A COMPARISON OF TERMINATION CRITERIA FOR A*OMP

Nazim Burak Karahanoglu[1,2], *Hakan Erdogan*[1]

[1]Department of Electronics Engineering, Sabanci University, Istanbul, Turkey
[2]Information Technologies Institute, TUBITAK-BILGEM, Kocaeli, Turkey

ABSTRACT

Heuristic search has recently been utilized for compressed sensing signal recovery problem by the A* Orthogonal Matching Pursuit (A*OMP) algorithm. A*OMP employs A* search on a tree with an OMP-based evaluation of the branches, where the search is terminated when the desired path length is achieved. The algorithm employs effective pruning techniques and cost models which make the tree search practical. Here, we propose two important extensions of A*OMP: We first introduce a novel dynamic cost model that reduces the search time. Second, we modify the termination criterion by stopping the search when ℓ_2 norm of the residue is small enough. Following the restricted isometry property, this termination criterion is more appropriate for our purposes. We demonstrate the improvements in terms of both reconstruction accuracy and computation times via a wide range of simulations.

Index Terms— Compressed sensing, A*OMP, A* termination criterion, A* auxiliary functions

1. INTRODUCTION

The fundamental problem of Compressed Sensing (CS) signal recovery is solving the minimization problem

$$\mathbf{x} = \arg\min \|\mathbf{x}\|_0 \quad s.t. \quad \mathbf{y} = \mathbf{\Phi}\mathbf{x}, \tag{1}$$

where $\mathbf{x}$ is a K-sparse signal of length N, $\mathbf{\Phi}$ is the observation matrix, also called the dictionary, of size $M \times N$ and $\mathbf{y}$ is the observation of length M where $M < N$. Many reconstruction algorithms have been suggested for solving (1), whose direct solution is intractable. These can be broadly categorized as [1] convex relaxation, greedy pursuits, Bayesian framework and nonconvex optimization.

Starting historically with Basis Pursuit (BP) [2], convex relaxation [3, 4, 5, 6] replaces the l_0 minimization in (1) with l_1 minimization, which can be solved via linear programming. Greedy pursuit algorithms such as Orthogonal Matching Pursuit (OMP) [7], Compressive Sampling Matching Pursuit (CoSaMP) [8] and Subspace Pursuit (SP) [9] employ iterative mechanisms which find approximate solutions by solving a stagewise constrained residue minimization problem.

The authors recently suggested a semi-greedy approach, A* Orthogonal Matching Pursuit (A*OMP) [10, 11]. This method solves the CS reconstruction problem with A* search [12, 13], on a tree whose branches are evaluated similar to OMP. Via pruning techniques and appropriate cost models, A*OMP was shown to improve the reconstruction significantly in many scenarios [10]. Note that A*OMP does not exploit any tree-based structured sparsity, as, for example, tree-based OMP [14] does, but covers all possible sparse representations. It is a general recovery algorithm that can be directly compared to algorithms such as to BP, OMP and SP.

In this work, we develop a modified version of the A*OMP algorithm, the AMul-A*OMP$_e$, that not only improves the reconstruction rates significantly over [10], but also effectively shortens the run time of the search. First, we define a novel dynamical cost model, namely the adaptive-multiplicative (AMul) cost model, that allows faster search without sacrificing the accuracy via relaxing the cost model parameter. As a result of relaxing the cost model parameter, the search is able to find the solution by opening fewer nodes and terminates faster. Second, we employ a termination criterion based on the residue of the observed vector, instead of the length-based termination criterion of A*OMP in [10]. Such a termination criterion is sometimes applied for OMP-type algorithms, however it appears for the first time in the concept of A*OMP. Moreover, most works in literature does not make a clear distinction of the termination criteria. However, this choice significantly affects the performance of OMP-type algorithms. In fact, residue-based termination is more appropriate for OMP-type CS recovery, as it is actually more accorded with the Restricted Isometry Property (RIP) [3, 15]. We discuss this issue in Section 4 for A*OMP, while our conclusions hold also for OMP. This is also justified via the recovery experiments, where the residue-based termination criterion improves the recovery for both OMP and A*OMP. The simulations involving sparse signals with different nonzero coefficient distributions in noiseless and noisy observation scenarios demonstrate the improved reconstruction capability of the AMul-A*OMP$_e$ in shorter run times.

