Approximate Bit-wise MAP Detection for Greedy Sparse Signal Recovery Algorithms

Author(s):
Chae, Jeongmin; Hong, Song-Nam

Publication Date:
2020-02-26

Permanent Link:
https://doi.org/10.3929/ethz-b-000403246

Rights / License:
In Copyright - Non-Commercial Use Permitted
Approximate Bit-wise MAP Detection
For Greedy Sparse Signal Recovery Algorithms

Jeongmin Chae and Song-Nam Hong
Ajou University, Suwon, Korea,
email: {jmchae92 and snhong}@ajou.ac.kr

Abstract—A greedy algorithm is a fascinating choice in support recovery problem due to its easy implementation and lower complexity compared with other optimization-based algorithms. In this paper, we present a novel greedy algorithm, referred to as bit-wise maximum a posteriori (MAP) detector. In the proposed method, for each iteration, one includes the best index to a target support in the sense of maximizing a posteriori probability given an observation, support indices previously chosen, and a priori information on a sparse vector. In other words, the proposed method employs statistical information on a given sparse recovery system while the other greedy-based algorithms (e.g., orthogonal matching pursuit (OMP)) uses the correlation values in magnitude. We remark that the proposed method has much lower complexity than the (vector-wise) MAP, where the complexity of the former is linear with a sparsity level but the latter is exponential. We further reduce the complexity of the proposed method by efficiently computing a posteriori probability for each iteration. Via simulations, we demonstrate that the proposed method can outperform the other greedy algorithms based on correlations, by exploiting statistical information properly.

Index Terms—Sparse vector recovery, compressed sensing, MAP detector, greedy algorithm.

I. INTRODUCTION

An inverse problem is widely studied in which a vector signal \( x \in \mathbb{R}^N \) is recovered from a set of linear noisy measurements \( y = Ax + z \), with an \( M \times N \) measurement matrix \( A \). In particular when \( M < N \) (i.e., under-determined system), the above problem has infinite solutions and thus it can be solved if some additional a priori information on \( x \) is available. In [1], [2], it has been proved that \( x \) can be exactly reconstructed with the a priori knowledge on the sparsity of \( x \) (i.e., \( \|x\|_0 = K \) with \( K \ll N \)), where \( K \) is referred to as the sparsity level. Also, the optimal sparse vector can be obtained by solving \( \ell_0 \)-minimization problem such as

\[
x^* = \arg\min_x \|x\|_0 \quad \text{s.t.} \quad \|y - Ax\|_2 \leq \eta,
\]

(1)

where \( \|x\|_0 \) is introduced to ensure the sparsity of \( x \). In general, the above \( \ell_0 \)-minimization is known to be NP-hard. Leveraging the idea of convex optimization, a well-established method, called LASSO, was proposed in which \( \ell_1 \)-norm is used as a convex-relaxation of \( \ell_0 \)-norm [3], [4]. LASSO can solve the sparse signal recovery problem with stability while it has polynomial bounded computational complexity.

A greedy approach seems to be attractive due to its lower complexity than convex-based algorithms in sparse signal reconstruction. The key idea of greedy-based algorithms is to estimate the support of a sparse signal vector in a sequential fashion, where for each iteration, one index is added to a target support by solving a sub-optimization problem. Since the sub-optimization problem has much lower complexity than the overall sparse signal recovery problem, the greedy approach can significantly reduce the computational complexity. Orthogonal matching pursuit (OMP) [5]–[7] is the most famous greedy approach where for each iteration, it identifies the best support index in the sense of correlations between column vectors in the measurement matrix and the residual vector. In addition, to overcome the downside of OMP, numerous advanced greedy algorithms have been proposed such as Compressive Sampling Matching Pursuit (CoSaMP) [9], Subspace Pursuit (SP) [10] and generalized OMP [11].

The fundamental concept of such advanced algorithms lies in the selection of multiple support indices for each iteration, which can decrease the probability for estimating incorrect support indices. In the above greedy algorithms, they only rely on the order statistics of the correlation values in magnitude to estimate support. However, it may not be optimal in the sense of support detection in probability depending on the statistical distributions of sparse signal vector and noise. Inspired by this, a greedy algorithm, named Bayesian matching pursuit (BMP), has been proposed in [12].

