© 2009 Society for Industrial and Applied Mathematics

CONCENTRATION OF RANDOM DETERMINANTS AND PERMANENT ESTIMATORS[*]

KEVIN P. COSTELLO[†] AND VAN VU[‡]

Abstract. We show that the absolute value of the determinant of a matrix with random independent (but not necessarily i.i.d.) entries is strongly concentrated around its mean. As an application, we show that Godsil–Gutman and Barvinok estimators for the permanent of a strictly positive matrix give subexponential approximation ratios with high probability. A positive answer to the main conjecture of the paper would lead to polynomial approximation ratios in the above problem.

Key words. random matrix, determinant, singular values

AMS subject classifications. 15A52, 60C05

DOI. 10.1137/080733784

1. Introduction. Let A be an $n \times n$ square matrix. We denote by $\det A$ and $\operatorname{per} A$ its determinant and permanent, respectively, which are defined by

$$\det A = \sum_{\sigma} (-1)^{sgn\sigma} \prod_{i=1}^{n} a_{i\sigma i}, \qquad \operatorname{per} A = \sum_{\sigma} \prod_{i=1}^{n} a_{i\sigma i},$$

where the sum is taken over all permutations in S_n and a_{ij} denotes the (i,j) entry of A.

In this paper, we focus on a random matrix A whose entries are independent (but not necessarily independently and identically distributed (i.i.d.)) random variables with mean zero. The size of A (which we denote by n) should be thought of as tending to infinity and all asymptotic notation will be used under this assumption.

Our main concern is the following basic question.

QUESTION 1.1. *How is $|\det A|$ distributed?* A special case is when the entries of A are i.i.d. Gaussian (with variance one). In this case, it is known that $\log |\det A|$ satisfies the central limit theorem.

THEOREM 1.2. *Let A be the random matrix of size n whose entries are i.i.d. Gaussian with variance one. Then*

$$\frac{\log(|\det A|) - \frac{1}{2}\log((n-1)!)}{\sqrt{\frac{\log n}{2}}}$$

converges weakly to the standard Gaussian variable $N(0,1)$.

This statement is easy to verify, as one can write

$$|\det A| = \prod_{i=1}^{n} d_i,$$

where d_i is the distance from the ith row vector of A to the subspace spanned by the first $i-1$ rows. As A has i.i.d. Gaussian entries, the random variables d_i are independent. Furthermore, their distributions can be computed explicitly, and the theorem follows from Lyapunov's central limit theorem and a routine calculation. (We include the details in the appendix for the reader's convenience.)

The situation with general random matrices is considerably more complicated. In [6], Girko claimed that Theorem 1.2 still holds if the entries are no longer Gaussian, but still i.i.d. with mean zero and variance one. We believe that this statement is true, but could not understand Girko's proof. On the other hand, it seems possible that one can give an alternative proof using recent developments in the field.

In this paper, instead of limiting distribution, we focus on tail inequalities, which are usually very useful in probabilistic combinatorics and related fields. As an illustration, we present an application concerning the problem of computing the permanent using determinant estimators. A consequence of our main result shows that one can use a determinant estimator to estimate the permanent of a matrix of size n with positive entries within a subexponential factor $\exp(n^{2/3}\log n)$ with high probability. If Conjecture 1.5 holds, then the approximation will typically be within a polynomial factor $n^{O(1)}$.

To start, we note an old observation of Turán that if the entries of A are i.i.d. with mean 0 and variance 1, then $\mathbf{E}(|\det A|^2) = \mathbf{E}(\det A^2) = n!$. Combining this with Theorem 1.2, we obtain the following corollary.

COROLLARY 1.3. *Let A be the random matrix of size n whose entries are i.i.d. Gaussian with variance one. Then with probability tending to one,*

$$(1) \qquad \det A^2 = n^{-1+o(1)}\mathbf{E}(\det A^2).$$

Remark 1.4. Here and elsewhere the relation $a_n = O(b_n)$ indicates that the ratio a_n/b_n is bounded above in absolute value as n tends to infinity, and the relation $a_n = o(b_n)$ indicates that it is tending to 0. In particular, the corollary above gives both an upper bound and a lower bound on the ratio between the squared determinant and its expectation.

We believe that a similar result holds for all random matrices having independent entries with mean zero and bounded variances.

CONJECTURE 1.5. *Let $c \le C$ be positive constants. Let A be the random matrix of size n whose entries are independent random variables with mean zero and variances between c and C. Then with probability tending to one,*

$$(2) \qquad |\det A| = n^{O(1)}\mathbf{E}|\det A|, \;\; \det A^2 = n^{O(1)}\mathbf{E}(\det A^2).$$

This conjecture looks highly nontrivial. As a first step, we consider the case when the entries of A are scaled Bernoulli random variables (namely, the ij entry takes values $\pm c_{ij}$ each with probability one half). Our experience is that this is usually the hardest case and its understanding would lead to the solution of the general case. Our main result is the following theorem.

THEOREM 1.6. *Let $0 < c < C$ and $B > 0$ be fixed. Let A be a random $n \times n$ matrix matrix whose entries a_{ij} take values $\pm c_{ij}$ with probability $1/2$, independently, where $c \le |c_{ij}| \le C$. Then with probability $1 - n^{-B}$,*

$$|\det A| = \exp(O(n^{2/3}\log n))\mathbf{E}(|\det A|),$$

and

$$\det(A^2) = \exp(O(n^{2/3}\log n))\mathbf{E}(\det A^2).$$

Here the hidden constants in the O notation may depend on c, C, and B.

In the case $c = C = 1$ (i.e., the entries of A are i.i.d. Bernoulli), a better bound $\exp(O(\sqrt{n \log n}))$ was recently proved in [17]. The approach in [17], however, does not extend to random matrices with entries having different variances. In the present approach, it seems to require some new ideas in order to significantly reduce the constant $2/3$.

If one assumes that the entries of A are Gaussian (with different variances c_{ij}^2), then a weaker bound ($\exp(\epsilon n)$ for any positive ϵ) was proved by Friedland, Rider, and Zeitouni [5]. Our Theorem 1.6 also holds for this case, with the same proof (see section 8). Thus we obtain an improvement for the main result of [5].

2. Computing permanents. Let us now consider $\det M$ and $\mathrm{per} M$ from the computational point of view. It is not hard to compute $\det M$. In fact, there are effective algorithms to compute the whole spectra of M. The problem of computing $\mathrm{per} M$, on the other hand, is notoriously hard, and has been a challenge in theoretical computer science for many years.

