

Article

Direction-of-Arrival Estimation via Sparse Bayesian Learning Exploiting Hierarchical Priors with Low Complexity

Ninghui Li ¹, Xiaokuan Zhang ², Fan Lv ¹ and Bin Feng Zong ^{2,*}

¹ Graduate School, Air Force Engineering University, Xi'an 710051, China; li6667778882022@163.com (N.L.); lf995407681@163.com (F.L.)

² Air Defence and Antimissile School, Air Force Engineering University, Xi'an 710051, China; ezxk@sina.com

* Correspondence: zongbinfeng@163.com

Abstract: For direction-of-arrival (DOA) estimation problems in a sparse domain, sparse Bayesian learning (SBL) is highly favored by researchers owing to its excellent estimation performance. However, traditional SBL-based methods always assign Gaussian priors to parameters to be solved, leading to moderate sparse signal recovery (SSR) effects. The reason is Gaussian priors play a similar role to l_2 regularization in sparsity constraint. Therefore, numerous methods are developed by adopting hierarchical priors that are used to perform better than Gaussian priors. However, these methods are in straitened circumstances when multiple measurement vector (MMV) data are adopted. On this basis, a block-sparse SBL method (named BSBL) is developed to handle DOA estimation problems in MMV models. The novelty of BSBL is the combination of hierarchical priors and block-sparse model originating from MMV data. Therefore, on the one hand, BSBL transfers the MMV model to a block-sparse model by vectorization so that Bayesian learning is directly performed, regardless of the prior independent assumption of different measurement vectors and the inconvenience caused by the solution of matrix form. On the other hand, BSBL inherited the advantage of hierarchical priors for better SSR ability. Despite the benefit, BSBL still has the disadvantage of relatively large computation complexity caused by high dimensional matrix operations. In view of this, two operations are implemented for low complexity. One is reducing the matrix dimension of BSBL by approximation, generating a method named BSBL-APPR, and the other is embedding the generalized approximate message passing (GAMP) technique into BSBL so as to decompose matrix operations into vector or scale operations, named BSBL-GAMP. Moreover, BSBL is able to suppress temporal correlation and handle wideband sources easily. Extensive simulation results are presented to prove the superiority of BSBL over other state-of-the-art algorithms.

Keywords: direction-of-arrival estimation; sparse Bayesian learning; hierarchical priors; block sparse



Citation: Li, N.; Zhang, X.; Lv, F.; Zong, B. Direction-of-Arrival Estimation via Sparse Bayesian Learning Exploiting Hierarchical Priors with Low Complexity. *Sensors* **2024**, *24*, 2336. <https://doi.org/10.3390/s24072336>

Academic Editor: Andrea Trucco

Received: 6 March 2024

Revised: 1 April 2024

Accepted: 4 April 2024

Published: 6 April 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

DOA estimation has advanced obviously due to several technique leaps in the last four decades, and the attained achievements are widely applied to communications, radar, sonar, and navigation. To be specific in communication [1,2], DOA estimation is essential in channel estimation, wireless communications, microphone localization, vehicular communications [3], Reconfigurable Intelligent Surfaces (RIS), and corresponding research focuses, including the RIS-based vehicle DOA estimation method [4–6].

Among the technique leaps in DOA estimation, compressed sensing (CS) and sparse recovery (SR) have played important roles in the last decade [7,8]. Compared with traditional algorithms based on beamforming or subspace techniques [9–13], sparsity-based estimators have achieved technique leaps since CS and SR can mitigate the requirements for high signal-to-noise ratios (SNRs) and abundant snapshots [14]. Moreover, it has been proven that sparsity-based estimators have remarkable advantages, such as good robustness to correlation and high estimation accuracy.

In recent years, relevant studies on DOA estimation mainly focus on sparse array configuration [15–17] or sparsity-based method innovation [18,19]. For the latter, three categories can be summarized out of massive documents. (i) The first is based on norm optimization [20–28], (ii) the second is exploiting basis pursuit [29,30], and (iii), and the third is utilizing sparse Bayesian learning (SBL) [31–47]. In particular, SBL-based methods are popular and highly favored owing to incomparable comprehensive performance beyond both norm optimization and basis pursuit [48].

From a Bayesian perspective, sparse Bayesian learning is a probabilistic method that achieves l_p -norm minimization by assigning sparse priors to the signal of interests for sparse signal recovery (SSR). In the single measurement vector (SMV) model, SBL can retain a desirable property of the l_0 -norm diversity measure, i.e., the global minimum is achieved at the maximally sparse solution [48] and produced a more limited constellation of local minima. In practice, the most used is the MMV model, so MSBL is developed [49]. Theoretically, in [49], the adopted empirical Bayesian prior plays an important role in estimating a convenient posterior distribution over candidate basis vectors based on the concept of automatic relevance determination. That implies that the used priors assigned to signals of interest can enforce a common sparsity profile and consistently place the prominent posterior mass on the appropriate region for sparse recovery. In other words, the adopted priors dominate the sparsity performance of SBL and indicate the ability to carry out l_p -norm minimization. Despite this, it is still unclear which prior is the best for sparse recovery, but a hierarchical Bayesian framework representing Laplace priors has been proven to be prominent in [50]. Therefore, hierarchical priors are widely attractive, and many corresponding works are developed and presented. For instance, hierarchical synthesis lasso (HSL) priors for representing the same small subset of features are created for enforcing a proper sparsity profile of signal vectors [46], and hierarchical priors are adopted to consider unknown mutual coupling and off-grid errors. No matter how hierarchical priors are used, SBL has to face the difficulty caused by the MMV model. To be specific, SBL needs prior assumption, i.e., the uncorrelation between different measurement vectors, while the assumption may not be satisfied in practice. Moreover, the solution of matrix form is not always easy to handle and is even prohibitive due to the large complexity during the learning process. On this basis, vectorizing the MMV model seems to be an optimal selection since the solution can be transformed into the vector form. A successful realization is developed in [32], and the used block-sparse model indeed contributes much to the whole designed algorithm, although the main focus is the real-valued transformation. Regrettably, that work adopts Gaussian priors rather than hierarchical priors, so its sparsity performance is bound to be limited, no matter how impressive and excellent the running efficiency is. Recently, there have been a few documents that show the combination of SBL and deep learning (DL) [51,52], attracting widespread attention among numerous researchers. With big data and artificial intelligence, DL gradually arises and applies to image processing, signal processing, classification, recognition, etc. The obvious advantages of DL are its adaptability to complicated practical cases and its high running efficiency. In signal processing, many researchers use derived iterative processes to design corresponding neural networks so that the proposed algorithms operate with low complexity burdens, and the designed networks are named deep unrolled neural networks. For example, an SBL-based algorithm is unfolded into a layer-wise structure with a set of introduced trainable parameters in [51], which is beneficial for channel estimation. In addition, a model-driven DL detector is developed based on variational Bayesian inference in [52]. Based on deep unrolled networks, the detector is able to capture channel features that may be important but neglected by model-based methods, including SBL-based methods. Although the algorithms based on both SBL and DL are not applied to DOA estimation, the disadvantages lie in the inexplicability caused by data-driven DL techniques. In fact, the lack of theoretical guarantee is always constraining the development and application of DL. Collectively, SBL-based methods face a conflict between prominent sparsity performance achieved by hierarchical priors with low computational complexity. Methods based on both SBL and DL are pending further development, especially in theoretical explicability.

In this paper, we creatively solve the contradiction between sparsity performance with complexity burdens in SBL-based methods. On the one hand, hierarchical priors are still adopted to indirectly enhance sparsity. On the other hand, the increased complexity essentially caused by the hierarchical priors in MMV models is tactfully reduced by two operations based on the transformed block-sparse model. In fact, the balance of both sparsity and complexity is the main innovation point in this paper. Not only that, but the block-sparse model provides great convenience for complexity reduction by decreasing matrix dimensions in terms of matrix operation properties, and hierarchical priors create an opportunity to embed the generalized approximate message passing (GAMP) technique for solving marginal distributions so as to reduce complexity greatly.

To be specific, as the thick and inevitable MMV model restricts the SBL to some extent, we directly vectorize the MMV model, resulting in a block-sparse model that is convenient to carry out Bayesian learning. Unluckily, the vectorization expands the model dimensions, leading to large complexity in later Bayesian learning. Despite this, block-sparse Bayesian learning, named BSBL, is still analytically derived and developed. Moreover, it is worth noting that the complexity of BSBL is reluctantly acceptable owing to our reasonable design of the iterative process. In order to further achieve complexity reduction, two operations are introduced. One is based on matrix operation properties to decrease the matrix dimensions. Specifically, we analyze the complexity of each iterative formula and select the one containing operations of large computation burdens. Later, the selected formulas of BSBL are simplified and approximated by terms of operational properties of Kronecker products and some reasonable preconditions, so the faster version is named BSBL-APPR. The other is based on the famous generalized approximate message passing (GAMP) technique. GAMP is developed to solve approximate marginal posteriors, which are exactly applicable in BSBL since BSBL is derived by iterative hyperparameters originating from marginal distributions. Therefore, the GAMP technique is able to be embedded into BSBL, and the only additional work is to derive the iterative process of GAMP so that GAMP is useful in our block-sparse model. Moreover, GAMP is able to decompose the high dimensional matrix operations into vector or scale operations, which achieves complexity reduction well. BSBL with embedded GAMP is named BSBL-GAMP. In addition, many SBL-based methods consider little about the intractable wideband cases. Since wideband sources are able to be regarded as a superposition of many narrowband sources, we extend the proposed BSBL to wideband cases. The whole algorithm for wideband cases is derived and finished. Last but not least, the temporal correlation, which is not often considered in SBL, is modeled in the block-sparse model. Therefore, all the above methods are able to suppress temporal correlation.

In summary, the contributions of this paper are as follows:

- Hierarchical priors are adopted to enhance sparsity, and a block-sparse model is generated to carry out Bayesian learning easily. Hierarchical priors play an important role in l_p -norm optimization and outperform Gaussian priors in sparsity constraint, indirectly resulting in better sparsity performance. In the MMV case, the equivalently transformed block-sparse model laid a foundation for complexity reduction. Combining hierarchical priors with the block-sparse model allows for the balance of sparsity and complexity;
- Two operations are created to reduce complexity based on the block-sparse model. One exploits matrix operation properties to approximate high-dimension operations of derived formulas in the iterative process, while the other leverages the GAMP technique to simplify the iteration for computing marginal distributions, so that the matrix operations are decomposed into vector or scale operations;
- For wideband sources appearing in practice, the proposed BSBL is extended to be applicable in terms of the decomposition of wideband signals into narrowband ones. Moreover, the temporal correlation is considered by introducing a temporally correlated matrix into our data model. The designed iterative process of BSBL is able to be robust to temporal correlation.

The rest of this paper is organized as follows. In Section 2, the DOA estimation problem is abstracted from radar detection. Furthermore, the DOA estimation is equivalently transformed into an SSR problem based on the exploited block-sparse model. In Section 3, the traditional SBL based on our model is briefly introduced and derived, and its defects are presented by simple analysis. In Section 4, the proposed BSBL, BSBL-APPR, BSBL-GAMP, and BSBL for wideband cases are analytically derived, and corresponding iterative processes are presented. In Section 5, the performance of the proposed methods is evaluated comprehensively. In Section 6, conclusions are drawn.

For the sake of convenience, the notations are listed in Table 1.

Table 1. List of notations.

Symbol	Description
$(\cdot)^T$	Transpose
$(\cdot)^*$	Complex conjugate
$(\cdot)^H$	Hermitian transpose
\mathbb{R}	Set of real numbers
\mathbb{C}	Set of complex numbers
$\mathcal{CN}(\mathbf{x} \boldsymbol{\mu}, \boldsymbol{\Sigma})$	\mathbf{x} Obeys a complex Gaussian distribution with mean $\boldsymbol{\mu}$ and variance $\boldsymbol{\Sigma}$
\otimes	Kronecker product
\odot	Hadamard product
\mathbf{I}_N	$N \times N$ identity matrix
$\mathbf{0}$	Vector with all zero elements
<i>const</i>	Constant
w.r.t.	With respect to
$p(a b)$	Conditional probability density distribution of variable a w.r.t. variable b
$p(a; b)$	Probability density distribution of variable a w.r.t. parameter b
$q(\cdot)$	Probability density distribution
$\langle \cdot \rangle_{q(\cdot)}$	Expectation with respect to $q(\cdot)$
$diag(\cdot)$	Transforming matrix/vector into vector/matrix diagonally
i.i.d.	Independent and identically distributed
$vec(\cdot)$	Vectorization function
$unvec(\cdot)$	Matrixing function
$[\cdot]$	Top integral function
$\text{mod}(a, b)$	Function to find the remainder of a divided by b
\mathbf{A}^n	Matrix \mathbf{A} to the power of n
$tr(\cdot)$	Function to obtain the trace of a matrix
$\mathbf{A}_{i \cdot}, \mathbf{A}_{\cdot j}$	The i -th row of matrix \mathbf{A} and the j -th column of matrix \mathbf{A}
$A_{i,j}$	The element in the i -th row and j -th column of matrix \mathbf{A}
$\ \cdot \ _p$	Obtain l_p norm for each row of a matrix
$\ \cdot \ _{p,q}$	Obtain l_p norm after finding l_q norm
$E(\cdot)$	Expectation
$\mathbf{1}_{L \times L}$	$L \times L$ Matrix with all elements 1
$ \cdot $	Function to find absolute value or determinant
$\arg(\cdot)$	Function to find phase

2. Problem Reformulation

In radar or sonar detection, for communication, localization, and navigation, the antenna array receiver can receive different signals from various directions. Taking far-field radar detection as an example, adjacent antenna sensors have the same phase difference due to the plane electromagnetic wave if every adjacent couple has the same distance. Thus, a steer vector can be abstracted as $\mathbf{a}(\theta) = [1, \exp(-j2\pi d \sin(\theta)/\lambda), \dots, \exp(-j2\pi(N-1)d \sin(\theta)/\lambda)]^T \in \mathbb{C}^{N \times 1}$. θ is the direction of the source signal. λ is the wavelength. d is the distance between adjacent sensors. N is the number of sensors. For all the K sources from different directions $\{\theta_{k=1}^K\}$, a steer matrix is yielded as $\mathbf{A} = [\mathbf{a}(\theta_1), \dots, \mathbf{a}(\theta_K)] \in \mathbb{C}^{N \times K}$. \mathbf{A} is vital in DOA estimation because it implies the directions of all the sources based on the whole antenna array composed of N sensors. Radio frequency signals were received by an antenna array, whose sensors transfer individually received signals to independent channels, in which down conversion is conducted to produce intermediate-frequency signals, i.e., baseband signals. Later, using prior signals to execute matched filtering at time l , the receiver can gain a complex reflection factor of all the K sources, i.e., $\mathbf{s}(l) = [s_1, \dots, s_k, \dots, s_K]^T \in \mathbb{C}^{K \times 1}$. s_k and $\forall k = 1, \dots, K$ are the products of complex reflection coefficients and Doppler shifts. Generally, radar takes pulse accumulation to enhance signal processing ability when the echo pulses are in the same coherent processing interval (CPI), in which little fluctuation happens between different originally received signals and processed signals. In other words, sources are motionless, and the processing signals are nearly consistent during a CPI. Last but not least, the noises of N channels in a CPI are indispensable and assumed to be mutually independent. Overall, the above entire process is briefly shown in Figure 1, where the signal model is expressed as

$$\mathbf{X} = \mathbf{AS} + \mathbf{N}. \quad (1)$$

where $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_L] \in \mathbb{C}^{N \times L}$ is the ideal data of N channels in a CPI containing L snapshots. $\mathbf{S} = [\mathbf{s}(1), \dots, \mathbf{s}(l), \dots, \mathbf{s}(L)] \in \mathbb{C}^{K \times L}$ is the complex reflection factor matrix of the K sources in a CPI. $\mathbf{N} = [\mathbf{n}_1, \dots, \mathbf{n}_L]$ is the assumed noise matrix with each entry \mathbf{n}_l obeying an i.i.d. Gaussian distribution denoted as $\mathcal{CN}(0, \alpha^{-1})$.

