Fast Inverse-Free Sparse Bayesian Learning via Relaxed Evidence Lower Bound Maximization

Huiping Duan, Linxiao Yang, Jun Fang, and Hongbin Li

***Abstract*—Sparse Beyesian learning is a popular approach for sparse signal recovery, and has demonstrated superior performance in a series of experiments. Nevertheless, the sparse Bayesian learning algorithm involves a matrix inverse at each iteration. Its associated computational complexity grows significantly with the problem size, which hinders its application to many practical problems even with moderately large datasets. To address this issue, in this letter, we develop a fast inverse-free sparse Bayesian learning method. Specifically, by invoking a fundamental property for smooth functions, we obtain a relaxed evidence lower bound (relaxed-ELBO) that is computationally more amiable than the conventional ELBO used by sparse Bayesian learning. A variational expectation-maximization (EM) scheme is then employed to maximize the relaxed-ELBO, which leads to a computationally efficient inverse-free sparse Bayesian learning algorithm. Simulation results show that the proposed algorithm has a fast convergence rate and achieves lower reconstruction errors than other state-of-the-art fast sparse recovery methods in the presence of noise.**

***Index Terms*—Compressed sensing, inverse-free sparse Bayesian learning (SBL), relaxed evidence lower bound (relaxed-ELBO).**

I. Introduction

COMPRESSED sensing is a recently emerged technique for signal sampling and data acquisition, which enables to recover sparse signals from undersampled linear measurements

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

where $\boldsymbol{A} \in \mathbb{R}^{M\times N}$ is the sampling matrix with $M \ll N$, $\boldsymbol{x}$ denotes an N-dimensional sparse signal, and $\boldsymbol{w}$ denotes the additive noise. Such a problem has been extensively studied and a variety of algorithms, e.g., the orthogonal matching pursuit algorithm [1], the basis pursuit (BP) method [2], and the iterative reweighted ℓ_1 and ℓ_2 algorithms [3], were proposed. Besides these methods, another important class of compressed sensing techniques that have received significant attention are Bayesian methods, among which sparse Bayesian learning (SBL) [4], [5], [8] is considered as one of the most popular compressed sensing methods. SBL was found to outperform the greedy methods and the BP method in a series of experiments, e.g., [6] and [7]. Despite its superior performance, a major drawback of the SBL method is that it requires to compute an inverse of an $M \times M$ matrix at each iteration, and thus has a cubic complexity in terms of the problem size. This high computational cost prohibits its application to many practical problems. To address this issue, a fast SBL scheme was proposed in [9] to alleviate this drawback. The key idea is to perform the type-II maximum likelihood estimation in a sequential manner, i.e., maximizing the marginal likelihood function with respect to a single sparsity parameter α_n at each time, assuming other sparsity parameters are fixed. In [10], the idea of sequential maximization was adapted to the variational Bayesian framework, where authors considered the maximization of the variational lower bound with respect to a single factor $q(\alpha_n)$. Besides the above efforts, recently the approximate message passing (AMP) [11] and generalized AMP (GAMP) [12] techniques were developed in a message-passing-based framework to compute approximate marginal posteriors with a very low complexity, and have been successfully applied to solve compressed sensing problems, e.g., [13] and [14].

In this letter, we develop an inverse-free SBL method that has a very low computational complexity. The rationale behind our proposed method is to find a relaxed evidence lower bound (relaxed-ELBO) that can help circumvent the matrix inverse operation. A variational expectation-maximization (EM) scheme is developed to maximize the relaxed-ELBO, which leads to a computationally efficient inverse-free SBL algorithm. Note that our algorithm development is similar to [8], with the key difference that the ELBO is replaced by a relaxed-ELBO, which results in an inverse-free algorithm.

II. Overview of Variational Inference

We first provide a review of the variational Bayesian inference. Let $\boldsymbol{\theta} \triangleq \{\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_I\}$ denote the hidden variables in the hierarchical model. The objective of Bayesian inference is to find the posterior distribution of the latent variables given the observed data, i.e., $p(\boldsymbol{\theta}|\boldsymbol{y})$. The computation of $p(\boldsymbol{\theta}|\boldsymbol{y})$, however, is usually intractable. To address this difficulty, in variational inference, the posterior distribution $p(\boldsymbol{\theta}|\boldsymbol{y})$ is approximated by a variational distribution $q(\boldsymbol{\theta})$ that has a factorized form as [15]

$$q(\boldsymbol{\theta}) = \prod_{i=1}^{I} q_i(\boldsymbol{\theta}_i). \tag{2}$$

Variational inference finds $q(\boldsymbol{\theta})$ by minimizing the Kullback–Leibler (KL) divergence from the variational distribution $q(\boldsymbol{\theta})$ to the posterior distribution $p(\boldsymbol{\theta}|\boldsymbol{y})$. The KL divergence minimization is equivalent to maximizing the evidence lower bound (ELBO), a lower bound on the logarithm of the marginal

