Binary Matrices for Compressed Sensing

Weizhi Lu, Tao Dai, and Shu-Tao Xia

***Abstract*—For an $m \times n$ binary matrix with d nonzero elements per column, it is interesting to identify the minimal column degree d that corresponds to the best recovery performance. Consider this problem is hard to be addressed with currently known performance parameters, we propose a new performance parameter, the average of *nonzero* correlations between normalized columns. The parameter is proved to perform better than the known *coherence* parameter, namely the maximum correlation between normalized columns, when used to estimate the performance of binary matrices with high compression ratios n/m and low column degrees d. By optimizing the proposed parameter, we derive an ideal column degree $d = \lceil \sqrt{m} \rceil$, around which the best recovery performance is expected to be obtained. This is verified by simulations. Given the ideal number d of nonzero elements in each column, we further determine their specific distribution by minimizing the coherence with a greedy method. The resulting binary matrices achieve comparable or even better recovery performance than *random* binary matrices.**

***Index Terms*—Compressed sensing, measurement matrix, binary matrix, column degree, column correlation, optimal construction.**

I. Introduction

COMPRESSED sensing is a novel sampling theory [1], [2], which has found applications in a variety of areas [3]. Given a measurement matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ with $m < n$. Compressed sensing states that a signal $\mathbf{x} \in \mathbb{R}^n$ with at most $k\,(< n)$ nonzero elements can be recovered from $\mathbf{y} = \mathbf{A}\mathbf{x}$ by solving the minimization problem

$$\min \|\mathbf{x}\|_0 \text{ subject to } \mathbf{y} = \mathbf{A}\mathbf{x}, \tag{1}$$

or

$$\min \|\mathbf{x}\|_1 \text{ subject to } \mathbf{y} = \mathbf{A}\mathbf{x}. \tag{2}$$

In practice, we usually hope to recover a signal $\mathbf{x}$ with large sparsity k and the performance is mainly determined by the measurement matrix $\mathbf{A}$. In the paper, we study the optimal construction of $\{0,1\}$-binary measurement matrix, whose sparse structure is beneficial to storage and computation. For ease of analysis, throughout the paper binary matrices are assumed to be of a uniform column degree d, namely with exactly d nonzero elements per column.

Manuscript received October 17, 2016; revised May 20, 2017 and September 3, 2017; accepted September 5, 2017. Date of publication September 28, 2017; date of current version November 13, 2017. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Ian Vaughan Clarkson. This work was supported in part by the National Natural Science Foundation of China (61771273, 61371078) and in part by the Basic Research Program of Shenzhen (JCYJ20170307153635551). *(Corresponding author: Tao Dai.)*

The authors are with the Division of Information Science and Technology, Graduate School at Shenzhen, Tsinghua University, Guangdong 518055, China (e-mail: wzlusd@sz.tsinghua.edu.cn; dait14@mails.tsinghua.edu.cn; xiast@sz.tsinghua.edu.cn).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSP.2017.2757915

For a binary matrix with given size $m \times n$, we hope not only to optimize the recovery performance by maximizing the guaranteed signal sparsity k, but also to sparsify the structure of binary matrix by minimizing the column degree d, so as to reduce the computation and storage requirements. Currently, little attention has been devoted to reducing the column degree d [4]–[7], while most research has been focused on improving the guaranteed signal sparsity k by minimizing the known *coherence* parameter μ_m, directly or indirectly [8]–[12].

Definition 1 (The coherence parameter μ_m [13], [14]): Given a matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$. The coherence of $\mathbf{A}$ is defined as

$$\mu_m = \max_{i \neq j} \frac{|\mathbf{A}_i^\top \mathbf{A}_j|}{\|\mathbf{A}_i\|_2 \|\mathbf{A}_j\|_2} \tag{3}$$

where $\mathbf{A}_i$ denotes the i-th column of $\mathbf{A}$.

The coherence μ_m means the maximum absolute correlation between normalized columns of $\mathbf{A}$. It has been proved that a k-sparse signal $\mathbf{x}$ can be recovered from $\mathbf{y} = \mathbf{A}\mathbf{x}$ by (1) or (2), if

$$k < \frac{1}{2}\left(\frac{1}{\mu_m} + 1\right). \tag{4}$$

It is easy to see that the smaller the coherence μ_m, the higher the upper bound on the guaranteed signal sparsity k. There exists a lower bound for the coherence, $\mu_m \geq \sqrt{\frac{n-m}{m(n-1)}} \approx \frac{1}{\sqrt{m}}$, known as the Welch bound [15]. The bound is proved to be approached by a few binary matrices with special sizes [11], [16], while the column degrees of these matrices are not guaranteed to be minimized. For a binary matrix with given size, it seems hard to analytically derive a column degree d such that the lower bound of μ_m can be achieved. In this sense, we propose a new performance parameter μ_α to estimate the underlying optimal column degree d, namely the minimal column degree that corresponds to the best recovery performance.

Definition 2 (The proposed parameter μ_α): For a matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$, define

$$\mu_\alpha = \frac{1}{z} \sum_{i \neq j} \frac{|\mathbf{A}_i^\top \mathbf{A}_j|}{\|\mathbf{A}_i\|_2 \|\mathbf{A}_j\|_2}$$

where $z = |\{(i,j) : \mathbf{A}_i^\top \mathbf{A}_j \neq 0, i \neq j \in [n]\}|$ and $\mathbf{A}_i$ denotes the i-th column of $\mathbf{A}$.

The parameter μ_α denotes the average of *nonzero* absolute correlations between normalized columns of $\mathbf{A}$. Compared to the coherence μ_m, the advantages of μ_α are twofold. First, μ_α is proved to be able to substitute the parameter μ_m in (4) and offer more accurate performance estimations for binary matrices with high compression ratios n/m and low column degrees d. Such kind of matrices is attractive for practical applications in terms of its sparse structure and competitive performance [4]. Second, μ_α can be used to estimate the optimal column degree d. More precisely, for an $m \times n$ binary matrix with column degree d, we can derive a lower bound of μ_α. The lower bound tends to first decrease and then increase with the increase of d, achieving its minimal value at $d = \lceil \sqrt{m} \rceil$. This implies that the binary matrix should obtain its best performance at $d = \lceil \sqrt{m} \rceil$, since a smaller μ_α will lead to a higher bound for the guaranteed signal sparsity k. This conjecture is roughly verified in our simulations.

Given the number d of nonzero elements in each column, we need to further determine their specific distribution by optimization. Consider the difficulty of directly minimizing (optimizing) the proposed parameter μ_α, we propose to minimize the coherence μ_m, which empirically can also lead to a small μ_α. To the best of our knowledge, no efficient method has yet been proposed to minimize the coherence of binary matrix with given size. As stated in [11], [16], the Welch bound can be approached with some algebraic and geometric methods, and however, these methods are complicated and can only generate a few matrices with special sizes. In the paper, we propose a greedy method to minimize the coherence of binary matrix with *arbitrarily* given size and column degree. This method is developed by associating the binary matrix with a bipartite graph. For clarity, two major contributions of the paper are summarized as follows:

- We propose a new performance parameter μ_α (in Definition 2), which is proved to perform better than the *coherence* parameter μ_m, when used to estimate the performance of binary matrices with high compression ratios n/m and low column degrees d. By analyzing the lower bound of μ_α, we derive an ideal column degree $d = \lceil \sqrt{m} \rceil$. Experiments show that the best recovery performance tends to be obtained around $d = \lceil \sqrt{m} \rceil$.
- Given the ideal number d of nonzero elements in each column, we need to further determine their specific distribution by optimization. Consider the difficulty of directly minimizing (optimizing) the parameter μ_α, we propose a greedy method to minimize the coherence μ_m. Empirically, this method can also lead to a small μ_α.

