THE SPARSE CIRCULAR LAW, REVISITED

ASHWIN SAH, JULIAN SAHASRABUDHE, AND MEHTAAB SAWHNEY

Abstract. Let A_n be an $n \times n$ matrix with iid entries distributed as Bernoulli random variables with parameter $p = p_n$. Rudelson and Tikhomirov, in a beautiful and celebrated paper, show that the distribution of eigenvalues of $A_n \cdot (pn)^{-1/2}$ is approximately uniform on the unit disk as $n \to \infty$ as long as $pn \to \infty$, which is the natural necessary condition.

In this paper we give a much simpler proof of this result, in its full generality, using a perspective we developed in our recent proof of the existence of the limiting spectral law when pn is bounded. One feature of our proof is that it avoids the use of ε-nets entirely and, instead, proceeds by studying the evolution of the singular values of the shifted matrices $A_n - zI$ as we incrementally expose the randomness in the matrix.

1. Introduction

For an $n \times n$ matrix M, we define the *spectral distribution* of M to be the probability measure μ_M that puts a point mass of equal weight on each eigenvalue of M:

$$\mu_M = n^{-1} \sum \delta_\lambda.$$

The study of the spectral distribution of *random* matrices M goes back to the seminal work of Wigner [19] in the 1950s who showed that the spectral distribution of random *symmetric* matrices (so called *Wigner* matrices) converges to the, so-called, semi-circular law in the large n limit, after an appropriate rescaling.

Determining the limiting spectral distribution for matrices with *iid* entries proved to be substantially more difficult and was only resolved by Tao and Vu [18] after a long succession of important papers going back to the 1960s [1, 4, 6, 8, 12, 13, 17]. They showed that the spectral measure for such matrices, after appropriate rescaling, tends to the *circular law* as $n \to \infty$, which is the probability measure that is uniform on the unit disc in $\mathbb{C}$.

Theorem 1.1 (Tao and Vu)**.** *Let ξ be a complex random variable with mean* 0 *and variance* 1*. For each n, let A_n be a random matrix with iid entries distributed as ξ. If we put $A_n^* = A_n \cdot n^{-1/2}$ then the spectral measure $\mu_{A_n^*}$ converges to the circular law in probability.*

Actually Tao and Vu also proved Theorem 1.1 for the stronger notion of almost sure convergence but we express their theorem in terms of convergence *in probability* as we will be exclusively interested in this form of convergence. Indeed, we say a sequence of measures μ_n converges to the circular law *in probability* if for all $s, t \in \mathbb{R}$, we have

$$\lim_{n\to\infty} \mu_n((-\infty, s) \times (-\infty, t)) = \frac{1}{\pi} \int_{-\infty}^{s} \int_{-\infty}^{t} \mathbb{1}_{x^2+y^2 \leqslant 1} \, dydx.$$

While Theorem 1.1 gives us a very good understanding of the limiting spectral law of the spectrum of "dense" matrices, it does not tell us anything about matrices where the non-zero entries are sparse. Of particular interest are *iid Bernoulli matrices* where all entries are iid and distributed as[1] $\mathrm{Ber}(p)$ for $p = p_n \to 0$.

Sah was supported by the PD Soros Fellowship. Sah and Sawhney were supported by NSF Graduate Research Fellowship Program DGE-2141064. Part of this work was conducted when Sawhney was visiting Cambridge with support from the Churchill Scholarship.

[1]Here $\mathrm{Ber}(p)$ denotes, as is standard, a $\{0, 1\}$-Bernoulli random variable taking 1 with probability p.

This sparse setting was considered by Götze and Tikhomirov [8], who proved that the limiting spectral distribution is still the circular law when $p > n^{-1/4+\varepsilon}$. Tao and Vu [17] improved this range to cover all $p > n^{-1+\varepsilon}$, and Basak and Rudelson further improved this to cover all $p > \omega(n^{-1}(\log n)^2)$.

Then, in a difficult and celebrated paper, Rudelson and Tikhomirov [15] proved that $pn \to \infty$ is sufficient for convergence to the circular law, which is also the natural necessary condition: if pn is bounded it is easy to see that the limiting measure must have a large atom at zero.

In fact, they prove a more general result that allows for the non-zero entries to be replaced with iid copies of a random variable ξ of variance 1. Indeed let $A_n \sim \Delta_n(p,\xi)$ indicate that A_n is a $n \times n$ random matrix where each entry is an iid copy of $\mathrm{Ber}(p) \cdot \xi$.

Theorem 1.2 (Rudelson and Tikhomirov). *Let ξ be a real random variable with $\mathbb{E}\,\xi^2 = 1$, let $np \to \infty$ and $p \to 0$ and for each n, let $A_n \sim \Delta_n(\xi, p)$. If we put $A_n^* = A_n \cdot (pn)^{-1/2}$ then $\mu_{A_n^*}$ converges to the circular law in probability.*

This long line of results leaves open the case of $p = d/n$, for constant $d > 0$, which has proven to be the most difficult and subtle case. In our paper [16], we complete this program by proving the existence of the limiting measure in this case.

Theorem 1.3. *For $d > 0$ and each n, let A_n be an $n \times n$ matrix with iid entries distributed as $\mathrm{Ber}(d/n)$. There exists a distribution μ_d on $\mathbb{C}$ so that μ_{A_n} converges to μ_d, in probability.*

In this paper we give a new and considerably shorter proof of Theorem 1.2 based on the method which we used in [16] to prove Theorem 1.3.

The method and "philosophy" of our proof is considerably different from that of Rudelson and Tikhomirov. We completely avoid any direct use of ε-nets and, instead, favour of a more "dynamic" approach, where we track the evolution of the point processes defined by the singular values of the shifted matrices $A - zI$ as we expose a new rows and columns. To pull this analysis off we need only to rely on a few "quasi-randomness" conditions on the graph defined by the non-zero entries.

One added advantage of our approach is that it allows us to effortlessly generalize Theorem 1.2 of Rudelson and Tikhomirov to allow for complex random variables ξ with unit variance. It appears the work of Rudelson and Tikhomirov would not generalize to this complex case without significant new ideas. Thus our main theorem here is the following.

Theorem 1.4. *Let ξ be a complex random variable with $\mathbb{E}\,|\xi|^2 = 1$, $np \to \infty$ and $p \to 0$ and for each n, let $A = A_n \sim \Delta_n(\xi, p)$. If we put $A_n^* = A \cdot (pn)^{-1/2}$ then $\mu_{A_n^*}$ converges to the circular law, in probability.*

We now turn to sketch the proof of Theorem 1.4 and setup the remainder of the paper.

2. Description of method

To establish the convergence of the spectral law to the circular law it is enough to prove the convergence of the logarithmic potential of the spectral law to the logarithmic potential of the circular law. For us, we may use Girko's "hermitization" method (see e.g. [2]) to express the logarithmic potential of the spectral law as the (random) function

$$U_n(z) = -\frac{1}{n}\sum_{j=1}^{n} \log\big(\sigma_j\big(A_z^*\big)\big), \tag{1}$$

where we set $d = pn$, here and throughout the paper, and define

$$A_z^* = A^* - zI_n = d^{-1/2}A - zI_n.$$

Here we have also used the notation $\sigma_1(M) \geqslant \sigma_2(M) \geqslant \cdots \geqslant \sigma_m(M)$, to denote the (right) singular values of the $n \times m$ matrix M. Our main task is to prove that for all $z \in \mathbb{C} \setminus \{0\}$ we have

$$\lim_n U_n(z) = U^\circ(z) = \begin{cases} -\log|z|, & \textit{if } |z| \geqslant 1; \\ (1-|z|^2)/2, & \textit{if } |z| \leqslant 1, \end{cases} \tag{2}$$

with probability 1, where U° is the log potential of the circular law. Once we have proved this we can simply appeal to the general theory of logarithmic potentials to conclude the convergence in probability $\mu_n \to \mu^\circ$.

Given the expression (1), there are traditionally two steps to establish (2). First, one shows that the limit of the empirical distribution of the *singular values* $\sigma_j(A_z)$ is "what it should be" by a (fairly standard) method of moments and truncation argument: this tells us that the "bulk" of the sum in (1) converges to what it is supposed to. Second, one shows that the small singular values $\sigma_n, \sigma_{n-1}, \ldots$ don't spoil the bulk convergence of this sum by getting too small and dominating the sum (1). Since we have good tools for establishing the first step these days, it is this second step that represents the core challenge and, in particular, one essentially needs to prove bounds of the type

$$\mathbb{P}\big(\sigma_{n-k}(A_z^*) \leqslant \exp(-\varepsilon n/k)\big) = o(1), \tag{3}$$

for any $\varepsilon > 0$ and all $k = 1, 2, \ldots$.

Now, *heuristically* we expect that typically $\sigma_{n-k} = \Theta(kd^{1/2}n^{-1})$, and thus (3) may not appear to be a particularly difficult obstacle to overcome, as it represents an extremely abnormal behaviour, heuristically. However, obtaining bounds of this type has recently represented *the* significant challenge in this area. Indeed, bounds of this type represent one of the main achievements of the work Tao and Vu [18] in their work on the circular law for dense matrices. For sparse matrices, the challenge is greater still as there is less "randomness" to use. For their sparse circular law, Rudelson and Tikhomirov [15], develop a whole toolbox of sophisticated techniques for constructing ε-nets to prove singular value estimates of the type[2] (3).