Section 2 provides a brief summary of the A*OMP approach, referring the reader to [10] for details. In Section 3, the novel AMul cost model is introduced. We discuss the residue-based termination criterion in Section 4. Finally, Section 5 demonstrates the empirical recovery performance of AMul-A*OMP$_e$ in a range of simulations. We conclude with a short summary in Section 6.

2. A*OMP IN SHORT

A*OMP is an iterative approach that operates on a best-first search tree. The search tree is represented by the set $\mathbf{S} = \{\mathbf{s}_i\}$, where $\mathbf{s}_i$ are the paths (branches of the tree). Nodes of the tree represent dictionary atoms, and each $\mathbf{s}_i$ is a collection of atoms, which represent a candidate support for $\mathbf{x}$. The corresponding coefficients, $\mathbf{c}_i$, are obtained by the orthogonal projection of $\mathbf{y}$ onto the support $\mathbf{s}_i$. The approximation of $\mathbf{y}$ by the path $\mathbf{s}_i$ is given by $\hat{\mathbf{y}}_i = \mathbf{s}_i\mathbf{c}_i$, where the notation $\mathbf{s}_i$ is abused to represent the matrix consisting of the atoms included in the particular path i. The residue of $\mathbf{s}_i$ is defined as $\mathbf{r}_i = \mathbf{y} - \hat{\mathbf{y}}_i$. Each $\mathbf{s}_i$ is also assigned a cost $F(\mathbf{s}_i)$, which is computed using $\|\mathbf{r}_i\|_2$.

Let's now summarize A*OMP as proposed in [10]: A*OMP is initialized with I candidate paths, each consisting of a single node. These I nodes are selected as the dictionary atoms which have the highest inner-product with $\mathbf{y}$. At each iteration, the algorithm first selects the best path $\mathbf{s}_b$ among $\mathbf{S}$ with the minimum cost criterion. After $\mathbf{s}_b$ selected, A*OMP finds the B dictionary atoms that have the highest inner-product with $\mathbf{r}_b$, and creates B candidate paths by expanding $\mathbf{s}_b$ with each of these individually. Each candidate path is added to the tree if it has not been considered in previous iterations. For each new $\mathbf{s}_i$, $\mathbf{r}_i$ is computed by orthogonal projection of $\mathbf{y}$ onto $\mathbf{s}_i$. Finally, the search tree is pruned such that only the best P paths (with minimum cost) remain at the end of the iteration. Following [10], the search is terminated when the best path $\mathbf{S}_b$ has the desired length K. We refer to this version as A*OMP$_K$.

3. A NOVEL ADAPTIVE-MULTIPLICATIVE COST FUNCTION FOR A*OMP

The cost function of A*OMP plays a major role in selection of the best path at each iteration. In best-first search, selection of the best path requires comparison of paths with different lengths, which necessitates an auxiliary cost function [12, 13]. For this purpose, [10] introduces three different structures, additive, adaptive and multiplicative (Mul) models. Here, we introduce another novel dynamic cost model, which is based on the multiplicative one.

In [10], the Mul cost model is defined as

$$F_{\text{Mul}}(\mathbf{S}_i^l) = \alpha_{\text{Mul}}^{K-l} \left\|\mathbf{r}_i^l\right\|_2 \tag{2}$$

where the superscripts in $\mathbf{S}_i^l$ and $\mathbf{r}_i^l$ denote the path length and α_{Mul} is a constant in $(0, 1]$. According to this model, each unexplored node is assumed to decrease the $\left\|\mathbf{r}_i^l\right\|_2$ by a constant rate α_{Mul}, hence total degradation is modeled by the term $\alpha_{\text{Mul}}^{K-l}$.

In [10], we have observed that the adaptive cost model performs better recovery than its nonadaptive counterpart, the adaptive cost model. Inspired by this observation, we define a novel adaptive extension of the multiplicative cost model, in which the decrement in $\left\|\mathbf{r}_i^l\right\|_2$ is governed by the decrement obtained via the addition of the latest node to $\mathbf{S}_i^l$:

$$F_{\text{AMul}}(\mathbf{s}_i^l) = (\alpha_{\text{AMul}} \frac{\left\|\mathbf{r}_i^l\right\|_2}{\left\|\mathbf{r}_i^{l-1}\right\|_2})^{K-l} \left\|\mathbf{r}_i^l\right\|_2 , \tag{3}$$

where $\mathbf{r}_i^{l-1}$ is the residue after the first $l-1$ nodes in path $\mathbf{S}_i^l$ and α_{AMul} is a constant in $(0, 1]$ as above. We refer to this model as adaptive-multiplicative (AMul), and the corresponding version of A*OMP$_e$ as AMul-A*OMP$_e$.

