

# RANDOM MATRICES HAVE SIMPLE SPECTRUM

TERENCE TAO AND VAN VU

ABSTRACT. Let  $M_n = (\xi_{ij})_{1 \leq i, j \leq n}$  be a real symmetric random matrix in which the upper-triangular entries  $\xi_{ij}, i < j$  and diagonal entries  $\xi_{ii}$  are independent. We show that with probability tending to 1,  $M_n$  has no repeated eigenvalues. As a corollary, we deduce that the Erdős-Renyi random graph has simple spectrum asymptotically almost surely, answering a question of Babai.

## 1. INTRODUCTION

Let  $n$  be an asymptotic parameter going to infinity; we allow all mathematical objects in the discussion below to depend on  $n$  unless explicitly declared to be fixed. Asymptotic notation such as  $o(1)$ ,  $O()$ , or  $\ll$  will always be understood to be with respect to the asymptotic limit  $n \rightarrow \infty$ , for instance  $X \ll Y$  denotes the claim that  $X \leq CY$  for sufficiently large  $n$  and for a fixed  $C$  independent of  $n$ .

In this paper, we study the spectrum of the following general random matrix model.

**Definition 1.1** (A general model). We consider real symmetric random matrices  $M_n$  of the form  $M_n = (\xi_{ij})_{1 \leq i, j \leq n}$ , where the entries  $\xi_{ij}$  for  $i \leq j$  are jointly independent with  $\xi_{ji} = \xi_{ij}$ , the upper-triangular entries  $\xi_{ij}, i < j$  have distribution  $\xi$  for some real random variable  $\xi$  (which may depend on  $n$ ). The diagonal entries  $\xi_{ii}, 1 \leq i \leq n$  can have an arbitrary real distribution (and can be correlated with each other), but are required to be independent of the upper diagonal entries  $\xi_{ij}, 1 \leq i < j \leq n$ .

Important classes of matrices covered by this model include real symmetric Wigner matrix ensembles (e.g. random symmetric sign matrices) and the adjacency matrix of an Erdős-Rényi random graph  $G(n, p)$  (note that we permit the diagonal entries to be identically zero). Notice that we do not require the distribution  $\xi$  to have zero mean, or even to be absolutely integrable. As a matter of fact, our proofs do not require the entries to be iid, either; see Section 5 for details. However, for sake of simplicity, we make the iid assumption in the main sections of this paper.

---

T. Tao is supported by a Simons Investigator grant, the James and Carol Collins Chair, the Mathematical Analysis & Application Research Fund Endowment, and by NSF grant DMS-1266164.

V. Vu is supported by NSF grant DMS 1307797 and AFORS grant FA9550-12-1-0083.

The spectrum of a symmetric matrix is real, and we say that it is simple if all eigenvalues have multiplicity one. This paper deals with the following basic question

*Is it true that the spectrum of a random matrix is simple with high probability ?*

It is easy to see that if the distribution  $\xi$  is continuous, then the spectrum is simple with probability 1. On the other hand, the discrete case is far from trivial. In particular, the following conjecture of Babai has been open since the 1980s [2].

**Conjecture 1.2.** *With probability  $1 - o(1)$ ,  $G(n, 1/2)$  has a simple spectrum.*

In [1], Babai, Grigoriev and Mount showed that the notorious graph isomorphism problem is in  $P$  within the class of graphs with simple spectrum. Conjecture 1.2, if holds, implies that most graphs belong to this class. From universality results [3] on the gap between adjacent eigenvalues, one can show that with probability  $1 - o(1)$ , that *most* (i.e.  $(1 - o(1))n$ ) of the eigenvalues of  $G(n, 1/2)$  are simple; however, the error terms in these universality results do not appear to be strong enough to resolve Babai's conjecture completely.

Our main result provides a positive answer to the question above. We say that a real-valued random variable  $\xi$  is *non-trivial* if there is a fixed  $\mu > 0$  (independent of  $n$ ) such that

$$(1) \quad \mathbf{P}(\xi = x) \leq 1 - \mu$$

for all  $x \in \mathbb{R}$ . In particular, any random variable independent of  $n$  is non-trivial if it is not deterministic (i.e. it does not take a single value almost surely). The distribution  $\xi$  for the adjacency matrix of  $G(n, p)$  will be non-trivial if  $p$  stays bounded away from both 0 and 1 (and in particular if  $0 < p < 1$  is a fixed value such as  $1/2$ ).

**Theorem 1.3** (Simple Spectrum). *Let  $M_n$  be a random matrix of the form in Definition 1.1 whose upper triangular entries have non-trivial distribution for some fixed  $\mu > 0$ . Then for every fixed  $A > 0$  and  $n$  sufficiently large (depending on  $A, \mu$ ), the spectrum of  $M_n$  is simple with probability at least  $1 - n^{-A}$ .*

In the case when  $M_n$  is the adjacency matrix of  $G(n, 1/2)$ ,  $\mu = 1/2$  and Theorem 1.3 implies

**Corollary 1.4.** *Conjecture 1.2 holds.*

The rest of the paper is devoted to the proof of Theorem 1.3. One can easily extend Theorem 1.3 (with the same proof) to more general models where the entries are independent but not iid and also to random Hermitian matrices; see Section 5.

## 2. FROM MULTIPLE EIGENVALUES TO STRUCTURED EIGENVECTORS

The first step in our proof is to reduce the non-simple spectrum problem to a problem about the structure of eigenvectors.

