GROUP SPARSE BAYESIAN LEARNING VIA EXACT AND FAST MARGINAL LIKELIHOOD MAXIMIZATION

Zeqiang Ma, Wei Dai†, Yimin Liu* and Xiqin Wang**

*Department of Electronic Engineering, Tsinghua University, Beijing, 100084, PRC
† Department of Electrical and Electronic Engineering, Imperial College London, London, SW72AZ, UK

ABSTRACT

This paper concerns sparse Bayesian learning (SBL) problem for group sparse signals. Group sparsity means that the signal components can be divided into groups, and the entries in one group are simultaneously zero or nonzero. In SBL, each group is controlled by a hyper-parameter. The marginal likelihood maximization (MLM) problem is to maximize the marginal likelihood of a given hyper-parameter by fixing all others. The main contribution of this paper is to solve the MLM problem by finding roots of a polynomial. Hence the global minimum of the marginal likelihood can be found efficiently. Furthermore, most large matrix inverses involved in MLM are replaced with the singular value decompositions of much smaller matrices, which substantially reduces the computational complexity. The proposed method is significantly different from the popular expectation maximization techniques in the literature where multiple iterations are required for MLM and the convergence to global optimum of marginal likelihood is not guaranteed.

Index Terms— Group sparse recovery, marginal likelihood maximization, sparse Bayesian learning.

1. INTRODUCTION

While intensive research has been done in the topics of compressed sensing and sparse recovery during the past decade [1], nowadays much research focuses on sparse signals with particular structures. A popular one is the so called group sparse signals which arise in many signal processing and machine learning applications. In this model, the signal components can be divided into multiple groups, and the entries in each group are either all zero or all nonzero. It has been shown that by exploiting the group structure properly, the signal recovery performance can be largely improved. Various algorithms have been designed to recover group sparse signals, including convex optimization approaches in [2, 3, 4] and greedy method in [5], to name a few.

This paper concerns the problem of recovering group sparse signals using sparse Bayesian learning (SBL). SBL was originally proposed in the machine learning society [6] and then adopted in the signal processing society [7]. It models the signal components as independent Gaussian random variables with mean zero and unknown variance. By inferring the unknown variance from the data, SBL produces sparse estimates of the signal. One advantage of SBL is that compared with other methods, it is more robust to the ill-conditionness of the sensing matrix.

One efficient way to solve the SBL problem is via fast marginal likelihood maximization (FMLM) [8]. Let hyper-parameter α_i be the inverse of the variance of the i-th signal component. The idea of FMLM is to maximize the marginal likelihood of α_i by fixing all other α_j's, $j \neq i$. It turns out that the global optimal α_i admits a closed-form solution, which allows low complexity implementations.

However, the extension of FMLM to group sparse signals does not exist in literature as the corresponding marginal likelihood function becomes much more complicated. Instead, expectation maximization (EM) techniques have been used [9, 10, 11]. Consider again the case that α_i is updated while all other α_j's, $j \neq i$, are fixed. EM methods result in only a refinement of α_i rather than the globally optimal value. Similar approaches can be also found in [6, 12]. In all these approaches, (i) multiple iterations are required for convergence; and (ii) there is no guarantee that the converged value is the globally optimal one.

The main contribution of this work is to extend the original FMLM to recover group sparse signals. Similar to the original FMLM, the central problem is to maximize the marginal likelihood of α_i. While the marginal likelihood involves matrices parameterized by α_i, we show that its maximization can be reduced to finding roots of a polynomial, which can be solved efficiently by existing tools. Thus the global optimum of the marginal likelihood can be found. At the same time, most large matrix inverses in MLM are replaced with singular value decompositions of much smaller matrices. The overall complexity is substantially reduced.

The rest of this paper is organized as follows. Section 2 specifies the signal model and SBL inference problem. Section 3 presents the main results to solve MLM problem. Simulation results are provided in Section 4. Section 5 includes final remarks and summary.

This work is partially supported by RS-NSFC Cost Share Programme, and UDRC Phase 1 Carry-On Work.