In this paper, we take a more direct route to proving (2) that is substantially different and considerably simpler. Instead of directly working with the singular values, we look to *compare* U_n with a "truncated" version of the log potential of a principal minor of A^*. To elaborate on this, let $\varepsilon \to 0$ sufficiently slowly and set $m = (1-\varepsilon)n$. We will understand $A_{m,z}^*$ to be the top left $m \times m$ principal minor of A_z^*. The main objective of the proof will be to compare $U_n(z)$ with the *truncated* sum

$$T_n(z) = -\frac{1}{n} \sum_{j=1}^{(1-\varepsilon/4)m} \log\big(\sigma_j\big(A_{m,z}^*\big)\big).$$

The point here is that the sum $T(z)$ is much easier to deal with than the corresponding log potential: the smallest $\varepsilon m/4$ singular values have been removed from the sum. Thus a simple variant of the trace moment method is sufficient to establish

$$T_n(z) = U^\circ(z) + o(1)$$

with high probability as $n \to \infty$. The core of the proof, therefore, lies in making the comparison between U_n and T_n, and in particular showing $U_n(z) \leqslant T_n(z) + o(1)$, which we achieve dynamically: we "build up" $A_{n,z}^*$ from $A_{m,z}^*$ by alternately adding rows and columns. Simultaneously, we build up U_n from T_n by taking more singular values into our sum, when possible. For this, we index time in *half-integer* steps $t \in [m, n]$ so that at each integer time t we have that A_t^* is

[2]In fact, both Tao and Vu [18] well as Rudelson and Tikhomirov [15] obtain better estimates on the least singular value than required, but this is not relevant to our discussion here.

a $t \times t$ matrix and at time $t + 1/2$, we define $A^*_{t+1/2}$ to be the $t \times (t+1)$ matrix which is A^*_{t+1} with the bottom-most row deleted. Thus we build $A^*_{n,z}$ from $A^*_{m,z}$ as

$$A^*_{m,z} \to A^*_{m+1/2,z} \to \cdots \to A^*_{n-1/2,z} \to A^*_{n,z}.$$

The point of this is that each row and column addition has the crucial property that it "pushes" all the singular values up and thus we aim to take more singular values into our sum as these values get "pushed" throughout the process. Let us define the sum

$$T_{r,t}(z) = -\frac{1}{n}\sum_{j=1}^{r} \log\big(\sigma_j(A^*_{t,z})\big).$$

Our crucial lemma tells us that we can take on singular values rather often: if $A_{t,0}$ satisfies some quasi-randomness conditions and $r < t$ then

$$\mathbb{P}\big(T_{r+1,t+1/2}(z) \leqslant T_{r,t}(z) + \delta_{r+1}\big) = 1 - o_{d\to\infty}(1), \tag{4}$$

where the probability is only over the new row/column being added and δ_r is a sequence with the property that

$$\sum_{r=(1-\varepsilon/4)m}^{n} \delta_r = o(1). \tag{5}$$

However this is not the end of the story; every time we add a column to our matrix we "create" a new singular value at 0 that needs to be taken under control in the following steps. Thus we need to ensure that our random process has sufficient drift to ensure that we can take on all new singular values by the end of the process.

To see this is the case, consider the random process $r(t)$, which is defined as the number of singular values that we have taken on by time t. It makes sense to consider the "height" of this process defined by $h(t) = t - r(t)$, which represents the number of singular values we don't have in our possesion at time t. Thus the goal of our work can be phrased as showing that $h(n) = 0$ with high probability. Note that at the start of this process, we have

$$h(m) = m - (1 - \varepsilon/4)m = \varepsilon m/4$$

and at each step we have a downward drift of $(1/2 - o(1))$ with each row or column addition. Since there are $2\varepsilon n$ row or column additions we expect the total drift to be εn which is significantly larger than $\varepsilon n/4$, and should give $h(n) = 0$ whp. Of course the wrinkle is that we need to ensure that these quasi-randomness conditions, mentioned above, occur sufficiently often.

To prove (4) we are led to study the structure of the vector space spanned by the $\lceil t \rceil - r$ smallest singular vectors of M (left-singular vectors for integral t and right-singular vectors for half-integral t). In particular, we will need to show that if X is a new row or column of our matrix and $P_{h,M}$ is the orthogonal projection onto the space spanned by the h smallest singular vectors of the appropriate side then for "quasi-random" M we have

$$\mathbb{P}_X\big(\|P_{\lceil t\rceil - r,M}X\|_2 < \exp(-n\delta_r)\big) = O(\varepsilon + (\log_{(2)} d)^{-1/2}). \tag{6}$$

While theorems of this type can be quite challenging, here we can get away with a very weak notion of quasi-randomness, based on the number of rows that have a unique non-zero entry in a set of columns. From this quasi-randomness condition, we can deduce the following basic structural information about vectors $v \in \mathbb{C}^t$ that are near-singular vectors of M:

$$|\{i\colon v_i \geqslant \exp(-n\delta_r)n^{-1/2}\}| \geqslant (n/(2d))\log_{(2)} d. \tag{7}$$

This in, in turn, allows us to deduce (6).

In Section 4 we define the notion of a "unique neighbourhood expansion" which our quasi-randomness notion is based on. In Section 5 we use this quasi-randomness property to derive (7).

In Section 6 we prove that if the kernel has the property (7) then we can deduce (6). Then in Section 7 we prove (4), showing how the process evolves in a single step. In Section 8 we give the simple analysis of this random walk. In Section 9 we show the convergence of T_n to U° and complete the proof of Theorem 1.4.

3. A few preliminaries

In what follows we fix ξ to be a complex random variable with

$$\mathbb{E}\,|\xi|^2 = 1.$$

We regard ξ as fixed throughout the paper and allow various quantities to depend on ξ. Given such a random variable ξ, we define $\beta = \beta(\xi) \leqslant 1$ to be such that

$$\max_y \mathbb{P}\big(\|\xi - y\| < \beta\big) \leqslant 1 - \beta.$$

Note that either $\beta(\xi) > 0$ or $\beta(\mathrm{Ber}(1/2)\cdot\xi) > 0$ and thus adjusting p and rescaling if necessary we may assume that $\beta(\xi) > 0$.

We define $\Delta_{n,m}(p,\xi)$ to be the probability space of all $n\times m$ matrices of A where $A_{i,j} = \delta_{i,j}\xi_{i,j}$, all of the $\xi_{i,j}$ and $\delta_{i,j}$ are independent, and $\delta_{i,j} \sim \mathrm{Ber}(p)$ and $\xi_{i,j}$ is distributed as ξ. We define $\Delta_n(\xi,p) = \Delta_{n,n}(\xi,p)$. We define $\mathrm{Col}_n(\xi,p)$ to be the random column vector $X \in \mathbb{R}^n$ where X_i are independent and distributed as $\mathrm{Ber}(p)\cdot\xi$. We define $\mathrm{Row}_n(\xi,p)$ similarly. For a matrix M we also define $\|M\|_{\mathrm{HS}}^2 = \sum_{i,j}|M_{i,j}|^2$.

We define for easy reference the quantities δ_r mentioned in the proof outline. We define

$$\delta_r = \begin{cases} n^{-1}(\log(n/(n-r+1)))^2 & \text{for } r < n(1-d^{-1/4}); \\ Cn^{-1}(\log d)^8(\log(n/(n-r+1)))^8 & \text{for } r \geqslant n(1-d^{-1/4}). \end{cases}$$

As we required in (5) it is not hard to check that $\sum_{r=(1-\varepsilon/4)n}^{n}\delta_r = o(1)$. To save on clutter, it also makes sense to define

$$\eta_r = \exp(-n\delta_r).$$

In addition to the fixing of ξ, we make a few more global assumptions throughout the paper. We assume throughout n is sufficiently large and that $Cn^{-1} \leqslant p \leqslant 1/2$, where C is a sufficiently large constant depending only on ξ. We assume $t \in \frac{1}{2}\mathbb{Z}\cap[m,n]$ and $m = (1-\varepsilon)n$. Throughout, for $t\in\mathbb{N}$, A_t will be our $t\times t$ random matrix for time t and $A_{t+1/2}$ will be a $t\times(t+1)$ random matrix, obtained as submatrices of A. Throughout we will define $A_{t,z} = A_t - zI$, where I is either a square identity or $(t-1/2)\times(t+1/2)$ identity matrix depending if we are at an integer or half-integer time. As in the previous section, define $A^*_{t,z} = d^{-1/2}A_t - zI$. We also allow ourselves the convention that $d = pn$ and that $z\in\mathbb{C}$. All whp statements are meant with respect to $n\to\infty$ and therefore $d\to\infty$.

4. Unique neighbourhood expansions and quasi-randomness properties

The goal of this section is to define the quasi-randomness event $\mathcal{E}_r$ that will allow us to show that we can take on the rth singular value into our sum, with high probability, assuming $r \leqslant \lfloor t\rfloor$. We define this event $\mathcal{E}_r$ in three parts

$$\mathcal{E}_r(\xi,p) = \mathcal{U}_r(\xi,p)\cap\mathcal{B}(p)\cap\mathcal{Q}(\xi,p)\cap\mathcal{R}(\xi,p).$$

Even before defining $\mathcal{E}_r$ we state the main lemma of this section which is essentially the only thing we need to carry forward in the paper.

Lemma 4.1. *If $r \geqslant t - n/d^{1/4}$ then*

$$\mathbb{P}(A_t \in \mathcal{E}_r) \geqslant 1 - \exp(-d^{1/2}(\lceil t\rceil - r + 1)).$$

We now define $\mathcal{U}_r$. For this, let B be an $m \times \ell$ matrix and let $S \subseteq [\ell]$ be a set of columns of B. We define $U(S) \subseteq [m]$, a subset of the rows of B, in two parts. We first define

$$U(S) \setminus S = \big\{ i \in [m] \setminus S \colon B_{i,j} \neq 0 \text{ for a unique } j \in S \text{ and } |B_{i,j}| \geqslant \beta \big\}.$$

We then define

$$U(S) \cap S = \big\{ i \in [m] \cap S \colon B_{i,j} = 0 \text{ for all } j \in S \big\}.$$

We fix $\alpha(x) = (\log(n/x))^{-2}$ and say $B \in \mathcal{U}_r$ if for all subsets of columns S with

$$c^*(\lceil t \rceil - r + 1) \leqslant |S| \leqslant (n/(2d)) \log_{(2)} d \quad \text{ we have } \quad |U(S)| \geqslant \alpha(|S|) d |S|,$$

where $c^* > 0$ is the absolute constant appearing in the statement of Lemma 6.3. (It is not heuristically important what this parameter is precisely, only the scale of the lower limit on $|S|$.)

Lemma 4.2. *If $r \geqslant t - n/d^{1/4}$ then*

$$\mathbb{P}(A_t \in \mathcal{U}_r) \geqslant 1 - \exp(-2d^{1/2}(\lceil t \rceil - r + 1)).$$

Now define $A_t \in \mathcal{B}$ if for all subsets $S \subseteq [\lfloor t \rfloor]$ of the rows, we have

$$\frac{1}{|S|} \sum_{i \in S} \sum_{j=1}^{\lceil t \rceil} \delta_{i,j} = O(d + \log(n/|S|)),$$

and analogously for the columns.

Furthermore, define $A_t \in \mathcal{Q}$ if there are at most

$$2dn/H^2 + (\log n)^2 \text{ entries of } A_t \text{ which are } > 8H/\beta \text{ in magnitude}$$

for each $H \in [1, n^4]$, and additionally the maximum value of A_t is at most n^3.

Finally define $A_t \in \mathcal{R}$ if for all $\ell \geqslant 1$, if we put $L = (dn/(\beta \ell))^5$ then

$$\Big| \Big\{ i \colon \sum_{j=1}^{\lceil t \rceil} |(A_t)_{i,j}| > L/\beta \Big\} \Big| \leqslant \alpha(\ell) d\ell/4$$

and similarly for columns.

Lemma 4.3. *We have*

$$\mathbb{P}(A_t \in \mathcal{B} \cap \mathcal{Q} \cap \mathcal{R}) \geqslant 1 - n^{-3}.$$

The proof of all of these lemmas are entirely standard union bound computations and are deferred to Appendix A.

