NON-ASYMPTOTIC GUARANTEES FOR CORRELATION-AWARE SUPPORT DETECTION

Ali Koochakzadeh, and Piya Pal

University of California, San Diego
Electrical and Computer Engineering Department
La Jolla, CA 92037

ABSTRACT

This paper considers the problem of sparse support recovery in Multiple Measurement Vector (MMV) models, where the support size (K) can exceed the dimension (M) of individual measurement vectors. Existing results in this regime mostly establish asymptotic performance guarantees, where the number of measurement vectors $L \to \infty$. In this paper, we develop non-asymptotic guarantees (finite L), and demonstrate that it is possible to recover supports of size $K = \mathcal{O}(M^2)$ provided the sparse signals are statistically uncorrelated. In particular, the probability of detecting a wrong support is shown to approach zero exponentially fast in L even when $K > M$, for appropriately designed measurement matrices. Our analysis is based on a simple least squares estimation of signal powers, followed by hard thresholding to detect the support.

Index Terms— Joint Support Recovery, Multiple Measurement Vectors, Khatri-Rao Product, Sparse Bayesian Learning, Correlation.

1. INTRODUCTION

Consider a set of jointly sparse signals $\mathbf{x}[l] \in \mathbb{F}^N$ (for $l = 0, \cdots, L-1$)[1] whose nonzero elements are indexed by the set $\mathcal{S} = \{s_1, \cdots, s_K\}$, i.e.,

$$x_i[l] \neq 0 \Leftrightarrow i \in \mathcal{S}.$$

The goal in sparse support recovery is to identify $\mathcal{S}$ from the measurements

$$\mathbf{y}[l] = \mathbf{A}\mathbf{x}[l] + \mathbf{w}[l]. \quad (1)$$

where $\mathbf{A} \in \mathbb{F}^{M \times N}$, $\mathbf{x}[l] \in \mathbb{F}^N$, $\mathbf{w}[l] \in \mathbb{F}^M$. This problem naturally arises in many signal processing applications, especially when the support itself contains meaningful physical information. They include Directions-of-Arrival of far field electromagnetic waves [1, 2], detection of unoccupied frequency bands in cognitive radio [3], compressed DNA microarrays for bio-sensing [4].

The problem of sparse support recovery (with or without estimating $\mathbf{x}[l]$) has been widely studied in compressed sensing literature [5, 6, 7, 8, 9, 10]. Both necessary and sufficient conditions for support recovery have been established for single measurement vector ($L = 1$) as well as multiple measurement vector (MMV) models ($L > 1$). These results indicate that in SMV models $M = \Omega(K \log N)$ measurements are necessary and sufficient for accurate support detection. However, a common feature of most of these results is that the sparse signal $\mathbf{x}[l]$ is modeled as a (unknown) deterministic quantity and statistical priors on $\mathbf{x}[l]$ (such as its correlation structure) are not fully exploited. In contrast, Tang et al. [11] considers a MMV model with statistically uncorrelated signals and derives both upper and lower bounds on the probability of error. Although the (non zero) signals are assumed to be uncorrelated, the derivation of the upper bound does not fully exploit this structure. Hence, their results do not guarantee successful support detection when $K > M$.

In recent work [12], using the same signal model as [11], we showed that it is possible to recover supports of size $K = \mathcal{O}(M^2)$ for appropriate measurement matrices, as long as the non-zero signals have equal power and the detector knows K. In this paper, we relax both conditions and show that is possible to recover supports of size $K = \mathcal{O}(M^2)$ even for sources with unequal power, and without the knowledge of K. Unlike [11, 12], we do not impose specific distribution on the measurements and only assume them to be bounded real-valued random variables. Using a simple least squares estimate of the source powers, followed by hard-thresholding, we are able to recover sparse supports of size $K > M$, with overwhelming probability (with respect to L).

2. SIGNAL MODEL

In this paper, we consider the MMV model introduced in (1) with L measurement vectors. We make the following statistical assumptions on the signal and noise:

(A1) Non-zero elements of the signal $\mathbf{x}[l]$ are uncorrelated, i.e. $\mathbb{E}(\mathbf{x}[l]\mathbf{x}[l]^H) = \mathbf{P}$, where $\mathbf{P} =$

This work was supported in parts by the NSF CAREER award (ECCS 1553954), and the University of California, San Diego.

