# AN ITERATIVE BAYESIAN ALGORITHM FOR BLOCK-SPARSE SIGNAL RECONSTRUCTION

*M. Korki*[†] *J. Zhang*[†] *C. Zhang*[†] *H. Zayyani*[⋆]

[†]Faculty of Science, Engineering and Technology, Swinburne Univ. of Technology, Hawthorn, Australia
[⋆]Department of Electrical and Computer Engineering, Qom Univ. of Technology, Qom, Iran

## ABSTRACT

This paper presents a novel iterative Bayesian algorithm, Block Iterative Bayesian Algorithm (Block-IBA), for reconstructing block-sparse signals with unknown block structures. Unlike the other existing algorithms for block sparse signal recovery which assume the cluster structure of the non-zero elements of the unknown signal to be independent and identically distributed (i.i.d.), we use a more realistic Bernoulli-Gaussian hidden Markov model (BGHMM) to capture the burstiness (block structure) of the impulsive noise in practical applications such as Power Line Communication (PLC). The Block-IBA iteratively estimates the amplitudes and positions of the block-sparse signal based on Expectation-Maximization (EM) algorithm which is also optimized with the steepest-ascent method. Simulation results show the effectiveness of our algorithm for block-sparse signal recovery.

***Index Terms***— Block-sparse, iterative Bayesian algorithm, steepest-ascent.

## 1. INTRODUCTION

Compressed sensing aims to recover the sparse signal from underdetermined systems of linear equations. If the structure of the signal is exploited, the better recovery performance can be achieved. A block-sparse signal, in which the nonzero samples manifest themselves as clusters, is one of the important structured sparsity. Block-sparsity has a wide range of applications in multiband signals [1], audio signals [2], structured compressed sensing [3], and the multiple measurement vector (MMV) model [4].

The mathematical model of block-sparse signal reconstruction is given by

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

where $\mathbf{\Phi} \in \mathbb{R}^{N\times M}$ $(N < M)$ is a known measurement matrix, $\mathbf{y} \in \mathbb{R}^N$ is the available measurement vector, and $\mathbf{n} \in \mathbb{R}^N$ is the corrupting noise. we aim to estimate the original unknown signal $\mathbf{w} \in \mathbb{R}^M$ with the cluster structure

$$\mathbf{w} = [\underbrace{w_1, \ldots, w_{d_1}}_{\mathbf{w}^T[1]}, \ldots, \underbrace{w_{d_{g-1}+1}, \ldots, w_{d_g}}_{\mathbf{w}^T[g]}]^T \tag{2}$$

where $\mathbf{w}[i]$ denotes the $i$th block with length $d_i$ which are not necessarily identical. In the block partition (2), only $k \ll g$ vectors $\mathbf{w}[i]$ have nonzero Euclidean norm. Given the *a priori* knowledge of block partition, a few algorithms such as Block-OMP [5], mixed $\ell^2/\ell^1$norm-minimization [6], group LASSO [7] and model-based CoSaMP [8], work effectively in the block-sparse signal recovery. These algorithms require the knowledge of the block structure (e.g. the location and the lengths of the blocks). However, in many applications, the prior knowledge about the block structure is often unavailable. For instance, the accurate tree structure of the coefficients for the clustered sparse representation of the images is unknown *a priori*, and the impulsive noise in Power Line Communication (PLC) occures in bursts with the unknown locations and lengths [9]. Hence, designing an adaptive method for estimating the block partition and recovering the clustered-sparse signal simoltanously remains a challenge. To address this challenge, some algorithms requiring very little *a priori* information have been proposed recently [10]-[13]. However, all these algorithms use the i.i.d. model to describe the cluster structure of the non-zero elements of the unknown signal, which restricts their applicability and performance because many practically important signals, e.g. the impulsive noise in PLC, do not satisfy the i.i.d. condition. It is therefore necessary to develop reconstruction algorithms for block-sparse signals using more realistic signal model.

In this paper, we propose a novel iterative Bayesian algorithm (Block-IBA) using a Bernoulli-Gaussian hidden Markov model (BGHMM) [9] for the block-sparse signals. This model better captures the burstiness (block structure) of the impulsive noise and hence is more realistic for practical applications such as PLC.