For a symmetric matrix  $M_n$  (either deterministic or random) of size  $n$ , write

$$(2) \quad M_n = \begin{pmatrix} M_{n-1} & X \\ X^* & \xi_{nn} \end{pmatrix}$$

where  $X = (x_1, \dots, x_{n-1}) \in \mathbb{R}^{n-1}$  is the column vector. We need the following (deterministic) lemma:

**Lemma 2.1.** *Let  $M_n$  be a real symmetric matrix of the form (2). Assume that the spectrum of  $M_n$  is not simple. Then  $X$  is orthogonal to an eigenvector of  $M_{n-1}$ .*

*Proof.* We can change basis in  $\mathbb{R}^{n-1}$  so that the standard basis  $e_1, \dots, e_{n-1}$  is an orthonormal eigenbasis of  $M_{n-1}$ ;  $M_{n-1}$  is now a diagonal matrix with entries  $\lambda_1, \dots, \lambda_{n-1}$ . If  $M_n$  has a multiple eigenvalue, then one of the  $\lambda_i$  is an eigenvalue of  $M_n$ . Assume (without loss of generality) that it is  $\lambda_1$  and let  $v = (v_1, \dots, v_n)$  be a corresponding eigenvector. The first row of the equation  $M_n v = \lambda_1 v$  implies that  $x_1 v_n = 0$ . If  $x_1 = 0$  then  $X$  is orthogonal to  $e_1$ . If  $v_n = 0$ , then  $v' := (v_1, \dots, v_{n-1})$  is an eigenvector of  $M_{n-1}$  and the last row of the equation  $M_n v = \lambda_1 v$  implies that  $v' \cdot X = \lambda_1 v_n = 0$ , proving the lemma.  $\square$

In view of this lemma, Theorem 1.3 clearly follows from the following statement.

**Proposition 2.2.** *Let the notation and hypotheses be as in Theorem 1.3, and expand  $M_n$  as (2). Let  $E_1$  be the event that  $X$  is orthogonal to a non-trivial eigenvector of  $M_{n-1}$ . Then  $\mathbf{P}(E_1) \ll n^{-A}$ .*

It remains to prove this proposition. The crucial point here is that  $X$  and  $M_{n-1}$  are independent of each other. Fix a constant  $A > 0$  and call a vector  $v \in \mathbb{R}^n$  *rich* if we have

$$\sup_{x \in \mathbb{R}} \mathbf{P}(X \cdot v = x) \geq n^{-A}$$

where  $X \in \mathbb{R}^n$  is a random vector whose entries are iid copies of  $\xi$ . Let  $E_{2,n-1}$  be the event that an eigenvector of  $M_{n-1}$  is rich. We have

$$\mathbf{P}(E_1) = \mathbf{P}(E_1 | \bar{E}_{2,n-1}) \mathbf{P}(\bar{E}_{2,n-1}) + \mathbf{P}(E_1 | E_{2,n-1}) \mathbf{P}(E_{2,n-1}) \leq n^{-A} + \mathbf{P}(E_{2,n-1}),$$

by the definition of  $E_1$  and  $E_{2,n-1}$ . To prove Proposition 2.2, it therefore suffices to prove

**Proposition 2.3** (Rich eigenvectors are rare). *Let the notation and hypotheses be as above. Then  $\mathbf{P}(E_{2,n-1}) \ll n^{-A}$ .*

In fact we will prove the stronger claim

$$(3) \quad \mathbf{P}(E_{2,n}) \ll \exp(-cn)$$

for some fixed  $c > 0$  (depending on  $\sigma$ ), where  $E_{2,n}$  is the event that an eigenvector of  $M_n$  is rich; Proposition 2.3 then follows by replacing  $n$  with  $n - 1$ .

The key ingredient of our proof of Proposition 2.3 is the so-called *inverse Littlewood-Offord theory*, introduced in [8] (see [5] for a survey) which established almost

completely the structure of rich vectors. On the other hand, one expects that eigenvectors of a random matrix must look random, and thus should not attain any rigid structure. This explains the intuition behind Proposition 2.3. The actual proof, however, requires some novel ideas and delicate arguments, and will be the subject of the next two sections.

### 3. INVERSE LITTLEWOOD-OFFORD THEORY

We recall the definition of a (symmetric) *generalized arithmetic progression* (GAP):

**Definition 3.1.** A set  $P \subset \mathbb{R}$  is a *symmetric GAP of rank  $r$*  if it can be expressed in the form

$$P = \{m_1 g_1 + \cdots + m_r g_r : -M_i \leq m_i \leq M_i, m_i \in \mathbf{Z} \text{ for all } 1 \leq i \leq r\}$$

for some  $r \geq 0$ ,  $g_1, \dots, g_r \in \mathbb{R}$  and some real numbers  $M_1, \dots, M_r$ .

It is convenient to think of  $P$  as the image of an integer box  $B := \{(m_1, \dots, m_r) \in \mathbf{Z}^r : -M_i \leq m_i \leq M_i\}$  under the linear map

$$\Phi : (m_1, \dots, m_d) \mapsto m_1 g_1 + \cdots + m_d g_d.$$

The numbers  $g_i$  are the *generators* of  $P$ , the numbers  $M_i$  are the *dimensions* of  $P$ . We refer to  $r$  as the *rank* of  $P$ . We say that  $P$  is *proper* if this map is one to one. We define by  $\prod_{i=1}^r (2\lfloor M_i \rfloor + 1)$  the volume of  $P$ . For more discussion about GAPs (including non-symmetric GAPs, which we will not use here), see [7].

For a vector  $V = (v_1, \dots, v_n)$ , define the *concentration probability*

$$p_\xi(V) := \sup_{x \in \mathbb{R}} \mathbf{P} \left( \sum_{i=1}^n \xi_i v_i = x \right),$$

where  $\xi_i$  are iid copies of  $\xi$ . In the notation of the previous section, a vector  $V$  is then rich precisely when  $p_\xi(V) \geq n^{-A}$ . Abusing the notation slightly, we also think of  $V$  as a (multi)-set, as the ordering of the coordinates plays no role.

