PARAMETER IDENTIFIABILITY IN SPARSE BAYESIAN LEARNING

Piya Pal and P. P. Vaidyanathan

Dept. of Electrical Engineering, MC 136-93
California Institute of Technology, Pasadena, CA 91125, USA
E-mail: piyapal@caltech.edu, ppvnath@systems.caltech.edu

ABSTRACT

The problem of parameter identifiability in linear underdetermined models is addressed, where the observed data vectors follow a multivariate Gaussian distribution. The problem is underdetermined because the dimension of parameters characterizing the distribution of the data is larger than the dimension of the observed vectors. Such models arise frequently in Bayesian Compressive sensing and Sparse Bayesian Learning problems, where the parameter vector to be estimated, is assumed to be sparse. We establish explicit conditions for parameter identifiability in such models, by relating the ambient dimension of the hyperparameter space and that of the data. We establish a crucial result that in such underdetermined models, even without requiring the parameter to be sparse, it is possible to guarantee unique identifiability of the parameters as long as these two dimensions satisfy a certain condition. When such a condition is violated, the unconstrained statistical model is no more identifiable and additional constraints in the form of sparsity need to be enforced to recover the true parameter.

***Index Terms* —** Parameter Identifiability, Cramer Rao Bound, Maximum Likelihood Technique, Sparse Bayesian Learning.

1. INTRODUCTION

Consider a linear underdetermined model

$$\mathbf{y} = \mathbf{A}\mathbf{x}_0 + \mathbf{n} \tag{1}$$

where $\mathbf{A} \in \mathbb{R}^{M\times N}$ is a *known* measurement matrix and $\mathbf{x}_0 \in \mathbb{R}^{N\times 1}, N \gg M$ is a random Gaussian vector with zero mean and a diagonal covariance matrix $\mathbf{\Gamma}_0 \in \mathcal{R}^{N\times N}$, which has $D < N$ non zero elements. The noise term $\mathbf{n}$ follows $\mathcal{N}(\mathbf{0}, \sigma_n^2\mathbf{I})$ and is usually assumed uncorrelated from $\mathbf{x}$. This model lies at the heart of a family of Bayesian techniques for Compressive sensing [3], the most popular among them being Sparse Bayesian Learning (SBL) [1, 2, 4, 5]. SBL computes a MAP estimate of $\mathbf{x}_0$ from multiple realizations of the data $\mathbf{y}$ by *learning* the sparse hyperparameter $\mathbf{\Gamma}_0$ of its distribution. It provides a Bayesian alternative to the conventional l_1 minimization based approaches to sparse signal reconstruction.

Theoretical analysis of the SBL framework has been the subject of a number of papers [2, 4, 5]. The problem of hyperparameter learning is generally formulated as a Maximum Likelihood Estimation problem, which is highly non convex. However, under suitable conditions, SBL can provably recover the true $\mathbf{\Gamma}_0$ when it is sparse [4, 5]. These guarantees are algorithmic in nature. They firstly establish conditions under which the SBL cost function exhibits unique global minimum at the true parameter value, and then provide guarantees that Expectation Maximization (EM) based algorithms can indeed converge to the true solution. However, there has been limited statistical analysis of the SBL framework. In [7], a Cramer Rao bound analysis for parameter estimation using the SBL signal model is performed. More recently, a hypothesis testing approach has been proposed in [10], using which the authors have derived explicit expressions for asymptotic Mean Squared Error (MSE) of estimation, at least for large data size.

