

Improving A* OMP: Theoretical and empirical analyses with a novel dynamic cost model

Nazim Burak Karahanoglu^{a,b,*}, Hakan Erdogan^b

^a TUBITAK BILGEM, Kocaeli 41470, Turkey

^b Faculty of Engineering and Natural Sciences, Sabanci University, Istanbul 34956, Turkey

ARTICLE INFO

Article history:

Received 29 October 2014

Received in revised form

5 June 2015

Accepted 12 June 2015

Available online 23 June 2015

Keywords:

Compressed sensing

A* orthogonal matching pursuit

Restricted isometry property

Adaptive-multiplicative cost model

ABSTRACT

Best-first search has been recently utilized for compressed sensing (CS) by the A* orthogonal matching pursuit (A*OMP) algorithm. In this work, we concentrate on theoretical and empirical analyses of A*OMP. We present a restricted isometry property (RIP) based general condition for exact recovery of sparse signals via A*OMP. In addition, we develop online guarantees which promise improved recovery performance with the residue-based termination instead of the sparsity-based one. We demonstrate the recovery capabilities of A*OMP with extensive recovery simulations 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 for different nonzero element distributions as well as recovery rates and average error, reveal not only the superior recovery accuracy of A*OMP, but also the improvements with the residue-based termination and the AMul cost model. Comparison of the run times indicates the speed up by the AMul cost model. We also demonstrate a hybrid of OMP and A*OMP to accelerate the search further. Finally, we run A*OMP on sparse images to illustrate its recovery performance for more realistic coefficient distributions.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

A* orthogonal matching pursuit (A*OMP) [1] aims at combination of best-first tree search with the orthogonal matching pursuit (OMP) algorithm [2] for the compressed sensing problem. It incorporates the A* search technique [3,4] to increase the efficiency of the search over the tree representing the hypotheses space consisting of sparse candidates. Dedicated cost models have been proposed in order to make the search intelligently guided and tractable. The empirical investigation in [1] implies significant recovery improvements over conventional compressed sensing methods. The recently introduced adaptive-multiplicative

cost model has led to further speed and accuracy improvements in the preliminary empirical findings of [5].

This paper addresses a detailed theoretical and experimental study of A*OMP algorithm particularly when the number of elements in the sparse solution is not restricted to the sparsity level K of the underlying signal. This is obtained by enforcing a residue-based termination criterion. We analyze the theoretical performance of the algorithm using the restricted isometry property. The analyses cover two possible cases for the length of the returned solution, namely when it is restricted to K nonzero elements, and when more than K elements are allowed. In addition, the impacts of the adaptive-multiplicative cost model and the residue-based termination criterion on the recovery speed and accuracy are evaluated via comprehensive simulations involving different signal statistics, phase transitions and images in comparison to conventional recovery algorithms.

* Corresponding author at: TUBITAK BILGEM, Kocaeli 41470, Turkey.

E-mail addresses: burak.karahanoglu@tubitak.gov.tr

(N.B. Karahanoglu), haerdogan@sabanciuniv.edu (H. Erdogan).

1.1. Compressed sensing

The fundamental goal of compressed sensing (CS) is to unify data acquisition and compression by observing a lower dimensional vector $\mathbf{y} = \Phi \mathbf{x}$ instead of the signal \mathbf{x} , where $\Phi \in \mathbb{R}^{M \times N}$ is the (generally random) measurement matrix, $\mathbf{y} \in \mathbb{R}^M$, and $\mathbf{x} \in \mathbb{R}^N$. The dimensionality reduction follows $M < N$, as a result of which \mathbf{x} cannot be directly solved back from \mathbf{y} . Alternatively, assuming \mathbf{x} is K -sparse (i.e., it has at most K nonzero components), or compressible, \mathbf{x} can be recovered under certain conditions by solving

$$\min \|\mathbf{x}\|_0 \quad \text{subject to } \mathbf{y} = \Phi \mathbf{x} \quad (1)$$

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

As the direct solution of (1) is intractable, approximate solutions have emerged in the CS literature. Convex optimization algorithms [6–9] relax (1) by replacing $\|\mathbf{x}\|_0$ with its closest convex approximation $\|\mathbf{x}\|_1$. Greedy algorithms [2,10–14] provide simple and approximate solutions via iterative residue minimization. Other recovery schemes include Bayesian methods [15,16], nonconvex approaches [17–20], iterative reweighted methods [21–23], etc.

1.2. A* orthogonal matching pursuit

OMP is among the most acknowledged greedy algorithms for sparse recovery [24–26]. It aims at iterative detection of the support, i.e., the set of indices corresponding to nonzero coefficients, of \mathbf{x} . At each iteration, OMP identifies the best match to the residue of \mathbf{y} among the atoms, i.e., columns of Φ , by choosing the index of the atom with maximum correlation. That is, OMP is structurally based on continuous expansion of a single hypothesis, represented by a single iteratively expanded support estimate, or path.

On the other hand, simultaneous evaluation of multiple hypotheses for the sparse support may improve recovery over the single-path algorithms like OMP. Multiple hypotheses may be represented by a search tree, and the recovery problem can be efficiently solved by sophisticated best-first search techniques. To promote this idea, the authors have introduced the A*OMP algorithm [1,5] which is an iterative semi-greedy approach that utilizes A* search on a multiple hypotheses search tree in order to find an approximation of (1). The nodes of the tree contain indices of the selected atoms, and the paths represent the candidate support sets for \mathbf{x} . Fig. 1 illustrates the evaluation of such a tree in comparison to OMP. At each iteration, A*OMP first selects the best path with the minimum cost criterion which depends on the ℓ_2 norm of the path residue. Then, the best path is expanded by exploring B of its child nodes with maximum correlation to the path residue. That is, B new candidate support sets are appended to the tree. The filled nodes in Fig. 1 indicate the child nodes explored per step, which are referred to as $\Delta\mathcal{T}$ in the rest.

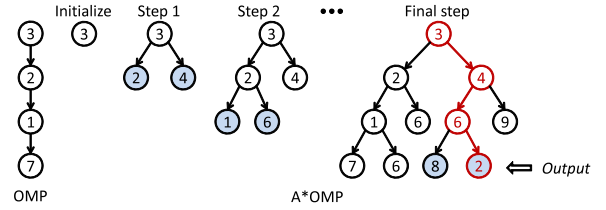


Fig. 1. OMP vs. evaluation of A*OMP search tree.

In Fig. 1, although the OMP solution is among the hypotheses in the tree, A*OMP returns a different support set. In fact, if OMP were successful, A*OMP would also return the same solution. This typical example of OMP failure, where the best-first search identifies the true solution by simultaneous evaluation of multiple hypotheses, illustrates how the multiple path strategy improves the recovery.

Despite this simple illustration, combining the A* search with OMP is not straightforward. It necessitates properly defined cost models which enable the A* search to perform the stage-wise residue minimization in an intelligent manner, and effective pruning techniques which make the algorithm tractable. Various structures are introduced in [1,5] for the cost model, which is vital for the comparison of paths with different lengths. Pruning strategies, which enable a complexity-accuracy trade-off together with the cost model, are detailed in [1]. Below, we provide a summary of A*OMP, and refer the interested reader to [1,5] for the details.

1.2.1. Notation

Before the summary, we clarify the notation in this paper. We define \mathcal{S} as the set of all paths in the search tree. \mathcal{T} is the true support of \mathbf{x} . \mathcal{T}^i , \mathbf{r}^i , l^i and $f(\mathcal{T}^i)$ denote the support estimate, residue, length and cost of the i th path, respectively. $\hat{\mathbf{x}}^i$ is the estimate of \mathbf{x} given by the i th path. The best path at a certain step is referred to as b . As mentioned above, $\Delta\mathcal{T}$ represents the set of indices selected during the expansion of b , i.e., the indices of the B largest magnitude elements in $\Phi^* \mathbf{r}^b$, where Φ^* denotes the conjugate of Φ , ϕ_j is the j th column of Φ , $\Phi_{\mathcal{J}}$ denotes the matrix composed of the columns of Φ indexed by the set \mathcal{J} . Similarly, $\mathbf{x}_{\mathcal{J}}$ is the vector of the elements of \mathbf{x} indexed by \mathcal{J} . K_{\max} is the maximum number of allowable nodes along a path in the A*OMP search tree. We say that path i complete if $l^i = K_{\max}$.

