Efficient Recovery of Block Sparse Signals by an Improved Algorithm of Block-StOMP

Boxue Huang[1] Tong Zhou[2]

Abstract Many problems arisen in research fields like systems biology and signal processing can be formulated as problems of block sparse signal recovery. Generally, it is known that to pursue the sparsest solution of an underdetermined system of linear equations is non-deterministic polynomial (NP)-hard. To solve block sparse recovery problems, the algorithm of block stagewise orthogonal matching pursuit (Block-StOMP) has been proposed to recover block sparse signals from compressed measurements, which is a greedy algorithm with satisfactory practical performance and some particularly interesting theoretical properties. In this paper, we propose an improved version of Block-StOMP, termed mBlock-StOMP. Specifically, mBlock-StOMP uses the estimated TDR (true discovery rate) to prune support sets of each stage in order to decrease FAR (false alarm rate) and pursue high recovery accuracy. Moreover, rigorous theoretical analysis for mBlock-StOMP is given in this paper. Compared with Block-StOMP, simulation results demonstrate that mBlock-StOMP outperforms Block-StOMP in terms of reconstruction accuracy without increasing computational burden significantly.

Key words Block sparse, compressed sensing, phase diagram, underdetermined system

Citation Boxue Huang, Tong Zhou. Efficient recovery of block sparse signals by an improved algorithm of Block-StOMP. *Acta Automatica Sinica*, 2017, **43**(9): 1607−1618

DOI 10.16383/j.aas.2017.e150116

1 Introduction

The research fields of signal processing, systems biology, and large-scale system analysis have attracted more and more attention in recent years [1]−[6]. They are involved with the problem of recovering sparse signals from underdetermined systems of linear equations as follows [7]:

$$\Phi x = y \tag{1}$$

where $\Phi \in \mathbb{R}^{n\times N}$ $(n < N)$ and $y \in \mathbb{R}^n$ are known, $x \in \mathbb{R}^N$ is the sparse signal to be recovered. The matrix Φ is called measurement matrix. Finding the sparsest solution to (1) can be written as the following ℓ_0 minimization problem:

$$\min_{x\in\mathbb{R}^N} \|x\|_0 \quad \text{s.t. } \Phi x = y. \tag{2}$$

Unfortunately, it is an NP-hard problem to find the solution to (2) in general. However, many algorithms have been proposed to deal with (2) when the measurement matrix and the sparsity satisfy certain conditions [8]. There are two main classes of algorithms to tackle this problem. One is the class of convex optimization algorithms. They are used to solve the following ℓ_1 minimization problem which is a convex relaxation version of (2):

$$\min_{x\in\mathbb{R}^N} \|x\|_1 \quad \text{s.t. } \Phi x = y. \tag{3}$$

As (3) is a convex optimization problem, many standard optimization tools can be used to obtain its solution. The basis pursuit (BP) algorithm [9] is a well-known convex optimization algorithm to solve (3).

Manuscript received July 23, 2015; accepted June 12, 2016.
This work was supported in part by the National Basic Research Program of China (973 Program) (2012CB316504, 2009CB320602), and by the National Natural Science Foundation of China (61174122, 61021063, 60721003 and 60625305), and also by the Specialized Research Fund for the Doctoral Program of Higher Education, China (20110002110045).
Recommended by Associate Editor Xuegong Zhang.
1. Institute of Electronics, Chinese Academy of Sciences, Beijing 100190, China 2. Department of Automation and TNList, Tsinghua University, Beijing 100084, China

The other one is the class of greedy algorithms. For example, the matching pursuit (MP) algorithm [10], the orthogonal matching pursuit (OMP) algorithm [11], and the stagewise orthogonal matching pursuit (StOMP) algorithm belong to this class. Generally, convex optimization algorithms can achieve global optimum but have huge computational costs. Contrarily, greedy algorithms are fast and well suited to large-scale applications.

In this paper, efficient recovery methods to solve a special kind of block sparse signal recovery problems are proposed. In our problem, x is a sparse vector with a special structure. Specifically, Φ and x can be expressed as follows:

$$\Phi = [\underbrace{\phi_1,\ldots,\phi_d}_{\Phi[1]},\underbrace{\phi_{d+1},\ldots,\phi_{2d}}_{\Phi[2]},\ldots,\underbrace{\phi_{N-d+1},\ldots,\phi_N}_{\Phi[M]}]$$

$$x = [\underbrace{x_1,\ldots,x_d}_{x^T[1]},\underbrace{x_{d+1},\ldots,x_{2d}}_{x^T[2]},\ldots,\underbrace{x_{N-d+1},\ldots,x_N}_{x^T[M]}]^T$$

where $N = Md$. Furthermore, x is called k block sparse if it satisfies:

$$\|x\|_{2,0} = \sum_{\ell=1}^{M} I\left(\|x[\ell]\|_2 > 0\right) \le k. \tag{4}$$

Dealing with multi-band signals, or gene expression level measurements is involved with the reconstruction of block sparse signals. Additionally, source localization in sensor networks, multiple-input multiple-output (MIMO) channel equalization, and magnetoencephalography, are also associated with the recovery of block-sparse signals [12]−[14].

There have been many algorithms proposed to recover block sparse signals. For example, in [12] and [15], the BP algorithm is generalized properly to give rise to a mixed ℓ_2/ℓ_1-norm minimization recovery algorithm. Besides, the compressive sampling matching pursuit (CoSaMP) algorithm and the iterative hard thresholding (IHT) algorithm are extended to the block sparse case with provable recovery guarantees and robust properties [13]. Moreover, [12]

proposed the algorithms of block matching pursuit (BMP) and block orthogonal matching pursuit (BOMP), which are the block versions of the MP algorithm and the OMP algorithm, respectively. These two algorithms are both provided theoretical guarantees in terms of block coherence. Moreover, it has been established in [16] that the block version of the StOMP algorithm, termed Block-StOMP, is derived and provided theoretical guarantees in terms of large system limits and phase transition curve.

It can be found that the algorithm of Block-StOMP determines the index of the nonzero blocks of x through hard thresholding. Inspired by the results of [18], Block-StOMP is improved and mBlock-StOMP is proposed to further enhance reconstruction efficiency with pruning support sets of each stage. Many special problems appeared when we tried to derive the convergence of mBlock-StOMP, which are presented and settled eventually in this paper.

The main contributions of this paper can be summarized as follows:

1) The algorithm of Block-StOMP with rigorous theoretical guarantees is modified successfully with improved accuracy of recovery.

2) Theoretical guarantees are provided for mBlock-StOMP from the perspective of extending a conditioned Gaussian model to a conditioned chi-square model.

3) The predicted phase transition curve of mBlock-StOMP is derived and shown in this paper. By depicting the phase diagram and the predicted transition curve of mBlock-StOMP, we see that mBlock-StOMP outperforms Block-StOMP.

Experimental simulations show that mBlock-StOMP has much better performance than Block-StOMP when applied into block sparse signals reconstruction, which evidently shows the significant effectiveness of support sets pruning.

The remainder of this paper is organized as follows. Section 2 starts with reviewing Block-StOMP, and then describes how to improve Block-StOMP. Theoretical guarantees for mBlock-StOMP are presented in Section 3. Section 4 includes the analysis of the predicted phase transition curve. Corresponding experimental simulations and the performance improvement caused by support sets pruning are also reported in Section 4. Finally, Section 5 concludes the whole paper.

Throughout the paper, we denote vectors by lowercase letters, e.g., a, and matrices by uppercase letters, e.g., Φ. For a given set T, $|T|$ and T^c denote its cardinality and complement, respectively. I_N denotes an $N \times N$ identity matrix. For a given matrix D, D^T, D^{-1} denote its transpose and inverse respectively. The Euclidean norm of $\boldsymbol{r}$ is denoted by $\|\boldsymbol{r}\|_2$.

2 An Improved Algorithm of Block-StOMP

2.1 Review of Block-StOMP

From the analysis of Block-StOMP [17], we get to know that the usual set of block matching filter coefficients can be viewed as being sampled from conditional chi-squares distribution and conditional non-central chi-squares distribution, which are corresponding to null case and non-null case respectively. Block-StOMP sets an appropriate threshold value to select the locations of nonzero entries of unknown block sparse signal, and then subtracts the least-square fit to produce a new residual. After a few iterations, a block sparse solution will be achieved.

We consider $\boldsymbol{y} = \Phi\boldsymbol{x}$, where Φ is an $n \times N$ matrix with each column sampled from uniform sphere ensemble (USE), and $\boldsymbol{x}$ is a k block sparse vector, nonzero entries of which follow a symmetric distribution on $\{-1, 1\}$. Denote the support set of $\boldsymbol{x}$ and its block support set as $T = \{i : x(i) \neq 0\}$ and $T' = \left\{i : \|x[i]\|_2 \neq 0\right\}$, respectively. In the algorithm procedure of Block-StOMP, the stage counter s begins at $s = 1$ with initial solution $\boldsymbol{x}_0 = 0$, initial residual $\boldsymbol{r}_0 = \boldsymbol{y}$ and initial support set $I_0 = \varnothing$. At the sth stage, we assume that the sequence of solution is $\boldsymbol{x}_0, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_s$, and the sequence of the estimated support set of vector $\boldsymbol{x}$ is $I_0, I_1, \ldots, I_s$. The maximum number of iteration stages is S. Let I'_{s-1} denote the set of the locations of nonzero blocks of $\boldsymbol{x}_{s-1}$.

Suppose Block-StOMP operates in S stages, and the sequence of residuals is denoted as $\boldsymbol{r}_0, \boldsymbol{r}_1, \ldots, \boldsymbol{r}_S$. The procedure in Table I will be performed at the sth stage of Block-StOMP, where $\Phi[j] = \left(\phi_{(j-1)d+1}, \ldots, \phi_{jd}\right)$, and $\sigma_{s-1} = \|\boldsymbol{r}_{s-1}\|_2/\sqrt{n}$, t_s is a threshold parameter.

TABLE I
BLOCK-STOMP ALGORITHM

Input: measurement matrix Φ, measurement vector $\boldsymbol{y}$ according to $\boldsymbol{y} = \Phi\mathbf{x}$, and block-sparsity level k;
Output: the estimate of $\boldsymbol{x}$

Procedure:
The following steps will be performed repeatedly if the conditions $s < S$ and $\|\boldsymbol{r}_s\|_2 > \varepsilon$ and $J_s \neq \varnothing$ are all satisfied, until the stopping criterion becomes true.
Step 1: Matched Filtering.
$\boldsymbol{c}_s(j) = \left\| \frac{\Phi^T[j]\boldsymbol{r}_{s-1}}{\sigma_{s-1}} \right\|_2^2, j = 1, 2, \ldots, M$.
Step 2: Hard Thresholding.
$J'_s = \{j : \boldsymbol{c}_s(j) > t_s\}$;
$J_s = \{j : 1 + [J'_s(i) - 1]\, d \leq j \leq J'_s(i)d, i = 1, 2, \ldots, |J'_s|\}$.
Step 3: Support Set Update. $I_s = I_{s-1} \cup J_s$.
Step 4: Projection and Pursuit.
$(\boldsymbol{x}_s)_{I_s} = \left(\Phi_{I_s}^T \Phi_{I_s}\right)^{-1} \Phi_{I_s}^T \boldsymbol{y}$.
Step 5: Residual Update. $\boldsymbol{r}_s = \boldsymbol{y} - \Phi\boldsymbol{x}_s$.