A common characteristic of the existing results for SBL is that the guarantees are restricted to the case when the number (D) of non zero elements satisfies $D \leq M$, and they do not extend easily to the case when $D > M$. Further, the problem of parameter identifiability [14] has not been systematically addressed so far. In this paper, we address this important question under the common theme of *parameter identifiability* in the SBL framework. The primary goal of most SBL-based approaches is to form a MAP estimate of $\mathbf{x}_0$ and the recovery of $\mathbf{\Gamma}_0$ is only a secondary step to achieve this. We distinguish between these two estimation problems and explicitly state conditions under which $\mathbf{\Gamma}_0$ is identifiable, although $\mathbf{x}_0$ may not be. We prove that the hyperparameter $\mathbf{\Gamma}_0$ is identifiable as long as N and M follow an implicit relation, via the so-called Khatri-Rao product of the measurement matrix $\mathbf{A}$. We further demonstrate that the Fisher Information matrix is non singular in this regime and hence the Maximum Likelihood (ML) algorithm for hyperparameter learning asymptotically attains the Cramer Rao Bound and recovers the true $\mathbf{\Gamma}_0$. The remarkable fact is that in this case, the model is identifiable for both $D \leq M$ and $D > M$, i.e. the level of sparsity can be potentially **larger** than the dimension of the data vectors. However, when the conditions for identifiability are violated, the FIM becomes singular and the ML algorithm for SBL can no longer uniquely recover the true $\mathbf{\Gamma}_0$. We show that in this case, additional constraints on the parameter (such as sparsity) need to be enforced to uniquely identify the true solution. To this end, we propose a l_1 minimization based technique to recover the true $\mathbf{\Gamma}_0$ among the multiple global minima of the SBL objective.

The paper is organized as follows. In Sec. 2, we discuss parameter identifiability in generic underdetermined models without explicit sparsity constraints and establish conditions under which the FIM is non singular. In Sec. 3, we use these results to evaluate the performance of SBL and derive explicit relations between the dimensions of hyperparameter space and the data, such that SBL can provably recover the true $\mathbf{\Gamma}_0$. When the model becomes non identifiable, it is possible to use extra information in form of sparsity to recover the true parameter. Section 4 establishes the validity of our proposed results via a number of numerical experiments.

Notations: Boldface Uppercase and Lowercase letters respectively represent matrices and vectors. The symbol $\odot$ represents Khatri-Rao product whereas $\otimes$ represent Kronecker

Work supported in parts by the ONR grant N00014-11-1-0676, and the California Institute of Technology.

product. For a matrix $\mathbf{A}$, $|\mathbf{A}|$, $\text{Tr}(\mathbf{A})$ and $\text{KRank}(\mathbf{A})$ represent its determinant, trace and Kruskal-Rank respectively.

2. IDENTIFABILITY IN THE NON-SPARSE MODEL

In this section, we consider the linear Gaussian underdetermined observation model given by (1). In our subsequent analysis, we will assume that σ_n^2 is known. Our goal is to study conditions on M, N and $\mathbf{A}$ under which it is possible to estimate unknown parameters $\mathbf{\Theta}_0$ of the pdf $p_{\mathbf{Y}}(y; \mathbf{\Theta})$ using the measurement $\mathbf{y}$. Obviously, the parameter $\mathbf{\Theta}$ to be estimated depends on what we consider unknown in characterizing the pdf of $\mathbf{y}$. In this regard, we will distinguish between two cases

1. **Type I Estimation:** In this case $\mathbf{x}_0$ is treated as an unknown deterministic quantity to be estimated, and hence $\mathbf{\Theta} = \mathbf{x}_0$. The pdf of $\mathbf{y}$ is given by $p_{\mathbf{Y}}^{(I)}(\mathbf{y}) = \mathcal{N}(\mathbf{A}\mathbf{x}, \sigma_n^2\mathbf{I})$. We describe the associated CRB as Type I CRB or as conditional CRB, as it is more popularly known [9].

2. **Type II Estimation:** In this case, we assume $\mathbf{x}_0$ to be a random vector drawn from the multivariate Gaussian distribution $\mathcal{N}(\mathbf{0}, \mathbf{\Gamma}_0)$ ($\mathbf{\Gamma}_0$ being a diagonal matrix). We obtain $p_{\mathbf{Y}}(\mathbf{y})$ by marginalizing with respect to the pdf of $\mathbf{x}_0$ as

$$p_{\mathbf{Y}}^{(II)}(\mathbf{y}; \mathbf{\Gamma}_0) = \mathcal{N}(\mathbf{0}, \mathbf{A}\mathbf{\Gamma_0}\mathbf{A}^H + \sigma_n^2\mathbf{I}) \tag{2}$$

Naturally, the unknown parameter to be estimated is given by $\mathbf{\Theta} = \mathbf{\Gamma}_0$.

