Combinatorics, Probability and Computing (2023), **32**, pp. 316–325
doi:10.1017/S0963548322000256

ARTICLE

Sparse recovery properties of discrete random matrices

Asaf Ferber[1], Ashwin Sah[2], Mehtaab Sawhney[2] and Yizhe Zhu[1]*

[1]Department of Mathematics, University of California, Irvine, CA, USA and [2]Department of Mathematics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
*Corresponding author. Email: yizhe.zhu@uci.edu

(Received 6 July 2022; revised 7 September 2022; accepted 8 September 2022; first published online 4 October 2022)

Abstract

Motivated by problems from compressed sensing, we determine the threshold behaviour of a random $n \times d \pm 1$ matrix $M_{n,d}$ with respect to the property 'every s columns are linearly independent'. In particular, we show that for every $0 < \delta < 1$ and $s = (1 - \delta)n$, if $d \le n^{1+1/2(1-\delta)-o(1)}$ then with high probability every s columns of $M_{n,d}$ are linearly independent, and if $d \ge n^{1+1/2(1-\delta)+o(1)}$ then with high probability there are some s linearly dependent columns.

Keywords: compressed sensing; discrete random matrix

2020 MSC Codes: Primary: 60B20

1. Introduction

Compressed sensing is a modern technique of data acquisition, which is at the intersection of mathematics, electrical engineering, computer science, and physics, and has grown tremendously in recent years. Mathematically, we define an unknown signal as a vector $\boldsymbol{x} \in \mathbb{R}^d$, and we have access to *linear measurements*: that is, for any vector $\boldsymbol{a} \in \mathbb{R}^d$, we have access to $\boldsymbol{a} \cdot \boldsymbol{x} = \sum_{i=1}^{d} a_i x_i$. In particular, if $\boldsymbol{a}^{(1)}, \ldots \boldsymbol{a}^{(n)} \in \mathbb{R}^d$ are the measurements we make, then we have an access to the vector $\boldsymbol{b} := A\boldsymbol{x}$, where

$$A := \begin{pmatrix} - & \boldsymbol{a}^{(1)} & - \\ & \vdots & \\ - & \boldsymbol{a}^{(n)} & - \end{pmatrix}.$$

The tasks of compressed sensing are: (*i*) to recover $\boldsymbol{x}$ from A and $\boldsymbol{b}$ as accurately as possible, and (*ii*) doing so in an efficient way. In practice, one would like to recover a high dimensional signal (that is, d is large) from as few measurements as possible (that is, n is small). In this regime, for an arbitrary vector $x \in \mathbb{R}^d$ the problem is ill-posed: for any given $\boldsymbol{b}$, the solution of $\boldsymbol{b} = A\boldsymbol{x}$, if it exists, forms a (translation of) linear subspace of dimension at least $d - n$, and therefore there is no way to uniquely recover the original $\boldsymbol{x}$.

Ferber was supported by NSF grants DMS-1954395 and DMS-1953799, NSF Career DMS-2146406, and Sloan's fellowship. Sah and Sawhney were supported by NSF Graduate Research Fellowship Program DGE-1745302. Sah was supported by the PD Soros Fellowship. Zhu was supported by NSF-Simons Research Collaborations on the Mathematical and Scientific Foundations of Deep Learning.

A key quantity to look at to guarantee the success of (unique) recovery is the *sparsity* of the vector $\boldsymbol{x}$, and we say that a vector is s-*sparse* if its *support* is of size at most s. That is, if

$$|\mathrm{supp}(\boldsymbol{x})| = \{i : x_i \neq 0\} \leq s.$$

A neat observation is that having at most one s-sparse solution to $A\boldsymbol{x} = \boldsymbol{b}$ for every $\boldsymbol{b}$ is equivalent to saying that A is $2s$-*robust* (that is, every $2s$ columns of A are linearly independent). Indeed, if we have two s-sparse vectors $\boldsymbol{x} \neq \boldsymbol{y}$ such that $A\boldsymbol{x} = A\boldsymbol{y}$ then $\boldsymbol{x} - \boldsymbol{y}$ is a nonzero $2s$-sparse vector in the kernel of A. For the other direction, if there is a nonzero $2s$-sparse vector in the kernel of A, one can split its support into two disjoint sets of size at most s each and consider the vectors restricted to these sets, one of which is multiplied by -1.

If we take A to be a random Gaussian matrix A (or any other matrix drawn from some 'nice' continuous distribution), then we clearly have that with probability one A is s-robust for $n = s$ and any $d \in \mathbb{N}$ (and in particular, one can uniquely recover $s/2$-sparse vectors). Moreover, in their seminal work, Candes and Tao [3] showed that it is possible to efficiently reconstruct $\boldsymbol{x}$ with very high accuracy by solving a simple linear programme if we take $n = O(s\log(d/s))$.

In this paper, we are interested in the compressed sensing problem with integer-valued measurement matrices and with entries of magnitude at most k. Integer-valued measurement matrices have found applications in measuring gene regulatory expressions, wireless communications, and natural images [1, 4, 12], and they are quick to generate and easy to store in practice [13, 14]. Under this setting, for integer-valued signal $\boldsymbol{x}$, we can have exact recovery even if we allow some noise $\boldsymbol{e}$ with $\|\boldsymbol{e}\|_\infty < 1/2$ (for more details, see [10]).

