Block-Sparsity Pattern Recovery from Noisy Observations

Jun Fang
National Key Laboratory on Communications
University of Electronic Science and Technology of China
Chengdu 610054, China
Email: Jun.Fang@ieee.org

Hongbin Li
Department of Electrical and Computer Engineering
Stevens Institute of Technology
Hoboken, NJ 07030
Email: hongbin.li@stevens.edu

***Abstract*— We study the problem of recovering the sparsity pattern of block-sparse signals from noise-corrupted measurements. A simple, efficient recovery method, namely, a block-version of the orthogonal matching pursuit (OMP) method, is considered in this paper and its behavior for recovering the block-sparsity pattern is analyzed. We provide sufficient conditions under which the block-version of the OMP can successfully recover the block-sparse representations in the presence of noise. Our analysis reveals that exploiting block-sparsity can improve the recovery ability and lead to a guaranteed recovery for a higher sparsity level. Numerical results are presented to corroborate our theoretical claim.**

***Index Terms*— Block-sparsity, orthogonal matching pursuit, compressed sensing.**

I. Introduction

In this paper, we consider the problem of recovering block-sparse signals whose nonzero elements appear in fixed blocks. Block-sparse signals arise naturally. For example, the atomic decomposition of multi-band signals [1] or audio signals [2] usually results in a block-sparse structure in which the nonzero coefficients occur in clusters. Recovery of block-sparse signals has been extensively studied in [3]–[5], in which the recovery behaviors of the basis pursuit (BP), or ℓ_1-constrained QP, and the orthogonal matching pursuit (OMP) algorithms were analyzed via the restricted isometry property (RIP) [4], [5] and the mutual coherence property [3]. Their analyses [3]–[5] revealed that exploiting block-sparsity yields a relaxed condition which can guarantee recovery for a higher sparsity level as compared with treating block-sparse signals as conventional sparse signals. Nevertheless, most of these studies focused on noiseless scenarios. In practice, measurements are inevitably contaminated with noise and underlying uncertainties. It is therefore important to analyze the effect of measurement noise on the block-sparsity pattern recovery, e.g. under what conditions the exact sparsity pattern can be recovered, and does exploiting block-sparsity still lead to a guaranteed recovery for a higher sparsity level? These questions will be addressed in this paper. Specifically, we consider a block version of the OMP algorithm and study its behavior for recovering block-sparsity pattern in the presence of noise. A comparison with the theoretical results for the conventional OMP algorithm [6] is presented to highlight the benefits of exploiting block-sparsity property.

This work was supported in part by the National Science Foundation under Grant ECCS–0901066, the National Science Foundation of China under Grant 61172114, the Program for New Century Excellent Talents in University (China) under Grant NCET-09-0261, and the Fundamental Research Funds for the Central Universities (China) under Grant ZYGX2010J011.

II. Problem Formulation

We consider the problem of recovering a block-sparse signal $\mathbf{x} \in \mathbb{R}^n$ from noise-corrupted measurements

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{w} \tag{1}$$

where $\mathbf{A} \in \mathbb{R}^{m\times n}$ $(m < n)$ is the measurement matrix with unit-norm columns, and $\mathbf{w}$ is an arbitrary and unknown vector of errors. To define block-sparsity, as in [3], we model $\mathbf{x}$ as a concatenation of equal-length blocks

$$\mathbf{x} = [\mathbf{x}_1^T \ \ \mathbf{x}_2^T \ \ \ldots \ \ \mathbf{x}_L^T]^T \tag{2}$$

where $\mathbf{x}_l \triangleq [x_{(l-1)d+1} \ \ \ldots \ \ x_{ld}]^T$ is a d-dimensional vector. Clearly, the vector $\mathbf{x}$ has a dimension $n = Ld$, and the vector is called block K-sparse if its block component $\mathbf{x}_l$ has nonzero Euclidean norm for at most K indices l. Similarly, the measurement matrix $\mathbf{A}$ can be expressed as a concatenation of column-block matrices $\{\mathbf{A}_l\}_{l=1}^L$

$$\mathbf{A} = [\mathbf{A}_1 \ \ \mathbf{A}_2 \ \ \ldots \mathbf{A}_L] \tag{3}$$

where $\mathbf{A}_l \in \mathbb{R}^{m\times d}$. Also, we assume that the number of rows of $\mathbf{A}$ is an integer multiples of d, i.e. $m = Rd$ with R an integer. The conventional coherence metric of the measurement matrix $\mathbf{A}$ is defined as

