SPARSE SUPPORT RECOVERY VIA COVARIANCE ESTIMATION

Lekshmi Ramesh and Chandra R. Murthy

Dept. of ECE, Indian Institute of Science, Bangalore $560\,012$, India

ABSTRACT

We consider the problem of recovering the common support of a set of k-sparse signals $\{\mathbf{x}_i\}_{i=1}^L$ from noisy linear underdetermined measurements of the form $\{\Phi\mathbf{x}_i + \mathbf{w}_i\}_{i=1}^L$ where $\Phi \in \mathbb{R}^{m \times N}$ ($m < N$) is the sensing matrix and $\mathbf{w}_i$ is the additive noise. We employ a Bayesian setup where we impose a Gaussian prior with zero mean and a common diagonal covariance matrix $\boldsymbol{\Gamma}$ across all $\mathbf{x}_i$, and formulate the support recovery problem as one of covariance estimation. We develop an algorithm to find the approximate maximum-likelihood estimate of $\boldsymbol{\Gamma}$ using a modified reweighted minimization procedure. Empirically, we find that the proposed algorithm succeeds in exactly recovering the common support with high probability in the $k < m$ regime with L of the order of m and in the $k \geq m$ regime with larger L. The key advantage of the proposed algorithm is that its complexity is independent of L, unlike existing sparse support recovery algorithms.

Index Terms— Sparse support recovery, covariance estimation, multiple measurement vectors

1. INTRODUCTION

The problem of recovering joint-sparse signals from multiple measurement vectors (MMVs) has received much attention in the recent literature, as it has a variety of applications in next generation communication systems such as channel estimation [1], distributed source coding [2], cooperative spectrum sensing [3], and sparse event detection in wireless sensor networks [4], to name a few. Formally, we are given observations $\mathbf{y}_i \in \mathbb{R}^m$ generated according to the linear model

$$\mathbf{y}_i = \Phi\mathbf{x}_i + \mathbf{w}_i, \quad i \in [L], \tag{1}$$

where $\Phi \in \mathbb{R}^{m \times N}(m < N)$ is the measurement matrix, $\mathbf{x}_i \in \mathbb{R}^N$, $i \in [L]$, are a set of k-sparse vectors, i.e., they have at most k nonzero entries, and $\mathbf{w}_i \stackrel{iid}{\sim} \mathcal{N}(0, \sigma^2 I)$ is additive noise. Further, $\mathbf{x}_i$ are assumed to be *jointly sparse vectors*. That is, they have a common support, where the support of a vector is defined as the set of indices corresponding to its nonzero entries. The goal of the MMV sparse vector/support recovery problem is to determine the $\mathbf{x}_i$'s or their common support from $\mathbf{y}_i$ and Φ. In this work, we propose a novel covariance estimation based algorithm for the MMV sparse support recovery problem.

Sparse signal/support recovery based on covariance matching is a relatively new theme in the MMV literature. Algorithms in this category include Co-LASSO [5] and RD-CMP [6]. The M-SBL algorithm, introduced in [7], can also be interpreted as performing covariance matching based on the log-det Bregman divergence metric [8]. These algorithms are based on a Bayesian modeling approach, and have been empirically shown to significantly outperform the so-called Type-I MMV algorithms such as the $\ell_{1,2}$ penalty algorithm [9], simultaneous OMP [10], etc.

Most of the recovery algorithms for the MMV case focus on the $k < m$ regime. It was shown recently in [5] that support recovery is possible even in the $k \geq m$ regime. The authors propose a LASSO-based approach to recover the common support using correlation among the $\mathbf{x}_i$. The authors empirically showed that support recovery is possible (with large L) for support size $k \geq m$ and conjectured that k can be as large as $O(m^2)$ (it is assumed that $m^2 \leq N$). Similar observations were made in [6] and [11] also, where M-SBL and RD-CMP were shown to successfully recover the support even when $k \geq m$, respectively. A theoretical understanding of the success of M-SBL was recently developed in [8].