The next theorem determines the structure of rich vectors/sets, asserting that such vectors mostly lie inside a GAP  $P$  of bounded rank.

**Theorem 3.2** (Structure theorem for rich vectors). *Let  $\delta < 1$  and  $A$  be positive constants. There are constant  $d_0 = d_0(\delta, A) \geq 1$  and  $C_0 = C_0(\delta, A)$  such that the following holds. Assume that*

$$p_\xi(V) \geq n^{-A}.$$

*Then for any  $n^\delta \leq m \leq n$ , there exists a proper symmetric GAP  $Q$  of rank  $r \leq r_0$  with volume at most  $C_0 p_\xi(v_1, \dots, v_n)^{-1} m^{-r/2}$  such that  $P$  contains all but at most  $m$  elements of  $V$  (counting multiplicities).*

*Proof.* See [6, Theorem 2.1]. This theorem extended earlier results in [8, 9]; see [5] for a survey.  $\square$

For our purpose, we are going to need the following refinement of Theorem 3.2, in which the GAP  $P$  not only contains most of the rich vector  $V$ , but also contains a large subset of  $V$  that does not concentrate too strongly.

**Theorem 3.3** (A finer structure theorem for rich vectors). *Let  $0 < \varepsilon < 1/4$  and  $A > 0$  be fixed, and let  $V = (v_1, \dots, v_n)$  be a rich (multi-) set. Set  $d_0 = d_0(1/2, A)$  and  $C_0 = C_0(1/2, A)$  from Theorem 3.2. Then there are (multi-)sets  $W' \subset W \subset V$  where  $|W| \geq n - n^{1-\varepsilon/4}$ ,  $|W'| \leq \varepsilon n$ , a parameter  $p \geq n^{-A}$ , and a GAP  $P$  of rank  $d \leq d_0$  and volume at most  $2C_0 p^{-1} n^{-d/2}$  such that the following holds:*

- $W \subset P$ .
- $p_\xi(W') \leq n^{d_0 \varepsilon} p$ .

We now prove this theorem. We first establish the following proposition:

**Proposition 3.4.** *Let  $A \geq 0$  be fixed, and set  $d_0 := d_0(1/2, A)$  and  $C_0 := C_0(1/2, A)$ . Let  $P$  be a proper GAP of rank  $d \leq d_0$  and volume at most  $2C_0 p^{-1} n^{-d/2}$  for some  $p \geq n^{-A}$ . Let  $v_1, \dots, v_n \in P$  (allowing repetitions). Assume  $n$  sufficiently large depending on  $A$ . Then one of the following statements holds:*

- (i) (Stability) *There are indices  $1 \leq i_1 < \dots < i_k \leq n$  with  $k \leq \varepsilon n$  such that the (multi-)set  $V' := \{v_{i_1}, \dots, v_{i_k}\}$  satisfies*

$$p_\xi(V') \leq n^{d_0 \varepsilon} p.$$

- (ii) (Concentration) *There exists a GAP  $P'$  of some rank  $d' \leq d_0$  and volume at most  $(n^{\varepsilon/2} p)^{-1} n^{-d'/2}$  which contains at least  $n - n^{1-\varepsilon/3}$  elements of  $\{v_1, \dots, v_n\}$ .*

*Proof.* Assume that the stability conclusion (i) fails. Let  $I := \{i_1, \dots, i_k\}$  be a subset of  $\{1, \dots, n\}$  of cardinality  $k := \lfloor \varepsilon n \rfloor$  to be chosen later, and set  $V_I := \{v_{i_1}, \dots, v_{i_k}\}$ . Then

$$p_\xi(V_I) > n^{d_0 \varepsilon} p \geq n^{-A}.$$

Applying Theorem 3.2 with  $m := n^{1-\varepsilon/2}$  and  $\delta = 1/2$  (and  $n$  replaced by  $k$ ), we obtain a proper symmetric GAP  $P_I$  of some rank  $d_I \leq d_0$  and volume at most

$$C_0 n^{-d_0 \varepsilon} p^{-1} k^{-(1-\varepsilon/2)d'/2} \leq (n^{\varepsilon/2} p)^{-1} n^{-d'/2}$$

which contains at least  $k - n^{1-\varepsilon/2}$  elements of  $V_I$ .

At present,  $P_I$  is unrelated to  $P$ , but we can make  $P_I$  “commensurate” with  $P$  as follows. Write

$$P_I = \{n_1 w_1 + \dots + n_{d_I} w_{d_I} : |n_i| \leq N_i \text{ for all } 1 \leq i \leq d_I\}$$

and let  $\Sigma \subset \mathbf{Z}^{d_I}$  be the set of  $d_I$ -tuples  $(n_1, \dots, n_{d_I}) \in \mathbf{Z}^{d_I}$  with  $|n_i| \leq N_i$  for all  $1 \leq i \leq d_I$  with  $n_1 w_1 + \dots + n_{d_I} w_{d_I} \in P$ . We say that  $P_I$  has *full rank* in  $P$  if  $\Sigma$  spans  $\mathbb{R}^{d_I}$  as a real vector space. We claim that we may assume without loss of generality that  $P_I$  is of full rank in  $P$ . Indeed, if this is not the case, then  $\Sigma$  is

contained in a hyperplane, which by symmetry we may take to be given by some equation  $x_{d_I} = a_1x_1 + \dots + a_{d_I-1}x_{d_I-1}$ . But then every element  $n_1w_1 + \dots + n_{d_I}w_{d_I}$  in  $P_I \cap P$  can be rewritten as  $\sum_{j=1}^{d_I-1} n_j(w_j + a_jw_{d_I})$  and so one may replace  $P_I$  with a rank  $d_I - 1$  GAP  $P'_I$  of volume at most that of  $P_I$ , such that  $P_I \cap P = P'_I \cap P$ . By the principle of infinite descent, we may iterate this procedure until we replace  $P_I$  with a functionally equivalent GAP which is of full rank in  $P$ .

