Support Knowledge-Aided Sparse Bayesian Learning for Compressed Sensing

Jun Fang*, Yanning Shen†, Fuwei Li*, Hongbin Li‡, and Zhi Chen *
*National Key Laboratory on Communications, University of Electronic Science and Technology of China
Chengdu 611731, China Email: JunFang@uestc.edu.cn
†Department of Electrical and Computer Engineering, University of Minnesota
Minneapolis, MN 55455 USA Email: shenx513@umn.edu
‡Department of Electrical and Computer Engineering, Stevens Institute of Technology
Hoboken, NJ 07030 USA Email: Hongbin.Li@stevens.edu

***Abstract*— In this paper, we study the problem of sparse signal recovery when partial but partly erroneous prior knowledge of the signal's support is available. Based on the conventional sparse Bayesian learning framework, we propose an improved hierarchical prior model. The proposed modeling constitutes a three-layer hierarchical form. The first two layers, similar to the conventional sparse Bayesian learning, place a Gaussian-inverse-Gamma prior on the signal, while the third layer is newly added, with a prior placed on the parameters $\{b_i\}$, where $\{b_i\}$ are parameters characterizing the sparsity-controlling hyperparameters $\{\alpha_i\}$. Such a modeling enables to automatically learn the true support from partly erroneous information through learning the values of the parameters $\{b_i\}$. A variational Bayesian inference algorithm is developed based on the proposed prior model. Numerical results are provided to illustrate the performance of the proposed algorithm.**

***Index Terms*— Compressed sensing, sparse Bayesian learning, prior support knowledge.**

I. Introduction

We study the problem of sparse signal recovery when prior information on the signal's partial support is available. In practice, prior information about the support region of the sparse signal may be obtained from the support estimate of the previous time instant. This is particularly the case for time-varying sparse signals whose support changes slowly over time. For example, in the real-time dynamic MRI reconstruction, it was shown that the support of a medical image sequence undergoes a small variation with the support changes (number of additions and removals) less than 2% of the support size. The problem of sparse signal recovery with partial support information was studied in several independent and parallel works [1]–[3]. It has been observed by extensive experiments [1]–[3] that the sparse recovery performance can be significantly improved through exploiting the prior support knowledge. Nevertheless, this performance improvement can only be achieved when the prior knowledge of the signal's partial support is fairly accurate. Existing methods, e.g. [1]–[3], suffer from severe recovery performance degradation or even recovery failure in the presence of inaccurate prior knowledge. In practice, however, the estimate of the signal's support inevitably incurs errors. In this paper, we propose an improved hierarchical prior model. The proposed modeling constitutes a three-layer hierarchical form. The first two layers, similar to the conventional sparse Bayesian learning, place a Gaussian-inverse-Gamma prior on the signal, while the third layer is newly added with a prior placed on the parameters $\{b_i\}$. Such a modeling enables to distinguish the true support from erroneous support through learning the values of $\{b_i\}$. We resort to the variational inference methodology to perform the Bayesian inference on the proposed three-stage hierarchical model, and develop a new sparse Bayesian learning method which has the ability to distinguish the true support from the erroneous information.

This work was supported in part by the National Science Foundation of China under Grant 61201274, and the National Science Foundation under Grant ECCS-0901066.

II. Hierarchical Prior Model

We consider the problem of recovering a sparse signal $\boldsymbol{x} \in \mathbb{R}^n$ from noise-corrupted measurements

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{w} \tag{1}$$

where $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ $(m < n)$ is the measurement matrix, and $\boldsymbol{w}$ is the additive multivariate Gaussian noise with zero mean and covariance matrix $\sigma^2\boldsymbol{I}$. Suppose we have partial but partly erroneous knowledge of the support of the sparse signal $\boldsymbol{x}$. The prior knowledge P can be divided into two parts: $P = S \cup E$, where S denotes the subset containing correct information about the support and E denotes the error subset. If we let T denote the true support of $\boldsymbol{x}$ and T^c denote the complement of the set T, i.e. $T \cup T^c = \{1, 2, \ldots, n\}$, then we have $S \subset T$, and $E \subset T^c$. Note that the only prior information we have is P. The partition of S and E is unknown to us.

