad (98)$$

which is the desired result.

VII. Proof of Lemma 15

Since $\Psi_l \subset T$ and $\beta^l = \max_{j\in T\setminus\Psi_l}|\phi'_j\mathbf{r}_{\mathbf{\Psi}_l}|$, we have

$$\begin{aligned}\max_{j\in T\setminus\Psi_l}|\phi'_j\mathbf{r}_{\mathbf{\Psi}_l}| &= \|\mathbf{\Phi}'_{\mathbf{T}\setminus\mathbf{\Psi}_l}\mathbf{r}_{\mathbf{\Psi}_l}\|_{\infty} \\ &\ge \frac{1}{\sqrt{|T\setminus\Psi_l|}}\|\mathbf{\Phi}'_{\mathbf{T}\setminus\mathbf{\Psi}_l}\mathbf{P}^{\perp}_{\mathbf{\Psi}_l}\mathbf{y}\|_{\mathbf{2}} \quad (99)\\ &= \frac{1}{\sqrt{|T\setminus\Psi_l|}}\|\mathbf{\Phi}'_{\mathbf{T}\setminus\mathbf{\Psi}_l}\mathbf{P}^{\perp}_{\mathbf{\Psi}_l}(\mathbf{\Phi}_{\mathbf{T}}\mathbf{x}_{\mathbf{T}}+\mathbf{v})\|_{\mathbf{2}} \quad (100)\\ &= \frac{1}{\sqrt{K-i-l}}\|\mathbf{\Phi}'_{\mathbf{T}\setminus\mathbf{\Psi}_l}\mathbf{P}^{\perp}_{\mathbf{\Psi}_l}\mathbf{\Phi}_{\mathbf{T}\setminus\mathbf{\Psi}_l}\mathbf{x}_{\mathbf{T}\setminus\mathbf{\Psi}_l} \\ &\quad+\mathbf{\Phi}'_{\mathbf{T}\setminus\mathbf{\Psi}_l}\mathbf{P}^{\perp}_{\mathbf{\Psi}_l}\mathbf{v}\|_{\mathbf{2}} \quad (101)\\ &\ge \frac{1}{\sqrt{K-i-l}}\|\mathbf{\Phi}'_{\mathbf{T}\setminus\mathbf{\Psi}_l}\mathbf{P}^{\perp}_{\mathbf{\Psi}_l}\mathbf{\Phi}_{\mathbf{T}\setminus\mathbf{\Psi}_l}\mathbf{x}_{\mathbf{T}\setminus\mathbf{\Psi}_l}\|_{\mathbf{2}} \\ &\quad-\frac{1}{\sqrt{K-i-l}}\|\mathbf{\Phi}'_{\mathbf{T}\setminus\mathbf{\Psi}_l}\mathbf{P}^{\perp}_{\mathbf{\Psi}_l}\mathbf{v}\|_{\mathbf{2}} \quad (102)\end{aligned}$$

where $\mathbf{P}^{\perp}_{\mathbf{\Psi}_l} = \mathbf{I} - \mathbf{\Phi}_{\mathbf{\Psi}_l}(\mathbf{\Phi}'_{\mathbf{\Psi}_l}\mathbf{\Phi}_{\mathbf{\Psi}_l})^{-\mathbf{1}}\mathbf{\Phi}'_{\mathbf{\Psi}_l}$ and (99) is from the inequality $\|\mathbf{u}\|_{\infty} \ge \frac{\mathbf{1}}{\sqrt{\|\mathbf{u}\|_{\mathbf{0}}}}\|\mathbf{u}\|_{\mathbf{2}}$ for any vector $\mathbf{u}$, and (102)

is from the triangle inequality. Recalling from (62) and (66) in Appendix C, the first component in (102) is bounded as

$$\|\mathbf{\Phi}'_{\mathbf{T}'\setminus\mathbf{\Psi}_1}\mathbf{P}^{\perp}_{\mathbf{\Psi}_1}\mathbf{\Phi}_{\mathbf{T}\setminus\mathbf{\Psi}_1}\mathbf{x}_{\mathbf{T}\setminus\mathbf{\Psi}_1}\|_{\mathbf{2}} > \frac{\mathbf{1}-\mathbf{2}\delta_{\mathbf{K}}}{\mathbf{1}-\delta_{\mathbf{K}}}\|\mathbf{x}_{\mathbf{T}\setminus\mathbf{\Psi}_1}\|_{\mathbf{2}}, \quad (103)$$