2. BAYESIAN MODEL AND INFERENCE

Consider the general model of sparse recovery

$$\boldsymbol{y} = \boldsymbol{\Phi}\boldsymbol{x} + \boldsymbol{n}, \tag{1}$$

where $\boldsymbol{y} \in \mathbb{R}^{M\times 1}$ is the observed vector, $\boldsymbol{\Phi} \in \mathbb{R}^{M\times N}$ is the measurement matrix, $\boldsymbol{x} \in \mathbb{R}^{N\times 1}$ is a sparse vector, $\boldsymbol{n} \in \mathbb{R}^{M\times 1}$ is the white Gaussian noise with variance σ^2 and it is typically assumed that $M < N$. Furthermore, assume that $\boldsymbol{x} = [\boldsymbol{x}_1^T, \cdots, \boldsymbol{x}_{N_g}^T]^T$ is a group sparse vector. That is, the signal components can be grouped into N_g groups given by $\boldsymbol{x}_i \in \mathbb{R}^{K_i}, i = 1, \cdots, N_g$, where K_i is the size of i-th group; and the entries in each group $\boldsymbol{x}_i$ are either zero or nonzero simultaneously.

The Bayesian model in [9, 10] is used. Suppose that the entries in the i-th group are jointly Gaussian distributed with the prior probability density function (PDF)

$$p(\boldsymbol{x}_i|\alpha_i, \boldsymbol{B}_i) = \mathcal{N}\big(\boldsymbol{0}, \frac{1}{\alpha_i}\boldsymbol{B}_i\big), \tag{2}$$

where $\alpha_i \in \mathbb{R}^+ \bigcup\{\infty\}$ is an unknown hyper-parameter, and $\boldsymbol{B}_i \in \mathbb{R}^{K_i\times K_i}$ is an appropriate covariance matrix and given a priori. It is clear that $\boldsymbol{x}_i = \boldsymbol{0}$ with probability one if $\alpha_i = \infty$ and otherwise if $\alpha_i \in \mathbb{R}^+$. Let $\boldsymbol{\alpha} = [\alpha_1, \cdots, \alpha_{N_g}]^T$. Then the prior PDF of the overall signal vector $\boldsymbol{x}$ is given by $p(\boldsymbol{x}|\boldsymbol{\alpha}) = \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma}_0)$, where the covariance matrix $\boldsymbol{\Sigma}_0$ is block diagonal and its i-th block is given by $\frac{1}{\alpha_i}\boldsymbol{B}_i$. In some applications, it is also assumed that the noise variance is unknown. For this case, another parameter $\beta = \sigma^{-2}$ is introduced.

The ultimate task is to infer $\boldsymbol{x}$ from $\boldsymbol{y}$. Note that given hyper-parameters $\boldsymbol{\alpha}$ and β, this inference problem is straightforward. By Bayes' rule, the posterior PDF of $\boldsymbol{x}$ is given by $p(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{\alpha}, \boldsymbol{B}, \beta) = \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where the posterior mean and covariance matrix are given by

$$\boldsymbol{\mu} = \beta\boldsymbol{\Sigma}\boldsymbol{\Phi}^T\boldsymbol{y}, \tag{3}$$

$$\boldsymbol{\Sigma} = \left(\boldsymbol{\Sigma}_0^{-1} + \beta\boldsymbol{\Phi}^T\boldsymbol{\Phi}\right)^{-1}, \tag{4}$$

respectively. One can then set $\hat{\boldsymbol{x}} = \boldsymbol{\mu}$.