We develop a sparse signal recovery algorithm which has the ability to distinguish the correct support from erroneous information and thus can exploit the prior support information in a more constructive way. To this objective, we will propose a new hierarchical sparse Bayesian learning (SBL) model which allows to learn the correct information from the partly erroneous knowledge. Before we proceed, we first provide a brief overview of the hierarchical model for the conventional SBL method.

A. Overview of Hierarchical Models for Conventional SBL

In the conventional sparse Bayesian learning framework, a two-layer hierarchical prior model was proposed to promote the sparsity of the solution. In the first layer, $\boldsymbol{x}$ is assigned a Gaussian prior distribution

$$p(\boldsymbol{x}|\boldsymbol{\alpha}) = \prod_{i=1}^{n} p(x_i|\alpha_i) \tag{2}$$

where $p(x_i|\alpha_i) = \mathcal{N}(x_i|0, \alpha_i^{-1})$, and $\boldsymbol{\alpha} \triangleq \{\alpha_i\}$, the inverse variance (precision) of the Gaussian distribution, are non-negative hyperparameters. The second layer specifies Gamma distributions as hyperpriors over the hyperparameters $\{\alpha_i\}$, i.e.

$$p(\boldsymbol{\alpha}) = \prod_{i=1}^{n} \text{Gamma}(\alpha_i|a, b) = \prod_{i=1}^{n} \Gamma(a)^{-1} b^a \alpha_i^{a-1} e^{-b\alpha_i} \tag{3}$$

where $\Gamma(a) = \int_0^\infty t^{a-1}e^{-t}dt$ is the Gamma function, the parameters a and b used to characterize the Gamma distribution are chosen to be very small values, e.g. 10^{-4}, in order to provide non-informative hyperpriors over $\{\alpha_i\}$. As discussed in [5], using a non-informative hyperprior allows α_i to become arbitrarily large. As a consequence, the associated coefficient x_i will be driven to zero, thus yielding a sparse solution. This mechanism is also referred to as the "automatic relevance determination" mechanism which tends to switch off most of the coefficients that are deemed to be irrelevant, and only keep very few relevant coefficients to explain the data.

B. Proposed Hierarchical Models

When the value of the parameter b is relatively large, e.g. $b = 1$, it can be readily observed from (3) that the hyperpriors are no longer non-informative and now they encourage small values of $\{\alpha_i\}$. In this case, an arbitrarily large value of α_i is prohibited. As a result, the formulation of the two layer hierarchical model does not result in a sparsity-encouraging prior and therefore does not necessarily lead to a sparse solution. This fact, however, inspires us to develop a new way to incorporate the prior support information into the sparse Bayesian learning framework. Specifically, instead of using a common parameter b for all hyperparameters $\{\alpha_i\}$, we hereby employ an individual parameter b_i for each hyperparameter α_i, i.e.

$$p(\boldsymbol{\alpha}) = \prod_{i=1}^{n} \text{Gamma}(\alpha_i|a, b_i) = \prod_{i=1}^{n} \Gamma(a)^{-1} b_i^a \alpha_i^{a-1} e^{-b_i\alpha_i} \tag{4}$$

Such a formulation allows us to assign different priors to different coefficients. If a partial knowledge of the signal's support, P, is available, then the associated parameters of $\{b_i\}$ can be set to a relatively large value, say 1, in order to place a non-sparsity-encouraging prior on the corresponding coefficients, whereas the rest parameters of $\{b_i\}$ are still assigned a small value, say 10^{-4}, to encourage sparse coefficients, that is,

$$b_i = \begin{cases} 1 & i \in P \\ 10^{-4} & \text{otherwise} \end{cases} \tag{5}$$

The above modified hierarchical model effectively integrates the prior support information into the sparse Bayesian learning framework. Nevertheless, the modified two-layer hierarchical model which assigns fixed values to $\{b_i\}$ still lacks flexibility to learn and adapt to the true situation. To address this issue, we partition the parameters $\{b_i\}$ into two subsets: $\{b_i, \forall i \in P\}$, and $\{b_i, \forall i \in P^c\}$, where P^c denote the complement of P, i.e. $P \cup P^c = \{1, 2 \ldots, n\}$. For $\{b_i, \forall i \in P^c\}$, the parameters are still considered to be deterministic and assigned a very small value, i.e.