5. Spreadness of near-kernel vectors from graph quasi-randomness

The goal of this section is to prove the following lemma, which says that if $A_t \in \mathcal{E}_r$ and v is "close" to the small right-singular vectors of $A_{t,z}$ then it has $(n/(2d)) \log_{(2)} d \gg n/d$ coordinates that are bounded away from zero by $\eta_r n^{-1/2}$. The point here is that each new row and column has random support of size d, on average. So, crucially, the intersection of these supports is typically $\gg 1$. For this lemma we define the notation, for $v \in \mathbb{C}^n$ and $x \geqslant 0$, $\lambda(v; x) = |\{i \colon |v_i| \geqslant x\}|$.

Lemma 5.1. *For $t \geqslant r \geqslant n(1 - d^{-1/4})$, let $A_t \in \mathcal{E}_r$ and $1 \leqslant |z| \leqslant d$. Put $k = c^*(\lceil t \rceil - r + 1)$, where c^* is as in Section 4, and let $v \in \mathbb{C}^{\lceil t \rceil}$ satisfy*

$$\|A_{t,z} v\|_2 \leqslant d^{1/2} \eta_r \quad \text{and} \quad \lambda(v; k^2 n^{-5/2}) \geqslant k.$$

Then

$$\lambda(v; \eta_r d^{1/2} k^{-1/2}) \geqslant (n/(2d)) \log_{(2)} d.$$

We prove this lemma by using the unique neighbourhood expansions. This link starts to become apparent with the following simple observation.

For this we define a few uses of notation. For $v \in \mathbb{C}^n$ and $S \subseteq [n]$, we let $v_S \in \mathbb{C}^n$ be the vector with $(v_S)_i = v_i$ for $i \in S$ and 0 otherwise. It is also useful to define, for a vector $v = (v_1, \dots, v_n)$, the vector $v^* = (v_1^*, \dots, v_n^*)$ where the entries of v^* are the $|v_i|$, but have been permuted so that $v_1^* \geqslant \cdots \geqslant v_n^*$.

Observation 5.2. *For $\ell \leqslant t$, let $v \in \mathbb{C}^{\lceil t \rceil}$ and let S be the set of the ℓ largest coordinates of v in absolute value. If $|z| \geqslant 1$ then $|(A_{t,z}v_S)_i| \geqslant v_\ell^* \beta$ for all $i \in U(S)$.*

Proof. We consider two cases. If $i \in U(S) \setminus S$ there is unique $j \in S$ with $(A_{t,z})_{i,j} \neq 0$. For this j, we additionally have $|\xi_{i,j}| \geqslant \beta$. Thus the observation is easily proved in this case.

For $i \in U(S) \cap S$ we have $(A_{t,z})_{i,j} = 0$ for all $j \in S$ with $j \neq i$. So

$$|(A_{t,z}v_S)_i| = |(A_{t,z})_{i,i}v_i| = |(A - zI)_{i,i}v_i| = |z||v_i| \geqslant v_\ell^* \beta,$$

which proves the observation. ☐

Recall that above we defined the function $\alpha(x) = (\log(n/x))^{-2}$. Here we define the function

$$g(x) = \left\lceil \frac{d\alpha(x)x}{C'(d + \log(n/x))} \right\rceil, \tag{8}$$

where C' is a sufficiently large constant. (It is chosen based on the $O(\cdot)$ in the definition of the event $\mathcal{B}$.)

We prove Lemma 5.1 by iterating the following lemma. This lemma says that if the mass of v is clustered on fewer than $(n/(2d))\log_{(2)} d$ coordinates then $A_{t,z}v$ has many large coordinates and is therefore is not close to the small singular vectors.

Lemma 5.3. *For $r \leqslant t$ let $A_t \in \mathcal{E}_r$ and let $c^*(\lceil t \rceil - r + 1) \leqslant \ell \leqslant (n/(2d))\log_{(2)} d$, where c^* is as in Section 4. If $v \in \mathbb{C}^{\lceil t \rceil}$ satisfies*

$$v_{\ell+g(\ell)}^* \leqslant v_\ell^* \left(\frac{\beta \ell}{dn}\right)^7 .$$

and $1 \leqslant |z| \leqslant d$ then

$$\lambda\big(A_{t,z}v\,;\, \beta v_\ell^*/2\big) \geqslant \alpha(\ell)d\ell. \tag{9}$$

Proof. Let S be the set of the ℓ largest coordinates of v in absolute value. Write $v = x + y$, where $x = v_S$ and $y = v_{S^c}$ and note that by Observation 5.2 we have

$$|(A_{n,z}x)_i| \geqslant |(A_{n,z}x)_i| - |(A_{n,z}y)_i| \geqslant v_\ell^*\beta - |(A_{n,z}y)_i|,$$

for all $i \in U(S)$. Since $A_t \in \mathcal{E}_r$ we have that $|U(S)| \geqslant \alpha(\ell)d\ell$ and thus it suffices to prove

$$\big|\{i \in U(S) \colon |(A_{n,z}y)_i| > \beta v_\ell^*/2\}\big| < \alpha(\ell)d\ell/2. \tag{10}$$

To prove (10), let B be the set in (10) and define $S^* \supseteq S$ to be the set of indices of the $\ell + g(\ell) - 1$ largest coordinates of v in magnitude. If $i \in B$ then either row i has unusually large magnitude *or* row i has a non-zero entry in $S^* \setminus S$. More precisely, define

$$B_1 = \left\{ i \colon \sum_{j=1}^{\lceil t \rceil} |\delta_{i,j}\xi_{i,j}| \geqslant L/\beta \right\} \quad \text{and} \quad B_2 = \left\{ i \colon \sum_{j \in S^* \setminus S} \delta_{i,j} > 0 \right\},$$

where we have set $L = (dn/(\beta\ell))^5$. We now claim that $B \subseteq B_1 \cup B_2$. To see this, assume $i \notin B_1 \cup B_2$ and observe that

$$|(A_{n,z}y)_i| \leqslant \sum_{j \notin S} |\delta_{i,j}\xi_{i,j} - z\mathbb{1}_{j=i}||v_j| = \sum_{j \notin S^*} |\delta_{i,j}\xi_{i,j} - z\mathbb{1}_{j=i}||v_j|,$$

since $i \notin B_2$. Using that $|v_j| \leqslant v^*_\ell L^{-1}(\beta\ell/(dn))^2$ for $j \notin S^*$, we see the above is at most

$$\left(\frac{\beta\ell}{dn}\right)^2 \cdot \frac{v^*_\ell}{L} \cdot \left(d + \sum_{j\notin S^*} |\delta_{i,j}\xi_{i,j}|\right) \leqslant \left(\frac{\beta}{d}\right)^2 \cdot \frac{v^*_\ell}{L} \cdot (d + \beta^{-1}L) \leqslant \beta v^*_\ell/2,$$

where we have used that $|z| \leqslant d$, $\ell \leqslant n$, and $i \notin B_1$. Thus $B \subseteq B_1 \cup B_2$.

To conclude (10), we just need to show $|B_1|, |B_2| < \alpha(\ell)d\ell/4$. Since $A_t \in \mathcal{E}_r$, the definition of event $\mathcal{B}$ tells us

$$|B_2| \leqslant \sum_{i\in S^*\setminus S} \sum_j \delta_{i,j} \leqslant O(g(\ell)(d + \log(n/g(\ell)))) \leqslant \alpha(\ell)d\ell/4.$$

Recall that in the definition of g, C' is chosen sufficiently large. Additionally, the definition of event $\mathcal{R}$ tells us $|B_1| \leqslant \alpha(\ell)d\ell/4$. This completes the proof. □

We now iterate Lemma 5.3 to obtain Lemma 5.1. To understand how many times we need to iterate Lemma 5.3, we need the following basic numerical fact. For this, we think of k as in Lemma 5.1 and we define the sequence $(k_t)_{t\geqslant 0}$, by setting $k_0 = k$ and then defining

$$k_i = k_{i-1} + g(k_{i-1}),$$

for all $i \geqslant 1$.

Fact 5.4. *Let τ be the minimum value for which $k_\tau > (n/(2d))\log_{(2)} d$. Then $\tau = O((\log(n/k))^4)$.*

We prove this fact in Appendix A and now jump to the proof of Lemma 5.1.

Proof of Lemma 5.1. We let k_t and τ be as above. We claim that for all $i \in [\tau]$ we have

$$v^*_{k_i} \geqslant v^*_{k_{i-1}} \cdot \delta, \tag{11}$$

where $\delta = (\beta k/(dn))^7$. For a contradiction assume $i \in [\tau]$ is the smallest failure of this inequality. We will then apply Lemma 5.3 to show that this contradicts the assumption that v is close to the small singular vectors.

Indeed, the failure of (11) allows us to apply Lemma 5.3 to v to learn

$$\lambda\big(A_{t,z}v\,;\,\beta v^*_{k_{i-1}}/2\big) \geqslant \alpha(k_{i-1})dk_{i-1} \geqslant k(\log(n/k))^{-2}. \tag{12}$$

Since i is the minimum such value for which (11) fails, we have

$$v^*_{k_{t-1}} \geqslant \delta^{7(i-1)}v^*_{k_0} \geqslant \delta^{7(i-1)}k^2n^{-5/2}, \tag{13}$$

where the last inequality holds by the given lower bound on $v^*_{k_0} = v^*_k$. Now (12) and (13) taken together imply

$$\|A_{t,z}v\|_2^2 \geqslant k(\log(n/k))^{-2} \cdot \beta^2\delta^{14(i-1)}k^4n^{-5}/4 \geqslant \delta^{14i}.$$

Since $i \leqslant \tau \leqslant O((\log(n/k))^4)$, we see that this contradicts the assumption

$$\|A_{t,z}v\|_2^2 \leqslant \eta_r^2 d \leqslant \exp(-C(\log(n/k))^8(\log d)^8),$$

where we have used that $k = c^*(\lceil t\rceil - r + 1)$ and $k \leqslant n/d^{1/4}$. So in fact (11) holds for all $i \in [\tau]$, as claimed.

Now iterating (11) we obtain the desired result, using that we have $\tau = O((\log(n/k))^4)$ and $k_\tau \geqslant (n/(2d))\log_{(2)} d$. □

6. Projection anti-concentration

As discussed in Section 2, our ability to "take on" the singular value σ_r depends on the magnitude of the projection of our new row or column onto the space spanned by the vectors $u_n, \ldots, u_r$, corresponding to the smallest singular directions $\sigma_n, \ldots, \sigma_r$ (on the appropriate side). Here we prove that it is unlikely for this projection to be small assuming $A_t \in \mathcal{E}_r$ (or $A_t^\dagger \in \mathcal{E}_r$). Recall that for a $t \times t$ or $(t-1/2) \times (t+1/2)$ matrix M, we let $P_{r,M}$ be the orthogonal projection onto the $\lceil t \rceil - r + 1$ smallest right-singular directions of M.