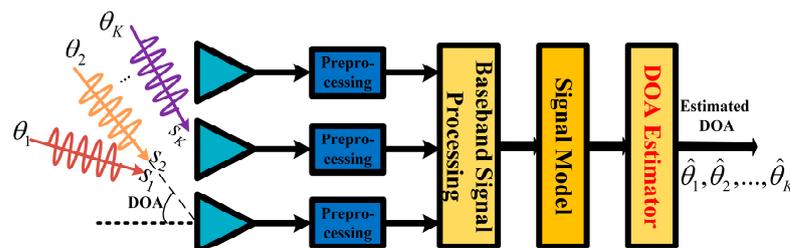


Figure 1. Flowchart of radar signal processing for DOA estimation.

Let $\{\bar{\theta}_{m=1}^M\}$ be M -grid sampling that covers spatial range $[-\pi/2, \pi/2]$. If sources are located on the grid exactly, (1) can be transformed into a sparse model expressed as

$$\mathbf{X} = \bar{\mathbf{A}}\mathbf{P} + \mathbf{N}. \quad (2)$$

where $\bar{\mathbf{A}} = [\mathbf{a}(\bar{\theta}_1), \dots, \mathbf{a}(\bar{\theta}_M)] \in \mathbb{C}^{N \times M}$ is the sparsely extended manifold matrix with the angle set $\{\bar{\theta}_{m=1}^M\}$ generated by the M -grid sampling. As $K \ll M$ holds in most cases, $\mathbf{P} \in \mathbb{C}^{M \times L}$ is the zero-padding version of \mathbf{S} , with each row representing a potential source that maps the M -grid spatial angular sampling.

Patently, processed data in a CPI are equivalently transformed into (2), which are typically sparse data in the MMV model. Now, the objective is to solve the sparse \mathbf{P} with known data \mathbf{X} and steer matrix $\bar{\mathbf{A}}$. It is worth emphasizing two mainstream methods to handle (2). One is dealing with (2) directly, which is more difficult but has less computa-

tional complexity. The other is solving the vectorization (2), which is simple and intuitive but has large complexity caused by the vectorization. In this paper, the latter is selected because we are able to reduce the complexity tactfully. Without loss of generality, (2) is vectorized as

$$\mathbf{x} = \Phi \mathbf{p} + \mathbf{n}. \quad (3)$$

where $\mathbf{x} = \text{vec}(\mathbf{X}^T) \in \mathbb{C}^{NL \times 1}$, $\Phi = \bar{\mathbf{A}} \otimes \mathbf{I}_L \in \mathbb{C}^{NL \times ML}$, $\mathbf{p} = \text{vec}(\mathbf{P}^T) \in \mathbb{C}^{ML \times 1}$, and $\mathbf{n} = \text{vec}(\mathbf{N}^T) \in \mathbb{C}^{NL \times 1}$. After the vectorization, a block-sparse vector \mathbf{p} is yielded since the original \mathbf{P} contains many zero rows. In addition, the matrix dimensions grow after the vectorization, leading to large complexity, but the problem will be solved by our approximation eventually. Based on the block-sparse model shown in (3), the DOA estimation is transformed into a sparse recovery problem, i.e., to solve \mathbf{p} with known \mathbf{x} and Φ .

3. Canonical SBL Method

To solve \mathbf{p} in (3), the canonical SBL method is introduced and briefly shown as follows. According to (3), the likelihood is

$$p(\mathbf{x}|\mathbf{p}; \alpha) \sim \mathcal{CN}(\Phi \mathbf{p}, \alpha^{-1} \mathbf{I}_{NL}). \quad (4)$$

The prior of \mathbf{p} is supposed to be

$$p(\mathbf{p}; \gamma_i, \forall i = 1, \dots, M) \sim \mathcal{CN}(\mathbf{0}, \Sigma_0). \quad (5)$$

where γ_i is a hyperparameter representing a potential source, and Σ_0 is expressed as

$$\Sigma_0 = \begin{bmatrix} \gamma_1 \mathbf{I}_L & & \\ & \ddots & \\ & & \gamma_M \mathbf{I}_L \end{bmatrix}. \quad (6)$$

Based on the Bayesian formula, the posterior of \mathbf{p} is

$$p(\mathbf{p}|\mathbf{x}; \alpha, \gamma_i, \forall i) = \frac{p(\mathbf{x}|\mathbf{p}; \alpha) p(\mathbf{p}; \gamma_i, \forall i)}{\int p(\mathbf{x}|\mathbf{p}; \alpha) p(\mathbf{p}; \gamma_i, \forall i) d\mathbf{p}}. \quad (7)$$

The posterior of (7) is rigorously solved as a Gaussian distribution with mean and covariance as follows:

$$\mu_{\mathbf{p}0} = \alpha \Sigma_{\mathbf{p}0} \Phi^H \mathbf{x}. \quad (8)$$

$$\Sigma_{\mathbf{p}0} = \Sigma_0 - \Sigma_0 \Phi^H (\alpha^{-1} \mathbf{I} + \Phi \Sigma_0 \Phi^H)^{-1} \Phi \Sigma_0. \quad (9)$$

The likelihood, prior, and posterior are uniquely determined by the hyperparameter set $\Theta = \{\gamma_i, \alpha, \forall i\}$. According to the maximum a posterior (MAP) criterion, the Expectation–Maximization (EM) algorithm [53] is used to maximize $p(\mathbf{p}|\mathbf{x}; \Theta)$. Here, \mathbf{p} is treated as a hidden variable to obtain the relationship between the hyperparameter set (i.e., Θ^{new}) and the old one (i.e., Θ^{old}) by maximizing the following term.

$$Q(\Theta^{new}) = E_{\mathbf{p}|\mathbf{x}, \Theta^{old}}[\ln p(\mathbf{p}|\mathbf{x}; \Theta)]. \quad (10)$$

Omitting the specific derivation, the final iteration solutions of the hyperparameters are expressed as

$$\gamma_i^{new} = \frac{\text{tr}[\Sigma_{\mathbf{p}0}^i + \mu_{\mathbf{p}0}^i (\mu_{\mathbf{p}0}^i)^H]}{L}, \forall i. \quad (11)$$

$$\alpha^{new} = \frac{NL}{\|\mathbf{x} - \Phi \mu_{\mathbf{p}0}\|_2^2 + \text{tr}(\Sigma_{\mathbf{p}0} \Phi^H \Phi)}. \quad (12)$$

where $\Sigma_{\mathbf{p}0}^i = \Sigma_{\mathbf{p}0}([(i-1)L+1:iL], [(i-1)L+1:iL])$ and $\mu_{\mathbf{p}0}^i = \mu_{\mathbf{p}0}([(i-1)L+1:iL])$.

The canonical SBL algorithm uses (8), (9), (11), and (12) iteratively to estimate γ until convergence, and the final $\gamma = [\gamma_1, \dots, \gamma_M]^T$ is regarded as the solved \mathbf{p} . However, recalling the whole process, two limitations are present: (i) the single Gaussian priors cannot enhance sparsity well and (ii) the computational complexity $O(M^3L^3)$, dominated by (12), is usually unacceptable in practice. Consequently, we develop an SBL-based method to resolve (3) in this paper.

4. Proposed Methods

4.1. BSBL

Without loss of generality, the SBL-based methods are required to construct a Bayesian framework and then complete the corresponding Bayesian inference to develop an iterative algorithm.

4.1.1. Bayesian Framework

Bayesian framework is composed of prior distributions of observed data and unknown variables.

Remark 1. According to the MAP criterion, priors are essential for SBL-based methods because the iterative process to be constructed is based on the derivatives of different variables so as to ensure the maximal posterior. Therefore, the prior distributions need to be clarified first.

In this paper, the prior distribution (i.e., likelihood) of the observed data \mathbf{x} is similar to (4), i.e.,

$$p(\mathbf{x}|\mathbf{p}; \alpha, \gamma) \sim \mathcal{CN}(\Phi\mathbf{p}, \alpha^{-1}\mathbf{I}_{NL}). \quad (13)$$

For better sparsity performance, we propose hierarchical priors containing Gaussian and Gamma priors. The reason for selecting Gamma priors is that the Gamma distribution is the conjugate prior to the inverse variance of the Gaussian distribution [54]. As usual, the prior \mathbf{p} obeys the i.i.d. complex Gaussian distribution, i.e.,

$$p(\mathbf{p}|\gamma) \sim \mathcal{CN}(\mathbf{0}, \Sigma). \quad (14)$$

where $\gamma = [\gamma_1, \dots, \gamma_M]^T$, $\gamma^{-1} = [\gamma_1^{-1}, \dots, \gamma_M^{-1}]^T$, and $\Sigma = \text{diag}(\gamma^{-1}) \otimes \mathbf{B}$. Note that $\mathbf{B} \in \mathbb{C}^{L \times L}$ representing temporally correlated level is not equal to \mathbf{I}_L in the canonical SBL method. Therefore, the proposed method will be able to suppress temporal correlation, which will be tested and verified in Section 5. \mathbf{B} is generally modeled as a Toeplitz matrix, i.e.,

$$\mathbf{B} = \begin{pmatrix} 1 & \beta & \dots & \beta^{L-1} \\ \beta & 1 & \dots & \beta^{L-2} \\ \vdots & \vdots & \ddots & \vdots \\ \beta^{L-1} & \beta^{L-2} & \dots & 1 \end{pmatrix}. \quad (15)$$

where β is the complex correlation coefficient with the amplitude $|\beta| \in [0, 1]$ and phase $\arg(\beta) \in [0, 2\pi]$.

γ obeys a Gamma distribution, i.e.,

$$p(\gamma; a, b) = \prod_{m=1}^M \frac{b^a}{\Gamma(a)} \gamma_m^{a-1} e^{-b\gamma_m}. \quad (16)$$

where $\Gamma(a) = \int_0^\infty x^{a-1} \exp(-x) dx$. a and b are the shape parameter and scale parameter. Then, a Gamma prior is applied to α so that

$$p(\alpha; c, d) = \frac{c^d}{\Gamma(c)} \alpha^{c-1} e^{-d\alpha}. \quad (17)$$

where c and d are the corresponding shape and scale parameters, respectively.

It is worth emphasizing that the true prior distribution of \mathbf{p} can be solved with Student's t-distribution, which promotes sparsity better than the traditional Gaussian distribution [47].

4.1.2. Bayesian Inference

Bayesian inference is necessary for the eventual iterative algorithm, and the crux is to deduce the posterior. Unfortunately, according to our Bayesian framework, the posterior is intractable. However, we just need to maximize the posterior by maximizing the evidence procedure, regardless of the analytical closed form of the posterior. Coincidentally, OGSBI [54] provides an example to maximize evidence, i.e., the marginal probability of the observed data \mathbf{x}

$$p(\mathbf{x}) = \int \int \int p(\mathbf{x}|\mathbf{p}; \alpha, \gamma) p(\mathbf{p}|\gamma) p(\gamma; a, b) p(\alpha; c, d) d\mathbf{p} d\gamma d\alpha. \quad (18)$$

But, (18) is still intractable, so maximizing evidence seems to help little. However, variational inference is able to achieve it. It is necessary to explain variational inference before later derivation. Variational inference defines a function as a mapping that takes a function as the input and returns the value of the function as the output [55]. In the entropy field, the function is a probability distribution. When variational inference is applied to (18), the parameter vector (i.e., unknown stochastic variables) no longer appears because the parameters are absorbed into new probability distributions. Thus, (18) can be converted to an addressable form.

To be specific, we adopt variational Bayesian inference (VBI) [56] to address (18) by introducing a distribution $q(\Theta)$, where $\Theta = \{\mathbf{p}, \gamma, \alpha\}$ is the parameter set of unknown variables. The introduced $q(\Theta)$ can simplify (18) and allow the logarithmic form of (18) to be divided into two parts, i.e.,

$$\ln p(\mathbf{x}) = \underbrace{\int q(\Theta) \ln \frac{p(\mathbf{x}, \Theta)}{q(\Theta)} d\Theta}_{F(q, \Theta)} - \underbrace{\int q(\Theta) \ln \frac{p(\Theta|\mathbf{x})}{q(\Theta)} d\Theta}_{KL(q||p)}. \quad (19)$$

where $p(\mathbf{x}, \Theta) = p(\mathbf{x}|\mathbf{p}; \alpha, \gamma) p(\mathbf{p}|\gamma) p(\gamma; a, b) p(\alpha; c, d)$ is the product of priors and $p(\Theta|\mathbf{x})$ is the posterior. The specific derivation from (18) to (19) is complicated [55], so it is omitted here. $F(q, \Theta)$ is the lower bound of $\ln p(\mathbf{x})$ because $KL(q||p) \geq 0$ is the Kullback–Leibler divergence between $q(\Theta)$ and the posterior $p(\Theta|\mathbf{x})$.

The significance of (19) is transforming the intractable $\ln p(\mathbf{x})$ into an approximated tractable $F(q, \Theta)$, so that maximizing $\ln p(\mathbf{x})$ is approximately equal to maximizing $F(q, \Theta)$. The lower bound $F(q, \Theta)$ is a functional in terms of $q(\Theta)$. In other words, $F(q, \Theta)$ is a mapping that takes as input a function $q(\Theta)$ and returns the value of the function as the output. Similar to the function derivative, maximizing $F(q, \Theta)$ requires some optimization over specific forms of $q(\Theta)$. In Bayesian inference, the commonly used form is factorization [56].

As performed in [56], to achieve the maximized $F(q, \Theta)$, $q(\Theta)$ is factorized into independent parts, i.e.,

$$q(\Theta) = \prod_i q_i(\Theta_i) = q(\mathbf{p})q(\gamma)q(\alpha). \quad (20)$$

where $q(\mathbf{p})$, $q(\gamma)$, and $q(\alpha)$ correspond marginal distributions of the hidden variables \mathbf{p} , γ , and α . $q_i(\Theta_i)$ is expressed as

$$q_i(\Theta_i) = \frac{\exp\langle \ln p(\mathbf{x}, \Theta) \rangle_{j \neq i}}{\int \exp\langle \ln p(\mathbf{x}, \Theta) \rangle_{j \neq i} d\Theta_i}. \quad (21)$$

(21) is hard to compute, but its logarithmic form is easy to obtain, i.e.,

$$\ln q_i(\Theta_i) = \langle \ln p(\mathbf{x}, \Theta) \rangle_{j \neq i} + \text{const.} \quad (22)$$

Using (22), marginal distributions $q(\mathbf{p})$, $q(\gamma)$, and $q(\alpha)$ can be solved.

