

## ON THE PROBABILITY THAT A RANDOM $\pm 1$ -MATRIX IS SINGULAR

JEFF KAHN, JÁNOS KOMLÓS, AND ENDRE SZEMERÉDI

### 1. INTRODUCTION

**1.1. The problem.** For  $M_n$  a random  $n \times n$   $\pm 1$ -matrix (“random” meaning with respect to uniform distribution), set

$$P_n = \Pr(M_n \text{ is singular}).$$

The question considered in this paper is an old and rather notorious one: What is the asymptotic behavior of  $P_n$ ?

It seems often to have been conjectured that

$$(1) \quad P_n = (1 + o(1)) n^2 / 2^{n-1},$$

that is, that  $P_n$  is essentially the probability that  $M_n$  contains two rows or two columns which are equal up to a sign. This conjecture is perhaps best regarded as folklore. It is more or less stated in [14] and is mentioned explicitly, as a standing conjecture, in [20], but has surely been recognized as the probable truth for considerably longer. (It has also been conjectured ([17]) that  $P_n / (n^2 2^{-n}) \rightarrow \infty$ .)

Of course the guess in (1) may be sharpened, e.g., to

$$P_n - 2 \binom{n}{2} 2^{-(n-1)} \sim 2 \binom{n}{4} \left(\frac{3}{8}\right)^n,$$

the right-hand side being essentially the probability of having a minimal row or column dependency of length 4.

Despite the availability of the natural guess (1), upper bounds on  $P_n$  have not been easy to come by. That  $P_n \rightarrow 0$  was shown by Komlós in 1963 (but published somewhat later [12]). This is a discrete analogue of the fact that the variety of singular real matrices has Lebesgue measure 0, and should be quite

---

Received by the editors May 31, 1991 and, in revised form, January 28, 1994.

1991 *Mathematics Subject Classification*. Primary 15A52, 11L03; Secondary 15A36, 11C20.

*Key words and phrases*. Random matrices, exponential sums, Littlewood-Offord Problem.

Research for the first author was supported in part by NSF grant NSFDM8-9003376 and AFOSR grants AFOSR-89-0512, AFOSR-90-0008.

Research for the second author was supported in part by the Hungarian National Foundation for Scientific Research #1905.

obvious. It is somewhat surprising that no trivial proof is known. A simpler proof (based, like the original proof, on Sperner's theorem ([24] or, e.g., [7])), giving the bound  $P_n = O(1/\sqrt{n})$ , was offered in [14] (see also [1], XIV.2).

Here we give an exponential bound.

**Theorem 1.** *There is a positive constant  $\varepsilon$  for which  $P_n < (1 - \varepsilon)^n$ .*

We prove this with  $\varepsilon = .001$  for  $n \geq n_0$ . While this could be improved somewhat, a proof of (1) seems to require substantial new ideas.

Let  $m_{ij}$ ,  $i, j \geq 1$ , be chosen at random from  $\{\pm 1\}$  independently of each other (an infinite random matrix). Let  $M_n$  be the finite matrix  $(m_{ij})_{1 \leq i, j \leq n}$ .

Our return to the estimation of  $P_n$  was motivated in part by a question proposed by Benji Weiss: Is it true that  $\sum P_n < \infty$ ?

The point of the question is that an affirmative answer (as provided by our Theorem 1) implies, via the Borel-Cantelli Lemma, that with probability 1 only finitely many of the  $M_n$  are singular.

A few additional applications and extensions are mentioned in Section 4. Of the extensions, the most interesting is perhaps Corollary 4(b), which, improving a result of Odlyzko [20], says that for an appropriate constant  $C$ ,  $n - C$  random  $\{\pm 1\}$ -vectors are not only (a.s.) independent, but in fact span no other  $\{\pm 1\}$ -vectors.

The problem of estimating  $P_n$  turns out to be closely related to questions arising in various other areas, e.g., geometry (Füredi [4]), threshold logic (Zuev [27]), and associative memories (Kanter-Sompolinsky [11]). Consequences of Theorem 1 for some of these are also discussed in Section 4.

For more on the by now vast literature on random matrices see, e.g., Girko [6] or Mehta [16]. See also Odlyzko [20] for a few problems more or less related to the present work.

In the remainder of this section we sketch the main points in the proof of Theorem 1. Let us in particular draw the reader's attention to Theorem 2, which is central to the proof of Theorem 1, and seems also to be of independent interest.

**1.2. Linear algebra.** For  $\underline{a} \in \mathbf{R}^n - \{0\}$ , let

$$p(\underline{a}) = \Pr(\underline{\varepsilon}^t \underline{a} = 0),$$

where  $\underline{\varepsilon}$  is drawn uniformly from  $\{\pm 1\}^n$ , and denote by  $E_{\underline{a}}$  the event " $M\underline{a} = 0$ ", where  $M = M_n$  is a random  $n \times n$   $\pm 1$ -matrix. Thus,  $\Pr(E_{\underline{a}}) = [p(\underline{a})]^n$ , and  $P_n = \Pr(\cup\{E_{\underline{a}} : \underline{a} \in \mathbf{Z}^n - \{0\}\})$ .

Of course "Boole's inequality"  $P_n \leq \sum \Pr(E_{\underline{a}})$  gives nothing here. For  $\underline{a}$ 's with very small  $p(\underline{a})$ , the following trivial lemma gives a usable lower bound. (This is just the case  $k = n$  of Lemma 2 below, but we prove it separately both by way of illustration and because it plays a special role in what follows.)

**Lemma 1.** For any  $p_0, 0 < p_0 < 1,$

$$Pr(\cup\{E_{\underline{a}} : p(\underline{a}) \leq p_0\}) \leq np_0.$$

*Proof.* The inequality is implied by the following observation. Let  $L_i$  denote the event “the  $i$ th row of  $M$  is a linear combination of the other  $n - 1$  rows”. Then,  $Pr([\cup\{E_{\underline{a}} : p(\underline{a}) \leq p_0\}] \cap L_i) \leq p_0.$  Indeed,  $p_0$  is an upper bound on the probability of this event conditioned on any values in the other  $n - 1$  rows, since for such matrices the set  $\{\underline{a} : p(\underline{a}) \leq p_0, M\underline{a} = \underline{0}\}$  is determined already by those  $n - 1$  rows. Taking total probability gives the bound  $p_0.$   $\square$

To deal with large  $p(\underline{a})$ 's, we must somehow exploit dependencies among the  $E_{\underline{a}}$ 's. A framework for doing so, based on the idea that linearly dependent  $\underline{a}$ 's tend to be annihilated by the same  $M$ 's, is given by the following lemma. (For  $S \subset \mathbf{R}^n, \dim(S)$  is the dimension of the subspace spanned by  $S.$ )

**Lemma 2.** Let  $S$  be a subset of  $\mathbf{R}^n - \{0\}, k = \dim(S),$  and  $p(S) = \max\{p(\underline{a}) : \underline{a} \in S\}.$  Then,

$$Pr(\cup\{E_{\underline{a}} : \underline{a} \in S\}) \leq \binom{n}{k-1} p(S)^{n-k+1}.$$

This is proved in Section 3.3. The factor  $\binom{n}{k-1}$  is somewhat wasteful. We will eventually substitute for Lemma 2 a more technical variant (Lemma 4) which gives a slightly better value of  $\epsilon$  in Theorem 1.