Hence the key of inference is to estimate the hyperparameters $\boldsymbol{\alpha}$ and β. The focus of this paper is the inference of $\boldsymbol{\alpha}$ as the inference of β has been solved in [6]. Consider the logarithmic likelihood of $\boldsymbol{\alpha}$ given by

$$\begin{aligned}\mathcal{L}(\boldsymbol{\alpha}) &= \log p(\boldsymbol{y}|\boldsymbol{\alpha}, \beta) \\ &= -\frac{1}{2}\left[N\log(2\pi) + \log|\boldsymbol{C}| + \boldsymbol{y}^T\boldsymbol{C}^{-1}\boldsymbol{y}\right],\end{aligned} \tag{5}$$

where

$$\boldsymbol{C} = \frac{1}{\beta}\boldsymbol{I} + \boldsymbol{\Phi}\boldsymbol{\Sigma}_0\boldsymbol{\Phi}^T, \tag{6}$$

is the covariance matrix of $\boldsymbol{y}$. A maximization of the log-likelihood is used to infer $\boldsymbol{\alpha}$, which is referred to as type-II maximum likelihood method [13].

3. A FAST AND EXACT MARGINAL LIKELIHOOD MAXIMIZATION METHOD

It is difficult to solve the log-likelihood maximization problem (5) due to its non-convexity. In this section, we extend the original marginal likelihood maximization (MLM) method in [6, 8] for group sparse signals and refer to the method as group MLM (GMLM). The basic idea is to perform the maximization with respect to α_i by fixing all other α_j's, $j \neq i$ (hence the term marginal likelihood). For simplificity, in this paper we assume that $\boldsymbol{B}_i = \boldsymbol{I}_K$ where $\boldsymbol{I}_K$ is a $K \times K$ identity matrix and hence $\boldsymbol{\Sigma}_0 = \text{diag}(\boldsymbol{\alpha}) \otimes \boldsymbol{I}_K$. The presented results can be easily extended to the general case described in Section 2. The details would be presented in the journal version which will appear later.

The marginal likelihood is computed as follows. Let $\boldsymbol{\Phi}_i$ be the submatrix of $\boldsymbol{\Phi}$ that consists of the columns corresponding the i-th group, i.e., the columns indexed by $\{(i-1)K+1, \cdots, iK\}$. Decompose the matrix $\boldsymbol{C}$ as

$$\boldsymbol{C} = \boldsymbol{C}_{-i} + \frac{1}{\alpha_i}\boldsymbol{\Phi}_i\boldsymbol{\Phi}_i^T. \tag{7}$$

where the first term $\boldsymbol{C}_{-i} = \frac{1}{\beta}\boldsymbol{I} + \sum_{j\neq i}\frac{1}{\alpha_j}\boldsymbol{\Phi}_j\boldsymbol{\Phi}_j^T$ contains all the terms that are independent of α_i and the second term includes all the terms related to it. By using Woodbury matrix identity [14], the objective function (5) can be also decomposed into two parts

$$\mathcal{L}(\boldsymbol{\alpha}) = \mathcal{L}(\boldsymbol{\alpha}_{-i}) + \frac{1}{2}\ell(\alpha_i), \tag{8}$$

where $\mathcal{L}(\boldsymbol{\alpha}_{-i})$ is independent of α_i, and

$$\begin{aligned}\ell(\alpha_i) := &\log|\alpha_i\boldsymbol{I}| - \log|\alpha_i\boldsymbol{I} + \boldsymbol{\Phi}_i^T\boldsymbol{C}_{-i}^{-1}\boldsymbol{\Phi}_i| \\ &+ \boldsymbol{y}^T\boldsymbol{C}_{-i}^{-1}\boldsymbol{\Phi}_i(\alpha_i\boldsymbol{I} + \boldsymbol{\Phi}_i^T\boldsymbol{C}_{-i}^{-1}\boldsymbol{\Phi}_i)^{-1}\boldsymbol{\Phi}_i^T\boldsymbol{C}_{-i}^{-1}\boldsymbol{y}.\end{aligned} \tag{9}$$