2.1. Identifiability and the relations between M and N

The identifiability of a parameter $\mathbf{\Theta}$ is defined as

Definition 1. *(see, for example, [14]) The parameter $\mathbf{\Theta}$ is said to be identifiable if $p_{\mathbf{Y}}(\mathbf{y}; \mathbf{\Theta}_1) \neq p_{\mathbf{Y}}(\mathbf{y}; \mathbf{\Theta}_2), \quad \forall \mathbf{\Theta}_1 \neq \mathbf{\Theta}_2$.*

This immediately proves the following lemma for the Type I estimation problem, by observing when two different values of $\mathbf{x}$ can map to same $\mathbf{A}\mathbf{x}$:

Lemma 1. *The parameter $\mathbf{x}$ in Type I estimation problem is non identifiable if $N > M$.*

The observation model for $N > M$ is said to be *underdetermined*, reflecting the fact that multiple values of $\mathbf{x}$ can lead to the same $\mathbf{y}$. The non identifiability of $\mathbf{x}$ implies that **there exists no consistent estimator for** $\mathbf{x}$. However, this underdetermined model frequently arises in sparse estimation problems. The problem of identifiability is alleviated in this case by *imposing constraints* on the parameter $\mathbf{x}$. In particular, it is assumed that $\mathbf{x}$ is sparse, i.e., only D out of N elements of $\mathbf{x}$ is non zero and typically $D \leq M$. Thus the *effective dimension of the non zero components of $\mathbf{\Theta}$ is reduced* by this assumption, rendering the constrained model identifiable under suitable conditions on $\mathbf{A}$. We will revisit the role played by sparsity again in Sec. 3.

For underdetermined models ($N > M$), a constraint on the parameter vector $\mathbf{\Theta}$ (e.g., in the form of sparsity) seems to be necessary for Type I Estimation to be meaningful. However, is the same true for Type II estimation, where we marginalize the unknown $\mathbf{x}$? The answer, surprisingly, is no. This is proved by the following theorem

Theorem 1. *The parameterization given by $\mathbf{\Gamma}_0 :\rightarrow p_{\mathbf{Y}}^{(II)}(\mathbf{y}; \mathbf{\Gamma}_0)$ is identifiable if $N = \text{Rank}(\mathbf{A} \odot \mathbf{A})$.*

Proof. If the model given by $p_{\mathbf{Y}}^{(II)}(\mathbf{y}; \mathbf{\Gamma})$ is non-identifiable, then $\mathbf{A}\mathbf{\Gamma}_1\mathbf{A}^T = \mathbf{A}\mathbf{\Gamma}_2\mathbf{A}^T$ for some $\mathbf{\Gamma}_1 \neq \mathbf{\Gamma}_2$. This can be rewritten in the vectorized form as $\Big(\mathbf{A} \odot \mathbf{A}\Big)(\gamma_1 - \gamma_2) = \mathbf{0}$ where γ represents the vector containing the diagonal elements of $\mathbf{\Gamma}$. Since $N = \text{Rank}(\mathbf{A} \odot \mathbf{A})$, $\mathbf{A} \odot \mathbf{A}$ is full column rank, implying $\gamma_1 = \gamma_2$, contradicting our assumption that the model is non identifiable. □

Defining $\mathbf{A}_{KR} \triangleq \mathbf{A} \odot \mathbf{A}$, it has been shown that for suitable $\mathbf{A}$, $\text{Rank}(\mathbf{A}_{KR}) = O(M^2)$ [13]. Hence N can be as large as $O(M^2)$ and yet, the model remains identifiable. This result shows that when the parameter of interest is $\mathbf{\Gamma}$, the underdetermined model is **identifiable even without imposing any sparsity constraint** on the unknown parameter. This serves as an important distinction between Type I and Type II estimation in the underdetermined model.

2.2. Cramer Rao Bound For the Underdetermined Model

The non-singularity of the Fisher Information matrix plays an important role in deciding if it is possible to obtain unbiased estimator of $\mathbf{\Theta}$ with finite variance from the model $p_{\mathbf{Y}}(\mathbf{y}; \mathbf{\Theta})$. The following theorem establishes the conditions under which Fisher Information matrix corresponding to $p_{\mathbf{Y}}^{(II)}(\mathbf{y}; \mathbf{\Theta})$ becomes singular:

Theorem 2. *The Fisher Information matrix $\mathbf{J}^{(II)}(\mathbf{\Gamma})$ characterizing the pdf $p_{\mathbf{Y}}^{(II)}(\mathbf{y}; \mathbf{\Gamma})$ is given by*

$$\mathbf{J}^{(II)}(\mathbf{\Gamma}) = \mathbf{A}_{KR}^T(\mathbf{R_{yy}}^{-T} \otimes \mathbf{R_{yy}}^{-1})\mathbf{A}_{KR}$$

It is singular iff $N > \text{Rank}(\mathbf{A}_{KR})$.

