Localization of More Sources Than Sensors via Jointly-Sparse Bayesian Learning

Article *in* IEEE Signal Processing Letters · February 2014

DOI: 10.1109/LSP.2013.2294862

Ozgur Balkan
Swartz Center for Computational Neuroscience

Ken Kreutz-Delgado
University of California, San Diego

Scott Makeig
University of California, San Diego

Localization of More Sources Than Sensors via Jointly-Sparse Bayesian Learning

Ozgur Balkan*, *Student Member, IEEE,* Kenneth Kreutz-Delgado, *Fellow, IEEE,* and Scott Makeig

***Abstract*—We analyze the jointly-sparse signal recovery problem in the regime where the number of sources k is larger than the number of measurements M. We show that the support set of sources can still be recovered with sparse bayesian learning (M-SBL) even if $k \geq M$. We provide sufficient conditions on the dictionary and sources which theoretically guarantee support set recovery in the noiseless case of M-SBL. We validate our sufficient conditions with experiments and also demonstrate that M-SBL outperforms M-CoSaMP, the algorithm recently used to localize more sources than sensors. Finally, we experimentally show robustness of the approach in the presence of noise.**

***Index Terms*—Simultaneous sparse approximation, Sparse Bayesian Learning, Source Localization**

I. Introduction

SPARSE signal recovery from limited number of measurements has found a large number of applications in diverse fields of engineering including but not limited to neural networks, data compression and biomedical source localization [1]. Typically, the underlying signal model is

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

where $\mathbf{y} \in \mathbb{R}^{M\times 1}$ is the observed vector, $\mathbf{A} = [\mathbf{a_1} \ldots \mathbf{a_N}] \in \mathbb{R}^{M\times N}$ is the known overcomplete dictionary $(M < N)$, $\mathbf{x} \in \mathbb{R}^{N\times 1}$ is the unknown sparse source vector (number of nonzero values $k < M$), and $\mathbf{e} \in \mathbb{R}^{M\times 1}$ is the noise vector. The goal is to find the original source vector $\mathbf{x}$ from $\mathbf{y}$ given the dictionary $\mathbf{A}$. Depending on the application, it is usually sufficient to find the set of nonzero indices of the vector $\mathbf{x}$, which we denote as the support set S. This corresponds to finding the locations of sources in array processing problems. Accuracy of localization is enhanced with the use of multiple signals that are assumed to be jointly sparse [2]. In the jointly-sparse model the source matrix $\mathbf{X} \in \mathbb{R}^{N\times L}$ consists of $k < M$ rows that are nonzero. The data matrix $\mathbf{Y} \in \mathbb{R}^{M\times L}$ is obtained similarly as

$$\mathbf{Y} = \mathbf{A}\mathbf{X} + \mathbf{E} \tag{2}$$

where $\mathbf{E} \in \mathbb{R}^{M\times L}$ is the noise matrix. In this problem, L different time samples are collected consecutively in time and the nonzero rows of the source matrix $\mathbf{X}$ carry the time courses of the k sources that are assumed to be active in that time window. The goal is to recover the indices of these nonzero rows from data $\mathbf{Y}$.

Previous efforts to solve the jointly-sparse problem [3], [4], [5] have assumed that the number of sources k is smaller than the number of measurements M, as generally all other joint sparse solvers do [6]. Recently, it was observed that $k < M$ is not a necessary condition, and more sources than sensors can be localized using M-CoSaMP [7], [8].

In this paper, we show that multiple measurement sparse bayesian learning (M-SBL) [9] can also recover the support set in the noiseless case when $M \leq k \leq N$. Moreover, we provide exact recovery conditions for M-SBL in that regime, conditions which are missing from previous reports on M-CoSaMP and other algorithms. We also experimentally show that M-SBL outperforms the suggested M-CoSaMP algorithm under non-ideal conditions. We assume that mixing at the sensors is instantaneous (no time delay between sources and sensors) and the environment is anechoic. Therefore, we use the time domain representation of the problem instead of the frequency domain.