Manuscript received January 4, 2017; revised February 17, 2017; accepted March 24, 2017. Date of publication April 7, 2017; date of current version April 24, 2017. This work was supported in part by the National Science Foundation of China under Grant 61522104, Grant U1530154, and Grant 61201274, and the 111 Project B17008. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Qing Ling. *(Corresponding author: Jun Fang.)*

H. Duan is with the School of Electronic Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China (e-mail: huipingduan@uestc.edu.cn).

L. Yang and J. Fang are with the National Key Laboratory of Science and Technology on Communications, University of Electronic Science and Technology of China, Chengdu 611731, China (e-mail: JunFang@uestc.edu.cn).

H. Li is with the Department of Electrical and Computer Engineering, Stevens Institute of Technology, Hoboken, NJ 07030 USA (e-mail: Hongbin.Li@stevens.edu).

Color versions of one or more of the figures in this letter are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/LSP.2017.2692217

probability of the observed data $\ln p(\boldsymbol{y})$. To see this, note that the marginal probability of the observed data can be decomposed into two terms

$$\ln p(\boldsymbol{y}) = L(q) + \mathrm{KL}(q||p) \tag{3}$$

where

$$L(q) = \int q(\boldsymbol{\theta}) \ln \frac{p(\boldsymbol{y}, \boldsymbol{\theta})}{q(\boldsymbol{\theta})} d\boldsymbol{\theta} \tag{4}$$

and

$$\mathrm{KL}(q||p) = -\int q(\boldsymbol{\theta}) \ln \frac{p(\boldsymbol{\theta}|\boldsymbol{y})}{q(\boldsymbol{\theta})} d\boldsymbol{\theta} \tag{5}$$

where $\mathrm{KL}(q||p)$ is the KL divergence between $p(\boldsymbol{\theta}|\boldsymbol{y})$ and $q(\boldsymbol{\theta})$, and $L(q)$ is the ELBO. Since $\mathrm{KL}(q||p) \geq 0$, it follows that $L(q)$ is a rigorous lower bound on $\ln p(\boldsymbol{y})$. Moreover, notice that the left-hand side of (3) is independent of $q(\boldsymbol{\theta})$. Therefore, maximizing $L(q)$ is equivalent to minimizing $\mathrm{KL}(q||p)$, and thus the posterior distribution $p(\boldsymbol{\theta}|\boldsymbol{y})$ can be approximated by the variational distribution $q(\boldsymbol{\theta})$ through maximizing $L(q)$. The ELBO maximization can be conducted in an alternating fashion for each latent variable, which leads to [15]

$$q_i(\boldsymbol{\theta}_i) = \frac{\exp(\langle \ln p(\boldsymbol{y}, \boldsymbol{\theta}) \rangle_{k \neq i})}{\int \exp(\langle \ln p(\boldsymbol{y}, \boldsymbol{\theta}) \rangle_{k \neq i}) d\boldsymbol{\theta}_i} \tag{6}$$

where $\langle \cdot \rangle_{k \neq i}$ denotes an expectation with respect to the distributions $q_k(\boldsymbol{\theta}_k)$ for all $k \neq i$.

III. Inverse-Free SBL

In the variational sparse Bayesian learning framework, $\boldsymbol{x}$ is assigned a two-layer hierarchical prior. In the first layer, $\boldsymbol{x}$ is assigned a Gaussian prior distribution characterized by $\boldsymbol{\alpha}$, i.e.,

$$p(\boldsymbol{x}|\boldsymbol{\alpha}) = \prod_{n=1}^{N} p(x_n|\alpha_n) = \prod_{n=1}^{N} \mathcal{N}(x_n|0, \alpha_n^{-1}) \tag{7}$$

where $\boldsymbol{\alpha} \triangleq \{\alpha_n\}$ are nonnegative hyperparameters controlling the sparsity of the signal $\boldsymbol{x}$. The second layer specifies Gamma distributions as hyperpriors over the hyperparameters $\{\alpha_n\}$, i.e.,

$$p(\boldsymbol{\alpha}) = \prod_{n=1}^{N} \mathrm{Gamma}(\alpha_n|a, b) = \prod_{n=1}^{N} \Gamma^{-1}(a) b^a \alpha_n^{a-1} e^{-b\alpha_n}$$

where $\Gamma(a) = \int_0^\infty t^{a-1} e^{-t} dt$ is the Gamma function. Also, $\boldsymbol{w}$ is assumed to follow a Gaussian distribution with zero mean and covariance matrix $(1/\gamma)\boldsymbol{I}$. To learn γ, a Gamma hyperprior is placed over γ, i.e.,