$$\mu \triangleq \max_{i\neq j} |\mathbf{a}_i^T \mathbf{a}_j| \tag{4}$$

where $\mathbf{a}_i$ denotes the ith column of $\mathbf{A}$. This coherence metric, albeit useful, is not sufficient to characterize the block-structure of the sparse signal. To exploit the block-sparsity property, we define the block-coherence μ_{B} and sub-coherence ν (these two concepts were firstly introduced in [3]):

$$\begin{aligned} \mu_{\text{B}} &\triangleq \max_{i,j\neq i} \quad \frac{1}{d}\rho(\mathbf{A}_i^T\mathbf{A}_j) \\ \nu &\triangleq \max_{l} \max_{i,j\neq i} \quad |\mathbf{a}_i^T\mathbf{a}_j|, \qquad \mathbf{a}_i, \mathbf{a}_j \in \mathbf{A}_l \end{aligned} \tag{5}$$

where $\rho(\mathbf{X})$ denotes the spectral norm of $\mathbf{X}$, which is defined as the square root of the maximum eigenvalue of $\mathbf{X}^T\mathbf{X}$, i.e. $\sqrt{\lambda_{\max}(\mathbf{X}^T\mathbf{X})}$. Related properties of the block-coherence μ_{B} can be found in [3]. We see that μ_{B} quantifies the coherence between blocks of $\mathbf{A}$, while the coherence within blocks is characterized by the sub-coherence ν.

The objective of this paper is to identify sufficient conditions on the measurement matrix $\mathbf{A}$ (in terms of the block-coherence μ_{B} and the sub-coherence ν), as well as the signal vector $\mathbf{x}$ and the error vector $\mathbf{w}$, under which the block-sparsity pattern can be recovered from the noisy measurements. We are particularly interested in analyzing the recovery ability of a block-version of the orthogonal matching pursuit (OMP). OMP is a simple greedy approximation algorithm developed in [7], [8]. Despite its simplicity, OMP is a provably good approximation algorithm which achieves performance close to Lasso in certain scenarios [9], [10]. In the following, we briefly summarize the block-version of the OMP, which is also termed as block-OMP (BOMP). This BOMP is a slight variant of the original BOMP that was introduced in [3] for noiseless scenarios.

BOMP Algorithm:

1) Initialize the residual $\mathbf{r}_0 = \mathbf{y}$, the index set $S_0 = \varnothing$.
2) At the tth step ($t \geq 1$), we choose the block that is best matched to $\mathbf{r}_{t-1}$ according to
$$i_t = \arg\max_i \|\mathbf{A}_i^T \mathbf{r}_{t-1}\|_2 \tag{6}$$
3) Augment the index set and the matrix of chosen blocks: $S_t = S_{t-1} \cup \{i_t\}$ and $\mathbf{\Psi}^{(t)} = [\mathbf{\Psi}^{(t-1)} \;\; \mathbf{A}_{i_t}]$. We use the convention that $\mathbf{\Psi}^{(0)}$ is an empty matrix.
4) Solve a least squares problem to obtain a new signal estimate $\mathbf{x}_t = \arg\min_{\mathbf{x}} \|\mathbf{y} - \mathbf{\Psi}^{(t)}\mathbf{x}\|_2$
5) Calculate the new residual as $\mathbf{r}_t = \mathbf{y} - \mathbf{\Psi}^{(t)}\mathbf{x}_t = \mathbf{y} - \mathcal{P}_{\mathbf{\Psi}^{(t)}}\mathbf{y}$, where $\mathcal{P}_{\mathbf{\Psi}^{(t)}} = \mathbf{\Psi}^{(t)}(\mathbf{\Psi}^{(t)})^\dagger$ is the orthogonal projection onto the column space of $\mathbf{\Psi}^{(t)}$, and † stands for the pseudo-inverse.
6) If $\|\mathbf{r}_t\|_2 \geq \epsilon$, return to Step 2; otherwise stop.

