MORE LIMITING DISTRIBUTIONS FOR EIGENVALUES OF WIGNER MATRICES

BY SIMONA DIACONU

1Department of Mathematics, Stanford University, sdiaconu@stanford.edu

The Tracy-Widom distributions are among the most famous laws in probability theory, partly due to their connection with Wigner matrices. In particular, for $A = \frac{1}{\sqrt{n}}(a_{ij})_{1 \leq i,j \leq n} \in \mathbb{R}^{n \times n}$ symmetric with $(a_{ij})_{1 \leq i,j \leq n}$ i.i.d. standard normal, the fluctuations of its largest eigenvalue $\lambda_1(A)$ are asymptotically described by a real-valued Tracy-Widom distribution $TW_1 : n^{2/3}(\lambda_1(A) - 2) \Rightarrow TW_1$. As it often happens, Gaussianity can be relaxed, and this results holds when $\mathbb{E}[a_{11}] = 0, \mathbb{E}[a_{11}^2] = 1$, and the tail of $a_{11}$ decays sufficiently fast: $\lim_{x \to \infty} x^4 \mathbb{P}(|a_{11}| > x) = 0$, whereas when the law of $a_{11}$ is regularly varying with index $\alpha \in (0, 4)$, $c_0(n)n^{1/2-\alpha/\alpha} \lambda_1(A)$ converges to a Fréchet distribution for $c_0 : (0, \infty) \to (0, \infty)$ slowly varying and depending solely on the law of $a_{11}$. This paper considers a family of edge cases, $\lim_{x \to \infty} x^4 \mathbb{P}(|a_{11}| > x) = c \in (0, \infty)$, and unveils a new type of limiting behavior for $\lambda_1(A)$: a continuous function of a Fréchet distribution in which 2, the almost sure limit of $\lambda_1(A)$ in the light-tailed case, plays a pivotal role:

$$f(x) = \begin{cases} 2, & 0 < x < 1 \\ x + \frac{1}{x}, & x \geq 1 \end{cases}.$$ 

1. Introduction. Wigner matrices have been an object of intensive study in mathematics ever since Eugene Wigner proposed them in 1955 as a tool for understanding the organization of heavy nuclei and showed their empirical spectral distribution converges to the semicircle law ([19]). Such matrices are generally square with entries in $\mathbb{R}$, $\mathbb{C}$, or $\mathbb{H}$, and satisfy certain symmetry conditions: the focus hereafter is the real-valued symmetric case. Let $A = \frac{1}{\sqrt{n}}(a_{ij})_{1 \leq i,j \leq n} \in \mathbb{R}^{n \times n}$ be a symmetric matrix with i.i.d. entries on its upper triangular component for which $\mathbb{E}[a_{11}] = 0, \mathbb{E}[a_{11}^2] = 1$, and denote by $\lambda_1(A) \geq \lambda_2(A) \geq ... \geq \lambda_n(A)$ its eigenvalues. When $a_{11}$ is Gaussian, it is well-known that the edge (i.e., a fixed number of the largest or smallest) eigenvalues of $A$ exhibit fluctuations described by Tracy-Widom distributions (see, for instance, the seminal paper [18]) and several universality results, meant to deal with the case in which $a_{11}$ is not normally distributed, have been discovered (e.g., Tao and Vu [17]).

Consequently, a natural question is what can substitute the Gaussianity assumption in such results. It must be mentioned that this condition cannot be completely dispensed with: as a finite second moment of the entries is necessary for the convergence of the empirical spectral distribution of $A$ to the semicircle law, the fourth moment is crucial for the asymptotic behavior of $\lambda_1(A)$ (Bai and Yin [5] showed a finite fourth moment is required if the largest eigenvalue has an almost sure deterministic limit; sample covariance matrices with the number of samples proportional to their dimension represent another instantiation of this phenomenon: when the fourth moment is finite, the largest eigenvalue tends almost surely to a constant, whereas when the former is infinite, the latter tends to infinity with probability one: see Bai and Yin [4], Bai et al. [3]). Furthermore, if $a_{11}$ is heavy-tailed (i.e., its law is regularly varying of index $\alpha \in (0, 4)$: at a high level, this says $\mathbb{P}(|a_{11}| > x)$ decays like $x^{-\alpha}$, and in particular, its $\alpha$-moment is infinite), then a new behavior emerges: the edge eigenvalues,
properly normalized, fluctuate according to a Poisson point process (Soshnikov [16] studied \(\alpha \in (0, 2)\), and Auffinger et al. [1] extended this result to \(\alpha \in (0, 4)\)).

The question of finding optimal conditions under which the edge eigenvalues can be described by a Tracy-Widom distribution received a fair amount of attention and was completely answered in a paper of Lee and Yin [11]: this occurs if and only if
\[
\lim_{x \to \infty} x^4 \mathbb{P}(\lambda_1 \geq x) = 0.
\]
It must be noticed there had been several publications prior to this result, proving an \(\alpha\)-finite moment of the underlying distribution suffices (for symmetric distributions, Ruzmaikina [12] obtained \(\alpha > 18\), and later this was improved to \(\alpha > 12\) by Khorunzhiy [9]). This paper is concerned with a family of edge cases, distributions with
\[
\lim_{x \to \infty} x^4 \mathbb{P}(\lambda_1 \geq x) = c > 0,
\]
and the main result is:

**Theorem 1.** Suppose \(A = \frac{1}{\sqrt{n}} (a_{ij})_{1 \leq i,j \leq n} \in \mathbb{R}^{n \times n}\) is a symmetric matrix for which \((a_{ij})_{1 \leq i,j \leq n}\) are i.i.d. and the distribution of \(a_{11}\) is symmetric with
\[
\mathbb{E}[a_{11}^2] = 1, \quad \lim_{x \to \infty} x^4 \mathbb{P}(\lambda_1 \geq x) = c \in (0, \infty).
\]
Then as \(n \to \infty\),
\[
\lambda_1(A) \Rightarrow f(\zeta_c),
\]
where
\[
f(x) = \begin{cases} 2, & 0 < x < 1, \\ x + \frac{1}{x}, & x \geq 1 \end{cases},
\]
and \(\zeta_c > 0\) has a Fréchet distribution with shape and scale parameters 4, \((\frac{c}{4})^{1/4}\), respectively: for all \(x > 0\), \(\mathbb{P}(\zeta_c \leq x) = \exp(-\frac{cx^{-4}}{2})\).

Several observations are in order regarding the convergence stated in (1). First, the limiting distribution arises at the collision of heavy- and light-tailed regimes. More precisely, it inherits the Fréchet fluctuations \(\zeta_c\) from the Poisson point process characterizing the extrema of heavy-tailed i.i.d. random variables, whereas 2 is a vestige from light-tailed distributions since the convergence of the empirical spectral distribution of \(A\) to the semicircle law holds as long as \(a_{11}\) has its second moment finite (subsection 1.1 expounds on this phenomenon). Moreover, the symmetry assumption could likely be weakened to being centered using an analysis similar to the one undertaken in [8] (see discussion below (13)).

Second, the function \(f\) is tightly related to a sequence of polynomials \((s(d, X))_{d \in \mathbb{N}}\), whose coefficients are nonnegative and have a combinatorial description: specifically, \(s(d, X)\) has degree \(2d - 2\), and for \(x > 0\),
\[
f(x) = \lim_{d \to \infty} s(d, x)^{1/2d}
\]
(conditionally on an event \(E(x)\), a trace will be roughly \(\lambda_1^{2d}(\frac{1}{\sqrt{n}} A)\) and of order \(s(d, x)\) by counting). The strategy adopted here has been oftentimes employed for getting a hold of the largest eigenvalue of a symmetric random matrix \(M\) : controlling by careful counting
\[
\mathbb{E}[\text{tr}(M^{p})] = \mathbb{E}[\sum_{1 \leq i \leq n} \lambda_i^p(M)]
\]
for large integers \(p\) (e.g., Bai and Yin [5], Benaych-Georges and Péché [6], Auffinger et al. [1]), an approach whose by-product is the above definition of \(f\). Nevertheless, in the current situation, the classical choice \(M = (a_{ij} \chi_{|a_{ij}| \leq c(n)})\), for a suitable \(c(n)\), falls short due to the heavy tail of \(a_{11}\). To illustrate how this occurs, suppose the goal is bounding \(||A||\), and let
$p$ be even so that $||M||^p \leq tr(M^p)$. After truncating $A$ (to ensure all moments are finite), two incompatible constraints on $p$ emerge: it must be large to annihilate the contribution of the other eigenvalues (since the empirical spectral distribution of $A$ converges to the semicircle law, such trace would be at least of order $n \cdot 2^p$: thus, to eliminate $n, p$ should grow faster than $\log n$), but also small to deal with the non-negligible terms, which are numerous because the moments of $a_{11}$ grow fast.

This failure suggests the necessity of twisting this method to adapt it to the present context: an ideal substitute of $tr(M^p)$ would be on the one hand, lighter than what is meant to replace, and on the other hand, amenable to combinatorics. In light of these observations, a promising candidate is

$$tr((S + Q)^p) - tr(Q^p)$$

where $S, Q$ are symmetric with (C1) $||S + Q - A||$ small, and (C2) $S$ very sparse (say, $O(n)$ non-zero entries) inasmuch as C1 would allow switching from $A$ to $S + Q$, while C2 would ensure a considerable overlap between the eigenvalues of $S + Q$ and those of $Q$, the difference above hence generating plenty of cancellations (see Lemma 7 for a rigorous statement).

The desired convergence concerning $A$ is thus justified by constructing such a proxy $S + Q$, further analyzed with the aid of the counting technique developed by Sinai and Soshnikov in [13]. Some modifications are anew indispensable: although both situations share the family of cycles dominating the considered expectations, in the current setting, there exist several types of comparable contributions (these underlie the sequence of polynomials $s(d, X)$ mentioned earlier), whereas in the framework of [13], each dominating cycle generates the same value. Furthermore, the expectation in this case is not unconditional (a conditioning is employed to freeze the largest entries of $A$).

Third, for $k$ fixed, the joint distribution of the $k$ largest eigenvalues of $A$ can be determined reasoning as in Soshnikov [16]. Theorem 1.2 of [16] states that for regularly varying distributions $a_{11}$ with index $\alpha \in (0, 2)$, the limiting law of the positive eigenvalues of $A$ (appropriately normalized) is given by an inhomogeneous Poisson point process $N$ on $[0, \infty)$ with intensity $\rho(x) = \frac{\alpha}{x^{\alpha+1}}$: the ingredients behind this result are the behavior of $\lambda_1(A)$, the Cauchy interlacing inequalities, and the theory on extrema of random variables in the domain of attraction of $\alpha$-laws (see, for instance, Theorem 2.3.1 in Leadbetter et al. [10]). In the present situation, the intensity is $\rho(x) = \frac{\alpha}{x^{\alpha+1}}$, and the convergence of the positive eigenvalues of $A$ is not to the point process itself, but rather to $f(N)$ (see end of subsection 3.3).

The remainder of the paper contains the proof of Theorem 1: subsections 1.1 and 1.2 present the rival forces behind the object of interest and the matrix decomposition leading to a proxy as described previously; section 2 gathers the necessary tools for showing

$$\limsup_{n \to \infty} \mathbb{P}(||A|| \geq x) \leq \mathbb{P}(f(\xi_c) \geq x) \leq \liminf_{n \to \infty} \mathbb{P}(||A|| \geq x)$$

(2)

for all $x \in \mathbb{R}$; section 3 consists of proving (2) and justifying why an analogous chain of inequalities holds when $||A||$ is replaced by $\lambda_1(A)$. Lastly, it must be mentioned Theorem 1 can be used to obtain the asymptotic behavior of the largest eigenvalue of one-rank perturbations of Wigner matrices with heavy-tailed entries ([7]).

1.1. A Lower Bound. This subsection presents a preliminary inequality $\lambda_1(A)$ satisfies: for $t > 0$,

$$\lambda_1(A) \geq \max (\max A, 2) - t$$

(3)

with probability tending to one as $n$ tends to infinity, where $\max A := \frac{1}{\sqrt{n}} \max_{1 \leq i \leq j \leq n} |a_{ij}|$. Although this result is not directly employed to prove Theorem 1, it displays the two essential
quantities underlying both the operator norm of \( A \) and its largest eigenvalue. Henceforth, an event \( E = E_n \) is said to hold with high probability if \( \lim_{n \to \infty} \mathbb{P}(E_n) = 1 \).

On the one hand, since \( A \) is a normalized Wigner matrix whose entries have variance \( \frac{1}{n} \), its empirical spectral distribution converges almost surely to the semicircle law \( \rho \) (Theorem 2.5 in Bai and Silverstein [2]). Thus, almost surely
\[
\liminf_{n \to \infty} \lambda_1(A) \geq 2
\]
as \( \rho \) assigns a positive mass to any neighborhood of 2, from which for \( t > 0 \), with high probability
\[
\lambda_1(A) \geq 2 - t.
\]

On the other hand,
\[
\max A = \frac{1}{\sqrt{n}} \max_{1 \leq i, j \leq n} |a_{ij}| \Rightarrow \zeta_c
\]
because for \( s > 0 \) as \( n \to \infty \),
\[
\mathbb{P}\left( \frac{1}{\sqrt{n}} \max_{1 \leq i, j \leq n} |a_{ij}| < s \right) = \left( 1 - \mathbb{P}(|a_{11}| \geq s \sqrt{n}) \right)^{\frac{s^2 n}{2}} = \exp(-c(s \sqrt{n})^{-4}(1+o(1)) \cdot \frac{n^2 + n}{2}).
\]
Let \( |a_{i_0 j_0}| = \max_{1 \leq i, j \leq n} |a_{ij}| \text{ with } i_0 \leq j_0 \) : (5) and
\[
\max_{1 \leq i \leq n} |a_{ii}| \xrightarrow{p} 0
\]
((5) implies \( \frac{1}{\sqrt{n}} \max_{1 \leq i \leq n} |a_{ii}| \Rightarrow \zeta_c \) yield \( i_0 < j_0 \) with high probability. Consider a unit vector \( v \in \mathbb{R}^n \) with \( |v_{i_0}| = |v_{j_0}| = \frac{1}{\sqrt{2}} \) and \( v_{i_0} v_{j_0} a_{i_0 j_0} \geq 0 \). Therefore,
\[
v^T A v = \frac{1}{\sqrt{n}} \left( |a_{i_0 j_0}| + \frac{a_{i_0 j_0}}{2} \right) \geq \frac{1}{\sqrt{n}} \max_{1 \leq i, j \leq n} |a_{ij}| - \frac{1}{\sqrt{n}} \max_{1 \leq i \leq n} |a_{ii}|,
\]
which in conjunction with (6) gives for \( t > 0 \),
\[
\lambda_1(A) \geq \max A - t.
\]

with high probability. The desired bound (3) ensues from (4) and (7).

1.2. A Matrix Decomposition. In light of (5), showing for \( M, \epsilon > 0 \),
\[
\lim_{n \to \infty} \mathbb{P}(\|A\| > f(M) + 2\epsilon \mid \max A \leq M) = 0,
\]
(8)
\[
\lim_{n \to \infty} \mathbb{P}(\|A\| < f(\max A) - 2\epsilon) = 0
\]
suffices to justify (2).

One core ingredient for both (8) and (9) is a decomposition of \( A \) into three matrices \( A_s, A_m, A_b \), with small, medium, and big entries, respectively. The last component, already sparse, is further split into two matrices, one of them being considerably sparser than \( A_b \). Next, it is proved that with high probability \( \|A_m\| \) is negligible, while the first component of \( A_b \) contributes at most \( \epsilon \). Thus, in an operator norm sense, the sum of \( A_s \) and the sparser component of \( A_b \) differs from \( A \) by at most \( \epsilon \), making the former a proxy for the latter.