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

Let $\boldsymbol{\theta} \triangleq \{\boldsymbol{x}, \boldsymbol{\alpha}, \gamma\}$ denote the hidden variables in the above hierarchical model, and let the variational distribution be expressed as $q(\boldsymbol{\theta}) = q_x(\boldsymbol{x}) q_\alpha(\boldsymbol{\alpha}) q_\gamma(\gamma)$. Maximizing the ELBO yields a procedure that involves updates of the approximate posterior distributions for hidden variables $\boldsymbol{x}$, $\boldsymbol{\alpha}$, and γ in an alternating fashion [4], [5]. Specifically, $q_x(\boldsymbol{x})$ is updated according to a Gaussian distribution with its mean and covariance matrix given, respectively, by

$$\boldsymbol{\mu} = \gamma \boldsymbol{\Phi} \boldsymbol{A}^T \boldsymbol{y} \tag{9}$$

$$\boldsymbol{\Phi} = (\gamma \boldsymbol{A}^T \boldsymbol{A} + \boldsymbol{D})^{-1} \tag{10}$$

where $\boldsymbol{D}$ is a diagonal matrix with its nth diagonal element equal to $\langle \alpha_n \rangle$. We see that the update of the posterior distribution $q_x(\boldsymbol{x})$ involves computing an $N \times N$ matrix inverse, which can be converted to an $M \times M$ matrix inverse by resorting to the Woodbury identity. Thus, the variational SBL algorithm has a computational complexity scaling as $\mathcal{O}(M^3)$, which makes the application of the SBL method to large datasets impractical.

To address this issue, we propose to maximize a relaxed ELBO, i.e., a lower bound on $L(q)$. Note that the ELBO for SBL can be expressed as

$$\begin{aligned} L(q) &= \int q(\boldsymbol{\theta}) \ln \frac{p(\boldsymbol{y}, \boldsymbol{\theta})}{q(\boldsymbol{\theta})} d\boldsymbol{\theta} \\ &= \int q(\boldsymbol{\theta}) \ln \frac{p(\boldsymbol{y}|\boldsymbol{x}, \gamma) p(\boldsymbol{x}|\boldsymbol{\alpha}) p(\boldsymbol{\alpha}) p(\gamma)}{q(\boldsymbol{\theta})} d\boldsymbol{\theta}. \end{aligned} \tag{11}$$

To obtain a relaxed ELBO, we recall the following well known and fundamental property for a smooth function [16]. Note that this property also plays a critical role in developing the fast first-order algorithm in [16].

Lemma 1: Let $f : \mathbb{R}^n \to \mathbb{R}$ be a continuously differentiable function with Lipschitz continuous gradient and Lipschitz constant $T(f)$. Then, for any $T \geq T(f)$, we have

$$f(\boldsymbol{u}) \leq f(\boldsymbol{v}) + (\boldsymbol{u} - \boldsymbol{v})^T \nabla f(\boldsymbol{v}) + \frac{T}{2} \|\boldsymbol{u} - \boldsymbol{v}\|_2^2 \tag{12}$$

for any $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^n$.

Invoking Lemma 1, a lower bound on $p(\boldsymbol{y}|\boldsymbol{x}, \gamma)$ can be obtained as

$$\begin{aligned} p(\boldsymbol{y}|\boldsymbol{x}, \gamma) &= \frac{\gamma^{N/2}}{\sqrt{2\pi}} \exp\left(-\frac{\gamma}{2} \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_2^2\right) \\ &\geq \frac{\gamma^{N/2}}{\sqrt{2\pi}} \exp\left(-\frac{\gamma}{2} g(\boldsymbol{x}, \boldsymbol{z})\right) \triangleq F(\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{z}, \gamma) \end{aligned} \tag{13}$$

where

$$g(\boldsymbol{x}, \boldsymbol{z}) \triangleq \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{z}\|_2^2 + 2(\boldsymbol{x} - \boldsymbol{z})^T \boldsymbol{A}^T (\boldsymbol{A}\boldsymbol{z} - \boldsymbol{y}) + \frac{T}{2} \|\boldsymbol{x} - \boldsymbol{z}\|_2^2. \tag{14}$$

Clearly, the inequality in (13) holds for any $\boldsymbol{z}$, and the inequality becomes equality when $\boldsymbol{x} = \boldsymbol{z}$.

