THE CIRCULAR LAW FOR RANDOM MATRICES WITH INTRA-ROW DEPENDENCE

CHRIS CONNELL† AND PAWAN PATEL

ABSTRACT. We consider the problem of determining the limiting spectral distribution for random matrices whose row distributions are permitted to have limited dependence. We assume mild moment conditions and give an extension of the Marčenko-Pastur theorem for this context. The main new feature here are geometric conditions on the distributions which allow us to extend the circular law to this setting.

1. INTRODUCTION

A central point of interest in the theory of random matrices is spectral universality, i.e. the degree to which the eigenvalues of a matrix randomly chosen from a given ensemble will follow a particular density law independent of the choice of matrix and sometimes, within certain limits, on choices that govern the ensemble such as the distributions for the entries. For a broad survey from a historical perspective, see [DF]. A key example of universality comes from the Tracy-Widom distribution [TW98] and the far reaching results of Tao and Vu [TV08, TV09] on the circular law. Like most of the results obtained until recently, these have focused on the classical setting where one assumes independence of the entries.

More recently, a number of authors have attacked various generalizations and analogues (see for example, [BY93, BGCD, BDn, BV, Cha, CTV06, D], [EGP, ERS]+10, GNR, HAF, KLLW, LW, Lou, Mal, TV10, TV12, TV14, TV15, Yao). Among these generalizations a number of recent results have begun to explore universality under the allowance for a (necessarily) limited amount of dependence between the entries (e.g. [AGL+08, Ada11, AC15, LAP, Woo]). Among the latter category include the original Marčenko-Pastur paper ([MP67]), where some dependence within rows was allowed, but for a spherically uniform distribution. This was generalized by Pajor and Pastor ([PP07]) to allow for an arbitrary isotropic log-concave distribution.

In this paper we are concerned with exploring the limits to which dependence can be allowed in the current best approach to achieving the circular law. As in several recent results, we use the Tao-Vu replacement principle ([TV08]) along with a generalization of the Marčenko-Pastur Law to our situation to handle the middling and large eigenvalues. Combined with our geometric conditions on the row distributions, we are able to obtain the appropriate bounds on the lowest singular values of the random matrices to obtain the circular distribution in our dependent case.

We need three main assumptions on the distribution of the entries of our random matrices in order to obtain the required Marčenko-Pastur law. These are only used for this purpose, and without these assumptions our results would still follow if the Marčenko-Pastur component can be guaranteed by other means.

For an ensemble of \( n \times N_n \) random matrices \( A_n \), denote by \( \text{Mom}_k(A_n) \) the expected value of the \( k \)-fold Kronecker (tensor) product \( A_n \otimes \cdots \otimes A_n \).

2010 Mathematics Subject Classification. Primary 15B52; Secondary 60F15.

† Supported in part by Simons Foundation grant #210442.
Assumptions on the Matrix Ensemble. Let \((N_n)\) be an increasing sequence of positive integers with \(\lim_{n \to \infty} \frac{n}{N_n} \in (0, \infty)\). We assume that the distributions on our ensembles of \(n \times N_n\) random matrices \(A_n\) with entries \(x_{ij}^{(n)}\) satisfy:

(A1) for every \(k \in \mathbb{N}\), \(\sup_n \max_{1 \leq i \leq n, 1 \leq j \leq N_n} \mathbb{E}[|x_{ij}^{(n)}|^k] < \infty\);

(A2) for every \(k \in \mathbb{N}\), the sum of all terms in \(\text{Mom}_{2k}(A_n)\) with at least one \(x_{ij}^{(n)}\) appearing with a power of 1 is of size \(o_k(n^{k+1})\);

(A3) for every \(\epsilon > 0\):

\[
\lim_{n \to \infty} \frac{1}{n} \sum_{i \leq n} \mathbb{P} \left( \left| \frac{1}{N_n} \sum_{j=1}^{N_n} (x_{ij}^{(n)})^2 - 1 \right| \geq \epsilon \right) = 0
\]

and

\[
\lim_{n \to \infty} \frac{1}{N_n} \sum_{j \leq N_n} \mathbb{P} \left( \left| \frac{1}{n} \sum_{i=1}^{n} (x_{ij}^{(n)})^2 - 1 \right| \geq \epsilon \right) = 0.
\]

The first and third of these conditions appear in Adamczak. The third assumption requires that the rows (resp. columns) of the random matrix \(A_n\) have Euclidean norm, normalized by \(\sqrt{N_n}\) (resp. \(\sqrt{n}\)) converge in probability to 1. This is necessary if one expects to have a universality result of this kind.

The main difference between our version and earlier results is assumption A2, which allows for more dependence in that it only requires a particular sum of the 2k-th moment of the random vector to decay sufficiently quickly. Previous results used a stronger assumption, namely that for every \(n, i, j\) the \(\mathbb{E}(x_{ij}^{(n)} | \mathcal{F}_{ij}) = 0\) where \(\mathcal{F}_{ij}\) is the \(\sigma\)-field generated by \(\{x_{kl}^{(n)} : (k, l) \neq (i, j)\}\).

Our main result is the following. (For the definition of \(\|v_1^{(n)}, \ldots, v_n^{(n)}\|_{d, \delta}\), see Section 2)

**Theorem 1.1.** Let \(A_n\) be a sequence of \(n \times n\) random matrices with independent rows \(X_1^{(n)}, \ldots, X_n^{(n)}\) defined on a common probability space and satisfying assumptions A1-A3. Assume that for each \(n\) and \(i, d \leq n\), and all \(\delta > 0\), the probability measures \(v_i^{(n)}\) for \(X_i^{(n)}\) have uniformly bounded \(\|v_1^{(n)}, \ldots, v_n^{(n)}\|_{d, \delta}\). Then almost surely the spectral measure of \(\mu_{\frac{1}{n}A_n}\) converges weakly to the uniform distribution on the unit disk in \(\mathbb{C}\).

**Remark 1.2.** The hypotheses are quite close to necessary in a certain sense, although there still seems to be room to slightly weaken the hypotheses A1-A3.

Also, one only needs \(\delta \geq n^{-\frac{1}{2}+\beta}\) for almost sure convergence and \(\delta \geq n^{-\frac{1}{2}+\beta}\) for convergence in probability. (Here \(\beta > 0\) is any small number.) Moreover the bound on the \(\|\cdot\|_{d, \delta}\) is only needed for \(d < n - n^{0.99}\), but this constraint is not very restrictive to begin with (see Proposition 2.9).

2. Least Singular Value

While our results are stated in terms of arbitrary (Radon) probability measures for the distributions of the rows of our random matrices, it is convenient to work with absolutely continuous measures on \(\mathbb{C}^n\). The general case is recovered by passing to weak-* limits.
For \( f \in L^1_{\text{loc}}(\mathbb{C}^m) \) and any bounded Borel set \( E \subset \mathbb{C}^m \). We define \( \text{Av}_E(f) = \frac{1}{\text{vol}(E)} \int_E f(x) dx \). If not explicitly specified, then the dimension \( m \) in the definition will be the minimal one for which \( E \) belongs to an \( m \)-dimensional affine subspace, in case it belongs to a larger copy of \( \mathbb{C}^n \).

Let \( \omega_n = \frac{n^n}{\Gamma(n+1)} \) denote the volume of the unit ball in \( \mathbb{C}^n \). The volume of the sphere \( S^{2n-1}(r) \) of dimension \( 2n-1 \) and radius \( r \) in \( \mathbb{C}^n \) is then \( 2n\omega_n r^{2n-1} = \frac{2r^{2n-1}n^n}{\Gamma(n)} \). Moreover, let \( r_n = (2n\omega_n)^{1/n} \) so that \( \text{Vol}(S^{2n-1}(r_n)) = 1 \). More generally, let \( S_A(r) \), resp. \( B_A(r) \), denote the sphere, resp. ball, of radius \( r \) around 0 in the subspace \( A \). For any set \( E \subset \mathbb{C}^n \) we denote by \( x + E \) the translation of \( E \) by \( x \in \mathbb{C}^n \).

Finally, let \( \text{Gr}_{n,d} \) represent the Grassmanian of all \( d \)-dimensional linear subspaces of \( \mathbb{C}^n \). We define the following norms.

**Definition 2.1.** For any \( W \in \text{Gr}_{n,n-d} \) with orthogonal subspace \( W^\perp \in \text{Gr}_{n,d} \) and any Borel function \( f : \mathbb{C}^n \to [0, \infty) \), we set

\[
\| f \|_{W,\delta,1} = \int_W \text{Av}_{x+B_{W^\perp}(\delta)}(|f|) \, dx
\]

and

\[
\| f \|_{d,\delta,1} = \sup_{W \in \text{Gr}_{n,n-d}} \| f \|_{W,\delta,1}
\]

**Definition 2.2.**

\[
\| f \|_{d,\delta,2} = (n-d)\omega_{n-d} \int_0^\infty \| f \|_{L^\infty\left(\bigcup_{t \in [t+\delta]} S^{2n-1}(r)\right)} t^{n-d-1} \, dt
\]

Given the that \( f \) is a probability measure on \( \mathbb{C}^n \), Def. 2.1 is the probability that the projection of a random vector drawn from \( f \) to a fixed \( n-d \) dimensional subspace, \( W \), will be less than \( \delta \). This quantity is related to the probability that a singular value for a random matrix with independent rows, distributed according to \( f \), is small. Taking the supremum over \( W \in \text{Gr}_{n,d} \) yields, in some sense, the worst subspaces \( W \) where a projection is likely to be small causing a singular value for the random matrix to also be small. Thus, this is a quantity one would like to control. Def 2.2 allows one to express Def. 2.1 without reference to any subspace, without too much loss of precision. It yields a condition that is easier to check in that the role of \( W \) disappears. Moreover, it still allows for \( f \) to have a pole of order at most \( d-1 \) at the origin. We have the following two lemmas:

**Lemma 2.3.**

\[
\| f \|_{d,\delta,1} \leq \| f \|_{d,\delta,2}
\]

and

\[
\| f \|_{W,\delta,1} \leq \| f \|_{1} \frac{1}{\omega_d \delta^d}
\]
Proof. The second inequality follows from the definition of the average, Fubini and the fact that $\text{Vol}(B_{W^\perp}(\delta)) = \omega_d \delta^d$.

