Abstract—A reliable support detection is essential for a greedy algorithm to reconstruct a sparse signal accurately from compressed and noisy measurements. This paper proposes a novel support detection method for greedy algorithms, which is referred to as maximum a posteriori (MAP) support detection. Unlike the existing support detection method that identifies support indices with the largest correlation value in magnitude per iteration, the proposed method selects them with the largest likelihood ratios computed under the true and null support hypotheses by simultaneously exploiting the distributions of sensing matrix, sparse signal, and noise. Leveraging this technique, MAP-Orthogonal Matching Pursuit (MAP-OMP) is first presented to show the vantages of exploiting the proposed support detection method, and the sufficient condition for the perfect signal recovery is derived when the sparse signal is binary. Subsequently, a set of iterative greedy algorithms, called MAP-generalized Orthogonal Matching Pursuit (MAP-gOMP), MAP-Compressive Sampling Matching Pursuit (MAP-CoSaMP), and MAP-Subspace Pursuit (MAP-SP) are presented to demonstrate how the proposed support detection method can be applied into the existing greedy algorithms. From the empirical results, it is shown that the proposed greedy algorithms with highly reliable support detection can be better, faster, and easier to implement than basis pursuit via linear programing.

I. INTRODUCTION

Compressive sensing (CS) [1], [2] is a technique to reconstruct sparse signals from compressed measurements. CS has received great attention due to its broad application areas including imaging, radar, and communication systems. The fundamental theory of CS guarantees to recover a high dimensional signal vector from linear measurements that are far fewer in number than the signal’s dimension, provided that the sparsity of the signal (i.e., number of nonzero elements) is smaller than a certain fraction of the number of measurements.

Denoting the sparse signal vector and the compressed sensing matrix as $x \in \mathbb{R}^N$ and $\Phi \in \mathbb{R}^{M \times N}$, respectively, with $M < N$, the optimal sparse recovery solution can theoretically be obtained by solving the $\ell_0$-minimization problem

$$\min \|x\|_0 \text{ subject to } y = \Phi x . \tag{1}$$

In practice, unfortunately, solving this problem is NP-hard and computationally infeasible once the signal dimensions increase.

There has been extensive work on designing computationally efficient sparse signal recovery algorithms. Basis Pursuit (BP) [3]–[5] is a representative sparse signal recovery algorithm via convex optimization. Relaxing the $\ell_0$-minimization problem to a $\ell_1$-minimization problem, it has been shown that the sparse signal recovery problem can be solved with stability and uniform guarantees using linear programming at the expense of the runtimes that are polynomially bounded in computation complexity.

Approaches based on greedy algorithms are also popular because their algorithm runtimes are faster than that of BP even if stability and guarantees are challenging to be proven in general. The underlying idea of greedy algorithms is to estimate the nonzero elements of a sparse vector successively. Orthogonal matching pursuit (OMP) is a well-known such algorithm [6]–[10]. The key idea of OMP is to estimate one coordinate of the non-zero element in signal $x$ that has the maximum absolute correlation between the column vector in the sensing matrix and the residual vector per iteration. By subtracting the contribution from the measurement vector $y$, the algorithm updates the entire support of $x$ in an iterative fashion. Although this algorithm is simple to implement, it is vulnerable to error propagation effect. This is because the OMP algorithm is not capable of removing incorrectly estimated supports once those are added to the support set during the iterations, which leads to significant degradation in the error performance.

Several other advanced greedy algorithms have been proposed to overcome the error propagation effect, which include Stagewise Orthogonal Matching Pursuit (StOMP) [11], iterative hard thresholding (IHT) [12], generalized OMP (gOMP) [13], Compressive Sampling Matching Pursuit (CoSaMP) [14], and Subspace Pursuit (SP) [15]. The underlying principle of these advanced greedy algorithms is the selection of multiple support indices per iteration, leading to a decrease in the probability of estimating incorrect support elements. For example, in each iteration, StOMP [11] identifies multiple support indices such that the correlation value in magnitude between the current residual vector and the corresponding column vector of $\Phi$ exceeds a predefined threshold. Similarly, gOMP [13] chooses multiple supports that provide $L$ largest correlation in magnitude per iteration, where $L$ is a fixed parameter given in the algorithm. CoSaMP [14] and SP [15] also identify multiple support indices per iteration but differ from StOMP and gOMP in that they perform a two-stage sparse signal estimation approach that allows to add or remove new support candidates adaptively. The common problem of all prior greedy algorithms in [6], [7], [11], [14], [15] is that...
they rely on the order statistics of the correlation value in magnitude for the support estimation.

Depending on statistical distributions of sensing matrix, sparse signal, and noise, however, the selection of the index with the largest correlation value may not be optimal in the sense of support detection probability. With this motivation, greedy algorithms called Bayesian matching pursuit were proposed in [19]–[23]. The key idea of Bayesian matching pursuit is the use of distributions of the sparse signal and noise in the support detection step. For example, fast Bayesian matching pursuit (FBMP) [20] performs sparse signal estimation via model selection, assuming a Gaussian distribution for the sparse vector. Similarly, in [22], [23] assuming the elements of a sparse signal are Bernoulli-Gaussian mixed variables, and a given deterministic sensing matrix, the algorithms jointly update a support index and the corresponding signal element at each iteration in order to maximize the increase of a local likelihood function. Although these approaches show a better sparse recovery performance compared to conventional matching pursuit algorithms in the presence of noise, they are limited to use in certain distributions of $x$ like Bernoulli-Gaussian, and there are no provable performance guarantees.

In this paper, we continue the same spirit of harnessing the statistical distributions of sparse signal, sensing matrix, and noise for the support detection in greedy algorithms. Our main contribution is to propose a novel support detection method for greedy algorithms, which is referred to as maximum a posteriori (MAP) support detection. The key difference with prior work in [19]–[23] is that the proposed method estimates supports with the largest log MAP-ratio values computed under the true and null support hypotheses in each iteration by incorporating the distributions of the sensing matrix, the sparse signal, and noise jointly. Specifically, assuming the sensing matrix has elements that are drawn from independent and identically distributed (IID) Gaussian random variables, and the sparse signal has non-zero elements that are distributed according to a continuous distribution, i.e., $N(0, \frac{1}{M})$. Then, the measurement equation is given by

$$y = \Phi x + w$$

where $y \in \mathbb{R}^M$ and $w \in \mathbb{R}^M$ are the measurement and noise vector, respectively. All entries of the noise vector are assumed to be IID Gaussian random variables with zero mean and variance $\sigma_w^2$, $\mathcal{N}(0, \sigma_w^2)$. Throughout this paper, the difference between two sets $T$ and $S$ is denoted by $T \setminus S$. We use the subscript notations $x_{|S}$ and $\Phi_{|S}$ to denote that vector $x$ and matrix $\Phi$ are being restricted to only elements or columns in set $S$.

### III. MAP-OMP

In this section, we first illustrate a novel greedy algorithm called MAP-OMP considering a binary sparse signal $x \in \{0, 1\}^N$. Then, we derive a bound that provides a sufficient condition for perfect signal recovery to demonstrate provable performance guarantees of the proposed algorithm.

#### A. Algorithm

Similar to the OMP algorithm [6], MAP-OMP is a greedy algorithm that sequentially finds support indices and estimates the signal representation within a certain number of iterations. The core difference between the proposed MAP-OMP algorithm and the prior OMP algorithms lies in the selection rule of the support index per iteration. Unlike the OMP algorithms, MAP-OMP chooses the index based on a maximum likelihood hypothesis test by leveraging statistical property of the sensing matrix and the sparse signal.

We begin by providing Lemmas that are required for explaining the MAP-OMP algorithm. Lemma 1 provides the distribution of the inner product between two (atom) dictionary vectors generated by IID Gaussian random variable. Lemma 2 yields the distribution of the 2-norm of each dictionary vector. From the empirical results, it is shown that the proposed modified algorithms provide significant gains in the perfect recovery performance compared to that of the existing greedy algorithms as well as the $\ell_1$-minimization algorithm via BP.

