

RANDOM SYMMETRIC MATRICES ARE ALMOST SURELY NONSINGULAR

KEVIN P. COSTELLO, TERENCE TAO, and VAN VU

Abstract

Let Q_n denote a random symmetric $(n \times n)$ -matrix, whose upper-diagonal entries are independent and identically distributed (i.i.d.) Bernoulli random variables (which take values 0 and 1 with probability $1/2$). We prove that Q_n is nonsingular with probability $1 - O(n^{-1/8+\delta})$ for any fixed $\delta > 0$. The proof uses a quadratic version of Littlewood-Offord-type results concerning the concentration functions of random variables and can be extended for more general models of random matrices.

1. Introduction

Let A_n denote a random $(n \times n)$ -matrix, whose entries are independent and identically distributed (i.i.d.) Bernoulli random variables that take values 0 and 1 with probability $1/2$. A basic question is the following.

Question 1.1

Is it true that A_n is almost surely (a.s.) nonsingular?

Here and later, we say that an event holds a.s. if it holds with probability tending to 1 as n tends to infinity.

Question 1.1 was answered affirmatively by Komlós [6] in 1967. Later, Komlós generalized the result (to more general models of random matrices) in [7] and also simplified the proof (see [1]). In a recent article, Tao and Vu [10] found a different proof that leads to a sharp estimate on the absolute value of the determinant of A_n .

Another popular model of random matrices is that of random symmetric matrices; this is one of the simplest models that has nontrivial correlations between matrix entries. Let Q_n denote a random symmetric $(n \times n)$ -matrix whose upper-diagonal entries $(q_{ij}, 1 \leq i \leq j \leq n)$ are i.i.d. Bernoulli random variables. It is natural to ask the following question.

Question 1.2

Is it true that Q_n is a.s. nonsingular?

DUKE MATHEMATICAL JOURNAL

Vol. 135, No. 2, © 2006

Received 10 May 2005. Revision received 15 March 2006.

2000 *Mathematics Subject Classification*. Primary 15A52; Secondary 05D40.

Tao's work partially supported by a Packard Foundation grant.

Vu's work partially supported by a Sloan Foundation fellowship and a National Science Foundation Career grant.

As far as we can trace, this question was first posed by Weiss in the early 1990s. Despite its obvious similarity to Question 1.1, we do not know of any partial results concerning this question prior to this article. A significant new difficulty is that the symmetry ensures that the determinant $\det(Q_n)$ is a quadratic function of each row, as opposed to $\det(A_n)$, which is a linear function of each row.

The goal of the current article is to give an affirmative answer to Question 1.2.

THEOREM 1.3

Q_n is a.s. nonsingular. More precisely,

$$p_n := \mathbf{P}(Q_n \text{ is singular}) = O(n^{-1/8+\delta})$$

for any positive constant δ . (The implicit constant in the $O()$ notation is, of course, allowed to depend on δ .)

Remark 1.4

The exponent $-1/8 + \delta$ can be improved somewhat by tightening the calculation and applying more technical arguments (e.g., it can be improved to $-1/4 + \delta$ fairly easily; see [2] or [13, Chapter 7]). However, to improve the bound to an exponential bound (in the spirit of [5]) seems to require new ideas (see Section 7).

The rest of the article is organized as follows. In Section 2, we present our approach and the key lemmas. The lemmas are discussed in Sections 3–5. Section 6 is devoted to the generalization of the result to other models of random matrices. We conclude with Section 7, which contains several open questions.

Notation. In the whole article, we assume that n is large, whenever needed. The asymptotic notation is used under the assumption that $n \rightarrow \infty$. $\mathbf{E}$ and $\mathbf{Var}$ denote expectation and variance, respectively; $\log$ denotes the logarithmic with natural base.

2. The approach and main lemmas

As mentioned in the introduction, there are now three different proofs of Komlós's 1967 result on the nonsingularity of A_n . The simpler ones are [1] and [10]. But the original (and longest) proof, from [6], is what really inspires us. The key difference between these proofs lies in the ways in which one generates A_n . In the proofs from [1] and [10], one builds up A_n by exposing the row vectors one by one and making use of the independence of these vectors. This approach, unfortunately, is no longer effective for Q_n , as the last few rows are almost deterministic once one has exposed all rows above them. In [6], one builds up A_n by taking A_{n-1} and adding a (random) row and a (random) column. This idea turns out to be useful for the consideration of Q_n . However, for Q_n the additional row and column are not independent. They are transposes of each other, and this has become the main obstacle. We have managed to

overcome this obstacle by developing a *quadratic* variant of Littlewood-Offord-type results concerning the concentration of random variables (see Section 4).

The basic strategy is to relate the rank of Q_n with the rank of Q_{n+1} . Assume that we get Q_{n+1} by adding a new column and its transpose as a new row to Q_n . Our starting point is the simple observation

$$\text{rank}(Q_n) \leq \text{rank}(Q_{n+1}) \leq \text{rank}(Q_n) + 2. \quad (1)$$

We refine this by showing that if Q_n is singular (so that $\text{rank}(Q_n) < n$), then $\text{rank}(Q_{n+1})$ will equal $\text{rank}(Q_n) + 2$ with high probability; similarly, if Q_n is nonsingular (so that $\text{rank}(Q_n) = n$), then $\text{rank}(Q_{n+1})$ will equal $\text{rank}(Q_n) + 1 = n + 1$ with high probability. These two results together are then easily combined with an inductive argument to show that $\text{rank}(Q_n) = n$ with high probability.