$$b_i = 10^{-4} \qquad \forall i \in P^c \tag{6}$$

For $\{b_i, \forall i \in P\}$, instead of assigning a fixed large value, we model them as random parameters and place hyperpriors over $\{b_i, \forall i \in P\}$. Since $\{b_i, \forall i \in P\}$ are expected to be positive values, suitable priors over $\{b_i, \forall i \in P\}$ are also Gamma distributions:

$$\text{Gamma}(b_i|p, q) = \Gamma(p)^{-1} q^p b_i^{p-1} e^{-qb_i} \qquad \forall i \in P \tag{7}$$

where p and q are parameters characterizing the Gamma distribution. Their choice will be specified later in this paper. In doing this way, the modeling constitutes a three-layer hierarchical form which allows to learn the parameters $\{b_i, \forall i \in P\}$ in an automatic manner from the data, and therefore has the ability to distinguish the correct support from erroneous information.

III. Variational Bayesian Inference

We now proceed to perform variational Bayesian inference based on the proposed hierarchical model. For notational convenience, define

$$\gamma \triangleq \sigma^{-2}$$

Following the conventional sparse Bayesian learning framework [5], we place a Gamma hyperprior over γ:

$$p(\gamma) = \text{Gamma}(\gamma|c, d) = \Gamma(c)^{-1} d^c \gamma^{c-1} e^{-d\gamma} \tag{8}$$

where the parameters c and d are set to small values, e.g. $c = d = 10^{-4}$.

Let $\boldsymbol{\theta} \triangleq \{\boldsymbol{x}, \boldsymbol{\alpha}, \gamma, \bar{\boldsymbol{b}}\}$, where $\bar{\boldsymbol{b}} \triangleq \{b_i, \forall i \in P\}$ are hidden variables as well since they are assigned hyperpriors and need to be learned. We assume posterior independence among the hidden variables $\boldsymbol{x}$, $\boldsymbol{\alpha}$, γ, and $\bar{\boldsymbol{b}}$, i.e.

$$\begin{aligned} p(\boldsymbol{x}, \boldsymbol{\alpha}, \gamma, \bar{\boldsymbol{b}}|\boldsymbol{y}) \approx & q(\boldsymbol{x}, \boldsymbol{\alpha}, \gamma, \bar{\boldsymbol{b}}) \\ = & q_x(\boldsymbol{x}) q_\alpha(\boldsymbol{\alpha}) q_\gamma(\gamma) q_{\bar{b}}(\bar{\boldsymbol{b}}) \end{aligned} \tag{9}$$

With this mean field approximation, the posterior distribution of each hidden variable can be computed by minimizing the Kullback-Leibler (KL) divergence while keeping other variables fixed using their most recent distributions, which gives

$$\begin{aligned} \ln q_x(\boldsymbol{x}) &= \langle \ln p(\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{\alpha}, \gamma, \bar{\boldsymbol{b}}) \rangle_{q_\alpha(\boldsymbol{\alpha}) q_\gamma(\gamma) q_{\bar{b}}(\bar{\boldsymbol{b}})} + \text{constant} \\ \ln q_\alpha(\boldsymbol{\alpha}) &= \langle \ln p(\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{\alpha}, \gamma, \bar{\boldsymbol{b}}) \rangle_{q_x(\boldsymbol{x}) q_\gamma(\gamma) q_{\bar{b}}(\bar{\boldsymbol{b}})} + \text{constant} \\ \ln q_\gamma(\gamma) &= \langle \ln p(\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{\alpha}, \gamma, \bar{\boldsymbol{b}}) \rangle_{q_x(\boldsymbol{x}) q_\alpha(\boldsymbol{\alpha}) q_{\bar{b}}(\bar{\boldsymbol{b}})} + \text{constant} \\ \ln q_{\bar{b}}(\bar{\boldsymbol{b}}) &= \langle \ln p(\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{\alpha}, \gamma, \bar{\boldsymbol{b}}) \rangle_{q_x(\boldsymbol{x}) q_\alpha(\boldsymbol{\alpha}) q_\gamma(\gamma)} + \text{constant} \end{aligned}$$

Details of this Bayesian inference scheme are provided below.