### II. Problem Statement

We consider a sparse signal detection problem from compressed and noisy measurement. Let us denote a $N$ dimensional input signal vector by $x \in \mathbb{R}^N$. We assume that the input vector is $K$-sparse, i.e., $\|x\|_0 = K \ll N$ and the sparsity level $K$ is known a priori. This prior information can be estimated accurately in some applications. We denote the true support set by $T \subset \{1, \ldots, N\}$ and $|T| = K$. The non-zero entries of $x$ are distributed according to a continuous distribution, i.e., $p(x) = \prod_{k \in T} p_k(x_k)$. Furthermore, we denote the sensing matrix consisting of $N$ column vectors by $\Phi \in \mathbb{R}^{M \times N}$,

$$\Phi = [a_1, a_2, \ldots, a_N]$$

where $a_n$ denotes the $n$-th dictionary vector whose entries are drawn from an IID Gaussian random distribution with zero mean and variance $\frac{1}{M}$, i.e., $\mathcal{N}(0, \frac{1}{M})$. Then, the measurement equation is given by

$$y = \Phi x + w$$

where $y \in \mathbb{R}^M$ and $w \in \mathbb{R}^M$ are the measurement and noise vector, respectively. All entries of the noise vector are assumed to be IID Gaussian random variables with zero mean and variance $\sigma_w^2$, $\mathcal{N}(0, \sigma_w^2)$. Throughout this paper, the difference between two sets $T$ and $S$ is denoted by $T \setminus S$. We use the subscript notations $x_{|S}$ and $\Phi_{|S}$ to denote that vector $x$ and matrix $\Phi$ are being restricted to only elements or columns in set $S$. 

#### A. Algorithm

Similar to the OMP algorithm [6], MAP-OMP is a greedy algorithm that sequentially finds support indices and estimates the signal representation within a certain number of iterations. The core difference between the proposed MAP-OMP algorithm and the prior OMP algorithms lies in the selection rule of the support index per iteration. Unlike the OMP algorithms, MAP-OMP chooses the index based on a maximum likelihood hypothesis test by leveraging statistical property of the sensing matrix and the sparse signal.

We begin by providing Lemmas that are required for explaining the MAP-OMP algorithm. Lemma 1 provides the distribution of the inner product between two (atom) dictionary vectors generated by IID Gaussian random variable. Lemma 2 yields the distribution of the 2-norm of each dictionary vector.
Lemma 1. Suppose that all the elements of $a_n$ for $n \in [1 : N]$ are drawn from IID Gaussian distribution with zero mean and variance $\frac{1}{M}$. Then, the distribution of $\frac{a_n^T a_n}{\|a_n\|^2}$ is Gaussian with zero mean and variance $\frac{1}{2M}$, i.e., $\frac{a_n^T a_n}{\|a_n\|^2} \sim \mathcal{N}(0, \frac{1}{2M})$.

Proof: See Appendix A

Lemma 2. The distribution of the norm $\|a_n\|$ is

$$f_{\|a_n\|}(x) = \frac{2^{1 - \frac{M}{2}} M^{\frac{M}{2}} x^{M-1} e^{-\frac{M}{2} x^2}}{\Gamma \left(\frac{M}{2}\right)}.$$  

(5)

Proof: See Appendix B

Lemma 3. The norm $\|a_n\|$ of each dictionary vector for $n \in [1 : N]$ concentrates to one asymptotically as $M$ goes to infinity,

$$\lim_{M \to \infty} \mathbb{P} \left( |\|a_n\| - 1| \geq \epsilon \right) = 0$$  

(6)

for some positive $\epsilon > 0$.

Proof: See Appendix C

By leveraging these Lemmas, we explain the proposed algorithm. In the $k$-th iteration, the algorithm produces $N$ correlation values $\{z_{1k}, z_{2k}, \ldots, z_{Nk}\}$ by computing the inner product between the residual vector $r^{k-1}$ updated in the $(k-1)$-th iteration and the $n$-th column vector $a_n$, i.e., $z_{nk} = a_n^T r^{k-1}$ for $n \in [1 : N]$. Under the premise that the algorithm has found the elements of the true support and their corresponding signal estimates perfectly in all previous iterations, the residual vector is

$$r^{k-1} = \sum_{\ell \in T \setminus S^{k-1}} a_{\ell} x_{\ell}$$  

(7)

where $S^{k-1} \subseteq T$ and $|S^{k-1}| = k - 1$. Then, the inner product value $z_{nk} = a_n^T r^{k-1}$ can be expressed as a linear combination of the remaining non-zero elements and their corresponding support as follows:

$$z_{nk} = \|a_n\|^2 x_n + \sum_{\ell \in T \setminus \{S^{k-1}, \{a_n\}\}} \frac{a_n^T a_{\ell} x_{\ell}}{\|a_n\|^2} + \frac{a_n^T w}{\|a_n\|^2}.$$  

(8)

Using (8), the proposed MAP-OMP algorithm performs the hypothesis test with two hypotheses corresponding to $x_n = 0$ and $x_n = 1$, respectively, as follows:

$$\mathcal{H}_0 : z_{nk} = \sum_{\ell \in T \setminus \{S^{k-1}\}} \frac{a_n^T a_{\ell}}{\|a_n\|^2} x_{\ell} + \frac{a_n^T w}{\|a_n\|^2}$$  

(9)

$$\mathcal{H}_1 : z_{nk} = \|a_n\|^2 x_n + \sum_{\ell \in T \setminus \{S^{k-1}, \{a_n\}\}} \frac{a_n^T a_{\ell}}{\|a_n\|^2} x_{\ell} + \frac{a_n^T w}{\|a_n\|^2},$$  

(10)

where $\mathcal{H}_0$ is the null hypothesis such that the $n$-th column vector $a_n$ is not the support, i.e., $x_n = 0$ ($n \notin T$), and $\mathcal{H}_1$ is the alternate hypothesis indicating that the $n$-th column vector is a non-zero support and the corresponding signal value is 1, i.e., $x_n = 1$ ($n \in T$). These two hypotheses in (9) and (10) involve multiple levels of randomness, namely,

1. The randomness associated with the inner product between two distinct vectors $\frac{a_n^T a_{\ell}}{\|a_n\|^2}$ (unit norm) and $a_{\ell}$; this is distributed as a Gaussian random variable, i.e., $\frac{a_n^T a_{\ell}}{\|a_n\|^2} \sim \mathcal{N}(0, \frac{1}{2M})$ for $\ell \neq n$ as shown in Lemma 1 (See Appendix).

2. The randomness associated with the effective noise $a_n^T w$; this is Gaussian with zero mean and variance $\sigma_w^2$, i.e., $\frac{a_n^T w}{\|a_n\|^2} \sim \mathcal{N}(0, \sigma_w^2)$, as $w$ is isotropically distributed in $\mathbb{R}^M$.

3. The randomness associated with the sum of independent Gaussian random variables, $z_{nk} = \sum_{\ell \in T \setminus \{S^{k-1}, \{a_n\}\}} \frac{a_n^T a_{\ell}}{\|a_n\|^2} + \frac{a_n^T w}{\|a_n\|^2}$; this is also Gaussian with zero mean and variance $\mathbb{E} \left[ z_{nk}^2 \right] = \frac{K-(k-1)}{M} + \sigma_w^2$ as $\frac{a_n^T w}{\|a_n\|^2}$ and $\frac{a_n^T a_{\ell}}{\|a_n\|^2}$ are mutually independent Gaussian random variables for $\ell \neq j$.