The proposed Block-IBA reconstructs the supports and the amplitudes of block-sparse signal $\mathbf{w}$ using an expectation maximization (EM) algorithm when its block structure is completely unknown. In the expectation step (E-step) the amplitudes of the signal $\mathbf{w}$ are estimated iteratively whereas in the maximization step (M-step) the supports of the signal $\mathbf{w}$ are estimated iteratively. To this end, we utilize a steepest-ascent algorithm after converting the estimation problem of discrete supports to a continuous maximization problem. Although using the steepest-ascent algorithm exists in the literature for recovering the sparse signals (e.g. [14]), investigation of this method is unavailable in the literature of block-sparse signal recovery. As a result the proposed Block-IBA algorithm offers more reconstruction accuracy than the existing state-of-the-art algorithms for the block-sparse signals which are not i.i.d. This is verified by the experimental results, where the block-sparse signal comprises a large number of narrow blocks.

## 2. SIGNAL MODEL

In this paper, the linear model of (1) is considered as the measurement process. The measurement matrix $\mathbf{\Phi}$ is assumed known beforehand and also its columns are normalized to have unit norms. Furthermore, we model the noise in model (1) as a stationary, additive white Gaussian noise (AWGN) process, with $\mathbf{n} \sim \mathcal{N}\left(0, \sigma_n^2 \mathbf{I}_N\right)$. To model the block-sparse sources $\mathbf{w}$, we introduce two hidden random processes, $\mathbf{s}$ and $\boldsymbol{\theta}$ [14]–[16]. The binary vector $\mathbf{s} \in \{0,1\}^M$ describes the support of $\mathbf{w}$, denoted $\mathcal{S}$, while the vector $\boldsymbol{\theta} \in \mathbb{R}^M$ represents the amplitudes of the active elements of $\mathbf{w}$. Hence, each element of the source vector $\mathbf{w}$ can be characterized as

$$w_i = s_i \cdot \theta_i \tag{3}$$

where $s_i = 0$ results in $w_i = 0$ and $i \notin \mathcal{S}$, while $s_i = 1$ results in $w_i = \theta_i$ and $i \in \mathcal{S}$. Hence, in vector form we can show that

$$\mathbf{w} = \mathbf{S}\boldsymbol{\theta}, \qquad \mathbf{S} = \mathrm{diag}(\mathbf{s}) \in \mathbb{R}^{M \times M}. \tag{4}$$

To model the block-sparsity of the source vector $\mathbf{w}$, we assume that $\mathbf{s}$ is a stationary first-order Markov process defined by two transition probabilities: $p_{10} \triangleq \Pr\{s_{i+1} = 1 | s_i = 0\}$ and $p_{01} \triangleq \Pr\{s_{i+1} = 0 | s_i = 1\}$ [15]. Also, in the steady-state operation of Markov process we assume $p \triangleq \Pr\{s_i = 0\}$. Therefore, the two parameters $p$ and $p_{10}$ completely describe the state process of the Markov chain. As a result, the remaining transition probability can be determined as $p_{01} = \frac{p \cdot p_{10}}{(1-p)}$. The length of the blocks of the block-sparse signal is determined by parameter $p_{01}$, namely, the average number of consecutive samples of ones is specified by $1/p_{01}$ in the Markov chain.

Because the $\mathbf{s}$ vector is a stationary first-order Markov process with the two transition probabilities $p_{10}$ and $p_{01}$, $p(\mathbf{s})$ is given by

$$p(\mathbf{s}) = p(s_1) \prod_{i=1}^{M-1} p(s_{i+1}|s_i) \tag{5}$$

where $p(s_1) = p^{(1-s_1)}(1-p)^{s_1}$ and

$$p(s_{i+1}|s_i) = \begin{cases} (1-p_{10})^{(1-s_{i+1})} p_{10}^{s_{i+1}}, & s_i = 0, \\ p_{01}^{(1-s_{i+1})} (1-p_{01})^{s_{i+1}}, & s_i = 1. \end{cases} \tag{6}$$