To get the first inequality, note that
\[
\|f\|_{W,\delta,1} = \int_W \text{Av}_{x+B_{W^\perp}(\delta)}(f) \, dx
\]
\[
= (n-d)\omega_{n-d} \int_0^\infty \text{Av}_{S_w(t) \times B_{W^\perp}(\delta)}(f) \, t^{n-d-1} dt
\]
\[
\leq \|f\|_{L^\infty(y+B_{W^\perp}(\delta))} \|L^1(W)
\leq (n-d)\omega_{n-d} \int_0^\infty \|f\|_{L^\infty(S_w(t) \times B_{W^\perp}(\delta))} \, t^{n-d-1} dt
\]

For any $t > 0$, and $W \in \text{Gr}_{n,n-d}$,
\[
S_w(t) \times B_{W^\perp}(\delta) \subset \bigcup_{t' \in [t,\sqrt{n^2+\delta^2}]} S^{n-1}(t')
\]
Hence the essential sup of $f$ on $S_w(t) \times B_{W^\perp}(\delta)$ will be achieved on $S^{n-1}(t')$ for some $t' \in [t, \sqrt{n^2+\delta^2}] \subset [t, t + \delta]$. \hfill \qed

**Lemma 2.4.** (Small Eigenvalue Lemma) Let $f_X$ be the PDF for the random variable $X$ taking values in $\mathbb{C}^n$. For any $W \in \text{Gr}_{n,n-d}$ we have
\[
\mathbb{P}(\|\text{proj}_W(X)\| \leq \delta) \leq \omega_d \|f_X\|_{W,\delta,1} \delta^d.
\]

**Proof.** For fixed $n-d$-dimensional subspace $W \subset \mathbb{C}^n$ we let
\[
H_W(\delta) = \{x \in \mathbb{C}^n : \|\text{proj}_{W^\perp}(x)\| \leq \delta\}
\]
Note that $H_W(\delta)$ is just the $\delta$-neighborhood of $W$. So we may write $H_W(\delta) = W \times B_{W^\perp}(\delta)$.

Hence we may write,
\[
\mathbb{P}(\|\text{proj}_{W^\perp}(X)\| \leq \delta) = \int_{H_W(\delta)} f_X(x) \, dx
\]
\[
= \int_W \int_{B_{W^\perp}(\delta)} f_X(y+z) \, dz \, dy
\]
\[
= \omega_d \delta^d \int_W \text{Av}_{y+B_{W^\perp}(\delta)}(f) \, dy.
\]

Taking suprema over $W$ in the previous lemma, we have the following.

**Corollary 2.5.** If $X$ has probability distribution $f_X$ on $\mathbb{C}^n$ then
\[
\mathbb{P}(\inf_{W \in \text{Gr}_{n,n-d}} \|\text{proj}_{W^\perp}(X)\| \leq \delta) \leq \omega_d \|f_X\|_{d,\delta,1} \delta^d.
\]

Note that since $\|f_X\|_1 = 1$ we have from Lemma 2.3 that $\|f_X\|_{d,\delta,1} \leq \frac{1}{\omega_d \delta^d}$. However, we want to exploit the $\delta$-decay in Lemma 2.12 below, so we would like a bound $\|f_X\|_{d,\delta,1} \leq C$ which is independent of $n$ or $\delta$.

We first describe some examples where this doesn’t happen. That is, where $\|f_X\|_{1,\delta,1} \geq \frac{C}{\delta^d \omega_d}$, and is therefore unsuitable for the estimates we need.
Example 2.6. Suppose that $X$ has iid entries each of which is a Bernoulli variable $f_i$ with point masses at $\pm 1$. Then consider any choice of codimension $d$ plane $W^\perp$ which passes through the origin and through $\frac{1}{2}\delta$ of the vertices, $1, -1^n$, of the $n$-cube of sidelength 2. Such a plane can be chosen as the span of any $n - d$ independent vectors with entries $\pm 1$. In this case, $\|fx\|_{W,\delta,1} \leq \frac{1}{2^d}\delta^d$.

A more obvious problem is the following.

Example 2.7. Consider a random vector $X \in \mathbb{C}^n$ whose PDF $f_X$ is concentrated completely in an $\epsilon$ neighborhood of $W^\perp = \mathbb{C}^{n-d}$, i.e. it’s support is $\mathbb{C}^{n-d} \times B_{\mathbb{C}^d}(\epsilon)$. We could even have $f_X$ be bounded, and then

$$\|fx\|_{d,\delta,1} \geq \|fx\|_{W,\delta,1} = \frac{\delta^d}{\epsilon^d} \|fx\|_1 = \frac{1}{\omega_d \delta^d}.$$ 

In particular, this is bounded but arbitrarily badly as $\epsilon \to 0$.

We can generalize this last example considerably to obtain differing behaviors.

Example 2.8. Now consider a fixed subspace $W$ of dimension $d$ and $f_X : W^\perp \times W \to [0,\infty)$ of the form $f_X(x,y) = C_{X^* + B_W(g(||x||))}$ for some function $g : (0,\infty) \to (0,\infty)$. Employing polar coordinates, the condition that $\|fx\|_1 = 1$ becomes the condition

$$C = \left((n-d)\omega_{n-d}\omega_d\int_0^\infty r^{n-d-1}g(r)^d dr\right)^{-1}$$

and we may evaluate

$$\|fx\|_{W,\delta,1} = \frac{1}{\omega_d \delta^d} \left[1 - C(n-d)\omega_{n-d}\omega_d\int_0^{g^{-1}(\delta)} r^{n-d-1}(g(r)^d - \delta^d) dr\right].$$

Now we can specialize to the case

$$g(r)^d = r^{1+d-n} \begin{cases} 1 & r \leq 1 \\ r^{-\alpha} & r > 1 \end{cases}$$

for some choice of $\alpha \in (0,\infty)$. We may explicitly compute $\frac{1}{C} = (n-d)\omega_{n-d}\omega_d(1 + \frac{1}{\alpha})$ and for $\delta < 1$, $g^{-1}(\delta) = \delta^{\frac{d}{n+\alpha-a}}$. Another explicit computation gives

$$\omega_d \delta^d \|fx\|_{W,\delta,1} = \frac{1}{\alpha} + \frac{1}{n-d} \delta^{\frac{dn}{n+\alpha-a}}.$$ 

As we need to take $\delta = n^{-\frac{\beta}{2}} - \beta$ for some small $\beta > 0$ (see Theorem 2.13), if we take $\alpha$ sufficiently close to 0 then the above decays slower than $n^{-\frac{\beta}{2}}$ and our necessary estimates fail. On the other hand for sufficiently large $\alpha$ the right hand side approaches $\frac{\delta^d}{n-d}$ and this is eventually smaller than $n^{-\frac{\beta}{2}} = O(n^{-\frac{\beta}{2}})$, and the required estimates succeed.

We now describe some general cases where $\|f\|_{d,\delta,1}$ is bounded independently of $\delta$. For $0 \leq d \leq n$, let $Gf(a,d)$ be the affine Grassmanian of all $d$-dimensional affine spaces of $\mathbb{C}^n$. Define the
generalized Radon transform of \( f \in L^1(\mathbb{C}^n) \) as the measurable function \( \mathcal{R}_d(f) : \text{Graff}_{n,d} \to \hat{\mathbb{C}} \) given by

\[
\mathcal{R}_d(f)(W) = \int_W f.
\]

(Here the measure is the Lebesgue measure and we must both allow for infinite values and accept that the Radon Transform is not invertible on all of \( L^1(\mathbb{C}^n) \).)

**Proposition 2.9.** If \( f_X \) satisfies any of the following conditions,

1. The row \( X \) consists of independent entries with PDF \( f_i \) (not necessarily i.d.) and \( \|f_i\|_\infty < \frac{C^2}{\sqrt{d}} \).

2. \( \left\| t^{n-d-1} \sup_{S^{n-1}(t)} f_X \right\|_{L^1((0,\infty))} < \frac{C}{(n-d)\omega_{n-d}} \).

3. \( \|\mathcal{R}_{n-d}(f_X)\|_\infty \leq C \).

Then \( \|f_X\|_{d,\delta,1} < C \).

**Proof.** For the first condition note that \( f_X(x) = \prod f_i(x_i) \), and suppose each is bounded by \( A = \frac{C^2}{\sqrt{d}} \).

For any choice of \( W^\perp \), and \( y \in B_{W^\perp}(\delta) \), there is a projection \( P \) onto one of the \( \binom{n}{n-d} \) choices of \( n-d \)-coordinate planes, say the first \( n-d \) coordinates, so that for all \( w \in W, d(P(w), w) \leq \|P(w)\| \). In other words, \( y + W \) is the graph of a linear map \( L : \mathbb{C}^{n-d} \to \mathbb{C}^d \) followed by a translation where \( \|L\|_{\text{op}} \leq 1 \). The volume distortion of the corresponding graph map \( (I, L) : \mathbb{C}^{n-d} \to \mathbb{C}^n \) is then

\[
\sqrt{\det(I + L^*L)} \leq \sqrt{2}^d.
\]

We may then write for any \( y \in B_{W^\perp}(\delta) \),

\[
\begin{align*}
\int_W f_X(y + w)dw &= \int_W \prod_{i=1}^n f_i(w_i)dw \\
&\leq \int_{R^{n-d}} \prod_{i=1}^{n-d} f_i(x_i) \prod_{i=n-d+1}^n f_i((Lx)_i) \sqrt{\det(I + L^*L)}dw \\
&\leq \int_{R^{n-d}} \prod_{i=1}^{n-d} f_i(x_i) A^d \sqrt{2}^d dw \\
&\leq \sqrt{2}^d A^d = C.
\end{align*}
\]

Here we have used that the \( f_i \) are individually PDF’s. Finally, taking the average over \( y \in B_{W^\perp}(\delta) \) does not change this.

