DISTRIBUTION AGNOSTIC STRUCTURED SPARSITY RECOVERY ALGORITHMS

Tareq Y. Al-Naffouri[1,2], *Mudassir Masood*[1]

[1]King Abdullah University of Science and Technology, Thuwal, Makkah, Saudi Arabia
[2]King Fahd University of Petroleum and Minerals, Dhahran, Saudi Arabia.
{tareq.alnaffouri, mudassir.masood}@kaust.edu.sa

ABSTRACT

We present an algorithm and its variants for sparse signal recovery from a small number of its measurements in a distribution agnostic manner. The proposed algorithm finds Bayesian estimate of a sparse signal to be recovered and at the same time is indifferent to the actual distribution of its non-zero elements. Termed Support Agnostic Bayesian Matching Pursuit (SABMP), the algorithm also has the capability of refining the estimates of signal and required parameters in the absence of the exact parameter values. The inherent feature of the algorithm of being agnostic to the distribution of the data grants it the flexibility to adapt itself to several related problems. Specifically, we present two important extensions to this algorithm. One extension handles the problem of recovering sparse signals having block structures while the other handles multiple measurement vectors to jointly estimate the related unknown signals. We conduct extensive experiments to show that SABMP and its variants have superior performance to most of the state-of-the-art algorithms and that too at low-computational expense.

1. INTRODUCTION

Recently, the problem of sparse signal recovery has attracted huge interest mainly due to the advent of compressed sensing and in part due to the possibility of simplifying signal processing in many applications. A typical sparse signal recovery problem is set up in the form of linear regression

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{w} \tag{1}$$

where an N-dimensional *sparse* signal $\mathbf{x}$ is recovered from its M possibly noisy linear combinations $\mathbf{y}$. The possibility of simplifying signal processing arises from the fact that a sparse signal could be recovered from a very small number of observations which defy the requirements imposed by the well-known Nyquist theorem. Thus in this setting $M \ll N$ and the linear regression model (1) basically refers to an under-determined system of equations. Moreover, the matrix $\mathbf{A} \in \mathbb{C}^{M\times N}$ is the sensing matrix which defines the linear combinations and $\mathbf{w}$ is the additive white Gaussian noise.

Several algorithms have been proposed for solving an under-determined system of linear equations utilizing the fact that the desired data vector (signal $\mathbf{x}$ in our case) is sparse. Recently, the framework of compressed sensing (CS) has been proposed in which Donoho *et al* [1] and Candes *et al* [2] proved that the sparse signal $\mathbf{x}$ could be recovered by solving the ℓ_1 minimization problem

$$\widehat{\mathbf{x}} = \arg\min_{\mathbf{x}} \parallel \mathbf{x} \parallel_1 \text{ subject to } \parallel \mathbf{y} - \mathbf{A}\mathbf{x} \parallel_2^2 < \delta. \tag{2}$$

Since N is usually large such an approach which has a computational complexity of $\mathcal{O}(N^3)$ quickly becomes unrealistic. Several faster approaches have been developed. These include, most famously, the approaches based on greedy algorithms. This category includes methods like Orthogonal Matching Pursuit (OMP) [3], projection pursuit [4], and the method proposed by Haupt *et al* [5] to name a few. An inherent feature of these algorithms is that they do not take into consideration any *a priori* information related to the distribution of the unknown signal and the noise. The only *a priori* information utilized is that of sparsity.

Furthermore, there exist algorithms based on Bayesian inference which takes into account the *a priori* information about the distribution of unknown signal as well as the noise. However, a problem is that, in real world scenarios, the distribution of unknown signals is most likely to be unknown. Thus the Bayesian approaches resort to some kind of assumption about the distribution of sparse signal $\mathbf{x}$. This assumption, which is usually Gaussian, could be wrong and thus could be detrimental to the accuracy of the recovered signal. Examples of algorithms from this category include those proposed by Larsson and Selen [6], Schniter *et al* [7] and Babacan *et al* [8] from which the former two assume a Gaussian while the last one assumes a Laplacian prior on the non-zero elements of the unknown sparse signal $\mathbf{x}$.

The algorithms mentioned above deal with a general scenario of recovering sparse signals without taking into account additional *a priori* information. It is obvious that if additional information is incorporated, a better estimate of the unknown sparse signal could be made. More recently, there has been a growing trend of taking advantage of the structure of the sparse signal recovery problem at hand. This opens up an important dimension of the sparse signal recovery algorithms as utilizing this additional *a priori* information guarantees better recovery performance. The structure utilized by these methods may include the following:

1. *structure in the sparse signal* ($\mathbf{x}$): signal might have

a block structure in which non-zero elements occur in groups (e.g., [9, 10, 11])

2. *structure in measurement vector* ($\mathbf{y}$): presence of multiple measurement vectors of a single phenomenon from different sensors (e.g., [12, 13, 14])

3. *structure in the measurement matrix* ($\mathbf{A}$): measurement matrix might have a well-defined structure. For example, it could be a Fourier matrix or a Toeplitz matrix. (e.g., [15, 16, 17])

4. *global structure*: problem rendering itself for distributed recovery where the information about the measurement matrix and the corresponding observations is distributed over several nodes/sensors (e.g., [18, 19, 20, 21])

A natural block structure might exist in the sparse signal where the few non-zero elements appear in groups. For example, an ideal sparse channel consisting of a few multi path components could be represented in a block sparse structure [22]. Some other interesting situations where block sparsity arises include gene expression analysis [23], time series data analysis involving lagged variables forming a block, multiple measurement vector (MMV) [24], Peak-to-average-power-ratio (PAPR) reduction in OFDM [25], neural activity [26] and seismic data analysis [27].

Several algorithms have been proposed taking into account the knowledge of the block structure. The foundational work in this respect was [9] which proposed the group-LASSO algorithm. However, it has limited applicability as compared to other algorithms as it makes some assumptions on the dictionary being used. Block-OMP [11] is an extension of the classical orthogonal matching pursuit algorithm (OMP [3]). It was proposed by Eldar *et al* where they used the concept of block coherence to extend the OMP algorithm. Another algorithm by Eldar called mixed ℓ_2/ℓ_1-norm recovery algorithm proposed in [24] extended the basis pursuit (BP) method to tackle block sparsity. Similarly, extensions of the CoSAMP algorithm [28] and IHT [29] were used to propose an algorithm called Block-CoSAMP [30] which has provable recovery guarantees and robustness properties. The LaMP algorithm proposed in [31] used Markov Random Fields model to capture the structure of sparse signal. They demonstrated that their algorithm performed well using fewer number of measurements.