In [17], the statistical behavior of block matching filter coefficients $\left\| \frac{\Phi^T[j]\boldsymbol{r}_s}{\sigma_s} \right\|_2^2$ is analyzed, and the following results are obtained:

1) Null case:

$$\mathcal{L}\left(\left\| \Phi^T[j]\, \boldsymbol{r}_s \right\|_2^2 \middle| j \notin T' \cup I'_{s-1} \right) \approx \mathcal{L}\left(\bar{Z}_s \middle| \left|\bar{Z}_i\right| < t_s \overline{\sigma}_i^2,\ i = 1, \ldots, s-1 \right)$$

where

$$\bar{Z}_s/\overline{\sigma}_s^2 \sim \chi_d^2,\ \ \overline{\sigma}_s^2 = p.\lim_{n\to\infty} \frac{\|\boldsymbol{r}_s\|_2^2}{n}.$$

2) Non-null case:

$$\mathcal{L}\left(\left\|\Phi^T[j]\,\boldsymbol{r}_s\right\|_2^2 \middle| j \in T' \cap I'^c_{s-1}\right) \approx \mathcal{L}\left(\bar{X}_s \middle| \left|\bar{X}_i\right| < t_s\bar{\sigma}_i^2,\ i=1,\ldots,s-1\right)$$

where $\bar{X}_s/\overline{\sigma}_s^2 \sim \chi^2_{d,\bar{\lambda}}, \bar{\lambda} = d\left(\bar{\mu}_s/\bar{\sigma}_s\right)^2, \bar{\mu}_s = p.\lim_{n\to\infty} \dfrac{\|\boldsymbol{r}_s\|_2^2}{k_s d}$ and $\overline{\sigma}_s^2 = p.\lim_{n\to\infty} \dfrac{k_s d-1}{k_s d}\dfrac{\|\boldsymbol{r}_s\|_2^2}{n}$

The above conditioned chi-squares model is exploited in [17] to model the statistical behavior of the block matching filter coefficients in the above two cases. It derives the boundary of phase transition curve from the perspective of large-system limits.

In [18], a modified version of StOMP was proposed which can be generalized to the block-sparse case. Denote $\boldsymbol{\phi}_i, \boldsymbol{\phi}_j$ as two columns of the measurement matrix Φ. If the angle between $\boldsymbol{\phi}_i$ and $\boldsymbol{r}_{s-1}$ is close to the angle between $\boldsymbol{\phi}_j$ and $\boldsymbol{r}_{s-1}$, then $|\langle\boldsymbol{\phi}_i, \boldsymbol{r}_{s-1}\rangle|$ will be close to $|\langle\boldsymbol{\phi}_j, \boldsymbol{r}_{s-1}\rangle|$. Consequently, it will be hard to separate these two columns $\boldsymbol{\phi}_i, \boldsymbol{\phi}_j$ from each other through matched filtering at the sth stage of StOMP. Similarly, if this kind of situation happens to $\left\|\frac{\Phi^T[i]\boldsymbol{r}_{s-1}}{\sigma_{s-1}}\right\|_2^2$ and $\left\|\frac{\Phi^T[j]\boldsymbol{r}_{s-1}}{\sigma_{s-1}}\right\|_2^2$ in the block sparse case, it will be difficult to separate them from each other by the steps of matched filtering and hard thresholding.

From this perspective, at each stage of Block-StOMP, the support set which is achieved through matched filtering and hard thresholding is just a rough estimate. Based on the above analysis, an improved algorithm which can obtain a more accurate support set at each stage is needed.

2.2 An Improved Algorithm of Block-StOMP

In this subsection, an improved algorithm of Block-StOMP, termed mBlock-StOMP, is proposed. Notably, the coordinates in I_s are called discoveries, the coordinates in $I_s \cap T$ are called true discoveries and the coordinates in $I_s \cap T^c$ are called false alarms. Firstly, some definitions are introduced:

$$\begin{cases} k_s = k - \sharp \text{ true discoveries prior to stage } s \\ n_s = n - \sharp \text{ discoveries prior to stage } s \\ N_s = N - \sharp \text{ discoveries prior to stage } s \\ \alpha_s = P\left\{\left\|\frac{\Phi^T[j]\boldsymbol{r}_s}{\sigma_s}\right\|_2^2 > t_s | j \in T'^c \cap I'^c_{s-1}\right\} \\ \beta_s = P\left\{\left\|\frac{\Phi^T[j]\boldsymbol{r}_s}{\sigma_s}\right\|_2^2 > t_s | j \in T' \cap I'^c_{s-1}\right\} \\ \boldsymbol{h}_s = \left(\rho_s, d_s, \nu_s, \sigma_s^2\right) = \left(\dfrac{k_s}{n}, \dfrac{n_s}{n}, \dfrac{N_s}{n}, \dfrac{\|\boldsymbol{r}_s\|_2^2}{n}\right) \end{cases}$$

where α_s, β_s are called false alarm rate (FAR) and true discovery rate (TDR), respectively. In [17], the convergence of the state vector $\boldsymbol{h}_s$ at the sth stage is proved:

$$\boldsymbol{h}_s \to \bar{\boldsymbol{h}}_s = \left(\bar{\rho}_s, \bar{d}_s, \bar{\nu}_s, \bar{\sigma}_s^2\right).$$

Moreover, a chi-square random variable $\bar{Z}_s$ and a non-central chi-square random variable $\bar{X}_s$ are defined. Specifically,

$$\frac{\bar{Z}_s}{\bar{\sigma}_s^2} \sim \chi_d^2, \quad \frac{\bar{X}_s}{\bar{\sigma}_s^2} \sim \chi^2_{d,\lambda}$$

where $\lambda = d(\bar{\mu}_s/\bar{\sigma}_s)^2$ is the non-central parameter. Let $\tilde{Z}_s$ denote the random variable $\bar{Z}_s$ conditioned on $\left\{\bar{Z}_s < t_s\bar{\sigma}_i^2, \quad i=1,\ldots,s-1\right\}$. Similarly, let $\tilde{X}_s$ denote the random variable $\bar{X}_s$ conditioned on $\left\{\bar{X}_s < t_s\bar{\sigma}_i^2, \quad i=1,\ldots,s-1\right\}$.

With the above definitions, the following results about the convergence of TDR and FAR are derived in [17].

$$\begin{cases} \alpha_s \to \bar{\alpha}_s = P\left\{\tilde{Z}_s \geq t_s\bar{\sigma}_s^2\right\} \\ \beta_s \to \bar{\beta}_s = P\left\{\tilde{X}_s \geq t_s\bar{\sigma}_s^2\right\}. \end{cases}$$

Based on the above results, in the limit quantities update formulas for the algorithm of Block-StOMP, when the system size n tends to infinity, the dimension-normalized variable ν_s can be updated according to the following equation:

$$\bar{\nu}_{s+1} = \bar{\nu}_s - \bar{\beta}_s\bar{\rho}_s d - \bar{\alpha}_s(\bar{\nu}_s - \bar{\rho}_s d). \tag{5}$$

Straightly, it holds

$$p.\lim_{n\to\infty}\left(\frac{N_{s+1}}{n}\right) = p.\lim_{n\to\infty}\left(\frac{N_s - \Delta N_s}{n}\right)$$

where $\Delta N_s = \beta_s k_s d + \alpha_s(N_s - k_s d)$, and it is the cardinality of the support set J_s. Since ΔN_s indices selected by Block-StOMP at the sth stage are composed of $\beta_s k_s d$ true discoveries and $\alpha_s(N_s - k_s d)$ false alarms, the ratio of true discoveries can be further increased if only $\beta_s k_s d$ true discoveries are included in the support set at the sth stage.

In the following problem, we consider an $n\times N$ matrix Φ with each column sampled from USE. Since $\|\boldsymbol{\phi}_i\|_2 = 1$, $\langle\boldsymbol{\phi}_i, \boldsymbol{\phi}_j\rangle \sim N(0, 1/n)$ for any $i \neq j$, we get to know that $p.\lim_{n\to\infty} \Phi^T\Phi = I_N$. It is easy to obtain

$$p.\lim_{n\to\infty} \Phi^T\Phi\boldsymbol{x} = p.\lim_{n\to\infty} \Phi^T\boldsymbol{y}$$

then,

$$p.\lim_{n\to\infty} \boldsymbol{x} = p.\lim_{n\to\infty} \begin{bmatrix} \boldsymbol{\phi}_1^T\boldsymbol{y} \\ \boldsymbol{\phi}_2^T\boldsymbol{y} \\ \vdots \\ \boldsymbol{\phi}_N^T\boldsymbol{y} \end{bmatrix}. \tag{6}$$

From the perspective of large system limits, the absolute values of matched filter coefficients corresponding to non-zero entries of $\boldsymbol{x}$ are larger than those corresponding to zero entries. On the other hand, at the sth stage of Block-StOMP, the new approximation $\boldsymbol{x}_s$ supported in I_s is $(\boldsymbol{x}_s)_{I_s} = \left(\Phi_{I_s}^T\Phi_{I_s}\right)^{-1}\Phi_{I_s}^T\boldsymbol{y}$. Similarly, we can get

$$p.\lim_{n\to\infty}(\boldsymbol{x}_s)_{I_s} = p.\lim_{n\to\infty} \begin{bmatrix} \boldsymbol{\phi}_{j_1}^T\boldsymbol{y} \\ \boldsymbol{\phi}_{j_2}^T\boldsymbol{y} \\ \vdots \\ \boldsymbol{\phi}_{j_{m_s}}^T\boldsymbol{y} \end{bmatrix}. \tag{7}$$

where $I_s = I_{s-1} \cup J_s$ and $I_s = \{j_1, j_2, \ldots, j_{m_s}\}$.

We denote $(\boldsymbol{x}_s)_{J_s}$ as the new approximation $\boldsymbol{x}_s$ supported in J_s. With (6) and (7), it is shown that J_s obtained at

the sth stage can be further pruned according to the amplitudes of the entries of $(\boldsymbol{x}_s)_{J_s}$. Specifically, a new J_s is derived as the set of $\beta_s k_s d$ indices corresponding to the largest amplitude entries of $(\boldsymbol{x}_s)_{J_s}$. In the next, some particular procedures are proposed to estimate β_s and k_s.

It is shown in [1] that, in the null case, the statistical behavior of $\langle \boldsymbol{\phi}_j, \boldsymbol{r}_s \rangle$ is approximated as follows:

$$L\left(\langle \boldsymbol{\phi}_j, \boldsymbol{r}_s \rangle \,|j \notin T \cup I_{s-1}\right)$$

$$\approx L\left(\bar{Z}_s \,\big|\, |\bar{Z}_i| < t\bar{\sigma}_i, i = 1, 2, \ldots, s-1\right) \approx L(\hat{Z}_s)$$

where $\hat{Z}_s \sim N\left(0, {\sigma_s^{(1)}}^2\right)$.