We now turn to the details. Let us fix a small positive constant $\epsilon > 0$. We allow the implicit constants to depend on ϵ , and we assume that n is sufficiently large, depending on ϵ . Set

$$N := n^{1-\epsilon}. \quad (2)$$

Definition 2.1

Given m vectors $\{v_1, v_2, \dots, v_m\}$, a *linear combination* of the v_i is a vector $v = c_1 v_1 + \dots + c_m v_m$, where the c_i are real numbers. We say that a linear combination *vanishes* if v is the zero vector. A vanishing linear combination has *degree* k if exactly k among the c_i are nonzero. We call a singular $(n \times n)$ -matrix *normal* if its row vectors do not admit a nontrivial vanishing linear combination with degree less than N . Otherwise, we call the matrix *abnormal*.

Remark 2.2

We use the terms *normal* and *abnormal* only when the matrix in question is singular. These terms are not defined (and we do not need them) for nonsingular matrices.

In Section 3, we prove that most singular matrices are normal.

LEMMA 2.3

The probability that Q_n is singular and abnormal is $O((2/3)^n)$.

In Section 5, we prove the following.

LEMMA 2.4

Let A be a (deterministic) $n \times n$ singular normal matrix, and let A' be the $(n+1 \times n+1)$ -matrix formed by augmenting A by a random vector of length $n+1$ and its transpose.

Then

$$\mathbf{P}(\text{rank}(A') - \text{rank}(A) < 2) = O(N^{-1/2}),$$

and thus

$$\mathbf{P}(\text{rank}(A') - \text{rank}(A) = 2) = 1 - O(N^{-1/2}).$$

Intuitively, these two lemmas state that, in most cases, augmenting a singular matrix by a random vector and its transpose increases the rank by exactly 2. Note that by Bayes's identity, Lemma 2.4 automatically generalizes to matrices A , which are random instead of deterministic, as long as the random vector that is augmenting A is independent of A .

We now develop analogues of Lemmas 2.3 and 2.4 for nonsingular matrices.

Definition 2.5

A row of an $n \times n$ nonsingular matrix is called *good* if its exclusion leads to an $((n-1) \times n)$ -matrix whose column vectors admit a nontrivial vanishing linear combination with degree at least N . (In fact, there is exactly one such combination, up to scaling, as the rank of this $((n-1) \times n)$ -matrix is $n-1$.) Otherwise, a row is *bad*. We say that an $n \times n$ nonsingular matrix A is *perfect* if every row in A is a good row. If a nonsingular matrix is not perfect, we call it *imperfect*.

Remark 2.6

We use the terms *perfect* and *imperfect* only when the matrix in question is nonsingular. These terms are not defined for singular matrices.

In Section 3, we prove that most nonsingular matrices are perfect.

LEMMA 2.7

The probability that Q_n is both nonsingular and imperfect is $O((2/3)^n)$.

In Section 5, we prove the following analogue of Lemma 2.4.

LEMMA 2.8

Let A be a (deterministic) nonsingular perfect symmetric matrix of size n , and let A' be the $((n+1) \times (n+1))$ -matrix formed by augmenting A by a random $(n+1)$ -vector of zeros and ones, and its transpose. Then

$$\mathbf{P}(\text{rank}(A') = \text{rank}(A)) = O(N^{-1/8})$$

for any positive constant δ , where the implicit constant can, of course, depend on δ . In particular, since

$$n = \text{rank}(A) \leq \text{rank}(A') \leq n + 1,$$

we see that

$$\mathbf{P}(\text{rank}(A') - \text{rank}(A) = 1) = 1 - O(N^{-1/8}).$$

The last two lemmas are the nonsingular counterparts of the first two. Together, they state that if a matrix already has full rank, augmenting it typically produces another matrix of full rank. Again, we can automatically generalize Lemma 2.8 to the case when A is random and independent of the augmenting row.

Proof of Theorem 1.3

Let us assume these lemmas for the moment and conclude the proof of Theorem 1.3.

Consider a random matrix Q_n . We embed it into a sequence $\{Q_1, Q_2, \dots\}$ of random matrices, where Q_{n+1} is formed from Q_n by adding a random vector of zeros and ones (independent of Q_n) of length $n + 1$ as the last column and setting its transpose as the last row.

Define the (somewhat artificial) random variable X_n by setting $X_n = 0$ if Q_n is nonsingular (and thus $\text{rank}(Q_n) = n$) and setting $X_n = (1.1)^{n - \text{rank}(Q_n)}$ otherwise. Thus X_n ranges between 0 and $(1.1)^n$. We have the following decay estimate for the expectation $\mathbf{E}(X_n)$ of X_n .

LEMMA 2.9

We have $\mathbf{E}(X_{n+1}) \leq 0.99\mathbf{E}(X_n) + O(N^{-1/8})$.