Lemma 6.1. *For $t - 2\varepsilon n \leqslant r \leqslant \lceil t \rceil$, let $A_t \in \mathcal{E}_r$, let $1 \leqslant |z| \leqslant d$, and put $M = A_{t,z}$. If $X \sim \mathrm{Row}_{\lceil t \rceil}(\xi, p)$ and $\sigma_r(M) \leqslant d^{1/2}\eta_r$ then for all $w \in \mathbb{C}^{\lceil t \rceil}$,*

$$\mathbb{P}_X\big(\|P_{r,M}(X+w)^\dagger\|_2 < d^{1/2}\eta_r\big) = O(\varepsilon + (\log_{(2)} d)^{-1/2}).$$

The remainder of this section is devoted to a proof of this lemma. The proof is broken into different regimes, when the co-dimension is large, meaning $h = \lceil t \rceil - r + 1 > nd^{-1/4}$, and when it is small, meaning $h \leqslant nd^{-1/4}$. We warm up by proving the large h case, since this is significantly easier and does not need $A_t \in \mathcal{E}_r$. When h is small we will have to appeal to the results that we proved in the previous section about matrices $A_t \in \mathcal{E}_r$.

6.1. Proof of the large h case.

We take care of the case of large h with the following.

Lemma 6.2. *For $h \geqslant nd^{-1/4}$, let $X \sim \mathrm{Row}_{\lceil t \rceil}(\xi, p)$ and let P be an orthogonal projection onto an h-dimensional subspace. Then for all $\kappa > 0$ and $w \in \mathbb{C}^n$ we have*

$$\mathbb{P}_X\big(\|P(X+w)^\dagger\|_2 < \kappa \cdot d^{1/2}h^{3/2}n^{-3/2}\big) = O(\kappa + d^{-1/4}).$$

The following lemma of Litvak, Lytova, Tikhomirov, Tomczak-Jaegermann, and Youssef [11, Lemma 4.3] gives a decent basis for this space.

Lemma 6.3. *Let $V \subseteq \mathbb{C}^n$ be a k-dimensional $\mathbb{C}$-vector space. There exists an orthonormal basis B of V so that for all $v \in B$, we have $v^*_{c^*k} \geqslant c^* k^{1/2} n^{-1}$, where $c^* > 0$ is an absolute constant.*

We also need the following anti-concentration inequality, which is a straightforward consequence of the classical inequality due to Lévy–Kolmogorov–Rogozin [10, 14].

Lemma 6.4. *Let $X \sim \mathrm{Col}_n(\xi, p)$, and let $v \in \mathbb{C}^n$ satisfy $v^*_k \geqslant \rho$. Then*

$$\max_{y \in \mathbb{C}} \mathbb{P}\big(|\langle X, v\rangle - y| \leqslant r\big) \leqslant \frac{C_\beta}{(kp)^{1/2}} \cdot \frac{r}{\rho},$$

for all $r \geqslant \beta\rho/\sqrt{2}$. Here we can take $C_\beta = O(\beta^{-3/2})$.

We include the simple deduction of this lemma in Appendix A. We are now prepared to prove Lemma 6.2, which takes care of Lemma 6.1 for large h.

Proof of Lemma 6.2. Let V be the image of P. By Lemma 6.3 there exists an orthonormal basis B of V so that $v^*_{c^*h} \geqslant c^* h^{1/2} n^{-1}$ for each $v \in B$. Write $\|P(X+w)^\dagger\|_2^2 = \sum_{v \in B} |\langle v, X+w\rangle|^2$ and note that

$$\mathbb{1}(\|P(X+w)^\dagger\|_2 \leqslant \kappa \cdot d^{1/2}h^{3/2}n^{-3/2}) \leqslant \frac{2}{h}\sum_{v \in B} \mathbb{1}(|\langle v, X+w\rangle| \leqslant 2\kappa \cdot d^{1/2}hn^{-3/2}),$$

since if $\|P(X+w)^\dagger\|_2 < \kappa \cdot d^{1/2}h^{3/2}n^{-3/2}$ then at most $h/2$ inner products on the right-hand-side are $> 2\kappa \cdot d^{1/2}hn^{-3/2}$. Taking expectations gives

$$\mathbb{P}(\|P(X+w)^\dagger\|_2 \leqslant \kappa \cdot d^{1/2}h^{3/2}n^{-3/2}) \leqslant \frac{2}{h}\sum_{v \in B} \mathbb{P}(|\langle v, X+w\rangle| \leqslant 2\kappa \cdot d^{1/2}hn^{-3/2}) = O(\kappa + d^{-1/4}),$$

where we applied Lemma 6.4 to each term in the sum with $k = c^*h$, $\rho = c^*h^{1/2}n^{-1}$ and $r = \max\{2\kappa \cdot d^{1/2}hn^{-3/2}, \beta\rho/\sqrt{2}\}$. Indeed, if $r = \beta\rho/\sqrt{2}$ we may apply Lemma 6.4 to obtain the upper bound of $O((kp)^{-1/2}) = O(d^{-3/8})$. On the other hand, if $r = 2\kappa \cdot d^{1/2}hn^{-3/2}$ then we can apply the lemma and obtain an upper bound of $O(\kappa)$. □

6.2. **Proof of the small h case.** The proof of Lemma 6.1 is similar to the proof of Lemma 6.2 but we will additionally need to appeal to our results that tell us the small singular vectors are unstructured.

Proof of Lemma 6.1. After applying Lemma 6.2 and using $h \leqslant 2\varepsilon n$, we may assume $1 \leqslant h < nd^{-1/4}$ here. We let V be the subspace spanned by the h smallest right-singular directions of M. As before, we apply Lemma 6.3 to find an orthonormal basis B of V so that

$$v^*_{c^*h} \geqslant c^*h^{1/2}n^{-1} \tag{14}$$

for each $v \in B$. Again write $\|P_{M,r}X\|_2^2 = \sum_{v\in B} |\langle v, X\rangle|^2$ and again note we have

$$\mathbb{P}(\|P_{r,M}X\|_2 \leqslant \eta_r d^{1/2}) \leqslant \frac{2}{h}\sum_{v\in B} \mathbb{P}(|\langle v, X\rangle| \leqslant 2\eta_r (d/h)^{1/2}). \tag{15}$$

Now fix $v \in B$ and express $v = \sum_{i=1}^h c_i w_i$ where u_i are unit vectors associated with the least right-singular directions of M and $\sum_{i=1}^h |c_i|^2 = 1$. We use $\sigma_{\lceil t\rceil - h+1}(A_{t,z}) = \sigma_r(A_{t,z}) \leqslant d^{1/2}\eta_r$ to see

$$\|A_{t,z}v\|_2^2 = \langle v, A_{t,z}^\dagger A_{t,z} v\rangle = \sum_{i=1}^h\sum_{j=1}^h c_i c_j \langle u_i, A_{t,z}^\dagger A_{t,z} u_j\rangle = \sum_{i=1}^h |c_i|^2 \sigma^2_{\lceil t\rceil - i+1} \leqslant d\eta_r^2, \tag{16}$$

where we have used that u_i are orthogonal eigenvectors of $A_{t,z}^\dagger A_{t,z}$, by definition. Due to (14) and (16) we may apply Lemma 5.1 to see

$$v^*_\ell \geqslant \eta_r d^{1/2} h^{-1/2} \quad \text{where} \quad \ell = (n/(2d)) \log_{(2)} d.$$

Thus the expected intersection of the support of X with the coordinates of v with $|v_i| \geqslant \eta_r d^{1/2}h^{-1/2}$ is $p\ell = (1/2)\log_{(2)} d \to \infty$. In particular, we can use Lemma 6.4 to see

$$\mathbb{P}(|\langle v, X\rangle| \leqslant \eta_r (d/h)^{1/2}) = O((\log_{(2)} d)^{-1/2}). \tag{17}$$

with the choice of $r = \rho = \eta_r d^{1/2}h^{-1/2}$. Applying (17) to each term in (15) concludes the proof of Lemma 6.1. □

7. A step in the process

The following crucial lemma tells us that each row or column addition allows us to bring a new singular value under our control, with probability $1 - o_{d\to\infty}(1)$.

Lemma 7.1. *For $t - 2\varepsilon n \leqslant r < \lceil t\rceil$, let $A_t^\dagger \in \mathcal{E}_{r+1}$ if t is integral and let $A_t \in \mathcal{E}_{r+1}$ if t is half-integral, and $d^{-1/2} \leqslant |z| \leqslant d^{1/2}$. Then*

$$\mathbb{P}\big(T_{r+1,t+1/2}(z) \leqslant T_{r,t}(z) + \delta_{r+1}\big) = 1 - O(\varepsilon + (\log_{(2)} d)^{-1/2}),$$

where the probability is over the new row or column.

To prove this we will have to track how the singular values evolve when we add a row or column to A_t. In particular we use the following basic linear-algebraic lemma along with Lemma 6.1, the main lemma from the previous section.

Lemma 7.2. *Let M be an $n \times m$ matrix and let M' be an $(n+1) \times m$ obtained by adding the row X to M. For $r < m$, we have*

$$\prod_{i=1}^{r+1} \sigma_i(M') \geqslant \|PX^\dagger\|_2 \cdot \prod_{i=1}^{r} \sigma_i(M),$$

where P is the orthogonal projection onto the span of the $m-r$ smallest right-singular vector of M.

Proof. Let Q be any $(r+1)\times(n+1)$ matrix such that $QQ^\dagger = I_{r+1}$. By Courant–Fischer applied to $M'M'^\dagger$, we have that $s_k(QM'M'^\dagger Q^\dagger) \leqslant \sigma_k(M')^2$ for all k and therefore

$$\det(QM'M'^\dagger Q^\dagger) \leqslant \prod_{i=1}^{r+1} \sigma_i(M')^2.$$

We now choose Q such that $\det(QM'M'^\dagger Q^\dagger)$ is exactly the RHS of the claimed inequality. Let Q' be an $r \times n$ matrix with rows corresponding to biggest r unit left-singular vectors of M. Q is obtained by adding an extra row and column to Q' which are all zeros, except for the bottom right entry which is 1. It is trivial to verify using orthogonality of singular vectors that $QQ^\dagger = I_{r+1}$.

Note that QM' is an $(r+1) \times m$ matrix. The first r rows of QM' are exactly the first r right-singular vectors of M' with the i-th largest singular vector scaled by $\sigma_i(M')$; this is most easily seen by using the singular value decomposition of M'. The final row of QM' is exactly X. Therefore using the base times height formula for determinants, we have that

$$\det(QM'M'^\dagger Q^\dagger) = \operatorname{dist}(X, \operatorname{span}_{\mathbb{C}}(\{e_i Q'M\}_{1\leqslant i\leqslant r}))^2 \prod_{i=1}^{r} \sigma_i(M)^2 = \|PX^\dagger\|_2^2 \cdot \prod_{i=1}^{r} \sigma_i(M)^2,$$

which completes the proof. □

We also require Cauchy interlacing (for singular values).