Step 1: $\ln q(\mathbf{p})$ satisfies

$$\ln q(\mathbf{p}) = \langle \ln p(\mathbf{x}|\mathbf{p}, \alpha) p(\mathbf{p}; \gamma) \rangle_{q(\gamma)q(\alpha)} + \text{const.} \quad (23)$$

Utilizing (13), (14), and (23), $q(\mathbf{p})$ can be solved as a Gaussian distribution, with mean and variance given by

$$\boldsymbol{\mu}_{\mathbf{p}} = \langle \alpha \rangle \boldsymbol{\Sigma}_{\mathbf{p}} \boldsymbol{\Phi}^H \mathbf{x}. \quad (24)$$

$$\boldsymbol{\Sigma}_{\mathbf{p}} = (\langle \alpha \rangle \boldsymbol{\Phi}^H \boldsymbol{\Phi} + \langle \boldsymbol{\Sigma} \rangle^{-1})^{-1}. \quad (25)$$

Please refer to Appendix A for the proof. In general, (24) and (25) are equivalently transformed into being less complex according to the properties of the matrix in [46].

$$\boldsymbol{\mu}_{\mathbf{p}} = \langle \boldsymbol{\Sigma} \rangle \boldsymbol{\Phi}^H (\langle \alpha \rangle^{-1} \mathbf{I}_{NL} + \boldsymbol{\Phi} \langle \boldsymbol{\Sigma} \rangle \boldsymbol{\Phi}^H)^{-1} \mathbf{x}. \quad (26)$$

$$\boldsymbol{\Sigma}_{\mathbf{p}} = \langle \boldsymbol{\Sigma} \rangle - \langle \boldsymbol{\Sigma} \rangle \boldsymbol{\Phi}^H (\langle \alpha \rangle^{-1} \mathbf{I}_{NL} + \boldsymbol{\Phi} \langle \boldsymbol{\Sigma} \rangle \boldsymbol{\Phi}^H)^{-1} \boldsymbol{\Phi} \langle \boldsymbol{\Sigma} \rangle. \quad (27)$$

Step 2: $\ln q(\gamma)$ satisfies

$$\ln q(\gamma) = \langle \ln p(\mathbf{p}; \gamma) p(\gamma; a, b) \rangle_{q(\mathbf{p})q(\alpha)} + \text{const.} \quad (28)$$

Using (14), (16), and (28), $q(\gamma)$ is identified as a Gamma distribution, whose shape parameter, the m -th scale parameters, and the m -th element of the mean are as follows:

$$\bar{a} = a + \frac{1}{2}. \quad (29)$$

$$\bar{b}_m = b + \frac{1}{2} \langle \mathbf{p}_m^H \mathbf{B}^{-1} \mathbf{p}_m \rangle. \quad (30)$$

$$\langle \gamma_m \rangle = \frac{\bar{a}}{\bar{b}_m}. \quad (31)$$

where $\mathbf{p}_m \in \mathbb{C}^{L \times 1}$ is the m -th entry of $\mathbf{p} = [\mathbf{p}_1^T, \dots, \mathbf{p}_M^T]^T$ and $m = 1, 2, \dots, M$. Please refer to Appendix B for the proof. (30) can be simplified further, i.e.,

$$\bar{b}_m = b + \frac{1}{2} \|(\mathbf{B}^{-1} + \mathbf{1}_{L \times L} - \mathbf{I}_L) \odot (\boldsymbol{\mu}_{\mathbf{p}_m} \boldsymbol{\mu}_{\mathbf{p}_m}^H)\|_{1,1} + \frac{1}{2} \text{diag}(\mathbf{B}^{-1} \odot \boldsymbol{\Sigma}_{\mathbf{p}_m}). \quad (32)$$

where $\boldsymbol{\mu}_{\mathbf{p}_m}$ is the m -th entry of $\boldsymbol{\mu}_{\mathbf{p}} = [\boldsymbol{\mu}_{\mathbf{p}_1}^T, \dots, \boldsymbol{\mu}_{\mathbf{p}_M}^T]^T$, $\boldsymbol{\Sigma}_{\mathbf{p}_m} = \langle \gamma_m^{-1} \rangle \mathbf{B}$, and $m = 1, 2, \dots, M$. Please refer to Appendix C for details of the derivation.

Step 3: $\ln q(\alpha)$ satisfies

$$\ln q(\alpha) = \langle \ln p(\mathbf{x}|\mathbf{p}, \alpha) p(\alpha; c, d) \rangle_{q(\mathbf{p})q(\gamma)} + \text{const.} \quad (33)$$

Using (13), (17), and (33), $q(\alpha)$ is also solved as a Gamma distribution with shape parameter \bar{c} , scale parameter \bar{d} , and mean $\langle \alpha \rangle$ as follows:

$$\bar{c} = c + \frac{NL}{2}. \quad (34)$$

$$\bar{d} = d + \frac{1}{2} \langle (\mathbf{x} - \boldsymbol{\Phi} \mathbf{p})^H (\mathbf{x} - \boldsymbol{\Phi} \mathbf{p}) \rangle = d + \frac{1}{2} (\|\mathbf{x} - \boldsymbol{\Phi} \boldsymbol{\mu}_{\mathbf{p}}\|_2^2 + \text{tr}(\boldsymbol{\Sigma}_{\mathbf{p}} \boldsymbol{\Phi}^H \boldsymbol{\Phi})). \quad (35)$$

$$\langle \alpha \rangle = \frac{\bar{c}}{\bar{d}}. \quad (36)$$

Please refer to Appendix D for the proof. For the convenience of iteration, (35) is further simplified as

$$\bar{d} = d + \frac{1}{2}(\|\mathbf{x} - \Phi \mu_{\mathbf{p}}\|_2^2 + \text{tr}(\Sigma_{\mathbf{p}} \Phi^H \Phi)) = d + \frac{1}{2}\|\mathbf{x} - \Phi \mu_{\mathbf{p}}\|_2^2 + \frac{1}{2}\langle \alpha \rangle^{-1}[ML - \text{tr}(\Sigma_{\mathbf{p}} \langle \Sigma \rangle^{-1})]. \quad (37)$$

Overall, $q(\mathbf{p})$, $q(\boldsymbol{\gamma})$, and $q(\alpha)$ are solved, and, alternately, are updating (26), (27), (31), (36) until some convergence criterion is satisfied.

Remark 2. According to [57], maximizing the lower bound $F(q, \Theta)$ guarantees convergence of the iterative optimization since each iteration leads to a nondecreasing value of $F(q, \Theta)$. Therefore, the proposed method must converge at some point.

4.1.3. Off-Grid Correction

Recalling (2), \mathbf{P} with angle set $\{\bar{\theta}_{m=1}^M\}$ is yielded by spatial discretization, causing estimation errors inevitably if sources are off grid. On this basis, the array steering vector of the i -th source is Taylor expanded around the nearest sampling grid denoted as $\bar{\theta}_{m_i}$, i.e.,

$$\mathbf{a}(\theta_i) \approx \mathbf{a}(\bar{\theta}_{m_i}) + \mathbf{a}_1(\bar{\theta}_{m_i})\Delta\theta_i. \quad (38)$$

where $\Delta\theta_i = \theta_i - \bar{\theta}_{m_i} \in [-\frac{1}{2}\|\bar{\theta}_m - \bar{\theta}_{m+1}\|, \frac{1}{2}\|\bar{\theta}_m - \bar{\theta}_{m+1}\|]$ and $\mathbf{a}_1(\bar{\theta}_{m_i}) = \mathbf{a}'(\bar{\theta}_{m_i})$. Patently, the final objective is to solve $\Delta\theta_i$. Following the above similar principle based on Taylor expansion, Φ can be extended as

$$\Phi(\boldsymbol{\lambda}) = \Phi + \Phi' \text{diag}(\boldsymbol{\lambda}). \quad (39)$$

where $\boldsymbol{\lambda} = [\lambda_1, \dots, \lambda_{ML}]^T$ with the i -th element is $\lambda_i \in [-\frac{1}{2}\|\bar{\theta}_m - \bar{\theta}_{m+1}\|, \frac{1}{2}\|\bar{\theta}_m - \bar{\theta}_{m+1}\|]$ and $\forall i = 1, \dots, ML$. To solve $\boldsymbol{\lambda}$, (3) can be used, i.e.,

$$E(\mathbf{x} - \Phi(\boldsymbol{\lambda})\mathbf{p}) = \mathbf{x} - \Phi(\boldsymbol{\lambda})\mu_{\mathbf{p}} = \mathbf{0}. \quad (40)$$

Combining (40) with (39), the following equation holds.

$$\boldsymbol{\lambda} \odot \mu_{\mathbf{p}} = [(\Phi')^H \Phi']^{-1} (\Phi')^H (\mathbf{x} - \Phi \mu_{\mathbf{p}}). \quad (41)$$

Using (41), $\boldsymbol{\lambda}$ is solved and is still block sparse; thus, we squeeze it as $\boldsymbol{\lambda}^{squ} = [\lambda_1^{squ}, \dots, \lambda_M^{squ}]^T$, where the m -th element $\lambda_m^{squ} = \text{sum}\{\boldsymbol{\lambda}([(m-1)L+1 : mL])\} / L$. Here, if λ_m^{squ} exceeds the interval $[-\frac{1}{2}\|\bar{\theta}_m - \bar{\theta}_{m+1}\|, \frac{1}{2}\|\bar{\theta}_m - \bar{\theta}_{m+1}\|]$, $\pm \frac{1}{2}\|\bar{\theta}_m - \bar{\theta}_{m+1}\|$ will be assigned to it. Letting the preliminary estimated DOA value vector be $\boldsymbol{\theta}^{est1} = [\hat{\theta}_1, \dots, \hat{\theta}_k, \dots, \hat{\theta}_K]^T$, the final DOA vector, denoted as $\boldsymbol{\theta}^{est2}$, is solved as follows:

$$\boldsymbol{\theta}^{est2} = \boldsymbol{\theta}^{est1} + \boldsymbol{\lambda}^{squ,est}. \quad (42)$$

where $\boldsymbol{\lambda}^{squ,est} = [\lambda_{m_1}^{squ}, \dots, \lambda_{m_k}^{squ}, \dots, \lambda_{m_K}^{squ}]^T$, $\lambda_{m_k}^{squ}$ is the corresponding angle compensation for $\hat{\theta}_k$. Using (42), grid errors are eliminated.

Overall, the whole BSBL algorithm is completed and summarized in Algorithm 1.

Algorithm 1. The proposed BSBL algorithm.

Initialization

- (i) set the first iterative number $k = 0$, $\mathbf{p}^{(0)} = \|\bar{\mathbf{A}}^H \mathbf{X}\|_2$.
- (ii) assign a , b , c and d very small values (ensure uninformative distributions).
- (iii) preset error tolerance ε and maximal iterative number k_{\max} .

Repetition

- while** ($\|\mathbf{p}^{(k+1)} - \mathbf{p}^{(k)}\|_2 / \|\mathbf{p}^{(k)}\|_2 > \varepsilon$ or $k < k_{\max}$) **do**
- (i) compute $\mu_{\mathbf{p}}$ and $\Sigma_{\mathbf{p}}$ according to (26) and (27).
- (ii) compute $\langle \Sigma \rangle$ and $\langle \alpha \rangle$ according to (31) and (36).
- (iii) Regard γ as $\mathbf{p}^{(k)}$.
- (iv) $k = k + 1$.

end while

Refinement

- (i) use the final \mathbf{p} to obtain θ^{est1} .
- (ii) use the final $\mu_{\mathbf{p}}$ to solve λ according to (41).
- (iii) obtain θ^{est2} according to (42).

Output The final DOA values.

Compared with the canonical SBL method, BSBL achieves better sparsity performance and lower computational complexity. According to the maximal number of complex multiplications, the complexity of BSBL, dominated by (26) (or rather $\Phi \langle \Sigma \rangle \Phi^H$), is $O(M^2NL^3)$, less than $O(M^3L^3)$ of the canonical SBL method.

However, BSBL will still suffer heavy computational burdens when L or M is large. Therefore, we must seek some techniques to reduce computational complexity.

4.2. BSBL-APPR

Obviously, the large complexity is mainly caused by high-dimensional matrix operations that contain massive useless zero (or near-zero) operations. Automatically, the simplest perspective is to exploit some approximation to shrink the matrix dimensions, generating the first faster version called BSBL-APPR.

Recalling the entire algorithm, the high dimensions are essentially yielded by computing (26), (27), and (37). For $\mu_{\mathbf{p}}$ in (26), the corresponding approximation is

$$\begin{aligned}
 \mu_{\mathbf{p}} &= \langle \Sigma \rangle \Phi^H (\langle \alpha \rangle^{-1} \mathbf{I}_{NL} + \Phi \langle \Sigma \rangle \Phi^H)^{-1} \mathbf{x} \\
 &= (\langle \Gamma^{-1} \rangle \otimes \mathbf{B}) (\bar{\mathbf{A}} \otimes \mathbf{I}_L)^H [\langle \alpha \rangle^{-1} \mathbf{I}_{NL} + (\bar{\mathbf{A}} \otimes \mathbf{I}_L) (\langle \Gamma^{-1} \rangle \otimes \mathbf{B}) (\bar{\mathbf{A}} \otimes \mathbf{I}_L)^H]^{-1} \text{vec}(\mathbf{X}^T) \\
 &\approx (\langle \Gamma^{-1} \rangle \bar{\mathbf{A}}^H \otimes \mathbf{B}) [(\langle \alpha \rangle^{-1} \mathbf{I}_N + \bar{\mathbf{A}} \langle \Gamma^{-1} \rangle \bar{\mathbf{A}}^H)^{-1} \otimes \mathbf{B}] \text{vec}(\mathbf{X}^T) \\
 &= [\langle \Gamma^{-1} \rangle \bar{\mathbf{A}}^H (\langle \alpha \rangle^{-1} \mathbf{I}_N + \bar{\mathbf{A}} \langle \Gamma^{-1} \rangle \bar{\mathbf{A}}^H)^{-1} \otimes \mathbf{B}^2] \text{vec}(\mathbf{X}^T) \\
 &= \text{vec}(\mathbf{B}^2 \mathbf{X}^T [(\langle \Gamma^{-1} \rangle \bar{\mathbf{A}}^H (\langle \alpha \rangle^{-1} \mathbf{I}_N + \bar{\mathbf{A}} \langle \Gamma^{-1} \rangle \bar{\mathbf{A}}^H)^{-1}]^T).
 \end{aligned} \tag{43}$$

where $\Gamma^{-1} = \text{diag}(\gamma^{-1})$. The derivation process follows Kronecker-Product properties, i.e., $(\mathbf{A} \otimes \mathbf{B})^H = \mathbf{A}^H \otimes \mathbf{B}^H$, $(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = \mathbf{AC} \otimes \mathbf{BD}$, and $\text{vec}(\mathbf{ABC}) = (\mathbf{C}^T \otimes \mathbf{A}) \text{vec}(\mathbf{B})$. The approximation exactly holds if $\langle \alpha \rangle^{-1} = 0$ or $\mathbf{B} = \mathbf{I}_L$. To be specific, the approximation is reasonable if high SNRs or low correlation coefficient levels are adopted. In fact, the two conditions (or at least one) are easy to meet in practice.

