RECOVERY OF BLOCK SPARSE SIGNALS USING THE FRAMEWORK OF BLOCK SPARSE BAYESIAN LEARNING

Zhilin Zhang and Bhaskar D. Rao

ECE Department, University of California at San Diego, La Jolla, CA 92093-0407, USA
{z4zhang,brao}@ucsd.edu

ABSTRACT

In this paper we study the recovery of block sparse signals and extend conventional approaches in two important directions; one is learning and exploiting intra-block correlation, and the other is generalizing signals' block structure such that the block partition is not needed to be known for recovery. We propose two algorithms based on the framework of block sparse Bayesian learning (bSBL). One algorithm, directly derived from the framework, requires a priori knowledge of the block partition. Another algorithm, derived from an expanded bSBL framework using the generalization method, can be used when the block partition is unknown. Experiments show that they have superior performance to state-of-the-art algorithms.

Index Terms— Sparse Signal Recovery, Compressed Sensing, Sparse Bayesian Learning, Block Sparse Model, Cluster Structure

1. INTRODUCTION

The basic problem in compressed sensing and sparse signal recovery is to recover a sparse signal, a signal with only a few non-zero elements, from a small number of its linear measurements. A trend in the field is to exploit structure of signals for better performance. In this paper we study the recovery of block sparse signals. Its mathematical model is given by [1–4]

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

where the signal $\mathbf{x}$ has the block structure

$$\mathbf{x} = [\underbrace{x_1, \cdots, x_{d_1}}_{\mathbf{x}_1^T}, \cdots, \underbrace{x_{d_{g-1}+1}, \cdots, x_{d_g}}_{\mathbf{x}_g^T}]^T \tag{2}$$

Here, $\mathbf{y} \in \mathbb{R}^{M\times 1}$ is a measurement vector consisting of M available measurements. $\mathbf{\Phi} \in \mathbb{R}^{M\times N}(M \ll N)$ is a known matrix. $\mathbf{x} \in \mathbb{R}^{N\times 1}$ is the block sparse signal which we want to recover. $\mathbf{v}$ is an unknown noise vector. In the block partition (2), $d_1, \cdots, d_g$ are not necessarily identical. Among the g blocks, only a few blocks are nonzero. The model (1)-(2) is called the block sparse model [4]. Recent works [2,3] have shown that if such block partition can be exploited, then under certain conditions, the number of measurements required to recover $\mathbf{x}$ can be further reduced.

Given the block partition, a number of algorithms have been proposed, such as Group Lasso [1], Mixed ℓ_2/ℓ_1 Program [4], Block-OMP [3], and Block-CoSaMp [2]. A summary of requirements and abilities of typical algorithms is given in Table 1. However, in many applications the block partition is unknown. So, how to adaptively estimate the block partition while recovering the sparse signal is a challenge. Some algorithms have been proposed, such as DGS [5], and CluSS-MCMC [6]. But in some cases these algorithms may perform even poorer than some general CS algorithms which do not exploit the underlying block partition. Recently, it is found in other related models such as the multiple measurement vector (MMV) model, exploiting (inter-vector) correlation structure can greatly improve algorithms' performance [7]. Unfortunately, for the block sparse model considered here, there is no algorithm exploiting correlation structure in each block (i.e. intra-block correlation).

Based on the framework of block sparse Bayesian learning (bSBL) [7], we derive two algorithms which can *learn and exploit intra-block correlation* and have superior performance to state-of-the-art algorithms. One algorithm, directly derived from the bSBL framework, requires a priori knowledge of the block partition. Another algorithm, derived from an expanded bSBL framework, is based on a weaker assumption on the block structure but actually can be used when the block partition is unknown. The two algorithms show that intra-block correlation is greatly helpful to improve recovery performance. Besides, the second one shows that the expanded bSBL framework is a promising framework when the block partition is unknown.