Note that the amplitude vector $\boldsymbol{\theta}$ has also a Gaussian distribution with $\boldsymbol{\theta} \sim \mathcal{N}\left(0, \sigma_\theta^2 \mathbf{I}_M\right)$. Therefore, from (3) it is obvious that $p(w_i|s_i, \theta_i) = \delta(w_i - s_i\theta_i)$, where $\delta(\cdot)$ is the Dirac delta function. If we follow the marginalization rule to remove $s_i$ and $\theta_i$, we can find the probability density of the sources as

$$p(w_i) = p\delta(w_i) + (1-p)\mathcal{N}\left(w_i; 0, \sigma_\theta^2\right) \tag{7}$$

where $\sigma_\theta^2$ is the variance of $\boldsymbol{\theta}$. Equation (7) shows that the distribution of the sources is a Bernoulli-Gaussian hidden Markov model (BGHMM) which is utilized to implicitly express the block sparsity of the signal model due to the point-mass distribution at $w_i = 0$ and the hidden variables $s_i$.

## 3. BLOCK ITERATIVE BAYESIAN ALGORITHM

In this section, we propose an algorithm to estimate the unknown original signal $\mathbf{w}$ by estimating its components ($\mathbf{s}$ and $\boldsymbol{\theta}$ in (4)) iteratively. We follow a two-step approach to estimate the $\mathbf{w}$ vector.

In the first step, we estimate the amplitude vector $\boldsymbol{\theta}$ based on the known estimation of $\mathbf{s}$ vector and the mixing observation vector $\mathbf{y}$. We call this expectation step (E-step). Hence, if we assume that the estimated vector $\mathbf{s}$, referred to as $\widehat{\mathbf{s}}$, is known, we obtain the MMSE estimation of $\boldsymbol{\theta}$ (referred to as $\widehat{\boldsymbol{\theta}}$) which is given as

$$\widehat{\boldsymbol{\theta}} = \sigma_\theta^2 \widehat{\mathbf{S}} \mathbf{\Phi}^T \mathbf{\Sigma}_{\widehat{\mathbf{s}}}^{-1} \mathbf{y} \tag{8}$$

where

$$\mathbf{\Sigma}_{\widehat{\mathbf{s}}} = E\left[\mathbf{y}\mathbf{y}^T \mid \widehat{\mathbf{s}}\right] = \sigma_n^2 \mathbf{I}_N + \sigma_\theta^2 \mathbf{\Phi} \widehat{\mathbf{S}} \mathbf{\Phi}^T. \tag{9}$$

Following the Sparse Bayesian Learning (SBL) framework [17], we consider a Gaussian prior distribution for amplitude vector $\boldsymbol{\theta}$, i.e. $p(\boldsymbol{\theta}; \gamma_i) \sim \mathcal{N}\left(0, \mathbf{\Sigma_0^{-1}}\right)$ where $\mathbf{\Sigma_0} = \mathrm{diag}(\gamma_1, \gamma_2, \cdots, \gamma_M)$. Furthermore, $\gamma_i$ are the non-negative elements of the hyperparameter vector $\boldsymbol{\gamma}$, that is, $\boldsymbol{\gamma} \triangleq \{\gamma_i\}$. We assume Gamma distributions as hyperpriors over the hyperparameters $\{\gamma_i\}$:

$$p(\boldsymbol{\gamma}) = \prod_{i=1}^{M} \mathrm{Gamma}(\gamma_i \mid a, b) = \prod_{i=1}^{M} \Gamma(a)^{-1} b^a \gamma_i^{a-1} e^{-b\gamma_i}$$

where $\Gamma(a) = \int_0^\infty t^{a-1} e^{-t} dt$ is the Gamma function and $a = b = 10^{-4}$. From (4), we can rewrite the linear model of (1) as $\mathbf{y} = \mathbf{\Phi}\mathbf{S}\boldsymbol{\theta} + \mathbf{n} = \mathbf{\Psi}\boldsymbol{\theta} + \mathbf{n}$ where $\mathbf{\Psi} = \mathbf{\Phi}\mathbf{S}$. It can be shown that the posterior $p\left(\boldsymbol{\theta} \mid \mathbf{y}; \boldsymbol{\gamma}, \sigma_n^2\right)$ is a multivariate Gaussian with the following mean and covariance

$$\boldsymbol{\mu}_\theta = \sigma_n^{-2} \mathbf{\Sigma}_\theta \mathbf{\Psi}^T \mathbf{y} \tag{10}$$

$$\mathbf{\Sigma}_\theta = \left(\sigma_n^{-2} \mathbf{\Psi}^T \mathbf{\Psi} + \mathbf{\Sigma}_0\right)^{-1}. \tag{11}$$