Let \( \delta_1, \delta_2 \in (0, \frac{1}{2\sqrt{2}}), \delta_2 \in (0, \frac{1}{\sqrt{2}}) \) be fixed constants and
\[
A_s = \frac{1}{\sqrt{n}} (a_{ij} \chi_{|a_{ij}| \leq n^{1/4 - \delta_1}}), \quad A_m = \frac{1}{\sqrt{n}} (a_{ij} \chi_{n^{1/4 - \delta_1} < |a_{ij}| \leq n^{3/8 + \delta_2}}), \quad A_b = \frac{1}{\sqrt{n}} (a_{ij} \chi_{|a_{ij}| > n^{3/8 + \delta_2}}),
\]

\( A_s \) and especially \( A_m \) are expected to be small, while \( A_b \) is much larger. The sum of the first two
\[
\lim_{n \to \infty} \mathbb{P}(\|A\| > f(M) + 2\epsilon \mid \max A \leq M) = 0
\]

(8)
\[
\lim_{n \to \infty} \mathbb{P}(\|A\| < f(\max A) - 2\epsilon) = 0
\]

(9) suffices to justify (2).

One core ingredient for both (8) and (9) is a decomposition of \( A \) into three matrices \( A_s, A_m, A_b \), with small, medium, and big entries, respectively. The last component, already sparse, is further split into two matrices, one of them being considerably sparser than \( A_b \). Next, it is proved that with high probability \( \|A_m\| \) is negligible, while the first component of \( A_b \) contributes at most \( \epsilon \). Thus, in an operator norm sense, the sum of \( A_s \) and the sparser component of \( A_b \) differs from \( A \) by at most \( \epsilon \), making the former a proxy for the latter.

Let \( \delta_1, \delta_2 \in (0, \frac{1}{2\sqrt{2}}), \delta_2 \in (0, \frac{1}{\sqrt{2}}) \) be fixed constants and
\[
A_s = \frac{1}{\sqrt{n}} (a_{ij} \chi_{|a_{ij}| \leq n^{1/4 - \delta_1}}), \quad A_m = \frac{1}{\sqrt{n}} (a_{ij} \chi_{n^{1/4 - \delta_1} < |a_{ij}| \leq n^{3/8 + \delta_2}}), \quad A_b = \frac{1}{\sqrt{n}} (a_{ij} \chi_{|a_{ij}| > n^{3/8 + \delta_2}}),
\]
for which

\[ A = A_s + A_m + A_b. \]

Theorem 2.1 of Benaych-Georges and Péché [6], stated below, is employed next to bound \( ||A_m|| \) (as well as \( ||A_s|| \) in later subsections).

**Theorem 2** (Benaych-Georges and Péché [6]). Suppose \( \tilde{A} = (a_{ij})_{1 \leq i, j \leq n} \) is a symmetric real-valued random matrix with at most \( n^\mu \) non-zero entries on each row, \( (a_{ij})_{1 \leq i, j \leq n} \) i.i.d., of variance one, with distribution symmetric and regularly varying of index \( \alpha > 2 \). Then for \( A_n = (a_{ij} \chi_{|a_{ij}| \leq n^\nu}) \) and any constants \( \gamma, \gamma', \gamma'' > 0 \) with \( \frac{\mu}{2} \leq \gamma', \frac{\mu}{2} + \gamma + \gamma'' < \gamma' \),

\[ \mathbb{E}[tr(A_n^{2s_n})] \leq L(n) n^{1+2\gamma} s_n^{-3/2} (2n^{\gamma'})^{2s_n} \]

for a slowly varying function \( L, \) and all \( s_n \in \mathbb{N}, s_n \leq n^{\gamma''}. \)

Notice Theorem 2 holds for matrices of the form \( A_n = (a_{ij} \chi_{n^{-\mu} < |a_{ij}| \leq n^\nu}) \) too since a lower bound on the entries can only decrease the moments of the random variables appearing in the trace and the number of non-zero elements among them. Consider \( A_m := \sqrt{n} A_m = (a_{ij} \chi_{n^{1/(4-\delta)} < |a_{ij}| \leq n^{3/(8+\delta) + 2}}) \). The probability of \( \tilde{A}_m \) having at least \( l \) non-zero entries on a given row is at most

\[ \binom{n}{l} (2cn^{1/4-\delta})^{-4l} \leq \frac{n^l}{l!} \cdot (2c)^{-4l} n^{-l+4l \delta} = (2c)^{-4l} n^{4\delta l} \leq (2c/e)^{-4l} n^{-4l \delta} \]

for \( l = n^{8\delta_1}, n \geq n(c) \). Hence for \( E_0 \), the event that each row of \( \tilde{A}_m \) has at most \( l \) non-zero entries,

\[ \mathbb{P}(E_0^c) \leq n(2c/e)^{-4l} n^{-4l \delta} = o(1). \]

Conditioning on \( E_0 \), Theorem 2 yields

\[ \mathbb{E}[tr(\tilde{A}_m^{2s_n}) \mid E_0] \leq L(n) n^{1+2\gamma} s_n^{-3/2} (2n^{\gamma'})^{2s_n} \]

for

\[ \mu = 8\delta_1, \gamma = \frac{3}{8} + \delta_2, \gamma'' = \delta_3, \gamma' = \frac{\mu}{4} + \gamma + 2\gamma'' = 2\delta_1 + \frac{3}{8} + \delta_2 + 2\delta_3 \leq \left( \frac{\mu}{2}, \frac{1}{2} \right) = (4\delta_1, \frac{1}{2}) \]

(subsection 2.2 presents in detail why such conditional expectations can replace their unconditional counterparts at the cost of a factor \( c(c) \) for \( \epsilon > 0 \), which can be evidently absorbed by \( L \)). Chebyshev’s inequality then gives for \( \delta_4 > 0 \) and \( n \) sufficiently large,

\[ \mathbb{P}(tr(\tilde{A}_m^{2s_n}) \geq (2n^{\gamma'})^{2s_n} n^{2s_n \delta_4}) \leq \mathbb{P}(E_0^c) + n^{2+2\gamma} n^{-2s_n \delta_4} = o(1) \]

by choosing \( s_n = [n^{\gamma''}] \). Since \( \gamma' < \frac{1}{2}, \delta_4 = \frac{1}{2}(\frac{1}{2} - \gamma') > 0 \) entails

\[ ||A_m|| \leq n^{-1/2}(tr(\tilde{A}_m^{2s_n}))^{1/2} n^{-\gamma' + \delta_4 - \frac{1}{2}} = o(1) \]

with high probability.

Proceed now with the split of \( A_b \). Let \( E_1 \) be the event that the non-zero entries of \( A_b \) are off-diagonal and any two lie on different rows: by a union bound,

\[ \mathbb{P}(E_1^c) \leq n \cdot 2c(n^{3/8+\delta_2})^{-4} + n^2 \cdot n \cdot (2c(n^{3/8+\delta_2})^{-4})^2 = o(1) \]

for \( n \geq n(c) \). For \( \kappa > 0 \) and a sequence \( m = m_n \rightarrow \infty \), let

\[ A_b = \frac{1}{\sqrt{n}} (a_{ij} \chi_{n^{3/8+\delta_2} \leq |a_{ij}| \leq \kappa}) + \frac{1}{\sqrt{n}} (a_{ij} \chi_{|a_{ij}| > \kappa}) := A_{b,\kappa} + A_{B,\kappa}, \]
and \( E_2 \) the event that \( A_{B,\kappa} \) has at most \( 2m \) non-zero entries. Then

\[
\mathbb{P}(E_2^c) \leq \left( \frac{n^2}{m} \right) \cdot (2c(\kappa \sqrt{n})^{-4})^m \leq \frac{n^{2m}}{m!} \cdot (2c\kappa^{-4})^{m-2m} = \frac{(2c\kappa^{-4})^{m-2m}}{m!} = o(1)
\]

(at least \( m \) elements of size at least \( \kappa \sqrt{n} \) must exist among the \( \frac{n^2+n}{2} \leq n^2 \) i.i.d. random variables \((a_{ij})_{1 \leq i \leq j \leq n}\)). Moreover, when \( E_1 \) occurs, \( A_{\kappa} \) has at most one non-zero entry per row and so

\[
||A_{\kappa}|| \leq \max_{1 \leq i,j \leq n} |(A_{\kappa})_{ij}| \leq \kappa.
\]

In virtue of (10) and (11), for any \( \kappa \leq \epsilon \) and fixed sequence \( m = m_n \to \infty \), (8) and (9) ensue from

\[
\lim_{n \to \infty} \mathbb{P}(||A_\kappa|| > f(M) + \epsilon | E_2, \max A \leq M) = 0,
\]

\[
\lim_{n \to \infty} \mathbb{P}(||A_\kappa|| < f(\max A) - \epsilon | E_2) = 0,
\]

where \( A_\kappa := A_s + A_{B,\kappa} \). These two limits are the subject of the forthcoming section.

2. Conditional Operator Norms. Identities (8') and (9') are justified by analyzing

\[
tr(A_{2p}^2) - tr(A_{2p}^2) = tr((A_s + A_{B,\kappa})^{2p}) - tr(A_{2p}^2)
\]

for large integers \( p \). Roughly speaking, Weyl's inequalities and the sparsity of \( A_{B,\kappa} \) entail this difference grows at the same rate as \( ||A_\kappa||^{2p} \). Furthermore, by conditioning on the appropriate events, its expectation can be squeezed between \( s(p, \max A) \) and \( s(p, M) \) (up to constant powers of \( p \)), where \( s : \mathbb{N} \times (0, \infty) \to [0, \infty) \) is the polynomial function yielding the corresponding conditional expectations, whenever \( m = m_n, p = p_n, n \) grow to infinity at completely different rates \( (m \leq \log \log p, p \leq \log \log n \) suffice). Henceforth such a growth hierarchy is implicitly assumed.

Let us introduce the notation needed for the conditionings to come. Denote by \( S = S_{n,m} \) the set of subsets \( S \subset \{(i,j) : 1 \leq i \leq j \leq n\} \) with the following properties:

(a) \(|S| \leq m\);
(b) any \((i,j) \in S\) has \( i < j\);
(c) all pairwise distinct elements \((i_1,j_1), (i_2,j_2)\) of \( S \) satisfy \( \{i_1,j_1\} \cap \{i_2,j_2\} = \emptyset\).

Consider the events

\[
E(S, \kappa, M) = \{ \max_{i \leq j, (i,j) \notin S} |a_{ij}| \leq \kappa \sqrt{n} < \min_{(i,j) \in S} |a_{ij}| \leq \max_{(i,j) \in S} |a_{ij}| \leq M \sqrt{n} \},
\]

\[
E^B(S, \kappa, M) = E(S, \kappa, M) \cap \{a_{ij}, (i,j) \in S\}
\]

(the first requires the positions of the non-zero entries of \( A_{B,\kappa} \) to be the elements of \( S \cup \{(j,i) : (i,j) \in S \} \), while the second also fixes their values). Clearly, \( (E(S, \kappa, M))_{S \in S} \) are pairwise disjoint, and

\[
E_1 \cap E_2 \cap E_3 \subset \bigcup_{S \in S} E(S, \kappa, M)
\]

where

\[
E_3 = \{ \max_{1 \leq i,j \leq n} |a_{ij}| \leq M \sqrt{n} \}.
\]

Since \( E_1 \cap E_2 \) has probability tending to one as \( n \to \infty \), a sufficient condition for (8') is

\[
\lim_{n \to \infty} \mathbb{P}(||A_\kappa|| > f(M) + \epsilon | E(S, \kappa, M)) = 0
\]
uniformly in $S \in \mathcal{S}$ (i.e., the bounds involve solely $M, m, n$). Inequality (9') ensues from a similar uniform convergence in $S \in \mathcal{S}, S \neq \emptyset$:

$(9'') \quad \lim_{n \to \infty} \mathbb{P}(\|A_n\| < f(\max A) - \epsilon \mid E^R(S, \kappa, M)) = 0$

(this yields)

$$\limsup_{n \to \infty} \mathbb{P}(\|A_n\| < f(\max A) - \epsilon) \leq \mathbb{P}(\max A \notin [\kappa, M]);$$

then use (5), and let $\kappa \to 0, M \to \infty$). For the sake of simplicity, denote conditioning on these events by $\ast, \ast\ast$, respectively (i.e., by an abuse of notation,

$$\mathbb{P}_\ast(||A_n\| > f(M) + \epsilon) := \mathbb{P}(||A_n\| > f(M) + \epsilon \mid E(S, \kappa, M)),$$

$$\mathbb{P}_{\ast\ast}(||A_n\| < f(\max A) - \epsilon) := \mathbb{P}(||A_n\| < f(\max A) - \epsilon \mid E^R(S, \kappa, M)),$$

for $S \in \mathcal{S}$ fixed).

The means of bounding the conditional probabilities in (8'') and (9'') is computing

$$E_\ast[tr(A_{2p}^\kappa) - tr(A_{2p}^\kappa)], \quad E_{\ast\ast}[tr(A_{2p}^\kappa) - tr(A_{2p}^\kappa)],$$

by employing the combinatorial technique behind the proof of Theorem 2, pioneered by Sinai and Soshnikov in [13], and subsequently used in several contexts (e.g., Sinai and Soshnikov [14], Soshnikov [15], Auffinger et al. [1]).

In the rest of this section,

• 2.1 presents in detail the combinatorial method of Sinai and Soshnikov in [13];

• 2.2 uses this counting device on the first conditional expectation above;

• 2.3 connects the functions $s$ and $f : \lim_{p \to \infty} s(M, p)^{1/2p} = f(M)$ for $M > 0$.

2.1. Large Moments. Suppose $p \in \mathbb{N}$ and $B = (b_{ij})_{1 \leq i, j \leq n} \in \mathbb{R}^{n \times n}$ is a symmetric random matrix for which $(b_{ij})_{1 \leq i, j \leq n}$ are i.i.d. and $b_{11}$ has a symmetric distribution with $\mathbb{E}[b_{11}^2] \leq 1, \mathbb{E}[b_{11}^4] \leq L(n)n^{\delta(2l - 4)}, 2 \leq l \leq p, \delta > 0$, and $L : \mathbb{N} \to [1, \infty), L(n) < n^{2\delta}$. The content of this subsection is

$$\mathbb{E}[tr(B^{2p})] \leq L(n)2^{2p}p! \sum_{(n_1, \ldots, n_p)} n_1^{1+\sum_{1 \leq k \leq p} n_k+2\delta} \prod_{1 \leq k \leq p} \left(\frac{k!}{n_k k!}\right) \prod_{2 \leq k \leq p} (2k)^{k n_k}$$

where $n_1, \ldots, n_p$ are non-negative integers with $\sum_{1 \leq k \leq p} k n_k = p$. This inequality and $||B||^{2p} \leq tr(B^{2p})$ provide some non-trivial information about $||B||$ as long as $\delta$ and $p$ are small enough, in which case the sum on the right-hand side is bounded by a simple expression (these computations are included at the end of this subsection).