The sufficient conditions for exact support recovery for M-SBL in the regime $k \geq M$ are twofold. The first condition requires the orthogonality of the active sources, that is $\mathbf{X_S}\mathbf{X_S^T} = \mathbf{\Lambda}$, where $\mathbf{\Lambda}$ is any diagonal matrix and $\mathbf{X_S} \in \mathbb{R}^{k\times L}$ is the matrix of active sources. This condition is not very far from ideal when the sources are independent, such as the case of audio sources or brain sources [10], and given enough snapshots L, we will then have $\mathbf{X_S}\mathbf{X_S^T} \approx \mathbf{\Lambda}$. The second condition imposes a constraint on the sensing dictionary $\mathbf{A}$. For deterministic dictionaries, this condition is easier to check than the spark, or RIP constraints [11], [12]. For random dictionaries, it is equivalent to a dictionary size constraint $N \leq \frac{M(M+1)}{2}$.

The outline of the paper is as follows: In Section II, we summarize the M-SBL algorithm. In Section III, we provide a theoretical analysis of guaranteed support recovery for any k in the noiseless case. We perform noiseless and noisy experiments in Section IV and provide conclusions in Section V.

II. M-SBL

In the M-SBL framework, the i-th row of $\mathbf{X}$, denoted as $\mathbf{x_{i.}}$, has an L-dimensional zero mean gaussian prior whose variance is controlled by a hyperparameter γ_i. That is,

$$p(\mathbf{x_{i.}};\gamma) \triangleq \mathcal{N}(0, \gamma_i \mathbf{I}). \tag{3}$$

This work was supported by The Swartz Foundation (Old Field, NY). O. Balkan and K. Kreutz-Delgado are with the Department of Electrical and Computer Engineering, University of California San Diego, La Jolla CA, 92093 USA e-mail: (obalkan@ucsd.edu, kreutz@eng.ucsd.edu). S. Makeig is with Swartz Center for Computational Neuroscience at UCSD email: (smakeig@ucsd.edu).

We also have

$$\mathbf{p}(\mathbf{y_{.j}}|\mathbf{x_{.j}}) \triangleq \mathcal{N}(\mathbf{A}\mathbf{x_{.j}}, \sigma^{\mathbf{2}}\mathbf{I}) \tag{4}$$

$$\mathbf{p}(\mathbf{Y}|\mathbf{X}) = \prod_{j=1}^{L} \mathbf{p}(\mathbf{y_{.j}}|\mathbf{x_{.j}}) \tag{5}$$

assuming gaussian noise with known variance σ^2, with $\mathbf{x_{.j}}$ denoting the j-th column of $\mathbf{X}$. Integrating out the unknown sources $\mathbf{X}$, we derive $p(\mathbf{Y};\gamma)$, the marginal likelihood of data $\mathbf{Y}$ given the hyperparameters $\gamma \in \mathbb{R}^N$, which is to be maximized. We apply the $-2\log(\cdot)$ transformation to arrive at the M-SBL cost function to be minimized [9], [13], [14],

$$\begin{aligned}\mathcal{L}(\gamma) &\triangleq -2\log(p(\mathbf{Y};\gamma)) = -2\log\int p(\mathbf{Y}|\mathbf{X})p(\mathbf{X};\gamma)d\mathbf{X}\\ &\equiv \log|\Sigma| + \frac{1}{L}\sum_{t=1}^{L}\mathbf{y_{.t}}^T\Sigma^{-1}\mathbf{y_{.t}}\end{aligned} \tag{6}$$

with $\Sigma \triangleq (\mathbf{A}\boldsymbol{\Gamma}\mathbf{A}^{\mathbf{T}} + \sigma^{\mathbf{2}}\mathbf{I})$, $\Gamma \triangleq \mathrm{diag}(\gamma)$. A common method to reach a local minimum of (6) is to use the EM procedure [9], however the updates may be slow depending on the size of the problem. Reference [13] provides the following fixed point update that is both fast and guaranteed to decrease the likelihood function at every step,

$$\gamma_i^{(k+1)} = \frac{\gamma_i^{(k)}}{\sqrt{\mathbf{a_i^T}(\Sigma^{(k)})^{-1}\mathbf{a_i}}}\frac{\|\mathbf{Y^T}(\Sigma^{(k)})^{-1}\mathbf{a_i}\|_2}{\sqrt{n}} \tag{7}$$

Note that with this update scheme, for $\gamma^{(0)} \geq 0$, the converged solution is nonnegative, $\gamma^* \geq 0$. M-SBL can also learn the noise parameter σ^2, however we do not discuss it in this paper since our theoretical analysis considers the the noiseless case, where $\sigma^2 \to 0$. After convergence, support set $\hat{S}$ is extracted from the solution $\hat{\gamma}$ by $\hat{S} = \{i, \hat{\gamma}_i \neq 0\}$.

