Improving A*OMP: Theoretical and Empirical Analyses With a Novel Dynamic Cost Model

Nazim Burak Karahanoglu, Student Member, IEEE, and Hakan Erdogan, Member, IEEE

Abstract—Best-first search has been recently utilized for the compressed sensing (CS) signal recovery problem by A* orthogonal matching pursuit (A*OMP). In this work we mainly concentrate on the theoretical analysis of A*OMP. First of all, we develop a restricted isometry property (RIP)-based condition for exact recovery of sparse signals via A*OMP. In addition, we present a theoretical foundation for the improved recovery performance with the residue-based termination instead of the sparsity-based one. We support our findings with extensive experiments using the adaptive-multiplicative (AMul) cost model, which effectively compensates for the path length differences in the search tree. The presented results, involving phase transitions as well as recovery rates and average error for noisy and noise-free sparse signals with different nonzero element distributions, not only reveal the superior recovery accuracy of A*OMP, but also demonstrate the improvements promised by the residue-based termination criterion. In addition, comparison of run times indicate the speed up by the AMul cost model. We also present a theoretical foundation for the improved recovery guarantees of many sparse recovery algorithms [3], [9]–[12], iterative reweighted methods [22]–[24], etc.

B. Restricted Isometry Property

Restricted isometry property (RIP) [3] has been acknowledged as an important means for obtaining theoretical guarantees in sparse signal recovery and approximation problems. A matrix $\Phi$ is said to satisfy the $L$-RIP if there exists a restricted isometry constant (RIC) $\delta_L \in (0, 1)$ satisfying

$$(1 - \delta_L)\|x\|_2^2 \leq \|\Phi x\|_2^2 \leq (1 + \delta_L)\|x\|_2^2$$

for all $x$ such that $\|x\|_0 \leq L$.

Some families of random matrices, such as those with independent and identically distributed Gaussian or Bernoulli entries and random selections from discrete Fourier transform, satisfy the $L$-RIP with high probabilities if $L$, $M$ and $N$ satisfy some specific conditions [25], [26]. Exploiting this property of such dictionaries, RIP has been utilized for theoretical guarantees of many sparse recovery algorithms [3], [9]–[12], [26]–[34].

C. A* Orthogonal Matching Pursuit

A* orthogonal matching pursuit (A*OMP) [1], [35] is an iterative semi-greedy approach that utilizes best-first search to find an approximation of $x$. A*OMP performs A* search [36], [37] on a search tree where each node addresses a dictionary atom, and each path is a support estimate for $x$. A*OMP possesses effective pruning mechanisms and appropriate cost models in order to make this search tractable.

Before summarizing A*OMP, let us first clarify the notation in this paper: We define $S$ as the set of all paths in the search tree, $T$ is the correct support of $x$, $T^i$, $r^i$, $b^i$ and $f(T^i)$ denote the support estimate, residue, length and cost of the $i$th path, respectively. $\hat{x}^i$ is the estimate of $x$ by the $i$th path. The best path at a certain time during the search is referred to as $b$. $\Delta T$ represents the set of indices selected during the expansion of path $b$, i.e., the indices of $B$ largest magnitude elements in $\Phi^* r^b$, where $\Phi^*$ denotes the conjugate of $\Phi$. $\phi_j$ is the $j$th column of $\Phi$, that is, $\Phi = [\phi_1 \phi_2 \ldots \phi_N]$, $\Phi_{\mathcal{J}}$ denotes the matrix composed of the columns of $\Phi$ indexed by the index set $\mathcal{J}$. Similarly, $x_{\mathcal{J}}$ is the vector of the elements of $x$ indexed by the set $\mathcal{J}$. 

I. INTRODUCTION

A. Compressed Sensing

The fundamental goal of compressed sensing (CS) is to unify data acquisition and compression. This is accomplished by observing a lower dimensional set of measurements

$$y = \Phi x$$

instead of the signal $x \in \mathbb{R}^N$ itself, where $\Phi \in \mathbb{R}^{M \times N}$ is the (mostly random) dictionary and $y \in \mathbb{R}^M$ is the observation vector. The dimensionality reduction is achieved by selecting $M < N$, as a result of which $x$ cannot be directly solved back from $y$. Alternatively, assuming it is $K$-sparse (i.e. it has at most $K$ nonzero components), or compressible, $x$ can be recovered by

$$x = \arg \min \|x\|_0 \quad s.t. \quad y = \Phi x,$$  \hspace{1cm} (1)

where $\|x\|_0$ denotes the number of nonzero elements in $x$.

As the direct solution of (1) is infeasible, a number of approximate solutions have emerged in CS literature. Among these, convex optimization algorithms [2]–[4] relax (1) by replacing $\|x\|_0$ with its closest convex approximation $\|x\|_1$. On the other hand, greedy algorithms [5]–[15] provide simple and approximate solutions via iterative residue minimization. Other CS recovery schemes include Bayesian methods [16]–[18], nonconvex approaches [19]–[21], iterative reweighted methods [22]–[24], etc.
A*OMP initializes the search tree with $I$ paths, each consisting of a single node. These nodes represent the indices of the $I$ largest magnitude elements in $\Phi^* y$. At each iteration, the algorithm first selects the best path $b$ with the minimum cost criterion, which we discuss below. Then, $\Delta T$ is chosen as the indices of the $B$ largest magnitude elements in $\Phi^* r^p$. This implies $B$ candidate paths, each of which expands $b$ with a single index in $\Delta T$. Each candidate path is added to the tree only if an equivalent path has not been explored during the previous stages of the search (equivalent path pruning [35]). For each new path $i$, $r^i$ is computed as the projection error of $y$ onto $\Phi_{T_i}$, and the cost, which is a function of $r^i$, is calculated. Finally, the tree is pruned such that only the best $P$ paths with minimum cost remain. Selection and expansion of the best path are repeated until either $|T^b| \geq K_{\text{max}}$, or $\|r^b\|_2 \leq \varepsilon \|y\|_2$. The pseudo-code for the A*OMP algorithm is given in Algorithm 1.

Termination parameters in Algorithm 1, $K_{\text{max}}$ and $\varepsilon$, provide means for employing different termination criteria. In [35], the search is terminated when the best path has $K$ nodes, i.e., $K_{\text{max}} = K$ and $\varepsilon = 0$. We call this sparsity-based termination, and denote the corresponding variant of A*OMP by A*OMP$_{K}$. Another alternative is the residue-based termination such as in [11], where $K_{\text{max}} > K$ and problem-specific $\varepsilon$ is selected small enough based on the noise level. This version, which runs until the approximation error in $y$ gets small enough, is referred to as A*OMP$_{\varepsilon}$. Preliminary results in [11] indicate that A*OMP$_{\varepsilon}$ yields not only better recovery but also faster termination than A*OMP$_{K}$. Note that with the flexibility on the choice of $K_{\text{max}}$ and $\varepsilon$, Algorithm 1 commonly represents both versions of A*OMP.$^*$

To incorporate the best path selection using the minimum cost criterion, A*OMP should appropriately compare the costs of paths with different lengths. This rises the need for proper cost models which can effectively compensate for the path length differences. For this purpose, some novel models have been proposed in [35] and [11]. This issue is addressed in Section III in more details.

D. Outline and Contributions

The manuscript at hand concentrates on the theoretical and empirical analyses of the A*OMP algorithm with a novel dynamic cost model. Our main contribution is the theoretical analysis of A*OMP. For this purpose, we employ a method similar to the analyses of the orthogonal matching pursuit (OMP) algorithm presented in [29], [30] and [38]. The results we develop include not only exact recovery guarantees, but also theoretical comparison of two different termination criteria. The former states a RIP condition for exact recovery of sparse signals from noise-free measurements via A*OMP. The latter provides a theoretical understanding of the improvements in the recovery performance when the residue-based termination is employed instead of the sparsity-based one. In addition to the theoretical results, we discuss a novel dynamic cost model which significantly enhances the efficiency of the search, before finally providing extensive empirical recovery analyses of A*OMP. These empirical results demonstrate two important aspects: First, the residue-based termination criterion improves not only the accuracy but also the speed of the A*OMP recovery. Second, the adaptive-multiplicative cost model reduces the number of explored paths, which further accelerates the search.