Our contributions: We propose a novel greedy algorithm, named bit-wise MAP detector, for sparse signal recovery problem. The key idea of the proposed algorithm is that for each iteration, one adds the best index to a target support in the sense of maximizing a posteriori probability given an observation, support indices previously chosen, and a priori information on a sparse signal vector. Namely, the proposed method needs to solve bit-wise MAP detection for each iteration, which has much lower complexity than the (vector-wise) MAP detection. This is because the complexity of the former is linear with the sparsity level while the latter is exponential. Unfortunately, the complexity to solve bit-wise MAP detection problem is still expensive since it requires the marginalization of joint probability mass function (PMF) with a large-size random vector. We address this problem by presenting an efficient way to compute a good proxy (i.e., lower-bound) of a posteriori probability. Via simulations, we demonstrate that the proposed method can outperform the other greedy algorithms based on correlations, by exploiting statistical information properly.
II. PRELIMINARIES

In this section we will provide some useful notations and state the sparse signal recovery problem.

A. Notations

We provide some notations which will be used throughout the paper. Let \([N] \triangleq \{1, ..., N\}\). We use \(\hat{X}\) and \(x\) to denote a random vector and its values, respectively. Also, for a vector \(x \in \mathbb{R}^N\), \(x_i\) denotes the \(i\)-th component of \(x\) for \(i \in [N]\). Similarly for a matrix \(B \in \mathbb{R}^{M \times N}\) the \((i, j)\)-th component of \(B\) is denoted as \(B_{i,j}\). The diagonal approximation of a square matrix \(S\) is denoted by \(\text{diag}(S)\), where \(\text{diag}(S)\) denotes the diagonal matrix whose \(i\)-th diagonal component is \(S_{i,i}\). For any positive \(K \leq N\), we let \(\Omega\) denote the set of all length-\(N\) binary vectors with the sparsity level \(K\), i.e.,

\[\Omega \triangleq \{x \in \{0, 1\}^N : \|x\|_0 = K\}.\]  

(2)

Given an index subset \(I \subseteq [N]\), we define the subset of \(\Omega\) as

\[\Omega_I \triangleq \{x \in \{0, 1\}^N : \|x\|_0 = K, x_i = 1 \text{ for } i \in I\}, \]

(3)

where \(|\Omega_I| = \binom{N-I}{K-I}\). Also, given a vector \(x \in \mathbb{R}^N\), \(S(x)\) represents its support containing the indices of non-zero locations of \(x\) such as

\[S(x) \triangleq \{i \mid x_i \neq 0, i \in [N]\}.\]  

(4)

As an extension, we also define \(S(\Omega_I) \triangleq \{S(x) \mid x \in \Omega_I\}\). Given two PMFs \(p(x)\) and \(q(x)\), the Kullback-Leibler (KL) divergence is denoted as \(\text{D}_{\text{KL}}[p]\parallel q\). For two probability distributions \(p(x)\) and \(q(x)\). Also, for \(0 \leq a \leq 1\), Bern\(\{a\}\) represents a Bernoulli distribution with \(P(X = 1) = a\).

B. Problem Formulation

We consider a \(N\)-dimensional binary sparse signal recovery problem from a noisy observation. Let \(x \in \{0, 1\}^N\) denote a \(K\)-sparse binary signal vector (i.e., \(\|x\|_0 = K\)). Then, the measurement vector \(y \in \mathbb{R}^M\) is obtained as

\[y = Ax + z,\]  

(5)

where \(A = [a_1, a_2, \ldots, a_N] \in \mathbb{R}^{M \times N}\) represents a fixed measurement matrix and \(z \in \mathbb{R}^M\) follows the zero-mean white Gaussian distribution, namely, \(Z \sim \mathcal{N}(0, \Sigma I)\). Throughout the paper, it is assumed that the sparsity level \(K\) is given as a priori information and additionally, the marginal PMFs of the sparse signal vector \(\hat{X} = (X_1, ..., X_N)^T\) (denoted by \(p_i(a)\)) are given as

\[p_i(a) \triangleq P(X_i = a) = a \quad \text{for } i \in [N] \text{ and } a \in \{0, 1\}.\]  

(6)

It is noticeable that in the case of no priori knowledge on the distribution of \(\hat{X}\), the marginal PMFs can be assigned as uniform distribution (i.e., \(p_i(1) = 0.5\) for \(i \in [N]\)).