III. Block-Sparsity Pattern Recovery Analysis

Let $\mathbf{x}_{\mathrm{nz}}$ denote a Kd dimensional column vector constructed by stacking the nonzero block components $\mathbf{x}_l, \forall\{l|\mathbf{x}_l \neq \mathbf{0}\}$, $\mathbf{A}_{\mathrm{nz}} \in \mathbb{R}^{m\times Kd}$ denote a submatrix of $\mathbf{A}$ constructed by concatenating the column-blocks $\mathbf{A}_l, \forall\{l|\mathbf{x}_l \neq \mathbf{0}\}$, i.e. the blocks corresponding to the nonzero $\mathbf{x}_l$, and let $\mathbf{A}_{\mathrm{z}} \in \mathbb{R}^{m\times(L-K)d}$ stand for a submatrix of $\mathbf{A}$ constructed by concatenating the column-blocks $\mathbf{A}_l$ corresponding to zero $\mathbf{x}_l$. For notational convenience, let $I_1 = \{l_1, l_2, \ldots, l_K\}$ denote a set of indices for which $\mathbf{x}_{l_i} \neq \mathbf{0}$, and $I_2 = \{l_{K+1}, l_{K+2}, \ldots, l_L\}$ denote a set of indices for which $\mathbf{x}_{l_i} = \mathbf{0}$. Therefore we can write

$$\begin{aligned}
\mathbf{x}_{\mathrm{nz}} &\triangleq \begin{bmatrix} \mathbf{x}_{l_1}^T & \mathbf{x}_{l_2}^T & \ldots & \mathbf{x}_{l_K}^T \end{bmatrix}^T \\
\mathbf{A}_{\mathrm{nz}} &\triangleq \begin{bmatrix} \mathbf{A}_{l_1} & \mathbf{A}_{l_2} & \ldots & \mathbf{A}_{l_K} \end{bmatrix} \\
\mathbf{A}_{\mathrm{z}} &\triangleq \begin{bmatrix} \mathbf{A}_{l_{K+1}} & \mathbf{A}_{l_{K+2}} & \ldots & \mathbf{A}_{l_L} \end{bmatrix}
\end{aligned}$$

The measurements can therefore be written as

$$\mathbf{y} = \mathbf{A}_{\mathrm{nz}}\mathbf{x}_{\mathrm{nz}} + \mathbf{w} \tag{7}$$

We can decompose the error vector $\mathbf{w}$ into $\mathbf{w} = \mathcal{P}_{\mathbf{A}_{\mathrm{nz}}}\mathbf{w} + \mathcal{P}_{\mathbf{A}_{\mathrm{nz}}}^{\perp}\mathbf{w}$, where $\mathcal{P}_{\mathbf{A}_{\mathrm{nz}}} = \mathbf{A}_{\mathrm{nz}}\mathbf{A}_{\mathrm{nz}}^\dagger$ denotes the orthogonal projection onto the subspace spanned by the columns of $\mathbf{A}_{\mathrm{nz}}$, and $\mathcal{P}_{\mathbf{A}_{\mathrm{nz}}}^{\perp} = \mathbf{I} - \mathcal{P}_{\mathbf{A}_{\mathrm{nz}}}$ is the orthogonal projection onto the null space of $\mathbf{A}_{\mathrm{nz}}^T$. We can further write

$$\begin{aligned}
\mathbf{y} = \mathbf{A}_{\mathrm{nz}}\mathbf{x}_{\mathrm{nz}} + \mathbf{w} &= \mathbf{A}_{\mathrm{nz}}\mathbf{x}_{\mathrm{nz}} + \mathcal{P}_{\mathbf{A}_{\mathrm{nz}}}\mathbf{w} + \mathcal{P}_{\mathbf{A}_{\mathrm{nz}}}^{\perp}\mathbf{w} \\
&= \mathbf{A}_{\mathrm{nz}}(\mathbf{x}_{\mathrm{nz}} + \mathbf{A}_{\mathrm{nz}}^\dagger\mathbf{w}) + \mathcal{P}_{\mathbf{A}_{\mathrm{nz}}}^{\perp}\mathbf{w} \\
&\triangleq \mathbf{A}_{\mathrm{nz}}\tilde{\mathbf{x}}_{\mathrm{nz}} + \tilde{\mathbf{w}}
\end{aligned} \tag{8}$$