A well-known observation that relates the problem of computing the permanent to that of determinant is the following. Let u_{ij} be independent random variables with mean zero and variance one. Given a matrix M with entries a_{ij}, define a random matrix A with entries $\sqrt{a_{ij}} u_{ij}$. Then, using linearity of expectation, it is easy to verify that

$$(3) \qquad\qquad \mathbf{E}(\det A^2) = \mathrm{per}(M).$$

If $\det A^2$ is strongly concentrated around its mean, then (3) leads to the following very simple algorithm: Given M, create a random sample of A. Compute $\det A^2$ and output it as an estimator for $\mathrm{per} A$. The core of the analysis is then to bound the degree of concentration of $\det A^2$ around its expectation.

We mention here that in the case when M has nonnegative entries, the famous work of Jerrum and Sinclair [10] and Jerrum, Sinclair, and Vigoda [11] gives a fully polynomial randomized approximation scheme for the problem using the Markov-chain Monte Carlo approach. Theoretically, this result is as good as it gets. On the other hand, the determinant estimator approach is still of interest, thanks to its simplicity and implementability. (The Markov chain algorithm requires running time $\Theta(n^7)$.)

In [7], Godsil and Gutman proposed setting u_{ij} to be i.i.d. Bernoulli random variables. Following the literature, we call this algorithm the Godsil–Gutman estimator. This is perhaps the simplest estimator. On the other hand, its analysis seems nontrivial. To illustrate this, let us consider the case when M is the all-one matrix. Clearly $\mathrm{per} M = n!$. On the other hand, it is already not easy to prove that with high probability $\det A \neq 0$ (this was first done by Komlós [12]). Effective bounds on $|\det A|$ have only recently become known (see [17]).

If one forces u_{ij} to have a continuous distribution, the situation is more favorable. For instance, it is trivial that $\det A \neq 0$ with probability one. By setting u_{ij} to be i.i.d. Gaussian variables, Barvinok [4] showed that one can approximate the permanent of a nonnegative matrix within a factor of c^n, for some constant $0 < c < 1$. A problem with using Gaussian (or continuous) distribution is that in practice the implementation involves a truncated version of each variable. If the goal function (which is a function of many random variables) has a small Lipschitz coefficient, then this routine is effective. However, if its Lipschitz coefficient is large, then one needs to use a very

fine approximation, and this increases the complexity of the input and would raise some challenges in implementation.

It is known that if one allows the matrix to have zero entries, then determinant estimators do not necessarily give a good approximation to the permanent. For example, Barvinok gave an example where the permanent is 2^n but the Godsil–Gutman estimator almost always returns 0, and another where his own estimator will almost surely perform no better than an $\exp(O(n))$ approximation. On the other hand, Friedland, Rider, and Zeitouni [5] showed that if the entries are strictly bounded from above and below by positive constants, then Barvinok's estimator gives an approximation factor $\exp(\varepsilon n)$, for any fixed $\varepsilon > 0$.

As a consequence of Theorems 1.6 and 4.1, we obtain the following improvement.

THEOREM 2.1. *Let A be a (deterministic) square matrix of size n with entries between c and C, where c and C are positive constants. Then both the Godsil–Gutman and Barvinok estimators approximate $\mathrm{per} A$ within a factor of $O(\exp(n^{2/3}\log n))$ with probability tending to one.*

If Conjecture 1.5 holds, then one can improve the approximation factor to $n^{O(1)}$.

It remains a tantalizing problem to analyze the determinant estimator for the case when the entries of A are not nonnegative real numbers. Notice that (3) still holds in this case, but no effective algorithm is known.

3. The main ideas. We start with the well-known identity

$$(4) \qquad \det A^2 = \det(AA^T) = \prod_{i=1}^{n} \sigma_i^2,$$

where $0 \le \sigma_1 \le \sigma_2 \le \cdots \le \sigma_n$ are the singular values of A.

If one could show that each singular value σ_i is very strongly concentrated around some nonzero value, then $\det A^2$ would be so as well. Unfortunately, such a result is not available. In [2], it was shown, via Talagrand's inequality, that the largest singular values are strongly concentrated, but the degree of concentration decreases rather quickly as the index decreases.

To overcome this obstacle, we will follow the approach in [5], which is based on the fact that, roughly speaking, the counting measure generated by the σ_i is strongly concentrated. This fact was proved by Guionnet and Zeitouni in an earlier paper [8], also using Talagrand's inequality. Guionnet and Zeitouni's result asserts that (after a proper normalization by a factor $1/\sqrt{n}$) any fixed interval, with high probability, contains the right number of singular values. This enables one to show that the product of most of the singular values is close to the expectation.

The main technical barrier of this approach arises at the end of the spectrum. The Guionnet–Zeitouni result does not reveal any information about the few smallest singular values. In [5], the authors needed to exploit the Gaussian assumption (following an approach of Bai [3]) in order to take care of these singular values. This technique, however, is not applicable for discrete distributions such as the Bernoulli distribution. In particular, it does not even show that a random matrix with discrete entries is nonsingular with high probability.

The proof of Theorem 1.6 requires two new ingredients. The first is a lower bound on the smallest singular value σ_n. In [15], it was shown that for many models of random matrices σ_n is at least n^{-C}, for some constant C. While the models in [15] do not include the type of random matrices we consider here, we are able to modify the proof, without too many difficulties, to treat our case.

To continue, naturally one would try to use the uniform bound n^{-C} for all singular values which have not been treated by the concentration result. These will be singular values which are less by some threshold $\varepsilon(n)$. It is now critical to estimate the number of such singular values. The value of $\varepsilon(n)$ will be too small for the concentration result of Guionnet and Zeitouni to give information about this number. The second main ingredient of our proof is a method that provides a good bound. This is based on a simple, but useful, identity (discovered in [19]) which gives a relation between the singular values σ_i and the distances d_i.

4. A more general theorem and the main lemmas. We will actually prove the following more general case of Theorem 1.6, where we merely require the entries to be bounded and have bounded variance instead of being Bernoulli.

THEOREM 4.1. *Let $K > 0, B > 0$, and $0 < c < 1$ be fixed. Let A be a random $n \times n$ matrix whose entries a_{ij} are random variables satisfying*
- $c \leq \mathbf{Var}(a_{ij}) \leq \frac{1}{c}$,
- $\mathbf{P}(|a_{ij} - E(a_{ij})| \leq K) = 1$.