Hence, the MAP estimate of $\boldsymbol{\theta}$ is

$$\widehat{\boldsymbol{\theta}} = \left(\widehat{\boldsymbol{\Psi}}^T\widehat{\boldsymbol{\Psi}} + \sigma_n^2\boldsymbol{\Sigma}_0\right)^{-1}\widehat{\boldsymbol{\Psi}}^T\mathbf{y} \tag{12}$$

where $\widehat{\boldsymbol{\Psi}} = \boldsymbol{\Phi}\widehat{\mathbf{S}}$. Using the EM algorithm [17], an estimate for hyperparameter $\boldsymbol{\gamma}$ is $\hat{\gamma}_i = \frac{1+2a}{(\hat{\mu}_{\theta,i})^2+\hat{\Sigma}_{\theta,ii}+2b}$ where $\hat{\mu}_{\theta,i}$ denotes the $i$th entry of $\boldsymbol{\mu_\theta}$ in (10) and $\hat{\Sigma}_{\theta,ii}$ denotes the $i$th diagonal element of the covariance matrix $\boldsymbol{\Sigma_\theta}$ in (11).

We call the second step of our approach Maximization step (M-step). In this step, we find the estimate of $\mathbf{s}$ with the assumption of known vector $\widehat{\boldsymbol{\theta}}$ and the observation vector $\mathbf{y}$. Therefore, we can write the MAP estimate of $\mathbf{s}$ as

$$\widehat{\mathbf{s}} = \underset{\mathbf{s}}{\operatorname{argmax}}\,\mathcal{L}(\mathbf{s}) = \underset{\mathbf{s}}{\operatorname{argmax}}(\log(p(\mathbf{s})p(\mathbf{y} \mid \mathbf{s}, \widehat{\boldsymbol{\theta}}))). \tag{13}$$

The maximization is performed over all $2^M$ possible sets of $\mathbf{s}$ vectors, which is intractable when $M$ is large. This is because the computation should be done over the discrete space. One way around this exhaustive search is to convert the maximization problem into a continuous form. Therefore, in this section we propose a method to convert the problem into a continuous maximization and apply a steepest-ascent algorithm to find the maximum value. To this end, we model the elements of $\mathbf{s}$ vector as a Gaussian Mixture (GM) with two Gaussian variables centered around 0 and 1 with sufficiently small variances. Hence, each discrete element of $\mathbf{s}$ vector, i.e. $s_i$ can be given as

$$p(s_1) \approx p\,\mathcal{N}\left(0, \sigma_0^2\right) + (1-p)\,\mathcal{N}\left(1, \sigma_0^2\right), \tag{14}$$

Moreover, the other elements of $\mathbf{s}$ vector, i.e. $s_{i+1}$ ($i = 1, \cdots, M-1$) can be expressed as

$$p(s_{i+1}) \approx \begin{cases} (1-p_{10})\mathcal{N}\left(0,\sigma_0^2\right) + p_{10}\mathcal{N}\left(1,\sigma_0^2\right), & s_i = 0, \\ p_{01}\mathcal{N}\left(0,\sigma_0^2\right) + (1-p_{01})\mathcal{N}\left(1,\sigma_0^2\right), & s_i = 1. \end{cases} \tag{15}$$

In order to find the global maximum of (13) we decrease the variance $\sigma_0^2$ in each iteration of the algorithm gradually, which averts the local maximum of (13). To calculate the log posterior $\mathcal{L}(\mathbf{s})$ in (13), we calculate the prior probability $p(\mathbf{s})$ and the likelihood $p\left(\mathbf{y} \mid \mathbf{s}, \widehat{\boldsymbol{\theta}}\right)$ which are given as