Similar to block sparsity a closely related problem of jointly recovering multiple unknown sparse vectors having same support from multiple measurement vectors (MMV) has gained increased attention. This problem could be viewed as recovering an unknown row-sparse matrix $\mathbf{X} \in \mathbb{C}^{N \times L}$ from an observation matrix $\mathbf{Y} \in \mathbb{C}^{M \times L}$. Some of the applications where multiple observations could be utilized include equalization of sparse communication channels [32, 33], blind source separation [34], imaging of brain using magnetoencephalography (MEG) and electroencephalography (EEG) [35, 26, 36] and multivariate regression [14].

Several algorithms have been proposed taking into account the case of multiple measurement vectors. Most of the foundational and important work in this respect is the extension of already developed SMV algorithms. For example, simultaneous orthogonal matching pursuit (S-OMP) [12], MMV-orthogonal matching pursuit (M-OMP) and MMV-focal underdetermined system solver (M-FOCUSS) [13], multiple sparse Bayesian learning (M-SBL) [37] and multivariate group LASSO [14]. Another class of algorithms exploit the properties of the unknown sparse signals such as correlation and structure. For example, auto-regressive sparse Bayesian learning (AR-SBL) [38], and orthogonal subspace matching pursuit (OSMP) and subspace-augmented multiple signal classification (SA-MUSIC) algorithms proposed by Lee *et al* [39] utilize some of the inherent properties of the unknown signals for recovery.

Most of these algorithms mentioned above belong to the category of convex relaxation algorithms which are agnostic to support distribution[1] and hence demonstrate robust performance. Algorithms considering the problem of Bayesian support recovery are not as common.

In this paper we pursue a Bayesian approach which combines the features of the distribution agnostic greedy algorithms and the Bayesian approaches [40]. Thus on the one hand, the proposed algorithm is Bayesian acknowledging the noise statistics and the signal sparsity rate, while on the other hand, the approach is agnostic to the signal support statistics (making it especially useful when these statistics are unknown or non-Gaussian). Specifically, the advantages of our approach are as follows

1. The approach provides a Bayesian estimate even when the support prior is non-Gaussian/unknown.

2. The approach does not require the parameters of signal distribution (whether Gaussian or not) to be estimated.

3. The approach utilizes the prior Gaussian statistics of the additive noise and the sparsity rate of the signal.

4. The approach has low complexity due to its greedy and recursive approach.

We extend the proposed approach to solve problems having the first two types of structures mentioned above i.e., block sparsity and presence of multiple measurement vectors.

The rest of the paper is organized as follows. In Section 2 we discuss in detail our algorithm for sparse signal recovery. In Section 3 we present the variants of our algorithm for block sparse signal recovery and joint recovery of sparse signals utilizing multiple measurement vectors. In the end we conclude in Section 4.

[1] In the paper we use the term support distribution to refer to the distribution of the active elements of the unknown signal $\mathbf{x}$.

2. SUPPORT AGNOSTIC BAYESIAN MATCHING PURSUIT (SABMP)

In this paper we consider the estimation of a sparse vector $\mathbf{x} \in \mathbb{C}^N$, from an observation vector $\mathbf{y} \in \mathbb{C}^M$ obeying the linear regression model

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{w}. \tag{3}$$

Here $\mathbf{A} \in \mathbb{C}^{M\times N}$ is a known regression matrix and $\mathbf{w} \sim \mathcal{CN}(\mathbf{0}, \sigma_{\mathbf{w}}^2\mathbf{I}_M)$ is the additive white Gaussian noise vector. We shall assume that $\mathbf{x}$ has a sparse structure and is modeled as $\mathbf{x} = \mathbf{x}_A \circ \mathbf{x}_B$ where $\circ$ indicates element-by-element multiplication. The vector $\mathbf{x}_A$ consists of elements that are drawn from some unknown distribution[2] and the entries of $\mathbf{x}_B$ are drawn i.i.d. from a Bernoulli distribution with success probability λ. The sparsity parameter λ controls the sparsity of $\mathbf{x}$.

We pursue an MMSE estimate of $\mathbf{x}$ given observation $\mathbf{y}$ as follows

$$\widehat{\mathbf{x}}_{\text{MMSE}} \triangleq \mathbb{E}[\mathbf{x}|\mathbf{y}] = \sum_{\mathcal{S}} p(\mathcal{S}|\mathbf{y})\mathbb{E}[\mathbf{x}|\mathbf{y},\mathcal{S}], \tag{4}$$

where the sum is executed over all possible 2^N support sets of $\mathbf{x}$. Given the support $\mathcal{S}$, (3) becomes $\mathbf{y} = \mathbf{A}_{\mathcal{S}}\mathbf{x}_{\mathcal{S}} + \mathbf{w}$ where $\mathbf{A}_{\mathcal{S}}$ is a matrix formed by selecting columns of $\mathbf{A}$ indexed by support $\mathcal{S}$. Similarly, $\mathbf{x}_{\mathcal{S}}$ is formed by selecting entries of $\mathbf{x}$ indexed by $\mathcal{S}$. Since the distribution of $\mathbf{x}$ is unknown, it is difficult or even impossible to compute the expectation $\mathbb{E}[\mathbf{x}|\mathbf{y},\mathcal{S}]$. Thus, the best we can do is to replace it with the best linear unbiased estimate (BLUE)[3]

$$\mathbb{E}[\mathbf{x}|\mathbf{y},\mathcal{S}] \leftarrow \left(\mathbf{A}_{\mathcal{S}}^{\mathsf{H}}\mathbf{A}_{\mathcal{S}}\right)^{-1}\mathbf{A}_{\mathcal{S}}^{\mathsf{H}}\mathbf{y}. \tag{5}$$