In Section II, we discuss the adaptive-multiplicative cost model, which is an adaptive extension of the multiplicative cost model developed in [35]. This cost model allows for more flexibility when choosing the cost model parameter. As a result, it significantly accelerates the recovery by reducing the number of nodes explored during the search.

We present our main theoretical contributions in Section III We first develop a RIP-based condition for the success of a single A*OMP iteration in Section III-B. This result forms a basis for the rest of the theoretical analysis. In Section III-C we derive a RIP condition for exact recovery via A*OMP$_{K}$. As intuitively expected, this condition turns out to be less restrictive than the one developed in [29] and [30] for the exact recovery of OMP in exactly $K$ steps. In Section III-D we develop conditions for exact recovery of a sparse signal via A*OMP$_{\varepsilon}$. In addition, we show that exact recovery guarantees of A*OMP$_{K}$ represent a special case of these conditions. Section III-F compares the recovery conditions of A*OMP$_{\varepsilon}$ and A*OMP$_{K}$. This clarifies that A*OMP$_{\varepsilon}$ possesses a less restrictive exact recovery condition than A*OMP$_{K}$.

Section IV presents extensive empirical analyses of A*OMP$_{\varepsilon}$ in comparison to basis pursuit (BP) [2], subspace pursuit (SP) [10], OMP [6], iterative hard thresholding (IHT) [11], iterative support detection (ISD) [22] and smoothed $\ell_0$ (SL0) [19]. The most important contribution of these simulations is the phase transition graphs which are obtained by running a set of computationally expensive experiments

---

**Algorithm 1: A* ORTHOGONAL Matching Pursuit**

Initialize: $T^i = \emptyset$, $r^i = y \forall i = 1, 2, ..., P$

$\Delta T = \arg \max_{\mathcal{J}, |\mathcal{J}|=1} \sum_{j \in \mathcal{J}} \langle \phi_j, y \rangle$

for $i = 1$ to $I$ do

$n = i$th index in $\Delta T$

$T^i = T^i \cup \{n\}$

$r^i = y - \langle y, \phi_n \rangle \phi_n$

end for

$i = 1, 2, ..., P$

while $(l^b < K_{\text{max}})$ & $(\|r^b\|_2 > \varepsilon \|y\|_2)$ do

$\Delta T = \arg \max_{\mathcal{J}, |\mathcal{J}|=B} \sum_{j \in \mathcal{J}} |\langle \phi_j, r^b \rangle|$

$T^b = T^b \cup \{j\}$

$b = \arg \min_{i \in 1, 2, ..., P} |f(T^i)|$

$p = \arg \max_{i \in 1, 2, ..., P} f(T^i)$

end while

return $T^b$
involving different signal characteristics. These reveal the recovery abilities of \( A^*\text{OMP}_e \). We also investigate recovery rates and average recovery error for noisy and noise-free cases. Comparison of run times illustrates the acceleration of the algorithm with the AMul cost function and the residue-based termination. Moreover, we test a hybrid of OMP and \( A^*\text{OMP}_e \) for further speed up. Finally, \( A^*\text{OMP}_e \) is demonstrated on a sparse image.

The theoretical guarantees developed in this paper have some requirements which we call the optimality conditions. As explained in Section III, it is not practical to configure the search such that these conditions are theoretically guaranteed. The validity of these conditions in practice can be justified by simple intuitive reasoning as discussed in Section III-E. In addition, we present strong empirical evidence supporting this claim in Section IV.C via investigation of optimal paths throughout the search.

Some parts of this work have been presented at EU-SIPCO’2012 [1]. These include introduction of the AMul cost model and some preliminary experiments. In addition to more insightful discussions of these issues, the most important contribution of this work is the theoretical analysis of \( A^*\text{OMP} \). Empirical analyses are also significantly enriched over the preliminary analyses in [1] by phase transitions which require an extensive computational power, demonstration on an image, the hybrid approach and the optimality analyses.

II. THE ADAPTIVE-MULTIPLICATIVE COST FUNCTION FOR \( A^*\text{OMP} \)

In order to handle paths with different lengths, \( A^*\text{OMP} \) requires properly defined cost functions. The choice of the cost function plays a major role in the performance of the algorithm, especially in the complexity-accuracy trade-off. Among the structures introduced in [35], the multiplicative (Mul) cost function is the most important contribution of this work is the theoretical analysis of \( A^*\text{OMP} \).

Note that, we replace \( K \) in [35] with \( K_{\text{max}} \) in order to employ the model with different termination criteria. The simulations in [35] demonstrate that decreasing \( \alpha_{\text{Mul}} \) improves recovery accuracy, while the search also takes longer.

Adaptive cost models can adapt themselves to the actual decrement in the residue. Being motivated by the empirical improvements with adaptive cost structures in [35], we define a dynamic extension of the Mul model, which is called the adaptive-multiplicative (AMul) cost model:

\[
\alpha_{\text{AMul}}(T^i) = \alpha_{\text{Mul}} \left( \frac{r_i^j}{r_{i-1}^j} \right)^{K_{\text{max}} - l^i},
\]

where \( r_i^j \) denotes the residue after the first \( l \) nodes of the path \( i \), and \( \alpha_{\text{AMul}} \in (0, 1] \) is the cost model parameter similar to the Mul model.

The AMul cost model relies on the following assumption: Each unexplored node reduces the residue by a rate proportional to the decay occurred during the last expansion of the path. This rate is modeled by the auxiliary term

\[
\alpha_{\text{AMul}} \frac{r_i^j}{r_{i-1}^j}^2 / \frac{r_i^{j-1}}{r_{i-1}^{j-1}}^2
\]

while the exponent \( K_{\text{max}} - l^i \) extends the auxiliary function to all unexplored nodes along path \( i \). The motivation behind this choice can be intuitively explained as follows: As the search is expected to select nodes in the order of descending absolute inner-products with \( y \), a node is more likely to reduce \( \|r^j\|_2 \) less than its ancestors do. Note that, this condition may obviously be violated for some particular nodes. However, the auxiliary term of the cost function is mostly computed over a set of nodes instead of a single one. Hence, it is practically sufficient if the decay of the residue obeys this assumption over a set of nodes. Moreover, the tree usually contains multiple paths which might lead to the correct solution if chosen. Therefore, that some particular paths violate this assumption does not actually harm the recovery performance. Hence, AMul cost model needs to be valid only on average. That is, any particular sequence of nodes may violate it, however, we expect it to hold in general and lead the search to the correct solution. This behavior is indeed similar to the other cost models introduced in [35].

As for the cost model parameter \( \alpha \), the adaptive structure of the AMul cost model allows for a larger choice than the Mul model does. This reduces the auxiliary term on average and makes the search favor longer paths. Consequently, the search explores fewer nodes and terminates faster. This speed up is demonstrated in Section IV.D.

The validity of the Mul and AMul cost functions with different choices of \( \alpha \) is also addressed by the empirical optimality analyses presented in Section IV.C. The results of these analyses indicate that these cost models are still useful in practice though the assumptions behind them may not be completely valid.

III. THEORETICAL ANALYSIS OF \( A^*\text{OMP} \)

A. Preliminaries

We first concentrate on some preliminary results which are necessary for the analysis in the rest of this chapter. Though some of these are already well-known in the CS community, we believe it is worth to present them shortly for the sake of completeness.