$$\begin{aligned} p(\mathbf{s}) &= p(s_1)\prod_{i=1}^{M-1} p(s_{i+1} \mid s_i) \\ &\propto p\exp\left(\frac{-s_1^2}{2\sigma_0^2}\right) + (1-p)\exp\left(\frac{-(s_1-1)^2}{2\sigma_0^2}\right) \\ &\quad\times \prod_{i=1}^{M-1}\left\{[p_{01} + (1-p_{10})]\exp\left(\frac{-s_{i+1}^2}{2\sigma_0^2}\right)\right. \\ &\quad\left. + [p_{10} + (1-p_{01})]\exp\left(\frac{-(s_{i+1}-1)^2}{2\sigma_0^2}\right)\right\}, \end{aligned} \tag{16}$$

$$p\left(\mathbf{y} \mid \mathbf{s}, \widehat{\boldsymbol{\theta}}\right) = \frac{1}{\left(\sqrt{2\pi\sigma_n^2}\right)^M}\exp\left(-\frac{\left\|\mathbf{y} - \boldsymbol{\Psi}\widehat{\boldsymbol{\theta}}\right\|_2^2}{2\sigma_n^2}\right) \tag{17}$$

where $\boldsymbol{\Psi} = \boldsymbol{\Phi}\mathbf{S}$. Hence, the log posterior $\mathcal{L}(\mathbf{s})$ in the M-step is simplified as

$$\mathcal{L}(\mathbf{s}) \propto \log(p(s_1)) + \sum_{i=1}^{M-1}\log(p(s_{i+1} \mid s_i)) - \frac{\left\|\mathbf{y} - \boldsymbol{\Psi}\widehat{\boldsymbol{\theta}}\right\|_2^2}{2\sigma_n^2}. \tag{18}$$

The first derivative of (18) can be written as

$$\begin{aligned} \frac{\partial\mathcal{L}(\mathbf{s})}{\partial\mathbf{s}} &\propto \frac{\partial}{\partial\mathbf{s}}\log(p(s_1)) + \frac{\partial}{\partial\mathbf{s}}\sum_{i=1}^{M-1}\log(p(s_{i+1} \mid s_i)) \\ &\quad - \frac{1}{2\sigma_n^2}\frac{\partial}{\partial\mathbf{s}}(\mathbf{y} - \boldsymbol{\Phi}\mathbf{S}\widehat{\boldsymbol{\theta}})^T(\mathbf{y} - \boldsymbol{\Phi}\mathbf{S}\widehat{\boldsymbol{\theta}}). \end{aligned} \tag{19}$$

Define $\mathbf{g}(\mathbf{s}) \triangleq -\sigma_0^2\frac{\partial}{\partial\mathbf{s}}\log(p(s_1)) - \sigma_0^2\frac{\partial}{\partial\mathbf{s}}\sum_{i=1}^{M-1}\log(p(s_{i+1} \mid s_i)) = \mathbf{g_1}(\mathbf{s}) + \mathbf{g_2}(\mathbf{s})$ and $\mathbf{n}(\mathbf{s}) \triangleq (\mathbf{y} - \boldsymbol{\Phi}\mathbf{S}\widehat{\boldsymbol{\theta}})^T(\mathbf{y} - \boldsymbol{\Phi}\mathbf{S}\widehat{\boldsymbol{\theta}})$. Then, the two scalar functions $g_1(s_1)$ and $g_2(s_{i+1})$ ($i = 1, 2, \cdots, M-1$) can be given as

$$g_1(s_1) = \frac{ps_1e^{\left(\frac{-s_1^2}{2\sigma_0^2}\right)} + (1-p)(s_1-1)e^{\left(\frac{-(s_1-1)^2}{2\sigma_0^2}\right)}}{pe^{\left(\frac{-s_1^2}{2\sigma_0^2}\right)} + (1-p)e^{\left(\frac{-(s_1-1)^2}{2\sigma_0^2}\right)}},$$

$$g_2(s_{i+1}) = \frac{q_1s_{i+1}e^{\left(\frac{-s_{i+1}^2}{2\sigma_0^2}\right)} + q_2(s_{i+1}-1)e^{\left(\frac{-(s_{i+1}-1)^2}{2\sigma_0^2}\right)}}{q_1e^{\left(\frac{-s_{i+1}^2}{2\sigma_0^2}\right)} + q_2e^{\left(\frac{-(s_{i+1}-1)^2}{2\sigma_0^2}\right)}}$$

where $q_1 = p_{01} + (1-p_{10})$, $q_2 = p_{10} + (1-p_{01})$ and $e^x = \exp(x)$. It can be shown that (see the complete proof in [14])

$$\frac{\partial\mathbf{n}(\mathbf{s})}{\partial\mathbf{s}} = 2\cdot\operatorname{diag}(\boldsymbol{\Phi}^T\boldsymbol{\Phi}\mathbf{S}\widehat{\boldsymbol{\theta}} - \boldsymbol{\Phi}^T\mathbf{y})\cdot\widehat{\boldsymbol{\theta}}. \tag{20}$$