1.2.2. Brief overview of A*OMP

A*OMP initializes the search tree with I paths of a single node each. These nodes represent the indices of the I largest magnitude elements in $\Phi^* \mathbf{y}$. At each iteration, the algorithm first selects the best path b among the incomplete paths in the search tree with minimum cost. Then, $\Delta\mathcal{T}$ is chosen as the indices of the B largest magnitude elements in $\Phi^* \mathbf{r}^b$. This implies B candidate paths, each of which expands b with a single index in $\Delta\mathcal{T}$. Each candidate path is opened unless an equivalent path has been explored before (equivalent path pruning [1]). For each new path i , \mathbf{r}^i is given by the projection error of \mathbf{y} onto $\Phi_{\mathcal{T}^i}$, and the cost $f(\mathcal{T}^i)$ is computed as a function of \mathbf{r}^i . Finally, all but the best P paths with minimum cost are pruned (tree size

¹ A more general model involves a structured dictionary Ψ for sparse representation of \mathbf{x} , i.e., $\mathbf{x} = \Psi \mathbf{z}$ where \mathbf{z} is sparse and \mathbf{x} is not. In this more general case, observation model can be written as $\mathbf{y} = \Phi \Psi \mathbf{z}$. For simplicity, we omit Ψ and treat \mathbf{x} as sparse.

Table 1
Comparison of A*OMP mechanisms.

| | Initial version [1,5] | Modified version |
|----------------------|---|-------------------------------|
| Best path selection | Among <i>all</i> paths | Among <i>incomplete</i> paths |
| Termination criteria | (The residue is small enough) or (The best path is complete) | The residue is small enough |

pruning [1]). Selection and expansion of the best path are repeated until either some path i satisfies $\|\mathbf{r}^i\|_2 \leq \varepsilon \|\mathbf{y}\|_2$, or all P paths are complete. The pseudo-code for A*OMP is given in Algorithm 1.

Algorithm 1. A* orthogonal matching pursuit.

```

1: Input:  $\Phi, \mathbf{y}$ 
2: Define:  $P, I, B, K_{\max}, \varepsilon, \{\alpha_{\text{Mul}} \text{ or } \alpha_{\text{AMul}}\}$ 
3: Initialize:  $\mathcal{T}^1 = \emptyset, \mathbf{r}^i = \mathbf{y} \ \forall i = 1, 2, \dots, P, b = \{1\}$ 
4:  $\Delta\mathcal{T} = \arg \max_{\mathcal{J}, |\mathcal{J}|=1} \sum_{j \in \mathcal{J}} |\langle \phi_j, \mathbf{y} \rangle|$ 
5: for  $i=1$  to  $I$  do  $\triangleright I$  paths of length 1
6:    $\mathcal{T}^i = \{i\text{th index in } \Delta\mathcal{T}\}, \mathbf{r}^i = \mathbf{y} - \langle \mathbf{y}, \phi_{\mathcal{T}^i} \rangle \phi_{\mathcal{T}^i}$ 
7: end for
8: while  $b \neq \emptyset$  do
9:    $\Delta\mathcal{T} = \arg \max_{\mathcal{J}, |\mathcal{J}|=B} \sum_{j \in \mathcal{J}} |\langle \phi_j, \mathbf{r}^b \rangle|$   $\triangleright B$  children of  $b$ 
10:   $\hat{\mathcal{T}} = \mathcal{T}^b$ 
11:   $p = b$   $\triangleright$  first to be replaced
12:  for  $i=1$  to  $B$  do  $\triangleright$  expansion
13:     $\hat{\mathcal{T}} = \hat{\mathcal{T}} \cup \{i\text{th index in } \Delta\mathcal{T}\}$   $\triangleright$  candidate path
14:     $\mathbf{z} = \arg \min_{\mathbf{z}} \|\mathbf{y} - \Phi_{\hat{\mathcal{T}}} \mathbf{z}\|_2$   $\triangleright$  orthogonal projection
15:     $\hat{\mathbf{r}} = \mathbf{y} - \Phi_{\hat{\mathcal{T}}} \mathbf{z}$   $\triangleright$  update residue
16:    if  $(\|\hat{\mathbf{r}}\|_2 \leq \varepsilon \|\mathbf{y}\|_2)$   $\triangleright$  check residue
17:      return  $\hat{\mathcal{T}}$   $\triangleright$  terminate
18:    end if
19:    if  $f(\hat{\mathcal{T}}) < f(\mathcal{T}^p)$  and  $(\hat{\mathcal{T}} \notin \mathcal{S})$   $\triangleright$  pruning
20:       $\mathcal{T}^p = \hat{\mathcal{T}}, \mathbf{r}^p = \hat{\mathbf{r}}$ 
21:    end if
22:     $p = \arg \max_{i \in 1, 2, \dots, P} f(\mathcal{T}^i)$   $\triangleright$  worst path (replaced next)
23:  end for
24:   $b = \arg \min_{i \in 1, 2, \dots, P, f < K_{\max}} f(\mathcal{T}^i)$   $\triangleright$  best incomplete path
25: end while
26:  $b = \arg \min_{i \in 1, 2, \dots, P} f(\mathcal{T}^i)$   $\triangleright$  best (complete) path
27: return  $\mathcal{T}^b$ 

```

1.2.3. Termination criteria

The outline presented above is slightly different from the introduction of A*OMP in [1,5]. The best path selection and termination mechanisms of A*OMP are modified in order to improve the theoretical guarantees of the algorithm. In [1,5], the best path is selected among *all* paths in the tree. Accordingly, the search terminates either when

- (i) the best path is complete ($i^b = K_{\max}$), or
- (ii) the residue is small enough ($\|\mathbf{r}\|_2 < \varepsilon \|\mathbf{y}\|_2$).

Actually, (i) appears here as a consequence of the best path choice involving complete paths in addition to the incomplete ones. Since the best path is chosen using the cost function, this termination criterion depends on the cost model. To guarantee exact recovery with this scheme, the

cost model should exhibit some sense of optimality, i.e., it should assign potentially true paths lower costs than false complete paths. This is necessary to ensure that some false complete path does not become the best path. However, optimality of the cost model is analytically hard to guarantee. This becomes an obstacle for obtaining theoretical guarantees independent of the cost models.

To overcome this problem, the dependency of the termination on the cost function, i.e., termination criterion (i), should be removed. For this purpose, we modify A*OMP as follows:

- (i) The best path is selected among the *incomplete* paths in the tree (Algorithm 1, line 24).
- (ii) The search terminates when the residue is small enough (Algorithm 1, lines 16 and 17).

To ensure termination, the best complete path is returned as the solution only when all paths are complete, but none of them satisfies the termination criterion on the residue² (Algorithm 1, lines 26 and 27). These modifications are listed in Table 1. In this structure, termination does not directly rely on the cost model, hence stronger exact recovery guarantees may be obtained. The results in this paper are obtained with the modified version of the algorithm. Note that, based on the authors' experience, this modification does not have a significant effect on the empirical performance of the algorithm. Yet, it is critical for the theoretical analysis. The rest of this paper concentrates on the modified version of the algorithm without explicit referral.

The termination parameters in Algorithm 1, K_{\max} and ε , can be adjusted for different termination behavior. In [1], each path is limited to K nodes, i.e., $K_{\max} = K$. We call this sparsity-based termination, and denote by A*OMP_K. Another alternative is the residue-based termination such as in [5], where more than K nodes are allowed along a path by setting $K_{\max} > K$ and ε is selected small enough based on the noise level. This version is referred to as A*OMP _{ε} . Preliminary results in [5] indicate that A*OMP _{ε} yields not only better recovery but also faster termination than A*OMP_K. With the flexibility on choosing K_{\max} and ε , we apply both termination criteria in Algorithm 1.

1.2.4. Cost models

To select the best path, A*OMP should compare the costs of paths with different lengths. This necessitates proper cost models which can compensate for the differences in path

² Note that this already indicates a recovery failure.

lengths. Some novel models have been proposed in [1,5]. In this work, we employ the multiplicative (Mul) [1] and adaptive-multiplicative (AMul) [5] models, following their superior recovery capabilities demonstrated in [1,5].

The Mul cost model relies on the expectation that unexplored nodes decrease $\|\mathbf{r}^i\|_2$ by a constant rate $\alpha_{\text{Mul}} \in (0, 1)$:

$$f_{\text{Mul}}(\mathcal{T}^i) = \alpha_{\text{Mul}}^{K_{\text{max}} - l^i} \|\mathbf{r}^i\|_2.$$

Note that we replace K in [1] with K_{max} to allow for different termination criteria. [1] demonstrates that decreasing α_{Mul} improves recovery accuracy, while the search gets slower.

The AMul model is a dynamic extension of the Mul model:

$$f_{\text{AMul}}(\mathcal{T}^i) = \left(\alpha_{\text{AMul}} \frac{\|\mathbf{r}_l^i\|_2}{\|\mathbf{r}_{l-1}^i\|_2} \right)^{K_{\text{max}} - l^i} \|\mathbf{r}_l^i\|_2 \quad (2)$$

where \mathbf{r}_l^i denotes the residue after the first l nodes of the path i , and $\alpha_{\text{AMul}} \in (0, 1]$ is the cost model parameter.

