

Correlated Sparse Bayesian Learning for Recovery of Block Sparse Signals with Unknown Borders

Doğan, Didem; Leus, Geert

DOI

[10.1109/OJSP.2024.3360914](https://doi.org/10.1109/OJSP.2024.3360914)

Publication date

2024

Document Version

Final published version

Published in

IEEE Open Journal of Signal Processing

Citation (APA)

Doğan, D., & Leus, G. (2024). Correlated Sparse Bayesian Learning for Recovery of Block Sparse Signals with Unknown Borders. *IEEE Open Journal of Signal Processing*, 5, 421-435.
<https://doi.org/10.1109/OJSP.2024.3360914>

Important note

To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy

Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.

Correlated Sparse Bayesian Learning for Recovery of Block Sparse Signals With Unknown Borders

DIDEM DOGAN ^{ID} AND GEERT LEUS ^{ID} (Fellow, IEEE)

Faculty of Electrical Engineering, Mathematics, and Computer Science, Delft University of Technology, 2628 CD Delft, The Netherlands

CORRESPONDING AUTHOR: DIDEM DOGAN (e-mail: d.doganbaskaya-1@tudelft.nl).

This work was supported by the TOUCAN project under Grant 17208 of the research programme TTW-OTP which is financed by the Dutch Research Council (NWO).

ABSTRACT We consider the problem of recovering complex-valued block sparse signals with unknown borders. Such signals arise naturally in numerous applications. Several algorithms have been developed to solve the problem of unknown block partitions. In pattern-coupled sparse Bayesian learning (PCSBL), each coefficient involves its own hyperparameter and those of its immediate neighbors to exploit the block sparsity. Extended block sparse Bayesian learning (EBSBL) assumes the block sparse signal consists of correlated and overlapping blocks to enforce block correlations. We propose a simpler alternative to EBSBL and reveal the underlying relationship between the proposed method and a particular case of EBSBL. The proposed algorithm uses the fact that immediate neighboring sparse coefficients are correlated. The proposed model is similar to classical sparse Bayesian learning (SBL). However, unlike the diagonal correlation matrix in conventional SBL, the unknown correlation matrix has a tridiagonal structure to capture the correlation with neighbors. Due to the entanglement of the elements in the inverse tridiagonal matrix, instead of a direct closed-form solution, an approximate solution is proposed. The alternative algorithm avoids the high dictionary coherence in EBSBL, reduces the unknowns of EBSBL, and is computationally more efficient. The sparse reconstruction performance of the algorithm is evaluated with both correlated and uncorrelated block sparse coefficients. Simulation results demonstrate that the proposed algorithm outperforms PCSBL and correlation-based methods such as EBSBL in terms of reconstruction quality. The numerical results also show that the proposed correlated SBL algorithm can deal with isolated zeros and nonzeros as well as block sparse patterns.

INDEX TERMS Block sparse signals, correlated sparse Bayesian learning, expectation-maximization (EM) method, compressive sensing.

I. INTRODUCTION

Block sparsity has been observed for signals in a wide range of applications, such as the cluster structure of scatterers in radar images [1], [2], [3], fetal ECG [4], ultrasound signals [5] and so on. The structured sparse model can be naturally exploited by including further the dependencies among sparse coefficients, such as the correlations between coefficients or dependence of the sparsity patterns. Under noisy environments or with very compressive measurements, algorithms properly leveraging such an underlying structure

could achieve a robust recovery compared to their counterparts which merely exploit the sparsity.

A number of algorithms have been proposed for block sparse signal recovery when the block partition is known a priori, including greedy pursuit algorithms like Model-based Compressive Sampling Matching Pursuit (CoSaMp) [6], Block Orthogonal Matching Pursuit (Block OMP) [7], and regularized convex optimizations, such as group Lasso [8], group basis pursuit [9], mixed ℓ_1/ℓ_2 programming [10] and block sparse Bayesian learning (BSBL) [11], [12]. These

algorithms require knowledge of the cluster pattern (block partition) a priori. However, prior knowledge of the block partition of sparse coefficients is practically unavailable. To address this problem, a particular structure is imposed on the support of sparse coefficients in the first category of algorithms [13], [14]. For instance, the Clustered Sparse Solver (Cluss) algorithm in [13] employs a hierarchical Bayesian “spike-and-slab” prior model to encourage the sparseness and promote the cluster patterns simultaneously. However, since the resulting posterior distribution of the Bayesian cluster sparse model cannot be analytically derived, Markov chain Monte Carlo (MCMC) sampling [15] has to be employed for Bayesian inference. Similarly, the Boltzmann machine is employed on the support of sparse coefficients in [14] to model the dependencies and an approximate model of the maximum a posterior (MAP) estimator is used to estimate hidden variables with exhaustive search.

In another category of algorithms, different block sparsity priors are imposed directly on the sparse coefficients. Algorithms such as extended block sparse Bayesian learning (EBSBL) [12], pattern-coupled sparse Bayesian learning (PCSBL) [16], cluster structured sparse Bayesian learning (CSBL) [17] and total variation regularized sparse Bayesian learning (TVSBL) [18] are evaluated under this category. EBSBL is an extension of the block sparse Bayesian learning algorithm which is designed for known block partitions. In EBSBL, it is assumed that the nonzero blocks are arbitrarily located and their size is unknown. Then the signal is partitioned into a number of overlapping and fully correlated blocks with user-defined block size. By expanding the overlapping blocks to a non-overlapping block structure, an extended set of fully correlated blocks is introduced for the unknown sparse coefficients. Based on this block structure, an expanded sensing matrix is constructed by adding redundant columns to the original sensing matrix. Similarly, the unknown coefficient vectors introduced for each block are concatenated as an augmented vector. Thereafter, the measurements are expressed as a linear combination of the expanded measurement matrix and concatenation of the block vectors. Then, the problem can be effectively solved by the traditional BSBL algorithm to find the augmented block vector. Finally, the unknown sparse coefficients can be computed by using the relation between the original sparse coefficients and the blocks. PCSBL, on the other hand, introduces a pattern-coupled hierarchical Gaussian prior for each coefficient involving its own hyperparameter and those of its immediate neighbors to exploit interactions between neighboring coefficients. A suboptimal solution is attained for the hyperparameters; however, the performance of PCSBL heavily depends on a proper selection of the hyperparameters. Extensions of PCSBL to two-dimensional cases are further addressed in [19], [20]. CSBL takes on a similar formula as the pattern-coupled prior used in PCSBL without relying on the hierarchical distribution over the hyperparameters. As a result, no proper hyperparameter selection is required in contrast

to PCSBL. For both the PCSBL and CSBL, the expectation-maximization (EM) is developed to learn the hidden variables and the unknown parameters. Lastly, in TVSBL, a block SBL method has been developed inspired by total variation (TV) denoising [18].

In the mentioned category, algorithms such as EBSBL, PCSBL, CSBL, and TVSBL exploit the EM algorithm in update rules. Instead of EM, the variational Bayesian inference can be exploited. As an alternative to PCSBL with EM, [21] develops an algorithm using variational Bayesian inference, and it has a noticeable performance for the MIMO channel estimation problem. Ref. [22] introduces a method that exploits variational Bayesian inference instead of EM, which can be considered an alternative version of EBSBL (BSBL). Although it performs similarly to BSBL, it is faster than BSBL as it is a covariance-free algorithm. Ref. [23] also presents both EM and variational Bayesian inference methods for Kalman smoothing, reporting that they have similar performance, but the variational Bayesian method is slower due to the high number of iterations required for convergence. Although there is no clear consensus on which family of methods is faster or has better performance, we prefer EM-based update rules for a fair comparison with the existing methods. However, the variational Bayesian method can also be exploited as future work.

In the second category, where different group-structured priors are imposed on the sparse coefficients, only a few existing algorithms consider intra-block correlation, i.e., the correlation among the elements within each block. In practical applications intra-block correlation widely exists in signals, such as physiological signals [4] and images [5]. In this work, we review several algorithms that explore and exploit intra-block correlation to improve performance. These algorithms are based on block sparse Bayesian learning (BSBL) and extended block sparse Bayesian learning (EBSBL) [12]. However, BSBL requires knowledge of the block partition and EBSBL suffers from several key drawbacks leading to high computational complexity and coherence, and a larger dictionary matrix. Note that high number of unknowns deteriorates the performance of EBSBL.

In this work, a new algorithm dealing with the problem of an unknown block partition of the correlated signal is proposed to alleviate the challenges of recently reported methods. This work is motivated by the disadvantages of EBSBL, where the interactions among neighboring coefficients are implicitly modeled by a linear transformation of the artificially constructed augmented vector. A new structured sparse prior can be derived based on the underlying relationship between the correlation matrices in the augmented EBSBL model and the original signal model. The proposed algorithm uses the fact that immediate neighboring sparse coefficients are correlated. It is also inspired by PCSBL [16] and CSBL [17] in the sense that it considers the relation between neighboring sparse coefficients. However, it is different than PCSBL and CSBL since they do not leverage the existing correlations among

the data. We naturally exploit these correlations instead of relating the hyperparameters. In this work, we only focus on the correlations between immediate neighbors. Hence, the unknown correlation matrix has a tridiagonal structure, which is different than the diagonal correlation matrix used in conventional SBL algorithms [24], [25]. The proposed algorithm with a tridiagonal correlation matrix is a simple extension of the classical sparse Bayesian learning algorithm which has a diagonal correlation matrix. On the other hand, our algorithm improves the group sparsity performance as it does not ignore the correlation with the neighbors.

While classical SBL assumes the sources are uncorrelated, EBSBL assumes the different blocks share a common correlation structure [12]. In this work, we claim that if there is an intra-block correlation in the data, modeling an inter-element correlation can already trigger a grouping effect. In other words, without building a new block-based data model, assuming inter-element correlation in the classical SBL model already promotes group sparsity. It is also shown that there is a grouping effect even when the true data does not have intra-block correlations but only contains the block structure without correlations.

There are three contributions of this paper, which can concisely be summarized as follows. First, a new structured sparse recovery algorithm is proposed, which can be considered a better alternative to EBSBL. The algorithm avoids the high dictionary coherence in EBSBL and also reduces the unknowns of EBSBL, making it computationally more efficient. Second, we provide an analysis of the relation between our algorithm and a particular case of EBSBL. The intuitions given here can be used to enhance the approximate update rule for the proposed algorithm with a tridiagonal structure. The third contribution of our paper is the ability of the proposed algorithm to tackle irregular sparsity patterns where the sparse vector contains both block sparse and isolated coefficients. Once there are isolated zeros and nonzeros in the data, the group sparsity algorithms might not perform well as their assumptions enforce only the grouping effect. However, the proposed algorithm is also able to deal with isolated zeros and isolated nonzeros as well as block sparse patterns.

The rest of the paper is organized as follows. Section II provides a review of classical sparse Bayesian learning and extended block SBL algorithms. A tridiagonal correlation based prior on the sparse coefficients is derived from the classical SBL algorithm in Section III. Section IV discusses the relationship of the proposed method to EBSBL. Comparisons of the proposed method with the state-of-the-art are shown in Section V. Conclusions are drawn in Section VI.

A. NOTATION

Throughout the paper, bold symbols in small and capital fonts are used for vectors and matrices, respectively. $\|\mathbf{x}\|_2^2$ denotes the l_2 -norm of vector \mathbf{x} . \mathbf{x}_i denotes the i -th block of \mathbf{x} . And the i -th element of \mathbf{x} is either denoted by $\mathbf{x}(i)$, $(\mathbf{x})_i$ or x_i . Furthermore, $(\mathbf{A})_{ij}$ and A_{ij} represent the element in i th row and j th column of an A matrix. For matrix \mathbf{A} , \mathbf{A}^H

and \mathbf{A}^{-1} denote the Hermitian and the inverse of the matrix, respectively. $\text{tr}(\mathbf{A})$ is the trace of a matrix \mathbf{A} . Notation $\text{diag}(\mathbf{A})$ denotes a column vector composed of the diagonal elements of a matrix \mathbf{A} . $\text{rank}(\mathbf{A})$ denotes the rank of matrix \mathbf{A} . $|\mathbf{A}|$ is the determinant of the matrix \mathbf{A} . $\mathcal{CN}(\cdot)$ denotes the multivariate complex Gaussian distribution.