In the non-null case, it is approximated as follows:

$$L\left(\langle \boldsymbol{\phi}_j, \boldsymbol{r}_s \rangle \,|j \in T \cup I_{s-1}^c\right)$$

$$\approx L\left(\bar{X}_s \,\big|\, |\bar{X}_i| < t\bar{\sigma}_i, i = 1, 2, \ldots, s-1\right) \approx L(\hat{X}_s)$$

where $\hat{X}_s \sim N\left(\mu_s, {\sigma_s^{(2)}}^2\right)$.

From the literature of statistics, the coefficients of the matched filter output at the same stage are statistically independent in the null case, while they are statistically independent as the problem size n tends to infinity in the non-null case. We consider the following quantity in the null case, whose statistical behavior is approximated by chi-square distribution with d degrees of freedom:

$$\left\| \frac{\boldsymbol{\Phi}^T[j]\,\boldsymbol{r}_s}{\sigma_s^{(1)}} \right\|_2^2 = \sum_{i=1}^{d} \left(\frac{\langle \boldsymbol{\phi}_{(j-1)d+1}, \boldsymbol{r}_s \rangle}{\sigma_s^{(1)}} \right)^2 \sim \chi_d^2. \tag{8}$$

In the non-null case, the above quantity follows approximately non-central chi-square distribution with d degrees of freedom and non-central parameter $\hat{\lambda}$:

$$\left\| \frac{\boldsymbol{\Phi}^T[j]\,\boldsymbol{r}_s}{\sigma_s^{(2)}} \right\|_2^2 = \sum_{i=1}^{d} \left(\frac{\langle \boldsymbol{\phi}_{(j-1)d+1}, \boldsymbol{r}_s \rangle}{\sigma_s^{(2)}} \right)^2 \sim \chi_{d,\hat{\lambda}}^2. \tag{9}$$

Based on (8) and (9), the TDR at the sth stage can be estimated by the following formula:

$$\begin{aligned}
\hat{\beta}_s &= \left\{ \left\| \frac{\boldsymbol{\Phi}^T[j]\,\boldsymbol{r}_{s-1}}{\sigma_{s-1}^{(1)}} \right\|_2^2 > t_s \right\} \\
&= P\left\{ \left\| \frac{\boldsymbol{\Phi}^T[j]\boldsymbol{r}_{s-1}}{\sigma_{s-1}^{(2)}} \right\|_2^2 \cdot \left(\frac{\sigma_{s-1}^{(2)}}{\sigma_{s-1}^{(1)}} \right)^2 > t_s \right\} \\
&= P\left\{ \left\| \frac{\boldsymbol{\Phi}^T[j]\boldsymbol{r}_{s-1}}{\sigma_{s-1}^{(2)}} \right\|_2^2 > t_s \left(\frac{\sigma_{s-1}^{(1)}}{\sigma_{s-1}^{(2)}} \right)^2 \right\} \\
&= P\left\{ \chi_{d,\hat{\lambda}}^2 > t_s \left(\frac{\sigma_{s-1}^{(1)}}{\sigma_{s-1}^{(2)}} \right)^2 \right\}
\end{aligned}$$

where $\hat{\lambda} = \sum_{i=1}^{d} \left(\frac{\mu_{s-1}}{\sigma_{s-1}^{(2)}} \right)^2$.

On the other hand, we can take the cardinality of J_s' obtained in the second step of Block-StOMP (hard thresholding) as the estimate of k_s. On the basis of the estimates of β_s and k_s, the set J_s can be pruned according to the amplitude of each entry of the solution $\boldsymbol{x}_s$ at the sth stage.

Suppose mBlock-StOMP operates in S stages, and the sequence of residuals is denoted as $\boldsymbol{r}_0, \boldsymbol{r}_1, \ldots, \boldsymbol{r}_s$. The procedure of mBlock-StOMP is shown in Table II.

TABLE II

mBLOCK-STOMP ALGORITHM

Input: measurement matrix Φ, measurement vector y according to $\boldsymbol{y} = \Phi\boldsymbol{x}$, and block-sparsity level k.

Output: the estimate of the block-sparse signal $\boldsymbol{x}$.

Procedure:

The following procedure will be performed repeatedly if the conditions $s < S$ and $\|\boldsymbol{r}_s\| > \varepsilon$ and $J_s \neq \varnothing$ are all satisfied, otherwise it stops.

Step 1: Matched Filtering.

$\boldsymbol{c}_s(j) = \left\| \frac{\Phi^T[j]\boldsymbol{r}_{s-1}}{\sigma_{s-1}^{(1)}} \right\|_2^2, j = 1, 2, \ldots, M;$

Step 2: Hard Thresholding.

$\tilde{J}_s' = \{j : \boldsymbol{c}_s(j) > t_s\}$, where

$\tilde{J}_s = \left\{ i : 1 + \left[\tilde{J}_s'(j) - 1\right] d \le i \le \tilde{J}_s'(i) d, j = 1, 2, \ldots, \left|\tilde{J}_s'\right| \right\}$

Step 3: Estimate True Discovery Rate $\hat{\beta}_s$.

To calculate $\hat{\beta}_s$, we assume that matched filter coefficients chosen in Step 2 follow Gaussian distribution with nonzero mean:

$\langle \boldsymbol{\phi}_j, \boldsymbol{r}_{s-1} \rangle \sim N\left(\mu_{s-1}, {\sigma_{s-1}^{(2)}}^2\right)$

The estimates of $\mu_{s-1}, \sigma_{s-1}^{(2)}$ can be obtained through maximum likelihood method. Thus, $\hat{\beta}_s$ can be calculated as follows:

$$\hat{\beta}_s = \text{Prob}\left\{ \chi^2_{d,\sum_{i=1}^{d}\left(\frac{\mu_{s-1}}{\sigma_{s-1}^{(2)}}\right)^2} > t_s \cdot \frac{{\sigma_{s-1}^{(1)}}^2}{{\sigma_{s-1}^{(2)}}^2} \right\}$$

and we take the cardinality of $\tilde{J}_s'$ as the estimate of k_s, denoted as $\hat{k}_s$.

Step 4: Pruning $\tilde{J}_s$.

First, update set according to $\tilde{I}_s = I_{s-1} \cup \tilde{J}_s$, and then use Least-squares, $(\boldsymbol{x}_s)_{\tilde{I}_s} = \left(\Phi_{\tilde{I}_s}^T \Phi_{\tilde{I}_s}\right)^{-1} \Phi_{\tilde{I}_s}^T \boldsymbol{y}$; based on the above solution, we can get $(\boldsymbol{x}_s)_{\tilde{J}_s}$ and $(\boldsymbol{x}_s')_{\tilde{J}_s'}$, corresponding to $\tilde{J}_s$, which can be written as $(\boldsymbol{x}_s')_{\tilde{J}_s'} = \left((\boldsymbol{x}_s')_{\tilde{J}_s'}[1], \ldots, (\boldsymbol{x}_s')_{\tilde{J}_s'}[\hat{k}_s]\right)$

where $(\boldsymbol{x}_s')_{\tilde{J}_s'}[j] = \sum_{i=(j-1)d+1}^{jd} ((\boldsymbol{x}_s)_{\tilde{J}_s}(i))^2, j = 1, 2, \ldots, \hat{k}_s$.

Then, sort the elements of $(\boldsymbol{x}_s')_{\tilde{J}_s'}$ in descending order based on their amplitudes, and we take the first $\hat{k}_s \times \hat{\beta}_s$ indices as the components of J_s', and

$J_s = \{ i : 1 + [J_s'(j) - 1]\, d \le i \le J_s'(i) d, j = 1, 2, \ldots, |J_s'| \}.$

Step 5: Support Set Update. $I_s = I_{s-1} \cup J_s$.

Step 6: Projection and Pursuit.

$(\boldsymbol{x}_s)_{I_s} = \left(\Phi_{I_s}^T \Phi_{I_s}\right)^{-1} \Phi_{I_s}^T \boldsymbol{y}$.

Step 7: Residual Update. $\boldsymbol{r}_s = \boldsymbol{y} - \Phi\boldsymbol{x}_s$.

In Table II, $\Phi[j] = \left(\boldsymbol{\phi}_{(j-1)d+1}, \ldots, \boldsymbol{\phi}_{jd}\right)$, $\sigma_{s-1}^{(1)} = \|\boldsymbol{r}_{s-1}\|_2/\sqrt{n}$, and t_s is the α quantile of Chi-square distribution with d degrees of freedom.

Next, we will give a simple example showing that the procedure in Table II works in a specific case.

We generate an $n \times N$ measurement matrix Φ with $n = 100, N = 400, M = 200$, and $d = 2$. The columns of Φ are sampled from USE independently, while $\boldsymbol{x}$ has $k = 20$

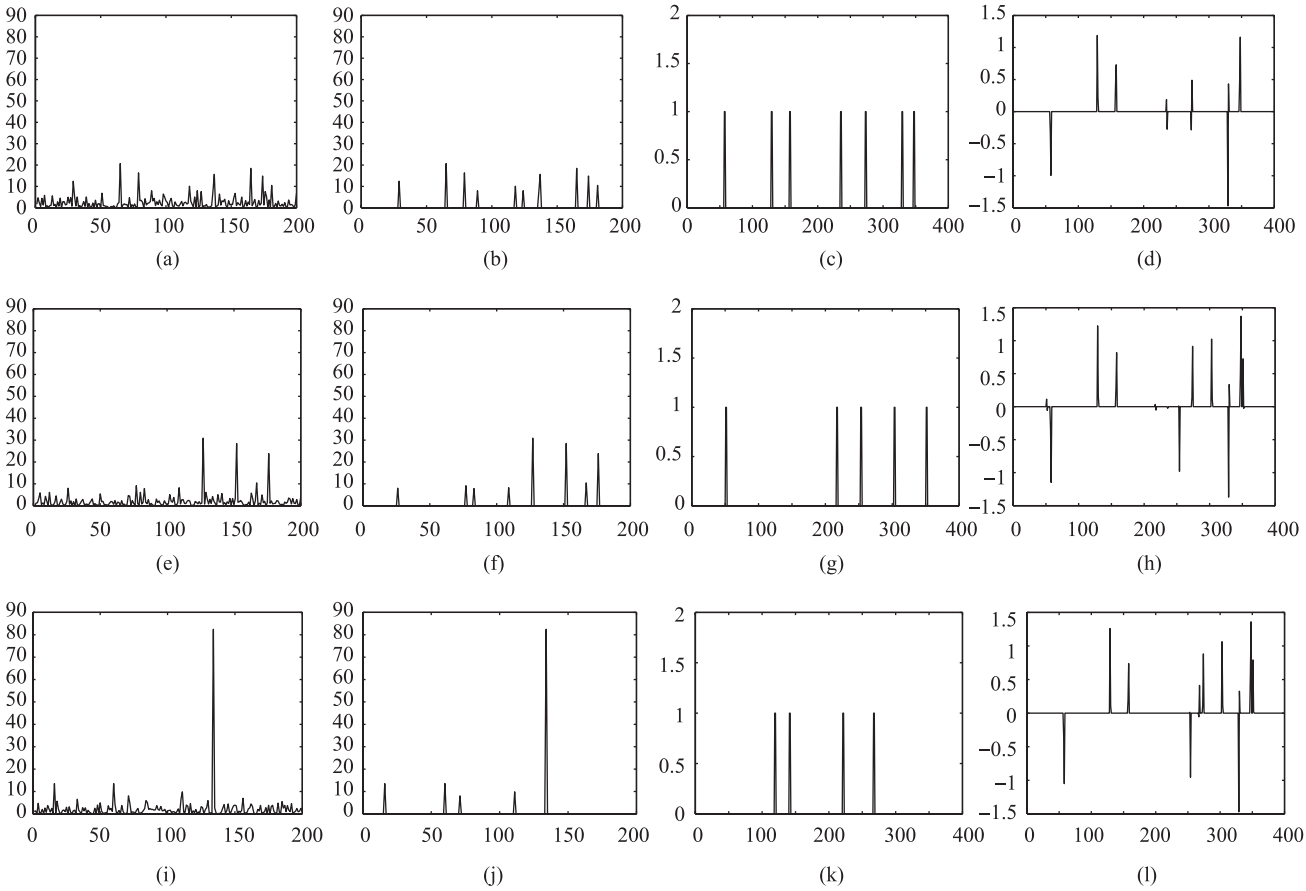


Fig. 1. Progression of the mBlock-StOMP algorithm. Panels (a), (e), and (i) (the 1st column): Successive block matched filtering outputs. Panels (b), (f) and (j) (the 2nd column): Successive hard thresholding results. Panels (c), (g), and (k) (the 3rd column): Successive pruned support sets. Panels (d), (h), and (l) (the 4th column): Successive approximate solutions.