In this work, we show that a maximum likelihood (ML) based technique that also uses the correlation information among $\mathbf{x}_i$ successfully recovers the common support in the $k < m$ case with "small" L (of the order of m) and requires a larger L in the $k \geq m$ case. A key distinction between our technique and M-SBL is that our technique estimates the hyperparameters using the sample covariance matrix of the observations. M-SBL, on the other hand, uses the entire sequence of samples in every iteration of the expectation maximization procedure, making it slow when L is large. Further, we explicitly account for the approximation error in constructing the covariance of the observation from a finite number of measurements. We derive the statistics of this error term and model it using a joint-Gaussian distribution, whose covariance turns out to be dependent on the sparse signal statistics. The makes the resulting support recovery problem one of nonconvex optimization, and we propose a modified reweighted minimization based iterative solution. Finally, we empirically illustrate the performance of our algorithm and compare it with state-of-the-art algorithms from the literature. We find that the algorithm works well with L of the order m when $k < m$, and requires larger L when $k > m$.

The outcome is an algorithm that provides competitive performance and whose complexity is virtually independent of L, the number of MMVs.

2. PROBLEM FORMULATION

We consider the MMV problem where we obtain observations $\mathbf{y}_i, i \in [L]$, as in (1), where the sparse vectors $\mathbf{x}_i$ have a common support $T \subseteq [N]$ with $|T| \leq k$. We further assume that the nonzero entries of $\mathbf{x}_i$ are uncorrelated. To capture the latent structure in $\mathbf{x}_i$, we impose the following prior:

$$p(\mathbf{x}_i; \boldsymbol{\gamma}) = \prod_{j=1}^{N} \frac{1}{\sqrt{2\pi\gamma_j}} e^{-\frac{\mathsf{x}_{ij}^2}{2\gamma_j}}, \qquad (2)$$

where $\mathbf{x}_{ij}$ denotes the jth entry of $\mathbf{x}_i$ and $\gamma_j \geq 0$ denotes the common variance of $\mathbf{x}_{ij}, i \in [L]$. In other words, we have $\mathbf{x}_i \overset{iid}{\sim} \mathcal{N}(0, \Gamma)$ where $\Gamma = \text{diag}(\boldsymbol{\gamma})$, with $\gamma \triangleq [\gamma_1, \gamma_2, \ldots, \gamma_N]^T$. This type of prior to model sparsity was first introduced in [12]. The observations $\mathbf{y}_i$ are therefore distributed as $\mathcal{N}(0, \Phi\Gamma\Phi^\top + \sigma^2 I)$. The goal is to estimate the common support T from $\{\mathbf{y}_i\}_{i=1}^{L}$.

We observe that under the prior model above, $T = \text{supp}(\mathbf{x}_i) = \text{supp}(\boldsymbol{\gamma})$, since $\gamma_j = 0$ if and only if $x_{ij} = 0$ almost surely. Hence, support recovery from MMVs is equivalent to recovering the support of $\boldsymbol{\gamma}$. As we will see later, our simulations show that under our prior model, we can recover the support even when $k \geq m$.

3. GAUSSIAN-APPROXIMATION BASED SUPPORT ESTIMATION

Let $\Sigma \in \mathbb{R}^{m \times m}$ denote the covariance matrix of the observations. Then, in the noiseless case (i.e., when $\sigma^2 = 0$), we have $\Sigma = \Phi\Gamma\Phi^\top$, which can be rewritten as

$$\text{vec}(\Sigma) = (\Phi \odot \Phi)\boldsymbol{\gamma}, \qquad (3)$$

where $\odot$ denotes the Khatri-Rao product [13]. Given the formulation above, the goal is to estimate the sparse non-negative vector $\boldsymbol{\gamma}$ from Σ. In [5], the authors formulate the support recovery problem as the following convex problem:

$$\begin{aligned} \min_{\boldsymbol{\gamma}} \ &\|\boldsymbol{\gamma}\|_1 \\ \text{s.t. } &(\Phi \odot \Phi)\boldsymbol{\gamma} = \text{vec}(\Sigma). \end{aligned} \qquad (4)$$

This model is analyzed in [14], and conditions under which the model is identifiable are derived. If we had access to the true covariance matrix Σ (which corresponds to the $L \to \infty$ case), then we could work with the system of equations in (3) to recover the support of $\boldsymbol{\gamma}$ which, in turn, would give us the common support of $\mathbf{x}_i$'s. For finite L, we use an estimate of Σ, the sample covariance matrix $\hat{\Sigma} \triangleq \frac{1}{L} \sum_{i=1}^{L} \mathbf{y}_i \mathbf{y}_i^\top$. It