Fact 7.3. *Let M be an $n \times m$ matrix and let M' be M with a row added. Then*

$$\sigma_m(M) \leqslant \sigma_m(M') \leqslant \sigma_{m-1}(M) \leqslant \sigma_{m-1}(M') \leqslant \cdots \leqslant \sigma_1(M) \leqslant \sigma_1(M').$$

We now prove Lemma 7.1.

Proof of Lemma 7.1. We only need to put the pieces together that we have already built up. Note that $A^*_{t,z} = d^{-1/2}A - zI = d^{-1/2}(A - zd^{1/2}I) = d^{-1/2}A_{t,zd^{1/2}}$. If $\sigma_{r+1}(A^*_{t,z}) \geqslant \eta_{r+1}$ then we have

$$\prod_{i=1}^{r+1} \sigma_i(A^*_{t+1/2,z}) \geqslant \prod_{i=1}^{r+1} \sigma_i(A^*_{t,z}) \geqslant \eta_{r+1} \prod_{i=1}^{r} \sigma_i(A^*_{t,z}),$$

where the first inequality holds by interlacing. Thus we are done in this case.

Otherwise $\sigma_{r+1}(A^*_{t,z}) < \eta_{r+1}$, in which case we use that $A_t \in \mathcal{E}_{r+1}$. Let us focus on the situation where t is half-integral and we are adding a row. The integral case is similar except we apply the relevant argument to $A_t^\dagger$, so we omit it. By Lemma 7.2,

$$\prod_{i=1}^{r+1} \sigma_i(A^*_{t+1/2,z}) \geqslant \|P_{r+1,A^*_{t,z}}(d^{-1/2}X + w)^\dagger\|_2 \cdot \prod_{i=1}^{r} \sigma_i(A^*_{t,z}),$$

where $X \sim \operatorname{Row}_{\lceil t\rceil}(\xi, p)$ is the new row added and w is all 0 except perhaps a nonzero element corresponding to $-z$ on the diagonal (when t is half-integral). We may now apply Lemma 6.1

with r replaced by $r+1$ to see

$$\mathbb{P}_X\big(\|P_{r+1,A^*_{t,z}}(d^{-1/2}X+w)\|_2<\eta_{r+1}\big)=\mathbb{P}_X\big(\|P_{r+1,A_{t,zd^{1/2}}}(X+d^{1/2}w)\|_2<d^{1/2}\eta_{r+1}\big)$$
$$=O(\varepsilon+(\log_{(2)}d)^{-1/2})$$

as desired. □

8. Analysis of the process

Recall that we index time in half-integer steps from the interval $[m,n]$ so that at each integer time t we have that A_t is a $t\times t$ matrix and at time $t+1/2$, $A_{t+1/2}$ is the $t\times(t+1)$ matrix which is A_{t+1} with the bottom-most row deleted. Recall $r(t)$ can be defined by setting $r(m)=(1-\varepsilon/4)m$ and

$$r(t+1/2)=\begin{cases} r(t)+1 & \text{if } T_{r(t)+1,t+1/2}(z)\leqslant T_{r(t),t}(z)+\delta_{r(t)+1};\\ r(t) & \text{otherwise.}\end{cases}$$

In what follows, if M is an $n\times m$ matrix, we define $\sigma_r(M)=0$ if $r>m$. We track the evolution of the random variable $h(t)=t-r(t)$. The main goal of this section is to show that, whp, we have taken *all* singular values into our sum by the end of the process.

Lemma 8.1. *For all $z\in\mathbb{C}\setminus\{0\}$, we have*

$$\mathbb{P}(h(n)=0)=1-o(1).$$

We shall also see that this immediately implies the following result, which is the "hard" direction of our two-sided comparison of $T_n(z)$ and $U_n(z)$.

Lemma 8.2. *For all $z\in\mathbb{C}\setminus\{0\}$, we have $U_n(z)\leqslant T_n(z)+o(1)$ with probability $1-o(1)$.*

To prove these lemmas we study how $h(t)$ evolves. For convenience when analyzing cases $h(t)=0$, however, we define the slightly modified function

$$h^*(t):=\begin{cases} h(t) & \text{if } h(t)\neq 1/2;\\ 0 & \text{if } h(t)=1/2.\end{cases}$$

At the start of the process we have

$$h^*(m)\leqslant m-(1-\varepsilon/4)m=\varepsilon m/4<\varepsilon n/4,$$

by definition. For all $t<n$, we have

$$h^*(t+1/2)\leqslant h^*(t)+1/2+(1/2)\mathbb{1}_{h^*(t)=0},$$

and finally, if $A_t^\dagger\in\mathcal{E}_{r(t)+1}$ for t integral or $A_t\in\mathcal{E}_{r(t)+1}$ for t half-integral, we know from Lemma 7.1 that

$$\mathbb{P}\big(h^*(t+1/2)\leqslant h^*(t)-(1/2)\mathbb{1}_{h^*(t)>0}\big)=1-o_{\substack{d\to\infty\\ \varepsilon\to 0}}(1).$$

Of course, for this to be useful, we need to guarantee the required event sufficiently often. One difficulty here is that our bounds on the failure of $\mathcal{E}_r$ are not sufficiently strong to ensure that our matrix always satisfies the appropriate condition. To get around this we use a simple idea: if at time t we have $h^*(t)<\lfloor(n-t)/8\rfloor$ then we don't worry about certifying that new singular values are taken into our sum, as we are already "over-achieving" at such a time. Thus we only need to union bound over all pairs (t,r) where $r\leqslant t-\lfloor(n-t)/8\rfloor$. Thus, for all times in $(1/2)\mathbb{Z}$, we define

$$\mathcal{Q}_t=\begin{cases}\bigcap_{r=1}^{t-\lfloor(n-t)/8\rfloor}\{A_t\in\mathcal{E}_r\} & \text{if } t \text{ half-integral;}\\ \bigcap_{r=1}^{t-\lfloor(n-t)/8\rfloor}\{A_t^\dagger\in\mathcal{E}_r\} & \text{if } t \text{ integral;}\end{cases}\qquad\text{and then set}\qquad \widetilde{\mathcal{Q}}_t=\bigcap_{t'\leqslant t}\mathcal{Q}_t,$$

where the latter intersection is over all $t \in (1/2)\mathbb{Z} \cap [m, t]$. Of course the key point here is that $\widetilde{\mathcal{Q}}$ holds with high probability.

Lemma 8.3. $\mathbb{P}(\widetilde{\mathcal{Q}}_n) = 1 - o(1)$.

Proof. We apply Lemma 4.1 to bound $\mathbb{P}\big(\widetilde{\mathcal{Q}}^c\big)$ above by

$$\sum_t \sum_{r=1}^{t-\lfloor (n-t)/8 \rfloor} \mathbb{P}(A_t \in \mathcal{E}_r) \leqslant \sum_{t,r} \exp(-d^{1/2}(\lceil t \rceil - r + 1)) \leqslant \sum_{k \geqslant 1} 2\exp(-kd^{1/4}),$$

which tends to zero as d tends to infinity. □

To avoid the property $\widetilde{\mathcal{Q}}_n$ running interference with the independence in the row/column revelation process, we couple $h^*(t)$ to a simpler process. For this let $\mathcal{F}_t$ be the σ-algebra corresponding to the matrix A_t. Now define the random variables $X_t = h^*(t)\mathbb{1}_{\widetilde{\mathcal{Q}}_t}$. This definition is crafted so that on $\widetilde{\mathcal{Q}}_n$, *if* $X_n = 0$ *then* $h^*(n) = 0$. We now prove the following simple probabilistic lemma which shows that X_t has sufficient downward drift to ensure $\mathbb{P}(X_n = 0)$ with high probability. Such lemmas originate in the work of Costello, Tao, and Vu [3] on singularity of symmetric random matrices, and have been used more recently to study singularity and rank in sparse random matrices [5, 7].

Lemma 8.4. *Let* $(\mathcal{F}_s)_{s=0}^T$ *be a filtration and let* $(Y_s)_{s=0}^T$ *be a sequence of random variables for which* Y_s *is* $\mathcal{F}_s$*-measurable,* $Y_s \in (1/2)\mathbb{Z}_{\geqslant 0}$*,* $Y_0 \leqslant T/8$*,* $Y_{s+1} \leqslant Y_s + 1$*, and such that*

$$\mathbb{P}(Y_{s+1} \leqslant Y_s - (1/2)\mathbb{1}_{Y_s > 0} \,|\, \mathcal{F}_s) \geqslant 1 - q,$$

whenever $Y_s \geqslant \lfloor (T-s)/16 \rfloor$*. Then*

$$\mathbb{P}(Y_T = 0) \geqslant 1 - 4q^{1/8}. \tag{18}$$

Proof. We proceed by considering an appropriate exponential moment: define the random variables $Z_s = q^{(T-s)/16} q^{-Y_s/2}$ and note

$$\mathbb{P}(Y_T \geqslant 1/2) = \mathbb{P}(Z_T \geqslant q^{-1/4}) \leqslant q^{1/4} \cdot \mathbb{E} Z_T.$$

We now show that $\mathbb{E} Z_T \leqslant 4$ by bounding how much the expectation moves in each step. Indeed, for each $s \leqslant T - 1$ we claim we have

$$\mathbb{E}[Z_{s+1} \,|\, \mathcal{F}_s\,] \leqslant 1 + 2q^{1/8} Z_s. \tag{19}$$

To see this, first consider the case $Y_s < \lfloor (T-s)/16 \rfloor$. Here $Y_{s+1} \leqslant \lfloor (T-s)/16 \rfloor + 1/2$ and thus

$$Z_{s+1} \leqslant q^{(T-s-1)/16 - 1/2\lfloor (T-s)/16 \rfloor - 1/4} \leqslant 1.$$

Otherwise we have $Y_s \geqslant \lfloor (T-s)/16 \rfloor$ and we calculate

$$\mathbb{E}[Z_{s+1} \,|\, \mathcal{F}_s\,] \leqslant q^{(T-s-1)/16}\big(q^{-1/2(Y_s - (1/2)\mathbb{1}_{Y_s > 0})} + q \cdot q^{-1/2(Y_s+1)}\big) \leqslant 1 + 2q^{1/8} Z_s.$$

This establishes (19). We now iteratively apply (19) to see

$$\mathbb{E}[Z_T] \leqslant 1 + 2q^{1/8} + (2q^{1/8})^2 + \cdots + (2q^{1/8})^{T-1} + (2q^{1/8})^T \mathbb{E}[Z_0] \leqslant 2 + (2q^{1/8})^T \leqslant 4.$$

For the inequality we used the fact we may assume $q \leqslant 1/2^{16}$, otherwise the conclusion at (18) is trivial. □