Clearly,

$$\mathbb{E}[tr(B^{2p})] = \sum_{(i_0, i_1, \ldots, i_{2p-1})} \mathbb{E}[b_{i_0 i_1, i_1 i_2 \ldots i_{2p-2} i_{2p-1} i_0}].$$

Let $i := (i_0, i_1, \ldots, i_{2p-1}, i_0)$ and $b_i := b_{i_0 i_1, i_1 i_2 \ldots i_{2p-2} i_{2p-1} i_0}$. Interpret $i$ as a directed cycle with vertices among $\{1, 2, \ldots, n\}$ and call $(i_{2k-1}, i_k)$ its $k^{th}$ edge for $1 \leq k \leq 2p$, where $i_{2p} := i_0$; for $u, v \in \{1, 2, \ldots, n\}, (u, v)$ is a directed edge from $u$ to $v$, whereas $uv$ is undirected, the former are the building blocks of the cycles underlying the trace in (12), while the latter determine their expectations; in particular, $uv = vu$. Call $i$ an even cycle if each undirected edge appears an even number of times in it; using symmetry, $\mathbb{E}[b_i] = 0$ unless $i$ is an even cycle.
The crux of the technique developed by Sinai and Soshnikov in [13] is a change of summation in (12), from even cycles \(i\) to \(p\)-tuples of non-negative integers \(n_1, n_2, \ldots, n_p\) satisfying \(\sum_{1 \leq k \leq p} k n_k = p\). This is achieved by mapping each such cycle to a tuple of this type, and bounding from above the sizes of the preimages of this transformation and the expectations of their elements. For \(i\), call an edge \((i_k, i_{k+1})\) and its right endpoint \(i_{k+1}\) marked if an even number of copies of \(i_k^\prime\) precedes it; i.e., if \(\{t \in \mathbb{Z} : 0 \leq t \leq k - 1, i_{t+1} = i_{k+1}\}\) has even size, and pair each unmarked edge with its last marked copy (i.e., for \((i_k, i_{k+1})\) unmarked, pair it with \((i_t, i_{t+1})\), where \(t' = \max\{t \in \mathbb{Z} : 0 \leq t \leq k - 1, i_{t+1} = i_{k+1}\}\)). As it will soon become apparent, the analysis of such cycles, and consequently of the trace, relies on this pairing. Each even cycle \(i\) has \(p\) marked edges, and any vertex \(j \in \{1, 2, \ldots, n\}\) of \(i\), apart perhaps from \(i_0\), is marked at least once (the first edge of \(i\) containing \(j\) is of the form \((i, j)\) since \(i_0 \neq j\), and no earlier edge is adjacent to \(j\)). For \(0 \leq k \leq p\), denote by \(N_i(k)\) the set of \(j \in \{1, 2, \ldots, n\}\) marked exactly \(k\) times in \(i\) with \(n_k := |N_i(k)|\). Then

\[
\sum_{0 \leq k \leq p} n_k = n, \quad \sum_{1 \leq k \leq p} k n_k = p.
\]

Let \(C(p)\) be the set of pairwise non-isomorphic even cycles of length \(2p\), with \(n_1 = p\), and the first vertex unmarked (call two cycles \(i, j\) of length \(2p\) isomorphic if \(i_s = i_t \iff j_s = j_t\) for all \(0 \leq s, t \leq 2p\), and \(b_{p,t}\) the number of vertices \(v\) of multiplicity \(t\) in \(i \in C(p)\): i.e., \(|\{0 \leq j \leq 2p, i_j = v\}| = t\). The elements of \(C(p)\), called simple even cycles in [13], are the first-order terms in (12) when the entries of \(B\) have light tails, and play a crucial role in the current framework too (see subsection 2.2). Additionally, when the law of \(a_{11}\) is not symmetric, more cycles than the elements of \(C(p)\) have nonzero expectations: it could be shown these are not first-order terms by arguing as in subsection 2.2 in [8], where (12) is analyzed for \(B\) having centered Bernoulli distributed entries. Since this case does not add any insight into the problem at hand, it is not pursued in this paper.

Having constructed a \(p\)-tuple of non-negative integers \((n_1, n_2, \ldots, n_p)\), \(\sum_{1 \leq k \leq p} k n_k = p\) from an even cycle \(i\), the final task is obtaining upper bounds on the number of such cycles mapped to a given tuple (steps 1 – 4) and their individual contributions (step 5). In what follows, \((n_1, n_2, \ldots, n_p)\) remains fixed, and \(i\) is any even cycle mapped to it by the procedure described above.

Step 1. Map \(i\) to a Dyck path \((s_1, s_2, \ldots, s_{2p})\), where \(s_k = +1\) if \((i_{k-1}, i_k)\) is marked, and \(s_k = -1\) if \((i_{k-1}, i_k)\) is unmarked. The number of such paths is the Catalan number \(C_p = \frac{1}{p+1} \left(\begin{array}{c} 2p \\ p \end{array}\right)\).

Step 2. Once the positions of the marked edges in \(i\) are chosen (i.e., a Dyck path), establish the order of their marked vertices. There are at most

\[
\frac{p!}{\prod_{1 \leq k \leq p} (k!)^{n_k}} \cdot \frac{1}{\prod_{1 \leq k \leq p} n_k!}
\]

possibilities as each is a partition of a set of size \(p\) in \(n_1 + \ldots + n_p\) subsets with \(n_k\) of them of size \(k\).

Step 3. Select the distinct vertices appearing in \(i\),

\[V(i) := \cup_{0 \leq k \leq 2p-1} \{i_k\},\]

one at a time, by reading the edges of \(i\) in order, starting at \((i_0, i_1)\). There are at most

\[n^{1+\sum_{1 \leq k \leq p} n_k}\]

such sets because \(|V(i)| \leq 1 + \sum_{1 \leq k \leq p} n_k\) (recall that any vertex of \(i\), except perhaps from \(i_0\), is marked at least once).
Step 4. Choose the remaining vertices of \( i \) from \( V(i) \), by reading anew the edges of \( i \) in order, beginning at \((i_0, i_1)\) (step 3 only established the first appearance of each element of \( V(i) \) in \( i \)). Observe that only the right ends of the unmarked edges have yet to be decided: the first edge \((i_0, i_1)\) is fixed as \( i_0, i_1 \) have already been chosen (\( i_1 \) is marked); by induction, any subsequent edge has its left end fixed, and therefore only its right end has yet to be chosen. This yields that marked edges are fully labeled: step 2 determines their positions in \( i \), while step 3 appoints their right endpoints.

The number of possibilities in this case is at most \( \prod_{2 \leq k \leq p} (2kn) \):

**Lemma 1.** If \( v \in N_i(k) \), then the number of unmarked edges of the form \((v, u)\) is at most

\[
\begin{cases} 
2k, & k \geq 1 \\
1, & k = 1
\end{cases}
\]

**Proof.** Let \((t_j, v) = (i_{n_j}, i_{n_j+1}), 1 \leq j \leq k \) for \( n_1 < n_2 < \ldots < n_k \) be the marked edges with right endpoints \( v \), and \( u_j \) the number of unmarked edges of the type \((v, u)\) with index (i.e., position in \( i \)) in \([n_j, n_{j+1}]\), where \( n_{k+1} := 2p \). Because there is no edge of the latter type preceding \((t_1, v)\), and \((t_j, v), 1 \leq j \leq k \) are marked, the statement above is equivalent to

\[
\begin{align*}
&\sum_{1 \leq j \leq k} u_j \leq 2k, \quad k \geq 1 \\
&u_1 \leq 1, \quad k = 1
\end{align*}
\]

For \( 1 \leq j \leq k \), denote by \( S_j \) the set of marked edges adjacent to \( v \), with index \( n < n_j \), and unmarked counterparts of index \( n' \geq n_j \). Take \( a_j := |S_j| \), fix an integer \( j \in [1, k] \), and suppose

\[
(i_{n_j}, i_{n_j+1}, \ldots, i_{n_{j+1}}, i_{n_{j+1}+1}) = (t_j, v, s_1, *, s_1, v, s_2, \ldots, s_l_j, v, s_{l_j+1}, *, t_{j+1}, v)
\]

where * are sequences of vertices that do not contain \( v \). Then the edges \((\tilde{s}_1, v), \ldots, (\tilde{s}_{l_j}, v)\) are unmarked and their marked counterparts are among \((t_j, v), (v, s_1), \ldots, (v, s_{l_j})\), and the elements of \( S_j \). This gives

\[
a_{j+1} \leq (a_j + l_j + 2) - (u_j + l_j) = a_j + 2 - u_j
\]

because \( S_{j+1} \subset S_j \cup \{(t_j, v), (v, s_1), \ldots, (v, s_{l_j}), (v, s_{l_j+1})\} \), and at least \( u_j + l_j \) elements of the latter set are not in the former since among them, \( u_j \) are unmarked and \( l_j \) are the marked counterparts of \((\tilde{s}_1, v), \ldots, (\tilde{s}_{l_j}, v)\).

Hence, \( u_j \leq a_j - a_{j+1} + 2 \) for all \( 1 \leq j \leq k \), from which

\[
\sum_{1 \leq j \leq k} u_j \leq 2k + a_1 - a_{k+1}.
\]

If \( v \neq i_0 \), then \( a_1 = 0 \) and so \( \sum_{1 \leq j \leq k} u_j \leq 2k \). Else, \( v = i_0, a_1 \leq 1 \) (if \( i_0 = i_1 \), then \( n_1 = 0, a_1 = 0 \); suppose next \( i_0 \neq i_1 \); if \( i_j \neq i_0 \) for \( 2 \leq j < n_1 \), then \( S_1 = \{(i_0, i_1)\} \), \( a_1 = 1 \); otherwise, consider \( m \) minimal with \( 2 \leq m < n_1, i_m = i_0 \); then \( i_{m-1} = i_1 \) because otherwise \((i_{m-1}, i_m) = (i_{m-1}, i_0)\) would be marked; ignore \((i_0, i_1, \ldots, i_{m-1})\), and proceed with an analogous analysis for \((i_m, \ldots, i_{n_i})\); this clipping does not affect the pairs of marked edges adjacent to \( i_0 \), and the process is iterated finitely many times), and

\[
a_{k+1} \leq (a_k + 1 + l_k) - (u_k + l_k) = a_k + 1 - u_k,
\]

which again yields \( \sum_{1 \leq j \leq k} u_j \leq 2k \). To justify this last inequality, notice that in this situation,

\[
(i_{n_k}, i_{n_k+1}, \ldots, i_{n_2}) = (t_k, v, s_1, *, s_1, v, s_2, \ldots, s_{l_k-1}, v, s_{l_k}, *, s_{l_k}, v),
\]
the edges \((s_1, v), \ldots, (s_{l-1}, v)\) are unmarked, and \(S_{k+1} \subset S_k \cup \{(t, v), (v, s_1), \ldots, (v, s_l)\}\) with at least \(u_k + l_k\) elements contained in the latter set but not in the former.

Suppose next \(k = 1\). If \(v \neq i_0\), then

\[
i = (i_0, *, t_1, v, s_1, *, s_1, v, s_2, *, \ldots, s_1, v, s_{l+1}, *, i_0).
\]

Clearly \((s_1, v), \ldots, (s_l, v), (v, s_{l+1})\) are unmarked, which implies the other \(l + 1\) edges containing \(v\) are marked. If \(v = i_0\), then

\[
i = (v, s_0, *, t_1, v, s_1, *, s_1, v, s_2, *, \ldots, s_1, v, s_{l+1}, *, s_{l+2}, v);
\]

\((s_1, v), \ldots, (s_l, v), (s_{l+2}, v)\) are unmarked, and so there is exactly another unmarked edge containing \(v\) (which is of the form \((v, u)\) because \((t_1, v)\) is marked).

Step 5. Bound the expectation generated by \(i\). For any undirected edge \(e = uv\), denote by \(2k(e)\) the number of times \(e\) appears in \(i\). The assumption on the moments of \(b_{11}\) implies

\[
\mathbb{E}[b_i] \leq \prod_{e \in i, k(e) \geq 2} L(n) n^{8(2k(e)-4)} = L(n) n^{8(\sum k(e))_2} k(e) - 2E
\]

where \(E := \{|e \in i, k(e) \geq 2\}\). Any edge \(e = uv\) with \(k(e) \geq 2\), except possibly for \(i_0 i_1\), has either \(u\) or \(v\) in \(N_i(k)\) for some \(k \geq 2\): \(k(uv) \geq 2\) entails either the desired conclusion or \(u, v \in N_i(1)\), in which case \(k(uv) = 2\), and there are two marked copies of \(uv\) in \(i\), \((u, v), (v, u)\); suppose without loss of generality they appear in this order; then \((u, v)\) is the first edge of \(i\) (otherwise, for the vertex \(t\) preceding this apparition of \(u\), the edge \((t, u)\) is marked). This observation gives

\[
\sum_{k(e) \geq 2} k(e) \leq E + 1 + \sum_{k \geq 2} k n_k,
\]

whereby for \(n > n(\delta)\),

\[
\mathbb{E}[b_i] \leq L(n) n^{26(1+\sum_{k \geq 2} k n_k) - E} \leq L(n) n^{26\sum k \geq 2 k n_k}
\]

using that when \(E = 0\), the left-hand side is at most 1, and \(L(n) < n^{2\delta} \text{ for } E \geq 1\).

Putting together steps 1 – 5 yields

\[
\mathbb{E}[tr(B^{2p})] \leq L(n) C_p p! \sum_{n_1, \ldots, n_p} n^{1+\sum s_{i \leq p} n_k + 2\delta \sum k \geq 2 k n_k} \prod_{1 \leq k \leq p} \frac{1}{(k!)^{n_k n_k}} \prod_{2 \leq k \leq p} (2k)^{k n_k},
\]

where the summation is over \(p\)-tuples of nonnegative integers \((n_1, n_2, \ldots, n_p)\) with \(\sum_{1 \leq k \leq p} k n_k = p\).

Lastly, an upper bound can be computed when \(\delta = 1/4 - \delta, p \leq n^{\delta}, n \geq n(\delta, c)\):

\[
\mathbb{E}[tr(B^{2p})] \leq 2^{2p} L(n) n^{p+1} e^8.
\]

From above,

\[
\mathbb{E}[tr(B^{2p})] \leq n L(n) C_p p! \sum_{n_1, \ldots, n_p} n^{\sum_{1 \leq k \leq p} n_k + (1/2 - 2\delta)(p - n_1)} \prod_{1 \leq k \leq p} \frac{1}{(k!)^{n_k n_k}} \prod_{2 \leq k \leq p} (2k)^{k n_k} =
\]

\[
= n L(n) n^{(1/2 - 2\delta)(p - n_1)} \sum_{n_1, \ldots, n_p} n^{(1/2 + 2\delta) n_1 + \sum_{2 \leq k \leq p} n_k} \prod_{1 \leq k \leq p} \frac{1}{(k!)^{n_k n_k}} \prod_{2 \leq k \leq p} (2k)^{k n_k}.
\]
As \( n_1 = p - \sum_{k \geq 2} k n_k \) and \( p! \leq n_1 ! p^{n_1} = n_1 ! p^{\sum_{k \geq 2} k n_k} \), the last expression is at most
\[
n L(n)n^p C_p \sum_{(n_1, \ldots, n_p)} n \sum_{2 \leq k \leq p} (1 - \frac{k(1/2 + 2 \delta_1)}{2}) n_k \prod_{2 \leq k \leq p} \frac{p^{k n_k}}{(k!)^{n_k} n_k!} \prod_{2 \leq k \leq p} (2k)^{k n_k} \leq n L(n)n^p C_p \sum_{(n_1, \ldots, n_p)} n^{-2 \delta_1} \sum_{2 \leq k \leq p} k n_k \prod_{2 \leq k \leq p} \frac{p^{k n_k}}{(k!)^{n_k} n_k!} \prod_{2 \leq k \leq p} (2k)^{k n_k}
\]
employing \( \frac{1 - k(1/2 + 2 \delta_1)}{k} \leq \frac{1 - 2(1/2 + 2 \delta_1)}{2} = -2 \delta_1 \) for \( k \geq 2 \). Since \( C_p \leq 2^{2p} \), \( k! \geq (ke^{-1})^k \), and \( 2epn^{-2 \delta_1} \leq 2en^{-\delta_1} \leq 1 \), the above sum is upper bounded by
\[
2^{2p} L(n)n^{p+1} \sum_{(n_2, \ldots, n_p)} \prod_{2 \leq k \leq p} \frac{(pn^{-2 \delta_1})^{k n_k}}{(ke^{-1})^{kn_k} n_k!} = 2^{2p} L(n)n^{p+1} \sum_{(n_2, \ldots, n_p)} \prod_{2 \leq k \leq p} \frac{(2epn^{-2 \delta_1})^{k n_k}}{n_k!} \leq 2^{2p} L(n)n^{p+1} \exp \left( \sum_{2 \leq k \leq p} 2epn^{-2 \delta_1} \right) = 2^{2p} L(n)n^{p+1} \exp(2ep(p-1)n^{-2 \delta_1}) \leq 2^{2p} L(n)n^{p+1} e^8.
\]

2.2. Large Conditional Moments. This subsection proves
\[
(16)
\]
\[
E_s[tr(A^p) - tr(A^2 p)] \leq 2mc(k, c) \cdot (M^{2p} + n^{-\delta} (\max (2, M))^{2p} (2m)^{2p} (2p)^{16p^2} + s(p, M))
\]
for \( \delta = 1/4 - \delta_1, p \leq \sqrt{\log n}, m \leq n^{1/2}, n \geq n(\delta_1, c) \), and \( s : \mathbb{N} \times (0, \infty) \to [0, \infty) \) given by
\[
(17)
s(p, M) = \sum_{1 \leq l \leq p-1} M^{2l} \sum_{1 \leq t \leq p-l+1, 0 \leq l \leq \min(\frac{1}{2}, l)} \left( \frac{l - l_0 + t - 1}{l - l_0} \right) \left( \frac{t}{2l_0} \right) b_{p-l, t}.
\]

In the classical case underlying (12), the sole contributors to the trace are the even cycles, which are in turn mapped to tuples \( (n_1, n_2, \ldots, n_p) \) with \( \sum_{1 \leq k \leq p} k n_k = p \). In the current situation, this remains true, and the change of summation contains essentially one additional parameter: the non-zero entries of \( A_{B, s} \), a matrix whose sparsity (encoded by \( S \)) is vital towards obtaining (16).