4. The randomness associated with the norm of the column vector $\|a_n\|$; this is a scaled Chi-distribution with $M$ degrees of freedom, i.e., $f_{\|a_n\|}(x) = 2^{1 - \frac{M}{2}} M^{\frac{M}{2}} x^{M-1} e^{-\frac{M}{2} x^2} / \Gamma \left(\frac{M}{2}\right)$ as shown in Lemma 2.

Using these facts, the conditional distribution of $z_{nk}$ under the null hypothesis is given by

$$\mathbb{P}(z_{nk} | x_n = 0) = \frac{1}{\sigma_0 \sqrt{2\pi}} \exp \left( -\frac{|z_{nk}|^2}{2\sigma_0^2} \right)$$  

(11)

where $\sigma_0 = \sqrt{\frac{K-(k-1)}{M} + \sigma_w^2}$. Similarly, under the hypothesis of $x_n = 1$ and $\|a_n\|_2 = u$, the conditional distribution of $z_{nk}$ is Gaussian with mean $u$ and variance $\frac{K-(k-1)+1}{M} + \sigma_w^2$, i.e.,

$$\mathbb{P}(z_{nk} | x_n = 1, \|a_n\|_2 = u) = \frac{1}{\sigma_1 \sqrt{2\pi}} \exp \left( -\frac{|z_{nk} - u|^2}{2\sigma_1^2} \right),$$  

(12)

where $\sigma_1 = \sqrt{\frac{K-(k-1)+1}{M} + \sigma_w^2}$. From Lemma 2, by marginalizing the conditional distribution in (12) with respect to $u$, we obtain the conditional distribution under the hypothesis of $x_n = 1$ as

$$\mathbb{P}(z_{nk} | x_n = 1) = \int_0^\infty \mathbb{P}(z_{nk} | x_n = 1, \|a_n\|_2 = u) f_{\|a_n\|_2}(u) du = \int_0^\infty e^{-\frac{|z_{nk} - u|^2}{2\sigma_1^2}} \frac{1}{\sigma_1 \sqrt{2\pi}} 2^{1 - \frac{M}{2}} M^{\frac{M}{2}} u^{M-1} e^{-\frac{M}{2} u^2} \frac{1}{\Gamma \left(\frac{M}{2}\right)} du$$  

(13)

This conditional distribution is intractable to analyze due to the integral expression. From Lemma 3, however, we know that $\|a_n\|_2$ approaches one almost surely as $M \to \infty$. Using this fact, the conditional distribution under the hypothesis of $x_n = 1$ simplifies to

$$\mathbb{P}(z_{nk} | x_n = 1) \approx \frac{1}{\sigma_1 \sqrt{2\pi}} \exp \left( -\frac{|z_{nk} - 1|^2}{2\sigma_1^2} \right).$$  

(14)
Leveraging the conditional probability density functions in (11) and (14), the a posteriori probabilities (MAP) ratio for a given observation $z_n^k$ is

$$\Lambda(z_n^k) = \ln \left( \frac{\mathbb{P}(n \in T | z_n^k)}{\mathbb{P}(n \notin T | z_n^k)} \right)$$

$$= \ln \left( \frac{1}{\sigma_0 \sqrt{2\pi}} \exp \left( -\frac{|z_n^k|^2}{2\sigma_0^2} \right) \right) + \ln \left( \frac{\mathbb{P}(n \in T)}{\mathbb{P}(n \notin T)} \right)$$

where (a) follows from the Bayes’ rule and (b) comes from the assumption that the $K$ non-zero supports are uniformly distributed from 1 to $N$. This log likelihood ratio carries reliability information about how the $n$-th column vector in the sensing matrix is likely to belong to the true support in the $k$-th iteration. Accordingly, at iteration $k = 1, \ldots, K - 1$, the proposed MAP-OMP algorithm selects index $J^k$ that maximizes $\Lambda(z_n^k)$, namely,

$$J^k = \arg \max_{n \in [1:N]} \Lambda(z_n^k)$$

$$= \arg \max_{n \in [1:N]} \frac{(z_n^k)^2}{2N} - \frac{(z_n^k - 1)^2}{2\sigma_w^2}$$

Once index $J^k$ is selected, MAP-OMP estimates the new sparse representation $\hat{x}^k$ using the updated support set $S^k = S^{k-1} \cup \{J^k\}$. Since the signal is assumed to be a binary, the new sparse representation is set to be one, namely,

$$\hat{x}^k_{S^k} = 1.$$ (17)

Lastly, to remove the contribution of $\hat{x}^k_{S^k}$, we update the new residual signal such that

$$r^k = r^{k-1} - A_{S^k} \hat{x}^k.$$ (18)

To obtain more insight on the proposed support detection method, it is instructive to consider certain special cases.

**Noise-Free Case:** Let us consider the case of noise-free compressive sensing, i.e., $\sigma_w^2 = 0$. The log MAP-ratio boils down to

$$\Lambda(z_n^k) = \frac{M(z_n^k)^2}{2(K - k + 1)} - \frac{M(z_n^k - 1)^2}{2(K - k)} + \frac{1}{2} \ln \left( \frac{K - k + 1}{K - k} \right) + \ln \left( \frac{K}{N - K} \right).$$ (19)

This expression clearly shows that the MAP-ratio in the $k$-th iteration is a function of the relevant system parameters—the dimension of the measurement vector $M$ and the sparsity level $K$. One key property of the proposed algorithm is that it updates the log MAP-ratio adaptively, since the variances of the conditional probability density functions decrease under the premise that the algorithm successively estimates the signal at each iteration. For the noise-free case, in the last iteration $k = K$, we slightly need to modify the computation of the ratio, as $\mathbb{P}(z_n^k | n \in T) = 1$. Accordingly, the modified ratio in the last iteration for the noise-free case is given by

$$\Lambda(z_n^K) = M(z_n^K)^2 + \ln \left( \frac{K}{N - K} \right).$$ (20)

**High Noise Power Case:** Let us consider the high noise power scenario, i.e., $\sigma_w^2 \gg \frac{K}{M}$. In this case, the MAP-ratio in (15) is approximated as

$$\Lambda(z_n^K) \approx \frac{(z_n^K)^2}{2\sigma_w^2} - \frac{(z_n^K - 1)^2}{2\sigma_w^2} = \frac{2z_n^K - 1}{2\sigma_w^2}.$$ (21)

From this, we are able to observe that the selection of the largest index of the MAP-ratio is equivalent to the selection of the largest index of the correlation value $z_n^K$ in the high noise power regime, namely,

$$\arg \max_n \Lambda(z_n^K) = \arg \max_n z_n^K.$$ (22)

Therefore, the conventional OMP algorithms that select the largest correlation value $z_n^K$ is the optimal in the sense of the MAP detection strategy for the high noise power regime. For the cases of low noise power and noise-free, however, the selection of the largest absolute value of $z_n^K$ for the support detection is not optimal. This fact clearly exhibits the benefits of the proposed MAP-OMP against the conventional OMP algorithm in [6].

**B. Asymptotic Analysis for Exact Recovery**

In this section, we derive a lower bound of the required measurements for the exact support recovery when the proposed MAP-OMP is applied for the binary sparse signal. Unlike the prior analysis approaches that rely on the Restricted Isometry Property (RIP) [4], [7], [8] or an information theoretical analysis tool in [18], we directly compute a lower bound of the success probability that the proposed algorithm identifies the $K$-sparse binary signal within $K$ number of iterations. Utilizing this, a lower bound of the required measurements is derived to reconstruct the signal perfectly as the signal dimension approaches infinity. The following theorem shows the main analysis result.

**Theorem 1.** Let $x \in \{0,1\}^N$ be a sparse binary signal vector with sparsity level $K \ll N$, and the noise variance be $\sigma_w^2 = \frac{\epsilon}{M}$. Then, the proposed MAP-OMP algorithm recovers the sparse vector perfectly with noisy measurements within $K$ number of iterations, provided that the number of measurements scales as

$$M = \mathcal{O}\left( (K + \alpha) \ln(N) \right),$$ (23)

for some $\epsilon > 0$ as $N$ goes to infinity.