Algorithm 1 Approximate Bit-wise MAP Detector

| Input: Measurement matrix \(A \in \mathbb{R}^{M \times N}\), noisy observation \(y \in \mathbb{R}^M\), sparsity level \(K\), and noise level \(\sigma^2\). |
| Output: Support \(\hat{Z}^{(K)} = \{i_1, ..., i_K\}\). |
| 1: Initialization \(\hat{Z}^{(0)} = \phi\) |
| 2: for \(k = 1 : K\) do |
| 3: Find the \(k\)-th support index \(i_k\) by taking the solution of |
| \[i_k = \arg \max_{i \in [N] \setminus \hat{Z}^{(k-1)}} \lambda \left(i_k | \hat{Z}^{(k-1)}\right),\]  

(7)

where the objective function is defined in (10). |
| 4: Update the support \(\hat{Z}^{(k)} = \hat{Z}^{(k-1)} \cup \{i_k\}\). |
| 5: end for |

From the above model, we will investigate the maximum a posteriori (MAP) support recovery problem, which can be mathematically formulated as

\[\hat{Z} = \arg \max_{\Omega \in \Omega} \log P(S(\hat{X}) = \Omega | y).\]  

(7)

Unfortunately, it is too complex to solve the above problem due to its combinatorial nature. Specifically, we need to check the objective function (a posteriori probability) with the \(\binom{N}{K}\) plausible candidates, which requires an exponential complexity with the sparsity level \(K\). In the following sections, we will address the above complexity problem by introducing a novel greedy approach.

III. THE PROPOSED BIT-WISE MAP SUPPORT DETECTOR

In this section, we propose a novel greedy approach to solve the support recovery problem in (7) efficiently. In the proposed method, \(K\) support indices (i.e., non-zero components of a sparse signal vector \(x\)) are derived in a sequential way via bit-wise MAP detection. Specifically, from the chain rule, the objective function (7) can be factorized as

\[\log P(S(\hat{X}) = \hat{Z}^{(K)} | y) = \sum_{k=1}^{K} \log P(\{i_k \in S(\hat{X}) | \hat{Z}^{(k-1)} \subset S(\hat{X}), y\}) + \Phi(\hat{Z}^{(k-1)}),\]  

(8)

where the above index sets are defined as \(\hat{Z}^{(K)} = \{i_1, ..., i_K\}\) and \(\hat{Z}^{(k)} = \{i_1, ..., i_k\} \subset \hat{Z}^{(K)}\) for \(k = 1, ..., K - 1\), with \(\hat{Z}^{(0)} = \phi\). In the proposed greedy approach, we find a support \(\hat{Z}^{(K)} = \{i_1, ..., i_K\}\) sequentially, by finding a local optimal solution based on the previously chosen solutions. This is mathematically formulated as

\[i_k = \arg \max_{i \in [N] \setminus \hat{Z}^{(k-1)}} \Phi(i_k | \hat{Z}^{(k-1)}).\]  

(9)

This problem is referred to as bit-wise MAP detection, which has much lower complexity than vector-wise MAP detection in (7) since the complexity of the former is linear with the sparsity level \(K\) while the that of the latter is exponential. Although the proposed greedy approach significantly reduces the computation complexity, it still suffers from the expensive complexity for computing a posteriori probability (i.e.,
Φ(ik|Z(k−1))). This is due to the marginalization of a large-scale random vector, which requires the summations of all possible K sparse vector signals x ∈ ΩZ(k).

To address the complexity problem, we will derive a good proxy (which is simply computable) of the objective function in (9), which is given as

\[
\Lambda \left( i_k | \hat{Z}^{(k-1)} \right) \triangleq \sum_{i=1}^{N} -D_{KL}(\text{Bern}(\mu_i^{(k)}), p_i) + \frac{1}{\sigma^2} \text{tr}(A^T \text{AR}^{(k)}),
\]

where

\[
\mu_i^{(k)} = \begin{cases} 
1, & i \in \hat{Z}^{(k-1)} \cup \{i_k\} \\
\lambda_k, & \text{else}
\end{cases}
\]

and

\[
R_{i,j}^{(k)} = \begin{cases} 
1, & i,j \in \hat{Z}^{(k-1)} \cup \{i_k\} \\
\lambda_k \lambda_{k+1}, & i,j \notin \hat{Z}^{(k-1)} \cup \{i_k\} \\
\lambda_k, & \text{else}.
\end{cases}
\]

This is in fact a lower bound on the objective function Φ(ik|Z(k−1)) which is obtained by using the concavity of log function and Jensen’s inequality (see Section IV for details). As expected, the objective function in (10) will be further simplified when a priori distribution on a sparse signal vector is unknown, since a priori term is removed. Based on this, the proposed greedy algorithm is described in Algorithm 1. In Section V, it will be demonstrated that the proposed proxy function performs very well.