Combining (11) and (13), a relaxed ELBO can eventually be obtained as

$$L(q) \geq \tilde{L}(q, \boldsymbol{z}) = \int q(\boldsymbol{\theta}) \ln \frac{G(\boldsymbol{y}, \boldsymbol{\theta}, \boldsymbol{z})}{q(\boldsymbol{\theta})} d\boldsymbol{\theta} \tag{15}$$

where

$$G(\boldsymbol{y}, \boldsymbol{\theta}, \boldsymbol{z}) \triangleq F(\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{z}, \gamma) p(\boldsymbol{x}|\boldsymbol{\alpha}) p(\boldsymbol{\alpha}) p(\gamma).$$

The relaxed ELBO can be further expressed as

$$\begin{aligned} \tilde{L}(q, \boldsymbol{z}) &= \int q(\boldsymbol{\theta}) \ln \frac{G(\boldsymbol{y}, \boldsymbol{\theta}, \boldsymbol{z})}{q(\boldsymbol{\theta})} d\boldsymbol{\theta} \\ &= \int q(\boldsymbol{\theta}) \ln \frac{G(\boldsymbol{y}, \boldsymbol{\theta}, \boldsymbol{z}) h(\boldsymbol{z})}{q(\boldsymbol{\theta}) h(\boldsymbol{z})} d\boldsymbol{\theta} \\ &= \int q(\boldsymbol{\theta}) \ln \frac{\tilde{G}(\boldsymbol{y}, \boldsymbol{\theta}, \boldsymbol{z})}{q(\boldsymbol{\theta})} d\boldsymbol{\theta} - \ln h(\boldsymbol{z}) \end{aligned} \tag{16}$$

where $\tilde{G}(\boldsymbol{y}, \boldsymbol{\theta}, \boldsymbol{z}) \triangleq G(\boldsymbol{y}, \boldsymbol{\theta}, \boldsymbol{z}) h(\boldsymbol{z})$, $h(\boldsymbol{z})$ is a normalizing term to make $\tilde{G}(\boldsymbol{y}, \boldsymbol{\theta}, \boldsymbol{z})$ a rigorous distribution, and $h(\boldsymbol{z})$ is given by

$$h(\boldsymbol{z}) \triangleq \frac{1}{\int G(\boldsymbol{y}, \boldsymbol{\theta}, \boldsymbol{z}) d\boldsymbol{\theta} d\boldsymbol{y}}. \tag{17}$$

Recall $q(\boldsymbol{\theta}) = q_x(\boldsymbol{x}) q_\alpha(\boldsymbol{\alpha}) q_\gamma(\gamma)$. Our objective is to maximize the relaxed ELBO $\tilde{L}(q, \boldsymbol{z})$ with respect to $q_x(\boldsymbol{x})$, $q_\alpha(\boldsymbol{\alpha})$, $q_\gamma(\gamma)$, as well as with respect to the parameter $\boldsymbol{z}$. This naturally leads to a variational EM algorithm. In the E-step, the posterior distribution approximations are computed in an alternating fashion for each hidden variable, with other variables fixed. In the M-step, $\tilde{L}(q, \boldsymbol{z})$ is maximized with respect to $\boldsymbol{z}$, given $q(\boldsymbol{\theta})$ fixed. Details of the Bayesian inference are provided below.

A. E-Step

1) Update of $q_x(\boldsymbol{x})$: Recalling (6) and ignoring those terms that are independent of $\boldsymbol{x}$, the approximate posterior distribution $q_x(\boldsymbol{x})$ can be computed by

$$\begin{aligned}\ln q_x(\boldsymbol{x}) &\propto \langle \ln \tilde{G}(\boldsymbol{y},\boldsymbol{\theta},\boldsymbol{z})\rangle_{q_\alpha(\boldsymbol{\alpha}),q_\gamma(\gamma)} \\ &\propto \langle \ln F(\boldsymbol{y},\boldsymbol{x},\boldsymbol{z},\gamma) + \ln p(\boldsymbol{x}|\boldsymbol{\alpha})\rangle_{q_\alpha(\boldsymbol{\alpha}),q_\gamma(\gamma)} \\ &\propto -\boldsymbol{x}^T\left(\frac{T\langle\gamma\rangle}{2}\boldsymbol{I}+\boldsymbol{\Lambda}\right)\boldsymbol{x} + \\ &\quad \langle\gamma\rangle\boldsymbol{x}^T(2\boldsymbol{A}^T(\boldsymbol{A}\boldsymbol{z}-\boldsymbol{y})-T\boldsymbol{z}) \end{aligned} \tag{18}$$