**Proof:** Without loss of generality, we assume that the first $K$ columns are the true supports, i.e., $x_n = 1$ for $n \in [1:K]$, i.e., $T = \{1,2,\ldots,K\}$ and the remaining $N - K$ columns
are the zero supports. Furthermore, we denote \( E_s^k \) to be the success recovery probability event in the \( k \)-th iteration. Then, the success recovery probability of the \( K \) sparse signal within \( K \) number of iterations is given by
\[
P_s = P \left( \bigcap_{k=1}^{K} E_s^k \right) = P(E_s^1)P(E_s^2|E_s^1) \times \cdots \times P(E_s^K|E_s^{K-1}, \ldots, E_s^1), \tag{24}\]
where the equality comes from the probability chain rule. To prove that \( P_s \) approaches one asymptotically as \( N \to \infty \), it suffices to check that the algorithm correctly identifies the column of the true support in the \( k \)-th iteration conditioned that all the prior iterations recover the true supports successfully, i.e., \( P(E_s^k|E_s^{k-1}, \ldots, E_s^1) = 1 - o \left( \frac{1}{K} \right) \) as \( N \to \infty \) for any \( k \in [1:K] \).

To detect the support correctly in the \( k \)-th iteration of the proposed algorithm, the maximum of \( \Lambda(z_k^\ell) \) for \( \ell \in T \setminus S^k \) should be larger than the maximum of \( \Lambda(z_k^n) \) for \( n \in T^c = \{K + 1, \ldots, N\} \), which is
\[
P(E_s^k|E_s^{k-1}, \ldots, E_s^1) \geq P \left[ \max_{\ell \in T \setminus S^k} \Lambda(z_k^\ell) \geq \max_{n \in T^c} \Lambda(z_k^n) \right]
= \prod_{n=1}^{N-K} P \left[ \Lambda(z_k^n) \leq \Lambda(z_k^\ell) \right]
= (1 - P \left[ \Lambda(z_k^\ell) < \Lambda(z_k^n) \right])^{N-K}, \tag{25}\]
where the first equality follows from the fact that \( \{ \Lambda(z_{K+1}^\ell), \ldots, \Lambda(z_N^\ell) \} \) are mutually independent and \( \{z_{K+1}^\ell, \ldots, z_N^\ell\} \) are IID Gaussian random variables with zero mean and variance \( \sigma_0^2 \). To this end, we need to compute the probability that \( \Lambda(z_k^\ell) \) is less than \( \Lambda(z_k^n) \) as follows:
\[
P \left[ \Lambda(z_k^n) > \Lambda(z_k^\ell) \right] = P \left[ \frac{(z_k^n)^2 - (z_k^\ell - 1)^2}{2\sigma_0^2} < \frac{(z_k^\ell)^2 - (z_k^n - 1)^2}{2\sigma_0^2} \right]
= \frac{-\lambda}{2\sigma_0^2} \left( \frac{(z_k^n)^2 - (z_k^\ell - 1)^2}{2\sigma_0^2} \right) \left( \frac{(z_k^\ell)^2 - (z_k^n - 1)^2}{2\sigma_0^2} \right)
\leq \min_{\Lambda \geq 0} \left[ \frac{\lambda}{2\sigma_0^2} \left( \frac{(z_k^n)^2 - (z_k^\ell - 1)^2}{2\sigma_0^2} \right) \right] e^{-\frac{-\lambda}{2\sigma_0^2} \left( \frac{(z_k^\ell)^2 - (z_k^n - 1)^2}{2\sigma_0^2} \right)} \tag{26}\]
where the last inequality follows from Markov’s inequality and the independence of \( z_k^\ell \) and \( z_k^n \). Since \( z_k^n \) given \( x_N = 0 \) is distributed as in (11), the first term in (26) is calculated as
\[
\mathbb{E} \left[ e^{-\frac{-\lambda}{2\sigma_0^2} \left( \frac{(z_k^n)^2 - (z_k^\ell - 1)^2}{2\sigma_0^2} \right)} \right] = \int_{-\infty}^{\infty} e^{-\frac{-t^2}{2\sigma_0^2}} \lambda \left( \frac{t^2}{2\sigma_0^2} - \frac{(t-1)^2}{2\sigma_0^2} \right) \sqrt{2\pi} \sigma_0 dt
= \frac{e^{2\lambda} \left( \frac{1}{\sigma_0} \right)}{\sigma_0 \sqrt{\frac{1}{\sigma_0^2} + \frac{\lambda}{\sigma_1}}}.
\tag{27}\]
Similarly, using the distribution of \( z_k^\ell \) given \( x_\ell = 1 \) in (12), the second term in (26) is computed as
\[
\mathbb{E} \left[ e^{-\frac{-\lambda}{2\sigma_1^2} \left( \frac{(z_k^\ell)^2 - (z_k^n - 1)^2}{2\sigma_1^2} \right)} \right] = \int_{-\infty}^{\infty} e^{-\frac{-t^2}{2\sigma_1^2}} \frac{-\lambda}{2\sigma_1^2} \left( \frac{t^2}{2\sigma_1^2} - \frac{(t-1)^2}{2\sigma_1^2} \right) \sqrt{2\pi} \sigma_1 dt
= \frac{e^{2\lambda} \left( \frac{1}{\sigma_1} \right)}{\sigma_1 \sqrt{\frac{1}{\sigma_1^2} + \frac{\lambda}{\sigma_1}}}.
\tag{28}\]
Plugging \( \lambda = \frac{1}{2} > 0 \), the probability that the MAP-ratio under the zero support is greater than that under the non-zero support is upper bounded by
\[
P \left[ \Lambda(z_k^n) > \Lambda(z_k^\ell) \right] \leq \frac{e^{2\lambda} \left( \frac{1}{\sigma_0^2} + \frac{1}{\sigma_1^2} \right)}{\frac{1}{2} \left( \frac{1}{\sigma_0^2} + \frac{1}{\sigma_1^2} \right)}.
\tag{29}\]
Since \( \sigma_0^2 = K-k+1+\alpha \) and \( \sigma_1^2 = K-k+1+\frac{M}{2\alpha} \) in the \( k \)-th iteration, this error upper bound is further simplified as
\[
P \left[ \Lambda(z_k^n) > \Lambda(z_k^\ell) \right] \leq \frac{e^{2\lambda} \left( \frac{1}{\sigma_0^2} + \frac{1}{\sigma_1^2} \right)}{\frac{1}{2} \left( \frac{1}{\sigma_0^2} + \frac{1}{\sigma_1^2} \right)} \leq \frac{e^{2\lambda} \left( \frac{1}{\sigma_0^2} + \frac{1}{\sigma_1^2} \right)}{\frac{1}{2} \left( \frac{1}{\sigma_0^2} + \frac{1}{\sigma_1^2} \right)}.
\tag{30}\]
Plugging (30) into (25), we have a lower bound as follows:
\[
P(E_s^k|E_s^{k-1}, \ldots, E_s^1) \geq \left( 1 - e^{2\lambda} \left( \frac{1}{\sigma_0^2} + \frac{1}{\sigma_1^2} \right) \right)^{N-K} \tag{31}\]
From (31), we observe that the success probability in the first iteration is lower than that of any other remaining iterations, i.e., \( P(E_s^k) \leq P(E_s^k|E_s^{k-1}, \ldots, E_s^1) \) for \( \forall k \). It follows that the lower bound of the exact recovery probability is
\[
P_s = P(E_s^1)P(E_s^2|E_s^1) \times \cdots \times P(E_s^K|E_s^{K-1}, \ldots, E_s^1)
\geq \left( 1 - e^{2\lambda} \left( \frac{1}{\sigma_0^2} + \frac{1}{\sigma_1^2} \right) \right)^{K(N-K)}.
\tag{32}\]
Since we have assumed that \( M = (1+\epsilon)2(2K - 1 + 2\alpha) \ln(K(N-K)) \), the lower bound is rewritten as
\[
\ln(P_s) \geq K(N-K) \ln \left( \frac{1}{\ln(K(N-K))^{1+\epsilon}} \right).
\]
As \( N \) goes to infinity, we have
\[
\lim_{N \to \infty} \ln(P_s) \geq \lim_{N \to \infty} K(N-K) \ln \left( \frac{1}{\ln(K(N-K))^{1+\epsilon}} \right)
= \lim_{N \to \infty} \frac{-(1+\epsilon) \ln(K(N-K))}{1-(1+\epsilon)} = 0,
\tag{33}\]
where the second equality follows from L'Hospital’s rule. Consequently, we conclude that \( \lim_{N \to \infty} P_s = 1 \). From the facts that \( N > M > 2K \) (the condition for a unique sparse solution) and \( \ln(K(N-K)) = \ln(N-K)+\ln(K) \leq 2 \ln(N-K) \), it is possible that the \( K \) sparse binary signal is perfectly recovered within \( K \) number of iterations, if the number of measurements scales as, at least, \( M \geq (1+\epsilon)2(4K - 2 + 4\alpha) \ln(N-K) \) for some \( \epsilon > 0 \). Therefore, the scaling law of the required number of measurements becomes \( M = \mathcal{O} \left( (K + \alpha) \ln(N) \right) \), which completes the proof.
Theorem 1 shows the statistical guarantee of the proposed MAP-OMP algorithm for the binary signal. The guarantee is that the proposed MAP-OMP algorithm recovers the $K$-sparse binary signal perfectly with $K$ number of iterations, if the number of (noisy) measurements scales as $\mathcal{O}(k(1+\alpha)\ln(N))$. This measurement scaling law clearly exhibits that the required measurements should increase with the sparsity level $K$ and the normalized noise variance $\alpha$. This result backs the intuition that the measurements should increase with $K$. Meanwhile, the requirement measurements increase with $N$ logarithmically. This condition extends the existing statistical guarantee for OMP proven in [6] by incorporating noise effects.