In the noiseless case, references [9] and [13] show that exact sparse reconstruction is guaranteed if the number of active sources $k < M$ and the active set of sources $\mathbf{X_S}$ are sample-wise uncorrelated. In the next section, we show that exact noiseless source localization (support set recovery) with M-SBL is also guaranteed in the regime $k \geq M$ with an additional constraint on the dictionary.

III. Analysis

A. Local Minima

First, we summarize the local minima analysis carried out in [13]. We start by letting $\sigma^2 \to 0$ (as appropriate for the noiseless case) and obtain the limiting cost function,

$$\mathcal{L}(\gamma) = L\log|\mathbf{A}\boldsymbol{\Gamma}\mathbf{A^T}| + \sum_{t=1}^{L}\mathbf{y_{.t}}^T(\mathbf{A}\boldsymbol{\Gamma}\mathbf{A^T})^{-1}\mathbf{y_{.t}} \tag{8}$$

We denote the matrix of the support set columns by $\mathbf{A_S}$. Under the condition (i), we construct the following cost function around the neighborhood of a local minimum γ^* as in [13],

$$\begin{aligned}\mathcal{L}(\alpha,\beta) = L\log|\alpha\mathbf{A}\boldsymbol{\Gamma}^*\mathbf{A^T} + \beta\mathbf{U}\mathbf{U^T}| + \\ \sum_{t=1}^{L}\mathbf{y_{.t}^T}(\alpha\mathbf{A}\boldsymbol{\Gamma}^*\mathbf{A^T} + \beta\mathbf{U}\mathbf{U^T})^{-1}\mathbf{y_{.t}},\end{aligned} \tag{9}$$

where $\mathbf{U} = \mathbf{A_S}\boldsymbol{\Lambda}^{\frac{1}{2}}$ and $\boldsymbol{\Gamma}^* = \mathrm{diag}(\gamma^*)$. Note that the overall covariance has the structural form,

$$\exists\,\hat{\gamma} \in \mathbb{R}^N \quad \text{such that} \quad \alpha\mathbf{A}\boldsymbol{\Gamma}^*\mathbf{A^T} + \beta\mathbf{U}\mathbf{U^T} = \mathbf{A}\hat{\boldsymbol{\Gamma}}\mathbf{A^T} \tag{10}$$

where $\hat{\boldsymbol{\Gamma}} = \mathrm{diag}(\hat{\gamma})$. We put the constraints $\alpha, \beta \geq 0$ so that the elements of $\hat{\gamma}$ are guaranteed to be nonnegative. Note that when $\alpha = 1 and \beta = 0$, $\hat{\gamma} = \gamma^*$. For $\alpha = 1$ and for small $\beta > 0$, (i.e., when we add a small contribution to the total covariance from $\mathbf{U}$), we expect the cost function not to decrease since γ^* is by definition a local minima of (8). Thus the following conditions must hold,

$$\left.\frac{\partial\mathcal{L}(\alpha,\beta)}{\partial\alpha}\right|_{\alpha=1,\beta=0} = 0 \qquad \left.\frac{\partial\mathcal{L}(\alpha,\beta)}{\partial\beta}\right|_{\alpha=1,\beta=0} \geq 0 \tag{11}$$

The derivative of the cost w.r.t β at $\beta = 0$ is not necessarily zero since $\beta \geq 0$ due to the nonnegativity constraint that must be satisfied for the elements of $\hat{\gamma}$.