II. REVIEW OF SBL ALGORITHMS

Sparse signal recovery problems attempt to recover the unknown sparse coefficient vector $\mathbf{s} \in \mathbb{C}^n$ from noisy and distorted measurements $\mathbf{z} \in \mathbb{C}^m$. More specifically, we consider the model

$$\mathbf{z} = \mathbf{A}\mathbf{s} + \mathbf{n} \quad (1)$$

where $\mathbf{n} \in \mathbb{C}^m$ is the additive white noise and $\mathbf{A} \in \mathbb{C}^{m \times n}$ is the measurement matrix with $m \ll n$. A block structure in \mathbf{s} is commonly observed in practice, where elements of \mathbf{s} tend to be nonzero in multiple groups with unknown block sizes and arbitrary locations. Let us first give a detailed review of the considered structure in classical SBL. Such a detailed review is given to show the relation between our method to the classical SBL. Then we provide a brief review on the EBSBL [12].

A. CLASSICAL SBL

Using Bayesian inference to solve the linear problem in (1) involves determining the posterior distribution of the complex amplitudes \mathbf{s} from the likelihood and prior distribution. The conditional probability density function (PDF) for \mathbf{z} given the sources \mathbf{s} is complex Gaussian with noise variance σ^2 :

$$p(\mathbf{z}|\mathbf{s}; \sigma^2) = \frac{\exp(-\frac{1}{\sigma^2}\|\mathbf{z} - \mathbf{A}\mathbf{s}\|_2^2)}{(\pi\sigma^2)^m}. \quad (2)$$

The unknown coefficients s_l are assumed to be independent across different coefficients l and to follow a zero-mean complex Gaussian distribution:

$$p_l(s_l; \gamma_l) = \frac{1}{\pi\gamma_l} e^{-|s_l|^2/\gamma_l}. \quad (3)$$

Then the variances of the elements of \mathbf{s} can be stacked into $\boldsymbol{\gamma} = [\gamma_1, \dots, \gamma_n]^T$ and we get

$$p(\mathbf{s}; \boldsymbol{\gamma}) = \prod_{l=1}^n p_l(s_l; \gamma_l) = \mathcal{CN}(\mathbf{0}, \boldsymbol{\Gamma}) \quad (4)$$

with $\boldsymbol{\Gamma} = \text{diag}(\boldsymbol{\gamma})$. When the variance $\gamma_l = 0$, then $s_l = 0$ with probability 1. Hence the sparsity of the model is controlled by the hyper-parameters $\boldsymbol{\gamma}$. It has been shown that such a model enforces sparsity.

By using the likelihood in (2) and the prior in (4), the posterior PDF of \mathbf{s} can be found using the Bayes rule conditioned on $\boldsymbol{\gamma}$ and σ^2 and neglecting the denominator

$$p(\mathbf{s}|\mathbf{z}; \boldsymbol{\gamma}, \sigma^2) = \frac{p(\mathbf{z}|\mathbf{s}; \sigma^2)p(\mathbf{s}; \boldsymbol{\gamma})}{p(\mathbf{z}; \boldsymbol{\gamma}, \sigma^2)} \propto p(\mathbf{z}|\mathbf{s}; \sigma^2)p(\mathbf{s}; \boldsymbol{\gamma})$$

$$\propto \frac{e^{-\text{tr}((\mathbf{s}-\boldsymbol{\mu}_s)^H \boldsymbol{\Sigma}_s^{-1} (\mathbf{s}-\boldsymbol{\mu}_s))}}{\pi^m \det(\boldsymbol{\Sigma}_s)} = \mathcal{CN}(\boldsymbol{\mu}_s, \boldsymbol{\Sigma}_s). \quad (5)$$

Since both $p(\mathbf{z}|\mathbf{s}; \sigma^2)$ in (2) and $p(\mathbf{s}; \boldsymbol{\gamma})$ in (4) are Gaussians, their product (5) is Gaussian with posterior mean $\boldsymbol{\mu}_s$ and covariance $\boldsymbol{\Sigma}_s$ given by

$$\begin{aligned} \boldsymbol{\mu}_s &= E[\mathbf{s}|\mathbf{z}; \boldsymbol{\Gamma}, \sigma^2] = \boldsymbol{\Gamma} \mathbf{A}^H \boldsymbol{\Sigma}_z^{-1} \mathbf{z} \\ &= \sigma^{-2} (\sigma^{-2} \mathbf{A}^H \mathbf{A} + \boldsymbol{\Gamma}^{-1})^{-1} \mathbf{A}^H \mathbf{z} \end{aligned} \quad (6)$$

and

$$\begin{aligned} \boldsymbol{\Sigma}_s &= E[(\mathbf{s} - \boldsymbol{\mu}_s)(\mathbf{s} - \boldsymbol{\mu}_s)^H | \mathbf{z}; \boldsymbol{\Gamma}, \sigma^2] \\ &= \left(\frac{1}{\sigma^2} \mathbf{A}^H \mathbf{A} + \boldsymbol{\Gamma}^{-1} \right)^{-1} = \boldsymbol{\Gamma} - \boldsymbol{\Gamma} \mathbf{A}^H \boldsymbol{\Sigma}_z^{-1} \mathbf{A} \boldsymbol{\Gamma} \end{aligned} \quad (7)$$

where the covariance matrix of the measurements $\boldsymbol{\Sigma}_z$ is given by

$$\boldsymbol{\Sigma}_z = \sigma^2 \mathbf{I}_m + \mathbf{A} \boldsymbol{\Gamma} \mathbf{A}^H. \quad (8)$$

Here, to estimate the $\boldsymbol{\Gamma}$ and σ^2 , we perform expectation-maximization (EM) to maximize $p(\mathbf{z}; \boldsymbol{\Gamma}, \sigma^2)$. The actual EM formulation proceeds by treating the \mathbf{s} as a hidden variable and then by maximizing

$$Q(\boldsymbol{\Gamma}, \sigma^2) = E_{\mathbf{s}|\mathbf{z}; \boldsymbol{\Gamma}, \sigma^2}[\log p(\mathbf{z}, \mathbf{s}; \boldsymbol{\Gamma}, \sigma^2)] \quad (9)$$

with respect to the hyperparameters $\boldsymbol{\Gamma}$ and σ^2 to find their estimates, where $E_{\mathbf{s}|\mathbf{z}; \boldsymbol{\Gamma}, \sigma^2}[\cdot]$ denotes an expectation with respect to the posterior distribution of \mathbf{s} . By substituting the joint distribution $p(\mathbf{z}, \mathbf{s}; \boldsymbol{\Gamma}, \sigma^2) = p(\mathbf{z}|\mathbf{s}; \sigma^2)p(\mathbf{s}; \boldsymbol{\Gamma})$ into (9) we obtain

$$\begin{aligned} E_{\mathbf{s}|\mathbf{z}; \boldsymbol{\Gamma}, \sigma^2}[\log p(\mathbf{z}, \mathbf{s}; \boldsymbol{\Gamma}, \sigma^2)] &= E_{\mathbf{s}|\mathbf{z}; \sigma^2}[\log p(\mathbf{z}|\mathbf{s}; \sigma^2)] \\ &+ E_{\mathbf{s}|\mathbf{z}; \boldsymbol{\Gamma}}[\log p(\mathbf{s}; \boldsymbol{\Gamma})]. \end{aligned} \quad (10)$$

Ignoring the terms independent from $\boldsymbol{\Gamma}$, we can estimate the $\boldsymbol{\Gamma}$ by maximizing $E_{\mathbf{s}|\mathbf{z}; \boldsymbol{\Gamma}}[\log p(\mathbf{s}; \boldsymbol{\Gamma})]$. Starting from

$$\log p(\mathbf{s}; \boldsymbol{\Gamma}) = -\frac{1}{2} \log(|\boldsymbol{\Gamma}|) - \frac{1}{2} \mathbf{s}^H \boldsymbol{\Gamma}^{-1} \mathbf{s} \quad (11)$$

and using the fact that $\mathbf{s}^H \boldsymbol{\Gamma}^{-1} \mathbf{s} = \text{tr}(\boldsymbol{\Gamma}^{-1} \mathbf{s} \mathbf{s}^H)$ and $E[\mathbf{s} \mathbf{s}^H] = \boldsymbol{\Sigma}_s + \boldsymbol{\mu}_s \boldsymbol{\mu}_s^H$ we attain the following expression:

$$E_{\mathbf{s}|\mathbf{z}; \boldsymbol{\Gamma}}[\log p(\mathbf{s}; \boldsymbol{\Gamma})] = -\frac{1}{2} \log(|\boldsymbol{\Gamma}|) - \frac{1}{2} \text{tr}(\boldsymbol{\Gamma}^{-1} (\boldsymbol{\Sigma}_s + \boldsymbol{\mu}_s \boldsymbol{\mu}_s^H)). \quad (12)$$

This function is also called as Q function. Since we have a diagonal $\boldsymbol{\Gamma}$, we can maximize this function only for the diagonal elements in $\boldsymbol{\Gamma}$. Hence, we can take the derivative as follows:

$$\frac{\partial E_{\mathbf{s}|\mathbf{z}; \boldsymbol{\Gamma}}[\log p(\mathbf{s}; \boldsymbol{\Gamma})]}{\partial \gamma_l} = \frac{(\boldsymbol{\Sigma}_s + \boldsymbol{\mu}_s \boldsymbol{\mu}_s^H)_{ll}}{\gamma_l^2} - \frac{1}{\gamma_l} = 0. \quad (13)$$

Then, the closed-form solution for γ_l is given by

$$\gamma_l = (\boldsymbol{\Sigma}_s)_{ll} + (\boldsymbol{\mu}_s)_l (\boldsymbol{\mu}_s^H)_l. \quad (14)$$

Note that in classical SBL, the sources are assumed uncorrelated.

B. EXTENDED BLOCK SBL (EBSBL)

The BSBL method is proposed to solve the group sparsity problem with a known block partition [12] in which both the block size and the border locations of the blocks are known. Group sparsity is enforced by considering a separate covariance matrix per block, with a common structure over the different blocks and potentially a specific structure within every block (e.g., Toeplitz). The extension of this method, the extended BSBL (EBSBL) deals with unknown block partitions. EBSBL considers an extended set of hidden blocks \mathbf{x}_i with hidden block size h for $i = 1 \dots g$ ($g = n - h + 1$ is the number of blocks). More specifically, the unknown signal \mathbf{s} is represented as

$$\mathbf{s} = \sum_{i=1}^g \mathbf{E}_i \mathbf{x}_i \quad (15)$$

where $\mathbf{E}_i \in \mathbb{R}^{m \times h}$ contains an identity matrix from the i th row to the $(i+h-1)$ th row and zeros for the other entries. This idea allows for blocks of size h with unknown border locations but it can also handle blocks that might have a size different from h as discussed in [12]. Under (15), the model in (1) can be written as follows:

$$\mathbf{z} = \sum_{i=1}^g \mathbf{A} \mathbf{E}_i \mathbf{x}_i + \mathbf{n} = \boldsymbol{\Theta} \mathbf{x} + \mathbf{n} \quad (16)$$

where $\mathbf{x} = [\mathbf{x}_1^T \dots \mathbf{x}_g^T]^T$ and $\boldsymbol{\Theta} = [\boldsymbol{\Theta}_1 \dots \boldsymbol{\Theta}_g]$ with $\boldsymbol{\Theta}_i = \mathbf{A} \mathbf{E}_i$. The problem in (16) becomes a block sparsity problem with a known block partition and is solvable by BSBL. Specifically, \mathbf{x} is assumed to follow the given distribution:

$$p(\mathbf{x}; \{\beta_i, \mathbf{B}\}_{i=1}^g) = \mathcal{CN}(\mathbf{0}, \boldsymbol{\Gamma}_0), \quad (17)$$

where $\boldsymbol{\Gamma}_0 = \text{diag}(\beta_1 \mathbf{B}, \dots, \beta_g \mathbf{B})$ and where each block satisfies the parameterized multivariate Gaussian distribution of $p(\mathbf{x}_i; \beta_i, \mathbf{B}) = \mathcal{CN}(\mathbf{0}, \beta_i \mathbf{B})$ with β_i determining the degree of block sparsity. Then we can find the MAP estimate of \mathbf{x} using the given formulas in (6), (7) as follows:

$$\boldsymbol{\mu}_x = E[\mathbf{x}|\mathbf{z}; \boldsymbol{\Gamma}_0, \sigma^2] = \boldsymbol{\Gamma}_0 \boldsymbol{\Theta}^H (\sigma^2 \mathbf{I}_m + \boldsymbol{\Theta} \boldsymbol{\Gamma}_0 \boldsymbol{\Theta}^H)^{-1} \mathbf{z} \quad (18)$$

$$\begin{aligned} \boldsymbol{\Sigma}_x &= E[(\mathbf{x} - \boldsymbol{\mu}_x)(\mathbf{x} - \boldsymbol{\mu}_x)^H | \mathbf{z}; \boldsymbol{\Gamma}_0, \sigma^2] \\ &= \left(\frac{1}{\sigma^2} \boldsymbol{\Theta}^H \boldsymbol{\Theta} + \boldsymbol{\Gamma}_0^{-1} \right)^{-1}. \end{aligned} \quad (19)$$