Moreover, it may be worth noting that the idea of averaging *all* or *partial* column correlations has been proposed in recent literature to guide the matrix construction [17], [18] or analyze the signal recovery [19, Ch. 9]. However, these parameters are not guaranteed to offer reasonable performance estimations. Take the average of *all* column correlations as example, as analyzed in Section II-C, this parameter in fact cannot well reflect the performance difference of binary matrices with different column degrees.

The rest of the paper is organized as follows. In Section II, a new performance parameter μ_α is proposed to estimate the optimal column degree of binary matrix with given size. To optimize the structure of the binary matrix with given size and column degree, a greedy method is developed in Section III to minimize the coherence μ_m. Simulations are provided in Section IV. The paper is concluded in Section V.

II. The Optimal Column Degree of Binary Matrix

As shown in Theorem 3, this section aims to estimate the optimal column degree d by minimizing (optimizing) the parameter μ_α proposed in Definition 2. Prior to this, we need to first prove that the proposed parameter μ_α can offer a reasonable perform estimation, when applied in (4) to substitute the coherence parameter μ_m. For this purpose, we provide an ideal performance parameter μ_β in Definition 3, which is proved in Theorem 1 to be able to substitute the parameter μ_m in (4) and yield a higher bound for the guaranteed signal sparsity k. Then the parameter μ_β is employed as a baseline in Theorem 2 to compare the estimation accuracy of μ_α and μ_m, and μ_α is proved to be better when used to estimate the performance of binary matrices with high compression ratios n/m and low column degrees d.

A. An Ideal Performance Estimation Parameter μ_β

As we know, the null space property (NSP) [13], [14] provides a sufficient and necessary condition for compressed sensing by measuring the minimal number of linearly dependent columns (known as *spark*) in the measurement matrix. Here, we also borrow the concept of *spark* and provide an ideal parameter μ_β in Definition 3, which is proved in Theorem 1 to be able to replace the parameter μ_m in (4) and result in a higher bound for the guaranteed signal sparsity k. See the remark of Theorem 1 for more analysis.

Definition 3 (The parameter μ_β): For a matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$, define

$$\mu_\beta = \max_{i,\psi} \left\{ \frac{1}{\gamma - 1} \sum_{j\in\psi} \frac{|\mathbf{A}_i^\top \mathbf{A}_j|}{\|\mathbf{A}_i\|_2 \|\mathbf{A}_j\|_2} : i \in [n], \psi \subset [n] \setminus i, \right.$$
$$\left. |\psi| = \gamma - 1 \right\}$$

where $\gamma := \min\{\|\mathbf{x}\|_0 : \mathbf{Ax} = \mathbf{0}, \mathbf{x} \neq 0\}$ means the spark of $\mathbf{A}$ and $\mathbf{A}_i$ denotes the i-th column of $\mathbf{A}$.

Theorem 1: Suppose a matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$ with each column ℓ_2-normalized and without the same columns. A k-sparse vector $\mathbf{x}$ can be recovered from $\mathbf{y} = \mathbf{Ax}$ by (1), if $k < \frac{1}{2}(\mu_\beta^{-1} + 1)$ where μ_β is defined in Definition 3.

Proof: Suppose $\mathbf{A}_\phi$ is a submatrix of $\mathbf{A}$ with columns indexed by $\phi \subset \{1, 2, ..., n\}$. Let $\mathbf{G} = \mathbf{A}_\phi^\top \mathbf{A}_\phi$ which has the diagonal elements $g_{i,i} = 1$ and off-diagonal elements $|g_{i,j\neq i}| < 1$. Let γ be the spark of $\mathbf{A}$. With the definition of spark, we know the columns of $\mathbf{A}_\phi$ will be linearly dependent, if $|\phi| = \gamma$. Recall the columns of $\mathbf{A}_\phi$ are linearly independent if and only if all eigenvalues of $\mathbf{G}$ are positive [20]. When $|\phi| = \gamma$, $\mathbf{G}$ will have at least one eigenvalue equal to zero. Moreover, with the Gershgorin circle theorem [20], we know the i-th eigenvalue of $\mathbf{G}$ belongs to the interval $[g_{i,i} - r_i,\ g_{i,i} + r_i]$, where $r_i := \sum_{j=1; j\neq i}^{|\phi|} |g_{i,j}|$. Consequently, if $|\phi| = \gamma$, there must exist an index i such that $g_{i,i} - r_i \leq 0$, namely $1 - r_i \leq 0$, which leads to

$$1 - (\gamma - 1)\mu_\beta \leq 0 \tag{5}$$

where μ_β is defined as in Definition 3, equal to the maximal r_i, $i \in [|\phi|]$. Recall that $k < \gamma/2$ is a sufficient and necessary condition for the perfect signal recovery with (1) [13], [14]. Combining this condition with (5) proves the theorem. ■

Remark of Theorem 1: The definition of μ_β can be easily understood as follows. Compute the average of the $\gamma - 1$ largest (in magnitude) off-diagonal elements in each row of $\mathbf{A}^\top\mathbf{A}$, μ_β means the largest value among them. In contrast, the coherence μ_m denotes the maximum absolute off-diagonal element in $\mathbf{A}^\top\mathbf{A}$. This implies that $\mu_\beta \leq \mu_m$. In practice, we usually have $\mu_\beta \ll \mu_m$. This relation implies that (4) will offer a higher upper bound for the guaranteed signal sparsity k, when μ_m is substituted with μ_β. In other words, the parameter μ_β can provide a more accurate performance estimation compared to the coherence μ_m. Note that the value of μ_β is usually unavailable in practice because it involves the computation of the spark γ, which is NP-hard in general. As shown later, μ_β will only be employed as a baseline to compare the estimation accuracy of μ_m and μ_α.

B. The Proposed Parameter μ_α vs. the Coherence μ_m

For notational convenience, let us use $k(\mu_m)$ to denote the upper bound derived in (4); and similarly, use $k(\mu_\alpha)$ and $k(\mu_\beta)$ to denote the bounds derived when μ_m is replaced with μ_α and μ_β. As discussed in the remark of Theorem 1, we have $k(\mu_\beta) \geq k(\mu_m)$. This relation implies that compared to $k(\mu_m)$, $k(\mu_\beta)$ is more close to the *true* upper bound of the guaranteed signal sparsity k. The superiority of $k(\mu_\beta)$ over $k(\mu_m)$ enables $k(\mu_\beta)$ to be employed as a baseline to compare the accuracy of $k(\mu_m)$ and $k(\mu_\alpha)$; between them, the one more close to $k(\mu_\beta)$ is regarded as better.

In Theorem 2, we provide the condition that ensures $k(\mu_\alpha)$ better than $k(\mu_m)$. As detailed in the remark of Theorem 2, this condition can be satisfied by binary matrices with high compression ratios n/m and low column degrees d. This implies that compared to $k(\mu_m)$, $k(\mu_\alpha)$ can be used to better estimate the performance of binary matrices with high compression ratios n/m and low column degrees d.