**1.3. Subspaces.** In the proof of Theorem 1 we will try to cover  $\mathbf{Z}^n \setminus \{0\}$  by a small number of subspaces of moderate dimensions, and simply add up the bounds in Lemma 2 (or rather Lemma 4). That is, we will use the estimate

$$(2) \quad P_n = Pr(\cup\{E_{\underline{a}} : \underline{a} \in \mathbf{Z}^n \setminus \{0\}\}) \leq \sum_{i \geq 0} Pr(\cup\{E_{\underline{a}} : \underline{a} \in S_i\}),$$

where  $\{S_i\}$  is an appropriate cover of  $\mathbf{Z}^n \setminus \{0\},$  and use Lemma 4 to bound the summands. We will choose  $S_0 = \{\underline{a} : p(\underline{a}) \leq p_0\}$  ( $p_0 \approx (1 - \epsilon)^n$ ), so  $\dim(S_0) = n.$  But for  $i \neq 0, \dim(S_i)$  will be roughly  $\gamma n,$  with  $\gamma < 1$  a constant to be specified later.

It is perhaps most natural here to try to use  $S_i$ 's of the form  $S(I) = \cap\{\underline{\epsilon}^\perp : \underline{\epsilon} \in I\},$  where  $I$  is a set of linearly independent vectors from  $\{\pm 1\}^n,$  the idea being that if  $\underline{a} \in \mathbf{Z}^n$  satisfies  $\underline{\epsilon}^t \underline{a} = 0$  for many  $\underline{\epsilon} \in \{\pm 1\}^n,$  then  $\underline{a}$  should lie in many  $S(I)$ 's. While this seems not quite to work, something similar does lead to usable  $S_i$ 's. Namely, our subspaces will be of the above form  $S(I)$  with  $I$  a set of about  $(1 - \gamma)n$  linearly independent vectors from  $\{-1, 0, +1\}^n,$  each with exactly  $d$  non-zero components, for some  $d \approx \mu n$  with  $\mu$  a small constant.

To show that a moderate number of such subspaces can cover  $\mathbf{Z}^n,$  we use a probabilistic construction (Lemma 5).

**Definition.** Let  $V_d$  be the set of vectors  $\underline{\varepsilon} \in \{-1, 0, +1\}^n$  with exactly  $d$  non-zero coordinates. A  $d$ -sum in  $\underline{a}$  is an expression of the form  $\sum_{i=1}^n \varepsilon_i a_i$ , where  $\underline{\varepsilon} \in V_d$ . We write

$$\Sigma_d(\underline{a}) = \{\underline{\varepsilon} \in V_d : \underline{\varepsilon}^t \underline{a} = 0\}$$

and

$$\sigma_d(\underline{a}) = |\Sigma_d(\underline{a})|.$$

The analogue of  $p(\underline{a})$  for  $d$ -sums is

$$p_d(\underline{a}) = \sigma_d(\underline{a})/|V_d| = \sigma_d(\underline{a}) / \binom{n}{d} 2^d.$$

**1.4. A Halász-type inequality.** As noted above, we may place all  $\underline{a}$ 's for which  $p(\underline{a})$  is small— $p(\underline{a}) < p_0 \approx (1 - \varepsilon)^n$ , say—in a single set  $S_0$ , which by Lemma 1 contributes only  $np_0$  to the bound (2).

The crucial question posed by Lemma 2 (or Lemma 4) is: What can be said about  $\underline{a}$ 's for which  $p(\underline{a})$  is large?

For example, as observed by Erdős [2] in connection with the “Littlewood-Offord Problem”, Sperner’s Theorem ([24] or, e.g., [7]) implies that if  $p(\underline{a})$  is much bigger than  $n^{-1/2}$ , then  $\underline{a}$  has relatively small support.

A second example is given by a theorem of Sárközy and Szemerédi [23], which says that if  $a_1, \dots, a_n$  are distinct, then

$$p(\underline{a}) = O(n^{-3/2}).$$

(The precise bound here is a celebrated result essentially due to Stanley; see [25], [21].) So if  $p(\underline{a})$  is much bigger than  $n^{-3/2}$ , then  $\underline{a}$  must have many repeated entries. (Incidentally, one can use this with Lemma 2 to show  $P_n = O(n^{-3/2})$ , which already answers the question of Weiss mentioned earlier. This was in fact our starting point.)

For smaller values of  $p(\underline{a})$ , some deep theorems of Halász [8, 9] apply. They say, roughly, that if  $p(\underline{a})$  is much bigger than  $n^{-(2r+1)/2}$ , then there must be considerable duplication among the sums  $\pm a_{i_1} \pm \dots \pm a_{i_r}$ .

We give here a more abstract condition, which says that for  $d$  much less than  $n$ ,  $p(\underline{a}) = p_n(\underline{a})$  tends to be significantly less than  $p_d(\underline{a})$ . This is perhaps the most important step in the proof of Theorem 1.

In terms of random walks, the result says roughly the following. Let  $a_1, \dots, a_n$  be integers and  $\mu \in (0, \frac{1}{2})$ . Then the probability that a random walk with step sizes  $a_1, \dots, a_n$  returns to the origin at time  $n$  is less by a factor  $O(\sqrt{\mu})$  than the corresponding probability for the “lazy” walk which at the  $i$ th step moves  $a_i$  or  $-a_i$ , each with probability  $\mu$ , and otherwise remains where it is.

While this is certainly the case for ordinary random walks, it is somewhat surprising that such a relation between  $p_n$  and  $p_d$  can be established for random walks with arbitrary step sizes, since in such generality it is hopeless to determine, or even to give reasonable estimates for, individual values of  $p_n$ .

Let  $\text{supp}(\underline{a})$  be the number of non-zero components in  $\underline{a}$ .

**Theorem 2.** Let  $\lambda < 1$  be a positive number, and let  $k$  be a positive integer such that  $4\lambda k^2 < 1$ . If  $\underline{a} \in \mathbf{Z}^n - \{\underline{0}\}$ , then

$$(3) \quad p(\underline{a}) \leq \left[ \frac{1}{k(1-4\lambda k^2)} + \frac{1}{1-4\lambda} e^{-(1-4\lambda) \text{supp}(\underline{a})/(4k^2)} \right] Q_\lambda(\underline{a}),$$

where  $Q = Q_\lambda(\underline{a})$  is defined as

$$Q = \sum_{d=0}^n \binom{n}{d} (\lambda e^{-\lambda})^d (1 - \lambda e^{-\lambda})^{n-d} p_d(\underline{a}).$$

The choice  $k = (12\lambda)^{-1/2}$  leads to the following corollary.

**Theorem 3.** There exists (for each  $\lambda$ )  $K(\lambda)$  such that if  $(12\lambda)^{-1/2}$  is integral and  $\text{supp}(\underline{a}) \geq K(\lambda)$ , then

$$p(\underline{a}) < c_0 \sqrt{\lambda} Q,$$

where  $c_0 < 5.2$ .

