

Research Article Block Sparse Bayesian Recovery with Correlated LSM Prior

Juan Zhao 🕑, Xia Bai, Tao Shan, and Ran Tao

School of Information and Electronics, Beijing Institute of Technology, Beijing 100081, China

Correspondence should be addressed to Juan Zhao; juanzhao@bit.edu.cn

Received 5 March 2021; Accepted 24 August 2021; Published 6 October 2021

Academic Editor: Liangtian Wan

Copyright © 2021 Juan Zhao et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Compressed sensing can recover sparse signals using a much smaller number of samples than the traditional Nyquist sampling theorem. Block sparse signals (BSS) with nonzero coefficients occurring in clusters arise naturally in many practical scenarios. Utilizing the sparse structure can improve the recovery performance. In this paper, we consider recovering arbitrary BSS with a sparse Bayesian learning framework by inducing correlated Laplacian scale mixture (LSM) prior, which can model the dependence of adjacent elements of the block sparse signal, and then a block sparse Bayesian learning algorithm is proposed via variational Bayesian inference. Moreover, we present a fast version of the proposed recovery algorithm, which does not involve the computation of matrix inversion and has robust recovery performance in the low SNR case. The experimental results with simulated data and ISAR imaging show that the proposed algorithms can efficiently reconstruct BSS and have good antinoise ability in noisy environments.

1. Introduction

Compressed sensing (CS) [1] provides a new sampling and reconstruction paradigm, which can recover sparse signals from linear measurements:

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

where $\mathbf{\Phi} \in \mathbb{R}^{M \times N}$ (M < N) is the measurement matrix, $\mathbf{y} \in \mathbb{R}^{M}$ is the measurement vector, $\mathbf{x} \in \mathbb{R}^{N}$ is the sparse signal, and $\mathbf{n} \in \mathbb{R}^{M}$ is the additive noise. Many recovery algorithms have been presented to reconstruct sparse signals, including orthogonal matching pursuit (OMP) [2] and sparse Bayesian learning (SBL) [3].

In some signal processing applications such as ISAR imaging [4] and gene expression levels [5], there are many sparse signals with block structural features; i.e., nonzero elements are often clustered. Inducing structural a priori can largely improve the reconstruction performance. Therefore, to improve the reconstruction effect of block sparse signals (BSS), many algorithms are proposed. For instance, Block-OMP [6] and Block-StOMP [7] are OMP-based approaches. Meanwhile, some block recovery algorithms based on the Bayesian compressed sensing framework are presented, including block sparse Bayesian learning (BSBL) [8], Cluss-MCMC [9], model-based Bayesian CS via local beta process (MBCS-LBP) [10], and pattern-coupled sparse Bayesian learning (PC-SBL) [11]. Among these recovery algorithms, Bayesian algorithms have parameter learning ability and can be applied to recovery arbitrary signals with unknown sparse structures by flexibly imposing different sparse prior models.

In [12], Zhang et al. have proposed an expectationmaximization-based variational Bayesian (EM-VB) inference method, which utilizes the Laplacian scale mixture (LSM) model as a sparse prior; i.e., it is assumed that the sparse signal obeys the Laplacian prior because the Laplacian distribution can represent sparseness well. Based on this model, for the BSS with unknown block information, this paper proposes a block Bayesian recovery algorithm by inducing a correlated LSM prior model, which uses the dependence between neighboring elements of the BSS. Furthermore, to improve the computational efficiency of the proposed recovery algorithm, a fast version without matrix inversion is presented, which is suitable for noisy environments, especially in the low SNR case. The experimental results on simulated data and ISAR imaging show that the proposed algorithms have a good reconstruction effect on BSS and can resist noise in noisy environments.

The remainder of this paper is organized as follows. In Section 2, a correlated LSM prior model for BSS is given. Then, the proposed block Bayesian recovery algorithm and the fast version are derived in Section 3. Simulation experiments are presented in Section 4. Finally, we conclude this paper in Section 5.

2. Signal Model

In the framework of sparse Bayesian learning, for the measurement model shown in (1), the noise **n** is generally assumed to obey a Gaussian prior distribution $\mathcal{N}(\mathbf{n} | \mathbf{0}, \gamma^{-1}\mathbf{I}_M)$, and a Gamma distribution for the hyperparameter γ is

$$p(\gamma; a, b) = G(\gamma; a, b) = \frac{b^a}{\Gamma(a)} \gamma^{a-1} \exp(-b\gamma), \qquad (2)$$

where $\Gamma(a) = \int_0^{+\infty} x^{a-1} \exp(-x) dx$. The sparse signal **x** is usually assumed to obey a sparse prior distribution. In the LSM layered prior model [12], the sparse signal **x** is supposed to follow a Laplacian prior distribution:

$$p(\mathbf{x} \mid \boldsymbol{\lambda}) = \prod_{n=1}^{N} \operatorname{La}(x_n \mid 0, \lambda_n) = \prod_{n=1}^{N} \frac{1}{2\lambda_n} \exp\left(-\frac{|x_n|}{\lambda_n}\right), \quad (3)$$

where $\lambda_n (n = 1, \dots, N)$ is the scale parameter of the Laplacian distribution for each element in the signal. Since the Inverse-Gamma (IG) distribution is conjugated to Laplacian distribution, the LSM model assumes that the scale parameter λ_n obeys the IG distribution as follows:

$$p(\boldsymbol{\lambda}; c, d) = \prod_{n=1}^{N} \mathrm{IG}(\lambda_{n}; c, d) = \prod_{n=1}^{N} \frac{d^{c}}{\Gamma(c)} \lambda_{n}^{-c-1} \exp\left(-\frac{d}{\lambda_{n}}\right).$$
(4)