2. THE BSBL FRAMEWORK AND THE CLUSTER-SBL (TYPE I) ALGORITHM

Based on the block sparse model (1)-(2), we assume each block $\mathbf{x}_i \in \mathbb{R}^{d_i\times 1}$ satisfies a parameterized multivariate Gaussian distribution:

$$p(\mathbf{x}_i) \sim \mathcal{N}_x(\mathbf{0}, \gamma_i\mathbf{B}_i), \quad i = 1, \cdots, g$$

Here γ_i is a nonnegative parameter. When $\gamma_i = 0$, the i-th block becomes zero. During the learning procedure, most $\gamma_i(\forall i)$ tend to zero, due to the mechanism of automatic relevance determination [8]. $\mathbf{B}_i \in \mathbb{R}^{d_i\times d_i}$ is a positive definite matrix, capturing correlation structure of the i-th block. Further, by assuming the blocks are mutually uncorrelated, we have the prior of $\mathbf{x}$, given by $p(\mathbf{x}) \sim \mathcal{N}_x(\mathbf{0}, \mathbf{\Sigma}_0)$, where $\mathbf{\Sigma}_0$ is a block-diagonal matrix with each principal block given by $\gamma_i\mathbf{B}_i$. The noise vector is assumed to satisfy $p(\mathbf{v}) \sim \mathcal{N}_v(\mathbf{0}, \lambda\mathbf{I})$, where λ is a nonnegative scalar. Therefore the posterior of $\mathbf{x}$ is given by $p(\mathbf{x}|\mathbf{y}; \lambda, \{\gamma_i, \mathbf{B}_i\}_{i=1}^g) = \mathcal{N}(\boldsymbol{\mu}_x, \mathbf{\Sigma}_x)$ with $\boldsymbol{\mu}_x = \mathbf{\Sigma}_0\mathbf{\Phi}^T\left(\lambda\mathbf{I} + \mathbf{\Phi}\mathbf{\Sigma}_0\mathbf{\Phi}^T\right)^{-1}\mathbf{y}$ and $\mathbf{\Sigma}_x = (\mathbf{\Sigma}_0^{-1} + \frac{1}{\lambda}\mathbf{\Phi}^T\mathbf{\Phi})^{-1}$. Once the hyperparameters $\lambda, \{\gamma_i, \mathbf{B}_i\}_{i=1}^g$ are estimated, the Maximum-A-Posterior (MAP) estimate of $\mathbf{x}$ can be directly obtained from the mean of the posterior.

This model is directly compatible with the block sparse Bayesian learning (bSBL) framework, which was originally developed in [7] for learning the hyper-parameters $\lambda, \gamma_i, \mathbf{B}_i(\forall i)$ in the MMV model with inter-vector correlation. This is not surprising, since the MMV

The work was supported by NSF grant CCF-0830612 and CCF-1144258.

Table 1. Summary of requirements and abilities of some typical algorithms. '✓' means that a given algorithm needs the corresponding a priori information. 'YES' means a given algorithm is able to deal with the corresponding situation.

	need block partition?	need nonzero block (or element) number?	for noisy cases?	for noiseless cases?	exploit intra-block correlation?
Group Lasso [1]	✓		YES	YES	
Mixed ℓ_2/ℓ_1 [4]	✓		YES	YES	
Block-OMP [3]	✓			YES	
Block-CoSaMp [2]	✓	✓		YES	
DGS [5]		✓	YES	YES	
CluSS-MCMC [6]			YES	YES	
Cluster-SBL(Type I)	✓		YES	YES	**YES**
Cluster-SBL(Type II)			YES	YES	**YES**

model can be cast as a special case of a block sparsity model [4]. So, following the EM method used in [7], we can derive the following iterative algorithm