Theorem 2: Consider a matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$ with each column ℓ_2-normalized and without the same columns, $m < n$. Assume all rows of $\mathbf{A}^\top\mathbf{A}$ will become the same after the elements in each row are arranged in ascending order. With Definitions 2 and 3, we know the parameters μ_α and μ_β can be derived only with the elements in the i-th row of $\mathbf{A}^\top\mathbf{A}$, $\forall i \in [n]$; more precisely,

$$\mu_\alpha = \frac{1}{z}\sum_{j\in[n]\setminus i} |\mathbf{A}_i^\top \mathbf{A}_j|$$

where $z = |\{j : \mathbf{A}_i^\top\mathbf{A}_j \neq 0, j \in [n]\setminus i\}|$; and

$$\mu_\beta = \max\left\{\frac{1}{\gamma - 1}\sum_{j\in\psi} |\mathbf{A}_i^\top\mathbf{A}_j| : \psi \subset [n]\setminus i, |\psi| = \gamma - 1\right\}$$

where γ means the spark of $\mathbf{A}$. It can be proved that the following inequality

$$|k(\mu_\beta) - k(\mu_\alpha)| \leq |k(\mu_\beta) - k(\mu_m)| \quad (6)$$

always holds for the matrix $\mathbf{A}$ with $\gamma \geq z$, and also holds for the matrix $\mathbf{A}$ with $\gamma < z$, if

$$\gamma \geq \left(\frac{1}{2} + \frac{\mu_\alpha}{2\mu_m}\right)z + 1. \quad (7)$$

Proof: The proof is divided into two parts to separately study the cases of $\gamma \geq z$ and $\gamma < z$. Let us first consider the case of $\gamma \geq z$. With definitions, we know

$$\mu_m \geq \mu_\alpha \geq \mu_\beta > 0$$

which has the first equality attained as all z nonzero elements share the same magnitude, and the second equality obtained as $\gamma = z$. With the above result, we can simply derive

$$\frac{1}{\mu_\beta} - \frac{1}{\mu_\alpha} \leq \frac{1}{\mu_\beta} - \frac{1}{\mu_m}$$

which leads to (6).

Now let us move on to the case of $\gamma < z$. Consider the off-diagonal nonzero elements in a row of $\mathbf{A}^\top\mathbf{A}$. Recall their number is equal to z. With definitions, we know μ_α is the average of the z nonzero elements in magnitude, while μ_β is the average of the $\gamma - 1$ largest nonzero elements in magnitude. This implies

$$\mu_\beta \geq \mu_\alpha$$

which has equality attained as z nonzero elements have the same magnitude; and

$$\mu_m \geq \mu_\beta$$

for which the equality is obtained as the $\gamma - 1$ largest absolute elements have the same magnitude. To obtain (6), we only need

$$\frac{1}{\mu_\alpha} - \frac{1}{\mu_\beta} \leq \frac{1}{\mu_\beta} - \frac{1}{\mu_m}$$

namely,

$$\mu_\beta \leq \frac{2\mu_\alpha\mu_m}{\mu_\alpha + \mu_m}. \quad (8)$$

Next let us provide the condition of (8). Recall $\gamma < z$. With the definitions of μ_β and μ_α, we have $\mu_\beta(\gamma - 1) < \mu_\alpha z$, namely

$$\mu_\beta < \frac{\mu_\alpha z}{\gamma - 1}. \quad (9)$$

To derive (8), we can make the right-hand side of (8) not smaller than the right-hand side of (9), then derive

$$\gamma \geq \left(\frac{1}{2} + \frac{\mu_\alpha}{2\mu_m}\right)z + 1.$$

which is the condition that ensures (6) holding under $\gamma < z$. Then the proof is completed. ■

Remark of Theorem 2: This theorem demonstrates that compared to the coherence estimation $k(\mu_m)$, the proposed estimation $k(\mu_\alpha)$ will be more close to the *true* upper bound of the guaranteed signal sparsity k, if the measurement matrix $\mathbf{A}$ satisfies the following two structure properties:

1) All rows of $\mathbf{A}^\top\mathbf{A}$ will become the same after the elements in each row are arranged in ascending order.

2) The spark γ of $\mathbf{A}$ satisfies the constraint of (7), i.e., $\gamma \geq (\frac{1}{2} + \frac{\mu_\alpha}{2\mu_m})z + 1$. Recall that the parameter z denotes the number of the off-diagonal nonzero elements in each row of $\mathbf{A}^\top\mathbf{A}$, and the other two parameters hold the relation of $\mu_\alpha \leq \mu_m$ (usually, $\mu_\alpha \ll \mu_m$).

The two properties can be satisfied by binary matrices with high compression ratios and low column degrees, as analyzed below.

Let us first see the structure property required for $\mathbf{A}^\top\mathbf{A}$. Consider a binary matrix $\mathbf{A}$ with d nonzero elements randomly generated in each column. According to the law of large numbers, the elements in each row of $\mathbf{A}^\top\mathbf{A}$ should follow the same distribution with high probability, as the compression ratio n/m is sufficiently large. This implies that the first structure property can be well satisfied by random binary matrices with high compression ratios. Empirically, this property can also be approximately satisfied by nonrandom (deterministic) binary matrices with high compression ratios, such as the matrix we will construct later by minimizing the coherence μ_m. Note that in compressed sensing the binary matrix cannot have the same columns, such that the compression ratio has an upper bound $n/m \leq \binom{m}{d}/m$. Usually, this bound is very large, allowing for the existence of binary matrices with high compression ratios.

Then let us see the structure property required for the spark γ of $\mathbf{A}$. Here we need to borrow a result from the following proof of Theorem 3, that is, the parameter z tends to increase with nd/m. This implies that with the increase of nd/m, we will eventually derive $z > 2m$ (assuming $n > 2m$), such that $\gamma < (\frac{1}{2} + \frac{\mu_\alpha}{2\mu_m})z + 1$ (because $\gamma \leq m + 1$), violating (7). This fact implies that to hold (7), the column degree d cannot be too large, especially as the compression ratio n/m is large. Note that the bound derived in (7) contains slacks and the constraint on column degree seems not to be strict in practice.

Finally, it is noteworthy that the proposed estimation $k(\mu_\alpha)$ performs well even when the condition mention above cannot be well satisfied, e.g., when the compression ratio is not large as shown in our simulations. The robust performance may partially benefit from the advantage of μ_α on estimating *average*-case recovery. Compared to the coherence μ_m, the parameter μ_α can better reflect the overall orthogonality level of the submatrices of $\mathbf{A}$ [21], [22].

C. The Optimal Column Degree Estimated with Parameter μ_α

For a binary matrix with given size $m \times n$, we propose to estimate the optimal column degree d by minimizing (optimizing) the parameter μ_α. For this purpose, we derive the lower bound of μ_α in Theorem 3, which suggests that the minimal column degree that corresponds to the minimal μ_α should be $d = \lceil\sqrt{m}\,\rceil$, see the remark of Theorem 3 for more analysis and discussions.

Theorem 3: Suppose a binary matrix $\mathbf{A} \in \{0, 1/\sqrt{d}\}^{m \times n}$ with d nonzero elements per column and dn/m nonzero elements per row, where $1 < d < m < n$. The parameter μ_α provided in Definition 2 satisfies

$$\mu_\alpha \geq \begin{cases} 1/d, & d \leq \sqrt{m} \\ [2pt]d/m, & d > \sqrt{m} \end{cases} \tag{10}$$

when $n \to \infty$.