*Remark.* Set  $\mu = \lambda e^{-\lambda}$ . The weight function in  $Q$  is a binomial distribution which is highly concentrated around the expected value  $\mu n$ . Hence, typically, only the terms  $p_d(\underline{a})$  with  $d \approx \mu n$  matter. Thus, Theorem 3 roughly says the following. Let  $\mu > 0$  be small. If  $a_1, \dots, a_n$  are non-zero integers, and many (more than a  $(1 - \mu)^n$  proportion) of the  $2^n$  signed sums of the  $a$ 's are 0, then, for some  $d \approx \mu n$ , an even larger (by a factor  $\sqrt{n/d}$ ) proportion of the  $\binom{n}{d} 2^d$  signed sums of  $d$  terms are 0.

We just mention two illuminating examples:

*Example 1* (verified for us by Imre Ruzsa [22]). If  $a_i = i^\alpha$ ,  $\alpha$  a positive integer, then (for large enough  $d, n$ )  $p(\underline{a}) \sim cn^{-\alpha-1/2}$ , while  $p_d(\underline{a}) \sim cn^{-\alpha} d^{-1/2}$ .

*Example 2.* If the  $a_i$  are random integers chosen from the range  $\{1, 2, \dots, M\}$ , then  $p(\underline{a}) \sim c/(M\sqrt{n})$  and  $p_d(\underline{a}) \sim c/(M\sqrt{d})$ .

Now fix a positive constant  $\epsilon'$ , and set

$$q_\lambda(\underline{a}) = \max\{p_d(\underline{a}) : |d - \mu n| < \epsilon' n\}.$$

By the Chernoff bound,

$$(4) \quad \begin{aligned} Q_\lambda(\underline{a}) &\leq q_\lambda(\underline{a}) + \sum_{|d-\mu n| \geq \epsilon' n} \binom{n}{d} \mu^d (1-\mu)^{n-d} \\ &\leq q_\lambda(\underline{a}) + 2e^{-DIV(\mu, \epsilon')n}, \end{aligned}$$

where

$$\begin{aligned} DIV(\mu, \epsilon') &= \min\{D(\mu + \epsilon' || \mu), D(\mu - \epsilon' || \mu)\} \\ &= \begin{cases} D(\mu - \epsilon' || \mu) & \text{if } \mu \leq 1/2, \\ D(\mu + \epsilon' || \mu) & \text{if } \mu > 1/2, \end{cases} \end{aligned}$$

with  $D(x||\mu) = x \log(x/\mu) + (1 - x) \log((1 - x)/(1 - \mu))$ , the information theoretical divergence of  $x$  from  $\mu$ .

Thus, under the conditions of Theorem 3 we have, provided  $\text{supp}(\underline{a}) > K(\lambda)$ ,

$$(5) \quad p(\underline{a}) \leq c_0 \sqrt{\lambda} \left( q_\lambda(\underline{a}) + 2e^{-DIV(\mu, \varepsilon')^n} \right)$$

with  $c_0 < 5.2$ .

**1.5. Sketch of the proof of the main theorem.** The proof of Theorem 2, using ideas of Halász, is given in Section 2. In Section 3, we complete the proof of Theorem 1. The argument (ignoring  $\underline{a}$ 's of small support, which are easily handled directly) will go roughly as follows.

We fix a small  $\lambda$  (eventually  $1/108$ ) and  $\varepsilon$  somewhat smaller ( $.002$ ). Vectors  $\underline{a}$  with  $p(\underline{a}) < (1 - \varepsilon)^n$  are placed in  $S_0$ .

For the remaining vectors, as indicated earlier, we use  $S_i$ 's based on  $d$ -sums and having dimension  $\gamma n$ , where  $d$  takes various values in the vicinity of  $\lambda n$  and  $\gamma = \varepsilon n/d$ .

The crucial difference between  $d$ -sums and full sums is in the factor  $\sqrt{\lambda}$ : for given  $\sigma$ , the number of  $S_i$ 's used to cover  $\underline{a}$ 's with  $q_\lambda(\underline{a}) = p_d(\underline{a})$  and  $\sigma_d(\underline{a}) \approx \sigma$  behaves roughly like  $\binom{n}{d} 2^d / \sigma^{(1-\gamma)n} \approx p_d(\underline{a})^{-(1-\gamma)n}$ ; the binomial coefficient in Lemma 2 turns out not to be too important; and the factor  $(\sqrt{\lambda})^{(1-\gamma)n}$  from  $p^{n-k+1}$  is small enough to give the desired exponential bound.

## 2. PROOF OF THEOREM 2

Recall that

$$(6) \quad \prod_{i=1}^n \cos \alpha_i = 2^{-n} \sum_{\underline{\varepsilon}} \cos(\varepsilon_1 \alpha_1 + \dots + \varepsilon_n \alpha_n),$$

where the sum is over  $\underline{\varepsilon} \in \{\pm 1\}^n$ . This gives, for any  $\underline{a} \in \mathbf{R}^n$ ,

$$(7) \quad \begin{aligned} p(\underline{a}) &= 2^{-n} \frac{1}{2\pi} \int_0^{2\pi} \sum_{\underline{\varepsilon}} \cos((\varepsilon_1 a_1 + \dots + \varepsilon_n a_n)t) dt \\ &= \frac{1}{2\pi} \int_0^{2\pi} \prod_{i=1}^n \cos(a_i t) dt. \end{aligned}$$

*Remark.* The reader may notice that the integrand on the right-hand side of (7) is the Fourier transform of the distribution of  $\sum_{i=1}^n \varepsilon_i a_i$ , where  $\underline{\varepsilon}$  is chosen uniformly from  $\{\pm 1\}^n$ . This is not by accident. Esséen's concentration lemma [3] says that for any finite measure  $\mu$ ,

$$\sup_y \int_{|x-y| \leq 1} \mu(dx) \leq c \int_{|t| \leq 1} |\phi(t)| dt,$$

where  $\phi(t)$  is the Fourier transform  $\phi(t) = \int e^{itx} \mu(dx)$  and  $c$  is an absolute constant.

This remark may be used to generalize Theorem 1 to random matrices with arbitrary independent identically distributed non-degenerate entries. (For such a generalization of the result of [12], see [13].)

Returning to (7) and using the inequality  $|x| \leq e^{-(1-x^2)/2}$  together with  $1 - \cos^2 \alpha = (1 - \cos(2\alpha))/2$  and the integrality of  $a_i$  we have

$$p(\underline{a}) \leq \frac{1}{2\pi} \int_0^{2\pi} \exp \left\{ -\frac{1}{4} \sum_{i=1}^n (1 - \cos(a_i t)) \right\} dt.$$

Setting

$$f(t) = \frac{1}{4} \sum_{i=1}^n (1 - \cos(a_i t)),$$

we define

$$T(x) = \{t \in (0, 2\pi) : f(t) \leq x\} \quad \text{and} \quad g(x) = \frac{1}{2\pi} |T(x)|$$

( $|\cdot|$  stands for Lebesgue measure).