Using the orthogonality constraint $\mathbf{X_S}\mathbf{X_S}^{\mathbf{T}} = \Lambda$, the second term in (9) can be expressed as,

$$\begin{aligned}&\sum_{t=1}^{L}\mathbf{y_{.t}^T}(\alpha\mathbf{A}\boldsymbol{\Gamma}^*\mathbf{A^T} + \beta\mathbf{U}\mathbf{U^T})^{-1}\mathbf{y_{.t}}\\ &= \mathrm{tr}(\mathbf{Y}^T(\alpha\mathbf{A}\boldsymbol{\Gamma}^*\mathbf{A^T} + \beta\mathbf{U}\mathbf{U^T})^{-1}\mathbf{Y})\\ &= \mathrm{tr}(\mathbf{X_S^T}\boldsymbol{\Lambda}^{-\frac{1}{2}}\mathbf{U^T}(\alpha\mathbf{A}\boldsymbol{\Gamma}^*\mathbf{A^T} + \beta\mathbf{U}\mathbf{U^T})^{-1}\mathbf{U}\boldsymbol{\Lambda}^{-\frac{1}{2}}\mathbf{X_S})\\ &= \mathrm{tr}(\mathbf{U^T}(\alpha\mathbf{A}\boldsymbol{\Gamma}^*\mathbf{A^T} + \beta\mathbf{U}\mathbf{U^T})^{-1}\mathbf{U}\boldsymbol{\Lambda}^{-\frac{1}{2}}\mathbf{X_S}\mathbf{X_S^T}\boldsymbol{\Lambda}^{-\frac{1}{2}})\\ &= \mathrm{tr}(\mathbf{U^T}(\alpha\mathbf{A}\boldsymbol{\Gamma}^*\mathbf{A^T} + \beta\mathbf{U}\mathbf{U^T})^{-1}\mathbf{U}).\end{aligned} \tag{12}$$

B. Exact Support Recovery Conditions

Extending the results described in [9], [13], we carry out an analysis for the case where $k \geq M$. First, we define a function $f : \mathbb{R}^{M\times N} \to \mathbb{R}^{M(M+1)/2\times N}$:

Definition. $\mathbf{B} = f(\mathbf{A})$ means that the i-th column of $\mathbf{B}$ is given by $\mathbf{b_i} = \mathrm{vech}(\mathbf{a_i}\mathbf{a_i}^{\mathbf{T}})$. Vech is the half-vectorization function, which results in a column vector obtained by vectorizing only the lower triangular part of a matrix.

Theorem 1. Given a dictionary $\mathbf{A}$ and a set of observed signals $\mathbf{Y}$, M-SBL recovers the support set of any size k exactly in the noiseless setting, if the following conditions hold.

(i) The active sources $\mathbf{X_S}$ are orthogonal. Namely, $\mathbf{X_S}\mathbf{X_S}^{\mathbf{T}} = \Lambda$, Λ is a diagonal matrix.

(ii) $\mathrm{rank}(f(\mathbf{A})) = N$.

Proof. Assume that $k \geq M$[1]. We will show that the minimum of (6) is unique if above conditions hold. Moreover, the solution γ^* must have positive values only at indices $i \in S$, thus recovering the support set.

Carrying out the differentiation for the first condition of (11) gives,

$$\mathrm{tr}\left(\mathbf{U^T}(\mathbf{A}\boldsymbol{\Gamma}^*\mathbf{A^T})^{-1}\mathbf{U}\right) = LM. \tag{13}$$

Similarly, for the second condition in (11) we have,

$$\left.\frac{\partial\mathcal{L}(\alpha,\beta)}{\partial\beta}\right|_{\alpha=1,\beta=0} = \sum_{i=1}^{M}\left(L\lambda_i - \lambda_i^2\right) \geq 0, \tag{14}$$

[1]The case $k < M$ is described in reference [9].

where λ_i is the i-th eigenvalue of $\mathbf{U^T}(\mathbf{A\Gamma^*A^T})^{-1}\mathbf{U}$. Note that $\lambda_i \geq 0$ since $\mathbf{U^T}(\mathbf{A\Gamma^*A^T})^{-1}\mathbf{U}$ is a positive-semidefinite matrix of size $M \times M$ due to the nonnegativity of γ^*. Combining (13) and (14) we get,

$$\begin{aligned}\left.\frac{\partial \mathcal{L}(\alpha,\beta)}{\partial \beta}\right|_{\alpha=1,\beta=0} &= L^2M - \sum_{i=1}^{M}\lambda_i^2 \\ &\leq L^2M - \sum_{i=1}^{M}\bar{\lambda}^2 \\ &= L^2M - L^2M = 0, \end{aligned} \tag{15}$$

with $\bar{\lambda} = \frac{1}{M}\sum_{i=1}^{M}\lambda_i = LM/M = L$.