For notational convenience, define

$$\bar{\boldsymbol{S}}_i := \boldsymbol{\Phi}_i^T\boldsymbol{C}_{-i}^{-1}\boldsymbol{\Phi}_i, \quad \bar{\boldsymbol{q}}_i := \boldsymbol{\Phi}_i^T\boldsymbol{C}_{-i}^{-1}\boldsymbol{y}, \tag{10}$$

where $\bar{\boldsymbol{S}}_i \in \mathbb{R}^{K\times K}$ and $\bar{\boldsymbol{q}}_i \in \mathbb{R}^{K\times 1}$. The objective function $\ell(\alpha_i)$ becomes

$$\ell(\alpha_i) = \log|\alpha_i\boldsymbol{I}| - \log|\alpha_i\boldsymbol{I} + \bar{\boldsymbol{S}}_i| + \bar{\boldsymbol{q}}_i^T(\alpha_i\boldsymbol{I} + \bar{\boldsymbol{S}}_i)^{-1}\bar{\boldsymbol{q}}_i. \tag{11}$$

To maximize the marginal likelihood, one simply needs to maximize $\ell(\alpha_i)$ in (11) with respect to α_i.

A sequential algorithm [6, 8] can be built upon the marginal likelihood $\ell(\alpha_i)$. We outline it in Algorithm 1 by omitting many details. There are two computational challenges associated with step 1. Firstly, the definitions in (10) involves the matrix inverse $\boldsymbol{C}_{-i}$ (of dimension M) which needs to be computed N_g times as i varies from 1 to N_g. Secondly, the optimization of $\ell(\alpha_i)$ is difficult as the objective function involves matrices parameterized by α_i. We shall develop an exact GMLM method to address these two challenges.

Algorithm 1 GMLM Sequential Algorithm

In the t-th iteration, do

1. Scan i and find the i^* to maximize $\ell(\alpha_i^*) - \ell\left(\alpha_i^{t-1}\right)$ where α_i^* is the global maximizer of $\ell(\alpha_i)$.
2. Update hyper-parameter vector $\boldsymbol{\alpha}$: $\alpha_{i^*}^t = \alpha_{i^*}^*$ and $\alpha_j^t = \alpha_j^{t-1}$ for all $j \neq i^*$.
3. Update other parameters to allow next iteration.

3.1. The Exact MLM

The aforementioned two challenges are addressed by reducing the number of large matrix inverses and turning the optimization problem to a root finding problem.

The next lemma reduces the number of large matrix inverses.

Lemma 3.1. *Define*

$$\boldsymbol{S}_i := \boldsymbol{\Phi}_i^T \boldsymbol{C}^{-1} \boldsymbol{\Phi}_i, \quad \boldsymbol{q}_i := \boldsymbol{\Phi}_i^T \boldsymbol{C}^{-1} \boldsymbol{y}. \tag{12}$$

Write the singular decomposition form of $\boldsymbol{S}_i$ as $\boldsymbol{S}_i = \boldsymbol{V}_i diag(s_{i,k}) \boldsymbol{V}_i^T$, where $\boldsymbol{V}_i$ is the singular vector matrix and $s_{i,k}$'s are the singular values. Then $\bar{\boldsymbol{S}}_i$ and $\bar{\boldsymbol{q}}_i$ in (10) are given by

$$\begin{aligned} \bar{\boldsymbol{S}}_i &= \boldsymbol{V}_i diag\Big(\frac{\alpha_i s_{i,k}}{\alpha_i - s_{i,k}}\Big) \boldsymbol{V}_i^T, \\ \bar{\boldsymbol{q}}_i &= \boldsymbol{V}_i diag\Big(\frac{\alpha_i}{\alpha_i - s_{i,k}}\Big) \boldsymbol{V}_i^T \boldsymbol{q}_i. \end{aligned} \tag{13}$$

At the beginning of each iteration, $\boldsymbol{\alpha}^{t-1}$ is given and one can compute $\boldsymbol{C}$ and its inverse according to the definition in (6). Then $\boldsymbol{S}_i$'s, $\boldsymbol{q}_i$'s, $\bar{\boldsymbol{S}}_i$'s, and $\bar{\boldsymbol{q}}_i$'s can be computed without further large matrix inverse. This reduces N_g many large matrix inverse (of dimension M) into one large matrix inverse and N_g many small matrix operations (of dimension K). Hence the computational complexity is reduced from $O(N_g M^3)$ to $O(M^3 + N_g K^3) = O(M^3 + NK^2)$. In many application, K is a small constant and N_g, M, and N are in the same order. In this case, computational complexity is reduced from $O(M^4)$ to $O(M^3)$.