The first step is to understand when the compressed sensing problem is well-posed for given s, n, k, and d. Namely, for which values of s, n, k and d does an s-robust $n \times d$ integer-valued matrix with entries in $\{-k, \dots, k\}$ exist? For $s = n$, observe that if $d \geq (2k+1)^2 n$, then by the pigeonhole principle, one can find n columns for which their first two rows are proportional and therefore are not linearly independent. In particular, we have $d = O_k(n)$. In [10], Fukshansky, Needell, and Sudakov showed that there exists an s-robust A with $d = \Omega(\sqrt{k}n)$, using the result of Bourgain, Vu and Wood [2] on the singularity of discrete random matrices (in fact, the more recent result by Tikhomirov [17] gives a better bound for $k = 1$). Konyagin and Sudakov [15] improved the upper bound to $d = O(k\sqrt{\log kn})$, and they gave a deterministic construction of A when $d \geq \frac{1}{2}k^{n/(n-1)} > n$.

When $1 \leq s \leq n-1$ and $k = 2$, Fukshansky and Hsu [9] gave a deterministic construction such that $d \geq \left(\frac{n+2}{2}\right)^{1+\frac{2}{3s-2}}$. When $s = o(\log n)$, this implies we can take $d = \omega(n)$. This result hints that if we allow s to be 'separated away' from n, then one could take d to be 'very large'. A natural and nontrivial step to understanding the s-robustness property of matrices is to investigate the *typical* behaviour. For convenience, we will focus on the case $k = 1$ (even though our argument can be generalized to all fixed k), and we define, for all $n, d \in \mathbb{N}$, the random variable $M_{n,d}$ which corresponds to an $n \times d$ matrix with independent entries chosen uniformly from $\{\pm 1\}$. For $1 \leq s \leq n$, we would like to investigate the *threshold* behaviour of $M := M_{n,d}$ with respect to being s-robust. That is, we wish to find some $d^* := d(s, n)$ such that

$$\lim_{n\to\infty} \mathbb{P}[M \text{ is } s\text{-robust}] = \begin{cases} 0 & d/d^* \to \infty \\ 1 & d/d^* \to 0. \end{cases}$$

It is trivial to show (deterministically) that if $s = n$ and M is s-robust, then $d \leq 2n$. What if we allow s to be 'separated away' from n? That is, what if $s = (1-\delta)n$ for some $0 < \delta < 1$? It is not hard to show (and it follows from the proof of Lemma 3.3) that the probability for a random $n \times n$ matrix to have rank at least $(1-\delta)n$ is at least $1 - 2^{-\Omega(\delta^2 n^2)}$. Therefore, one could think that a typical $M_{n,d}$ might be $(1-\delta)n$-robust for some $d = 2^{n^{1-o(1)}}$. This turns out to be wrong as we show in the following simple theorem:

https://doi.org/10.1017/S0963548322000256 Published online by Cambridge University Press

Theorem 1.1. *For any fixed $0 < \delta < 1$ there exists $C > 0$ such that for sufficiently large $n \in \mathbb{N}$ the following holds. If $s = (1-\delta)n$ and $d \geq Cn^{1+1/(1-\delta)}$, then every ± 1 $n \times d$ matrix M is not s-robust.*

Proof. Given any $s/2$-subset of column vectors $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{s/2} \in \{\pm 1\}^n$ of M, by Spencer's 'six standard deviations suffice' [16], there exist some $x_1, \ldots, x_{s/2} \in \{\pm 1\}$ for which $\| \sum_{i=1}^{s/2} x_i \boldsymbol{v}_i \|_\infty \leq C' \sqrt{n}$ for a universal constant $C' > 0$ (a simple Chernoff bound suffices if one is willing to lose a $\sqrt{\log n}$ factor). Fix such a combination $\sum_{i=1}^{s/2} x_i \boldsymbol{v}_i$ for each $s/2$-subset of column vectors. Since there are at most $\left(3C'\sqrt{n}\right)^n$ integer-valued vectors in the box $[-C'\sqrt{n}, C'\sqrt{n}]^n$, and since

$$\binom{d}{s/2} \geq \left(\frac{d}{s}\right)^{s/2} = \left(\frac{Cn^{1/(1-\delta)}}{1-\delta}\right)^{(1-\delta)n/2} > \left(3C'\sqrt{n}\right)^n,$$

by the pigeonhole principle, as long as C is large enough, there are two $s/2$-subsets whose corresponding combination of column vectors are the same. Subtracting the corresponding combination of column vectors leads to a nonzero s-sparse kernel vector of M (since the indices of two $s/2$-subsets are not the same), proving the result. □

In our main result, we determine the (typical) asymptotic behaviour up to a window of $(\log n)^{\omega(1)}$.