1). Update of $q_x(\boldsymbol{x})$: The variational optimization of $q_x(\boldsymbol{x})$ can be calculated as follows by ignoring the terms that are independent of $\boldsymbol{x}$:

$$\begin{aligned}\ln q(\boldsymbol{x}) \propto & \langle \ln p(\boldsymbol{y}|\boldsymbol{x},\gamma) + \ln p(\boldsymbol{x}|\boldsymbol{\alpha}) \rangle_{q_\alpha(\boldsymbol{\alpha}) q_\gamma(\gamma)} \\ \propto & -\frac{\langle\gamma\rangle}{2}(\boldsymbol{y}-\boldsymbol{A}\boldsymbol{x})^T(\boldsymbol{y}-\boldsymbol{A}\boldsymbol{x}) - \frac{1}{2}\boldsymbol{x}^T\langle \boldsymbol{D}\rangle \boldsymbol{x} \end{aligned} \quad (10)$$

We can easily verify that $q(\boldsymbol{x})$ follows a Gaussian distribution with its mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Phi}$ given respectively as

$$\begin{aligned}\boldsymbol{\mu} = & \langle\gamma\rangle \boldsymbol{\Phi}\boldsymbol{A}^T\boldsymbol{y} \\ \boldsymbol{\Phi} = & \Big(\langle\gamma\rangle \boldsymbol{A}^T\boldsymbol{A} + \langle \boldsymbol{D}\rangle\Big)^{-1}\end{aligned} \quad (11)$$

2). Update of $q_\alpha(\boldsymbol{\alpha})$: Similarly, the approximate posterior $q_\alpha(\boldsymbol{\alpha})$ can be computed as

$$\begin{aligned}\ln q_\alpha(\boldsymbol{\alpha}) \propto & \langle \ln p(\boldsymbol{x}|\boldsymbol{\alpha}) + \ln p(\boldsymbol{\alpha}|a,\boldsymbol{b}) \rangle_{q_x(\boldsymbol{x}) q_{\bar{b}}(\bar{\boldsymbol{b}})} \\ = & \sum_i^n \big\langle (a-0.5)\ln\alpha_i - (0.5x_i^2 + b_i)\alpha_i \big\rangle_{q_x(\boldsymbol{x}) q_{\bar{b}}(\bar{\boldsymbol{b}})} \\ \overset{(a)}{=} & \sum_{i\in P} \big\{ (a+0.5)\ln\alpha_i - (\langle b_i\rangle + 0.5\langle x_i^2\rangle)\alpha_i \big\} \\ & + \sum_{i\in P^c} \big\{ (a+0.5)\ln\alpha_i - (b_i + 0.5\langle x_i^2\rangle)\alpha_i \big\}\end{aligned} \quad (12)$$

where in (a), the terms inside the summation are partitioned into two subsets P and P^c because $\{b_i, i \in P^c\}$ are deterministic parameters whose values are given in (6), while $\{b_i, i \in P\}$ are latent variables and thus we need to perform the expectation over these hidden variables. The posterior $q(\boldsymbol{\alpha})$ has a form of a product of Gamma distributions

$$q(\boldsymbol{\alpha}) = \prod_{i=1}^n \text{Gamma}(\alpha_i|\tilde{a},\tilde{b}_i) \quad (13)$$

with the parameters $\tilde{a}$ and $\tilde{b}_i$ given by

$$\tilde{a} = a + 0.5 \quad (14)$$

$$\tilde{b}_i = \begin{cases} \langle b_i\rangle + 0.5\langle x_i^2\rangle & i \in P \\ b_i + 0.5\langle x_i^2\rangle & i \in P^c \end{cases} \quad (15)$$

3). Update of $q_\gamma(\gamma)$: The approximate posterior $q_\gamma(\gamma)$ can be computed as

$$\begin{aligned}\ln q_\gamma(\gamma) \propto & \langle \ln p(\boldsymbol{y}|\boldsymbol{x},\gamma) + \ln p(\gamma|c,d) \rangle_{q_x(\boldsymbol{x})} \\ \propto & \Big(\frac{m}{2} + c - 1\Big)\ln\gamma \\ & - \left(\frac{1}{2}\langle (\boldsymbol{y}-\boldsymbol{A}\boldsymbol{x})^T(\boldsymbol{y}-\boldsymbol{A}\boldsymbol{x})\rangle_{q_x(\boldsymbol{x})} + d\right)\gamma\end{aligned} \quad (16)$$