Now we show how to maximize the marginal likelihood $\ell(\alpha_i)$. Note that the singular value decomposition of $\bar{\boldsymbol{S}}_i$ can be easily obtained from (13). Denote the singular values of $\bar{\boldsymbol{S}}_i$ by $\bar{s}_{i,k}$'s, $k = 1, 2, \cdots, K$. The marginal likelihood $\ell(\alpha_i)$ can be written as

$$\ell(\alpha_i) = \sum_{k=1}^{K} \Big(\log(\alpha_i) - \log(\alpha_i + \bar{s}_{i,k}) + \frac{c_{i,k}^2}{\alpha_i + \bar{s}_{i,k}} \Big), \tag{14}$$

where $c_{i,k}$ is the k-th entry of $\boldsymbol{c}_i = \boldsymbol{V}_i \bar{\boldsymbol{q}}_i$. Its derivative with respect to α_i is given by

$$\ell'(\alpha_i) = \frac{\sum_{k=1}^{K} \left((\bar{s}_{i,k} - c_{i,k}^2)\alpha_i + \bar{s}_{i,k}^2 \right) \prod_{t \neq k} (\alpha_i + \bar{s}_{i,t})^2}{\alpha_i \prod_{k=1}^{K} (\alpha_i + \bar{s}_{i,k})^2}. \tag{15}$$

The local maximizers or minimizers of $\ell(\alpha_i)$ correspond to where the derivative equals to zero. Note that the denominator is always positive. One only needs to find the roots of the numerator. As the numerator is a polynomial of degree $2K - 1$, there are $2K - 1$ roots (possible complex). Since α_i was introduced in the statistical model to describe the variance of $\boldsymbol{x}_i$, one can define the feasible set of α_i as

$$\mathcal{A}_i = \left\{ \alpha_i \in \mathbb{R}^+ : \ell'(\alpha_i) = 0 \right\} \bigcup \{+\infty\}.$$

Here, the point $+\infty$ has to be included as it is feasible and it may correspond to the global optimum of $\ell(\alpha_i)$ (see [6] for a specific example). Then the global optimal α_i can be identified by

$$\alpha_i^* = \arg \max_{\alpha_i \in \mathcal{A}_i} \ell(\alpha_i),$$

where $\ell(+\infty) = 0$ according to the definition of $\ell(\alpha_i)$ in (11). Note that the root finding problem can be solved by using companion matrix of size $K \times K$. The overall complexity of the root finding step is $O(N_g K^3)$ (for all $i = 1, 2, \cdots, N_g$) which is the same as that for computing $\bar{S}_i$'s and does not change the previous complexity analysis in terms of order of magnitude.

3.2. Other Compuational Simplifications

Further computational simplification is possible. At each iteration, the matrix $\boldsymbol{C}^{-1}$ is needed. Note that the matrix $\boldsymbol{C}$ defined in (6) can be viewed as a function of $\boldsymbol{\alpha}$. Since the $\boldsymbol{\alpha}$'s in two consecutive iterations only differ in one entry, one may use Woodbury matrix identity to simplify the updates of $\boldsymbol{C}^{-1}$.

When the noise variance $\sigma^2 = \beta^{-1}$ is fixed (e.g. given a priori), one can directly update $\boldsymbol{C}^{-1}$ by observing that $\boldsymbol{C}^t = \boldsymbol{C}^{t-1} + \left(\alpha_{i^*}^t - \alpha_{i^*}^{t-1}\right) \boldsymbol{\Phi}_{i^*} \boldsymbol{\Phi}_{i^*}^T$, where i^* is the index of the updated α in the t-th iteration.