Proof: To derive μ_α, with its definition we need to derive the sum and the number of the nonzero elements in the off-diagonal of $\mathbf{A}^\top\mathbf{A}$. Let us first calculate the sum. Suppose $a_{r,i}$ is the element of $\mathbf{A}$ in the r-th row and the i-th column and $\mathbf{A}_i$ is the i-th column of $\mathbf{A}$. The sum of the off-diagonal nonzero elements within the i-th row of $\mathbf{A}^\top\mathbf{A}$ can be expressed as

$$\sum_{j\in[n]\backslash i} \mathbf{A}_i^\top \mathbf{A}_j = \sum_{j\in[n]\backslash i}\sum_{r\in[m]} a_{r,i}a_{r,j} = \sum_{r\in[m]} a_{r,i} \sum_{j\in[n]\backslash i} a_{r,j}$$
$$= \frac{d}{\sqrt{d}} \times \frac{(dn/m-1)}{\sqrt{d}} = dn/m - 1.$$

This implies that the sum of the off-diagonal nonzero elements is the same for each row of $\mathbf{A}^\top\mathbf{A}$. Then the sum of the nonzero elements in the off-diagonal of $\mathbf{A}^\top\mathbf{A}$ is derived as

$$\sum_{i\neq j\in[n]\backslash i} \mathbf{A}_i^\top \mathbf{A}_j = m(dn/m - 1). \tag{11}$$

Next let us see the number of nonzero elements in the off-diagonal of $\mathbf{A}^\top\mathbf{A}$, denoted by z. Let z_i be the number of the off-diagonal *nonzero* elements within the i-th row of $\mathbf{A}^\top\mathbf{A}$. Let us first see the solution of z_i. Recall $\mathbf{A}_i^\top\mathbf{A}_{j\neq i} \in \{0, 1/d, ..., s/d\}$, where $s \in [d]$ is a constant. Then we can define

$$\lambda_q = \frac{1}{z_i}\left|\left\{j : \mathbf{A}_i^\top \mathbf{A}_{j\in[n]\backslash i} = q/d\right\}\right|$$

where $q \in [s]$, $\sum_{q\in[s]} \lambda_q = 1$. It follows that

$$\sum_{j\in[n]\backslash i} \mathbf{A}_i^\top \mathbf{A}_j = z_i \sum_{q\in[s]} \lambda_q \frac{q}{d}.$$

Combining this result with (11),

$$z_i = d\,(dn/m - 1)\,/\sum_{q\in[s]} \lambda_q q.$$

Since $\sum_{q\in[s]} \lambda_q q \geq \sum_{q\in[s]} \lambda_q = 1$,

$$z_i \leq d\,(dn/m - 1) \tag{12}$$

where the equality is attained for $q = 1$. With the definition of z_i, we have $z_i \leq n - 1$. Combining this result with (12),

$$z_i \leq \min\{d\,(dn/m-1)\,, n-1\}\,.$$

Moreover, with the relation between z_i and z, we can easily derive

$$z \leq m(\min\{d\,(dn/m-1)\,, n-1\}). \tag{13}$$

Since μ_α is the ratio of (11) to (13), we can write

$$\mu_\alpha = \frac{m(dn/m-1)}{z}$$
$$\geq \frac{m(dn/m-1)}{m(\min\{d\,(dn/m-1)\,, n-1\})}$$
$$= \frac{dn/m-1}{\min\{d\,(dn/m-1)\,, n-1\}}$$

which leads to (10) as $n \to \infty$. The theorem is proved. ∎

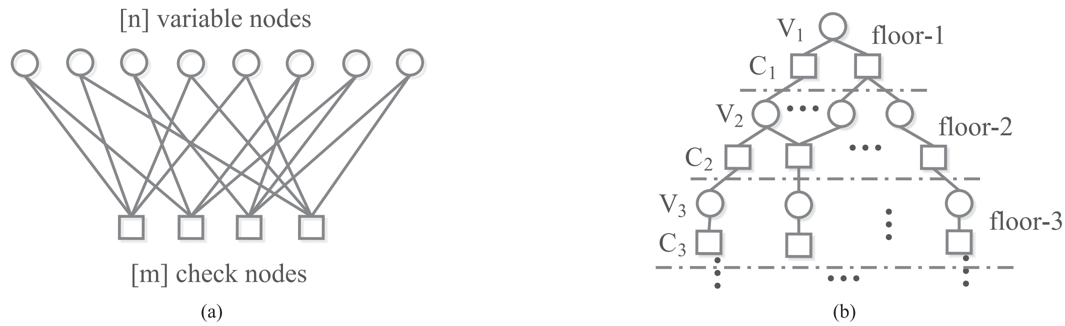

Fig. 1. A bipartite graph with m check nodes and n variable nodes in (a) and a tree expanded from a variable node in (b). The check nodes are denoted with squares and the variable nodes are denoted with circles.

Remark of Theorem 3: From (10), it is easy to see that with the increase of column degree d, the lower bound of μ_α will first decrease and then increase, achieving its minimal value at $d = \lceil\sqrt{m}\,\rceil$. Accordingly, the recovery performance should become first better and then worse, obtaining its best performance at $d = \lceil\sqrt{m}\,\rceil$. This performance trend is roughly verified in our simulations. After fixing the number d of nonzero elements in each column, we need to further determine their specific distribution. This problem will be addressed in the next section.

It is noteworthy that with (11), we can easily derive another commonly used performance parameter, the average of *all* column correlations [17], [19, Ch. 9], which is equal to $(dn/m - 1)/(n - 1) \approx d/m$. This value implies that a binary matrix with given size should become worse with the increase of column degree d; however, this conjecture does not coincide with the true performance provided in our simulations.

III. The Optimal Construction of Binary Matrix

This section aims to optimize the structure of binary matrix with given size and column degree. Consider the parameter μ_α is hard to be directly minimized, we propose to minimize the coherence μ_m, and empirically, this will also lead to a small μ_α. For analytical purposes, the coherence of binary matrix is formulated as $\mu_m = s/d$, where d is the column degree and s denotes the maximum correlation between distinct columns. This formulation implies two ways to minimize the coherence μ_m.

- One way is to minimize the value of s while fixing the value of d. The resulting matrix is what we aim to obtain, simply called the binary matrix with the minimal s for given d.
- The other way is to maximize the value of d while fixing the value of s. The resulting matrix is called the binary matrix with the maximal d for given s. In fact, such matrix can also be viewed as a binary matrix with the minimal s for given d, as detailed in Section III-B.

For easier realization, we adopt the second approach, that is to construct the binary matrix with the maximal d for given s. Prior to introducing the construction, we provide the basic knowledge of bipartite graph, which will be employed to describe and analyze the structure of binary matrix.

A. Preliminaries on Bipartite Graph

In this part, we introduce two concepts, bipartite graph and tree. By associating the binary matrix with bipartite graph, we show that the column correlation can be simply characterized with a tree.

1) Bipartite Graph: Fig. 1(a) provides an example of bipartite graph, which consists of m check nodes, n variable nodes, and a number of edges between two kinds of nodes. Such graph can be associated with a binary matrix $\mathbf{A} \in \{0, 1\}^{m\times n}$, whose $(i,\, j)$-th element is nonzero if and only if the i-th check node is connected with the j-th variable node. For notational convenience, let us use $C = \{c_1, ..., c_m\}$ to denote the set of m check nodes and use $V = \{v_1, ..., v_n\}$ to denote the set of n variable nodes. Suppose Γ is a subset of variable (check) nodes, the set of check (variable) nodes connecting to at least one element of Γ is denoted by $\mathcal{N}(\Gamma)$.