**Lemma 1 (Monotonicity of the RIC):** Let \( R, S \) be positive integers such that \( R > S \). Then,

\[
\delta_R \geq \delta_S.
\]

**Lemma 2 (Consequence of RIP):** Let \( \mathcal{I} \subset \{1, 2, \ldots, N\} \). For any arbitrary vector \( z \in \mathbb{R}^{|\mathcal{I}|} \), RIP directly leads to

\[
(1 - \delta_{|\mathcal{I}|}) \|z\|_2 \leq \|\Phi_{\mathcal{I}} \Phi \mathcal{J} \|_2 \leq (1 + \delta_{|\mathcal{I}|}) \|z\|_2.
\]

**Lemma 3 (Lemma 1 in [10]):** Let \( \mathcal{I}, \mathcal{J} \subset \{1, 2, \ldots, N\} \) such that \( \mathcal{I} \cap \mathcal{J} = \emptyset \). For any arbitrary vector \( z \in \mathbb{R}^{|\mathcal{J}|} \),

\[
\|\Phi_{\mathcal{I}} \Phi \mathcal{J} z\|_2 \leq \delta_{|\mathcal{I}|+|\mathcal{J}|} \|z\|_2.
\]

**Lemma 4 (Direct consequence of Corollary 2 in [38]):** For positive integers \( K \) and \( B \),

\[
\delta_{K+B} > \frac{\delta_3(K/2)}{3},
\]
where $[z]$ denotes the ceiling of $z$, i.e. the smallest integer greater than or equal to $z$.

Proof: Corollary 2 of (3) states that Lemma 4 holds for $B = 1$. By monotonicity of the RIC, $\delta_{K+B} \geq \delta_{K+1}$ when $B > 1$. Hence, Lemma 4 also holds for $B > 1$. \hfill \blacksquare

Lemma 5: Assume $K \geq (3 + 2\sqrt{B})^2$. There exists at least one positive integer $n_c < K$ such that

$$\frac{3\sqrt{B}}{\sqrt{K} + \sqrt{B}} \leq \frac{\sqrt{B}}{\sqrt{K} - n_c + \sqrt{B}}. \quad (3)$$

Moreover, $n_c$ values which satisfy (3) are bounded by

$$K > n_c \geq \frac{8K + 4\sqrt{BK} - 4B}{9}. \quad (4)$$

Proof: Set $K - n_c = sK$ where $0 < s < 1$, and replace this into (3):

$$\frac{3\sqrt{B}}{\sqrt{K} + \sqrt{B}} \leq \frac{\sqrt{B}}{sK + \sqrt{B}}.$$ It can trivially be shown that $s$ is bounded by

$$0 < s \leq \left(\frac{\sqrt{K} - 2\sqrt{B}}{3\sqrt{K}}\right)^2.$$ Then, we obtain the lower bound for $n_c$ as

$$n_c = (1 - s)K \geq \frac{8K + 4\sqrt{BK} - 4B}{9}. \quad (5)$$

Since $n_c < K$, we get $sK = K - n_c \geq 1$. This translates as

$$K \geq \frac{1}{s} \geq \left(\frac{3\sqrt{K}}{\sqrt{K} - 2\sqrt{B}}\right)^2,$$

from which we deduce the assumption $K \geq (3 + 2\sqrt{B})^2$. Combining this with (5), we conclude that $n_c$ values bounded by (4) satisfy (3) when $K \geq (3 + 2\sqrt{B})^2$. \hfill \blacksquare

B. Success Condition of an A*OMP Iteration

Let us define success of an A*OMP iteration as $\Delta T$ containing at least one correct index, i.e. $\Delta T \cap \{T - T^b\} \neq \emptyset$. Then, we can establish a RIP condition for the success of an iteration given the number of correct and incorrect indices in $T^b$ as follows:

Theorem 1: Let $n_c = |T^b \cap \bar{T}|$ and $n_f = |T^b - \bar{T}|$. When this path is expanded, at least one index in $\Delta T$ is in the support of $x$, i.e. $\Delta T \cap \{T - T^b\} \neq \emptyset$ if $A$ satisfies RIP with

$$\delta_{K+n_f+B} < \min \left(\frac{\sqrt{B}}{\sqrt{K} - n_c + \sqrt{B}}, \frac{1}{2}\right). \quad (6)$$

Proof:

Remember that $\Delta T$ is defined as

$$\Delta T = \arg \max_{J \cup \{J\} = B} \sum_{j \in J} |\langle \phi_j, r^b \rangle|,$$

which is equivalent to

$$\Delta T = \arg \max_{\mathcal{J}, |\mathcal{J}| = B} \| \Phi_{\mathcal{T}}^* r^b \|_2. \quad (7)$$

Since $r^b$ is the orthogonal projection error of $y$ onto $\Phi_T^*$, $r^b = \Phi_T r^b$. Therefore, $\langle \phi_i, r^b \rangle = 0$ for all $i \in T^b$. Hence,

$$\| \Phi_{T \cup T^b}^* r^b \|_2^2 = \sum_{i \in T \cup T^b} |\langle \phi_i, r^b \rangle|^2 = \sum_{i \in T - T^b} |\langle \phi_i, r^b \rangle|^2. \quad (8)$$

We observe that (8) has only $K - n_c$ nonzero terms. Combining (7), (8) and the norm inequality, we obtain

$$\| \Phi_{\Delta T}^* r^b \|_2^2 = \max_{\mathcal{J}, |\mathcal{J}| = B} \| \Phi_{\mathcal{T}}^* r^b \|_2^2 \geq c \| \Phi_{\mathcal{T} \cup \mathcal{T}^b}^* r^b \|_2^2; \quad \text{where } c \text{ is defined as}$$

$$c \equiv \min \left(\sqrt{1 - \frac{B}{K - n_c}}, 1\right).$$

Then, the residue can be written as

$$r^b = y - \Phi_T x_T = \Phi_T x_T - \Phi_T \hat{x}_T^b = \Phi_{T \cup T^b} z.$$

where $z \in \mathbb{R}^{K+n_f}$. Using Lemma 2 and 9, we write

$$\| \Phi_{\Delta T}^* r^b \|_2^2 \leq \delta_{K+n_f+B} \|z\|_2 \leq c(1 - \delta_{K+n_f}) \|z\|_2. \quad (10)$$

Now, suppose that $\Delta T \cap T = \emptyset$. Then,

$$\| \Phi_{\Delta T}^* r^b \|_2^2 = \| \Phi_{\Delta T}^* \Phi_T z_b \|_2 \leq \delta_{K+n_f+B} \|z\|_2$$

by Lemma 5. Clearly, this never occurs if

$$c(1 - \delta_{K+n_f}) \|z\|_2 > \delta_{K+n_f+B} \|z\|_2$$

or equivalently

$$\frac{\delta_{K+n_f+B}}{c} + \delta_{K+n_f} < 1. \quad \text{or equivalently} \quad \delta_{K+n_f+B} < \frac{c}{1 + c}.$$ (11)

Following monotonicity of RIC, $\delta_{K+n_f+B} \geq \delta_{K+n_f}$. Hence, (12) is satisfied when $(\frac{1}{c} + 1) \delta_{K+n_f+B} < 1$, or equivalently

$$\delta_{K+n_f+B} < \frac{c}{1 + c} < \min \left(\frac{\sqrt{B}}{\sqrt{K} - n_c + \sqrt{B}}, \frac{1}{2}\right).$$

by definition of $c$. This guarantees that $\Delta T \cap T \neq \emptyset$. Moreover, since $\langle \phi_i, r^b \rangle = 0$ for all $i \in T^b$, we know that $\Delta T \cap T = \emptyset$. Hence, we conclude $\Delta T \cap \{T - T^b\} \neq \emptyset$, that is the A*OMP iteration is successful when (6) holds. \hfill \blacksquare

Theorem 1 does not directly imply any exact recovery guarantees. However, it is used below as a basis for exact recovery of A*OMP. Note that we assume $\sqrt{B} \leq \sqrt{K - n_c}$ in the rest of this paper and skip the term $\frac{1}{2}$ in Theorem 1 for simplicity. This assumption can be justified by the fact that $B$ is chosen small, such as 2 or 3, in practice.
C. Exact Recovery Conditions for $A^*\text{OMP}_K$

Exact recovery via $A^*\text{OMP}_K$ requires some conditions on the best path selection in addition to Theorem 1. For this purpose, we introduce the following definitions:

**Definition 1 (Complete path):** Path $i$ is called complete if $l^i = K_{\text{max}}$. (For $A^*\text{OMP}_K$, this translates as $l^i = K$ since $K_{\text{max}} = K$.)