Then with probability $1 - n^{-B}$,

$$|\det A| = \exp(O(n^{2/3} \log n))\mathbf{E}(|\det A|)$$

and

$$\det A^2 = \exp(O(n^{2/3} \log n))\mathbf{E}(\det A^2),$$

where the constant implicit in the O notation depends on K, B, and c.

Remark 4.2. The uniform boundedness condition can be replaced by the condition that all of the entries have a Gaussian distribution; see section 8.

Recall (4),

$$(\det A)^2 = \det(AA^T) = \prod_{\sigma \in \mathrm{spec}AA^T} \sigma = \prod_{i=1}^{n} \sigma_i^2,$$

where $0 \leq \sigma_1 \leq \sigma_2 \leq \cdots \leq \sigma_n$ are the singular values of A.

We will start in a similar way as in [5]. Let ϵ be a parameter to be determined later (which may depend on n). We estimate (4) by dividing the spectrum into two parts, writing $|\det A| = (\det_{trunc} A)(\det_{small} A)$, where

$$\det_{trunc} = \left(\prod_{\sigma \in \mathrm{spec}(AA^T)} \max\{\sigma, \epsilon^2\} \right)^{1/2},$$

$$\det_{small} = \left(\prod_{\sigma \in \mathrm{spec}(AA^T)} \min\{\sigma\epsilon^{-2}, 1\} \right)^{1/2}.$$

We first show that $\det_{trunc} A$ and $\det_{trunc} A^2$ are strongly concentrated around their means.

LEMMA 4.3. *There is a constant $c_0 > 0$ dependent only on c (the bound on the variance of the entries of A) such that*

$$\det_{trunc} A = \exp(O(n\epsilon^{-2} \log n))\mathbf{E}(\det_{trunc} A)$$

and

$$\det_{trunc}A^2 = \exp(O(n\epsilon^{-2}\log n))\mathbf{E}(\det_{trunc}A^2)$$

with probability $1 - O(n^{-c_0 \log n})$.

The proof of this lemma is presented in section 3.

To handle $\det_{small}A$, notice that

(5) $$1 \geq \det_{small}A \geq \min\{1, (\sigma_n(A)\epsilon^{-1})^{s_\epsilon(A)}\},$$

where $\sigma_n(A)$ is the smallest singular value of A and $s_\epsilon(A)$ denotes the number of singular values of A which are at most ϵ. We can therefore bound $\det_{small}A$ from below by using the following two lemmas.

LEMMA 4.4. *For any $B > 0$,*

$$\mathbf{P}(\sigma_n(A) < n^{-4B-7}) \leq n^{-B}.$$

We remark that $-4B - 7$ is pretty far from being optimal and can be improved, but doing so would not affect our final results in any essential way. This lemma is a variant of many results proved in [18] (see also [16]). However, [18] required that the distributions of the entries of A to be dominated in a certain Fourier analytic sense by a single common distribution. Our matrices do not satisfy this assumption. However, we are able to modify the proof, without too many difficulties, to obtain the desired result.

LEMMA 4.5. *Let $r \geq \log^4 n$, and assume $c^2 \leq \mathbf{Var}(a_{ij}) \leq \frac{1}{c^2}$. Then*

$$\mathbf{P}\left(\sigma_{2r}(A) \leq \frac{rc}{2\sqrt{n-r}}\right) = o(n^{-\log n}).$$

The above two lemmas combine to show that no singular value of A is likely to be so small as to have too large an effect on the determinant, and that not many singular values will have to be handled by $\det_{small}$.

Let us for now assume the previous two lemmas to be true. By taking $r = \frac{3\epsilon\sqrt{n}}{c^2}$ in Lemma 4.5, we see that with high probability $s_\epsilon(A) = O(n^{1/2}\epsilon)$. Combining this with Lemma 4.5 and the bounds in (5), we see that for any $B > 0$ we have with probability $1 - n^{-B+o(1)}$ that

$$\det_{small}A = \exp(O(n^{1/2}\epsilon\log n)),$$

which therefore implies that with the same probability

(6) $$\det_{trunc}A \geq |\det A| \geq \exp(O(n^{1/2}\epsilon\log n))\det_{trunc}A.$$

(Note again that the use of O in the lower bound here indicates an exponent bounded in magnitude.) Now let us fix $\epsilon = n^{1/6}$. Combining the second half of the above inequality with Lemma 4.3, we see that with probability $1 - n^{-B+o(1)}$ we have

$$|\det A| \geq \exp(O(n^{2/3}\log n))\mathbf{E}(\det_{trunc}A).$$

Taking expectations, we find that

(7) $$\mathbf{E}(\det_{trunc}A) \geq \mathbf{E}(|\det A|) \geq (1+o(1))\exp(O(n^{2/3}\log n))\mathbf{E}(\det_{trunc}A).$$

The first half of Theorem 4.1 follows from combining (6), (7), and Lemma 4.3. The second half follows from the identical argument being applied to $\det A^2$.

5. Proof of Lemma 4.3. As in [5], we begin with the spectral concentration results of Guionnet and Zeitouni, in particular the following special case of Corollary 1.8(a) in [8].

THEOREM 5.1. *Let Y be an $n \times n$ matrix whose entries are independent random variables each having support on a compact set of diameter at most K, and let $Z = Y^T Y$. Let $\lambda_1, \ldots, \lambda_n$ be the eigenvalues of Z, and let f be an increasing, convex function such that $g(x) = f(x^2)$ has Lipschitz norm $|g|_L$. Then for any $\delta > \delta_0 := \frac{2K\sqrt{\pi}|g|_L}{n}$,*

$$\mathbf{P}\left(\left|\sum_{i=1}^{n} f(\lambda_i) - \mathbf{E}\left(\sum_{i=1}^{n} f(\lambda_i)\right)\right| > 2\delta n\right) \le 4\exp\left(-\frac{(\delta - \delta_0)^2 n}{K^2 |g|_L^2}\right).$$