2) Tree: From each variable node, as shown in Fig. 1(b), we can expand a tree by forward traversing all nodes achievable through edges. For ease of analysis, the tree is divided into multiple floors, each floor containing one layer of variable nodes and one layer of check nodes. Here we use V_i and C_i to denote the sets of variable nodes and check nodes within the i-th floor of the tree, $i \geq 1$. The unique variable node in the set V_1, namely the node the tree starts with, is often called the *root* variable node. As can be seen in Fig. 1(b), the tree structure allows to be explicitly described with two relations, $V_i = \mathcal{N}(C_{i-1}) \setminus V_{i-1}$ and $C_i = \mathcal{N}(V_i) \setminus C_{i-1},\ i \geq 2$. Following them, the tree is expanded in the following matrix construction. Note that a tree usually does not contain all variable nodes or check nodes, i.e. $\sum_i |V_i| \leq |V| = n$ and $\sum_i |C_i| \leq |C| = m$.

3) Column Correlation: The column correlation of binary matrix can be easily described and analyzed with a tree. Suppose that the variable node v_i (check node c_i) corresponds to the i-th column (row). The correlation between the i-th and j-th columns is equal to the number of check nodes connecting to both v_i and v_j. Let us see a tree expanded from variable node v_i, such as the example shown in Fig. 1(b), where v_i is the unique node in V_1. It can be seen that the *root* variable node v_i correlates with every variable node $v_j \in V_2$ (in the 2nd floor). Connect the *root* variable node v_i with a check node $c_k \in C \setminus C_1$ (outside the 1st floor), the correlations between the root variable node v_i and other variable nodes are likely to be changed. The change follows the rule detailed in Property 1. This rule will be considered in the following matrix construction for constraining the column correlations.

Property 1: Consider a tree expanded from variable node v_i, see the example shown in Fig. 1(b), where v_i is the unique node in V_1. Connect the *root* variable node v_i with a check node c_k outside of the 1st floor, namely $c_k \in C \setminus C_1$, it can be observed:

- The correlation between the root variable node v_i and the variable node $v_j \in \mathcal{N}(c_k)$ will be increased by 1, if c_k lies in the 2nd floor, namely $c_k \in C_2$.
- New correlations equal to 1 will occur between the root variable node v_i and the variable node $v_j \in \mathcal{N}(c_k)$, if c_k is outside of the first 2 floors, namely $c_k \in C \setminus \{C_1 \cup C_2\}$.

B. The Structure of the Binary Matrix with the Maximal d for Given s

In Theorem 4, we provide an upper bound for the column degree d, which shows the tendency of growing with s. This trend implies that a binary matrix with the maximal d for given s is also a binary matrix with the minimal s for given d. More precisely, suppose that the upper bound of d is derived as $d = d^*$ under $s = s^*$, then the lower bound of s should be derived as $s = s^*$ under $d = d^*$.

Theorem 4: [23, Eqs. (3.3) and (3.7)] Suppose a binary matrix $\mathbf{A} \in \{0,1\}^{m \times n}$ with d nonzero elements per column. Let the coherence $\mu_m = s/d$, where $s \geq 1$ is an integer. Then,

$$d \leq \frac{m}{2n} + \sqrt{\frac{m(n-1)s}{n} + \frac{m^2}{4n^2}}. \tag{14}$$

By the way, with (14) we can derive a lower bound for the coherence μ_m, i.e., $\mu_m = \frac{s}{d} \gtrsim \frac{1}{\sqrt{m}}$. With the increase of n/m (fix d), the lower bound of μ_m tends to grow, resulting in a worse performance guarantee. As stated in [12], this is the major limitation of binary matrices.

In Property 2, we study the bipartite graph structure of an $m \times n$ binary matrix with the maximal d for given s. The tree expanded from each variable node is expected to contain m check nodes in its first 2 floors.

Property 2: Suppose a binary matrix $\mathbf{A} \in \{0,1\}^{m \times n}$ with d nonzero elements per column and with the maximum correlation between distinct columns equal to s, $1 \leq s < d < m < n$. Assume that the value of s will certainly be increased, if one element with value 1 is added into the matrix $\mathbf{A}$. Let us consider the bipartite graph associated with $\mathbf{A}$, which consists of m check nodes and n variable nodes. From each variable node, we can expand a tree as shown in Fig. 1(b). The tree will contain m check nodes exactly in its first 2 floors, namely $C_1 \cup C_2 = C$, $C_1 \neq \emptyset$ and $C_2 \neq \emptyset$.

Proof: Let us consider a tree expanded from variable node v_i, see the example shown in Fig. 1(b). For contradiction, assume that the tree contains less than m check nodes in its first 2 floors, namely $C_1 \cup C_2 \subsetneqq C$. In this case, let us connect the *root* variable node v_i with a check node c_k outside the first 2 floors, namely $c_k \in C \setminus \{C_1 \cup C_2\}$; according to Property 1, this connection will only introduce column correlations equal to 1, not greater than the maximum correlation s (≥ 1) we preset. This result contradicts with our previous assumption that the value of s will be increased if a new nonzero element (namely a new edge) is added into the matrix (namely the graph). The contradiction implies that the tree must contain m check nodes in its first 2 floors, namely $C_1 \cup C_2 = C$. Since each column of $\mathbf{A}$ has d nonzero elements, we have the number of check nodes in the 1st floor equal to d, namely $|C_1| = d$. Recall that $d < m = |C|$. To hold $C_1 \cup C_2 = C$, we have $C_2 \neq \emptyset$. ■

Algorithm 1: Construction of Binary Matrix with the Maximal d for Given s (Equivalently, with the Minimal s for Given d).

Initialization: Given a set C of m check nodes and a set V of n variable nodes. Let E be the set of the edges placed between two kinds of nodes and s be the maximal column correlation we preset. Repeat the procedure below until no edges can be added.

for $v_i \in V$ **do**
  Define a set $E^* = \emptyset$.
  From the variable node v_i, expand a tree of 2 floors using the edges in E; namely finding the sets C_1, V_2 and C_2.
  Define $V_{2a} \in V_2$ such that $|\mathcal{N}(v_j) \cap C_1| < s$, $v_j \in V_{2a}$; and define $V_{2b} = V_2 \setminus V_{2a}$. See Fig. 2.
  Define $C_{2a} \in C_2$ such that $\mathcal{N}(c_k) \cap V_{2b} = \emptyset$, $c_k \in C_{2a}$; and define $C_{2b} = C_2 \setminus C_{2a}$. See Fig. 2.
  if $C \setminus \{C_1 \cup C_2\} \neq \emptyset$ **then**
    Connect v_i with a check node $c_k \in C \setminus \{C_1 \cup C_2\}$ and then add the new edge (v_j, c_k) to E^*.
  end if
  if $C \setminus \{C_1 \cup C_2\} = \emptyset$ and $C_{2a} \neq \emptyset$ **then**
    Connect v_i with a check node $c_k \in C_{2a}$ and then add the new edge (v_j, c_k) to E^*.
  end if
  if $C \setminus \{C_1 \cup C_2\} = \emptyset$ and $C_{2a} = \emptyset$ **then**
    End the program and output E.
  end if
end for
$E = E \cup E^*$.

C. The Construction of the Binary Matrix with the Maximal d for Given s

For an $m \times n$ binary matrix with the maximum column correlation s previously fixed, we propose an iterative method to maximize the column degree d. At each iteration, only one nonzero element is added into each column, in a way that will not introduce column correlations greater than s. The method is detailed below.