The posterior in (4) can be written using the Bayes rule as

$$p(\mathcal{S}|\mathbf{y}) = \frac{p(\mathbf{y}|\mathcal{S})p(\mathcal{S})}{p(\mathbf{y})}. \tag{6}$$

The probability, $p(\mathbf{y})$, is a normalizing factor common to all posteriors and hence can be ignored. Since the elements of $\mathbf{x}$ are activated according to the Bernoulli distribution with success probability λ, we have

$$p(\mathcal{S}) = \lambda^{|\mathcal{S}|}(1-\lambda)^{N-|\mathcal{S}|}. \tag{7}$$

It remains to evaluate the likelihood $p(\mathbf{y}|\mathcal{S})$. Since, by virtue of $\mathbf{x}_{\mathcal{S}}$, $\mathbf{y}$ is also non-Gaussian/unknown, we are motivated to eliminate the non-Gaussian component to estimate $p(\mathbf{y}|\mathcal{S})$. This is done by projecting $\mathbf{y}$ onto the orthogonal complement space of $\mathbf{A}_{\mathcal{S}}$. This is done by multiplying $\mathbf{y}$ by the projection matrix $\mathbf{P}_{\mathcal{S}}^{\perp} = \mathbf{I} - \mathbf{P}_{\mathcal{S}} = \mathbf{I} - \mathbf{A}_{\mathcal{S}}\left(\mathbf{A}_{\mathcal{S}}^{\mathsf{H}}\mathbf{A}_{\mathcal{S}}\right)^{-1}\mathbf{A}_{\mathcal{S}}^{\mathsf{H}}$. This leaves us with $\mathbf{P}_{\mathcal{S}}^{\perp}\mathbf{y} = \mathbf{P}_{\mathcal{S}}^{\perp}\mathbf{w}$, which is Gaussian with a zero mean and covariance

$$\mathbf{K} = \mathbb{E}[(\mathbf{P}_{\mathcal{S}}^{\perp}\mathbf{w})(\mathbf{P}_{\mathcal{S}}^{\perp}\mathbf{w})^{\mathsf{H}}] = \mathbf{P}_{\mathcal{S}}^{\perp}\sigma_{\mathbf{w}}^2\mathbf{P}_{\mathcal{S}}^{\perp\,\mathsf{H}} = \sigma_{\mathbf{w}}^2\mathbf{P}_{\mathcal{S}}^{\perp}. \tag{8}$$

This allows us to write the likelihood in a simplified form as follows

$$p(\mathbf{y}|\mathcal{S}) \simeq \exp\left(-\frac{1}{2\sigma_{\mathbf{w}}^2}\left\|\mathbf{P}_{\mathcal{S}}^{\perp}\mathbf{y}\right\|_2^2\right). \tag{9}$$

To evaluate the sum in (4) is a challenging task when N is large. Therefore, we approximate the posterior over a few support sets corresponding to significant posteriors, yielding

$$\widehat{\mathbf{x}}_{\text{AMMSE}} = \sum_{\mathcal{S}\in\mathcal{S}^d} p(\mathcal{S}|\mathbf{y})\mathbb{E}[\mathbf{x}|\mathbf{y},\mathcal{S}]. \tag{10}$$

where $\mathcal{S}^d$ is the set of supports corresponding to significant posteriors. Next, we propose a greedy algorithm to find $\mathcal{S}^d$. For convenience, we represent the posteriors in the log domain. In this regard, we define a dominant support selection metric, $\nu(\mathcal{S})$, to be used by the greedy algorithm as

$$\begin{aligned}\nu(\mathcal{S}) &\triangleq \ln p(\mathbf{y}|\mathcal{S})p(\mathcal{S})\\ &= \ln\exp\left(-\frac{1}{2\sigma_{\mathbf{w}}^2}\left\|\mathbf{P}_{\mathcal{S}}^{\perp}\mathbf{y}\right\|_2^2\right) + \ln\left(\lambda^{|\mathcal{S}|}(1-\lambda)^{N-|\mathcal{S}|}\right)\\ &= \frac{1}{2\sigma_{\mathbf{w}}^2}\left\|\mathbf{A}_{\mathcal{S}}(\mathbf{A}_{\mathcal{S}}^{\mathsf{H}}\mathbf{A}_{\mathcal{S}})^{-1}\mathbf{A}_{\mathcal{S}}^{\mathsf{H}}\mathbf{y}\right\|_2^2 - \frac{1}{2\sigma_{\mathbf{w}}^2}\left\|\mathbf{y}\right\|_2^2\\ &\quad + |\mathcal{S}|\ln\lambda + (N-|\mathcal{S}|)\ln(1-\lambda).\end{aligned} \tag{11}$$

2.1. A Greedy and Recursive Approach to Find the Most Probable Supports

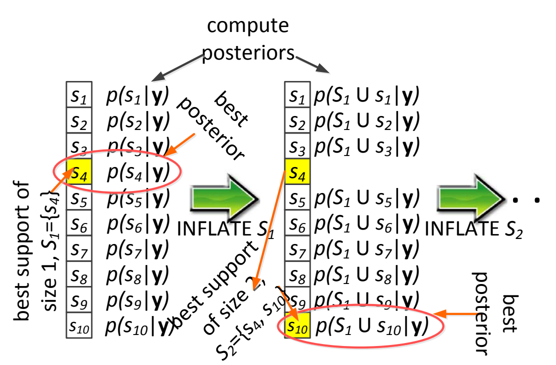


Fig. 1: Inflating the support in a greedy and recursive manner by finding the best location at each stage in terms of posterior.

To determine the set of dominant supports, $\mathcal{S}^d$, required to evaluate $\widehat{\mathbf{x}}_{\text{AMMSE}}$ in (10) we search for the optimal support in a greedy manner. We first start by finding the best support $i_1^\star$ of size 1, which involves evaluating $\nu(\mathcal{S})$ for $\mathcal{S} = \{1\},\ldots,\{N\}$. Next, we look for the

[2]The distribution may be unknown or known with unknown parameters or even Gaussian. Our developments are agnostic with regard to signal statistics.