IV. GOOD PROXY OF A POSTERIOR PROBABILITIES

In this section, we will explain how to derive the proxy function in (10) from the objective function in (9). Note that we will use the notations C0, C1, C2, and C3 in the below in order to indicate the constant terms which does not impact on the bit-wise MAP optimization in (9). Then, our goal is to efficiently compute the following posteriori probability for a given index set Z = Z(k−1) ∪ \{i_k\} = \{i_1, ..., i_{k-1}, i_k\):

\[
\Phi \left( i_k | \hat{Z}^{(k-1)} \right) - C_0 = \log P(Z \in S(X) | y) = \log P(X_{i_1} = 1, ..., X_{i_{k-1}} = 1, X_{i_k} = 1 | y) = \log \left( \frac{\prod_{x \in \Omega_{Z(k-1)}} \frac{P_X(x) f_{Y|x}^1(y|x)}{f_Y^1(y)} }{\prod_{x \in \Omega_{Z(k-1)}} \frac{P_X(x) f_{Y|x}^0(y|x)}{f_Y^0(y)} } \right),
\]

where p_X and f_{X|Y} denote the joint PMF and conditional PDF, respectively, and ΩZ is defined in Section I. We first provide some definitions which will be used throughout this section.

**Definition 1.** We define a length-N auxiliary random vector U which takes the values in the set ΩZ uniformly. Its joint PMF is denoted by q(ΩZ) where each element in ΩZ can occur with probability 1/|ΩZ| = 1/(N−K) since |Z| = k. Then, its marginal PMF can be easily obtained as Uj ~ Bern(λk), j /∈ I and Uj ~ Bern(1), j ∈ I.

From Definition 1, (13) can be written as

\[
\log P(Z \subset S(X) | y) = C_1
\]

where the last inequality follows the Jensen’s inequality due to the concavity of log function.

A. The computation of a priori part

In this subsection, we will compute the a priori part in (14). From the a priori probability p_j for j ∈ [N], we first approximate the joint PMF of X as p_X(x) ≈ \prod_{i=1}^{N} p_j(x_i).

Then, we have:

\[
\mathbb{E}_{q(ΩZ)} \left[ \log p_X(U) \right] = \mathbb{E}_{q(ΩZ)} \left[ \sum_{j=1}^{N} \log p_j(U_j) \right] = \sum_{j=1}^{N} \log p_j(1) + \sum_{j \in [N] \setminus I} \mathbb{E}_{q(ΩZ)} [\log p_j(U_j)] .
\]

Leveraging the marginal PMFs of U_j’s in (15), we have:

\[
\mathbb{E}_{q(ΩZ)} [\log p_j(U_j)] = \lambda_k \log p_j(1) + (1 - \lambda_k) \log p_j(0) = -\mathbb{H}_2(\text{Bern}(\lambda_k)) - D_{KL}(\text{Bern}(\lambda_k)||p_j) ,
\]

where \( \mathbb{H}_2 \) and \( D_{KL} \) denote the binary entropy function and KL divergence, respectively. By plugging (16) into (15), we have:

\[
\mathbb{E}_{q(ΩZ)} \left[ \log p_X(U) \right] - C_2 = \sum_{j \in [N] \setminus I} \log p_j(1) - \sum_{j \in [N] \setminus I} \mathbb{D}_{KL}(\text{Bern}(\lambda_k)||p_j) .
\]

B. The computation of likelihood part

In this subsection, we will compute the likelihood part in (14). We first introduce a binary random vector V for the ease of exposition, which is defined as V = AU. Using this, we have:

\[
\mathbb{E}_{q(ΩZ)} \left[ \log f_{Y|X}(y|U) \right] = \mathbb{E}_{q(ΩZ)} \left[ \log \prod_{j=1}^{M} f_{V_j|X}(y_j|U) \right] = \sum_{j=1}^{M} \mathbb{E}_{q(ΩZ)} \left[ \log \left( \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left( -\frac{(y_j - V_j)^2}{2\sigma^2} \right) \right) \right] .
\]
Focusing on the interesting terms depending on \( i_k \), we have

\[
\begin{align*}
\mathbb{E}_{\Omega(\Omega_2)} \left[ \log f_{\bar{Y}|\bar{X}}(\bar{y}|\bar{U}) \right] - C_3 \\
= \frac{1}{\sigma^2} \left( \sum_{j=1}^{M} y_j \mathbb{E}[V_j] - \frac{1}{2} \mathbb{E}[V_j^2] \right) \\
= \frac{1}{\sigma^2} \mathbf{y}^T \mathbf{A} \mathbb{E}[ar{U}] - \frac{1}{2\sigma^2} \text{tr} (\mathbf{A} \mathbb{E}[\bar{U} \bar{U}^T] \mathbf{A}^T) \\
= \frac{1}{\sigma^2} \mathbf{y}^T \mathbf{A} \mathbb{E}[ar{U}] - \frac{1}{2\sigma^2} \text{tr} (\mathbf{A}^T \mathbf{A} \mathbb{E}[\bar{U} \bar{U}^T]), \quad (18)
\end{align*}
\]