The purpose of making this full rank reduction is that it cuts down on the number of possible  $P_I$ . Indeed, to specify  $P_I$ , one needs to specify the rank  $d_I$ , the dimensions  $N_1, \dots, N_{d_I}$ , and the generators  $w_1, \dots, w_{d_I}$ . As  $P_I$  has rank at most  $d_0$  and volume at most  $n^{O(1)}$  we have  $O(n^{O(1)})$  choices for  $d_I, N_1, \dots, N_{d_I}$ . To specify the generators, it suffices to choose  $d_I$  linearly independent elements  $(n_1, \dots, n_{d_I})$  of  $\Sigma$ , and their representatives  $n_1w_1 + \dots + n_{d_I}w_{d_I}$  in  $P$ . As  $P$  has volume  $O(n^{O(1)})$ , we see that the total number of choices here is also  $O(n^{O(1)})$ . Thus we see that there are at most  $O(n^{O(1)})$  choices for  $P_I$ .

Applying the pigeonhole principle, we conclude that there exists a fixed GAP  $P'$  of rank  $d' \leq d_0$  and volume at most  $(n^\varepsilon p)^{-1} n^{-d'/2}$  such that, when  $I$  is chosen uniformly at randomly from all subsets of size  $k := \lfloor \varepsilon n \rfloor$  of  $\{1, \dots, n\}$  then with probability  $\gg n^{-O(1)}$ , at least  $k - n^{1-\varepsilon/2}$  of the elements of  $P_I$  lie in  $P'$ . A routine application of the Chernoff inequality then shows that at least  $n - n^{1-\varepsilon/3}$  of the  $v_1, \dots, v_n$  lie in  $P'$ . This gives the desired concentration conclusion (ii).  $\square$

To prove Theorem 3.3, we apply Proposition 3.4 iteratively, as follows.

- (i) Let  $V$  be a rich vector, and set  $p_1 := p_\xi(V)$ , thus  $p_1 \geq n^{-A}$  by hypothesis. By Theorem 3.2 (with  $m = n^{1-\varepsilon/2}$ ), we may find a GAP  $P_1$  of rank  $d_1 \leq d_0$  and volume at most  $C_0 p^{-1} n^{-d_1/2}$  which contains all but at most  $n^{1-\varepsilon/2}$  elements of  $V$ . Set  $V_1 := V \cap P_1$  and  $n_1 := |V_1|$ , thus  $n_1 \geq n - n^{1-\varepsilon/2}$ . Initialize  $i = 1$ , thus  $V_i, P_i, n_i, p_i$  have all been defined.
- (ii) If there exist a subset  $V'_i$  of  $V_i$  of size  $k_i := \lfloor \varepsilon n_i \rfloor$  such that  $p_\xi(V'_i) \leq n_i^{d_0 \varepsilon} p_i$ , then we set  $W', W, p, P$  equal to  $V'_i, V_i, p_i, P_i$  respectively, and STOP. Otherwise, if no such subset  $V'_i$  exists, we move on to step (iii).
- (iii) If  $p_i < n_i^{-A}$  then STOP. Otherwise, by Proposition 3.4 with  $n$  replaced by  $n_i$ , we may find a set  $V_{i+1} \subset V_i$  of size  $n_{i+1} := |V_{i+1}| \geq n_i - n_i^{1-\varepsilon/3}$  contained in a GAP  $P_{i+1}$  of rank  $d_{i+1}$  at most  $d_0$  and volume at most  $p_{i+1}^{-1} n_{i+1}^{-d_{i+1}/2}$ , where  $p_{i+1} := n_i^{\varepsilon/2} p_i$ . Thus  $V_{i+1}, P_{i+1}, n_{i+1}, p_{i+1}$  have all been defined.
- (iv) Increment  $i$  to  $i + 1$  and return to step (ii).

Note that after each successfully completed loop, the probability  $p_i$  increases by a multiplicative factor of  $n_i^{\varepsilon/2}$ , while  $n_i$  only decreases by an additive factor of  $n_i^{1-\varepsilon/3}$ ; since  $p_1$  is initially at least  $n^{-A}$ , we see that after  $O(1)$  steps  $p_i$  will exceed 1, at which point we must terminate at step (ii). Thus the above algorithm can only run for at most  $O(1)$  steps. The same analysis also shows inductively that  $p_i \geq n_i^{-A}$  for all  $i = O(1)$ , so one never terminates at step (iii), and so must instead terminate

at step (ii). It is then routine that  $W', W, p, P$  obey all the properties required for Theorem 3.3.

#### 4. PROOF OF PROPOSITION 2.3

We can now conclude the proof of Proposition 2.3, and more precisely the stronger bound (3). One can show that for any given rich vector  $v$ , the probability that  $v$  is an eigenvector is very small. Ideally, we would like to conclude by bounding the number of rich vectors, using the inverse theorems, and then apply the union bound (this will explain the entropy estimates below). However, this strategy does not work straightforwardly, and we will need to introduce an additional twist to see it through. In this section, all implied constants in the  $O()$  notation can depend on the fixed quantity  $A$  (but not on the quantity  $\varepsilon$  to be introduced shortly).