[3]This is essentially minimum-variance unbiased estimator (MVUE) which renders the estimate (6) itself an MVU estimate. The linear MMSE would have been a more faithful approach of the MMSE but that would depend on the second-order statistics of the support, defying our support agnostic approach.

procedure GREEDY($\mathbf{A}, \mathbf{y}, \lambda, \sigma^2_{\mathbf{w}}$)
  initialize $L \leftarrow \{1, 2, \ldots, N\}$, $i \leftarrow 1$
  initialize empty sets $\mathcal{S}_{max}$, $\mathcal{S}^d$, $p(\mathcal{S}^d|\mathbf{y})$, $\mathbb{E}[\mathbf{x}|\mathbf{y}, \mathcal{S}^d]$
  $L_i \leftarrow L$
  while $i \leq P$ **do**
    $\Omega \leftarrow \{\mathcal{S}_{max} \cup \{\alpha_1\}, \mathcal{S}_{max} \cup \{\alpha_2\}, \cdots, \mathcal{S}_{max} \cup \{\alpha_{|L_i|}\} \mid \alpha_k \in L_i\}$
    compute $\{\nu(\mathcal{S}_k) \mid \mathcal{S}_k \in \Omega\}$
    find $\mathcal{S}^\star \in \Omega$ **such that** $\nu(\mathcal{S}^\star) \geq \max_j \nu(\mathcal{S}_j)$
    $\mathcal{S}^d \leftarrow \{\mathcal{S}^d, \mathcal{S}^\star\}$
    compute $p(\mathcal{S}^\star|\mathbf{y}), \mathbb{E}[\mathbf{x}|\mathbf{y}, \mathcal{S}^\star]$
    $p(\mathcal{S}^d|\mathbf{y}) \leftarrow \{p(\mathcal{S}^d|\mathbf{y}), p(\mathcal{S}^\star|\mathbf{y})\}$
    $\mathbb{E}[\mathbf{x}|\mathbf{y}, \mathcal{S}^d] \leftarrow \{\mathbb{E}[\mathbf{x}|\mathbf{y}, \mathcal{S}^d], \mathbb{E}[\mathbf{x}|\mathbf{y}, \mathcal{S}^\star\}$
    $\mathcal{S}_{max} \leftarrow \mathcal{S}^\star$
    $L_{i+1} \leftarrow L \setminus \mathcal{S}^\star$
    $i \leftarrow i + 1$
  end while
  return $\mathcal{S}^d, p(\mathcal{S}^d|\mathbf{y}), \mathbb{E}[\mathbf{x}|\mathbf{y}, \mathcal{S}^d]$
end procedure

Algorithm 1: The Greedy Algorithm

optimal support of size 2 in a greedy manner and look for the point $i_2^\star \neq i_1^\star$ such that $\mathcal{S}_2 = \{i_1^\star, i_2^\star\}$ maximizes $\nu(\mathcal{S}_2)$. We continue in this manner until we reach $\mathcal{S}_P = \{i_1^\star, \ldots, i_P^\star\}$. The value of P is selected to be slightly larger than the expected number of nonzero elements in the constructed signal such that $\Pr(|\mathcal{S}| > P)$ is sufficiently small[4]. The process explained above is illustrated in Fig. 1. The figure shows how the support is inflated based on the best non-zero location detected at each stage.

It is also important to note that our approach while inflating the support utilizes the computation of $\nu(\mathcal{S}_j)$ for the computation of $\nu(\mathcal{S}_{j+1})$ in an order-recursive manner. This greatly reduces the computational complexity of our approach and makes it one of the fastest among the state-of-the-art algorithms. We now present a formal algorithmic description of our greedy algorithm in Algorithm 1.

2.2. Refinement of the Estimated Signal

The only parameters required by the SABMP algorithm are the noise variance, $\sigma^2_{\mathbf{w}}$, and the sparsity rate, λ. The proposed SABMP method can bootstrap itself and does not require the user to provide any initial estimate of λ and $\sigma^2_{\mathbf{w}}$. Instead the method starts by finding initial estimates of these parameters which are used to compute the dominant support selection metric $\nu(\mathcal{S})$ in (11). These parameters are refined by repeated execution of the greedy algorithm. The repetition continues until a predetermined criterion has been satisfied. Specifically, this process continues until the estimate of λ changes by less than a prespecified factor (eg., we use 2% in simulations), or until a predetermined maximum number of iterations have been performed. Fig. 2 shows the process in the form of a flow chart.

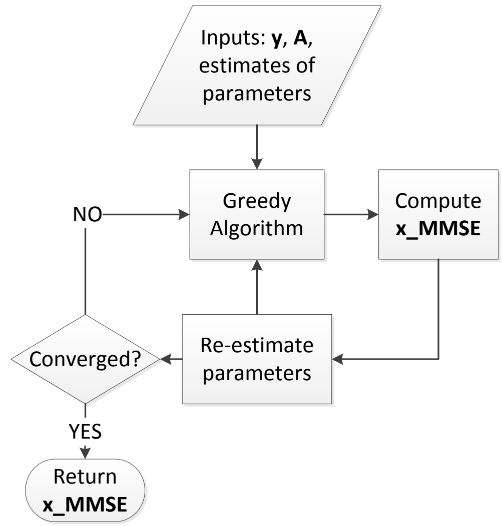


Fig. 2: Refinement of parameters and signal estimated by repeated execution of the greedy algorithm.

2.3. Simulation Results (Signal estimation performance comparison for varying sparsity parameter λ)

We now demonstrate the performance of our algorithm. Experiments were conducted for signals whose active elements are drawn from non-Gaussian (uniform non-i.i.d.) distributions.

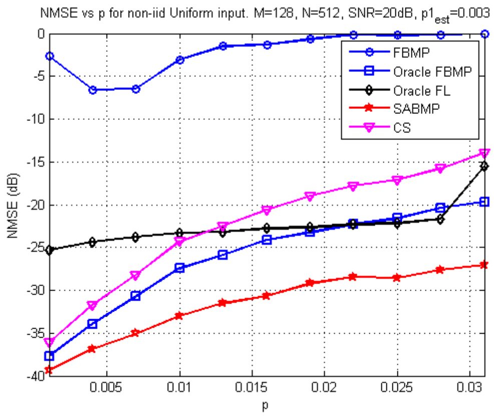
The size of the sensing/measurement matrix $\mathbf{A}$ was selected to be 128×512 where the elements were drawn i.i.d. from zero mean complex Gaussian distribution. In addition, the columns of $\mathbf{A}$ were normalized to the unit norm. The noise had a zero mean and was white and Gaussian, $\mathcal{CN}(\mathbf{0}, \sigma^2_{\mathbf{n}}\mathbf{I}_M)$, with $\sigma^2_{\mathbf{n}}$ determined according to the desired signal-to-noise ratio (SNR).