In the above model, each hyperparameter λ_n controls the corresponding signal element x_n individually and each signal element is considered to be independent. Considering that nonzero elements of BSS appear in clusters, this requires a more appropriate model for BSS. For the BSS whose structure prior information is unknown, the PC-SBL algorithm [11] assumes that the hyperparameters of adjacent elements have a certain influence on its sparsity. Inspired by the PC-SBL, we assume that the block sparse signal **x** obeys the following correlated Laplacian prior distribution, i.e.,

$$p(\mathbf{x} \mid \boldsymbol{\lambda}) = \prod_{n=1}^{N} \operatorname{La}(x_n \mid 0, \lambda_{n-1}, \lambda_n, \lambda_{n+1})$$

=
$$\prod_{n=1}^{N} \frac{(1/\lambda_n) + \beta(1/\lambda_{n-1}) + \beta(1/\lambda_{n+1})}{2} \exp \left(5 \right)$$

$$\cdot \left(-\left(\frac{1}{\lambda_n} + \beta \frac{1}{\lambda_{n-1}} + \beta \frac{1}{\lambda_{n+1}}\right) |x_n| \right),$$

where the parameter $\beta \in [0, 1]$ indicates the degree of correlation between adjacent elements in the signal. It can be seen from (5) that the element x_n is affected by its own hyperparameter λ_n and the neighboring ones λ_{n-1} and λ_{n+1} . For the elements at both ends x_1 and x_N , let $\lambda_0 = 0$ and $\lambda_{N+1} = 0$. The model (5) makes use of the feature of the block sparse signal, in which the scale parameters λ still obey the IG distribution shown in (4).

3. Block Bayesian Recovery Algorithms

In Bayesian inference, given observation **y**, it needs to derive the posterior probability density for all unknown parameters $p(\mathbf{x}, \lambda, \gamma | \mathbf{y}) \propto p(\mathbf{y} | \mathbf{x}, \gamma)p(\mathbf{x} | \lambda)p(\lambda)p(\gamma)$. Variational Bayesian inference is a widely used method to approximately solve the maximization of a posteriori, which assumes that the variables **x**, λ , and γ are independent of each other. Let $\boldsymbol{\theta} = \{\mathbf{x}, \lambda, \gamma\}$, and then

$$p(\mathbf{\theta} \mid \mathbf{y}) \approx q(\mathbf{\theta}) = \prod_{i} q(\mathbf{\theta}_{i}) = q(\mathbf{x})q(\mathbf{\lambda})q(\mathbf{\gamma}).$$
 (6)

For each of these latent variables, the approximate posterior distribution may be computed in an alternating manner as follows:

$$q(\mathbf{\theta}_{i}) = \frac{\exp\left(\langle \ln p(\mathbf{y}, \mathbf{\theta}) \rangle_{q(\mathbf{\theta}_{k}), k \neq i}\right)}{\int \exp\left(\langle \ln p(\mathbf{y}, \mathbf{\theta}) \rangle_{q(\mathbf{\theta}_{k}), k \neq i}\right) d\mathbf{\theta}_{i}},$$
(7)

where $\langle \cdot \rangle_{q(\mathbf{\theta}_k),k\neq i}$ represents the expected operation with respect to the distributions $q(\mathbf{\theta}_k), k\neq i$. According to (7), the proposed reconstruction algorithm is derived by alternately learning the updating rules of these latent variables.

Firstly, the approximate posterior distribution $q(\mathbf{x})$ is

$$\ln q(\mathbf{x}) \propto \langle \ln p(\mathbf{y} \mid \mathbf{x}, \gamma) + \ln p(\mathbf{x} \mid \boldsymbol{\lambda}) \rangle_{q(\boldsymbol{\lambda})q(\gamma)}, \qquad (8)$$

where $p(\mathbf{y} | \mathbf{x}, \gamma)$ is a Gaussian distribution $\mathcal{N}(\mathbf{\Phi}\mathbf{x}, \gamma^{-1}\mathbf{I}_M)$ and $p(\mathbf{x} | \boldsymbol{\lambda})$ is a Laplacian distribution shown by (5). Since these two distributions are not conjugated, a direct solution is difficult. Similar to [12], let

$$L(\mathbf{x}) = -\langle \ln p(\mathbf{y} \mid \mathbf{x}, \gamma) + \ln p(\mathbf{x} \mid \boldsymbol{\lambda}) \rangle_{q(\boldsymbol{\lambda})q(\gamma)} \propto \frac{\langle \gamma \rangle}{2} \|\mathbf{y} - \mathbf{\Phi}\mathbf{x}\|_{2}^{2} + \sum_{n=1}^{N} \left\langle \frac{1}{\lambda_{n}} + \beta \frac{1}{\lambda_{n-1}} + \beta \frac{1}{\lambda_{n+1}} \right\rangle |x_{n}|.$$
(9)