Using this  $f$  and  $g$ , we can rewrite our estimate as

$$(8) \quad p(\underline{a}) \leq \frac{1}{2\pi} \int_0^{2\pi} e^{-f(t)} dt = \int_0^{2\pi} \int_{f(t)}^{\infty} \frac{1}{2\pi} e^{-x} dx dt = \int_0^{\infty} e^{-x} g(x) dx.$$

The following inequality of Halász ([8], see also [9]) is at the heart of our proof. For any  $x > 0$  and positive integer  $k$ ,

$$(9) \quad g(x) \leq g(k^2 x)/k$$

provided  $g(k^2 x) < 1$ , which certainly holds if  $k^2 x \leq \text{supp}(\underline{a})/4$  since  $\frac{1}{2\pi} \int_0^{2\pi} f(t) dt = \text{supp}(\underline{a})/4$  and  $f$  is not constant.

For the convenience of the reader, we sketch Halász's proof of (9). For a fixed integer  $k \geq 2$ , let  $T^*(x) = \{t_1 + \dots + t_k : t_i \in T(x)\}$  (addition modulo  $2\pi$ ). Then (9) follows from

$$(10) \quad T^*(x) \subset T(k^2 x),$$

$$(11) \quad \frac{1}{2\pi} |T^*(x)| \geq \min\{kg(x), 1\}.$$

The set containment (10) follows from

$$1 - \cos(\alpha) = 2 \sin^2 \left( \frac{\alpha}{2} \right)$$

and

$$\sin^2 \left( \sum_{i=1}^k \alpha_i \right) \leq \left( \sum_{i=1}^k |\sin \alpha_i| \right)^2 \leq k \sum_{i=1}^k \sin^2 \alpha_i.$$

For the proof of (11), see [8]. (Alternatively, it is an easy consequence of the Cauchy-Davenport Theorem (e.g., Halberstam-Roth [10])).

We return to the proof of Theorem 2. Let us fix a positive number  $\lambda < 1$ . First we use Chernoff’s method to show

$$(12) \quad g(x) \leq e^{4\lambda x} Q.$$

By Markov’s inequality,

$$g(x) = \frac{1}{2\pi} \left| \left\{ t \in (0, 2\pi) : \exp \left\{ \lambda \sum_i \cos(a_i t) \right\} \geq \exp \{ \lambda(n - 4x) \} \right\} \right| \\ \leq \exp \{ -\lambda(n - 4x) \} \frac{1}{2\pi} \int_0^{2\pi} \exp \left\{ \lambda \sum_i \cos(a_i t) \right\} dt.$$

Using the inequality

$$e^{\lambda z} \leq e^\lambda - \lambda(1 - z) \quad \text{for } |z| \leq 1,$$

recalling  $\mu = \lambda e^{-\lambda}$ , and then using (6), we have

$$\exp \left\{ \lambda \sum_i \cos(a_i t) \right\} \leq \prod (e^\lambda - \lambda + \lambda \cos(a_i t)).$$

Thus,

$$g(x) \leq e^{\lambda x} \frac{1}{2\pi} \int_0^{2\pi} \prod (1 - \lambda e^{-\lambda} + \lambda e^{-\lambda} \cos(a_i t)) dt \\ = e^{\lambda x} \sum_d \sum_{i_1 < \dots < i_d} \mu^d (1 - \mu)^{n-d} \frac{1}{2\pi} \int_0^{2\pi} \prod_{j=1}^d \cos(a_{i_j} t) dt \\ = e^{\lambda x} \sum_d \mu^d (1 - \mu)^{n-d} \frac{\sigma_d(\underline{a})}{2^d} = e^{\lambda x} \sum_d \binom{n}{d} \mu^d (1 - \mu)^{n-d} p_d(\underline{a}),$$

proving (12).

Now, let  $k$  be a positive integer with  $4\lambda k^2 < 1$ . Let us write  $S = \text{supp}(\underline{a})/(4k^2)$  and split the estimate (8) as

$$p(\underline{a}) \leq \int_0^\infty e^{-x} g(x) dx = \int_0^S e^{-x} g(x) dx + \int_S^\infty e^{-x} g(x) dx.$$

We start with the second integral. By (12),

$$\int_S^\infty e^{-x} g(x) dx \leq \int_S^\infty e^{-x} e^{4\lambda x} Q dx = \frac{Q}{1 - 4\lambda} e^{-(1-4\lambda)S}.$$

In the domain of the first integral we have  $k^2 x \leq k^2 S = \text{supp}(\underline{a})/4$ . Thus (9) applies, and with (12) yields

$$\int_0^S e^{-x} g(x) dx \leq \int_0^S \frac{1}{k} g(k^2 x) e^{-x} dx \leq \frac{Q}{k} \int_0^S e^{-(1-4\lambda k^2)x} dx \leq \frac{Q}{k(1 - 4\lambda k^2)},$$

proving Theorem 2.  $\square$

### 3. PROOF OF THEOREM 1

We assume throughout that  $n$  is large enough to support our approximations. We generally treat large real numbers as integers without comment; if the reader prefers, replacing each such number by its floor, say, removes this imprecision without affecting any of the arguments.

**3.1.  $a$ 's with many 0's.** We first dispose of the easy case of  $\underline{a}$ 's with many 0's. The following observation is from [14] (see also [1], p. 348, Lemma 10).

For all  $K$ ,

$$(13) \quad \begin{aligned} &Pr(\cup\{E_{\underline{a}} : \text{supp}(\underline{a}) \leq K\}) \\ &\leq \sum_{k=2}^K \binom{n}{k} \binom{n}{k-1} \left[ 2^{-k} \binom{k}{\lfloor k/2 \rfloor} \right]^{n-k+1}. \end{aligned}$$

In particular,

$$(14) \quad Pr\left(\cup\left\{E_{\underline{a}} : \text{supp}(\underline{a}) < n - 3\frac{n}{\log_2 n}\right\}\right) < n^3 2^{-n}.$$

*Remark.* This can easily be improved to the true value  $(1 + o(1))n^2 2^{-n}$ , e.g., by substituting  $\binom{n-1}{k-2}$  for  $\binom{n}{k-1}$  in (13). Thus, vectors  $\underline{a}$  with at least  $3n/\log_2 n$  0's do not obstruct a proof of (1).

**3.2. An Odlyzko-type lemma.** We need one more easy observation, which generalizes Theorem 2 of [20]. (Recall that  $V_d$  is the set of  $\{-1, 0, +1\}$ -vectors with exactly  $d$  non-zero coordinates.)