The second statement amounts to \( \|f_X\|_{d,\delta,2} < C \), and the statement follows by Lemma 2.3. The last condition states that for any \( W \in \text{Graff}_{n,n-d} \), every translate of \( W \) has \( f_X \) integral bounded by \( C \). Hence

\[
\|f_X\|_{d,\delta,1} = \frac{1}{\omega_d \delta^d} \int_{B_{W^\perp}(\delta)} \int_W f_X(x + y)dxdy \leq \frac{1}{\omega_d \delta^d} \int_{B_{W^\perp}(\delta)} Cdy = C.
\]

\[\square\]

It will turn out that Proposition 2.8 gives sufficient criteria for the singular values of a random matrix whose rows are independently drawn from \( f \) to be small enough for the Circular Law to possibly hold. However, Example 2.6 demonstrates that when all of these conditions fail the Circular Law may still hold. Indeed, in the case of a random matrix of i.i.d. Bernoulli entries,
Tao and Vu (TV07) have showed the circular law holds. On the other hand, there is no bound of the form (1),(2) or (3) from Proposition 2.8 for a Bernoulli random vector, even with independent entries.

The issue at hand, of course, is that the Bernoulli random vector has a support measure with atoms that have unbounded Lebesgue integrals on lower dimensional slices. By taking the supremum over $W$ in in Definition 2.1 to arrive at Definition 2.2, we select the worst case $W$ for our purposes and, in the case of unbounded $f$, are doomed. To deal with this, we may instead take the expectations over $W$.

**Definition 2.10.**

$$\|f\|_{d,\delta} = \mathbb{E}\left[\|f\|_{W,\delta,1}\right]$$

where the expectation is taken over $W$ spanned by $n - d$ vectors $X_1, \ldots, X_{n-d}$ with the corresponding joint expectation induced from $f_{X_1} = f$. (We will usually assume that the vectors are chosen independently so that the joint distribution simplifies.)

More generally if we have $n$ vectors $X_1, \ldots, X_n$ in $\mathbb{C}^n$ randomly chosen via corresponding independent distributions $f_1, \ldots, f_n$ then we define

$$\|f_1, \ldots, f_n\|_{d,\delta} = \sup \mathbb{E}\left[\|f_i\|_{W,\delta,1}\right]$$

Where the sup is over all $i$ and all measures $\mu_i$ and $\mu_i'$ on $\text{Gr}_{n,n-d}$. Here $\mu_i$ for $1 \leq k \leq \binom{n-1}{n-d}$ is the measure induced on $\text{Gr}_{n,n-d}$ from the distributions $f_{i_1}, \ldots, f_{i_n}$ with $i \notin \{i_1, \ldots, i_{n-d}\}$ and $\mu_i'$ for $1 \leq l \leq \binom{n-1}{d}$ is the pushforward of $\mu_i$ on $\text{Gr}_{n,d}$ under the map $W \mapsto W^\perp$. More specifically, for $E \subset \text{Gr}_{n,n-d}$, $\mu_i(E)$ is the probability that the $n-d$ vectors with distributions $f_{i_1}, \ldots, f_{i_{n-d}}$ span a subspace in $E$ where the choice of indices come from the $k-th$ permutation. (Note that the probability that the span is lower dimensional is zero.) In what follows, a “random subspace” will mean one chosen with respect to one of these distributions.

**Remark 2.11.** Recall, that the map $V \mapsto V^\perp$ induces an isometry between $\text{Gr}_{n,d}$ and $\text{Gr}_{n,n-d}$. However this map does not necessarily push forward the measure $\mu_i$ on $\text{Gr}_{n,d}$ to any of the $\mu_i$ on $\text{Gr}_{n,n-d}$. This is why we had to use sup over the $\mu_i'$ as well.

**Lemma 2.12.** (Strong Small Eigenvalue Lemma) Let $X \in \mathbb{C}^n$ be a random vector with distribution $f_X : \mathbb{C}^n \rightarrow [0,\infty)$ and for $W \in \text{Gr}_{n,n-d}$ chosen at random. Then,

$$\mathbb{P}\left(\|\text{proj}_W(X)\| \leq \delta\right) \leq \omega_d\delta^d \|f_X\|_{d,\delta}.$$

(Here the probability on the left is over both $X$ and $W$.) In particular, if $X_1, \ldots, X_n$ are rows of an $n \times n$ matrix, with possibly distinct distributions $f_{X_i}$ and $\sigma$ is a permutation of $\{1, \ldots, n\}$ Then for $1 \leq j \leq d$,

$$\mathbb{P}\left(\|\text{proj}_W(X_{\sigma(j)})\| \leq \delta\right) \leq \omega_d\delta^d \|f_{X_1}, \ldots, f_{X_n}\|_{d,\delta}.$$

where the left hand side is the probability over $X_{\sigma(j)}$ and all $(n-d)$-subspaces $W$ of the form $W = X_{\sigma(d+1)} \wedge \cdots \wedge X_{\sigma(n)}$ with their corresponding induced distribution.

**Proof.** Taking expectations in $W$ on both sides of the inequality from Lemma 2.4 we have

$$\omega_d\delta^d \mathbb{E}_W[\|f_X\|_{W,d,1}] \geq \mathbb{E}_W[\mathbb{P}_{X|W}(\|\text{proj}_W(X)\| \leq \delta|W)] = \mathbb{P}_{X,W}(\|\text{proj}_W(X)\| \leq \delta),$$

where the last equality follows by Fubini.
The last statement follows from taking \( X = X_{\sigma(j)} \) and the distribution on \( \text{Gr}_{n,n-d} \) induced by the map \((X_{\sigma(d+1)}, \ldots, X_{\sigma(n)}) \mapsto W = X_{\sigma(d+1)} \wedge \ldots \wedge X_{\sigma(n)} \) (which is well defined off of a measure zero subset). \( \square \)

If \( \omega_d \delta^d \| f \|_{d,\delta} \leq C(n, \delta) \) we will need to bound the case that \( \delta = n^{-\frac{5}{2}} - \beta \) for some small \( \beta \), by the decay rate \( C(n, n^{-\frac{5}{2}} - \beta) \leq Cn^{-\frac{5}{2}} \) for some universal constant \( C \). (See Theorem 2.13 below.)

The following theorem demonstrates that when \( \| f \|_{d,\delta} \) is suitably bounded, then the singular values of a random matrix with iid rows drawn from \( f \) are almost surely nonzero.

**Theorem 2.13. (Least Singular Value):**

Let \( A_n \) be a random matrix with rows \( X_i \) drawn from multivariate distributions \( f_i \) and suppose that \( \| f_1, \ldots, f_n \|_{1,\delta} \) is bounded for \( \delta = \frac{1}{n^{\frac{5}{2}} \log(n)} \) independently of \( n \). Then the smallest singular value of \( A_n \) is almost surely greater than \( \frac{1}{n^{\frac{5}{2}} \log(n)} \) as \( n \to \infty \).

**Proof.** Denote the rows of \( A_n \) by \( X_i \) and lowest singular value of \( A_n \) by \( \sigma_n \). Recall that

\[
\sigma_n \geq \frac{1}{\sqrt{n}} \min_{i \leq n} (\text{dist}(X_i, H_i))
\]

where \( H_i = \text{span}\{X_j\}_{j \neq i} \). Thus, given \( \delta > 0 \),

\[
\mathbb{P}(\sigma_n \leq \delta) \leq \sum_{i=1}^{n} \mathbb{P}(\text{dist}(X_i, H_i) \leq \delta \sqrt{n})
\]

\[
= \sum_{i=1}^{n} \mathbb{E}_{H_i}[\mathbb{P}_{X_i|H_i}(\text{dist}(X_i, H_i) \leq \delta \sqrt{n}|H_i)]
\]

\[
\leq n \max_{i} \mathbb{E}_{H_i}[\mathbb{P}_{X_i|H_i}(\text{dist}(X_i, H_i) \leq \delta \sqrt{n}|H_i)]
\]

\[
\leq n \max_{i} E_{W \in Gr_{n,n-1}} [\mathbb{P}_{X_i|W}(\|\text{proj}_{W^\perp}(X_i)\| \leq \delta \sqrt{n}|W)]
\]

\[
\leq n^3 \delta \max_{i} E_{W \in Gr_{n,n-1}} [\|f_i\|_{W,\delta,1}]
\]

\[
\leq n^3 \delta \omega_1 \|f_1, \ldots, f_n\|_{1,\delta}
\]

Note here that the measure on the Grassmanian on the fourth and fifth line will depend on \( i \) unless the vectors are identically distributed.

Setting \( \delta = \frac{1}{n^{\frac{5}{2}} \log(n)} \), the last line is \( O(\frac{1}{n \log(n)}) \) and the result follows from the Borel-Cantelli Lemma. \( \square \)

**Remark 2.14.** Note that the above theorem does not rely on the independence of the rows, but the case where each \( f_i \) limits sufficiently fast to the same single Dirac measure shows that the condition on the norm is an essential hypothesis.

We now give a family of examples where the rows are equidistributed, i.e. \( f_i = f_X \), and \( \|f_X\|_{1,\delta,1} \geq \frac{C}{\delta \omega_d} \), but where we still have \( \|f_X\|_{d,\delta} < C \) uniformly in \( n \) provided \( \delta = o(n^{-1}) \). Hence the asymptotic bound on the lowest singular value still applies.
Example 2.15. Suppose that $X$ has iid entries each of which is a Bernoulli variable distribution with point masses of weight $p$ at 0 and $(1 - p)$ at 1. The resulting distribution in $C^n$ is supported on the vertices of the $n$-cube with side-length 1 whose vertices are all binary vectors of length $n$. Now consider any choice of codimension $d$ plane $W \perp$ which passes through the origin and is spanned by any choice of $n - d$ distinct coordinate vectors $e_i$. (Note these are admissible in that each has a positive probability of being a row of a random $n \times n$ matrix.) In this case, $W \perp$ contains a total of $2^{n-d}$ vertices, for a total mass of $p^d$. In this case, $\|f_X\|_{1,\delta,1} \geq \|f_X\|_{W,\delta,1} = \frac{p^d}{\omega_{s\delta}^d}$, which explodes as $\delta \to 0$. Hence we cannot achieve a useful bound for this “worst case” choice of $W \perp$.