in which $\boldsymbol{\Lambda} \triangleq \text{diag}(\langle\alpha_1\rangle,\ldots,\langle\alpha_N\rangle)$, $\langle\alpha_n\rangle$ denotes the expectation of α_n with respect to $q_\alpha(\boldsymbol{\alpha})$. Clearly, $q_x(\boldsymbol{x})$ follows a Gaussian distribution with its mean and covariance matrix given, respectively, as

$$\boldsymbol{\mu} = \langle\gamma\rangle\boldsymbol{\Sigma}\left(\boldsymbol{A}^T\boldsymbol{A}\boldsymbol{z}-\boldsymbol{A}^T\boldsymbol{y}-\frac{T}{2}\boldsymbol{z}\right) \tag{19}$$

$$\boldsymbol{\Sigma} = \left(\frac{T\langle\gamma\rangle}{2}\boldsymbol{I}+\boldsymbol{\Lambda}\right)^{-1}. \tag{20}$$

2) Update of $q_\alpha(\boldsymbol{\alpha})$: The approximate posterior $q_\alpha(\boldsymbol{\alpha})$ can be obtained as

$$\begin{aligned}\ln q_\alpha(\boldsymbol{\alpha}) &\propto \langle \ln \tilde{G}(\boldsymbol{y},\boldsymbol{\theta},\boldsymbol{z})\rangle_{q_x(\boldsymbol{x})} \\ &\propto \langle \ln p(\boldsymbol{x}|\boldsymbol{\alpha}) + \ln p(\boldsymbol{\alpha})\rangle_{q_x(\boldsymbol{x})} \\ &\propto \sum_{n=1}^{N}\left\{\left(a-\frac{1}{2}\right)\ln\alpha_n - \left(\frac{\langle x_n^2\rangle}{2}+b\right)\alpha_n\right\} \end{aligned} \tag{21}$$

where $\langle x_n^2\rangle$ denotes the expectation of x_n^2 with respect to $q_x(\boldsymbol{x})$. Hence, $\boldsymbol{\alpha}$ has a form of a product of Gamma distributions

$$q_\alpha(\boldsymbol{\alpha}) = \prod_{n=1}^{N}\text{Gamma}(\alpha_n;\tilde{a},\tilde{b}_n) \tag{22}$$

in which the parameters $\tilde{a}$ and $\tilde{b}_n$ are, respectively, given as

$$\tilde{a} = a+\frac{1}{2} \qquad \tilde{b}_n = b+\frac{1}{2}\langle x_n^2\rangle. \tag{23}$$

3) Update of $q_\gamma(\gamma)$: Similarly, variational optimization of $q_\gamma(\gamma)$ yields

$$\begin{aligned}\ln q_\gamma(\boldsymbol{\gamma}) &\propto \langle \ln \tilde{G}(\boldsymbol{y},\boldsymbol{\theta},\boldsymbol{z})\rangle_{q_x(\boldsymbol{x})} \\ &\propto \langle \ln F(\boldsymbol{y},\boldsymbol{x},\boldsymbol{z},\gamma) + \ln p(\gamma)\rangle_{q_x(\boldsymbol{x})} \\ &\propto \left(c-1+\frac{M}{2}\right)\ln\gamma - \left(\frac{1}{2}\langle g(\boldsymbol{x},\boldsymbol{z})\rangle + d\right)\gamma. \end{aligned} \tag{24}$$

Thus, γ follows a Gamma distribution

$$q_\gamma(\gamma) = \text{Gamma}(\gamma;\tilde{c},\tilde{d}) \tag{25}$$

with the parameters $\tilde{c}$ and $\tilde{d}$ given as

$$\begin{aligned}\tilde{c} &= c+\frac{M}{2} \\ \tilde{d} &= d+\frac{1}{2}\langle g(\boldsymbol{x},\boldsymbol{z})\rangle. \end{aligned} \tag{26}$$