where $\tilde{\mathbf{x}}_{\mathrm{nz}} \triangleq \mathbf{x}_{\mathrm{nz}} + \mathbf{A}_{\mathrm{nz}}^\dagger\mathbf{w}$, and $\tilde{\mathbf{w}} \triangleq \mathcal{P}_{\mathbf{A}_{\mathrm{nz}}}^{\perp}\mathbf{w}$. Equation (8) decomposes the measurements into two mutually orthogonal components: a signal component $\mathbf{A}_{\mathrm{nz}}\tilde{\mathbf{x}}_{\mathrm{nz}}$ and a noise component $\tilde{\mathbf{w}}$. The reason for doing so is that even the exact signal support (block-sparsity pattern) is known, there is no way to separate the noise projection term $\mathbf{A}_{\mathrm{nz}}^\dagger\mathbf{w}$ from the true signal $\mathbf{x}_{\mathrm{nz}}$. Hence it is more convenient to carry out our analysis based on (8) instead of (7).

Our main results are summarized as follows.

Theorem 1: Let

$$\omega \triangleq \|\mathbf{A}^T\tilde{\mathbf{w}}\|_{2,\infty} = \max_l \|\mathbf{A}_l^T\tilde{\mathbf{w}}\|_2 \tag{9}$$

denote the maximum correlation between the column block $\mathbf{A}_l$ and the residual noise component $\tilde{\mathbf{w}}$. Let

$$x_{\mathrm{b,min}} \triangleq \min_{l\in I_1} \|\tilde{\mathbf{x}}_l\|_2 \tag{10}$$

the minimum ℓ_2-norm of the non-zero signal block components. Suppose that the following conditions are satisfied

$$\begin{aligned}
&\text{(i)} \quad 1 - (d-1)\nu - (2K-1)d\mu_{\mathrm{B}} > 0 \\
&\text{(ii)} \quad \frac{[1 - (d-1)\nu - (2K-1)d\mu_{\mathrm{B}}]^2}{1 - (d-1)\nu - (K-1)d\mu_{\mathrm{B}}} > \frac{\omega}{x_{\mathrm{b,min}}}
\end{aligned} \tag{11}$$

then we can guarantee that the BOMP algorithm selects indices from I_1 throughout the first K iterations. If the error tolerance ϵ is chosen such that the algorithm stops at the end of iteration K, then the BOMP recovers the exact block-sparsity pattern.

Proof: Proof is omitted here due to space limitations but can be found in our full version manuscript (available at http://arxiv.org/abs/1109.5430). ■

Theorem 1 is a generalization of the results presented in [3] which considered block-sparse signal recovery from noise-free measurements. To see this, for the noiseless case, we have $\omega = 0$, and hence the condition (11) is simplified as

$$1 - (d-1)\nu - (2K-1)d\mu_{\mathrm{B}} > 0 \tag{12}$$

which is exactly the recovery condition provided in [3] for block-sparse signal recovery. On the other hand, for the noisy case, the success of the BOMP algorithm not only depends on the block-coherence μ_{B} and the sub-coherence ν, but also depends on the ratio of the maximum correlation (between the column block $\mathbf{A}_l$ and the residual noise component $\tilde{\mathbf{w}}$) to the minimum ℓ_2-norm of the nonzero signal block components $\tilde{\mathbf{x}}_l, \forall l \in I_1$. The importance of the minimum nonzero signal component in sparsity pattern recovery has been highlighted in [11], [12]. In particular, [11] showed that both the sufficient and necessary conditions require control of the minimum nonzero signal component. Our result suggests that, for block-sparse signal recovery, the minimum ℓ_2-norm of the nonzero signal block components, instead of the minimum magnitude of an entry, is the key quantity that controls the block subset selection.

Also, we observe that the left-hand side of the second condition in (11) is strictly less than one. Therefore the ratio $\omega/x_{\text{b,min}}$ cannot be greater than one, otherwise the condition cannot be met, irrespective of the choice of the sub-coherence ν and the block-coherence μ_{B}. The deterministic condition (11), however, guarantees recovery of the sparsity pattern under the worst-case scenario and therefore is very pessimistic. If we take a probabilistic analysis (as in [13]) that ensures a probabilistic recovery, the condition can be significantly relaxed. This could be a direction of our future study.