Likewise, for $\Sigma_{\mathbf{p}}$ in (27), the approximation is

$$\begin{aligned}
 \Sigma_{\mathbf{p}} &= \langle \Sigma \rangle - \langle \Sigma \rangle \Phi^H (\langle \alpha \rangle^{-1} \mathbf{I}_{NL} + \Phi \langle \Sigma \rangle \Phi^H)^{-1} \Phi \langle \Sigma \rangle \\
 &= \langle \Gamma^{-1} \rangle \otimes \mathbf{B} - (\langle \Gamma^{-1} \rangle \otimes \mathbf{B}) (\bar{\mathbf{A}} \otimes \mathbf{I}_L)^H [\langle \alpha \rangle^{-1} \mathbf{I}_{NL} + (\bar{\mathbf{A}} \otimes \mathbf{I}_L) (\langle \Gamma^{-1} \rangle \otimes \mathbf{B}) (\bar{\mathbf{A}} \otimes \mathbf{I}_L)^H]^{-1} (\bar{\mathbf{A}} \otimes \mathbf{I}_L) (\langle \Gamma^{-1} \rangle \otimes \mathbf{B}) \\
 &\approx \langle \Gamma^{-1} \rangle \otimes \mathbf{B} - (\langle \Gamma^{-1} \rangle \bar{\mathbf{A}}^H \otimes \mathbf{B}) [(\langle \alpha \rangle^{-1} \mathbf{I}_N + \bar{\mathbf{A}} \langle \Gamma^{-1} \rangle \bar{\mathbf{A}}^H)^{-1} \otimes \mathbf{B}] (\bar{\mathbf{A}} \langle \Gamma^{-1} \rangle \otimes \mathbf{B}) \\
 &= \langle \Gamma^{-1} \rangle \otimes \mathbf{B} - \langle \Gamma^{-1} \rangle \bar{\mathbf{A}}^H (\langle \alpha \rangle^{-1} \mathbf{I}_N + \bar{\mathbf{A}} \langle \Gamma^{-1} \rangle \bar{\mathbf{A}}^H)^{-1} \bar{\mathbf{A}} \langle \Gamma^{-1} \rangle \otimes \mathbf{B}^3.
 \end{aligned} \tag{44}$$

Similarly, for \bar{d} in (37), its approximation is

$$\begin{aligned} \bar{d} &= d + \frac{1}{2}(\|\mathbf{x} - \Phi\boldsymbol{\mu}_p\|_2^2 + \text{tr}(\boldsymbol{\Sigma}_p\Phi^H\Phi)) \\ &\approx d + \frac{1}{2}\|\mathbf{x} - \Phi\boldsymbol{\mu}_p\|_2^2 + \frac{1}{2}\text{tr}\{(\langle\Gamma^{-1}\rangle \otimes \mathbf{B})(\bar{\mathbf{A}}^H\bar{\mathbf{A}} \otimes \mathbf{I}) - [\langle\Gamma^{-1}\rangle\bar{\mathbf{A}}^{-1}(\langle\alpha\rangle^{-1}\mathbf{I}_N + \bar{\mathbf{A}}\langle\Gamma^{-1}\rangle\bar{\mathbf{A}}^H)^{-1}\bar{\mathbf{A}}\langle\Gamma^{-1}\rangle \otimes \mathbf{B}^3](\bar{\mathbf{A}}^H\bar{\mathbf{A}} \otimes \mathbf{I})\} \\ &= d + \frac{1}{2}\|\mathbf{x} - \Phi\boldsymbol{\mu}_p\|_2^2 + \frac{1}{2}\text{tr}\{(\langle\Gamma^{-1}\rangle\bar{\mathbf{A}}^H\bar{\mathbf{A}} \otimes \mathbf{B}) - [\langle\Gamma^{-1}\rangle\bar{\mathbf{A}}^{-1}(\langle\alpha\rangle^{-1}\mathbf{I}_N + \bar{\mathbf{A}}\langle\Gamma^{-1}\rangle\bar{\mathbf{A}}^H)^{-1}\bar{\mathbf{A}}\langle\Gamma^{-1}\rangle\bar{\mathbf{A}}^H \otimes \mathbf{B}^3]\}. \end{aligned} \quad (45)$$

After the approximation, BSBL-APPR is completed. Its concrete iterative steps are omitted here since the process is the same as BSBL, except for the calculation of $\boldsymbol{\mu}_p$, $\boldsymbol{\Sigma}_p$, and \bar{d} by (43)–(45).

Since the approximation operations have been performed, the computational complexity of BSBL-APPR, dominated by $\bar{\mathbf{A}}\langle\Gamma^{-1}\rangle\bar{\mathbf{A}}^H$, is $O(M^2N)$ less than $O(M^2NL^3)$, which theoretically verifies the higher efficiency of BSBL-APPR.

Even so, the complexity of BSBL-APPR still seems to be intolerable when dense sampling is adopted, i.e., M is large enough. Additionally, there exist several operations with many zero (or near-zero) elements, e.g., $\bar{\mathbf{A}}\langle\Gamma^{-1}\rangle\bar{\mathbf{A}}^H$ and $\langle\Gamma^{-1}\rangle \otimes \mathbf{B}$. For lower complexity, the most effective method is to decompose matrix operations into vector and even scalar operations so as to selectively avoid useless computation. Fortunately, a technique, named generalized approximate message passing (GAMP), exactly achieves that.

4.3. BSBL-GAMP

GAMP is a technique developed to solve approximate marginal posteriors with low complexity based on the central limit theorem [58]. To briefly explain the principle of the GAMP technique, (3) is rewritten in scalar form.

$$x_i = \Phi_i \mathbf{p} + n_i, \forall i = 1, \dots, NL. \quad (46)$$

where x_i and n_i are the i -th entries of \mathbf{x} and \mathbf{n} . Let $z_i = \Phi_i \mathbf{p}$, $\mathbf{p} = [p_1, \dots, p_j, \dots, p_{ML}]^T$, and $j = 1, \dots, ML$. Given the known measurement matrix Φ and the observed vector \mathbf{x} , the objective is to obtain the estimation of \mathbf{p} .

Each x_i is connected to p_j by Φ , and vice versa. x_i and p_j are defined as the input node and the output node, respectively. The association between them is called an edge. Input nodes and output nodes pass messages to each other along the edges. The original message passing (MP) technique is to keep passing messages (i.e., probability distributions) with respect to p_j until convergence. Based on MP, GAMP is the extension for low complexity since it passes only important messages that mainly affect the approximated marginal posteriors of \mathbf{p} .

Overall, the GAMP technique is selected to speed up our algorithm for two motivations. (i) It passes messages from one node to another, enforcing scalar operations with low complexity. (ii) It can also compute approximate marginal posteriors of \mathbf{p} , which allows GAMP to be embedded into the proposed BSBL.

To apply GAMP to our used block-sparse data model, we must derive it again. Along the line of the common procedures of GAMP, there are two important approximate marginal posteriors to consider. One is

$$p(z_i|\mathbf{x}, v_i^z, \tau_i^z, \boldsymbol{\eta}) = \frac{p(x_i|z_i, \boldsymbol{\eta})\mathcal{CN}(z_i|v_i^z, \tau_i^z)}{\int_z p(x_i|z_i, \boldsymbol{\eta})\mathcal{CN}(z_i|v_i^z, \tau_i^z)dz}. \quad (47)$$

for approximating $p(z_i|\mathbf{x}, \boldsymbol{\eta})$, $i = 1, 2, \dots, NL$, where $\boldsymbol{\eta} = \{\boldsymbol{\gamma}, \alpha\}$. v_i^z and τ_i^z are quantities to be updated. The other is

$$p(p_j|\mathbf{x}, v_j^p, \tau_j^p, \boldsymbol{\eta}) = \frac{p(p_j|\boldsymbol{\eta})\mathcal{CN}(p_j|v_j^p, \tau_j^p)}{\int_p p(p_j|\boldsymbol{\eta})\mathcal{CN}(p_j|v_j^p, \tau_j^p)dp}. \quad (48)$$

for approximating $p(p_j|\mathbf{x}, \boldsymbol{\eta})$, $j = 1, 2, \dots, ML$, where v_j^p and τ_j^p are quantities to be updated. Given the model and priors, it is easy to obtain $p(x_i|z_i, \boldsymbol{\eta}) = \mathcal{CN}(x_i|z_i, \alpha^{-1})$ and $p(p_j|\boldsymbol{\eta}) = \mathcal{CN}(p_j|0, \gamma_{j_1}^{-1}\beta_{j_2, j_2})$, where β_{j_2, j_2} is the j_2 -th row, j_2 -th column element of \mathbf{B} , and $(j_1, j_2) = \begin{cases} (m, L), & \text{if } j = mL \\ (\lceil j/L \rceil, \text{mod}(j, L)), & \text{else} \end{cases}, \forall m = 1, \dots, M. j_1 = 1, \dots, M, j_2 = 1, \dots, L$. Similar to (7), (47) and (48) are easily identified as Gaussian distributions.

$$p(z_i|\mathbf{x}, v_i^z, \tau_i^z, \boldsymbol{\eta}) = \mathcal{CN}(z_i|\mu_i^z, \phi_i^z). \quad (49)$$

$$p(p_j|\mathbf{x}, v_j^p, \tau_j^p, \boldsymbol{\eta}) = \mathcal{CN}(p_j|\mu_j^p, \phi_j^p). \quad (50)$$

where $\mu_i^z = \frac{v_i^z + \alpha x_i \tau_i^z}{1 + \alpha \tau_i^z}$ and $\phi_i^z = \frac{\tau_i^z}{1 + \alpha \tau_i^z}$ are the mean and variance of $p(z_i|\mathbf{x}, v_i^z, \tau_i^z, \boldsymbol{\eta})$, while $\mu_j^p = \frac{v_j^p \beta_{j_2, j_2}}{\beta_{j_2, j_2} + \gamma_{j_1} \tau_j^p}$ and $\phi_j^p = \frac{\tau_j^p \beta_{j_2, j_2}}{\beta_{j_2, j_2} + \gamma_{j_1} \tau_j^p}$ are the mean and variance of $p(p_j|\mathbf{x}, v_j^p, \tau_j^p, \boldsymbol{\eta})$. Please refer to Appendix E for details of the derivation. Then, it is required to determine two scalar functions denoted as $g_{in}(\cdot)$ and $g_{out}(\cdot)$, where $g_{in}(\cdot)$ is equal to the posterior mean μ_j^p of p_j , i.e.,

$$g_{in}(v_j^p, \tau_j^p, \boldsymbol{\eta}) = \mu_j^p = \frac{v_j^p \beta_{j_2, j_2}}{\beta_{j_2, j_2} + \gamma_{j_1} \tau_j^p}. \quad (51)$$

The corresponding posterior variance is

$$\tau_j^p \frac{\partial}{\partial v_j^p} g_{in}(v_j^p, \tau_j^p, \boldsymbol{\eta}) = \phi_j^p = \frac{\tau_j^p \beta_{j_2, j_2}}{\beta_{j_2, j_2} + \gamma_{j_1} \tau_j^p}. \quad (52)$$

$g_{out}(\cdot)$ satisfies

$$g_{out}(v_i^z, \tau_i^z, \boldsymbol{\eta}) = \frac{1}{\tau_i^z} (\mu_i^z - v_i^z) = \frac{1}{\tau_i^z} \left(\frac{v_i^z + \alpha x_i \tau_i^z}{1 + \alpha \tau_i^z} - v_i^z \right). \quad (53)$$

The corresponding posterior variance is

$$\tau_i^z \frac{\partial}{\partial v_i^z} g_{out}(v_i^z, \tau_i^z, \boldsymbol{\eta}) = \frac{-\alpha \tau_i^z}{1 + \alpha \tau_i^z}. \quad (54)$$

So far, the derivation of GAMP is completed. Note that the variance of (52) becomes a vector (equivalent to a diagonal matrix), while $\boldsymbol{\Sigma}_p$ of BSBL (or BSBL-APPR) is still a normal matrix. From this perspective, BSBL-GAMP is generated by the most thorough approximation, resulting in the least complexity. As shown in Algorithm 2, the GAMP algorithm is summarized.

Intuitively, only $\boldsymbol{\mu}_p$ and diagonalized $\boldsymbol{\Sigma}_p$ could be updated in the GAMP algorithm. For hyperparameter α , its update process will still lead to relatively large complexity if computed by (36). Here, α is rewritten as

$$\langle \alpha \rangle = \frac{c + NL/2}{d + \frac{1}{2} \|\mathbf{x} - \boldsymbol{\Phi} \boldsymbol{\mu}_p\|_2^2 + \sum_{i=1}^M (\boldsymbol{\Sigma}_p)_i (\boldsymbol{\Phi}^H \boldsymbol{\Phi})_i}. \quad (55)$$

Overall, the second faster version, called BSBL-GAMP, is eventually yielded by embedding the GAMP algorithm into BSBL. Specific steps are summarized in Algorithm 3.

Algorithm 2. The proposed GAMP algorithm.**Initialization**

- (i) set $\hat{s}_i = 0, k = 0$, input γ and α .
- (ii) use (22) to compute $\mu_{\mathbf{p}} = \{\mu_j^p\}_{j=1}^{NL}$.
- (iii) use (23) to compute $\text{diag}(\Sigma_{\mathbf{p}}) = \{\phi_j^p\}_{j=1}^{NL}$.
- (iv) preset error tolerance ε .

Repetition

$\forall i = 1, \dots, NL, j = 1, \dots, ML, j_1 = 1, \dots, M, j_2 = 1, \dots, L$.

- (i) $\hat{z}_i = \sum_j \Phi_{i,j} \mu_j^p(k)$.
- (ii) $\tau_i^z = \sum_j \Phi_{i,j}^2 \phi_j^p(k)$.
- (iii) $v_i^z = \hat{z}_i - \tau_i^z \hat{s}_i$.
- (iv) $\hat{s}_i = [(v_i^z + \alpha x_i \tau_i^z) / (1 + \alpha \tau_i^z) - v_i^z] / \tau_i^z$.
- (v) $\tau_i^s = \alpha / (1 + \alpha \tau_i^z)$.
- (vi) $\tau_j^p = (\sum_i \Phi_{i,j}^2 \tau_i^s)^{-1}$.
- (vii) $v_j^p = \mu_j^p(k) + \tau_j^p (\sum_i \Phi_{i,j} \hat{s}_i)$.
- (viii) $\mu_j^p(k+1) = v_j^p \beta_{j_2, j_2} / (\beta_{j_2, j_2} + \gamma_{j_1} \tau_j^p)$.
- (ix) $\phi_j^p(k+1) = \tau_j^p \beta_{j_2, j_2} / (\beta_{j_2, j_2} + \gamma_{j_1} \tau_j^p)$.
- (x) $k = k + 1$.

Terminate $\|\mu_j^p(k+1) - \mu_j^p(k)\|_2 \leq \varepsilon$.

Output The final $\mu_{\mathbf{p}}$ and $\Sigma_{\mathbf{p}}$.

Algorithm 3. The proposed GAMP-BSBL algorithm.**Initialization.**

- (i) set the first iterative number $k = 0$, error tolerance ε , maximal iterative number k_{\max} .
- (ii) set $\mathbf{p}^{(0)} = \|\bar{\mathbf{A}}^H \mathbf{X}\|_2$.
- (iii) assign a, b, c and d very small values.
- (iv) compute γ and α with (31) and (55), respectively.

Repetition.

- (i) compute $\mu_{\mathbf{p}}$ and $\Sigma_{\mathbf{p}}$ according to the above GAMP.
- (ii) compute γ and α according to (31) and (55), respectively.
- (iii) update $k = k + 1$.
- (iv) regard γ as $\mathbf{p}^{(k)}$.

Terminate $\|\mathbf{p}^{(k+1)} - \mathbf{p}^{(k)}\|_2 / \|\mathbf{p}^{(k)}\|_2 \leq \varepsilon$ or $k = k_{\max}$

Refinement

- (i) use the final \mathbf{p} to obtain θ^{est1} .
- (ii) use the final $\mu_{\mathbf{p}}$ to solve λ according to (41).
- (iii) obtain θ^{est2} according to (42).

Output the final DOA values.