### C. Generalization to Non-Binary Signal

So far, we have assumed a binary sparse signal. In some applications, however, the element in the zero-support can be an arbitrary value drawn from a continuous probability distribution $f_x(u)$. In this subsection, we present a modified MAP-OMP algorithm for the sparse signal whose non-zero element is distributed according to a distribution $f_x(u)$.

Recall that the crux idea of MAP-OMP algorithm is the identification of the support element by performing hypothesis testing using the correlation value $z_n^k$ under the null hypothesis $H_0: z_n^k = \sum_{\ell \in T} \frac{{\mathbf{a}}^T_n {\mathbf{a}}_\ell}{\|\mathbf{a}_\ell\|_2^2} x_\ell + \frac{{\mathbf{a}}^T_n \mathbf{w}}{\|\mathbf{a}_n\|_2}$ as

$$H_0: z_n^k = \sum_{\ell \in T, |S^{k-1}\cup\{n\}|} \frac{{\mathbf{a}}^T_n {\mathbf{a}}_\ell}{\|\mathbf{a}_\ell\|_2^2} x_\ell + \frac{{\mathbf{a}}^T_n \mathbf{w}}{\|\mathbf{a}_n\|_2},$$

$$H_1: z_n^k = \|\mathbf{a}_n\|_2 x_n + \sum_{\ell \in T, |S^{k-1}\cup\{n\}|} \frac{{\mathbf{a}}^T_n {\mathbf{a}}_\ell}{\|\mathbf{a}_\ell\|_2^2} x_\ell + \frac{{\mathbf{a}}^T_n \mathbf{w}}{\|\mathbf{a}_n\|_2},$$

where $x_n$ is distributed as $f_x(u)$. The exact characterization of the distribution for $z_n^k$ under the null hypothesis is challenging as it highly depends on the signal distribution $f_x(u)$. To facilitate simplified calculations, the distribution of $z_n^k$ is approximated using Gaussian distribution with the first and second order moments matching. From Lemma 1, recall that $\frac{{\mathbf{a}}^T_n {\mathbf{a}}_\ell}{\|\mathbf{a}_\ell\|_2^2}$ and $\frac{{\mathbf{a}}^T_n \mathbf{w}}{\|\mathbf{a}_n\|_2}$ are distributed as $\mathcal{N}(0, \frac{K-k-1}{M})$ and $\mathcal{N}(0, \sigma_n^2)$. Furthermore, since $\mathbb{E}[x_\ell] = \mu$ and $\mathbb{E}[x_\ell^2] = \sigma_n^2$ for $\ell \in T$, the first and second moments of $z_n^k$ are

$$\mathbb{E} \left[ z_n^k \mid x_n = 0 \right] = \sum_{\ell \in T, |S^{k-1}\cup\{n\}|} \mathbb{E} \left[ \frac{{\mathbf{a}}^T_n {\mathbf{a}}_\ell}{\|\mathbf{a}_\ell\|_2^2} x_\ell \right] + \mathbb{E} \left[ \frac{{\mathbf{a}}^T_n \mathbf{w}}{\|\mathbf{a}_n\|_2} \right] = 0 \quad (35)$$

and

$$\mathbb{E} \left[ (z_n^k)^2 \mid x_n = 0 \right] = \sum_{\ell \in T, |S^{k-1}\cup\{n\}|} \mathbb{E} \left[ \left( \frac{{\mathbf{a}}^T_n {\mathbf{a}}_\ell}{\|\mathbf{a}_\ell\|_2^2} \right)^2 x_\ell^2 \right] + \mathbb{E} \left[ \left( \frac{{\mathbf{a}}^T_n \mathbf{w}}{\|\mathbf{a}_n\|_2} \right)^2 \right] = \left( K - k + 1 \right) \sigma_n^2 + \sigma_n^2.$$

(36)

Accordingly, the approximated distribution of $z_n^k$ is given by

$$\mathbb{P}(z_n^k \mid x_n = 0) \approx \frac{1}{\sigma_n \sqrt{2\pi}} \exp \left( -\frac{(z_n^k)^2}{2\sigma_n^2} \right), \quad (37)$$

where $\sigma_n = \sqrt{(K-k+1)\sigma_n^2 + \sigma_n^2}$. Similarly, conditioning the hypothesis of $x_n = u$, the approximated distribution of $z_n^k$ is given by

$$\mathbb{P}(z_n^k \mid x_n = u) \approx \frac{1}{\sigma_n \sqrt{2\pi}} \exp \left( -\frac{|z_n^k - u|^2}{2\sigma_n^2} \right), \quad (38)$$

where $\sigma_1 = \sqrt{(K-k+1)\sigma_n^2 + \sigma_n^2}$. Utilizing the approximated distributions in (36) and (38), the log MAP-ratio is obtained by marginalizing with respect to the distribution $f_x(u)$, namely,

$$\Lambda(z_n^k) \approx \ln \left( \frac{\int_{-\infty}^{\infty} \mathbb{P}(z_n^k \mid x_n = u) f_x(u) du}{\mathbb{P}(z_n^k \mid x_n = 0)} \right) + \ln \left( \frac{K}{N-K} \right). \quad (39)$$

Therefore, the proposed MAP support detection for the non-binary signal is to select the support index such that

$$\arg \max_{n \in [1:N]} \Lambda(z_n^k) \approx \arg \max_{n \in [1:N]} \ln \left( \frac{\int_{-\infty}^{\infty} \mathbb{P}(z_n^k \mid x_n = u) f_x(u) du}{\mathbb{P}(z_n^k \mid x_n = 0)} \right). \quad (40)$$

To provide a more transparent interpretation of the expression in (40), we consider the following two cases of interest.