With the equivalence between binary matrix and bipartite graph, we know that adding a nonzero element to the i-th column is equivalent to adding an edge to the i-th variable node v_i. To avoid generating column correlations greater than s, the edge can be added in the following way. First, let us expand a tree of 2 floors (as suggested by Property 2) from variable node v_i, and then connect the *root* variable node v_i with a check node c_k selected according to the following rule:

- A check node $c_k \in C \setminus \{C_1 \cup C_2\}$ (namely outside the tree) can be selected, if $C \setminus \{C_1 \cup C_2\} \neq \emptyset$, namely not all check nodes are included in the tree.

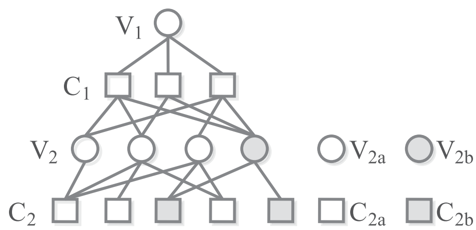

Fig. 2. An example for the sets V_{2a}, V_{2b}, C_{2a} and C_{2b} defined in Algorithm 1. Here the maximum column correlation $s = 3$.

- Otherwise, we can select a check node $c_k \in C_2$ (namely in the 2nd floor), whose every adjacent variable node $v_j \in \mathcal{N}(c_k)$ is required to connect less than s check nodes in the 1st floor, namely $|N(v_j) \cap C_1| < s$.

According to Property 1, the above rule will not introduce column correlations greater than s. Note that at the initial of the matrix construction, the tree we can expand only contains a root variable node and thus the sets C_1 and C_2 are empty. For clarity, the whole construction process is sketched in Algorithm 1.

Recall that a binary matrix with the maximal d for given s (constructed with Algorithm 1) can also be viewed as a binary matrix with the minimal s for given d. The latter is what we aim to obtain. Unfortunately, Algorithm 1 cannot output arbitrary column degree d, because as shown in (14), the upper bound of d cannot continuously grow with s. Given a column degree d that cannot be directly output by Algorithm 1, the minimal s can be identified in an indirect manner. To be specific, suppose that $d = d_{s^*}$ and $d = d_{s^*+1}$ are the maximal column degrees derived with Algorithm 1 under $s = s^*$ and $s = s^* + 1$, and moreover $d_{s^*+1} > d_{s^*} + 1$. With (14), we know the minimal $s = s^* + 1$ for any given $d \in (d_{s^*}, d_{s^*+1})$. Such structure can be constructed with Algorithm 1 by setting the parameter $s = s^* + 1$ and the iteration number to be d.

IV. Simulations

In this section, we aim to examine the behavior of the proposed parameter μ_α on performance estimation, and test the recovery performance of the binary matrices we construct by minimizing the coherence μ_m. Using Algorithm 1, we construct three groups of binary matrices with different sizes, whose structure parameters are detailed in Table I. For comparison, we also investigate the performance of *random* binary matrices, whose two structure parameters μ_α and μ_m are derived with the average of 10^4 random samples.

The simulation setting is as follows. The known orthogonal matching pursuit algorithm [24] is used for signal recovery. The recovery accuracy is measured with $1 - \|\mathbf{x} - \hat{\mathbf{x}}\|_2 / \|\mathbf{x}\|_2$. Each result is an average of 10^5 simulation runs. Sparse signals have nonzero elements randomly drawn from $N(0, 1)$. Random binary matrices are randomly generated in each simulation.

A. Performance Estimation with μ_α vs. with μ_m

The recovery performance of binary matrix is contrasted with the coherence parameter μ_m in Figs. 3–5 and contrasted with the proposed parameter μ_α in Figs. 6–8. Recall that a smaller μ_m (and μ_α) is expected to correspond to a better recovery performance. Our simulations show that the performance change of binary matrix over varying column degree d can be reasonably estimated by μ_α, rather than by μ_m, especially as the column degree d is small (e.g., $d \le \sqrt{m}$). This verifies the advantage of μ_α over μ_m.

According to Theorem 3, the minimal value of μ_α should be obtained at $d = \lceil \sqrt{m} \rceil$. This is confirmed in Figs. 6–8. As expected with μ_α, the best recovery performance is derived around $d = \lceil \sqrt{m} \rceil = 8$, $m = 50$, despite the varying of compression ratio. Note that the best performance cannot be guaranteed to be exactly derived at $d = \lceil \sqrt{m} \rceil$, because the performance estimation based on μ_α is not perfect and its reliability is only guaranteed for binary matrices with high compression ratios n/m and low column degrees d. However, as shown in our simulations, the proposed estimation performs well even as the compression ratio is not large. As discussed in the remark of Theorem 2, the robust performance may be partially owing to the advantage of μ_α on estimating *average*-case recovery.

B. Our Matrices vs. Random Matrices

Figs. 3–5 show that our binary matrices can obtain much smaller μ_m than random binary matrices with the same column degrees. This verifies the effectiveness of Algorithm 1 on minimizing the coherence μ_m. As shown in Figs. 6–8, our matrices also obtain slightly smaller μ_α than random matrices at low column degrees, but the advantage quickly disappears as the column degree increases. Coinciding with the changing trend of μ_α, our matrices present slightly better recovery performance than random matrices at low column degrees, and comparable performance at high column degrees.

TABLE I
The Binary Matrices With the Minimal s for Given d. Recall s Denotes the Maximum Correlation Between Columns and d Denotes the Column Degree.

<table>
<tr><td colspan="16">Matrix Size 50×250</td></tr>
<tr><td>d</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td><td>7</td><td>8</td><td>9</td><td>10</td><td>11</td><td>12</td><td>13</td><td>14</td><td>15</td><td>16</td></tr>
<tr><td>s</td><td colspan="2">1</td><td colspan="2">2</td><td colspan="2">3</td><td colspan="2">4</td><td colspan="2">5</td><td>6</td><td colspan="2">7</td><td>8</td><td>9</td></tr>
</table>

<table>
<tr><td colspan="16">Matrix Size 50×500</td></tr>
<tr><td>d</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td><td>7</td><td>8</td><td>9</td><td>10</td><td>11</td><td>12</td><td>13</td><td>14</td><td>15</td><td>16</td></tr>
<tr><td>s</td><td>1</td><td colspan="3">2</td><td colspan="2">3</td><td colspan="2">4</td><td>5</td><td colspan="2">6</td><td>7</td><td>8</td><td colspan="2">9</td></tr>
</table>

<table>
<tr><td colspan="16">Matrix Size 50×1000</td></tr>
<tr><td>d</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td><td>7</td><td>8</td><td>9</td><td>10</td><td>11</td><td>12</td><td>13</td><td>14</td><td>15</td><td>16</td></tr>
<tr><td>s</td><td>1</td><td colspan="2">2</td><td colspan="2">3</td><td colspan="2">4</td><td colspan="2">5</td><td>6</td><td colspan="2">7</td><td>8</td><td>9</td><td>10</td></tr>
</table>

V. Concluding Remarks

This paper has studied the optimal construction of binary matrix with arbitrarily given size. The research consists of two phases. In the first phase, a novel performance parameter, the average of nonzero column correlations is proposed to estimate the optimal column degree, namely the minimal column degree that corresponds to the best recovery performance. Compared to the known *coherence* parameter, the proposed parameter is proved to offer more accurate performance estimations for binary matrices with high compression ratios and low column

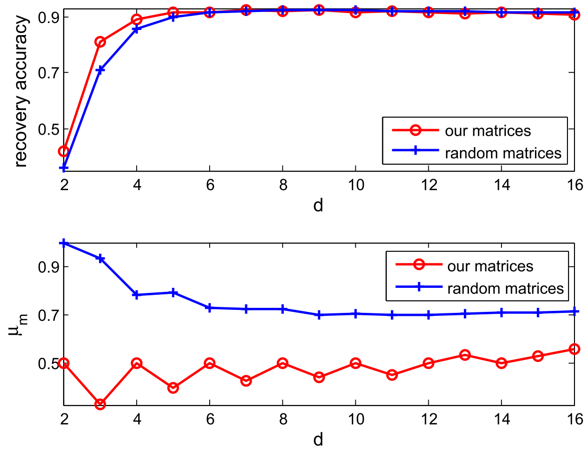

Fig. 3. Upside: Recovery accuracy of 50×250 binary matrix with column degree d on recovering signals of sparsity $k = 12$. Downside: The value of μ_m of each matrix.