Proof

For any $0 \leq j \leq n$, let A_j be the event where Q_n has rank $n - j$ and where A_n is neither abnormal (if $j > 0$) nor imperfect (if $j = 0$). By Bayes's identity and Lemmas 2.3 and 2.7, we have

$$\mathbf{E}(X_n) = \sum_{j=0}^n (1.1)^j \mathbf{P}(A_j) + O\left((1.1)^n \left(\frac{2}{3}\right)^n\right)$$

and

$$\mathbf{E}(X_{n+1}) = \sum_{j=0}^n \mathbf{E}(X_{n+1} | A_j) \mathbf{P}(A_j) + O\left((1.1)^n \left(\frac{2}{3}\right)^n\right).$$

Now let us condition on the event A_0 ; thus Q_n is nonsingular. From Lemma 2.4, we see that Q_{n+1} has rank n with probability $O(N^{-1/8})$ and rank $n+1$ otherwise. Thus

$$\mathbf{E}(X_{n+1}|A_0) = O(N^{-1/8}).$$

Now let $1 \leq j \leq n$, and condition on the event A_j ; thus Q_n is singular with rank $n-j$. From Lemmas 2.3 and 2.4, we see that Q_{n+1} has rank $n-j+2$ with probability $1 - O(N^{-1/2})$ and has rank $n-j$ or $n-j+1$ otherwise. Thus

$$\begin{aligned} \mathbf{E}(X_{n+1}|A_j) &\leq \mathbf{E}(2^{n+1-\text{rank}(Q_{n+1})}|A_j) \\ &\leq (1.1)^{j-1} + O(N^{-1/2})(1.1)^{j+1} \\ &\leq 0.99(1.1)^j \end{aligned}$$

if $N = n^{1-\epsilon}$ is large enough. Putting all these estimates together, and noting that

$$O\left((1.1)^n \left(\frac{2}{3}\right)^n\right) = O(N^{-1/8}),$$

we obtain the claim. □

From Lemma 2.9 and an easy induction, we see that

$$\mathbf{E}(X_n) = O(N^{-1/8})$$

for all large n . From Markov's inequality, we then see that

$$\mathbf{P}(Q_n \text{ singular}) = \mathbf{P}(X_n \geq 1) = O(N^{-1/8}).$$

Theorem 1.3 then follows from the definition of N . □

It remains to prove Lemmas 2.3, 2.4, 2.7, and 2.8. This is done over the next few sections. Among these lemmas, the first three are variants of lemmas from [6] and are relatively simple. The proof of Lemma 2.8 is somewhat more complicated and requires a new machinery, which is discussed in Section 4.

3. Proofs of Lemmas 2.3 and 2.7

The two proofs are similar and rely on the following simple observation from [6] (which has also been used in [5], [10], [12]).

LEMMA 3.1

Let H be a linear subspace in $\mathbb{R}^n$ of dimension at most $d \leq n$. Then it contains at most 2^d vectors from $\{0, 1\}^n$.

Proof

The space H is spanned by the row vectors of a $d' \times n$ full-rank matrix, where $d' := \dim(H) \leq d$. This matrix has at least one nonsingular $(d' \times d')$ -minor; thus there exists a set of d' coordinates of $\mathbb{R}^n$ which can be used to parameterize H . But in $\{0, 1\}^n$, these coordinates take only $2^{d'} \leq 2^d$ values, and the claim follows. $\square$

Proof of Lemma 2.3

For any $1 \leq d \leq n$, let $g(n, d)$ be the probability that the row vectors of Q_n admit a nontrivial vanishing linear combination of degree d . For $d = 1$, we have the easy bound $g(n, 1) \leq n2^{-n}$ since $g(n, d)$ is simply the probability that one of the rows of Q_n is entirely zero. Now take $d \geq 2$. To bound $g(n, d)$ from above, notice that by symmetry and the union bound, we have the crude estimate

$$g(n, d) \leq \binom{n}{d} h(n, d) \leq n^d h(n, d),$$

where $h(n, d)$ is the probability that the first d rows admit such combination. This means that if we fix the first $d - 1$ row vectors, then the d th row vector lies in the subspace spanned by these vectors. The same claim is true if we delete the first $d - 1$ columns from Q_n . (We need to do this, as Q_n is symmetric.) The remaining entries in the d th row vector are now distributed independently in $\{0, 1\}^{n-d-1}$, and so by Lemma 3.1, the probability of lying in the span of the first $d - 1$ row vectors is at most $2^{d-1}/2^{n-d-1}$. Thus the probability that Q_n is singular and abnormal is at most

$$\sum_{d=1}^N g(n, d) \leq n2^{-n} + \sum_{d=2}^N \frac{n^{d+1}2^{d-1}}{2^{n-d-1}} = o\left(\left(\frac{2}{3}\right)^n\right)$$

by definition (2) of N . $\square$

Proof of Lemma 2.7

Let $b(n)$ be the probability that the last row of Q_n is bad. By symmetry and the union bound, the probability that Q_n is nonsingular and imperfect is at most $nb(n)$. We can bound $b(n)$ using the same argument as in the proof of Lemma 2.7, with a slight modification; the column vectors have length $n - 1$, so we need to replace n by $n - 1$, but this does not affect the bound. We omit the details. $\square$

4. A quadratic Littlewood-Offord inequality

Let us start with the following classical result, proved by Erdős, which strengthens an earlier result of Littlewood and Offord [8].

THEOREM 4.1 (Linear Littlewood-Offord inequality; see [3])

Let $z_1, \dots, z_n$ be i.i.d. random variables that take values 0 and 1 with probability $1/2$. Let $a_1, \dots, a_n$ be real deterministic coefficients with $|a_i| \geq 1$ for at least $k \geq 1$ values of i . Then for any interval $I \subset \mathbf{R}$ of length 1, we have

$$\mathbf{P}\left(\sum_{i=1}^n a_i z_i \in I\right) = O(k^{-1/2}),$$

where the implied constant is absolute.