Next, $\|\mathbf{\Phi}'_{\mathbf{T}\setminus\mathbf{\Psi}_1}\mathbf{P}^{\perp}_{\mathbf{\Psi}_1}\mathbf{v}\|_{\mathbf{2}}$ is upper bounded as

$$\begin{aligned}\|\mathbf{\Phi}'_{\mathbf{T}\setminus\mathbf{\Psi}_1}\mathbf{P}^{\perp}_{\mathbf{\Psi}_1}\mathbf{v}\|_{\mathbf{2}} &\leq \|\mathbf{\Phi}'_{\mathbf{T}\setminus\mathbf{\Psi}_1}\|_{\mathbf{2}}\,\|\mathbf{P}^{\perp}_{\mathbf{\Psi}_1}\mathbf{v}\|_{\mathbf{2}} && (104)\\ &\leq \sqrt{1+\delta_{|T\setminus\Psi_l|}}\|\mathbf{P}^{\perp}_{\mathbf{\Psi}_1}\mathbf{v}\|_{\mathbf{2}} && (105)\\ &\leq \sqrt{1+\delta_{|T\setminus\Psi_l|}}\|\mathbf{v}\|_{\mathbf{2}} && (106)\\ &\leq \sqrt{1+\delta_{K-i-l}}\|\mathbf{v}\|_{\mathbf{2}}. && (107)\end{aligned}$$

By combining (102), (103), and (107), we obtain the desired result as

$$\begin{aligned}\beta^l \geq& \frac{1}{\sqrt{K-i-l}}[\|\mathbf{\Phi}'_{\mathbf{T}\setminus\mathbf{\Psi}_1}\mathbf{P}^{\perp}_{\mathbf{\Psi}_1}\mathbf{\Phi}_{\mathbf{T}\setminus\mathbf{\Psi}_1}\mathbf{x}_{\mathbf{T}\setminus\mathbf{\Psi}_1}\|_{\mathbf{2}}\\ &-\|\mathbf{\Phi}'_{\mathbf{T}\setminus\mathbf{\Psi}_1}\mathbf{P}^{\perp}_{\mathbf{\Psi}_1}\mathbf{v}\|_{\mathbf{2}}] && (108)\\ >& \frac{1}{\sqrt{K-i}}\left[\frac{1-2\delta_K}{1-\delta_K}\|\mathbf{x}_{\mathbf{T}\setminus\mathbf{\Psi}_1}\|_{\mathbf{2}}-\sqrt{\mathbf{1}+\delta_{\mathbf{K-i}}}\|\mathbf{v}\|_{\mathbf{2}}\right] && (109)\end{aligned}$$

Since $1 \leq i \leq K-1$, we obtain the desired result.

VIII. Proof of Lemma 16

From the definition of $\alpha^l = \min_{j\in D}|\phi'_j\mathbf{r}_{\mathbf{\Psi}_1}|$, we have

$$\sqrt{|D|}\alpha^l \leq \sqrt{\sum_{j\in D}|\phi'_j\mathbf{r}_{\mathbf{\Psi}_1}|^{\mathbf{2}}} = \|\mathbf{\Phi}'_{\mathbf{D}}\mathbf{r}_{\mathbf{\Psi}_1}\|_{\mathbf{2}} \quad (110)$$