On the other hand, for the case $p = \frac{1}{2}$ at least, the main result of [TV07] states that the probability that a random matrix with iid $p = \frac{1}{2}$-Bernoulli $\{-1,1\}$-entries is singular is $(\frac{3}{4} + o(1))^n$. A standard procedure using row and column operations produces from an $(n+1) \times (n+1)$ $\{-1,1\}$-matrix a new matrix with $n \times n \{0,1\}$-lower submatrix and first column $e_1$ (see e.g. [Orr05]). Hence the probability that a random iid $\frac{1}{2}$-Bernoulli $\{0,1\}$-matrix is singular is also $(\frac{3}{4} + o(1))^{n+1} \leq (\frac{3}{4} + o(1))^n$.

A plane $P$ spanned by linearly independent $\{0,1\}$-vectors $v_1, \ldots, v_{n-d}$ contains the $\{0,1\}$-vector $v$ if and only if for all choices of $\{0,1\}$-vectors $w_1, \ldots, w_{d-1}$ the $n \times n \{0,1\}$-matrix $A_n = [v_1, \ldots, v_{n-1}, v, w_1, \ldots, w_{d-1}]$ is singular.

The condition that $v_1 \wedge \cdots \wedge v_{n-d} \neq 0$ is the condition that there be $\{0,1\}$-vectors $w_1, \ldots, w_d$ such that $B_n = [v_1, \ldots, v_{n-d}, w_1, \ldots, w_d]$ be nonsingular. Note that a subset of this last condition is the case that the $(n-d) \times (n-d)$ minor, denoted $C_{n-d}$, be nonsingular. The probability that $C_{n-d}$ be nonsingular is the same that $A_{n-d}$ be nonsingular. So the probability that $v \in P$ for a randomly chosen plane $P \in \text{Gr}_{n-d}(C^n)$ can be estimated as

$$
\mathbb{P}(v \in P) \leq \frac{\mathbb{P}(\det(A_n) = 0)}{1 - \mathbb{P}(\det(B_n) = 0 \forall w_1, \ldots, w_d)} \leq \frac{\mathbb{P}(\det(A_n) = 0)}{\mathbb{P}(\det(A_{n-d}) \neq 0)} \leq \frac{(\frac{3}{4} + o(1))^n}{1 - (\frac{3}{4} + o(1))^{n-d}} = (\frac{3}{4} + o(1))^n.
$$

Now if we let $P = W \perp$ be the plane spanned by $n - d$ randomly chosen rows of our random $n \times n$ matrix we note that the nearest distance to the plane $W \perp$ of a vertex not in the plane is the distance of the origin to the standard $n$-simplex, namely $\frac{1}{n}$. Hence, letting $W \perp$ vary over all choices of rows, and for $\delta < \frac{1}{n}$ we obtain that $\|f_X\|_{d,\delta} \leq (\frac{3}{4} + o(1))^n$. Since we will be taking $\delta < n^{-\frac{1}{2}}$ we obtain the desired bound. Note the same estimate still holds even when $d$ is allowed to grow in $n$ provided $d = o(n)$.

Lastly we consider a case where we allow coordinate-wise dependency. Suppose the random Bernoulli vector $X$ has coordinate wise probability $\frac{1}{2}$ of being 0 or 1 but has symmetric $n \times n$ covariance matrix $\text{Cov}(X) = [C_{ij}]$ with entries $C_{ij} \in [-\frac{1}{4}, \frac{1}{4}]$ and $C_{ii} = \frac{1}{4}$, i.e. the joint pairwise probabilities on coordinates $i$ and $j$ being 1 are $p_{ij} = C_{ij} + \frac{1}{4}$, but are otherwise independent. The vector $W \cdot X$, where $W$ is the Whitening Matrix such that $W^TW = C^{-1}$, has covariance the identity matrix. Provided the entries of the off-diagonal entries $C_{ij}$ are uniformly bounded away from $\frac{1}{4}$, and $\delta = o(\frac{1}{n})$, then we will have identical probability of $X$ being in the $\delta$-neighborhood of $W^{-1}P$ as for $WX$ in some $O(\delta)$-neighborhood of $P$. In particular $\|f_X\|_{d,\delta} \leq (\frac{3}{4} + o(1))^n$ provided $\delta = o(\frac{1}{n}).$
3. Generalized Marčenko-Pastur

In this section, we discuss a key ingredient of the Circular Law: a corresponding limiting law for symmetric random matrices. Given $A_n$, the shifted and symmetrized version of $A_n$ is

$$H_n = H_n(z) = \left( \frac{1}{\sqrt{n}}A_n - zI \right)\left( \frac{1}{\sqrt{n}}A_n - zI \right)^*$$

for $z \in \mathbb{C}$.

One desires that the spectral distribution of $H_n$ converges in the large $n$ limit to a deterministic measure, independent of the entries of $A_n$. Or, as in our case, independent of the probability measure in $\mathbb{C}^n$ (resp. $\mathbb{C}^n$) that the rows of $A_n$ are independently drawn from.

The classical version of this result is the Marčenko-Pastur law and does not assume that $A_n$ is square to begin with. Moreover, the Marčenko-Pastur law deals with iid entries and with the particular case of shifting by $z = 0$. We state it here for the reader.

**Theorem 3.1. (Marčenko-Pastur Law)** Let $A_n$ be a $n \times N_n$ random matrix with iid entries that satisfy $\mathbb{E}[x_{ij}] = 0$ and $\mathbb{E}[x_{ij}^2] = 1$ and suppose that $\frac{N_n}{n} \to c \in (0, 1]$ as $n \to \infty$. Denote by $\mu_n$ the spectral distribution of $\frac{1}{n}A_nA_n^*$. Then, $\mu_n \to \mu$ almost surely, where $\mu$ is a deterministic measure given by

$$\frac{d\mu}{dx} = \frac{1}{2\pi xy} \sqrt{(b-x)(x-a)} 1_{a \leq x \leq b}$$

where $a = (1 - \sqrt{y})^2$ and $b = (1 + \sqrt{y})^2$.

The Marčenko-Pastur Law has many methods of proof: combinatorial, methods using the Stieltjes (Cauchy) transform, and methods using free probability. The combinatorial proof is the most readily generalized to our setting and has been utilized by other authors to extend the Marčenko-Pastur Law to have certain amounts of dependency (Adamczak) and we will follow similar suit, but with some different assumptions.

Let $(N_n)_{n \geq 1}$ be a sequence of positive integers such that $\lim_{n \to \infty} n/N_n = y \in (0, \infty)$. Recall we have the following assumptions:

(A1) for every $k \in \mathbb{N}$, $\sup_{n} \max_{1 \leq i \leq n, 1 \leq j \leq N_n} \mathbb{E}[|x_{ij}^{(n)}|^k] < \infty$;

(A2) for every $k \in \mathbb{N}$, the sum of all terms in $\text{Mom}_2k(A_n)$ with at least one $x_{ij}$ appearing with a power of 1 is of size $o_k(n^{k+1})$;

(A3) for every $\epsilon > 0$:

$$\lim_{n \to \infty} \frac{1}{n} \sum_{i \leq n} \mathbb{P} \left( \left| \frac{1}{N_n} \sum_{j=1}^{N_n} (x_{ij}^{(n)})^2 - 1 \right| \geq \epsilon \right) = 0$$

and

$$\lim_{n \to \infty} \frac{1}{N_n} \sum_{j \leq N_n} \mathbb{P} \left( \left| \frac{1}{n} \sum_{i=1}^{n} (x_{ij}^{(n)})^2 - 1 \right| \geq \epsilon \right) = 0.$$
We need the following theorem for proof of the Circular Law. We will defer the proof of this result to Subsection 3.2.

**Theorem 3.2 (cf. Theorem 2.4 [Ada11]).** Assume that $N_n = n$ and $A_n$ is a sequence of random matrices with rows independently drawn from a probability measure in $C^n$ given by $f^{(n)}$. Assume that $A_n$ satisfies assumption A1-A3. Then for any $k \in \mathbb{N}$,
\[
\lim_{n \to \infty} \frac{1}{n} \mathbb{E}[\text{tr} H_n^k] = \mu_k(|z|^2),
\]
where $\mu_k(|z|^2)$ is a function depending only $|z|^2$ and not on the distribution of $H_n$.

**Corollary 3.3 (cf. Corollary 2.5 [Ada11]).** Let $A_n$ be as in Theorem 3.2 and let $L_n(z)$ be the spectral measure of $H_n(z)$. For every $z \in \mathbb{C}$, $L_n(z)$ converges weakly to a non-random measure which does not depend on the distribution of the rows of $A_n$.

**Proof.** The proof of this corollary is verbatim identical to the proof of Cor 2.5 of [Ada11] except for the replacement of Theorem 2.4 of [Ada11] with Theorem 3.2 above. □

3.1. **Combinatorial Trees.** In what follows, we will use certain combinatorial structures to keep track of distinct classes of terms. A detailed background for these structures can be found in Chapter 3 of [BS10]. We will closely follow the notation of [Ada11].

Let $T = (V, E, r)$ be a rooted tree. A $\Gamma$-tree is a rooted tree having the following structure:

- The set $V$ is partitioned into two sets $S$ and $O$, denoting special and ordinary vertices.
- Every edge adjacent to a special vertex is given an orientation so that
  - For any $u, w \in S$ such that on the path $u = v_0 v_1 \ldots v_m = w$ connecting $u$ and $w$, we have that $v_1, \ldots, v_{m-1} \in O$. If $m$ is odd, then the orientations of the first and the last edge on this path are the same. That is to say, one has $(u \rightarrow v_1$ and $v_{m-1} \rightarrow w)$ or $(v_1 \rightarrow u$ and $w \rightarrow v_{m-1})$. If $m$ is even, then the orientation of the first and the last edge in the path are opposite.
  - If $r \in O$, then for any $u \in S$ such that $u$ is the only special vertex on the path $r = v_0 v_1 \ldots v_m = u$, one has $v_{m-1} \rightarrow u \iff m$ is odd.