Roughly speaking, the theorem says that linear random sums cannot concentrate on small intervals if the coefficients of the underlying linear form are large.

Remark 4.2

There are a number of far-reaching generalizations and interesting refinements of Theorem 4.1 (see, e.g., [4] and references therein). We mention some rather trivial ones here (which we need later). First, we can replace the unit interval I with any other interval of length $O(1)$ (at the cost of changing the implied constant in $O(k^{-1/2})$, of course) by covering such an interval with unit intervals. Similarly, we may scale the constraint $|a_i| \geq 1$ and replace it with $|a_i| \geq c$ for some other $c > 0$, again at the cost of letting the implied constant depend on c . Finally, one can replace the distribution of the z_i with the distribution $\mathbf{P}(z_i = 0) = \alpha$, $\mathbf{P}(z_i = 1) = \beta$, $\mathbf{P}(z_i = -1) = \gamma$, where α, β, γ are nonnegative constants summing up to 1 and $\alpha < 1$. The implied constant then, of course, depends on α, β, γ .

To conclude the proof of Theorem 1.3, we need to generalize Theorem 4.1 in a direction different from what has been done before. Instead of considering a linear form, we are going to consider a *quadratic* form of the z_i . (In fact, our method works for polynomials of any fixed degree by iterating the argument in Theorem 4.3.) Consider random variables z_i as in Theorem 4.1 and define the quadratic random variable

$$Q := \sum_{1 \leq i, j \leq n} c_{ij} z_i z_j. \quad (3)$$

The main result of this section is the following quadratic generalization of Theorem 4.1.

THEOREM 4.3 (Quadratic Littlewood-Offord inequality)

Let the quadratic random variable Q be as in (3), let $\{1, \dots, n\} = U_1 \cup U_2$ be any nontrivial partition, and let S be any nonempty subset of U_1 . For each $i \in S$, let d_i be the number of indices $j \in U_2$ such that $|c_{ij}| \geq 1$. Suppose that $d_i \geq 1$ for each $i \in S$.

Then for any interval I of length 1, we have

$$\mathbf{P}(Q \in I) = O\left(|S|^{-1/2} + |S|^{-1} \sum_{i \in S} d_i^{-1/2}\right)^{1/4}.$$

The implied constant is absolute.

It is unlikely that the bound on the right-hand side is the best possible, but for us, any bound that decays to zero when the number of large coefficients c_{ij} goes to infinity suffices.

The proof of Theorem 4.3 is lengthy and is given later. Assuming it for the moment, we have the following corollary.

COROLLARY 4.4

Let Q be as in (3), and suppose that there is a set $U \subseteq \{1, \dots, n\}$ of cardinality $|U| \geq m \geq 2$ such that for each $i \in U$, there are m indices $j \in \{1, \dots, n\}$, where $|c_{ij}| \geq 1$. Then for any interval I of length 1,

$$\mathbf{P}(Q \in I) \leq O(m^{-1/8}).$$

The implied constant is absolute.

Proof

Without loss of generality, we may take m to be even. Let U_1 be an arbitrary subset of U of cardinality $m/2$, and write $U_2 := \{1, \dots, n\} \setminus U_1$. Then for any $i \in U_1$, there exist at least $m/2$ indices $j \in U_2$ for which $|c_{ij}| \geq 1$. Applying Theorem 4.3 with $S := U_1$, we conclude that

$$\begin{aligned} \mathbf{P}(Q \in I) &= O\left(\left(\frac{m}{2}\right)^{-1/2} + \left(\frac{m}{2}\right)^{-1} \sum_{i \in U_1} \left(\frac{m}{2}\right)^{-1/2}\right)^{1/4} \\ &= O(m^{-1/8}), \end{aligned}$$

as desired. □

By rescaling Corollary 4.4, we obtain the following discrete version.

COROLLARY 4.5

Let Q be as in (3), and suppose that there are at least m indices i such that for each i , there are m indices j , where $|c_{ij}| \neq 0$. Then

$$\mathbf{P}(Q = 0) = O(m^{-1/8}),$$

where the implied constant is absolute.

We use this corollary to establish Lemma 2.8.

4.6. Proof of Theorem 4.3

We now prove Theorem 4.3. As a first attempt to prove this theorem, one may try to view the quadratic form Q as a linear form

$$Q = \sum_{i=1}^n Q_i z_i, \quad (4)$$

where the coefficients Q_i are themselves linear-form random variables $Q_i := \sum_{j=1}^n c_{ij} z_j$. Thus one may hope to obtain Theorem 4.3 from two applications of Theorem 4.1. Unfortunately, there is a serious obstruction to this strategy because the coefficients $Q_1, \dots, Q_n$ are not independent of the variables $z_1, \dots, z_n$. However, we can get around this obstacle by the following decoupling lemma, which relies on the Cauchy-Schwarz inequality.

LEMMA 4.7 (Decoupling lemma)

Let X and Y be random variables, and let $E = E(X, Y)$ be an event depending on X and Y . Then

$$\mathbf{P}(E(X, Y)) \leq \mathbf{P}(E(X, Y) \wedge E(X', Y) \wedge E(X, Y') \wedge E(X', Y'))^{1/4},$$

where X' and Y' are independent copies of X and Y , respectively.