$$\begin{aligned}
\boldsymbol{\mu}_x &\leftarrow \boldsymbol{\Sigma}_0\boldsymbol{\Phi}^T\big(\lambda\mathbf{I}+\boldsymbol{\Phi}\boldsymbol{\Sigma}_0\boldsymbol{\Phi}^T\big)^{-1}\mathbf{y}\\
\boldsymbol{\Sigma}_x &\leftarrow \boldsymbol{\Sigma}_0-\boldsymbol{\Sigma}_0\boldsymbol{\Phi}^T\big(\lambda\mathbf{I}+\boldsymbol{\Phi}\boldsymbol{\Sigma}_0\boldsymbol{\Phi}^T\big)^{-1}\boldsymbol{\Phi}\boldsymbol{\Sigma}_0\\
\lambda &\leftarrow \frac{\|\mathbf{y}-\boldsymbol{\Phi}\boldsymbol{\mu}_x\|_2^2+\lambda[N-\mathrm{Tr}(\boldsymbol{\Sigma}_x\boldsymbol{\Sigma}_0^{-1})]}{M}\\
\gamma_i &\leftarrow \frac{1}{d_i}\mathrm{Tr}\big[\mathbf{B}_i^{-1}\big(\boldsymbol{\Sigma}_x^i+\boldsymbol{\mu}_x^i(\boldsymbol{\mu}_x^i)^T\big)\big]\quad\forall i
\end{aligned}$$

where in the last learning rule $\boldsymbol{\mu}_x^i$ is the corresponding i-th block in $\boldsymbol{\mu}_x$ (with the size $d_i\times 1$), and $\boldsymbol{\Sigma}_x^i$ is the corresponding i-th principal diagonal block in $\boldsymbol{\Sigma}_x$ (with the size $d_i\times d_i$). Once the algorithm converges, the estimate of $\mathbf{x}$ is given by $\boldsymbol{\mu}_x$.

Similarly, using the EM method we can derive a learning rule for each $\mathbf{B}_i$. However, assigning each block with a different $\mathbf{B}_i$ can result in overfitting. When blocks have the same size, an effective strategy to prevent overfitting is to constrain all the blocks to have the same correlation structure [7]. In this case, we can constrain $\mathbf{B}_i=\mathbf{B}(\forall i)$. And its learning rule is obtained as follows:

$$\mathbf{B}\leftarrow\frac{1}{g}\sum_{i=1}^{g}\frac{\boldsymbol{\Sigma}_x^i+\boldsymbol{\mu}_x^i(\boldsymbol{\mu}_x^i)^T}{\gamma_i}. \tag{3}$$

However, the algorithm's performance can be improved by further constraining the matrix $\mathbf{B}$. The idea is to find a positive definite and symmetric matrix $\widehat{\mathbf{B}}$ such that $\widehat{\mathbf{B}}$ is close to $\mathbf{B}$ especially in the elements along the main diagonal and the main sub-diagonal. Based on this idea, a possible form of $\widehat{\mathbf{B}}$ is given by

$$\begin{aligned}
\widehat{\mathbf{B}} &\triangleq \mathrm{Toeplitz}([1,r,\cdots,r^{d-1}])\\
&= \begin{bmatrix} 1 & r & r^2 & \cdots & r^{d-1}\\ r & 1 & r & \cdots & r^{d-2}\\ \vdots & \vdots & \ddots & & \vdots\\ r^{d-1} & r^{d-2} & \cdots & & 1\end{bmatrix}
\end{aligned} \tag{4}$$

Here $r\triangleq\mathrm{sign}(\frac{m_1}{m_0})\min\{|\frac{m_1}{m_0}|,0.99\}$, where m_0 is the average of the elements along the main diagonal and m_1 is the average of the elements along the main sub-diagonal of the matrix $\mathbf{B}$ in (3). 0.99 is a bound such that r has a reasonable value. The form in (4) is equivalent to modeling elements in a block as a first order AR process. A similar constraint on $\mathbf{B}$ was used in our MMV work [9], which shows good results. When blocks have different sizes, we still can adopt the above idea. First, using the EM method we can derive the learning rule for each $\mathbf{B}_i$, and then obtain r by averaging over corresponding elements from all the matrices $\mathbf{B}_i(\forall i)$, and finally form $\widehat{\mathbf{B}}_i\triangleq\mathrm{Toeplitz}([1,r,\cdots,r^{d_i-1}])$ for each block.

We denote the above algorithm by **Cluster-SBL (Type I)**.

3. THE EXPANDED BSBL FRAMEWORK AND THE CLUSTER-SBL (TYPE II) ALGORITHM

In many applications the block partition (2) is not available. Here we generalize the conventional block sparse model with the unknown block partition. Based on it, we derive another algorithm, called **Cluster-SBL (Type II)**, which requires little a priori knowledge on the block structure. Note that based on this generalized model, many existing algorithms can also be used. But our proposed algorithm is more advantageous as shown in experiments.