**Lemma 3.** *If  $S$  is a  $D$ -dimensional subspace of  $\mathbf{R}^n$ , then*

$$(15) \quad |S \cap V_d| \leq F(D, d) := \sum_{i=D-n+d}^d \binom{D}{i} 2^i.$$

*Proof.* Without loss of generality, the set of restrictions of vectors in  $S$  to the coordinates  $\{1, \dots, D\}$  is of dimension  $D$ . Thus different vectors in  $S \cap V_d$  have different restrictions to these coordinates, each a  $\{-1, 0, 1\}$ -vector with between  $D - n + d$  and  $d$  nonzero coordinates. This gives (15).  $\square$

The case  $D = n$  is Odlyzko's result, which we state for future reference as

**Corollary 1.** *For  $V$  a subspace of  $\mathbf{R}^n$  and  $\underline{r}$  chosen uniformly at random from  $\{\pm 1\}^n$ ,*

$$Pr(\underline{r} \in V^\perp) \leq 2^{-\dim(V)}.$$

**3.3. Back to Lemma 2.** As mentioned earlier, a little more care with Lemma 2 eventually gives a somewhat better  $\varepsilon$  in Theorem 1.

**Lemma 4.** *Suppose the  $k$ -dimensional subset  $S$  of  $\mathbf{R}^n$  and numbers  $p, \varepsilon''$  satisfy*

$$(16) \quad p(S) \leq p,$$

*together with the technical conditions*

$$(17) \quad p < \varepsilon'' < 1/2,$$

$$(18) \quad p^{-1}2^{-\varepsilon''n} < n^{-2}.$$

*Then (for large enough  $n$ ),*

$$Pr(\cup\{E_{\underline{a}} : \underline{a} \in S\}) < \binom{n}{\varepsilon''n} p^{n-k+1}.$$

**Proofs of Lemmas 2 and 4.** Let  $r_1, \dots, r_n$  be the rows of  $M$ . A necessary condition for the event  $E_S := \cup\{E_{\underline{a}} : \underline{a} \in S\}$  is that there be at most  $k - 1$  indices  $j \in [n]$  for which

$$(19) \quad \dim\left(S \cap \bigcap_{l \leq j} r_l^\perp\right) < \dim\left(S \cap \bigcap_{l < j} r_l^\perp\right).$$

Call the event in (19)  $F_j$ . For  $I \subset [n]$ , let  $H_I$  be the event  $\{S \cap \bigcap_{i \in I} r_i^\perp \neq \{0\}\}$ .

*Proof of Lemma 2.* The discussion to this point implies

$$Pr(E_S) \leq \sum \left\{ Pr\left(H_{[n] \setminus J} \cap \bigcap_{j \in J} \overline{F}_j\right) : J \subset [n], |J| = n - k + 1 \right\}.$$

But for any  $J \subset [n]$ ,

$$(20) \quad Pr\left(H_{[n] \setminus J} \cap \bigcap_{j \in J} \overline{F}_j\right) \leq Pr\left(\bigcap_{j \in J} \overline{F}_j \mid H_{[n] \setminus J}\right) \leq p(S)^{|J|}.$$

To see this, fix (and condition on) rows  $r_j, j \notin J$ , satisfying  $H_{[n] \setminus J}$ . Then  $\overline{F}_j$  implies that  $r_j$  lies in  $\langle S \cap \bigcap_{l < j} r_l^\perp \rangle^\perp$ , so in particular is orthogonal to any given  $\underline{a} \in S \cap \bigcap_{l < j} r_l^\perp$ . Since the latter occurs with probability at most  $p(S)$ , and the rows are chosen independently, we have (20). The lemma follows.  $\square$

*Proof of Lemma 4.* We just elaborate the preceding proof a little. Note we may assume

$$(21) \quad \varepsilon''n \leq k \leq n - \varepsilon''n,$$

since otherwise the conclusion follows from Lemma 2.

Given  $I \subset [n]$  with  $|I| \leq k - 1$ , set  $J = [n] \setminus I$ ,

$$G_I = H_I \cap \bigcap_{i \in I} F_i, \quad F_I = G_I \cap \bigcap_{j \in J} \bar{F}_j.$$

Thus

$$(22) \quad \Pr(E_S) \leq \sum \{ \Pr(F_I) : I \subset [n], |I| \leq k - 1 \}.$$

For  $j \in J$  let  $t(j) = |I \setminus [j]|$ . Our basic inequality is

$$(23) \quad \Pr(F_I) \leq \Pr \left( \bigcap_{j \in J} \bar{F}_j \mid G_I \right) \leq \prod_{j \in J} \min \{ 2^{-t(j)-1}, p(S) \} =: f(I)$$

To see this, fix rows  $r_i$ ,  $i \in I$ , satisfying  $G_I$ . Then  $\bar{F}_j$  requires that

$$(24) \quad r_j \in \left\langle S \cap \bigcap_{i < j} r_i^\perp \right\rangle^\perp.$$

Now  $G_I$  implies that

$$\dim \left\langle S \cap \bigcap_{i < j} r_i^\perp \right\rangle \geq \dim \left\langle S \cap \bigcap_i r_i^\perp \right\rangle + t(j) \geq t(j) + 1,$$

so by Corollary 1, (24) occurs with probability at most  $2^{-t(j)-1}$ . But (24) also requires that  $r_j$  be orthogonal to any given  $a \in S \cap \bigcap_{i < j} r_i^\perp$ , so occurs with probability at most  $p(S)$ . This gives (23).

Consider first  $I$  of size  $k - 1$ . Set  $m = k - \epsilon''n$  and suppose  $|I \cap [m]| = i$ . Then  $t(j) \geq k - 1 - i$  for  $j \in J \cap [m]$ , and so, by (23),

$$f(I) \leq 2^{-(k-i)(m-i)} p(S)^{n-k-m+i+1}.$$

Letting  $I$  vary, this gives

$$(25) \quad \begin{aligned} & \sum \{ f(I) : I \subset [n], |I| = k - 1 \} \\ & < \sum_{i=0}^m \binom{m}{i} \binom{n-m}{k-1-i} 2^{-(k-i)(m-i)} p(S)^{n-k-m+i+1} \\ & < (1 + o(1)) \binom{n-m}{k-1-m} p^{n-k+1}, \end{aligned}$$

the second inequality by (16), (18).

For smaller  $I$ , notice that for any  $I \subset I' \subset [n]$ ,

$$f(I) \leq p^{|I' \setminus I|} f(I').$$

Thus (see (22))

$$\begin{aligned}
 \Pr(E_S) &\leq \sum_{t=0}^{k-1} \sum \{f(I) : I \subset [n], |I| = k - 1 - t\} \\
 &\leq \sum_{t=0}^{k-1} \frac{\binom{n}{k-1-t}}{\binom{n}{k-1}} p^t \sum \{f(I) : I \subset [n], |I| = k - 1\} \\
 (26) \quad &< (1 + o(1)) \sum_{t=0}^{k-1} \frac{\binom{n}{k-1-t}}{\binom{n}{k-1}} p^t \binom{n-m}{k-1-m} p^{n-k+1}
 \end{aligned}$$

$$(27) \quad < \binom{n}{\varepsilon'' n} p^{n-k+1}.$$

The inequality (26) is from (25), while (27) is a consequence of (17) and (21).  $\square$

**3.4. A random construction.** We can now construct the sets  $S_i$  for use in Lemma 4. Our basic parameters are  $\lambda$ ,  $\mu = \lambda e^{-\lambda}$ ,  $\varepsilon$ , and  $\varepsilon' = \alpha\mu$ , all small positive constants. We assume first of all that

$$(28) \quad 1 - \varepsilon > e^{-DIV(\mu, \varepsilon')}.$$

(As mentioned earlier, we also assume  $n$  is large.)

