



HAL
open science

Compressed Sensing Performance of Random Bernoulli Matrices with High Compression Ratio

Weizhi Lu, Weiyu Li, Kidiyo Kpalma, Joseph Ronsin

► **To cite this version:**

Weizhi Lu, Weiyu Li, Kidiyo Kpalma, Joseph Ronsin. Compressed Sensing Performance of Random Bernoulli Matrices with High Compression Ratio. *IEEE Signal Processing Letters*, 2015, 22 (8), pp.1074 - 1078. 10.1109/LSP.2014.2385813 . hal-01116031

HAL Id: hal-01116031

<https://hal.science/hal-01116031v1>

Submitted on 12 Feb 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Compressed Sensing Performance of Random Bernoulli Matrices with High Compression Ratio

Weizhi Lu, Weiyu Li, Kidiyo Kpalma and Joseph Ronsin

Abstract—This paper studies the sensing performance of random Bernoulli matrices with column size n much larger than row size m . It is observed that as the compression ratio n/m increases, this kind of matrices tends to present a performance floor regarding the guaranteed signal sparsity. The performance floor is effectively estimated with the formula $\frac{1}{2}(\sqrt{\pi m/2} + 1)$. To the best of our knowledge, it is the first time in compressed sensing, a theoretical estimation is successfully proposed to reflect the real performance.

Index Terms—compressed sensing, compression ratio, high dimension, random matrix, Bernoulli distribution

I. INTRODUCTION

Compressed sensing states that a sparse signal can be well acquired through an underdetermined system of linear equations [1]. This technique successfully transfers the cost of acquisition to the later recovery process, which can be briefly described as follows. Let $\mathbf{x} \in \mathbb{R}^n$ be a k -sparse signal with at most k nonzero elements, and $\mathbf{A} \in \mathbb{R}^{m \times n}$ be a sensing matrix with $m < n$. Then x can be recovered from a finite number of linear observations $\mathbf{y} = \mathbf{A}\mathbf{x}$. The recovery process can be formulated as an ℓ_0 -based minimization problem

$$\min \|\hat{\mathbf{x}}\|_0 \quad \text{subject to} \quad \mathbf{y} = \mathbf{A}\hat{\mathbf{x}} \quad (1)$$

which holds a unique solution, if \mathbf{x} is sparse enough such that the sensing matrix \mathbf{A} satisfies some well-known conditions, e.g. null space property (NSP) [2] and restricted isometry property (RIP) [3]. Note that, for the convenience of analysis, the columns of \mathbf{A} are typically normalized in the following study.

For compressed sensing, it is natural to seek a sensing matrix with high compression ratio n/m . Empirically, the sensing performance will inevitably degrade as the compression ratio increases. A question of practical interests then arise: how fast will the performance degrade with the increasing of compression ratio? This paper is developed to address this question for the random Bernoulli matrix, which is popular in compressed sensing and performs as well as Gaussian ones. Surprisingly, as will be shown in the final simulation, the random Bernoulli matrix approximately presents a ‘performance floor’ with the fast increasing of compression ratio. To be specific, the decreasing speed of guaranteed sparsity k is very slow, and can even be ignored in the setting where m is fixed while n tends to infinity. This property enables

the significant compression of high-dimensional signal with sparsity lower than the performance floor. Then it becomes interesting to theoretically evaluate the performance floor, namely the guaranteed sparsity k . Unfortunately, the exact performance estimation is still an open problem in compressed sensing. Currently, the guaranteed sparsity k is often simply estimated with