can be easily shown that the sample covariance matrix $\hat{\Sigma}$ is a sufficient statistic for estimating Γ, but note that the "exact" covariance matrix Σ is actually not available in practice. In this paper, we derive the statistics of the "noise" arising because of the finite sample approximation to Σ, and then find the ML estimate of γ. More precisely, the sample covariance matrix can be written as a noisy version of the true covariance matrix, i.e.,

$$\hat{\Sigma} = \Sigma + E, \qquad (5)$$

where E represents the noise/error matrix. Equivalently, vectorizing the matrices on either side of (5), we get

$$\mathbf{r} = (\Phi \odot \Phi)\boldsymbol{\gamma} + \mathbf{e}, \qquad (6)$$

where $\mathbf{r} \triangleq \text{vec}(\hat{\Sigma})$ and $\mathbf{e} \triangleq \text{vec}(E)$. We now proceed to find the approximate ML estimate of $\boldsymbol{\gamma}$. To that end, we first derive the statistics of the noise.

3.1. Noise Statistics

Our starting point is the following Lemma, which provides the mean and covariance of the vectorized noise $\mathbf{e}$.

Lemma 1. Consider $\{\mathbf{y}_i\}_{i=1}^{L}$ drawn i.i.d. from $\mathcal{N}(0, \Sigma)$. Let $\hat{\Sigma}$ denote the sample covariance matrix and $\mathbf{e} = \text{vec}(\hat{\Sigma} - \Sigma)$. Further, let $B = \text{cov}(\text{vec}(\mathbf{z}\mathbf{z}^\top))$ where $\mathbf{z} \sim \mathcal{N}(0, I)$ and let C be a matrix satisfying $\Sigma = CC^\top$. Then,

$$\mathbb{E}\mathbf{e} = 0, \ \ \text{cov}(\mathbf{e}) = \frac{1}{L}(C \otimes C)B(C \otimes C)^\top,$$

where $\otimes$ denotes the Kronecker product.

Proof. Let $E = \hat{\Sigma} - \Sigma$. The mean computation is straightforward:

$$\mathbb{E}(E) = \frac{1}{L}\sum_{i=1}^{L} \mathbb{E}\mathbf{y}_i \mathbf{y}_i^\top - \Sigma = 0.$$

The covariance of E is a tensor of size $m \times m \times m^2$. It can be computed as follows:

$$\begin{aligned} \text{cov}(E) &= \text{cov}\left(\frac{1}{L}\sum_{i=1}^{L} \mathbf{y}_i \mathbf{y}_i^\top - \Sigma\right) \\ &= \frac{1}{L}\text{cov}(\mathbf{y}_1\mathbf{y}_1^\top - \Sigma) = \frac{1}{L}\text{cov}(\mathbf{y}\mathbf{y}^\top), \end{aligned}$$

where we used the fact that $\left(\frac{\mathbf{y}_i\mathbf{y}_i^\top}{L} - \frac{\Sigma}{L}\right)$ are independent for $i = 1, ..., L$. We now represent $\mathbf{y}$ as $\mathbf{y} = C\mathbf{z}$, where $\mathbf{z} \sim \mathcal{N}(0, I)$ and $\Sigma = CC^\top$. We can then calculate the covariance matrix of the vectorized version of E as follows:

$$\begin{aligned} \text{cov}(\text{vec}(E)) &= \frac{1}{L}\text{cov}(\text{vec}(C\mathbf{z}\mathbf{z}^\top C^\top)) \\ &= \frac{1}{L}\text{cov}((C \otimes C)\text{vec}(\mathbf{z}\mathbf{z}^\top)) \\ &= \frac{1}{L}(C \otimes C)B(C \otimes C)^\top, \end{aligned}$$

where $B \triangleq \text{cov}(\text{vec}(\mathbf{z}\mathbf{z}^\top))$. $\blacksquare$

For our model, $\Sigma = \Phi\Gamma\Phi^\top + \sigma^2 I$. Letting $C = \Phi D^{\frac{1}{2}}$, with $D = (\Gamma + \sigma^2\Phi^\dagger\Phi^{\top\dagger})$, and using Lemma 1, we get

$$
\begin{aligned}
\mathrm{cov}(\mathrm{vec}(E)) &= \frac{1}{L}(\Phi D^{\frac{1}{2}} \otimes \Phi D^{\frac{1}{2}})B(\Phi D^{\frac{1}{2}} \otimes \Phi D^{\frac{1}{2}}) \\
&= \frac{1}{L}(\Phi \otimes \Phi)(D^{\frac{1}{2}} \otimes D^{\frac{1}{2}})B(D^{\frac{1}{2}} \otimes D^{\frac{1}{2}})(\Phi \otimes \Phi)^\top,
\end{aligned}
\tag{7}
$$

where the second step uses the property that $UV \otimes XY = (U \otimes X)(V \otimes Y)$. The $N^2 \times N^2$ covariance matrix B of $\mathrm{vec}(\mathbf{zz}^\top)$ can be computed explicitly for a given N and it can be verified that the entries of B lie in $\{0, 1, 2\}$.