Since \( A_s = A_{B, s} + A_{s} \), the left-hand side of (16) is a sum over cycles with contributions determined not only by their vertices, but also by whether their factors are entries of \( A_{B, s} \) or \( A_s \). Say \( a_{i j} \) belongs to \( A_{B, s}, A_s \) if \( (i, j) \in S, (i, j) \notin S \), respectively, where \( 1 \leq i, j \leq n \).

Then
\[
(18)
E_s[tr((A_s + A_{B, s})^{2p}) - tr(A^2 p)] = n^{-p} \sum_{(i_0, i_1, \ldots, i_{2p-1})} E_s[a_{i_0 i_1} a_{i_1 i_2} \ldots a_{i_{2p-1} i_0}]
\]
where all the entries appearing in the product belong either to \( A_s \) or \( A_{B, s} \), with at least one of them in the latter category. By independence, for any non-negative integers \( (p_{ij})_{1 \leq i \leq j \leq n} \),
\[
E_s\left[ \prod_{1 \leq i \leq j \leq n} a^{p_{ij}}_{ij} \right] = \prod_{1 \leq i \leq j \leq n, (i, j) \in S} E[a_{ij}^{p_{ij}} | \kappa \sqrt{n} < |a_{ij}| \leq M \sqrt{n}] \cdot \prod_{1 \leq i \leq j \leq n, (i, j) \notin S} E[a_{ij}^{p_{ij}} | \chi_{a_{ij}} | \leq n^\epsilon |a_{ij}| \leq \kappa \sqrt{n}].
\]

By symmetry, if some \( p_{ij} \) is odd, then the expectation is zero; else,
\[
E_s\left[ \prod_{1 \leq i \leq j \leq n} a^{p_{ij}}_{ij} \right] \leq c(\kappa, c) M^{\sum_{(i, j) \in S} p_{ij}} \cdot \prod_{1 \leq i \leq j \leq n, (i, j) \notin S} E[a_{ij}^{p_{ij}} | \chi_{a_{ij}} | \leq n^\epsilon].
\]
Choose the remaining vertices of $1 \leq \binom{m}{2}$. In other words, conditional moments can be replaced by unconditional ones for entries belonging to $A_s$, and by powers of $M$ for entries belonging to $A_{B_s}$, at a cost of a multiplicative factor $c(k, c)$. This observation is used when bounding the terms on the right-hand side of (18).

Keeping the terminology introduced in subsection 2.1, the above paragraph entails only even cycles contribute in (18). For such $i$, let $i', i''$ be the strings of 2p elements such that for $0 \leq t \leq 2p - 1$, if $a_{i_t, i_{t+1}}$ belongs to $A_s$, then $i'_t = (i_t, i_{t+1}), i''_t = \emptyset$; else, $i'_t = \emptyset, i''_t = (i_t, i_{t+1})$, and adopt $i = (i', i'')$ as a shorthand for this decomposition. Put differently, $i', i''$ record the edges of $i$ belonging to $A_s$ and $A_{B_s}$, respectively; moreover, by ignoring the empty set entries in these sequences, they can be naturally seen as subgraphs of $i$, an interpretation implicitly assumed henceforth. An important observation is that $i, i', i''$ share the property underlying even cycles: any undirected edge appears in each of them an even number of times (since no entry belongs to both $A_s$ and $A_{B_s}$). Steps 1' - 5' below consider the contributions of cycles $i$ for $i''$ fixed, while step 6' sums them over all such directed graphs.

Proceed with the first five steps: since $i''$ is fixed at this stage (denote its length by $2l$), the second summation in (18) is over $i'$ with $i = (i', i''_t)$, and as in the classical case, a change of summation is employed: from $i'$ to tuples $(n'_1, n'_2, \ldots, n'_{p-l})$ with $\sum_{1 \leq k \leq p-l} \kappa n'_k = p - l$. For any even cycle $i$, let $N'_i(k)$ be the set of vertices of $i$ appearing as right endpoints of marked edges of $i'$ exactly $k$ times, and $n'_k := |N'_i(k)|$ for $1 \leq k \leq p - l$ (since $i'$ and $i''$ share no undirected edge, marking them either separately or jointly in $i$ leads to the same configuration of marked edges). In what follows, $i = (i', i''_l)$ is an even cycle with $(n'_1, n'_2, \ldots, n'_{p-l})$ fixed and $1 \leq l \leq p - 1$. Although steps 1 - 5 do not generally hold when $(n_1, n_2, \ldots, n_p)$ is replaced by $(n'_1, n'_2, \ldots, n'_{p-l})$, they can be modified and still yield useful bounds.

Step 1'. Map the marked edges of $i'$ to a Dyck path of length $2p - 2l$. The number of such paths is at most $C_{p-l} = \frac{1}{p-l} \binom{2p-2l}{p-l}$.

Step 2'. Select the order of the marked vertices in $i'$: the number of possibilities is at most

$$\frac{(p - l)!}{\prod_{1 \leq k \leq p-l} (k!)^{n'_k}} \cdot \frac{1}{\prod_{1 \leq k \leq p-l} n'_k^{l_k}}.$$

Step 3'. Choose the distinct vertices of $i'$, $V(i')$, one at a time by reading its edges in order: the number of possibilities is at most

$$2m \cdot n^{\sum_{1 \leq k \leq p-l} n'_k}.$$

Each vertex of $i'$ is $i_0$, marked at least once, or some endpoint of an edge in $i''_t$; hence, only the first two categories, whose union has size at most $1 + \sum_{1 \leq k \leq p-l} n'_k$, are yet to be chosen. Since $l > 0$, let $v = i_t$ be the first vertex in $i$ appearing also in $i''$ (i.e., $i''$ contains some edge adjacent to $v$, and $t$ is minimal). If $v = i_0$, then it can be chosen in at most $2m$ ways ($uv \in i''$ yields $(\min(u, v), \max(u, v)) \in S$, and $|S| \leq m$), and the desired bound follows. Otherwise, $(u, v) := (i_{t-1}, i_t)$ is marked in $i'$ by the definitions of $t$ and $v$, this edge belongs to $i'$ and contains the first apparition of $v$ in $i$, and so there are at most $2m \cdot n^{\sum_{1 \leq k \leq p-l} n'_k - 1}$ possibilities ($v$ is both an element of a fixed set of size at most $2m$, and of $N'_i(k)$, for some $1 \leq k \leq p - l$).

Step 4'. Choose the remaining vertices of $i'$ among the ones selected in step 3'. At this stage, $(i_0, i_1)$ is fully determined since it is either in $i''$ or marked in $i'$. The same rationale as in step 4 shows only the right endpoints of the unmarked edges of $i'$ have yet to be chosen, which can be done in at most $(2l + 1)!d^l \prod_{2 \leq k \leq p-l} (2k + 2l)^k n'_k$ ways:
LEMMA 2. If $v \in N_1^i(k)$, then the number of unmarked edges of the form $(v, u)$ is at most
$$\begin{cases} 2k + 2l, & k > 1 \\ 1, & k = 1 \\ 2l + 2, & v \in E(i) \end{cases},$$
where $|E(i)| \leq 4l$.

**Proof.** If $v \in N_1^i(k)$, then $v \in N_1^i(k')$ with $k \leq k' \leq k + l$ ($i''$ contains $l$ marked edges). Lemma 1 then gives the result above for $k > 1$. For $k = 1$, there is at most one possibility unless there exists an edge in $i''$ containing $v$ (else, the proof of Lemma 1 for this case is still valid); let $E(i)$ be the set of such vertices. For $v \in E(i)$, there are at most $2(l + 1)$ such edges, and $|E(i)| \leq 2 \cdot 2l = 4l$ ($i''$ contains $2l$ edges).

**Step 5'.** Let $E' = \{|e \in i': k(e) > 2\}$. For $a = a_{11}1_{[a_{11}] \leq n^4}, q \in \mathbb{N}, q \geq 2,$
$$\mathbb{E}[a^2] \leq 1, \quad \mathbb{E}[a^{2q}] \leq c(\delta, c) \log n \cdot n^{6(2q-4)} := L(n)n^{6(2q-4)},$$
and so
$$\mathbb{E}[a_{i'}] \leq L(n)^{(p-l)/2} \prod_{e \in V, k(e) > 2} n^{6(2k(e)-4)} = L(n)^{(p-l)/2}n^{2\delta(\sum_{e \in V, k(e) > 2} k(e)-2E')}$$
as there are at most $(p-l)/2$ pairwise distinct undirected edges in $i'$, each appearing at least four times in $i'$. Because every edge $e = uv$ with $k(e) > 2$ has either $u \in N_1^i(k)$ or $v \in N_1^i(k)$ for some $k \geq 2$,
$$\sum_{e \in V, k(e) > 2} k(e) \leq E' + \sum_{k \geq 2} kn'_{k},$$
providing
$$\mathbb{E}[a_{i'}] \leq L(n)^{(p-l)/2}n^{2\delta(\sum_{k \geq 2} kn'_{k}-E')} \leq L(n)^{(p-l)/2}n^{2\delta(\sum_{k \geq 2} kn'_{k})}.$$}

Furthermore, an overall saving of some power of $n$ is possible unless $i$ has a very special form.

**Lemma 3.** For any even cycle $i$ with $l < p$, at least one of the following occurs:

1. a factor of $n^{2\delta}$ can be saved in step $5'$:
   $$\mathbb{E}[a_{i'}] \leq L(n)^{(p-l)/2}n^{2\delta(\sum_{k \geq 2} kn'_{k})},$$

2. a factor of $n/(2m)$ can be saved in step $3'$:
   $$(2m)^2 \cdot n \sum_{1 \leq k \leq n_{i_0} - 1} n_{i_0}^2 - 1,$$

3. $n_{i_0} = p - l, i_0$ is unmarked in $i'$, $i''$ contains a unique undirected edge $vw$ with $v \in N_1^i(1) \cup \{i_0\}$ and $w \neq i_0$ unmarked in $i'$.

**Proof.** Since
$$\mathbb{E}[a_{i'}] \leq L(n)^{(p-l)/2}n^{2\delta(\sum_{k \geq 2} kn'_{k}-E')},$$
(1) follows unless $E' = 0$ and $n'_{i_0} = p - l$; if $E' > 0$, then it is clear; if $E' = 0, n'_{i_0} \neq p - l$, then $\sum_{k \geq 2} n'_{i_0} \geq 1$ and the desired inequality holds too. What is left is the case $E' = 0, n'_{i_0} = p - l$.

If $i''$ contains at least two distinct undirected edges, then (II) holds. Suppose the condition is satisfied. Without loss of generality, assume each cluster of edges in $i''$ has size at most one (such a block is fully determined by its length and its first edge because any two distinct undirected edges of $i''$ share no vertex; therefore, if the first edge of the cluster is $(u, v)$,
then its edges are \((u, v), (v, u), (u, v), \ldots\), which can be compressed to \(u, (u, v)\) for even, odd length, respectively, without affecting \(i'\). Each undirected edge \(e = vw\) of \(i''\) is adjacent to either a marked vertex in \(i'\) or \(i_0\) : take \(s = \min\{0 \leq t \leq 2p, i_t \in \{v, w\}\}\); if \(s = 0\), then \(e\) is adjacent to \(i_0\); else, \(s > 0\), and \((i_{s-1}, i_s) \in i'\), because otherwise \(i_{s-1} \in \{v, w\}\), and is marked since \(i_s\) is the first apparition of a vertex in \(i\). This observation implies \((II)\).

If \(i''\) contains solely one undirected edge \(vw\), then \((II)\) holds unless \((III)\) is satisfied. In this case, if \(i_0\) is marked in \(i'\), then \((II)\) holds since \(\sum_{k \geq 1} n'_k\) can replace \(1 + \sum_{k \geq 1} n'_k\) in step \(3\). If after compressing the clusters either both \(v\) and \(w\) are marked or one is \(i_0\) and the other marked, then again some saving is possible and \((II)\) ensues. Else, \((III)\) holds using \(\{v, w\} \cap (N_i(1) \cup \{i_0\}) \neq \emptyset\).

In conclusion, merging steps \(3'\) and \(5'\) yields the overall contribution of cycles of type \((I)\) and \((II)\) is at most

\[
2m \cdot n^{2\delta} \left(\sum_{k \geq 2} n'_k + \sum_{k \geq 1} n'_k L(n) (p - l)/2\right) \cdot (n^{-2\delta} + 2m \cdot n^{-1}) \leq 4m \cdot n^{2\delta} \cdot n^{2\delta} \sum_{k \geq 2} n'_k + \sum_{k \geq 1} n'_k,
\]

for \(p \leq \sqrt{\log n}\) and \(n\) large enough.

Putting steps \(1' - 5'\) together, the computations at the end of the previous subsection can be used with the substitutions \(p \rightarrow p - l, n'_k \rightarrow n_k, 2k \rightarrow 2k + 2l,\) and \(\delta = 1/4 - \delta_1\), \(p \leq \sqrt{\log n}\) \((8ep^2n^{-2\delta_i}\) replaces \(2epm^{-2\delta_i}\) from \(\frac{2k+2l}{\log n} \leq 4pe\)). Thus, this sum is upper bounded by

\[
(20) \quad c(\kappa, c) \cdot 4mn^{-\delta} 2^{2p-2l} \cdot (2l + 2l)!^4 e^{16}
\]

for \(p \leq \sqrt{\log n}\) and \(n\) sufficiently large.

**Step 6'.** The conditional expectation coming from \(i''\) is not larger than \(M^{2l}\), and there are at most \((\frac{2p}{2l}) (2m)^{2l}\) such directed cycles for \(1 \leq l \leq p - 1\), and \(2|S| \leq 2m\) for \(l = p\) (cycles with all edges belonging to \(A_{B,\kappa}\) are fully determined by their first edge). Hence, using \((20)\) and \((19)\), the overall contribution of cycles of types \((I)\) and \((II)\) is at most

\[
(21) \quad 2m \cdot M^{2p} + c(\kappa, c) \cdot 4mn^{-\delta} e^{16} \sum_{1 \leq l \leq p - 1} \left(\frac{2p}{2l}\right) (2m)^{2l} \cdot 2^{2p-2l} \cdot (2l + 2l)!^4 M^{2l}.
\]

Consider now the cycles of type \((III)\) : they generate a term less or equal than

\[
(22) \quad c(\kappa, c) \cdot 2m \sum_{1 \leq l \leq p - 1} M^{2l} \sum_{1 \leq t \leq p - l + 1, 0 \leq l_0 \leq \min\{\frac{1}{2}, l\}} \binom{l - l_0 + t - 1}{l - l_0} \binom{t}{2l_0} b_{p-l,t}.
\]

To see this, let \(vw\) be the edge appearing in a fixed \(i''\) of length \(2l\). Map the compressed version of any cycle \(i\) of type \((III)\) to an element of \(j \in C(p - l)\) by replacing each cluster \(\{v, w, (v, v), (w, w)\}\) by a new vertex \(\rho\) (note \(\rho\) is marked in \(j\) exactly when \(v\) and \(w\) are in \(i'\), hence, this procedure generates cycles in \(C(p - l)\)). It is shown next that the preimage of any \(j \in C(p - l)\) with \(\rho\) a fixed vertex in it of multiplicity \(t\), and the first cluster of \(i\) containing \(v\), has size in the following interval

\[
[\sum_{0 \leq l_0 \leq \min\{\frac{1}{2}, l\}} \binom{l - l_0 + t - 1}{l - l_0} \binom{t - 1}{2l_0}, \sum_{0 \leq l_0 \leq \min\{\frac{1}{2}, l\}} \binom{l - l_0 + t - 1}{l - l_0} \binom{t}{2l_0}],
\]

whereby \((22)\) is fully justified since any element of \(C(l)\) contains \(l\) pairwise distinct undirected edges (this ensues by induction and the recursive description of \(C(l)\) : see proof of \((25)\) in subsection 2.3.3), which together with this mapping gives \(E[a_i] \leq 1\) (although the upper bound suffices for \((22)\), the lower bound comes into play in subsection 3.2.2).