$$k < \frac{1}{2}(\mu_m(\mathbf{A})^{-1} + 1), \quad (2)$$

where the parameter $\mu_m(\mathbf{A})$ represents the *maximum* absolute correlation between distinct columns of \mathbf{A} [4]. However, as will be shown later, the formula (2) is only a sufficient condition for perfect recovery, such that the estimated k is usually much smaller than the real value we can achieve in practice. In this paper, we will prove that the formula (2) can be modified to be a sufficient and necessary condition in the limit, if the maximum correlation $\mu_m(\mathbf{A})$ is replaced with the *average* absolute correlation between distinct columns of \mathbf{A} , denoted by $\mu_a(\mathbf{A})$. This improved estimation allows us to propose a simple formula $\frac{1}{2}(\sqrt{\pi m/2} + 1)$ to effectively approximate the performance floor of random Bernoulli matrices with fixed row size m . To the best of our knowledge, it is the first time that a theoretical estimation is reported being able to reflect the real sensing performance. Thus the contribution of this paper is of both practical and theoretical interests.

The rest of the paper is organized as follows. In the next section, by analyzing the proof process of formula (2), we demonstrate how the sufficient and necessary condition is approached by the average correlation. In section III, we first calculate the average correlation of random Bernoulli matrix, then estimate its performance floor. The numerical evidence is illustrated and discussed in section IV. Finally, this paper is concluded in section V.

II. ESTIMATION METHODS BASED ON AVERAGE CORRELATION VS. MAXIMUM CORRELATION

This section demonstrates how the formula (2) is modified to be a sufficient and necessary condition in the limit by analyzing its proof process shown in the following Lemma 2. To simplify the proof of Lemma 2, we first review an important conclusion in Lemma 1.

Lemma 1. [5] *A k -sparse signal \mathbf{x} can be uniquely recovered from $\mathbf{y} = \mathbf{A}\mathbf{x}$ with the ℓ_0 -minimization problem as shown in formula (1), if and only if the sparsity $k < \text{Spark}(\mathbf{A})/2$, where $\text{Spark}(\mathbf{A})$ denotes the smallest number of linearly dependant columns of \mathbf{A} .*

W. Lu, K. Kpalma and J. Ronsin are with INSA de Rennes, IETR, UMR CNRS 6164, F-35708, France (e-mail: weizhi.lu, kidiyo.kpalma and joseph.ronsin@insa-rennes.fr).

W. Li is with CREST-ENSAI, Bruz, F-35170, France (e-mail: weiyu.li@ensai.fr).

Lemma 2. [6] If

$$k < \frac{1}{2}(\mu_m(\mathbf{A})^{-1} + 1),$$

a k -sparse signal \mathbf{x} is the unique recovery of $\mathbf{y} = \mathbf{A}\mathbf{x}$ with the ℓ_0 -minimization problem as shown in formula (1).

Proof. We first need to prove a critical property, that is $\text{Spark}(\mathbf{A}) \geq 1 + \mu_m(\mathbf{A})^{-1}$. Let subset $\psi \subseteq \{1, 2, \dots, n\}$ with cardinality denoted as $|\psi|$, \mathbf{A}_ψ be a submatrix of \mathbf{A} with columns indexed by ψ , and \mathbf{A}_ψ^T be the transpose of \mathbf{A}_ψ . Consider the Gram matrix $\mathbf{G} = \mathbf{A}_\psi^T \mathbf{A}_\psi \in \mathbb{R}^{|\psi| \times |\psi|}$. Clearly its diagonal elements $g_{ii} = 1$, and off-diagonal elements $g_{ij} \leq \mu_m(\mathbf{A}) \leq 1$, $i \neq j$. Recall that the columns of \mathbf{A}_ψ are linearly independent if and only if Gram matrix \mathbf{G} has positive determinant, equivalently each eigenvalue is positive. With Gershgorin circle theorem [7], the i -th eigenvalue of \mathbf{G} is bounded in the interval $[g_{ii} - r_i, g_{ii} + r_i]$, where $r_i = \sum_{j=1, j \neq i}^{|\psi|} |g_{ij}|$. To render all eigenvalues positive, we only require $1 - (|\psi| - 1)\mu_m(\mathbf{A}) > 0$. In other words, for any $|\psi| < \mu_m(\mathbf{A})^{-1} + 1$, the columns of \mathbf{A}_ψ are linear independent. From the definition of Spark, it can be easily deduced that $\text{Spark}(\mathbf{A}) \geq \mu_m(\mathbf{A})^{-1} + 1$. By merging the result with Lemma 1, we immediately derive the condition $k < \frac{1}{2}(\mu_m(\mathbf{A})^{-1} + 1)$ for the unique solution based on ℓ_0 -minimization. \square