For the noiseless case, we have $D = \Gamma$ and we can further simplify (7) by exploiting the structure of B. Specifically, it can be shown that B can be expressed as $I_{m^2} + Q$, where Q denotes a permutation matrix and I_{m^2} denotes the $m^2 \times m^2$ identity matrix. Using this fact and the structure of $\Gamma^{\frac{1}{2}} \otimes \Gamma^{\frac{1}{2}}$, we get

$$
\begin{aligned}
W &\triangleq \mathrm{cov}(\mathrm{vec}(E)) \\
&= \frac{1}{L}(\Phi \otimes \Phi)(\Gamma^{\frac{1}{2}} \otimes \Gamma^{\frac{1}{2}})(I_{m^2} + Q)(\Gamma^{\frac{1}{2}} \otimes \Gamma^{\frac{1}{2}})(\Phi \otimes \Phi)^\top \\
&= \frac{1}{L}(\Phi \otimes \Phi)B(\Gamma \otimes \Gamma)(\Phi \otimes \Phi)^\top.
\end{aligned}
\tag{8}
$$

In the next section, we use these statistics to derive an approximate ML estimate of $\boldsymbol{\gamma}$.

3.2. Maximum Likelihood Estimation of γ

We consider the model derived in the previous section:

$$
\mathbf{r} = A\boldsymbol{\gamma} + \mathbf{e}, \tag{9}
$$

where $A \triangleq (\Phi \odot \Phi)$. We seek the ML estimate of $\boldsymbol{\gamma}$ from $\mathbf{r}$. Note that the statistics of the noise $\mathbf{e}$ also depends on $\boldsymbol{\gamma}$.

Since $\mathbf{r}$, $A\boldsymbol{\gamma}$ and $\mathbf{e}$ are vectorized versions of $m \times m$ symmetric matrices, they lie in an $\frac{m(m+1)}{2}$ dimensional subspace of $\mathbb{R}^{m^2}$. We therefore restrict our attention to the $\frac{m(m+1)}{2}$ linearly independent equations in (9). This can be done by pre-multiplying (9) by a projection matrix $P \in \mathbb{R}^{\frac{m(m+1)}{2} \times m^2}$, formed using a subset of the rows of I_{m^2} that picks the $\frac{m(m+1)}{2}$ independent entries. Thus,

$$
\mathbf{r}_P = A_P\boldsymbol{\gamma} + \mathbf{e}_P,
$$

where $\mathbf{r}_P \triangleq P\mathbf{r}$, $A_P \triangleq PA$, and $\mathbf{e}_P \triangleq P\mathbf{e}$. Further, we approximate the distribution of $\mathbf{e}_P$ by $\mathcal{N}(0, W_P)$, where $W_P = PWP^\top$ and W is the noise covariance matrix derived in the previous section. This Gaussian approximation is motivated from the fact that the noise vector $\mathbf{e}$ is a sum of i.i.d. random vectors, i.e.,

$$
\mathbf{e} = \frac{1}{L}\left(\sum_{i=1}^{L} \mathrm{vec}(\mathbf{y}_i\mathbf{y}_i^\top - \mathbb{E}\mathbf{y}_1\mathbf{y}_1^\top)\right) \triangleq \frac{1}{L}\sum_{i=1}^{L}\mathbf{u}_i.
$$

So, from the central limit theorem, as $L \to \infty$, $\frac{1}{L}\sum_{i=1}^{L}\mathbf{u}_i \xrightarrow{d} \mathcal{N}(0, W)$. Using this, the approximate ML estimate of $\boldsymbol{\gamma}$, which we denote $\boldsymbol{\gamma}_{\mathrm{ML}}$, can be found by solving the following optimization problem:

$$
\boldsymbol{\gamma}_{\mathrm{ML}} = \underset{\boldsymbol{\gamma} \geq 0}{\mathrm{argmax}}\ p(\mathbf{r}_P; \boldsymbol{\gamma}), \tag{10}
$$

where

$$
p(\mathbf{r}_P; \boldsymbol{\gamma}) = \frac{1}{(2\pi)^{\frac{m(m+1)}{4}}|W_P|^{\frac{1}{2}}}e^{-\frac{(\mathbf{r}_P - A_P\boldsymbol{\gamma})^\top W_P^{-1}(\mathbf{r}_P - A_P\boldsymbol{\gamma})}{2}}.
$$