where $D = \arg\max_{|I|=K-i, I\subset\Omega\setminus T'}\|\mathbf{\Phi}'_{\mathbf{I}}\mathbf{r}_{\mathbf{\Psi}^{\mathbf{l}}}\|_{\mathbf{2}}$. Using the triangle inequality, we have

$$\begin{aligned}\|\mathbf{\Phi}'_{\mathbf{D}}\mathbf{r}_{\mathbf{\Psi}_1}\|_{\mathbf{2}} &= \|\mathbf{\Phi}'_{\mathbf{D}}\mathbf{P}^{\perp}_{\mathbf{\Psi}_1}\mathbf{y}\|_{\mathbf{2}} && (111)\\ &= \|\mathbf{\Phi}'_{\mathbf{D}}\mathbf{P}^{\perp}_{\mathbf{\Psi}_1}(\mathbf{\Phi}_{\mathbf{T}}\mathbf{x}_{\mathbf{T}}+\mathbf{v})\|_{\mathbf{2}} && (112)\\ &= \|\mathbf{\Phi}'_{\mathbf{D}}\mathbf{P}^{\perp}_{\mathbf{\Psi}_1}\mathbf{\Phi}_{\mathbf{T}\setminus\mathbf{\Psi}_1}\mathbf{x}_{\mathbf{T}\setminus\mathbf{\Psi}_1}+\mathbf{\Phi}'_{\mathbf{D}}\mathbf{P}^{\perp}_{\mathbf{\Psi}_1}\mathbf{v}\|_{\mathbf{2}} && (113)\\ &\leq \|\mathbf{\Phi}'_{\mathbf{D}}\mathbf{P}^{\perp}_{\mathbf{\Psi}_1}\mathbf{\Phi}_{\mathbf{T}\setminus\mathbf{\Psi}_1}\mathbf{x}_{\mathbf{T}\setminus\mathbf{\Psi}_1}\|_{\mathbf{2}}\\ &\quad +\|\mathbf{\Phi}'_{\mathbf{D}}\mathbf{P}^{\perp}_{\mathbf{\Psi}_1}\mathbf{v}\|_{\mathbf{2}} && (114)\end{aligned}$$

Recalling from (81) in Appendix D, $\|\mathbf{\Phi}'_{\mathbf{D}}\mathbf{P}^{\perp}_{\mathbf{\Psi}_1}\mathbf{\Phi}_{\mathbf{T}\setminus\mathbf{\Psi}_1}\mathbf{x}_{\mathbf{T}\setminus\mathbf{\Psi}_1}\|_{\mathbf{2}}$ is upper bounded as

$$\begin{aligned}\|\mathbf{\Phi}'_{\mathbf{D}}\mathbf{P}^{\perp}_{\mathbf{\Psi}_1}\mathbf{\Phi}_{\mathbf{T}\setminus\mathbf{\Psi}_1}\mathbf{x}_{\mathbf{T}\setminus\mathbf{\Psi}_1}\|_{\mathbf{2}} &\leq \|\mathbf{\Phi}'_{\mathbf{D}}\mathbf{\Phi}_{\mathbf{T}\setminus\mathbf{\Psi}_1}\mathbf{x}_{\mathbf{T}\setminus\mathbf{\Psi}_1}\|_{\mathbf{2}}\\ &\quad +\|\mathbf{\Phi}'_{\mathbf{D}}\mathbf{P}_{\mathbf{\Psi}_1}\mathbf{\Phi}_{\mathbf{T}\setminus\mathbf{\Psi}_1}\mathbf{x}_{\mathbf{T}\setminus\mathbf{\Psi}_1}\|_{\mathbf{2}} && (115)\\ &< \frac{\delta_{2K-i}}{1-\delta_K}\|\mathbf{x}_{\mathbf{T}\setminus\mathbf{\Psi}_1}\|_{\mathbf{2}}. && (116)\end{aligned}$$