Remark 4.8

This lemma is a probabilistic analogue of the well-known result in extremal graph theory: if a bipartite graph connecting n and m vertices contains at least cnm edges for some $0 \leq c \leq 1$, then it also contains at least $c^4 n^2 m^2$ copies of the four-cycle C_4 , where we include degenerate four-cycles. Indeed, the two results are easily shown to be equivalent. This decoupling lemma also plays the role of the van der Corput lemma used in Weyl's estimation of exponential sums with quadratic (or, more generally, polynomial) phases; indeed, it is quite likely that one could obtain an estimate very similar to Theorem 4.3 by means of these techniques (combined with Esséen's concentration inequality). However, we have chosen a more elementary combinatorial approach here.

Proof of Lemma 4.7

By discretizing, we can reduce to the case when X takes a finite number of values $x_1, \dots, x_n$ and Y takes a finite number of values $y_1, \dots, y_m$. From Bayes's identity,

we have

$$\mathbf{P}(E(X, Y)) = \sum_{i=1}^n \mathbf{P}(E(x_i, Y))\mathbf{P}(X = x_i)$$

and

$$\mathbf{P}(E(X, Y) \wedge E(X, Y')) = \sum_{i=1}^n \mathbf{P}(E(x_i, Y))^2 \mathbf{P}(X = x_i),$$

and hence, by the Cauchy-Schwarz inequality,

$$\mathbf{P}(E(X, Y)) \leq \mathbf{P}(E(X, Y) \wedge E(X, Y'))^{1/2}.$$

Similarly, we have

$$\mathbf{P}(E(X, Y) \wedge E(X, Y')) = \sum_{j=1}^m \sum_{j'=1}^m \mathbf{P}(E(X, y_j) \wedge E(X, y_{j'}))\mathbf{P}(Y = y_j)\mathbf{P}(Y = y_{j'})$$

and

$$\begin{aligned} & \mathbf{P}(E(X, Y) \wedge E(X, Y') \wedge E(X', Y) \wedge E(X', Y')) \\ &= \sum_{j=1}^m \sum_{j'=1}^m \mathbf{P}(E(X, y_j) \wedge E(X, y_{j'}))^2 \mathbf{P}(Y = y_j)\mathbf{P}(Y = y_{j'}), \end{aligned}$$

so by the Cauchy-Schwarz inequality again,

$$\mathbf{P}(E(X, Y) \wedge E(X, Y')) \leq \mathbf{P}(E(X, Y) \wedge E(X, Y') \wedge E(X', Y) \wedge E(X', Y'))^{1/2}.$$

Combining these two applications of Cauchy and Schwarz, we obtain the claim. $\square$

We return to the task of proving Theorem 4.3. Let $Z \in \{0, 1\}^n$ be the random variable $(z_1, \dots, z_n)$. Consider the quadratic form $Q(Z) = Q(z_1, \dots, z_n)$ defined by (3), and fix a nontrivial partition $\{1, \dots, n\} = U_1 \cup U_2$ and a nonempty subset S of U_1 . Let I be an interval of length 1. We need to prove that

$$\mathbf{P}(Q(Z) \in I)^4 = O\left(|S|^{-1/2} + |S|^{-1} \sum_{i \in S} d_i^{-1/2}\right).$$

Define $X := (z_i)_{i \in U_1}$, and define $Y := (z_i)_{i \in U_2}$. We can write $Q(Z) = Q(X, Y)$. Let z'_i be an independent copy of z_i , and set $X' := (z'_i)_{i \in U_1}$ and $Y' := (z'_i)_{i \in U_2}$. Applying Lemma 4.7, we see that it suffices to show that

$$\mathbf{P}(Q(X, Y), Q(X, Y'), Q(X', Y), Q(X', Y') \in I) = O\left(|S|^{-1/2} + |S|^{-1} \sum_{i \in S} d_i^{-1/2}\right).$$

A simple calculation shows that the random variable

$$R := Q(X, Y) - Q(X', Y) - Q(X, Y') + Q(X', Y')$$

can be written as

$$\begin{aligned} R &= \sum_{i \in U_1} \sum_{j \in U_2} c_{ij} (z_i - z'_i)(z_j - z'_j) \\ &= \sum_{i \in U_1} R_i w_i, \end{aligned}$$

where for $i \in U_1$, w_i is the random variable $w_i := z_i - z'_i$ and R_i is the random variable

$$R_i := \sum_{j \in U_2} c_{ij} w_j.$$

We have eliminated the coupling problem in factorization (4) because the random variables $(R_i)_{i \in U_1}$ are independent of the random variables $(w_i)_{i \in U_1}$.

Consider the four events $Q(X, Y) \in I$, $Q(X', Y) \in I$, $Q(X, Y') \in I$, and $Q(X', Y') \in I$. If all of these hold, then R lies in the interval $J := 2I - 2I$ of length 4. Thus it suffices to show that

$$\mathbf{P}(R \in J) = O\left(|S|^{-1/2} + |S|^{-1} \sum_{i \in S} d_i^{-1/2}\right).$$

Recall that for each $i \in U_1$, d_i is the number of coefficients $j \in U_1$ for which $|c_{ij}| \geq 1$. For each $i \in S \subseteq U_1$, we may apply Theorem 4.1 (and Remark 4.2) to the random variable R_i to obtain

$$\mathbf{P}(|R_i| < 1) = O(d_i^{-1/2}). \quad (5)$$

By the union bound, we thus have the crude estimate

$$\mathbf{P}(|R_i| \geq 1 \text{ for all } i \in S) = 1 - O\left(\sum_{i \in S} d_i^{-1/2}\right).$$