Theorem 1.2. *For any fixed $0 < \delta < 1$, let $n \in \mathbb{N}$ be sufficiently large, let $s = (1-\delta)n$, and let $\varepsilon = \omega(\log \log n / \log n)$. We have that:*

1. *If $d \leq n^{1+1/(2-2\delta)-\varepsilon}$ then with high probability $M_{n,d}$ is s-robust.*
2. *If $d \geq n^{1+1/(2-2\delta)+\varepsilon}$ then with high probability $M_{n,d}$ is not s-robust.*

We believe that by optimizing our bounds/similar methods, one would be able to push the bounds in Theorem 1.2 up to a constant factor of $n^{1+1/(2-2\delta)}$ (though we did not focus on this aspect). It would be interesting to obtain the $1 + o(1)$ multiplicative threshold behaviour.

2. Proof outline

We first outline the proof of Theorem 1.2. We will prove part (1) of Theorem 1.2 over $\mathbb{F}_p$ for some prime $p = e^{\omega(\log^2 n)}$ to be chosen later (a stronger statement). Our strategy, at large, is to generate M as

$$M = \begin{pmatrix} M_1 \\ M_2 \end{pmatrix}$$

where $M_1 = M_{n_1,d}$ and $M_2 = M_{n_2,d}$, with $n_1 \approx n$ and $n_2 = o(n)$. The proof consists of the following two phases:

(1) **Phase 1:** Given any nonzero vector $\boldsymbol{a} \in \mathbb{F}_p^d$, we let

$$\rho_{\mathbb{F}_p}(\boldsymbol{a}) = \max_{x \in \mathbb{F}_p} \mathbb{P}\left[\sum_{i=1}^{d} a_i \xi_i = x\right], \tag{2.1}$$

where the ξ_is are i.i.d. Rademacher random variables. In this phase, we will show that

(a) M_1 is with high probability such that for all nonzero $\boldsymbol{a} \in \mathbb{F}_p^d$, if $|\text{supp}\, \boldsymbol{a}| \leq s := (1-\delta)n$ and $M_1 \boldsymbol{a} = \boldsymbol{0}$, then $\rho_{\mathbb{F}_p}(\boldsymbol{a}) = e^{-\omega(\log^2 n)}$, and

(b) M_1 is with high probability such that every s-subset of its columns has rank $s - o(s)$.

https://doi.org/10.1017/S0963548322000256 Published online by Cambridge University Press

(2) **Phase 2:** Conditioned on the above properties, we will use the extra randomness of M_2 to show that for a specific set of s columns, after exposing M_2, the probability that it does not have full rank is $o\left(1/\binom{d}{s}\right)$, and hence a simple union bound will give us the desired result.

In this strategy, it turns out that **Phase 1(a)** is the limiting factor, that is, ruling out structured kernel vectors.

For the proof of the upper bound in Theorem 1.2, we exploit this observation. We show using the second-moment method that it is highly likely that some $2\lfloor(1-\delta)n/2\rfloor$ columns sum to the zero vector (corresponding to an all 1s, highly structured kernel vector).

3. Proof of the lower bound in Theorem 1.2

In this section we prove Theorem 1.2. Let (say) $p \approx e^{\log^3 n}$ be a prime, let $d = n^{1+1/(2-2\delta)-\varepsilon}$ and $s = (1-\delta)n$ as given, and $n_1 = (1-\beta)n$ where $\beta = \omega(1/\log n)$ and $\beta = o(\log\log n/\log n)$. As described in Section 2, our proof consists of two phases, each of which will be handled separately.

3.1 Phase 1: no sparse structured vectors in the kernel of M_1

Our first goal is to prove the following proposition.

Proposition 3.1. *$M_{n_1,d}$ is with high probability such that for every $(1-\delta)n$-sparse vector $\boldsymbol{a} \in \mathbb{F}_p^d \setminus \{\mathbf{0}\}$, if $M_1\boldsymbol{a} = \mathbf{0}$ then $\rho_{\mathbb{F}_p}(\boldsymbol{a}) = e^{-\omega(\log^2 n)}$.*

In order to prove the above proposition, we need some auxiliary results.

Lemma 3.2. *$M_{n_1,d}$ is with high probability $n/\log^4 n$-robust over $\mathbb{F}_p$.*

Proof. Observe that for any $\boldsymbol{a} \in \mathbb{F}_p^d \setminus \{\mathbf{0}\}$ we trivially have that $\mathbb{P}[M_1\boldsymbol{a} = \mathbf{0}] \le 2^{-n_1} = 2^{-\Theta(n)}$. Since there are at most

$$\binom{d}{n/\log^4 n} p^{n/\log^4 n} \le \left(\frac{edp\log^4 n}{n}\right)^{n/\log^4 n} = 2^{o(n)}$$

$n/\log^4 n$-sparse vectors $\boldsymbol{a} \in \mathbb{F}_p^d$, by a simple union bound we obtain that the probability for such an $\boldsymbol{a}$ to satisfy $M_1\boldsymbol{a} = \mathbf{0}$ is $o(1)$. This completes the proof. □

In particular, by combining the above lemma with the Erdős-Littlewood-Offord inequality [5], we conclude that if $\boldsymbol{a} \in \mathbb{F}_p^d$ is $(1-\delta)n$-sparse and $M_1\boldsymbol{a} = \mathbf{0}$, then $\rho_{\mathbb{F}_p}(\boldsymbol{a}) = O(\log^2 n/n^{1/2})$. However, to prove Proposition 3.1, we need a stronger estimate.