The inequalities (14) and (15) together imply $\left.\frac{\partial \mathcal{L}(\alpha,\beta)}{\partial \beta}\right|_{\alpha=1,\beta=0} = 0$ and thus the eigenvalues all satisfy $\lambda_1 = \lambda_2 = \ldots = \lambda_M = L$.

We proceed by showing the following: If $\lambda_1 = \lambda_2 = \ldots = \lambda_M = L$, then we must have $\mathbf{A\Gamma^*A^T} = \frac{1}{L}\mathbf{UU^T}$. This is clear after we write $\mathbf{U^T}(\mathbf{A\Gamma^*A^T})^{-1}\mathbf{UV} = L\mathbf{V}$, where $\mathbf{V}$ is a $k \times M$ matrix of M eigenvectors. This equality is equivalent to $\left(\mathbf{A\Gamma^*A^T}\right)^{-1}\mathbf{UV} = L(\mathbf{UU^T})^{-1}\mathbf{UV}$, which leads to

$$\mathbf{A\Gamma^*A^T} = \frac{1}{L}\mathbf{UU^T} = \frac{1}{L}\mathbf{A_S\Lambda A_S^T} = C. \tag{16}$$

where C is a constant matrix.

We now show that there is exactly one solution $\mathbf{\Gamma}^*$ that satisfies (16) if (ii) holds. Denoting the j-th column of $\mathbf{A_S}$ as $\mathbf{A_S}_j$, a particular solution is,

$$\gamma_i^* = \begin{cases} \frac{1}{L}\mathbf{\Lambda}_{jj}, & \text{if } i \in S, \text{where } \mathbf{a_i} = \mathbf{A_S}_j \\ 0, & \text{otherwise} \end{cases}. \tag{17}$$

Denoting $\mathbf{B} = f(\mathbf{A})$ as defined above, and vectorizing both sides of (16), we get,

$$\begin{aligned} \text{vech}\Big(\sum_{i=1}^{N}\gamma_i^*\mathbf{a_i}\mathbf{a_i}^\mathbf{T}\Big) &= \text{vech}(C) \\ \text{or}\quad \mathbf{B}\gamma^* &= \text{vech}(C) \end{aligned} \tag{18}$$

Finally we observe that if (ii) is satisfied, the null space of $\mathbf{B}$ is trivial and thus (18) has a unique solution.

□

C. Remarks

Theoretically, condition (i) of Theorem 1 is satisfied in the limit $L \to \infty$. However, for large L, we can get very close to sample-wise orthogonality ($\mathbf{X_S X_S^T} \approx \mathbf{\Lambda}$) for independent sources, which is adequate for exact recovery in practical applications. The fact that perfect, asymptotic sample-wise orthogonality is not a hard requirement for good performance is demonstrated in the experiments.

The second condition can be related to the size of a random dictionary. It can be seen that for (ii) to be satisfied for a random dictionary $\mathbf{A}$, we must have $N \leq M(M+1)/2$. For deterministic dictionaries, the same equality is a necessary condition and can be satisfied even if the columns of $\mathbf{A}$ are coherent. Using this inequality we can estimate the maximum resolution of the source space given the number of sensors M, and vice versa. For instance, in an EEG source localization problem if the desired resolution is $N = 10000$ voxels on the cortex, at least $M = 142$ sensors are needed for condition (ii) to be satisfied. However, as we see in the experiments below the conditions of Theorem 1 are not necessary conditions.

IV. Experiments

A. Theorem Validation

In this section, we validate Theorem 1 by experimenting with source activations that satisfy condition (i). For this purpose, we create k random source activations and perform a whitening transformation using eigenvalue decomposition to ensure that $\mathbf{X_S X_S^T} = \mathbf{I}$ at each trial. We create random dictionaries of various sizes and for each trial we randomly choose different support sets (source locations). We assign the number of sources to the values $k = N/4$ and $k = N/2$, which ensures that for almost all of the parameter settings we have more sources than sensors. We create the data matrix as $\mathbf{Y} = \mathbf{AX}$. We perform 200 iterations of M-SBL and select the k largest elements of the converged hyperparameter vector γ^* as the support set. For a single trial, a success ratio of source localization is calculated as,

$$r = \frac{|s^* \cap s|}{k} \tag{19}$$

where s^* is the resulting estimated support set and s is the true support set. After performing 100 trials for each parameter setting, we calculate the mean support recovery ratio and plot it in Figure 1.