The maximum a posteriori (MAP) estimate of the signal **x** can be obtained by $\widehat{\mathbf{x}}_{MAP} = \arg \min_{\mathbf{x}} \{L(\mathbf{x})\}$. The derivative of $L(\mathbf{x})$ is

$$\nabla_{\mathbf{x}} L(\mathbf{x}) = (\langle \gamma \rangle \mathbf{\Phi}^T \mathbf{\Phi} + \mathbf{D}) \mathbf{x} - \langle \gamma \rangle \mathbf{\Phi}^T \mathbf{y}, \tag{10}$$

where

$$\mathbf{D} = \operatorname{diag}\left[\left(\left\langle \frac{1}{\lambda_n} \right\rangle + \beta \left\langle \frac{1}{\lambda_{n-1}} \right\rangle + \beta \left\langle \frac{1}{\lambda_{n+1}} \right\rangle\right) \frac{1}{|x_n|}\right], \quad (11)$$

and diag $[\cdot]$ denotes a diagonal matrix with the elements in the bracket. Let the derivative be equal to zero and get the approximate MAP estimate:

$$\widehat{\mathbf{x}}_{\text{MAP}} = \langle \gamma \rangle \left(\langle \gamma \rangle \mathbf{\Phi}^T \mathbf{\Phi} + \mathbf{D} \right)^{-1} \mathbf{\Phi}^T \mathbf{y}.$$
(12)

So the posterior distribution can be approximated by using the second-order Taylor expansion around $\widehat{\mathbf{x}}_{MAP}$, i.e.,

$$\ln q(\mathbf{x}) \approx \ln q(\widehat{\mathbf{x}}_{\text{MAP}}) + \frac{1}{2} (\mathbf{x} - \mathbf{x} \wedge_{\text{MAP}})^T H(\widehat{\mathbf{x}}_{\text{MAP}}) (\mathbf{x} - \widehat{\mathbf{x}}_{\text{MAP}}),$$
(13)

where $H(\hat{\mathbf{x}}_{MAP}) \approx -(\langle \gamma \rangle \Phi^T \Phi + \mathbf{D})$. After similar simplification in [12], $q(\mathbf{x})$ can be approximated to obey the Gaussian distribution $\mathcal{N}(\mathbf{x} | \mathbf{\mu}, \mathbf{\Sigma}_o)$ with the mean $\mathbf{\mu} = \langle \gamma \rangle \mathbf{\Sigma}_o \Phi^T \mathbf{y}$ and covariance matrix $\mathbf{\Sigma}_o = (\langle \gamma \rangle \Phi^T \Phi + \mathbf{D})^{-1}$. Due to that, there exists some approximation in the above derivation of the posterior distribution $q(\mathbf{x})$; the sparsity of the signal may be underestimated. So a parameter $\alpha \in [0.5, 1]$ is introduced into the computation of the covariance matrix. Thus, $q(\mathbf{x})$ is approximated to be the following Gaussian distribution:

$$q(\mathbf{x}) \approx \mathcal{N}(\mathbf{x} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}), \tag{14}$$

where the mean and covariance matrix, respectively, are

$$\boldsymbol{\mu} = \langle \boldsymbol{\gamma} \rangle \boldsymbol{\Sigma} \boldsymbol{\Phi}^T \mathbf{y},$$

$$\boldsymbol{\Sigma} = \left(\langle \boldsymbol{\gamma} \rangle \boldsymbol{\Phi}^T \boldsymbol{\Phi} + \alpha \mathbf{D} \right)^{-1}.$$
 (15)

Secondly, the approximate posterior distribution $q(\lambda)$ is

$$\ln q(\mathbf{\lambda}) \propto \langle \ln p(\mathbf{\lambda}) + \ln p(\mathbf{x} \mid \mathbf{\lambda}) \rangle_{q(\mathbf{x})} = \ln p(\mathbf{\lambda}) + \langle \ln p(\mathbf{x} \mid \mathbf{\lambda}) \rangle_{q(\mathbf{x})}.$$
(16)

From (4) and (5), we have

$$\ln p(\lambda) \propto \sum_{n=1}^{N} \left\{ (c+1) \ln \frac{1}{\lambda_n} - \frac{d}{\lambda_n} \right\},$$
(17)

$$\langle \ln p(\mathbf{x} \mid \boldsymbol{\lambda}) \rangle_{q(\mathbf{x})} \propto \sum_{n=1}^{N} \left\{ \ln \left(\frac{1}{\lambda_{n}} + \beta \frac{1}{\lambda_{n-1}} + \beta \frac{1}{\lambda_{n+1}} \right) - \left(\frac{1}{\lambda_{n}} + \beta \frac{1}{\lambda_{n-1}} + \beta \frac{1}{\lambda_{n+1}} \right) \langle |x_{n}| \rangle \right\}$$

$$= \sum_{n=1}^{N} \left\{ \ln \left(\frac{1}{\lambda_{n}} + \beta \frac{1}{\lambda_{n-1}} + \beta \frac{1}{\lambda_{n+1}} \right) - \left(\langle |x_{n}| \rangle + \beta \langle |x_{n-1}| \rangle + \beta \langle |x_{n+1}| \rangle \right) \frac{1}{\lambda_{n}} \right\}$$

$$\geq \sum_{n=1}^{N} \left\{ \ln \frac{1}{\lambda_{n}} - \left(\langle |x_{n}| \rangle + \beta \langle |x_{n-1}| \rangle + \beta \langle |x_{n+1}| \rangle \right) \frac{1}{\lambda_{n}} \right\}.$$

$$(18)$$

So $q(\mathbf{\lambda})$ can be approximated as

$$\ln q(\boldsymbol{\lambda}) \propto \sum_{n=1}^{N} \left\{ \left[(c+1)+1 \right] \ln \frac{1}{\lambda_{n}} - \left[\left(\langle |x_{n}| \rangle + \beta \langle |x_{n-1}| \rangle + \beta \langle |x_{n+1}| \rangle \right) + d \right] \frac{1}{\lambda_{n}} \right\}.$$
(19)