Proof. The closed form for the Fisher Information Matrix follows from the derivations of CRB for uncorrelated sources in Appendix A of [8]. The matrix $\mathbf{R_{yy}}$ is full rank and hence the rank of $\mathbf{R}^{-T} \otimes \mathbf{R}^{-1}$ is M^2. Hence the rank of $\mathbf{J}^{(II)}(\mathbf{\Gamma}) \in \mathcal{R}^{N \times N}$ is equal to the rank of $\mathbf{A}_{KR} \in \mathcal{R}^{M^2 \times N}$ which is rank deficient iff $N > \text{Rank}(\mathbf{A}_{KR})$ □

In other words, even for an underdetermined model with $N = O(M^2)$, the Fisher Information matrix continues to be non singular as long as $N = \text{Krank}(\mathbf{A}_{KR})$. The Cramer-Rao bound thereby exists and imposes a lower bound on the MSE of any unbiased estimator for $\mathbf{\Gamma}_0$. This also proves the existence of asymptotically (in the number of observed vectors) consistent estimators (in particular, the Maximum Likelihood estimator that asymptotically attains the CRB) in the underdetermined model.

The discussion so far shows that in underdetermined models, without introducing any constraint on the parameter space, it is still possible to obtain non-singular unconstrained FIM. This indicates that an unconstrained Maximum Likelihood Estimator (MLE) can successfully recover the parameter $\mathbf{\Gamma}$ even when $N > M$. We now focus on a special class of algorithms, popularly known as Sparse Bayesian Learning (SBL) that forms the basis of Bayesian Compressive Sensing [3–6], and evaluate its performance with this newfound perspective.

3. WHEN DOES SBL WORK?

Sparse Bayesian learning considers the same signal model as the Type II estimation problem, however, we assume that the true $\mathbf{\Gamma_0}$ characterizing the pdf of $\mathbf{x}_0$ has only D non zero elements on its diagonal. This also implies the resulting $\mathbf{x}_0$ is D-sparse. Using the Gaussian prior on $\mathbf{x}_0$, it computes a MAP estimate of $\mathbf{x}_0$, which explicitly depends on the hyperparameter $\mathbf{\Gamma}_0$ characterizing the pdf of $\mathbf{x}_0$. Since, $\mathbf{\Gamma}_0$ is unknown, the SBL framework estimates $\mathbf{\Gamma}_0$ as an intermediate step by solving the Type II ML estimation problem [1, 4] (assuming L i.i.d realizations of $\mathbf{y}$):

$$\min_{\mathbf{\Gamma}} \log|\mathbf{A}\mathbf{\Gamma}\mathbf{A}^H + \sigma_n^2\mathbf{I}| + \frac{1}{L}\sum_{l=1}^{L}\mathbf{y}_l^T\left(\mathbf{A}\mathbf{\Gamma}\mathbf{A}^H + \sigma_n^2\mathbf{I}\right)^{-1}\mathbf{y}_l \quad (3)$$

In SBL literature, no discussion so far exists to distinguish the cases when $\mathbf{x}_0$ is identifiable, and when only $\mathbf{\Gamma}_0$ is identifiable (although $\mathbf{x}$ may not be). In fact, the Type II ML estimate (of $\mathbf{\Gamma}_0$) is used as a crucial intermediate step to compute the MAP estimate of $\mathbf{x}_0$ and the sparsity of $\mathbf{x}_0$ is solely determined by the sparsity of $\mathbf{\Gamma}_0$. The performance analysis of SBL has been the focus of a body of recent work [4, 5]. They establish conditions under which (3) recovers the true $\mathbf{\Gamma}_0$. *However, the standing assumption in all these results is :* $D < M$. The analysis framework does not easily extend to the case when $D > M$. We now aim to understand the behavior of SBL when D can be potentially much larger than M, by considering two regimes: $N = \text{Rank}(\mathbf{A}_{KR})$ and $N > \text{Rank}(\mathbf{A}_{KR})$:

3.1. The case when $M < N = \textbf{Rank}(\mathbf{A}_{KR})$

It is to be noted that (3) is the standard ML estimation problem for $\mathbf{\Gamma}$ and it does not explicitly impose any sparsity enforcing constraint on $\mathbf{\Gamma}$. So the performance of (3) can be analyzed using standard statistical signal processing tools, at least in the regime when the Fisher Information matrix is non singular. The following theorem [15] characterizes the asymptotic performance of the ML estimate, solving (3).