Therefore using (20), (19), and the definitions of $\mathbf{g}(\mathbf{s})$ and $\mathbf{n}(\mathbf{s})$, the expression for obtaining the sequence of optimal solution of (13) using steepest-ascent method ($\mathbf{s}^{(k+1)} = \mathbf{s}^{(k)} + \mu\frac{\partial\mathcal{L}(\mathbf{s})}{\partial\mathbf{s}}\Big|_{\mathbf{s}=\mathbf{s}^{(k)}}$) can be given as

$$\mathbf{s}^{(k+1)} = \mathbf{s}^{(k)} + \frac{\mu}{\sigma_0^2}\mathbf{g}(\mathbf{s}) + \frac{\mu}{\sigma_n^2}\operatorname{diag}(\boldsymbol{\Phi}^T(\boldsymbol{\Psi}\widehat{\boldsymbol{\theta}} - \mathbf{y}))\cdot\widehat{\boldsymbol{\theta}}. \tag{21}$$

Note that in the computation we decrease $\sigma_0$ in the consecutive iterations to guarantee the global maxima of (13). Hence, for each iteration we have $\sigma_0^{(i+1)} = \alpha\sigma_0^{(i)}$, where $\alpha$ is selected in the range $[0.6, 1]$. If the columns of $\boldsymbol{\Phi}$ are normalized to have unit norms, the range for step size $\mu$ can be expressed as

$$0 < \mu < \frac{2}{\frac{1}{\sigma_0^2} + \frac{MM^{*2}}{\sigma_n^2}} \tag{22}$$

where $M^* = \sigma_\theta Q^{-1}(\frac{1-\sqrt[M]{0.99}}{2})$ and $Q^{-1}(\cdot)$ is the inverse Gaussian Q-function.

We initialize the proposed Block-IBA with the minimum $\ell^2$-norm solution and we use a decreasing threshold ($Th$) to activate the sparse samples of the signal $\mathbf{w}$. The value of $Th$ specifies the number of nonzero elements in $\mathbf{s}$ vector. To obtain an estimate of $\sigma_\theta$, the method of moments estimator is appealing. We assume that the $\mathbf{\Phi}$ matrix has the columns with the unit norms and its elements have a uniform distribution between [-1,1]. From (1), we know that $y_j = \sum_{i=1}^{M} \varphi_{ji} w_i + n_j$, and by neglecting the noise power we have $\mathbb{E}\left(y_j^2\right) = M\mathbb{E}\left(\varphi_{ji}^2\right)\mathbb{E}\left(w_i^2\right)$. Moreover, we know that $\sum_{j=1}^{N} \varphi_{ji}^2 = 1$, hence $\mathbb{E}\left(\varphi_{ji}^2\right) = 1/N$. Finally, as $\mathbb{E}(w_i^2) = (1-p)\sigma_\theta^2$, we can obtain a simple update for $\sigma_\theta$ as

$$\hat{\sigma}_\theta = \sqrt{\frac{N\mathbb{E}(y_j^2)}{M(1-\hat{p})}}. \tag{23}$$

we can also express the following update equations for the rest of parameters

$$\hat{\sigma}_n = \frac{\|\mathbf{y} - \mathbf{\Phi}\widehat{\mathbf{w}}\|_2}{\sqrt{N}}, \tag{24}$$

$$\hat{p} = \frac{\|\mathbf{s}\|_0}{M}, \tag{25}$$

$$\hat{p}_{01} = \frac{\sum_{i=1}^{M-1} s_i\left(1 - s_{i+1}\right)}{\sum_{i=1}^{M-1} s_i}. \tag{26}$$