Simplifying (10), we get

$$
\boldsymbol{\gamma}_{\mathrm{ML}} = \underset{\boldsymbol{\gamma} \geq 0}{\mathrm{argmin}}\ \log|W_P| + (\mathbf{r}_P - A_P\boldsymbol{\gamma})^\top W_P^{-1}(\mathbf{r}_P - A_P\boldsymbol{\gamma}).
\tag{11}
$$

The objective function in (11) is nonconvex in $\boldsymbol{\gamma}$ since W_P also depends on $\boldsymbol{\gamma}$, and is difficult to optimize directly. In the next section, we propose a heuristic technique to solve the optimization problem.

3.3. Modified Reweighted Minimization

In this section, we propose a modified reweighted minimization approach to solve (11). We fix W_P, solve the resulting convex non-negative quadratic problem, re-compute W_P using the new $\boldsymbol{\gamma}$, and iterate.

Now, to solve the convex non-negative quadratic program (NNQP)

$$
\underset{\boldsymbol{\gamma} \geq 0}{\mathrm{argmin}}\ (\mathbf{r}_P - A_P\boldsymbol{\gamma})^\top W_P^{-1}(\mathbf{r}_P - A_p\boldsymbol{\gamma}),
$$

we use the iterative technique of [15], which gives the following entry-wise update for $\boldsymbol{\gamma}$ in the $(i+1)^{\mathrm{th}}$ iteration:

$$
\gamma_j^{(i+1)} = \gamma_j^{(i)}\left(\frac{-b_j + \sqrt{b_j^2 + 4(Q^+\gamma^{(i)})_j(Q^-\gamma^{(i)})_j}}{2(Q^+\gamma^{(i)})_j}\right),
$$

where $\mathbf{b} = -A_P^\top W_P^{-1}\mathbf{r}_P$, $Q = A_P^\top W_P^{-1}A_P$, $Q^+ = \max(Q, 0)$, $Q^- = \max(-Q, 0)$, with $\max(A, 0)$ representing the entry-wise maximum of the elements of A and 0.

Thus, our approach is as follows: we approximate the noise covariance W_P by its zeroth-order Taylor expansion around a previous estimate of $\boldsymbol{\gamma}$ and then minimize the resulting cost function over $\boldsymbol{\gamma}$ keeping W_P fixed. This can be viewed as an iterative reweighted minimization [16] technique where we only consider the zeroth order term in the Taylor expansion, since gradient computation is difficult. The steps are summarized in Algorithm 1. Also, the computational complexity of the different steps in the algorithm is summarized in Table 1. Note that the complexity is independent of L, the number of MMVs.

Algorithm 1 Modified Reweighted NNQP (MRNNQP)

1: Input: Measurement matrix Φ, vectorized sample co-variance $\mathbf{r}$, initial value $\boldsymbol{\gamma}^{(0)} = (1, \ldots, 1)^\top$, $\Gamma^{(0)} = \mathrm{diag}(\boldsymbol{\gamma}^{(0)})$, $i = 1$
2: **While** (not converged) **do**
3:     $W_P^{(i)} \leftarrow \frac{1}{L} P(\Phi \otimes \Phi) B(\Gamma^{(i-1)} \otimes \Gamma^{(i-1)})(\Phi \otimes \Phi)^\top P^\top$
4:     $\mathbf{b}^{(i)} \leftarrow -A_P^\top {W_P^{(i)}}^{-1} \mathbf{r}_P$
5:     $Q^{(i)} \leftarrow A_P^\top {W_P^{(i)}}^{-1} A_P$
6:     $\boldsymbol{\gamma}^{(i)} \leftarrow \mathrm{NNQP}(Q^{(i)}, \mathbf{b}^{(i)})$
7:     $\Gamma^{(i)} \leftarrow \mathrm{diag}(\boldsymbol{\gamma}^{(i)})$
8:     $i \leftarrow i + 1$
9: **end While**
10: Output: support of $\boldsymbol{\gamma}^{(i)}$