Now we focus our attention on the analysis of the proof of Lemma 2. First, we need to show how the maximum correlation $\mu_m(\mathbf{A})$ is involved. From the proof, it can be observed that the following inequality

$$1 - \sum_{j=1, j \neq i}^{|\psi|} |g_{ij}| > 0 \quad (3)$$

must hold to ensure the Gram matrix $\mathbf{A}_\psi^T \mathbf{A}_\psi$ being positive definite, where g_{ij} with $i \neq j$ denotes the correlation between the i -th and j -th columns of submatrix \mathbf{A}_ψ . Considering $|g_{ij}| \leq \mu_m(\mathbf{A})$, the condition in formula (3) is then simply relaxed to

$$1 - (|\psi| - 1)\mu_m(\mathbf{A}) > 0. \quad (4)$$

This relaxation process leads to the sufficient but unnecessary property of Lemma 2, which makes the estimation of Lemma 2 far away from the real performance.

To reflect the real performance, it is necessary to reduce the relaxation error between $\sum_{j=1, j \neq i}^{|\psi|} |g_{ij}|$ and $(|\psi| - 1)\mu_m(\mathbf{A})$, such that the sufficient and necessary condition can be approached for Lemma 2. To this end, we propose to replace the maximum correlation $\mu_m(\mathbf{A})$ with the average correlation $\mu_a(\mathbf{A})$. In this case, as proved in Lemma 3, the relaxation error will be close to zero with high probability as the submatrix size $|\psi|$ increases, if the average correlation of submatrix \mathbf{A}_ψ with high probability can be approximated by the average correlation of \mathbf{A} . According to the law of large numbers, the condition above should be well satisfied by the random Bernoulli matrix, if the matrix column size is sufficiently large. Thus in the following part, we are motivated

to verify the feasibility of the proposed estimation through evaluating the performance of random Bernoulli matrix with high compression ratio.

Lemma 3. Suppose $a_i \in [0, t]$ is arbitrarily distributed with mean of $e < t$ and variance of σ^2 . Then for k elements i.i.d drawn from the distribution, we have $\Pr(\sum_{i=1}^k a_i = ke) \rightarrow 1$ and $\Pr(\sum_{i=1}^k a_i = kt) \rightarrow 0$, if $k \rightarrow \infty$.

Proof. Suppose a binary distribution as below

$$a'_i = \begin{cases} 0 & \text{with probability } 1 - p \\ t & \text{with probability } p \end{cases}$$

where $p = e/t$. Then it follows that $\Pr(a_i = t) \leq \Pr(a'_i = t)$, and

$$\Pr(\sum_{i=1}^k a_i = kt) \leq \Pr(\sum_{i=1}^k a'_i = kt) = p^k = \left(\frac{e}{t}\right)^k$$

According to the law of large numbers, it is known that

$$\lim_{k \rightarrow \infty} \Pr\left(\left|\sum_{i=1}^k a_i - ke\right| < k\varepsilon\right) \geq 1 - \frac{\sigma^2}{k\varepsilon^2}$$

where ε is an arbitrarily small positive constant. Then the conclusion of the lemma can be easily derived. \square

III. AVERAGE COLUMN CORRELATION OF RANDOM BERNOULLI MATRIX

To evaluate the performance floor, this section calculates the average column correlation of random Bernoulli matrices in Lemma 4. According to the law of large numbers, the average column correlation of random Bernoulli matrix with $n \gg m$ should be equivalent to the expected value of the absolute correlation between two arbitrary Bernoulli vectors. Therefore in Lemma 4 we only calculate the expected value mentioned above.