In the AMul cost model, each unexplored node is assumed to decrease $\left\|\mathbf{r}_i^l\right\|_2$ by the rate $\alpha_{\text{AMul}}\left\|\mathbf{r}_i^l\right\|_2/\left\|\mathbf{r}_i^{l-1}\right\|_2$. This is justified by the following discussion: In general, we expect the search to select nodes in a descending order with respect to their inner-products with $\mathbf{y}$ (or similarly with respect to the corresponding nonzero values of $\mathbf{x}$). Hence, a node is expected to yield less degradation in $\left\|\mathbf{r}_i^l\right\|_2$ than its ancestors do. The term $\|\mathbf{r}_i^l\|_2/\|\mathbf{r}_i^{l-1}\|_2$ represents this degradation, and α_{AMul} is a relaxation factor less than 1. Accumulation of this effect over all missing nodes finally leads to the power $K-l$. Similar to other cost models given in [10], this assumption is also valid on average, i.e. any particular sequence of nodes may violate this, however, we expect it to hold in general and lead the search to the correct solution.

The simulations in Section 5 indicate that AMul cost model makes it possible to select α closer to 1 than Mul cost model allows. Increasing α reduces the effect of unopened nodes on the cost function and makes the search favor longer paths. As a result, the search opens fewer nodes and terminates faster. The simulation results in Section 5 demonstrate the efficiency of AMul cost model. These indicate that the AMul model not only provides better reconstruction than BP, SP and OMP but also terminates faster than the Mul cost model.

4. RESIDUE-BASED TERMINATION CRITERION FOR A*OMP

As defined in [10], A*OMP$_K$ returns the first $\mathbf{s}_b$ with length K. An alternative is terminating the search when $\mathbf{s}_b$ satisfies $\|\mathbf{r_b}\|_2 \leq \varepsilon\|\mathbf{y}\|_2$, i.e. the residue falls below a threshold. We refer to this variant A*OMP$_e$. ε is selected with respect to the noise level in a noisy problem, or very small in a noiseless scenario. A*OMP$_e$ is free to select more than K nodes on a path, up to a practical limit $K_{\text{max}} > K$. (That is, even if an $\mathbf{s}_b$ with K_{max} nodes does not satisfy the termination criterion, the search is terminated.) K_{max} might be selected with respect to the number of observations M. In this work, we set it as $M/2$, while, in practice, this bound is only reached when the reconstruction fails.

A*OMP$_e$ has some solid advantages over A*OMP$_K$: First, no a priori knowledge of K is necessary. This is beneficial when we either do not know K, or K is dynamically varying. Second, in case an $\mathbf{s}_b$ with length K does not satisfy the criterion $\|\mathbf{r}_b\|_2 > \varepsilon\|\mathbf{y}\|_2$, A*OMP$_e$ still continues searching for another, possibly longer, path that fulfills the this criterion. In other words, in cases A*OMP$_K$ fails, i.e. it returns a path of length K with some nonzero residue, A*OMP$_e$ still has the chance to find the correct path, as the search is not stopped at the first suboptimal point. This, as empirically shown in Section 5, improves the recovery accuracy.

That A*OMP$_e$ can, and indeed in practice frequently does, return a path longer than the actual sparsity level K may at first sound as if the recovery fails, however this is not true. This can be simply understood by a simple consequence of the RIP. Assume that we deal with noise-free observations, and set $\varepsilon = 0$ accordingly. If the search can find a path, say $\mathbf{s}_b$, with vanishing residue, i.e. $\|\mathbf{r_b}\|_2 = 0$, and the dictionary $\mathbf{\Phi}$ satisfies the $(K{+}K_{\text{max}})$-RIP, then $\mathbf{s}_b$ is ensured to be the correct solution. (Remember that the actual sparsity level is K, and the length of $\mathbf{s}_b$ is upper bounded at K_{max}, hence $(K{+}K_{\text{max}})$-RIP is sufficient.) That is, $\mathbf{s}_b$ consists of T, the correct support of $\mathbf{x}$ and some additional atoms, $\mathbf{s}_b \backslash T$. Following $(K{+}K_{\text{max}})$-RIP, the orthogonal projection of $\mathbf{y}$ onto the subspace $\mathbf{s}_b$ correctly identifies the nonzero entries in the correct support T, while setting other entries corresponding to the set $\mathbf{s}_b \backslash T$ to 0. Hence, that the returned path is longer than K does not indicate a recovery failure. In contrast, as the residue-based termination is the actual one in accordance with the RIP, it is more optimal than the length-based termination criterion.