In Fig. 3a and 3b, NMSE and mean runtime are plotted, respectively, for different values of sparsity parameter λ. The plot shows the performance comparison of the proposed SABMP with FBMP [7], oracle-FBMP (oracle-aided version of FBMP), oracle-FL [8] and the CS algorithm [41]. The value of SNR selected for these experiments was 10 dB and the results were averaged over $K = 250$ trials.

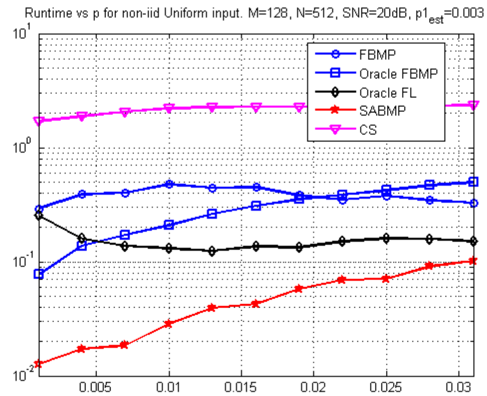
Results demonstrate the superiority of SABMP over all other algorithms. The figures suggest that, 1) even though true parameters were provided to all other algorithms, they were still outperformed by SABMP, and, 2) the huge gap in performance of FBMP and SABMP is due, in part, to the superior parameter estimation and refinement capability of SABMP.

Runtime graph of Fig. 3 depict that SABMP has an added advantage of being faster than other algorithms.

[4] $|\mathcal{S}|$, i.e., support of the constructed signal, follows the binomial distribution $\mathcal{B}(N, \lambda)$, which can be approximated by the Gaussian distribution $\mathcal{N}(N\lambda, N\lambda(1-\lambda))$ if $N\lambda > 5$. For this case, $\Pr(|\mathcal{S}| > P) = \frac{1}{2}\text{erfc}\frac{P - N\lambda}{\sqrt{2N\lambda(1-\lambda)}}$.



(a) NMSE vs λ.



(b) Runtime vs λ.

Fig. 3: NMSE and average runtime vs λ graphs for Uniform non-i.i.d. input.

3. UTILIZING STRUCTURE OF SPARSITY RECOVERY PROBLEMS

The proposed SABMP algorithm could be extended to handle the presence of special structure present in the problem at hand. We, specifically, present here the extensions of SABMP that relate to the recovery of block-sparse signals and the one that could take advantage of the multiple measurement vectors.

3.1. Block SABMP

The SABMP algorithm could be extended to solve the problem of block-sparse signal recovery. The block-sparse signal recovery problem could be divided into two categories:

1. known block partition
2. unknown block partition

In the known block partition case we know exactly where the blocks *could possibly* occur in the sparse signal. Moreover, we also have the knowledge of the size of the blocks. This could be visualized as shown in Fig. 4 .

On the other hand, in the unknown block partition case we do not have any *a priori* knowledge about the blocks. The blocks could be of any size and could occur anywhere in the signal. Such a signal is also illustrated in Fig. 4

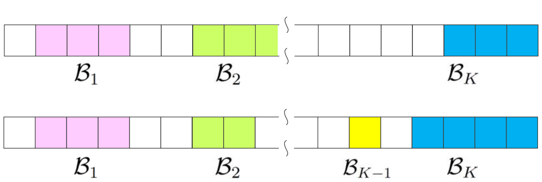


Fig. 4: Examples of known (top) and unknown (bottom) block partition block-sparse signals.

The block SABMP algorithm that we will focus on in this paper deals with the known block partition case. In this case, $\mathbf{x}$ is modeled as $\mathbf{x} = \mathbf{x}_A \circ \mathbf{x}_B$ where $\circ$ indicates element-by-element multiplication. The vector $\mathbf{x}_A$ models the support distribution and consists of elements that are drawn from some unknown distribution and $\mathbf{x}_B$ is a block-structured binary vector with K blocks of size C each as shown below:

$$\mathbf{x}_B = [\mathbf{0}^\mathsf{T}\, \mathbf{x}_{\mathcal{B}_1}^\mathsf{T}\, \mathbf{0}^\mathsf{T}\, \mathbf{x}_{\mathcal{B}_2}^\mathsf{T}\, \mathbf{0}^\mathsf{T} \cdots \mathbf{x}_{\mathcal{B}_K}^\mathsf{T}\, \mathbf{0}^\mathsf{T}]^\mathsf{T}, \tag{12}$$

where $\mathcal{B}_i$, $i = 1, \ldots K$ refer to the size C supports of each block. Equivalently, we can write $\mathbf{x}_B$ as[5]

$$\mathbf{x}_B = \mathbf{x}_b \otimes \mathbf{1}_C \tag{13}$$

where $\mathbf{1}_C$ is a $C \times 1$ vector of 1's and $\mathbf{x}_b$ is a $K \times 1$ binary vector. Block sparsity requires that only a few among the K blocks in $\mathbf{x}_B$ are non-zero. Pursuing the same methodology as discussed in Sec. 2 we conclude that the dominant support selection metric $\nu(\mathcal{S})$, to be used by the greedy algorithm, is

$$\begin{aligned}
\nu(\mathcal{S}) &\triangleq \ln p(\mathbf{y}|\mathcal{S})p(\mathcal{S}) \\
&= \ln \exp\Big(\frac{-1}{2\sigma_\mathbf{w}^2} \left\|\mathbf{P}_\mathcal{S}^\perp \mathbf{y}\right\|^2\Big) + \ln\big(\lambda^{|\mathcal{S}|/C}(1-\lambda)^{K-|\mathcal{S}|/C}\big) \\
&= \frac{1}{2\sigma_\mathbf{w}^2} \left\|\mathbf{A}_\mathcal{S}(\mathbf{A}_\mathcal{S}^\mathsf{H}\mathbf{A}_\mathcal{S})^{-1}\mathbf{A}_\mathcal{S}^\mathsf{H}\mathbf{y}\right\|^2 - \frac{1}{2\sigma_\mathbf{w}^2}\left\|\mathbf{y}\right\|^2 \\
&\quad + \frac{|\mathcal{S}|}{C}\ln\lambda + \Big(K - \frac{|\mathcal{S}|}{C}\Big)\ln(1-\lambda).
\end{aligned} \tag{14}$$

This dominant support selection metric for block-structured signals could now be used to perform greedy estimation. The process is similar to that discussed in Sec. 2.1.