**Definition 2 (Optimal path):** Path $i$ is called optimal if $T^i \subset T$.

**Definition 3 (Optimal pruning):** Pruning is defined as optimal if it does not remove all optimal paths from the tree.

**Definition 4 (Optimal cost condition):** The optimal cost condition is defined as

$$F(\hat{T}^i) < F(T^i), \forall \hat{T}^i \in S^{\text{opt}}, \forall T^j \in \{S_K - S^{\text{opt}}\},$$

where $S_K$ and $S^{\text{opt}}$ denote the sets of all complete paths, and all optimal paths, respectively.

In words, the optimal cost condition assures that the cost of an optimal path is lower than that of any suboptimal complete path. Once satisfied, this guarantees that $A^*\text{OMP}_K$ either terminates at an optimal path or there are no optimal paths in the search tree.

Theorem 2 exploits these definitions to develop exact recovery guarantees for $A^*\text{OMP}_K$:

**Theorem 2:** Assume that the optimal cost condition holds and the pruning is optimal. Set $I \geq B$. Then, $A^*\text{OMP}_K$ perfectly recovers any $K$-sparse signal from noise-free measurements if the observation matrix $F$ satisfies RIP with

$$\delta_{K+B} < \frac{\sqrt{B}}{\sqrt{K} + \sqrt{B}}. \quad (13)$$

**Proof:** Let us start with the initialization. Replacing $n_c = n_f = 0$ into Theorem 1 (13) assures success of the first iteration since $I \geq B$.

Next, consider that $A^*\text{OMP}_K$ selects an optimal path of length $l$, where $n_c = l$ and $n_f = 0$. By Theorem 1 expansion of this path is successful if

$$\delta_{K+B} < \frac{\sqrt{B}}{\sqrt{K-l} + \sqrt{B}}. \quad (14)$$

which is already satisfied when (13) holds.

Now, there exists a number of optimal paths after initialization. Moreover, expanding an optimal path introduces at least one longer optimal path, and by assumption pruning cannot remove all of the optimal paths. Altogether, these guarantee existence of at least one optimal path in the tree at any iteration. Under these conditions, the optimal cost assumption assures selection of optimal paths before termination. Consequently, the search cannot terminate at a suboptimal path. Instead, optimal paths evolve until the search terminates at an optimal complete path. Therefore, we conclude that exact recovery of any $K$-sparse signal is guaranteed via $A^*\text{OMP}_K$ if (13), and the other assumptions are satisfied.

To get a better understanding of Theorem 2, the optimality assumptions, which may at first seem unnatural, should be discussed in detail. We postpone this until Section III-E where we discuss these assumptions together with their $A^*\text{OMP}_e$ counterparts. In addition, we provide an empirical analysis of optimal paths in Section V to support the validity of these conditions in practice. For now, note that these assumptions are not only reasonable but also necessary in practice, since no efficient cost model may guarantee them for all $K$-sparse signals.

We observe that the RIP condition $\delta_{K+1} < \frac{1}{\sqrt{K+1}}$, which guarantees exact recovery via OMP in exactly $K$ steps, can be obtained as a special case of Theorem 2 when $B = I = 1$. Moreover, when the bounds for OMP and $A^*\text{OMP}_K$ are compared, (13) is clearly less restrictive, which explains the improved recovery accuracy of $A^*\text{OMP}_K$ over OMP.

D. Exact Recovery with $A^*\text{OMP}_e$

In order to develop recovery guarantees for $A^*\text{OMP}_e$, the definitions of the previous section should be extended to cover for the residue-based termination which typically employs $K_{\text{max}} > K$. First, note that path $i$ is now complete if $l^i = K_{\text{max}} > K$. In addition, we introduce the following definitions:

**Definition 5 (Potentially-optimal path):** Path $i$ is called potentially-optimal ($p$-optimal) if it satisfies $K + n_f \leq K_{\text{max}}$.

**Definition 6 (Potentially-optimal pruning):** Pruning is defined as $p$-optimal if it does not remove all $p$-optimal paths from the search tree.

**Definition 7 (Potentially-optimal cost condition):** The $p$-optimal cost condition is defined as

$$F(\hat{T}^i) < F(T^i), \forall \hat{T}^i \in S^{p-\text{opt}}, \forall T^j \in \{S_{K_{\text{max}}} - S^{p-\text{opt}}\},$$

where $S_{K_{\text{max}}}$ and $S^{p-\text{opt}}$ denote the set of all complete paths, and the set of all $p$-optimal paths, respectively.

Next, we state the following lemma which we exploit later while deriving the recovery condition for $A^*\text{OMP}_e$:

**Lemma 6:** Let $i$ be a $p$-optimal path with $n_c$ correct and $n_f$ incorrect indices. If path $i$ satisfies

$$\delta_{K+n_f+B} < \frac{\sqrt{B}}{\sqrt{K-n_c} + \sqrt{B}}. \quad (15)$$

expansion of path $i$ by its best $B$ children introduces at least one $p$-optimal path with $n_c + 1$ correct indices. Moreover, all $p$-optimal paths introduced by the expansion of path $i$ satisfy (15).

**Proof:** By Theorem 1 expansion of path $i$ is successful when (15) holds. Hence, it introduces at least one path, say $j$, with $n_c + 1$ correct and $n_f$ incorrect indices. It is clear that path $j$ is also $p$-optimal. Moreover, the upper bounds which (15) imposes on the RIC for the paths $i$ and $j$ are related as

$$\frac{\sqrt{B}}{\sqrt{K-n_c} + \sqrt{B}} < \frac{\sqrt{B}}{\sqrt{K-n_c} + 1 + \sqrt{B}}$$

where the left and right hand sides are the upper bounds for path $i$ and $j$, respectively. Because the upper bound is larger for path $j$, and path $i$ satisfies (15), path $j$ should also satisfy (15).

1This definition depends on the observation that a $p$-optimal path may evolve into a superset of the correct support after a series of expansions carried out up to a maximum of $K_{\text{max}}$ nodes.
Now the recovery condition of \( A^*\text{OMP}_e \) can be presented as follows:

**Theorem 3:** Set \( \varepsilon = 0 \), and \( K_{\text{max}} \leq M - K \). Let \( \Phi \) be full rank. Assume that the p-optimal cost condition holds and the pruning is p-optimal. Then, \( A^*\text{OMP}_e \) perfectly recovers a \( K \)-sparse signal from noise-free measurements if the search, at any step, expands a path which satisfies \( K + n_f \leq K_{\text{max}} \) and

\[
\delta_{K+n_f+B} < \frac{\sqrt{B}}{\sqrt{K-n_c}+\sqrt{B}}
\]

**Proof:** First, by \( K + n_f \leq K_{\text{max}} \), the best path \( b \) at this certain iteration is p-optimal. Via Lemma 6 (16) guarantees that at least one child of \( b \) will be p-optimal. Combining this with p-optimal pruning, existence of p-optimal paths is guaranteed until termination of the search.

On the other hand, as \( \varepsilon = 0 \), termination of the search requires that the residue vanishes. Since \( K_{\text{max}} \leq M - K \) and \( \Phi \) is full rank, residue may vanish if and only if \( T \) is a subset of the support estimate. Hence, the search terminates either when the recovery is successful with \( ||\Phi^b||_2 = 0 \), or a complete but not p-optimal path is selected as the best path, where \( ||\Phi^b||_2 > 0 \). By the p-optimal condition, the latter cannot happen once the existence of p-optimal paths is guaranteed by Lemma 6 and p-optimal pruning. Therefore, we conclude that p-optimal paths must be iteratively chosen for expansion.

Now, via Lemma 6 expansion of a p-optimal path satisfying (16) introduces at least a new p-optimal path satisfying (16) with a larger number of correct indices. It is obvious that iterative expansion of such paths leads to a superset of \( T \) with maximum \( K_{\text{max}} \) elements. For such a path, the residue is null and the search terminates. Finally, the orthogonal projection of \( y \) onto this set gives the correct solution.

Note that the condition \( \varepsilon = 0 \), which is stated in Theorem 3, may at first seem annoying. In addition, the readers may question why we do not aim at providing guarantees for such conditions. In fact, the answer is quite simple: There exists hardly any “practical” cost model which may provide such theoretical guarantees. Moreover, these assumptions are not only necessary but also quite reasonable in practice. We discuss these issues in more detail below, where we refer to both optimality and p-optimality assumptions as optimality for clarity of the presentation, while the discussion is valid for both.