Ideally, we would like to apply this theorem with f taken to be the logarithm, so that $\sum f(\lambda_i) = \log \det A^2$. The difficulty is that the logarithm is not Lipschitz. To overcome this problem, we follow [5] and truncate the logarithm. Write

$$\log^\epsilon x = \max\{2\log\epsilon, \log x\},$$

where $\log^\epsilon(0)$ is defined to be $2\log\epsilon$. Note that we have

$$(8) \qquad \log(\det_{trunc} A) = \frac{1}{2} \sum_{\sigma \in \text{spec}(AA^T)} \log^\epsilon(\sigma).$$

Although $\log^\epsilon(x^2)$ now has finite Lipschitz constant $\frac{1}{\epsilon}$ (this was the purpose of truncating the logarithm), it is not convex. However, it can easily be written as the difference of two convex Lipschitz functions, so the above theorem applies, and we have for some absolute constants C_0 and C_1 and any $\delta \ge \delta_0 := C_0 \epsilon^{-1}/n$ that

$$(9) \qquad \mathbf{P}\left(|\log(\det_{trunc} A) - \mathbf{E}(\log(\det_{trunc} A))| > \delta n\right) \le 4\exp\left(-C_1 \frac{\epsilon^2 n \delta^2 c^2}{16}\right).$$

Taking $\delta = \frac{\log n}{\epsilon \sqrt{n}}$, we see that for some constant c_0 we have

$$(10) \qquad \mathbf{P}\left(|\log(\det_{trunc} A) - \mathbf{E}(\log(\det_{trunc} A))| > \frac{\sqrt{n}\log n}{\epsilon}\right) = O(n^{-c_0 \log n}).$$

This would be exactly the result we wanted, if only the expectation and the logarithm were switched on the left-hand side of (9). Following [5], we now write

$$U(A) = \log(\det_{trunc} A) - \mathbf{E}(\log(\det_{trunc} A)).$$

We know $\mathbf{E}(U) = 0$, and by Jensen's inequality we have

$$1 \le \mathbf{E}(e^U) \le \mathbf{E}(e^{|U|}) \le 1 + \int_0^\infty e^t \mathbf{P}(|U| > t)dt.$$

It follows from the above equation and (9) that

$$1 \le \mathbf{E}(e^U) = \frac{\mathbf{E}(\det_{trunc} A)}{\exp\left(\mathbf{E}(\log(\det_{trunc} A))\right)} \le exp\left(O\left(\frac{n}{\epsilon^2}\right)\right),$$

and the first half of Lemma 4.3 follows by taking logarithms and combining with (10). The second half follows from the identical calculation applied to e^{2U}.

Remark 5.2. If we had only required that the truncated determinant concentrate *somewhere*, the argument above would have given a stronger bound (roughly $\exp(\epsilon^{-1}n^{1/2}\log n)$). The dominant term in our bound came from showing that the "somewhere" was close to the actual expectation.

Also, we did not at any point use our lower bound on the variance of the entries. In particular, this truncated determinant will be concentrated around its expectation even if we allow most of the entries of A to be nonrandom. However, it is not true in general that $\det_{trunc}A$ will be close to $\det A$.

6. Proof of Lemma 4.4. We begin by first reducing from the general case back to the case of Bernoulli matrices. To do so, we will use the idea of Bernoulli decomposition from a paper of Aizenman et. al. [1]. In that paper, it is shown that for any random variable X that is nondegenerate (not taking on any single value with probability 1), we can find a $p \in (0,1)$ and functions $f(t)$ and $g(t)$ such that we have the following:

- If t is uniform on $[0,1]$, and ϵ is a Bernoulli variable independently equal to 1 (with probability p) or 0 (with probability $1-p$), then $f(t) + g(t)\epsilon$ has the same distribution as X.
- $\inf g(t) > 0$.

Recall that we are assuming that our entries are both uniformly bounded in magnitude by K and bounded below in variance by c. It follows from the methods of [1] (see Remark 2.1(i) there) that in this case we can find a Bernoulli decomposition $a_{ij} = f_{ij}(t_{ij}) + g(t_{ij})\epsilon_{ij}$ of every entry of A in which the $g(t_{ij})$ have a uniform lower bound $\beta = \beta(K,c)$ for all values of i,j, and t, and for which the p_{ij} in the decompositions are uniformly bounded away from 0 and 1.

We now view our matrix as being formed in two steps. First, we expose t_{ij} for each entry. At this point every entry can be viewed as having a shifted Bernoulli distribution. Next we expose the ϵ_{ij}. It follows by taking expectations over all possible values of t_{ij} that it suffices to show the following lemma.

LEMMA 6.1. *Let $0 < q < \frac{1}{2}$ and $B,C,c > 0$ be fixed. Let A be a matrix whose entries are independent random variables distributed as $a_{ij} = m_{ij} + \epsilon_{ij}n_{ij}$, where $|m_{ij}| < n^{1/8}$ and $c < n_{ij} < C$, and furthermore the ϵ_{ij} satisfy*

$$q < \mathbf{P}(\epsilon_{ij} = 1) = 1 - \mathbf{P}(\epsilon_{ij} = -1) < 1 - q.$$

Then for sufficiently large n we have

$$\mathbf{P}(\sigma_n(A) < n^{-4B-7}) \leq n^{-B}.$$

Remark 6.2. The form of this theorem is very similar to that of the smoothed analysis of the smallest singular value in [18]. The key difference here is that we no longer require the n_{ij} to be identical.

Proving Lemma 6.1 is equivalent to bounding the probability that for some unit vector v we have $||Av|| \leq n^{-4B-7}$. We will do this by dividing the vectors into two classes, which should be thought of as "structured" and "unstructured," for an appropriate definition of "structured" depending both on A and on B.

DEFINITION 6.3. *A vector v is* rich *if there is some i for which*

$$\sup_{z} \mathbf{P}\left(\left|\sum_{j=1}^{n} a_{ij}v_j - z\right| < n^{-4B-13/2}\right) \geq n^{-B-1}.$$

Otherwise v is poor.

Equivalently, a poor vector is one for which no individual coordinate of Av is too concentrated. Lemma 4.4 would be an immediate consequence of the following two lemmas.