3.2. Simulation Results (Block SABMP)

To demonstrate the performance of the known block partition case block-SABMP algorithm we compare it with the known block partition version of cluster-SBL algorithm

[5] Our algorithm applies to the general case when the C sized blocks could be placed arbitrarily within $\mathbf{x}_B$. However, due to space limitation we focus on the special case (13).

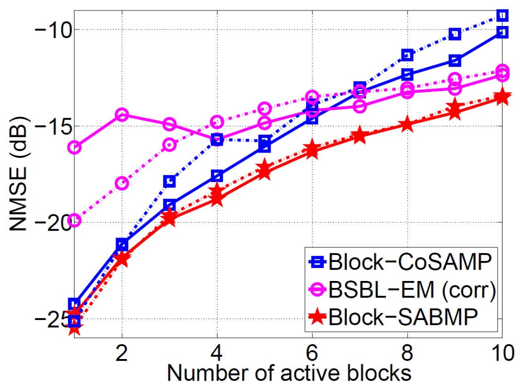


Fig. 5: Block-sparsity: NMSE vs sparsity rate. Distribution of non-zero elements: Gaussian (solid), non-Gaussian (dotted)

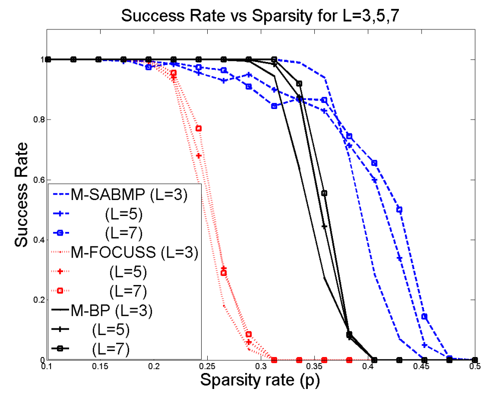


Fig. 6: MMV: Signal recovery success rate.

[42] and Block-CoSaMP [30]. These methods were selected for comparisons as these have been shown to be robust and outperform many of the state-of-the-art algorithms. The results shown in Fig. 5 demonstrate that the proposed method outperforms the other two algorithms.

3.3. Multiple Measurement Vector (MMV) SABMP (M-SABMP)

In this section we show how the SABMP algorithm could be extended to use multiple measurement vectors for joint sparse signal recovery. In this case we basically consider the estimation of a row-sparse matrix, $\mathbf{X} \in \mathbb{C}^{N\times L}$, from multiple observation vectors represented as a matrix $\mathbf{Y} \in \mathbb{C}^{M\times L}$, obeying the linear regression model,

$$\mathbf{Y} = \mathbf{A}\mathbf{X} + \mathbf{W}. \tag{15}$$

Here $\mathbf{A} \in \mathbb{C}^{M\times N}$ is a known regression/sensing matrix[6] and $\mathbf{W}$ is a matrix representing a collection of additive white Gaussian noise vectors following $\mathcal{CN}(\mathbf{0}, \sigma^2_{\mathbf{w}}\mathbf{I}_M)$. To assist in algorithm development in the following, we will represent the matrices $\mathbf{X}, \mathbf{Y}$ and $\mathbf{W}$ as collection of column vectors $\mathbf{X} = [\mathbf{x}_1\, \mathbf{x}_2\, \dots\, \mathbf{x}_L]$, $\mathbf{Y} = [\mathbf{y}_1\, \mathbf{y}_2\, \dots\, \mathbf{y}_L]$ and $\mathbf{W} = [\mathbf{w}_1\, \mathbf{w}_2\, \dots\, \mathbf{w}_L]$ respectively, wherever needed. The formulation is valid for $L = 1$ (SMV) as well as $L > 1$ (MMV) case. We shall assume that each unknown sparse vector $\mathbf{x}_i$ is modeled as $\mathbf{x}_i = \mathbf{x}_{A_i} \circ \mathbf{x}_B$ where $\mathbf{x}_{A_i}$ and $\mathbf{x}_{A_j}$ are independent and $\circ$ indicates element-by-element multiplication. The vector $\mathbf{x}_{A_i}$ models the support distribution and consists of elements that are drawn from some unknown distribution and $\mathbf{x}_B$ is a binary vector whose entries are drawn i.i.d. from a Bernoulli distribution with success probability λ. Recall that all vectors $\mathbf{x}_i$'s have exactly the same sparse structure due to $\mathbf{X}$ being row-sparse.

Following a method similar to that presented in Sec. 2 we find that the dominant support selection metric $\nu(\mathcal{S})$ in this case is given by

$$\begin{aligned}\nu(\mathcal{S}) &\triangleq \ln p(\mathbf{Y}|\mathcal{S})p(\mathcal{S}) \\ &= \ln \exp\Big(\frac{-1}{2\sigma^2_{\mathbf{w}}}\sum_{i=1}^{L}\left\|\mathbf{P}^{\perp}_{\mathcal{S}}\mathbf{y}_i\right\|^2\Big) + \ln(\lambda^{|\mathcal{S}|}(1-\lambda)^{N-|\mathcal{S}|}) \\ &= \frac{1}{2\sigma^2_{\mathbf{w}}}\sum_{i=1}^{L}\Big[\left\|\mathbf{A}_{\mathcal{S}}(\mathbf{A}^{\mathsf{H}}_{\mathcal{S}}\mathbf{A}_{\mathcal{S}})^{-1}\mathbf{A}^{\mathsf{H}}_{\mathcal{S}}\mathbf{y}_i\right\|^2 - \left\|\mathbf{y}_i\right\|^2\Big] \\ &\quad + |\mathcal{S}|\ln\lambda + (N - |\mathcal{S}|)\ln(1-\lambda). \end{aligned} \tag{16}$$

$\nu(\mathcal{S})$ is then used to construct the row-sparse matrix $\mathbf{X}$ by following the greedy process discussed in detail for the single measurement vector (Sec. 2.1).