It can be easily verified that $q(\gamma)$ follows a Gamma distribution

$$q(\gamma) = \text{Gamma}(\gamma|\tilde{c},\tilde{d}) \quad (17)$$

where

$$\begin{aligned}\tilde{c} = & \frac{m}{2} + c \\ \tilde{d} = & d + \frac{1}{2}\langle (\boldsymbol{y}-\boldsymbol{A}\boldsymbol{x})^T(\boldsymbol{y}-\boldsymbol{A}\boldsymbol{x})\rangle_{q_x(\boldsymbol{x})}\end{aligned} \quad (18)$$

in which

$$\langle (\boldsymbol{y}-\boldsymbol{A}\boldsymbol{x})^T(\boldsymbol{y}-\boldsymbol{A}\boldsymbol{x})\rangle_{q_x(\boldsymbol{x})} = \|\boldsymbol{y}-\boldsymbol{A}\boldsymbol{\mu}\|_2^2 + \text{tr}\Big\{\boldsymbol{A}^T\boldsymbol{A}\boldsymbol{\Phi}\Big\}$$

4). Update of $q_{\bar{b}}(\bar{\boldsymbol{b}})$: The variational optimization of $q_{\bar{b}}(\bar{\boldsymbol{b}})$ yields:

$$\begin{aligned}\ln q_{\bar{b}}(\bar{\boldsymbol{b}}) \propto & \langle \ln p(\boldsymbol{\alpha}|a,\boldsymbol{b}) + \ln p(\bar{\boldsymbol{b}}|p,q) \rangle_{q_\alpha(\boldsymbol{\alpha})} \\ \propto & \sum_{i\in P}\{-b_i\langle\alpha_i\rangle + (p-1)\ln b_i - q b_i\}\end{aligned} \quad (19)$$

from which we can readily arrive at

$$q(\bar{\boldsymbol{b}}) = \prod_{i\in P} \text{Gamma}(b_i|p,\tilde{q}_i) \quad (20)$$

where

$$\tilde{q}_i = q + \langle\alpha_i\rangle$$

In summary, the variational Bayesian inference consists of successive update of the approximate posterior distributions for hidden variables $\boldsymbol{x}$, $\boldsymbol{\alpha}$, $\boldsymbol{\gamma}$, and $\bar{\boldsymbol{b}}$. Some of the expectations and moments used during the update are summarized as

$$\begin{aligned}\langle\alpha_i\rangle = \frac{\tilde{a}}{\tilde{b}_i} \quad \langle\gamma\rangle = \frac{\tilde{c}}{\tilde{d}} \\ \langle x_i^2\rangle = \mu_i^2 + \phi_{i,i} \quad \langle b_i\rangle = \frac{p}{\tilde{q}_i}\end{aligned}$$

where μ_i denotes the ith element of $\boldsymbol{\mu}$, and $\phi_{i,i}$ denotes the ith diagonal element of $\boldsymbol{\Phi}$. We now summarize our algorithm as follows.

Partial Support Aided-SBL with Support Learning

1. Given the current approximate posterior distributions of $q_\alpha(\boldsymbol{\alpha})$, $q_\gamma(\gamma)$ and $q_{\bar{b}}(\bar{\boldsymbol{b}})$, update $q_x(\boldsymbol{x})$ according to (11).
2. Given $q_x(\boldsymbol{x})$, $q_\gamma(\gamma)$ and $q_{\bar{b}}(\bar{\boldsymbol{b}})$, update $q_\alpha(\boldsymbol{\alpha})$ according to (13)–(15).
3. Given $q_x(\boldsymbol{x})$, $q_\alpha(\boldsymbol{\alpha})$, and $q_{\bar{b}}(\bar{\boldsymbol{b}})$, update $q_\gamma(\gamma)$ according to (17)–(18).
4. Given $q_x(\boldsymbol{x})$, $q_\alpha(\boldsymbol{\alpha})$ and $q_\gamma(\gamma)$, update $q_{\bar{b}}(\bar{\boldsymbol{b}})$ according to (20).
5. Continue the above iteration until $\|\boldsymbol{\mu}^{(t)} - \boldsymbol{\mu}^{(t-1)}\|_2 \leq \epsilon$, where ϵ is a prescribed tolerance value. Choose $\hat{\boldsymbol{\mu}}^{(t)}$ as the estimate of the sparse signal.