First, obtaining theoretical guarantees for the optimality conditions is neither trivial nor practical. The former follows from the fact that we cannot exactly predict the decay of the residue along unexplored nodes since we know neither the magnitudes of the nonzero elements, nor the order which the algorithm will select them. Yet, the latter is more vital: Since the actual decay in the residue cannot be properly known, no cost model may guarantee the optimality condition for all \( K \)-sparse signals unless it explores too many paths during the search. In practice, the cost model should be “cautious” while selecting paths for expansion, i.e. it should explore as few nodes as possible. That is, the risk of making suboptimal decisions should be undertaken for the sake of tractability. Therefore, practical implementations of \( A^*\text{OMP} \) have to deal with cost models which do not guarantee optimality, such as the ones we employ. Because of the same reason, we do not attempt at providing guarantees for the optimality conditions.

Furthermore, pruning is also unavoidable for the algorithm to be tractable. Since it removes all but the \( P \) paths with minimum cost, the optimality of the pruning strictly depends on the selected cost model. Because of the reasons discussed above, optimality of neither the pruning nor cost models may hardly be analytically guaranteed for practical choices.

Though hardly guaranteed as being analytically optimal, there is still some intuitive explanation why the optimality conditions are reasonable in practice. First of all, there are \( K! \) possible paths which represent the correct solution since we are not interested in a particular ordering of the nodes. That is, there exists mostly a large number of optimal paths in the tree in case (13) holds, while it is sufficient that not all but only one of them satisfies the optimality assumptions. Moreover, since they potentially represent a solution that exactly satisfies \( y = \Phi x \), optimal paths tend to have very small costs. This makes them not only less likely to be pruned during the search, but also less likely to have higher costs than suboptimal complete paths.

In addition to this intuitive understanding, the optimality conditions can also be justified empirically by analyzing the number of optimal paths during recovery. For this purpose, we present the average number of optimal paths explored during the recovery simulations in Section [VC]. These graphs illustrate the validity of the optimality assumptions in practice.
F. Comparison of Termination Criteria

Though Theorem 2 can be obtained a special case of Theorem 3 we have not yet clarified whether it is possible to satisfy Theorem 3 despite failure of Theorem 2. In other words, we question whether the search may find a p-optimal path satisfying (10) even when (13) fails. We address this issue in the following theorem.

Theorem 4: Assume $K \geq (3 + 2\sqrt{B})^2$. If $1 \leq n_f + B \leq \lceil K/2 \rceil$ and $n_c$ satisfies (14) at some intermediate iteration, (10) becomes less restrictive than (13). Hence, it is possible to satisfy Theorem 3 even when Theorem 2 is violated.

Proof: Assume that

$$\delta_{K+n_f+B} \geq \frac{3\sqrt{B}}{\sqrt{K} + \sqrt{B}}$$

(17)

Since $n_f + B \leq \lceil K/2 \rceil$, we can write $3\lceil K/2 \rceil \geq K + n_f + B$. Following monotonicity of RIC, we obtain

$$\delta_{\lceil K/2 \rceil} \geq \frac{3\sqrt{B}}{\sqrt{K} + \sqrt{B}}.$$ 

Then, by Lemma 4

$$\delta_{K+B} > \frac{\sqrt{B}}{\sqrt{K} + \sqrt{B}}$$

which clearly contradicts (13).

On the other hand, Lemma 5 yields

$$\frac{3\sqrt{B}}{\sqrt{K} + \sqrt{B}} \leq \frac{\sqrt{B}}{\sqrt{K} - n_c + \sqrt{B}}$$

for $n_c$ satisfying (15) when $K \geq (3 + 2\sqrt{B})^2$. That is, there exists some range of $\delta_{K+n_f+B}$ such that

$$\frac{3\sqrt{B}}{\sqrt{K} + \sqrt{B}} \leq \delta_{K+n_f+B} \leq \frac{\sqrt{B}}{\sqrt{K} - n_c + \sqrt{B}}.$$ 

Therefore, when the parameters $K$, $n_f$ and $n_c$ satisfy the necessary conditions, there exists some $\delta_{K+n_f+B}$ which fulfill (16), though (13) fails for $\delta_{K+B}$. Hence, it is possible for the search to find a p-optimal path and satisfy Theorem 3 despite Theorem 2 is violated.

Theorem 4 clarifies that A*OMP allows online recovery guarantees for a range of sparse signals for which A*OMP has no guarantees. This reveals that the residue-based termination is more optimal for recovery of sparse signals from noise-free observations than its sparsity-based counterpart.

IV. EMPIRICAL ANALYSES

A. Experimental Setup

Below, we demonstrate the recovery performance of A*OMP in comparison to A*OMP, BP, SP, OMP, IHT, ISD and SL0 in various scenarios. In the simulations, we employ the AMul and Mul cost models which are denoted as AMul-A*OMP and Mul-A*OMP, respectively. Unless given explicitly, the setup is as follows: We set $I = 3$, $B = 2$ and $P = 200$. For A*OMP, $\varepsilon$ is set to $10^{-6}$ in the noiseless case, while it is selected with respect to the noise level in noisy scenarios. This value of $\varepsilon$ is shared by OMP, which also runs until $\|\hat{y}\|_2 \leq \varepsilon\|y\|_2$. We select $\alpha_{\text{Mul}} = 0.8$ for Mul-A*OMP, $\alpha_{\text{Mul}} = 0.9$ for Mul-A*OMP, and $\alpha_{\text{Mul}} = 0.97$ for AMul-A*OMP. Each test is repeated over a randomly generated set of sparse samples. For each sample, $\Phi$ is drawn from the Gaussian distribution with mean zero and standard deviation $1/N$. The average normalized mean-squared-error (ANMSE) is defined as

$$\text{ANMSE} = \frac{1}{L} \sum_{i=1}^{L} \frac{\|\hat{x}_i - x_i\|_2^2}{\|x_i\|_2^2}$$

(18)

where $\hat{x}_i$ is the recovery of the $i$th test vector $x_i$, and $L$ is the number of test samples. For the noisy case, we specify the distortion ratio as $10\log_{10}(\text{ANMSE})$, in order to better relate the recovery distortion to the signal-to-noise ratio (SNR).

The nonzero entries of the test samples are selected from four random ensembles. The nonzero entries of the Gaussian sparse signals are drawn from the standard Gaussian distribution. The nonzero entries of the uniform sparse signals are distributed uniformly in $[-1, 1]$, while those of the binary sparse signals are set to one. The last ensemble is the Constant Amplitude Random Sign (CARS) sparse signals where the nonzero elements have unit magnitude with random sign.

We perform the A*OMP simulations using the AStarOMP software. AStarOMP implements the A* search by an efficient trie structure, where the nodes of each path are ordered with decreasing priorities proportional to their inner products with the observation vector. This ordering maximizes the number of shared nodes between paths. Consequently, the trie structure allows not only a more compact representation of the search tree, but also faster modifications (i.e. insertion of nodes and equivalent path detection.) As for the orthogonal projection, AStarOMP uses the QR factorization method.

B. Exact Recovery Rates and Reconstruction Error

The first simulation set involves the exact recovery rates and ANMSE for the Gaussian, uniform and binary sparse signals. For this case, we set $N = 256$ and $M = 100$, while $K \in [10, 50]$ and $K_{\text{max}} = 55$. Each test case consists of 500 vectors.

Fig. 1 and 2 depict recovery results for the Gaussian and uniform sparse signals. We observe that A*OMP variants yield better ANMSE than A*OMP at identifying smaller magnitude coefficients, which hardly change the ANMSE, however increase the exact recovery rates. In comparison to the others, A*OMP variants perform significantly better. At the best, A*OMP provides exact recovery until $K = 40$ and $K = 35$ for the Gaussian and uniform ensembles, respectively. These breakpoints are clearly the best among the involved algorithms.