Theorem 3. *Suppose the data $\{\mathbf{y}_l\}_{l=1}^L$ is generated from the pdf (2) characterized by the parameter $\mathbf{\Gamma}_0$ with D out of N non zero elements. If $N = Rank(\mathbf{A}_{KR})$, then, the solution $\mathbf{\Gamma}_{ML}(L)$ to (3) satisfies:*

1. *Consistency:* $\lim_{L\to\infty}\mathbf{\Gamma}_{ML}(L) = \Gamma_0$
2. *Asymptotic Efficiency: Asymptotically in L, $\mathbf{\Gamma}_{ML}(L)$ is unbiased and attains the Cramer-Rao Bound.*

The preceding result establishes that $\mathbf{\Gamma}_{ML}(L)$ indeed recovers the true $\mathbf{\Gamma}_0$ for large L. In particular, it also shows that the objective in (3) becomes unimodal with $L \to \infty$, as long as $N = \text{Rank}(\mathbf{A}_{KR})$. However, the same *cannot be claimed for finite* L. In fact, for finite L, the objective in (3), in general, can have multiple local and global minima and the success of any algorithm will depend on initial conditions. However, we now consider a special case where even for finite L, one can guarantee recovery of the true $\mathbf{\Gamma}$ by a modified version of (3). The special case assumes $\mathbf{X}_L$ has *orthogonal rows* and absence of noise. A proof can be found in [15]. It is to be noted [5] also uses a similar assumption on $\mathbf{X}_L$ to prove that their implementation of SBL can recover the true $\mathbf{\Gamma}_0$. However that result only applies to the case $D \leq M$. In contrast, our result holds for $D > M$ as long as $N = \text{Rank}(\mathbf{A}_{KR})$

Theorem 4. *[15] Consider the noiseless MMV model $\mathbf{Y}_L = \mathbf{A}\mathbf{X}_L$ where $\mathbf{X}_L$ satisfies $\frac{1}{L}\mathbf{X}_L\mathbf{X}_L^H = \mathbf{\Gamma}_0$ where $\mathbf{\Gamma}_0$ is a diagonal matrix with $D > M$ non zero elements. Then, $\mathbf{\Gamma}_0$ is the unique minimizer to the following problem provided $N = Rank(\mathbf{A}_{KR})$:*

$$\min_{Rank(\mathbf{\Gamma})\geq M} \mathcal{L}^{(L)}(\mathbf{\Gamma}) \triangleq \log|\mathbf{A}\mathbf{\Gamma}\mathbf{A}^H| + \frac{1}{L}Tr\{(\mathbf{Y}_L\mathbf{Y}_L^H)(\mathbf{A}\mathbf{\Gamma}\mathbf{A}^H)^{-1}\}$$

The preceding theorems convey the following important message:

- Whether or not it is possible to recover sparse support of size $D > M$ actually depends on the relation of N and $\text{Rank}(\mathbf{A}_{KR})$.
- As long as $N = \text{Rank}(\mathbf{A}_{KR})$, it is possible to recover S, at least for large L, **irrespective of whether** $D > M$ **or** $D < M$.

3.2. The case when $N > \textbf{Rank}(\mathbf{A}_{KR})$

Recall from Theorem 1 that in this regime, the model described by $p_{\mathbf{Y}}(\mathbf{y};\mathbf{\Gamma}_0)$ can be non-identifiable. Hence, even the limiting log likelihood function $\mathcal{L}^{\infty}(\mathbf{\Gamma})$ can exhibit multiple global minimma. However, it can be verified that any global minimizer $\mathbf{\Gamma}^*$ in this case satisfies

$$\mathbf{A}\mathbf{\Gamma}^*\mathbf{A}^H = \mathbf{A}\mathbf{\Gamma}_0\mathbf{A}^H \quad (4)$$