The AMul cost model relies on the following assumption: each unexplored node would reduce $\|\mathbf{r}^i\|_2$ by a rate proportional to the decay occurred during the last expansion of the path i . This rate is modeled by the auxiliary term $\alpha_{\text{AMul}} \|\mathbf{r}_l^i\|_2 / \|\mathbf{r}_{l-1}^i\|_2$, and the exponent $K_{\text{max}} - l^i$ extends this to all unexplored nodes along path i . The motivation is intuitive: since the search is expected to select nodes with descending correlation to \mathbf{Y} , a node is expected to reduce $\|\mathbf{r}^i\|_2$ less than its ancestors do. Note that this condition may be violated for a particular node. However, the auxiliary term is mostly computed over a number of nodes instead of a single one. Hence, it is practically sufficient if this assumption holds for groups of nodes. Moreover, the tree usually contains multiple paths which may lead to the correct solution. The fact that some of these paths violate this assumption does not actually harm the recovery. This behavior is similar to the other cost models in [1]. The empirical results in Section 3 indicate that these cost models are useful in practice.

The adaptive structure of the AMul model allows for a larger α than the Mul model. This reduces the auxiliary term. Consequently, the search favors longer paths, explores fewer nodes and terminates faster as demonstrated in Section 3.

In the rest, we identify the cost model employed by A*OMP with an appropriate prefix. AMul–A*OMP and Mul–A*OMP denote the use of AMul and Mul cost models, respectively.

1.2.5. Relations of A*OMP to recent proposals

One of the first and trivial combinations of the tree search with matching pursuit type algorithms has been suggested in [27] in 2001. Two strategies have been considered for exploring a tree with branching factor K , that is where each node has K children only.³ MP:K has a depth-first nature, that

is the candidate paths are explored one by one. The algorithm first explores a complete path up to the maximum depth. If this path does not yield the desired solution, the tree is backtracked and other candidates are explored sequentially until the solution is found. The other variant, MP:M-L, is based on breadth-first search. It processes all leaf nodes at a certain depth at once by exploring K children of each leaf and keeps the best M among the new candidates. The process is repeated until tree depth becomes L , and the path with the lowest residual is returned.

This idea has recently been revisited in [28], where the algorithm is referred to as multipath matching pursuit (MMP). As in [27], breadth-first (MMP-BF) and depth-first (MMP-DF) strategies have been evaluated to explore a search tree with branching factor L . For tractability, MMP-DF sets a limit on the number of sequentially explored paths. As a novel contribution, [28] provides RIP-based theoretical guarantees for MMP. Note that these guarantees are applicable when the number of new paths per level or the number of explored paths are not limited, i.e., when no pruning is applied.

Though both ideas are based on exploring a search tree, these algorithms are fundamentally different than A*OMP, where the tree search is guided by adaptive cost models in an intelligent manner. Instead, MMP-BF and MMP-DF are rather unsophisticated techniques where tree search follows a predefined order. We compare A*OMP and MMP via recovery simulations in Section 3. MMP-DF is chosen among the two variants, since it is referred to as the practical one in [28].

1.3. Outline and contributions

This paper at hand concentrates on detailed analyses of the sparse signal recovery performance of A*OMP. Particularly, we concentrate on the variant AMul–A*OMP_e which extends the general form in [1] by the novel AMul cost model from [5] and residue-based termination. We present new theoretical and empirical results to demonstrate the superiority of this variant not only over the A*OMP variants in [1], but also over some conventional sparse recovery methods. Note that AMul–A*OMP_e has only been preliminarily tested in [5] by a set of limited simulations, which are far away from providing enough evidence to generalize its performance. This paper presents a detailed empirical investigation of AMul–A*OMP_e, without which the performance analyses would not be complete. These simulations significantly enrich the findings of [5] by previously unpublished results which include phase transition comparisons for different signal statistics, demonstration on an image, a faster hybrid approach and optimality analyses. The results reveal not only the superior recovery accuracy of AMul–A*OMP_e, but also the improvements in the speed of the algorithm due to the residue-based termination and the AMul cost model.

On the other hand, our theoretical findings not only include RIP-based exact recovery guarantees for exact recovery of sparse signals via A*OMP_K and A*OMP_e, but also provide means for comparison of different termination criteria. The former states RIP conditions for the exact recovery of sparse signals from noise-free measurements,

³ In this context, K is not related to the sparsity level of \mathbf{x} as before. We denote the branching factor as K in order to be consistent with [27].

while the latter addresses the recovery improvements when the residue-based termination is employed instead of the sparsity-based one. For the analyses, we employ a method similar to the analyses of the OMP algorithm in [26,29]. In Section 2.3, we develop a RIP condition for the success of a single A*OMP iteration, which forms a basis for the following theoretical analyses. In Section 2.4, we derive a general recovery condition for exact recovery via A*OMP_K. As intuitively expected, this condition is less restrictive than the K-step OMP recovery condition [26]. Section 2.5 presents a very similar general condition for exact recovery via A*OMP_e. In addition, we establish an online recovery condition for exact recovery of a signal with A*OMP_e. Section 2.7 compares the general and online recovery conditions, clarifying that the latter is less restrictive. This suggests that A*OMP_e possesses stronger recovery capabilities than A*OMP_K.

Section 3 compares the recovery accuracy of AMul–A*OMP_e to other A*OMP variants, basis pursuit (BP) [6], subspace pursuit (SP) [10], OMP [2], iterative hard thresholding (IHT) [11], iterative support detection (ISD) [21], smoothed ℓ_0 (SLO) [17], MMP [28], and forward-backward pursuit (FBP) [12]. The main contribution is the phase transitions which are obtained by computationally expensive experiments for different signal types. These generalize the strong recovery capability of AMul–A*OMP_e over a wide range of N , M and K . We investigate the recovery rates and average recovery error as well. Run times illustrate the acceleration with the AMul cost model and the residue-based termination. An hybrid of A*OMP_e and OMP demonstrates how the recovery speed can be improved without losing the accuracy. The sparse image recovery problem represents a more realistic coefficient distribution than the other artificial examples.

2. Theoretical analysis of A*OMP

In this section, we develop theoretical guarantees for signal recovery with A*OMP. We first visit the restricted isometry property and then provide some related preliminary lemmas. Then, we concentrate recovery with A*OMP.

2.1. Restricted isometry property

Restricted isometry property (RIP) [7] provides an important means for theoretical guarantees in sparse recovery problems. A matrix Φ is said to satisfy the L -RIP if there exists a restricted isometry constant (RIC) $\delta_L \in (0, 1)$ satisfying

$$(1 - \delta_L) \|\mathbf{x}\|_2^2 \leq \|\Phi \mathbf{x}\|_2^2 \leq (1 + \delta_L) \|\mathbf{x}\|_2^2, \quad \forall \mathbf{x}: \|\mathbf{x}\|_0 \leq L.$$

Some random matrices, such as Gaussian or Bernoulli matrices, satisfy the L -RIP with high probabilities if L , M and N satisfy some specific conditions [30,31]. Exploiting this property, RIP has been utilized to obtain recovery guarantees for sparse recovery algorithms [7,29,31–33].

2.2. Preliminaries

We now present some preliminary lemmas based on RIP:

Lemma 1 (Monotonicity of the RIC, Dai and Milenkovic [10]). Let R and S be positive integers such that $R > S$. Then, $\delta_R \geq \delta_S$.

Lemma 2 (Lemma 2, Wang and Shim [26]). Let $\mathcal{I} \subset \{1, 2, \dots, N\}$ and $|\mathcal{I}|$ denote the cardinality of \mathcal{I} . For any arbitrary vector $\mathbf{z} \in \mathbb{R}^{|\mathcal{I}|}$, RIP directly leads to

$$(1 - \delta_{|\mathcal{I}|}) \|\mathbf{z}\|_2 \leq \|\Phi_{\mathcal{I}}^* \Phi_{\mathcal{I}} \mathbf{z}\|_2 \leq (1 + \delta_{|\mathcal{I}|}) \|\mathbf{z}\|_2.$$

Lemma 3 (Lemma 1, Dai and Milenkovic [10]). Let $\mathcal{I}, \mathcal{J} \subset \{1, 2, \dots, N\}$ such that $\mathcal{I} \cap \mathcal{J} = \emptyset$. For any arbitrary vector $\mathbf{z} \in \mathbb{R}^{|\mathcal{I}|}$

$$\|\Phi_{\mathcal{I}}^* \Phi_{\mathcal{J}} \mathbf{z}\|_2 \leq \delta_{|\mathcal{I}|+|\mathcal{J}|} \|\mathbf{z}\|_2.$$

Lemma 4. Let K and B be positive integers, and $\lceil z \rceil$ denote the smallest integer greater than or equal to z . Then,

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

Proof. Corollary 2 of [29] states that Lemma 4 holds for $B = 1$. By Lemma 1, $\delta_{K+B} \geq \delta_{K+1}$ for $B > 1$. Hence, Lemma 4 also holds for $B > 1$. \square

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. We set $K - n_c = sK$ and replace into (3):

$$\frac{3\sqrt{B}}{\sqrt{K} + \sqrt{B}} \leq \frac{\sqrt{B}}{\sqrt{sK} + \sqrt{B}}$$

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 $n_c < K$, $sK = K - n_c \geq 1$. 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 result with (5) completes the proof. \square