Lemma 4. Suppose \mathbf{v} and \mathbf{w} are two distinct normalized column vectors of random Bernoulli matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ with i.i.d elements being $\pm \frac{1}{\sqrt{m}}$ equiprobably, and $f(\mathbf{v}, \mathbf{w}) = |\mathbf{v}^T \mathbf{w}|$ denotes the correlation between them, then the expected value of f is derived with the following two forms:

$$1) \quad \mathbb{E}(f) = \frac{2}{m} \frac{1}{2^m} \left\lceil \frac{m}{2} \right\rceil \binom{m}{\lceil \frac{m}{2} \rceil} \quad (5)$$

$$2) \quad \lim_{m \rightarrow \infty} \mathbb{E}(\sqrt{m}f) = \sqrt{\frac{2}{\pi}} \quad (6)$$

where $\lceil * \rceil$ denotes the minimum integer not smaller than $*$.

Proof. First, following from

$$f = |\mathbf{v}^T \mathbf{w}| = \left| \sum_{i=1}^m (\mathbf{v}_i \mathbf{w}_i) \right|,$$

f is equivalently written as

$$f = \frac{1}{m} \left| \sum_{i=1}^m \mathbf{z}_i \right|,$$

where \mathbf{z}_i being ± 1 equiprobably. Then with combination probability, the expected value of f can be formulated as

$$\mathbb{E}(f) = \frac{1}{m} \frac{1}{2^m} \sum_{i=0}^m \binom{m}{i} |m - 2i|$$

where $\binom{m}{i} := \frac{m!}{(m-i)!i!}$. With

$$\binom{m}{i} |m - 2i| = \begin{cases} m \binom{m-1}{0} & \text{if } i = 0 \\ m \binom{m-1}{m-i-1} - m \binom{m-1}{i-1} & \text{if } 1 \leq i \leq \frac{m}{2} \\ m \binom{m-1}{i-1} - m \binom{m-1}{m-i-1} & \text{if } \frac{m}{2} < i < m \\ m \binom{m-1}{m-1} & \text{if } i = m \end{cases}$$

one can further derive that

$$\sum_{i=0}^m \binom{m}{i} |m - 2i| = \begin{cases} 2m \binom{m-1}{\frac{m}{2}-1} & \text{if } m \text{ is even} \\ 2m \binom{m-1}{\frac{m-1}{2}} & \text{if } m \text{ is odd} \end{cases}$$

Finally, with $\binom{m-1}{i-1} = \frac{i}{m} \binom{m}{i}$, it follows that

$$\sum_{i=0}^m \binom{m}{i} |m - 2i| = 2 \lceil \frac{m}{2} \rceil \binom{m}{\lceil \frac{m}{2} \rceil}$$

The first conclusion of the theorem is thus obtained as

$$\mathbb{E}(f) = \frac{2}{m} \frac{1}{2^m} \lceil \frac{m}{2} \rceil \binom{m}{\lceil \frac{m}{2} \rceil}$$

We now turn to proving the second conclusion. According to Stirling's approximation:

$$m! = \sqrt{2\pi m} \left(\frac{m}{e}\right)^m \exp(\lambda_m), \quad 1/(12m+1) < \lambda_m < 1/(12m)$$

$\mathbb{E}(f)$ can be described as

$$\mathbb{E}(f) = \frac{1}{2^m} \frac{m!}{\frac{m}{2}! \frac{m}{2}!} = \sqrt{\frac{2}{\pi m}} \exp(\lambda_m - 2\lambda_{\frac{m}{2}})$$

if m is even; otherwise,

$$\begin{aligned} \mathbb{E}(f) &= \frac{m+1}{m} \frac{1}{2^m} \frac{m!}{\frac{m+1}{2}! \frac{m-1}{2}!} \\ &= \sqrt{\frac{2}{\pi m}} \left(\frac{m^2}{m^2-1}\right)^{\frac{m}{2}} \exp(\lambda_m - \lambda_{\frac{m+1}{2}} - \lambda_{\frac{m-1}{2}}) \end{aligned}$$