Therefore, $q(\lambda)$ obeys an Inverse-Gamma distribution:

$$p(\boldsymbol{\lambda}; \tilde{c}, \tilde{d}) = \prod_{n=1}^{N} \mathrm{IG}(\lambda_n; \tilde{c}, \tilde{d}_n), \qquad (20)$$

with

$$\tilde{c} = c + 1,$$

$$\tilde{d}_n = \left(\langle |x_n| \rangle + \beta \langle |x_{n-1}| \rangle + \beta \langle |x_{n+1}| \rangle \right) + d.$$
(21)

We can also obtain

$$\left\langle \frac{1}{\lambda_n} \right\rangle = \frac{\tilde{c}}{\tilde{d}_n} = \frac{c+1}{\left(\langle |x_n| \rangle + \beta \langle |x_{n-1}| \rangle + \beta \langle |x_{n+1}| \rangle \right) + d}, \quad (22)$$

where $\langle |x_n| \rangle$ can be computed as follows according to [13]:

$$\langle |x_n| \rangle = \sqrt{\frac{2\Sigma_{n,n}}{\pi}} \exp\left(-\frac{\mu_n^2}{2\Sigma_{n,n}}\right) + |\mu_n| \operatorname{erf}\left(\sqrt{\frac{\mu_n^2}{2\Sigma_{n,n}}}\right),$$
(23)

and erf $(x) = (2/\sqrt{\pi}) \int_0^x e^{-t^2} dt$ is the error function.

Thirdly, the approximate posterior distribution of the noise parameter $q(\gamma)$ is

$$\ln q(\gamma) \propto \langle \ln p(\gamma) + \ln p(\mathbf{y} | \mathbf{x}, \gamma) \rangle_{q(\mathbf{x})} \propto \left(a - 1 + \frac{M}{2} \right) \ln \gamma - \left(\frac{\langle \| \mathbf{y} - \mathbf{\Phi} \mathbf{x} \|_2^2 \rangle}{2} + b \right) \gamma.$$
(24)



Algorithm 1: Block EM-VB.

Input: y, Φ , ε , Max_iter, β , α Initialize: μ^0 , λ , γ , a, b, c, dWhile $\|\mu^{t+1} - \mu^t\|_2 > \varepsilon$ or $t \le Max_i$ ter do Update: (1) Compute the mean μ_i and the variance σ_i^2 ($i = 1, 2, \dots, N$), sequentially by (35). (2) Compute λ^{-1} according to (22) and (23). (3) Compute γ via (27) and (36). Output: $\hat{\mathbf{x}} = \mu$





FIGURE 1: The support recovery rate versus the sparsity in the noiseless case when (a) M = 30 and (b) M = 40.

(25)

So $q(\gamma)$ obeys the Gamma distribution:

 $q(\gamma) = G(\gamma; \tilde{a}, \tilde{b}),$

with

$$\tilde{a} = a + \frac{M}{2},$$

$$\tilde{b} = \frac{\langle \|\mathbf{y} - \mathbf{\Phi}\mathbf{x}\|_{2}^{2} \rangle}{2} + b,$$
(26)



FIGURE 2: The success rate vs. the sparsity level in the noiseless case when (a)M = 30 and (b) M = 40.



FIGURE 3: The success rate vs. the number of measurements in the noiseless case when (a) K = 20 and (b) K = 25.

and we can obtain

$$\langle \gamma \rangle = \frac{\tilde{a}}{\tilde{b}} = \frac{a + (M/2)}{\left(\left\langle \| \mathbf{y} - \mathbf{\Phi} \mathbf{x} \|_2^2 \right\rangle / 2 \right) + b},\tag{27}$$

where $\langle \|\mathbf{y} - \mathbf{\Phi}\mathbf{x}\|_2^2 \rangle = \|\mathbf{y} - \mathbf{\Phi}\boldsymbol{\mu}\|_2^2 + \mathrm{Tr}[\mathbf{\Sigma}\mathbf{\Phi}^T\mathbf{\Phi}]$ and $\mathrm{Tr}(\cdot)$ is the trace of a matrix.

Therefore, the whole process of the proposed algorithm is summarized in Algorithm 1, where ε is the preset error that can be tolerated. The proposed algorithm can be regarded as an extension of the EM-VB algorithm for the recovery of BSS, which is termed the Block EM-VB algorithm. It has an additional parameter α except for the block parameter β . When $\alpha = 1$ and $\beta = 0$, the Block EM-VB algorithm reduces to the EM-VB. The parameters α in (15) and β in (22) have great influences on the recovery performance of the Block EM-VB. It is appropriate to set $\alpha < 1$ to void underestimation of the support set if the sparsity of the signal is larger, while larger α can suppress nonzero signal elements and has certain antinoise capacity in noisy environments. For the choice of β , it is similar that larger β may enhance the influence between adjacent elements and could suppress the nonzero values to make the signal more sparse in noisy environments.

Remark 1. From (23), it can be seen that the computation of $\langle |x_n| \rangle$ is related to the erf (*x*) involving integral operation. In



FIGURE 4: The reconstructed signal by different algorithms when N = 100, M = 50, K = 25, and SNR = 10 dB. (a) Original signal, (b) reconstructed signal by PC-SBL with NMSE = 0.2362, (c) reconstructed signal by Block EM-VB with NMSE = 0.2184, and (d) reconstructed signal by Fast Block EM-VB with NMSE = 0.1092.