2.3. Success condition of an A*OMP iteration

We define the success of an A*OMP iteration as $\Delta \mathcal{T}$ containing at least one correct index, i.e., $\Delta \mathcal{T} \cap$

$\{\mathcal{T} - \mathcal{T}^b\} \neq \emptyset$. The following theorem guarantees the success of an iteration:

Theorem 1. Let $n_c = |\mathcal{T}^b \cap \mathcal{T}|$ and $n_f = |\mathcal{T}^b - \mathcal{T}|$. When b is expanded, at least one index in $\Delta\mathcal{T}$ is in the support of \mathbf{x} , i.e., $\Delta\mathcal{T} \cap \{\mathcal{T} - \mathcal{T}^b\} \neq \emptyset$ if Φ 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. $\Delta\mathcal{T}$ can be defined as

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

\mathbf{r}^b is the residue from the orthogonal projection of \mathbf{Y} onto $\Phi_{\mathcal{T}^b}$. Therefore, $\mathbf{r}^b \perp \Phi_{\mathcal{T}^b}$, i.e., $\langle \phi_i, \mathbf{r}^b \rangle = 0$ if $i \in \mathcal{T}^b$. Hence,

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

(8) has only $K - n_c$ nonzero terms. Combining (8) and (7), we can write

$$\|\Phi_{\Delta\mathcal{T}}^* \mathbf{r}^b\|_2 = \max_{\mathcal{J}, |\mathcal{J}|=B} \|\Phi_{\mathcal{J}}^* \mathbf{r}^b\|_2 \geq c \|\Phi_{\mathcal{T} \cup \mathcal{T}^b}^* \mathbf{r}^b\|_2 \quad (9)$$

where the inequality holds since \mathcal{J} maximizes $\|\Phi_{\mathcal{J}}^* \mathbf{r}^b\|_2$, and c defines a scaling proportional to the number of nonzero terms:

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

Next, the residue can be written as

$$\mathbf{r}^b = \mathbf{y} - \Phi_{\mathcal{T}^b} \hat{\mathbf{x}}_{\mathcal{T}^b} = \Phi_{\mathcal{T}} \mathbf{x}_{\mathcal{T}} - \Phi_{\mathcal{T}^b} \hat{\mathbf{x}}_{\mathcal{T}^b} = \Phi_{\mathcal{T} \cup \mathcal{T}^b} \mathbf{z} \quad (10)$$

where $\mathbf{z} \in \mathbb{R}^{K+n_f}$. Using Lemma 2, (9) and (10), we write

$$\|\Phi_{\Delta\mathcal{T}}^* \mathbf{r}^b\|_2 \geq c \|\Phi_{\mathcal{T} \cup \mathcal{T}^b}^* \Phi_{\mathcal{T} \cup \mathcal{T}^b} \mathbf{z}\|_2 \geq c(1 - \delta_{K+n_f}) \|\mathbf{z}\|_2.$$

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

$$\|\Phi_{\Delta\mathcal{T}}^* \mathbf{r}^b\|_2 = \|\Phi_{\Delta\mathcal{T}}^* \Phi_{\mathcal{T} \cup \mathcal{T}^b} \mathbf{z}\|_2 \leq \delta_{K+n_f+B} \|\mathbf{z}\|_2$$

by Lemma 3. Clearly, this never occurs if

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

or equivalently

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

Following Lemma 1, $\delta_{K+n_f+B} \geq \delta_{K+n_f}$. Hence, (11) is satisfied when $((1/c)+1)\delta_{K+n_f+B} < 1$, or equivalently

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

This guarantees that $\Delta\mathcal{T} \cap \mathcal{T} \neq \emptyset$. Moreover, since $\langle \phi_i, \mathbf{r}^b \rangle = 0$ for all $i \in \mathcal{T}^b$, $\Delta\mathcal{T} \cap \mathcal{T}^b = \emptyset$. Hence, we conclude $\Delta\mathcal{T} \cap \{\mathcal{T} - \mathcal{T}^b\} \neq \emptyset$, that is the A*OMP iteration is successful. \square

Below, Theorem 1 is used as a basis for exact recovery. Note that we assume $\sqrt{B} \leq \sqrt{K-n_c}$ in the rest and skip the term $\frac{1}{2}$ in Theorem 1 for simplicity. This can be justified by the fact that B is chosen small (such as 2 or 3) in practice.

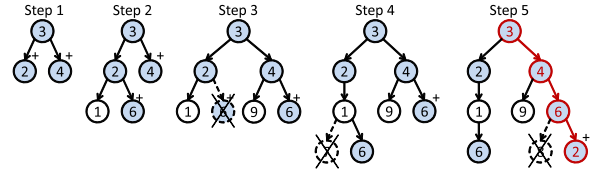


Fig. 2. Optimality during the search. The true support is $\{3, 4, 6, 2\}$.

2.4. Exact recovery conditions for A*OMP_K

First, let us introduce some definitions:

Optimal path: Path i is said to be *optimal* if $\mathcal{T}^i \subset \mathcal{T}$.

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

Fig. 2 illustrates the optimality notion with a typical example. Indices from the true support are shown by the filled nodes. The optimal paths, which contain filled nodes only, are designated with '+'. Since $K_{\max} = K = 4$, the paths become complete with four nodes. The nodes which are pruned due to $P=3$ are crossed. Step 3 exemplifies the optimal pruning. Though an optimal path is pruned, there still remains another optimal one. Expanding this optimal path in Step 5, the search obtains the true solution represented by the complete and optimal path $\{3, 4, 6, 2\}$.

Now, we present the exact recovery condition for A*OMP_K:

Theorem 2. Set $\varepsilon = 0$ and $K_{\max} = K$. Let Φ be full rank and $2K \leq M$ hold.⁴ Assume that pruning is optimal. Then, A*OMP_K perfectly recovers all K -sparse signals from noise-free measurements if Φ satisfies RIP with

$$\delta_{K+B} < \frac{\sqrt{B}}{\sqrt{K} + \sqrt{B}}. \quad (12)$$

Proof. Let us start with the initialization. With $n_c = n_f = 0$, Theorem 1 assures success of the first iteration.

Next, consider A*OMP_K selects an optimal path of length l , i.e., $n_c = l$, at some step. By Theorem 1, expansion of this path is successful if

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

which is already satisfied when (12) holds.

Now, there exists some optimal paths after initialization. Moreover, expanding an optimal path introduces at least one longer optimal path, and by assumption pruning cannot remove all of these. Altogether, these guarantee the existence of at least one optimal path in the tree at any iteration.

On the other hand, the criterion $\varepsilon = 0$ requires that the residue should vanish for termination. Since $2K \leq M$ and Φ is full rank, the residue may vanish if and only if \mathcal{T} is a subset of the support estimate.⁵ Therefore, the search must terminate at a complete optimal path containing \mathcal{T} unless there remain no optimal paths in the search tree. Together

⁴ $2K \leq M$ is a global condition for the uniqueness of all K -sparse solutions. Hence this condition is necessary for any sparse recovery algorithm.

⁵ As linearly dependent subsets should contain at least $M+1$ columns of Φ , any other solution should be at least $(M-K+1)$ -sparse.

Correct support: {1,4,5,9}, $K_{\max} = 6$

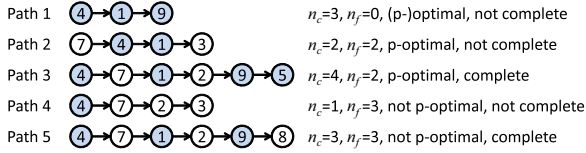


Fig. 3. Completeness and p -optimality with respect to n_c and n_f .

with the existence of at least one optimal path, this guarantees exact recovery.

Note that the condition $\varepsilon = 0$, stated in Theorem 2 for the sake of theoretical correctness, translates into a very small ε in practice to account for the numerical computation errors.

We observe that the K -step exact recovery condition of OMP, $\delta_{K+1} < 1/(\sqrt{K}+1)$, is a special case of Theorem 2 when $B = I = 1$. Moreover, when the bounds for OMP and A^*OMP_K are compared, (12) is clearly less restrictive, which explains the improved recovery accuracy of A^*OMP_K .

Theorem 2 is closely related to the theoretical analysis of MMP in [28]. It can be trivially shown that Theorem 2 is also applicable to MMP. Moreover, it implies a better, i.e., less restricted, recovery condition when compared to [28], where the condition is stated as $\delta_{K+B} < 1/(\sqrt{K}+2\sqrt{B})$. \square

2.5. Exact recovery with A^*OMP_e

We extend the definitions in the previous section for A^*OMP_e where $K_{\max} > K$. First, note that path i is now complete if $l^i = K_{\max} > K$. Next, we introduce the following definitions:

Potentially optimal path: A path is said to be *potentially optimal* (p -optimal) if $n_f \leq K_{\max} - K$. A p -optimal path can be expanded into a superset of \mathcal{T} with at most K_{\max} nodes. Note that an optimal path is a special case where $K_{\max} = K$.