After iteratively finding the hyperparameters, finally, the estimate of the unknown signal \mathbf{s} is given by

$$\hat{\mathbf{s}} = \sum_{i=1}^g \mathbf{E}_i \boldsymbol{\mu}_{x_i}. \quad (20)$$

EBSBL is designed to cope with block-sparse recovery under the assumption of an unknown block partition. However, it suffers from several disadvantages. First, it leads to a higher

computational complexity, as the augmented vector \mathbf{x} is of size $h \times (n - h + 1)$, which is almost h times the size of the original signal \mathbf{s} . More importantly, since the expanded measurement matrix Θ is constructed by adding redundant columns to the original measurement matrix \mathbf{A} , dictionary Θ will be of high coherence. It affects the efficiency of the sparse coefficient estimation in SBL [17].

III. CORRELATED SPARSE BAYESIAN LEARNING ALGORITHM

As we mentioned before, classical SBL assumes the sources are uncorrelated, whereas EBSBL assumes the different blocks share a common correlation structure, which includes intra-block correlation or not. Related to the latter, it has been shown that if the data contains an intra-block correlation, the performance of EBSBL improves if it is taken into account. On the other hand, the performance of EBSBL ignoring intra-block correlation does not change with the amount of intra-block correlation in the data. The claim we make in this work is that if there is an intra-block correlation in the data, just modeling an inter-element correlation can already trigger a grouping effect. In other words, assuming inter-element correlation in the classical SBL model already promotes group sparsity without the need for building a new block-based data model first. Experimental results (shown later on) show that this grouping effect is even there when the true data has no intra-block correlations.

In this section, we, therefore, discuss how the classical SBL can be extended for correlated sources, where we assume for simplicity that only neighboring elements are correlated.

A. PRIORS ON THE SOURCES

In this section, the complex coefficients s_l , which were assumed to be independent and uncorrelated in the classical SBL, are assumed to be correlated with their immediate neighbors. In other words, \mathbf{s} has the following distribution:

$$p(\mathbf{s}; \mathbf{\Gamma}) = \mathcal{CN}(\mathbf{0}, \mathbf{\Gamma}) \quad (21)$$

with the following tri-diagonal structure for $\mathbf{\Gamma}$:

$$\mathbf{\Gamma} = E[\mathbf{ss}^H] = \begin{bmatrix} \Gamma_{11} & \Gamma_{12} & 0 & \dots & 0 \\ \Gamma_{21} & \Gamma_{22} & \Gamma_{23} & \dots & 0 \\ 0 & \Gamma_{32} & \Gamma_{33} & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & \dots & \dots & \Gamma_{mm} \end{bmatrix}. \quad (22)$$

Hence, we assume to have nonzero elements in the tri-diagonal elements of $\mathbf{\Gamma}$, $\Gamma_{l'l'}$ where $l' = l, l' = l + 1, l' = l - 1$, by ignoring the other elements of $\mathbf{\Gamma}$. Note that $\mathbf{\Gamma}$'s diagonal elements represent the power of the coefficients and $\Gamma_{ll} \geq 0$. When the variance $\Gamma_{ll} = 0$, then $s_l = 0$. Hence, the sparsity of the model is controlled by the diagonal elements of $\mathbf{\Gamma}$. For that problem, the likelihood is given by

$$p(\mathbf{s}|\mathbf{z}; \mathbf{\Gamma}, \sigma^2) = \mathcal{CN}(\boldsymbol{\mu}_s, \boldsymbol{\Sigma}_s), \quad (23)$$

which is similar to the formulation in (21). Based on the likelihood in (23) and the prior of \mathbf{s} in (21), it is easy to show that the posterior of \mathbf{s} is a Gaussian with mean and covariance

$$\hat{\mathbf{s}} = \boldsymbol{\mu}_s = \mathbf{\Gamma} \mathbf{A}^H (\sigma^2 \mathbf{I}_m + \mathbf{A} \mathbf{\Gamma} \mathbf{A}^H)^{-1} \mathbf{z}, \quad (24)$$

$$\boldsymbol{\Sigma}_s = \left(\frac{1}{\sigma^2} \mathbf{A}^H \mathbf{A} + \mathbf{\Gamma}^{-1} \right)^{-1}, \quad (25)$$

where $\mathbf{\Gamma}$ is a tri-diagonal matrix but the inverse $\mathbf{\Gamma}^{-1}$ does not have a simple structure as in the diagonal case. However, a fast inverse of this tri-diagonal matrix is available with recursive methods [26].

B. DISCUSSION

Once we insert (22) into (24), while attaining $\boldsymbol{\mu}_s$ or $\hat{\mathbf{s}}$, we can see how the structure of $\mathbf{\Gamma}$ affects the relation with the neighboring elements. For the computation of $\boldsymbol{\mu}_s$, we can see that in each row there are contributions from the sub-diagonals (the correlations with the neighbors) in the multiplication of the $\mathbf{\Gamma}$ and $\mathbf{A}^H (\sigma^2 \mathbf{I}_m + \mathbf{A} \mathbf{\Gamma} \mathbf{A}^H)^{-1} \mathbf{z}$. Here, in addition to Γ_{ll} , both $\Gamma_{l(l-1)}$ and $\Gamma_{l(l+1)}$ contribute to $(\boldsymbol{\mu}_s)_l$.

For the correlated block sparse data, the correlations with the neighboring elements (correlations in the sub-diagonals) become nonzero inside the group, and they become zero outside the group or for the corner elements of the group. The nonzero elements on the sub-diagonals enforce the neighboring elements to be nonzero in the group. Likewise, the corners of the groups can be clearly separated as the diagonals for the corner elements are nonzero, but the correlation with the neighboring zero element which is in the sub-diagonal is zero. If we focus on a single nonzero element surrounded by zeros (isolated nonzero element), its autocorrelation is nonzero, but the elements in the sub-diagonal are zero. Similarly, for a zero element inside a nonzero group (isolated zero elements), the correlations in the sub-diagonal become zero and are not affected by the neighboring nonzeros.

Once these elements of the $\mathbf{\Gamma}$ are used in (24), it can be seen that the contributions are only between consecutive nonzero elements. If there is a zero neighbor, then it does not have any contribution. Furthermore, if the nonzero element is a corner element or an isolated nonzero element, it also does not have a contribution to the zero neighbors. Only the nonzero elements in the group contribute to each other as given in (24) and provide a grouping effect. As a result, the proposed algorithm does not have any exponential decay around the corners of the groups or isolated zero and nonzero coefficients. Hence, unlike pattern coupling approaches [16], [17], the proposed algorithm tackles both isolated nonzero and zero elements in addition to the block sparse patterns.

C. ESTIMATION OF $\mathbf{\Gamma}$

By following the derivations from (9)–(12) to derive the EM-based update rule for $\mathbf{\Gamma}$, we attain the minimization function for $\mathbf{\Gamma}$. Since we assume a structure over $\mathbf{\Gamma}$, we can minimize the function in (12) only for the tri-diagonal elements in (22). Note that we have complex-valued data and $\mathbf{\Gamma}$ is Hermitian

symmetric but not symmetric, and thus each entry in the matrix is considered as a unique entry. Hence, we can take the derivative of all tridiagonal entries and set them to zero, leading to

$$\frac{\partial E_{\mathbf{s}|\mathbf{z};\Gamma}[p(\mathbf{s}; \Gamma)]}{\partial \Gamma_{ll'}} = [\Gamma^{-1}(\boldsymbol{\Sigma}_{\mathbf{s}} + \boldsymbol{\mu}_{\mathbf{s}}\boldsymbol{\mu}_{\mathbf{s}}^H)\Gamma^{-1} - \Gamma^{-1}]_{ll'} = 0$$

for $l = l', l = l' + 1$ and $l = l' - 1$. (26)

Note that while Γ has a particular tridiagonal structure, Γ^{-1} does not have a simple particular structure, unlike the diagonal version of Γ . Hence, finding a closed-form solution to this problem is tricky as the $\Gamma_{ll'}$ terms are entangled. Alternatively, we can use iterative algorithms to maximize (12), such as gradient-ascent [27] or fixed-point iterations [28]. However, they might be time-consuming as we need to take the inverse of Γ several times during the update steps of the iterative algorithms. Instead of solving the problem with iterative methods, we propose an approximation for the solution of Γ .

To propose an approximation for the tridiagonal Γ , we first consider the update rule for Γ without any structure, and then try to relate this to the solution for the tridiagonal Γ . If we consider all the correlations in Γ without any structure, the derivative of (12) with respect to Γ is expressed as follows:

$$\frac{\partial E_{\mathbf{s}|\mathbf{z};\Gamma}[p(\mathbf{s}; \Gamma)]}{\partial \Gamma} = \Gamma^{-1}(\boldsymbol{\Sigma}_{\mathbf{s}} + \boldsymbol{\mu}_{\mathbf{s}}\boldsymbol{\mu}_{\mathbf{s}}^H)\Gamma^{-1} - \Gamma^{-1}. \quad (27)$$

Setting this derivative to zero, we obtain the following closed-form solution

$$\bar{\Gamma} = \boldsymbol{\Sigma}_{\mathbf{s}} + \boldsymbol{\mu}_{\mathbf{s}}\boldsymbol{\mu}_{\mathbf{s}}^H. \quad (28)$$

Here $\bar{\Gamma}$ notation is used for the full Γ . However, using all the correlations does not lead to a sparse solution. In classical SBL, the sparse coefficients are generally assumed to be uncorrelated and result in a diagonal Γ . Note that the Γ estimated in (14) corresponds to the diagonal of $\bar{\Gamma}$ and is also the closed-form solution to the problem in (13). Likewise, in the tridiagonal case, although it is not a closed-form solution to the problem in (26), we can use the following approximation

$$\Gamma_{ll'} \approx (\boldsymbol{\Sigma}_{\mathbf{s}})_{ll'} + (\boldsymbol{\mu}_{\mathbf{s}})_l(\boldsymbol{\mu}_{\mathbf{s}}^H)_{l'}$$

for $l = l', l = l' + 1$ and $l = l' - 1$ (29)

as a solution for the tridiagonal elements. Note that this solution contains the elements in the diagonal and main sub-diagonals of $\bar{\Gamma}$. The intuition behind such an update rule is using the neighboring correlations that come from the full correlation matrix $\bar{\Gamma}$. Since extracting a tridiagonal submatrix from the correlation matrix $\bar{\Gamma}$ preserves the relation between neighboring elements, we embrace this approach. However, the convergence of such an update rule cannot be guaranteed as the proposed Γ matrix is not guaranteed to be positive definite anymore. Then, $\log(|\Gamma|)$ might be undefined, and it might become complex valued at certain points, and it cannot guarantee the increment of the Q function in (12) during EM steps.