Suppose that we are in the event  $E_{2,n}$ , thus  $M_n$  has an eigenvector  $V = (v_1, \dots, v_n)$  which is rich. Let  $\varepsilon > 0$  be a small fixed quantity (depending on  $A$ ) to be chosen later. Applying Theorem 3.3, we can find (multi-)sets  $W' \subset W \subset V$  where  $|W| \geq n - n^{1-\varepsilon/4}$ ,  $|W'| \leq \varepsilon n$ , as well as a parameter  $p \gg n^{-A}$ , and a GAP  $P$  of rank  $d = O(1)$  and volume at most  $O(p^{-1}n^{-d/2})$  such that  $W \subset P$  and

$$p_\xi(W') \ll n^{d_0\varepsilon} p.$$

By rounding  $p$  to the nearest multiple of  $n^{-A}$ , we may assume that  $p$  is an integer multiple of  $n^{-A}$ , which must then be of size at most  $O(n^{O(d/2)})$  since otherwise  $P$  would have volume less than 1.

Write  $W = \{v_{i_1}, \dots, v_{i_{n'}}\}$  and  $W' = \{v_{j_1}, \dots, v_{j_k}\}$ , where  $k \leq \varepsilon n$  and  $n - n^{1-\varepsilon/4} \leq n' \leq n$ . From Stirling's formula we observe that the total number of possibilities for  $p, d, n', k, i_1, \dots, i_{n'}$  and  $j_1, \dots, j_k$  is at most  $\exp(O(n\varepsilon \log \frac{1}{\varepsilon}))$ . Thus, if we let  $E_{2,n,p,d,n',k,i_1,\dots,i_{n'},j_1,\dots,j_k}$  denote the event that  $M_n$  has a rich eigenvector obeying the above assertions, then by the union bound we will obtain (3) if we can show that

$$\mathbf{P}(E_{2,n,p,d,n',k,i_1,\dots,i_{n'},j_1,\dots,j_k}) \ll \exp(-cn)$$

for some fixed  $c > 0$  independent of  $\varepsilon$ , and for sufficiently small  $\varepsilon$ .

Let us work now with a single choice of  $n, p, d, n', k, i_1, \dots, i_{n'}, j_1, \dots, j_k$ , and abbreviate  $E_{2,n,p,d,n',k,i_1,\dots,i_{n'},j_1,\dots,j_k}$  as  $E_3$ , thus our task is to show that

$$(4) \quad \mathbf{P}(E_3) \ll \exp(-cn).$$

By symmetry we may assume that  $i_l = l$  for  $l = 1, \dots, n'$ , and that  $j_l = l$  for  $l = 1, \dots, k$ . Thus on the event  $E_3$ , we now have

$$(5) \quad v_1, \dots, v_{n'} \in P$$

and

$$(6) \quad p_\xi(v_1, \dots, v_k) \ll n^{O(\varepsilon)} p.$$

We cover  $E_3$  by the events  $E_3'$  and  $E_3''$ , where  $E_3'$  is the event that we can take  $d = 0$ , and  $E_3''$  is the event that we can take  $d > 0$ .

*Case 1:  $d = 0$ .* In this case,  $P$  is trivial and so  $v_1 = \dots = v_{n'} = 0$ . We now use a conditioning argument of Komlós [4]. Split

$$(7) \quad M_n = \begin{pmatrix} M_{n'} & B \\ B^* & C \end{pmatrix}$$

where  $M_{n'}$  is the top left  $n' \times n'$  minor of  $M_n$ ,  $B$  is the  $n' \times n - n'$  top right minor,  $B^*$  is the adjoint of  $B$ , and  $C$  is the bottom right  $n - n' \times n - n'$  minor. By hypothesis,  $M_n$  has an eigenvector  $V$  with the first  $n'$  coefficients vanishing, which implies from the eigenvector equation  $M_n V = \lambda V$  and (7) that the matrix  $B$  does not have full rank. Thus there exists  $n - n'$  rows of  $B$  which span a proper subspace  $H$  of  $\mathbb{R}^{n-n'}$  in which the remaining  $n' - (n - n')$  rows necessarily lie. The entropy cost of picking these  $n - n'$  rows is  $\binom{n'}{n-n'}$ . Now suppose we fix the position of these rows, as well as the precise values that the random matrix attains on these rows, so that  $H$  is now deterministic. The entries of  $B$  on the remaining rows remain identically distributed with law  $\xi$ . This can be seen by embedding  $H$  in a hyperplane, which can be written as a graph of one of the  $n$  coordinates of  $\mathbb{R}^n$  as a linear combination of the other  $n - 1$  coordinates, and then using (1), we conclude that each of these rows has an independent probability of at most  $1 - \mu$  of lying in  $H$ . Putting all this together, we conclude that

$$\mathbf{P}(E'_3) \leq \binom{n'}{n-n'} \times (1 - \mu)^{n' - (n - n')}$$

and hence from Stirling's formula and the size  $n' = n - O(n^{1-\varepsilon/4})$  of  $n$  that

$$(8) \quad \mathbf{P}(E'_3) \ll \exp(-cn)$$

for some fixed  $c > 0$  independent of  $\varepsilon$  (but depending on  $\mu$ ).