This use of the union is somewhat wasteful, and we can do better by invoking the second moment method. For each $i \in S$, let I_i be the indicator variable of the event $|R_i| \geq 1$; thus $I_i = 1$ when $|R_i| \geq 1$, and $I_i = 0$ otherwise. Thus (5) can be rewritten as

$$\mathbf{E}(I_i) = 1 - O(d_i^{-1/2}),$$

and hence, by linearity of expectation,

$$\mathbf{E}\left(\sum_{i \in S} I_i\right) = |S| - O\left(\sum_{i \in S} d_i^{-1/2}\right).$$

Also, since $d_i \geq 1$, we have at least one $j \in U_2$ for which $|c_{ij}| \geq 1$, which easily implies that $\mathbf{E}(I_i) \geq 1/2$. Thus we also have

$$\mathbf{E}\left(\sum_{i \in S} I_i\right) \geq \frac{|S|}{2}.$$

Next, we compute the variance of $\sum_{i \in S} I_i$:

$$\begin{aligned} \mathbf{Var}\left(\sum_{i \in S} I_i\right) &= \mathbf{E}\left(\left(\sum_{i \in S} I_i\right)^2\right) - \mathbf{E}\left(\sum_{i \in S} I_i\right)^2 \\ &\leq |S|^2 - \left(|S| - O\left(\sum_{i \in S} d_i^{-1/2}\right)\right)^2 \\ &= O\left(|S| \sum_{i \in S} d_i^{-1/2}\right). \end{aligned}$$

By Chebyshev's inequality, we conclude that

$$\mathbf{P}\left(\sum_{i \in S} I_i \leq \frac{1}{4}|S|\right) \leq \frac{16 \mathbf{Var}\left(\sum_{i \in S} I_i\right)}{|S|^2} = O\left(\frac{1}{|S|} \sum_{i \in S} d_i^{-1/2}\right).$$

Thus with probability $1 - O\left((1/|S|) \sum_{i \in S} d_i^{-1/2}\right)$, we have $|R_i| \geq 1$ for at least $|S|/4$ values of $i \in S$.

Now, let us temporarily condition the R_i to be fixed for all $i \in U_1$ and assume that $|R_i| \geq 1$ for at least $|S|/4$ values of $i \in S$. Applying Theorem 4.1 (and Remark 4.2) to $R = \sum_{i \in U_1} w_i R_i$ (treating the R_i as fixed coefficients), we have the conditional probability estimate

$$\mathbf{P}(R \in J | R_i \text{ fixed}) = O(|S|^{-1/2}).$$

By the preceding discussion and Bayes's identity, we thus have

$$\mathbf{P}(R \in J) = O(|S|^{-1/2}) + O\left(\frac{1}{|S|} \sum_{i \in S} d_i^{-1/2}\right),$$

as desired, thus concluding the proof of Theorem 4.3. □

5. Proofs of Lemmas 2.4 and 2.8

Proof of Lemma 2.4

Let A be a normal symmetric singular $(n \times n)$ -matrix of rank d . Let v_i be the i th row vector of A . Without loss of generality, we can assume that $v_1, \dots, v_d$ are linearly independent. Thus the last row vector v_n can be written as a linear combination of these vectors in a unique way:

$$v_n = \sum_{i=1}^d c_i v_i.$$

As A is normal, by definition at least N among the coefficients c_i are nonzero.

Consider the addition of a random $(0, 1)$ -column of length n to A . Each of the row vectors v_i receives a new (random) coordinate and becomes a new vector v'_i . Clearly, $v'_1, \dots, v'_d$ are still independent. If the new matrix A' fails to have a larger rank than A , then the last row v'_n must remain within the span of $v'_1, \dots, v'_d$. By considering the first n coordinates, the only way this can happen is if

$$v'_n = \sum_{i=1}^d c_i v'_i.$$

This implies that the last coordinate x_{n+1} of v'_n satisfies

$$x_{n+1} = \sum_{i=1}^d c_i y_{n+1}^i, \tag{6}$$

where y_{n+1}^i is the last coordinate of v'_i . Since x_{n+1} and y_{n+1}^i are i.i.d. $(0, 1)$ random variables and at least N of the c_i are nonzero, Theorem 4.1 (see also Remark 4.2) implies that the probability that (6) holds is $O(N^{-1/2})$. Thus we can conclude that with probability $1 - O(N^{-1/2})$, the new column increases the rank by 1. If adding the new column increases the rank by 1, then by the fact that A is symmetric, adding the column and its transpose as a new row increases the rank of A by 2 (regardless of the value of the last diagonal entry), concluding the proof. $\square$

Proof of Lemma 2.8

Let A be a perfect nonsingular symmetric matrix of order n . Let A' be the $(n+1) \times (n+1)$ symmetric matrix obtained from A by adding a new random $(0, 1)$ -column u of length $n+1$ as the $(n+1)$ st column and its transpose as the $(n+1)$ st row.

Let $x_1, \dots, x_{n+1}$ be the coordinates of u ; x_{n+1} is the lower-right diagonal entry of A' . The determinant for A' can be expressed as

$$(\det A)x_{n+1} + \sum_{i=1}^n c_{ij}x_i x_j,$$

where c_{ij} is the ij -cofactor of A . We can rewrite $\det A'$ as

$$Q(x_1, \dots, x_{n+1}) = (\det A)x_{n+1}^2 + \sum_{i=1}^n c_{ij}x_i x_j,$$

thanks to the fact that $x_{n+1}^2 = x_{n+1}$. We are going to bound the probability that $Q = 0$.