Our model is motivated by the physics of certain applications, but with proper interpretation it has considerable flexibility and generality. We *first assume that all the nonzero blocks are of equal size h and are arbitrarily located*. This model is consistent with communication channel modeling where an ideal sparse channel consisting of a few specular multi-path components has a discrete-time, bandlimited, baseband representation, which exhibits a block sparse structure with the block centers determined by the arbitrary arrival times of the multi-path components. Since the blocks are arbitrarily located they can overlap giving rise to larger unequal blocks, making the model quite flexible. So, *the assumption of equal block-size is not limiting*. However, we would like to add that though the resulting algorithm is not very sensitive to the choice of h, h still has to be properly chosen for good algorithmic performance. We will comment more on h in Section 4.

Given the identical block size h, there are $p\triangleq N-h+1$ possible blocks in $\mathbf{x}$, which overlap each other. The i-th block starts at the i-th element and ends at the $(i+h-1)$-th element of $\mathbf{x}$. All the nonzero entries in $\mathbf{x}$ lie in some of these blocks. Similar to Section 2, for the i-th block, we assume it satisfies a multivariate Gaussian distribution with the mean given by $\mathbf{0}$ and the covariance matrix given by $\gamma_i\mathbf{B}_i$, where $\mathbf{B}_i\in\mathbb{R}^{h\times h}$. So we have the prior of $\mathbf{x}$ as the form: $p(\mathbf{x})\sim\mathcal{N}_x(\mathbf{0},\boldsymbol{\Sigma}_0)$. Note that due to the overlapping of these blocks, here $\boldsymbol{\Sigma}_0$ is no longer a block diagonal matrix as in Section 2. It has the structure that each $\gamma_i\mathbf{B}_i$ lies along the principal diagonal of $\boldsymbol{\Sigma}_0$ and overlaps other $\gamma_j\mathbf{B}_j$ (see the left picture in Fig.1 for an illustration). Starting with this model we can develop algorithms for estimating the hyper-parameters $\lambda,\gamma_i,\mathbf{B}_i(\forall i)$. However, because the covariance matrix $\boldsymbol{\Sigma}_0$ is no longer a block diagonal matrix we cannot directly use the bSBL framework and need to make

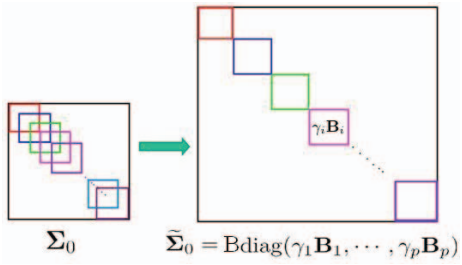

Fig. 1. Structures of $\boldsymbol{\Sigma}_0$ and $\widetilde{\boldsymbol{\Sigma}}_0$. Each color block corresponds to a possible nonzero block in $\mathbf{x}$.

some modification.

To facilitate the use of the bSBL framework, we expand the covariance matrix $\boldsymbol{\Sigma}_0$ as follows:

$$\widetilde{\boldsymbol{\Sigma}}_0 = \mathrm{Bdiag}(\gamma_1 \mathbf{B}_1, \cdots, \gamma_p \mathbf{B}_p) \in \mathbb{R}^{ph \times ph} \tag{5}$$

where $\mathrm{Bdiag}(\cdot)$ denotes a block diagonal matrix with principal diagonal blocks given by $\gamma_1 \mathbf{B}_1, \cdots, \gamma_p \mathbf{B}_p$. Note that now $\gamma_i \mathbf{B}_i$ does not overlap other $\gamma_j \mathbf{B}_j (i \neq j)$ (see the right picture in Fig.1). The expanded covariance matrix $\widetilde{\boldsymbol{\Sigma}}_0$ implies the decomposition of $\mathbf{x}$:

$$\mathbf{x} = \sum_{i=1}^{p} \mathbf{E}_i \mathbf{z}_i, \tag{6}$$

where $E\{\mathbf{z}_i\} = 0$, $E\{\mathbf{z}_i \mathbf{z}_j^T\} = \delta_{i,j} \gamma_i \mathbf{B}_i$ ($\delta_{i,j} = 1$ if $i = j$; otherwise, $\delta_{i,j} = 0$), and $\mathbf{z} \triangleq [\mathbf{z}_1^T, \cdots, \mathbf{z}_p^T]^T \sim \mathcal{N}_z(\mathbf{0}, \widetilde{\boldsymbol{\Sigma}}_0)$. $\mathbf{E}_i \in \mathbb{R}^{M \times h}$ is a zero matrix except that the part from its i-th row to $(i + h - 1)$-th row is replaced by the identity matrix $\mathbf{I}$. Then the original model (1) can be expressed as:

$$\mathbf{y} = \sum_{i=1}^{p} \boldsymbol{\Phi} \mathbf{E}_i \mathbf{z}_i + \mathbf{v} \triangleq \mathbf{A}\mathbf{z} + \mathbf{v}, \tag{7}$$

where $\mathbf{A} \triangleq [\mathbf{A}_1, \cdots, \mathbf{A}_p]$ and $\mathbf{A}_i \triangleq \boldsymbol{\Phi} \mathbf{E}_i$. Now we see the new model (7) is exactly a bSBL model. Directly following the development in Section 2, we can derive the algorithm as follows:

$$\begin{aligned}
\boldsymbol{\mu}_z &\leftarrow \widetilde{\boldsymbol{\Sigma}}_0 \mathbf{A}^T \big(\lambda \mathbf{I} + \mathbf{A} \widetilde{\boldsymbol{\Sigma}}_0 \mathbf{A}^T\big)^{-1} \mathbf{y} \\
\boldsymbol{\Sigma}_z &\leftarrow \widetilde{\boldsymbol{\Sigma}}_0 - \widetilde{\boldsymbol{\Sigma}}_0 \mathbf{A}^T \big(\lambda \mathbf{I} + \mathbf{A} \widetilde{\boldsymbol{\Sigma}}_0 \mathbf{A}^T\big)^{-1} \mathbf{A} \widetilde{\boldsymbol{\Sigma}}_0 \\
\gamma_i &\leftarrow \frac{\mathrm{Tr}\big[\mathbf{B}^{-1} \big(\boldsymbol{\Sigma}_z^i + \boldsymbol{\mu}_z^i (\boldsymbol{\mu}_z^i)^T\big)\big]}{h}, \quad \forall i \\
\mathbf{B} &\leftarrow \frac{1}{p} \sum_{i=1}^{p} \frac{\boldsymbol{\Sigma}_z^i + \boldsymbol{\mu}_z^i (\boldsymbol{\mu}_z^i)^T}{\gamma_i} \\
\lambda &\leftarrow \frac{\|\mathbf{y} - \mathbf{A}\boldsymbol{\mu}_z\|_2^2 + \lambda [ph - \mathrm{Tr}(\boldsymbol{\Sigma}_z \widetilde{\boldsymbol{\Sigma}}_0^{-1})]}{M}
\end{aligned}$$

where $\boldsymbol{\mu}_z^i$ is the corresponding i-th block in $\boldsymbol{\mu}_z$, and $\boldsymbol{\Sigma}_z^i$ is the corresponding i-th main diagonal block in $\boldsymbol{\Sigma}_z$. As in Section 2 we can further constrain the matrix $\mathbf{B}$ for better performance. After convergence, the original signal $\mathbf{x}$ is recovered by (6).

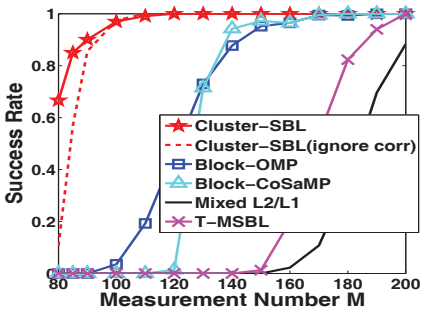

Fig. 2. Results of Experiment 1.