Theoretically, BSBL-GAMP contains only simple multiplication of vectors and linear operations, so it is undoubtedly the fastest algorithm. To be specific, its complexity dominated by (55) is $O(ML)$, much less than $O(M^2NL^3)$ of BSBL or $O(M^2N)$ of BSBL-APPR. For comparison, the computational complexity of all the narrowband algorithms involved in this paper is summarized in Table 2. Generally, $M \gg N, L$ holds; thus, BSBL-GAMP obviously has the least computational complexity. In contrast, BSBL-APPR seems to be also satisfactory since its complexity is smaller than others, except for IC-SPICE and RVM-DOA. Moreover, the complexity of BSBL is moderate.

Table 2. Complexity of various algorithms.

BSBL	BSBL-APPR	BSBL-GAMP	SBL [21]
$O(M^2NL^3)$	$O(M^2N)$	$O(ML)$	$O(M^2N^2)$
L1-SRACV	IC-SPICE	L1-SVD	HSL
$O(M^3N^3)$	$O(MN^3)$	$O(M^3)$	$O(2M^2N)$
RVM-DOA	RV-ON-SBL	ON-SBLRVM	SBLMC
$O(MNL)$	$O(M^3)$	$O(M^2NL)$	$O(M^3L)$

4.4. BSBL for Wideband Sources

Although BSBL, BSBL-APPR, and BSBL-GAMP are developed, they are only applicable in the case of narrowband sources. For wideband cases, we must extend BSBL further.

A way to deal with wideband sources is to separate the wideband spectrum into independent narrowband ones. Without loss of generality, (3) can be rewritten as the special case at the j -th frequency point $f_j, \forall j = 1, \dots, J$, i.e.,

$$\mathbf{x}_j = \Phi_j \mathbf{p}_j + \mathbf{n}_j. \quad (56)$$

In this model, it is worth emphasizing that we only care about the locations of non-zero elements in \mathbf{p}_j rather than the concrete values because different \mathbf{p}_j theoretically indicate the same location of some source. Consequently, considering all the frequency points, different \mathbf{p}_j can be unified as \mathbf{q} so that

$$\mathbf{y} = \Psi \mathbf{q} + \bar{\mathbf{n}}. \quad (57)$$

where $\mathbf{y} = [\mathbf{x}_1^T, \dots, \mathbf{x}_J^T]^T \in \mathbb{C}^{NLJ \times 1}$, $\Psi = [\Phi_1^T, \dots, \Phi_J^T]^T \in \mathbb{C}^{NLJ \times ML}$, and $\bar{\mathbf{n}} = [\mathbf{n}_1^T, \dots, \mathbf{n}_J^T]^T \in \mathbb{C}^{NLJ \times 1}$. Similar to the aforementioned derivation of BSBL, (58)–(65) are yielded as follows.

$$\boldsymbol{\mu}_{\mathbf{q}} = \langle \Sigma_w \rangle \Psi^H (\langle \alpha_w \rangle^{-1} \mathbf{I}_{NLJ} + \Psi \langle \Sigma_w \rangle \Psi^H)^{-1} \mathbf{y}. \quad (58)$$

$$\Sigma_{\mathbf{q}} = \langle \Sigma_w \rangle - \langle \Sigma_w \rangle \Psi^H (\langle \alpha_w \rangle^{-1} \mathbf{I}_{NLJ} + \Psi \langle \Sigma_w \rangle \Psi^H)^{-1} \Psi \langle \Sigma_w \rangle. \quad (59)$$

$$\bar{a}_w = a_w + \frac{1}{2}. \quad (60)$$

$$\bar{b}_{w,m} = b_w + \frac{1}{2} \|(\mathbf{B}^{-1} + \mathbf{1}_{L \times L} - \mathbf{I}_L) \odot (\boldsymbol{\mu}_{\mathbf{q}_m} \boldsymbol{\mu}_{\mathbf{q}_m}^H)\|_{1,1} + \frac{1}{2} \text{diag}(\mathbf{B}^{-1} \odot \Sigma_{\mathbf{q}_m}). \quad (61)$$

$$\langle \gamma_m^w \rangle = \frac{\bar{a}_w}{\bar{b}_{w,m}}. \quad (62)$$

$$\bar{c}_w = c_w + \frac{NLJ}{2}. \quad (63)$$

$$\bar{d}_w = d_w + \frac{1}{2} (\|\mathbf{y} - \Psi \boldsymbol{\mu}_{\mathbf{q}}\|_2^2 + \text{tr}(\Sigma_{\mathbf{q}} \Psi^H \Psi)). \quad (64)$$

$$\langle \alpha_w \rangle = \frac{\bar{c}_w}{\bar{d}_w}. \quad (65)$$

To be distinguished from BSBL in narrowband cases, the different parameters in wideband cases are with superscript or subscript w and \mathbf{q} . In particular, $\Sigma_w = \text{diag}((\gamma^w)^{-1}) \otimes \mathbf{B}$ is the variance of variable \mathbf{q} , where $\gamma^w = [\gamma_1^w, \dots, \gamma_M^w]^T$. In wideband cases, the off-grid correction is the same as BSBL, except for (41), which is replaced by

$$\lambda_w \odot \boldsymbol{\mu}_{\mathbf{q}} = [(\Psi')^H \Psi']^{-1} (\Psi')^H (\mathbf{y} - \Psi \boldsymbol{\mu}_{\mathbf{q}}). \quad (66)$$

So far, BSBL for wideband sources has been completed. The specific process is summarized in Algorithm 4.

Algorithm 4. BSBL for wideband sources.

Initialization.

- (i) set $k = 0$, $\mathbf{q}^{(0)} = \left\| \sum_{j=1}^J \overline{\mathbf{A}}_j^H \mathbf{X}_j \right\|_2 / J$.
- (ii) set a_w, b_w, c_w and d_w very small values.
- (iii) preset error tolerance ε and maximal iterative number k_{\max} .

Repetition

- while** ($\|\mathbf{q}^{(k+1)} - \mathbf{q}^{(k)}\|_2 / \|\mathbf{q}^{(k)}\|_2 > \varepsilon$ or $k < k_{\max}$) **do**:
- (i) compute $\boldsymbol{\mu}_{\mathbf{q}}$ and $\boldsymbol{\Sigma}_{\mathbf{q}}$ according to (58) and (59).
 - (ii) compute $\langle \boldsymbol{\Sigma}_w \rangle$ and $\langle \boldsymbol{\alpha}_w \rangle$ according to (62) and (65).
 - (iii) regard $\boldsymbol{\gamma}^w$ as $\mathbf{q}^{(k)}$.
 - (iv) $k = k + 1$.
- end while**

Refinement

- (i) Use the final \mathbf{q} to obtain $\boldsymbol{\theta}^{est1}$.
- (ii) Use the final $\boldsymbol{\mu}_{\mathbf{q}}$ to solve λ_w according to (66).
- (iii) Obtain $\boldsymbol{\theta}^{est2}$ according to (42).

Output The final DOA values.

5. Numerical Simulation

In this section, the superiority of our proposed algorithms will be proven comprehensively through three subsections, including extensive simulations. For simplicity, the proposed BSBL, BSBL-APPR, and BSBL-GAMP are collectively referred to as BSBLs. In the first and second subsections, the narrowband and wideband estimation performance of various algorithms is evaluated comprehensively. In the third subsection, the in-depth analysis of Bayesian performance is completed by comparison with other off-grid SBL-based methods.

5.1. Estimation Performance for Narrowband Sources

In this subsection, estimation performance is evaluated by the Root-Mean-Square Error (RMSE) expressed as

$$RMSE = \sqrt{\frac{1}{M_c K} \sum_{m_c=1}^{M_c} \sum_{k=1}^K (\hat{\theta}_{m_c, k} - \theta_k)^2}. \quad (67)$$

where M_c is the Monte Carlo number and K is the number of sources. $\hat{\theta}_{m_c, k}$ is the estimation value for the k -th source in the m_c -th trial and θ_k is the true angle of the k -th source. For clarity, we introduce four canonical SMV algorithms, i.e., l_1 -SVD [20], l_1 -SRACV [21], IC-SPICE [22], SBL [31], SS-ANM [25], and StrucCovMLE [26], as comparisons in following simulations. Before that, unless otherwise stated, baseline simulation conditions are $SNR = 20$ dB, $K = 3$ temporally correlated sources with a random DOA set $\{-20.53^\circ, 10.10^\circ, 43.01^\circ\}$, the number of sensors is $N = 8$, the number of snapshots is $L = 4$, the grid interval is 1° , the number of grids is $M = 180$, the Monte Carlo number is $M_c = 200$, and the temporal correlation coefficient is $\beta = 0$.

Remark 3. $\beta = 0$ and $SNR = 20$ dB (or at least one) are provided to ensure that BSBL-APPR performs normally.

Simulation 1 tests the ability of various algorithms to suppress temporal correlation. Specifically, the amplitudes and phases of temporally correlated coefficients uniformly vary from $[0, 1]$ and $[0, 2\pi]$, respectively. The results are shown in Figure 2. Obviously, the

RMSEs of BSBLs persistently remain low and fluctuate only slightly with the correlation coefficient varying, while others just struggle when the correlation coefficient is large. Particularly, IC-SPICE is able to impair the influence of temporal correlation to some extent, but it is still at a loss if the correlated level is high. Overall, the simulation results fully live up to our expectations that BSBLs are able to suppress temporal correlation effectively, which confirms that the considered temporal correlation modeled in the block-sparse model indeed plays an important role in improving the robustness of temporal correlation.

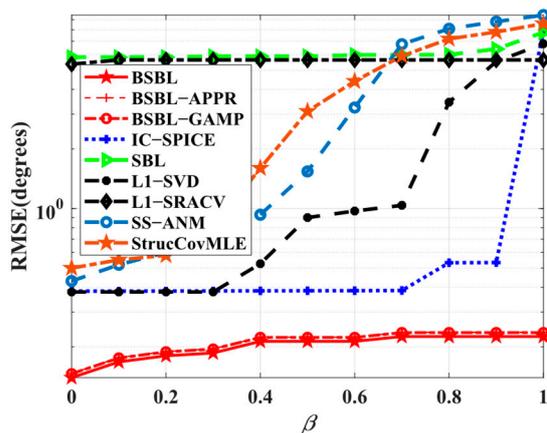


Figure 2. RMSE versus correlation coefficient.

Simulation 2 examines the dependence on the number of snapshots. In Figure 3, all the sparsity-based algorithms are collectively robust to various snapshots. In other words, all the algorithms seem to achieve SSR with only a few snapshots. Despite this, BSBLs are still commendable due to the realized lowest RMSEs. In fact, SBL itself enables finding global minima and smoothing out numerous local minima in some cases with a few snapshots [48]. For SBL-based methods, BSBLs undoubtedly inherit their advantages. Additionally, the used hierarchical priors can improve sparsity performance so that BSBLs perform best. Thus, BSBLs enable high estimated precision with a few snapshots.

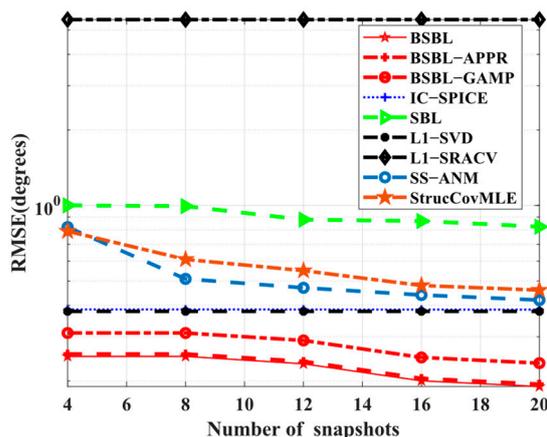


Figure 3. RMSE versus number of snapshots.

Simulation 3 focuses on the RMSE performance with respect to SNRs. Apparently, in Figure 4, all the algorithms work well if high SNRs are adopted, while only BSBLs maintain fewer RMSEs at low SNRs. The results can be explained by the fact that the sparsity-based algorithms rely on high SNRs to some extent, but SBL can reduce the dependency. Taking IC-SPICE as an example, it can achieve efficient iterative optimization under the condition of high SNRs but cannot seek the right global minima or even trap some fixed local optima. The original reason for this is that the used covariance matrix and the updated

parameters have large errors with ideal ones. However, SBL seems to work normally owing to convergence guarantee and gradual optima under the condition of existing data errors. Surprisingly, the proposed BSBLs have a similar ability in some way. On the whole, BSBLs are preferable, especially under the condition of low SNRs.

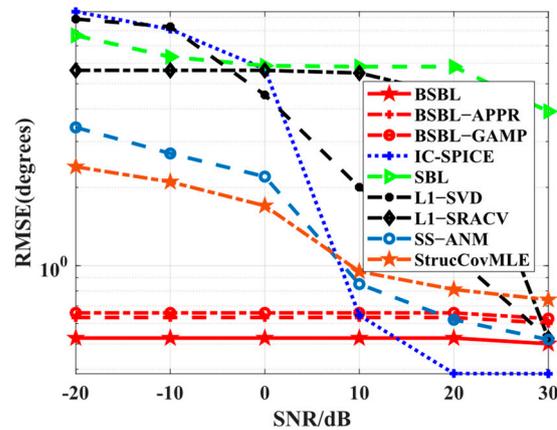


Figure 4. RMSE versus SNR.

Simulation 4 investigates the RMSE performance with respect to the number of sensors. Here, this simulation is executed at the number of snapshots $L = 20$ rather than $L = 4$ since l_1 -SVD cannot work normally when the number of sensors exceeds the number of snapshots. Intuitively, in Figure 5, the RMSE performance of BSBLs is still excellent, although BSBLs are inferior to IC-SPICE at $N = 4$. In fact, the results are related to the ability to solve underdetermined DOA estimation problems. SBL still seems to find the sparsest solutions, although the restricted isometry property (RIP) is not satisfied [48]. When the number of sensors is not large enough, i.e., the solution to be solved is not sparse enough, SBL still tries its best to realize global minima. Thus, BSBLs can handle underdetermined cases efficiently. In other words, BSBLs are able to realize SSR effectively on the condition of a few sensors.

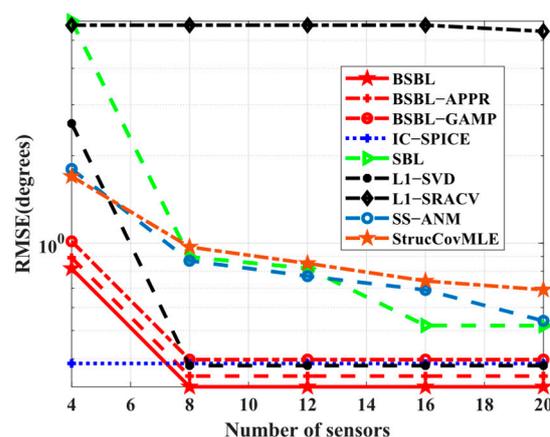


Figure 5. RMSE versus number of sensors at $L = 20$.

Simulation 5 tests the adaptability to wide grid intervals. In Figure 6, the results illustrate that both BSBLs and IC-SPICE gain highly accurate estimation values at refined grids, but only BSBLs reluctantly adapt to wide grid intervals, although all the algorithms suffer hardship at coarse grids. The coarse grids compel relatively large errors of the pre-estimated values, so the refined values have a larger bias. In the proposed BSBLs, the grid refinement has an effect on reducing the bias at each iteration. Therefore, BSBLs adapt to coarse grids to some extent.