Suppose first \(\rho \neq j_0\), and let \((s_0, \rho, s_1), \ldots, (s_{2t-2}, \rho, s_{2t-1})\) be the apparitions of \(\rho\) in \(j\) in increasing order: \((s_2, \rho), (s_4, \rho), \ldots, (s_{2t-2}, \rho), (\rho, s_{2t-1})\) are unmarked and so
(\rho, s_1), (\rho, s_3), \ldots, (\rho, s_{2t-3}) are marked. Denote by 2l_0 the number of clusters of size two in \( i \), an even cycle in the preimage of \( j \) (this number is even because \( i \) is): then
\[
0 \leq l_0 \leq \min\left(\frac{t}{2}, l\right)
\]
(each contains an edge of \( i^0 \)) and thus, the preimage has at most
\[
\left(\frac{l}{l-l_0+1}\right)\binom{t}{2l_0}
\]
elements since once the sizes of the clusters underlying \( \rho \) are fixed, by induction on \( 1 \leq k \leq t \), the \( k^{th} \) cluster \((s_{2k-2}, \rho, s_{2k-1})\) is fully determined: for \( k = 1 \), it is clear as \( \rho \in \{v, (v, w)\} \) and has fixed size; for \( k \geq 2 \), the counterpart of \((s_{2k-2}, \rho)\) has already been decided, yielding the first vertex of the cluster, and its size dictates whether it has a second vertex or not, which is then fully determined by the first. Lastly, choosing tuples of even integers \((2x_1, 2x_2, \ldots, 2x_t)\) with \( x_1 + \ldots + x_t = l - l_0, x_j \geq 0 \) can be done in
\[
\binom{l + l_0 - 1}{l - l_0}
\]
ways (for \( t, l \in \mathbb{N}, \) let \( \tilde{a}_{t, l} = |\{(2x_1, 2x_2, \ldots, 2x_t) : x_1 + \ldots + x_t = l, x_j \in \mathbb{Z}, x_j \geq 0\}|\); clearly, \( \tilde{a}_{1, l} = 1 \), and for \( t \geq 2 \), \( \tilde{a}_{t, l} = \tilde{a}_{t-1, l} + \tilde{a}_{t, l-1} \), as \( x_1 = 0 \) or \( x_1 > 0 \); this yields by induction on \( t + l, \tilde{a}_{t, l} = \binom{l + l_0 - 1}{l - l_0} \).

Conversely, select \( 2l_0 \) elements out of \( \{1, 2, \ldots, t - 1\} \) with \( 0 \leq l_0 \leq \min\left(\frac{t-1}{2}, l\right) \): call this set \( D \). Then there is an even cycle \( i \) in the preimage of \( j \) with the \( d^{th} \) apparition of \( \rho \) replaced by two vertices for \( d \in D \) and by one if \( d \notin D, d \neq t \). The last cluster is fully determined by the previous ones, while for the rest there are two possibilities: traverse the clusters from left to right; for the \( k^{th} \), the first edge, \((s_{2k-2}, \rho)\) is fixed (it is unmarked), while the second can be chosen in two ways because it is marked, generating two scenarios for \( \rho \), in which it has size one and two, respectively. Thus, it is always possible to decide the size of any of its apparitions, and \(|D|\) even ensures the cycle obtained is also even. Lastly, the case \( \rho = j_0 \) is analogous to \( \rho \neq j_0 \), the main difference being that the apparitions of \( \rho \) are \((\rho, s_1), (s_2, \rho, s_3), \ldots, (s_{2t-2}, \rho, s_{2t-1}), (s_{2t}, \rho)\). Finally, (16) ensues from (21) and (22).

2.3. Asymptotics of Large Conditional Moments. This subsection justifies for \( M > 0 \),
\[
\lim_{p \to \infty} s(p, M)^{1/2p} = \begin{cases} 
2, & M \leq 1 \\
\frac{M^2 + 1}{M}, & M \geq 1 
\end{cases}
\]
Recall that
\[
s(p, M) = \sum_{1 \leq l \leq p-1} M^{2l} \sum_{1 \leq t \leq p-l+1, 1 \leq l_0 \leq \min\left(\frac{t}{2}, l\right)} \binom{l - l_0 + t - 1}{l - l_0} \binom{t}{2l_0} b_{p-l,t}.
\]
For \( t \geq l + 2, \) \( b_{l,t} = 0 \) (any cycle in \( C(l) \) contains \( 2l + 1 \) vertices, out of which at least \( l + 1 \) are pairwise distinct: \( i_0 \) and the marked vertices), and \( b_{l+1,l+1} = 1 \): the sole element of \( C(l) \) with a vertex repeated \( l + 1 \) times is \((u_0, u_1, u_0, \ldots, u_0, u_1, u_0)\) with \( u_0, u_1, \ldots, u_l \) pairwise distinct \( (l \) vertices have multiplicity \( 1 \), and one has multiplicity \( l + 1 \); since the first vertex appears at least twice, its multiplicity is \( l + 1 \), and all the other vertices show up exactly once, yielding a unique cycle up to isomorphism).

(23) is concluded in two phases, first, computing the asymptotic behavior of \((b_{l,t})\):
\[
\frac{1}{4l} \cdot \left(\frac{2l + 1 - t}{l}\right) \leq b_{l,t} \leq (l + 1)^{120} \cdot \left(\frac{2l + 1 - t}{l}\right)
\]
for \( 1 \leq t \leq l + 1 \), and second, using binomial proxies for \( b_{l,t} \) in \( s(p, M) \) (any polynomial factor becomes negligible when \( p \to \infty \)).

Proceed with understanding the sizes of \((b_{l,t})\)_{1 \leq l \leq t \leq t+1}, for which two recursions, describing the sequence itself and \((C_l)_{t \geq 1}\), respectively, are crucial. Observe that
\[
|C(l)| = C_l,
\]
where $C_l$ is the $l^{th}$ Catalan number, $C_l = \frac{1}{l+1} \binom{2l}{l}$; on the one hand, steps 1, 2, 4 from sub-section 2.1 yield $|C(l)| \leq C_l$. On the other hand, recall the recursive characterization of the Catalan numbers:

$$C_{l+1} = \sum_{0 \leq a \leq l} C_a C_{l-a}, \quad C_0 = C_1 = 1.$$ 

Since $|C(1)| = 1$, justifying

$$|C(l+1)| \geq 2|C(l)| + \sum_{1 \leq a \leq l-1} |C(a)| \cdot |C(l-a)|$$

yields by induction $|C(l)| \geq C_l$. To do so, construct three types of cycles among the elements of $C(l+1)$:

(i) $i = (v_0, v_1, ..., v_{2l}, v_0) \in C(l)$ with an extra loop at $v_0: (v_0, u, v_0, v_1, ..., v_{2l}, v_0)$, and $u$ new (i.e., not among the vertices of $i$);

(ii) $i = (v_0, v_1, ..., v_{2l}, v_0) \in C(l)$ with an extra loop at $v_1: (v_0, v_1, u, v_1, v_2, ..., v_{2l}, v_0)$, and $u$ new;

(iii) $(u_0, u_1, u_2, S_1, u_2, u_1, S_2)$ with $(u_2, S_1, u_2) \in C_a, (u_0, u_1, S_2) \in C_{l-a}$, no vertex appearing in both, $u_0, u_1, u_2$ pairwise distinct, and $1 \leq a \leq l-1$.

Clearly, these three families are pairwise disjoint (by considering the second apparitions of the first three vertices of such cycles), and their union is a subset of $C_{l+1}$ of size

$$|C(l)| + |C(l)| + \sum_{1 \leq a \leq l-1} |C(a)| \cdot |C(l-a)|.$$

The proof of (25) is thus complete, and its crucial by-product is:

$$C(l+1) \text{ consists of three components: (i), (ii), and (iii).}$$

Returning to (24), (25) yields the desired chain of inequalities for $t = 1$:

$$C_{l-1} = |\{(v_0, u, v_0, v_1, ..., v_{2l}, v_0), (v_0, v_1, ..., v_{2l}, v_0) \in C(l-1)\}| \leq b_{l-1} \leq \sum_{1 \leq t \leq l+1} b_{l,t} = (2l+1)C_l,$$

from which

$$\frac{1}{4l-2} \cdot \binom{2l}{l} \leq b_{l-1} \leq \frac{2l+1}{l+1} \cdot \binom{2l}{l}.$$

Let now $2 \leq t \leq l+1$, and denote by $f_{l,t}$ the number of elements of $C(l)$ in which the first vertex appears exactly $t$ times; similarly, $s_{l,t}$ is the number of elements in $C(l)$ whose second vertex has multiplicity $t$. The aforementioned description of $C(l+1)$ gives

$$b_{l+1,t} = (f_{l,t-1} + f_{l,t} + s_{l,t-1} + s_{l,t}) + \sum_{1 \leq a \leq l-1} \left[ C_{l-a} b_{a,t} + C_a (b_{l-a,t} - s_{l-a,t} + s_{l-a,t-1}) \right]$$

since a vertex $v$ has multiplicity $t > 1$ if and only if

(i) its multiplicity in $i$ is $t$, $v = v_0$; (ii) its multiplicity in $i$ is $t$, $v \neq v_0$; (iii) $v \in (u_2, S_1, u_2)$ has multiplicity $t$, or $v \in (u_0, u_1, S_1)$ has multiplicity $t$, $v \neq u_1$.

By rearranging terms,

(26)

$$b_{l+1,t} = 2 \sum_{0 \leq a \leq l-1} C_a b_{l-a,t} + (f_{l,t-1} - f_{l,t}) + (s_{l,t-1} - s_{l,t}) + \sum_{1 \leq a \leq l-1} C_a (s_{l-a,t-1} - s_{l-a,t}).$$
LEMMA 4. For \( k \geq -1 \) and \( l \geq k + 1 \),
\[
f_{l,t-k} = \binom{l+k}{k} - \binom{l+k}{k+1}.
\]

PROOF. Reasoning as above, the description of \( C(l+1) \) gives for \( 2 \leq t \leq l+1 \),
\[
f_{l+1,t} = f_{l,t-1} + f_{l,t} + \sum_{1 \leq a \leq l-1} C_a f_{l-a,t},
\]
where \( f_{l,t} = 0 \) if \( t > l + 1 \). Proceed by induction on \( k \geq -1 \) for \( l \geq k + 1 \). The base case \( k = -1 \) is immediate from \( f_{l,t+1} = b_{l,t+1} = 1 \). Let \( k \geq 0 \) be fixed.

Use induction now on \( l \geq k + 1 \); the base case \( l = k + 1 \) is clear inasmuch as the first vertex of any even cycle has multiplicity at least two. Take next \( l \geq k + 2 \); since
\[
\binom{l+k+1}{k+1} - \binom{l+k+1}{k} = \binom{l+k}{k+1} + \binom{l+k}{k} - \binom{l+k}{k+1} \Rightarrow \binom{l+k}{k} = f_{l,t-k} + f_{l,t-k+1} + \sum_{1 \leq a \leq l-1} C_a f_{l-a,t-k+1},
\]
the summation is over \( 1 \leq a \leq k \).

In light of (27) rewritten for \( l \geq k + 1 \) as
\[
f_{l+1,t-k+1} = f_{l,t-k} + f_{l,t-k+1} + \sum_{1 \leq a \leq k} C_a f_{l-a,t-k+1},
\]
it suffices to prove that for \( l \geq k + 1 \),
\[
A(l,k) := \binom{l+k-1}{k-1} - \binom{l+k-1}{k-2} = \sum_{1 \leq a \leq k} C_a(\binom{l+k-1-2a}{k-a} - \binom{l+k-1-2a}{k-a-1}) := B(l,k).
\]

Because
\[
A(l,k) = A(l-1,k) + A(l-1,k-1), \quad B(l,k) = B(l-1,k) + B(l-1,k-1),
\]
\( B(l,k) - B(l-1,k) \) the summation is over \( 1 \leq a \leq k - 1 \) since the coefficient of \( C_k \) in \( B(l,k) \) is 1 from \( \binom{n}{m} = \binom{n-1}{m} + \binom{n}{m-1} \), showing \( A(l,k) = B(l,k) \) for \( l = k + 1 \) is sufficient as induction on \( l - k \geq 1 \) yields the desired identity (for \( d = l - k \) fixed, induction is used anew on \( l \geq k + 1 \), whose base case \( l = d + 1, k = 1 \) is immediate because \( f_{d+1,1} = 0 = \binom{2d+1}{d+1} - \binom{2d}{d} \)). In this situation, the desired result is
\[
\binom{2k}{k-1} - \binom{2k}{k-2} = \sum_{1 \leq a \leq k} C_a(\binom{2k-2a}{k-a} - \binom{2k-2a}{k-a-1}).
\]

Note the right-hand side term is
\[
\sum_{1 \leq a \leq k} C_a C_{k-a} = C_{k+1} - C_k = C_k \cdot \frac{3k}{k+2}.
\]
and
\[
\binom{2k}{k} - \binom{2k}{k-2} = C_k \cdot \binom{3k}{k+2} = \binom{2k}{k} \cdot \frac{3k}{(k+2)(k+1)}
\]
since simplifying this equation by \(k!(k-1)!/(2k)!\) turns it equivalent to
\[
\frac{1}{k+1} - \frac{k-1}{(k+1)(k+2)} = \frac{1}{k} \cdot \frac{3k}{(k+2)(k+1)} = \frac{3}{(k+1)(k+2)}.
\]
\[\square\]

Notice \(s\) and \(f\) are comparable, i.e.,
(29)  
\[f_{t-1,t} \leq s_{t,t} \leq f_{t,t+1} :\]
if \(i \in C(l - 1)\) has the first vertex of multiplicity \(t\), then \((u_0,i,u_0) \in C(l)\) has the second vertex of multiplicity \(t\), implying the first inequality. The upper bound ensues by induction on \(t + l\) from (27): using the description of \(C(l + 1)\), \(s_{l,1} = C_{l-1} = f_{l,2}\), and for \(2 \leq t \leq l + 1\),
\[s_{l+1,t} = s_{l,t-1} + \sum_{1 \leq a \leq t-1} C_a s_{l-a,t}.
\]
Now the growth of \((b_{l,t})_{2 \leq l \leq l+1}\) can be fully established.