4. EXPERIMENTS

We compared our proposed algorithms to T-MSBL [7] and most algorithms in Table 1. Note that T-MSBL, when used in the model (1)-(2), can be viewed as a special case of our proposed algorithms (block size is 1). In experiments, we also computed the 'oracle' result, which was the least-square estimate of the signal $\mathbf{x}$ given its true support (i.e. locations of nonzero elements). We chose MSE as a performance index in noisy cases, which was defined by $\|\mathbf{x}' - \mathbf{x}\|_2^2 / \|\mathbf{x}\|_2^2$, where $\mathbf{x}'$ was the estimate of $\mathbf{x}$. In noiseless cases we chose the *Success Rate* as a performance index, which was defined as the ratio of the number of successful trials to the number of total trials (For every set of experiment settings, our experiment consisted of 500 trials.). A successful trial was defined as the one when MSE $\leq 10^{-6}$.

Matlab codes of the proposed algorithms and other experiment results are available at: `http://dsp.ucsd.edu/~zhilin/BSBL.html`.

Experiment 1: Block partition is known in noiseless cases. First we compared algorithms at different number of measurements M. We generated a block sparse signal, whose length was 512. It consisted of 64 blocks with identical sizes, ten of which were nonzero (randomly chosen). For the k-th nonzero blocks ($k = 1, \cdots, 10$), its elements were generated from a multivariate Gaussian distribution: $\mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_k)$, where $\boldsymbol{\Sigma}_k$ was generated using the Matlab command: $\mathrm{Toeplitz}([1, \beta_k, \cdots, \beta_k^7])$. Note that β_k was the parameter measuring the intra-block correlation of the k-th nonzero block. In each trial, it was randomly and uniformly chosen from 0.9 to 0.99. The matrix $\boldsymbol{\Phi}$ was of the size $M \times 512$, where M varied from 80 to 200. Since the block partition was known to all the algorithms, we used Cluster-SBL (Type I). The result (Fig.2) shows Cluster-SBL (Type I) had the best performance. It required much fewer measurements than other algorithms to achieve the same performance.

We also plot the performance curve of Cluster-SBL (Type I) when it did not exploit the intra-block correlation (i.e. fixing $\mathbf{B}$ to the identity matrix). Its performance was poorer than when it exploited the correlation. This shows the benefits of exploiting intra-block correlation. For example, to achieve the same recovery performance, exploiting the correlation requires fewer measurements than ignoring the correlation; or given the same measurements, exploiting the correlation can achieve better recovery performance.

Experiment 2: Block partition is unknown in noisy cases. Since Cluster-SBL (Type II), Cluss-MCMC, and DGS can be used when block partition is unknown, we compared them in this experiment. $\boldsymbol{\Phi}$ was of the size 80×256. The number of nonzero ele-

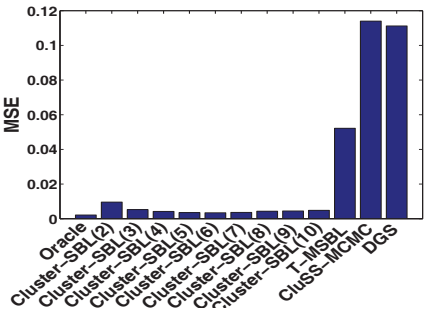

Fig. 3. Results of Experiment 2. The label 'Cluster-SBL(k)' denotes Cluster-SBL with $h = k$.

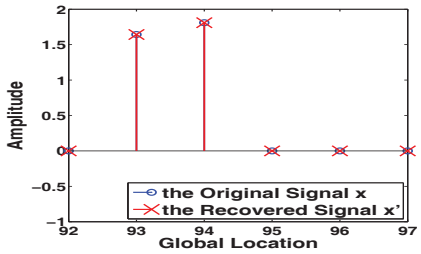

Fig. 4. A nonzero block (block size: 2) of $\mathbf{x}$ was recovered by some $\mathbf{z}_i$ (block size: 6). In $\mathbf{x}$ the values from the 92-th location to 97-th location were $[0, \mathbf{1.64269416}, \ \mathbf{1.80932747}, 0, 0, 0]$. In the recovered signal $\mathbf{x}'$ the corresponding values were $[2.0 \times 10^{-7}, \mathbf{1.64269411}, \ \mathbf{1.80932730}, 3.3 \times 10^{-7}, -7.7 \times 10^{-7}, -5.7 \times 10^{-7}]$.