non-zero entries and the non-zero entries of $\boldsymbol{x}$ are drawn from standard Gaussian distribution. The mBlock-StOMP algorithm is applied to this sparse reconstruction problem. The results are shown in Fig. 1. Figs. 1(a)−(l) depict each block matched filtering output, its hard thresholding, pruned support set and the evolving approximate solution.

In this example, Figs. 1(a)−(d) (the 1st row) show the results of the first stage of the procedure of mBlock-StOMP. Specifically, Fig. 1(a) depicts the block matched filter coefficients $\left\| \frac{\Phi^T[j]\boldsymbol{r}_0}{\sigma_0^{(1)}} \right\|_2^2, j = 1, 2, \ldots, M$. After hard thresholding, the remaining block matched filter coefficients which are greater than t_1 are shown in Figs. 1(b). The estimated TDR $\hat{\beta}_1 = 0.5531$. Furthermore, the support set $\tilde{J}_1$ corresponding to Figs. 1(b) is pruned based on $\hat{\beta}_1$. The pruned support set J_1 is depicted in Figs. 1(c). The approximate solution x_1 is shown in Figs. 1(d), $\|\boldsymbol{x}_1 - \boldsymbol{x}\|_2 / \|\boldsymbol{x}\|_2 = 0.5456$.

Figs. 1(e)−(h) (the 2nd row) show the results of the second stage of the procedure of mBlock-StOMP. Specifically, Figs. 1(e) depicts the block matched filter coefficients $\left\| \frac{\Phi^T[j]\boldsymbol{r}_1}{\sigma_1^{(1)}} \right\|_2^2, j = 1, 2, \ldots, M$. After hard thresholding, the remaining block matched filter coefficients which are greater than t_2 are shown in Fig. 1(f). The estimated TDR $\hat{\beta}_2 = 0.6148$. Furthermore, the support set $\tilde{J}_2$ corresponding to Fig. 1(f) is pruned based on $\hat{\beta}_2$. The pruned support set J_2 is depicted in Fig. 1(g). The approximate solution $\boldsymbol{x}_2$ is shown in Fig. 1(h), $\|\boldsymbol{x}_2 - \boldsymbol{x}\|_2 / \|\boldsymbol{x}\|_2 = 0.1473$.

Fig. 1(i)−(l) (the 3rd row) show the results of the third stage of the procedure of mBlock-StOMP. Specifically, Fig. 1(i) depicts the block matched filter coefficients $\left\| \frac{\Phi^T[j]\boldsymbol{r}_2}{\sigma_2^{(1)}} \right\|_2^2, j = 1, 2, \ldots, M$. After hard thresholding, the remaining block matched filter coefficients which are greater than t_3 are shown in Fig. 1(j). The estimated TDR $\hat{\beta}_3 = 0.7389$. Furthermore, the support set $\tilde{J}_3$ corresponding to Fig. 1(j) is pruned based on $\hat{\beta}_3$. The pruned support set J_3 is depicted in Fig. 1(k). The approximate solution $\boldsymbol{x}_3$ is shown in Fig. 1(l), $\|\boldsymbol{x}_3 - \boldsymbol{x}\|_2 / \|\boldsymbol{x}\|_2 = 5.8636 \times 10^{-16}$.

As we can see, after 3 stages an approximate solution is achieved which matches well the true sparse signal $\boldsymbol{x}$. In fact, the relative error at the end of the third stage is reduced to 5.8636×10^{-16}, which shows that a mere three stages were required to obtain an accuracy of 16 decimal digits.

3 Theoretical Analysis of mBlock-StOMP

The problem suite $S(k, n, d, N; USE, \pm 1)$ is considered in this section. A rigorous theoretical analysis of mBlock-StOMP is presented. Specifically, each column of Φ is sam-

pled from USE, $\boldsymbol{x}$ is a block sparse signal with d entries in each block, and $\boldsymbol{y}$ is obtained through $\boldsymbol{y}=\boldsymbol{\Phi x}$. Moreover, $\boldsymbol{x}$ has $\boldsymbol{k}$ nonzero blocks in random positions with nonzero entries following a symmetric distribution on $\{-1,1\}$. We fix $\rho,\delta\in(0,1)$, and $k=\lfloor n\rho\rfloor$, $N=\lfloor n/\delta\rfloor$, where $\lfloor\cdot\rfloor$ represents the floor function. Furthermore, an augmented state vector is defined as follows:

$$\boldsymbol{g}_s=\left(\rho_s,d_s,\nu_s,\sigma_s^2,\mu_s\right)=\left(\frac{k_s}{n},\frac{n_s}{n},\frac{N_s}{n},\frac{\|\boldsymbol{r}_s\|_2^2}{n},\frac{\|\boldsymbol{r}_s\|_2^2}{k_sd}\right).$$

3.1 Main Results

With the above definitions, we state main results of this paper in the form of the following three theorems, and their proofs are given in Appendix A and Appendix B.

Theorem 1: (large-system limits) For $s=1,2,\ldots,S+1$, there are constants $\bar{\sigma}_s,\bar{\rho}_s,\bar{d}_s,\bar{\nu}_s,\bar{\mu}_s$ depending on ρ and δ, so that $\boldsymbol{g}_s\to\bar{\boldsymbol{g}}_s=\left(\bar{\rho}_s,\bar{d}_s,\bar{\nu}_s,\bar{\sigma}_s^2,\bar{\mu}_s\right)$.

To investigate the convergence of α_s,β_s, we introduce chi-square random variable $\bar{Z}_s$ with distribution parameter $\bar{\sigma}_s^2$ defined by Theorem 1, and non-central chi-square random variable $\bar{X}_s$ with $\bar{\mu}_s,\bar{\sigma}_s^2$ defined by Theorem 1. Specifically,

$$\frac{\bar{Z}_s}{\bar{\sigma}_s^2}\sim\chi_d^2,\quad\frac{\bar{X}_s}{\bar{\sigma}_s^2}\sim\chi_{d,\lambda}^2$$

where $\lambda=\sum_{i=1}^{d}\left(\bar{\mu}_i/\bar{\sigma}_i\right)^2$. Let $\tilde{X}_s$ denote the random variable $\bar{X}_s$ conditioned on $\left\{\bar{X}_s<t\bar{\sigma}_i,i=1,\ldots,s-1\right\}$.

In the null case, we present the following theorem:

Theorem 2: (false alarm rate) For the sth stage of the procedure of mBlock-StOMP, and $w\in\mathbb{R}$, when $n\to\infty$, the following result can be obtained:

$$P\left(\left\|\boldsymbol{\Phi}\left[j\right]^T\boldsymbol{r}_s\right\|_2^2\le w\,|\,j\in T'^c\cup I'^c_{s-1}\right)\to P\left\{\bar{Z}_s\le w\right\}$$

then it holds,

$$\alpha_s\to\bar{\alpha}_s=P\left\{\bar{Z}_s\ge t_s\bar{\sigma}_s^2\right\}.$$

In the non-null case, the comparable result is shown as follows:

Theorem 3: (true discovery rate) For the sth stage of the procedure of mBlock-StOMP, and $w\in\mathbb{R}$, when $n\to\infty$, the following result can be obtained:

$$P\left(\left\|\boldsymbol{\Phi}\left[j\right]^T\boldsymbol{r}_s\right\|_2^2\le w\,|\,j\in T'\cap I'^c_{s-1}\right)\to P\left\{\tilde{X}_s\le w\right\}$$

then it holds,

$$\beta_s\to\bar{\beta}_s=P\left\{\tilde{X}_s\ge t_s\bar{\sigma}_s^2\right\}.$$

The proof for Theorem 3 is almost the same as of Theorem 3 in [17], which is thus omitted here.

To clarify the relationship among the above theorems, some notations have to be introduced. Denote Theorem 1_1 as Theorem 1 when $s=1$, and Theorem 1_2 when $s=2$, etc. Similarly, we denote Theorem 2_1 as Theorem 2, and Theorem 3_1 as Theorem 3 when $s=1$, etc. With these notations, Theorem 1 can be divided into many little theorems $1_1,1_2,\ldots,1_{S+1}$, and Theorem 2 and 3 can be divided into theorems $2_1,2_2,\ldots,2_S$ and $3_1,3_2,\ldots,3_S$ respectively. In the subsequent subsections of proofs, Theorem 2_1 and Theorem 3_1 are proved based on Theorem 1_1, and Theorem 1_2 is proved based on Theorem 1_1 ,Theorem 2_1, and Theorem 3_1, etc.

In general, when $s>1$, the proof of Theorem 1_s requires that Theorem 1_ℓ, Theorem 2_ℓ and Theorem 3_ℓ be proved for $1\le\ell<s$.

4 Numerical Simulations

In this section, we perform numerical simulations to observe the empirical performance of mBlock-StOMP. We adopt the same performance metrics with [17] to show the significant effectiveness of mBlock-StOMP.

4.1 Performance Analysis by Phase Transition

In this subsection, the reconstruction performance of mBlock-StOMP is evaluated by phase diagram and phase transition.

Phase diagram for mBlock-StOMP is shown in Fig. 2(b). The diagram displays a rapid transition from perfect reconstruction to perfect disagreement. Moreover, Fig. 2 displays a grid of $\delta-\rho$ values, with δ, ρ ranging through 25 equivalently spaced points in the interval $[0.05,1]$ and $[0.05/d,1/d]$ respectively. Here $N=800,d=2$. Each point on the grid shows the mean number of coordinates at which original and reconstruction differ by more than 10^{-4}, averaged over 100 independent realizations of the standard problem suite $S_{\rm st}(k,n,d,N)$. Specifically, each column of $\boldsymbol{\Phi}$ is sampled from USE, and $\boldsymbol{x}$ is a block sparse vector with k nonzero blocks in random positions, and these nonzero entries follow the standard Gaussian distribution. Take Fig. 2(a) for example, it can be seen that this diagram displays a phase transition. For small ρ, it seems that highly accurate reconstruction is obtained, while for large ρ reconstruction fails. The transition from success to failure occurs at different ρ for different values of δ.