[1] Depending on the context $\mathbb{F}$ can be the field of real $\mathbb{R}$ or complex $\mathbb{C}$ numbers.

$\operatorname{diag}(p_1, \cdots, p_N)$ is a diagonal matrix, and $\{\mathbf{x}[l]\}_{l=0}^{L-1}$ are independent and identically distributed (i.i.d.) random vectors. Moreover $p_i = 0, i \neq \mathcal{S}$, consistent with the fact that the L vectors $\mathbf{x}[l], 0 \leq l \leq L-1$ share a common support $\mathcal{S}$.

(A2) Signal $\mathbf{x}[l]$ and noise $\mathbf{w}[l]$ are uncorrelated, i.e, $\mathbb{E}(\mathbf{x}[l]\mathbf{w}[l]^H) = \mathbf{0}$.

(A3) The noise $\mathbf{w}$ is white, i.e. $\mathbb{E}(\mathbf{w}[l]\mathbf{w}[l]^H) = \sigma^2\mathbf{I}$, and $\{\mathbf{w}[l]\}_{l=0}^{L-1}$ are i.i.d. random variables. We assume σ^2 is known.

(A4) The signal and noise are *bounded random variables*, i.e., $\|\mathbf{x}[l]\|_2 \leq C_x$, $\|\mathbf{w}[l]\|_2 \leq C_w$, where $C_x, C_w > 0$ are positive constants.

(A5) The measurement matrix $\mathbf{A}$ satisfies $\operatorname{rank}(\mathbf{A}^* \odot \mathbf{A}) = N$.

Assumptions (**A1-A3**) are typical in the context of Sparse Bayesian Learning (SBL) [13], line spectrum estimation and so forth. However, unlike SBL, (**A4**) further enforces the signal and noise to be bounded random variables. This assumption simplifies the error analysis of our proposed detector in the regime $K > M$, and ensures that the error decays exponentially fast in L. Unlike SBL, we do not consider any particular distribution for the measurements. The following remark immediately follows from the assumption (**A4**):

Remark 1. *Under assumption (**A4**), we have* $\|\mathbf{y}\| \leq C_y$

$$C_y = \sigma_{\max}(\mathbf{A})C_x + C_w \tag{2}$$

where $\sigma_{\max}(\mathbf{A})$ *denotes the maximum singular value of the measurement matrix* $\mathbf{A}$.

Remark 2. *Based on assumptions (**A1-A3**), one can write the covariance matrix of the measurement vectors as*

$$\mathbf{R} := \mathbb{E}(\mathbf{y}[l]\mathbf{y}[l]^H) = \mathbf{A}\mathbf{P}\mathbf{A}^H + \sigma^2\mathbf{I}.$$

The vectorized form of the covariance matrix can be written as

$$\operatorname{vec}(\mathbf{R}) = (\mathbf{A}^* \odot \mathbf{A})\mathbf{p} + \sigma^2 \operatorname{vec}(\mathbf{I})$$

where $\mathbf{p} = [p_1, \cdots, p_N]^T$ is a sparse vector with support $\mathcal{S}$. The goal of support recovery in MMV models is to detect the common support $\mathcal{S}$ from the measurements $\mathbf{Y} = [\mathbf{y}[0], \cdots, \mathbf{y}[L-1]]$. Let $\psi(\mathbf{Y})$ denote a detector that returns a candidate support. The probability of detecting a wrong support, given $\mathcal{S}$ is the true support, can be expressed as

$$p_{e|\mathcal{S}} = \mathbb{P}(\psi(\mathbf{Y}) \neq \mathcal{S}|\mathcal{S})$$

It has been empirically demonstrated that SBL is capable of detecting supports of size larger than M, but no theoretical guarantees exist. In this paper, we propose a simple detector $\psi_{\text{LS}}(\mathbf{Y})$ (that *does not know the support size* K), and compute upper bounds on the probability of error $p_{e|\mathcal{S}}$ of this detector. Before presenting our results, we review existing results that consider support recovery in the regime $K > M$ but only provide partial guarantees.