**Lemma 5.** For \(2 \leq t \leq l + 1\),
\[f_{t,t} \leq b_{l,t} \leq (l+1)^{120} \cdot f_{l,t}.
\]
**Proof.** The lower bound is evident. Focus next on the second inequality. Recall (26):
\[b_{l+1,t} = 2 \sum_{0 \leq a \leq l-1} C_a b_{l-a,t} + (f_{l,t-1} - f_{l,t}) + (s_{l,t-1} - s_{l,t}) + \sum_{1 \leq a \leq l-1} C_a (s_{l-a,t-1} - s_{l-a,t}).
\]
For the time being, let \(\Sigma_s = \sum_{1 \leq a \leq l-1} C_a (s_{l-a,t-1} - s_{l-a,t})\). Subtract \(f_{l+1,t}\) from both sides
\[b_{l+1,t} - f_{l+1,t} = 2 \sum_{0 \leq a \leq l-1} C_a b_{l-a,t} + (f_{l,t-1} - f_{l,t}) + (s_{l,t-1} - s_{l,t}) + \Sigma_s - f_{l+1,t},
\]
use (27)
\[b_{l+1,t} - f_{l+1,t} = 2 \sum_{0 \leq a \leq l-1} C_a b_{l-a,t} - 2 f_{l,t} - \sum_{1 \leq a \leq l-1} C_a f_{l-a,t} + (s_{l,t-1} - s_{l,t}) + \Sigma_s,
\]
change \(b_{l-a,t}\) to \(b_{l-a,t} - f_{l-a,t}\)
\[b_{l+1,t} - f_{l+1,t} = 2 \sum_{0 \leq a \leq l-1} C_a (b_{l-a,t} - f_{l-a,t}) + \sum_{1 \leq a \leq l-1} C_a f_{l-a,t} + (s_{l,t-1} - s_{l,t}) + \Sigma_s,
\]
dispense with the second summation
\[b_{l+1,t} - f_{l+1,t} = 2 \sum_{0 \leq a \leq l-1} C_a (b_{l-a,t} - f_{l-a,t}) + (f_{l+1,t} - f_{l,t-1} - f_{l,t}) + (s_{l,t-1} - s_{l,t}) + \Sigma_s.
\]
Consider now \(\Sigma_s\) : (29) implies
\[\sum_{1 \leq a \leq l-1} C_a (s_{l-a,t-1} - s_{l-a,t}) \leq \sum_{1 \leq a \leq l-1} C_a f_{l-a,t} \leq f_{l+1,t},
\]
whereby \( b_{l+1,t} - f_{l+1,t} \) is at most
\[
2 \sum_{0 \leq a \leq l-1} C_a(b_{l-a,t} - f_{l-a,t}) + (f_{l+1,t} - f_{l,t} - f_{l,t-1}) + (f_{l,t} - f_{l-1,t-1}) + f_{l+1,t},
\]
from which
\[
(30) \quad b_{l+1,t} - f_{l+1,t} \leq 2 \sum_{0 \leq a \leq l-1} C_a(b_{l-a,t} - f_{l-a,t}) + 2f_{l+1,t}.
\]

Induction on \( k \geq -1 \) entails
\[
b_{l,k} - f_{l,k} \leq (l+1)^{120} \cdot f_{l,k}.
\]
When \( k = -1, \) \( b_{l,l-1} - f_{l,l-1} = 0. \) Let now \( k \geq 0 \) be fixed, and suppose \( l \geq 4 \) (if \( l \leq 3, \) then \( b_{l,l} = (2l+1)C_l < 2^{120}. \) In light of \( (30) \) with \( t = l - k + 1 \geq 2, \) it suffices to show
\[
2 \sum_{0 \leq a \leq l-1} C_a(l - a + 1)^{120} f_{l-a,t} + 2f_{l+1,t} \leq (l + 2)^{120} \cdot f_{l+1,t}.
\]
to see this, note that
\[
\frac{f_{l,t-k+1}}{f_{l+1,t-k+1}} = \frac{k+1}{k} \cdot \frac{(l+k-1)!}{(l+1)!} = \frac{(k+1)(l+1)}{(l+k)(l+k+1)} \leq \frac{l+1}{(\sqrt{l+1} + \sqrt{l})^2} \leq \frac{l+1}{4 + \frac{2}{l}},
\]
employing \( \frac{x}{(x+\alpha)(x+\beta)} \leq \frac{1}{(\sqrt{\alpha}+\sqrt{\beta})^2} \) for \( x, \alpha, \beta > 0. \) This inequality implies for \( 0 \leq a \leq \frac{l}{2} - 1 \)
\[
\frac{f_{l-a,t-k+1}}{f_{l+1,t-k+1}} \leq \left( \frac{1}{4} + \frac{2}{l} \right) \cdots \left( \frac{1}{4} + \frac{2}{l-a} \right) \leq \left( \frac{1}{4} + \frac{4}{l+2} \right)^{a+1},
\]
from which
\[
2 \sum_{0 \leq a \leq l/2 - 1} C_a(l - a + 1)^{120} f_{l-a,t} \leq 2(l + 1)^{120} \cdot f_{l+1,t} \sum_{0 \leq a \leq l/2 - 1} \left( \frac{1}{4} + \frac{4}{l+2} \right)^{a+1} \leq 2(l + 1)^{120} \cdot f_{l+1,t} \cdot \frac{1/4 + 4/(l + 2)}{3/4 - 4/(l + 2)} = 2(l + 1)^{120} \cdot f_{l+1,t} \cdot \frac{l + 18}{3l - 10},
\]
while
\[
2 \sum_{l/2 - 1 < a \leq l} C_a(l - a + 1)^{120} f_{l-a,t} \leq 2(l/2 + 2)^{120} \cdot (f_{l+1,t} - f_{l,t-1} - f_{l,t}) \leq 2(l/2 + 2)^{120} \cdot f_{l+1,t}.
\]
The claim follows from
\[
2 \left( \frac{l + 1}{l + 2} \right)^{120} \cdot \frac{l + 18}{3l - 10} + \left( \frac{l + 4}{2l + 4} \right)^{120} \cdot 2 + \frac{2}{(l + 2)^{120}} \leq 1.
\]
the left-hand side is at most
\[
\left( \frac{5}{6} \right)^{120} \cdot 2 + \left( \frac{2}{3} \right)^{120} \cdot 2 \leq 2 \cdot \frac{25}{6^{120}} \leq 25 \cdot \left( \frac{5}{6} \right)^{120} \leq \frac{25}{1 + 120 \cdot 1/5} = 1.
\]

To finalize \( (24), \) note the result is clear for \( t = l + 1, \) and for \( t = l - k + 1, 1 \leq k \leq l - 1, \)
\[
f_{l,t-k+1} = \frac{l - k}{k} \left( \frac{l + k - 1}{k - 1} \right) = \frac{l - k}{l + k} \left( \frac{l + k}{l} \right) \in \left[ \frac{1}{2l} \left( \frac{l + k}{l} \right) \right].
\]
which in conjunction with Lemma 5 entails the desired inequalities.

Having completed (24), proceed with the last missing piece of this subsection, (23). Let \( x! := (\lfloor x \rfloor)! \) for \( x \geq 0 \): Stirling’s formula yields
\[
C_1 \sqrt{n} \cdot \left( \frac{n}{e} \right)^n \leq n! \leq C_2 \sqrt{n} \cdot \left( \frac{n}{e} \right)^n
\]
for universal constants \( C_1, C_2 > 0 \) and all \( n \geq 0 \), from which
\[
\lim_{p \to \infty} s(p, M)^{1/p} = \sup M^{2x} \frac{(x-y+z)^{x-y+z}}{(x-y)^{x-y}(2y)^{2y(z-2y)^{z-2y}} \cdot \frac{(2-2x-z)^{2-2x-z}}{(1-x)^{1-x}(1-x-z)^{1-x-z}}}
\]
over \( 0 \leq x \leq 1, 0 \leq z \leq 1-x, 0 \leq y \leq \min \left( \frac{x}{y}, x \right) \) : take \( l = px, l_0 = py, t = pz \), and so
\[
\left( M^2(l - l_0 + t) \left( \frac{t}{l - l_0} \right) \left( 2p - 2l - t \right) ^{1/p} \right) = \left( M^2 \left( px - py + pz \right) \left( \frac{pz}{2py} \right) \left( 2p - 2px - pz \right) ^{1/p} \right) = M^{2x} \frac{(x-y+z)^{x-y+z}}{(x-y)^{x-y}(2y)^{2y(z-2y)^{z-2y}} \cdot \frac{(2-2x-z)^{2-2x-z}}{(1-x)^{1-x}(1-x-z)^{1-x-z}} + o(M)
\]
with the last term tending to 0 as \( p \to \infty \) uniformly in \( x, y, z \). (23) is a consequence of:

**Lemma 6.** For \( M > 0 \) fixed, let \( h : D \to (0, \infty) \),
\[
h(x, y, z) = M^{2x} \cdot \frac{(x-y+z)^{x-y+z}}{(x-y)^{x-y}(2y)^{2y(z-2y)^{z-2y}} \cdot \frac{(2-2x-z)^{2-2x-z}}{(1-x)^{1-x}(1-x-z)^{1-x-z}}}
\]
where \( D = \{(x, y, z) \in \mathbb{R}^3 : 0 \leq x \leq 1, 0 \leq z \leq 1-x, 0 \leq y \leq \min \left( \frac{x}{y}, x \right) \} \) and \( 0^0 := 1 = \lim_{x \to 0+} x^x \). Then
\[
\sup_{(x, y, z) \in D} h(x, y, z) = \begin{cases} 4, & M \leq 1, \\ \frac{(M^2+1)^2}{M^2}, & M \geq 1 \end{cases}
\]

**Proof.** For \( a + bx > 0 \),
\[
\frac{\partial}{\partial x} ((a + bx)^{(a+bx)}) = \pm (a + bx)^{(a+bx)} \cdot (b \log (a + bx) + b),
\]
from which the partial derivatives of \( h \) satisfy
\[
\frac{1}{h} \cdot \frac{\partial h}{\partial x} = \log \frac{(x-y+z)^{x-y+z}}{(2y)^{2y(z-2y)^{z-2y}} \cdot \frac{(2-2x-z)^{2-2x-z}}{(1-x)^{1-x}(1-x-z)^{1-x-z}}},
\]
\[
\frac{1}{h} \cdot \frac{\partial h}{\partial y} = \log \frac{(x-y+z)(1-x-z)}{(2y)(2-2x-z)},
\]
\[
\frac{1}{h} \cdot \frac{\partial h}{\partial z} = 2 \log M + \log \frac{(x-y+z)(1-x-z)}{(x-y)(2-2x-z)^2},
\]
as the formula above gives
\[
\frac{1}{h} \cdot \frac{\partial h}{\partial y} = -\log (x + z - y) - 1 + \log (x-y) + 1 - 2\log 2y + 1 + 2(\log (z-2y)+1)
\]
and similarly for \( \frac{\partial h}{\partial y}, \frac{\partial h}{\partial x} \). Moreover,
\[
\frac{(z-2y)^2(x-y)}{(2y)^2(x+z-y)} = 1 + \frac{(z-2y)^2 - 4y^2(x-y) - 4zy^2}{(2y)^2(x+z-y)} = 1 + \frac{z^2x - (z^2 + 4xz)y}{(2y)^2(x+z-y)},
\]
\[
\frac{(x - y + z)(1 - x - z)}{(z - 2y)(2 - 2x - z)} = 1 + \frac{(x + 3y)(1 - x) - z(y + 1)}{(z - 2y)(2 - 2x - z)}
\]
reveal the zeros of \( \frac{\partial h}{\partial y}, \frac{\partial h}{\partial z} \), which will be subsequently employed.

Let \((x_0, y_0, z_0) \in \text{arg sup} h(x, y, z)\): then each coordinate either makes some constraining inequality equality or has its partial derivative zero. The analysis below uses this observation to find \((x_0, y_0, z_0)\) and thus \(\sup h\) (each case proves \(\sup h \leq f^2(M)\), and the third also entails \(\sup h \geq f^2(M)\)). For simplicity, drop the subscripts, and denote such a point by \((x, y, z)\).

If \(z = 0\), then \(h(x, y, z) = M^{2x+2-2x} \leq \max(4, M^2) \leq \max(4, (\frac{M^2+1}{M^2})^2 \chi_{M>1})\).

If \(z = 1 - x\), then \(h(x, y, z) = M^{2x} \frac{1}{x^2(1-x)^{1-x}}, \) whose supremum, attained at \(x = \frac{M^2}{M^2+1}\) as the derivative is \(h \cdot (2 \log M - \log \frac{x}{1-x})\), is \(\frac{1}{1-x} = M^2 + 1 \leq \max(4, (\frac{M^2+1}{M^2})^2 \chi_{M>1})\).

Else, \(\frac{\partial h}{\partial z} = 0\), from which \(z = x(1-x)\). Then \(h(x, y, z) = M^{2x} \frac{(2-x)^2+2x}{(1-x)^{1-x}}\), whose critical point is given by \(\log M^2 - \log \frac{2-x}{1-x} = 0\) or \(x = \frac{M^2-2}{M^2-1}\) for \(M \geq \sqrt{2}\), yielding \(h(x, y, z) = (\frac{(2-x)^2}{1-x}) = \frac{M^4}{M^2-1}\). The supremum in this case is \(\max(4, M^2), \frac{M^4}{M^2-1} \chi_{M>\sqrt{2}}\) \leq \max(4, (\frac{M^2+1}{M^2})^2 \chi_{M>1})\):

\[
(q + 1)q^3 - q^2 - q = q^2 + 1 > 0.
\]

Case 2: \(y = x\). Then \(2x \leq z \leq 1 - x\), and

\[
h(x, y, z) = M^{2x} \frac{z^2}{(2x)^{2x}(z - 2x)^{2x}} \cdot (1 - x)^{1-x}(1 - x - z)^{1-x-z}.
\]

If \(z = 2x\), then

\[
h(x, y, z) = M^{2x} \frac{(2 - 4x)^{2-4x}}{(1-x)^{1-x}(1-3x)^{1-3x}}.
\]

Its critical points satisfy

\[
\log M^2 + \log \frac{1-x}{(1-x)^3}{2-4x} = 0,
\]
giving \(h(x, y, z) = \frac{(2-4x)^2}{(1-x)(1-3x)} = M \sqrt{\frac{1-3x}{1-x}} \leq M < \max(4, (\frac{M^2+1}{M^2})^2 \chi_{M>1})\), while \(x \in \{0, \frac{1}{3}\}\) yields \(\max(4, M^{2/3}) \leq \max(4, M) \leq \max(4, (\frac{M^2+1}{M^2})^2 \chi_{M>1})\) as \(\frac{(q^2+1)^2}{q^2} > q^2 > q\) for \(q \geq 1\).

If \(z = 1 - x\), then

\[
h(x, y, z) = M^{2x} \frac{(1-x)^{1-x}}{(2x)^{2x}(1-3x)^{1-3x}}.
\]

At the endpoints, this function is 1, \(M^{2/3}\), and its critical points solve

\[
\log M^2 + \log \frac{(1-3x)^3}{(2x)^2(1-x)} = 0,
\]
yielding

\[
h(x, y, z) = \frac{1-x}{1-3x} = M^{2/3} \frac{(1-x)^{2/3}}{2x} = 4 + \frac{11x - 3}{1-3x} \leq
\]
\[
\leq \max(4, M^{2/3}(4/3)^{2/3}) \leq \max(4, \frac{(M^2 + 1)^2}{M^2} \chi_{M > 1})
\]

using that either \( x < \frac{3}{11} \) or \( x \geq \frac{3}{11} \), and \( M^{2/3}(4/3)^{2/3} < M^2 \) for \( M \geq 6 \).

Else, \( z = \frac{4x(1-x)}{x+1}, x \leq \frac{1}{3} \), and

\[
h(x, y, z) = M^{2x} \frac{1}{(2y)^2y(z-2y)^{-2y}} \cdot \frac{(2x-2z)^{2-2x}}{(1-x)^{1-x}(1-x-z)^{1-x}} = M^{2x} \frac{(2-2x)^{2-2x}}{(1+x)^{1+x}(1-3z)^{1-3x}}.
\]

At the endpoints, \( h \) is \( 4, M^{2/3} \), and its critical points satisfy

\[
\log M^2 + \log \frac{(1-3x)^3}{(2-2x)^2(1+x)} = 0,
\]

giving

\[
h(x, y, z) = \frac{(2-2x)^2}{(1-3x)(x+1)} = M^2 \cdot \frac{(1-3x)^2}{(1+x)^2} \leq M^2 \leq \max(4, \frac{(M^2 + 1)^2}{M^2} \chi_{M > 1}).
\]

Case 3: \( y = \frac{x^2}{x+y} \).

Since \( 0 \leq y \leq \min(\frac{2}{x}, \frac{1}{x}) \) for all \( z, x \geq 0 \) \((x, y, z) \in D \) implies \( y = 0 \) if \( z = x = 0 \), the constraints on \( z \) are \( 0 \leq z \leq 1 - x \). If \( z = 0 \), then \( z = y = 0 \), a case already considered. If \( z = 1 - x \), then \( y = \frac{x-x^2}{1+3x} \), and

\[
h(x, y, z) = M^{2x} \frac{(x-y+z)(1-x)(1-x-z)}{(x-y)(2-2x-z)^2} = M^{2x} \frac{(1+x)^2}{(4x^2)^x(1-x^2)^{1-x}} = M^{2x} \frac{(1+x)^{1+x}}{(2x)^{2x}(1-x)^{1-x}}.
\]

The critical point is given by

\[
\log M^2 + \log \frac{1-x^2}{4x^2} = 0,
\]
or \( x^2 = \frac{M^2}{M^2+4}, \) and is a global maximum with value

\[
\frac{1+x}{1-x} = \frac{(M + \sqrt{M^2 + 4})^2}{4} \leq \max(4, \frac{(M^2 + 1)^2}{M^2} \chi_{M > 1})
\]

if \( M^2 \leq 3/2 \), then \( \frac{(M + \sqrt{M^2 + 4})^2}{4} \leq 4 \); else, \( M^2 \geq 3/2 \), and \( \frac{(M+\sqrt{M^2+4})^2}{4} < \frac{(M^2+1)^2}{M^2} \) as \( M \sqrt{M^2+4} < M^2 + 2 \).