The following lemma asserts that every subset of s columns in M_1 has large rank. It will be crucial in Phase 2.

Lemma 3.3. *Let $t = \omega(\log n)$. Then, with high probability $M_1 = M_{n_1,d}$ is such that every subset of s columns contains at least $s - t$ linearly independent columns.*

Proof. Consider the event that one such subset has rank at most $s-t$. There are $\binom{d}{s} \le d^s \le n^n$ possible choices of columns. For each such choice, there are at most $2^s \le 2^n$ ways to choose a spanning set of $r \le s-t$ columns. Such a subset has span containing at most 2^s many $\{\pm 1\}$ vectors (indeed, consider a full-rank $r \times r$ sub-block; any $\{\pm 1\}$ vector in the span of the columns is determined by its value on these r coordinates), so the probability that the remaining at least $t = \omega(\log n)$ columns are in the span is at most $(2^s/2^{n_1})^t \le (2^{-(\delta-\beta)n})^t = o(n^{-n})$. Taking a union bound, the result follows. □

Next, we state a version of Halász's inequality [11, Theorem 3] as well as a 'counting inverse Littlewood-Offord theorem' as was developed in [7].

Definition 3.4. Let $\boldsymbol{a} \in \mathbb{F}_p^n$ and $k \in \mathbb{N}$. We define $R_k^*(\boldsymbol{a})$ to be the number of solutions to

$$\pm a_{i_1} \pm a_2 \pm \ldots \pm a_{i_{2k}} \equiv 0 \mod p$$

with $|\{i_1, \ldots, i_{2k}\}| > 1.01k$.

Theorem 3.5. ([7, Theorem 1.4]). *Given an odd prime p, integer n, and vector $\boldsymbol{a} = (a_1, \ldots, a_n) \in \mathbb{F}_p^n \setminus \{\boldsymbol{0}\}$, suppose that an integer $0 \leq k \leq n/2$ and positive real L satisfy $30L \leq |supp(\boldsymbol{a})|$ and $80kL \leq n$. Then*

$$\rho_{\mathbb{F}_p}(\boldsymbol{a}) \leq \frac{1}{p} + C_{3.5} \frac{R_k^*(\boldsymbol{a}) + ((40k)^{0.99} n^{1.01})^k}{2^{2k} n^{2k} L^{1/2}} + e^{-L}.$$

We denote $\boldsymbol{b} \subset \boldsymbol{a}$ if $\boldsymbol{b}$ is a subvector of $\boldsymbol{a}$ and let $|\boldsymbol{b}|$ be the size of the support of a vector $\boldsymbol{b}$.

Theorem 3.6. ([7, Theorem 1.7]). *Let p be a prime, let $k, n \in \mathbb{N}$, $s \in [n]$ and $t \in [p]$. Define $\boldsymbol{B}_{k,m,\geq t}(s, d)$ as the following set:*

$$\left\{ \boldsymbol{a} \in \mathbb{F}_p^d : |\boldsymbol{a}| \leq s, \text{ and } R_k^*(\boldsymbol{b}) \geq t \cdot \frac{2^{2k} \cdot |\boldsymbol{b}|^{2k}}{p} \text{ for every } \boldsymbol{b} \subseteq \boldsymbol{a} \text{ with } |\boldsymbol{b}| \geq m \right\},$$

We have

$$|\boldsymbol{B}_{k,m,\geq t}(s, d)| \leq \binom{d}{s} \left(\frac{m}{s}\right)^{2k-1} (1.01t)^{m-s} p^s.$$

We now are in position to prove Proposition 3.1. The proof is quite similar to the proofs in [6–8].

Proof of Proposition 3.1. Let $k = \log^3 n$ and $m = n/\log^4 n$, $p \approx e^{\log^3 n}$.

First we use Lemma 3.2 to rule out vectors $\boldsymbol{a}$ with a support of size less than $n/\log^4 n$. Next, let (say) $L = n/\log^{10} n$ and let $\sqrt{L} \leq t \leq p$.