In the next, we will show that based on the derived theoretical results, the transition curve of mBlock-StOMP from success to failure can be accurately predicted. The quantities of $\bar{\alpha}_s,\bar{\beta}_s,\bar{\rho}_s,\bar{\nu}_s,\bar{d}_s$ can be approximately calculated and updated on the basis of their initial values and the following iterative formulas:

$$\begin{cases}\tilde{d}_{s+1}=\tilde{d}_s-\tilde{\beta}_s\cdot\tilde{\rho}_s\cdot d\\ \tilde{\nu}_{s+1}=\tilde{\nu}_s-\tilde{\beta}_s\cdot\tilde{\rho}_s\cdot d\\ \tilde{\rho}_{s+1}=\tilde{\rho}_s-\tilde{\beta}_s\cdot\tilde{\rho}_s\\ \tilde{\alpha}_s=P\left\{\left\|\frac{\boldsymbol{\Phi}^T[j]\boldsymbol{r}_s}{\tilde{\sigma}_s}\right\|_2^2>t_s\right\}=P\left\{\chi_d^2>t_s\right\}\\ \tilde{\beta}_s=P\left\{\chi_d^2>t_s\tilde{\sigma}_s^2\right\}\end{cases}\tag{10}$$

where $\tilde{\sigma}_s=\sqrt{(k_sd)/n_s}$. The above heuristic computation formulas of $\bar{\alpha}_s,\bar{\beta}_s$ in the block sparse case were proposed by [17].

It is shown in [17] that there are two thresholding strategies to determine the value of t_s: control false alarm rate (CFAR), and control false discovery rate (CFDR). According to the CFAR thresholding, t_s is chosen as the $\left[1-\frac{\delta(1-d\rho)}{S(1-d\delta\rho)}\right]$ quantile of chi-square distribution with d degrees of freedom. If the CFDR thresholding strategy is chosen, the number of false discoveries cannot exceed a

fixed fraction q of all discoveries. As it is reported in [1] that it is more appropriate to use the strategy of CFAR in the noiseless case, we use this thresholding strategy when investigating the phase diagram and the predicted phase transition curve. In the experimental simulations, the CFDR thresholding is used when the measurement noise level is high. It will be elaborated in the next section.

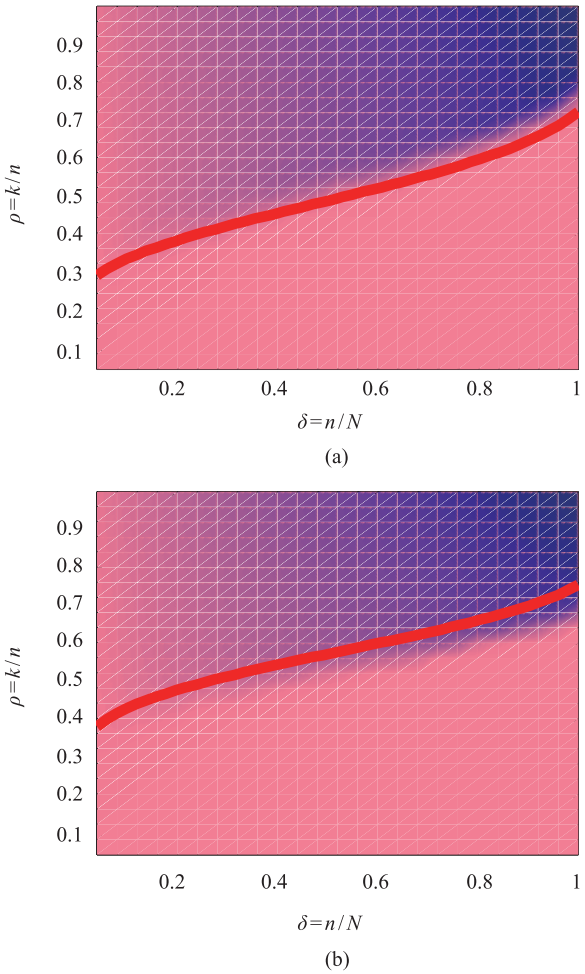


Fig. 2. Phase diagrams and predicted phase transition curves. (a) Block-StOMP. (b) mBlock-StOMP.

Assume that $\tilde{\rho}_s, \tilde{d}_s, \tilde{\nu}_s$ are initialized as $\{k/n, 1, N/n\}$, then $\tilde{\alpha}_s, \tilde{\beta}_s, \tilde{\rho}_s, \tilde{d}_s, \tilde{\nu}_s$ can be updated according to (22). With different values of $\delta - \rho$, we achieve a predicted phase transition curve, which can be shown as an overlaid curve in Fig. 2 (b). The calculation announces success, at or before stage S,

$$\bar{d}_s \geq \eta, \quad \bar{\rho}_s \leq \eta\rho$$

where $\eta = 0.001$. Otherwise, it announces failure. Thus, the phase transition curve can be plotted according to the above results. It can be seen from Fig. 2(b) that the agreement between the simulation result and the predicted transition curve is reasonably good.

The phase diagram and phase transition of Block-StOMP is shown in Fig. 2(a). With the comparison between Figs. 2(a) and (b), the predicted phase transition curve of mBlock-StOMP significantly outperforms that of Block-StOMP when δ is not very large. Moreover, we see that the performance difference between Block-StOMP and mBlock-StOMP becomes less and less with the increment of the value of δ.

4.2 Performance Comparison With Other Recovery Algorithms

In this subsection, we adopt the following metrics to evaluate the effectiveness of different recovery algorithms: frequency of successful reconstruction, number of iterations and running time. To calculate the frequency of successful reconstruction, the strategy mentioned in [1] is used. Denote $\boldsymbol{x}_s$ and $\boldsymbol{x}$ respectively as the estimate and real value of the signal to be recovered. We declare that the sparse signal x is reconstructed successfully if $\boldsymbol{x}_s$ satisfies the following formula:

$$relerr = \frac{\|\boldsymbol{x}_s - \boldsymbol{x}\|_2}{\|\boldsymbol{x}\|_2} \leq \epsilon_x$$

where ϵ_x takes the fixed value of 10^{-4} in the following numerical simulations.

In this paper, we compare mBlock-StOMP with the original algorithm of Block-StOMP and four other recovery algorithms: the OMP algorithm, the generalized orthogonal matching pursuit (gOMP-2) algorithm [17], the StOMP algorithm, and the BOMP algorithm.

Suppose block-sparsity levels are from 1 to 35. We perform 500 trials for each block sparsity, and then compute the corresponding frequency of successful reconstruction for each recovery algorithm. In each trial, we generate an $n \times N$ measurement matrix Φ with $n = 100, N = 400$, and $d = 2$. The columns of Φ are sampled from USE independently, while the non-zero entries of $\boldsymbol{x}$ are drawn from standard Gaussian distribution. In the next, we investigate the effectiveness of the above mentioned algorithms in the noiseless and noisy settings respectively.

We consider that Φ and $\boldsymbol{y}$ are both contaminated by white Gaussian noise in the noisy setting. Let $\hat{\Phi}$ and $\hat{\boldsymbol{y}}$ denote the noisy measurement matrix and the noisy observation vector respectively. Then, $\hat{\Phi}$ and $\hat{\boldsymbol{y}}$ obey the following constraints:

$$\begin{cases} \boldsymbol{y} = \Phi\boldsymbol{x} \\ \hat{\boldsymbol{y}} = \boldsymbol{y} + \tilde{\boldsymbol{y}} \ \hat{\Phi} = \Phi + \tilde{\Phi}, \end{cases}$$

where the elements of $\tilde{\boldsymbol{y}}$ and $\tilde{\Phi}$ are independently identical distribution. $N(0, \eta^2)$ random variables. The vector $\boldsymbol{x}$ is to be recovered based on known $\hat{\Phi}$ and $\hat{\boldsymbol{y}}$.

Due to space limitations, we have considered only four different noise levels, and performed experimental simulations with $\eta \in \{0, 5 \times 10^{-6}, 1 \times 10^{-5}, 2 \times 10^{-5}\}$.

To enhance recovery performance, the CFAR thresholding is used to choose the value of hard thresholding when the noise level is moderate, and t_s is chosen as the $\left[1 - \frac{\delta(1-d\rho)}{S(1-d\delta\rho)}\right]$ quantile of chi-square distribution with d degrees of freedom.

However, the strategy of CFDR thresholding is utilized instead of CFAR thresholding when the noise level becomes larger.

Fig. 3 shows each algorithm's performance in terms of the frequency of successful reconstruction under different noise levels. We find that the performance of Block-StOMP is improved a lot after pruning support sets. From Fig. 3, we