Now consider the case that $\sigma^2 = \beta^{-1}$ is not given and needs to be updated. By setting $\frac{\partial \mathcal{L}(\boldsymbol{\alpha})}{\partial \sigma^2} = 0$, one obtains

$$\beta^{(new)} = \frac{M - \sum_{n=1}^{N_g} \left(K - \alpha_n \mathrm{tr}(\boldsymbol{\Sigma}_n) \right)}{\|\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{\mu}\|_2^2},$$

where $\boldsymbol{\Sigma}_n$ is the n-th block of the matrix $\boldsymbol{\Sigma}$. This means at each iteration, one needs to update $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$, $\boldsymbol{C}$, $\boldsymbol{S}_i$'s and $\boldsymbol{q}_i$'s. Again by using the fact that only one element in $\boldsymbol{\alpha}$ has been changed, one can simplify the computations. However, much more details are involved and will be provided only in the journal version of this paper.

3.3. Detailed Relation to Prior Work

Our work is largely motivated by the original FMLM [12, 8], which concerns only standard sparse signals, i.e., sparse signals without group structures. Over there the matrices involved in (11) become scalars which makes the MLM problem much easier. At the same time, very similar techniques are used in our work and [8] to reduce the number of large matrix inverses. The difference is that the connections between singular value decompositions are essential for group sparse signals (Lemma 3.1) while there is no need to resort to such decompositions for standard sparse signals [8].

In [9, 10, 11], different techniques including BSBL-EM, BSBL-BO and BSBL-ℓ_1 are developed to refine the estimation of the hyper-parameter α_i for group sparse signals. All these techniques originate from EM mechanism. The basic form used for α_i refinement is the same: $\alpha_i^t = g(\alpha_i^{t-1}, \boldsymbol{y}, \boldsymbol{\Phi}, \boldsymbol{\alpha}_{-i}^{t-1})$, where the g function involves large matrix inverses and α_i appears at both sides of the equation. Hence multiple iterations are needed for convergence and there is no guarantee the global optimality of the converged value. As a comparison, the global optimal α_i is obtained by finding roots of a polynomial in our work.

4. NUMERICAL RESULTS

Consider a group sparse signal recovery problem (1). In all the simulations presented here, the measurement matrix is generated from the standard Gaussian random matrix ensemble.

In [11] the authors compared the performance of the three SBL algorithms with other non-SBL algorithms, showing that the SBL based algorithms outperform non-SBL algorithms and BSBL-BO has the fastest convergence speed compared to BSBL-EM, BSBL-l_1. In this work we focus on the performance comparison within the SBL family and hence choose the best algorithm BSBL-BO for comparison.

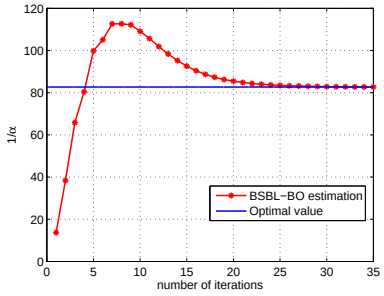
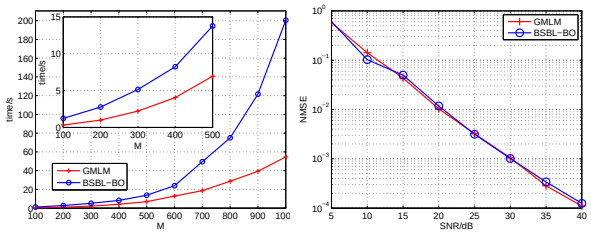


Fig. 1. Convergence rate.

Figure 1 illustrates the convergence rate of MLM. We randomly pick a specific $\boldsymbol{\alpha}$, and update α_i while fixing all other α_j's, $j \neq i$. As discussed before, BSBL-BO is based on EM technique and multiple iterations are required for solving MLM. According to the simulations, it needs roughly 25 iterations to get close to the globally optimal value. As a comparison, the method in this work gives the globally optimal α_i in one-shot. In this example, $M = 400, N = 800, K = 4, N_g = 200, \text{SNR} = 20dB$ and 50 groups of all the 200 groups are nonzero.