FIGURE 5: The NMSE vs. the sparsity level when SNR = 10 dB and (a) M = 40 and (b) M = 50.

practice, to reduce the complexity, erf (x) can be calculated by utilizing some approximation equations such as

erf
$$(x) \approx \left\{ 1 - e^{-4x^2/\pi} \left[1 + \frac{8x^4}{\pi} \left(\frac{1}{3} - \frac{1}{\pi} \right) \right] \right\}^{1/2}$$
. (28)

From the process of the Block EM-VB algorithm, it can be seen that its complexity is almost the same as that of the PC-SBL.

The proposed Block EM-VB algorithm involves the matrix inversion shown in (15), which is the main computa-

tional complexity. It is better to consider the fast version of the proposed algorithm. To void computation of the inverse of the matrix, we can use the following approximate posterior distribution, which is expressed as

$$p(\mathbf{\theta} \mid \mathbf{y}) \approx q(\mathbf{\theta}) = \prod_{i=1}^{N} q(x_i) q(\mathbf{\lambda}) q(\mathbf{\gamma}).$$
 (29)

That is to say, it assumes the independence on the posterior of each coefficient element of the signal. Similarly, by using (7), we can alternately learn the updating rules of these latent variables.



FIGURE 6: The NMSE vs. the number of measurements when SNR = 10 dB and (a) K = 20 and (b) K = 25.

Firstly, the approximate posterior distribution $q(x_i)$ $(i = 1, 2, \dots, N)$ is

$$\ln q(x_i) \propto \langle \ln p(\mathbf{y} \mid \mathbf{x}, \gamma) + \ln p(x_i \mid \lambda_{i-1}, \lambda_i, \lambda_{i+1}) \rangle_{q(\gamma)q(\lambda)q(x_j)_{j\neq i}},$$
(30)

where $p(\mathbf{y} | \mathbf{x}, \gamma) = \mathcal{N}(\mathbf{\Phi}\mathbf{x}, \gamma^{-1}\mathbf{I}_M)$ and $p(x_i | \lambda_{i-1}, \lambda_i, \lambda_{i+1}) = (((1/\lambda_i) + \beta(1/\lambda_{i-1}) + \beta(1/\lambda_{i+1}))/2) \exp(-(((1/\lambda_i) + \beta(1/\lambda_{i+1}))/2) + \beta(1/\lambda_{i+1}))/2)$. Let

$$\begin{split} l(x_{i}) &= -\langle \ln p(\mathbf{y} \mid \mathbf{x}, \gamma) + \ln p(x_{i} \mid \lambda_{i-1}, \lambda_{i}, \lambda_{i+1}) \rangle_{q(\gamma)q(\lambda)q(x_{j})_{j+i}} \propto \frac{\langle \gamma \rangle}{2} \langle ||\mathbf{y} - \mathbf{\Phi}\mathbf{x}||_{2}^{2} \rangle_{q(x_{j})_{j+i}} \\ &+ \left\langle \frac{1}{\lambda_{i}} + \beta \frac{1}{\lambda_{i-1}} + \beta \frac{1}{\lambda_{i+1}} \right\rangle |x_{i}| \propto \frac{\langle \gamma \rangle}{2} \langle -2\mathbf{y}^{T} \mathbf{\Phi}\mathbf{x} + \mathbf{x}^{T} \mathbf{\Phi}^{T} \mathbf{\Phi}\mathbf{x} \rangle_{q(x_{j})_{j+i}} \\ &+ \left\langle \frac{1}{\lambda_{i}} + \beta \frac{1}{\lambda_{i-1}} + \beta \frac{1}{\lambda_{i+1}} \right\rangle |x_{i}| \propto \frac{\langle \gamma \rangle}{2} \left(-2\mathbf{y}^{T} \varphi_{i} x_{i} + \varphi_{i}^{T} \varphi_{i} x_{i}^{2} + 2x_{i} \varphi_{i}^{T} \sum_{j\neq i} \varphi_{j} \langle x_{j} \rangle \right) \\ &+ \left\langle \frac{1}{\lambda_{i}} + \beta \frac{1}{\lambda_{i-1}} + \beta \frac{1}{\lambda_{i+1}} \right\rangle |x_{i}|, \end{split}$$

$$(31)$$

where φ_i is the *i*-th column of Φ . The derivative of $l(x_i)$ is

$$\nabla_{x_{i}} l(x_{i}) = \frac{\langle \gamma \rangle}{2} \left(-2\mathbf{y}^{T} \boldsymbol{\varphi}_{i} + 2\boldsymbol{\varphi}_{i}^{T} \boldsymbol{\varphi}_{i} x_{i} + 2\boldsymbol{\varphi}_{i}^{T} \sum_{j \neq i} \boldsymbol{\varphi}_{j} \langle x_{j} \rangle \right)$$

$$+ \left\langle \frac{1}{\lambda_{i}} + \beta \frac{1}{\lambda_{i-1}} + \beta \frac{1}{\lambda_{i+1}} \right\rangle \frac{x_{i}}{|x_{i}|}.$$
(32)

Let $\nabla_{x_i} l(x_i) = 0$ and obtain the following approximate MAP estimate:

$$\tilde{x}_{i} = \langle \gamma \rangle \left(\langle \gamma \rangle \varphi_{i}^{T} \varphi_{i} + \left\langle \frac{1}{\lambda_{i}} + \beta \frac{1}{\lambda_{i-1}} + \beta \frac{1}{\lambda_{i+1}} \right\rangle \frac{1}{|x_{i}|} \right)^{-1} \left(\mathbf{y} - \sum_{j \neq i} \varphi_{j} \langle x_{j} \rangle \right)^{T} \varphi_{i}.$$
(33)



FIGURE 7: The comparison of different algorithms under different SNR when N = 100, K = 20, and M = 50.