Then Theorem 2 (see (5)) guarantees that for any  $\underline{a}$  with

$$(29) \quad q_\lambda(\underline{a}) > (1 - \varepsilon)^n \quad \text{and} \quad \text{supp}(\underline{a}) > K(\lambda),$$

we have the crucial inequality

$$(30) \quad p(\underline{a}) < 5.2\sqrt{\lambda} q_\lambda(\underline{a}).$$

Let us fix (temporarily) two integers  $d$  and  $\sigma$  with

$$(31) \quad |d - \mu n| < \varepsilon' n$$

and

$$(32) \quad (1 - \varepsilon)^n N \leq \sigma \leq N,$$

where  $N = N_d = |V_d| = \binom{n}{d} 2^d$ .

For such  $d$  and  $\sigma$ , we also define the sets

$$M(d, \sigma) = \{\underline{a} \in \mathbf{Z}^n - \{0\} : \text{supp}(\underline{a}) > K(\lambda), q_\lambda(\underline{a}) = p_d(\underline{a}), \text{ and } \sigma_d(\underline{a}) = \sigma\}.$$

As mentioned above, vectors  $\underline{a}$  in any of the sets  $M(d, \sigma)$  satisfy (30).

Define in addition  $\delta = d/n$ ,  $\gamma = \varepsilon/\delta$ , and  $D = (1 - \gamma)n$ . We will choose the parameters so that  $d \leq D/2$ , that is,

$$(33) \quad \gamma \leq 1 - 2\delta,$$

implying that  $F(D, d) < 2\binom{D}{d} 2^d$ . Note also that, since  $\gamma < 1$ ,

$$(34) \quad (1 - \gamma)^\delta < 1 - \varepsilon.$$

We will cover  $M(d, \sigma)$  by a number of sets  $S_i$ , each consisting of  $\underline{a}$ 's which are orthogonal to some  $D$  linearly independent vectors from  $V_d$ .

**Lemma 5.** *There exist  $m < (1 + o(1)) \left(\frac{N}{\sigma}\right)^D \log \binom{N}{\sigma}$  and  $W_1, \dots, W_m$ , each a set of  $D$  linearly independent vectors from  $V_d$ , such that any  $\sigma$ -subset of  $V_d$  contains at least one of the  $W_i$ .*

*Remark.* If we don't require that the elements of  $W_i$  be independent, then Lemma 5 becomes a special case of a hypergraph covering result of Lovász [15] (see also Füredi [5] for a survey of this and related topics). In the present situation we use Lemma 3 to show that the independence requirement doesn't really cause any trouble.

*Proof.* Fix  $\Sigma \subset V_d$  with  $|\Sigma| = \sigma$ , and set  $q = \sigma/N$ . Let  $w_1, \dots, w_D$  be chosen uniformly and independently from  $V_d$ , and set

$$F = \{w_1, \dots, w_D \text{ are linearly independent elements of } \Sigma\}.$$

We show that

$$Pr(F) = (1 - o(1))q^D.$$

Since  $Pr(w_1, \dots, w_D \in \Sigma) = q^D$ , it's enough to show

$$(35) \quad Pr(F|w_1, \dots, w_D \in \Sigma) = 1 - o(1).$$

This probability is just  $P := Pr(v_1, \dots, v_D \text{ are linearly independent})$ , where  $v_1, \dots, v_D$  are drawn uniformly and independently from  $\Sigma$ . By Lemma 3,

$$\begin{aligned} P &> 1 - \sum_{i=1}^D Pr(v_i \in \langle v_1, \dots, v_{i-1} \rangle) \\ &\geq 1 - \sum_{i=1}^D F(i-1, d)/\sigma \\ &> 1 - DF(D, d)/\sigma. \end{aligned}$$

But, writing  $(x)_s$  for  $x(x-1)\dots(x-s+1)$  and using (32), (33),

$$F(D, d)/\sigma < 2 \frac{(D)_d}{(n)_d(1-\varepsilon)^n} < 2 \frac{(1-\gamma)^d}{(1-\varepsilon)^n} = 2 \left[ \frac{(1-\gamma)^\delta}{1-\varepsilon} \right]^n.$$

So (35) follows from (34).

This gives the lemma: For appropriate  $m < (1 + o(1)) \left(\frac{N}{\sigma}\right)^D \log \binom{N}{\sigma}$ , if  $W_1, \dots, W_m$  are uniformly and independently chosen  $D$ -subsets of  $V_d$ , then the expected number of  $\sigma$ -subsets of  $V_d$  containing no independent  $W_i$  is

$$\binom{N}{\sigma} (1 - (1 - o(1))q^D)^m < 1,$$

so in particular there exist  $W_i$ 's as in the statement of the lemma.  $\square$

3.5. **Defining the  $S_i$ 's.** With notation as in Section 3.4, let  $W_1, \dots, W_m$  be as in Lemma 5 and set

$$S_j = \langle W_j \rangle^\perp \cap M(d, \sigma).$$

Suppose also that the constant  $\varepsilon''$  satisfies

$$(36) \quad \varepsilon'' > -\log_2(1 - \varepsilon).$$

Then applying Lemma 4 with  $S = S_j$  and  $k = D$ ,  $p = 5.2\sqrt{\lambda}\sigma/N$ , and using (30) we have

$$Pr(\cup\{E_{\underline{a}} : \underline{a} \in S_j\}) < \binom{n}{\varepsilon''n} (5.2\sqrt{\lambda}\sigma/N)^D$$

and (since  $m < (N/\sigma)^D N$ )

$$(37) \quad \begin{aligned} Pr(\cup\{E_{\underline{a}} : \underline{a} \in M(d, \sigma)\}) &< N \binom{n}{\varepsilon''n} (5.2\sqrt{\lambda})^D \\ &< \exp_2[(H_2(\delta) + \delta + H_2(\varepsilon'') + (1 - \gamma) \log_2(5.2\sqrt{\lambda}))n] =: m_d. \end{aligned}$$

Thus we have

$$(38) \quad Pr(\cup\{E_{\underline{a}} : \text{supp}(\underline{a}) > K(\lambda), q_\lambda(\underline{a}) > (1 - \varepsilon)^n\}) < \sum N_d m_d,$$

where the sum is over  $d$  satisfying (31).

The factor  $N$  in (38) is much more than is necessary, and, though this makes only a small difference in our final bound, we modify the argument as follows to reduce it.