## 4. EMPIRICAL EVALUATION

This section presents the experimental result to demonstrate the performance of the Block-IBA. The experiment is conducted for 400 independent simulation runs. In each simulation run the elements of the matrix $\mathbf{\Phi}$ are chosen from a uniform distribution in [-1,1] with columns normalized to unit $\ell^2$-norm. The size of the matrix is $N = 192$ and $M = 512$. Block-sparse sources $\mathbf{w_{gen}}$ are synthetically generated using BGHMM in (7) with parameters $p = 0.9$ and $\sigma_\theta = 1$. The measurement vector $\mathbf{y}$ is constructed by $\mathbf{y} = \mathbf{\Phi}\mathbf{w_{gen}} + \mathbf{n}$ where $\mathbf{n}$ is zero-mean Gaussian noise with a variance tuned to a specified value of SNR which is $\mathrm{SNR}\,(\mathrm{dB}) \triangleq 20\log_{10}\left(\|\mathbf{\Phi}\mathbf{w_{gen}}\|_2 \,/\, \|\mathbf{n}\|_2\right)$. We set $\mathrm{SNR} = 15\mathrm{dB}$, $\alpha = 0.98$, and $\mu = 10^{-7}$.

The Normalized Mean Square Error (NMSE) $\mathrm{NMSE} \triangleq \frac{\|\widehat{\mathbf{w}} - \mathbf{w_{gen}}\|_2^2}{\|\mathbf{w_{gen}}\|_2^2}$ is used as a performance metric where $\widehat{\mathbf{w}}$ is the estimate of the true signal $\mathbf{w_{gen}}$. We compare the proposed Block-IBA with some recently developed algorithms for block sparse signal reconstruction, such as the block sparse Bayesian learning algorithm (BSBL) [12], the expanded block sparse Bayesian learning algorithm (EBSBL) [12], the Boltzman machine-based greedy pursuit algorithm (BM-MAP-OMP) [11], the cluster-structured MCMC algorithm (CluSS-MCMC) [10], and the pattern-coupled sparse Bayesian learning algorithm (PC-SBL) [13]. Recall from Section 2 that the block size and the number of blocks of $\mathbf{w}$ are proportional to $1/p_{01}$. That is, when $p_{01}$ is small $\mathbf{w}$ comprises small number of blocks with big sizes and vice versa. Hence, we vary the value of $p_{01}$ between 0.09 and 0.9 to obtain the NMSE for various algorithms. The results of NMSE versus $p_{01}$ is shown in Fig. 1. Two observations can be made from Fig. 1. First, the Block-IBA with Gamma distribution as hyperprior over hyperparameter $\boldsymbol{\gamma}$ (denoted as Gamma-Block-IBA) outperforms the Block-IBA with MMSE estimation (denoted as Block-IBA). Second, for $p_{01} \geq 0.36$ Gamma-Block-IBA outperforms all other algorithms. This is because for $p_{01} \geq 0.36$ the samples of $\mathbf{s}$ vector inside each block follow the first-order Markov chain process more accurately and they tend to be more non-i.i.d.

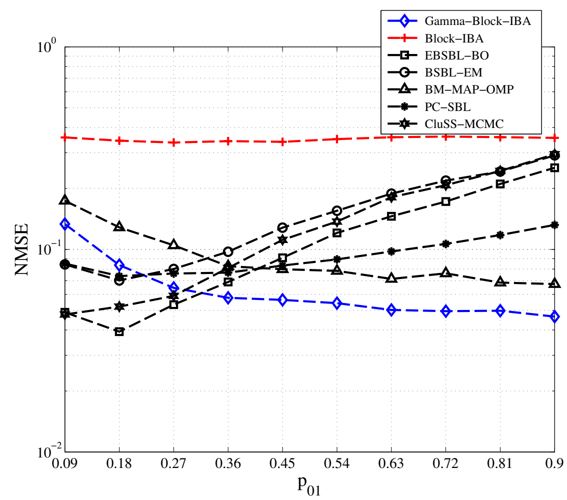


**Fig. 1.** NMSE versus $p_{01}$ for the Block-IBA and other algorithms. The results are averaged over 400 trials.

## 5. CONCLUSION

In this paper, we have developed a novel Block-IBA using an iterative expectation-maximization (EM) algorithm to recover the block-sparse signal whose structure of block sparsity is completely unknown. Unlike the other existing algorithms we have modeled the cluster pattern of the signal using Bernoulli-Gaussian hidden Markov model (BGHMM) which is a more realistic model in practical applications. We have optimized the M-step of the EM algorithm with the steepest-ascent method. Experimental results show that the Block-IBA outperforms other existing algorithms (for block sparse signal reconstruction) when the block-sparse signal comprises a large number of blocks with short lengths.