To guarantee the positive definiteness of Γ and obtain a generalized update rule for Γ , we multiply the sub-diagonals by a parameter β to reduce their values:

$$\begin{aligned} \Gamma_{ll'} &\approx (\boldsymbol{\Sigma}_{\mathbf{s}})_{ll'} + (\boldsymbol{\mu}_{\mathbf{s}})_l(\boldsymbol{\mu}_{\mathbf{s}}^H)_{l'} \text{ for } l = l' \\ \Gamma_{ll'} &\approx \beta((\boldsymbol{\Sigma}_{\mathbf{s}})_{ll'} + (\boldsymbol{\mu}_{\mathbf{s}})_l(\boldsymbol{\mu}_{\mathbf{s}}^H)_{l'}) \\ &\text{for } l = l' + 1 \text{ and } l = l' - 1. \end{aligned} \quad (30)$$

Still, the correlation between neighboring elements is preserved which causes the group sparsity thanks to the relation between neighboring zeros and nonzeros. For values of $\beta \in [0, 0.5]$, we empirically observe that Γ stays positive definite and the Q function (12) does not result in complex values. Here, the most efficient β is observed to be 0.5. Although the update rule proposed for the sub-diagonal elements is intuitive, it provides a significant performance improvement for block sparse and isolated elements.

To provide theoretical bounds of β for the positive definiteness of Γ , we use the following proposition from [29]:

Proposition 1: Let $\Gamma_t = \begin{bmatrix} a_1 & b_1 & 0 & \dots & 0 \\ b_1^* & a_2 & b_2 & \dots & 0 \\ 0 & b_2^* & a_3 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & \dots & \dots & a_n \end{bmatrix}$ be a

Hermitian symmetric tridiagonal matrix with diagonal entries positive and real. If

$$b_i b_i^* < \frac{1}{4} a_i a_{i+1} \frac{1}{\cos^2(\frac{\pi}{n+1})} \quad (31)$$

then Γ_t is positive definite. Using this proposition, we can introduce the following theorem.

Theorem 1: For $\beta^2 < \frac{1}{4} \frac{1}{\cos^2(\frac{\pi}{n+1})}$, Γ in (30) is positive definite.

Proof: First, we prove this for the second part of (30) which includes $\boldsymbol{\mu}_{\mathbf{s}}\boldsymbol{\mu}_{\mathbf{s}}^H$. Assuming that $\boldsymbol{\mu}_{\mathbf{s}} = \mathbf{c}$, we have the following

$$\begin{aligned} \text{tridiag}(\boldsymbol{\mu}_{\mathbf{s}}\boldsymbol{\mu}_{\mathbf{s}}^H) &= \text{tridiag}(\mathbf{c}\mathbf{c}^H) \\ &= \begin{bmatrix} c_1 c_1^* & \beta c_1 c_2^* & 0 & \dots & 0 \\ \beta c_2 c_1^* & c_2 c_2^* & \beta c_2 c_3^* & \dots & 0 \\ 0 & \beta c_3 c_2^* & c_3 c_3^* & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & c_n c_n^* \end{bmatrix} \end{aligned} \quad (32)$$

where $\text{tridiag}(\cdot)$ is defined as the extraction of the tridiagonal part of a given matrix and multiplication of the sub-diagonals by β . For $\beta^2 < \frac{1}{4} \frac{1}{\cos^2(\frac{\pi}{n+1})}$, we can write $\beta^2 c_i c_{i+1}^* c_{i+1} c_i^* < \frac{1}{4} c_i c_i^* c_{i+1} c_{i+1}^* \frac{1}{\cos^2(\frac{\pi}{n+1})}$ using the *Proposition 1*. Therefore, for the values of $\beta^2 < \frac{1}{4} \frac{1}{\cos^2(\frac{\pi}{n+1})}$, we can conclude that $\text{tridiag}(\cdot)$ results in a positive definite matrix from rank-one matrices and $\text{tridiag}(\boldsymbol{\mu}_{\mathbf{s}}\boldsymbol{\mu}_{\mathbf{s}}^H)$ is positive definite.

Now, we generalize this approach for any positive definite matrix. Note that all positive definite matrices can be written as $\mathbf{C}\mathbf{C}^H = \sum_{i=1}^m \mathbf{c}_i \mathbf{c}_i^H$ which is a summation of multiple rank-one matrices. Here, \mathbf{c}_i is the i th column of \mathbf{C} . As $\text{tridiag}(\cdot)$ is a linear operator, it can be written as follows:

$$\text{tridiag}(\mathbf{C}\mathbf{C}^H) = \sum_{i=1}^m \text{tridiag}(\mathbf{c}_i \mathbf{c}_i^H). \quad (33)$$

Since $\text{tridiag}(\cdot)$ results in a positive definite matrix for each $\mathbf{c}_i \mathbf{c}_i^H$, and the summation of positive definite matrices is also a positive definite, $\text{tridiag}(\mathbf{C}\mathbf{C}^H)$ is positive definite.

Note that Σ_s is positive definite for a positive definite Γ . Therefore, $\text{tridiag}(\Sigma_s)$ is positive definite. As a final step, it can be concluded that,

$$\Gamma = \text{tridiag}(\Sigma_s) + \text{tridiag}(\mu_s \mu_s^H) \quad (34)$$

is positive definite since it is a summation of positive definite matrices. ■

Remark 1: From *Theorem 1*, $-\frac{1}{2 \cos(\frac{\pi}{n+1})} < \beta < \frac{1}{2 \cos(\frac{\pi}{n+1})}$ guarantees the positive definiteness of Γ . To choose a safe boundary for β , we keep it in the range $[0, 0.5]$ for values of $n > 50$. Note that we keep β positive so as not to change the sign of the correlation between the neighboring elements.

Remark 2: Note that it is important to show the proposed update rule increases the Q function (12) in every iteration. Because then we can draw from the generalized EM theory [30]: any hyperparameter update rule that ensures that the Q function is non-decreasing in each EM iteration will ensure convergence of the EM iterations to a local maximum or saddle point. Once we consider the case $\beta \in [0, 0.5]$ for (30), we empirically observe that the Q function in (12) increases in each iteration, and it has a higher increase in each EM iteration compared to the one in (14). Although the theoretical convergence proof is unavailable, we have never encountered a case where the algorithm's convergence is not satisfied with a high number of trials of simulations and various problem models.

Remark 3: The choice of β seems slightly important for the recovery performance, as demonstrated by our simulation results. Although our simulations suggest that choosing a non-zero β mostly improves the performance compared to the setting with $\beta = 0$, the best choice of β appears to be around 0.5. However, for cases where the signal structure is unknown, the β parameter might be adjusted for the structure of the data by setting it to another value in the range $\beta \in [0, 0.5]$.

D. ESTIMATION OF THE NOISE PARAMETER σ^2

To estimate σ^2 , we maximize $E_{\mathbf{s}|\mathbf{z}, \Gamma, \sigma^2}[p(\mathbf{z}, \mathbf{s}; \Gamma, \sigma^2)]$ with respect to σ^2 [16]. Now, we only focus on the σ^2 related terms on the right-hand side of (10) as follows:

$$E_{\mathbf{s}|\mathbf{z}, \sigma^2}[\log p(\mathbf{z}|\mathbf{s}; \sigma^2)] \propto -m \log \sigma^2 - \frac{E_{\mathbf{s}|\mathbf{z}, \sigma^2}[\|\mathbf{z} - \mathbf{A}\mathbf{s}\|_2^2]}{\sigma^2}$$

Algorithm 1: Correlated SBL.

```

Initialize  $\sigma^2 = 1$ ,  $\text{diag}(\Gamma) = \mathbf{1}$ ,  $\epsilon_{\min} = 0.001$ ,
 $J_{\max} = 100$ 
while  $j < J_{\max}$  and  $\epsilon_{\min} < \epsilon$  do
   $j = j + 1$ ,  $\mu_s^{\text{old}} = \mu_s^{\text{new}}$ ,  $\Gamma_{ll'}^{\text{old}} = \Gamma_{ll'}^{\text{new}}$ ,
   $\Gamma = \text{tridiag}(\Gamma_{ll'}^{\text{new}})$ 
  compute  $\Sigma_s = (\frac{1}{\sigma^2} \mathbf{A}^H \mathbf{A} + \Gamma^{-1})^{-1}$ 
  compute  $\mu_s^{\text{new}} = \Gamma \mathbf{A}^H \Sigma_z^{-1} \mathbf{z}$ 
  update  $\Gamma_{ll'}^{\text{new}}$  with (30)
  update  $(\sigma^2)^{\text{new}}$  with (37)
   $\epsilon = \|\mu_s^{\text{new}} - \mu_s^{\text{old}}\|_2$ 
end while
 $\hat{\mathbf{s}} = \mu_s^{\text{new}}$ 

```

$$= -m \log \sigma^2 - \frac{\|\mathbf{z} - \mathbf{A}\mu_s\|_2^2 + \text{tr}(\Sigma_s \mathbf{A}^H \mathbf{A})}{\sigma^2}. \quad (35)$$

The second equality can be derived as follows:

$$\begin{aligned} E_{\mathbf{s}|\mathbf{z}, \sigma^2}[\|\mathbf{z} - \mathbf{A}\mathbf{s}\|_2^2] &= \mathbf{z}^H \mathbf{z} - 2E[\mathbf{s}^H \mathbf{A}^H \mathbf{z}] + E[\mathbf{s}^H \mathbf{A}^H \mathbf{A} \mathbf{s}] \\ &= \mathbf{z}^H \mathbf{z} - 2\mu_s^H \mathbf{A}^H \mathbf{z} + \mu_s^H \mathbf{A}^H \mathbf{A} \mu_s + \text{tr}(\Sigma_s \mathbf{A}^H \mathbf{A}) \\ &= \|\mathbf{z} - \mathbf{A}\mu_s\|_2^2 + \text{tr}(\Sigma_s \mathbf{A}^H \mathbf{A}). \end{aligned} \quad (36)$$

Then, we set the derivative of (35) with respect to σ^2 to 0 and we obtain the update for σ^2 as

$$\hat{\sigma}^2 = \frac{\|\mathbf{z} - \mathbf{A}\mu_s\|_2^2 + \text{tr}(\Sigma_s \mathbf{A}^H \mathbf{A})}{m}. \quad (37)$$

The estimate of \mathbf{s} is finally given by μ_s in (6) by iteratively calculating the $\Gamma_{ll'}$'s in (29) and σ^2 in (37) till convergence. The iterative steps of the proposed algorithm are given in Algorithm 1. For the convergence, we use the stopping criterion that the maximum absolute error of two successive estimates of \mathbf{s} is smaller than a threshold, or the number of iterations exceeds the maximum number of iterations.

E. COMPUTATIONAL COMPLEXITY

The computational complexity of the proposed algorithm is similar to the classical SBL, PCSBL, and CSBL. Here, the main computational task at each iteration is to calculate the covariance matrix Σ_s as it requires computing the inverse of an $n \times n$ matrix. By using the matrix inversion lemma [31], this matrix inversion can be converted to an $m \times m$ matrix inversion. Hence the computational complexity is of order $\mathcal{O}(m^3)$. However, for the computation of the tridiagonal Γ , there is a slight increase in the computational cost, but its effect on the overall computational complexity is negligible.