Partition  $[(1 - \varepsilon)^n N, N]$  into intervals of the form  $I = \{\sigma : (1 + 1/n)^t < \sigma \leq (1 + 1/n)^{t+1}\}$  and use Lemma 5 to cover  $\cup_{\sigma \in I} M(d, \sigma)$  rather than an individual  $M(d, \sigma)$ . This has essentially no effect on any of the above calculations, and allows us to replace (38) by, for example,

$$(39) \quad Pr(\cup\{E_{\underline{a}} : \text{supp}(\underline{a}) > K(\lambda), q_\lambda(\underline{a}) > (1 - \varepsilon)^n\}) < n^2 \sum m_d.$$

Finally, set

$$\begin{aligned} S_0 &= \{\underline{a} \in \mathbf{Z}^n - \{0\} : q_\lambda(\underline{a}) \leq (1 - \varepsilon)^n \text{ OR } \text{supp}(\underline{a}) \leq K(\lambda)\} \\ &= \mathbf{Z}^n - \{0\} - \cup_{d, \sigma} M(d, \sigma). \end{aligned}$$

By (5) and (28), the conditions  $q_\lambda(\underline{a}) \leq (1 - \varepsilon)^n$ ,  $\text{supp}(\underline{a}) > K(\lambda)$ , with our eventual choice  $\lambda = 1/108$ , imply

$$(40) \quad p(\underline{a}) < (1 - \varepsilon)^n.$$

Thus, by Lemma 1 and (14)

$$Pr(\cup\{E_{\underline{a}} : \underline{a} \in S_0\}) < n(1 - \varepsilon)^n + n^3 2^{-n},$$

and finally,

$$(41) \quad P_n < n^2 \sum m_d + n(1 - \varepsilon)^n + n^3 2^{-n}.$$

**3.6. Choosing the parameters.** It remains to set the parameters. Essentially this amounts to choosing  $\lambda$  and  $\alpha = \varepsilon'/\mu$ , the other values then being dictated by (28), (31), (33), and (36).

A convenient, if not quite optimal choice, is  $\lambda = 1/108$  ( $k = 3$ ) ( $\mu = \lambda e^{-\lambda}$ ),  $\alpha = .5$  ( $\varepsilon' = .5\mu$ ),  $\varepsilon'' = .01$ , and  $\varepsilon = 0.002$  (i.e., something a little bigger than 0.001).

It is then straightforward to check that for any  $(1 - \alpha)\mu \leq \delta \leq (1 + \alpha)\mu$  and  $\gamma = \varepsilon/\delta$ , the expression

$$H_2(\delta) + \delta + H_2(\varepsilon'') + (1 - \gamma) \log_2(5.2\sqrt{\lambda}) = \frac{1}{n} \log_2 m_d$$

in (37) is less than  $\log_2(1 - \varepsilon)$ . (Its values at the extremes  $\delta = (1 \pm \alpha)\mu$  are less than  $\log_2(1 - \varepsilon)$ , and its second derivative with respect to  $\delta$  is positive between the extremes.)

Thus the bound in (41) is essentially equal to its second term and we have Theorem 1.  $\square$

#### 4. CONCLUDING REMARKS

It would be of considerable interest to say more about the distribution of  $\det(M_n)$ . Viewed “up close”, the distribution is not very nice—for instance it’s easy to see that  $\det(M_n)$  is always divisible by  $2^{n-1}$ —but it seems reasonable to expect some kind of limit distribution. The log-normal law for random determinants (Girko [6], Theorem 6.4.1) doesn’t apply here, for the entries don’t satisfy  $Em_{ij}^4 = 3$ .

It is not hard to see, based on the above results, that for any  $b$ ,

$$(42) \quad Pr(\det(M_n) = b) < (1 - \varepsilon)^n.$$

(Briefly, this is because: If we define  $p^*(\underline{a}) = \max_c Pr(\underline{\varepsilon}^t \underline{a} = c)$ , then the bound of Theorem 2 applies to  $p^*(\underline{a})$ . (Multiplying the integrands in (7) by  $\cos(ct)$  gives  $Pr(\underline{\varepsilon}^t \underline{a} = c)$  in place of  $p(\underline{a})$ , and the rest of the proof goes through as is.) This implies, as in the proof of Theorem 1, that with probability at least  $1 - (1 - \varepsilon)^n$  the first  $n - 1$  rows,  $r_1, \dots, r_{n-1}$ , of  $M$  annihilate a unique  $\underline{a}$ , which satisfies  $p^*(\underline{a}) > (1 - \varepsilon)^n$ . On the other hand, given any such  $r_1, \dots, r_{n-1}$  and  $\underline{a}$ ,  $Pr(\det(M) = b) \leq p^*(\underline{a})$  for any  $b$ .)

However, (42) should be far from the truth, which we believe to be that (except when  $b = 0$ ) the probability in (42) is  $\exp[-\Omega(n \log n)]$ .

It is also not hard to see that (37) together with (13) implies the following extensions.

**Corollary 2.** For any  $\gamma > 0$  there are constants  $C$  and  $\varepsilon^* > 0$  so that

$$Pr(M_n \underline{a} = \underline{0} \text{ for some } \underline{a} \text{ with } \text{supp}(\underline{a}) > C \text{ and } p(\underline{a}) > (1 - \varepsilon^*)^n) < \gamma^n.$$

**Corollary 3.** For every  $\gamma > 0$  there is a constant  $C$  such that

$$Pr(\text{rank}(M_n) < n - C) < \gamma^n.$$

**Corollary 4.** *There is a constant  $C$  so that if  $r \leq n - C$  and  $\underline{v}_1, \dots, \underline{v}_r$  are chosen (uniformly, independently) at random from  $\{\pm 1\}^n$ , then*

- (a)  $Pr(\underline{v}_1, \dots, \underline{v}_r \text{ are linearly dependent}) = (1 + o(1)) 2 \binom{r}{2} 2^{-n},$
- (b)  $Pr(\langle \underline{v}_1, \dots, \underline{v}_r \rangle \cap \{\pm 1\}^n \neq \{\underline{v}_1, \dots, \underline{v}_r\}) = (1 + o(1)) 4 \binom{r}{3} (\frac{3}{4})^n.$

(The precise error term in (a) is  $(1 + o(1)) 8 \binom{r}{4} (\frac{3}{8})^n$ , while (b) has an error term  $O((\frac{7}{10})^n)$ .)

Part (b) improves a result of Odlyzko [20]—studied in connection with a question on associative memories [11]— which gives the same bound provided  $r < n - 10n/\log n$ . As he observes, the error term  $O((\frac{7}{10})^n)$  is not best possible.

We conjecture that the conclusions of Corollary 3 hold provided  $r \leq n - 1$ , but expect that proving this will be about the same as proving (1).

Denote by  $T_n$  the number of *threshold functions* of  $n$  variables, that is, functions  $f: \{\pm 1\}^n \rightarrow \{\pm 1\}$  of the form

$$f(\underline{x}) = \text{sgn}(a_0 + a_1 x_1 + \dots + a_n x_n)$$

with  $a_i \in \mathbf{R}$ . The behavior of  $\log T_n$  was considered beginning in the late 1950s by various authors who established the bounds  $\binom{n}{2} < \log_2 T_n < n^2$ . (See Muroga [18] for details, related results, and references.) More recently, Zuev [27] showed, using results from [26] and [20], that  $\log_2 T_n \sim n^2$ . His precise bound is

$$(43) \quad T_n \geq \binom{2^n}{n - 10n/\log n} 2^{-(n-10n/\log n)} = \exp_2[n^2 - 10n^2/\log n - O(n \log n)],$$

whereas an upper bound is

$$(44) \quad 2 \sum_{i=0}^n \binom{2^n - 1}{i} = \exp_2[n^2 - n \log_2 n + O(n)].$$

Using Corollary 4(b) in place of [20] improves the lower bound (43) to  $\exp_2[n^2 - n \log_2 n - O(n)]$ . Moreover, if the conjecture that one may replace  $r \leq n - C$  by  $r \leq n - 1$  in the corollary is true, then a slight elaboration of the argument of [27] gives the asymptotics, not just of  $\log T_n$ , but of  $T_n$  itself: it would be asymptotic to the left-hand side of (44).