Given the orientation of paths between special vertices, we can partition $O$ into two sets $U$ and $D$. Let $u \in O$ and $r = v_0 v_1 \ldots v_m = u$ be a path from the root $r$ to $u$:

- If $r \in O$ and $v_1, \ldots, v_{m-1} \in O$, then $u \in D \iff m$ is odd. Otherwise, $u \in U$.
- If $v_l$ is the last special vertex on the path, then $u \in D \iff (m - l$ is odd and $v_l \rightarrow v_{l+1}$) or $(m - l$ is even and $v_{l+1} \rightarrow v_l$).

Note that every edge which has ends that are ordinary vertices must have one end in $D$ and one end in $U$ and we can assign to each edge an orientation $u \rightarrow v$ where $u \in U$ and $v \in D$. We write $e = (u \rightarrow v)$.

Let $A_n$ be a sequence of random $n \times n$ matrices with $A_n = [a_{ij}^{(n)}]$ and let $T$ be a $\Gamma$-tree. Let $I_T^n$ be the set of functions $i = (i_v)_{v \in V} : V \rightarrow \{1, \ldots, n\}$ such that if $A$ is one of the sets $D \cup S$, $U \cup S$ then
∀u, v ∈ A, u ≠ v  ⇒  i_u ≠ i_v. And, for every e = (u → v) ∈ E, i_u ≠ i_v.

We define

\[ \xi_n(T) = n^{-|E|-1} \mathbb{E} \left( \sum_{i \in I^n_e} \prod_{e = (u \rightarrow v) \in E} (x_{i_{u,v}}^{(n)})^2 \right) \]

We will prove the following proposition necessary to prove Theorem 3.2.

**Proposition 3.4.** Let \( N_n = n \) and let \( A_n \) be as in Theorem 3.2. Then for every \( \Gamma \)-tree \( T \),

\[ \lim_{n \to \infty} \xi_n(T) = 1 \]

**Proof.** We prove by induction on the size of the tree. If \(|V| = 1\) then it is clear that \( \xi_n(T) = 1 \) for all \( n \).

Suppose that the proposition holds for all trees of size \( s \leq m - 1 \) and that \( T \) is a tree of size \( m \). Consider an arbitrary leaf \( w \) of the tree \( T \), where \( w \) is not the root and let \( x \) be the unique neighbor of \( w \).

We consider the case where \( w \in D \) (The proof when \( w \in U \) follows similarly so we omit it). Let \( \tilde{T} = (\tilde{V}, \tilde{E}, r) \) be the tree obtained from \( T \) by deleting \( w \) along with the edge \( e = (x \rightarrow w) \).

Let \( \tilde{I}^n \) be the set of multi-indices \( i_\psi : \tilde{V} \to \{1, ..., n\} \) which can be extended to a multi-index \( i_V = (i_\psi, i_w) \in I^n_T \). Denote \( U(i_\psi) = \prod_{e = (u \rightarrow v) \in E} (x_{i_{u,v}}^{(n)})^2 \). We have

\[ \xi_n(T) = n^{-|U|-|D|} \sum_{i_\psi \in \tilde{I}^n} \sum_{i_w \in I^n} \mathbb{E} \left( (x_{i_{x,w}}^{(n)})^2 U(i_\psi) \right) \]

For large enough \( n \), \( I^n_T = \tilde{I}^n \), and there are only \(|D| - 1\) choices for \( i_w \) such that \((i_\psi, i_w) \notin I^n_T \). By A1, we have that \( \mathbb{E}(x_{ij}^{(n)}^2) \) is bounded for all \( i, j \), independent of \( n \). Thus, by generalized Hölder’s inequality, for every such \( i_w \) we have that \( \mathbb{E}((x_{i,w}^{(n)})^2 U(i_\psi)) \) is bounded by a number independent of \( n \). Thus, for large enough \( n \), we have

\[ \xi_n(T) = n^{-|U|-|D|} \sum_{i_\psi \in \tilde{I}^n} \left( \sum_{i_w = 1}^n \mathbb{E} \left( (x_{i_{x,w}}^{(n)})^2 U(i_\psi) \right) + O_T(1) \right) \]

\[ = n^{-|U|-|D|} \sum_{i_\psi \in \tilde{I}^n} \sum_{i_w = 1}^n \mathbb{E} \left( (x_{i_{x,w}}^{(n)})^2 U(i_\psi) \right) + o_T(1) \]

where the constant depends on \( T \) and where in the last inequality we use the fact that \(|I^n_T| = n(n-1)...(n-|U|+1)n...(n-|D|+2) = O_T(n^{|U|+|D|-1})\).
Notice that,
\[
\left| n^{-|U|-|D|} \sum_{i_\varphi \in I_T^u} \sum_{i_{i_{\bar{\varphi}i_{\bar{\varphi}}} \in I_T^u}^{n} \mathbb{E} \left( (x_{i_{i_{\bar{\varphi}i_{\bar{\varphi}}}}^{(n)})^2 U(i_{\varphi}) \right) - \xi_n(T) \right|\n\leq n^{-|U|-|D|+1} \sum_{i_\varphi \in I_T^u} \mathbb{E} \left( n^{-1} \sum_{i_{i_{\bar{\varphi}i_{\bar{\varphi}}} = 1}^{n} (x_{i_{i_{\bar{\varphi}i_{\bar{\varphi}}}}}^{(n)})^2 - 1 \right) U(i_{\bar{\varphi}}) \right).\n\]
Moreover, using that \( y \leq \max \{ \epsilon, y \} \) for any \( \epsilon > 0 \) and then applying Cauchy-Schwarz, we have for every \( i_\varphi \in I_T^u \) that,
\[
\mathbb{E} \left( n^{-1} \sum_{i_{i_{\bar{\varphi}i_{\bar{\varphi}}} = 1}^{n} (x_{i_{i_{\bar{\varphi}i_{\bar{\varphi}}}}}^{(n)})^2 - 1 \right) U(i_{\bar{\varphi}}) \right) \leq \epsilon \mathbb{E} U(i_{\bar{\varphi}}) + \left\| n^{-1} \sum_{i_{i_{\bar{\varphi}i_{\bar{\varphi}}} = 1}^{n} (x_{i_{i_{\bar{\varphi}i_{\bar{\varphi}}}}}^{(n)})^2 - 1 \right\|_2 \mathbb{P} \left( \left| n^{-1} \sum_{j = 1}^{n} (x_{i_{i_{\bar{\varphi}i_{\bar{\varphi}}}}}^{(n)})^2 - 1 \right| \geq \epsilon \right)^{1/2}
\]
By A1, the triangle inequality in \( L_p \), and generalized Holder’s inequality, we have that
\[
\mathbb{E} U(i_{\bar{\varphi}}) \text{ and } \left\| n^{-1} \sum_{i_{i_{\bar{\varphi}i_{\bar{\varphi}}} = 1}^{n} (x_{i_{i_{\bar{\varphi}i_{\bar{\varphi}}}}}^{(n)})^2 - 1 \right\|_2
\]
are bounded by some constant \( C \), depending only on \( T \) and the bounds from A1. Thus, we get that
\[
|\xi_n(T) - \xi_n(\bar{T})| \leq \epsilon + C n^{-|U|-|D|+1} \sum_{i_\varphi \in I_T^u} (\epsilon + \mathbb{P}(\left| \sum_{j = 1}^{n} (x_{i_{i_{\bar{\varphi}i_{\bar{\varphi}}}}}^{(n)})^2 - n \right| \geq \epsilon n)^{1/2})
\]
Since for each \( i_x \) there are at most \( n^{|U|+|D|-2} \) multi-indices \( i_{\varphi \setminus (x)} \) such that \( i_\varphi = (i_{\varphi \setminus (x)}, i_x) \in I_T^u \), we get that
\[
|\xi_n(T) - \xi_n(\bar{T})| \leq (C + 1) \epsilon + C n \sum_{i_{i_{\bar{\varphi}i_{\bar{\varphi}}} = 1}^{n} \mathbb{P}(\left| \sum_{j = 1}^{n} (x_{i_{i_{\bar{\varphi}i_{\bar{\varphi}}}}}^{(n)})^2 - n \right| \geq \epsilon n)^{1/2}
\]
Using the Cauchy-Schwarz inequality and the first part of assumption A3 (note: for \( w \in U \), we simply use the second part of assumption A3 here), we get that
\[
\xi_n(T) - \xi_n(\bar{T}) = o_T(1)
\]
Thus, we have that
\[
\lim_{n \to \infty} \xi_n(T) = 1
\]
for trees of size \( m \) and the proof follows by induction.

\[\square\]

3.2. \( \Delta \) Graphs. To prove Theorem 3.2, we begin by introducing the notion of \( \Delta \) graphs. Here we follow closely the work outlined in Adamczak ([Ada11]) with slight modifications for our alternative assumptions.

For two sequences of integers \( i = (i_1, \ldots, i_k) \) and \( j = (j_1, \ldots, j_k) \), we define a \( \Delta \)-graph \( \Delta = G(i, j) \) as a bipartite graph \( (I_i, I_j, E) \) such that \( I_i = \{i_1, \ldots, i_k\} \) (the upper indices) and \( I_j = \{j_1, \ldots, j_k\} \) (the lower indices) and the set \( E \) of edges consisting of \( k \) directed edges from \( i_u \) to \( j_u \) and \( k \) directed
edges from $j_u$ to $i_{u+1}$, where we set $i_{k+1} = i_1$. We also label the edges from 1 to $2k$ in the order of $(i_1, j_1), (j_1, i_2), (i_2, j_2), ..., (i_k, j_k), (j_k, i_1)$. Note that $I_i$ and $I_j$ may not be disjoint, but their common elements are treated as different objects when considered as upper and lower vertices of the graph.

We would also like to partition into classes of up and down edges. An edge will be called perpendicular if its two end vertices are equal and skew if they are distinct. For any $\Delta$-graph $\Delta$, let $UP(\Delta)$ denote the up edges, $DP(\Delta)$ denote the down edges, and $S(\Delta)$ denote the skew edges.