Consider a fixed $\boldsymbol{a} \in \boldsymbol{B}_{k,m,\geq t}(s, d) \setminus \boldsymbol{B}_{k,m,\geq 2t}(s, d)$ and we wish to bound the probability that $M_1 \boldsymbol{a} = \boldsymbol{0}$. By definition, there is a set $S \subseteq \operatorname{supp}(\boldsymbol{a})$ of size at least m such that

$$R_k^*(\boldsymbol{a}|_S) < 2t \cdot \frac{2^{2k}|S|^{2k}}{p}. \tag{3.1}$$

Since the rows are independent and since $\rho_{\mathbb{F}_p}(\boldsymbol{a}) \leq \rho_{\mathbb{F}_p}(\boldsymbol{a}|_S)$, the probability that $M_1 \boldsymbol{a} = \boldsymbol{0}$ is at most $\rho_{\mathbb{F}_p}(\boldsymbol{a}|_S)^{n_1}$. Furthermore, by Theorem 3.5 and the given conditions, which guarantee $30L \leq m \leq |\operatorname{supp}(\boldsymbol{a}|_S)|$ and $80kL \leq m \leq |S|$, and by $\sqrt{L} \leq t \leq p$, we have

$$\begin{aligned}
\rho_{\mathbb{F}_p}(\boldsymbol{a}|_S) &\leq \frac{1}{p} + C_{3.5} \frac{R_k^*(\boldsymbol{a}|_S) + ((40k)^{0.99}|S|^{1.01})^k}{2^{2k}|S|^{2k}L^{1/2}} + e^{-L} \\
&\leq \frac{1}{p} + \frac{2C_{3.5}t}{p\sqrt{L}} + \frac{10^k C_{3.5}}{L^{1/2}} \left(\frac{k}{|S|}\right)^{0.99k} + e^{-L} \\
&\leq \frac{Ct}{p\sqrt{L}}
\end{aligned} \tag{3.2}$$

for all sufficiently large n by equation (3.1). All in all, taking a union bound over all the possible choices of $\boldsymbol{a}$ (Theorem 3.6), and using the fact that $s = (1-\delta)n$ and $n_1 = (1-\beta)n$ with

$\beta = \omega(1/\log n)$, we obtain the bound

$$\begin{aligned}\binom{d}{s}\left(\frac{m}{s}\right)^{2k-1}(1.01t)^{m-s}p^s\left(\frac{Ct}{p\sqrt{L}}\right)^{n_1} &\leq \left(\frac{ed}{s}\right)^s(1.01t)^m\left(\frac{p}{1.01t}\right)^s\left(\frac{Ct}{p\sqrt{L}}\right)^{(1-\beta)n}\\ &\leq \left(\frac{ed}{(1-\delta)n}\right)^{(1-\delta)n}2^{o(n)}\left(\frac{1.01t}{p}\right)^{(\delta-\beta)n}\left(\frac{C(\log n)^5}{\sqrt{n}}\right)^{(1-\beta)n}\\ &= o(1/p)\end{aligned}$$

on the probability M_1 has such a kernel vector for sufficiently large n. Here we used the bounds $d \leq n^{1+1/(2-2\delta)-\varepsilon}$, $\varepsilon = \omega(\log\log n/\log n)$ and $\beta = o(\varepsilon)$. Union bounding over all possible values of t shows that there is an appropriately small chance of having such a vector for any $t \geq \sqrt{L}$.

Finally, note that $B_{k,m,\geq p}(s,d)$ is empty and thus the above shows that kernel vectors $\boldsymbol{a}$ cannot be in $\boldsymbol{B}_{k,m,\geq\sqrt{L}}(s,d)$. A similar argument as in equations (3.1) and (3.2) shows that

$$\rho_{\mathbb{F}_p}(\boldsymbol{a}) \leq \frac{C'}{p},$$

and the result follows. □

3.2 Phase 2: boosting the rank using M_2

Here we show that, conditioned on the conclusions of Proposition 3.1 and Lemma 3.3, after exposing M_2 with high probability $M = \begin{pmatrix} M_1 \\ M_2 \end{pmatrix}$ is s-robust.

To analyse the probability that a given subset of s columns is not of full rank, we will use the following procedure:

Fix any subset of s columns in M_1, and let $C := (\boldsymbol{c}_1, \ldots, \boldsymbol{c}_s)$ be the submatrix in M_1 that consists of those columns. We reveal M_2 according to the following steps:

(1) Let $I \subseteq [s]$ be the largest subset of indices such that the columns $\{\boldsymbol{c}_i \mid i \in I\}$ are linearly independent. By Lemma 3.3 we have that $T := |I| \geq s - t = (1-\delta)n - t$, where $t = \omega(\log n)$. Without loss of generality we may assume that $I := \{\boldsymbol{c}_1, \ldots, \boldsymbol{c}_T\}$ and $T \leq s-1$ (otherwise we have already found s independent columns of M). By maximality, we know that $\boldsymbol{c}_{T+1}$ can be written (uniquely) as a linear combination of $\boldsymbol{c}_1, \ldots, \boldsymbol{c}_T$. That is, there exists a unique combination for which $\sum_{i=1}^T x_i\boldsymbol{c}_i = \boldsymbol{c}_{T+1}$. In particular, this means that

$$\sum_{i=1}^T x_i\boldsymbol{c}_i - \boldsymbol{c}_{T+1} = 0,$$

and hence the vector $\boldsymbol{x} = (0, \ldots, x_1, \ldots, x_T, -1, \ldots, 0) \in \mathbb{F}_q^d$ is $(T+1)$-sparse and satisfies $M_1\boldsymbol{x} = 0$. Since $T+1 \leq s$, by Proposition 3.1 we know that $\rho_{\mathbb{F}_p}(\boldsymbol{x}) = 2^{-\omega(\log^2 n)}$.