**Example 1** (Uniformly Distributed Signal): One basic case is the scenario where the elements of the transmit signal are drawn from a uniform distribution between 0 and 1, i.e., $f_x(u) = 1$ for $0 \leq u \leq 1$. Then, the MAP-ratio expression in (40) becomes

$$\Lambda_L(z_n^k) \approx \ln \left( \frac{\frac{\sigma_1 \sqrt{\pi}}{2} \text{Erf} \left( \frac{|z_n^k - u|}{\sigma_1} \right) + \text{Erf} \left( \frac{z_n^k}{\sigma_1} \right)}{\frac{1}{\sqrt{2\pi}\sigma_0} \exp \left( -\frac{(z_n^k)^2}{2\sigma_0^2} \right)} \right), \quad (41)$$

where $\text{Erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt$.

**Example 2** (Sparse Signal with Finite Alphabet): Another popular case of interest is one where the non-zero entry of $x$ is uniformly selected from the elements of a finite set of alphabet $C = \{c_1, \ldots, c_L\}$ as considered in [24, 25]. For example, each pixel of a bitmap image file is capable of storing 8 different colors when the 3-bit per pixel (8bpp) format is used. In this application, the finite set can be given as $C = \{0, 1, \ldots, 7\}$. In this case, the log-MAP is computed as follows:

$$\Lambda_C(z_n^k) \approx \ln \left( \frac{\sum_{\ell=1}^{L} \mathbb{P}(z_n^k \mid x_n = c_\ell) \mathbb{P}(x_n = c_\ell)}{\mathbb{P}(z_n^k \mid x_n = 0)} \right)$$

$$= \ln \left( \frac{1}{L} \sum_{\ell=1}^{L} \frac{1}{\sqrt{2\pi}\sigma_0} \exp \left( -\frac{(z_n^k - c_\ell)^2}{2\sigma_0^2} \right) \right) \quad (42)$$

It is worth noting that when $L = 1$, this MAP-ratio approximation in (42) recovers the exact MAP-ratio for the binary signal case given in (15).

Using the approximated log-MAP ratio, we provide a modified MAP-OMP algorithm as in Table I. The key difference with the MAP-OMP algorithm for the binary signal is that the MAP-ratio is computed differently depending on the sparse signal distribution. Furthermore, the algorithm estimates the sparse signal using a least square solution in each iteration similar to the conventional OMP algorithm.
TABLE I
MAP-OMP ALGORITHM (NON-BINARY SIGNAL)

1) Initialization:
\[ k := 0, z^0 = 0 \]
\[ r^0 := y \text{ (the current residual)} \]
\[ S^0 := \{ \} \]

2) Repeat until a stopping criterion is met
i) \[ k := k + 1 \]
ii) Compute the current proxy:
\[ \hat{x}^k = \sum_{\Omega} \left( z_k \right) w_n \]
\[ z_k = \left( \frac{\sum_{\Omega} w_n}{\sum_{\Omega} w_n} \right) \frac{z_k}{\sum_{\Omega} w_n} \]
\[ \text{for } n \in \{ 1 : |N| \} \]
iii) Select the largest index of MAP-ratio:
\[ \Lambda^k := \text{arg max}_{\Omega} \left\{ \Lambda^k \left( z_k \right) \right\} \text{ for } d \in \{ U, C \} \]
iv) Merge the support set:
\[ S^k := S^{k-1} \cup J^k \]
v) Update sparse signal:
\[ \hat{x}^k := \text{arg min}_{\Omega} \| \Phi \hat{x}^k \hat{x} - y \|_2 \]
w) Update the residual for next round:
\[ r^k := r^{k-1} - \Phi \hat{x}^k \hat{x} \]

B. MAP-CoSaMP

CoSaMP is an effective iterative sparse signal recovery algorithm [14]. It was shown to yield the same sparse signal recovery performance guarantees as $l_1$-norm minimization even with less computational complexity. The main idea of CoSaMP is that, in the first step, it estimates a large support set with $L$ largest correlation values in magnitude and obtains a least square solution based on it, where $L$ is typically chosen between $K \leq L \leq 2K$. In the next step, the algorithm reduces the cardinality of the support set back to the desired sparsity level of $K$ using pruning, and acquires a sparse solution again based on the reduced support.

We modify this algorithm by incorporating the proposed support detection technique. Unlike the conventional CoSaMP algorithm, MAP-CoSaMP adds $2K$ support candidates with $2K$ largest MAP-ratio values to the support set $S^k$ per iteration. Once the least square solution is obtained based on the corresponding support set $\hat{x}_{S^k} = \Phi_{S^k} \hat{x}$, an approximation to the signal is updated by selecting the $K$ largest coordinates using pruning. Finally, the residual is updated using the approximated signal estimate. The algorithm is described in Table III.

C. MAP-SP Algorithm

SP is a two-step iterative algorithm for sparse recovery [15]. Similar to CoSaMP, the SP algorithm identifies the current estimate of support set by greedily adding multiple indices with the largest correlation in magnitude. The main difference between CoSaMP and SP lies in the second step. While CoSaMP applies a pruning technique using the estimated sparse signal in the first stage to maintain the required sparsity level without performing the second least-square estimation. Whereas, the SP algorithm updates the sparse solution by solving a least square problem based on the reduced support in the second stage.

Applying the proposed MAP support detection method, we modify this algorithm by changing the support set identification stage. The proposed MAP-SP algorithm selects $2K$ support indices with the largest MAP-ratio values in each iteration. The modified algorithm is summarized in Table IV.

IV. EXTENSION TO THE OTHER GREEDY ALGORITHMS

One vantage of the proposed MAP support detection method is, in fact, directly applicable to many other greedy sparse signal recovery algorithms that exploit the proposed support detection method.

A. MAP-gOMP

gOMP [13] is a simple yet effective algorithm that improves the performance of OMP. The key idea of gOMP is the selection of multiple support indices with the largest correlation in magnitude at each iteration; thereby, it reduces the misdetection probability compared to that of OMP. Similar to the gOMP algorithm, MAP-gOMP is a greedy algorithm that sequentially finds multiple support indices and estimates the signal representation within the certain number of iterations. The core difference lies in the selection rule of the support indices per iteration. Unlike the gOMP algorithm, MAP-gOMP chooses $L$ support indices with the largest log-MAP ratio values instead of the largest correlation in magnitude. The proposed MAP-gOMP is summarized in Table II.

TABLE III
MAP-CoSaMP ALGORITHM

1) Initialization:
\[ k := 0, z^0 = 0 \]
\[ r^0 := y \text{ (the current residual)} \]
\[ S^0 := \{ \} \]
\[ \Omega^0 := \{ \} \]

2) Repeat until a stopping criterion is met
i) Compute the current proxy:
\[ \hat{x}^k = \text{arg min}_{\Omega} \{ \Lambda^k \} \text{ for } d \in \{ U, C \} \]
ii) Select the $K$ largest indices of MAP-ratio:
\[ \Lambda^k := \text{arg max}_{\Omega} \{ \Lambda^k \} \text{ for } d \in \{ U, C \} \]
iii) Merge the support set:
\[ S^k := S^{k-1} \cup \Omega^k \]
iv) Perform a least-squares signal estimation:
\[ \hat{x}_{S^k} := \text{arg min}_{\Omega} \| \Phi_{S^k} \hat{x} - y \|_2, \hat{x}_{S^k} = 0 \]
v) Prune $\Lambda^k$ and update $r$ for next round:
\[ \Omega := \text{arg max}_{\Omega} \{ |\Lambda^k| \} \text{ for } \{ |\Lambda^k| \} \]
\[ r = y - \Phi_{S^k} \hat{x}_{S^k} \]

TABLE II
MAP-gOMP ALGORITHM

1) Initialization:
\[ k := 0, z^0 = 0 \]
\[ r^0 := y \text{ (the current residual)} \]
\[ S^0 := \{ \} \]

2) Repeat until a stopping criterion is met
i) \[ k := k + 1 \]
ii) Compute the current proxy:
\[ z_k = \sum_{\Omega} \left( \frac{\sum_{\Omega} w_n}{\sum_{\Omega} w_n} \right) \frac{z_k}{\sum_{\Omega} w_n} \]
\[ \text{for } n \in \{ 1 : |N| \} \]
iii) Select the $L$ largest indices of MAP-ratio:
\[ \Lambda^k := \text{arg max}_{\Omega} \{ \Lambda^k \} \text{ for } d \in \{ U, C \} \]
iv) Merge the support set:
\[ S^k := S^{k-1} \cup \Lambda^k \]
v) Perform a Least-Squares Signal Estimation:
\[ \hat{x}^k := \text{arg min}_{\Omega} \| \Phi \hat{x}^k \hat{x} - y \|_2 \]
w) Update the residual for next round:
\[ r^k := r^{k-1} - \Phi \hat{x}^k \hat{x} \]
Fig. 1. Performance comparison of perfect reconstruction probability for the binary signal with noise-free measurements.