B. Performance Comparison

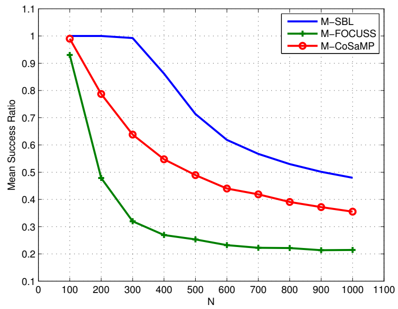
In this part, we compare the performance of M-SBL to M-CoSaMP and M-FOCUSS [2] when the conditions (i) and (ii) of Theorem 1 are violated. We create random sources of length $L = 1024$. The sources are independent yet the outer product is not exactly diagonal ($\mathbf{X_S X_S^T} \neq \mathbf{\Lambda}$). The random dictionaries we create also violate (ii), namely $N \geq M(M+1)/2$.

Figure 2(a) shows the source recovery performance of all algorithms with varying resolution number N. The number of sensors $M = 20$ is fixed, and $k = N/5$. Figure 2(b), shows the performance of the algorithms with fixed M, N but varying number of sources k. In both cases we see that M-SBL outperforms the other algorithms. In Figure 2(c), we see the positive effect of having more snapshots L, on the recovery of locations for M-SBL, as $\mathbf{X_S X_S^T}$ approaches to a diagonal matrix.

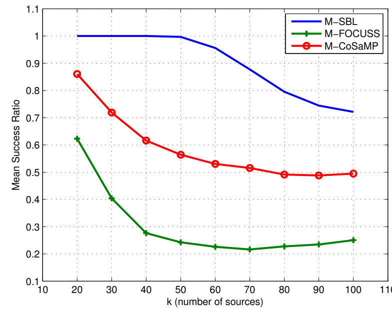
Finally, we test the algorithms under various noise levels. Although our analysis is based on the noiseless case, we observe that M-SBL still outperforms M-CoSaMP in low noise as well as high noise scenarios. See Table I.

V. Conclusion

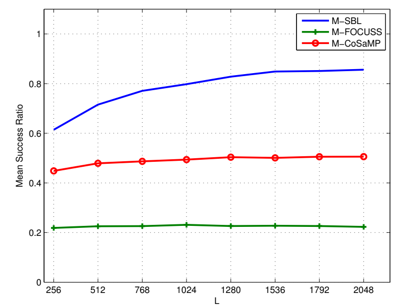
In this paper, we showed that it is possible to perform localization (support recovery) of more sources than sensors using M-SBL applied directly on the time domain data. We theoretically provided sufficient conditions for exact recovery, conditions which were missing for previously suggested joint



(a) Changing resolution N, $M = 20, k = N/5, L = 1024$.

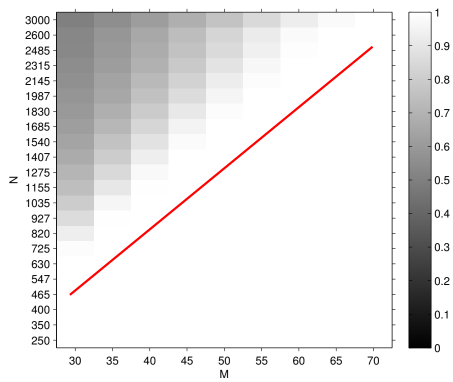


(b) Varying the number of sources k, $M = 20, N = 500, L = 1024$.

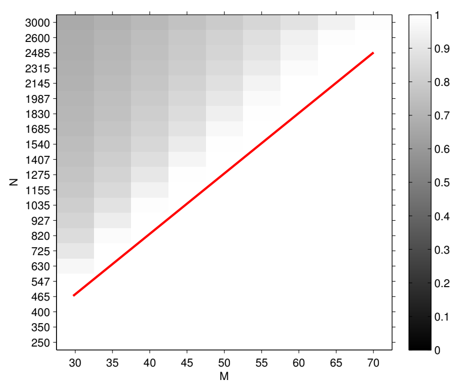


(c) Varying the source duration L. Other parameters are fixed to $M = 20, N = 500, k = 80$.