(2) Expose the row vector of dimension $T+1$ from M_2 below the matrix $(\boldsymbol{c}_1, \ldots, \boldsymbol{c}_{T+1})$. We obtain a matrix of size $(n_1+1) \times (T+1)$. Denote the new row as $(y_1, \ldots, y_{T+1})$.

(3) If the new matrix is of rank $T+1$, then consider this step as a 'success', expose the entire row and start over from (1). Otherwise, consider this step as a 'failure' (As we failed to

https://doi.org/10.1017/S0963548322000256 Published online by Cambridge University Press

increase the rank) and observe that if $\begin{bmatrix} \boldsymbol{c}_1 & \dots & \boldsymbol{c}_{T+1} \\ y_1 & \dots & y_{T+1} \end{bmatrix}$ is not of full rank, then we must have

$$x_1y_1 + x_2y_2 + \ldots - y_{T+1} = 0.$$

The probability to expose such a vector y is at most $\rho_{\mathbb{F}_p}(\boldsymbol{x}) = e^{-\omega(\log^2 n)}$.

(4) All in all, the probability for more than $\beta n - t$ failures is at most $\binom{\beta n}{t}\left(e^{-\omega(\log^2 n)}\right)^{\beta n - t} = e^{-\omega(n\log n)} = o\left(\binom{d}{s}^{-1}\right)$. Therefore, by the union bound we obtain that with high probability M is s-robust.

This completes the proof.

4. Proof of the upper bound in Theorem 1.2

We first perform preliminary computations to compute a certain correlation. This boils down to estimating binomial sums. Let ξ_i, ξ_i' be independent Rademacher variables and define

$$\alpha(n, m) = \frac{\mathbb{P}[\xi_1 + \cdots + \xi_n = \xi_1 + \cdots + \xi_m + \xi_{m+1}' + \cdots + \xi_n' = 0]}{\mathbb{P}[\xi_1 + \cdots + \xi_n = 0]^2}.$$

Clearly $\alpha(n, m) \le \alpha(n, n) \le 10\sqrt{n}$ by [5].

Lemma 4.1. *Fix $\lambda > 0$. If n is even and $0 \le m \le (1-\varepsilon)n$ we have*

$$\alpha(n, m) = 1 + O(m/(\varepsilon n)).$$

Proof. We have

$$\alpha(n, m) \le \frac{\sup_k \mathbb{P}[\xi_1 + \cdots + \xi_{n-m} = k]}{\mathbb{P}[\xi_1 + \cdots + \xi_n = 0]^2} \le \frac{2^{-(n-m)}\binom{n-m}{\lfloor (n-m)/2 \rfloor}}{2^{-n}\binom{n}{n/2}} = 1 + O(m/(n-m)).$$

□

We will also need a more refined bound when m is small.

Lemma 4.2. *If n is even and $0 \le m \le n^{1/2}$, we have*

$$\alpha(n, m) = 1 + O(m^2/n^2).$$

Proof. Using the approximation $1 - x = \exp(-x - x^2/2 + O(x^3))$ for $|x| \le 1/2$ we see that if y is an integer satisfying $1 \le y \le x/2$ then

$$\begin{aligned} x(x-1)\cdots(x-y+1) &= x^y \exp\left(-\sum_{i=0}^{y-1}\frac{i}{x} - \sum_{i=0}^{y-1}\frac{i^2}{2x^2} + O\left(\frac{y^4}{x^3}\right)\right) \\ &= x^y \exp\left(-\frac{y(y-1)}{2x} - \frac{y(y-1)(2y-1)}{12x^2} + O\left(\frac{y^4}{x^3}\right)\right). \qquad (4.1) \end{aligned}$$

We now apply this to the situation at hand. We see $\alpha(n, m)$ is equal to

https://doi.org/10.1017/S0963548322000256 Published online by Cambridge University Press

$$\begin{aligned}
&\frac{2^{-(2n-m)}\sum_{k=0}^{m}\binom{m}{k}\binom{n-m}{n/2-k}^2}{2^{-2n}\binom{n}{n/2}^2}\\
&=2^m\sum_{k=0}^{m}\binom{m}{k}\left(\frac{(n/2)(n/2-1)\cdots(n/2-k+1)\times(n/2)(n/2-1)\cdots(n/2-(m-k)+1)}{n(n-1)\cdots(n-m+1)}\right)^2\\
&=2^m\sum_{k=0}^{m}\binom{m}{k}\left(\frac{(n/2)^m e^{-\frac{k(k-1)}{n}-\frac{k(k-1)(2k-1)}{3n^2}-\frac{(m-k)(m-k-1)}{n}-\frac{(m-k)(m-k-1)(2m-2k-1)}{3n^2}+O(m^4/n^3)}}{n^m e^{-\frac{m(m-1)}{2n}-\frac{m(m-1)(2m-1)}{12n^2}+O(m^4/n^3)}}\right)^2\\
&=2^{-m}\sum_{k=0}^{m}\binom{m}{k}\exp\left(-\frac{m^3-4mk(m-k)+n(2k-m)^2-nm}{2n^2}+O(m^2/n^2)\right)\\
&=2^{-m}\sum_{k=0}^{m}\binom{m}{k}\left(1-\frac{m^3-4mk(m-k)-nm}{2n^2}+O(m^2/n^2)\right)\left(1-\frac{(2k-m)^2}{2n}+O\left(\frac{(2k-m)^4}{n^2}\right)\right)\\
&=2^{-m}\sum_{k=0}^{m}\binom{m}{k}\left(1-\frac{m^3-4mk(m-k)-nm}{2n^2}\right)\left(1-\frac{(2k-m)^2}{2n}\right)+O(m^2/n^2).
\end{aligned}$$