3. REVIEW OF CORRELATION-AWARE TECHNIQUES FOR RECOVERING SUPPORTS OF SIZE $K = \mathcal{O}(M^2)$

In compressed sensing, existing guarantees for sparse support recovery are mostly relevant in the regime $K < M$. The only algorithms, which, under certain restrictive assumptions, theoretically or experimentally show possibility of recovering supports of size $K > M$, are Sparse Bayesian Learning [13, 14], and Correlation-Aware LASSO (Co-LASSO) [15]. We now briefly review these results and elaborate more on the role of correlation awareness in recovering supports of size $K = \mathcal{O}(M^2)$.

1. *Sparse Bayesian Learning:* The authors in [14] show that the MSBL algorithm is capable of recovering supports of size $K > M$ under the following assumptions: 1) The measurements are assumed to be noiseless. 2) Non-zero rows of $\mathbf{X} = [\mathbf{x}[0], \mathbf{x}[1], \cdots, \mathbf{x}[L-1]]$ are orthogonal. Although these conditions may not be satisfied in practice, their numerical results show that even under a noisy setting MSBL is able to recover supports of size $K > M$.

2. *Correlation-Aware Support Recovery:* In our earlier work in [15], we showed that if we have access to the exact covariance matrix $\mathbf{R}$ (which happens when $L \to \infty$), then, under assumptions **(A1-A3)** and **(A5)**, it is possible to recover sparse supports of size $K = \mathcal{O}(M^2)$, by solving the following ℓ_1 minimization problem:

$$\min_{\mathbf{p} \geq \mathbf{0}} \|\mathbf{p}\|_1 \quad \text{subject to} \quad (\mathbf{A}^* \odot \mathbf{A})\mathbf{p} = \operatorname{vec}(\mathbf{R})$$

For finite L, we can only compute an estimate of $\mathbf{R}$. In this case, we proposed a variation of LASSO [15] namely (Co-LASSO) for joint support recovery, and showed that it can recover $\mathcal{S}$ as long as $K < \frac{1}{2}(1 + \frac{1}{\mu^2})$. Here $\mu \leq 1$ is the mutual coherence of $\mathbf{A}$ defined as

$$\mu = \max_{i \neq j} \frac{|\mathbf{a}_i^H \mathbf{a}_j|}{\|\mathbf{a}_i\|_2 \|\mathbf{a}_j\|_2}$$

This result showed that by merely exploiting the lack of correlation between sparse signals, one can recover larger supports compared to traditional coherence-based guarantees in compressed sensing (which require $K < \frac{1}{2}(1 + \frac{1}{\mu})$) [16]. However, in presence of finite L, these guarantees are rather weak and only apply in the regime $K < M$.

3. *Existence of Cramér Rao Bound, when $N = \mathcal{O}(M^2)$:* In past work [17], we showed that the Cramér Rao Bound (CRB) for estimating source powers in a MMV model (1) exist, even when $K = \mathcal{O}(M^2)$, as long as $\mathrm{rank}(\mathbf{A} \odot \mathbf{A}) = N$. This condition is obeyed by almost all choices of $\mathbf{A}$ if $N \leq \frac{M^2+M}{2}$. In this setting, as $L \to \infty$, the CRB goes to zero at the rate $1/L$. Since Maximum Likelihood (ML) Estimates asymptotically attain the CRB, this automatically shows that MSBL can recover the vector $\mathbf{p}$ as $L \to \infty$ since it solves a maximum likelihood problem.

Most of aforementioned results provide asymptotic guarantees (i.e. when $L \to \infty$). No non-asymptotic guarantees currently exist for support recovery in the regime $K = O(M^2)$ that can ensure $p_{e|\mathcal{S}}$ decays exponentially fast in L. In the next section, we will address this issue by proposing a detector which is based on solving a simple least squares problem followed by a thresholding step.