**Definition 3.5.** Pairs $(i, j)$ and $(i', j')$ are isomorphic if there exist functions $f : I_i \rightarrow I_{i'}$ and $g : I_j \rightarrow I_{j'}$, such that for $u = 1, ..., k$ one has:

- $f(i_u) = i'_u, g(j_u) = j'_u$
- $f(i_u) = g(j_u) \iff i_u = j_u$
- $f(i_{u+1}) = g(j_u) \iff i_{u+1} = j_u$

**Definition 3.6.** $G(i, j)$ and $G(i', j')$ are isomorphic if and only if $(i, j)$ and $(i', j')$ are isomorphic. We write $G(i, j) \sim G(i', j')$ when the two graphs are isomorphic.

Let $\Delta(k)$ be a set of representatives of isomorphism classes of $\Delta$-graphs $G(i, j)$ with $i = (i_1, ..., i_k)$, $j = (j_1, ..., j_k)$, and $i, j \in \{1, ..., 2k\}$. Any graph based on two sequences of length $k$ is isomorphic to a graph in $\Delta(k)$.

**Definition 3.7.** Given $\Delta \in \Delta(k)$, we define $I^n_\Delta$ to be the set of all indices $i : V(\Delta) \rightarrow \{1, ..., n\}$ such that

- for any two upper indices $v, w$, we have $i_v \neq i_w$
- for any two lower indices $v, w$, we have $i_v \neq i_w$
- for any edge $i_{u(e)} = i_{d(e)} \iff e$ is perpendicular

Denote by $W_n = \frac{1}{\sqrt{n}} A_n - z I = (w_{ij})$, so that $H_n = W_n W_n^*$ where we now suppress the dependence on $n$ when we write the entries for ease of notation; we will also suppress this dependence for the entries $x_{ij}$. We now prove Theorem 3.2.
Proof. (of Theorem 3.2) We have

\[
\frac{1}{n} \mathbb{E}[\text{tr} H_n^k] = \frac{1}{n} \sum_{i_1, \ldots, i_k = 1}^{n} \sum_{j_1, \ldots, j_k = 1}^{n} \mathbb{E} w_{i_1j_1} \overline{w}_{i_2j_2} \cdots w_{i_kj_k} \overline{w}_{i_1j_1} = \frac{1}{n} \sum_{\Delta \in \Delta(k)} \sum_{\mathbf{i} \in I_\Delta^n} \mathbb{E} \left( \prod_{e \in S(\Delta)} w_{\mathbf{i}(\mathbf{e})} \prod_{e \in UP(\Delta)} \overline{w}_{\mathbf{i}(\mathbf{e})} \prod_{e \in DP(\Delta)} w_{\mathbf{i}(\mathbf{e})} \right)
\]

where \( \alpha = 1 + |S(\Delta)|/2 \). For a fixed \( \Delta \), let \( \Delta' \) be the graph obtained by replacing each pair of vertices connected with a perpendicular edge by one vertex and removing all corresponding perpendicular edges, while keeping all skew edges so that \( \Delta' \) is connected and has \( |S(\Delta)| \) edges. For this fixed \( \Delta \), each term in the sum over \( \mathbf{i} \in I_\Delta^n \) above is bounded by some constant in \( k, z \) as we have all bounded moments of the individual \( x_{ij} \).

As \( |I_\Delta^n| \leq n^{|V(\Delta')}| \), the graphs \( \Delta \) such that \( \Delta' \) has fewer than \( 1 + S(\Delta)/2 \) vertices have no asymptotic contribution. Note that in the case of \( z = 0 \), these are entries of \( \text{Mom}_k(A_n) \) that have \( x_{ij} \) terms with powers greater than or equal to 2, but not all equal to 2.

Moreover, for skew edges \( e = (v, w) \) of multiplicity 1, \( (w, v) \) is not an edge of \( \Delta \) and so the corresponding variable \( x_{i(\mathbf{e})j(\mathbf{e})} \) appears in the product exactly once. By assumption A2, the sum of these terms is also asymptotically negligible. Note that in the case of a random matrix with mean zero iid entries, these terms vanish automatically.

We are left with the graphs \( \Delta \) for which each skew edge \( e \), treated as an undirected edge, appears only twice and \( \Delta' \) has at least \( 1 + |S(\Delta)|/2 \) vertices. Let \( \Delta'' \) be the graph formed by identifying up and down edges of \( \Delta' \) that share the same endpoints. This implies that number of edges \( a \) of \( \Delta'' \) is at most \( |S(\Delta)|/2 \). If \( b \) is the number of vertices of \( \Delta'' \), we have that \( b \geq a + 1 \). Moreover, since \( \Delta'' \) is connected we have that \( b = a + 1 = |S(\Delta)|/2 + 1 \), and \( \Delta'' \) is a tree. Since the cycle in \( \Delta' \) inherited from \( \Delta \) corresponds to a walk in \( \Delta'' \) which goes through every vertex and returns to the starting vertex, it means all skew edges in \( \Delta \) appear exactly twice. We also have that among the perpendicular edges connected any two vertices of \( \Delta \), there are equal numbers of up and down edges. Thus, we can write

\[
\frac{1}{n} \mathbb{E}[\text{tr} H_n^k] = \sum_{\Delta \in \Delta(k); \Delta'' \text{is a tree}} |z|^{2|UP(\Delta)|/n} \sum_{\mathbf{i} \in I_\Delta^n} \mathbb{E} \prod_{e \in S(\Delta)} x_{\mathbf{i}(\mathbf{e})} + o_{k,z}(1)
\]
To each $\Delta$ such that $\Delta''$ is a tree, we can assign a $\Gamma$-tree $T(\Delta)$, where the special vertices are obtained by merging vertices of $\Delta$ connected by perpendicular edges and the orientation of edges is always from up to down. Using Proposition 3.4, we have

$$\frac{1}{n} \mathbb{E} \text{tr} H_n^k = \sum_{\Delta \in \Delta'(k): \Delta'' \text{is a tree}} |z|^{2|UP(\Delta)|} \xi_n(T(\Delta)) + o_k(z(1))$$

which completes the proof. \hfill \Box

4. Proof of Circular Law

In this section we prove the main theorem of this paper

**Theorem 4.1.** Let $A_n$ be a sequence of $n \times n$ random matrices with independent rows $X_1^{(n)}, \ldots, X_n^{(n)}$ defined on a common probability space and satisfying assumptions A1-A3. Assume that for each $n$ and $i, d \leq n$, and all $\delta > 0$, the probability measures $\nu_i^{(n)}$ for $X_i^{(n)}$ have uniformly bounded $\|\nu_i^{(n)}\|_{d, \delta}$. Then almost surely the spectral measure of $\mu_{\frac{1}{\sqrt{n}} A_n}$ converges weakly to the uniform distribution on the unit disk in $\mathbb{C}$.

To prove this theorem, we will use the following replacement principle for random matrices by Tao and Vu.

**Theorem 4.2 (Tao-Vu Replacement Principle).** Suppose for each $n$ that $A_n, B_n \in M_n(\mathbb{C})$ are ensembles of random matrices defined on a common probability space. Assume that

1. \[
\frac{1}{n} \left\| A_n \right\|^2 + \frac{1}{n^2} \left\| B_n \right\|^2
\]

is almost surely bounded

2. for almost all complex numbers $z$,

\[
\frac{1}{n} \log |\det(\frac{1}{\sqrt{n}} A_n - zI)| - \frac{1}{n} \log |\det(\frac{1}{\sqrt{n}} B_n - zI)|
\]

converges almost surely to zero.

Then $\mu_{\frac{1}{\sqrt{n}} A_n} - \mu_{\frac{1}{\sqrt{n}} B_n}$ converges almost surely to 0.

To use the replacement principle, note that if the rows of random matrices are distributed according to the multivariate Gaussian random variables with independent coordinates having mean zero and finite second moment, then the $\left\| v_1^{(n)}, \ldots, v_n^{(n)} \right\|_{d, \delta}$ is bounded and assumptions A1-A3 are satisfied. As matrices of this kind are known to have spectral distribution converging to the uniform distribution on the unit disk in $\mathbb{C}$, we need only show that the replacement principle holds for matrices of the kind described in Theorem 4.1.
We first verify the second condition in Theorem 4.2. We wish to show that for any \( z \in \mathbb{C} \), with probability one,
\[
\frac{1}{n} \log |\det(\frac{1}{\sqrt{n}}A_n - zI)| - \frac{1}{n} \log |\det(\frac{1}{\sqrt{n}}B_n - zI)| \to 0
\]
Denote the rows of \( \frac{1}{\sqrt{n}}A_n - zI \) by \( Z_1, Z_2, ..., Z_n \) and the rows of \( \frac{1}{\sqrt{n}}B_n - zI \) by \( Y_1, Y_2, ..., Y_n \). Denote by \( V_i \) the span of \( Z_1, Z_2, ..., Z_{i-1} \) and by \( U_i \) the span of \( Y_1, Y_2, ..., Y_{i-1} \). We then have that
\[
\frac{1}{n} \log |\det(\frac{1}{\sqrt{n}}A_n - zI)| = \frac{1}{n} \sum_{i=1}^{n} \log \text{dist}(Z_i, V_i)
\]
and
\[
\frac{1}{n} \log |\det(\frac{1}{\sqrt{n}}B_n - zI)| = \frac{1}{n} \sum_{i=1}^{n} \log \text{dist}(Y_i, U_i)
\]
and we wish to show that
\[
\frac{1}{n} \sum_{i=1}^{n} \log \text{dist}(Z_i, V_i) - \frac{1}{n} \sum_{i=1}^{n} \log \text{dist}(Y_i, U_i) \to 0
\]

Now, recall the following identity.