Another advantage over A*OMP$_K$ follows from the following corollary: A*OMP$_e$ can cope with some additional misidentified nodes as these do not harm the recovery when the correct support is a subset of the final solution and a certain RIP is satisfied. That is, A*OMP$_e$ can correct for misidentified nodes in later stages.As we can afford addition of some misidentified components to the solution, we can relax the auxiliary function parameter α, decreasing the competence of shorter paths. This reduces the computation times without sacrificing reconstruction accuracy (see Section 5).

5. EXPERIMENTAL RESULTS

In this section, the recovery performance of A*OMP$_e$ is demonstrated in comparison to A*OMP$_K$, BP, SP and OMP in noiseless and noisy scenarios. We utilize the AMul and Mul cost models, comparing them in terms of both reconstruction performance and run times. As for OMP, we also use two versions: OMP$_K$ terminates after K steps, while

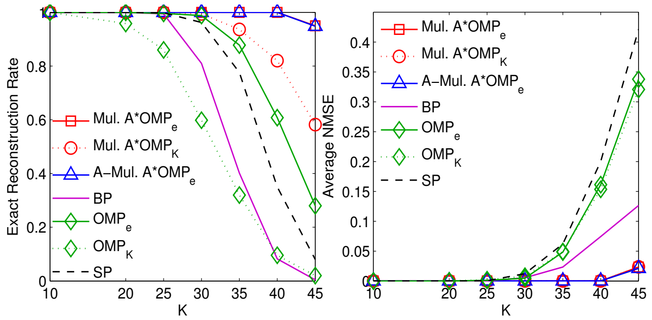

Fig. 1. Recovery results over sparsity for normally distributed non-zero coefficients.

OMP$_e$ runs until $\|\mathbf{r}\|_2 \leq \varepsilon\|\mathbf{y}\|_2$. For all tests, A*OMP parameters are selected as $I = 3$, $B = 2$, $P = 200$. ε is set to $9{\times}10^{-7}$ in the noiseless case, while it is selected with respect to the noise level in noisy scenarios. K_{max} is selected as either 50 or $K + 10$, whichever is greater. Each test is repeated over 500 randomly generated samples of length $N = 256$. Nonzero entries are selected as standard normally distributed, uniformly distributed in $[-1, 1]$ or equal to one (binary). $M = 100$ observations were taken from each vector, drawing an individual $\mathbf{\Phi}$ from normal distribution with mean 0 and standard deviation $1/N$. The recovery results are given in terms of the Average Normalized Mean-Squared-Error (ANMSE) and exact reconstruction rates. ANMSE is defined as

$$ANMSE = \frac{1}{500}\sum_{i=1}^{500}\frac{\|\mathbf{x}_i - \hat{\mathbf{x}}_i\|_2^2}{\|\mathbf{x}_i\|_2^2} \quad (4)$$

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

As for the cost model parameter α, we select different values for A*OMP variants. $\alpha_{\text{Mul}} = 0.8$ is selected for Mul-A*OMP$_K$. For Mul-A*OMP$_e$, this is relaxed as $\alpha_{\text{Mul}} = 0.9$. Using AMul-A*OMP$_e$, a further relaxation to $\alpha_{\text{AMul}} = 0.97$ is used. As mentioned in Section 3, when α is increased, less nodes are opened and the search terminates faster. Using the modified termination criterion and AMul cost model, α can be selected very close to 1 without sacrificing the recovery accuracy, as demonstrated below.

The simulations for A*OMP were performed using the AStarOMP software developed by the authors. The AStarOMP software incorporates a trie structure to implement the A* search tree in an efficient way. The orthogonalization over the residue is solved using the QR factorization. This software, and its MATLAB version, are available at http://myweb.sabanciuniv.edu/karahanoglu/research/.