Similarly, $q(x_i)$ is approximated to obey the following Gaussian distribution:

$$q(x_i) \approx \mathcal{N}(x_i \mid \mu_i, \sigma_i^2), \qquad (34)$$

where

$$\mu_{i} = \langle \gamma \rangle \sigma_{i}^{2} \left(\mathbf{y} - \sum_{j \neq i} \varphi_{j} \langle x_{j} \rangle \right)^{T} \varphi_{i},$$

$$\sigma_{i}^{2} = \left(\langle \gamma \rangle \varphi_{i}^{T} \varphi_{i} + \left\langle \frac{1}{\lambda_{i}} + \beta \frac{1}{\lambda_{i-1}} + \beta \frac{1}{\lambda_{i+1}} \right\rangle \frac{\alpha}{|x_{i}|} \right)^{-1},$$
(35)

in which a parameter α is also introduced to avoid underestimation of the sparsity of the signal. Thus, $\langle x_i \rangle = \mu_i$ $(i = 1, 2, \dots, N)$ can be computed in a sequential manner.

Secondly, the approximate posterior distribution $q(\lambda)$ still obeys an Inverse-Gamma distribution shown in (20) and the computation of $\langle \lambda^{-1} \rangle$ is the same as (22).

Thirdly, the approximate posterior distribution $q(\gamma)$ obeys the Gamma distribution $G(\gamma; \tilde{a}, \tilde{b})$ and we can obtain $\langle \gamma \rangle = \tilde{a}/\tilde{b} = (a + M/2)/(\langle ||\mathbf{y} - \mathbf{\Phi}\mathbf{x}||_2^2)/2 + b)$, where

$$\langle \| \mathbf{y} - \mathbf{\Phi} \mathbf{x} \|_{2}^{2} \rangle = \| \mathbf{y} - \mathbf{\Phi} \boldsymbol{\mu} \|_{2}^{2} + \operatorname{Tr} \left[\boldsymbol{\Sigma} \mathbf{\Phi}^{T} \mathbf{\Phi} \right]$$

$$= \| \mathbf{y} - \mathbf{\Phi} \boldsymbol{\mu} \|_{2}^{2} + \sum_{i=1}^{N} \sigma_{i}^{2} \boldsymbol{\varphi}_{i}^{T} \boldsymbol{\varphi}_{i},$$
(36)

and $\boldsymbol{\mu} = (\mu_1, \mu_2, \cdots, \mu_N)^T$ and $\boldsymbol{\Sigma} \triangleq \text{diag} [\sigma_i^2]$.

Compared with the Block EM-VB algorithm, the above algorithm has low computational complexity due to without matrix inversion in each iteration. It can be called Fast Block EM-VB, and its process is summarized in Algorithm 2. In noisy environments, the correlations between adjacent signal elements may be weakened, especially in the low SNR case, so it is appropriate to assume independence on the posterior of each coefficient element, which implies that the Fast Block EM-VB is suitable to recover the BSS under the low SNR case.

4. Simulation Experiments

In this section, some simulation experiments are carried out to demonstrate the performances of the proposed Block EM-VB algorithm and its fast version. A comparison with other algorithms such as EM-VB [12] and PC-SBL [11] is also given.

4.1. Performance Analysis via Simulated Data. In the following simulation, let the length of the sparse signal be N = 100, and an arbitrary block Gaussian sparse signal is randomly generated with its nonzero entries randomly distributed in B = 3 blocks, and the measurement matrix is a random Gaussian matrix. The number of Monte Carlo simulations is 200. The parameters in EM-VB and in the Block EM-VB and its fast version are set as $a = b = c = d = 10^{-6}$, Max_iter = 200, and $\varepsilon = 10^{-5}$. The parameters α and β in the Block EM-VB and its fast version will be set adaptively according to the noiseless and noisy cases because their performances are sensitive to the choice of α and β . The parameters of PC-SBL are set the same as those in [11].

First, we discuss the influences of two parameters α and β on the performance of the proposed Block EM-VB and make a comparison with the EM-VB, which is a special case of the Block EM-VB when $\alpha = 1$ and $\beta = 0$. To demonstrate the effect of these two parameters, the performances of the Block EM-VB with $\alpha = 0.7$ and $\beta = 0$ and $\alpha = 1$ and $\beta = 0.5$ are also given. The support recovery rate is used to evaluate the performance of the Block EM-VB algorithm with different parameters. The recovered support of the sparse signal is defined as supp ($\hat{\mathbf{x}} = \{i, |\hat{x}_i| > 0.001\}$, and then the support



FIGURE 8: The average runtimes vs. the length of signal N.

recovery rate is defined by $|\operatorname{supp}(\widehat{\mathbf{x}}) \cap \operatorname{supp}(\mathbf{x})| | | \operatorname{supp}(\widehat{\mathbf{x}}) \cup \operatorname{supp}(\mathbf{x})|$, where $|\cdot|$ denotes the number of elements in a set. If the overlap between the estimated support and the true support is more, the recovery rate is closer to 1. Figure 1 plots the support recovery rates of different algorithms versus the sparsity level *K* when the number of measurements is M = 30 and M = 40, respectively. It can be seen that the parameters of the Block EM-VB algorithm have an important influence on the recovery performance. The appropriate parameter $\alpha < 1$ can avoid the underestimation of the support of the sparse signals, and the block parameter $\beta > 0$ is helpful to recover the block sparse signals. Thus, the proposed Block EM-VB algorithm with appropriate parameters has better performance than the EM-VB.