In the third line, we used equation (4.1) and in the fourth line, we simplified the expression and used $k\le m\le n^{1/2}$ to subsume many terms into an error of size $O(m^2/n^2)$. The fifth line used $\exp(x)=1+x+O(x^2)$ for $|x|\le 1$ and the sixth line uses $2^{-m}\binom{m}{k}(2k-m)^4\le 2m^2\exp(-(2k-m)^2/100)$. Finally, this sum equals

$$\alpha(n,m)=1-\frac{3nm^2-3m^3+2m^2}{4n^3}+O(m^2/n^2)=1+O(m^2/n^2).$$

□

We are ready to prove the upper bound in Theorem 1.2.

Proof of the upper bound in Theorem 1.2. We are given $\delta\in(0,1)$ and $\varepsilon=\omega(\log\log n/\log n)$, with $d=n^{1+1/(2-2\delta)+\varepsilon}$. Let $s=2\lfloor(1-\delta)n/2\rfloor$. We consider an $n\times d$ random matrix with independent Rademacher entries and wish to show it is not s-robust with high probability. We may assume $\varepsilon<1/2$ as increasing d makes the desired statement strictly easier.

For an s-tuple of columns labelled by the index set $S\subseteq[d]$, let X_S be the indicator of the event that these columns sum to the zero vector. Let $X=\sum_{S\in\binom{[d]}{s}}X_S$, and let $(\xi_1,\dots,\xi_d)$ be a vector of independent Rademachers. We have

$$\mathbb{E}X=\binom{d}{s}\mathbb{E}X_{[s]}=\binom{d}{s}\mathbb{P}[\xi_1+\cdots+\xi_s=0]^n=\binom{d}{s}\left(2^{-s}\binom{s}{s/2}\right)^n$$

and

$$\begin{aligned}
\operatorname{Var}X=\mathbb{E}X^2-(\mathbb{E}X)^2&=\sum_{S,T\in\binom{[d]}{s}}\left(\mathbb{P}\Big[\sum_{i\in S}\xi_i=\sum_{j\in T}\xi_j=0\Big]^n-\mathbb{P}[\xi_1+\cdots+\xi_s=0]^{2n}\right)\\
&=(\mathbb{E}X)^2\cdot\frac{1}{\binom{d}{s}^2}\sum_{S,T\in\binom{[d]}{s}}\left(\frac{\mathbb{P}\Big[\sum_{i\in S}\xi_i=\sum_{j\in T}\xi_j=0\Big]^n}{\mathbb{P}[\xi_1+\cdots+\xi_s=0]^{2n}}-1\right)\\
&=(\mathbb{E}X)^2\sum_{m=0}^{s}\frac{\binom{s}{m}\binom{d-s}{s-m}}{\binom{d}{s}}\cdot(\alpha(s,m)^n-1).
\end{aligned}$$

For every $\eta > 0$ and $m \le c_\eta n^{1/2}$, where c_η is a sufficiently small absolute constant in terms of η, we see $|\alpha(s,m)^n - 1| \le \eta$ by Lemma 4.2. For $c_\eta n^{1/2} < m \le (1-\varepsilon/8)s$ we have $\alpha(s,m)^n \le \exp(O(m/\varepsilon))$ by Lemma 4.1. For this range we have, since $m/s \ge n^{\delta/2} s/d$,

$$\frac{\binom{s}{m}\binom{d-s}{s-m}}{\binom{d}{s}} \le (s+1)\mathbb{P}[\mathrm{Bin}(s, s/d) \ge m] \le \exp(-\, sD(m/(2s)||s/d)) \le \exp(-\, m(\delta/4)\log n)$$

by Chernoff–Hoeffding (the fact that $\mathrm{Bin}(n,p)$ exceeds nq for $q \ge p$ with probability at most $\exp(-\, nD(q||p))$, where this is the KL-divergence). Thus

$$\sum_{m=c\sqrt{n}}^{(1-\varepsilon)s} \frac{\binom{s}{m}\binom{d-s}{s-m}}{\binom{d}{s}} \cdot (\alpha(s,m)^n - 1) \le \sum_{m=c\sqrt{n}}^{(1-\varepsilon)s} \exp(O(m/\varepsilon)) \cdot \exp(-\, m(\delta/4)\log n) = o(1)$$

as $\varepsilon = \omega(\log\log n/\log n)$.