*Case 2.  $d \geq 1$ .* As in Case 1, we split  $M_n$  using (7). Write  $V' := (v_1, \dots, v_{n'})$  (viewed as a column vector), then from the eigenvalue equation  $M_n V = \lambda V$  and (7), we see that  $M_{n'} V'$  lies in the space spanned by  $V'$  and the  $n - n'$  columns of  $B$ . We expand the GAP  $P$  as

$$P = \{n_1 w_1 + \dots + n_d w_d : |n_i| \leq N_i\}$$

for some  $w_1, \dots, w_d \in \mathbb{R}$  and  $N_1, \dots, N_d \geq 1$ . By (5), we have

$$(9) \quad V' = w_1 V'_{(1)} + \dots + w_d V'_{(d)}$$

where each  $V'_{(i)} \in \mathbb{R}^{n'}$  is a vector whose entries all lie in  $[-N_i, N_i] \cap \mathbf{Z}$ . In particular, if we let  $H$  be the subspace of  $\mathbb{R}^{n'}$  spanned by  $V'_{(1)}, \dots, V'_{(d)}$  and the columns of  $B$ , then  $V'$  lies in  $H$ , and  $H$  has dimension at most  $d + (n - n') = O(n^{1-\varepsilon/4})$ .

The total number of possibilities for each vector  $V'_{(i)}$  is at most  $(2N_i + 1)^n$ . By Theorem 3.3,  $P$  has volume at most  $O(p^{-1} n^{-d/2})$ , so the total number of possibilities for  $V'_{(1)}, \dots, V'_{(d)}$  is (very crudely) at most  $O(p^{-1} n^{-d/2})^n$ . Thus, by paying this as an entropy cost, we may assume that  $V'_{(1)}, \dots, V'_{(d)}$  are fixed. For the rest of the argument, we condition on the minors  $B, C$  in (7), so the subspace  $H$  defined previously is now deterministic, while the matrix  $M_{n'}$  remains random (and is of the form in Definition 1.1, with  $n$  replaced by  $n'$ ). The real numbers  $w_1, \dots, w_d$  are also random and may potentially depend on  $M_{n'}$ .

We now split  $M_{n'}$  further as

$$(10) \quad M_{n'} = \begin{pmatrix} M_k & D \\ D^* & E \end{pmatrix}$$

where  $M_k$  is the top left  $k \times k$  minor of  $M_{n'}$  (or of  $M_n$ ),  $D$  is a  $k \times (n' - k)$  matrix, and  $E$  is a  $(n' - k) \times (n' - k)$  matrix. We aim to bound the probability

$$(11) \quad \mathbf{P}(M_{n'}V' \in H)$$

(conditioning on  $B, C$  as mentioned above); any bound we obtain on this probability, multiplied by the previously mentioned entropy cost of  $O(p^{-1}n^{-d/2})^n$ , will provide a bound on  $\mathbf{P}(E_3'')$  by Fubini's theorem and the union bound.

To illustrate our ideas, let us first consider the toy case when  $H = \{0\}$  and the generators  $w_i, 1 \leq i \leq d$  are fixed, which would make  $V'$  deterministic. We split

$$(12) \quad V' = \begin{pmatrix} V'' \\ V''' \end{pmatrix}$$

where  $V''$  is the column vector  $(v_1, \dots, v_k)$ , and  $V'''$  is the column vector  $(v_{k+1}, \dots, v_{n'})$ . Expanding the condition  $M_{n'}V' \in H$  using (10) and (12) and extracting the lower  $n' - k$  entries of  $M_{n'}V'$ , we see that

$$D^*V'' = w$$

where  $w \in \mathbb{R}^{n'-k}$  is the vector  $w := -EV'''$ . If we condition  $M_k$  and  $E$  to be deterministic, then  $w$  becomes deterministic also, while each entry of  $D^*$  remains independent with distribution  $\xi$ , and by (6) each entry of  $D^*V''$  will match its corresponding entry of  $w$  with an independent probability  $O(n^{O(\varepsilon)}p)$ . Multiplying these probabilities and integrating out the conditioning, we obtain the bound

$$\mathbf{P}(M_{n'}V' \in H) \ll O(n^{O(\varepsilon)}p)^{n'-k}$$

for (11), which as mentioned earlier would give an upper bound for  $\mathbf{P}(E_3'')$  of the form

$$O(p^{-1}n^{-d/2})^n \times O(n^{O(\varepsilon)}p)^{n'-k}$$

which simplifies to

$$n^{-\frac{d}{2}n + O(\varepsilon n)} (1/p)^{k + (n - n')};$$

since  $k + (n - n') = O(\varepsilon n)$  and  $p \geq n^{-A}$ , this can be bounded by  $\exp(-cn)$  for some fixed  $c > 0$  independent of  $\varepsilon$ , with plenty of room to spare, if  $\varepsilon$  is chosen small enough.

There are two problems with this toy argument. Firstly, in general  $H$  is not  $\{0\}$ . However, we will be able to (morally) reduce to the  $H = \{0\}$  case by a linear projection argument that incurs a tolerable extra entropy loss (using the fact that the dimension of  $H$  is only  $O(n^{1-\varepsilon})$ ). The second problem is that the number of choices for the generators  $w_i$  is potentially infinite or even uncountable, so the entropy loss here is unacceptable. We are going to avoid this problem by not counting the number of  $w_i$ , but a finite set of representatives, which is linear algebraically equivalent.

We turn to the details. We split

$$V'_{(i)} = \begin{pmatrix} V''_{(i)} \\ V'''_{(i)} \end{pmatrix}$$

for  $i = 1, \dots, d$ , where  $V''_{(i)} \in \mathbb{R}^k$  and  $V'''_{(i)} \in \mathbb{R}^{n'-k}$ . Expanding the condition  $M_{n'}V' \in H$  using (10), (9) and extracting the bottom  $n-k$  coefficients, we conclude that

$$(13) \quad w_1(D^*V''_{(1)} + EV'''_{(1)}) + \dots + w_d(D^*V''_{(d)} + EV'''_{(d)}) \in H_1$$

where  $H_1 \subset \mathbb{R}^{n'-k}$  is the projection of  $H$  to  $\mathbb{R}^{n'-k}$ . With our current conditioning,  $H_1$  is a deterministic subspace of  $\mathbb{R}^{n'-k}$  of some dimension  $d_1 = O(n^{1-\varepsilon/4})$ . Meanwhile, from (6), and (9) we have

$$(14) \quad p_\xi(w_1V''_1 + \dots + w_dV''_d) \leq n^{C\varepsilon}p.$$

for some fixed constant  $C$  (independent of  $\varepsilon$ ).