Table 1. Computational complexity of MRNNQP

Operation	Computational complexity
Computing W_P	$O(m^4 N^2)$
Computing Q	$O(m^2 N^2)$
Computing b	$O(m^4 N)$
Computing W_P^{-1}	$O(m^6)$

4. DISCUSSION

The statistics of the noise term $\mathbf{e}$ depends on L as well as on the parameter $\boldsymbol{\gamma}$ that has to be estimated, as can be seen from (8). As a result of this parameter-dependent noise, the maximum likelihood cost function is nonconvex in $\boldsymbol{\gamma}$ and difficult to optimize. The Co-LASSO algorithm [5], which also uses the sample covariance matrix to estimate $\boldsymbol{\gamma}$, does not account for the statistics of the noise/error arising because of the difference between the true covariance and its finite sample based estimate. Therefore, the algorithm performs well only when L is large, i.e., when the noise term is negligible. As we illustrate in the next section, the proposed algorithm performs well at a much smaller L. Also, under our generative model for the inputs, the $\ell_{1,2}$ penalty algorithm [9] and simultaneous OMP [10] perform poorly in the $k > m$ regime.

Another interesting feature of our algorithm is that the key step, namely, Step 6 in Algorithm 1, involves solving a non negative quadratic program. In particular, no sparsity-promoting penalty is required. Similar observations were made in [17], where the authors note that a non negative least squares program can be used for recovering non negative sparse vectors without explicit sparsity-inducing regularization, under certain conditions on the measurement matrix.

5. SIMULATION RESULTS

For a given set of (N, m, k, L) values, we generate the following: an $m \times N$ measurement matrix Φ with entries $\Phi_{ij} \overset{iid}{\sim}$

(a) $N=40$, $m=20$, $k=25$ (b) $N=70$, $m=20$, $L=50$, 1000
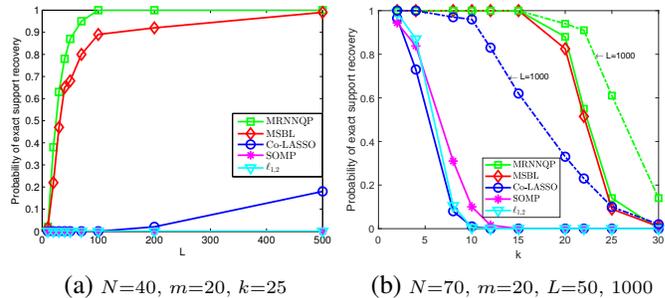
Fig. 1. Support recovery performance of different algorithms.

$\mathcal{N}(0, \frac{1}{m})$, a support $T \subset [N]$ with $|T| = k$ chosen uniformly at random from $\binom{N}{k}$ possibilities, $\gamma \in \{0,1\}^N$ with support T, $\{\mathbf{x}_i\}_{i=1}^L$ drawn independently from $\mathcal{N}(0, \Gamma)$. For each trial, The algorithm is provided with Φ and $\{\mathbf{y}_i\}_{i=1}^L$ generated according to the linear model (1). We run the algorithm 200 times, and a trial is declared successful if the algorithm exactly recovers the true support. The objective value decreases as the iterations proceed and stabilizes after about 20 iterations. Hence, we stop the algorithm after 20 iterations.

Figures $1(a)$ and $1(b)$ show the probability of successful recovery of the proposed algorithm, the Co-LASSO approach from [5], the M-SBL algorithm [7], simultaneous OMP (SOMP) [10], and the $\ell_{1,2}$ penalty algorithm [9], as a function of L and k, respectively. Both the proposed algorithm and M-SBL, which use a maximum likelihood based approach to estimate $\boldsymbol{\gamma}$ show similar performance, with the proposed algorithm performing slightly better in the low L regime. The Co-LASSO approach requires much larger L for reliable support recovery, while SOMP and $\ell_{1,2}$ minimization perform well only in the $k < m$ regime. Thus, our proposed algorithm provides competitive performance with the attractive benefit that its complexity is independent of L, the number of MMVs.

6. CONCLUSIONS

We proposed a novel algorithm to find the common support of a set of k-sparse signals from their noisy underdetermined linear measurements. The algorithm is based on matching the sample covariance of the measurements with the covariance induced by a parametrized Gaussian prior. We proposed a modified iterative reweighted minimization procedure for covariance matching, that uses a non-negative quadratic program as its inner optimization problem. Empirical results suggest that exact support recovery is possible in the $k < m$ regime with L of the order of m and in the $k \geq m$ regime with higher L. Future work could involve convergence analysis of the algorithm and extending the approach to handle other kinds of structure, such as intra- and inter-vector correlation, and slowly varying support.