4. A LEAST SQUARES THRESHOLDING BASED SUPPORT DETECTOR

We propose the following simple detector based on least-squares method:

$$\psi(\mathbf{Y}; \tau, \mathbf{A}, \sigma^2) = \{i | \hat{p}_i \geq \tau, \hat{\mathbf{p}} = \phi(\mathbf{Y}; \mathbf{A}, \sigma^2)\} \tag{3}$$

where τ is a predefined threshold, and

$$\phi(\mathbf{Y}; \mathbf{A}, \sigma^2) = (\mathbf{A}^* \odot \mathbf{A})^\dagger \,\mathrm{vec}(\hat{\mathbf{R}} - \sigma^2 \mathbf{I}) \tag{4}$$

is the least square estimator of the vector of source powers $\mathbf{p}$, where $\hat{\mathbf{R}}$ denotes the sample covariance matrix, defined as

$$\hat{\mathbf{R}} = \frac{1}{L} \sum_{l=1}^{L} \mathbf{y}[l] \mathbf{y}[l]^H \tag{5}$$

Inspite of its simplicity, we will now show that this detector can recover supports of size $K = \mathcal{O}(M^2)$ with overwhelming probability. [2] We first state some preliminary lemmas:

Lemma 1. *The estimator* (4) *is unbiased, i.e* $\mathbb{E}(\hat{\mathbf{p}}) = \mathbf{p}$.

Proof. Let $\hat{\mathbf{p}} = \phi(\mathbf{Y}; \mathbf{A}, \sigma^2)$. We have

$$\begin{aligned}
\mathbb{E}(\hat{\mathbf{p}}) &= \mathbb{E}((\mathbf{A}^* \odot \mathbf{A})^\dagger \,\mathrm{vec}(\hat{\mathbf{R}} - \sigma^2 \mathbf{I})) \\
&= (\mathbf{A}^* \odot \mathbf{A})^\dagger \,\mathrm{vec}(\mathbb{E}(\hat{\mathbf{R}}) - \sigma^2 \mathbf{I})) \\
&= (\mathbf{A}^* \odot \mathbf{A})^\dagger \,\mathrm{vec}(\mathbf{R} - \sigma^2 \mathbf{I})) \qquad (6) \\
&= (\mathbf{A}^* \odot \mathbf{A})^\dagger \,\mathrm{vec}(\mathbf{A}\mathbf{P}\mathbf{A}^H) \\
&= (\mathbf{A}^* \odot \mathbf{A})^\dagger (\mathbf{A}^* \odot \mathbf{A}) \mathbf{p} \\
&= \mathbf{p} \qquad (7)
\end{aligned}$$

where (6) follows from the fact that $\mathbb{E}(\hat{\mathbf{R}}) = \frac{1}{L} \sum_{l=1}^{L} \mathbb{E}(\mathbf{y}[l]\mathbf{y}[l]^H) = \mathbf{R}$, and (7) holds due to assumption **(A5)**. □

Lemma 2. *The estimator* (4) *can be also be written as*

$$\hat{p}_i = \frac{1}{L} \sum_{l=0}^{L-1} \sum_{j=1}^{N} b_{ij} (|\mathbf{a}_j^H \mathbf{y}[l]|^2 - \sigma^2 \|\mathbf{a}_j\|^2) \tag{8}$$

where $\mathbf{B} := [b_{ij}] = \left((\mathbf{A}^* \odot \mathbf{A})^H (\mathbf{A}^* \odot \mathbf{A})\right)^{-1}$.

Proof. Following the definition of matrix $\mathbf{B}$, one can write the estimator (3) as

$$\begin{aligned}
\hat{\mathbf{p}} &= \mathbf{B} (\mathbf{A}^* \odot \mathbf{A})^H \,\mathrm{vec}(\hat{\mathbf{R}} - \sigma^2 \mathbf{I}) \\
&= \mathbf{B}\mathbf{J}^H (\mathbf{A}^* \otimes \mathbf{A})^H \,\mathrm{vec}(\hat{\mathbf{R}} - \sigma^2 \mathbf{I}) \\
&= \mathbf{B}\mathbf{J}^H \,\mathrm{vec}(\mathbf{A}^H (\hat{\mathbf{R}} - \sigma^2 \mathbf{I}) \mathbf{A}) \\
&= \mathbf{B} \,\mathrm{diag}(\mathbf{A}^H (\hat{\mathbf{R}} - \sigma^2 \mathbf{I}) \mathbf{A}) \qquad (9) \\
&= \frac{1}{L} \sum_{l=0}^{L-1} \mathbf{B} \,\mathrm{diag}(\mathbf{A}^H (\mathbf{y}[l]\mathbf{y}[l]^H - \sigma^2 \mathbf{I}) \mathbf{A}) \qquad (10)
\end{aligned}$$

where $\mathbf{J} \in \mathbb{R}^{N^2 \times N}$ is an appropriate column selection matrix, $\mathrm{diag}(\mathbf{X})$ (with matrix argument $\mathbf{X}$) returns a column vector containing the diagonal entries of the matrix $\mathbf{X}$. The equation (9) follows by exploiting the structure of the matrix $\mathbf{J}$, and (10) follows from the definition of $\hat{\mathbf{R}}$ in (5) and changing the order of summations. □