Finally for $(1-\varepsilon/8)s \le m \le s$ we have

$$\sum_{m=(1-\varepsilon)s}^{s} \frac{\binom{s}{m}\binom{d-s}{s-m}}{\binom{d}{s}} \cdot (\alpha(s,m)^n - 1) \le \sum_{m=(1-\varepsilon)s}^{s} \frac{\binom{s}{m}\binom{d-s}{s-m}}{\binom{d}{s}} (10\sqrt{n})^n \le 2^s \frac{\binom{d}{\varepsilon s/8}}{\binom{d}{s}} (10\sqrt{n})^n.$$

Thus

$$\sum_{m=(1-\varepsilon)s}^{s} \frac{\binom{s}{m}\binom{d-s}{s-m}}{\binom{d}{s}} \cdot (\alpha(s,m)^n - 1) \le \left(\frac{10s}{\varepsilon d}\right)^{(1-\varepsilon/8)s} (10\sqrt{n})^n \le (n^{-\frac{1}{2-2\delta}-\varepsilon/2})^{(1-\varepsilon/8)(1-\delta)n} (10\sqrt{n})^n,$$

since $d = n^{1+1/(2-2\delta)+\varepsilon}$ and $s = 2\lfloor (1-\delta)n/2 \rfloor$ along with $\varepsilon = \omega(\log\log n/\log n)$. We see that this is $o(1)$. Thus

$$\operatorname{Var} X \le (\mathbb{E}X)^2 \cdot \Big(\eta + o(1) + o(1)\Big) \le 2\eta(\mathbb{E}X)^2$$

for n sufficiently large, and thus $X > 0$ with probability at least $1 - 2\eta$. □

References

[1] Abdi, A., Fekri, F. and Zhang, H. (2019) Analysis of sparse-integer measurement matrices in compressive sensing. In *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, pp. 4923–4927.

[2] Bourgain, J., Vu, V. H. and Wood, P. M. (2010) On the singularity probability of discrete random matrices. *J. Funct. Anal.* **258**(2) 559–603.

[3] Candes, E. J. and Tao, T. (2006) Near-optimal signal recovery from random projections: universal encoding strategies? *IEEE Trans. Inform. Theory* **52**(12) 5406–5425.

[4] Eldar, Y. C., Haimovich, A. M. and Rossi, M. (2014) Spatial compressive sensing for MIMO radar. *IEEE Trans. Signal Process.* **62**(2) 419–430.

[5] Erdös, P. (1945) On a lemma of Littlewood and Offord. *Bull. Am. Math. Soc.* **51**(12) 898–902.

[6] Ferber, A. and Jain, V. (2019) Singularity of random symmetric matrices—a combinatorial approach to improved bounds. *Forum Math. Sigma* **7**(e22) 29.

[7] Ferber, A., Jain, V., Luh, K. and Samotij, W. (2021) On the counting problem in inverse Littlewood-Offord theory. *J. Lond. Math. Soc.* **103**(4) 1333–1362.

[8] Ferber, A., Luh, K. and McKinley, G. (2021) Resilience of the rank of random matrices. *Combin. Probab. Comput.* **30**(2) 163–174.

[9] Fukshansky, L. and Hsu, A. (2022) overing point-sets with parallel hyperplanes and sparse signal recovery. *Discrete Comput. Geom.* **14**(2) 79.

[10] Fukshansky, L., Needell, D. and Sudakov, B. (2019) An algebraic perspective on integer sparse recovery. *Appl. Math. Comput.* **340**(1–2) 31–42.

[11] Halász, G. (1977) Estimates for the concentration function of combinatorial number theory and probability. *Period. Math. Hungar.* **8**(3-4) 197–211.

https://doi.org/10.1017/S0963548322000256 Published online by Cambridge University Press

[12] Haseyama, M., He, Z. and Ogawa, T. (2010) The simplest measurement matrix for compressed sensing of natural images. In *2010 IEEE International Conference on Image Processing*, IEEE, pp. 4301–4304.

[13] Iwen, M. A. (2014) Compressed sensing with sparse binary matrices: instance optimal error guarantees in near-optimal time. *J. Complex.* **30**(1) 1–15.

[14] Jiang, Y., Liu, X.-J., Xia, S.-T. and Zheng, H.-T. (2015) Deterministic constructions of binary measurement matrices from finite geometry. *IEEE Trans. Signal Process.* **63**(4) 1017–1029.

[15] Konyagin, S. and Sudakov, B. (2020) An extremal problem for integer sparse recovery. *Linear Algebra Appl.* **586** 1–6.

[16] Spencer, J. (1985) Six standard deviations suffice. *Trans. Am. Math. Soc.* **289**(2) 679–706.

[17] Tikhomirov, K. (2020) Singularity of random Bernoulli matrices. *Ann. Math.* **191**(2) 593–634.

Cite this article: Ferber A, Sah A, Sawhney M, and Zhu Y (2023). Sparse recovery properties of discrete random matrices. *Combinatorics, Probability and Computing* **32**, 316–325. https://doi.org/10.1017/S0963548322000256

https://doi.org/10.1017/S0963548322000256 Published online by Cambridge University Press