Figure 1 and 2 depict the recovery performance for sparse signals with non-zero entries from standard normal distribution and uniform distribution in $[-1, 1]$, respectively. In both cases, all A*OMP versions provide significantly better recon-

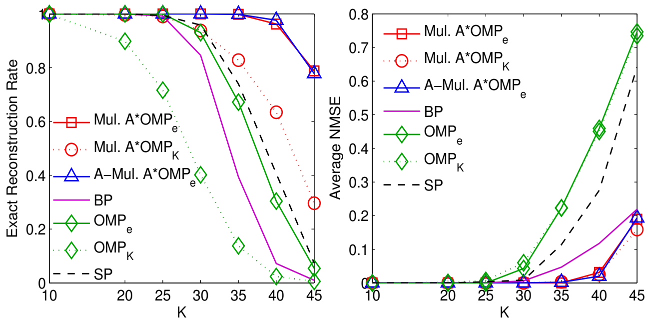

Fig. 2. Recovery results over sparsity for uniformly distributed non-zero coefficients.

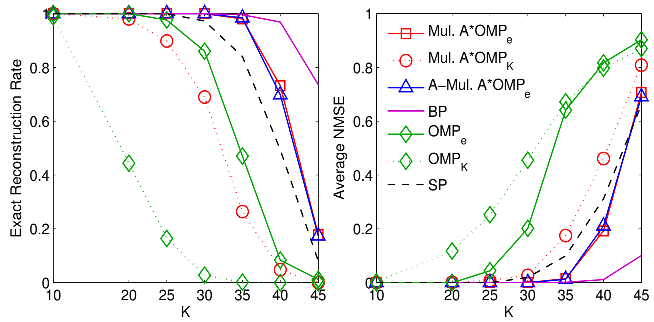

Fig. 4. Recovery results over sparsity for binary non-zero coefficients. Note that both graphs share the same legend.

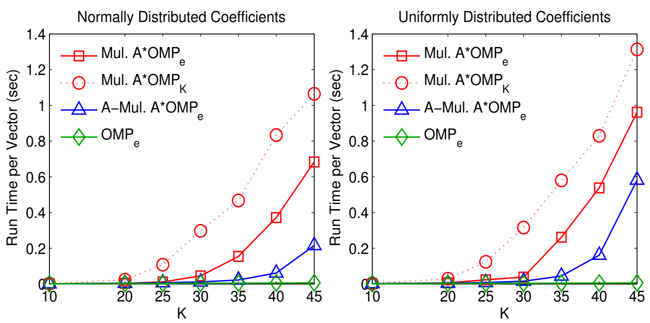

Fig. 3. Average run-time of A*OMP per vector using the AStarOMP software.

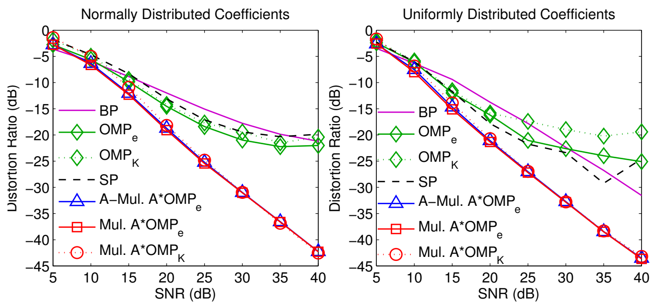

Fig. 5. Average distortion over SNR in the noisy recovery scenario. K is 30 and 25 for normally and uniformly distributed entries, respectively.

struction results than BP, SP and OMP. We also observe that Mul-A*OMP$_K$, Mul-A*OMP$_e$ and AMul-A*OMP$_e$ all yield very close ANMSE. However, the difference between them is clearly visible in the exact recovery rates: The modified termination criterion of A*OMP$_e$ improves exact recovery significantly. We may conclude that, for this case A*OMP$_e$ is better at recovering coefficients with smaller magnitudes, which do not change the reconstruction error significantly, however improve exact recovery. This fact is also visible in the results for OMP: Though OMP$_e$ and OMP$_K$ yield similar ANMSE, the exact recovery rate of OMP$_e$ is significantly better than OMP$_K$ and even better than SP.

According to Figures 1 and 2, there is no significant difference between reconstruction performances Mul-A*OMP$_e$ and AMul-A*OMP$_e$. The difference is, however, clear in Figure 3, which depicts the average run time per vector for A*OMP variants and OMP$_e$ in the two simulations mentioned above. OMP$_e$ is naturally the fastest. It is clear that both error-based termination and the novel AMul cost model decrease the run times of the A* search due to using a larger α value. As a result, AMul-A*OMP is significantly faster than other A*OMP variants. Note that we do not compare the run times for BP and SP, as we run these in MATLAB while A*OMP and OMP run times are obtained using the stand-alone AStarOMP software.