LEMMA 6.4.

$$\mathbf{P}(||Av|| \leq n^{-4B-7} \text{ for some poor } v) \leq \frac{1}{2}n^{-B}.$$

LEMMA 6.5.

$$\mathbf{P}(||Av|| \leq n^{-4B-7} \text{ for some rich } v) \leq \frac{1}{2}n^{-B}.$$

6.1. Proof of Lemma 6.4. We adapt an argument from [13] (see also [18]). Let E be the event that for some poor unit vector v we have $||Av|| \leq n^{-4B-7}$. If E holds, then the least singular value of A is at most n^{-4B-7}, so the same must hold for A^T. For $1 \leq j \leq n$, let F_j be the event that there exists a unit vector $w = (w_1, \ldots, w_n)^T$, which simultaneously satisfies

$$||w^T A|| \leq n^{-4B-7}, \quad |w_j| \geq \frac{1}{\sqrt{n}}.$$

Since every w has at least one coordinate at least $n^{-1/2}$ in magnitude, we have

$$\mathbf{P}(E) \leq \sum_{i=1}^{n} \mathbf{P}(E \wedge F_j).$$

Now let j be fixed. Let $A_1, \ldots, A_n$ be the rows of A. We will condition on all of the rows except row j. If E is to hold, there must be a poor v such that

$$\left(\sum_{i=1}^{n} |A_i \cdot v|^2 \right)^{1/2} = ||Av|| \leq n^{-4B-7}.$$

It follows that if $\mathbf{P}(E|A_1, \ldots, A_{j-1}, A_{j+1}, \ldots, A_n)$ is nonzero, then there is a poor u such that

(11) $$\left(\sum_{i \neq j} |A_i \cdot u|^2 \right)^{1/2} \leq n^{-4B-7}.$$

Conversely, by our assumptions on w we have that if F_j holds, then

$$\left\| \sum_{i \neq j} w_i A_i \right\| \leq n^{-4B-7}.$$

Taking inner products with u and using the triangle inequality, we conclude

$$|w_j||A_j \cdot u| \leq \sum_{i \neq j} |w_i||A_i \cdot u| + n^{-4B-7}.$$

Combining the above with (11), the Cauchy–Schwarz inequality, and our assumption on $|w_j|$, we obtain that if both E and F_j hold, then

$$|A_j \cdot u| \leq 2n^{-4B+13/2}.$$

On the other hand, since u is poor and A_j and u are independent, we have that

$$\mathbf{P}(|A_j \cdot u| \leq 2n^{-4B+13/2}|A_1, \ldots, A_{j-1}, A_{j+1}, \ldots, A_n) \leq n^{-B-1}.$$

Combining the above, we see that

$$\mathbf{P}(E \wedge F_j|A_1, \ldots, A_{j-1}, A_{j+1}, \ldots, A_n) \leq n^{-B-1},$$

regardless of our choice of the remaining $n-1$ rows. It follows that $\mathbf{P}(E \wedge F_j) \leq n^{-B-1}$ for every j, and therefore $\mathbf{P}(E) \leq n^{-B}$.

6.2. Proof of Lemma 6.5. Let J be the unique integer satisfying $2B + 2 < J \leq 2B + 3$, and let $\delta = (B+1)/J$. Let $\gamma > 0$ be a constant chosen to be sufficiently small that $\delta + 3\gamma < 1/2$ and $(4B+7)\gamma < \frac{1}{2}$. Finally, we let $D = 2 + 2\gamma$.

Let v be a rich unit vector. We define

$$g(j) := \sup_{i,z} \mathbf{P}(|A_i^T v - z| < n^{-4B-13/2+Dj}).$$

Clearly $0 \leq g(j) \leq 1$, and $g(j)$ is an increasing function in j. The assumption that v is rich is equivalent to the statement that $g(0) \geq n^{-B-1}$. It follows from the pigeonhole principle that for some $0 < j \leq J - 1$ we have

$$g(j+1) \leq n^\delta g(j).$$

For $0 \leq j \leq J - 1$ and $1 \leq k \leq \lceil \frac{(A+1)}{\gamma} \rceil$, we define $\Omega_{j,k}$ to be the collection of rich v satisfying both

$$g(j+1) \leq n^\delta g(j) \quad \text{and} \quad g(j) \in [n^{-k\gamma}, n^{-(k-1)\gamma}].$$

Since every rich v is contained in some $\Omega_{j,k}$, and there are only a bounded number of pairs (j,k), it suffices to prove that for every fixed j and k we have

$$(12) \qquad \mathbf{P}(\|Av\| \leq n^{-4B-7} \text{ for some } v \in \Omega_{j,k}) = o(n^{-B}).$$

Our goal will now be to construct a β-net for each $\Omega_{j,k}$; that is, a set V_0 such that any point in $\Omega_{j,k}$ is within (Euclidean) distance β of some point in V_0. Assuming that for sufficiently small β the net is not too large, we will then be able to obtain Lemma 6.5 by a union bound. We begin bounding the size of the net with the following result, a special case of [16, Thm. 3.2, see also Remark 2.8].

THEOREM 6.6. *Let $0 < q < \frac{1}{2}$, and let $x_1, \ldots, x_n$ be independent random variables taking on values in $\{1, -1\}$ and satisfying*

$$q \leq \mathbf{P}(x_j = 1) \leq 1 - q.$$

Let $0 < \delta < 1$ be fixed, and let p and β be chosen to satisfy $p = n^{-O(1)}$ and $\beta > \exp(-n^{-\delta/2})$. Then the set of vectors $(v_1, \ldots, v_n)$ satisfying

$$\sup_{z \in \mathbf{C}} \mathbf{P}\left(\left|\sum_{i=1}^n v_i x_i - z\right| < \beta\right) < p$$

has a β-net in the l_∞ norm of size at most $n^{-(1/2+\delta)n}p^{-n} + \exp(o(n))$.

For any particular i, we have

$$(13) \qquad \mathbf{P}\left(\left|\sum_{j=1}^{n}(m_{ij} + n_{ij}\epsilon_{ij})v_j - z\right| < \beta\right) = \mathbf{P}\left(\left|\sum_{j=1}^{n}\epsilon_{ij}\widetilde{v}_j - \widetilde{z}\right| < \beta\right),$$

where $\widetilde{v}_j = v_j n_{ij}$ and $\widetilde{z} = z - \sum_j m_{ij}v_j$. For any particular coordinate of Av, Theorem 6.6 gives an upper bound on the minimal size of a β-net for the set of $\widetilde{v}$ for which the right-hand side of (13) holds with probability at least p. By taking an affine transformation $v_j \rightarrow \frac{v_j}{n_{ij}}$ of the case $\beta = n^{-4B-13/2+Dj}, p = n^{-k\gamma}$ of this net and taking the union of the resulting net for each coordinate, we obtain the following modified version of Theorem 6.6.