To facilitate our analysis, for the rest of this paper, we will further assume that $\mathbb{F} = \mathbb{R}$, i.e., all random variables and the measurement matrix $\mathbf{A}$ are real valued.

Lemma 3. *Given any* $i \in \{1, \cdots, N\}$ *and* $\eta > 0$*, it holds that*

$$\mathbb{P}(|\hat{p}_i - p_i| > \eta) \leq 2e^{-\beta_i L \eta^2}$$

where β_i *is a constant (specified in the proof).*

Proof. Using the result of Lemma 2, we can write

$$\hat{p}_i = \frac{1}{L} \sum_{l=0}^{L-1} (z_l^{(i)} - \hat{e}_i)$$

where $\hat{e}_i = \sum_{j=1}^{N} b_{ij} \sigma^2 \|\mathbf{a}_j\|^2$, and

$$z_l^{(i)} = \sum_{j=1}^{N} b_{ij} |\mathbf{a}_j^H \mathbf{y}[l]|^2.$$

[2] Although the MMV model is underdetermined ($N > M$), under assumption (**A5**), $\mathbf{A}^* \odot \mathbf{A}$ is tall and has full column-rank. Hence it is reasonable to estimate $\mathbf{p}$ using least squares method. Assumption (**A5**) continues to hold in the regime $M < N < (M^2 + M)/2$ for almost all $\mathbf{A} \in \mathbb{R}^{M \times N}$, and it serves as a necessary condition for existence of CRB [15, 17].

Next, we show that each $|z_l^{(i)}|$ is bounded. We have

$$|z_l^{(i)}| \leq \sum_{j=1}^{N} |b_{ij}||\mathbf{a}_j^H \mathbf{y}[l]|^2 \tag{11}$$

$$\leq \sum_{j=1}^{N} |b_{ij}|\|\mathbf{a}_j\|^2 \|\mathbf{y}[l]\|^2 \tag{12}$$

$$\leq C_y^2 \sum_{j=1}^{N} |b_{ij}|\|\mathbf{a}_j\|^2 := C_z^{(i)} \tag{13}$$

From Lemma 1 we know that $\mathbb{E}(\hat{p}_i) = p_i$. Therefore, using Hoeffding Inequality [18], we obtain

$$\mathbb{P}(|\hat{p}_i - p_i| > \eta) \leq 2e^{-\frac{L\eta^2}{2(C_z^{(i)})^2}}$$

which concludes the proof by choosing $\beta_i = \frac{1}{2(C_z^{(i)})^2}$. □

Equipped with the above lemmas, we are now ready to state our main result:

Theorem 1. *Under assumptions* **(A1-A5)***, the probability of error* $p_{e|\mathcal{S}}$ *of the detector* (3) *with* $\tau = \frac{p_{\min}}{2}$ *is upper bounded as*

$$p_{e|\mathcal{S}} \leq e^{-\beta p_{\min}^2 L/4 + \log(2N))}$$

where $p_{\min} := \min_{i \in \mathcal{S}} p_i$*, and* $\beta = \min_i \frac{1}{2(C_z^{(i)})^2}$*, with* $C_z^{(i)}$ *given by* (13).