Since $N > \text{Rank}(\mathbf{A}_{KR})$ there can be, in general, multiple $\mathbf{\Gamma}^*$ satisfying (4). This encourages us to use "sparsity" as a guiding factor to search for the true $\mathbf{\Gamma}_0$ among all $\mathbf{\Gamma}^*$ that satisfy (4). For the sake of developing theoretical guarantees, we assume $L \to \infty$ so that we know the ideal covariance matrix $\mathbf{R}_{yy}$. Then, $\mathbf{A}\mathbf{\Gamma}_0\mathbf{A}^H = \mathbf{R}_{\mathbf{yy}} - \sigma_n^2\mathbf{I}$. Substituting this in (4), we can say that all global minima belong to the set: $S^* = \{\boldsymbol{\gamma}^* : \quad \mathbf{A}_{KR}\boldsymbol{\gamma}^* = \text{vec}(\mathbf{R}_{\mathbf{yy}} - \sigma_n^2\mathbf{I})\}$. Hence we can formulate the following linear program to find the sparsest $\boldsymbol{\gamma}^*$ belonging to S^*:

$$\min_{\mathbf{x}} \|\mathbf{x}\|_1 \quad \text{s. t. } \mathbf{A}_{KR}\mathbf{x} = \text{vec}(\mathbf{R}_{yy} - \sigma_n^2\mathbf{I}) \quad (P1_{SBL})$$

The following theorem establishes the sufficient condition under which the solution to $(P1_{SBL})$ recovers the true $\mathbf{\Gamma}_0$:

Theorem 5. *If $N > Rank(\mathbf{A}_{KR})$ and $D \leq \frac{1}{2}KRank(\mathbf{A}_{KR})$, the solution $\mathbf{x}^*$ to $(P1_{SBL})$ satisfies $\mathbf{x}^* = vec(diag(\mathbf{\Gamma}_0))$*

We had proposed $(P1_{SBL})$ and its variations in a series of our recent work, as a means to perform "Correlation Aware Sparse Recovery" [13]. We now make an explicit connection to Sparse Bayesian Learning and justify why $(P1_{SBL})$ naturally becomes the algorithm of choice when the log likelihood function in SBL exhibits multiple global minima (i.e. when $N > \text{Rank}(\mathbf{A}_{KR})$, rendering the model *non identifiable*). Also, it is to be noted that in the regime $N > \text{Rank}(\mathbf{A}_{KR})$, the sparsity D needs to be upper bounded by the Kruskal rank of $\mathbf{A_{KR}}$ whereas, when $N = \text{Rank}(\mathbf{A}_{KR})$, any level of sparsity $D \leq N$ can be identified.

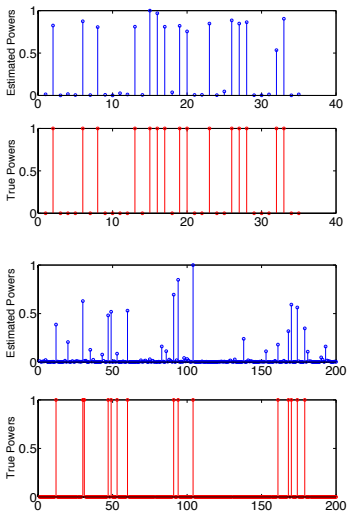

Fig. 1. Estimated power recovered by MSBL when $N = \text{Rank}(\mathbf{A}_{KR})$ (top), and Vector Representing the True Parameter (second from top). Here $M = 10, N = 35, D = 15, L = 300$. Estimated Power Recovered by MSBL when $N > \text{Rank}(\mathbf{A}_{KR})$ (third from top) and Vector representing the True Parameter (bottom). Here $M = 10, N = 200, D = 15, L = 300$.