LEMMA 6.7. *Let $x_1, \ldots, x_n$ be independent and have the form $x_i = m_i + \epsilon_i n_i$, where the m, n, ϵ are as in Lemma 6.1. Let $0 < \delta < 1$ be fixed. Then $\Omega_{j,k}$ has an $\frac{n^{-4B-13/2+Dj}}{c}$-net in the l_∞ norm of size at most $n^{1-(1/2+\delta)n}n^{k\gamma n} + \exp(o(n))$.*

Let V_0 be a net guaranteed by the above lemma, and consider any $v' \in V_0$ and $v \in \Omega_{j,k}$ such that $||v - v'||_\infty \leq \beta$. Our bounds on the m_{ij} and n_{ij} guarantee that (assuming n to be sufficiently large) the spectral norm of A satisfies $\sigma_n(A) < n$. Since

$$||Av'|| \leq ||Av|| + ||A(v - v')|| \leq ||Av|| + n^{1/2}\sigma_n(A)||v - v'||_\infty,$$

it follows that if $||Av|| \leq n^{-4B-7}$, then

$$||Av'|| \leq \left(1 + \frac{1}{c}\right)n^{-4B-4+Dj}.$$

It then follows that there must be at least $n - n^{1-\gamma}$ rows of A for which

$$(14) \qquad |A_i^T v'| \leq n^{-4B-9/2+Dj+\gamma}.$$

On the other hand, we also have for any i for which (14) holds that

$$|A_i^T v| \leq |X_i^T v'| + ||v - v'||_\infty \sum_{j=1}^{n}(m_{ij} + n_{ij})$$

$$\leq n^{-4B-9/2+Dj+\gamma} + \frac{1}{c}n^{-4B-9/2+Dj}\left(\frac{1}{c} + n^{1/8}\right)$$

$$\leq n^{-4B-7+D(j+1)},$$

where the last inequality comes from our definition of D.

It follows that

$$\mathbf{P}(|A_i^T v'| \leq n^{-4B-9/2+Dj+\gamma}) \leq \mathbf{P}(|A_i^T v| \leq n^{-4B-7+D(j+1)})$$
$$\leq n^\delta g(j)$$
$$\leq n^{\delta+\gamma-k\epsilon},$$

where for the last two inequalities we use the definition of $\Omega_{j,k}$.

This will be sufficient to handle the case where k is sufficiently large. For smaller k, we note that by our choice of γ and D we have $-4B - 9/2 + Dj + \gamma < -1$, so

$$\mathbf{P}(|A_i^T v'| \leq n^{-4B-9/2+Dj+\gamma}) < \mathbf{P}\left(|A_i^T v'| < \frac{1}{n}\right),$$

which can easily be checked to be at most $1 - q$.

Therefore

$$\mathbf{P}(|A_i^T v'| \leq n^{-4B-9/2+Dj+\gamma}) \leq \min(n^{\delta+\gamma-k\epsilon}, 1 - q).$$

Taking the union bound over all sets of $n - n^{1-\gamma}$ rows, we see that

$$\mathbf{P}(\|Av'\| \leq n^{-4B-4+Dj+\gamma}) \leq \min(n^{\delta+\gamma-k\epsilon}, 1 - q)^{n'} \binom{n}{n - n^{1-\gamma}}$$

for any particular v' in our net for $\Omega_{j,k}$. Taking the union bound over the entire net, we obtain that the probability that the left-hand side of (12) holds is at most

$$(n^{-(1/2+\delta)n+1} n^{k\gamma n} + \exp(o(n))) \min(n^{\delta+\gamma-k\epsilon}, 1 - q)^{n - n^{1-\gamma}} \binom{n}{n - n^{1-\gamma}},$$

which can be verified to be exponentially small by a routine calculation.

7. Proof of Lemma 4.5. As in the proof of Lemma 4.4, it suffices by Bernoulli decomposition to prove the following special case of this lemma.

LEMMA 7.1. *Let $0 < q < \frac{1}{2}$ and $B, C, c > 0$ be fixed. Let A be a matrix whose entries are independent random variables each having distribution $a_{ij} = m_{ij} + \epsilon_{ij} n_{ij}$, where $|m_{ij}| < n^{1/8}$ and $c < n_{ij} < C$, and assume furthermore the ϵ_{ij} satisfy*

$$q < \mathbf{P}(\epsilon_{ij} = 1) = 1 - \mathbf{P}(\epsilon_{ij} = -1) < 1 - q.$$

Then for $r \geq \log^4 n$,

$$\mathbf{P}\left(\sigma_{2r}(A) < \frac{rc}{2\sqrt{n-r}}\right) = o(n^{-\log n}).$$

To prove this lemma we are going to use the following lemma from [19].

LEMMA 7.2 (see [19]). *Let M be an $m \times n$ matrix ($m \leq n$). Let d_i be the distance from its ith row vector to the space spanned by the first $i-1$ rows and σ_i be its singular values. Then*

$$\sum_{i=1}^{m} d_i^{-2} = \sum_{i=1}^{m} \sigma_i^{-2}.$$

Recall that $\log^4 n < r < \frac{n}{2}$. By the interlacing inequalities for singular values (see, for example, Theorem 7.3.9 in [9]), we have that

$$\sigma_{2r}(A) \geq \sigma_r(A'),$$

where A' is the matrix formed by removing the last r columns from A.

To bound the right-hand side of this equation, we note that

$$\sigma_r(A')^{-2} \leq \frac{1}{r} \sum_{k=1}^{r} \sigma_k(A')^{-2}$$

$$\leq \frac{1}{r} \sum_{k=1}^{n-r} \sigma_k(A')^{-2}$$

(15)
$$= \frac{1}{r} \sum_{i=1}^{n-r} d_i^{-2},$$

where d_i denotes the distance from the ith column of A' to the span of the remaining columns, and for the last equality we use Lemma 7.2.

Informally, this states that if a matrix has many small singular values, it must have many columns which are very close to the subspace spanned by the other columns. Since r is becoming increasingly large, the codimension of this subspace is increasing as well, so this should become unlikely.