Potentially optimal pruning: Pruning is defined as p -optimal if it does not remove all p -optimal paths from the search tree.

Fig. 3 depicts some examples of p -optimality and completeness. Clearly, a path is p -optimal only if a superset of \mathcal{T} is among its extensions. Moreover, path 1 reveals that the optimal and p -optimal path notions are equal when $n_f = 0$.

Next, we state the following lemma:

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

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

holds for path i , the set of its extensions by each of its best B children contains at least one p -optimal path with n_c+1 correct indices. Moreover, (15) holds for all p -optimal paths in this set.

Proof. By Theorem 1, expansion of path i is successful when (15) holds. Hence, it introduces at least one p -optimal path, say j , with n_c+1 correct and n_f incorrect indices.

Moreover, the upper bounds from (15) 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 sides are for the upper bounds on the RIC's corresponding to the number of correct and false nodes in path i and j , respectively. Since the upper bound is larger for path j , and (15) holds for path i , (15) should also hold for path j . \square

Now an online recovery condition can be presented for A^*OMP_e :

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

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

Proof. As $K+n_f \leq K_{\max}$, the best path at this step, b , is p -optimal. Moreover, Lemma 6 and (15) guarantee p -optimality of at least one child of b . By assumption pruning cannot remove all p -optimal paths. Altogether, these guarantee the existence of at least one p -optimal path in the tree until termination. On the other hand, the criterion $\varepsilon = 0$ requires that the residue should vanish for termination. Since $K_{\max} \leq M - K$ and Φ is full rank, the residue may vanish if and only if the support estimate is a superset of \mathcal{T} . Therefore, the search must terminate at a p -optimal path containing the true support unless there remain no p -optimal paths in the search tree. As Φ is full rank, the orthogonal projection of \mathbf{Y} onto this set yields exactly \mathbf{x} . Combining with the existence of at least one p -optimal path, exact recovery of \mathbf{x} is guaranteed. \square

Note that the condition $\varepsilon = 0$, stated in Theorem 3 for the sake of theoretical correctness, translates into a very small ε in practice to account for the numerical computation errors.

Since Theorem 3 depends on the existence of a p -optimal path satisfying (15), it does not provide overall guarantees for all K -sparse signals as Theorem 2 does. In contrast, Theorem 3 implies online guarantees depending on the support estimates of the intermediate (i.e., neither complete nor empty) p -optimal paths. In fact, it is hardly possible to provide guarantees for the existence of such paths. Nonetheless, Theorem 3 can be generalized starting with the empty path:

Theorem 4. Set $\varepsilon = 0$, $I \geq B$ and $K_{\max} \leq M - K$. Let Φ be full rank. Assume pruning is optimal. Then, A^*OMP_e perfectly recovers all K -sparse signals from noise-free measurements if Φ satisfies RIP with

$$\delta_{K+B} < \frac{\sqrt{B}}{\sqrt{K}+\sqrt{B}}. \quad (16)$$

We omit the proof of Theorem 4, since it follows trivially from Theorem 3 by replacing $n_f = n_c = 0$.

Theorem 4 provides overall guarantees for all sparse signals as **Theorem 2**. We observe that both theorems require the same RIP condition for exact recovery of all sparse signals.

Although **Theorems 2 and 4** provide similar overall guarantees, A^*OMP_e also possesses the online guarantees of **Theorem 3**. **Section 2.7** presents an analytical comparison of the conditions in **Theorems 3 and 4**. This study states that **Theorem 3** may be satisfied even when **Theorem 4** fails. This reveals the importance of the online guarantees to comprehend the improved recovery accuracy of A^*OMP_e .

2.6. A note on pruning

Theorems 2 and 4 both rely on optimal/ p -optimal pruning, which is hard to prove analytically. Though this may be seen as a limitation of the theoretical findings, it is obvious that pruning is unavoidable for the tractability of the search. As an empirical justification, it is important to observe that the true solution may be reached along different paths in the tree. This is due to the fact that the ordering of the nodes is not important. Hence, the tree is subject to contain a large number of candidate solutions (optimal/ p -optimal paths), while it is enough for exact recovery when only one of these optimal paths is not pruned.

Moreover, the theoretical analysis of MMP in [28] is also subject to an equivalent assumption in practice. Though the authors analyse MMP without any limit on the number of explored paths, they acknowledge that this is impractical. Similar to A^*OMP , they limit the number of paths for the empirical evaluation of MMP. It is clear that the theoretical findings of [28], which do not address this pruning strategy, are practically meaningful only with an assumption on pruning.

2.7. On the validity of the online guarantees

To address the validity of the online condition in **Theorem 3**, we show that it can be satisfied when the overall guarantees in **Theorem 4** fail. The next theorem reveals that a p -optimal path satisfying (15) may be found even when (16) fails.

Theorem 5. Assume $K \geq (3+2\sqrt{B})^2$. If $1 \leq n_f + B \leq \lfloor K/2 \rfloor$ and n_c satisfies (4) at some intermediate iteration, (15) becomes less restrictive than (16).

Proof. Assume that

$$\delta_{K+n_f+B} \geq \frac{3\sqrt{B}}{\sqrt{K}+\sqrt{B}}. \quad (17)$$

Since $n_f + B \leq \lfloor K/2 \rfloor$, we can write $3\lfloor K/2 \rfloor \geq K + n_f + B$. By **Lemma 1**, we obtain

$$\delta_{3\lfloor K/2 \rfloor} \geq \frac{3\sqrt{B}}{\sqrt{K}+\sqrt{B}}.$$

Then, **Lemma 4** yields

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

which clearly contradicts (12). In contrast, **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 (4) when $K \geq (3+2\sqrt{B})^2$. That is, there exists some range of δ_{K+n_f+B} such that

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

Hence, under the given conditions, there exists some δ_{K+n_f+B} satisfying (15), but no δ_{K+B} satisfying (16). \square

Theorem 5 clarifies that **Theorem 3** may hold even when **Theorem 4** fails. In other words, A^*OMP_e possesses online guarantees for some sparse signals for which the overall guarantees fail. This explains why A^*OMP_e improves the recovery accuracy over A^*OMP_K , and reveals that the residue-based termination is more optimal for noise-free sparse signal recovery than its sparsity-based counterpart.

Note that **Theorem 5** is based on some n_f , n_c and K ranges for which it is provable. This is enough, since even a single supporting case already establishes the validity of **Theorem 3**. On the other hand, we expect it to be valid for a wider range. This intuition is also supported by the following simulations, where A^*OMP_e improves recovery in almost all cases.

3. Empirical analyses

We demonstrate A^*OMP in comparison to BP [6], SP [10], OMP [2], ISD [21], SLO [17], IHT [11], FBP [12] and MMP-DF [28] in various scenarios involving synthetically generated signals and images. We mainly concentrate on comparison with different algorithms. Regarding the impact of A^*OMP parameters such as B , P and α on the recovery performance, we refer the reader to [1], where the matter has been discussed with detailed empirical analysis. The numerical choices of the parameters in this work are mainly based on these findings, which we do not repeat here.

3.1. Experimental setup

Unless given explicitly, the setup is as follows: We set $L=3$, $B=2$ and $P=200$. For A^*OMP_e , ϵ is set to 10^{-6} . This ϵ is shared by OMP, which also runs until $\|\mathbf{r}\|_2 \leq \epsilon \|\mathbf{y}\|_2$. We select $\alpha_{Mul} = 0.8$ for Mul- A^*OMP_K , $\alpha_{Mul} = 0.9$ for Mul- A^*OMP_e and $\alpha_{AMul} = 0.97$ for AMul- A^*OMP_e . We employ FBP with $\alpha = 0.2M$ and $\beta = \alpha - 1$ as suggested in [12]. For MMP-DF, we set the branching factor $L=6$ following [28], and allow a maximum of 200 paths for a fair comparison with the A^*OMP variants where $P=200$. Each test is repeated over a randomly generated set of S sparse samples. For each sample, Φ is drawn from the Gaussian distribution with mean zero and standard deviation $1/N$. The nonzero entries of the test samples are selected from three random ensembles. The nonzero entries of the Gaussian sparse signals follow standard Gaussian distribution while those of the uniform sparse signals are distributed uniformly in $[-1, 1]$. The constant amplitude random sign (CARS) sparse signals have unit magnitude nonzero elements with random sign. The average normalized mean-squared-error (ANMSE) is defined as

$$ANMSE = \frac{1}{S} \sum_{i=1}^S \frac{\|\mathbf{x}_i - \hat{\mathbf{x}}_i\|_2^2}{\|\mathbf{x}_i\|_2^2} \quad (18)$$

where $\hat{\mathbf{x}}_i$ is the recovery of the i th test vector \mathbf{x}_i .

We perform the A*OMP recovery using the AStarOMP software.⁶ AStarOMP implements the A* search by an efficient trie⁷ structure [34], where the nodes are ordered with priorities proportional to their inner products with \mathbf{Y} . This maximizes the number of shared nodes between paths, allowing a more compact tree representation and faster tree modifications. The orthogonal projection is performed via the QR factorization.