Then we have $\lim_{m \rightarrow \infty} \mathbb{E}(\sqrt{m}f) = \sqrt{\frac{2}{\pi}}$, whenever m is even or odd. The proof is completed. \square

Note that the formula (6) in fact converges very fast, and can be satisfied with a relatively small m (on the order of tens). This implies that the expected value $\mathbb{E}(f)$ in formula (5) allows to be approximately written as $\mathbb{E}(f) = \sqrt{2/(\pi m)}$. For confirmation, the coincidence between $\mathbb{E}(f)$ and $\sqrt{2/(\pi m)}$ is illustrated in Figure 3. Then the average correlation $\mu_a(\mathbf{A})$ of random Bernoulli matrices with $n \gg m$ can also be approximated as $\sqrt{2/(\pi m)}$. In this case, the guaranteed sparsity on the performance floor of random Bernoulli matrices can be estimated with

$$k_{avr} = \frac{1}{2}(\sqrt{\pi m/2} + 1) \quad (7)$$

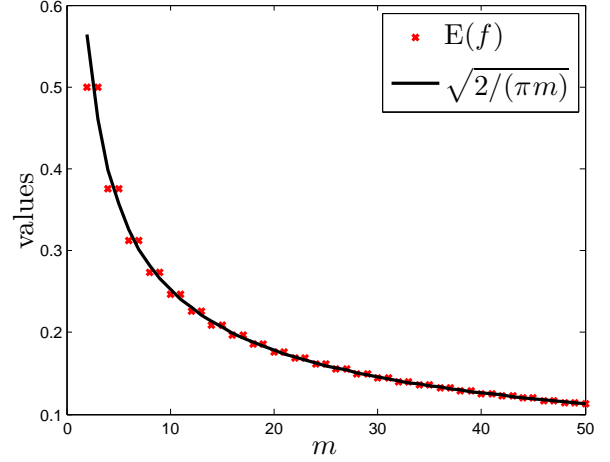


Figure 1: The values of $\mathbb{E}(f)$ in formula (5) and $\sqrt{2/(\pi m)}$ over varying m .

which is derived by replacing $\mu_m(\mathbf{A})$ with $\mu_a(\mathbf{A}) = \sqrt{2/(\pi m)}$ in formula (2). As it is expected in the former section, k_{avr} should be close to the real performance floor. Note that the maximum correlation of random Bernoulli matrices without same columns is equal to 1. For comparison, in the following simulation we also consider the performance floor estimated with $\mu_m(\mathbf{A}) = 1$, which is simply derived as

$$k_{max} = \frac{1}{2}(1 + 1) = 1. \quad (8)$$

by incorporating $\mu_m(\mathbf{A}) = 1$ into formula (2). As will be shown later, k_{avr} performs much better than k_{max} .

IV. NUMERICAL SIMULATIONS

In this section, we will show that the random Bernoulli matrix indeed approximately presents a performance floor, which can be effectively estimated with k_{avr} . To illustrate the performance floor, the guaranteed sparsity k of random Bernoulli matrices with fixed m and increasing n is derived in Table I. Before analyzing the data, we first briefly introduce the simulation setting. The sparse signal \mathbf{x} with sparsity k is randomly generated in each simulation, and recovered with subspace pursuit algorithm [8]. The recovery rate is measured with $1 - \|\hat{\mathbf{x}} - \mathbf{x}\|_2 / \|\mathbf{x}\|_2$. Note that here we only consider the largest k that can be recovered with rate larger than 0.99, because the perfect recovery is hard to be validated with simulation, unless we can enumerate all possible distributions of k nonzero elements. Each result in Table I is derived after 10000 simulation runs.