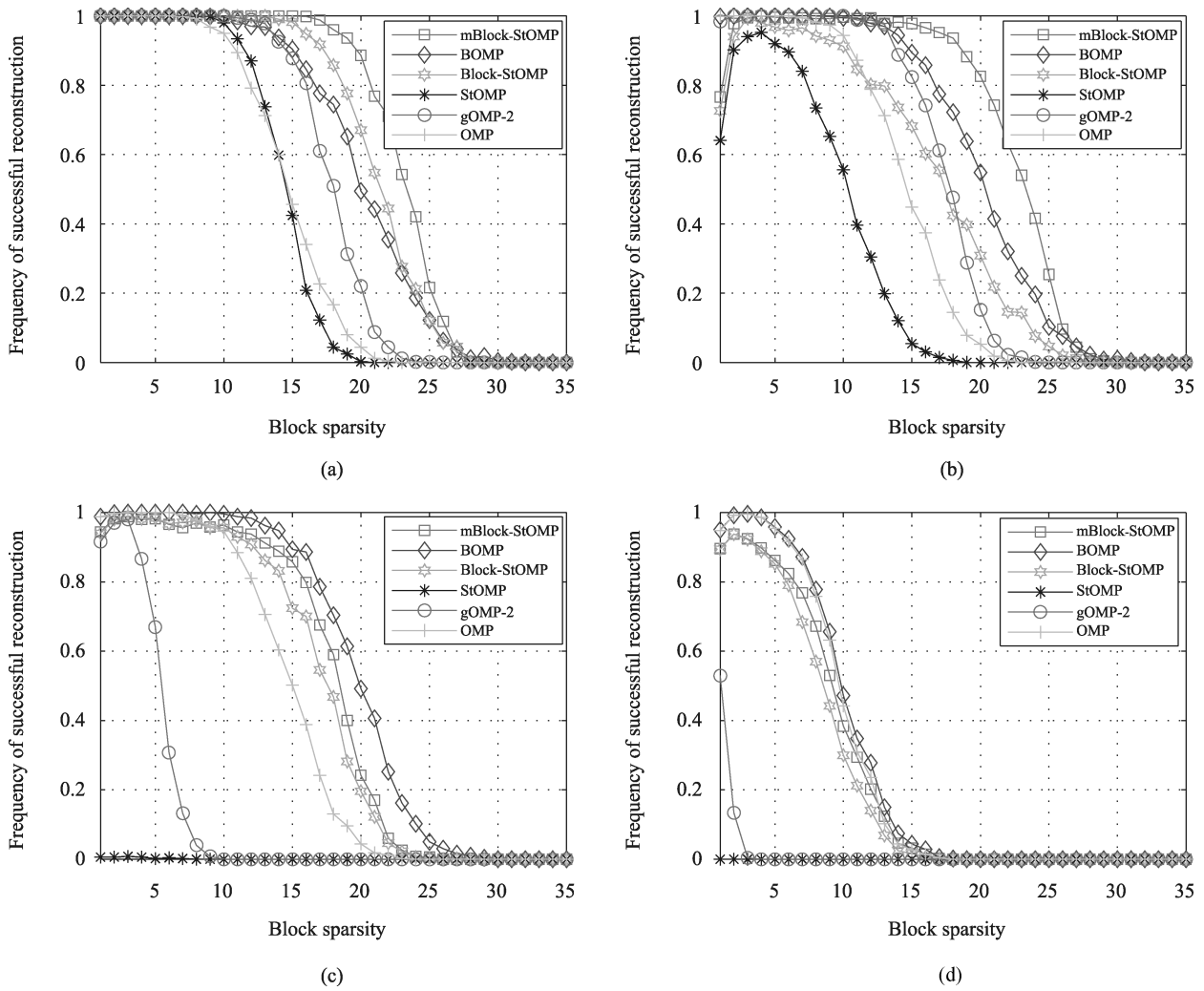


Fig. 3. Frequency of successful reconstruction as a function of block sparsity. (a) CFAR thresholding, $\eta = 0$, $S = 10$. (b) CFAR thresholding, $\eta = 5 \times 10^{-6}$, $S = 10$. (c) CFDR thresholding, $\eta = 1 \times 10^{-1}$, $q = 0.2$. (d) CFDR thresholding, $\eta = 2 \times 10^{-5}$, $q = 0.1$.

find that, using the CFAR thresholding, mBlock-StOMP performs best in terms of frequency of successful reconstruction when $\eta = 0$ in the noiseless case and $\eta = 5 \times 10^{-6}$ in the noisy case. We can see that mBlock-StOMP is robust to measurement noise when the noise level is moderate. Furthermore, it can be seen that BOMP tends to be more robust to measurement noise compared with mBlock-StOMP when the noise level is increased.

The curves of the number of iterations for each algorithm are presented in Fig. 4, and the curves of the running time for them are shown in Fig. 5. The running time is the physical running time in a personal computer. It is measured using the MATLAB(R2009a) program under Core i5-2400 64-bit processor and Windows 7 environment. It is obvious that mBlock-StOMP performs almost the same as Block-StOMP and outperforms all other algorithms in terms of the number of iterations. Moreover, mBlock-StOMP has less running time than BOMP, gOMP-2 and OMP when the block sparsity is more than a certain threshold value. Furthermore, mBlock-StOMP, with the CFDR thresholding, has almost the same running time as Block-StOMP and StOMP when the block sparsity is over one threshold value.

5 Conclusions

In this paper, an improved algorithm of the Block-StOMP algorithm, called mBlock-StOMP, is proposed. At each stage, similar to Block-StOMP, mBlock-StOMP performs matched filtering on the residual, and determines a rough support set of the sparse vector $\boldsymbol{x}$ through hard thresholding.

The novelty of mBlock-StOMP lies in that the TDR at each stage is estimated and used to prune the support set, and then a more accurate support set is obtained. Moreover, a rigorous convergence analysis is provided for mBlock-StOMP. In the numerical simulations, compared with other recovery algorithms, based on the original algorithm of Block-StOMP and some modifications, mBlock-StOMP has attractive performances in terms of the frequency of successful reconstruction, number of iterations and running time in the noiseless and noisy settings.

In our paper, the tools of phase diagram and phase transition are explored to rate the success of mBlock-StOMP. Based on the theoretical results of large system limits, the predicted phase transition curve of mBlock-StOMP is derived. Through numerical simulations, we find that the

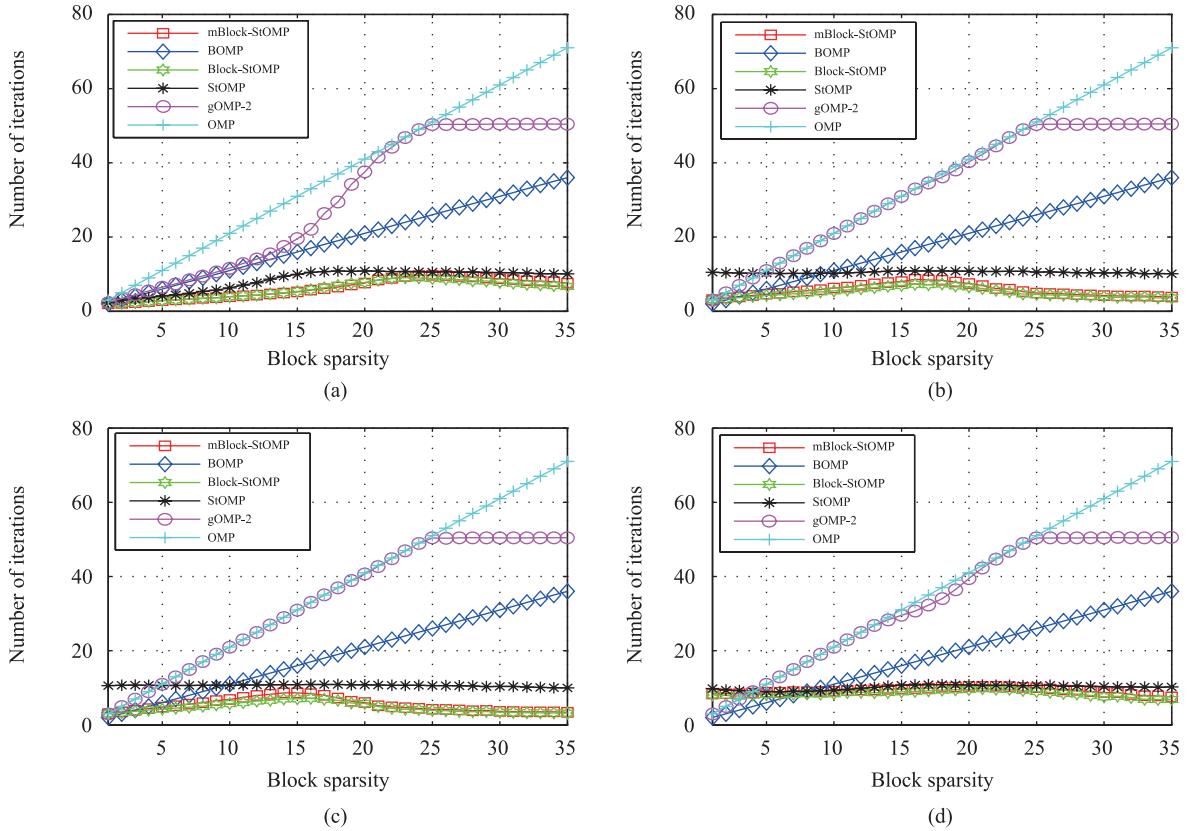


Fig. 4. Number of iterations as a function of block sparsity. (a) CFAR thresholding, $\eta = 0$, $S = 10$. (b) CFAR thresholding, $\eta = 5 \times 10^{-6}$, $S = 10$. (c) CFDR thresholding, $\eta = 1 \times 10^{-1}$, $q = 0.2$. (d) CFDR thresholding, $\eta = 2 \times 10^{-5}$, $q = 0.1$.

simulation result of the phase diagram of mBlock-StOMP matches the predicted phase transition curve quite well.

As many problems of network topological identification, like gene regulatory network identification, can be treated as block sparse signals recovery problems [6], mBlock-StOMP can be applied to these problems. It is our another line of research in the future.

Appendix

A. Proof of Theorem 1

The aforementioned augmented state vector is rewritten below:

$$\boldsymbol{g}_s = \left(\frac{k_s}{n}, \frac{n_s}{n}, \frac{N_s}{n}, \frac{\|\boldsymbol{r}_s\|_2^2}{n}, \frac{\|\boldsymbol{r}_s\|_2^2}{k_s d} \right), \ 1 \le s \le S+1.$$

According to the procedure of mBlock-StOMP, it is obvious to derive the following updating formulas:

$$\begin{cases} k_{s+1} = k_s - \dfrac{|J_s \cap T|}{d} \\ n_{s+1} = n_s - |J_s| \\ N_{s+1} = N_s - |J_s| \end{cases}$$

where, $J'_s = \left\{ j : \left\| \Phi[j]^T \boldsymbol{r}_s \right\|_2^2 > t\sigma_s^2, j \in T' \cap I'^c_{s-1} \right\}$, $J_s = \{ j : 1 + [J'_s(i) - 1]\, d \le j \le J'_s(i) d, i = 1, 2, \ldots, |J'_s| \}$.

Based on the above updating formulas, we derive that $\boldsymbol{g}_s = \boldsymbol{g}_\ell, \ell \le s \le S+1$ if the algorithm of mBlock-StOMP stops at the ℓ-th ($\ell < S$) stage. Define random vectors: $G_s = (\boldsymbol{g}_1, \ldots, \boldsymbol{g}_s)$ and $G'_s = (\boldsymbol{g}'_1, \ldots, \boldsymbol{g}'_s)$, then the distance measure between them is: $d(G_s, G'_s) = \max_{1 \le \ell \le s} d(\boldsymbol{g}_\ell, \boldsymbol{g}'_\ell)$. To prove Theorem 1, the large system limit of G_{S+1} has to be proved to exist, that is, if δ, ρ and d are given, $G_{S+1,n}(\delta, \rho, d)$ should satisfy:

$$P_n \left\{ d(G_{S+1,n}, \bar{G}_{S+1}) > \epsilon \right\} \to 0, n \to \infty. \tag{11}$$

The existence of a deterministic sequence $(\bar{\boldsymbol{g}}_s, s = 1, \ldots, S+1)$ is proved in the next. Moreover, this sequence satisfies: $\bar{\boldsymbol{g}}_{s+1} = \bar{\boldsymbol{g}}_s - \Delta_s\left(\bar{G}_s\right), s = 1, \ldots, S$, and for each $\epsilon > 0$,

$$P_n \left\{ d(\boldsymbol{g}_{s+1,n}, \bar{\boldsymbol{g}}_{s+1}) > 2\epsilon \,\middle|\, d(\boldsymbol{g}_{s,n}, \bar{\boldsymbol{g}}_s) < \epsilon \right\} \to 0, \ n \to \infty. \tag{12}$$

If the above results are achieved, (10) can be proved. Thereafter, for $1 \le i \le 5$, $\Delta_{s,i}\left(\bar{G}_s\right)$ is used to represent the i-th component of $\Delta_s\left(\bar{G}_s\right)$.