4. SIMULATIONS

We now conduct numerical examples to show how SBL (for MMV model) behaves in the two regimes, viz. $N = \text{Rank}(\mathbf{A}_{KR})$ and $N > \text{Rank}(\mathbf{A}_{KR})$. We generate $\mathbf{A}$ with i.i.d standard normal entries. We set $M = 10$ and consider a sparsity level of $D = 15 (> M)$ for these examples. It can be shown, with probability 1, rank of $\mathbf{A_{KR}}$ is N when $N \leq 0.5(M^2 - M) = 45$. For the first case, we consider $N = 35$ which is less than 45, and so $\mathbf{A}_{KR}$ is full column rank. We also generate $L = 300$ measurement vectors by generating $\mathbf{x}_l, l = 1, \cdots, L$ from a multivariate Gaussian distribution with zero mean and diagonal covariance matrix $\mathbf{\Gamma}_0$ with $D = 15$ non zero elements all equal to 1. Fig. 1 (second from top) shows the vector containing the diagonal elements of $\mathbf{\Gamma}_0$. Fig. 1 (top) shows the corresponding quantity for $\mathbf{\Gamma}^*$ which is the solution to the problem (3) using the MSBL algorithm in [5]. It can be seen that in this case, the SBL perfectly recovers the true support and $\mathbf{\Gamma}^*$ provides a very close estimate of $\mathbf{\Gamma}_0$. Now, let's increase N to 200 which is more than the rank of $\mathbf{A}_{KR}$. We plot the corresponding $\mathbf{\Gamma}^*$ recovered by MSBL in Fig 1 (third from top). It is clear that SBL fails to recover even the support of the true $\mathbf{\Gamma}_0$. The estimated $\mathbf{\Gamma}^*$ in this case is not even sparse and this shows that SBL can fail when the model is non identifiable (when $N > \text{Rank}(\mathbf{A}_{KR})$).

In the next experiment, we conduct Monte Carlo simulations to study the performance of MSBL and $(P1_{SBL})$ and compare them with relevant Cramer Rao Bounds. We let $M = 10$, $D = 15 > M$ and test for two values of N viz., $N = 40$, and $N = 100$. For $N = 40$, the SBL algorithm is guaranteed to recover the true $\mathbf{\Gamma}_0$ for large L. We test the performance by plotting the MSE of MSBL and the Cramer Rao Bound as a function of L in Fig. 2 For each L, we generate 500 realizations of the signal $\{\mathbf{x}_l\}_{l=1}^{L}$ and the noise at an SNR of 0 dB. We compute the MSE defined as $E(\|\mathbf{\Gamma}^* - \mathbf{\Gamma}_0\|^2)$, averaged over these 500 Monte Carlo runs and plot it in Fig. 2 (left). It can be clearly seen that for large L, the MSBL estimate asymptotically approaches the Cramer Rao Bound. It is to be noted that the sparsity D is larger than the dimension M of the measurements and yet, SBL is able to successfully recover the parameter $\mathbf{\Gamma}_0$. We next consider $N = 100$. In

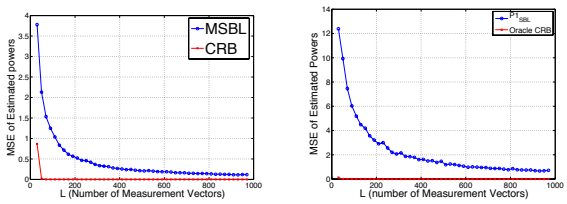

Fig. 2. (Left) MSE of MSBL algorithm and the corresponding unconstrained Cramer Rao Bound, plotted as a function of L. Here $M = 10, N = 40, D = 15$. (Right) MSE of $(P1_{SBL})$ and the Oracle Cramer Rao Bound, plotted as a function of L. Here $M = 10, N = 200, D = 15$.

this regime the MSBL algorithm fails to identify the true $\mathbf{\Gamma}_0$ owing to non identifiability. However, since $D < 45$, the $(P1_{SBL})$ algorithm is able to recover the true $\mathbf{\Gamma}_0$, especially for large L. For finite L, the equality constraint in $(P1_{SBL})$ is not quite valid and hence we employ a LASSO implementation of $(P1)_{SBL)}$. Also, the unconstrained FIM is singular in this regime and the Cramer Rao Bound is not defined. Hence, we resort to the so called "Genie-aided" or "Oracle" Cramer Rao Bound [12], which is computed given the side information that the support S is known. In Fig. 2 (right) we plot the MSE of $(P1_{SBL})$ and the Oracle CRB as a function of L. It can be seen that the MSE of $(P1_{SBL})$ decreases with L and comes closer to the Oracle CRB. However, it is not clear if there is a gap between the two curves even in the asymptotic regime. In future, it will be interesting to establish guarantees to show how closely the LASSO implementation of $(P1_{SBL})$ approaches Oracle CRB for a given M and N, as L increases.