**Lemma 4.3.** Let \( 1 \leq k \leq n \) and \( M \) be a full rank \( k \times n \) matrix with singular values \( \sigma_1(M) \geq ... \geq \sigma_k(M) > 0 \) and rows \( X_1, ..., X_k \in \mathbb{C}^n \). For each \( 1 \leq i \leq k \), let \( W_i \) be the hyperplane generated by the \( k - 1 \) vectors \( X_1, ..., X_{i-1}, X_{i+1}, ..., X_k \). Then,
\[
\sum_{j=1}^{k} \sigma_j(M)^{-2} = \sum_{j=1}^{k} \text{dist}(X_j, W_j)^{-2}
\]

By Theorem 2.15 and the Borel-Cantelli lemma, we have that
\[
\log \text{dist}(Z_i, V_i), \log \text{dist}(Y_i, U_i) \geq -C_1 \log n
\]
with probability one for some constant \( C_1 > 0 \). Furthermore, by A3 (or from Proposition 4.5 proved independently below), we have that with probability one there exists a constant \( C_2 > 0 \) such that for large \( n \) and \( i \leq n \),
\[
\log \text{dist}(Z_i, V_i), \log \text{dist}(Y_i, U_i) \leq C_2
\]

Thus, to show the second condition of the replacement principle, it suffices to show that
\[
\frac{1}{n} \sum_{i=1}^{n-n^{0.99}} \log \text{dist}(Z_i, V_i) - \frac{1}{n} \sum_{i=1}^{n-n^{0.99}} \log \text{dist}(Y_i, U_i) \to 0
\]

Following Tao and Vu in [TV10], we show this with the following two lemmas.

**Lemma 4.4.** (High-dimensional contribution) There exists a constant \( C \), such that for every \( \epsilon \in (0, 1/4) \) and every \( \delta \in (0, \epsilon^2) \), with probability one, for sufficiently large \( n \),
\[
\frac{1}{n} \sum_{(1-\delta)n \leq i \leq n-n^{0.99}} (|\log \text{dist}(Z_i, V_i)| + |\log \text{dist}(Y_i, U_i)|) \leq C\epsilon
\]
Proposition 4.5 (Lower tail bound). Let $1 \leq d \leq n$ and $c > 0$, and let $V$ be a random $d$-dimensional subspace of $\mathbb{C}^n$. Let $X$ be a row of $A_n$. Then,
\[
P(\text{dist}(\sqrt{n}X, V) \leq c\sqrt{n-d}) = O(c^d \omega_d (1 - \frac{d}{n})^{\frac{d}{2}})
\]
where the constant depends on $c$. Moreover, the right hand side is less than $O((\frac{d}{\sqrt{\pi c}})^{-\frac{d}{2}})$.

Proof. Recall the definition of $\|f_{X_1}, \ldots, f_{X_n}\|_{d,\delta}$ and the (non-measure preserving) homeomorphism given by $V \mapsto V^\perp$. Then we have,
\[
P_{X, V \in \text{Gr}_{n,d}}(\text{dist}(\sqrt{n}X, V) \leq c\sqrt{n-d}) \leq P_{X, W \in \text{Gr}_{n,n-d}}(\|\text{proj}_W(\sqrt{n}X)\| \leq c\sqrt{n-d}) 
\leq P_{X, W \in \text{Gr}_{n,n-d}}(\|\text{proj}_W(X)\| \leq c\sqrt{1 - \frac{d}{n}}) 
\leq \omega_d \left(c\sqrt{1 - \frac{d}{n}}\right)^d \|f_{X_1}, \ldots, f_{X_n}\|_{d,\delta} \sqrt{1 - \frac{d}{n}}
\]
The last inequality follows from Lemma 2.12.

Since, $\|f_{X_1}, \ldots, f_{X_n}\|_{d,\delta} < C$ for all $\delta > 0$, we obtain the bound $O(\omega_d c^d (1 - \frac{d}{n})^{\frac{d}{2}})$.

The last statement follows from the definition of $\omega_d$ followed by an application of Sterling’s estimate.

Now, since with probability one, $V_i$ is of dimension $i - 1$, and $Z_i$ and $V_i$ are independent of each other, the proposition implies that
\[
dist(Z_i, V_i) \geq \sqrt{n-i+1}
\]
for each $(1 - \delta)n \leq i \leq n - n^{0.99}$, with probability $1 - O(n^{-10})$, say. Setting $\delta$ sufficiently small, compared to $\epsilon$, taking logarithms and summing in $i$ and $n$, one obtains the result. \hfill \square

Lemma 4.6. (Low-dimensional contribution) There exists a constant $C$, such that for every $\epsilon \in (0, 1/4)$ and every $\delta_0 > 0$ such that for every $\delta \in (0, \delta_0)$, with probability one, for sufficiently large $n$,
\[
\left|\frac{1}{n} \sum_{1 \leq i \leq (1-\delta)n} (\log \text{dist}(Z_i, V_i) - \log \text{dist}(Y_i, U_i))\right| \leq C\epsilon
\]
Proof. Let \( n' = \lfloor (1 - \delta) \rfloor \) and let \( A_{nn'} \) be the matrix with rows \( \sqrt{n}Z_1, \ldots, \sqrt{n}Z_{n'} \), and let \( B_{nn'} \) be the matrix with rows \( \sqrt{n}Y_1, \ldots, \sqrt{n}Y_{n'} \). Expressing the determinant as product of singular values, we have that

\[
\frac{1}{n} \sum_{1 \leq i \leq (1-\delta)n} \log \text{dist}(Z_i, V_i) = \frac{1}{n} \sum_{i=1}^{n'} \log(\sigma_i(A_{nn'}))
\]

and similarly for \( Y_i, U_i, \) and \( B_{nn'} \). Thus, it suffices to show that

\[
\frac{1}{n} \sum_{i=1}^{n'} \log(\sigma_i(A_{nn'})) - \log(\sigma_i(B_{nn'})) = O(\epsilon)
\]

for all but finitely many \( n \). Which amounts to showing that

\[
\int_{R_\epsilon} \log t \, dv_{nn'}(t) = O(\epsilon)
\]

for all but finitely many \( n \), where \( dv_{nn'} = \mu_{\frac{1}{n} A_{nn'} A_{nn'}^*} - \mu_{\frac{1}{n} B_{nn'} B_{nn'}^*} \). We show this by dividing the region of \( t \) into parts.

(1) The region of very large \( t \):
Note that

\[
\int_{0}^{\infty} t \, dv_{nn'}(t) \leq \frac{1}{n'} \sum_{i=1}^{n'} \left( \frac{1}{n} \sigma_i(A_{nn'})^2 + \frac{1}{n} \sigma_i(B_{nn'})^2 \right)
\]

\[
= \frac{1}{n'} \sum_{i=1}^{n'} (|Z_i|^2 + |Y_i|^2) < C
\]

for some \( C \), with probability 1, for sufficiently large \( n \) and all \( i \), by A3. Thus, we have that

\[
\int_{R_\epsilon} |\log t| \, dv_{nn'}(t) \leq \epsilon
\]

for all but finitely many \( n \) and some \( R_\epsilon \) depending only on \( \epsilon \).

(2) The region of intermediate \( t \):
Consider the region \( t \in [\epsilon^4, R_\epsilon] \). First, recall the Cauchy Interlacing Property

**Lemma 4.7** (Cauchy Interlacing Property). Let \( A \) be a \( n \times n \) matrix with complex entries and let \( A' \) be the submatrix formed by the first \( n - k \) rows. Let \( \sigma_1(A), \ldots, \sigma_n(A) \) denote the singular values of \( A \), and similarly for \( A' \). Then,

\[
\sigma_i(A) \geq \sigma_i(A') \geq \sigma_{i+k}(A)
\]

for every \( 1 \leq i \leq n - k \)

Let \( \Psi \) be a smooth function which equals 1 on \([\epsilon^4, R_\epsilon]\) and is supported on \([\epsilon^4/2, 2R_\epsilon]\). Then, using Lemma 4.7, we have

\[
\int_{0}^{\infty} \Psi(t) \log t \, dv_{nn'}(t) = \int_{0}^{\infty} \Psi(t) \log t \, dv_{nn}(t) + O(\epsilon)
\]
if $\delta$ is sufficiently small, depending on $\epsilon$ and $\Psi$. By Corollary 3.3, we have that $\mu_{\frac{1}{n} A_n A_n^*}$ and $\mu_{\frac{1}{n} B_n B_n^*}$ converge to the same limit, and thus $\nu_{nn}$ converges to zero. Thus,

$$\int_{\epsilon^4}^{R} \log t d\nu_{nn'}(t) \leq \int_0^\infty \Psi(t) \log t d\nu_{nn'}(t) = O(\epsilon)$$

(3) The region of moderately small $t$:
Consider the region $t \in [\delta^2, \epsilon^4]$. We wish to show that

$$\int_{\delta^2}^{\epsilon^4} |\log t||d\nu_{nn'}(t)| = O(\epsilon)$$

By the triangle inequality and symmetry, it suffices to show that, with probability 1, one has

$$\int_{\delta^2}^{\epsilon^4} |\log t||d\mu_{\frac{1}{n} A_{nn'} A_{nn'}^*}(t)| = O(\epsilon)$$

for all but finitely many $n$. We can express the left hand side as

$$\frac{1}{n} \sum_{i=1}^{n'} f\left(\frac{1}{\sqrt{n}} \sigma_i(A_{nn'})\right)$$

where $f(t) = |\log t|1_{(\delta^2 \leq t^2 \leq \epsilon^4)}$.

As $f$ is less than $|\log \delta|$, if $\delta < \epsilon^2$, we may make the contribution for $i \geq (1-2\delta)n$ acceptable. Thus, it suffices to show that we have, almost surely,

$$\frac{1}{n} \sum_{1 \leq i \leq (1-2\delta)n} f\left(\frac{1}{\sqrt{n}} \sigma_i(A_{nn'})\right) = O(\epsilon)$$

for all but finitely many $n$.