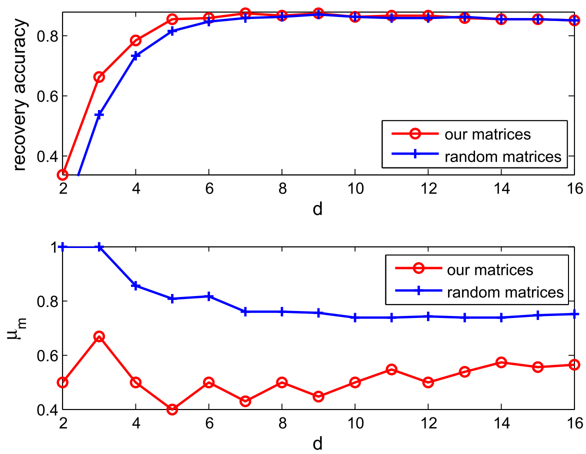

Fig. 4. Upside: Recovery accuracy of 50×500 binary matrix with column degree d on recovering signals of sparsity $k = 11$. Downside: The value of μ_m of each matrix.

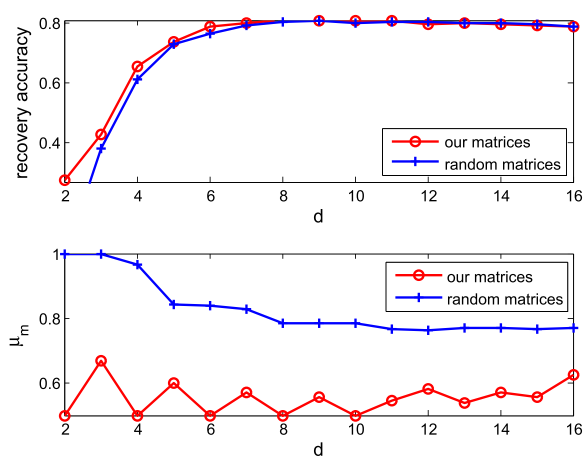

Fig. 5. Upside: Recovery accuracy of 50×1000 binary matrix with column degree d on recovering signals of sparsity $k = 10$. Downside: The value of μ_m of each matrix.

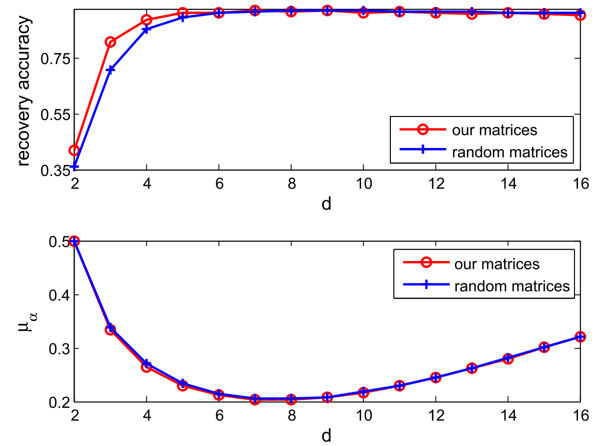

Fig. 6. Upside: Recovery accuracy of 50×250 binary matrix with column degree d on recovering signals of sparsity $k = 12$. Downside: The value of μ_α of each matrix.

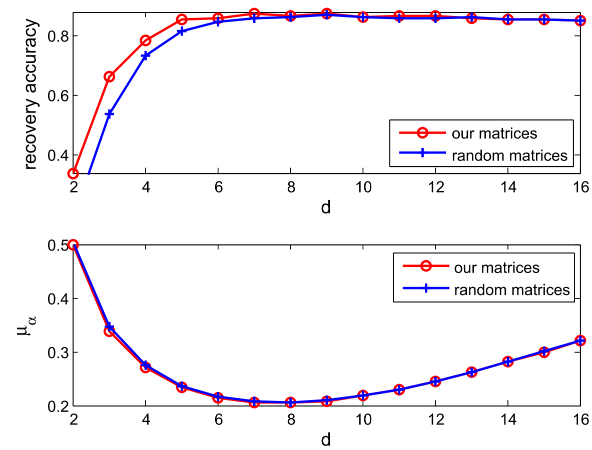

Fig. 7. Upside: Recovery accuracy of 50×500 binary matrix with column degree d on recovering signals of sparsity $k = 11$. Downside: The value of μ_α of each matrix.

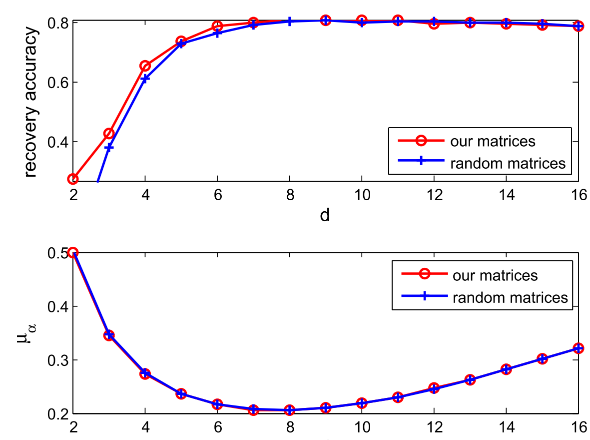

Fig. 8. Upside: Recovery accuracy of 50×1000 binary matrix with column degree d on recovering signals of sparsity $k = 10$. Downside: The value of μ_α of each matrix.

degrees. By minimizing (optimizing) the proposed parameter, we derive an ideal column degree $d = \lceil \sqrt{m} \rceil$, around which the best recovery performance is expected to be obtained. This is verified by simulations. In the second phase, we determine the distribution of the d nonzero elements in each column, by minimizing the *coherence* with a greedy method. Compared to random binary matrices, our matrices tend to present better performance at low column degrees, and comparable performance at high column degrees.

Moreover, it is worth noting that a $\{0, 1\}$-binary matrix can be simply transformed to a $\{0, \pm 1\}$-ternary matrix by introducing the plus-minus sign, randomly [25] or deterministically [26]. Compared to the binary matrix, the resulting ternary matrix usually has smaller column correlations and thus better recovery performance.