5. CONCLUSION

We have shown that the Type II ML estimation which lies at the centre of SBL framework, successfully recovers the true $\mathbf{\Gamma}_0$ as long as $N = \text{Rank}(\mathbf{A}_{KR})$, *even when the sparsity D is larger than M*. However, when $N > \text{Rank}(\mathbf{A}_{KR})$, the parametric model becomes **non identifiable** and the ML estimation stage fails. However, the SBL algorithm *as implemented as an iterative EM algorithm in [5]* might still successfully recover $\mathbf{\Gamma}_0$ in some cases. This can happen when the initial point is well chosen and/or when an intermediate pruning/thresholding stage [6, 10, 11] is used, which is different from just solving (3). In future, we aim to thoroughly study the effect of pruning in the regime $N > \text{Rank}(\mathbf{A}_{KR})$ and $D > M$ [15].

6. REFERENCES

[1] M. E. Tipping, "Sparse Bayesian learning and the relevance vector machine", *Journal of Machine Learning Research*, 2001.

[2] M. E. Tipping, and A. C. Faul. "Analysis of sparse Bayesian learning", *Proceedings of the Conference on Neural Information Processing Systems (NIPS)*, 2002.

[3] S. Ji, Y. Xue, and L. Carin, "Bayesian compressive sensing", *IEEE Transactions on Signal Processing*, vol. 56, no. 6, pp. 2346-2356, 2008.

[4] D. Wipf and B. D. Rao, "Sparse Bayesian Learning for Basis Selection", *IEEE Transactions on Signal Processing*, vol. 52, no. 8, pp. 2153-2164, August 2004.

[5] D. Wipf and B. D. Rao, "An empirical Bayesian strategy for solving the simultaneous sparse approximation problem", *IEEE Transactions on Signal Processing*, vol. 55, no. 7, pp. 3704-3716, July 2007.

[6] Zhilin Zhang, and B.D. Rao, "Sparse Signal Recovery With Temporally Correlated Source Vectors Using Sparse Bayesian Learning", *IEEE Journal of Selected Topics in Signal Processing*, vol.5, no.5, pp. 912-926, Sep. 2011.

[7] R. Prasad, and C. R. Murthy, "Cramr-Rao-Type Bounds for Sparse Bayesian Learning," *IEEE Transactions on Signal Processing*, vol. 61, no. 3, pp. 622-632, Feb. 2013.

[8] M. Jansson, B. Goransson, and B. Ottersten, "A subspace method for direction of arrival estimation of uncorrelated emitter signals", *IEEE Transactions on Signal Processing*, vol. 47, no. 4, pp. 945-956, Apr. 1999.

[9] P. Stoica, and A. Nehorai, "Performance study of conditional and unconditional direction-of-arrival estimation", *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 38, no. 10, pp. 1783-1795, Oct. 1990.

[10] M. Hurtado, C. H. Muravchik, and A. Nehorai, "Enhanced Sparse Bayesian Learning via Statistical Thresholding for Signals in Structured Noise," *IEEE Transactions on Signal Processing*, vol. 61, no. 21, pp. 5430-5443, Nov 2013.

[11] Qiu Kun, and A. Dogandzic, "Sparse Signal Reconstruction via ECME Hard Thresholding", *IEEE Transactions on Signal Processing*, vol. 60, no. 9, pp. 4551-4569, Sept. 2012.

[12] B. Babadi, N. Kalouptsidis, and V. Tarokh, "Asymptotic Achievability of the CramerRao Bound for Noisy Compressive Sampling", *IEEE Transactions on Signal Processing*, vol. 57, no. 3, pp. 1233-1236, March 2009.

[13] Piya Pal and P. P. Vaidyanathan, "Correlation-Aware Techniques for Sparse Support Recovery", *IEEE Statistical Signal Processing Workshop, 2012*, Ann Arbor, MI, Aug. 5-8, 2012.

[14] Steven M. Kay *Fundamentals of Statistical signal processing, Volume 1: Estimation theory*, Prentice Hall PTR, 1998.

[15] Piya Pal and P. P. Vaidyanathan, "A Statistical Analaysis of Sparse Bayesian Learning", *in preparation*.