Then, we make a comparison between the Block EM-VB and its fast version with the PC-SBL. In the noiseless case, the success rate is used to evaluate the performances of these different algorithms. When a trial satisfies $\|\mathbf{\hat{x}} \cdot \mathbf{x}\|_2^2 / \|\mathbf{x}\|_2^2 < 10^{-6}$, it is regarded as a successful trial. The success rate is defined as the percentage of successful trials in the total of independent trials. In the noisy case, the reconstruction performance of each algorithm is evaluated by the normalized mean square error (NMSE), where NMSE = $\|\mathbf{\hat{x}} \cdot \mathbf{x}\|_2^2 / \|\mathbf{x}\|_2^2$.

In the noiseless case, let $\alpha = 0.7$ and $\beta = 0.5$ in the Block EM-VB and $\alpha = 0.5$ and $\beta = 0.5$ in the Fast Block EM-VB. Figure 2 plots the success rate of individual recovery algorithm versus the sparsity level *K* when the number of measurement is M = 30 and M = 40, respectively. Then, let the sparsity K = 20 and K = 25, and the success rate of each algorithm versus the number of measurements *M* is shown in Figure 3. From these results, it is observed that the Block EM-VB is superior to the PC-SBL when the number of measurements is less or the sparsity is smaller. The Fast Block EM-VB is inferior to these two algorithms due to the independent assumption.

Then, we consider the noisy case where the Gaussian noise is added to the measurements with the signal-to-



FIGURE 9: Reconstructed images with different algorithms (the left column is the noiseless case, and the right column is the noisy case).

noise ratio (SNR) defined as $20 \log_{10}(||\Phi \mathbf{x}||_2/||\mathbf{n}||_2)$. It should be noted that the setting of parameter α in the proposed algorithms needs to consider the trade-off between removing the noise and maintaining the nonzero elements of the signal. A larger value of α tends to suppress noise while losing the nonzero values of the signal. Compared with the Block EM-VB, the fast version needs to select smaller α to ensure the recovery of nonzero elements of the sparse signal because of its inherent denoising ability. Let SNR = 10 dB and $\alpha = 0.85$ and $\beta = 0.5$ in the Block

EM-VB and $\alpha = 0.5$ and $\beta = 0.5$ in the Fast Block EM-VB. The reconstruction results of different algorithms when K = 25 and M = 50 are given in Figure 4. It can be seen that the Block EM-VB algorithm and its fast version have better reconstruction performance than the PC-SBL in the low SNR environment.

The NMSE of each algorithm versus the sparsity *K* in the case of SNR = 10 dB is shown in Figure 5, where the numbers of measurements are given as M = 40 and M = 50, respectively. Figure 6 plots the NMSE of individual recovery

algorithm versus the number of measurements M in the case of SNR = 10 dB, where K = 20 and K = 25, respectively. It can be found that the Block EM-VB algorithm is superior to the PC-SBL when the number of the measurements is larger or the sparsity is less, and the latter is better than the former when the number of the measurements is less and the sparsity is larger. It is also observed that the Fast Block EM-VB is better than the Block EM-VB when the number of measurements is less or the sparsity is larger. In addition, it is shown that the Fast Block EM-VB outperforms the PC-SBL when the sparsity is less or the number of the measurements is less or the number of the

The performances of each algorithm in different SNR cases are shown in Figure 7, where K = 20 and M = 50. For the Block EM-VB, the parameters $\beta = 0.5$ and α is set to vary with the SNR, i.e., $\alpha = 0.85$ when the SNR varies from 0 dB to 15 dB, $\alpha = 0.75$ when the SNR changes from 20 dB to 25 dB, and $\alpha = 0.73$ for the case of 30 dB. The parameters in the Fast Block EM-VB algorithm are still set as $\alpha = 0.5$ and $\beta = 0.5$. From Figure 7, it can be seen that the Block EM-VB algorithm and the fast version have good reconstruction performance in the case of low SNR when compared with the PC-SBL. The Fast Block EM-VB especially has robust recovery performance in respect of noise immunity.

Finally, the average runtimes of these algorithms versus the length of signals N by 5 independent trials are given in Figure 8, where M = N/2, K = N/10, SNR = 10 dB, and the number of nonzero blocks B = K/5. It validates that the Block EM-VB almost has considerable computational complexity as the PC-SBL and the Fast Block EM-VB has the highest computational efficiency compared with other recovery algorithms, which makes it have a potential advantage in practical application.

4.2. Application in ISAR Imaging. The inverse synthetic aperture radar (ISAR) imaging is appropriately implemented under the framework of sparse signal recovery due to the sparse characteristic of the target [14]. In this experiment, the "Yak-42" dataset is used, in which the number of range cells is 256 and the number of pulses is 256. 128 pulses are randomly sampled to simulate the sparse aperture data. Here, we use the MATLAB code provided in [14], where the PC-SBL adopts a pruning operation. For a fair comparison, the proposed Block EM-VB and the fast version also use a similar pruning operation and the parameters in these two algorithms are set as Max_iter = 200 and $\varepsilon = 10^{-6}$. Image entropy is usually used to measure image quality in ISAR imaging. The smaller image entropy means better reconstruction performance. The image entropy is defined as

Entropy =
$$-\sum_{i}\sum_{j}\left(\frac{|\mathbf{x}(i,j)|^2}{E}\right)\log\left(\frac{|\mathbf{x}(i,j)|^2}{E}\right)$$
, (37)

where $E = \sum_{i} \sum_{j} |\mathbf{x}(i, j)|^2$ is the energy of the radar image **x**.