Proof. For the detector specified by (3), consider any threshold τ such that $\tau < p_{\min}$. In this case, the probability of detecting a wrong support (given $\mathcal{S}$ is the true support) can be written as

$$p_{e|\mathcal{S}} = \mathbb{P}(\psi(\mathbf{Y}; \tau, \mathbf{A}, \sigma^2) \neq \mathcal{S}|\mathcal{S})$$

$$= \mathbb{P}\left(\bigcup_{i \in \mathcal{S}} \{\hat{p}_i < \tau\} \cup \bigcup_{i \notin \mathcal{S}} \{\hat{p}_i > \tau\}\right) \tag{14}$$

$$\leq \sum_{i \in \mathcal{S}} \mathbb{P}(\hat{p}_i < \tau) + \sum_{i \notin \mathcal{S}} \mathbb{P}(\hat{p}_i > \tau) \tag{15}$$

$$\leq \sum_{i \in \mathcal{S}} \mathbb{P}(|\hat{p}_i - p_i| > p_i - \tau) + \sum_{i \notin \mathcal{S}} \mathbb{P}(|\hat{p}_i| > \tau) \tag{16}$$

where (15) follows from the union bound, and (16) follows from the fact that $\hat{p}_i \leq \tau$ is equivalent to $\hat{p}_i - p_i \leq \tau - p_i$, which implies $|\hat{p}_i - p_i| \geq p_i - \tau$ [3] Using Lemma 3, we have

$$p_{e|\mathcal{S}} \leq \sum_{i \in \mathcal{S}} 2e^{-\beta_i L(p_i - \tau)^2} + \sum_{i \notin \mathcal{S}} 2e^{-\beta_i L \tau^2}$$

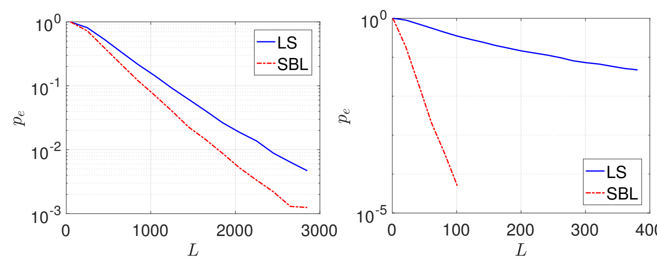
$$\leq 2Ke^{-\beta L(p_{\min} - \tau)^2} + 2(N - K)e^{-\beta L \tau^2}$$

where $\beta = \min_i \beta_i$. Substituting $\tau = \frac{p_{\min}}{2}$ concludes the proof. □

[3] Since $\tau < p_{\min}$, we have $\tau - p_i < 0$ for all $i \in \mathcal{S}$

5. SIMULATIONS

We now numerically validate that it is possible to obtain exponentially decaying probability of error for support recovery in the regime $K > M$. To this end, we consider two algorithms: i) MSBL [14], and ii) the proposed detector in (3). We consider a fixed measurement matrix $\mathbf{A} \in \mathbb{R}^{M \times N}$, $M = 7, N = 21$. The ith nonzero element of $\mathbf{x}[l]$ is chosen from the uniform distribution over $[-\sqrt{3p_i}, \sqrt{3p_i}]$ (which will ensure that $\mathbb{E}(x_i^2[l]) = p_i$). The elements of the noise vector $w_i[l]$ are i.i.d. and uniformly distributed in the range $[-\sqrt{3}\sigma, \sqrt{3}\sigma]$. We let $p_i = 1$, for $i \in \mathcal{S}$, and $\sigma = 0.1$. For the proposed least square detector, we set the threshold $\tau = \frac{1}{2}$. We use the same threshold τ to detect the support using MSBL. Fig. 1 shows the probability of error of both detectors as a function of L (log scale), for both $K > M$ and $K < M$. It is clear that the slope for both detectors is linear in L, indicating an exponential decay of $p_{e|\mathcal{S}}$ with respect to L. It can also be seen that MSBL has a better error exponent compared to the least squares detector, in both the regimes. It will be of interest in future to analyze the performance of MSBL, and characterize this error exponent.



(a) $K = 17 > M, N = 21$ (b) $K = 6 < M, N = 21$

Fig. 1. Probability of error of both detectors ("LS" denotes the proposed least squares detector, and "SBL" denotes the detector based on Sparse Bayesian Learning algorithm.)

6. CONCLUSION

In this paper, we considered the problem of joint support recovery of sparse signals in multiple measurement vector (MMV) models. For the first time, we provided non-asymptotic guarantees for recovering supports of size $K = \mathcal{O}(M^2)$, where M is the size of each measurement vector. Our detector is based on a simple least square estimator of source powers, followed by a hard thresholding operation. Assuming the sparse signals to be statistically uncorrelated bounded random variables, we can ensure that the probability of detecting a wrong support approaches zero exponentially fast in L even when $K > M$. This result holds for appropriately designed measurement matrices whose Khatri-Rao products satisfy certain rank constraints.