## 6. REFERENCES


[1] M. Mishali and Y. C. Eldar, "Blind multiband signal reconstruction: Compressed sensing for analog signals," *Signal Processing, IEEE Transactions on*, vol. 57, no. 3, pp. 993–1009, 2009.

[2] R. Gribonval and E. Bacry, "Harmonic decomposition of audio signals with matching pursuit," *Signal Processing, IEEE Transactions on*, vol. 51, no. 1, pp. 101–111, 2003.

[3] M. F. Duarte and Y. C. Eldar, "Structured compressed sensing: From theory to applications," *Signal Processing, IEEE Transactions on*, vol. 59, no. 9, pp. 4053–4085, 2011.

[4] Z. Zhilin and B. D. Rao, "Sparse signal recovery with temporally correlated source vectors using sparse bayesian learning," *Selected Topics in Signal Processing, IEEE Journal of*, vol. 5, no. 5, pp. 912–926, 2011.

[5] Y. C. Eldar, P. Kuppinger, and H. Bolcskei, "Block-sparse signals: Uncertainty relations and efficient recovery," *Signal Processing, IEEE Transactions on*, vol. 58, no. 6, pp. 3042–3054, 2010.

[6] Y. C. Eldar and M. Mishali, "Robust recovery of signals from a structured union of subspaces," *Information Theory, IEEE Transactions on*, vol. 55, no. 11, pp. 5302–5316, 2009.

[7] M. Yuan and Y. Lin, "Model selection and estimation in regression with grouped variables," *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, vol. 68, no. 1, pp. 49–67, 2006.

[8] R. G. Baraniuk, V. Cevher, M. F. Duarte, and C. Hegde, "Model-based compressive sensing," *Information Theory, IEEE Transactions on*, vol. 56, no. 4, pp. 1982–2001, 2010.

[9] M. Zimmermann and K. Dostert, "Analysis and modeling of impulsive noise in broad-band powerline communications," *Electromagnetic Compatibility, IEEE Transactions on*, vol. 44, no. 1, pp. 249–258, 2002.

[10] L. Yu, H. Sun, J. P. Barbot, and G. Zheng, "Bayesian compressive sensing for cluster structured sparse signals," *Signal Processing*, vol. 92, no. 1, pp. 259–269, 2012.

[11] T. Peleg, Y. C. Eldar, and M. Elad, "Exploiting statistical dependencies in sparse representations for signal recovery," *Signal Processing, IEEE Transactions on*, vol. 60, no. 5, pp. 2286–2303, 2012.

[12] Z. Zhang and B. D. Rao, "Extension of SBL algorithms for the recovery of block sparse signals with intra-block correlation," *Signal Processing, IEEE Transactions on*, vol. 61, no. 8, pp. 2009–2015, 2013.

[13] S. Yanning, D. Huiping, F. Jun, and L. Hongbin, "Pattern-coupled sparse bayesian learning for recovery of block-sparse signals," in *Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on*, pp. 1896–1900.

[14] H. Zayyani, M. Babaie-Zadeh, and C. Jutten, "An iterative bayesian algorithm for sparse component analysis in presence of noise," *Signal Processing, IEEE Transactions on*, vol. 57, no. 11, pp. 4378–4390, 2009.

[15] J. Ziniel and P. Schniter, "Dynamic compressive sensing of time-varying signals via approximate message passing," *Signal Processing, IEEE Transactions on*, vol. 61, no. 21, pp. 5270–5284, 2013.

[16] H. Zayyani, M. Babaie-Zadeh, and C. Jutten, "Bayesian Pursuit algorithm for sparse representation," in *Acoustics, Speech and Signal Processing (ICASSP), 2009 IEEE International Conference on*, pp. 1549-1552.

[17] M. E. Tipping, "Sparse bayesian learning and the relevance vector machine," *Journal of Machine Learning Research*, vol. 1, no. 3, pp. 211–244, 2001.