Now let i be fixed. To bound the probability that d_i is small, we first expose the subspace S_i spanned by the remaining $n - r - 1$ columns, then finally the remaining column. Let P denote the projection matrix onto that subspace, and let p_{ij} be the entries of P. Let $X_i = M_i + N_i = (m_{i1}, \ldots, m_{in}) + (\epsilon_{i1}n_{i1}, \ldots, \epsilon_{in}n_{in})$ be this final column. We first note that

$$
\begin{aligned}
\mathbf{E}(d_i^2|S_i) &= \mathbf{E}\left((M_i^T + N_i^T)(I - P)(M_i + N_i)\right) \\
&= \mathbf{E}(M_i^T(I - P)M_i) + \mathbf{E}(N_i^T(I - P)N_i) \\
&\geq \mathbf{E}(N_i^T(I - P)N_i),
\end{aligned}
$$

as the cross terms cancel due to the symmetric distribution of N_i (recall that each ϵ_{ij} is equally likely to be 1 or -1) and the matrix $(I - P)$ is positive semidefinite. We can bound this last expression by

$$
\begin{aligned}
\mathbf{E}(N_i^T(I - P)N_i) &= \sum_{j=1}^{n} n_{ij}^2 - \sum_{k=1}^{n}\sum_{l=1}^{n} \mathbf{E}(\epsilon_{ik}\epsilon_{il}n_{ik}n_{il}p_{kl}) \\
&= \sum_{j=1}^{n} n_{ij}^2(1 - p_{jj}) \\
&\geq \sum_{j=1}^{n} c^2(1 - p_{jj}) \\
&= c^2(n - Tr(P)) = c^2 r,
\end{aligned}
$$

as the terms with $k \neq l$ cancel by the independence of the ϵ_{ij}, and we again use how the variances of the entries of A are bounded away from 0.

Remark 7.3. If the entries of A were to have mean 0 and equal variance c, then the inequalities here would actually be equalities, and the expected square distance would be independent of S. This is what enables the arguments of [6, 17] (which are based on row-by-row exposure of the matrix in question), and why those arguments don't carry over to give an immediate estimate on the determinant of A.

In other words, a random vector from our distribution will on average be far away from any fixed $n - r - 1$ dimensional subspace. It remains to show that it will typically be far away.

It follows from Talagrand's inequality [14] and our bounds on the n_{ij} that if $M_{i,S}$ is the median value of d_i conditioned on S_i, then

$$
\mathbf{P}(|d_i - M_{i,S}| \geq t|S_i) \leq 4\exp\left(-\frac{t^2}{64C^2}\right).
$$

By an argument identical to that in [17], it can be shown that $|M_i - \sqrt{E(d_i^2)}| \leq \frac{C^2}{c^2}$. It therefore follows that for sufficiently large r,

$$
\mathbf{P}\left(d_i \leq \frac{c\sqrt{r}}{2}|S_i\right) \leq 4\exp\left(-\frac{rc^2}{300C^2}\right) = o(n^{-\log n - 1}),
$$

as by assumption $r > \log(n)^3$.

In particular, with probability $1 - o(n^{-\log n})$ every d_i will be at least $c\sqrt{r}/2$. Combining this with (15), we see

$$\mathbf{P}\left(\sigma_{2r}(A) \leq \frac{rc}{2\sqrt{n-r}}\right) \leq \mathbf{P}\left(\sum_{i=1}^{r} \sigma_r(A')^{-2} \geq \frac{4(n-r)}{c^2 r}\right)$$
$$\leq \mathbf{P}\left(\sum_{i=1}^{n-r} d_i^{-2} \geq \frac{4(n-r)}{c^2 r}\right)$$
$$= o(n^{-\log n}).$$

8. Concentration of determinants for Gaussian variables. Although we have focused mainly on the concentration of determinants for the case of matrices with uniformly bounded entries, our main results also hold in the case where every entry has a Gaussian distribution, assuming that the means of the entries are uniformly bounded and the variances of the entries are uniformly bounded above and below. In particular, this implies that our bounds hold for Barvinok's as well as Godsil and Gutman's estimator for the permanent.

The proof of Lemma 4.3 is exactly the same as before, except that we use Corollary 1.8b of [8] instead of Corollary 1.8a. For the remaining two lemmas, we again use the idea of Bernoulli decomposition. It can be explicitly checked that if X is a Gaussian variable satisfying $|\mathbf{E}(X)| < K$, $c < \mathbf{Var}(X) < C$, then X can be decomposed as

$$(16) \qquad\qquad X = f(t) + g(t)\epsilon,$$

where t is uniform on $[0,1]$ and ϵ is uniform on $\{-1,1\}$. Furthermore, we can do so in such a way that $g(t)$ is bounded uniformly from below, and the measure of the set of t for which $|f(t)| + |g(t)| > \log^3 n$ is $o(n^{-\log n})$.

We now expose t_{ij} for each entry of A. At this point every entry will have a Bernoulli distribution and (except for an exceptional set of probability $o(n^{-B})$ for any B) the mean and variance of the entries will be bounded by $\log^2 n$. Lemmas 4.4 and 4.5 now follow as before from Lemmas 6.1 and 7.1.

Appendix. Log-normality of the determinant of Gaussian matrices. In this appendix we will show that the determinant of a matrix whose entries are i.i.d. standard Gaussian random variables has the distribution given by Theorem 1.2. Our starting point is the formula

$$(17) \qquad\qquad |\det A| = \prod_{i=1}^{n} d_i,$$

where d_i is the distance from the ith row of A to the subspace spanned by the previous $i-1$ rows.

This formula is particularly useful for Gaussian vectors due to their rotational invariance: If x is a random vector whose coordinates are i.i.d. Gaussian variables having mean zero, then the distribution of the distance from x to a fixed subspace S is dependent only on the dimension of S. If the dimension is $n - k$, then the distribution of the square of the distance follows a chi-square distribution with k degrees of freedom. In particular, this implies that the distribution of the determinant is the same as that where we treat each of the variables in (17) as being independent and following a chi distribution. We will do so for the remainder of this appendix.

Taking logarithms in (17) and rearranging, we see that

$$2\log(|\det A|) - \log n! = \sum_{i=1}^{n} \log\left(\frac{d_i^2}{i}\right) = \sum_{i=1}^{n} \log\left(1 + \frac{d_i^2 - i}{i}\right).$$

Following the ideas of [6], we next perform a Taylor expansion on the right-hand side, writing
(18)

$$\frac{2\log(|\det A|) - \log n!}{\sqrt{2\log n}} = \frac{\sum_{i=1}^{n} \frac{d_i^2 - i}{i}}{\sqrt{2\log n}} - \frac{1}{2}\frac{\sum_{i=1}^{n}(\frac{d_i^2 - i}{i})^2}{\sqrt{2\log n}} + \frac{1}{3}\frac{\sum_{i=1}^{n}(\frac{d_i^2 - i}{i})^3}{\sqrt{2\log n}} + \frac{\sum_{i=1}^{n}\epsilon_i}{\sqrt{2\log n}}.$$

We examine the terms in order.