As pointed out by Füredi [4], Theorem 1 also gives some improvement in the bounds of that paper, namely, if  $n = O(d)$  and  $x_1, \dots, x_n$  are chosen uniformly and independently from  $\{\pm 1\}^d$ , then

$$Pr(\underline{0} \notin \text{conv}\{x_1, \dots, x_n\}) = \frac{h(n, d)}{2^{n-1}} + O(d^2(1 - \epsilon)^d),$$

where  $h(n, d) = \sum_{i=0}^d \binom{n-1}{i}$ .

## ACKNOWLEDGMENTS

We would like to thank József Beck and Gábor Halász for helpful conversations, and Volodia Blinovskiy for pointing out some errors in the manuscript.

## REFERENCES

- [1] Béla Bollobás, *Random graphs*, Academic Press, New York, 1985.
- [2] Pál Erdős, *On a lemma of Littlewood and Offord*, Bull. Amer. Math. Soc. **51** (1945), 898–902.
- [3] C. G. Esséen, *On the Kolmogorov-Rogozin inequality for the concentration function*, Z. Wahrsch. Verw. Gebiete **5** (1966), 210–216.
- [4] Zoltán Füredi, *Random polytopes in the  $d$ -dimensional cube*, Discrete Comput. Geom. **1** (1986), 315–319.
- [5] ———, *Matchings and covers in hypergraphs*, Graphs Combin. **4** (1988), 115–206.
- [6] V. L. Girko, *Theory of random determinants*, Math. Appl. (Soviet Ser.), vol. 45, Kluwer Acad. Publ., Dordrecht, 1990.
- [7] C. Greene and D. J. Kleitman, *Proof techniques in the theory of finite sets*, Studies in Combinatorics (G.-C. Rota, ed.), Math. Assoc. Amer., Washington, D.C., 1978.
- [8] Gábor Halász, *On the distribution of additive arithmetic functions*, Acta Arith. **27** (1975), 143–152.
- [9] ———, *Estimates for the concentration function of combinatorial number theory and probability*, Period. Math. Hungar. **8** (1977), 197–211.
- [10] H. Halberstam and K. F. Roth, *Sequences*, Vol. 1, Oxford Univ. Press, London and New York, 1966.
- [11] I. Kanter and H. Sompolinsky, *Associative recall of memory without errors*, Phys. Rev. (A) (3) **35** (1987), 380–392.
- [12] János Komlós, *On the determinant of  $(0, 1)$  matrices*, Studia Sci. Math. Hungar. **2** (1967), 7–21.
- [13] ———, *On the determinants of random matrices*, Studia Sci. Math. Hungar. **3** (1968), 387–399.
- [14] ———, Circulated manuscript, 1977.
- [15] László Lovász, *On the ratio of optimal integral and fractional covers*, Discrete Math. **13** (1975), 383–390.
- [16] Madan Lal Mehta, *Random matrices*, second ed., Academic Press, New York, 1991.
- [17] N. Metropolis and P. R. Stein, *A class of  $(0, 1)$ -matrices with vanishing determinants*, J. Combin. Theory **3** (1967), 191–198.
- [18] Saburo Muroga, *Threshold logic and its applications*, Wiley, New York, 1971.
- [19] A. M. Odlyzko, *On the ranks of some  $(0, 1)$ -matrices with constant row-sums*, J. Austral. Math. Soc. Ser. A **31** (1981), 193–201.
- [20] ———, *On subspaces spanned by random selections of  $\pm 1$  vectors*, J. Combin. Theory Ser. A **47** (1988), 124–133.
- [21] G. W. Peck, *Erdős conjecture on sums of distinct numbers*, Studies Appl. Math. **63** (1980), 87–92.
- [22] Imre Ruzsa, Private communication.
- [23] András Sárközy and Endre Szemerédi, *Über ein Problem von Erdős und Moser*, Acta. Arith. **11** (1965), 205–208.
- [24] E. Sperner, *Ein Satz über Untermenge einer endliche Menge*, Math. Z. **27** (1928), 544–548.
- [25] Richard P. Stanley, *Weyl groups, the hard Lefschetz theorem, and the Sperner property*, SIAM J. Alg. Discrete Math. **1** (1980), 168–184.
- [26] Thomas Zaslavsky, *Facing up to arrangements: Face-count formulas for partitions of space by hyperplanes*, Mem. Amer. Math. Soc., vol. 154, Amer. Math. Soc., Providence, RI, 1975.

- [27] Yu. A. Zuev, *Methods of geometry and probabilistic combinatorics in threshold logic*, Discrete Math. Appl. 2 (1992), 427–438.

**ABSTRACT.** We report some progress on the old problem of estimating the probability,  $P_n$ , that a random  $n \times n$   $\pm 1$ -matrix is singular:

**Theorem.** *There is a positive constant  $\varepsilon$  for which  $P_n < (1 - \varepsilon)^n$ .*

This is a considerable improvement on the best previous bound,  $P_n = O(1/\sqrt{n})$ , given by Komlós in 1977, but still falls short of the often-conjectured asymptotical formula  $P_n = (1 + o(1))n^2 2^{1-n}$ .

The proof combines ideas from combinatorial number theory, Fourier analysis and combinatorics, and some probabilistic constructions. A key ingredient, based on a Fourier-analytic idea of Halász, is an inequality (Theorem 2) relating the probability that  $\underline{a} \in \mathbf{R}^n$  is orthogonal to a random  $\underline{\varepsilon} \in \{\pm 1\}^n$  to the corresponding probability when  $\underline{\varepsilon}$  is random from  $\{-1, 0, 1\}^n$  with  $Pr(\varepsilon_i = -1) = Pr(\varepsilon_i = 1) = p$  and  $\varepsilon_i$ 's chosen independently.

(J. Kahn and J. Komlós) DEPARTMENT OF MATHEMATICS, RUTGERS UNIVERSITY,  
NEW BRUNSWICK, NEW JERSEY 08903

*E-mail address:* jkahn@math.rutgers.edu

*E-mail address:* komlos@math.rutgers.edu

(E. Szemerédi) DEPARTMENT OF COMPUTER SCIENCE, RUTGERS UNIVERSITY, NEW BRUNSWICK,  
NEW JERSEY 08903

*E-mail address:* szemered@cs.rutgers.edu

(J. Komlós and E. Szemerédi) HUNGARIAN ACADEMY OF SCIENCES, BUDAPEST, HUNGARY