Lastly, \( z = \frac{(x+y)(1-x)}{y+1} \) and so \( y = \frac{x-x^2}{3-3x-z} \), whereby \( z^2(x+1) + z \cdot 4x^2 + 4x^3 - 4x^2 = 0 \). The roots of this quadratic equation are \(-2x, \frac{2x-2x^2}{x+1}\). Hence \( z = \frac{2x-2x^2}{x+1} \) and \( y = \frac{x-x^2}{x+3} \) \((x, y, z) \in D \) for \( 0 \leq x \leq 1 \).

If \( x \in \{0, 1\} \), then \( y = z = 0 \), a situation already analyzed. Else, \( x \) is a critical point,

\[
2 \log M + \log \frac{(x-y+z)(1-x)(1-x-z)}{(x-y)(2-2x-z)^2} = 0
\]

\[
2 \log M + \log \left[ (1+\frac{z}{x-y}) \cdot \frac{1-\frac{x}{1-x}}{(2-\frac{x}{1-x})^2} \right] = 0,
\]

with \( \frac{x}{x-y} = \frac{(x+1)(1-x)}{(1+x)^2}, \frac{z}{x-y} = \frac{2x}{x+1} \), from which

\[
2 \log M + \log \frac{4}{(x+1)^2} \cdot \frac{1-x^2}{4} = 0,
\]
or \( x = \frac{M^2 - 1}{M^2 + 1} \) for \( M > 1 \). Since all the partial derivatives are zero,

\[
    h(x, y, z) = \frac{(2 - 2x - z)^2}{(1 - x)(1 - x - z)} = \frac{4}{1 - x^2} = \frac{(M^2 + 1)^2}{M^2}.
\]

\[ \square \]

### 3. Limiting Distributions of \( |A| \) and \( \lambda_1(A) \)

This last section completes the proof of Theorem 1 by concluding (2) (subsections 3.1 and 3.2 treat (8’’) and (9’’), respectively) and its analog for \( |A| \) replaced by \( \lambda_1(A) \) (subsection 3.3).

#### 3.1. The Upper Bound.

This subsection justifies (8’’). Two tools are used towards this: the bound on the conditional expectation \(* \) given by (16), and the following linear algebra result, which provides a connection between \( ||A_n||, ||A_\delta|| \) and \( tr((A_n + A_{B,n})^{2p}) - tr(A_{B,n}^{2p}) \).

For symmetric matrices, denote by \( \lambda_i(\cdot) \) the \( i \)th largest eigenvalue for \( i \geq 1 \), and by convention, \( \lambda_i(Q) \) implicitly assumes \( i \geq 1 \) (e.g., \( \lambda_{j+1-i}(Q) \) entails \( l \leq j \)) and \( \lambda_i(Q) = 0 \) for \( i > n \) when \( Q \in \mathbb{R}^{n \times n} \).

**Lemma 7.** Suppose \( S, Q \in \mathbb{R}^{n \times n} \) are symmetric matrices with \( \lambda_1(S) \geq 0 \), \( \lambda_{2m+1}(Q) = 0 \) for some integer \( m \in \{1, \frac{n}{2} - 1\} \). Then for \( p \in \mathbb{N} \),

\[
    ||S + Q||^{2p} - 7m \cdot ||S||^{2p} \leq tr((S + Q)^{2p}) - tr(S^{2p}) \leq 4m \cdot ||S + Q||^{2p}.
\]

**Proof.** Since \( S, Q \) are symmetric, for \( \bar{\sigma}_i = \lambda_i(S + Q), \sigma_i = \lambda_i(S), 1 \leq i \leq n \),

\[
    tr((S + Q)^{2p}) - tr(S^{2p}) = \sum_{1 \leq i \leq n} \bar{\sigma}_i^{2p} - \sum_{1 \leq i \leq n} \sigma_i^{2p}.
\]

Weyl’s inequalities

\[
    \lambda_k(S) + \lambda_{n+j-k}(Q) \leq \lambda_j(S + Q) \leq \lambda_j(S) + \lambda_{j+1-i}(Q),
\]

give when \( j \geq 2m + 1 \),

\[
    \lambda_j(S) + \lambda_n(Q) = \sigma_j \leq \bar{\sigma}_j \leq \sigma_{j-2m} = \lambda_j - 2m(S) + \lambda_{2m+1}(Q),
\]

the first inequality holding for all \( 1 \leq j \leq n \). Let \( t \geq 1 \) be minimal with \( \sigma_t \geq 0 \geq \sigma_{t+1} \).

The ensuing case-by-case analysis yields the statement of the lemma (1, 2, 3 cover the lower bound, and 4, 5 the upper bound).

1. \( t \leq 2m \):

\[
    tr((S + Q)^{2p}) - tr(S^{2p}) = \sum_{i \leq n} \bar{\sigma}_i^{2p} - \sum_{i \leq n} \sigma_i^{2p} \geq
\]

\[
    \geq \bar{\sigma}_1^{2p} + \sum_{1 < i \leq t} \sigma_i^{2p} + \sum_{t+1 \leq i \leq t+2m} \bar{\sigma}_i^{2p} + \sum_{t+2m < i \leq n-2m} \sigma_{i-2m}^{2p} + \sum_{n-2m < i \leq n} \bar{\sigma}_i^{2p} - \sum_{i \leq n} \sigma_i^{2p} \geq
\]

\[
    \geq \bar{\sigma}_1^{2p} + \sum_{n-2m < i \leq n} \bar{\sigma}_i^{2p} - \sigma_1^{2p} - \sum_{n-4m < i \leq n} \sigma_i^{2p} \geq \max_{1 \leq i \leq n} \bar{\sigma}_i^{2p} - (4m + 1) \max_{1 \leq i \leq n} \sigma_i^{2p}.
\]
2. $2m < t < n - 4m$:

$$tr((S + Q)^{2p}) - tr(S^{2p}) = \sum_{i \leq n} \sigma_i^{2p} - \sum_{i \leq n} \sigma_i^{2p} \geq$$

$$\geq \sum_{i \leq 2m} \sigma_i^{2p} + \sum_{2m < i \leq t} \sigma_i^{2p} + \sum_{t < i \leq t+2m} \sigma_i^{2p} + \sum_{t+2m < i \leq n-3m} \sigma_i^{2p} + \sum_{n-2m < i \leq n} \sigma_i^{2p} \geq \max_{1 \leq i \leq n} \sigma_i^{2p} - 6m \max_{1 \leq i \leq n} \sigma_i^{2p};$$

3. $n - 4m \leq t$:

$$tr((S + Q)^{2p}) - tr(S^{2p}) = \sum_{i \leq n} \sigma_i^{2p} - \sum_{i \leq n} \sigma_i^{2p} \geq$$

$$\geq \sum_{i \leq 2m} \sigma_i^{2p} - \sum_{2m < i \leq n} \sigma_i^{2p} \geq \max_{1 \leq i \leq n} \sigma_i^{2p} - (6m + 1) \max_{1 \leq i \leq n} \sigma_i^{2p};$$

4. $t \leq 2m$:

$$tr((S + Q)^{2p}) - tr(S^{2p}) = \sum_{i \leq n} \sigma_i^{2p} - \sum_{i \leq n} \sigma_i^{2p} \leq \sum_{i \leq 2m+t} \sigma_i^{2p} - \sum_{2m+t < i \leq n} \sigma_i^{2p} \leq 4m \max_{1 \leq i \leq n} \sigma_i^{2p};$$

5. $t > 2m$:

$$tr((S + Q)^{2p}) - tr(S^{2p}) = \sum_{i \leq n} \sigma_i^{2p} - \sum_{i \leq n} \sigma_i^{2p} \leq \sum_{i \leq 2m} \sigma_i^{2p} + \sum_{2m < i \leq n} \sigma_i^{2p} + \sum_{t < i \leq \min(t + 2m, n)} \sigma_i^{2p} + \sum_{\min(t + 2m, n) < i \leq n} \sigma_i^{2p} \leq 4m \max_{1 \leq i \leq n} \sigma_i^{2p}. $$

(8") can now be concluded. Conditional on $E(S, \kappa, M)$, Lemma 7 implies for $n$ sufficiently large,

$$tr((A_s + A_{B,s})^{2p}) - tr(A_s^{2p}) \geq ||A_s||^{2p} - 7m \cdot ||A_s||^{2p},$$

and so

$$ (tr((A_s + A_{B,s})^{2p}) - tr(A_s^{2p})) \leq 7m \cdot ||A_s||^{2p}. $$

Since $P(X > a) \leq E[X^p + X_{+}^p]a^{-p}$ for any real-valued random variable $X$ and $a > 0$,

$$\mathbb{P}_s(||A_s|| > f(M) + \epsilon) \leq (f(M) + \epsilon)^{-2p}(E_s(tr((A_s + A_{B,s})^{2p}) - tr(A_s^{2p})) + 7m \cdot E_s(||A_s||^{2p})).$$

Arguing as in subsection (2.2),

$$E_s(||A_s||^{2p}) \leq c(\kappa, \epsilon)E_s(||A_s||^{2p}),$$

and Theorem 2 for $\mu = 1, \gamma' = 1, \gamma = 1 - \delta_1, \gamma'' = \delta_1, s = ||n^{\gamma''}|| \geq n^{\gamma''}/2$ yields

$$E(||A_s||^{2p}) \leq (2 + \epsilon)^{2p} + E(||A_s||^{2p}) \cdot (2 + \epsilon)^{2p} - 2s \leq (2 + \epsilon)^{2p} + (2 + \epsilon)^{2p} \cdot n^3(2 + \epsilon)^{-n^{\gamma''}}.$$
with the first term dominated the second for $n$ large enough. Consequently,
\[
\mathbb{P}_s(||A_\kappa|| > f(M) + \epsilon) \leq (f(M) + \epsilon)^{-2p}(E_s[tr((A_s + A_{B,\kappa})^{2p}) - tr(A_{s}^{2p})]) + 14m(2+\epsilon)^{2p},
\]
whereby (16) and (23) for $\kappa = \epsilon$ yields
\[
\mathbb{P}_s(||A_\kappa|| > f(M) + \epsilon) \leq 2mc(\epsilon,c)(1 + c_1(M,\epsilon))^{-2p} + n^{-\delta/2}c(p) + 14m(1 + c_3(M,\epsilon))^{-2p}
\]
for some $c_i(M,\epsilon) > 0$, and $n, p$ sufficiently large. Given the growth hierarchy $m, p, n$ form, this last inequality entails $(8^*)$.

3.2. The Lower Bound. This subsection proves $(9^*)$ by justifying for $n \geq n(\delta, \kappa, M)$,

\begin{align}
(31) \quad & E_{ss}[tr(A_{s}^{2p}) - tr(A_{s}^{2p})] \geq \frac{c_0(\kappa, c)}{p}(1 - 2mn^{-1})^p(1 - n^{-2\delta})^p s(\max A, p), \\
(32) \quad & \text{Var}_{ss}(tr(A_{s}^{2p}) - tr(A_{s}^{2p})) \leq n^{-1/2}[(\max (M, 2))^{4p-2}(2m)^{4p}(4p)^{64p^2} + s(2p - 1, M)].
\end{align}

Begin with (31). The key observation is that anew solely even cycles contribute (in particular, the considered expectation is a sum of non-negative terms):

**Lemma 8.** Suppose a cycle $i$ contains some edge belonging to $A_s$ and is not even (its length might be odd). Then there exists an undirected edge belonging to $A_s$ appearing an odd number of times in $i$.

**Proof.** Compress the clusters of edges belonging to $A_{B,\kappa}$ to points or single edges (as in Lemma 3), and note this procedure leaves the edges of $i$ belonging to $A_s$ intact and does not change the parity of the cycle; i.e., this new cycle $i_c \neq \emptyset$ is not even and shares with $i$ its edges belonging to $A_s$ (including which ones are marked). If $i_c$ has no edge belonging to $A_{B,\kappa}$, then the conclusion follows. Else, there is an undirected edge $vw$ belonging to $A_{B,\kappa}$ appearing an odd number of times in $i_c$; let the indices of these edges be $1 \leq p_1 < p_2 < \ldots < p_k = \delta'$, where $p'$ is the length of $i_c$. A case-by-case analysis and $u \neq v$ show either $u$ or $v$ is adjacent to an odd number of edges belonging to $A_s$ (let $P_{odd}$ be the set of vertices of $i_c$ having this property), whereby the claim of the lemma ensues:

1. $p_1 > 1, p_{2k+1} < p' : v, w \in P_{odd}$;
2. $p_1 = 1, p_{2k+1} < p' : i \in P_{odd}$;
3. $p_1 > 1, p_{2k+1} = p' : i_{p'-1} \in P_{odd}$;
4. $p_1 = 1, p_{2k+1} = p' : i \in P_{odd}$ (as $i_{p-1} = i_1$).

Since the left-hand side of (31) is a sum of non-negative terms, the inequality follows from the description of cycles of type (III) in subsection 2.2 (see proof of (22)) and $S \neq \emptyset$; all cycles of type (III) with $|a_{ij}| = \max_{1 \leq i \leq j \leq n} |a_{ij}|$ have expectation at least $c_0(\kappa, c)(\max A)^{2l}(n^{-1}E_{a_{11}^2a_{11}^2|a_{11}|^2})^{p-1}$, where $2l$ is the multiplicity of $uv$ in the cycle, and there are at least $(n - 2m)^{p-1}$ possibilities for choosing the remaining vertices of $i$ because restricting them to $\{1, 2, \ldots, n\} - \{t : \exists s, (\min (s, t), \max (s, t)) \in S\}$, a set of size at least $n - 2m$, ensures no edge belonging to $A_{B,\kappa}$ is created by any such assignment.

Consider now (32). Clearly,
\[
\text{Var}_{ss}[tr(A_{s}^{2p}) - tr(A_{s}^{2p})] = \sum_{i,j} (E_{ss}[a_{ij} - E_{ss}[a_{ij}] - E_{ss}[a_{ij}]E_{ss}[a_{ij}]),
\]
where \(i, j\) are cycles of length \(2p\) containing at least one edge belonging to \(A_{B, \kappa}\). Proceed in the same vein as Sinai and Soshnikov [13] did when analyzing the variance of large moments of the trace of a Wigner matrix. By independence, the contribution of \((i, j)\) is non-zero iff \(i\) and \(j\) share at least one undirected edge belonging to \(A_s\), and every undirected edge in their union \(i \cup j\) appears an even number of times (if they share no edge, then they are independent; else, if there is an edge in the union appearing an odd number of times, then both terms are zero by Lemma 8 and symmetry).

A crucial step in [13] is mapping such pairs \((i, j)\) to even cycles \(\mathcal{P}\) of length \(2 \cdot 2p - 2 = 4p - 2\). Let \(i_{t-1}i_t = j_{s-1}j_s\) with \(t, s\) minimal in this order (i.e., consider only edges belonging to \(A_s\) and let \(t = \min \{1 \leq k \leq 2p, \exists 1 \leq q \leq 2p, i_{k-1}i_k = j_{q-1}j_q\}, s = \min \{1 \leq q \leq 2p, j_{q-1}j_q = i_{t-1}i_t\}\)). Then \(\mathcal{P}\) is obtained by gluing these two cycles along this common edge, which then gets erased. Put differently, \(\mathcal{P}\) traverses \(i\) up to \(i_{t-1}i_t\), which is then used as a bridge to switch to \(j\), traverse all of it, and get back to the rest of \(i\) upon returning to \(j_{s-1}j_s = i_{t-1}i_t\). More specifically, if \((i_{t-1}, i_t) = (j_{s-1}, j_s)\), then

\[
\mathcal{P} := (i_0, \ldots, i_{t-1}, j_{s-2}, \ldots, j_0, j_{2p-1}, \ldots, j_s, i_{t+1}, \ldots, i_{2p});
\]

else, \((i_{t-1}, i_t) = (j_s, j_{s-1})\), and

\[
\mathcal{P} := (i_0, \ldots, i_{t-1}, j_{s+1}, \ldots, j_{2p-1}, j_0, \ldots, j_{s-1}, i_{t+1}, \ldots, i_{2p}).
\]

Evidently, \(\mathcal{P}\) is an even cycle of length \(2 \cdot 2p - 2 = 4p - 2\).