IV. Simulation Results

We now carry out experiments to illustrate the performance of our proposed algorithm. The proposed algorithm is referred to as the support knowledge-aided sparse Bayesian learning with support learning (SA-SBL-SL). Also, by placing fixed values to $\{b_i\}$ according to (5), a Bayesian variational method can be readily developed (Details are omitted due to the space limitation) and is referred to as the support knowledge-aided

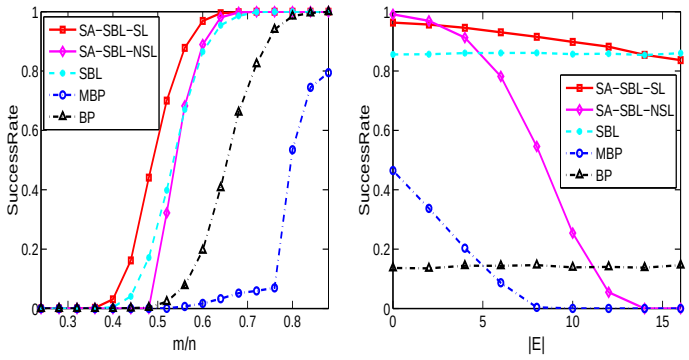

Fig. 1. (a). Success rates vs. the ratio m/n; (b). Success rates vs. the size of the error set.

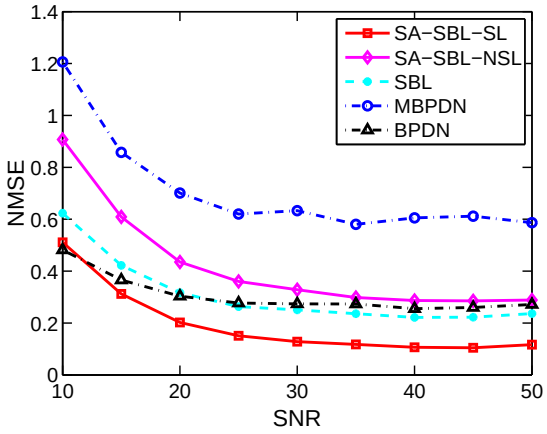

Fig. 2. NMSEs vs. the signal-to-noise ratio.

sparse Bayesian learning with no support learning (SA-SBL-NSL). In our simulations, the parameters p and q are set to be $p = 0.1$ and $q = 0.1$.

Suppose the K-sparse signal is randomly generated with the support set of the sparse signal randomly chosen according to a uniform distribution. The signals on the support set are independent and identically distributed (i.i.d.) Gaussian random variables with zero mean and unit variance. The measurement matrix $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ is randomly generated with each entry independently drawn from Gaussian distribution with zero mean and unit variance. The prior support information P consists of two subsets: $P = S \cup E$, where $S \subset T$ denotes the subset containing the correct information about the support, and $E \subset T^c$ is a subset comprised of false information. In our simulations, the partition of S and E is unknown. We compare our proposed algorithms with the conventional sparse Bayesian learning (SBL), the basis pursuit (BP) method, and the modified basis pursuit (MBP) method [1].