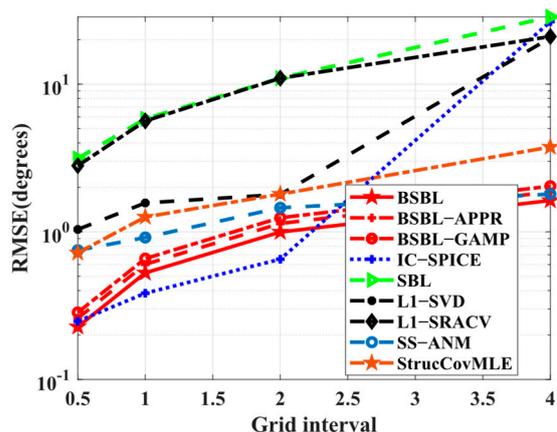


Figure 6. RMSE versus grid interval.

Simulation 6 examines the RMSE performance with respect to the number of sources. The different conditions are as follows: the DOA sets of sources are selected from $[-90^\circ, 90^\circ]$ randomly, and $L = 10$ is chosen to ensure the normal operation of l_1 -SVD. In Figure 7, all the RMSEs drop rapidly, but BSBLs seem to slow down the pace of performance degradation, which implies that BSBLs have the potential to locate more sources. In fact, the results are another proof of the excellent underdetermined DOA estimation ability of BSBLs in Simulation 4. More sources and fewer sensors play similar roles in decreasing the sparse level in sparse recovery theory, and BSBLs handle this case efficiently.

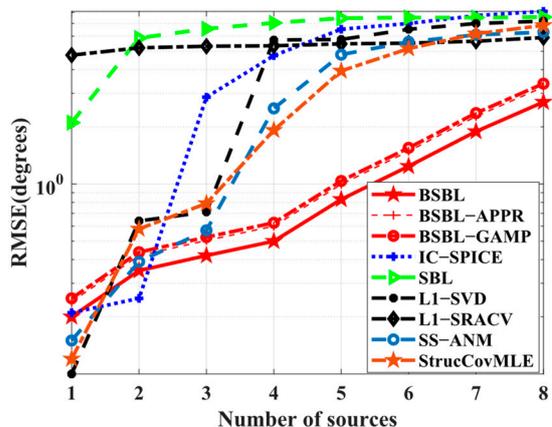


Figure 7. RMSE versus number of sources.

5.2. Estimation Performance for Wideband Sources

For comparison, JLZA-DOA [23], W-SpSF [24], W-SBL [36], GS-WSpSF [27], and ANM [28] are introduced. On behalf of BSBLs, only BSBL is adopted in the last two subsections for simplicity.

Simulation 7 tests the spectral performance of the above algorithms. Its conditions are two uncorrelated chirps with angles of -10° and 25° with a center frequency of 1 kHz and a bandwidth of 400 Hz from 0.8 kHz to 1.2 kHz, $SNR = 20\text{dB}$, the number of sensors is $N = 8$, the number of snapshots is $L = 4$, and the grid interval is 1° . Intuitively, in Figure 8, JLZA-DOA and W-SpSF arise explicit sidelobes around their spikes, while GS-WSpSF, ANM, W-SBL, and BSBL are excellent since their spectra are almost without sidelobes and the spikes are sharp. It is worth noting that the spikes of W-SBL seem to be very low at some frequency points. Specifically, W-SBL fluctuates intensely with varied frequencies, i.e., the signal energy non-uniformly leaks between different frequencies. GS-WSpSF and ANM are better; at least their spikes are visible and apparent over the range of all the frequency points. The proposed BSBL has the highest spikes and varies little over

the range of frequencies. Thus, three conclusions are drawn. (1) When sparse recovery is adopted, BSBL can ensure that equal energy is assigned between different frequency bins. (2) BSBL realizes the highest spikes and shows the best convergence effect, i.e., ensuring global minima.

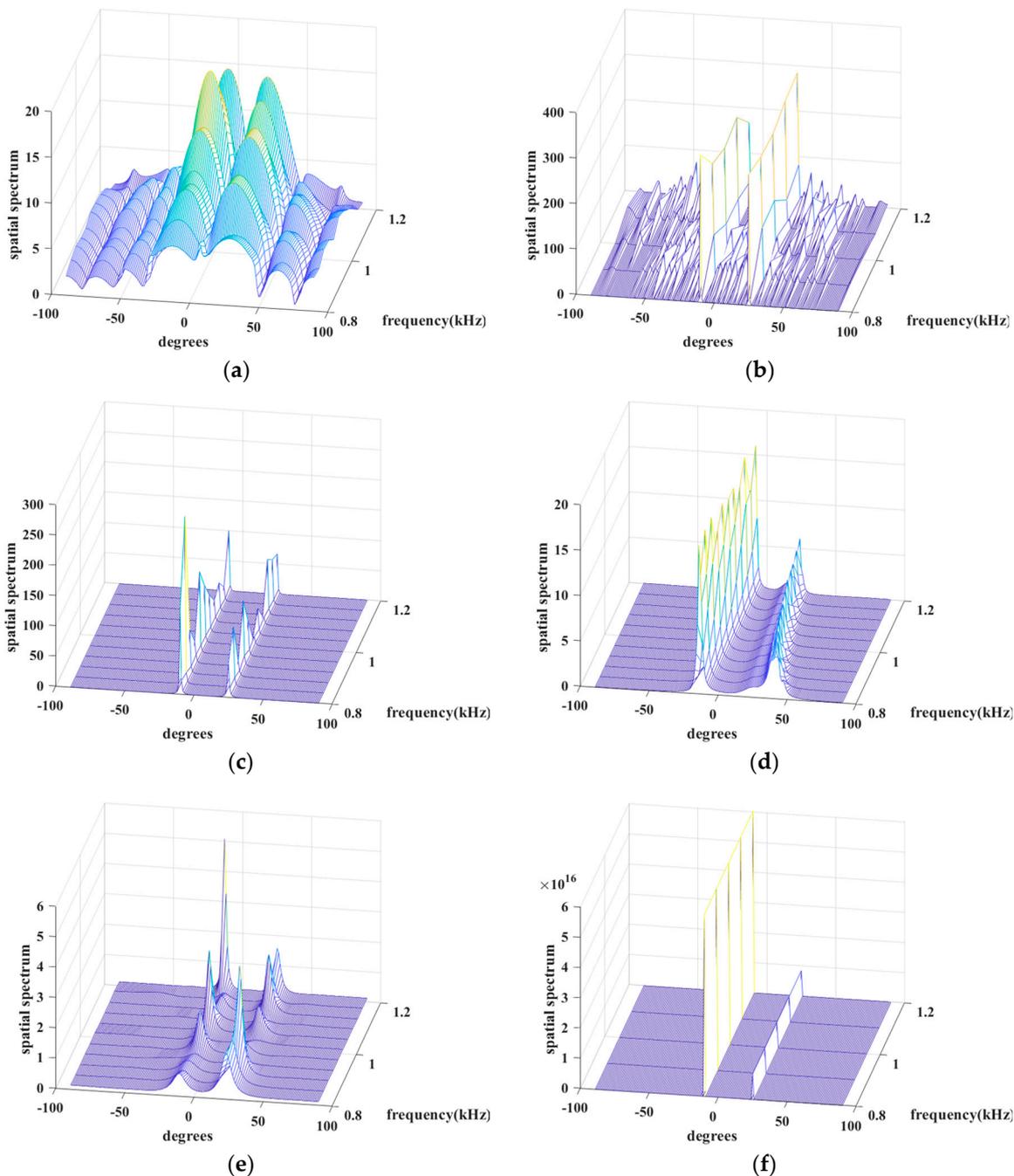


Figure 8. Spatial spectrum versus degree and frequency of (a) JLZA–DOA, (b) W–SpSF, (c) W–SBL, (d) GS–WSpSF, (e) ANM, (f) BSBL.

Simulation 8 examines the *RMSE* performance with respect to the number of sensors. The condition set is the same as baseline conditions, except for wideband sources. As expected, BSBL achieves excellent *RMSE* performance in Figure 9. Over the whole range of the number of snapshots, BSBL shows overwhelming advantages and outperforms others patently. In fact, DOA estimation for wideband sources is difficult, and one of the main reasons is many algorithms fail to realize accuracy estimation over the whole

range of frequency bins. Simulation 7 shows the special ability of BSBL to overcome this problem, and Simulation 8 seems to verify it again. On the one hand, BSBL for wideband sources maintains the superiority of narrowband BSBL, so it can obtain excellent estimation performance. On the other hand, BSBL extends the advantages of SBL to wideband sources, i.e., achieving robust sparse recovery with only a few snapshots for all the sources over the whole frequency band.

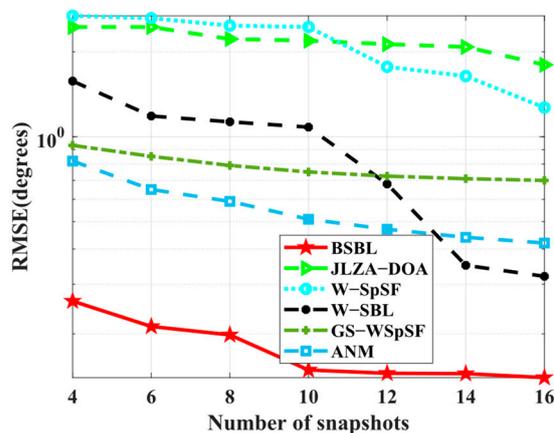


Figure 9. RMSE versus number of snapshots.

Simulation 9 tests the RMSE performance with respect to SNRs. In Figure 10, BSBL outperforms others and improves well when SNRs increase. It is worth noting that the results are different from the narrowband ones in Simulation 2. To be specific, BSBL seems to fluctuate intensely with the varied SNRs. BSBL for wideband sources cannot work well enough compared to narrowband source cases. We carefully analyze the reasons and find that BSBL cannot realize ensuring global minima at each frequency bin for sparse recovery. Despite this, the defects cannot obscure the virtues. BSBL still achieves impressive performance for wideband sources.

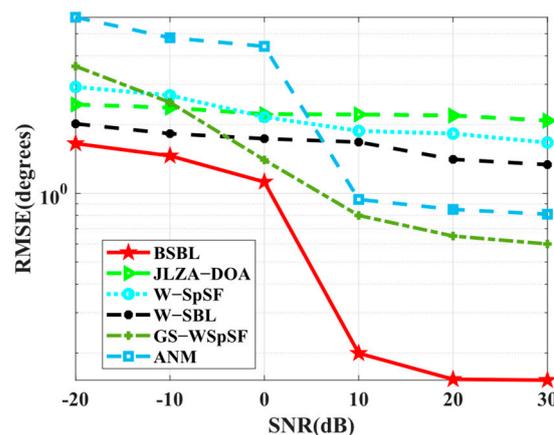


Figure 10. RMSE versus SNR.

5.3. Analysis of Sparse Bayesian Performance

According to the common perception of SBL, the elaborate Bayesian framework with substantial priors is regarded as a characteristic to enhance sparsity well because priors play a role of regularization in sparse recovery [55,59]. Here, we abstract several Bayesian frameworks from several off-grid SBL-based algorithms, such as RVM-DOA [37], RV-ON-SBL [47], ON-SBLRVM [43], SBLMC [39], and HSL [46], for comparison and analysis. As shown in Figure 11, RVM-DOA, RV-ON-SBL, and ON-SBLRVM are only imposed on Gaussian priors that have been proven of poor Bayesian performance.

The complicated Bayesian frameworks of HSL and BSBL are the same, except for the priors assigned to the unknown variables; thus, they may perform equally well. SBLMC has the most elaborate Bayesian framework composed of sufficient priors, so its sparsity performance will be perfect theoretically.

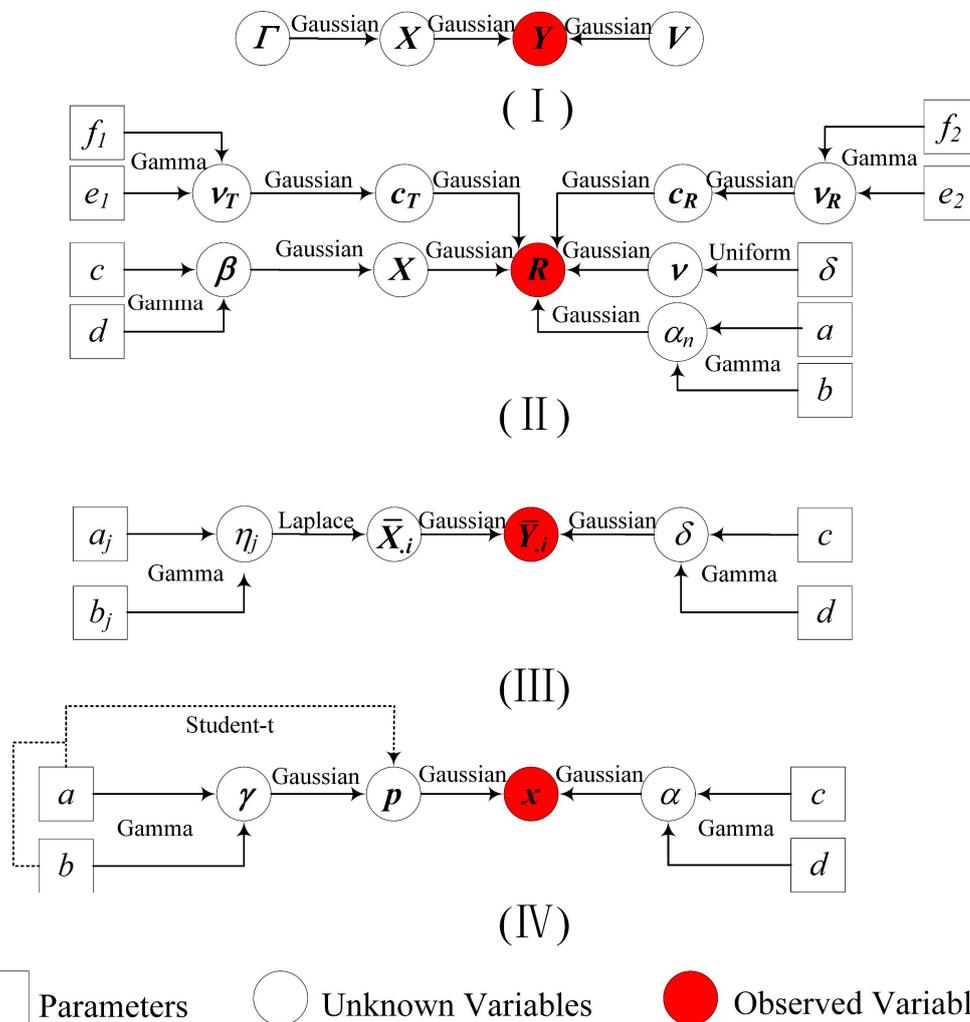


Figure 11. Directed acyclic graph representing the Bayesian frameworks of (I) RVM-DOA, RV-ON-SBL, ON-SBLRVM, (II) SBLMC, (III) HSL, (IV) BSBL.

To confirm the above conjectures, two more simulations are performed to test the estimation performance of these algorithms.

Simulation 10 tests the *RMSE* performance with respect to the number of snapshots. The conditions are the same as the baseline conditions. In Figure 12, the *RMSE* performance of RVM-DOA, RV-ON-SBL, and ON-SBLRVM are expectedly worse than others, but SBLMC seems to not meet our expectations, while BSBL and HSL achieve preminent *RMSE* performance. BSBL and HSL with moderately elaborate Bayesian frameworks outperform others, including the SBLMC with the most elaborate one. The result seems to violate the rule of Bayesian learning, which will be explained in the following text.