Here, the image is reconstructed by each range cell. Figures 9(a), 9(c), and 9(e) give the reconstruction results of these algorithms in the noiseless case, where we set

TABLE 1: The PSNR results of different algorithms.

	PC-SBL	Block EM-VB	Fast Block EM-VB
Noiseless	5.6384	5.3275	5.4201
Noisy	5.5457	5.0689	5.0059

TABLE 2: Comparison of the runtimes (s) of different algorithms.

	PC-SBL	Block EM-VB	Fast Block EM-VB
Noiseless	27.7130	7.8391	3.3582
Noisy	30.3940	31.6483	6.8828

 $\alpha = 1$ and $\beta = 1$ for the Block EM-VB and $\alpha = 0.8$ and $\beta = 1$ for the Fast Block EM-VB. In the noisy case, the data are corrupted by additive Gaussian noise and let $\alpha = 0.85$ and $\beta = 1$ for the Block EM-VB and $\alpha = 0.6$ and $\beta = 1$ for the Fast Block EM-VB. Figures 9(b), 9(d), and 9(f) demonstrate the results in the case of SNR ≈ 3 dB (the noise variance $\gamma^{-1} = 0.0025$). The entropy values of these algorithms are shown in Table 1. From these reconstruction results, it can be seen that the image obtained by the Block EM-VB and its fast version has better quality in the noisy case, which implies that the Block EM-VB and its fast version have strong noise immunity ability. Table 2 gives the corresponding runtimes of these algorithms, which demonstrates that the Fast Block EM-VB has the highest computational efficiency and can be used in real-time processing.

5. Conclusions

Considering the clustered structural features of nonzero elements of block sparse signals, this paper proposes the Block EM-VB algorithm for signal recovery, which is based on a correlated LSM model. Furthermore, a fast version of the Block EM-VB is presented, which can recover the block sparse signals with lower computational complexity because of no inversion in each iteration. Experimental results with simulation data and ISAR imaging demonstrate that the Block EM-VB and its fast version have good BSS reconstruction performance and noise tolerance capability, especially in the low SNR scenarios, which implies that the proposed algorithms can be potentially applied in various signal processing fields.

Data Availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Conflicts of Interest

There is no conflict of interest regarding the publication of this paper.

Acknowledgments

The work is supported by the National Natural Science Foundation of China under Grants 61771046, 61931015, and 61731023 and the Beijing Natural Science Foundation (L191004). The authors are grateful for their support of this research.

References

- D. L. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.
- [2] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Transactions on Information Theory*, vol. 53, no. 12, pp. 4655–4666, 2007.
- [3] M. E. Tipping, "Sparse Bayesian learning and the relevance vector machine," *Journal of Machine Learning Research*, vol. 1, pp. 211–244, 2001.
- [4] Lu Wang, Lifan Zhao, Guoan Bi, Chunru Wan, and Lei Yang, "Enhanced ISAR imaging by exploiting the continuity of the target scene," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 9, pp. 5736–5750, 2014.
- [5] F. Parvaresh, H. Vikalo, S. Misra, and B. Hassibi, "Recovering sparse signals using sparse measurement matrices in compressed DNA microarrays," *IEEE Journal of Selected Topics in Signal Processing*, vol. 2, no. 3, pp. 275–285, 2008.
- [6] Y. C. Eldar, P. Kuppinger, and H. Bolcskei, "Block-sparse signals: uncertainty relations and efficient recovery," *IEEE Transactions on Signal Processing*, vol. 58, no. 6, pp. 3042–3054, 2010.
- [7] B. X. Huang and T. Zhou, "Recovery of block sparse signals by a block version of StOMP," *Signal Processing*, vol. 106, pp. 231–244, 2015.
- [8] Z. Zhang and B. D. Rao, "Recovery of block sparse signals using the framework of block sparse Bayesian learning," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 3345–3348, Kyoto, Japan, 2012.
- [9] L. Yu, J.-P. Barbot, G. Zheng, and H. Sun, "Bayesian compressive sensing for cluster structured sparse signals," *Signal Processing*, vol. 92, no. 1, pp. 259–269, 2012.
- [10] L. Yu, H. Sun, G. Zheng, and J. Pierre Barbot, "Model based Bayesian compressive sensing via local beta process," *Signal Processing*, vol. 108, pp. 259–271, 2015.
- [11] J. Fang, Y. Shen, H. Li, and P. Wang, "Pattern-coupled sparse Bayesian learning for recovery of block-sparse signals," *IEEE Transactions on Signal Processing*, vol. 63, no. 2, pp. 360–372, 2015.
- [12] S. Zhang, Y. Liu, X. Li, and G. Bi, "Variational Bayesian sparse signal recovery with LSM prior," *IEEE Access*, vol. 5, pp. 26690–26702, 2017.
- [13] A. Winkelbauer, "Moments and absolute moments of the normal distribution," 2012, https://arxiv.org/abs/1209.4340.
- [14] H. Duan, L. Zhang, J. Fang, L. Huang, and H. Li, "Patterncoupled sparse Bayesian learning for inverse synthetic aperture radar imaging," *IEEE Signal Processing Letters*, vol. 22, no. 11, pp. 1995–1999, 2015.