When $s = 1$, obviously, we have

$$\boldsymbol{g}_1 = \left(\frac{k}{n}, 1, \frac{N}{n}, \frac{\|\boldsymbol{y}\|_2^2}{n}, \frac{\|\boldsymbol{y}\|_2^2}{kd} \right).$$

We assume $\boldsymbol{y} = \sum_{\ell=1}^{k} \pm_\ell \boldsymbol{\phi}_{i_\ell}$ in the case of $s = 1$, then $p.\lim_{n\to\infty} \|\boldsymbol{y}\|_2^2 / n = \lim_{n\to\infty} k_n / n = \rho$. Hence, we get the following lemma:

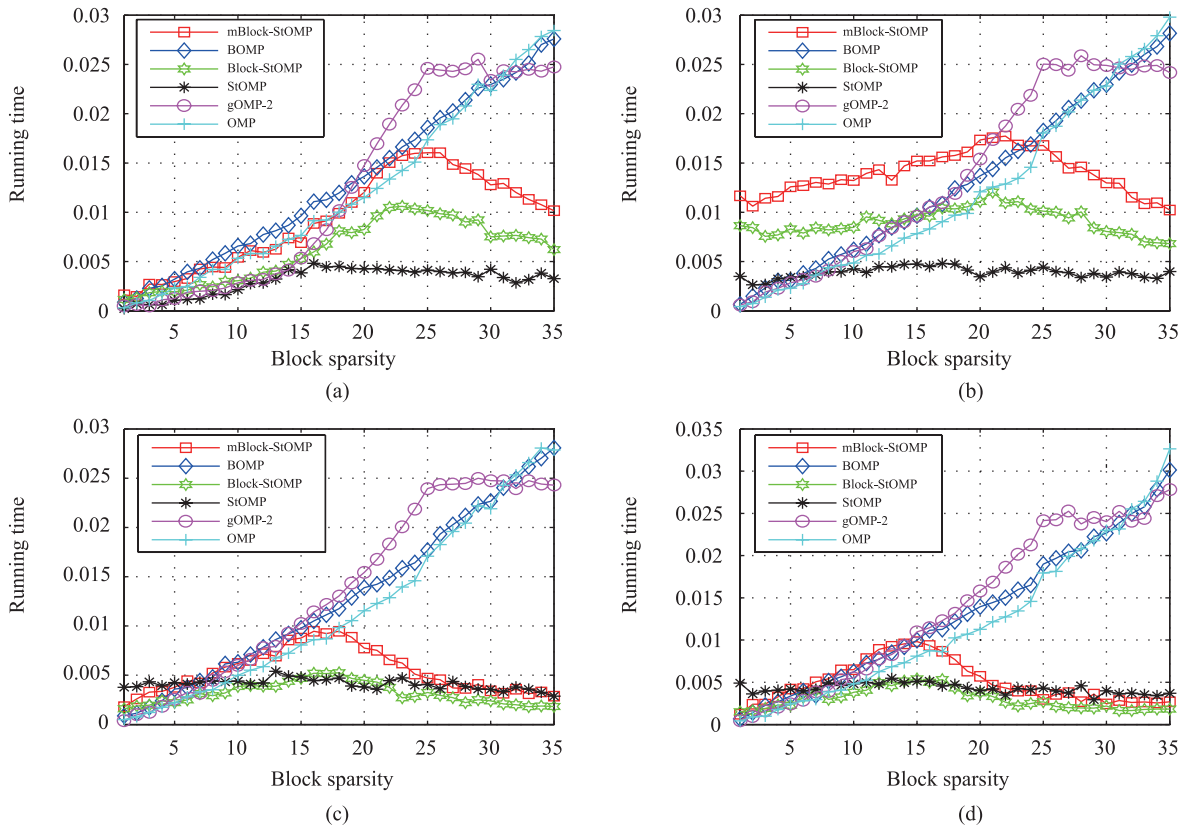


Fig. 5. Running time as a function of block sparsity. (a) CFAR thresholding, $\eta = 0$, $S = 10$. (b) CFAR thresholding, $\eta = 5 \times 10^{-6}$, $S = 10$. (c) CFDR thresholding, $\eta = 1 \times 10^{-1}$, $q = 0.2$. (d) CFDR thresholding, $\eta = 2 \times 10^{-5}$, $q = 0.1$.

Lemma 1: As $n \to \infty$, and $n/N_n \to \delta$, $k_n/n \to \rho$, it can be achieved that $p.\lim_{n\to\infty} \boldsymbol{g}_{1,n} = \bar{\boldsymbol{g}}_1(\delta, \rho) = (\rho, 1, \delta^{-1}, \rho, d^{-1})$.

To investigate the relationship between $\bar{\boldsymbol{g}}_{s+1}$ and $\bar{\boldsymbol{g}}_s$, we start with analyzing the first component of $\boldsymbol{g}_s$. According to the above analysis, k_{s+1} and k_s should satisfy the following equation:

$$k_{s+1} = k_s - \frac{|J_s \cap T|}{d} = k_s - |J'_s \cap T'|$$

$$|J'_s \cap T'| = \sum_{j \in T^c_{s-1} \cap T'} 1_{\left\{\|\Phi[j]^T \boldsymbol{r}_s\|_2^2 > t\sigma_s^2\right\}} = \sum_{\ell=1}^{k_s} V_{\ell,s}.$$

It is obvious that $(\Phi[j], j \in T')$ are all statistically independent, and the columns of $(\Phi[j], j \in I'^c_{s-1} \cap T')$ belong to $(\Phi[j], j \in I'^c_{s-1})$. It implies that random variables $\left(\left\|\Phi[j]^T \boldsymbol{r}_s\right\|_2^2, j \in I'^c_{s-1} \cap T'\right)$, conditional on I'_{s-1}, are independent and identically distributed, which are conditionally exchangeable random variables. Then we obtain the following lemma.

Lemma 2: Under the condition G_s, $(V_{\ell,s} : \ell = 1, 2, \ldots, k_s)$ are exchangeable random variables.

Suppose either that $s = 1$ or that $s > 1$ and (11) has been proved for $1 \le \ell \le s-1$. Let $\tilde{X}_s$ denote the conditioned non-central chi-square random variable defined by the sequence $(\bar{\mu}_\ell : 1 \le \ell \le s)$, $(\bar{\sigma}_\ell : 1 \le \ell \le s)$, and $(\lambda_\ell : 1 \le \ell \le s)$. Here, $\bar{\mu}_s = \bar{\boldsymbol{g}}_s(4)/\bar{\boldsymbol{g}}_s(1)$, $\bar{\sigma}_s = \sqrt{\bar{\boldsymbol{g}}_s(4)}$, $\lambda_s = d\left(\bar{\mu}_s/\bar{\sigma}_s\right)^2$.

Suppose that Theorem 1_ℓ, Theorem 2_ℓ and Theorem 3_ℓ, for $1 \le \ell < s$, are already proved, then the following lemma is addressed:

Lemma 3: Based on the above definitions, we obtain:

$$\lim_{n\to\infty} P_n\left(\left\|\Phi[j]^T \boldsymbol{r}_s\right\|_2^2 > t\sigma_s^2 \,|\, i \in T' \cap I'^c_{s-1}, T', I'_{s-1}\right) = P\left\{|\tilde{X}_s| > t\bar{\sigma}_s^2\right\}.$$

Lemma 4: $p.\lim_{n\to\infty} \mathrm{Cov}(V_{1,s}, V_{2,s}|G_s) = 0$.

Combining the above three lemmas, it is clear that

$$\frac{|J_s \cap T|}{nd} = \frac{|J'_s \cap T'|}{n} \to \frac{k_s}{n} P\left\{|\tilde{X}_s| > t\bar{\sigma}_s^2\right\}, n \to \infty.$$

Finally, we present Lemma 5:

Lemma 5: With the above three lemmas, we can derive the following equation

$$\frac{k_{s+1}}{n} = \frac{k_s}{n}\left(1 - P\left\{\left|\tilde{X}_s\right| > t\bar{\sigma}_s^2\right\}\right) + o_p(1), n \to \infty \quad (13)$$

where $P\left\{\left|\tilde{X}_s\right| > t\bar{\sigma}_s^2\right\}$ only depends on $\bar{G}_s$. Furthermore, on the condition that $k_s/n \to_p \bar{\boldsymbol{g}}_s(1)$ is proved at a previous stage of the argument, (12) can be rewritten as

$$\bar{\boldsymbol{g}}_{s+1}(1) = \bar{\boldsymbol{g}}_s(1) - \Delta_{s,1}(\bar{G}_s) \quad (14)$$

where $\Delta_{s,1}(\bar{G}_s) = \bar{g}_s(1) \cdot P\left\{\left|\tilde{X}_s\right| > t\bar{\sigma}_s^2\right\}$.

Similarly, for the second and third component of the state vector, the following equations can be derived:

$$\bar{\boldsymbol{g}}_{s+1}(2) = \bar{\boldsymbol{g}}_s(2) - \Delta_{s,2}(\bar{G}_s) \quad (15)$$

$$\bar{\boldsymbol{g}}_{s+1}(3) = \bar{\boldsymbol{g}}_s(3) - \Delta_{s,3}(\bar{G}_s) \quad (16)$$

In the next, we will give the proof that the fourth component of the state vector also satisfies the equation similar to (13)−(15). Based on the analysis of $\boldsymbol{r}_{s+1}$ and $\boldsymbol{r}_s$, we have the following equation:

$$\|\boldsymbol{r}_{s+1}\|_2^2 = \|\boldsymbol{r}_s\|_2^2 - \boldsymbol{c}_s^T \left(\Phi_{J_s}^T \Phi_{J_s}\right)^{-1} \boldsymbol{c}_s \tag{17}$$

where $\boldsymbol{c}_s = \Phi_{J_s}^T \boldsymbol{r}_s$. Define $\bar{\rho}_s = \bar{\boldsymbol{g}}_s(1)$, and $\Gamma_s(\bar{G}_s) = \bar{\rho}_s \int x \cdot 1_{\left\{j:\left\|\Phi[j]^T \boldsymbol{r}_s\right\|_2^2 > t\sigma_s^2\right\}} p_{\tilde{X}_s}(x)dx$.

Lemma 6: $\underset{n\to\infty}{p.\lim} \|\boldsymbol{c}_s\|_2^2 / n = \dfrac{\Gamma_s(\bar{G}_s)}{d}$.