Sparse binary signals are an interesting test case for our purposes, as this case is known to be particularly challenging for OMP-type algorithms. Recovery results for sparse binary signals are depicted in Figure 4. These show that the error-based termination criterion improves the A*OMP recovery. The performance of SP is significantly better than A*OMP$_K$, however, A*OMP$_e$ outperforms SP. BP is clearly better than other algorithms in this example. Note that, for this particular case, ℓ_0 norm of the correct solution is exactly equal to its ℓ_1 norm, which might be considered as an advantage for BP.

Figure 5 illustrates the recovery performance where the observation vectors are contaminated with white gaussian noise with different SNR levels. Here, K is selected as 30 and 25 for normally and uniformly distributed non-zero entries, respectively. The distortion ratio refers to the ANMSE in the decibel (dB) scale. We observe that A*OMP$_e$ is superior to others except for 5dB SNR where BP is slightly better. It results in improved accuracy over A*OMP$_K$ for lower SNR values. We compare the average A*OMP run times for this scenario in Figure 6. As for the other cases above, we observe that both error-based termination and AMul cost model decrease the run times efficiently due to using a larger

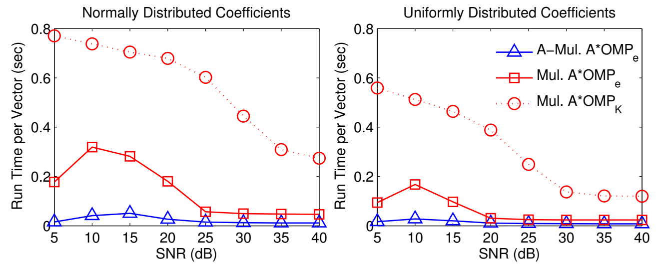

Fig. 6. Average run-time of A*OMP per vector in noisy recovery using the AStarOMP software.

α value. Consequently, AMul-A*OMP$_e$ is significantly faster than other A*OMP variants.

6. SUMMARY

A*OMP algorithm utilizes best-first search for the compressed sensing signal recovery problem. It incorporates effective cost models and pruning techniques to make the tree search practically possible. In this work, we introduced a novel A*OMP variant, AMul-A*OMP$_e$, which employs the residue-based termination criterion and the AMul cost model, which extends the Mul model in an adaptive manner. We have discussed that the residue-based termination criterion is capable of providing better recovery. Moreover, this holds not only for A*OMP but also for OMP, an important conclusion which is mostly underestimated in CS literature. This termination criterion is also reduces A*OMP search times. A further reduction of the search times is possible via the AMul cost model as it allows a higher choice of the cost model parameter α. The proposed modifications have been evaluated experimentally in noiseless and noisy scenarios including sparse signals with different non-zero coefficient distributions in terms of both the recovery results and the run times. These simulations have shown that AMul-A*OMP$_e$ promises not only better recovery rates, but also faster termination of the search.

7. REFERENCES

[1] J. A. Tropp and S. J. Wright, "Computational methods for sparse solution of linear inverse problems," *Proc. IEEE*, vol. 98, no. 6, pp. 948–958, June 2010.

[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. Candès and T. Tao, "Decoding by linear programming," *IEEE Trans. Inf. Theory*, vol. 51, no. 12, pp. 4203–4215, Dec. 2005.

[4] E. Candes and J. Romberg, "Practical signal recovery from random projections," in *Proc. SPIE Computational Imaging*, 2005, vol. 5674, pp. 76–86.

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

[6] D. L. Donoho and M. Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via ℓ_1 minimization," in *Proc. Natl. Acad. Sci.*, Los Alamitos, CA, Mar. 2003, vol. 100, p. 21972202.

[7] Y. C. Pati, R. Rezaiifar, 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.

[8] 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.

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

[10] 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.

[11] N. B. Karahanoglu and H. Erdogan, "Compressed sensing signal recovery via A* orthogonal matching pursuit," in *Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on*, May 2011, pp. 3732 –3735.

[12] 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.

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

[14] C. La and M.N. Do, "Tree-based orthogonal matching pursuit algorithm for signal reconstruction," in *Image Processing, 2006 IEEE International Conference on*, Oct. 2006, pp. 1277 –1280.

[15] 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.