3.2. Exact recovery rates and reconstruction error

The first simulation deals with the recovery of Gaussian sparse signals, where $N=256$, $M=100$, $K \in [10, 50]$, $K_{\max} = 55$ and $S=500$. The results are depicted in Fig. 4, which reveals that A*OMP performs significantly better than the other algorithms. A*OMP_e provides exact recovery until $K = 40$, which is clearly the best. We observe similar ANMSE among A*OMP variants, while the residue-based termination improves the exact recovery rates significantly. Evidently, A*OMP_e is better than A*OMP_K at identifying smaller magnitude entries, which hardly change the ANMSE, however increase the exact recovery rates. It is also important that all A*OMP variants yield higher recovery rates than MMP-DF. Especially, Mul–A*OMP_K and MMP-DF are interesting to compare since both return solutions with K indices. We observe that Mul–A*OMP_K yields significantly higher recovery rates than MMP-DF, which implies the effectiveness of the sophisticated search techniques employed in A*OMP.

As for the average run times,⁸ both the residue-based termination and the AMul cost model significantly accelerate A*OMP due to the relaxation of α to larger values. Since AMul–A*OMP_e can afford the largest α , it is the fastest A*OMP variant. This confirms the claim in Section 1.2 that increasing α reduces the number of explored nodes and accelerates A*OMP.

In addition to this example, [5] contains simulations for uniform and binary sparse signals and noisy measurements. These simulations indicate that AMul–A*OMP_e improves the recovery for uniform sparse signals and noisy cases as well.

3.3. 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 λ [35], 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 perform the recovery. Setting the

exact recovery criterion as $(\|\mathbf{x}_i - \hat{\mathbf{x}}_i\|_2) / \|\mathbf{x}_i\|_2 \leq 10^{-2}$, we count the exactly recovered samples. The phase transitions are then obtained as in [35]. For each λ , we employ a generalized linear model with logistic link to describe the exact recovery curve over ρ , and then find ρ which yields 50% exact recovery probability. Combination of these ρ values gives the phase transition curve.

Let us first concentrate on K_{\max} and define the normalized measure $\rho_{\max} = K_{\max}/M$. Once we identify the optimal ρ_{\max} over λ , we can set K_{\max} for particular λ and M . To find the optimal ρ_{\max} , we have run a number of simulations and observed that the recovery performance of AMul–A*OMP_e is quite robust to ρ_{\max} , with a perturbation up to 3%. Hence, the recovery accuracy is mostly independent of ρ_{\max} . Yet, based on our experience, we set $\rho_{\max} = 0.5 + 0.5\lambda$ taking into account both the accuracy and complexity of the search.

Fig. 5 depicts the phase transition curves. Clearly, AMul–A*OMP_e yields better phase transitions than the other algorithms for the Gaussian and uniform sparse signals. We also observe that FBP provides fast approximations with better accuracy than BP and the two other greedy competitors, SP and OMP for these two cases. This reveals that FBP is actually suitable to applications where speed is crucial. On the other hand, BP and ISD are the best performers for the CARS case, while AMul–A*OMP_e is the third best. We observe that BP is robust to the coefficient distribution, while the phase transitions for AMul–A*OMP_e and OMP exhibit the highest variation among distributions. This indicates that OMP-type algorithms are more effective when the nonzero elements span a wide magnitude range such as the Gaussian distribution. Moreover, if this range gets wide enough, even OMP can outperform BP. In parallel, the CARS ensemble is referred to as the most challenging case for the greedy algorithms in the literature [10,35]. This can be understood analytically by considering the span of the correlation between $\Phi_{\mathcal{T}}$ and \mathbf{Y} . The detailed analytical analysis in [36] state that this span gets smaller when the magnitudes of the nonzero elements get closer, and vice versa. When this span gets smaller (for the CARS ensemble in the limit), wrong indices are more likely to be selected by OMP, increasing the failure rate.

3.4. A hybrid approach for faster practical recovery

We have observed that OMP provides exact recovery up to some mid-sparsity range. Moreover, there are regions where AMul–A*OMP_e provides exact recovery while OMP recovery rates are also quite high. In these regions, we can accelerate the recovery without sacrificing the accuracy by a two-stage hybrid scheme. We first run OMP, and then AMul–A*OMP_e only if OMP fails. Assuming that $K + K_{\max}$ -RIP holds, a non-vanishing residue indicates OMP failure, and then AMul–A*OMP_e is run. This reduces the number of AMul–A*OMP_e runs and accelerates the recovery. Moreover, we use the order by which OMP chooses the vectors for setting the priorities of trie nodes. 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 trie modification costs.

According to the recovery results in Fig. 6, AMul–A*OMP_e and the hybrid approach yield identical exact recovery rates, while the latter is significantly faster.

⁶ Available at <http://myweb.sabanciuniv.edu/karahanoglu/research/>.

⁷ Trie is an ordered tree data structure in computer science. The ordering provides important advantages for the implementation of A*OMP, such as reducing the cost of addition/deletion of paths and finding equivalent paths.

⁸ OMP and A*OMP are tested using the AStarOMP software. The other algorithms are ignored as they run in MATLAB, which is slower.

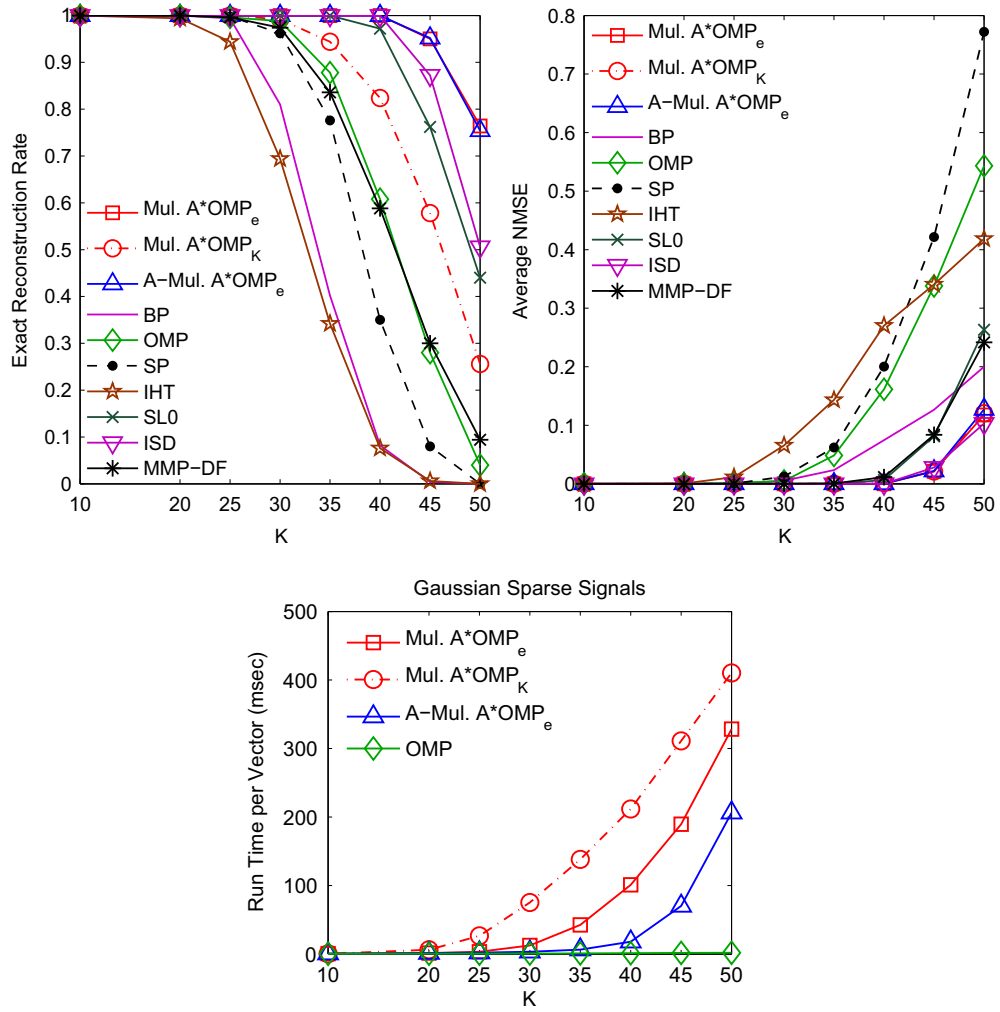
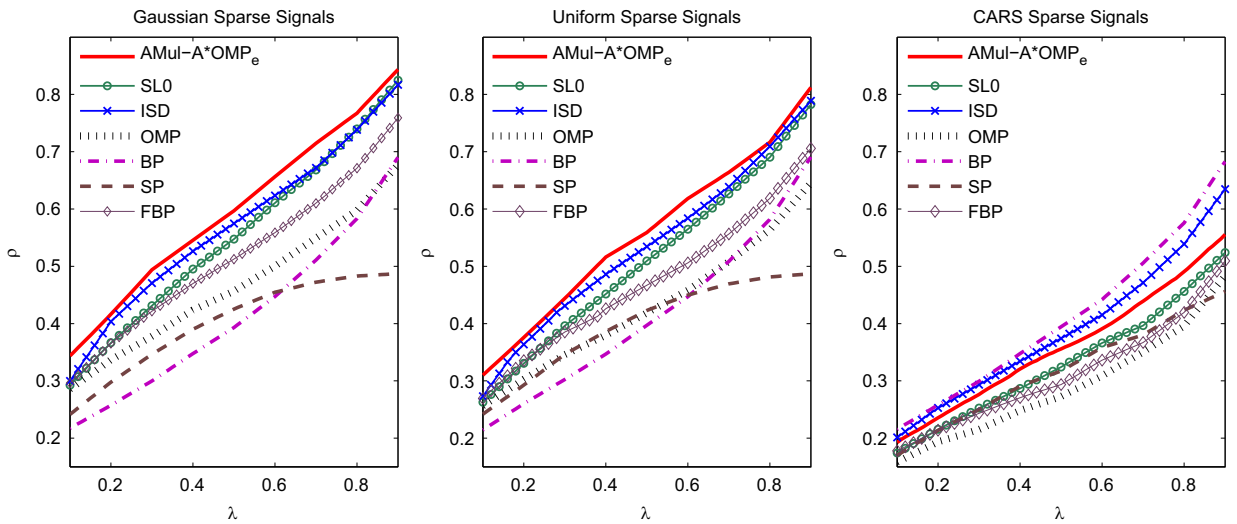