IV. RELATION TO EBSBL

To show the relation between our method and EBSBL, we consider EBSBL with $h = 2$ and hence $g = n - 1$. In such case, $s_i = \mathbf{x}_{i-1}(2) + \mathbf{x}_i(1)$, where (\cdot) shows the entry indice

in \mathbf{x}_i . Different than EBSBL, where there is a single \mathbf{B} for each block, we use different \mathbf{B}_i s for each block. We then have $E[\mathbf{x}_i \mathbf{x}_j^H] = \delta_{i,j} \beta_i \mathbf{B}_i$ ($\delta_{i,j} = 1$ if $i = j$; otherwise $\delta_{i,j} = 0$). Note that, to avoid ambiguities, we also take $\beta_i = 1$. In our approach, we assume each s_i is correlated to the neighboring elements s_{i-1} and s_{i+1} and ignore the other correlations. By interpreting the $E[\mathbf{s} \mathbf{s}^H]$ in (22) in terms of the \mathbf{B}_i matrices our entries in the tridiagonal $\mathbf{\Gamma}$ matrix are given as follows:

$$\begin{aligned} \Gamma_{ii} &= E[s_i s_i^H] \\ &= E[(\mathbf{x}_{i-1}(2) + \mathbf{x}_i(1))(\mathbf{x}_{i-1}(2)^* + \mathbf{x}_i(1)^*)] \\ &= E[\mathbf{x}_i(1)\mathbf{x}_i(1)^* + \mathbf{x}_{i-1}(2)\mathbf{x}_{i-1}(2)^*] \\ &= \mathbf{B}_i^{(11)} + \mathbf{B}_{i-1}^{(22)} \\ \Gamma_{i(i+1)} &= E[s_i s_{i+1}^H] \\ &= E[(\mathbf{x}_{i-1}(2) + \mathbf{x}_i(1))(\mathbf{x}_i(2)^* + \mathbf{x}_{i+1}(1)^*)] \\ &= E[\mathbf{x}_i(1)\mathbf{x}_i(2)^*] = \mathbf{B}_i^{(12)} \\ \Gamma_{(i+1)i} &= \Gamma_{i(i+1)}^* \text{ for } n > i > 1, \end{aligned} \quad (38)$$

where $\mathbf{B}_i^{(\cdot)}$ corresponds to the (\cdot) th entry of the \mathbf{B}_i matrix. Here, the intermediate terms in $E[s_i s_i^H]$ and $E[s_i s_{i+1}^H]$ are $E[(\mathbf{x}_{i-1}(2)\mathbf{x}_i(1)^* + \mathbf{x}_i(1)\mathbf{x}_{i-1}(2)^*)] = 0$ and $E[(\mathbf{x}_{i-1}(2)\mathbf{x}_i(2)^* + \mathbf{x}_i(1)\mathbf{x}_{i+1}(1)^* + \mathbf{x}_{i-1}(2)\mathbf{x}_{i+1}(1)^*)] = 0$, respectively as $E[\mathbf{x}_i \mathbf{x}_j^H] = 0$ for $i \neq j$.

The equations in (38) can also be written as follows:

$$E[s_{i:i+1} s_{i:i+1}^H] = \begin{bmatrix} \mathbf{B}_i^{(11)} + \mathbf{B}_{i-1}^{(22)} & \mathbf{B}_i^{(12)} \\ \mathbf{B}_i^{(21)} & \mathbf{B}_{i+1}^{(11)} + \mathbf{B}_i^{(22)} \end{bmatrix} \quad (39)$$

where $s_{i:j}$ represents the elements of \mathbf{s} from i to j . Therefore, our model can be interpreted as an alternative to the EBSBL for $h = 2$ when there are separate correlation matrices \mathbf{B}_i for each group.

To show the equivalence between the MAP estimates of $\hat{\mathbf{s}}$ in EBSBL and in the proposed method, we give an examination of the updating rule of $\hat{\mathbf{s}}$ in (20) that comes from (18), and the updating rule of $\hat{\mathbf{s}}$ in (24), respectively. Rewriting the first part of (18) and (24) as

$$\mathbf{\Gamma} \mathbf{A}^H = \begin{bmatrix} \mathbf{B}_1^{(11)} \mathbf{a}_1^H + \mathbf{B}_1^{(12)} \mathbf{a}_2^H \\ \mathbf{B}_1^{(21)} \mathbf{a}_1^H + (\mathbf{B}_1^{(22)} + \mathbf{B}_2^{(11)}) \mathbf{a}_2^H + \mathbf{B}_2^{(12)} \mathbf{a}_3^H \\ \vdots \\ \mathbf{B}_{n-2}^{(21)} \mathbf{a}_{n-2}^H + (\mathbf{B}_{n-2}^{(22)} + \mathbf{B}_{n-1}^{(11)}) \mathbf{a}_{n-1}^H + \mathbf{B}_{n-1}^{(12)} \mathbf{a}_n^H \\ \mathbf{B}_{n-1}^{(21)} \mathbf{a}_{n-1}^H + \mathbf{B}_{n-1}^{(22)} \mathbf{a}_n^H \end{bmatrix} \quad (40)$$

$$\mathbf{\Gamma}_0 \mathbf{\Theta}^H = \begin{bmatrix} \mathbf{B}_1^{(11)} \mathbf{a}_1^H + \mathbf{B}_1^{(12)} \mathbf{a}_2^H \\ \mathbf{B}_1^{(21)} \mathbf{a}_1^H + \mathbf{B}_1^{(22)} \mathbf{a}_2^H \\ \vdots \\ \mathbf{B}_{n-1}^{(11)} \mathbf{a}_1^H + \mathbf{B}_{n-1}^{(12)} \mathbf{a}_n^H \\ \mathbf{B}_{n-1}^{(21)} \mathbf{a}_1^H + \mathbf{B}_{n-1}^{(22)} \mathbf{a}_n^H \end{bmatrix}. \quad (41)$$

where \mathbf{a}_i represents the i th column of the \mathbf{A} matrix. It is straightforward to see that $\mathbf{\Gamma} \mathbf{A}^H = \sum_{i=1}^{n-1} \mathbf{E}_i \mathbf{\Gamma}_0 \mathbf{\Theta}^H$. Now, we need to show the equivalence of the second part of (18) and (24), which is given by $(\sigma^2 \mathbf{I}_N + \mathbf{\Theta} \mathbf{\Gamma}_0 \mathbf{\Theta}^H)^{-1} \mathbf{Z}$ and $(\sigma^2 \mathbf{I}_N + \mathbf{A} \mathbf{\Gamma} \mathbf{A}^H)^{-1} \mathbf{Z}$. Using the simple diagonal structure of $\mathbf{\Gamma}_0$ we attain the following:

$$\mathbf{\Theta} \mathbf{\Gamma}_0 \mathbf{\Theta}^H = \sum_{i=1}^{n-1} (\mathbf{a}_i \mathbf{B}_i^{(11)} \mathbf{a}_i^H + \mathbf{a}_{i+1} \mathbf{B}_i^{(21)} \mathbf{a}_i^H + \mathbf{a}_i \mathbf{B}_i^{(12)} \mathbf{a}_{i+1}^H + \mathbf{a}_{i+1} \mathbf{B}_i^{(22)} \mathbf{a}_{i+1}^H), \quad (42)$$

and using the tridiagonal structure of $\mathbf{\Gamma}$ we obtain

$$\begin{aligned} \mathbf{A} \mathbf{\Gamma} \mathbf{A}^H &= \mathbf{a}_1 \mathbf{B}_1^{(11)} \mathbf{a}_1^H + \mathbf{a}_2 \mathbf{B}_1^{(21)} \mathbf{a}_1^H \\ &+ \sum_{i=1}^{n-2} (\mathbf{a}_i \mathbf{B}_i^{(12)} \mathbf{a}_{i+1}^H + \mathbf{a}_{i+1} (\mathbf{B}_i^{(22)} + \mathbf{B}_{i+1}^{(11)}) \mathbf{a}_{i+1}^H \\ &+ \mathbf{a}_{i+2} \mathbf{B}_i^{(21)} \mathbf{a}_{i+1}^H) + \mathbf{a}_{n-1} \mathbf{B}_{n-1}^{(12)} \mathbf{a}_n^H + \mathbf{a}_n \mathbf{B}_{n-1}^{(22)} \mathbf{a}_n^H. \end{aligned} \quad (43)$$

It is easy to see the equivalence of the expressions in (42) and (43). Therefore, the MAP estimates of $\hat{\mathbf{s}}$ are the same for both algorithms.

The difference between the algorithms can be seen in the update rules for \mathbf{B}_i in EBSBL and the proposed algorithm. The update rules for EBSBL in [12] are given as follows:

$$\beta_i = \text{tr}[\mathbf{B}_i^{-1} (\mathbf{\Sigma}_{\mathbf{x}_i} + \boldsymbol{\mu}_{\mathbf{x}_i} \boldsymbol{\mu}_{\mathbf{x}_i}^H)], \quad (44)$$

$$\mathbf{B} = \frac{1}{g} \sum_{i=1}^g \frac{\mathbf{\Sigma}_{\mathbf{x}_i} + \boldsymbol{\mu}_{\mathbf{x}_i} \boldsymbol{\mu}_{\mathbf{x}_i}^H}{\beta_i}. \quad (45)$$

Normally, in EBSBL, β_i is also learned during the iterations but once we assume $\beta_i = 1$ we cancel that step and only have the iterations for \mathbf{B}_i and σ^2 . Our update rule can be considered a counterpart of the case without averaging the \mathbf{B}_i s in EBSBL as follows:

$$\mathbf{B}_i = \mathbf{\Sigma}_{\mathbf{x}_i} + \boldsymbol{\mu}_{\mathbf{x}_i} \boldsymbol{\mu}_{\mathbf{x}_i}^H, \quad (46)$$

where $\mathbf{\Sigma}_{\mathbf{x}_i} \in \mathbb{C}^{2 \times 2}$ corresponds to the i th diagonal block in $\mathbf{\Sigma}_{\mathbf{x}}$ in (19) and $\boldsymbol{\mu}_{\mathbf{x}_i} \in \mathbb{C}^2$ is the i th block of $\boldsymbol{\mu}_{\mathbf{x}}$ in (18). We already showed that $\boldsymbol{\mu}_{\mathbf{s}} = \sum_{i=1}^{n-1} \mathbf{E}_i \boldsymbol{\mu}_{\mathbf{x}}$. However, now we cannot obtain the tridiagonal part of $\boldsymbol{\mu}_{\mathbf{s}} \boldsymbol{\mu}_{\mathbf{s}}^H$, which is included in our update rule in (29), by the overlapping block diagonal sum of blocks in $\boldsymbol{\mu}_{\mathbf{x}} \boldsymbol{\mu}_{\mathbf{x}}^H$ in the way that we obtained (39). Another difference between our algorithm and EBSBL is in the update of $\mathbf{\Sigma}_{\mathbf{x}}$ in (19) and $\mathbf{\Sigma}_{\mathbf{s}}$ in (25) due to the inversion terms. While $\mathbf{\Gamma}_0$ in (19) is a block diagonal matrix, $\mathbf{\Gamma}$ in (25) is a tridiagonal matrix, and relating their inverses is not straightforward.

Besides, the outmost inverse term entangles the relationship between Σ_x and Σ_s .

Note that the number of unknown variables in EBSBL when $h = 2$ is almost two times the one required in the proposed algorithm. The reduced number of unknowns enhances the performance of the proposed algorithm. Lastly, the proposed method has around $\mathcal{O}(h^3) = \mathcal{O}(2^3)$ times lower computational complexity than EBSBL.

V. NUMERICAL RESULTS

In this section, we conduct numerical experiments to evaluate the performance of the proposed algorithm in comparison with the existing literature. The performance of the algorithms is examined for both synthetic and real data. The benchmark algorithms include SBL [24], BSBL [12], EBSBL [12], CSBL [17], PCSBL [16], and the proposed method. For EBSBL, we use EBSBL-BO: the bound-optimization presented in [12], as it is used as a reference method for comparison and it is stated that it has a similar performance to EBSBL-EM: the expectation-maximization method.

A. SYNTHETIC DATA AND SYNTHETIC SYSTEM MATRIX

The measurement matrix \mathbf{A} is randomly generated with each entry independently drawn from a normal distribution, and the columns are normalized to unit norm. Likewise, the nonzero coefficients of \mathbf{s} are drawn from a complex normal distribution. Complex Gaussian white noise is added with a signal-to-noise ratio of $SNR(dB) = 20 \log_{10}(\|\mathbf{A}\mathbf{s}\|_2/\|\mathbf{n}\|_2)$.

We use the success rate and the support recovery rate under the noiseless case for performance evaluation. On the other hand, the normalized mean squared error (NMSE) is used under noisy cases. The NMSE is calculated by averaging the normalized squared errors $\|\mathbf{s} - \hat{\mathbf{s}}\|_2^2/\|\mathbf{s}\|_2^2$. The success rate is defined as the percentage of successful trials in a total of T independent runs. A successful trial is defined as one with NMSE being less than 10^{-3} . A total number of $T = 100$ independent trials are conducted. Besides, for the identification of the true support of sparse signals, we consider the ‘‘pattern recovery success rate’’. Similar to the regular success rate, it is the ratio of the number of successful trials to the total number of independent runs. However, each trial is considered successful if the support of the block-sparse signal is recovered. A coefficient whose magnitude is less than 10^{-2} is assumed as a zero coefficient for the calculation of the pattern recovery success rates (but not for the regular success rate).