IV. Discussions

We note that in this paper, as in [3], block-sparsity is explicitly exploited to yield a more relaxed condition imposed on the measurement matrix, and therefore lead to a guaranteed recovery for a potentially higher sparsity level. If the block-sparse signal is treated as a conventional Kd-sparse vector without exploiting knowledge of the block-sparsity structure, sufficient conditions for exact sparsity pattern recovery using OMP are given in [6, Theorem 18] and can be formulated as (by combining the first and the third equation in [6, Theorem 18])

$$\begin{aligned} &\text{(i)} \quad 1 - 2Kd\mu > 0 \\ &\text{(ii)} \quad \frac{(1-2Kd\mu)^2}{1-Kd\mu} > \frac{\|\mathbf{A}^T\tilde{\mathbf{w}}\|_\infty}{x_{\text{min}}} \end{aligned} \tag{13}$$

where x_{min} denotes the minimum magnitude of the nonzero signal elements in $\tilde{\mathbf{x}}_{\text{nz}}$. When $d = 1$, block-sparsity reduces to conventional sparsity and we have $\nu = 0$, $\mu_{\text{B}} = \mu$. The condition (11) is simplified as

$$\begin{aligned} &\text{(i)} \quad 1 - (2K-1)d\mu > 0 \\ &\text{(ii)} \quad \frac{(1-(2K-1)d\mu)^2}{1-(K-1)d\mu} > \frac{\|\mathbf{A}^T\tilde{\mathbf{w}}\|_\infty}{x_{\text{min}}} \end{aligned} \tag{14}$$

which is the same as (13) except that $2K$ and K in the numerator and denominator are replaced by $2K-1$ and $K-1$, respectively (It can be easily verified that (14) is slightly loose than (13)). When $d > 1$, in the special case that the columns of $\mathbf{A}_l$ are orthonormal for each l, we have $\nu = 0$ and therefore the recovery condition (11) becomes

$$\begin{aligned} &\text{(i)} \quad 1 - (2K-1)d\mu_{\text{B}} > 0 \\ &\text{(ii)} \quad \frac{[1-(2K-1)d\mu_{\text{B}}]^2}{1-(K-1)d\mu_{\text{B}}} > \frac{\omega}{s_{\text{min}}} \end{aligned} \tag{15}$$

This recovery condition, (15), is less restrictive than (13) since we have

$$\begin{aligned} \frac{[1-(2K-1)d\mu_{\text{B}}]^2}{1-(K-1)d\mu_{\text{B}}} &> \frac{(1-2Kd\mu_{\text{B}})^2}{1-Kd\mu_{\text{B}}} \overset{(a)}{\geq} \frac{(1-2Kd\mu)^2}{1-Kd\mu} \\ &> \frac{\|\mathbf{A}^T\tilde{\mathbf{w}}\|_\infty}{x_{\text{min}}} \overset{(b)}{\geq} \frac{\omega}{x_{\text{b,min}}} \end{aligned} \tag{16}$$

where (a) comes from the fact that $1-2Kd\mu > 0$ and $\mu_{\text{B}} \leq \mu$ [3, Proposition 2], (b) follows from $\omega \leq \sqrt{d}\|\mathbf{A}^T\tilde{\mathbf{w}}\|_\infty$ and $x_{\text{b,min}} \geq \sqrt{d}x_{\text{min}}$. We see that through exploiting the block-sparsity, the sparsity pattern recovery condition is relaxed and we can guarantee a recovery of sparsity pattern with a higher sparsity level. A close examination of (16) reveals that this improvement comes from two aspects. First, the measurement matrix requires a less restrictive mutual coherence condition since $\mu_{\text{B}} \leq \mu$. Second, for the same signal, noise, and measurement matrix, the quantity $\omega/x_{\text{b,min}}$ is always smaller than or equal to $\|\mathbf{A}^T\tilde{\mathbf{w}}\|_\infty/x_{\text{min}}$, meaning that exploiting block-sparsity can improve the ability of detecting weak signals buried in noise.