In summary, the E-step involves update of the posterior approximations for hidden variables $\boldsymbol{x}$, $\boldsymbol{\alpha}$, and γ. Some of the expectations and moments used during the update are summarized as

$$\begin{aligned}&\langle\alpha_n\rangle = \frac{\tilde{a}}{\tilde{b}_n} \quad \langle\gamma\rangle = \frac{\tilde{c}}{\tilde{d}} \quad \langle x_n^2\rangle = \mu_n^2+\Sigma_{n,n} \\ &\langle g(\boldsymbol{x},\boldsymbol{z})\rangle = \|\boldsymbol{y}-\boldsymbol{A}\boldsymbol{z}\|_2^2 + 2(\boldsymbol{\mu}-\boldsymbol{z})^T\boldsymbol{A}^T(\boldsymbol{A}\boldsymbol{z}-\boldsymbol{y}) \\ &\qquad + \frac{T}{2}(\|\boldsymbol{\mu}-\boldsymbol{z}\|_2^2 + \text{Tr}(\boldsymbol{\Sigma})) \end{aligned} \tag{27}$$

where μ_n denotes the nth element of $\boldsymbol{\mu}$ and $\Sigma_{n,n}$ represents the nth diagonal element of $\boldsymbol{\Sigma}$.

B. M-Step

Substituting $q(\boldsymbol{\theta};\boldsymbol{z}^{\text{old}})$ into $\tilde{L}(q,\boldsymbol{z})$, an estimate of $\boldsymbol{z}$ can be found via the following optimization

$$\boldsymbol{z}^{\text{new}} = \arg\max_{\boldsymbol{z}} \langle \ln G(\boldsymbol{y},\boldsymbol{\theta},\boldsymbol{z})\rangle_{q(\boldsymbol{\theta};\boldsymbol{z}^{\text{old}})} \triangleq Q(\boldsymbol{z}|\boldsymbol{z}^{\text{old}}). \tag{28}$$

Taking the derivative of $Q(\boldsymbol{z}|\boldsymbol{z}^{\text{old}})$ with respect to $\boldsymbol{z}$ yields

$$\frac{\partial Q(\boldsymbol{z}|\boldsymbol{z}^{\text{old}})}{\partial \boldsymbol{z}} = \langle (T\boldsymbol{I}-2\boldsymbol{A}^T\boldsymbol{A})(\boldsymbol{z}-\boldsymbol{x})\rangle_{q(\boldsymbol{\theta};\boldsymbol{z}^{\text{old}})}. \tag{29}$$

Setting the derivative to zero gives the solution

$$\boldsymbol{z} = \boldsymbol{\mu} \tag{30}$$

which follows from the fact that $T > T(f) = 2\lambda_{\max}(\boldsymbol{A}^T\boldsymbol{A})$ and $T\boldsymbol{I}-2\boldsymbol{A}^T\boldsymbol{A}$ is a positive-definite matrix, where $\lambda_{\max}(\boldsymbol{X})$ denotes the largest eigenvalue of matrix $\boldsymbol{X}$.

For clarity, we summarize our proposed algorithm as follows.

1) Given the current estimate of $\boldsymbol{z}$, update the posterior approximations $q_x(\boldsymbol{x})$, $q_\alpha(\boldsymbol{\alpha})$, and $q_\gamma(\boldsymbol{\gamma})$ according to (19)–(20), (22), and (25).
2) Update the parameter $\boldsymbol{z}$ according to (30).
3) Continue the above iteration until $\|\boldsymbol{\mu}^{(t)}-\boldsymbol{\mu}^{(t-1)}\|_2 \le \epsilon$, where ϵ is a prescribed tolerance value.

Discussions: For our proposed method, we see that the update of $q_x(\boldsymbol{x})$ still requires to calculate an $N\times N$ matrix inverse. Nevertheless, the matrix to be inverted [cf., (20)] now becomes a diagonal matrix. Thus, our proposed method is free of any matrix inverse operations, and as a consequence, the computational complexity of our proposed method is substantially reduced. Note that the update of other hidden variables $\boldsymbol{\alpha}$, γ, and the parameter $\boldsymbol{z}$ only involves simple addition and multiplication operations.

IV. Simulation Results

We present simulation results to illustrate the performance of our proposed inverse-free sparse Bayesian learning method (referred to as IF-SBL).[1] Throughout our experiments, a, b, c, and d are all set to 10^{-10}. T is chosen to be slightly larger than the smallest Lipschitz constant, i.e., $T = \lambda_{\max}(2\boldsymbol{A}^T\boldsymbol{A})+10^{-10}$, since simulation results suggest that a smaller value of T leads to a faster convergence rate. We compare our method with the fast SBL (referred to as F-SBL) [9], the fast iterative shrinkage-thresholding algorithm (referred to as FISTA) [16], and an efficient GAMP-based method (referred to as SBL-GAMP) that is a simplified version of [14] without considering a block structure. The regularization parameter used in FISTA is carefully selected to achieve the best performance. Due to the large problem size, the conventional SBL [4] was not included in our experiments. Nevertheless, numerical results on small datasets suggest that its performance is similar to that of the F-SBL.