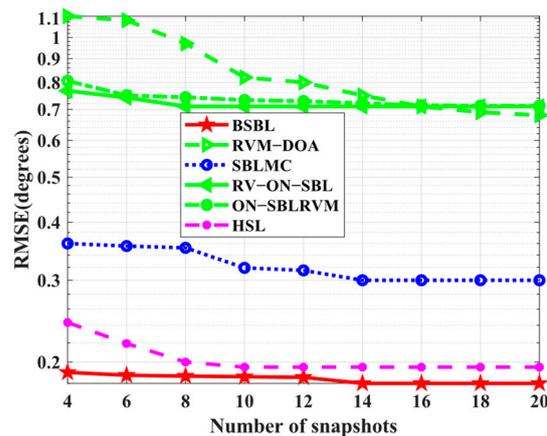


Figure 12. RMSE versus number of snapshots.

Simulation 11 tests the RMSE performance with respect to SNRs. In Figure 13, the proposed BSBL still seems to work best. It is worth noticing that the advantages of BSBL are not obvious, especially when SNRs are low. For all the SBL-based methods, BSBL has shown no more advantages than others on the condition of low SNRs because hierarchical priors improve sparsity if, and only if, SNRs are high. To be specific, Bayesian learning is able to find global minima, even at low SNRs, but the parameters yielded by hierarchical priors seem to update well only if SNRs are high.

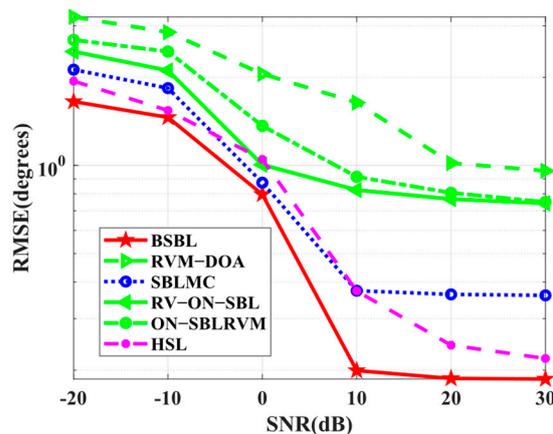


Figure 13. RMSE versus SNR.

Based on the simulation results, it can be seen that the complicated Bayesian frameworks, i.e., II, III, and IV in Figure 10, indeed achieve more excellent Bayesian performance than the canonical one, i.e., I in Figure 11. However, SBLMC with the most elaborate Bayesian framework has not met our initial expectation, which can be explained by the fact that (i) SBL belongs to machine learning, so the Bayesian framework with too many priors will yield massive iterative hyperparameters, leading to overfitting during the iterative process. (ii) SBLMC is developed in the presence of mutual coupling; thus, the involved additional hyperparameter iteration is bound to affect the key parameters related to DOA estimation.

It is worth emphasizing that BSBL achieves slightly better estimation performance than HSL. The result indicates that the indirectly induced Student's *t* priors, generated by Gaussian and Gamma priors, indeed express excellent sparsity performance. In fact, Student's *t* priors have preferable sparsity-inducing performance, which has been mentioned in [55,56].

Overall, the above three subsection simulation results sufficiently demonstrate the superiority of BSBLs. Understandably, BSBL leverages hierarchical Gaussian and Gamma

priors and uses VBI to complete the Bayesian inference so as to construct the corresponding iterative algorithm. Theoretically, the superiority is guaranteed by (1) indirect Student's t-distributions, which have an excellent sparsity-inducing ability [56], and (2) variational approximation for Bayesian inference shows better performance than maximum posterior (MAP) estimation adopted in many SBL-based methods [60]. In addition, two approximation operations have achieved impressive running efficiency beyond many state-of-the-art methods. Moreover, BSBL still performs well in wideband cases and outperforms other algorithms in smoothing the spectrum peaks and super resolutions. Last but not least, BSBL has suppressed temporal correlation efficiently owing to its tactful algorithm design.

6. Conclusions

In this paper, we develop a DOA estimator (i.e., BSBL) based on sparse Bayesian learning with hierarchical priors. Due to the unacceptable computational complexity caused by the vectorization of the MMV model, two approximation operations are creatively introduced, thereby yielding two faster versions of BSBL, i.e., BSBL-APPR and BSBL-GAMP. As expected, all the proposed BSBLs (including BSBL, BSBL-APPR, BSBL-GAMP) achieve excellent estimation performance. For narrowband source estimation, BSBLs show perfect sparsity performance owing to the designed hierarchical priors. Further, BSBLs inherit and even extend the advantages of SBL, such as sparse signal recovery guarantee, less dependency on numerous snapshots or high SNRs, and the ability to handle underdetermined DOA estimation. Moreover, BSBLs enable robustness to temporally correlated sources and adaptability to coarse grids, which owes to the considered temporal correlation and the used grid refinement. For wideband source estimation, BSBL almost maintains huge advantages, i.e., realizing highly accurate estimation among the whole frequency band, while others suffer performance reduction to varying degrees. However, in wideband cases, BSBL cannot retain the good performance as in narrowband cases if low SNRs are adopted, which is our goal to solve in the next study. For Bayesian performance, BSBL with a moderately elaborate Bayesian framework realizes the best estimation performance. Furthermore, BSBL can balance both sparsity and complexity. Specifically, BSBL achieves sharp spectrum spikes and avoids overfitting produced by too many parameters.

Overall, the proposed BSBLs tactfully combine the hierarchical priors and the block-sparse model that contribute much to complexity reduction, which is never achieved by other SBL-based methods. Moreover, BSBLs retain and extend the advantages of SBL. Most importantly, BSBL is more practical and applicable when sources are temporally correlated or wideband. Despite this, BSBL seems not to be perfect because its performance suffers a loss at low SNRs in wideband cases to some extent. Anyway, the proposed BSBLs are worth recommendation and praise.

Author Contributions: Conceptualization, N.L. and X.Z.; methodology, N.L.; software, F.L.; validation, X.Z.; writing—original draft preparation, N.L.; writing—review and editing, N.L.; visualization, F.L.; supervision, B.Z.; project administration, X.Z.; funding acquisition, X.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by [the Natural Science Basic Research Program of Shaanxi] grant number [2023-JC-YB-488].

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: No new data were created.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

$$\begin{aligned}
 (23) &= \left\langle -\frac{(\mathbf{x}-\Phi\mathbf{p})^H(\mathbf{x}-\Phi\mathbf{p})}{2}\alpha - \frac{\mathbf{p}^H\boldsymbol{\Sigma}^{-1}\mathbf{p}}{2} \right\rangle + const \\
 &= \left\langle -\frac{1}{2}\mathbf{p}^H(\alpha\Phi^H\Phi + \boldsymbol{\Sigma}^{-1})\mathbf{p} + \alpha\mathbf{p}^H\Phi^H\mathbf{x} \right\rangle + const \\
 &= -\frac{1}{2}\mathbf{p}^H(\langle\alpha\rangle\Phi^H\Phi + \langle\boldsymbol{\Sigma}^{-1}\rangle)\mathbf{p} + \langle\alpha\rangle\mathbf{p}^H\Phi^H\mathbf{x} + const.
 \end{aligned} \tag{A1}$$

The terms unrelated to \mathbf{p} are absorbed into the constant. Obviously, \mathbf{p} obeys a Gaussian distribution with mean $\boldsymbol{\mu}_p$ and variance $\boldsymbol{\Sigma}_p$, where $\boldsymbol{\Sigma}_p^{-1} = \langle\alpha\rangle\Phi^H\Phi + \langle\boldsymbol{\Sigma}^{-1}\rangle$ and $\boldsymbol{\Sigma}_p^{-1}\boldsymbol{\mu}_p = \langle\alpha\rangle\Phi^H\mathbf{x}$. Using linear transformation, (24) and (25) are yielded.

Appendix B

$$\begin{aligned}
 (28) &= \left\langle \ln|\boldsymbol{\Sigma}|^{-\frac{1}{2}} - \frac{\mathbf{p}^H\boldsymbol{\Sigma}^{-1}\mathbf{p}}{2} + \sum_{m=1}^M (\ln\gamma_m^{a-1} - b\gamma_m) \right\rangle + const \\
 &= \left\langle \ln|\text{diag}(\boldsymbol{\gamma})|^{\frac{1}{2}} + \ln|\mathbf{B}|^{-\frac{M}{2}} - \frac{\mathbf{p}^H(\text{diag}(\boldsymbol{\gamma})\otimes\mathbf{B}^{-1})\mathbf{p}}{2} \right\rangle + \sum_{m=1}^M [(a-1)\ln\gamma_m - b\gamma_m] + const \\
 &= \left\langle \sum_{m=1}^M [(\frac{1}{2} + a - 1)\ln\gamma_m - \frac{1}{2}\gamma_m\mathbf{p}_m^H\mathbf{B}^{-1}\mathbf{p}_m - b\gamma_m] \right\rangle + const \\
 &= \sum_{m=1}^M [(a + \frac{1}{2} - 1)\ln\gamma_m - (b + \frac{1}{2}\langle\mathbf{p}_m^H\mathbf{B}^{-1}\mathbf{p}_m\rangle)\gamma_m] + const.
 \end{aligned} \tag{A2}$$

The terms unrelated to $\boldsymbol{\gamma}$ are absorbed into the constant. Undoubtedly, γ_m obeys a Gamma distribution with shape parameter $a + \frac{1}{2}$ and scale parameter $b + \frac{1}{2}\langle\mathbf{p}_m^H\mathbf{B}^{-1}\mathbf{p}_m\rangle$.

Appendix C

Diagonal elements of \mathbf{B}^{-1} affect the second-order moment of \mathbf{p}_m , while the off-diagonal elements only affect the mean value. Therefore, two cases are each considered for the element $\beta_{i,j}^{-1}$ of \mathbf{B}^{-1} .

(i) $i = j$

$$\begin{aligned}
 \langle\mathbf{p}_m^H\mathbf{B}^{-1}\mathbf{p}_m\rangle &= \sum_{i=1}^L \beta_{i,i}^{-1} E(p_{m_i}^* p_{m_i}) = \sum_{i=1}^L \beta_{i,i}^{-1} [(\mu_{p_m})_i (\mu_{p_m})_i^* + (\boldsymbol{\Sigma}_{p_m})_{i,i}] \\
 &= \|(\text{diag}(\mathbf{B}_\Lambda^{-1}))^T [\boldsymbol{\mu}_{p_m} \odot \boldsymbol{\mu}_{p_m}^* + \text{diag}(\boldsymbol{\Sigma}_{p_m})]\|_{1,1} = \|\mathbf{B}_\Lambda^{-1} \odot [(\boldsymbol{\mu}_{p_m} \boldsymbol{\mu}_{p_m}^H) + \text{diag}(\text{diag}(\boldsymbol{\Sigma}_{p_m}))]\|_{1,1}.
 \end{aligned} \tag{A3}$$

where $\mathbf{B}_\Lambda^{-1} = \text{diag}(\text{diag}(\mathbf{B}^{-1}))$ and p_{m_i} and $(\mu_{p_m})_i$ are the i -th element of \mathbf{p}_m and $\boldsymbol{\mu}_{p_m}$, respectively. $(\boldsymbol{\Sigma}_{p_m})_{i,i}$ is the i -th diagonal element of $\boldsymbol{\Sigma}_{p_m}$.

(ii) $i \neq j$

$$\begin{aligned}
 \langle\mathbf{p}_m^H\mathbf{B}^{-1}\mathbf{p}_m\rangle &= \sum_{i=1}^L \sum_{j=1}^L \beta_{i,j}^{-1} E(p_{m_i}^* p_{m_j}) = \sum_{i=1}^L \sum_{j=1}^L \beta_{i,j}^{-1} (1 + \beta_{i,j}) E(p_{m_i}^*) E(p_{m_j}) \\
 &= \sum_{i=1}^L \sum_{j=1}^L (\beta_{i,j}^{-1} + 1) (\mu_{p_m})_i^* (\mu_{p_m})_j = \|(\mathbf{B}_\Lambda^{-1} + \mathbf{1}_{L \times L} - \mathbf{I}_L) \odot (\boldsymbol{\mu}_{p_m} \boldsymbol{\mu}_{p_m}^H)\|_{1,1}.
 \end{aligned} \tag{A4}$$

where $\mathbf{B}_\Lambda^{-1} = \mathbf{B}^{-1} - \mathbf{B}_\Lambda^{-1}$. Based on (i) and (ii), (32) is verified.

Appendix D

$$\begin{aligned}
 (33) &= \left\langle \ln|\alpha^{-1}\mathbf{I}_{NL}|^{-\frac{1}{2}} - \frac{(\mathbf{x}-\Phi\mathbf{p})^H(\mathbf{x}-\Phi\mathbf{p})}{2}\alpha + \ln\alpha^{c-1} - d\alpha \right\rangle + const \\
 &= \frac{NL}{2} \ln\alpha - \frac{\langle(\mathbf{x}-\Phi\mathbf{p})^H(\mathbf{x}-\Phi\mathbf{p})\rangle}{2}\alpha + (c-1)\ln\alpha - d\alpha + const \\
 &= (c + \frac{NL}{2} - 1)\ln\alpha - [d + \frac{\langle(\mathbf{x}-\Phi\mathbf{p})^H(\mathbf{x}-\Phi\mathbf{p})\rangle}{2}]\alpha + const.
 \end{aligned} \tag{A5}$$

The terms unrelated to α are absorbed into the constant. α obeys another Gamma distribution with shape parameter $c + \frac{NL}{2}$ and scale parameter $d + \frac{\langle (x - \Phi \mathbf{p})^H (x - \Phi \mathbf{p}) \rangle}{2}$.