It follows from standard facts about the chi-square distribution that $\frac{d_i^2 - i}{i}$ has mean 0, variance $\frac{2}{i}$, and fourth moment $\frac{12i+48}{i^3}$. It follows immediately from Lyapunov's central limit theorem that the first term on the right-hand side of (18) converges weakly to $N(0,1)$.

For the second term, we observe $(\frac{d_i^2 - i}{i})^2$ has mean $\frac{2}{i}$ and variance $\frac{12i+48}{i^3}$. In particular, the variance of $\sum(\frac{d_i^2 - i}{i})^2$ is $o(\log n)$. It follows that the second term converges to $\frac{-\log n}{\sqrt{2\log n}}$. Similarly, it follows from the moments of the chi-square distribution that the expectation and the variance of $\sum(\frac{d_i^2 - i}{i})^3$ are $O(1) = o(\log(n))$, so the third term converges weakly to zero.

The final term is slightly more complicated due to the singularity of the logarithm at 0. We first split the error term as $\epsilon_i = \epsilon_i' + \epsilon_i''$, where ϵ_i' is zero whenever $d_i^2 < \frac{i}{3}$, and ϵ_i'' is zero whenever $d_i^2 \geq \frac{i}{3}$. We will show separately that each of the contributions of each of these errors converge weakly to zero.

For the first error term (the case where d_i is large), we note that $|\epsilon_i'| = O(|d_i^2 - i|^4)$ and therefore (using the fourth moment given above) $\mathbf{E}(|\epsilon_i'|) = O(i^{-3})$. It follows that $\frac{\sum \epsilon_i'}{\sqrt{2\log n}}$ converges to zero, so the first part of our decomposition is negligible.

It can be checked by direct computation that $\mathbf{E}(|\log(d_i^2)|)$ is finite for any i. The same therefore also holds for $\mathbf{E}(|\epsilon_i''|)$, so it follows that for some function $s = s(n)$ diverging to infinity sufficiently slowly we have

(19)
$$\frac{\sum_{i=1}^{s} \epsilon_i''}{\sqrt{2\log n}} \overset{\text{weak}}{\to} 0.$$

From the fourth moment given above we know that $\mathbf{P}(d_i^2 < \frac{i}{3}) = O(\frac{1}{i^2})$. Since s diverges to infinity, it follows immediately that $\sum_{i=s}^{n} \epsilon_i''$ is almost surely zero. Combining this with (19), we see that the ϵ'' portion of our truncation error is also negligible.

The theorem follows from our bounds on each term in the Taylor expansion.

REFERENCES

[1] M. AIZENMAN, F. GERMINET, A, KLEIN, AND S. WARZEL, *On Bernoulli decompositions for random variables, concentration bounds, and spectral localization*, Probab. Theory Related Fields, 143 (2009), pp. 219–238.

[2] N. ALON, M. KRIVELEVICH, AND V. VU, *On the concentration of eigenvalues of random symmetric matrices*, Israel J. Math., 131 (2002), pp. 259–267.
[3] Z. D. BAI, *Circular law*, Ann. Probab., 25 (1997), pp. 494–529.
[4] A. BARVINOK, *Polynomial time algorithms to approximate permanents and mixed discriminants within a simply exponential factor*, Random Structures Algorithms, 14 (1999), pp. 29–61.
[5] S. FRIEDLAND, B. RIDER, AND O. ZEITOUNI, *Concentration of permanent estimators for certain large matrices*, Ann. Appl. Probab., 14 (2004), pp. 1559–1576.
[6] V. L. GIRKO, *A refinement of the central limit theorem for random determinants*, Teor. Veroyatnost. i Primenen, 42 (1997), pp. 63–73; translation in Theory Probab. Appl., 42 (1998), pp. 121–129.
[7] C. GODSIL AND I. GUTMAN, *On the matching polynomial of a graph*, in Algebraic Methods in Graph Theory, Vol. I, II (Szeged, 1978), North–Holland, Amsterdam–New York, 1981, pp. 241–249.
[8] A. GUIONNET AND O. ZEITOUNI, *Concentration of the spectral measure for large matrices*, Electron. Comm. Probab., 5 (2000), pp. 119–136.
[9] R. HORN AND C. JOHNSON, *Matrix Analysis*, Cambridge University Press, Cambridge, UK, 1985.
[10] M. JERRUM AND A. SINCLAIR, *Approximating the permanent*, SIAM J. Comput., 18 (1989), pp. 1149–1178.
[11] M. JERRUM, A. SINCLAIR, AND E. VIGODA, *A polynomial-time approximation algorithm for the permanent of a matrix with nonnegative entries*, J. ACM, 51 (2004), pp. 671–697.
[12] J. KOMLÓS, *On the determinant of* $(0, 1)$ *matrices*, Studia Sci. Math. Hungar., 3 (1967), pp. 7–22.
[13] M. RUDELSON, *Invertibility of random matrices: Norm of the inverse*, Ann. of Math. (2), 168 (2008), pp. 575–600.
[14] M. TALAGRAND, *A new look at independence*, Ann. Probab., 1 (1996), pp. 1–34.
[15] T. TAO AND V. H. VU, *Inverse Littlewood-Offord theorems and the condition number of random discrete matrices*, Ann. of Math. (2), 169 (2009), pp. 595–632.
[16] T. TAO AND V. VU, *Random matrices: The circular law*, Commun. Contemp. Math., 10 (2008), pp. 261–307.
[17] T. TAO AND V. VU, *On random* ± 1 *matrices: Singularity and determinant*, Random Structures Algorithms, 28 (2006), pp. 1–23.
[18] T. TAO AND V. VU, *On the condition number of a randomly perturbed matrix*, in Proceedings of the Thirty-Ninth Annual ACM Symposium on Theory of Computing, San Diego, CA, ACM, New York, 2007, pp. 248–255.
[19] T. TAO AND V. VU, *Random matrices: Universality of ESD and the circular law*, preprint available online at arXiv:0807.4898.