where

\[ \mathbf{A} \mathbb{E}[ar{U}] = \sum_{j=1}^{n} \mathbf{a}_j \mathbb{E}[U_j] = \sum_{j \in \mathcal{I}} \mathbf{a}_j + \lambda_k \sum_{j \notin \mathcal{N} \setminus \mathcal{I}} \mathbf{a}_j, \quad (19) \]

and the \((i,j)\)-element of the matrix \( \mathbb{E}[\bar{U} \bar{U}^T] \) is computed as

\[ \mathbb{E}[ar{U} \bar{U}^T]_{ij} = \begin{cases} 
1, & i, j \in \mathcal{I} \\
\lambda_k \lambda_{k+1} & i, j \notin \mathcal{N} \\
\lambda_k & \text{else.} 
\end{cases} \quad (20) \]

From (17), (18), (19), and (20), and eliminating the constant terms \( C_1, C_2, \) and \( C_3 \) we can easily derive our objective function in (10) for the bit-wise MAP detection problem.

**Remark 1.** In this paper, we only considered a binary sparse signal vector for support recovery problem. Yet, we would like to emphasize that the proposed method can straightforwardly extended to a more general case in which \( X_i \) follows a given probability distribution when \( i \) belongs to support. In this case, we only need to modify the computations of expectations in (19) and (20) where they should be performed by taking into account the probability distribution of \( X_i \).

**V. Numerical Results**

In this section we provide numerical results to show the superiority of the proposed bit-wise MAP detector. We used the reconstruction probability as a performance metric and considered OMP as benchmark method (see Remark 2 for the comparisons with the other greedy algorithms).

**No knowledge on a priori distribution:** We consider the case that a priori information on a sparse vector signal \( x \) is unknown (i.e., each component of \( x \) can be 1 with equal probability under the constraints of \( K \) sparsity). It is remarkable that in this case, a priori term in (10) of the proxy objective function is not used. Fig. 1 shows the reconstruction probabilities of the proposed bit-wise MAP detector and OMP as a function of SNRs. For the simulations, we considered the 50 × 120 measurement matrix \( \mathbf{A} \) (i.e., \( M = 50 \) and \( N = 120 \)) whose elements are drawn from I.D. Gaussian distribution with zero mean and unit variance. The proposed method shows the 4–5 times better reconstruction performances than OMP in the relative lower SNR regimes (e.g., \( 0 \sim 10 \) dBs). For the range of higher than 20dB, the proposed method achieves the two times higher reconstruction performance than OMP. Not surprisingly, the proposed method performs better in the relatively lower SNR regimes since in this case, the use of statistical information on noise does matter. We next evaluated the reconstruction probabilities of the proposed method and OMP as a function of sparsity levels (see Fig. 2). In this case, we considered a little bit larger measurement matrix (e.g., \( 80 \times 150 \) measurement matrix \( \mathbf{A} \)) to see the performances with a larger sparsity level (e.g., \( K = 30 \)). It was shown that the proposed method can successfully recover the sparse signal vector with 0.85 reconstruction probability even in high sparsity condition (e.g., \( K = 14 \)). Whereas, the reconstruction probability of OMP is lower than 0.4 after \( K = 14 \). These results demonstrated that the proposed algorithm can identify supports better than OMP even in high SNR regime, nonetheless, statistical information on noise gives smaller effect compared with relatively lower SNR regimes (as shown in Fig. 1).
VI. Conclusion

In this paper, we proposed a novel greedy algorithm where for each iteration, it finds the best support index by solving bit-wise maximum a posteriori (MAP) detection. Namely, the proposed method exploited the statistical distributions of a sparse signal vector and noise, differently from the existing greedy-based algorithms which rely on the correlation values in magnitude. Our major contribution is to introduce a good proxy function (which is simply evaluated) for the objective function of a bit-wise MAP detection problem (i.e., a selection function in the greedy algorithm), which enables the proposed method practical. Via simulation results, we demonstrated that the proposed method improves the reconstruction probability in all SNR regimes compared with the representative greedy algorithm, named OMP. Moreover, we showed that KL-divergence term, depending on a priori distribution on a sparse signal vector, performs quite well. Our ongoing work is to extend the proposed method for the case of general sparse signal vector with a certain probability distribution. Another interesting research direction is to consider sparse support recovery problems with multiple or quantized measurements.

Acknowledgment

This work was supported by Samsung Research Funding & Incubation Center of Samsung Electronics under Project Number SRFC-IT1902-00.

References