For the proposed method, $\epsilon = 0.001$ and $J_{\max} = 100$ in Algorithm 1 are used in our experiments. Similarly, for PCSBL and CSBL, we used the same values. On the other hand, the probability distribution relies on hierarchical parameters a and b in $p(\Gamma|a, b)$ in PCSBL. It should be noted that the choice of the hyperparameter a of PCSBL affects dramatically the algorithm’s performance [17]. The parameter a is always set to 0.5 to achieve its best performance in the following experiments. Also, the parameter $b = 10^{-10}$. To make a fair comparison, the block size h is first set to be 2 for both BSBL

and EBSBL. However, we also consider $h = 4$ for BSBL and EBSBL, as it is used for comparison in the literature [12], [16]. Note that the algorithms are modified to handle the complex data.

For the proposed algorithm, we use the update rule in (30) with $\beta = 0.5$. For this, we observe the behavior of the Q function in (12) during EM iterations both for SBL and the proposed method with different β values. For a sample realization, the objective function increase during EM iterations for all of them is shown in Fig. 1(b), and the reconstructed data for different methods are shown in 1(a) with a noiseless case. Note that the proposed method further increases the value of the objective function compared to classical SBL for values of $\beta \in [0, 0.5]$. Since the highest increment is observed with $\beta = 0.5$ among several realizations, we select $\beta = 0.5$ to test our algorithm.

In the numerical simulations, sparse signals with dimensionality of $n = 100$ and $K = 25$ nonzero coefficients are partitioned into $L = 5$ arbitrary blocks with random sizes and arbitrary locations. For these arbitrary groups, we generate them in the same way with [16]. Here, the group sizes are likely to be higher than two and the nonzero groups are apart from each other. Hence, in this setting, the chance of there being isolated zeros and nonzeros among the groups is very small. The sources are chosen to be both uncorrelated and correlated which matches our tridiagonal correlation assumption. The correlated sources are created as $\mathbf{s} = \mathbf{R}^{1/2}\mathbf{w}$ where \mathbf{w} is complex random noise with unit variance. We choose

$$\mathbf{R} = \begin{bmatrix} 1 & c & \dots & 0 \\ c & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & \dots & 1 \end{bmatrix} \quad (47)$$

where $c = 0.3$. \mathbf{R} is chosen to be a tridiagonal correlation matrix; hence, only the neighboring correlations are considered. As a second setting, the sources are chosen to be uncorrelated to see the robustness of the algorithms with such a setting and $c = 0$.

1) PERFORMANCE IN NOISELESS ENVIRONMENTS

The success rates of the exact recovery of different algorithms the noiseless case ($SNR = 100$ dB) are provided from the viewpoint of the size of the measurements m for the purpose of comparison. Fig. 2 gives the success rates of different algorithms against the size of the measurements m and the sparsity level K for both correlated and uncorrelated data, respectively. Simulation results with correlated complex-valued data in Figs. 2(a) and (b) show that the proposed method outperforms all other methods in terms of success rate. Note that the main counterpart of our method is EBSBL ($h = 2$) and there is a significant difference between EBSBL and the proposed one in terms of success rate. Furthermore, our method has lower computational complexity. Surprisingly, CSBL’s success rate

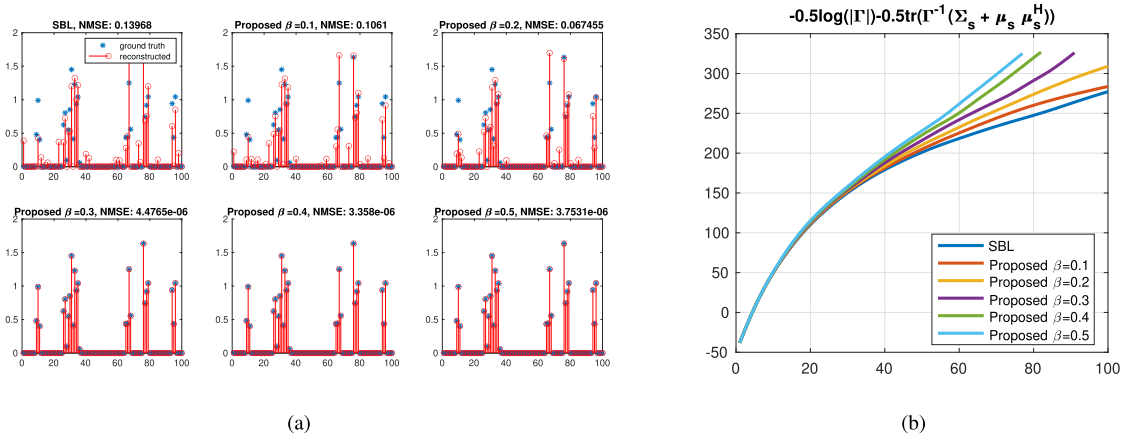


FIGURE 1. (a) Magnitudes of reconstructed group sparse data with $K = 25$, $m = 35$ and $n = 100$ for SBL and the proposed method with noiseless data and (b) the value of the Q function for SBL and the proposed one.

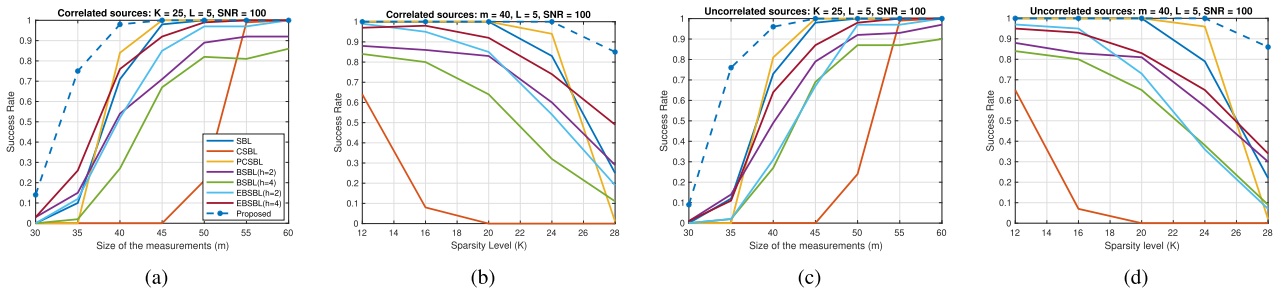


FIGURE 2. Success rate performance comparison of the sparse Bayesian learning algorithms with correlated (a) and uncorrelated (c) noiseless data for different sizes of the measurements m ; with correlated (b) and uncorrelated (d) noiseless data for different sparsity levels K .

is lower than other algorithms. It has been reported that CSBL performs similarly to PCSBL [17]. It still has good performance in terms of NMSE, but since the preselected threshold is set at 10^{-3} , its success rate is low. If we select it as 10^{-1} , the success rate of CSBL would be higher. However, for the noiseless case, 10^{-3} seems to be a good choice to evaluate the performances of the algorithms. On the other hand, BSBL might not be a suitable algorithm to test the performances of the data with varying block sizes and varying block partition locations. While BSBL performs well with the data that might fit into the considered block partition assumption, it does not perform well with the other data.

With uncorrelated data, again the proposed method has a higher success rate than the existing literature. Compared to the correlated case in Fig. 2, the EBSBL method has lower performance with uncorrelated data. This is expected as it directly uses the correlations. However, surprisingly our method's and PCSBL's performances did not change significantly. While we exploit the correlations of the neighboring elements, PCSBL assumes them uncorrelated and enforces the statistical dependence between sparsity patterns. Although these two methods utilize different motivations for correlated neighboring coefficients, they seem not to be affected by the level of the correlation.

In the support recovery rate graphs in Fig. 3, the performances of the algorithms are similar to the success rate performances given in Fig. 2. Although we observe small variations over the methods, again the proposed method outperforms the existing literature.

2) PERFORMANCE IN NOISY ENVIRONMENTS

The performances of different algorithms in terms of the NMSE against m and the sparsity level K are given in Fig. 4. Under noisy environments, the NMSEs of all algorithms consistently decrease as the size of the measurements m increases and the sparsity degree K decreases as shown in Fig. 4, where PCSBL with $a = 0.5$ achieves the lowest NMSE and proposed method achieves lower performance than the best performance of PCSBL. Note that the tested data consists of sparse groups and the chance of isolated zeros and nonzero elements in the data is very small. Different from the noiseless case, CSBL is better at dealing with noise than other correlation-based methods. However, in noiseless cases, it does not provide a perfect estimation of \mathbf{s} while other methods can produce a very low reconstruction error and high success rate.

Illustrative examples of sparse coefficient recovery of different algorithms with the size of the measurements m being

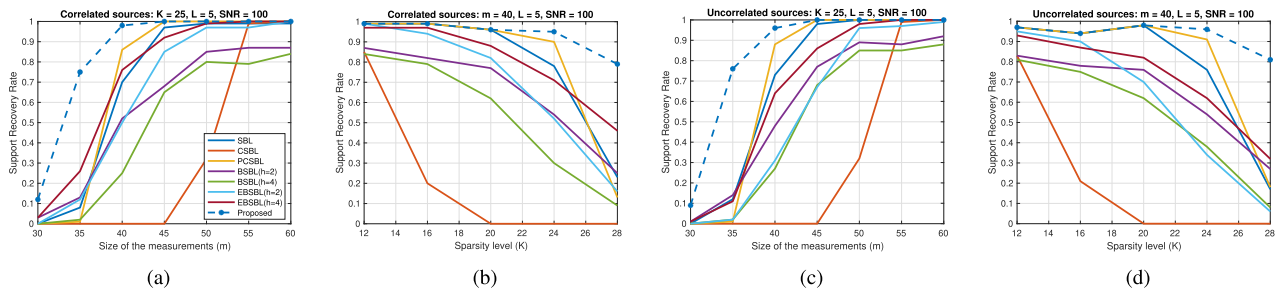


FIGURE 3. Support recovery rate performance comparison of the sparse Bayesian learning algorithms with correlated (a) and uncorrelated (c) noiseless data for different sizes of the measurements m ; With correlated (b) and uncorrelated (d) noiseless data for different sparsity levels K .

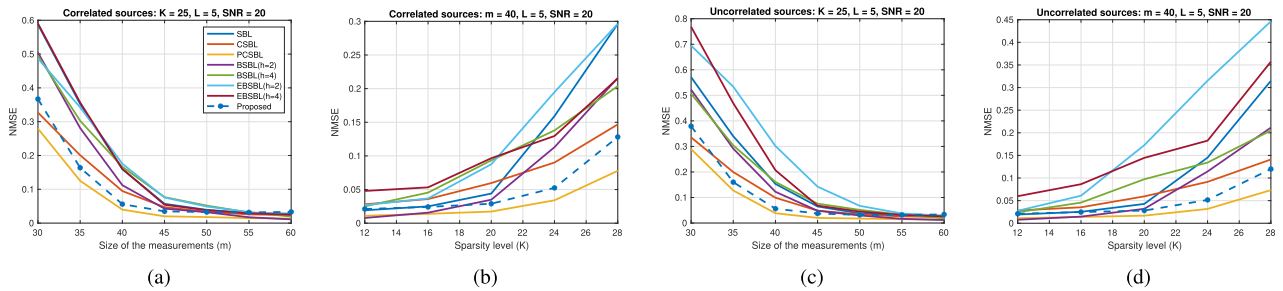


FIGURE 4. NMSE performance comparison of the sparse Bayesian learning algorithms with correlated (a) and uncorrelated (c) noisy data (20 dB SNR) for different sizes of the measurements m ; with correlated (b) and uncorrelated (d) noisy data (20 dB SNR) for different sparsity levels K .