3.4. Simulation Results (M-SABMP)

In order to demonstrate the performance of the proposed M-SABMP algorithm we compare it with the multiple measurement vectors versions of the FOcal Underdetermined System Solver (FOCUSS) [13] and Basis Pursuit Denoising (BPDN) [43]. It was shown in Sec. 2 that the SMV version of the proposed algorithm detects the unknown support with high accuracy and thus results in better signal reconstruction. In the MMV case, since more information is available the decisions made by M-SABMP are strengthened even further and thus are more reliable. In Fig. 6 we plot the rate of successfully reconstructing sparse signals using multiple measurements as a function of sparsity rate. Success rate is calculated as the ratio of the total number of successful recoveries and the total experiments. The success rate for each value of sparsity rate was averaged over a total of 200 experiment realizations. An experiment was declared successful if the resulting NMSE was ≤ -10dB. It is observed that the M-SABMP algorithm was able to recover signals with higher values of the sparsity rate which correspond to dense signals. Specifically, we notice that for $L = 3$ the proposed algorithm was able to recover signals having 35% non-zero elements with 100% accuracy. Moreover the graphs also show that both M-BPDN and M-FOCUSS fail completely at 40% and 30% sparsity rate respectively while this figure for M-SABMP is 45%.

[6] Our algorithm is capable of modeling the scenario where multiple sensing matrices could be used to sense the unknown sparse vectors. However, in this paper we focus on the case where sensing matrices are same.

4. CONCLUSION

In this paper we presented a sparse signal recovery algorithm which is Bayesian and at the same time is agnostic to the distribution of the active elements of the unknown signal. It was shown that the algorithm can be easily extended to recover block-sparse signals and to utilize multiple measurement vectors. Extensive simulations and comparisons show that the algorithm and its extensions outperform many state-of-the-art algorithms.

5. ACKNOWLEDGEMENT

The authors would like to acknowledge the support provided by King Fahd University of Petroleum and Minerals (KFUPM) and King Abdulaziz City for Science and Technology (KACST) through the Science and Technology Unit at KFUPM for funding this work through project number 09-ELE781-4 as part of the National Science, Technology and Innovation Plan.

6. REFERENCES

[1] D. L. Donoho, "Compressed sensing," *IEEE Trans. Info. Theory*, vol. 52, pp. 1289–1306, Apr. 2006.

[2] E. J. Candes, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Info. Theory*, vol. 52, pp. 489–509, Feb. 2006.

[3] Y. C. Pati, R. Rezaifar, and P. S. Krishnaprasad, "Orthogonal matching pursuit: Recursive function approximation with applications to Wavelet decomposition," in *Proc. 27th Asilomar Conf. Signals, Systems, Comput.*, Nov 1993, pp. 40–44.

[4] Peter J. Huber, "Projection Pursuit," *The Annals of Statistics*, vol. 13, no. 2, pp. 435–475, Jun. 1985.

[5] Jarvis Haupt and Robert Nowak, "Signal reconstruction from noisy random projections," *IEEE Trans. Info. Theory*, vol. 52, pp. 4036–4048, Sep. 2006.

[6] Erik G. Larsson and Yngve Selén, "Linear regression with a sparse parameter vector," *IEEE Trans. Signal Process.*, vol. 55, pp. 451–460, Feb 2007.

[7] P. Schniter, L.C. Potter, and J. Ziniel, "Fast Bayesian Matching Pursuit," in *Information Theory and Applications Workshop, 2008*, 27 Jan. – 1 Feb. 2008, pp. 326–333.

[8] S. D. Babacan, R. Molina, and A. K. Katsaggelos, "Bayesian Compressive Sensing Using Laplace Priors," *IEEE Trans. Image Process.*, vol. 19, no. 1, pp. 53 –63, Jan. 2010.

[9] M. Yuan and Y. Lin, "Model selection and estimation in regression with grouped variables," *Journal of The Royal Statistical Society Series B-Statistical Methodology*, vol. 68, no. Part 1, pp. 49–67, 2006.

[10] Mudassir Masood and Tareq Y. Al-Naffouri, "Support agnostic bayesian matching pursuit for block sparse signals," in *IEEE International Conf. on Acoustics Speech and Signal Process. (ICASSP), 2013*, May 2013.

[11] Yonina C. Eldar, Patrick Kuppinger, and Helmut Boelcskei, "Block-Sparse Signals: Uncertainty Relations and Efficient Recovery," *IEEE Trans. Signal Process.*, vol. 58, no. 6, pp. 3042–3054, Jun. 2010.

[12] J. A. Tropp, A. C. Gilbert, and M. J. Strauss, "Simultaneous sparse approximation via greedy pursuit," in *IEEE International Conf. on Acoustics, Speech, and Signal Process., 2005*, March 2005, vol. 5, pp. 721 – 724.

[13] S. F. Cotter, B. D. Rao, Kjersti Engan, and K. Kreutz-Delgado, "Sparse solutions to linear inverse problems with multiple measurement vectors," *IEEE Trans. Signal Process.*, vol. 53, no. 7, pp. 2477 – 2488, July 2005.

[14] Wainwright-M. J. Obozinski, G. and M. I. Jordan, "Support union recovery in high-dimensional multivariate regression," *The Annals of Statistics*, , no. 1, pp. 1–47.

[15] A. A. Quadeer and Tareq Y. Al-Naffouri, "Structure-based bayesian sparse reconstruction," *IEEE Trans. Signal Process.*, vol. 60, no. 12, pp. 6354–6367, 2012.

[16] Tareq Y. Al-Naffouri, A. A. Quadeer, and G. Caire, "Impulsive noise estimation and cancellation in dsl using orthogonal clustering," in *Information Theory Proceedings (ISIT), 2011 IEEE International Symposium on*, 2011, pp. 2841–2845.