Note that in Table I the compression ratio n/m exponentially increases, while the decreasing speed of k is very slow and can even be ignored compared with the fast increasing n/m . This implies that there indeed approximately exists a performance floor regarding the fixed row size m , as graphically illustrated in Figure 2. Note that the gradual performance degradation is inevitable with the exponentially increasing of n . In practice, it seems hard to present a definite range for the performance floor. Here we empirically suppose the performance floor occurs, as the sparsity k decreases in a

Table I: The largest k guaranteed by (m, n) -sized random Bernoulli matrices under the condition of recovery rate larger than 0.99. For each m , all k with decreasing step smaller than 2 are underlined in bold.

n/m	2^1	2^2	2^3	2^4	2^5	2^6	2^7	2^8	2^9	2^{10}	2^{11}	2^{12}	2^{13}
m	6^2	5	<u>3</u>	<u>3</u>	<u>2</u>	<u>2</u>	<u>2</u>	<u>1</u>	<u>1</u>	<u>1</u>	<u>1</u>	<u>1</u>	<u>1</u>
	8^2	12	9	7	<u>5</u>	<u>5</u>	<u>4</u>	<u>3</u>	<u>3</u>	<u>3</u>	<u>2</u>	<u>2</u>	<u>2</u>
	10^2	22	16	12	10	<u>8</u>	<u>7</u>	<u>6</u>	<u>5</u>	<u>5</u>	<u>4</u>	<u>4</u>	<u>3</u>
	12^2	34	24	19	16	13	<u>11</u>	<u>10</u>	<u>9</u>	<u>8</u>	<u>7</u>	<u>6</u>	<u>6</u>
	14^2	50	35	27	22	19	16	14	<u>12</u>	<u>11</u>	<u>10</u>	<u>9</u>	<u>9</u>
	16^2	68	50	38	30	26	22	19	17	<u>15</u>	<u>14</u>	<u>13</u>	<u>12</u>
												<u>12</u>	<u>11</u>

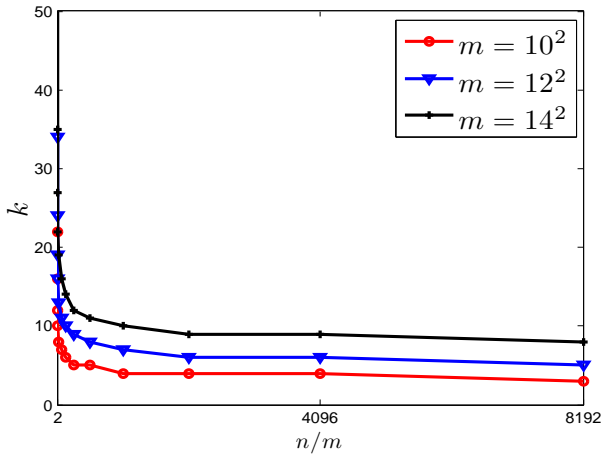


Figure 2: The performance curves of three random Bernoulli matrices from Table I, with $m=10^2$, 12^2 and 14^2 .

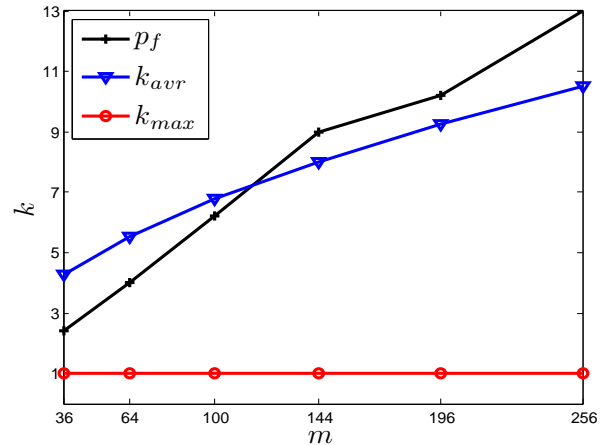


Figure 3: The performance floor measured with p_f , and the theoretical estimations k_{avr} in formula (7) and k_{max} in formula (8).