Fig. 2. Comparison of algorithms under different parameter settings when the guaranteed recovery conditions for M-SBL do not hold.



(a) $k = N/4$.



(b) $k = N/2$.

Fig. 1. Validation of Theorem 1. Sample-wise orthogonal sources. The line corresponds to dictionary sizes where $N = M(M+1)/2$. As the theorem suggests, below the line we have guaranteed support recovery. Perfect recovery is also observed above the line, also showing that conditions of Theorem 1 are not necessary conditions.

TABLE I
MEAN SUCCESS RATIO UNDER DIFFERENT NOISE LEVELS.
$M = 20, N = 500, k = 100, L = 100$

SNR (dB)	M-CoSaMP	M-SBL
10	0.4601	0.5962
20	0.5018	0.8159
30	0.5028	0.9172
Inf	0.5068	0.9335

sparse algorithms in the case of more sources than sensors [8], [2]. Our experiments also demonstrated the superior performance of M-SBL when the conditions of guaranteed recovery of Theorem 1 do not hold.

REFERENCES

[1] R.G. Baraniuk, E. Candes, M. Elad, and Y. Ma, "Applications of sparse representation and compressive sensing [scanning the issue]," *Proceedings of the IEEE*, vol. 98, no. 6, pp. 906–909, 2010.

[2] Shane F. Cotter, Bhaskar D. Rao, Kjersti Engan, Kenneth Kreutz-delgado, and Senior Member, "Sparse solutions to linear inverse problems with multiple measurement vectors," *IEEE Trans. Signal Processing*, pp. 2477–2488, 2005.

[3] David P. Wipf, Julia P. Owen, Hagai Attias, Kensuke Sekihara, and Srikantan S. Nagarajan, "Estimating the location and orientation of complex, correlated neural activity using meg," in *NIPS*, 2008, pp. 1777–1784.

[4] Karl Friston, Lee Harrison, Jean Daunizeau, Stefan Kiebel, Christophe Phillips, Nelson Trujillo-Barreto, Richard Henson, Guillaume Flandin, and Jrmie Mattout, "Multiple sparse priors for the m/eeg inverse problem," *NeuroImage*, vol. 39, no. 3, pp. 1104 – 1120, 2008.

[5] David P. Wipf, Julia P. Owen, Hagai T. Attias, Kensuke Sekihara, and Srikantan S. Nagarajan, "Robust bayesian estimation of the location, orientation, and time course of multiple correlated neural sources using meg," *NeuroImage*, vol. 49, no. 1, pp. 641 – 655, 2010.

[6] A. Rakotomamonjy, "Surveying and comparing simultaneous sparse approximation (or group-lasso) algorithms," *Signal Processing*, vol. 91, no. 7, pp. 1505 – 1526, 2011.

[7] Petros T Boufounos, Paris Smaragdis, and Bhiksha Raj, "Joint sparsity models for wideband array processing," in *SPIE Optical Engineering+ Applications*. International Society for Optics and Photonics, 2011, pp. 81380K–81380K.

[8] Jonathan Le Roux, Petros T Boufounos, Kang Kang, and John R Hershey, "Source localization in reverberant environments using sparse optimization," *ICASSP*, 2013.

[9] D.P. Wipf and B.D. Rao, "An empirical bayesian strategy for solving the simultaneous sparse approximation problem," *Signal Processing, IEEE Transactions on*, vol. 55, no. 7, pp. 3704 –3716, july 2007.

[10] Arnaud Delorme, Jason Palmer, Julie Onton, Robert Oostenveld, and Scott Makeig, "Independent eeg sources are dipolar," *PLoS ONE*, vol. 7, 02 2012.

[11] Michael Elad, *Sparse and redundant representations*, Springer, 2010.

[12] A.S. Bandeira, E. Dobriban, D.G. Mixon, and W.F. Sawin, "Certifying the restricted isometry property is hard," *Information Theory, IEEE Transactions on*, vol. 59, no. 6, pp. 3448–3450, 2013.

[13] D.P. Wipf, *Bayesian Methods For Finding Sparse Representations*, Ph.D. thesis, UC San Diego, 2006.

[14] Michael E Tipping, "Sparse bayesian learning and the relevance vector machine," *The Journal of Machine Learning Research*, vol. 1, pp. 211–244, 2001.