Fig. 2. Performance comparison of perfect reconstruction probability for the signal whose non-zero element is uniformly distributed between 0 and 1, i.e., $x_i \sim \text{Uni}[0, 1]$ with noise-free measurements.

V. Numerical Results

We provide empirical recovery performance of the proposed algorithms by means of simulations. We evaluate the empirical frequency (cumulative density function) of exact reconstruction for the proposed algorithms in both noise and noiseless cases and compare them with the conventional algorithms. In our simulation, we generate $M \times N$ ($M = 128$ and $N = 256$) sensing matrix whose elements are drawn independently from Gaussian distribution $\mathcal{N}(0, \frac{1}{2})$. Furthermore, we consider $K$-sparse vector $x$ whose support is uniformly distributed. Each non-zero element of $x$ is one for the binary signal and is randomly selected from $[0, 1]$ for the uniform signal. To obtain the empirical frequency of exact reconstruction, we perform 1,000 independent trials for each algorithm. For each trial, we perform iterations until the stopping criterion $\|x - \hat{x}\|_2^2 \leq 10^{-12}$ is satisfied except for gOMP and MAP-gOMP. For gOMP and MAP-gOMP, we perform $\min(K, \left\lfloor\frac{M}{2}\right\rfloor)$ number of iterations in each trial, where $L = 2$. To obtain the performance of BP, we use the CVX tool provided in MATLAB [20].

Table IV provides the empirical recovery performance of the proposed algorithms by means of simulations, where $M=128$ and $N=256$. The proposed MAP-ratio function in (41) for the simulations. Similar to the binary signal case, it is no wonder that the proposed MAP-gOMP, MAP-CoSaMP, and MAP-SP algorithms outperform than the existing sparse recovery algorithms, as the proposed support detection method reduces the mis-detection probability considerably. In particular, MAP-gOMP and MAP-SP are able to recover the signal with more than 95 % probability up to a sparsity level of 60, which is close to the maximum sparsity level ($\left\lfloor\frac{M}{2}\right\rfloor = 64$) that can be recovered with a unique solution guarantee.

We consider now a sparse image recovery example. As illustrated in Fig. 3 (the left-top figure), a binary sparse image with $37 \times 37$-pixel size is considered for the experiment. Applying linear random projection matrix $\Phi \in \mathbb{R}^{685 \times 1369}$ whose elements are drawn from $\mathcal{N}(0, \frac{1}{1369})$, we compress the binary image. As shown in Fig. 3 when the noise-free measurements are used for image reconstruction, we observe that the proposed MAP-gOMP and MAP-SP algorithms outperform than the other existing algorithms, which agrees with

\begin{table}[h]
\centering
\caption{MAP-SP Algorithm}
\begin{tabular}{l}
\hline
1) Initialization: \\
\quad $k := 0$, $S^0 := \emptyset$ \\
\quad $v^0 := y$ (the current residual) \\
\quad $S^0 := \{\emptyset\}$ and $\Omega^0 := \{\emptyset\}$ \\
2) Repeat until a stopping criterion is met \\
\quad i) Compute the current proxy: \\
\quad \quad $z_k^d = \frac{v_k^d}{\|v_k^d\|}$ for $n \in [1 : N]$. \\
\quad ii) Select the $K$ largest indices of MAP-ratio: \\
\quad \quad $\Omega^k =: \arg \max_{\Omega} \frac{\|z_k^d\|}{\|x\|}$ for $d \in \{U, C\}$. \\
\quad iii) Merge the support set: \\
\quad \quad $S^k = S^{k-1} \cup \Omega^k$. \\
\quad iv) Perform a Least-Squares Signal Estimation: \\
\quad \quad $b^k := \arg \min_b \|\Phi_b y - y\|_2$ \\
\quad v) Select the $K$ largest index in $S^k$: \\
\quad \quad $G := \arg \max_{G} |\{b^k\}|$ \\
\quad vi) Perform a LS using the updated $G$: \\
\quad \quad $\hat{S}^k := \arg \min_{\hat{S}} \|\Phi_{\hat{S}} x - y\|_2$. \\
\hline
\end{tabular}
\end{table}
the result shown in Fig. [1]. To demonstrate the effect of noisy measurement, we add Gaussian noise with zero mean and variance $\sigma^2_a = 0.005$. In this case, as depicted in Fig. [4], the proposed MAP-SP method is able to recover the image almost perfectly even in the presence of noise. Whereas, the image reconstruction performance of the proposed MAP-gOMP algorithm is degraded compared to the case of noise-free, which exhibits the noise sensitivity of the algorithm.

As can be seen in Table [V], the proposed algorithms achieve significant speedup compared to the existing algorithms in both the noise-free and noisy measurements cases. These speedup gains are mainly due to the fact that the proposed algorithms identify the true support set with less number of iterations, leading to the faster convergence rates than those of the existing algorithms. In particular, the runtimes of MAP-SP ($\approx 0.21$ sec) under the noise-free measurements speed up 157 times than that of BP ($\approx 33.22$ sec).

**VI. Conclusion**

We have presented a new support detection technique based on a MAP criterion for greedy sparse signal recovery algorithms. Using this method, we have proposed a set of greedy sparse signal recovery algorithms and established a theoretical signal recovery guarantee for a particular case. One major implication is that the joint use of the distributions of sensing matrix, sparse signal, and noise in support identification offers a tremendous recovery performance improvement over previous support detection approaches that ignore such statistical information. Our numerical results demonstrate that the greedy algorithms with highly reliable support detection provide significantly better sparse recovery performance than the linear programming approach.

An interesting direction for future study would be to explore the statistical guarantees of the proposed MAP-gOMP, MAP-CoMSaMP, and MAP-SP. Another possible research direction is to investigate the greedy algorithms when different statistical distributions of the sensing matrix are used.

---

**TABLE V**

| Algorithms          | Runtimes (Sec) | Speedup $\sigma^2_a = 0$ | Runtimes $\sigma^2_a = 0.005$ | Speedup $\sigma^2_a = 0.005$ |
|---------------------|----------------|---------------------------|-------------------------------|-------------------------------|
| gOMP                | 5.03           | 6.6x                       | 5.03                          | 7.1x                          |
| MAP-gOMP            | 2.26           | 14.6x                      | 4.81                          | 2.5x                          |
| SP                  | 12.97          | 2.3x                       | 14.5                          | 2.4x                          |
| MAP-SP              | 0.21           | 157.8x                     | 15.1                          | 2.4x                          |
| BP                  | 33.22          | baseline                   | 35.73                         | baseline                      |

Appendix A

**Proof of Lemma [1]**

Note that the distribution of each atom vector $a_n$ is rotationally invariant. This implies that for any unitary matrix $U \in \mathbb{R}^{M \times M}$, the distributions of $Ua_n$ and $a_n$ are identical. By selecting a unitary matrix $U$ so that $Ua_n = [1, 0, \ldots, 0]^T$, we can compute the cumulative distribution function of $\frac{a_n^T a_{\ell}}{||a_n||_2}$ as

$$
\mathbb{P} \left[ \frac{a_n^T a_{\ell}}{||a_n||_2} \leq x \right] = \mathbb{P} \left[ \frac{a_n^T U^T a_{\ell}}{||a_n||_2} \leq x \right] = \mathbb{P} \left[ a_{\ell}(1) \leq x \right] \quad (43)
$$

where $a_{\ell}(1)$ denotes the first component of $a_{\ell}$. As a result, $\frac{a_n^T a_{\ell}}{||a_n||_2}$ is IID Gaussian with zero mean and variance $\frac{1}{M}$.