We first consider the noiseless case. Fig. 1(a) plots the success rates of respective algorithms vs. the ratio m/n, where we set $K = 16$, $n = 50$, $|S| = 12$ and $|E| = 8$, $|S|$ and $|E|$ denote the cardinality (size) of the set S and E, respectively. The success rate is computed as the ratio of the number of successful trials to the total number of independent runs. A trial is considered successful if the normalized recovery error, i.e. $\|\boldsymbol{x} - \hat{\boldsymbol{x}}\|_2^2/\|\boldsymbol{x}\|_2^2$, is no greater than 10^{-6}, where $\hat{\boldsymbol{x}}$ denotes the estimate of the true signal $\boldsymbol{x}$. Results are averaged over 1000 independent runs, with the measurement matrix and the sparse signal randomly generated for each run. It can be seen that our proposed SA-SBL-SL method presents a substantial performance advantage over the SA-SBL-NSL and the SBL methods. The performance gain is primarily due to the fact that the SA-SBL-SL method is able to learn the true support from the partly erroneous knowledge. We also observe that when a considerable amount of errors are present in the prior knowledge, the methods SA-SBL-NSL and MBP have no advantage over the methods SBL and BP. To examine the behavior of the SA-SBL-SL method more thoroughly, we fix the number of elements in the set S and increase the number of elements in the error set E. Fig. 1(b) depicts the success rates vs. the number of elements in the error set E, where we set $m = 25$, $|S| = 12$ and $|E|$ varies from 1 to 15. As can be seen from Fig., when a fairly accurate knowledge is available, the SA-SBL-NSL achieves the best performance. This is an expected result since little learning is required at this point. Nevertheless, as the number of elements, $|E|$, increases, the SA-SBL-NSL suffers from substantial performance degradation. As compared with the SA-SBL-NSL, the SA-SBL-SL method provides stable recovery performance through learning the values of $\{b_i\}$, and outperforms all other algorithms when prior knowledge contains a considerable amount of errors. We, however, notice that the proposed SA-SBL-SL method is surpassed by the conventional SBL method when inaccurate information becomes dominant (e.g. $|E| = 15$), in which case even learning brings limited benefits and simply ignoring the error-corrupted prior knowledge seems the best strategy.

We now consider the noisy case where the measurements are contaminated by additive Gaussian noise. The normalized mean-squared errors (NMSEs) as a function of signal-to-noise ratio (SNR) are plotted in Fig 2, where we set $m = 25$, $n = 50$, $K = 16$, $|S| = 12$, and $|E| = 6$. The MBP-DN is a noisy version of the MBP method [4]. We observe that the conventional SBL and BP-DN methods outperform their respective counterparts: SA-SBL-NSL and MBP-DN. This, again, demonstrates that SA-SBL-NSL and MBP-DN methods are sensitive to prior knowledge inaccuracies. On the other hand, the proposed SA-SBL-SL method which takes advantage of the support learning presents superiority over both the conventional SBL as well as the SA-SBL-NSL method.

V. Conclusions

We developed an improved sparse Bayesian learning method which is able to accommodate the prior support knowledge and learn the true support from partly erroneous information. Numerical results show that our proposed algorithm achieves a significant performance improvement through learning the underlying true information from partly erroneous knowledge.

REFERENCES

[1] N. Vaswani and W. Lu, "Modified-CS: modifying compressive sensing for problems with partially known support," *IEEE Trans. Signal Processing*, no. 9, pp. 4595–4607, September 2010.

[2] M. A. Khajehnejad, W. Xu, A. S. Avestimehr, and B. Hassibi, "Weighted ℓ_1 minimization for sparse recovery with prior information," in *Proc. IEEE Int. Symp. Information Theory*, Seoul, Korea, June 28–July 3 2009.

[3] C. J. Miosso, R. von Borries, M. Argaez, L. Velazquez, C. Quintero, and C. M. Potes, "Compressive sensing reconstruction with prior information by iteratively reweighted least-squares," *IEEE Trans. Signal Processing*, no. 6, pp. 2424–2431, June 2009.

[4] W. Lu and N. Vaswani, "Regularized modified BPDN for noisy sparse reconstruction with partial erroneous support and signal value knowledge," *IEEE Trans. Signal Processing*, no. 1, pp. 182–196, January 2012.

[5] M. Tipping, "Sparse Bayesian learning and the relevance vector machine," *Journal of Machine Learning Research*, vol. 1, pp. 211–244, 2001.

[6] D. P. Wipf, "Bayesian methods for finding sparse representations," Ph.D. dissertation, University of California, San Diego, 2006.

[7] S. Ji, Y. Xue, and L. Carin, "Bayesian compressive sensing," *IEEE Trans. Signal Processing*, vol. 56, no. 6, pp. 2346–2356, June 2008.