[17] J. Haupt, W. U. Bajwa, G. Raz, and R. Nowak, "Toeplitz compressed sensing matrices with applications to sparse channel estimation," *IEEE Trans. Info. Theory*, vol. 56, no. 11, pp. 5862–5875, 2010.

[18] M. F. Duarte, Shriram Sarvotham, D. Baron, M. B. Wakin, and R. G. Baraniuk, "Distributed compressed sensing of jointly sparse signals," in *Signals, Systems and Computers, 2005. Conf. Record of the Thirty-Ninth Asilomar Conf. on*, 2005, pp. 1537–1541.

[19] J. F. C. Mota, J. Xavier, P. M. Q. Aguiar, and M. Puschel, "Distributed basis pursuit," *IEEE Trans. Signal Process.*, vol. 60, no. 4, pp. 1942–1956, 2012.

[20] D. Sundman, S. Chatterjee, and M. Skoglund, "A greedy pursuit algorithm for distributed compressed sensing," in *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on*, 2012, pp. 2729–2732.

[21] S. Patterson, Y. C. Eldar, and I. Keidar, "Distributed sparse signal recovery for sensor networks," in

Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on, 2013.

[22] Hwanjoon Kwon and Bhaskar D. Rao, "On the benefits of the block-sparsity structure in sparse signal recovery," in *IEEE International Conf. on Acoustics, Speech and Signal Process.*, 2012, pp. 3685–3688.

[23] Farzad Parvaresh, Haris Vikalo, Sidhant Misra, and Babak Hassibi, "Recovering Sparse Signals Using Sparse Measurement Matrices in Compressed DNA Microarrays," *IEEE Journal of Selected Topics in Signal Process.*, vol. 2, no. 3, pp. 275–285, June 2008.

[24] Yonina C. Eldar and Moshe Mishali, "Robust Recovery of Signals From a Structured Union of Subspaces," *IEEE Trans. Info. Theory*, vol. 55, no. 11, pp. 5302–5316, Nov. 2009.

[25] Abdullatif Al-Rabah, Mudassir Masood, and Tareq Y. Al-Naffouri, "Bayesian Recovery of Clipped OFDM for PAPR Reduction," *IEEE Trans. Signal Process.*, to be submitted.

[26] I. F. Gorodnitsky, J. S. George, and B. D. Rao, "Neuromagnetic Source Imaging with FOCUSS - A Recursive Weighted Minimum Norm Algorithm," *Electroencephalography and Clinical Neurophysiology*, vol. 95, no. 4, pp. 231–251, Oct. 1995.

[27] Syed R. Hussaini, Mudassir Masood, Tareq Y. Al-Naffouri, and Aldo Vesnaver, "Seismic deconvolution of non-gaussian sparse reflectivity," *IEEE Trans. Geoscience and Remote Sensing*, to be submitted.

[28] D. Needell and J. A. Tropp, "CoSaMP: Iterative signal recovery from incomplete and inaccurate samples," *Applied and Computational Harmonic Analysis*, vol. 26, no. 3, pp. 301–321, May 2009.

[29] Thomas Blumensath and Mike E. Davies, "Iterative hard thresholding for compressed sensing," *Applied and Computational Harmonic Analysis*, vol. 27, no. 3, pp. 265–274, Nov. 2009.

[30] Richard G. Baraniuk, Volkan Cevher, Marco F. Duarte, and Chinmay Hegde, "Model-Based Compressive Sensing," *IEEE Trans. on Info. Theory*, vol. 56, no. 4, pp. 1982–2001, Apr. 2010.

[31] V. Cevher, , M. F. Duarte, C. Hegde, and R. G. Baraniuk, "Sparse signal recovery using Markov Random Fields," in *NIPS*, Vancouver, B.C., Canada, 8–11 December 2008.

[32] I. J. Fevrier, S. B. Gelfand, and M. P. Fitz, "Reduced complexity decision feedback equalization for multipath channels with large delay spreads," *IEEE Trans. Commun.*, vol. 47, no. 6, pp. 927 –937, Jun 1999.

[33] S. F. Cotter and B. D. Rao, "Sparse channel estimation via matching pursuit with application to equalization," *IEEE Trans. Commun.*, vol. 50, no. 3, pp. 374 –377, Mar 2002.

[34] A. Cichocki, "Blind source separation: New tools for extraction of source signals and denoising," *SPIE*, pp. 11–25, 2005.

[35] D. Wipf and S. Nagarajan, "A unified Bayesian framework for MEG/EEG source imaging," *NeuroImage*, pp. 947–966, 2008.

[36] J. W. Phillips, R. M. Leahy, and J. C. Mosher, "Meg-based imaging of focal neuronal current sources," *IEEE Trans. Medical Imaging*, vol. 16, no. 3, pp. 338 –348, June 1997.

[37] D. P. Wipf and B. D. Rao, "An empirical bayesian strategy for solving the simultaneous sparse approximation problem," *IEEE Trans. Signal Process.*, vol. 55, no. 7, pp. 3704 –3716, July 2007.

[38] Zhilin Zhang and B. D. Rao, "Sparse signal recovery in the presence of correlated multiple measurement vectors," in *IEEE International Conf. on Acoustics, Speech and Signal Process. (ICASSP)*, March 2010, pp. 3986 –3989.

[39] Kiryung Lee, Y. Bresler, and M. Junge, "Subspace methods for joint sparse recovery," *IEEE Trans. Info. Theory*, vol. 58, no. 6, pp. 3613 –3641, June 2012.

[40] Mudassir Masood and Tareq Y. Al-Naffouri, "A Fast Non-Gaussian Bayesian Matching Pursuit Method for Sparse Reconstruction," *IEEE Trans. Signal Process.*, to appear.

[41] CVX Research Inc, "CVX: Matlab software for disciplined convex programming, version 1.22," .

[42] Zhilin Zhang and Bhaskar D. Rao, "Recovery of block sparse signals using the framework of block sparse bayesian learning," in *IEEE International Conf. on Acoustics Speech and Signal Process.*, 2012, pp. 3345–3348.

[43] E. van den Berg and M. P. Friedlander, "SPGL1: A solver for large-scale sparse reconstruction," June 2007, http://www.cs.ubc.ca/labs/scl/spgl1.