Appendix B

**Proof of Lemma [2]**

Recall that all elements of $a_n$ are Gaussian random variables with zero mean and variance $\frac{1}{M}$, and they are mutually independent. Thus,

$$
\mathbb{P} \left[ ||a_n|| \leq x \right] = \mathbb{P} \left[ \sqrt{\sum_{m=1}^{M} (a_n(m))^2} \leq x \right] = \mathbb{P} \left[ \sum_{m=1}^{M} (a_n(m))^2 \leq \frac{Mx}{\frac{1}{M}} \right] \quad (44)
$$
where (a) follows from the fact that $\sqrt{\sum_{m=1}^{M} \left(\frac{\mathbf{a}_m(m)}{\sqrt{M}}\right)^2}$ is Chi-distributed with $M$ degrees of freedom, since $\frac{\mathbf{a}_m(m)^2}{\sqrt{M}}$ is a normal Gaussian with zero mean and unit variance, and

$$\gamma(s, x) = \int_0^x t^{s-1} e^{-t} dt$$ denotes the lower incomplete gamma function. By taking the derivative with respect to $x$, we obtain the distribution of $\|\mathbf{a}_n\|_2$ as

$$f_{\|\mathbf{a}_n\|_2}(x) = \frac{1}{x} \frac{\Gamma(M/2)}{\Gamma(M/2)} \frac{\Gamma(1+M/2)}{\Gamma(1+M/2)} \frac{1}{\Gamma(M/2)} x^M e^{-Mx^2/2}.$$

Accordingly, the mean of the norm is

$$\mathbb{E}[\|\mathbf{a}_n\|_2] = \int_0^\infty \frac{x^M e^{-Mx^2/2}}{\Gamma(M/2) \Gamma(1+M/2)} dx = \frac{1}{\sqrt{M}} \frac{\Gamma(1+M/2)}{\Gamma(1/2)},$$

which completes the proof.

**APPENDIX C**

**PROOF OF LEMMA 3**

We commence by computing the probability that the absolute difference between the norm and its average is greater than or equal to a small value $\epsilon$, which is

$$P[\|\mathbf{a}_n\|_2 - \mathbb{E}[\|\mathbf{a}_n\|_2] \geq \epsilon] = P\left[\sum_{m=1}^{M} \frac{(\mathbf{a}_n(m))^2}{\sqrt{M}} - \frac{1}{\sqrt{M}} \frac{\Gamma(1+M/2)}{\Gamma(1/2)} \geq \epsilon\right] \leq \frac{\sum_{m=1}^{M} (\mathbf{a}_n(m))^2}{\epsilon^2}. $$

Since $\frac{1}{\sqrt{M}} \frac{\Gamma(1+M/2)}{\Gamma(1/2)}$ converges to one as $M$ goes to infinity, we conclude that

$$\lim_{M \to \infty} P[\|\mathbf{a}_n\|_2 - \mathbb{E}[\|\mathbf{a}_n\|_2] \geq \epsilon] = 0 $$(47)

for some $\epsilon > 0$. As a result, the norm of each column vector concentrates to the average $\mathbb{E}[\|\mathbf{a}_n\|_2] = \sqrt{\frac{1}{M} \frac{\Gamma(1+M/2)}{\Gamma(1/2)}}$ and it also converges to one for a large enough $M$ because

$$\lim_{M \to \infty} \sqrt{\frac{1}{M} \frac{\Gamma(1+M/2)}{\Gamma(1/2)}} = 1.$$ This completes the proof.

**REFERENCES**

[1] E. J. Candès and T. Tao, “Decoding by linear programming,” IEEE Trans. Inf. Theory, vol. 51, no. 12, pp. 4203 - 4215, Dec. 2005.

[2] E. J. Candès, 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.

[3] E. J. Candès and J. Romberg, “Sparsity and incoherence in compressive sampling,” Inverse problems, vol. 23, pp. 969, Apr. 2007.

[4] E. J. Candès, “The restricted isometry property and its implications for compressed sensing,” Comptes Rendus Mathematique, vol. 346, no. 9-10, pp. 589-592, Feb. 2008.

[5] S. S. Chen, “Basis Pursuit,” Ph.D. dissertation, Stanford Univ., Stanford, CA. Nov. 1995.

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

[7] 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.

[8] T. T. Cai and L. Wang, “Orthogonal matching pursuit for sparse signal recovery with noise,” IEEE Trans. Inf. Theory, vol. 57, no. 7, pp. 4680-4698, July 2011.

[9] T. Zhang, “Sparse recovery with orthogonal matching pursuit under RIP,” IEEE Trans. Inf. Theory, vol. 55, no. 9, pp. 6215-6221, Sept. 2011.

[10] E. Liu and V. N. Temlyakov, “The orthogonal super greedy algorithm and applications in compressed sensing,” IEEE Trans. Inf. Theory, vol. 58, no. 4, pp. 2040-2047, Apr. 2012.

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

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

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

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

[15] W. Dai and O. Milenkovic, “SubSAMP: iterative signal recovery for incomplete and inaccurate samples,” IEEE Trans. Inf. Theory, vol. 55, no. 5, pp. 2230-2249, May 2009.

[16] S. Verdu, “Multiuser detection,” Cambridge University Press, 1998.

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

[18] M. J. Wainwright, “Information-theoretic limits on sparsity recovery in the high-dimensional and noisy setting,” IEEE Trans. Inf. Theory, vol. 55, no. 12, pp. 5728-5741, Dec. 2009.

[19] S. Ji, Y. Xue, and L. Carin, “Bayesian compressive sensing,” IEEE Trans. Signal Process., vol. 56, no. 6, pp. 2346-2356, Jun. 2008.

[20] P. Schniter, L. C. Potter, and J. Ziniel, “Fast Bayesian matching pursuit,” in Proc. of IEEE Information Theory and Applications Workshop, pp. 326-333, Jan. 2008.

[21] H. Zayyani, M. Babaie-Zadeh, and C. Jutten, “An iterative bayesian algorithm for sparse component analysis in presence of noise,” IEEE Trans. Signal Processing, vol. 57, no. 11, pp. 4378-4390, Nov. 2009.

[22] C. Herzet and A. Dremeau, “Bayesian pursuit algorithms,” in Proc. IEEE European Signal Processing Conference (EUSIPCO), pp. 1474-1478, Aug. 2010.

[23] A. Dremeau, C. Herzet, L. Daudet, “Soft Bayesian pursuit algorithm for sparse representations,” in Proc. IEEE Statistical Signal Processing Workshop (SSP), pp. 341-344, Jan. 2011.

[24] Z. Tian, G. Leus, and V. Lottici, “Detection of sparse signals under finite-alphabet constraints,” in Proc. IEEE Int. Conf. Acoust. Speech Signal Process (ICASSP), pp. 2349-2352, Mar. 2009.

[25] A. K. Das and S. Vishwanath, “On finite alphabet compressive sensing,” in Proc. IEEE Int. Conf. Acoust. Speech Signal Process (ICASSP), pp. 5890-5894, Mar. 2013.

[26] M. Grant and S. Boyd, “CVX: Matlab software for disciplined convex programming,” version 2.1, http://cvxr.com/cvx, Mar. 2014.