7. REFERENCES

[1] R. Prasad, C. R. Murthy, and B. D. Rao, "Joint channel estimation and data detection in MIMO-OFDM systems: A sparse Bayesian learning approach," *IEEE Trans. on Sig. Proc.*, vol. 63, no. 20, pp. 5369–5382, Oct. 2015.

[2] M. F. Duarte, S. Sarvotham, D. Baron, M. B. Wakin, and R. G. Baraniuk, "Distributed compressed sensing of jointly sparse signals," in *Asilomar Conference on Signals, Systems and Computers*, Oct. 2005, pp. 1537–1541.

[3] J. A. Bazerque and G. B. Giannakis, "Distributed spectrum sensing for cognitive radio networks by exploiting sparsity," *IEEE Trans. on Sig. Proc.*, vol. 58, no. 3, pp. 1847–1862, Mar. 2010.

[4] J. W. Choi, B. Shim, Y. Ding, B. Rao, and D. I. Kim, "Compressed sensing for wireless communications: Useful tips and tricks," *IEEE Commun. Surveys Tutorials*, vol. PP, no. 99, pp. 1–1, 2017.

[5] P. Pal and P. P. Vaidyanathan, "Pushing the limits of sparse support recovery using correlation information," *IEEE Trans. on Sig. Proc.*, vol. 63, no. 3, pp. 711–726, 2015.

[6] S. Khanna and C. R. Murthy, "Rényi divergence based covariance matching pursuit of joint sparse support," in *Sig. Process. Advances in Wireless Commun. (SPAWC)*, Jul. 2017, pp. 1 – 5.

[7] D. P. Wipf and B. D. Rao, "An empirical bayesian strategy for solving the simultaneous sparse approximation problem," *IEEE Trans. on Sig. Proc.*, vol. 55, no. 7-2, pp. 3704–3716, 2007.

[8] S. Khanna and C. R. Murthy, "On the restricted isometry of the columnwise khatri-rao product," *CoRR*, vol. abs/1709.05789, 2017. [Online]. Available: http://arxiv.org/abs/1709.05789

[9] D. M. Malioutov, M. Çetin, and A. S. Willsky, "A sparse signal reconstruction perspective for source localization with sensor arrays," *IEEE Trans. on Sig. Proc.*, vol. 53, no. 8-2, pp. 3010–3022, 2005.

[10] J. A. Tropp, A. C. Gilbert, and M. J. Strauss, "Simultaneous sparse approximation via greedy pursuit," in *IEEE ICASSP, Philadelphia, Pennsylvania, USA, March*, 2005, pp. 721–724.

[11] O. Balkan, K. Kreutz-Delgado, and S. Makeig, "Localization of more sources than sensors via jointly-sparse bayesian learning," *IEEE Sig. Proc. Letters*, vol. 21, no. 2, pp. 131–134, 2014.

[12] D. P. Wipf and B. D. Rao, "Sparse bayesian learning for basis selection," *IEEE Trans. on Sig. Proc.*, vol. 52, no. 8, pp. 2153–2164, 2004.

[13] C. G. Khatri and C. R. Rao, "Solutions to some functional equations and their applications to characterization of probability distributions," *Sankhya, Series A*, vol. 30, pp. 167–180, 1968.

[14] P. Pal and P. P. Vaidyanathan, "Parameter identifiability in sparse bayesian learning," in *IEEE ICASSP, Florence, Italy*, 2014, pp. 1851–1855.

[15] F. Sha, L. K. Saul, and D. D. Lee, "Multiplicative updates for nonnegative quadratic programming in support vector machines," in *Advances in Neural Information Processing Systems*, 2002, pp. 1041–1048.

[16] E. J. Candès, M. B. Wakin, and S. P. Boyd, "Enhancing sparsity by reweighted ℓ_1 minimization," *Journal of Fourier Analysis and Applications*, vol. 14, no. 5, pp. 267–288, Dec. 2008.

[17] S. Foucart and D. Koslicki, "Sparse recovery by means of nonnegative least squares," *IEEE Signal Process. Lett.*, vol. 21, no. 4, pp. 498–502, 2014.