Appendix E

To obtain the mean and variance of variables (i.e., z_i and p_j), their distributions are transformed into logarithmic form as follows:

$$\begin{aligned} \ln (49) &= \ln [N(x_i|z_i, \alpha^{-1})N(z_i|v_i^z, \tau_i^z)] + const = -\frac{\alpha}{2}(x_i - z_i)^2 - \frac{1}{2\tau_i^z}(z_i - v_i^z)^2 + const \\ &= -\frac{1}{2}\left(\alpha + \frac{1}{\tau_i^z}\right)z_i^2 + \left(\alpha x_i + \frac{v_i^z}{\tau_i^z}\right)z_i + const = -\frac{1}{2\phi_i^z}z_i^2 + \frac{\mu_i^z}{\phi_i^z}z_i + const. \end{aligned} \quad (A6)$$

The final equation of (A6) is yielded based on the expansion of the Gaussian function. Patently, $\frac{1}{\phi_i^z} = \alpha + \frac{1}{\tau_i^z}$ and $\frac{\mu_i^z}{\phi_i^z} = \alpha x_i + \frac{v_i^z}{\tau_i^z}$ can be directly abstracted and solved by simple linear transformation, so $\mu_i^z = \frac{v_i^z + \alpha x_i \tau_i^z}{1 + \alpha \tau_i^z}$ and $\phi_i^z = \frac{\tau_i^z}{1 + \alpha \tau_i^z}$ are obtained. For (50), the logarithmic form is

$$\begin{aligned} \ln (50) &= \ln [N(p_j|0, \gamma_{j1}^{-1} \beta_{j2,j2})N(p_j|v_j^p, \tau_j^p)] + const = -\frac{\gamma_{j1}}{2\beta_{j2,j2}}p_j^2 - \frac{1}{2\tau_j^p}(p_j - v_j^p)^2 + const \\ &= -\frac{1}{2}\left(\frac{\gamma_{j1}}{\beta_{j2,j2}} + \frac{1}{\tau_j^p}\right)p_j^2 + \frac{v_j^p}{\tau_j^p}p_j + const = -\frac{1}{2\phi_j^p}p_j^2 + \frac{\mu_j^p}{\phi_j^p}p_j + const. \end{aligned} \quad (A7)$$

$$\text{Similarly, } \mu_j^p = \frac{v_j^p \beta_{j2,j2}}{\beta_{j2,j2} + \gamma_{j1} \tau_j^p} \text{ and } \phi_j^p = \frac{\tau_j^p \beta_{j2,j2}}{\beta_{j2,j2} + \gamma_{j1} \tau_j^p} \text{ are yielded.}$$

References

- Liu, X.; Xu, B.; Zheng, K.; Zheng, H. Throughput maximization of wireless-powered communication network with mobile access points. *IEEE Trans. Wirel. Commun.* **2022**, *22*, 4401–4415. [\[CrossRef\]](#)
- Liu, X.; Lin, Z.; Zheng, K.; Yao, X.W.; Liu, J. Optimal Time Allocation for Backscatter-Aided Relay Cooperative Transmission in Wire-less-Powered Heterogeneous CRNs. *IEEE Internet Things* **2023**, *10*, 16209–16224. [\[CrossRef\]](#)
- Mizmizi, M.; Linsalata, F.; Brambilla, M.; Morandi, F.; Dong, K.; Magarini, M.; Nicoli, M.; Khormuji, M.N.; Wang, P.; Pitaval, R.A.; et al. Fastening the initial access in 5G NR sidelink for 6G V2X networks. *Veh. Commun.* **2022**, *33*, 100402. [\[CrossRef\]](#)
- Chen, Z.; Chen, P.; Guo, Z.; Zhang, Y.; Wang, X. A RIS-based vehicle DOA estimation method with integrated sensing and communication system. *IEEE Trans. Intell. Transp. Syst.* **2023**, *1*, 1–13. [\[CrossRef\]](#)
- Chen, P.; Yang, Z.; Chen, Z.; Guo, Z. Reconfigurable intelligent surface aided sparse DOA estimation method with non-ULA. *IEEE Signal Process. Lett.* **2021**, *28*, 2023–2027. [\[CrossRef\]](#)
- Chen, P.; Chen, Z.; Zheng, B.; Wang, X. Efficient DOA estimation method for reconfigurable intelligent surfaces aided UAV swarm. *IEEE Trans. Signal Process.* **2022**, *70*, 743–755. [\[CrossRef\]](#)
- Shen, Q.; Liu, W.; Cui, W.; Wu, S. Underdetermined DOA estimation under the compressive sensing framework: A review. *IEEE Access* **2016**, *4*, 8865–8878. [\[CrossRef\]](#)
- Baraniuk, R.G. Compressive sensing [lecture notes]. *IEEE Signal Proc. Mag.* **2007**, *24*, 118–121. [\[CrossRef\]](#)
- Niwaria, K.; Jain, M. A Literature Survey Different Algorithms of MUSIC and MVDR DOA Estimation. *IJEMR* **2018**, *8*, 56–69.
- Li, F.; Liu, H.; Vaccaro, R. Performance analysis for DOA estimation algorithms: Unification, simplification, and observations. *IEEE Trans. Aerosp. Electron. Syst.* **1993**, *29*, 1170–1184. [\[CrossRef\]](#)
- Gershman, A.B.; Rübsumen, M.; Pesavento, M. One- and two-dimensional direction-of-arrival estimation: An overview of search-free techniques. *Signal Process.* **2010**, *90*, 1338–1349. [\[CrossRef\]](#)
- Krishnaveni, V.; Kesavamurthy, T.; Aparna, B. Beamforming for direction-of-arrival (DOA) estimation—A survey. *Int. J. Comput. Appl.* **2013**, *61*, 4–11. [\[CrossRef\]](#)
- Krim, H.; Viberg, M. Two decades of array signal processing research: The parametric method. *IEEE Signal Proc. Mag.* **1996**, *13*, 67–94. [\[CrossRef\]](#)
- Yang, Z.; Li, J.; Stoica, P.; Xie, L. Sparse methods for direction-of-arrival estimation. In *Academic Press Library in Signal Processing*; Academic Press: Cambridge, MA, USA, 2018; pp. 509–581.
- Aboumahmoud, I.; Muqaibel, A.; Alhassoun, M.; Alawsh, S. A Review of sparse sensor arrays for two-dimensional direction-of-arrival estimation. *IEEE Access* **2021**, *9*, 92999–93017. [\[CrossRef\]](#)
- Cohen, R.; Eldar, Y.C. Sparse array design via fractal geometries. *IEEE Trans. Signal Process.* **2020**, *68*, 4797–4812. [\[CrossRef\]](#)
- Patwari, A. Sparse linear antenna arrays: A review. In *Antenna Systems*; Intech Open: London, UK, 2021.
- Yang, Z.; Xie, L. On gridless sparse methods for multi-snapshot DOA estimation. In Proceedings of the 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Shanghai, China, 20–25 March 2016; pp. 3236–3240.

19. Chung, P.J.; Viberg, M.; Yu, J. DOA estimation methods and algorithms. In *Academic Press Library in Signal Processing*; Elsevier: Amsterdam, The Netherlands, 2014; Volume 3, pp. 599–650.
20. Malioutov, D.; Cetin, M.; Willsky, A. A sparse signal reconstruction perspective for source localization with sensor arrays. *IEEE Trans. Signal Process.* **2005**, *53*, 3010–3022. [[CrossRef](#)]
21. Yin, J.; Chen, T. Direction-of-arrival estimation using a sparse representation of array covariance vectors. *IEEE Trans. Signal Process.* **2011**, *59*, 4489–4493. [[CrossRef](#)]
22. Cai, S.; Wang, G.; Zhang, J.; Wong, K.-K.; Zhu, H. Efficient direction of arrival estimation based on sparse covariance fitting criterion with modeling mismatch. *Signal Process.* **2017**, *137*, 264–273. [[CrossRef](#)]
23. Hyder, M.M.; Mahata, K. Direction-of-Arrival Estimation Using a Mixed $l_2,0$ Norm Approximation. *IEEE Trans. Signal Process.* **2010**, *58*, 4646–4655. [[CrossRef](#)]
24. He, Z.-Q.; Shi, Z.-P.; Huang, L.; So, H.C. Underdetermined DOA estimation for wideband signals using robust sparse covariance fitting. *IEEE Signal Process. Lett.* **2014**, *22*, 435–439. [[CrossRef](#)]
25. Li, L.; Yu, C.; Li, Y.; Huang, Z.; Wu, Q. Joint squared-sine function and ANM-based DOA estimation with RIS. *IEEE Trans. Veh. Technol.* **2023**, *72*, 16856–16860. [[CrossRef](#)]
26. Pote, R.R.; Rao, B.D. Maximum likelihood-based gridless doa estimation using structured covariance matrix recovery and sbl with grid refinement. *IEEE Trans. Signal Process.* **2023**, *71*, 802–815. [[CrossRef](#)]
27. Wang, M.; Shen, Q.; Liu, W.; Cui, W. Wideband DOA Estimation with Frequency Decomposition via a Unified GS-WSpSF Framework. *IEEE Trans. Aerosp. Electron. Syst.* **2024**, 1–8. [[CrossRef](#)]
28. Wu, Y.; Wakin, M.B.; Gerstoft, P. Gridless DOA estimation with multiple frequencies. *IEEE Trans. Signal Process.* **2023**, *71*, 417–432. [[CrossRef](#)]
29. Aich, A.; Palanisamy, P. On-grid DOA estimation method using orthogonal matching pursuit. In Proceedings of the 2017 International Conference on Signal Processing and Communication (ICSPC), Coimbatore, India, 28–29 July 2017; pp. 483–487.
30. Ganguly, S.; Ghosh, I.; Ranjan, R.; Ghosh, J.; Kumar, P.K.; Mukhopadhyay, M. Compressive sensing based off-grid DOA estimation using OMP algorithm. In Proceedings of the 2019 6th International Conference on Signal Processing and Integrated Networks (SPIN), Noida, India, 7–8 March 2019; pp. 772–775.
31. Yang, J.; Liao, G.; Li, J. An efficient off-grid DOA estimation method for nested array signal processing by using sparse Bayesian learning strategies. *Signal Process.* **2016**, *128*, 110–122. [[CrossRef](#)]
32. Dai, J.; So, H.C. Real-valued sparse Bayesian learning for DOA estimation with arbitrary linear arrays. *IEEE Trans. Signal Process.* **2021**, *69*, 4977–4990. [[CrossRef](#)]
33. Zheng, R.; Xu, X.; Ye, Z.; Dai, J. Robust sparse Bayesian learning for DOA estimation in impulsive noise environments. *Signal Process.* **2020**, *171*, 107500. [[CrossRef](#)]
34. Mecklenbrauker, C.F.; Gerstoft, P.; Ollila, E. DOA M-estimation using sparse Bayesian learning. In Proceedings of the ICASSP 2022—2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Singapore, 23–27 May 2022; pp. 4933–4937.
35. Yang, Y.; Zhang, Y.; Yang, L.; Wang, Y. Wideband Direction-of-Arrival Estimation Based on Hierarchical Sparse Bayesian Learning for Signals with the Same or Different Frequency Bands. *Electronics* **2023**, *12*, 1123. [[CrossRef](#)]
36. Hu, N.; Sun, B.; Zhang, Y.; Dai, J.; Wang, J.; Chang, C. Underdetermined DOA estimation method for wideband signals using joint nonnegative sparse Bayesian learning. *IEEE Signal Process. Lett.* **2017**, *24*, 535–539. [[CrossRef](#)]
37. Liu, Z.-M.; Huang, Z.-T.; Zhou, Y.-Y. An Efficient Maximum Likelihood Method for Direction-of-Arrival Estimation via Sparse Bayesian Learning. *IEEE Trans. Wirel. Commun.* **2012**, *11*, 1–11. [[CrossRef](#)]
38. Wang, H.; Wan, L.; Dong, M.; Ota, K.; Wang, X. Assistant Vehicle Localization Based on Three Collaborative Base Stations via SBL-Based Robust DOA Estimation. *IEEE Internet Things J.* **2019**, *6*, 5766–5777. [[CrossRef](#)]
39. Chen, P.; Cao, Z.; Chen, Z.; Wang, X. Off-Grid DOA Estimation Using Sparse Bayesian Learning in MIMO Radar With Unknown Mutual Coupling. *IEEE Trans. Signal Process.* **2019**, *67*, 208–220. [[CrossRef](#)]
40. Dai, J.; So, H.C. Sparse Bayesian Learning Method for Outlier-Resistant Direction-of-Arrival Estimation. *IEEE Trans. Signal Process.* **2018**, *66*, 744–775. [[CrossRef](#)]
41. Wang, L.; Zhao, L.; Bi, G.; Wan, C.; Zhang, L.; Zhang, H. Novel Wideband DOA Estimation Based on Sparse Bayesian Learning with Dirichlet Process Priors. *IEEE Trans. Signal Process.* **2015**, *64*, 275–289. [[CrossRef](#)]
42. Liu, Z.-M.; Huang, Z.-T.; Zhou, Y.-Y. Sparsity-Inducing Direction Finding for Narrowband and Wideband Signals Based on Array Covariance Vectors. *IEEE Trans. Wirel. Commun.* **2013**, *12*, 1–12. [[CrossRef](#)]
43. Das, A.; Sejnowski, T.J. Narrowband and Wideband Off-Grid Direction-of-Arrival Estimation via Sparse Bayesian Learning. *IEEE J. Ocean. Eng.* **2017**, *43*, 108–118. [[CrossRef](#)]
44. Jian, M.; Gao, F.; Tian, Z.; Jin, S.; Ma, S. Angle-Domain Aided UL/DL Channel Estimation for Wideband mmWave Massive MIMO Systems with Beam Squint. *IEEE Trans. Wirel. Commun.* **2019**, *18*, 3515–3527. [[CrossRef](#)]
45. Wang, Q.; Zhao, Z.; Chen, Z.; Nie, Z. Grid Evolution Method for DOA Estimation. *IEEE Trans. Signal Process.* **2018**, *66*, 2374–2383. [[CrossRef](#)]
46. Yang, J.; Yang, Y. Sparse Bayesian DOA Estimation Using Hierarchical Synthesis Lasso Priors for Off-Grid Signals. *IEEE Trans. Signal Process.* **2020**, *68*, 872–884. [[CrossRef](#)]

47. Das, A. Real-Valued Sparse Bayesian Learning for Off-Grid Direction-of-Arrival (DOA) Estimation in Ocean Acoustics. *IEEE J. Ocean. Eng.* **2021**, *46*, 172–182. [[CrossRef](#)]
48. Wipf, D.P.; Rao, B.D. Sparse Bayesian learning for basis selection. *IEEE Trans. Signal Process.* **2004**, *52*, 2153–2216. [[CrossRef](#)]
49. Wipf, D.P.; Rao, B.D. An empirical Bayesian strategy for solving the simultaneous sparse approximation problem. *IEEE Trans. Signal Process.* **2007**, *55*, 3704–3716. [[CrossRef](#)]
50. Babacan, S.D.; Molina, R.; Katsaggelos, A.K. Bayesian compressive sensing using laplace priors. *IEEE Trans. Image Process.* **2009**, *19*, 53–63. [[CrossRef](#)] [[PubMed](#)]
51. Hu, Q.; Shi, S.; Cai, Y.; Yu, G. DDPG-driven deep-unfolding with adaptive depth for channel estimation with sparse Bayesian learning. *IEEE Trans. Signal Process.* **2022**, *70*, 4665–4680. [[CrossRef](#)]
52. Wan, Q.; Fang, J.; Huang, Y.; Duan, H.; Li, H. A variational Bayesian inference-inspired unrolled deep network for MIMO detection. *IEEE Trans. Signal Process.* **2022**, *70*, 423–437. [[CrossRef](#)]
53. Gupta, M.R.; Chen, Y. Theory and use of the EM algorithm. *Found. Trends Signal Process.* **2011**, *4*, 223–296. [[CrossRef](#)]
54. Yang, Z.; Xie, L.; Zhang, C. Off-grid direction of arrival estimation using sparse Bayesian inference. *IEEE Trans. Signal Process.* **2012**, *61*, 38–43. [[CrossRef](#)]
55. Bishop, C.M.; Nasrabadi, N.M. *Pattern Recognition and Machine Learning*; Springer: New York, NY, USA, 2006.
56. Tzikas, D.G.; Likas, A.C.; Galatsanos, N.P. The variational approximation for Bayesian inference. *IEEE Signal Process. Mag.* **2008**, *25*, 131–146. [[CrossRef](#)]
57. Bishop, C.M.; Tipping, M. Variational relevance vector machines. *arXiv* **2013**, arXiv:1301.3838.
58. Rangan, S. Generalized approximate message passing for estimation with random linear mixing. In Proceedings of the 2011 IEEE International Symposium on Information Theory Proceedings, St. Petersburg, Russia, 31 July–5 August 2011; pp. 2168–2172.
59. Wipf, D.; Palmer, J.; Rao, B.; Kreutz-Delgado, K. Performance evaluation of latent variable models with sparse priors. In Proceedings of the 2007 IEEE International Conference on Acoustics, Speech, and Signal Processing—ICASSP '07, Honolulu, HI, USA, 15–20 April 2007; pp. II-453–II-456.
60. Fox, C.W.; Roberts, S.J. A tutorial on variational Bayesian inference. *IEEE Signal Proc. Mag.* **2012**, *38*, 85–95. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.