step not greater than 1. The relevant results are underlined in bold in Table I. Note that, due to the limitation of computer memory, as shown in Table I, we cannot test enough samples n to describe the performance floor, especially as m increases. For instance, we obtain only five samples for the largest case of $m = 16^2$. Thus to present a uniform metric, the performance floor of each m is quantified only with the mean of the first five k among the results underlined in Table I. The quantified result is denoted with p_f . Note that the whole performance floor is hard to be accurately reflected with the metric p_f , because it only considers the first five largest values on the performance floor. To be specific, as m increases, p_f tends to behave better than the real performance floor, since with five samples it can only consider few relatively large k and ignores most other smaller k on the performance floor. In contrast, for relatively small m , p_f probably performs worse than the real performance floor, as the five samples are likely to include some relatively small k . In this case, it is reasonable to infer that the distance between p_f and the estimation k_{avr} will increase with the increasing of m , if k_{avr} is close to the real performance. This conjecture is validated by the following simulation results.

In Figure 3, we compare the two estimations k_{avr} and k_{max} against the performance floor measured with p_f . As it is expected, the estimation k_{avr} based on average correlation performs much better than the estimation k_{max} based on maximum correlation. Precisely, the estimation k_{avr} is very close to p_f while the estimation k_{max} is of no practical use.

The infeasibility of k_{max} results from the fact that it is derived by simply assuming all column correlations take the value of the maximum correlation. Obviously, this assumption is far from the truth and suffers from significant errors. Despite obtaining much better estimate than k_{max} , as illustrated in Figure 3, k_{avr} is still not perfect, which becomes smaller than p_f as m increases. As discussed before, this is partially due to the inaccuracy of p_f on measuring the real performance floor. Precisely, p_f tends to behave worse than the real performance floor when m is relatively small, but perform better as m increases. From this fact, it can be derived that the error between k_{avr} and the real performance floor should be smaller than the result shown in Figure 3.

V. CONCLUSION

This paper has shown that the random Bernoulli matrix approximately presents a performance floor with the increasing of compression ratio, which enables the significant compression of high-dimensional sparse signals. More importantly, we successfully estimated the performance floor by exploring the average correlation between distinct columns of random Bernoulli matrix, instead of the traditional maximum column correlation. By this means, the theoretical estimation is built on a sufficient and necessary condition derived in the limit, and its value is thus much closer to the real sensing performance compared with the traditional estimation only based on a sufficient condition. Empirically, the result of this paper is also valid for Gaussian random matrices.

REFERENCES

- [1] D. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.
- [2] A. Cohen, W. Dahmen, and R. DeVore, "Compressed sensing and best k -term approximation," *Journal of The American Mathematical Society*, vol. 22, pp. 211–231, 2009.
- [3] E. Candes and T. Tao, "Decoding by linear programming," *IEEE Transactions on Information Theory*, vol. 51, no. 12, pp. 4203 – 4215, dec. 2005.
- [4] D. Donoho and X. Huo, "Uncertainty principles and ideal atomic decomposition," *IEEE Transactions on Information Theory*, vol. 47, no. 7, pp. 2845–2862, Nov 2011.
- [5] R. Gribonval and M. Nielsen, "Sparse representations in unions of bases," *IEEE Transactions on Information Theory*, vol. 49, no. 12, pp. 3320–3325, Dec 2003.
- [6] D. L. Donoho and M. Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via l_1 minimization," *Proceedings of the National Academy of Sciences*, vol. 100, no. 5, pp. 2197–2202, 2003.
- [7] R. Horn and C. Johnson, *Matrix Analysis*. Cambridge university press, 1985.
- [8] W. Dai and O. Milenkovic, "Subspace pursuit for compressive sensing signal reconstruction," *IEEE Transactions on Information Theory*, vol. 55, no. 5, pp. 2230–2249, 2009.