7. REFERENCES

[1] Dmitry Malioutov, Müjdat Cetin, and Alan S Willsky, "A sparse signal reconstruction perspective for source localization with sensor arrays," *IEEE transactions on signal processing*, vol. 53, no. 8, pp. 3010–3022, 2005.

[2] Dongyang Xu, Nan Hu, Zhongfu Ye, and Ming Bao, "The estimate for doas of signals using sparse recovery method," in *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on*. IEEE, 2012, pp. 2573–2576.

[3] Zhi Tian and Georgios B Giannakis, "Compressed sensing for wideband cognitive radios," in *Acoustics, Speech and Signal Processing, 2007. ICASSP 2007. IEEE International Conference on*. IEEE, 2007, vol. 4, pp. IV–1357.

[4] Mona A Sheikh, Shriram Sarvotham, Olgica Milenkovic, and Richard G Baraniuk, "Dna array decoding from nonlinear measurements by belief propagation," in *Statistical Signal Processing, 2007. SSP'07. IEEE/SP 14th Workshop on*. IEEE, 2007, pp. 215–219.

[5] Alyson K Fletcher, Sundeep Rangan, and Vivek K Goyal, "Necessary and sufficient conditions for sparsity pattern recovery," *IEEE Transactions on Information Theory*, vol. 55, no. 12, pp. 5758–5772, 2009.

[6] Martin J Wainwright, "Information-theoretic limits on sparsity recovery in the high-dimensional and noisy setting," *IEEE Transactions on Information Theory*, vol. 55, no. 12, pp. 5728–5741, 2009.

[7] Kamiar Rahnama Rad, "Nearly sharp sufficient conditions on exact sparsity pattern recovery," *IEEE Transactions on Information Theory*, vol. 57, no. 7, pp. 4672–4679, 2011.

[8] Mehmet Akçakaya and Vahid Tarokh, "Shannon-theoretic limits on noisy compressive sampling," *IEEE Transactions on Information Theory*, vol. 56, no. 1, pp. 492–504, 2010.

[9] Cem Aksoylar, George K Atia, and Venkatesh Saligrama, "Sparse signal processing with linear and nonlinear observations: A unified shannon-theoretic approach," *IEEE Transactions on Information Theory*, vol. 63, no. 2, pp. 749–776, 2017.

[10] Sangjun Park, Nam Yul Yu, and Heung-No Lee, "An information-theoretic study for joint sparsity pattern recovery with different sensing matrices," *IEEE Transactions on Information Theory*, 2017.

[11] Gongguo Tang and Arye Nehorai, "Performance analysis for sparse support recovery," *IEEE transactions on information theory*, vol. 56, no. 3, pp. 1383–1399, 2010.

[12] Ali Koochakzadeh and Piya Pal, "On fundamental limits of correlation aware sparse support recovery," *Submitted to IEEE transactions on signal processing*, 2017.

[13] David P Wipf and Bhaskar D Rao, "Sparse bayesian learning for basis selection," *IEEE Transactions on Signal processing*, vol. 52, no. 8, pp. 2153–2164, 2004.

[14] Ozgur Balkan, Kenneth Kreutz-Delgado, and Scott Makeig, "Localization of more sources than sensors via jointly-sparse bayesian learning," *IEEE Signal Processing Letters*, vol. 21, no. 2, pp. 131–134, 2014.

[15] Piya Pal and PP Vaidyanathan, "Pushing the limits of sparse support recovery using correlation information," *IEEE Transactions on Signal Processing*, vol. 63, no. 3, pp. 711–726, 2015.

[16] Robert Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 267–288, 1996.

[17] Piya Pal and PP Vaidyanathan, "Parameter identifiability in sparse bayesian learning," in *Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on*. IEEE, 2014, pp. 1851–1855.

[18] Wassily Hoeffding, "Probability inequalities for sums of bounded random variables," *Journal of the American statistical association*, vol. 58, no. 301, pp. 13–30, 1963.