[1] Codes for our algorithm are available at http://www.junfang-uestc.net/codes/IF-SBL.rar

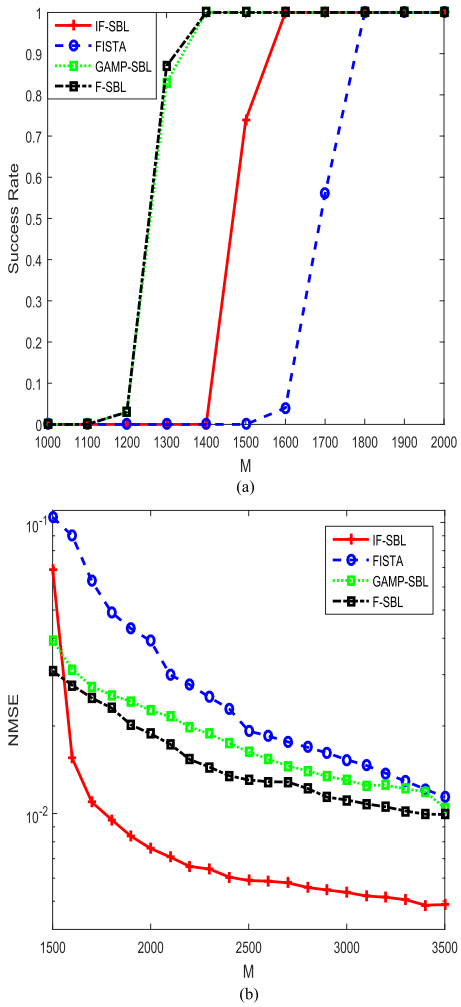

Fig. 1. (a). Success rates of respective algorithms versus M; (b). NMSEs of respective algorithms versus M.

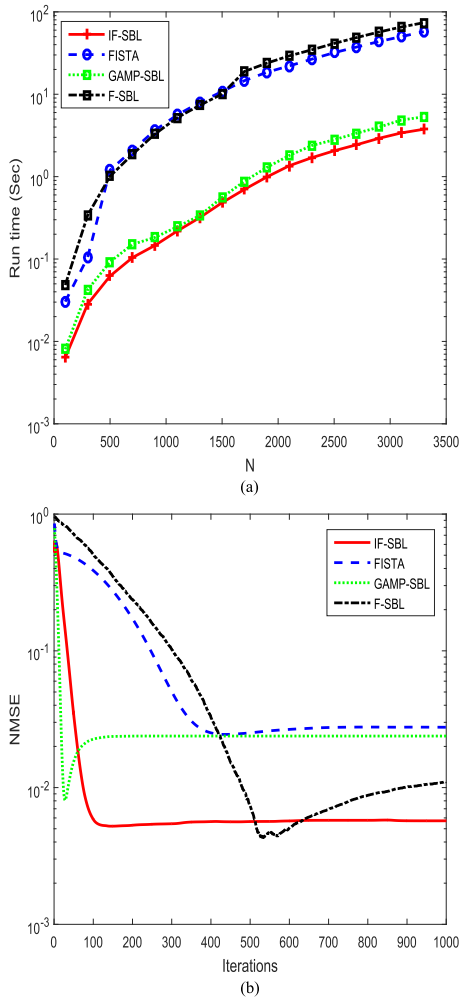

Fig. 2. (a). Average run times of respective algorithms versus N; (b). Convergence behavior of respective algorithms.

In our simulations, the sparse signal is randomly generated with its K nonzero entries generated according to a normal distribution. The measurement matrix $\boldsymbol{A} \in \mathbb{R}^{M\times N}$ is randomly generated with its entries independently drawn from a normal distribution. Fig. 1(a) depicts the recovery success rates of respective algorithms as a function of M, where $N = 5000$, $K = 500$, and the observation $\boldsymbol{y}$ in (1) contains no measurement noise. Results are averaged over 10^3 independent runs. We see that in the noiseless case, the F-SBL and the SBL-GAMP provides the best performance. Due to the relaxation made to the ELBO, the IF-SBL method incurs some performance loss. Nevertheless, our proposed IF-SBL method presents a clear performance advantage over the FISTA. Fig. 1(b) shows the normalized mean square errors (NMSEs) of respective algorithms in the presence of Gaussian noise, where we set the signal-to-noise ratio (SNR) to 20 dB. From Fig. 1(b), we see that the IF-SBL method exhibits better robustness against noise and achieves a lower reconstruction error for a moderate ratio M/N.