If the individual blocks $\mathbf{A}_l$ are, however, not orthonormal, then $\nu > 0$, and ν has to be small in order to result in a performance gain for block-sparsity recovery as compared with the conventional sparse recovery. We can also follow the orthogonalization approach [3] to analyze the general non-orthonormal case. We orthogonalize the individual blocks $\mathbf{A}_l = \tilde{\mathbf{A}}_l\mathbf{V}_l$, in which $\tilde{\mathbf{A}}_l$ consists of orthonormal columns, and $\mathbf{V}_l$ is an invertible matrix. The original dictionary can therefore be written as $\mathbf{A} = \tilde{\mathbf{A}}\mathbf{V}$, where $\mathbf{V}$ is a block-diagonal matrix with blocks $\mathbf{V}_l$. Clearly, orthogonalization preserves the block-sparsity level. The comparison that is meaningful here is between the recovery based on the original model without exploiting block-sparsity and the recovery based on the orthogonalized model taking block-sparsity into account. For the orthogonalized dictionary $\tilde{\mathbf{A}}$, we have $\nu(\tilde{\mathbf{A}}) = 0$. Therefore we are only concerned about the relation between μ before orthogonalization and μ_{B} after orthogonalization, which are denoted by $\mu(\mathbf{A})$ and $\mu_{\text{B}}(\tilde{\mathbf{A}})$ respectively. Although an exact relation between $\mu(\mathbf{A})$ and $\mu_{\text{B}}(\tilde{\mathbf{A}})$ is difficult to derive, it has been shown in [3] that if $d > RL/(L-R)$, then we have $\mu(\mathbf{A}) \geq \mu_{\text{B}}(\tilde{\mathbf{A}})$. Hence even for general dictionaries, exploiting block-sparsity still leads to a guaranteed sparsity pattern recovery for a potentially higher sparsity level by properly choosing the number of measurements to satisfy $d > RL/(L-R)$.

V. Numerical Results

We present numerical results to illustrate the sparsity pattern recovery performance of the BOMP algorithm. In the simulations, the dictionary is randomly generated with each entry independently drawn from Gaussian distribution with zero mean and unit variance. We then normalize each column of the dictionary to satisfy the unit-norm constraint. The dictionary is divided into consecutive blocks of length d. The support set of the block-sparse signal is randomly chosen according to a uniform distribution, and the signals on the support set

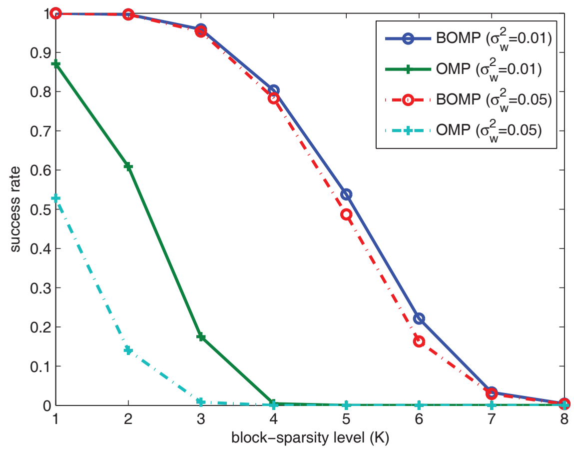

Fig. 1. Sparsity pattern recovery success rates of OMP and BOMP algorithms vs. block sparsity level, $m = 40$, $n = 400$, $d = 4$, and $L = 100$.

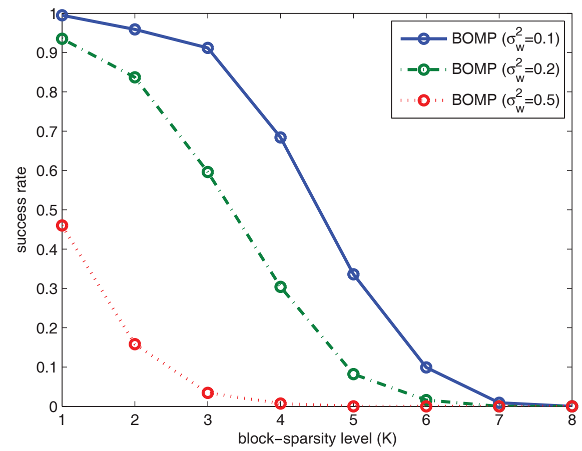

Fig. 2. Sparsity pattern recovery success rate of BOMP algorithm vs. block sparsity level, $m = 40$, $n = 400$, $d = 4$, and $L = 100$.

are i.i.d. Gaussian random variables with zero mean and unit variance. The measurement noise vector is randomly generated with each entry drawn from Gaussian distribution with zero mean and variance σ_w^2.