To reveal the benefits of the AMul cost model over the Mul model, we plot the average run time per vector in Fig. 3. The figure is limited to OMP and A*OMP, which are tested using

The code, documentation and binaries of AStarOMP are available at http://myweb.sabanciuniv.edu/karahanoglu/research/
Fig. 1. Recovery results over sparsity for the Gaussian sparse signals.

Fig. 2. Recovery results over sparsity for the uniform sparse signals.

the AStarOMP software. The other algorithms are ignored as they run in MATLAB, which is slower than AStarOMP. We observe that both residue-based termination and AMul cost model significantly accelerate A*OMP due to the relaxation of $\alpha$ to larger values. Since AMul-A*OMP $e$ can afford the largest $\alpha$, it is significantly faster than the other A*OMP variants. These findings confirm the claim in Section II that increasing $\alpha$ reduces the number of explored nodes and hence accelerates A*OMP. Note that decreasing $\alpha_{AMul}$ would further improve the recovery, but also increase the search time.

In Fig. 4 we illustrate the recovery performance for the binary sparse signals, which are known as the most challenging case for greedy algorithms [10], [14]. As expected, $\ell_1$ minimization is the best performer in this case. As above, A*OMP recovery is significantly improved with the residue-based termination. We observe that A*OMP $e$ outperforms the greedy alternatives, despite A*OMP $K$ performs worse than SP.

C. Empirical Investigation of Optimal and p-Optimal Paths

Having presented the recovery results for a range of $K$, we now concentrate on some specific sparsity levels to investigate the optimality/p-optimality conditions of Section III empirically. We evaluate the paths in the search tree after each iteration to determine the number of optimal/p-optimal ones, and plot the results versus the number of iterations after averaging over all test samples. In the simulations, we alter the cost model parameter $\alpha$ in order to analyse its effects on the optimality/p-optimality conditions. Other than that, we use the same settings and the same test data as above. Note that the plots extend to the maximum number of iterations encountered.
throughout each test and not to the average of them. Hence, they are not informative about the average number of search iterations.

The motivation behind these analyses is two-folds: First, we would like to make sure that the $A^\star$OMP algorithm is properly configured, i.e. it does not fail if there still exists some hypotheses which may lead to the true solution in the search tree. Since such hypotheses are represented by optimal/p-optimal paths, $A^\star$OMP should not fail if there is some optimal/p-optimal path in the search tree. Hence, these analyses provide important hints about the cost model and pruning techniques in addition to revealing the effects of cost model parameters on the recovery.

Second, the empirical number of optimal/p-optimal paths reveal whether the optimality/p-optimality conditions of Section III hold in practice. Imagine that $A^\star$OMP fails even if there are some optimal/p-optimal paths in the tree. Such an example would indicate that the optimality/p-optimality conditions are invalid, since these conditions necessitate the algorithm either to iterate until there remains no optimal/p-optimal paths, or to terminate at the true solution. Therefore, analyzing the number of optimal/p-optimal paths is extremely helpful for questioning the validity of the presented theoretical analyses. Moreover, by investigating the number of optimal/p-optimal paths per successful recovery, we may also justify our claim about the number of such paths being large, which increases the expectation that the optimality/p-optimality conditions hold.

Among simulations for a series of $K$ values, we present the results of two cases here: The number of average optimal/p-optimal paths per successful recovery is illustrated in Fig. 5 for $K = 30$, while those per failure are depicted in Fig. 6 for $K = 50$. The reason for these choices is the high number of successful recoveries and failures for $K = 30$ and $K = 50$, respectively. Note that, though they are not illustrated here, we have observed similar results for other choices of $K$ as well.

Fig. 5 reveals that there exists a high number of optimal/p-optimal paths per successful recovery. In addition, this number increases even further if the residue-based termination is employed instead of the sparsity based one.

Considering the average number of optimal/p-optimal paths per failure, we observe that both $\alpha$ and the termination criteria play important roles on satisfying the optimality/p-optimality conditions. When recovering $x$ via Mul-$A^\star$OMP$_K$ using $\alpha = 0.8$ or $\alpha = 0.85$, the failures occur only if there

---

Fig. 3. Average run-time with the AStarOMP software.

Fig. 4. Recovery results over sparsity for the binary sparse signals.
remains no optimal paths in the tree. On the other hand, if $\alpha \geq 0.9$, the cost model of Mul-A*OMP$_K$ becomes too aggressive due to the reduced contribution of the auxiliary function and misses some optimal paths. It is obvious that choosing $\alpha \geq 0.9$ does not preserve the optimality conditions with Mul-A*OMP$_K$ when $K = 50$. In addition, since the cost model depends on a specific $K$ for this type of termination criterion, smaller $\alpha$ values may also turn out to be suboptimal when $K$ is smaller. Contrarily, when the termination criterion is replaced with the residue-based one, i.e. Mul-A*OMP$_e$, we observe that choosing $\alpha = 0.9$ still preserves the p-optimality requirements. In addition, since the cost model depends on a fixed $K_{\text{max}}$, this holds for all $K$ values as well. However, choosing $\alpha = 0.95$ does still violate the p-optimality requirements. Finally, adopting the AMul cost model via AMul-A*OMP$_e$, we observe that $\alpha$ can be further relaxed to $\alpha = 0.95$ without violating the p-optimality requirements.

These results indeed lead to important observations about the cost models and termination criteria of A*OMP. First, we observe that the optimality can be satisfied in practice with properly chosen cost models and termination criteria. This observation reveals that the optimality/p-optimality conditions which are necessary for theoretical guarantees of A*OMP are not artificial. This finding clearly supports the validity of
the analytical results derived in Section III. In addition, we observe that larger $\alpha$ values may be used without violating the $p$-optimality conditions when the residue-based termination criterion and AMul cost model are employed. Consequently, AMul-A$^*$OMP$_e$ allows for much greed by relaxing $\alpha$, and can still provide high accuracy in less iterations than the other versions of the algorithm. This explains the improvements in the recovery performance observed with AMul-A$^*$OMP$_e$ in the last section.

\subsection*{D. Phase Transitions}

Empirical phase transitions provide important means for recovery analysis, since they reveal the recovery performance over the feasible range of $M$ and $K$. Consider the normalized measures $\lambda = M/N$ and $\rho = K/M$. The phase transition curve is mostly a function of $\lambda$ \cite{14}, hence it allows for a general characterization of the recovery performance.

To obtain the phase transitions, we fix $N = 250$, and alter $M$ and $K$ to sample the $\{\lambda, \rho\}$ space for $\lambda \in [0.1, 0.9]$ and $\rho \in [0, 1]$. For each $\{\lambda, \rho\}$ tuple, we randomly generate 200 sparse instances and run AMul-A$^*$OMP$_e$, OMP, BP, SP, ISD and SL0 algorithms for the recovery. Setting the exact recovery criterion as $\|x_i - \hat{x}_i\|_2 / \|x_i\|_2 \leq 10^{-2}$, we count the number of exactly recovered samples in each test. The phase transitions are then obtained using the methodology described in \cite{14}. For each $\lambda$, we employ a generalized linear model with logistic link to describe the exact recovery curve over $\rho$, and then find the $\rho$ value which yields 50% exact recovery probability.
Combination of the results over the whole $\lambda$ range leads to the empirical phase transition curve.

First, we discuss the choice of $K_{\max}$. Let us define $\rho_{\max} = K_{\max}/M$. This normalized measure helps us to identify the optimal $\rho_{\max}$ values over $\lambda$. Then, we can select $K_{\max} = \rho_{\max}M$ using the optimal $\rho_{\max}$ value for a particular $\lambda$. To find an optimal formulation for $\rho_{\max}$ as a function of $\lambda$, we have run a number of simulations, where we have observed that the recovery performance of AMul-$A^*\text{OMP}_\epsilon$ is quite robust to the choice of $K_{\max}$, with a perturbation up to $\%3$ in the phase transition curve. Hence, the recovery accuracy is mostly independent of $K_{\max}$. Yet, based on our experience, we propose to choose $\rho_{\max} = 0.5 + 0.5\lambda$ taking into account both the accuracy and complexity of the search. The phase transition curves below are obtained with this setting.