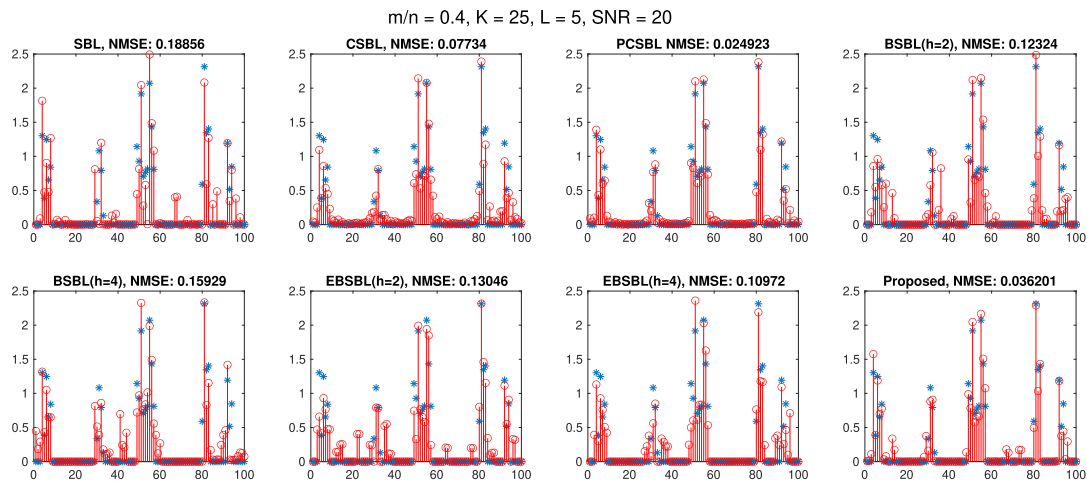


FIGURE 5. Magnitudes of the reconstructions of SBL algorithms with group sparse noisy data (20 dB SNR).

40 are given in Fig. 5, respectively. PCSBL and our proposed method provide the most accurate estimates of the original sparse coefficients with fewer measurements, especially for those significant elements inside blocks. By closely looking at CSBL and PCSBL, we observe smooth decay around the corners of the groups. This effect is stronger in CSBL due to the modeling difference between CSBL and PCSBL. On the other hand, although the proposed method is able to reconstruct sharp edges, the proposed one and other correlation-based methods suffer from reconstructing some off-group elements.

This is possibly arising from the effect of the noise correlation. In CSBL and PCSBL, we observe smooth edges and the boundaries of the groups are indistinct but we do not observe reconstructed elements that are not close to the groups.

In addition to testing with group sparse data, we test our algorithm for a new setting where each group has an isolated zero element inside the group. For such a scenario, the proposed method achieves the best performance in terms of NMSE with 20 dB SNR as shown in Figs. 6(a) and (b) for varying numbers of m and K respectively. The

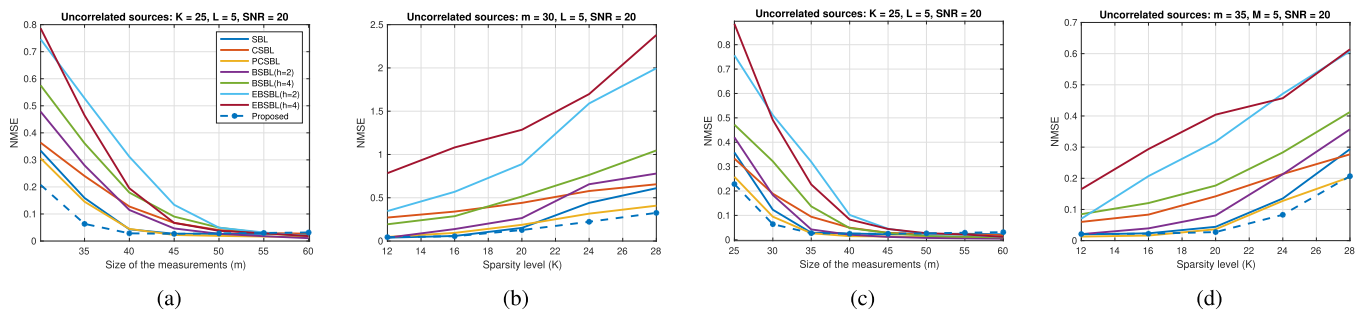


FIGURE 6. NMSE performance comparison of the SBL algorithms with the data that has isolated zeros for different (a) size of the measurements m ; (b) Sparsity level K and NMSE performance comparison of the data with mixed groups, isolated zeros, and isolated nonzeros for different (c) size of the measurements m ; (d) Sparsity levels K under 20 dB SNR.

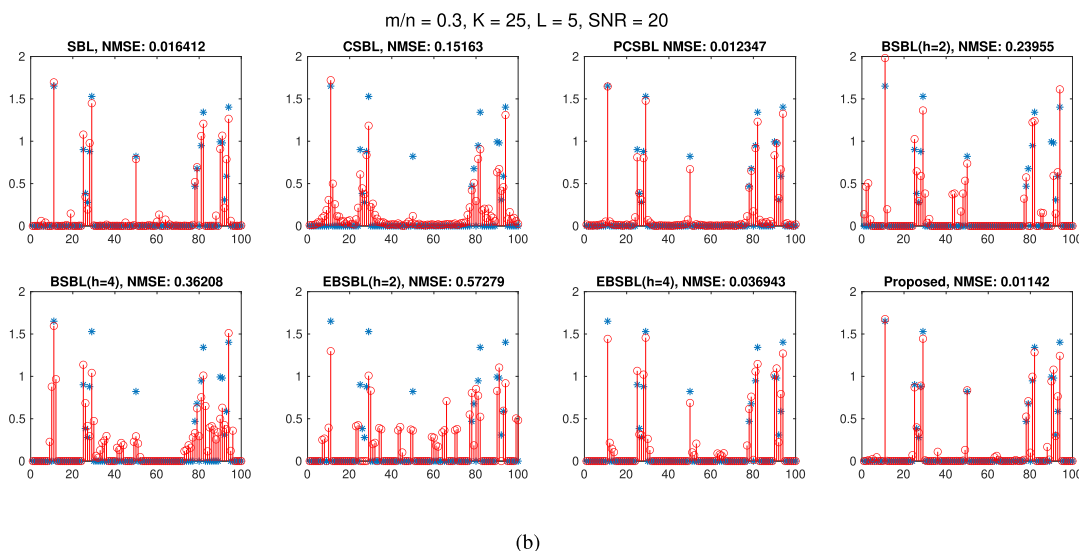
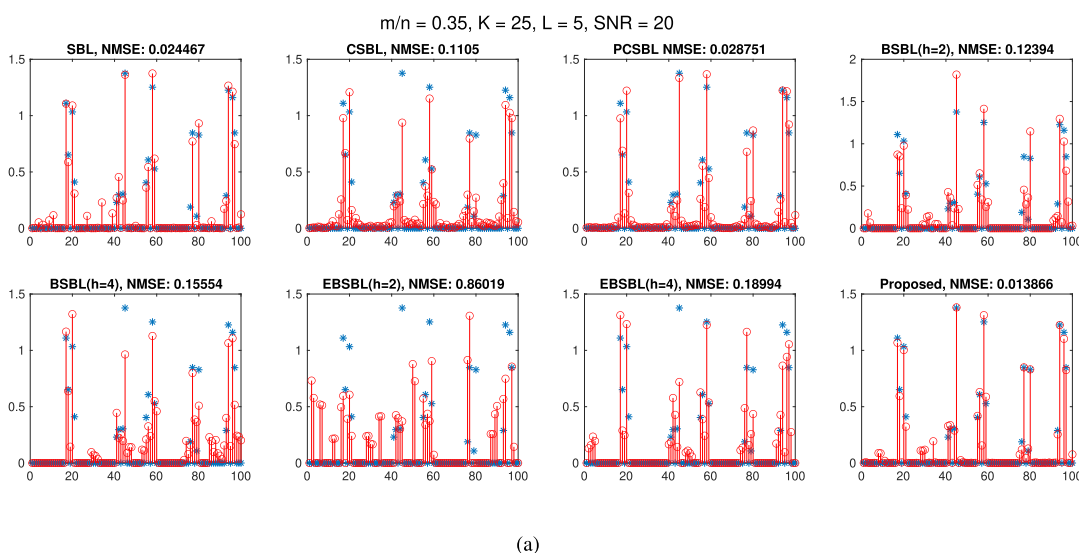


FIGURE 7. Magnitudes of the reconstructions of SBL algorithms with (a) noisy data (20 dB SNR) which has isolated zeros in each group (b) the data with mixed groups, isolated zeros, and isolated nonzeros.

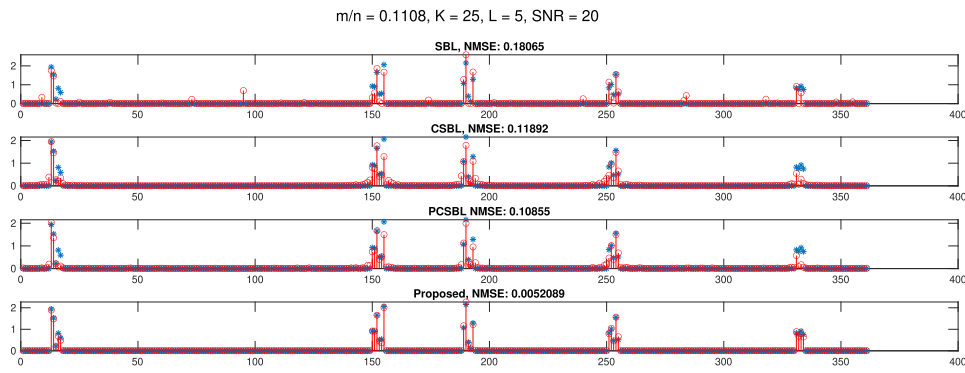


FIGURE 8. Magnitudes of the reconstructions of SBL algorithms with noisy group sparse data (20 dB SNR) for DOA and amplitude estimation.

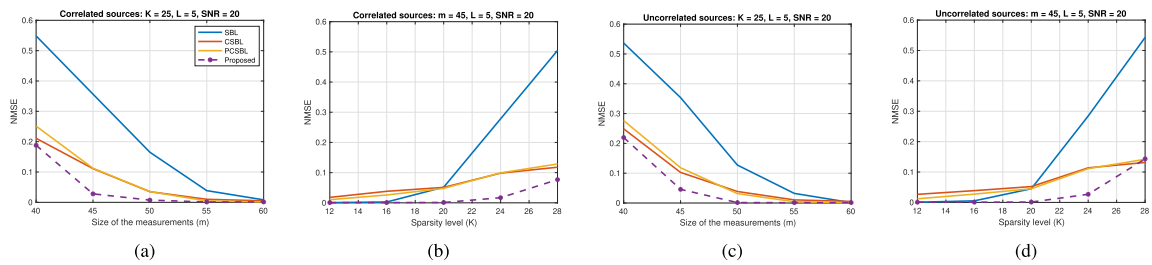


FIGURE 9. NMSE performance comparison of the SBL algorithms with correlated (a) and uncorrelated (c) noisy data (20 dB SNR) for different sizes of the measurements m ; with correlated (b) and uncorrelated (d) noisy data (20 dB SNR) for different sparsity levels K for a single snapshot problem.

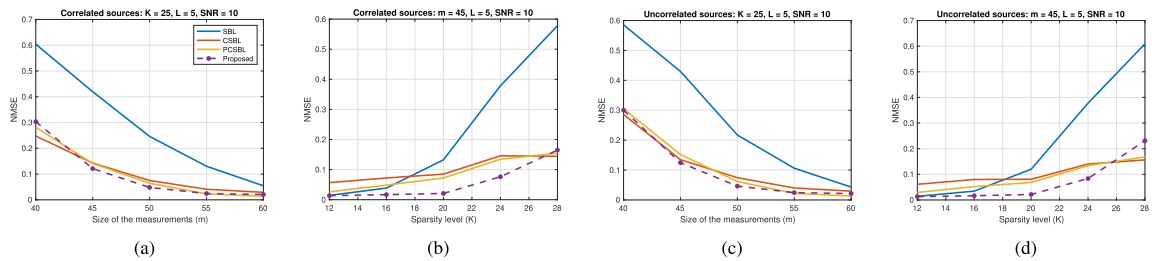


FIGURE 10. NMSE performance comparison of the SBL algorithms with correlated (a) and uncorrelated (c) noisy data (10 dB SNR) for different sizes of the measurements m ; with correlated (b) and uncorrelated (d) noisy data (10 dB SNR) for different sparsity levels K for a single snapshot problem.

reconstructions; are shown in Fig. 7(a) for a single realization. As CSBL and PCSBL algorithms consider contributions from the neighboring elements, the isolated zero element is affected by the surrounding nonzeros. On the other hand, the proposed algorithm considers the correlation with the neighbors in a multiplication form. When the sub-diagonal elements become zero they do not negatively affect the reconstruction of the isolated zero elements. Note that with such data, the classical SBL algorithm is also performing better than most of the other group sparsity-based algorithms. For BSBL and EBSBL, the performance varies significantly depending on different Monte Carlo realizations. Again it possibly arises from the block partition assumption of the algorithms.