In Fig. 2(a), we plot the average run times of respective algorithms as a function of N, where we set $M = N/2$, $K = N/10$, and SNR $= 20$ dB. In such a setup, all methods can provide reliable recovery results. We see that similar to the SBL-GAMP, the average run time required by the IF-SBL only incurs a slight increase as the dimension of the signal becomes large, whereas the F-SBL and the FISTA methods suffer from a dramatic increase in the average run time as N increases. The reason is that the IF-SBL has a much faster convergence rate than the other two methods, as observed from Fig. 2(b). In particular, the F-SBL method starts with an empty support set, and then iteratively includes new basis to fit the data. This process may take much longer to converge in the noisy case. From Fig. 2(b), we see that it takes the F-SBL more than 1000 iterations to converge, the FISTA about 600 iterations to converge, while the IF-SBL method converges within only 120 iterations. Also, the F-SBL method needs to perform a matrix inverse at each iteration. As the iterative process evolves, the dimension of the matrix to be inverted becomes larger and larger. Thus, it is no wonder that the average run time of the F-SBL method increases quickly as the signal dimension becomes large.

V. Conclusion

We developed a fast inverse-free SBL method in this letter. To circumvent the matrix inverse operation required by the conventional SBL method, we resort to a fundamental property for smooth functions and obtain a relaxed-ELBO. A variational EM scheme was then used to maximize the relaxed-ELBO, which leads to a computationally efficient inverse-free SBL algorithm. Simulation results showed that the proposed algorithm has a fast convergence rate and is more robust against noise than other state-of-the-art fast sparse signal recovery methods.

REFERENCES

[1] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4655–4666, Dec. 2007.

[2] E. Candés and T. Tao, "Decoding by linear programming," *IEEE Trans. Inf. Theory*, vol. 51, no. 12, pp. 4203–4215, Dec. 2005.

[3] D. Wipf and S. Nagarajan, "Iterative reweighted ℓ_1 and ℓ_2 methods for finding sparse solutions," *IEEE J. Sel. Topics Signal Process.*, vol. 4, no. 2, pp. 317–329, Apr. 2010.

[4] M. Tipping, "Sparse Bayesian learning and the relevance vector machine," *J. Mach. Learn. Res.*, vol. 1, pp. 211–244, 2001.

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

[6] Z. Zhang and B. 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, Sep. 2011.

[7] Z. Yang, L. Xie, and C. Zhang, "Off-grid direction of arrival estimation using sparse Bayesian inference," *IEEE Trans. Signal Process.*, vol. 61, no. 1, pp. 38–42, Jan. 2013.

[8] C. M. Bishop and M. E. Tipping, "Variational relevance vector machines," in *Proc. 16th Conf. Uncertainty Artif. Intell.*, Jun. 30–Jul. 4, 2000, pp. 46–53.

[9] M. Tipping and A. C. Faul, "Fast marginal likelihood maximisation for sparse Bayesian models," in *Proc. 9th Int. Workshop Artif. Intell. Statist.*, Key West, FL, USA, Jan. 3–6, 2003.

[10] D. Shutin, T. Buchgraber, S. R. Kulkarni, and H. V. Poor, "Fast variational sparse Bayesian learning with automatic relevance determination for superimposed signals," *IEEE Trans. Signal Process.*, vol. 59, no. 12, pp. 6257–6261, Dec. 2011.

[11] D. L. Donoho, A. Maleki, and A. Montanari, "Message passing algorithms for compressed sensing: I. Motivation and construction," in *Proc. Inf. Theory Workshop*, Cairo, Egypt, Jan. 2010.

[12] S. Rangan, "Generalized approximate message passing for estimation with random linear mixing," in *Proc. IEEE Int. Symp. Inf. Theory*, Saint Petersburg, Russia, Aug. 2011, pp. 2168–2172.

[13] S. Som and P. Schniter, "Compressive imaging using approximate message passing and a Markov-tree prior," *IEEE Trans. Signal Process.*, vol. 60, no. 7, pp. 3439–3448, Jul. 2012.

[14] J. Fang, L. Zhang, and H. Li, "Two-dimensional pattern-coupled sparse Bayesian learning via generalized approximate message passing," *IEEE Trans. Image Process.*, vol. 25, no. 6, pp. 2920–2930, Jun. 2016.

[15] D. G. Tzikas, A. C. Likas, and N. P. Galatsanos, "The variational approximation for Bayesian inference," *IEEE Signal Process. Mag.*, vol. 29, no. 6, pp. 131–146, Nov. 2008.

[16] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM J. Imag. Sci.*, vol. 2, no. 1, pp. 183–202, Mar. 2009.