Proof of Lemma 8.1. All that remains is to check the pieces fit together. Note that X_t is $\mathcal{F}_t$-measurable, by definition. Let $Y_s = X_{m+s/2}$ for $s \in [0, 2(n-m)] \cap \mathbb{Z}$. We have $T = 2\varepsilon n$ for our process and our starting point $Y_0 = X_m$ satisfies $Y_0 \leqslant m/4 \leqslant \varepsilon n/4 = T/8$, by definition. We now claim that

$$\mathbb{P}\big(X_{t+1/2} \leqslant X_t - (1/2)\mathbb{1}_{X_t > 0} \,|\, \mathcal{F}_t\big) \geqslant 1 - o_{d \to \infty}(1), \tag{20}$$

whenever $X_t \geqslant \lfloor (n-t)/8 \rfloor$. We check this inequality pointwise. If $A_t \in \mathcal{Q}_t^c$ then (20) holds, by definition, since then $X_{t+1/2} = X_t = 0$. On the other hand, if $A_t \in \mathcal{Q}_t$ then we apply Lemma 7.1 to see that

$$\mathbb{P}\big(X_{t+1/2} \leqslant X_t - (1/2)\mathbb{1}_{X_t>0} \,|\, \mathcal{Q}_t\big) = \mathbb{P}\big(h^*(t+1/2) \leqslant h^*(t) - (1/2)\mathbb{1}_{h^*(t)>0} \,|\, \mathcal{Q}_t\big) \geqslant 1 - o_{\substack{d\to\infty \\ \varepsilon\to\infty}}(1).$$

Thus X_t is a random process that satisfies the hypothesis of Lemma 8.4. We apply Lemma 8.3 and then Lemma 8.4 to see that

$$\mathbb{P}\big(h^*(n) > 0\big) = \mathbb{P}\big(h^*(n) > 0 \cap \widetilde{\mathcal{Q}}_n\big) + o(1) \leqslant \mathbb{P}\big(X_n > 0\big) + o(1) = o(1).$$

Using that $h^*(n) = 0$ implies $h(n) = 0$ (since n is integral), we are done. □

We now deduce the important consequence of Lemma 8.1, Lemma 8.2, which says that $U_n(z) \leqslant T_n(z) + o(1)$, for all $z \neq 0$.

Proof of Lemma 8.2. Assume $h(n) = 0$. Thus for each $(1-\varepsilon/4)m \leqslant r \leqslant n$, there exists $t(r) \in [m,n] \cap \frac{1}{2}\mathbb{Z}$ so that

$$T_{r,t(r)}(z) \leqslant T_{r-1,t(r)-1/2}(z) + \delta_r$$

and such that $t(\cdot)$ is a strictly increasing function. Additionally, by Fact 7.3 we have $T_{r,t}(z) \leqslant T_{r,t-1/2}(z)$.

Chaining these inequalities together gives

$$U_n(z) = T_{n,n}(z) \leqslant \cdots \leqslant T_{m,(1-\varepsilon/4)n}(z) + \sum_{r=(1-\varepsilon/4)m}^{n} \delta_r = T_n(z) + o(1),$$

and we're done. □

9. Completion of the proof

With the main difficulty of the proof behind us, we no longer need to consider A_t, and now simply think of $n \to \infty$. It is convenient to modify $T_n(z)$ to be a function of $(A^*_{n,z})$ instead of $(A^*_{m,z})$; define

$$T_n^{(1)}(z) = -\frac{1}{n}\sum_{i=1}^{(1-\varepsilon/4)m} \log(\sigma_i(A^*_{n,z}))$$
$$T_n^{(2)}(z) = -\frac{1}{n}\sum_{i=2\varepsilon n}^{(1-\varepsilon)n} \log(\sigma_i(A^*_{n,z})) - \frac{(1-\varepsilon/4)m - (1-3\varepsilon)n}{n}\log(\sigma_{m(1-\varepsilon/4)}(A^*_{m,z})).$$

We now relate $U_n(z)$, $T_n(z)$, $T_n^{(1)}(z)$, and $T_n^{(2)}(z)$.

Fact 9.1. *We have*

$$T_n(z) = -\frac{1}{n}\sum_{i=1}^{(1-\varepsilon/4)m} \log\big(\sigma_i(A^*_{m,z})\big) \leqslant T_n^{(2)}(z)$$
$$U_n(z) \geqslant -\frac{1}{m(1-\varepsilon/4)}\sum_{i=1}^{(1-\varepsilon/4)m} \log(\sigma_i(A^*_{n,z})) = \frac{T_n^{(1)}(z)}{(1-\varepsilon/4)(1-\varepsilon)}.$$

Proof. The second inequality follows from $\sigma_1(A^*_{n,z}) \geqslant \sigma_2(A^*_{n,z}) \geqslant \cdots \geqslant \sigma_n(A^*_{n,z})$. For the first inequality, note that

$$
\begin{aligned}
-\frac{1}{n}&\sum_{i=1}^{(1-\varepsilon/4)m} \log\left(\sigma_i(A^*_{m,z})\right) \\
&\leqslant -\frac{1}{n}\sum_{i=1}^{m-3\varepsilon n} \log\left(\sigma_i(A^*_{m,z})\right) - \frac{(1-\varepsilon/4)m-(1-3\varepsilon)n}{n}\log(\sigma_{(1-\varepsilon/4)m}(A^*_{m,z})) \\
&\leqslant T_n^{(2)}(z)
\end{aligned}
$$

where the final inequality follows from iterating Fact 7.3 a total of $2\varepsilon n$ times. ☐

Next we will require control on the Hilbert–Schmidt norm of $A^*_{n,z}$.

Fact 9.2. *With high probability,*

$$\|A^*_{n,z}\|^2_{\mathrm{HS}} \leqslant 4(|z|^2+1)n$$

Proof. Note that

$$\mathbb{E}\,\|A^*_{n,z}\|^2_{\mathrm{HS}} = \mathbb{E}\,\|d^{-1/2}A_n - zI_n\|^2_{\mathrm{HS}} \leqslant 2(\mathbb{E}\|d^{-1/2}A_n\|^2_{\mathrm{HS}} + \|zI_n\|^2_{\mathrm{HS}}) = 2(1+|z|^2)n.$$

Viewing the square of the Hilbert–Schmidt norm as the entry-wise sum of the squares, the result follows by Chernoff and the strong law of large numbers. ☐

Finally we require convergence of the truncated log potential; this is a fairly standard proof which we defer to the end of the section.

Lemma 9.3. *For $pn \to \infty$ and $p \to 0$, let $A_n \sim \Delta_n(\xi, p)$ and let $\varepsilon \to 0$ sufficiently slowly. Then $T_n^{(1)}(z)$ and $T_n^{(2)}(z)$ converge to $U^\circ(z)$ in probability.*

To finish the proof we state a simple criterion that allows us to jump from point-wise convergence of $U_n(z)$, plus a simple tightness condition, to the convergence of the spectral measures. The following lemma is a simple reworking of Theorem 2.1 in [18], which gives the following "abstract" criterion for convergence to a circular law.

Proposition 9.4. *For each n, let A_n be a random $n \times n$ matrix and let $\mu_n = \mu_{d^{-1/2}A_n}$ be the scaled spectral law of A_n. If $\mathbb{E}\,\|d^{-1/2}A_n\|_{\mathrm{HS}} = O(n^{1/2})$ and for almost all $z \in \mathbb{C}$ we have $U_{\mu_n}(z)$ converges to $U^\circ(z)$ in probability, then $\mu_n \to \mu^\circ$ in probability.*

We now prove Thoerem 1.2 given Lemma 9.3.

Proof of Theorem 1.4. By combining the first item of Fact 9.1, Lemma 8.2, and Lemma 9.3, we have

$$\mathbb{P}(U_n(z) \leqslant U^\circ(z) + o(1)) = 1 - o(1).$$

Combining the second item of Fact 9.1 and Lemma 9.3, we have

$$\mathbb{P}(U_n(z) \geqslant U^\circ(z) - o(1)) = 1 - o(1).$$

Therefore, Fact 9.2 allows us to invoke Proposition 9.4, which completes the proof. ☐

We now prove Lemma 9.3. Define $\nu_{n,z}$ to be the empirical measure of all of the singular values of the shifted matrix $A^*_{n,z}$. Since $T_n^{(1)}(z)$ and $T_n^{(2)}(z)$ cut out the smallest singular values, arguments based on just convergence of distribution can control these sums.

Define G_n to be an iid $n \times n$ random matrix with entries distributed as variance 1 complex Gaussians: $\frac{1}{\sqrt{2}}(Z_1 + Z_2 i)$, with $Z_1, Z_2 \sim N(0,1)$ iid. Let $\nu^G_{n,z}$ be the empirical measure of the singular values of the shifted matrix $G_{n,z} = n^{-1/2}G_n - zI_n$. It is well known that $\nu^G_{n,z}$ converges in probability to a limit ν^G_z, which satisfies the following properties.

Fact 9.5. *There exists an absolute constant $C > 0$ such that the following holds. For all $z \in \mathbb{C}$, ν_z^G is a measure on $[0, |z| + C]$ with $\nu_z([0,t)) \leqslant Ct$ for all t, and*

$$-\int \log t \, d\nu_z^G = U^\circ(z).$$

Next we establish convergence in distribution for the measures $\nu_{n,z}$; this is by now standard and we include a short sketch for completeness.

Fact 9.6. *For $pn \to \infty$ and $p \to 0$, let $A_n \sim \Delta_n(\xi, p)$. For all bounded continuous functions $\phi\colon \mathbb{R} \to \mathbb{R}$, we have*

$$\int \phi(t) \, d\nu_{n,z} \to \int \phi(t) \, d\nu_z^G$$

in probability.

Proof sketch. Fix $M \geqslant 1$, let $\xi_{i,j}^{(1)} = \xi_{i,j} \mathbb{1}_{|\xi_{i,j}| \leqslant M}$, $\xi_{i,j}^{(2)} = \xi_{i,j} \mathbb{1}_{|\xi_{i,j}| > M}$, and note that $\xi_{i,j} = \xi_{i,j}^{(1)} + \xi_{i,j}^{(2)}$. Let $A_n^{(k)}$ have entries $\delta_{i,j}\xi_{i,j}^{(k)}$ for $k \in \{1, 2\}$. By the Chernoff and the strong law of large numbers, with probability $1 - o(1)$ we have

$$\|A_n^{(2)}\|_{\mathrm{HS}}^2 \leqslant 2pn^2 \cdot \mathbb{E}|\xi_{i,j}^{(2)}|^2.$$

Furthermore, since the entries of $A_n^{(1)}$ are bounded by M, by a standard trace method argument we have that the empirical measure of the singular values of the matrix

$$(pn\mathbb{E}|\xi_{i,j}^{(1)}|^2)^{-1/2}(A_n^{(1)} - \mathbb{E}A_n^{(1)}) - zI_n$$

converges to ν_z^G in probability. Here we have used that the entries of $A_n^{(1)} - \mathbb{E}A_n^{(1)}$ are distributed as $\delta_{i,j}\xi_{i,j}^{(1)} - p\mathbb{E}\xi_{i,j}^{(1)}$ which has mean zero and variance $p\mathbb{E}|\xi_{i,j}^{(1)}|^2 - p^2(|\mathbb{E}\xi_{i,j}^{(1)}|)^2$. Therefore by the Hoffman–Wielandt inequality [9, Theorem 1] to bound deviations introduced by adding $A_n^{(2)}$, using that $\mathbb{E}A_n^{(1)}$ is a rank 1 matrix, and finally taking $M \to \infty$, the desired result follows. □

We are now in a position to easily prove Lemma 9.3, which will conclude the proof.