In order to apply Corollary 4.5, we next show that for each $1 \leq i \leq n$, many among the c_{ij} are not zero.

Since A is nonsingular, dropping the i th row (for any $1 \leq i \leq n$) results in an $(n-1 \times n)$ -matrix whose columns admit a unique (up to scaling) vanishing linear combination $\sum_{j=1}^n a_j u_j$. As A is perfect, at least N among the coefficients a_j are nonzero. For each j where $a_j \neq 0$, dropping both the i th row and the j th column must result in a full-rank matrix of order $n-1$. Thus $c_{ij} \neq 0$. We can conclude that for each $1 \leq i \leq n$, there are at least N indices j where $c_{ij} \neq 0$. The claim of the lemma follows by applying Corollary 4.5 with $m = N$. $\square$

6. More general results

6.1. Generalizations of Theorem 4.3

Theorem 4.3 and Corollary 4.4 can be extended to polynomials with arbitrary degree. One such extension reads as follows.

THEOREM 6.2

Let $z_1, z_2, \dots, z_n$ be i.i.d. random variables that take values 0 and 1 with probability $1/2$. Let k be a fixed positive integer. Let

$$f := \sum_{1 \leq i_1, i_2, \dots, i_k \leq n} c_{i_1, i_2, \dots, i_k} z_{i_1} z_{i_2}, \dots, z_{i_k},$$

where at least $n^{k-1}m$ of the coefficients $c_{i_1, i_2, \dots, i_k}$ are at least 1 in absolute value. Then for any interval I of length 1,

$$\mathbf{P}(f \in I) = O\left(\frac{1}{m^{a_k}}\right),$$

where $a_k = 2^{-(k^2+k)/2}$ and the implicit constant in O depends on k .

The proof proceeds via induction on k with the base case being the classical Littlewood-Offord lemma and the inductive step closely following that of Theorem 4.3, including

the use of the following generalization of the decoupling lemma (also proven by induction on k).

LEMMA 6.3 (Decoupling lemma)

Let $X_1, \dots, X_k$ be random variables, and let $E = E(X_1, \dots, X_k)$ be an event depending on the X_i . Then

$$\mathbf{P}(E(X_1, \dots, X_k)) \leq \mathbf{P}\left(\bigwedge_{S \subset \{1, \dots, k\}} E(X_1^S, \dots, X_k^S)\right)^{1/2^k},$$

where $X_i^S = X_i$ if $i \in S$ and $X_i^S = X'_i$, an independent copy of X_i if $i \notin S$.

Theorem 4.3 can also be extended to more general classes of variables by a nearly identical proof, with the main difference being that Theorem 4.1 must be replaced by [4, Theorem 4].

6.4. Generalizations of Theorem 1.3

We say that a random variable ξ has the ρ -property if

$$\max_{c \in \mathbb{R}} \mathbf{P}(\xi = c) \leq \rho.$$

Let ξ_{ij} , $1 \leq i, j \leq n$, be independent random variables. Assume that there is a constant $\rho < 1$ (not depending on n) such that for all $1 \leq i < j \leq n$, ξ_{ij} has the ρ -property. Observe that we do not require that ξ_{ij} be identically distributed, and furthermore, that we do not place any requirements on the diagonal elements of the matrix.

THEOREM 6.5

Let ξ_{ij} , $1 \leq i \leq j \leq n$, be as in Section 6.4. Let Q_n be the random symmetric matrix with upper-diagonal entries ξ_{ij} . Then Q_n is nonsingular with probability $1 - O(n^{-1/8+\delta})$, where the implicit constant depends only on ρ and δ .

To prove this result, it suffices to show that analogues of Lemmas 2.3, 2.4, 2.7, and 2.8 still hold for this more generalized model. Lemmas 2.3 and 2.7 (with $2/3$ replaced by any δ with $\rho < \delta < 1$) follow from the same argument as in Theorem 1.3, except that Lemma 3.1 must be replaced by the following.

LEMMA 6.6

Let H be a linear subspace in $\mathbb{R}^n$ of dimension at most $d \leq n$. Let v be a vector whose entries are independent random variables, all but one of which have the ρ -property. Then

$$\mathbf{P}(v \in H) < \rho^{n-d-1}.$$

Proof

As before, H can be parameterized by some set of d' coordinates, where $d' \leq d$. Once those coordinates of v are known, the remaining coordinates can each take on at most one value for all $(v \in H)$, giving a necessary set of $(n - d')$ independent events, $(n - d' - 1)$ of which have probability at most ρ . □

The proof of Lemma 2.4 also goes through, except that Theorem 4.1 must be replaced by the following rescaled version of the $d = 1$ case of [4, Theorem 4].

LEMMA 6.7

Let $z_1, \dots, z_n$ be independent random variables with the ρ -property. Let $a_1, \dots, a_n$ be real deterministic coefficients with $a_i \neq 0$ for at least k values of i . Then for any interval $c \in \mathbf{R}$, we have

$$\mathbf{P}\Big(\sum_{i=1}^n a_i z_i = c\Big) = O\big((1 + k)^{-1/2}\big),$$

where the implied constant depends only on ρ .