ments in the signal was fixed to 32, which were randomly put into 4 blocks. So each block had random size and random location. To more clearly see the effectiveness of our generalization model, we set intra-block correlation to zero. SNR was 25 dB. For Cluster-SBL (Type II), we set h to different values ranging from 2 to 10. The result (Fig.3) shows that Cluster-SBL (Type II) had much better performance than other compared algorithms. And its performance was almost the same when h chose values from 3 to 10. We also compared T-MSBL (can be viewed as Cluster-SBL (Type II) with $h = 1$). The result shows it also had better performance than Cluss-MCMC and DGS.

Experiment 3: Choice of the parameter h. In the algorithm development we assumed that all the blocks have the same size h, which is known. However, this assumption is not crucial for practical use. When the size of a nonzero block of $\mathbf{x}$, say $\mathbf{x}_j$, is larger or equal to h, it can be recovered by a set of (overlapped) $\mathbf{z}_i$ ($i \in \mathcal{S}$, $\mathcal{S}$ is a non-empty set). When the size of $\mathbf{x}_j$ is smaller than h, it can also be recovered by a $\mathbf{z}_i$ for some i. In this case, the elements of $\mathbf{z}_i$ with global locations different to those of elements of $\mathbf{x}_j$ are very close to zero. By global locations, we mean the indexes in $\mathbf{x}$.

Figure 4 shows a result from an experiment with the same experiment settings as Experiment 3 but without noise. A nonzero block of $\mathbf{x}$, consisting of two elements with global locations being 93 and 94, was recovered by Cluster-SBL (Type II) with $h = 6$. A segment in the recovered signal $\mathbf{x}'$, whose global locations were from 92 to 97, was constructed from an estimated $\mathbf{z}_i$ with some i. We can see the elements in the segment with global locations different from 93 and 94 had very small amplitudes.

Therefore, the parameter h should be better viewed as a regularization parameter, balancing the qualify of the recovered signal and the computational load (larger h requires more computational load). In Experiment 2 we have seen different values of h led to similar performance. So in practice we suggest choosing a small value for h to reduce computational load.

5. CONCLUSION

In this work we proposed two algorithms for the recovery of block sparse signals with or without a priori knowledge on block partition. A uniqueness of them is that they can data-adaptively exploit intra-block correlation. Although spatial correlation is widely studied in other signal processing fields, in the compressed sensing field there is not much work studying the (spatial) intra-block correlation. However, experiments showed that exploiting the intra-block correlation can further improve algorithms' performance. The recovery of block sparse signals with unknown block partition is a more difficult problem, for which few algorithms were proposed. We proposed an approach transforming the difficult problem to a much easier one. Thus many existing algorithms requiring to know block partition can be used. This approach also facilitates the exploitation of intra-block correlation. Experiments showed that our algorithm based on this approach has significant advantages.

6. REFERENCES

[1] M. Yuan and Y. Lin, "Model selection and estimation in regression with grouped variables," *J. R. Statist. Soc. B*, vol. 68, no. 1, pp. 49–67, 2006.

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

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

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

[5] J. Huang, X. Huang, and D. Metaxas, "Learning with dynamic group sparsity," in *2009 IEEE 12th International Conference on Computer Vision*, 2009, pp. 64–71.

[6] L. Yu, H. Sun, J. Barbot, and G. Zheng, "Bayesian compressive sensing for cluster structured sparse signals," *Signal Processing*, vol. 92, pp. 259–269, 2012.

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

[8] M. Tipping, "Sparse Bayesian learning and the relevance vector machine," *Journal of Machine Learning Research*, vol. 1, pp. 211–244, 2001.

[9] Z. Zhang and B. D. Rao, "Sparse signal recovery in the presence of correlated multiple measurement vectors," in *2010 IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP)*, 2010, pp. 3986–3989.