Proof:

$$\begin{aligned}
&\|\boldsymbol{c}_s\|_2^2 \\
&= \sum_{i \in I'^c_{s-1} \cap T'} \left\|\Phi[j]^T \boldsymbol{r}_s\right\|_2^2 1_{\left\{\left\|\Phi[j]^T \boldsymbol{r}_s\right\|_2^2 > t\sigma_s^2\right\}} \\
&\sim k_s \int x 1_{\left\{\left\|\Phi[j]^T \boldsymbol{r}_s\right\|_2^2 > t\sigma_s^2\right\}} p_{\tilde{X}_s}(x)dx \\
&\sim n\bar{\rho}_s \int \boldsymbol{x} \cdot 1_{\left\{\left\|\Phi[j]^T \boldsymbol{r}_s\right\|_2^2 > t\sigma_s^2\right\}} p_{\tilde{X}_s}(x)dx \sim n \cdot \Gamma_s(\bar{G}_s).
\end{aligned}$$

■

Lemma 7: Based on (16) and Lemma 6, then

$$n^{-1} \|\boldsymbol{r}_{s+1}\|_2^2 = n^{-1} \|\boldsymbol{r}_s\|_2^2 - \Delta_{s,4}(\bar{G}_s) + o_p(1), n \to \infty \tag{18}$$

where

$$\Delta_{s,4}(\bar{G}_s) = \frac{\frac{\Gamma_s(\bar{G}_s)}{d}}{\bar{d}_{s+1} + \frac{\Gamma_s(\bar{G}_s)}{(d\bar{\sigma}_s^2)}}$$

Proof: With (16) and Lemma 6, we get the following equation:

$$\underset{n\to\infty}{p.\lim} n^{-1} \boldsymbol{c}_s^T \left(\Phi_{J_s}^T \Phi_{J_s}\right)^{-1} \boldsymbol{c}_s = \frac{\frac{\Gamma_s(\bar{G}_s)}{d}}{\bar{d}_{s+1} + \frac{\Gamma_s(\bar{G}_s)}{(d\bar{\sigma}_s^2)}}.$$

Then we can derive (17). Eventually, we achieve the following equation:

$$\bar{\boldsymbol{g}}_{s+1}(4) = \bar{\boldsymbol{g}}_s(4) - \Delta_{s,4}(\bar{G}_s). \tag{19}$$

With (13)−(15) and (18), it shows that

$$\left(\frac{k_s}{n}, \frac{n_s}{n}, \frac{N_s}{n}, \frac{\|\boldsymbol{r}_s\|_2^2}{n}\right) \to \left(\bar{\rho}_s, \bar{d}_s, \bar{\nu}_s, \bar{\sigma}_s^2\right). \tag{20}$$

For the last component of the state vector $\boldsymbol{g}_s$, we have the following equation:

$$\mu_s = \frac{\|\boldsymbol{r}_s\|_2^2}{k_s d} = \frac{n}{\boldsymbol{k}_s d} \frac{\|\boldsymbol{r}_s\|_2^2}{n} = \frac{\frac{\|\boldsymbol{r}_s\|_2^2}{n}}{\frac{k_s d}{n}} = \frac{\boldsymbol{g}_s(4)}{\boldsymbol{g}_s(1)d}$$

With (19), it implies that:

$$p \lim_{n\to\infty} \mu_s = \frac{\bar{\boldsymbol{g}}_s(4)}{\bar{\boldsymbol{g}}_s(1)d} = \bar{\mu}_s.$$

Consequently, we obtain the following equation:

$$\bar{\boldsymbol{g}}_{s+1}(5) = \bar{\boldsymbol{g}}_s(5) - \Delta_{s,5}(\bar{G}_s) \tag{21}$$

where $\Delta_{s,5}(\bar{G}_s) = \bar{\mu}_s - \bar{\mu}_{s+1}$. Combining the above results completes the proof of Theorem 1. ■

B. Proof of Theorem 2

The proof of Theorem 1 is associated with a particular model for the statistical behavior of $\left\|\frac{\Phi^T[j]\boldsymbol{r}_s}{\sigma_s}\right\|_2^2$. To prove Theorem 2, a conditioned chi-square model is used to model the statistical behavior of $\left\|\frac{\Phi^T[j]\boldsymbol{r}_s}{\sigma_s}\right\|_2^2$ in the null case, and then the FAR is proven to be convergent when the problem size n tends to infinity.

Denote $\Psi_0 = (\boldsymbol{\psi}_{01}, \boldsymbol{\psi}_{02}, \ldots, \boldsymbol{\psi}_{0d})$, where $\boldsymbol{\psi}_{0i} \sim N(0, (1/n)I_n)$, and $\boldsymbol{\psi}_{0i}, i = 1, \ldots, d$ are statistically independent random vectors. Assume v_s is a determined vector which has the same definition as that in [1], and we define a random vector $a \in \mathbb{R}^d$ and a scalar random variable Z_s as follows:

$$\boldsymbol{a} = (a_1, a_2, \ldots, a_d)^T = \Psi_0^T \boldsymbol{v}_s,\ Z_s = \left\|\Psi_0^T \boldsymbol{v}_s\right\|_2^2 = \boldsymbol{a}^T \boldsymbol{a}.$$

Lemma 8: With the above definitions, it holds that:

$$a_i \sim N\left(0, \sigma_s^2\right),\ \frac{Z_s}{\sigma_s^2} \sim \chi_d^2$$

where $\sigma_s^2 = \dfrac{\|\boldsymbol{v}_s\|_2^2}{n}$.

The proof of the above lemma is elementary, thus omitted here. After normalization, $\Phi_0 = (\boldsymbol{\phi}_{01}, \boldsymbol{\phi}_{02}, \ldots, \boldsymbol{\phi}_{0d})$, where $\boldsymbol{\phi}_{0i} = \dfrac{\boldsymbol{\psi}_{0i}}{\|\boldsymbol{\psi}_{0i}\|_2}, i = 1, 2, \ldots, d$.

Define matrix $\mathrm{D} = \mathrm{diag}\left(\|\boldsymbol{\psi}_{01}\|_2^{-1}, \ldots, \|\boldsymbol{\psi}_{0d}\|_2^{-1}\right)$, then $\Phi_0 = \Psi_0 \cdot D$. Moreover, we define $W_s = \left\|\Phi_0^T \boldsymbol{v}_s\right\|_2^2 = \left\|D\,\Psi_0^T \boldsymbol{v}_s\right\|_2^2 = \boldsymbol{a}^T D^2 \boldsymbol{a}$.

Lemma 9: The sequence of $(W_s, s = 1, 2, \ldots, S)$ can be asymptotically well approximated by the sequence of $(Z_s, s = 1, 2, \ldots, S)$. More precisely, for a sequence ξ_n depending on n only,

$$\frac{E|W_s - Z_s|}{\sigma_s^2} \le d\xi_n \to 0, \quad n \to \infty$$

where $\sigma_s^2 = \|\boldsymbol{v}_s\|_2^2 / n$, $\xi_n = \left[3E\left(\|\boldsymbol{\psi}_{0i}\|_2^{-2} - 1\right)^2\right]^{1/2}$.

The proof of the above lemma can be referred to [17]. With the above notations, the statistical behavior of $\left\|\Phi[j]^T \boldsymbol{r}_s\right\|_2^2$ can be described as follows:

$$L\left(\left\|\Phi[j]^T \boldsymbol{r}_s\right\|_2^2 \left| j \in T'^c \cap I'^c_{s-1}\right.\right) = L\left(\left\|\boldsymbol{\phi}_0{}^T \boldsymbol{r}_s\right\|_2^2\right) \tag{22}$$

where $L(\cdot)$ denotes the probability law. Therefore, when n tends to infinity,

$$P\left(\left\|\boldsymbol{\phi}_0{}^T \boldsymbol{r}_s\right\|_2^2 < w \left| j \in T'^c \cup I'^c_{s-1}\right.\right) \to P(\bar{Z}_s \le w).$$

which completes the proof of Theorem 2.

References

1 D. L. Donoho, Y. Tsaig, I. Drori, and J. L. Starck, "Sparse solution of underdetermined systems of linear equations by stagewise orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol. 58, no. 2, pp. 1094−1120, Feb. 2012.

2 J. Xiong and T. Zhou, "Gene regulatory network inference from multifactorial perturbation data using both regression and correlation analyses," *PLoS One*, vol. 7, no. 9, Article ID e43819, Sep. 2012.

3 T. Zhou and Y. L. Wang, "Causal relationship inference for a large-scale cellular network," *Bioinformatics*, vol. 26, no. 16, pp. 2020−2028, Aug. 2010.

4 Y. L. Wang and T. Zhou, "A relative variation-based method to unraveling gene regulatory networks," *PLoS One*, vol. 7, no. 2, Article ID e31194, Feb. 2012.

5 W. H. Zhang, B. X. Huang, and T. Zhou, "An improvement on StOMP for sparse solution of linear underdetermined problems," in *Proc. the 32nd Chinese Control Conf.*, Xian, China, 2013, 1951−1956.

6 B. M. Sanandaji, T. L. Vincent, and M. B. Wakin, " Exact topology identification of large-scale interconnected dynamical systems from compressive observations," in *Proc. the 2011 American Control Conf.*, 2011, pp. 649−656.

7 D. L. Donoho, "For most large underdetermined systems of linear equations the minimal -norm solution is also the sparsest solution," *Commun. Pure Appl. Math.*, vol. 59, no. 6, pp. 797−829, Mar. 2006.

8 M. A. Davenport and M. B. Wakin, "Analysis of orthogonal matching pursuit using the restricted isometry property," *IEEE Trans. Inf. Theory*, vol. 56, no. 9, pp. 4395−4401, Sep. 2010.

9 S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM J. Sci. Comput.*, vol. 20, no. 1, pp. 33−61, Aug. 1998.

10 S. G. Mallat and Z. F. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Trans. Signal Processing*, vol. 41, no. 12, pp. 3397−3415, Dec. 1993.

11 J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4655−4666, Dec. 2007.

12 Y. C. 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, Jun. 2010.

13 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.

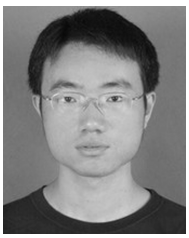
14 L. Yu, H. Sun, J. P. Barbot, and G. Zheng, "Bayesian compressive sensing for cluster structured sparse signals," *Signal Processing*, vol. 92, no. 1, pp. 259−269, Jan. 2012.

15 M. Stojnic, F. Parvaresh, and B. Hassibi, "On the reconstruction of block-sparse signals with an optimal number of measurements," *IEEE Trans. Signal Processing*, vol. 57, no. 8, pp. 3075−3085, Apr. 2009.

16 B. X. Huang and T. Zhou, "Recovery of block sparse signals by a block version of StOMP," *Signal Processing*, vol. 106, pp. 231−244, Jan. 2015.

17 J. Wang, S. Kwon and B. Shim, "Generalized orthogonal matching pursuit," *IEEE Trans. Signal Processing*, vol. 60, no. 12, pp. 6202 −6216, Dec. 2012.

18 W. H. Zhang, T. Zhou, and B. X. Huang, "Outlier deletion based improvement on the StOMP algorithm for sparse solution of large-scale underdetermined problems," *Sci. China Inf. Sci.*, vol. 57, no. 9, pp. 1−14, Sep. 2014.



Boxue Huang is a Senior Engineer of Institute of Electronics, Chinese Academy of Sciences. He received the bachelor degree from Huazhong University of Science and Technology in 2009 and the Ph.D. degree from Tsinghua University in 2016. His research interests include system identification, signal processing and systems biology. Corresponding author of this paper. E-mail: bxhuang@mail.ie.ac.cn

Tong Zhou is a Professor of control theory and control engineering, Tsinghua University. He received the B.S. and M.S. degrees from the University of Electronic Science and Technology of China, Chengdu, China, in 1984 and 1989, respectively, another M.S. degree from Kanazawa University, Ishikawa Prefecture, Japan, in 1991, and the Ph.D. degree from Osaka University, Osaka, Japan, in 1994. His current research interests include robust control, system identification, signal processing and their applications to real-world problems in molecular cell biology, spatio-temporal systems. Dr. Zhou was a recipient of the First-Class Natural Science Prize in 2003 from the Ministry of Education, China, and a recipient of the National Outstanding Youth Foundation of China Award in 2006. He has served as an Associate Editor of both the IEEE Transactions on Automatic Control and Automatica.
E-mail: tzhou@mail.tsinghua.edu.cn