A nearly identical decoupling argument now proves an analogue of Theorem 4.3 with d_i now taken for each i to be the number of j for which c_{ij} is nonzero. Corollary 4.5 (with the implied constant now depending only on ρ) and Lemma 2.8 now follow as in the arguments of Section 4.

7. Open questions

Let us conclude this article with a few open questions. From a quantitative point of view, there are two natural ways to strengthen Questions 1.1 and 1.2.

Problem 7.1

Give an estimate for the determinant.

Problem 7.2

Give an estimate for the probability that the matrix is singular.

In fact, Problem 7.1 seems to be the motivation of Komlós for his original article [6] (see the title of this article), which started this line of research. There are several partial results concerning the model A_n . In the rest of this section, it is more convenient to assume that the entries of A_n (and Q_n) take value 1 and -1 (rather than 1 and 0). Under this condition, Tao and Vu [10] showed that a.s. $\det A_n$ has absolute value $n^{(1/2-o(1))n}$. We conjecture that a similar bound holds for Q_n .

CONJECTURE 7.3

Almost surely, $|\det Q_n| = n^{(1/2-o(1))n}$.

Regarding Problem 7.2, Kahn, Komlós, and Szemerédi [5] proved that the singularity probability of A_n is $O(.999^n)$. This bound has recently been improved (see [12]) to $(3/4 + o(1))^n$. The conjectured bound is $(1/2 + o(1))^n$. We conjecture that the same bound holds for Q_n .

CONJECTURE 7.4

The probability that Q_n is singular is $(1/2 + o(1))^n$.

By considering the probability that the first two rows are equal, it is easy to see that $(1/2 + o(1))^n$ is a lower bound. (One can actually make a more precise conjecture similar to the case with A_n .)

Another problem is to measure how far the eigenvalues of Q_n are from zero. By Wigner's semicircle law, most of the eigenvalues of Q_n have order lying in an interval of length $\Theta(\sqrt{n})$. Based on this, one would expect there to be a gap of order $\Theta(n^{-1/2})$ between zero and the closest eigenvalue of Q_n . This is true in the case when the entries have normal distribution but seems very hard in the discrete setting here. A weaker result has been obtained recently for A_n (see [9], [11]).

The main obstacle in these questions is the fact that the row vectors of Q_n , unlike those of A_n , are not independent. In fact, if one exposes these vectors one by one, then the last few vectors are almost deterministic. The independence among the row vectors is critical in the previous articles [5], [10], and [12]. So, it seems that significant new ideas are required to attack these conjectures.

References

- [1] B. BOLLOBÁS, *Random Graphs*, Academic Press, New York, 1985. [MR 0809996](#)
[395](#), [396](#)
- [2] K. COSTELLO and V. VU, *Rank of random graphs*, preprint, 2006. [396](#)
- [3] P. ERDÖS, *On a lemma of Littlewood and Offord*, Bull. Amer. Math. Soc. **51** (1945),
898–902. [MR 0014608](#)
- [4] G. HALÁSZ, *Estimates for the concentration function of combinatorial number theory
and probability*, Period. Math. Hungar. **8** (1977), 197–211. [MR 0494478](#) [402](#),
[410](#), [411](#)
- [5] J. KAHN, J. KOMLÓS, and E. SZEMERÉDI, *On the probability that a random ± 1 matrix
is singular*, J. Amer. Math. Soc. **8** (1995), 223–240. [MR 1260107](#) [396](#), [400](#),
[412](#)

- [6] J. KOMLÓS, *On the determinant of $(0, 1)$ matrices*, *Studia Sci. Math. Hungar.* **2** (1967), 7–21. [MR 0221962](#) [395](#), [396](#), [400](#), [411](#)
- [7] ———, *On the determinant of random matrices*, *Studia Sci. Math. Hungar.* **3** (1968), 387–399. [MR 0238371](#) [395](#)
- [8] J. E. LITTLEWOOD and A. C. OFFORD, *On the number of real roots of a random algebraic equation, III*, *Rec. Math. [Mat. Sbornik] N.S.* **12(54)** (1943), 277–286. [MR 0009656](#) [401](#)
- [9] M. RUDELSON, *Invertibility of random matrices: Norm of the inverse*, preprint, [arXiv:math.FA/0507024](#) [412](#)
- [10] T. TAO and V. VU, *On random ± 1 matrices: Singularity and determinant*, *Random Structures Algorithms* **28** (2006), 1–23. [MR 2187480](#) [395](#), [396](#), [400](#), [411](#), [412](#)
- [11] ———, *Inverse Littlewood-Offord theorems and the condition number of random discrete matrices*, preprint, [arXiv:math.PR/0511215](#) [412](#)
- [12] ———, *On the singularity probability of random Bernoulli matrices*, to appear in *J. Amer. Math. Soc.*, preprint, [arXiv:math.CO/0501313](#) [400](#), [412](#)
- [13] ———, *Additive Combinatorics*, in preparation. [396](#)

Costello

Department of Mathematics, Rutgers University, 110 Frelinghuysen Road, Piscataway, New Jersey 08854, USA; kcostell@math.ucsd.edu

Tao

Department of Mathematics, University of California, Los Angeles, Los Angeles, California 90095-1555, USA; tao@math.ucla.edu

Vu

Department of Mathematics, Rutgers University, 110 Frelinghuysen Road, Piscataway, New Jersey 08854, USA; vanvu@math.rutgers.edu