Proof of Lemma 9.3. By the convergence in distribution established in Fact 9.6 as well as Fact 9.5 (applied with n replaced by m and z by $(1-\varepsilon)^{-1/2}z$) we have that $\sigma_{m-\varepsilon m/4}(A_{m,z}^*) \lesssim |z| + 1$ and $\sigma_{m-\varepsilon m/4}(A_{m,z}^*) \gtrsim \varepsilon$ whp. Given this, establishing that $T_n^{(1)}(z)$ and $T_n^{(2)}(z)$ each converge in probability to $U^\circ(z)$ boil down to essentially identical arguments; we prove convergence for $T_n^{(1)}(z)$ explicitly.

Let $\varepsilon' = 5\varepsilon/4 - \varepsilon^2/4$ and note that

$$T_n^{(1)}(z) = -\frac{1}{n}\sum_{i=1}^{(1-\varepsilon')n} \log(\sigma_i(A_{n,z}^*)) + o(1)$$

whp, where the error term arises due to rounding the top index and can be absorbed using Fact 9.5 and convergence in distribution of $\sigma_i(A_{n,z}^*)$.

Let $\tau_z(\varepsilon')$ denote the unique real number such that $\mathbb{P}_{X\sim\nu_z^G}[X \leqslant \tau_z(\varepsilon')] \geqslant \varepsilon'$ and $\mathbb{P}_{X\sim\nu_z^G}[X < \tau_z(\varepsilon')] \leqslant \varepsilon'$. Define $\nu_z^{G,\varepsilon'}$ by letting $X \sim \nu_z^G$, then outputting X conditional on $X > \tau_z(\varepsilon')$ with probability $(1-\varepsilon')^{-1}\mathbb{P}_{X\sim\nu_z^G}[X > \tau_z(\varepsilon')]$ and $\tau_z(\varepsilon')$ with probability $(1-\varepsilon')^{-1}(\mathbb{P}_{X\sim\nu_z^G}[X \leqslant \tau_z(\varepsilon')] - \varepsilon')$. By convergence in distribution (Fact 9.6) and some manipulation, we have that $\{\sigma_i(A_{n,z})\}_{1\leqslant i\leqslant(1-\varepsilon')n}$ converges in distribution to $\nu_z^{G,\varepsilon'}$.

Let $\psi\colon \mathbb{R} \to \mathbb{R}$ be a smooth nonnegative 1-bounded function which is 1 between $\varepsilon'/(2C)$ and $|z| + C + \varepsilon'^{-1}$ and which is 0 outside $\varepsilon'/(4C)$ and $|z| + C + 2\varepsilon'^{-1}$, where C as in Fact 9.5.

By Fact 9.6 and Fact 9.5, whp $\sigma_{(1-\varepsilon')n}(A_{n,z}^*) \geqslant \varepsilon'/(2C)$. Furthermore, since $\psi(t)\log(t)$ is a bounded continuous function, Fact 9.6 yields

$$-\frac{1}{n}\sum_{i=1}^{(1-\varepsilon')n} \psi(\sigma_i(A_{n,z}^*))\log(\sigma_i(A_{n,z}^*)) = -\int \psi(t)\log t \, d\nu_z^{G,\varepsilon'} + o(1)$$

whp. Next note that $\log(|x|+1)\mathbb{1}_{|x|\geqslant M} \leqslant \frac{x^2}{M}$. Thus for ε' sufficiently small and using Fact 9.2, whp

$$\left|\frac{1}{n}\sum_{i=1}^{(1-\varepsilon')n}\Big(\psi(\sigma_i(A_{n,z}^*))\log(\sigma_i(A_{n,z}^*)) - \log(\sigma_i(A_{n,z}^*))\Big)\right| \leqslant \frac{1}{n}\sum_{i=1}^{(1-\varepsilon')n} \mathbb{1}_{\sigma_i(A_{n,z}^*)\geqslant \varepsilon'^{-1}}\log(\sigma_i(A_{n,z}^*))$$
$$\leqslant \frac{\varepsilon'}{n}\sum_{i=1}^{n}\sigma_i(A_{n,z}^*)^2 = \frac{\varepsilon'}{n}\|A_{n,z}^*\|_{\mathrm{HS}}^2 \leqslant \varepsilon'^{1/2},$$

supposing that ε is sufficiently small in terms of z (which occurs for large n).

Finally, since ν_z^G is supported only on $[0,|z|+C]$ and $\nu_z^G([0,t]) \leqslant Ct$ by Fact 9.5,

$$\left|U^\circ(z) + \int \psi(t)\log t \, d\nu_z^{G,\varepsilon'}\right| = \left|-\int \log t \, d\nu_z^G + \int \log t \, d\nu_z^{G,\varepsilon'}\right| \leqslant \int |\log t|\mathbb{1}_{t<\varepsilon'/C} \, d\nu_z^G \leqslant \varepsilon'^{1/2}$$

and we are finished upon combining the four centered equations. □

Appendix A. Ommitted proofs

Proof of Lemma 4.2. Assume $t-1/2$ is integral so that A_t is a $t' \times (t'+1)$ matrix where $t' = t - 1/2$. The other case is strictly simpler so we omit it. We fix a set $S \subseteq [t'+1]$ of columns of size k. We note that

$$|U(S)| = \sum_{i\in[t']} \mathbb{1}(i \in U(S)),$$

where the sum is over the rows and thus is a sum of independent random variables. Thus we compute the expectation and then control the deviations. For $i \in [t']$ we have

(21) $$\mathbb{P}(i \in U(S)) = (1-p)^k \text{ if } i \in S \text{ and } \mathbb{P}(i\in U(S)) = \mathbb{P}(|\xi| \geqslant \beta)(1-p)^{k-1}pk \text{ if } i \notin S.$$

If k is such that $n/(2d) \leqslant k \leqslant (n/(2d))\log_{(2)} d$ we can just consider the contribution of $i \notin S$ and then use Chernoff. Indeed, picking up from (21) we have

$$pk(1-p)^{k-1} \geqslant (1/3)(1-p)^{(1/(2p))\log_{(2)} d} \geqslant (1/3)\exp(-\log_{(2)} d) \geqslant (\log d)^{-3/2},$$

where we have used that d is large and that $1-x \geqslant e^{-2x}$ for $x \leqslant 1/2$. Thus

$$\mathbb{E}|U(S)| \geqslant (t'-k)(\log d)^{-3/2} \geqslant 2(\log(n/k))^{-2}dk = 2\alpha(k)dk$$

and so by Chernoff, we have

(22) $$\mathbb{P}\big(|U(S)| \leqslant \alpha(k)dk\big) \leqslant \mathbb{P}\big(|U(S)| \leqslant (1/2)\mathbb{E}|U(S)|\big) \leqslant \exp(-cn/(\log d)^{3/2}).$$

We finish this range of k by simply union bounding over all subsets of size k: the probability there is a set S of size k with $|U(S)| < \alpha(k)dk$ in this range of k is at most

$$e^{-cn(\log d)^{-3/2}}\binom{n}{k} \leqslant \exp\big(-cn/(\log d)^{3/2} + (n/(2d))(\log d)\log_{(2)} d\big) \leqslant e^{-d^{2/3}k},$$

where the last inequality follows since $k \leqslant (n/(2d))\log_{(2)} d$ and d is sufficiently large.

We now turn to the trickier range $c^*(t-r+1) \leqslant k \leqslant n/(2d)$. For this, define $T = \alpha(k)dk$ and write

(23) $$\mathbb{P}(\exists S, |S| = k, |U(S)| < T) \leqslant \binom{n}{k}\mathbb{P}(|U(S)\cap S| < T)\cdot\mathbb{P}(|U(S)\setminus S| < T).$$

For the sake of notation we also set $\gamma = \mathbb{P}(|\xi| \geqslant \beta)$ and then define

$$q_1 = (1-p)^k \text{ and } q_2 = \gamma(1-p)^{k-1}pk.$$

Note that $1-q_1 \leqslant pk$ and $1-q_2 \leqslant 1-\gamma kp/4$. Let $k' = |S \cap [t']|$; note that $k' \in \{k-1, k\}$. Now $|U(S) \cap S|$ is distributed as binomial random variable $B(k', q_1)$ and $|U(S) \setminus S|$ is distributed as $B(t'-k', q_2)$. Therefore (23) is at most

$$\sum_{i,j<T} \binom{n}{k'}\binom{k'}{i}(1-q_1)^{k'-i} \cdot \binom{t'-k'}{j}(1-q_2)^{t'-k'-j}$$

$$\leqslant (T+1)^2 \left(\frac{en}{k'}\right)^{k'} 2^{k'}(k'p)^{k'-T}\left(\frac{en}{T}\right)^T (1-\gamma kp/4)^{n/2},$$

where we used that $k' + T < n/3$ and bounds on q_1, q_2. We bound this by

$$\begin{aligned}(4epnd)^{k'}\left(\frac{2en^2}{kdT}\right)^T e^{-\gamma kpn/8} &\leqslant (4ed^2)^k (n/k)^{5T} \exp(-\gamma kd/8) \\ &\leqslant (2epn)^k e^{5dk/\log(n/k)} \exp(-\gamma kd/8) \leqslant e^{-d^{2/3}k};\end{aligned}$$

the desired result follows after summing over all $k \geqslant c^*(t-r+1)$. □

Proof of Lemma 4.3. We first address $\mathcal{B}$. Note that for a given $i \in [\lfloor t \rfloor]$, we have

$$\mathbb{P}\left(\sum_{j=1}^{\lfloor t \rfloor} \delta_{i,j} > 2d + x\right) \leqslant \exp(-c(d+x))$$

by Chernoff. By Chernoff again, for every k we see

$$\#\left\{i \in [\lfloor t \rfloor] \colon \sum_{j=1}^{\lceil t \rceil} \delta_{i,j} \geqslant C\log(n/k)\right\} \leqslant n(k/n)^2 + (\log n)^2$$

with probability at least $1 - n^{-9}$. Additionally, every such sum is bounded by $d + O(\log n)$ with probability at least $1 - n^{-9}$. We can now see $\mathcal{B}$ holds with sufficiently good probability by considering for each possible size of S the sum of the largest rows (and similarly for the columns).