To show the effectiveness of the BOMP algorithm, we compare it with the OMP algorithm that does not take block-sparsity into account. Fig. 1 shows the sparsity pattern recovery success rate as a function of the block-sparsity level, K. The sparsity pattern recovery is considered successful only if the algorithm determines all the correct support indices in the first K steps for the BOMP or in the first Kd steps for the OMP, supposing the block-sparsity level, K, is known *a priori*. The results are averaged over 1000 Monte Carlo runs, with the dictionary, the signal, and the noise randomly generated for each run. From Fig. 1, we observe that for both the BOMP and the OMP algorithms, the success rate decreases as the block-sparsity level, K, increases. Also, it can be seen that the BOMP algorithm presents a significant performance improvement over the OMP. The result corroborate our theoretical claim that exploiting block-sparsity can lead to an improved recovery ability. Fig. 2 depicts the success rate of the BOMP algorithm under different noise power levels. We see that as the noise power increases, the recovery performance degrades. This observation is quite intuitive and coincides with our theoretical result since a higher noise power calls for a stricter requirement on the measurement matrix in order to satisfy the condition (11).

VI. Conclusion

We studied the problem of recovering the sparsity pattern of block-sparse signals from noise-corrupted measurements. Our results showed that even in the presence of noise, the block-sparsity pattern can still be completely recovered via a block-version of the OMP algorithm when certain conditions are satisfied. Also, our analysis revealed that exploiting block-sparsity can lead to a guaranteed recovery for a potentially higher sparsity level. This theoretical claim was also corroborated by our numerical results.

References

[1] M. Mishali and Y. C. Eldar, "Blind multi-band signal reconstruction: compressed sensing for analog signals," *IEEE Trans. Signal Processing*, vol. 57, no. 3, pp. 993–1009, Mar. 2009.
[2] R. Gribonval and E. Bacry, "Harmonic decomposition of audio signals with matching pursuit," *IEEE Trans. Signal Processing*, vol. 51, no. 1, pp. 101–111, Jan. 2003.
[3] Y. C. Eldar, P. Kuppinger, and H. Bölcskei, "Block-sparse signals: uncertainty relations and efficient recovery," *IEEE Trans. Information Theory*, vol. 58, no. 6, pp. 3042–3054, June 2010.
[4] Y. C. Eldar and M. Mishali, "Robust recovery of signals from a structured union of subspaces," *IEEE Trans. Information Theory*, vol. 55, no. 11, pp. 5302–5316, Nov. 2009.
[5] R. G. Baraniuk, V. Cevher, M. F. Duarte, and C. Hegde, "Model-based compressive sensing," *IEEE Trans. Information Theoy*, vol. 56, no. 4, pp. 1982–2001, Apr. 2010.
[6] J. A. Tropp, "Just relax: convex programming methods for identifying sparse signals in noise," *IEEE Trans. Information Theory*, vol. 52, no. 3, pp. 1030–1051, Mar. 2006.
[7] S. Chen, S. A. Billings, and W. Luo, "Orthogonal least squares methods and their application to non-linear system identification," *Int. J. Control*, vol. 50, no. 5, pp. 1873–1896, Nov. 1989.
[8] Y. C. Pati, R. Rezaiifar, and P. S. Krishnaprasad, "Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition," in *Proc. 27th Annu. Asilomar Conf. Signals, Systems, and Computers*, vol. 1, Pacific Grove, CA, Nov. 1993, pp. 40–44.
[9] J. A. Tropp, "Greed is good: algorithmic results for sparse approximation," *IEEE Trans. Information Theory*, vol. 50, no. 10, pp. 2231–2242, Oct. 2004.
[10] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Information Theory*, vol. 53, no. 12, pp. 4655–4666, Dec. 2007.
[11] M. J. Wainwright, "Information-theoretic limits on sparsity recovery in the high-dimensional and noisy setting," *IEEE Trans. Information Theory*, vol. 55, no. 12, pp. 5728–5741, Dec. 2009.
[12] A. K. Fletcher, S. Rangan, and V. K. Goyal, "Necessary and sufficient conditions for sparsity pattern recovery," *IEEE Trans. Information Theory*, vol. 55, no. 12, pp. 5758–5772, Dec. 2009.
[13] A. K. Fletcher and S. Rangan, "Orthogonal matching pursuit from noisy measurements: a new analysis," in *Twenty-Third Annual Conference on Neural Information Processing Systems*, Vancouver, B.C., Canada, 2009.