We next reduce the space  $H_1$  to the trivial space  $\{0\}$ . Recall that  $d_1 = O(n^{1-\varepsilon/4})$  is the dimension of  $H_1$ . By permuting the indices if necessary, we may assume that<sup>1</sup>  $H_1$  is a graph over the last  $d_1$  coordinates of  $\mathbb{R}^{n'-k}$ . In other words, we may express  $H_1$  as

$$H_1 = \{(L(Y), Y) : Y \in \mathbb{R}^{d_1}\}$$

for some (deterministic) linear map  $L : \mathbb{R}^{d_1} \rightarrow \mathbb{R}^{n'-k-d_1}$ . Equivalently, we have

$$H_1 = \{(X, Y) \in \mathbb{R}^{n'-k-d_1} \times \mathbb{R}^{d_1} : \tilde{L}(X, Y) = 0\}$$

where  $\tilde{L} : \mathbb{R}^{n'-k-d_1} \times \mathbb{R}^{d_1} \rightarrow \mathbb{R}^{n'-k-d_1}$  is the map

$$\tilde{L}(X, Y) := X - L(Y).$$

If we identify  $\mathbb{R}^{n'-k-d_1}$  with  $\mathbb{R}^{n'-k-d_1} \times \{0\}$ , then  $\tilde{L}$  is the identity map on  $\mathbb{R}^{n'-k-d_1}$  and has  $H_1$  as its kernel. Applying  $\tilde{L}$  to (13), we obtain

$$(15) \quad w_1\tilde{L}(D^*V''_1 + EV'''_1) + \dots + w_d\tilde{L}(D^*V''_d + EV'''_d) = 0.$$

We now condition  $M_k, E$  to be fixed; the only remaining random variables are the entries of the  $k \times n' - k$  matrix  $D$ , which are iid with distribution  $\xi$ . Let  $E_4$  denote the event that (15) holds for a given choice of deterministic data  $(M_k, E, B, C, V_{(i)}, d, H, d_1, H_1)$ ; we suppress the dependence of  $E_4$  on this data. If we can obtain an upper bound on the conditional probability that  $E_4$  occurs, then by multiplying this bound by the previous entropy cost of  $O(p^{-1}n^{-d/2})^n$  would give an upper bound on  $\mathbf{P}(E_3)$ .

We still need to control  $\mathbf{P}(E_4)$ . Now that  $H$  has been eliminated, the most significant remaining difficulty is the lack of control of the quantities  $w_1, \dots, w_d$ , which at present are arbitrary real numbers and can thus take an uncountable number of possible values. To resolve this difficulty we again use the conditioning arguments of Komlós [4]. Given any  $m \times d$  matrix  $M$  for any  $m$ , we say that  $M$  has *good kernel* if the kernel  $\ker(M) := \{w \in \mathbb{R}^d : Mw = 0\}$  contains a tuple  $(w_1, \dots, w_d)$  obeying both (15) and (14). If we form the  $(n' - k - d_1) \times d$  random matrix  $U$  with

<sup>1</sup>This does not incur any entropy cost, as  $H_1$  was already deterministic under our current conditioning.

the vectors  $\tilde{L}(D^*V_1'' + EV_1'''), \dots, \tilde{L}(D^*V_d'' + EV_d''')$  as columns, then clearly  $E_4$  is contained in the event that  $U$  has good kernel.

Trivially,  $U$  has rank at most  $d$ . As a consequence, one can select  $d$  rows from  $U$  whose row span is the same as that of  $U$ , or equivalently that the corresponding  $d \times d$  minor of  $U$  has the same kernel as that of  $U$ . The number of possible ways to select these rows is  $\binom{n'-k-d_1}{d}$ , which we crudely bound by  $n^d$ . By paying an entropy cost of  $n^d$  for the purposes of bounding  $\mathbf{P}(E_4)$ , we may thus assume that these row positions are deterministic, thus there are deterministic  $1 \leq l_1 < \dots < l_d \leq n' - k - d_1$ , and we need to bound the event that the  $d \times d$  minor  $U_d$  formed by the  $l_1, \dots, l_d$  rows of  $U$  has a good kernel.

We now condition on the  $l_1, \dots, l_d$  rows of the  $n' - k \times k$  matrix  $D^*$ , as well as the last  $d_1$  rows of the same matrix  $D^*$ , thus leaving at least  $n' - k - d_1 - d$  of the first  $n' - k - d_1$  rows of  $D^*$  random (with entries independently distributed with law  $\xi$ ). As  $\tilde{L}$  is the identity on  $\mathbb{R}^{n'-k-d_1}$ , we see that the  $l_1, \dots, l_d$  entries of  $\tilde{L}(D^*V_1'' + EV_1'''), \dots, \tilde{L}(D^*V_d'' + EV_d''')$  are now deterministic (they do not depend on the remaining random rows of  $D^*$ ). In other words, the minor  $U_d$  is now deterministic. If  $U_d$  does not have a good kernel, its contribution to  $\mathbf{P}(E_4)$  is zero. If instead  $U_d$  has a good kernel, then we may find a *deterministic* choice of  $w_1, \dots, w_d$  in the kernel of  $U_d$  that obeys both (15) and (14).