Recall that $n' = \lfloor (1-\delta)n \rfloor$. For any $0 < c < 1$, by Proposition 4.5 and the Borel-Cantelli lemma, we have with probability 1 that

$$\text{dist}(Z_i, \text{span}(Z_1, \ldots, Z_{i-1}, Z_{i+1}, \ldots, Z_{n'})) \geq c\sqrt{n-n'} = c\sqrt{\delta n}$$

for all but finitely many $n$ and all $1 \leq i \leq n'$. Thus, $\frac{1}{\sqrt{n}} \sigma_i(A_{nn'}) \geq c\sqrt{\delta}$, so that

$$\frac{1}{n} \sum_{i=1}^{n'} \left(\frac{1}{\sqrt{n}} \sigma_i(A_{nn'})\right)^{-2} \leq \frac{n' c}{n} = O(1)$$

Moreover, as $n' = \lfloor (1-\delta)n \rfloor$ and $\sigma_i(A_{nn'})$ is decreasing in $i$, one has that $\frac{1}{\sqrt{n}} \sigma_{\lfloor (1-2\delta)n \rfloor}(A_{nn'}) \geq c\sqrt{\delta}$.

Now, using Proposition 4.5 and the Borel-Cantelli lemma again, we have with probability one,

$$\text{dist}(Z_i, \text{span}(Z_1, \ldots, Z_{i-1}, Z_{i+1}, \ldots, Z_{n''})) \geq c\sqrt{n-n''}$$

for all but finitely many $n$, all $1 \leq i \leq n''$ and $\frac{n}{2} \leq n'' \leq n'$. Thus,

$$\left(\frac{1}{\sqrt{n}} \sigma_i(A_{nn''})\right)^{-2} \leq \frac{cn}{n-n''}$$
so that we have almost surely that 
\[ \frac{1}{n} \sum_{i=1}^{n'} \left( \frac{1}{\sqrt{n}} \sigma_i(A_{nn'}) \right)^{-2} = O\left( \frac{n}{n - n''} \right) \]
for all but finitely many \( n \) and all \( \frac{n}{2} \leq n'' \leq n' \).

Using the last \( n - n'' \) terms in the sum on the left hand side below, we get that
\[ \sum_{i=1}^{n''} \left( \frac{1}{\sqrt{n}} \sigma_i(A_{nn'}) \right)^{-2} \leq (n - n'') \left( \frac{1}{\sqrt{n}} \sigma_{2n''-n}(A_{nn''}) \right)^{-2} \]
we can conclude that
\[ \left( \frac{1}{\sqrt{n}} \sigma_{2n''-n}(A_{nn''}) \right) \geq c \frac{n-n''}{n} \]
for all but finitely many \( n \) and \( n/2 \leq n'' \leq n' \), for some constant \( c' > 0 \). Using Lemma 4.7, we can conclude that
\[(*) \quad \frac{1}{\sqrt{n}} \sigma_i(A_{nn'}) \geq c' n' - i \]
for all but finitely many \( n \) and all \( 1 \leq i \leq (1 - 2\delta)n \).

Now recall,
\[ \frac{1}{n} \sum_{1 \leq i \leq (1-\delta)n} f\left( \frac{1}{\sqrt{n}} \sigma_i(A_{nn'}) \right) \]
By (*), we see that the only terms in this sum that do not vanish are those with for which \( i = (1 - O(\epsilon^2))n \). For such terms, using (*) and the fact \( f(t) \leq -\log t \), we have that
\[ \frac{1}{n} \sum_{1 \leq i \leq (1-2\delta)n} f\left( \frac{1}{\sqrt{n}} \sigma_i(A_{nn'}) \right) = O(\epsilon) \]

(4) The region of small \( t \):
Consider \( t \leq \delta \). As in the region of moderately small \( t \), we need only show that
\[ \frac{1}{n} \sum_{i=1}^{n'} g\left( \frac{1}{\sqrt{n}} \sigma_i(A_{nn'}) \right) = O(\epsilon) \]
for all but finitely many \( n \), where \( g(t) = |\log t| 1_{(t^2 < \epsilon^2)} \).

By Proposition 4.5 and the Borel-Cantelli lemma, we that with probability 1, for large enough \( n \) and all \( i \leq n' \),
\[ \text{dist}(Z_i, \text{span}(Z_1, ..., Z_i, Z_{i+1}, ..., Z_{n'})) \geq c\sqrt{\delta}. \]
Again, using Proposition 4.5 and Lemma 4.3, we have that
\[ \frac{1}{n'} \sum_{i=1}^{n'} \left( \frac{1}{\sqrt{n}} \sigma_i(A_{nn'}) \right)^{-2} = O(\epsilon). \]
If \( \delta \) is small enough, we have \( g(t) \leq \epsilon/t^2 \), and the result follows.
REFERENCES

[AC15] Radostaw Adamczak and Djalil Chafai, Circular law for random matrices with unconditional log-concave distribution, Commun. Contemp. Math. 17 (2015), no. 4, 1550020. MR 3359233

[Ada11] Radostaw Adamczak, On the Marchenko-Pastur and circular laws for some classes of random matrices with dependent entries, Electron. J. Probab. 16 (2011), no. 37, 1068–1095. MR 2820070

[AGL+08] Radostaw Adamczak, Olivier Guédon, Alexander Litvak, Alain Pajor, and Nicole Tomczak-Jaegermann, Smallest singular value of random matrices with independent columns, C. R. Math. Acad. Sci. Paris 346 (2008), no. 15-16, 853–856. MR 2441920

[BDn] Pavel Bleher and Alfredo Deaño, Painlevé I double scaling limit in the cubic random matrix model, no. 2, 1650004, 58. MR 3493550

[BGCD] Florent Benaych-Georges and Thierry Cabanal-Duvillard, Marchenko-Pastur theorem and Bercovici-Pata bijections for heavy-tailed or localized vectors, no. 2, 685–715. MR 3069381

[BS10] Zhidong Bai and Jack W Silverstein, Spectral analysis of large dimensional random matrices, second edition ed., Springer Series in Statistics, Springer, 2010.

[BV] Alex Bloemendal and Bálint Virág, Limits of spiked random matrices II, no. 2, 313–340. MR 3449601

[EGP] Alan Edelman, A. Guionnet, and S. Péché, Beyond universality in random matrix theory, no. 3, 1275–1294.

[EKR06] Kevin P. Costello, Terence Tao, and Van Vu, Random symmetric matrices are almost surely nonsingular, Duke Math. J. 135 (2006), no. 2, 395–413. MR 2267289

[DF] Persi Diaconis and Peter J. Forrester, Hurwitz and the origins of random matrix theory in mathematics, no. 1, 1730001, 26. MR 3612265

[DI] Ioana Dumitriu and Tobias Johnson, The Marchenko-Pastur law for sparse random bipartite biregular graphs, no. 2, 313–340. MR 3449601

[CTV06] Kevin P. Costello, Terence Tao, and Van Vu, Random symmetric matrices are almost surely nonsingular, Duke Math. J. 135 (2006), no. 2, 395–413. MR 2267289

[DFA] Kjetil B. Halvorsen, Victor Ayala, and Eduardo Fierro, On the marginal distribution of the diagonal blocks in a blocked Wishart random matrix, Art. ID 5967218, 5. MR 3571980

[GW10] William P. Orrick, The maximal \{-1,1\}-determinant of order 15, Metrika 62 (2005), no. 2, 675–712. MR 3319140

[LAP] Haoyang Liu, Alexander Aue, and Debashis Paul, On the Marchenko-Pastur law for linear time series, no. 2, 675–712. MR 3319140

[LP] Philippe Loubaton, On the almost sure location of the singular values of certain Gaussian block-Hankel large random matrices, no. 4, 1339–1443. MR 3571248

[LW] Dang-Zheng Liu and Yanhui Wang, Universality for products of random matrices I: Ginibre and truncated unitary cases, no. 11, 3473–3524. MR 3556416

[Mal] Camille Male, The limiting distributions of large heavy Wigner and arbitrary random matrices, no. 1, 1–46. MR 3567500

[MP67] V A Marchenko and L A Pastur, Distribution of eigenvalues in certain sets of random matrices, Mat. Sb. (N.S.) 72 (1967), 507–536.

[Orr05] William P. Orrick, The maximal \{-1,1\}-determinant of order 15, Metrika 62 (2005), no. 2, 195–219.

[PP07] A. Pajor and L. Pastur, On the limiting empirical measure of eigenvalues of the sum of rank one matrices with log-concave distribution, Studia Math 195 (2007), no. 2009, 11–29.

[Tao07] Terence Tao and Van Vu, On the singularity probability of random Bernoulli matrices, J. Amer. Math. Soc. 20 (2007), no. 3, 603–628. MR 2291914

[Tao08] Terence Tao and Van Vu, Random matrices: the circular law, Commun. Contemp. Math. 10 (2008), no. 2, 261–307. MR 2409368
[TV09] Terence Tao and Van H. Vu, *Inverse Littlewood-Offord theorems and the condition number of random discrete matrices*, Ann. of Math. (2) 169 (2009), no. 2, 595–632. MR 2480613

[TV10] Terence Tao and Van Vu, *Random matrices: universality of ESDs and the circular law*, Ann. Probab. 38 (2010), no. 5, 2023–2065, With an appendix by Manjunath Krishnapur. MR 2722794

[TV12] , *Random covariance matrices: universality of local statistics of eigenvalues*, Ann. Probab. 40 (2012), no. 3, 1285–1315. MR 2962092

[TV14] , *Random matrices: the universality phenomenon for Wigner ensembles*, Modern aspects of random matrix theory, Proc. Sympos. Appl. Math., vol. 72, Amer. Math. Soc., Providence, RI, 2014, pp. 121–172. MR 3288230

[TV15] , *Random matrices: universality of local spectral statistics of non-Hermitian matrices*, Ann. Probab. 43 (2015), no. 2, 782–874. MR 3306005

[TW98] Craig A Tracy and Harold Widom, *Correlation functions, cluster functions, and spacing distributions for random matrices*, J. Statist. Phys 92 (1998), 5–6.

[Woo] Philip Matchett Wood, *Universality of the ESD for a fixed matrix plus small random noise: a stability approach*, no. 4, 1877–1896. MR 3573299

[Yao] Jianfeng Yao, *A note on a Marchenko-Pastur type theorem for time series*, no. 1, 22–28. MR 2863018

**INDIANA UNIVERSITY**

_E-mail address: cconnel1@indiana.edu_

**INDIANA UNIVERSITY**

_E-mail address: pawpatel@indiana.edu_