Fig. 4. Recovery results and average run time for the Gaussian sparse signals.

Fig. 5. Phase transitions of AMul-A*OMP_e, BP, SP, OMP, ISD and SL0 for Gaussian, uniform and CARS sparse signals.

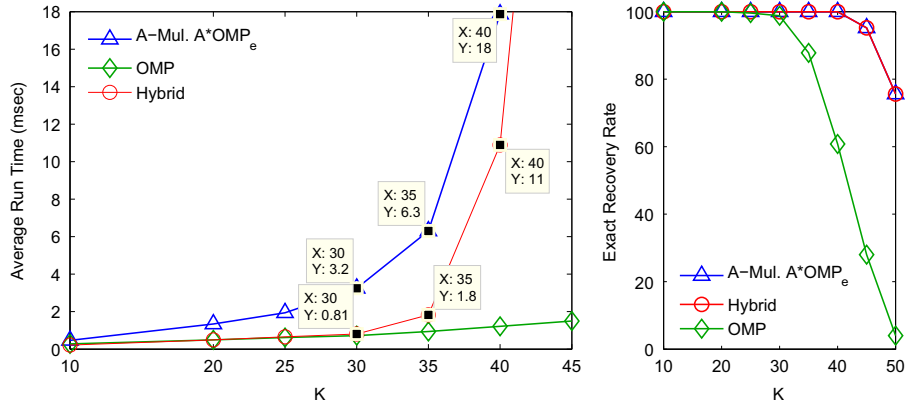


Fig. 6. Performance of the hybrid scheme for the Gaussian sparse vectors.

Table 2

PSNR values for images reconstructed using different algorithms. Maximum and minimum PSNR values are shown in bold and in italics, respectively. Mean PSNR values over the whole set of images are given in the last row of the table.

| Image | BP | OMP | SP | IHT | ISD | SLO | MMP-DF | AMul–A*OMP | |
|-----------|------|------|------|------|------|------|--------|------------|-------------|
| | | | | | | | | B=2 | B=3 |
| Bridge | 29.9 | 26.9 | 24.6 | 19.8 | 32.1 | 29.1 | 36.9 | 46.8 | 51.4 |
| Lena | 33.5 | 29.6 | 27.5 | 22.9 | 33.2 | 30.6 | 36.4 | 42.6 | 47.1 |
| Tracy | 40.6 | 36.8 | 33.9 | 27.6 | 39.4 | 38.2 | 43.8 | 52.5 | 56.8 |
| Pirate | 31.7 | 27.7 | 25.3 | 21.5 | 32.4 | 30.3 | 35.7 | 40.3 | 43.4 |
| Cameraman | 34.4 | 30.7 | 28.5 | 23 | 33.9 | 31.8 | 37.4 | 48.3 | 54.7 |
| Mandrill | 28.3 | 24.4 | 22.1 | 19.2 | 29.7 | 26.8 | 32.1 | 36.3 | 39.9 |
| Mean PSNR | 33.1 | 29.4 | 27 | 22.3 | 33.5 | 31.1 | 37.1 | 44.5 | 48.8 |

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*OMP_e only for those instances.

3.5. Demonstration on a sparse image

To illustrate AMul–A*OMP_e on a more realistic coefficient distribution, we demonstrate recovery of some commonly used 512×512 images including ‘Lena’, ‘Tracy’, ‘Cameraman’, etc. The recovery is performed in 8×8 blocks in order to break the problem into smaller and simpler subproblems. Each image is first preprocessed to obtain K -sparse blocks in the 2D Haar Wavelet basis Ψ by keeping the K largest magnitude wavelet coefficients for each block.⁹ We select $K=12$ for the image ‘Bridge’, and $K=14$ for the rest. $M=32$ observations are taken from each block. The entries of Φ are randomly drawn from the Gaussian distribution with mean 0 and standard deviation $1/N$. We set $l=3$, $P=200$ and $K_{\max}=20$. α_{AMul} is reduced to 0.85 in order to compensate the decrement in the auxiliary term of (2) due to smaller K_{\max} . Peak signal-to-noise ratio (PSNR) values obtained by employing different recovery algorithms are given in Table 2.

⁹ This formulation involves the structured dictionary Ψ for sparse representation of the images. The observation model becomes $\mathbf{y} = \Phi\Psi\mathbf{x}$, where the reconstruction basis is not Φ alone, but $\Phi\Psi$.

For each image, maximum and minimum PSNR values obtained are shown in bold and in italics, respectively. Mean PSNR for each algorithm is also given on the last row. AMul–A*OMP_e exhibits significant improvements over the competitors and yields the maximum PSNR for all images tested. Increasing B from 2 to 3 further improves PSNR, yielding improvements of 15.7 dB over BP, 15.3 dB over ISD, 17.7 dB over SLO, and 11.7 dB over MMP-DF on the average. As a visual example, we depict the reconstruction of the test image ‘Bridge’ using BP and AMul–A*OMP_e with $B=3$ in Fig. 7. In this case, BP yields 29.9 dB PSNR, while AMul–A*OMP_e improves the PSNR to 51.4 dB. Though not depicted in Fig. 7, AMul–A*OMP_e outperforms BP with 46.8 dB when $B=2$ as well. A detailed investigation of the recovered images reveals that AMul–A*OMP_e improves the recovery especially at detailed regions and boundaries.

4. Summary

The fundamental goal of this paper is a comprehensive analysis of sparse recovery using A*OMP, with a particular focus on the novel variant AMul–A*OMP_e. We have addressed this issue with emphasis on both theoretical and practical aspects.

We have presented a theoretical analysis of signal recovery with A*OMP. We have first derived a RIP condition for the success of an A*OMP iteration. Then, we have generalized this result for the exact recovery of all K -sparse signals from noise-free measurements both with

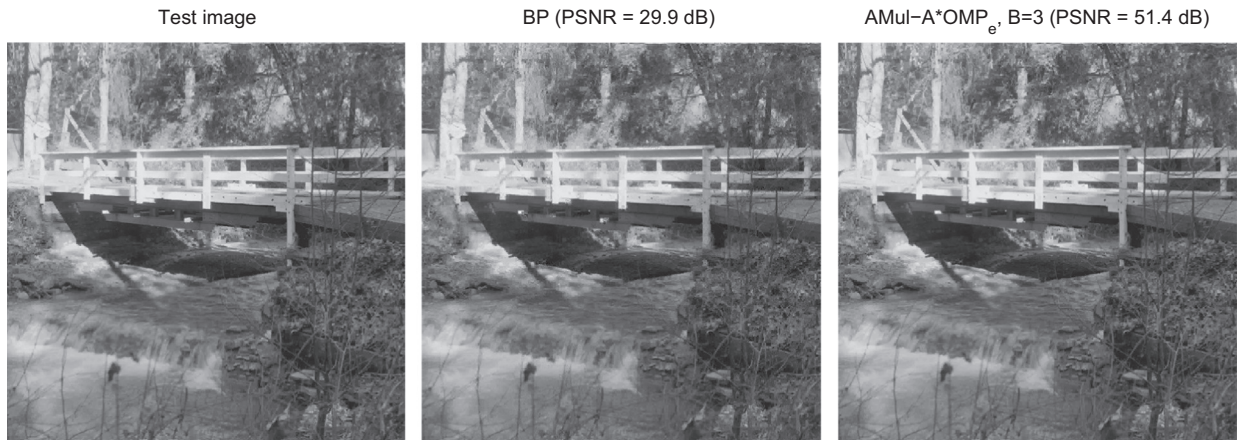


Fig. 7. Recovery of the image 'Bridge' using BP and AMul–A*OMP_e.