It is noteworthy that the above comparison does not reflect the complexity difference of the overall algorithms. In EM based methods, MLM cannot be solved in one-shot and therefore the algorithm does not try to find the global optimal value when updating a given hyper-parameter. Instead, only one EM iteration is used to update one hyper-parameter before the algorithm moves to the next hyper-parameter. The resulting algorithm flowchart is quite different from the one in Algorithm 1. Due to this difference, the direct complexity comparasion of the sequential GMLM in Algorithm 1 and BSBL-BO is not rigorous. Nevertheless for engineering purpose, it is always beneficial to numerically study the improvement from benchmark methods.



(a) Time cost vs dimension (b) NMSE vs SNR

Fig. 2. Performance comparison

Figure 2(a) compares the running-time of GMLM and BSBL-BO. The running time shown in the results are obtained from avaraging 100 realizations. In the simulations, $N = 2M$, the group size $K = 4$, and the sparsity level (the fraction of nonzero groups) is 0.25. From the figure we can see that the running time of GMLM is generally less than half of that of BSBL-BO, and it is further improved as the dimension of the problem increases. At the same time, both algorithms achieve almost identical reconstruction distortion in terms of normalized mean squared error.

5. CONCLUSION

This paper studies the SBL mechanism for group sparse signals. The main contribution is to find the exact solution of the MLM problem. This is achieved by translating the MLM problem to a root-finding problem. At the same time, mechanisms are developed for replacing most matrix inverses in MLM with the singular value decompositions of much smaller matrices. Our method extends the original FMLM designed only for standard sparse signals.

6. REFERENCES

[1] Emmanuel J Candès and Michael B Wakin, "An introduction to compressive sampling," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 21–30, 2008.

[2] Ming Yuan and Yi Lin, "Model selection and estimation in regression with grouped variables," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 68, no. 1, pp. 49–67, 2006.

[3] Yonina C Eldar and Moshe Mishali, "Robust recovery of signals from a structured union of subspaces," *IEEE Trans. Inf. Theory*, vol. 55, no. 11, pp. 5302–5316, 2009.

[4] Ehsan Elhamifar and René Vidal, "Block-sparse recovery via convex optimization," *IEEE Trans. Signal Process.*, vol. 60, no. 8, pp. 4094–4107, 2012.

[5] Yonina C Eldar, Patrick Kuppinger, and Helmut Bölcskei, "Block-sparse signals: Uncertainty relations and efficient recovery," *IEEE Trans. Signal Process.*, vol. 58, no. 6, pp. 3042–3054, 2010.

[6] Michael E Tipping, "Sparse Bayesian learning and the relevance vector machine," *The journal of machine learning research*, vol. 1, pp. 211–244, 2001.

[7] Shihao Ji, Ya Xue, and Lawrence Carin, "Bayesian compressive sensing," *IEEE Trans. Signal Process.*, vol. 56, no. 6, pp. 2346–2356, 2008.

[8] Michael E Tipping and Anita C Faul, "Fast marginal likelihood maximisation for sparse Bayesian models," in *Proceedings of the ninth international workshop on artificial intelligence and statistics*, 2003, vol. 1.

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

[10] Zhilin Zhang and Bhaskar D Rao, "Sparse signal recovery with temporally correlated source vectors using sparse Bayesian learning," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 5, pp. 912–926, 2011.

[11] Zhenhao Zhang and Bhaskar Rao, "Extension of SBL algorithms for the recovery of block sparse signals with intra-block correlation," *IEEE Trans. Signal Process.*, vol. 61, no. 8, pp. 2009–2015, 2013.

[12] M. E. Tipping and A. Faul, "Analysis of sparse Bayesian learning," *Advances in neural information processing systems*, vol. 14, pp. 383–389, 2002.

[13] James O Berger, *Statistical decision theory and Bayesian analysis*, Springer Science & Business Media, 2013.

[14] Max A Woodbury, "Inverting modified matrices," *Memorandum report*, vol. 42, pp. 106, 1950.