The empirical phase transition curves are depicted in Fig. 7. These indicate that AMul-$A^*\text{OMP}_\epsilon$ yields better phase transitions than the other algorithms for the Gaussian and uniform sparse signals. Contrarily, for the CARS case, BP and ISD perform better than the other algorithms involved, while AMul-$A^*\text{OMP}_\epsilon$ is the third best. Regarding the effect of the coefficient distribution on the performance, we observe that BP is robust, while the phase transition curves for AMul-$A^*\text{OMP}_\epsilon$ and OMP exhibit the highest variation among distributions. When the nonzero values cover a wide range, such as the Gaussian distribution, the performances of $A^*\text{OMP}$ and OMP are boosted. In contrast, nonzero values of equal magnitude unexpectedly turn out to be the most challenging case for these two. These observations indicate that OMP-type algorithms are more effective when the nonzero elements span a wide range of magnitudes. Moreover, when this range gets wide enough, even OMP can possess better phase transition than BP.

E. Recovery from Noisy Observations

In order to evaluate the empirical recovery performance from corrupted measurements, the observation vectors are contaminated with white Gaussian noise at different Signal-to-Noise Ratio (SNR) levels. Fig. 8 illustrates the recovery performance over SNR where $K = 30$ and $K = 25$ for the Gaussian and uniform sparse signals, respectively. We observe that $A^*\text{OMP}$ is superior to other algorithms except for 5dB SNR where BP is slightly better. In addition, $A^*\text{OMP}_\epsilon$ improves the recovery accuracy slightly over $A^*\text{OMP}_K$ for low SNR. Fig. 8 depicts the average $A^*\text{OMP}$ run times in this scenario. Similar to the previous examples, AMul-$A^*\text{OMP}_\epsilon$ is significantly faster than the other $A^*\text{OMP}$ variants.

F. A Hybrid Approach for Faster Practical Recovery

Based on the results above, it is possible to speed up the recovery from noise-free observations using a hybrid of OMP and AMul-$A^*\text{OMP}_\epsilon$. First, OMP provides exact recovery up to some mid-sparsity range. Moreover, there are regions where AMul-$A^*\text{OMP}_\epsilon$ provides exact recovery while OMP also yields quite high recovery rates. In these regions, we can accelerate the recovery without sacrificing the accuracy by a simple two stage hybrid scheme: We run OMP first, and then AMul-$A^*\text{OMP}_\epsilon$ only if OMP fails. This reduces the number of AMul-$A^*\text{OMP}_\epsilon$ runs and accelerates the algorithm if OMP failures can be properly identified. This is indeed not difficult: Assuming that $K + K_{\max}$-RIP holds, OMP fails when the residue does not vanish, after which the hybrid approach runs AMul-$A^*\text{OMP}_\epsilon$. Moreover, we use the order by which OMP chooses the vectors in consequent iterations for setting the priorities of trie nodes. That is, a vector OMP chooses first gets higher priority, and is placed at lower levels of the trie. This reduces not only the trie size but also the cost of equivalent path detection and insertions.

The recovery results for the hybrid approach are depicted in Fig. 9. We observe that AMul-$A^*\text{OMP}_\epsilon$ and the hybrid approach yield identical exact recovery rates, while the latter is significantly faster. This acceleration is proportional to the exact recovery rate of OMP. That is, the hybrid approach is faster where OMP is better. These results show that this approach is indeed able to detect the OMP failures, and run AMul-$A^*\text{OMP}_\epsilon$ only for those instances.

G. Demonstration on a Sparse Image

To illustrate AMul-$A^*\text{OMP}$, on a more realistic coefficient distribution, we demonstrate recovery of the $512 \times 512$ image "bridge", shown in Fig. 11 on upper left. The recovery is performed in blocks of size $8 \times 8$. The aim of block processing is to break the recovery problem into a number of smaller and simpler subproblems. The image is first preprocessed such that each $8 \times 8$ block is $K$-sparse in the 2D Haar Wavelet basis, $\Psi$, where $K = 12$, i.e. for each block only the 12 largest magnitude wavelet coefficients are kept. Note that in this case, the reconstruction dictionary is not the observation matrix $\Phi$ itself, but $\Phi\Psi$. From each block, $M = 32$ observations are taken, where the entries of $\Phi$ are randomly drawn from the Gaussian distribution with mean 0 and standard deviation $1/N$. The parameters are selected as $I = 3$, $P = 200$ and $K_{\max} = 20$. $\alpha_{\text{AMul}}$ is reduced to 0.85 in order to compensate the decrement in $K_{\max}$, which decreases the auxiliary term in (2). Fig. 11 shows the images recovered by BP and AMul-$A^*\text{OMP}_\epsilon$. We observe that BP provides a Peak Signal-to-Noise Ratio (PSNR) value of 29.9 dB, while AMul-$A^*\text{OMP}_\epsilon$ improves the recovery PSNR to 42.1 dB for $B = 2$ and to 49.3 dB for $B = 3$. A careful investigation of the recovered images yields that AMul-$A^*\text{OMP}_\epsilon$ improves the recovery especially at detailed regions and boundaries.

V. Summary

The fundamental goal of this manuscript is the theoretical analysis of the $A^*\text{OMP}$ algorithm. For this purpose, we have first derived a RIP condition for the success of an $A^*\text{OMP}$ iteration. Then, we have generalized this result to the exact recovery of all $K$-sparse signals from noise-free measurements with $A^*\text{OMP}_K$, where the termination is based on the sparsity level $K$. Next, we have extended this result to $A^*\text{OMP}_\epsilon$, which employs a residue-based termination criterion. In particular, we have stated that a $K$-sparse signal can be recovered with $A^*\text{OMP}_\epsilon$ when some online condition is satisfied. Interestingly, exact recovery guarantees of $A^*\text{OMP}_K$ represent a special case of this condition. This has led to the conclusion...
that $A^\star OMP_e$ enjoys at least the same general exact recovery guarantees with $A^\star OMP_K$ in addition to the online guarantees. Further comparison of the two has also revealed that the recovery condition of $A^\star OMP_e$ represents a less restrictive requirement than that of $A^\star OMP_K$. This result encourages utilising the residue-based termination criterion instead of the sparsity-based one for recovery from noise-free observations. In addition, we have also discussed the novel AMul cost model, which extends the Mul model in a dynamic manner. This model allows for a larger auxiliary model parameter, which accelerates the search without sacrificing the accuracy.

Lastly, we have demonstrated the empirical recovery performance of AMul-$A^\star OMP_e$ by extensive simulations, including sparse signals with different characteristics. The results
of these experiments support that AMul-A*OMPₖ possesses better recovery capabilities and shorter execution times than A*OMPₖ. A*OMP variants perform better recovery than BP, SP, IHT, OMP, SL0 and ISD for uniform and Gaussian sparse signals. With constant magnitude nonzero elements, such as the binary and CARS sparse signals, AMul-A*OMPₑ still provides better recovery accuracy than the greedy alternatives involved, however BP yields the most accurate recovery in this case. We have also demonstrated that the search can be significantly accelerated without sacrificing the accuracy via a hybrid approach, which first applies OMP, and then AMul-A*OMPₑ only if OMP fails. Finally, we have employed AMul-A*OMPₑ on a sparse image, where it improves the recovery significantly over BP.

The presented guarantees for A*OMPₖ and A*OMPₑ depend on the so-called optimality/p-optimality conditions, which can be hardly guaranteed for any practical cost model because of tractability issues. In Section IV.C we have discussed that these conditions may be justified by intuitive reasoning such as the high number of optimal/p-optimal paths, and small value of the cost function for such paths. More important in this context, we have given space to supporting empirical evidence in Section IV.C via analyses of optimal/p-optimal paths for successful recoveries and failures. According to these results, the search can properly identify optimal/p-optimal paths with an appropriately chosen cost model and termination criterion. Especially, the AMul cost model with residue-based termination has demonstrated strong empirical performance while also providing more greed via allowing for a larger α. Hence, AMul-A*OMPₑ turns out to be the most promising A*OMP variant, not only with high recovery accuracy in shorter run times but also with strong empirical evidence on the validity of the p-optimality conditions.