Finally, let us briefly discuss the practical application of the binary matrix we construct in the paper. In practice, a signal $\mathbf{x}$ is usually sparse over a sparsifying dictionary $\mathbf{D}$, rather than naturally sparse. In this case, compressed sensing can be formulated as $\mathbf{y} = \mathbf{P}\mathbf{x} = \mathbf{P}\mathbf{D}\mathbf{c}$, where $\mathbf{c}$ denotes a sparse coefficient vector and $\mathbf{P}$ denotes a projection matrix. The binary matrix we construct should be $\mathbf{A} = \mathbf{P}\mathbf{D}$. Consider the recovery performance is determined by $\mathbf{A}$, we should first fix $\mathbf{A}$ and then derive $\mathbf{P}$ with $\mathbf{A}$ and $\mathbf{D}$. It is apparent that $\mathbf{P}$ will be used for signal acquisition and $\mathbf{A}$ will be used for signal recovery. In the paper, we sparsify the structure of $\mathbf{A}$ to reduce the storage and computation cost of signal recovery. To reduce the complexity of signal acquisition, we can sparsify the structure of $\mathbf{P}$, which is worthy of future study.

ACKNOWLEDGMENT

The authors would like to thank Prof. F. Lazebnik for referring them to [23]. The authors would also like to thank the editor and reviewers for their constructive comments.

REFERENCES

[1] E. Candes and T. Tao, "Decoding by linear programming," *IEEE Trans. Inf. Theory*, vol. 51, no. 12, pp. 4203–4215, Dec. 2005.
[2] D. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.
[3] S. Qaisar, R. M. Bilal, W. Iqbal, M. Naureen, and S. Lee, "Compressive sensing: From theory to applications, a survey," *J. Commun. Netw.*, vol. 15, no. 5, pp. 443–456, Oct. 2013.
[4] R. Berinde and P. Indyk, "Sparse recovery using sparse random matrices," Tech. Rep. MIT-CSAIL-TR-2008-001, 2008.
[5] A. Dimakis, R. Smarandache, and P. Vontobel, "LDPC codes for compressed sensing," *IEEE Trans. Inf. Theory*, vol. 58, no. 5, pp. 3093–3114, May 2012.
[6] W. Lu, K. Kpalma, and J. Ronsin, "Sparse binary matrices of LDPC codes for compressed sensing," in *Proc. Data Compression Conf.*, Apr. 2012, pp. 405–405.
[7] X. J. Liu, S. T. Xia, and F. W. Fu, "Reconstruction guarantee analysis of basis pursuit for binary measurement matrices in compressed sensing," *IEEE Trans. Inf. Theory*, vol. 63, no. 5, pp. 2922–2932, May 2017.
[8] R. A. DeVore, "Deterministic constructions of compressed sensing matrices," *J. Complexity*, vol. 23, no. 4–6, pp. 918–925, 2007.
[9] S. Howard, A. Calderbank, and S. Searle, "A fast reconstruction algorithm for deterministic compressive sensing using second order reed-muller codes," in *Proc. Annu. Conf. Inf. Sci. Syst.*, Mar. 2008, pp. 11–15.
[10] A. Amini and F. Marvasti, "Deterministic construction of binary, bipolar, and ternary compressed sensing matrices," *IEEE Trans. Inf. Theory*, vol. 57, no. 4, pp. 2360–2370, Apr. 2011.
[11] S. Li and G. Ge, "Deterministic sensing matrices arising from near orthogonal systems," *IEEE Trans. Inf. Theory*, vol. 60, no. 4, pp. 2291–2302, Apr. 2014.
[12] R. R. Naidu, P. Jampana, and C. S. Sastry, "Deterministic compressed sensing matrices: Construction via euler squares and applications," *IEEE Trans. Signal Process.*, vol. 64, no. 14, pp. 3566–3575, Jul. 2016.
[13] R. Gribonval and M. Nielsen, "Sparse representations in unions of bases," *IEEE Trans. Inf. Theory*, vol. 49, no. 12, pp. 3320–3325, Dec. 2003.
[14] D. L. Donoho and M. Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via ℓ^1 minimization," *Proc. Natl. Acad. Sci.*, vol. 100, no. 5, pp. 2197–2202, 2003.
[15] L. Welch, "Lower bounds on the maximum cross correlation of signals (corresp.)," *IEEE Trans. Inf. Theory*, vol. 20, no. 3, pp. 397–399, May 1974.
[16] S. T. Xia, X. J. Liu, Y. Jiang, and H. T. Zheng, "Deterministic constructions of binary measurement matrices from finite geometry," *IEEE Trans. Signal Process.*, vol. 63, no. 4, pp. 1017–1029, Feb. 2015.
[17] M. Elad, "Optimized projections for compressed sensing," *IEEE Trans. Signal Process.*, vol. 55, no. 12, pp. 5695–5702, Dec. 2007.
[18] L. Zelnik-Manor, K. Rosenblum, and Y. C. Eldar, "Sensing matrix optimization for block-sparse decoding," *IEEE Trans. Signal Process.*, vol. 59, no. 9, pp. 4300–4312, Sep. 2011.
[19] P. Casazza and G. Kutyniok, *Finite Frames: Theory and Applications*. New york, NY, USA: Springer-Verlag, 2012.
[20] R. Horn and C. Johnson, *Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1985.
[21] R. Calderbank, S. Howard, and S. Jafarpour, "Construction of a large class of deterministic sensing matrices that satisfy a statistical isometry property," *IEEE J. Sel. Topics Signal Process.*, vol. 4, no. 2, pp. 358–374, Apr. 2010.
[22] A. Barg, A. Mazumdar, and R. Wang, "Restricted isometry property of random subdictionaries," *IEEE Trans. Inf. Theory*, vol. 61, no. 8, pp. 4440–4450, Aug. 2015.
[23] Z. Füredi and M. Simonovits, "The history of degenerate (bipartite) extremal graph problems," in *Erdös Centennial* (Bolyai Society Mathematical Studies). Berlin, Germany: Springer-Verlag, vol. 25, pp. 169–264, 2013.
[24] Y. Pati, R. Rezaiifar, and P. Krishnaprasad, "Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition," in *Proc. Asilomar Conf. Signals Syst. Comput.*, Nov. 1993, vol. 1, pp. 40–44.
[25] W. Lu, K. Kpalma, and J. Ronsin, "Semi-deterministic ternary matrix for compressed sensing," in *Proc. Eur. Signal Process. Conf.*, Sep. 2014, pp. 2230–2234.
[26] W. Lu and S. T. Xia, "Construction of ternary matrices with small coherence for compressed sensing," *Electron. Lett.*, vol. 52, no. 6, pp. 447–448, 2016.

Weizhi Lu received the B.E. and M.E. degrees from Shandong University, Jinan, China, in 2007 and 2010, respectively, and the Ph.D. degree from the National Institute of Applied Sciences, Rennes, France, in 2014, all in electronic engineering. He is currently in the Graduate School at Shenzhen, Tsinghua University, Guangdong, China, as a Postdoctoral Researcher. His research interests include compressed sensing and machine learning.

Tao Dai received the B.S. degree in electronic engineering from Xidian University, Xi'an, China, in 2014. He is currently working toward the Ph.D. degree in the Department of Computer Science and Technology, Tsinghua University, Guangdong, China. His research interests include the domains of sparse representation, machine learning, image processing, and computer vision (image restoration and image quality assessment).

Shu-Tao Xia received the B. S. degree in mathematics and the Ph.D. degree in applied mathematics from Nankai University, Tianjin, China, in 1992 and 1997, respectively.

Since January 2004, he has been in the Graduate School at Shenzhen, Tsinghua University, Guangdong, China. He is currently a full Professor there. From March 1997 to April 1999, he was in the research group of information theory, Department of Mathematics, Nankai University, Tianjin, China. From September 1997 to March 1998 and from August to September 1998, he visited the Department of Information Engineering, The Chinese University of Hong Kong, Hong Kong. His current research interests include coding and information theory, networking, and machine learning.