The rest of the calculation is similar to the toy case. Consider the  $i^{\text{th}}$  component of the vector equation (15) for this deterministic choice of  $w_1, \dots, w_d$ , where  $1 \leq i \leq n' - k - d''$  is not equal to any of the  $l_1, \dots, l_d$ . We can rewrite this component as

$$(16) \quad (w_1V_1'' + \dots + w_dV_d'') \cdot R_i = x_i$$

where  $R_i \in \mathbb{R}^k$  is the  $i^{\text{th}}$  row of  $D^*$ , and  $x_i \in \mathbb{R}$  is a deterministic quantity that does not depend on the remaining random rows in  $D^*$ . By (14), for each such  $i$ , the equation (16) holds with an independent probability of at most  $n^{C\varepsilon}p$ , and so the probability that (15) holds in full is at most  $O(n^{C\varepsilon}p)^{n'-k-d_1-d}$ . Taking into account the entropy cost of  $n^d$  mentioned earlier, we thus have

$$\mathbf{P}(E_4) \leq n^d \times O(n^{C\varepsilon}p)^{n'-k-d_1-d} \leq n^{O(\varepsilon n)}p^{n'-k-d_1-d}.$$

Paying the previously mentioned entropy cost of  $O(p^{-1}n^{-d/2})^n$ , we then have

$$\mathbf{P}(E_3'') \leq O(p^{-1}n^{-d/2})^n \times n^{O(\varepsilon n)}p^{n'-k-d_1-d} \leq n^{-\frac{d}{2}n+O(\varepsilon n)}(1/p)^{n-n'+k+d_1+d}.$$

Since  $p \geq n^{-A}$ ,  $d \geq 1$ , and  $n - n' + k + d_1 + d = O(\varepsilon n)$ , we conclude that

$$\mathbf{P}(E_3'') \ll \exp(-cn)$$

for some fixed  $c > 0$  independent of  $\varepsilon$  (with plenty of room to spare), if  $\varepsilon$  is small enough. Combining this with (8) we obtain (4). This concludes the proof of (3), and Theorem 1.3 follows.

## 5. CONCLUDING REMARKS

The assumption that the upper diagonal entries  $\xi_{ij}$ ,  $1 \leq i < j \leq n$  are iid is not essential; an inspection of the argument reveals that the proof continues to work if we assume that the  $\xi_{ij}$  are independent and there is a constant  $\mu > 0$  such that  $\mathbf{P}(\xi_{ij} = x) \leq 1 - \mu$  for any  $1 \leq i < j \leq n$  and  $x \in \mathbb{R}$ . Our main tool, Theorem 3.2, holds under this assumption; see [6]. The argument also easily extends to Hermitian random matrix models, in which the coefficients  $\xi_{ij}$  for  $i < j$  are allowed to be complex, and one imposes the condition  $\xi_{ji} = \overline{\xi_{ij}}$ . In other words, the above arguments can extend to show the following result:

**Theorem 5.1.** *For any fixed  $A, \mu \geq 0$  and sufficiently large  $n$  the following holds. Let  $\xi_{ij}, 1 \leq i < j \leq n$  be independent (complex or real) random variables such that  $\mathbf{P}(\xi_{ij} = x) \leq 1 - \mu$  for any  $1 \leq i < j \leq n$  and  $x \in \mathbb{R}$ . Let  $\xi_{ii}, 1 \leq i \leq n$  be real random variables that are independent of the  $\xi_{ij}, 1 \leq i < j \leq n$ . Set  $\xi_{ji} = \overline{\xi_{ij}}$  for  $1 \leq i < j \leq n$ . Then the spectrum of the matrix  $(\xi_{ij})_{1 \leq i < j \leq n}$  is simple with probability at least  $1 - n^{-A}$ .*

## REFERENCES

- [1] L. Babai, D. Grigoryev and D. Mount, Isomorphism of graphs with bounded eigenvalue multiplicity, *Proceedings of the 14th Annual ACM Symposium on Theory of Computing*, 310–324 (1982).
- [2] L. Babai, Private conversation, May 2012.
- [3] L. Erdős, A. Knowles, H.-T. Yau, J. Yin, Spectral statistics of Erdős-Rényi Graphs II: Eigenvalue spacing and the extreme eigenvalues, *Comm. Math. Phys.* **314** (2012), no. 3, 587–640.
- [4] J. Komlós, On the determinant of  $(0, 1)$  matrices, *Studia Sci. Math. Hungar.* **2** (1967), 7–21.
- [5] H. Nguyen and V. Vu, Small Ball Probability, Inverse Theorems, and Applications, *Erdős Centennial Proceeding*, Eds. L. Lovász et. al., Springer 2013.
- [6] H. Nguyen and V. Vu, Optimal Littlewood-Offord theorems, *Advances in Math.*, Vol. 226 6 (2011), 5298-5319.
- [7] T. Tao and V. Vu, *Additive Combinatorics*, Cambridge University Press, 2006.
- [8] T. Tao and V. Vu, Inverse Littlewood-Offord theorems and the condition number of random matrices, *Annals of Mathematics* (2) 169 (2009), no 2, 595-632.
- [9] T. Tao and V. Vu, A sharp inverse Littlewood-Offord theorem, *Random Structures Algorithms* 37 (2010), no. 4, 525-539.

DEPARTMENT OF MATHEMATICS, UCLA, LOS ANGELES CA 90095-1555

*E-mail address:* tao@math.ucla.edu

DEPARTMENT OF MATHEMATICS, YALE UNIVERSITY, NEW HAVEN 06520

*E-mail address:* van.vu@yale.edu