In addition, $\|\mathbf{\Phi}'_{\mathbf{D}}\mathbf{P}^{\perp}_{\mathbf{\Psi}_1}\mathbf{v}\|_{\mathbf{2}}$ is upper bounded as

$$\begin{aligned}\|\mathbf{\Phi}'_{\mathbf{D}}\mathbf{P}^{\perp}_{\mathbf{\Psi}_1}\mathbf{v}\|_{\mathbf{2}} &\leq \|\mathbf{\Phi}'_{\mathbf{D}}\|_{\mathbf{2}}\|\mathbf{P}^{\perp}_{\mathbf{\Psi}_1}\mathbf{v}\|_{\mathbf{2}} && (117)\\ &\leq \sqrt{1+\delta_{|D|}}\|\mathbf{P}^{\perp}_{\mathbf{\Psi}_1}\mathbf{v}\|_{\mathbf{2}} && (118)\\ &\leq \sqrt{1+\delta_{|D|}}\|\mathbf{v}\|_{\mathbf{2}}. && (119)\end{aligned}$$

By combining (110), (116), and (121), we get the desired result as

$$\begin{aligned}\alpha^l &\leq \frac{1}{\sqrt{|D|}}\left[\frac{\delta_{2K-i}}{1-\delta_K}\|\mathbf{x}_{\mathbf{T}\setminus\mathbf{\Psi}_1}\|_{\mathbf{2}}+\sqrt{\mathbf{1}+\delta_{\mathbf{|D|}}}\|\mathbf{v}\|_{\mathbf{2}}\right] && (120)\\ &< \frac{1}{\sqrt{K-i}}\left[\frac{\delta_{2K-i}}{1-\delta_K}\|\mathbf{x}_{\mathbf{T}\setminus\mathbf{\Psi}_1}\|_{\mathbf{2}}+\sqrt{\mathbf{1}+\delta_{\mathbf{K-i}}}\|\mathbf{v}\|_{\mathbf{2}}\right] && (121)\end{aligned}$$

Since $1 \leq i \leq K-1$, we obtain the desired result.

REFERENCES

[1] E. J. Candes, J. Romberg and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inform. Theory*, vol. 52, no. 2, pp. 489-509, Feb. 2006.

[2] E. J. Candes and T. Tao, "Decoding by linear programming," *IEEE Trans. Inform. Theory*, vol. 51, no. 12, pp. 4203-4215, Dec. 2005,

[3] E. Liu and V. N. Temlyakov, , "The orthogonal super greedy algorithm and applications in compressed sensing," *IEEE Trans. Inform. Theory*, vol. 58, no. 4, pp. 2040-2047, April 2012.

[4] E. J. Candes, "The restricted isometry property and its implications for compressed sensing," *Comptes Rendus Mathematique*, vol. 346, no. 9-10, pp. 589-592, May 2008.

[5] R. Tibshirani, "Regression shrinkage and selection via the Lasso," *J. Royal Stat. Soc. Series B*, vol. 58, no. 1, pp. 267-288, 1996.

[6] E. J. Candes and T. Tao, "The dantzig selector: statistical estimation when p is much larger than n," *The Annal. Stat.*, vol. 35, no. 6, pp. 2313-2351, Dec. 2007.

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

[8] 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, Sept. 2012.

[9] D. Needell and J. A. Tropp, "CoSaMP: iterative signal recovery from incomplete and inaccurate samples," *Commun. ACM*, vol. 53, no. 12, pp. 93-100, Dec. 2010.

[10] W. Dai and O. Milenkovic, "Subspace pursuit for compressive sensing signal reconstruction," *IEEE Trans. Inform. Theory*, vol. 55, no. 5, pp. 2230-2249, May 2009.

[11] T. Blumensath and M. E. Davies, "Iterative hard thresholding for compressed sensing," *Applied and Computational Harmonic Analysis*, vol. 27, no. 3, pp. 265-274, Nov. 2009.