For $\mathcal{Q}$, note that $\mathbb{E}|\xi_{i,j}|^2 \leqslant 1 + \beta^{-2} \leqslant 2\beta^{-2}$, so we have $\mathbb{P}(|\xi_{i,j}| > 8H/\beta) \leqslant 1/(32H^2)$, and hence $\mathbb{P}(\delta_{i,j}|\xi_{i,j}| > 8H/\beta) \leqslant p/(32H^2)$. By the Chernoff bound, there are at most $2pn^2/(32H^2) + (\log n)^2$ entries of A_t which are greater than $8H/\beta$ in magnitude with probability at least $1 - \exp(-(\log n)^2)$. Now we take a union bound over a dyadically separated set of $H \in [1, n^4]$ and appropriately adjust constants to obtain $\mathbb{P}(\mathcal{Q}) \leqslant 1 - 1/(3n^3)$.

For $\mathcal{R}$, note that we may assume $\ell \leqslant n/2$ (else the bound is vacuous) and thus $L/\beta \geqslant d^5\beta^{-5} \geqslant d^2\mathbb{E}|\xi|$. Furthermore note that

$$\mathbb{E}[\delta_{i,j}|\xi_{i,j}|] = p\mathbb{E}|\xi_{i,j}| \leqslant p\beta^{-1} \text{ and } \mathrm{Var}[\delta_{i,j}|\xi_{i,j}|] \leqslant \mathbb{E}[\delta_{i,j}^2|\xi_{i,j}|^2] \leqslant p(1+\beta^{-2}) \leqslant 2p\beta^{-2}.$$

Thus we have by Chebyshev's inequality

$$\mathbb{P}\left(\sum_{j=1}^{\lceil t \rceil} |(A_t)_{i,j}| \geqslant L/(2\beta)\right) \leqslant \frac{2pn\beta^{-2}}{(L/(2\beta) - pn\mathbb{E}[|\xi_{i,j}|])^2} \leqslant 2pn\beta^{-2} \cdot (L/(4\beta))^{-2} = 32dL^{-2}.$$

The desired result then follows by Chernoff applied to each row and considering a union bound over values of ℓ along an exponential sequence, for $\ell > n^{1/9}$. (When ℓ is sufficiently small we can use $\mathcal{Q}$ instead.) □

Proof of Fact 5.4. If $k \geqslant (n/(2d))\log_{(2)} d$ then the result is trivial, so we assume the opposite.

Now we consider how many steps it takes us to double the value of k_t in two different ranges. When $\ell \leqslant n\exp(-d)$ we see $d \leqslant \log(n/\ell)$ hence

$$g(\ell) \geqslant (cd\ell/2)(\log(n/\ell))^{-3}.$$

The right side is increasing in ℓ, thus it takes at most $O((\log(n/\ell))^3)$ steps for this recurrence to go from ℓ to a value which is at least size 2ℓ.

When $n\exp(-d) \leqslant \ell \leqslant (n/(2d))\log_{(2)} d$ we have $d \geqslant \log(n/\ell)$ so

$$g(\ell) \geqslant (c\ell/2)(\log(n/\ell))^{-2}.$$

Thus it takes at most $O((\log(n/\ell))^2)$ steps to go from ℓ to at least 2ℓ. Putting these two observations together we have

$$\tau \leqslant O\bigg(\sum_{a\geqslant 0}(\log(n/(2^a k)))^3\bigg) = O((\log(n/k))^4),$$

as desired. □

We now prove the required anti-concentration inequality Lemma 6.4; for the sake of simplicity we define for a (real or complex) random variable Γ,

$$\mathcal{L}(\Gamma, t) = \sup_{z\in\mathbb{C}} \mathbb{P}(|\Gamma - z| \leqslant t).$$

We require the following anti-concentration inequality due to Lévy–Kolmogorov–Rogozin [10, 14].

Lemma A.1. *Let $\xi_1, \dots, \xi_n$ be independent real-valued random variables. For any real numbers $r_1, \dots, r_n > 0$ and any real $r \geqslant \max_{1\leqslant i\leqslant n} r_i$, we have*

$$\mathcal{L}\bigg(\sum_{i=1}^n \xi_i, r\bigg) \leqslant \frac{Cr}{\sqrt{\sum_{i=1}^n (1 - \mathcal{L}(\xi_i, r_i))r_i^2}}$$

for an absolute constant $C > 0$.

We will also require the following basic observation regarding the distribution of certain complex random variables.

Lemma A.2. *If $\mathcal{L}(\xi, \beta) \leqslant 1 - \beta$ and $z \in \mathbb{C}\setminus\{0\}$. We have either*

$$\mathcal{L}(\mathrm{Re}(z\xi), \beta|z|/\sqrt{2}) \leqslant 1 - \beta/2 \quad \text{or} \quad \mathcal{L}(\mathrm{Im}(z\xi), \beta|z|/\sqrt{2}) \leqslant 1 - \beta/2.$$

Proof. Suppose not. Then there exist $x, y \in \mathbb{R}$ such that

$$|\mathrm{Re}(z\xi) - x| \leqslant \beta|z|/\sqrt{2}, \quad |\mathrm{Im}(z\xi) - y| \leqslant \beta|z|/\sqrt{2}$$

with probability greater than $1 - \beta$. This implies that

$$|z\xi - x - yi| \leqslant \beta|z|$$

with probability greater than $1 - \beta$; dividing by z we obtain a contradiction. □

We now prove Lemma 6.4.

Proof of Lemma 6.4. Without loss of generality, let $|v_1| \geqslant |v_2| \geqslant \cdots \geqslant |v_n|$ so that $|v_k| \geqslant \rho$ and let $\rho' = \beta\rho/\sqrt{2}$. By Lemma A.2, at least $k/2$ coordinates $j \in [k]$ one of

$$\mathcal{L}(\mathrm{Re}(v_j\xi), \rho') \leqslant 1 - \beta/2 \quad \text{or} \quad \mathcal{L}(\mathrm{Im}(v_j\xi), \rho') \leqslant 1 - \beta/2.$$

By multiplying ξ by $\sqrt{-1}$ if necessary, we may assume the first case holds, and let the set of such coordinates be $S \subseteq [k]$. This immediately implies that

$$\mathcal{L}(\mathrm{Re}(v_j\delta_j\xi_j), \rho') \leqslant 1 - \beta p/2$$

for all $j \in S$. By Lemma A.1, for appropriately chosen C we have

$$\mathcal{L}\bigg(\sum_{j=1}^{n} v_j\delta_j\xi_j, r\bigg) \leqslant \mathcal{L}\bigg(\sum_{j\in S} v_j\delta_j\xi_j, r\bigg) \leqslant \mathcal{L}\bigg(\sum_{j\in S}\mathrm{Re}(v_j\delta_j\xi_j), r\bigg)$$
$$\leqslant \frac{Cr}{\big(\rho'^2\sum_{i\in S}(1-\mathcal{L}(\mathrm{Re}(v_i\delta_i\xi_i),\rho'))\big)^{1/2}} \leqslant \frac{Cr}{\rho'((kp\beta)/2)^{1/2}} = \frac{2C\beta^{-3/2}r}{(kp)^{1/2}\rho}$$

where we have used $r \geqslant \beta\rho/\sqrt{2} = \rho'$. □

References

[1] Z. D. Bai, *Circular law*, Ann. Probab. **25** (1997), 494–529.

[2] Charles Bordenave and Djalil Chafaï, *Around the circular law*, Probab. Surv. **9** (2012), 1–89.

[3] Kevin P. Costello, Terence Tao, and Van Vu, *Random symmetric matrices are almost surely nonsingular*, Duke Math. J. **135** (2006), 395–413.

[4] Alan Edelman, *Eigenvalues and condition numbers of random matrices*, SIAM J. Matrix Anal. Appl. **9** (1988), 543–560.

[5] Asaf Ferber, Matthew Kwan, Ashwin Sah, and Mehtaab Sawhney, *Singularity of the k-core of a random graph*, Duke Mathematical Journal **1** (2023), 1–40.

[6] V. L. Girko, *The circular law*, Teor. Veroyatnost. i Primenen. **29** (1984), 669–679.

[7] Margalit Glasgow, Matthew Kwan, Ashwin Sah, and Mehtaab Sawhney, *The exact rank of sparse random graphs*, arXiv:2303.05435.

[8] Friedrich Götze and Alexander Tikhomirov, *The circular law for random matrices*, Ann. Probab. **38** (2010), 1444–1491.

[9] A. J. Hoffman and H. W. Wielandt, *The variation of the spectrum of a normal matrix*, Duke Math. J. **20** (1953), 37–39.

[10] A. Kolmogorov, *Sur les propriétés des fonctions de concentrations de M. P. Lévy*, Ann. Inst. H. Poincaré **16** (1958), 27–34.

[11] Alexander E. Litvak, Anna Lytova, Konstantin Tikhomirov, Nicole Tomczak-Jaegermann, and Pierre Youssef, *Circular law for sparse random regular digraphs*, J. Eur. Math. Soc. (JEMS) **23** (2021), 467–501.

[12] M. L. Mehta, *Random matrices and the statistical theory of energy levels*, Academic Press, New York-London, 1967.

[13] Guangming Pan and Wang Zhou, *Circular law, extreme singular values and potential theory*, J. Multivariate Anal. **101** (2010), 645–656.

[14] Boris A. Rogozin, *On the increase of dispersion of sums of independent random variables*, Teor. Verojatnost. i Primenen **6** (1961), 106–108.

[15] Mark Rudelson and Konstantin Tikhomirov, *The sparse circular law under minimal assumptions*, Geom. Funct. Anal. **29** (2019), 561–637.

[16] Ashwin Sah, Julian Sahasrabudhe, and Mehtaab Sawhney, *The limiting spectral law for sparse iid matrices*, Forthcoming.

[17] Terence Tao and Van Vu, *Random matrices: the circular law*, Commun. Contemp. Math. **10** (2008), 261–307.

[18] Terence Tao and Van Vu, *Random matrices: universality of ESDs and the circular law*, Ann. Probab. **38** (2010), 2023–2065, With an appendix by Manjunath Krishnapur.

[19] Eugene P. Wigner, *On the distribution of the roots of certain symmetric matrices*, Ann. of Math. (2) **67** (1958), 325–327.

Massachusetts Institute of Technology. Department of Mathematics.
Email address: asah@mit.edu

University of Cambridge. Department of Pure Mathematics and Mathematical Statistics.
Email address: jdrs2@cam.ac.uk

Massachusetts Institute of Technology. Department of Mathematics.
Email address: msawhney@mit.edu