REFERENCES

[1] N.B. Karahanoglu and H. Erdogan, “A comparison of termination criteria for A*OMP,” in European Signal Processing Conference, Bucharest, Romania, Aug. 2012.

[2] S. Chen, D. Donoho, and M. Saunders, “Atomic decomposition by basis pursuit,” SIAM J. on Sci. Comp., vol. 20, no. 1, pp. 33–61, 1998.

[3] E. Candés and T. Tao, “Decoding by linear programming,” IEEE Trans. Inf. Theory, vol. 51, no. 12, pp. 4203–4215, Dec. 2005.

[4] D. Donoho, “Compressed sensing,” IEEE Trans. Inf. Theory, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

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

[6] Y. C. Pati, R. Rezaifar, and P. S. Krishnaprasad, “Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition,” in Proc. 27th Asilomar Conference on Signals, Systems and Computers, Los Alamitos, CA, 1993, vol. 1, pp. 40–44.

[7] 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, 2007.

[8] J. A. Tropp, “Greed is good: Algorithmic results for sparse approximation,” IEEE Trans. Inf. Theory, vol. 50, no. 10, pp. 2231–2242, Oct. 2004.

[9] D. Needell and J. A. Tropp, “CoSaMP: Iterative signal recovery from incomplete and inaccurate samples,” Appl. Comp. Harmonic Anal., vol. 26, pp. 301–321, 2008.

[10] W. Dai and O. Milenkovic, “Subspace pursuit for compressive sensing signal reconstruction,” IEEE Trans. Inf. Theory, vol. 55, no. 5, pp. 2233–2249, May 2009.

[11] T. Blumensath and M.E. Davies, “Iterative hard thresholding for compressed sensing,” Appl. Comp. Harmonic Anal., vol. 27, no. 3, pp. 265–274, 2009.

[12] D. Needell and R. Vershynin, “Signal recovery from incomplete and inaccurate measurements via regularized orthogonal matching pursuit,” Selected Topics in Signal Processing, IEEE Journal of, vol. 4, pp. 310–316, Apr. 2010.

[13] T. Blumensath and M.E. Davies, “Gradient pursuits,” IEEE Trans. Signal Process., vol. 56, no. 6, pp. 2370–2382, June 2008.

[14] A. Maleki and D.L. Donoho, “Optimally tuned iterative reconstruction algorithms for compressed sensing,” Selected Topics in Signal Processing, IEEE Journal of, vol. 4, no. 2, pp. 330 –341, Apr. 2010.

[15] N. B. Karahanoglu and H. Erdogan, “Compressed sensing signal recovery via forward-backward pursuit,” 2012, submitted to Digital Signal Processing, available as http://arxiv.org/abs/1210.5626.

[16] D.P. Wipf, B.D. Rao, and S. Nagarajana, “Latent variable bayesian models for promoting sparsity,” IEEE Trans. Inf. Theory, to appear.

[17] Shihao Ji, Ya Xue, and L. Carin, “Bayesian compressive sensing,” IEEE Trans. Signal Process., vol. 56, no. 6, pp. 2346–2356, 2008.

[18] S. D. Babacan, R. Molina, and A. K. Katsaggelos, “Bayesian compressive sensing using laplace priors,” IEEE Trans. Image Process., vol. 19, no. 1, pp. 53–63, 2010.

[19] C. Jutten H. Mohimani, M. Babaie-Zadeh, “A fast approach for overcomplete sparse decomposition based on smoothed l0 norm,” IEEE Trans. Signal Process., vol. 57, no. 1, pp. 289–301, Jan. 2009.

[20] R. Chartrand, “Exact reconstruction of sparse signals via nonconvex minimization,” IEEE Signal Process. Lett., vol. 14, no. 10, pp. 707–710, 2007.

[21] R. Chartrand and Wotao Yin, “Iteratively reweighted algorithms for
Fig. 11. Recovery of the image “bridge” using BP and AMul-A*OMPₐ.

compressive sensing," in Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing ICASSP 2008, 2008, pp. 3869–3872.
[22] Y. Wang and W. Yin, “Sparse signal reconstruction via iterative support detection,” SIAM Journal on Imaging Sciences, vol. 3, no. 3, pp. 462–491, 2010.
[23] Emmanuel Candes, Michael Wakin, and Stephen Boyd, “Enhancing sparsity by reweighted ℓ_1 minimization,” Journal of Fourier Analysis and Applications, vol. 14, pp. 877–905, 2008.
[24] Ingrid Daubechies, Ronald DeVore, Massimo Fornasier, and C. Sinan Gunturk, “Iteratively reweighted least squares minimization for sparse recovery,” Comm. Pure Appl. Math., vol. 63, no. 1, pp. 1–38, Jan. 2010.
[25] M. Rudelson and R. Vershynin, “Sparse reconstruction by convex relaxation: Fourier and gaussian measurements,” in 40th Annual Conference on Information Sciences and Systems, 2006, pp. 207–212.
[26] E. Candès and T. Tao, “Near-optimal signal recovery from random projections: universal encoding strategies?,” IEEE Trans. Inf. Theory, vol. 52, no. 12, pp. 5406–5425, 2006.
[27] E. Candès, “The restricted isometry property and its implications for compressed sensing,” Comptes Rendus Mathematique, vol. 346, no. 9–10, pp. 589–592, May 2008.
[28] 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, sept. 2010.
[29] 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.
[30] Q. Mo and Y. Shen, “A remark on the restricted isometry property in orthogonal matching pursuit,” *IEEE Trans. Inf. Theory*, vol. 58, no. 6, pp. 3654–3656, 2012.

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

[32] S. Foucart, “Stability and robustness of weak orthogonal matching pursuits,” *Recent Advances in Harmonic Analysis and Applications, Springer Proceedings in Mathematics & Statistics*, vol. 25, pp. 395–405, 2013.

[33] J. Wang and B. Shim, “Improved recovery bounds of orthogonal matching pursuit using restricted isometry property,” 2012, submitted to *IEEE Trans. Inf. Theory*, available as http://arxiv.org/abs/1211.4293.

[34] E. Candés, J. Romberg, and T. Tao, “Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information,” *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.

[35] N. B. Karahanoglu and H. Erdogan, “A* orthogonal matching pursuit: Best-first search for compressed sensing signal recovery,” *Digital Signal Processing*, vol. 22, no. 4, pp. 555–568, July 2012.

[36] P. E. Hart, N. J. Nilsson, and B. Raphael, “A formal basis for the heuristic determination of minimum cost paths,” *IEEE Trans. Syst. Sci. Cybern.*, vol. 4, no. 12, pp. 100–107, July 1968.

[37] F. Jelinek, *Statistical Methods For Speech Recognition*, chapter 6, MIT Press, Cambridge, MA, USA, 1997.

[38] N. B. Karahanoglu and H. Erdogan, “Online recovery guarantees for orthogonal matching pursuit,” 2013, submitted to *IEEE Signal Process. Lett*, extended preprint available as http://arxiv.org/abs/1210.5991.

**Nazim Burak Karahanoglu** is a senior researcher at the Advanced Technologies Research Institute of the Scientific and Technological Research Council of Turkey (TUBITAK) in Kocaeli, Turkey. He received his B.S. degree in Electrical and Electronics Engineering from METU, Ankara in 2003, M.S. degree in Computational Engineering from the Friedrich-Alexander University of Erlangen-Nuremberg, Germany in 2006 and Ph.D. degree in Electronics Engineering from Sabanci University, Istanbul in 2013. He has been with TUBITAK since 2008. His research interests include compressed sensing, sparse signal recovery, and sonar signal processing.

**Hakan Erdogan** is an assistant professor at Sabanci University in Istanbul, Turkey. He received his B.S. degree in Electrical Engineering and Mathematics in 1993 from METU, Ankara and his M.S. and Ph.D. degrees in Electrical Engineering: Systems from the University of Michigan, Ann Arbor in 1995 and 1999 respectively. He was with the Human Language Technologies group at IBM T.J. Watson Research Center, NY between 1999 and 2002. He has been with Sabanci University since 2002. His research interests are in developing and applying probabilistic methods and algorithms for multimedia information extraction.