A*OMP_K, where the termination is based on the sparsity level K , and with A*OMP_e, which employs the residue-based termination criterion. We have observed that both A*OMP variants enjoy similar RIP-based general exact recovery guarantees. In addition, we have presented online guarantees for A*OMP_e, which can be satisfied even when the general guarantees fail. This has led to the conclusion that A*OMP_e is more advantageous for sparse recovery, since it enjoys some online guarantees in addition to the general recovery guarantees analogous to those of A*OMP_K. This encourages utilizing the residue-based termination instead of the sparsity-based one for recovery from noise-free observations.

We have demonstrated AMul–A*OMP_e in a wide range of recovery simulations involving sparse signals with different characteristics. According to the results, A*OMP performs better recovery than all the other candidates for uniform and Gaussian sparse signals. Among A*OMP variants, AMul–A*OMP_e promises the most accurate recovery and fastest execution times. For CARS sparse signals, AMul–A*OMP_e recovery is still better than the involved greedy alternatives, however BP is the most accurate algorithm in this case. We have also shown that the search can be significantly accelerated without sacrificing the accuracy via a hybrid approach, which first applies OMP, and then AMul–A*OMP_e only if OMP fails. Finally, we have employed AMul–A*OMP_e on sparse images, where it improves the recovery significantly over BP.

The AMul cost model with the residue-based termination has demonstrated strong empirical performance while also providing more greed due to the allowance for a larger α . Hence, AMul–A*OMP_e turns out to be the most promising A*OMP variant in this paper. As future work, it is worth to investigate different cost model structures which may improve speed and convergence of the algorithm in specific problems. For example, the cost model may be formulated to reflect the expected statistics of the signal of interest. Such a strategy would be problem-dependent, however it may guide the algorithm faster and more accurately to the desired solution. Combining A*OMP with sparsity models, i.e., signals with specific sparsity patterns, is another promising future work direction.

References

- [1] N.B. Karahanoglu, H. Erdogan, A* orthogonal matching pursuit: best-first search for compressed sensing signal recovery, *Digit. Signal Process.* 22 (July (4)) (2012) 555–568.
- [2] Y.C. Pati, R. Rezaifar, P.S. Krishnaprasad, Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition, in: *Proceedings of the 27th Asilomar Conference on Signals, Systems and Computers*, Los Alamitos, CA, vol. 1, 1993, pp. 40–44.
- [3] P.E. Hart, N.J. Nilsson, B. Raphael, A formal basis for the heuristic determination of minimum cost paths, *IEEE Trans. Syst. Sci. Cybern.* 4 (July (12)) (1968) 100–107.
- [4] F. Jelinek, *Statistical Methods For Speech Recognition*, MIT Press, Cambridge, MA, USA, 1997 (Chapter 6).
- [5] N.B. Karahanoglu, H. Erdogan, A comparison of termination criteria for A*OMP, in: *European Signal Processing Conference*, Bucharest, Romania, August 2012.
- [6] S. Chen, D. Donoho, M. Saunders, Atomic decomposition by basis pursuit, *SIAM J. Sci. Comput.* 20 (1) (1998) 33–61.
- [7] E. Candès, T. Tao, Decoding by linear programming, *IEEE Trans. Inf. Theory* 51 (December (12)) (2005) 4203–4215.
- [8] D. Donoho, Compressed sensing, *IEEE Trans. Inf. Theory* 52 (April (4)) (2006) 1289–1306.
- [9] J.A. Tropp, Algorithms for simultaneous sparse approximation. Part ii: convex relaxation, *Signal Process.* 86 (3) (2006) 589–602. (Sparse approximations in signal and image processing).
- [10] W. Dai, O. Milenkovic, Subspace pursuit for compressive sensing signal reconstruction, *IEEE Trans. Inf. Theory* 55 (May (5)) (2009) 2230–2249.
- [11] T. Blumensath, M.E. Davies, Iterative hard thresholding for compressed sensing, *Appl. Comput. Harmonic Anal.* 27 (3) (2009) 265–274.
- [12] N.B. Karahanoglu, H. Erdogan, Compressed sensing signal recovery via forward-backward pursuit, *Digit. Signal Process.* 23 (5) (2013) 1539–1548.
- [13] D. Sundman, S. Chatterjee, M. Skoglund, Distributed greedy pursuit algorithms, *Signal Process.* 105 (0) (2014) 298–315.
- [14] J.A. Tropp, A.C. Gilbert, M.J. Strauss, Algorithms for simultaneous sparse approximation. Part i: greedy pursuits, *Signal Process.* 86 (March (3)) (2006) 572–588.
- [15] S. Ji, Y. Xue, L. Carin, Bayesian compressive sensing, *IEEE Trans. Signal Process.* 56 (6) (2008) 2346–2356.
- [16] S.D. Babacan, R. Molina, A.K. Katsaggelos, Bayesian compressive sensing using Laplace priors, *IEEE Trans. Image Process.* 19 (1) (2010) 53–63.
- [17] H. Mohimani, M. Babaie-Zadeh, C. Jutten, A fast approach for over-complete sparse decomposition based on smoothed l0 norm, *IEEE Trans. Signal Process.* 57 (January (1)) (2009) 289–301.
- [18] M.R. Mohammadi, E. Fatemizadeh, M.H. Mahoor, Non-negative sparse decomposition based on constrained smoothed l0 norm, *Signal Process.* 100 (0) (2014) 42–50.
- [19] L.B. Montefusco, D. Lazzaro, S. Papi, A fast algorithm for nonconvex approaches to sparse recovery problems, *Signal Process.* 93 (9) (2013) 2636–2647.

- [20] T. Ince, A. Nacaroglu, N.I. Watsuji, Nonconvex compressed sensing with partially known signal support, *Signal Process.* 93 (1) (2013) 338–344.
- [21] Y. Wang, W. Yin, Sparse signal reconstruction via iterative support detection, *SIAM J. Imaging Sci.* 3 (3) (2010) 462–491.
- [22] E. Candès, M. Wakin, S. Boyd, Enhancing sparsity by reweighted ℓ_1 minimization, *J. Fourier Anal. Appl.* 14 (2008) 877–905.
- [23] J. Fang, Y. Shen, H. Li, Z. Ren, Sparse signal recovery from one-bit quantized data: an iterative reweighted algorithm, *Signal Process.* 102 (0) (2014) 201–206.
- [24] J.A. Tropp, A.C. Gilbert, Signal recovery from random measurements via orthogonal matching pursuit, *IEEE Trans. Inf. Theory* 53 (12) (2007) 4655–4666.
- [25] M.A. Davenport, M.B. Wakin, Analysis of orthogonal matching pursuit using the restricted isometry property, *IEEE Trans. Inf. Theory* 56 (September (9)) (2010) 4395–4401.
- [26] J. Wang, B. Shim, On the recovery limit of sparse signals using orthogonal matching pursuit, *IEEE Trans. Signal Process.* 60 (September (9)) (2012) 4973–4976.
- [27] S.F. Cotter, B.D. Rao, Application of tree-based searches to matching pursuit, in: 2001 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2001, pp. 3933–3936.
- [28] S. Kwon, J. Wang, B. Shim, Multipath matching pursuit, *IEEE Trans. Inf. Theory* 60 (May (5)) (2014) 2986–3001.
- [29] N.B. Karahanoglu, H. Erdogan, Online Recovery Guarantees for Orthogonal Matching Pursuit, Preprint (<http://arxiv.org/abs/1210.5991>), 2013.
- [30] M. Rudelson, R. Vershynin, Sparse reconstruction by convex relaxation: Fourier and Gaussian measurements, in: 40th Annual Conference on Information Sciences and Systems, 2006, pp. 207–212.
- [31] E. Candès, T. Tao, Near-optimal signal recovery from random projections: universal encoding strategies? *IEEE Trans. Inf. Theory* 52 (12) (2006) 5406–5425.
- [32] W. Xu, Y. Tian, J. Lin, Performance analysis of partial segmented compressed sampling, *Signal Process.* 93 (9) (2013) 2653–2663.
- [33] Y. Wang, J. Wang, Z. Xu, Restricted p-isometry properties of nonconvex block-sparse compressed sensing, *Signal Process.* 104 (0) (2014) 188–196.
- [34] R. Sedgewick, K. Wayne, *Algorithms*, 4th ed. Addison-Wesley Professional, Westford, Massachusetts, USA, 2011.
- [35] A. Maleki, D.L. Donoho, Optimally tuned iterative reconstruction algorithms for compressed sensing, *IEEE J. Sel. Top. Signal Process.* 4 (April (2)) (2010) 330–341.
- [36] N.B. Karahanoglu, Search-Based Methods for the Sparse Signal Recovery Problem in Compressed Sensing (Ph.D. thesis), Sabanci University, Istanbul, Turkey, January 2013.