[12] R. Garg and R. Khandekar, "Gradient descent with sparsification: an iterative algorithm for sparse recovery with restricted isometry property," *Proc. Annual International Conf. Machine Learning*, pp. 337-344, May 2009.

[13] S. Foucart, "Hard thresholding pursuit: an algorithm for compressive sensing," *SIAM J. Numerical Analysis*, vol. 49, no. 6, pp. 2543-2563, Dec. 2011.

[14] M. A. Khajehnejad, A. G. Dimakis, W. Xu and B. Hassibi, "Sparse recovery of nonnegative signals with minimal expansion," *IEEE Trans. Signal Process.*, vol. 59, no. 1, pp. 196-208, Jan. 2011.

[15] S. Qaisar, R. M. Bilal, W. Iqbal, M. Naureen and S. Lee, "Compressive sensing: from theory to applications, a survey," *J. Commun. and Networks*, vol. 15, no. 5, pp. 443-456, Oct. 2013.

[16] S. Kwon, J. Wang and B. Shim, "Multipath matching pursuit," *IEEE Trans. Inform. Theory*, vol. 56, no. 10, pp. 4867-4878, Oct. 2013.

[17] S. S. Chen, D. L. Donoho and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM J. Scientific Comput.*, pp. 33-61, 1998.

[18] W. Chen, M. R. D. Rodrigues, I. J. Wassell and L. Hanzo, "Projection design for statistical compressive sensing: a tight frame based approach," *IEEE Trans. Signal Process.*, vol. 61, no. 8, pp. 2016-2029, April 2013.

[19] T. Zhang, "Sparse recovery with orthogonal matching pursuit under RIP," *IEEE Trans. Inform. Theory*, vol. 57, no. 9, pp. 6215-6221, Sept. 2011.

[20] J. Wang, S. Kwon, and B. Shim, "Generalized orthogonal matching pursuit," *IEEE Trans. Signal Process.*, vol. 60, no. 12, pp. 6202-6216, Dec. 2012.

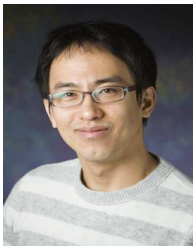
[21] G. D. Forney Jr., "The Viterbi algorithm," *Proc. IEEE*, pp. 268-278, March 1973.

[22] E. Viterbo and J. Boutros, "A universal lattice code decoder for fading channels," *IEEE Trans. Inf. Theory*, vol. 45, no. 5, pp. 1639-1642, July 1999.

[23] B. Shim and I. Kang, "Sphere decoding with a probabilistic tree pruning," *IEEE Trans. Signal Process.*, vol. 56, no. 10, pp. 4867-4878, Oct. 2008.

[24] J. Choi, B. Lee and B. Shim, "Iterative group detection and decoding for large MIMO systems," *J. Commun. and Networks*, vol 17. pp. 609-621, Dec. 2015.

Jaeseok Lee received the B.S. and Ph.D. degrees in the School of Information and Communication, Korea University, Seoul, Korea, in 2008 and 2015, respectively. Until 2015, he was with the Institute of New Media and Communications, Seoul National University, Seoul, Korea as a research engineer. He is currently a post doctoral researcher with the Department of Information and Communication Engineering, Daegu Gyeongbuk Institute of Science and Technology, Daegu, Korea. His research interests include sparse signal reconstruction and cyber-physical system security.



Jun Won Choi received the B.S. and M.S. degrees in Electrical and Computer Engineering, Seoul National University and earned Ph. D. degree in Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, respectively. In 2010, he joined Qualcomm, San Diego USA and participated in wireless communication system/algorithm design for commercializing LTE and LTE-A modem chipsets. Since 2013, he has been a faculty member of the Department of Electrical Engineering, Hanyang University and is leading signal processing and optimization research group. His research area includes wireless communications, signal processing, optimization, data analytic, and machine learning.

Byonghyo Shim 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 Institute of New Media and Communications and School 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).