We also tested these algorithms with a mixed scenario which consists of two isolated zero elements, two full group

elements, and one group with an isolated nonzero element. As shown in Figs. 6(c) and (d), we attain the best performance with the proposed method as it tackles the isolated nonzeros and isolated zeros more successfully compared to the existing literature. As the groups do not have a specific size, other correlation-based algorithms such as BSBL and EBSBL might fail to reconstruct the groups. We again demonstrate the reconstructions in Fig. 7(b) and observe a similar effect.

B. DOA AND AMPLITUDE ESTIMATION WITH SYNTHETIC DATA

Now, the proposed algorithm is tested on a DOA and amplitude estimation problem with block sparse sources. It should be noted that most analysis in the literature has been done with randomly designed sensing matrices [12], [16]. However, such

a random design is not realistic to evaluate the performance of these algorithms. Hence, we tested and compared their performance for a simple DOA and amplitude estimation problem. Here, we consider an array with various numbers of elements. The DOAs are on an angular grid $[-90 : 0.5 : 90]^\circ$, and $m = 361$. The noise is modeled as i.i.d. complex Gaussian. Here, we examine a scenario with $K = 25$ random sources in $L = 5$ random groups. DOA groups collect s_l values having random complex amplitudes. The sources are chosen to be correlated and $c = 0.5$.

The performances of different algorithms in terms of the NMSE against the size of the measurements m and the sparsity level K are given in Figs. 9 and 10 for 20 dB SNR and 10 dB SNR, respectively. With such a realistic setting where the system model \mathbf{A} is a realistic matrix, the best performance is attained by the proposed algorithm, especially for the 20 dB SNR case. The performance of the proposed algorithm is still comparable to or better than the state-of-the-art under the 10 dB SNR case. Note that the DOA matrix \mathbf{A} has a higher coherence compared to a randomly generated \mathbf{A} matrix. Illustrative examples of the sparse coefficient recovery of different algorithms with the size of the measurements m being 40 are given in Fig. 8. Here, the proposed method provides the most accurate estimates of the original sparse coefficients. By closely looking at CSBL and PCSBL, we again observe a smooth decay around the corners of the groups. On the other hand, the proposed method is able to reconstruct sharp edges.

VI. CONCLUSION AND DISCUSSION

In this paper, we have proposed a correlated sparse Bayesian learning algorithm for block sparse signals with arbitrary block sizes and locations under the Bayesian framework. This is a simpler alternative to EBSBL and we explain the underlying relationship between the proposed method and a particular case of EBSBL. The proposed algorithm uses the fact that immediate neighboring sparse coefficients are correlated. Unlike the diagonal correlation matrix in conventional SBL, the unknown correlation matrix has a tridiagonal structure to capture the correlation with neighbors. Due to the entanglement of the elements in the inverse tridiagonal matrix, instead of a direct closed-form solution, an approximate solution is proposed. The sparse reconstruction performance of the algorithm is evaluated with both correlated and uncorrelated block sparse coefficients. Results of comprehensive simulations demonstrate that the proposed algorithm outperforms CSBL and PCSBL and other correlation-based methods such as EBSBL in terms of reconstruction quality. The numerical results also show that the proposed correlated SBL algorithm is capable of recovering signals with both block patterns and isolated coefficients.

REFERENCES

- [1] L. Wang, L. Zhao, G. Bi, C. Wan, and L. Yang, "Enhanced ISAR imaging by exploiting the continuity of the target scene," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 9, pp. 5736–5750, Sep. 2014.

- [2] L. Wang, L. Zhao, G. Bi, and C. Wan, "Sparse representation-based ISAR imaging using Markov random fields," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 8, pp. 3941–3953, Aug. 2015.
- [3] L. Zhao, L. Wang, L. Yang, A. M. Zoubir, and G. Bi, "The race to improve radar imagery: An overview of recent progress in statistical sparsity-based techniques," *IEEE Signal Process. Mag.*, vol. 33, no. 6, pp. 85–102, Nov. 2016.
- [4] Z. Zhang, T.-P. Jung, S. Makeig, and B. D. Rao, "Compressed sensing for energy-efficient wireless telemonitoring of noninvasive fetal ECG via block sparse Bayesian learning," *IEEE Trans. Biomed. Eng.*, vol. 60, no. 2, pp. 300–309, Feb. 2013.
- [5] O. Lorintiu, H. Liebgott, and D. Friboulet, "Compressed sensing doppler ultrasound reconstruction using block sparse Bayesian learning," *IEEE Trans. Med. Imag.*, vol. 35, no. 4, pp. 978–987, Apr. 2016.
- [6] R. G. Baraniuk, V. Cevher, M. F. Duarte, and C. Hegde, "Model-based compressive sensing," *IEEE Trans. Inf. Theory*, vol. 56, no. 4, pp. 1982–2001, Apr. 2010.
- [7] Y. C. Eldar, P. Kuppinger, and H. Bolcskei, "Block-sparse signals: Uncertainty relations and efficient recovery," *IEEE Trans. Signal Process.*, vol. 58, no. 6, pp. 3042–3054, Jun. 2010.
- [8] X. Lv, G. Bi, and C. Wan, "The group lasso for stable recovery of block-sparse signal representations," *IEEE Trans. Signal Process.*, vol. 59, no. 4, pp. 1371–1382, Apr. 2011.
- [9] E. v. d. Berg and M. P. Friedlander, "Probing the pareto frontier for basis pursuit solutions," *SIAM J. Sci. Comput.*, vol. 31, no. 2, pp. 890–912, 2009.
- [10] Y. C. Eldar and M. Mishali, "Robust recovery of signals from a structured union of subspaces," *IEEE Trans. Inf. Theory*, vol. 55, no. 11, pp. 5302–5316, Nov. 2009.
- [11] Z. Zhang and B. D. Rao, "Sparse signal recovery with temporally correlated source vectors using sparse Bayesian learning," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 5, pp. 912–926, Sep. 2011.
- [12] Z. Zhang and B. D. Rao, "Extension of SBL algorithms for the recovery of block sparse signals with intra-block correlation," *IEEE Trans. Signal Process.*, vol. 61, no. 8, pp. 2009–2015, Apr. 2013.
- [13] L. Yu, H. Sun, J. P. Barbot, and G. Zheng, "Bayesian compressive sensing for clustered sparse signals," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2011, pp. 3948–3951.
- [14] A. Dremeau, C. Herzet, and L. Daudet, "Boltzmann machine and mean-field approximation for structured sparse decompositions," *IEEE Trans. Signal Process.*, vol. 60, no. 7, pp. 3425–3438, Jul. 2012.
- [15] C. Andrieu, N. d. Freitas, A. Doucet, and M. I. Jordan, "An introduction to MCMC for machine learning," *Mach. Learn.*, vol. 50, no. 1/2, pp. 5–43, 2003.
- [16] J. Fang, Y. Shen, H. Li, and P. Wang, "Pattern-coupled sparse Bayesian learning for recovery of block-sparse signals," *IEEE Trans. Signal Process.*, vol. 63, no. 2, pp. 360–372, Jan. 2015.
- [17] L. Wang, L. Zhao, S. Rahardja, and G. Bi, "Alternative to extended block sparse Bayesian learning and its relation to pattern-coupled sparse Bayesian learning," *IEEE Trans. Signal Process.*, vol. 66, no. 10, pp. 2759–2771, May 2018.
- [18] A. Sant, M. Leinonen, and B. D. Rao, "Block-sparse signal recovery via general total variation regularized sparse Bayesian learning," *IEEE Trans. Signal Process.*, vol. 70, pp. 1056–1071, 2022.
- [19] H. Duan, L. Zhang, J. Fang, L. Huang, and H. Li, "Pattern-coupled sparse Bayesian learning for inverse synthetic aperture radar imaging," *IEEE Signal Process. Lett.*, vol. 22, no. 11, pp. 1995–1999, Nov. 2015.
- [20] J. Fang, L. Zhang, and H. Li, "Two-dimensional pattern-coupled sparse Bayesian learning via generalized approximate message passing," *IEEE Trans. Image Process.*, vol. 25, no. 6, pp. 2920–2930, Jun. 2016.
- [21] J. Dai, A. Liu, and H. C. So, "Non-uniform burst-sparsity learning for massive MIMO channel estimation," *IEEE Trans. Signal Process.*, vol. 67, no. 4, pp. 1075–1087, Feb. 2019.
- [22] A. Rajoriya, A. Kumar, and R. Budhiraja, "Covariance-free variational Bayesian learning for correlated block sparse signals," *IEEE Commun. Lett.*, vol. 27, no. 3, pp. 966–970, Mar. 2023.
- [23] R. K. Chakraborty, G. Joseph, and C. R. Murthy, "Joint state and input estimation for linear dynamical systems with sparse control," 2023, *arXiv:2312.02082*.
- [24] M. E. Tipping, "Sparse Bayesian learning and the relevance vector machine," *J. Mach. Learn. Res.*, vol. 1, pp. 211–244, 2001.

- [25] P. Gerstoft, C. F. Mecklenbräuker, A. Xenaki, and S. Nannuru, "Multisnapshot sparse Bayesian learning for DOA," *IEEE Signal Process. Lett.*, vol. 23, no. 10, pp. 1469–1473, Oct. 2016.
- [26] G. Meurant, "A review on the inverse of symmetric tridiagonal and block tridiagonal matrices," *SIAM J. Matrix Anal. Appl.*, vol. 13, no. 3, pp. 707–728, 1992.
- [27] S. i. Amari, "Backpropagation and stochastic gradient descent method," *Neurocomputing*, vol. 5, no. 4, pp. 185–196, 1993.
- [28] H. F. Walker and P. Ni, "Anderson acceleration for fixed-point iterations," *SIAM J. Numer. Anal.*, vol. 49, no. 4, pp. 1715–1735, 2011.
- [29] M. Anelić and C. d. Fonseca, "Sufficient conditions for positive definiteness of tridiagonal matrices revisited," *Positivity*, vol. 15, pp. 155–159, 2011.
- [30] R. M. Neal and G. E. Hinton, *A View of the EM Algorithm That Justifies Incremental, Sparse, and Other Variants*. Dordrecht, The Netherlands: Springer, 1998, pp. 355–368.
- [31] D. J. Tylavsky and G. R. Sohie, "Generalization of the matrix inversion lemma," *Proc. IEEE*, vol. 74, no. 7, pp. 1050–1052, Jul. 1986.



DIDEM DOGAN received the M.Sc. degree in electrical engineering from Middle East Technical University, Ankara, Turkiye, in 2020. She is currently working toward the Ph.D. degree with the Signal Processing Systems Department, Delft university of Technology, Delft, The Netherlands. Her research interests include signal processing, with an emphasis in computational imaging and particularly the ultrasound imaging.



GEERT LEUS (Fellow, IEEE) received the M.Sc. and Ph.D. degrees in electrical engineering from the KU Leuven, Leuven, Belgium, in 1996 and 2000, respectively. He is currently a Full Professor with the Faculty of Electrical Engineering, Mathematics and Computer Science, Delft University of Technology, Delft, The Netherlands. He was the recipient the 2021 EURASIP Individual Technical Achievement Award, 2005 IEEE Signal Processing Society Best Paper Award, and 2002 IEEE Signal Processing Society Young Author Best Paper

Award. He was also the Member-at-Large of the Board of Governors of the IEEE Signal Processing Society, Chair of the IEEE Signal Processing for Communications and Networking Technical Committee, Chair of the EURASIP Technical Area Committee on Signal Processing for Multisensor Systems and an Editor in Chief of *EURASIP Journal on Advances in Signal Processing*. He is currently an Editor in Chief of *EURASIP Signal Processing*. He is the Fellow of EURASIP.