Since for the conditional expectation \(\mathbb{E}_{**}[\cdot]\) a similar split to the one in (19) occurs,

\[
|\mathbb{E}_{**}[a_1 \cdot a_j] - \mathbb{E}_{**}[a_1] \cdot \mathbb{E}_{**}[a_j]| \leq |\mathbb{E}_{**}[a_1 \cdot a_j]|.
\]

Hence

\[
\text{Var}_{**}[\text{tr}(A_{\kappa}^{2p}) - \text{tr}(A_s^{2p})] \leq 4p \cdot n^{-2p} \sum_{\mathcal{P}, 0 \leq r \leq 2p-1} \mathbb{E}_{**}[a_P a_P^2],
\]

where \(\mathcal{P} = (q_0, q_1, \ldots, q_{4p-3}, q_0)\) is an even cycle with at least two edges belonging to \(A_{B, \kappa}\) inasmuch as for any such \(\mathcal{P}\) and \(0 \leq r \leq 2p - 1\) there are at most \(4p\) pairs \((i, j)\) mapped to it \((i_{t-1} = q_r, i_t = q_{r+1}, i_t = q_{r+2p-1}, \text{ and it remains to choose whether } (i_{t-1}, i_t) = (j_{s-1}, j_s) \text{ or } (i_{t-1}, i_t) = (j_s, j_{s-1}), \text{ and the first vertex of } j, \text{ which can be done in at most } 2 \cdot 2p = 4p \text{ ways). Because } a_P a_P^2 \text{ belongs to } A_s, \quad \mathbb{E}_{**}[a_P a_P^2] \leq n^{2(1/4 - \delta_1)} \mathbb{E}_{**}[a_P],
\]

and reasoning as in subsection 2.2,

\[
\sum_{\mathcal{P}} \mathbb{E}_{**}[a_P] \leq c(\kappa, c) \cdot n^{2p-1} [(4m - 2) \cdot M^{4p-2} + 4mn^{-\delta_1} \sum_{1 \leq l \leq 2p-1} \left( \frac{4p - 2}{2l} \right) (2m)^{2l} \cdot 2^{4p-2-2l}((2l + 2)!)^4l \cdot M^{2l} + s(2p - 1, M)],
\]

yielding the conditional variance is at most

\[
(4p \cdot n^{-2p}) \cdot n^{2(1/4 - \delta_1)} \cdot c(\kappa, c) n^{2p-1} \cdot [(\max (M, 2))^4p-2(2m)^4p(4p)^64p^2 + s(2p - 1, M)].
\]

Now (9") can be concluded. Lemma 7 gives, conditional on \(E(S, M, \kappa)\),

\[
||A_{\kappa}||_{2p}^2 \geq \frac{1}{4m} (\text{tr}(A_{\kappa}^{2p}) - \text{tr}(A_s^{2p})).
\]
Since \(2 \leq f(\max A) \leq f(M)\), (31) yields for \(\epsilon \in (0, 1)\)
\[
\frac{1}{4m} \mathbb{E}_{**}[(tr(A_κ^{2p}) - tr(A_s^{2p})] \geq \frac{c_0(\kappa, \epsilon)}{4m} (1 - 2n^{-2\delta})^p (f(\max A) - \epsilon/2)^{2p} \geq 2(f(\max A) - \epsilon/2)^{2p}
\]
(subsection 2.3 entails \(s(p, \cdot)^{1/2p} \rightarrow f(\cdot)\) uniformly on compact subsets of \((0, \infty)\); thus, for all \(p \geq p(M, \epsilon, \kappa)\), \(s(max A, p) \geq (f(max A) - \epsilon/2)^{2p}\). Chebyshev’s inequality gives
\[
\mathbb{P}_{**}(||A_κ||^{2p} < (f(max A) - \epsilon)^{2p}) \leq \frac{Var_{**}[(tr(A_κ^{2p}) - tr(A_s^{2p})]}{16m^2(f(max A) - \epsilon)^{2p}} = o(1)
\]
using (32).

3.3. The Largest Eigenvalue. This subsection completes the proof of Theorem 1 by arguing (2) remains true when \(||A||\) is replaced by \(\lambda_1(A)\). The first inequality is immediate from (2), while for the second, in the same spirit as before, it suffices to show for \(\kappa = \delta > 0\) and \(0 < \epsilon < \frac{f(1+\delta)^{-2}}{8}\),
\[
\lim_{n \to \infty} \mathbb{P}_{**}(\lambda_1(A_κ) \leq f(\max A) - \epsilon, \max A \geq 1 + \delta) = 0
\]
(if \(\max A \leq 1\), then (3) implies the desired result). Consider the following modified version of Lemma 7:

**Lemma 9.** Suppose \(S, Q \in \mathbb{R}^{n \times n}\) are symmetric matrices with \(\lambda_{2m+1}(Q) = 0\) for some integer \(m \in [1, \frac{n}{2} - 1]\). Then for \(p \in \mathbb{N}\),
\[
tr(((S + Q)^{2p+1}) - tr(S^{2p+1}) \leq 2m \cdot (\lambda_1(S + Q))^{2p+1} + (\lambda_n(S + Q))^{2p+1} + 3m \cdot ||S||^{2p+1}.
\]

**Proof.** Keeping the notation from the proof of Lemma 7,
\[
tr((S + Q)^{2p+1}) - tr(S^{2p+1}) = \sum_{i \leq n} \bar{\sigma}_i^{2p+1} - \sum_{i \leq n} \sigma_i^{2p+1} \leq \sum_{i \leq 2m} \bar{\sigma}_i^{2p+1} + \sum_{2m < i \leq n - 1} \sigma_i^{2p+1} + \sum_{i \leq n} \sigma_i^{2p+1} - \sum_{i \leq n} \bar{\sigma}_i^{2p+1} \leq 2m \cdot \bar{\sigma}_1^{2p+1} + \bar{\sigma}_n^{2p+1} + (2m + 1) \cdot \left(\max_{1 \leq i \leq n} |\sigma_i|\right)^{2p+1}.
\]
\[\blacksquare\]

Conditional on ***, Lemma 9 gives
\[
tr(A_κ^{2p+1}) - tr(A_s^{2p+1}) \leq 2m \cdot (\lambda_1(A_κ))^{2p+1} + (\lambda_n(A_κ))^{2p+1} + 3m \cdot ||A_κ||^{2p+1};
\]
if also \(\lambda_1(A_κ) \leq f(\max A) - \epsilon < f(\max A) - \epsilon/2 \leq ||A_κ||, ||A_κ|| \leq 2 + \epsilon, \max A \geq 1 + \delta\), then
\[
tr(A_κ^{2p+1}) - tr(A_s^{2p+1}) \leq 2m \cdot (f(\max A) - \epsilon)^{2p+1} - (f(\max A) - \epsilon/2)^{2p+1} + 3m \cdot (2 + \epsilon)^{2p+1} \leq -\frac{(f(\max A) - \epsilon/2)^{2p+1}}{2}.
\]

Therefore,
\[
\mathbb{P}_{**}(\lambda_1(A_κ) \leq f(\max A) - \epsilon, \max A \geq 1 + \delta) \leq \mathbb{P}_{**}(||A_κ|| < f(\max A) - \epsilon/2) + \mathbb{P}_{**}(||A_κ|| > 2 + \epsilon) + \frac{4Var_{**}(tr(A_κ^{2p+1}) - tr(A_s^{2p+1}))}{(f(\max A) - \epsilon/2)^{4p+2}} = o(1)
\]
since Lemma 8 yields
\[ \mathbb{E}_{**}[\text{tr}(A_{n}^{2p+1}) - \text{tr}(A_{n}^{2p+1})] = 0 \]
(a cycle of odd length is not even and contains some edge belonging to \(A_{n}\)), and reasoning as for (32),
\[ \text{Var}_{**}(\text{tr}(A_{n}^{2p+1}) - \text{tr}(A_{n}^{2p+1})) \leq n^{-1/2}[(\max(M, 2))^{4p-2}(2m)^{4p}(4p)^{64p^2} + s(2p, M)]. \]
This completes the proof of Theorem 1.

Regarding the largest \(k\) eigenvalues of \(A\) for \(k \in \mathbb{N}\) fixed, a similar rationale to \(k = 1\) could be used, although the combinatorics would be more involved. Denote by \(\lambda_{i}(1_{1}^{\frac{1}{\sqrt{n}}}(\sqrt{A}))_{1 \leq i \leq n}, (\max_{j}(\sqrt{A}))_{1 \leq j \leq n}^{\frac{1}{\sqrt{n}}}\) the ordered statistics of \(|\lambda_{i}(\frac{1}{\sqrt{n}}(\sqrt{A}))_{1 \leq i \leq n}, (\frac{1}{\sqrt{n}}|a_{ij}|)_{1 \leq i, j \leq n}|\), respectively with \(\lambda(1)(\frac{1}{\sqrt{n}}(\sqrt{A})) = \frac{1}{\sqrt{n}}\|A\|, \max_{1}(A) = \max(A)\). Use induction on \(k\) to show
\[ \lambda_{k}(\frac{1}{\sqrt{n}}A) - f(\max_{k}(A)) \xrightarrow{p} 0, \]
which in conjunction with symmetry would imply
\[ \lambda_{n+1-k}(\frac{1}{\sqrt{n}}A) + f(\max_{k}(A)) \xrightarrow{p} 0, \]
whereby
\[ \lambda(1)(\frac{1}{\sqrt{n}}A) - f(\max_{k}(A)) \xrightarrow{p} 0. \]
The base case is Theorem 1; suppose the result holds for \(k \geq 1\), and consider next \(k + 1\). Similarly to the case \(k = 1\), prove first
\[ \lambda(1)(\frac{1}{\sqrt{n}}A) - f(\max_{k}(A)) \xrightarrow{p} 0, \]
and second justify this holds also for \(\lambda_{k+1}(\frac{1}{\sqrt{n}}A)\). Since the behavior of the largest \(k\) eigenvalues is known, consider
\[
\sum_{i_{1} < i_{2} < \ldots < i_{k+1}} \lambda_{i_{1}}^{2p}(\frac{1}{\sqrt{n}}A) \cdot \lambda_{i_{2}}^{2p}(\frac{1}{\sqrt{n}}A) \cdot \ldots \cdot \lambda_{i_{k+1}}^{2p}(\frac{1}{\sqrt{n}}A)
\]
for \(p \in \mathbb{N}\). The dominant term is
\[ \lambda_{(1)}^{2p}(\frac{1}{\sqrt{n}}A) \cdot \lambda_{(2)}^{2p}(\frac{1}{\sqrt{n}}A) \cdot \ldots \cdot \lambda_{(k+1)}^{2p}(\frac{1}{\sqrt{n}}A), \]
for which the induction hypothesis gives it is roughly
\[ f^{2p}(\max_{(1)}(A)) \cdot f^{2p}(\max_{(2)}(A)) \cdot \ldots \cdot f^{2p}(\max_{(k)}(A)) \cdot \lambda_{(k+1)}^{2p}(\frac{1}{\sqrt{n}}A); \]
(33) could be expressed using
\[ \text{tr}(\frac{1}{\sqrt{n}}A)^{2q} = \lambda_{1}^{2q}(\frac{1}{\sqrt{n}}A) + \lambda_{2}^{2q}(\frac{1}{\sqrt{n}}A) + \ldots + \lambda_{n}^{2q}(\frac{1}{\sqrt{n}}A) \]
for \(q \in \{p, 2p, \ldots, (k + 1)p\}\), by employing trace difference \(\text{tr}(A_{n}^{2q}) - \text{tr}(A_{n}^{2q})\) instead (to ensure they capture just the edge eigenvalues of \(A\), a rationale as in section 3 and the analysis from section 2 would give (33) is in (conditional) expectation
\[ f^{2p}(\max_{(1)}(A)) \cdot f^{2p}(\max_{(2)}(A)) \cdot \ldots \cdot f^{2p}(\max_{(k+1)}(A))(1 + o(1)) \]
and its variance small, whereby the desired result for $\lambda_{(k+1)}(\frac{1}{\sqrt{n}}A)$ ensues; finally,

$$\sum_{\{i_1,i_2,\ldots,i_{k+1}\}} \lambda_{i_1}^{2p}(\frac{1}{\sqrt{n}}A) \cdot \lambda_{i_2}^{2p}(\frac{1}{\sqrt{n}}A) \cdot \cdots \cdot \lambda_{i_{k+1}}^{2p}(\frac{1}{\sqrt{n}}A),$$

where the $k+1$ indices are pairwise distinct (i.e., they are the elements of a set), would yield $\lambda_k(\frac{1}{\sqrt{n}}A) < \lambda_{(k)}(\frac{1}{\sqrt{n}}A) - \epsilon$ occurs with small probability (else, (34) would be negative because $\lambda_{n-k}(\frac{1}{\sqrt{n}}A) = -\lambda_{(k+1)}(\frac{1}{\sqrt{n}}A)$ given the induction hypothesis, which should occur with small probability since its expectation would be 0 and its variance small).

Acknowledgments. The author would like to thank professors George Papanicolaou and Lenya Ryzhik for their feedback on the expository aspects of this paper.

REFERENCES

[1] A. AUFFINGER, G. BEN-AROUS, AND S. PÉCHÉ, Poisson convergence for the largest eigenvalues of heavy tailed random matrices, Ann. Inst. H. Poincaré Probab. Statist., Vol. 45, No. 3, 589 − 610, 2009.
[2] Z. D. BAI, AND J. SILVERSTEIN, Spectral Analysis of Large Dimensional Random Matrices, Springer Series in Mathematics, Second Edition, 2010.
[3] Z. D. BAI, J. SILVERSTEIN, AND Y. Q. YIN, A note on the largest eigenvalue of a large dimensional sample covariance matrix, J. Multivariate Anal., 26, 166 − 168, 1988.
[4] Z. D. BAI, AND Y. Q. YIN, Limit of the Smallest Eigenvalue of a Large Dimensional Sample Covariance Matrix, Ann. of Probab., Vol. 21, No. 3, 1275 − 1294, 1993.
[5] Z. D. BAI, AND Y. Q. YIN, Necessary and Sufficient Conditions for Almost Sure Convergence of the Largest Eigenvalue of a Wigner Matrix, Ann. of Probab., Vol. 16, No. 4, 1729 − 1741, 1988.
[6] F. BENAYCH-GEORGES, AND S. PÉCHÉ, Localization and Delocalization for Band Matrices, Ann. Inst. H. Poincaré Probab. Statist., Vol. 50, No. 4. 1385 − 1403, 2014.
[7] S. DIACONU, Finite Rank Perturbations of Heavy-Tailed Random Matrices, arXiv:2208.02756, 2022.
[8] S. DIACONU, Two CLTs for Sparse Random Matrices, in preparation, 2022.
[9] O. KHORUNZHY, High Moments of Large Wigner Random Matrices and Asymptotic Properties of the Spectral Norm, Random Operators and Stochastic Equations, 20, 25 − 68, 2012.
[10] M. R. LEADBETTER, G. LINDGREN, AND H. ROOTZÉN, Extremes and Related Properties of Random Sequences and Processes, Springer-Verlag, New York, 1983.
[11] J. O. LEE, AND J. YIN, A Necessary and Sufficient Condition for Edge Universality of Wigner Matrices, 2012, Duke Math. J., Vol. 163, No. 1. 117 − 173, 2014.
[12] A. RUSMAIKINA, Universality of the Edge Distribution of Eigenvalues of Wigner Random Matrices with Polynomially Decaying Distributions of Entries, Commun. Math. Phys., Vol. 261, Issue 2, 277 − 296, 2006.
[13] YA. SINAI, AND A. SOSNIKOV, Central Limit Theorem for Traces of Large Random Symmetric Matrices With Independent Matrix Elements, Bol. Soc. Brasil. Mat., Vol. 29, No. 1. 11 − 24, 1998.
[14] YA. SINAI, AND A. SOSNIKOV, A Refinement of Wigner’s Semicircle Law in a Neighborhood of the Spectrum Edge for Random Symmetric Matrices, Functional Analysis and Its Applications, Vol. 32, No. 2. 1998.
[15] A. SOSNIKOV, Universality at the edge of the spectrum in Wigner random matrices, Comm. Math. Phys., Vol. 207, No. 3, 697 − 733, 1999.
[16] A. SOSNIKOV, Poisson Statistics for the Largest Eigenvalue of Wigner Random Matrices with Heavy Tails, Comm. in Probab., 9, 82 − 91, 2004.
[17] T. TAO, AND V. VU, Random matrices: Universality of local eigenvalue statistics, Acta Math., Vol. 206, No. 1, 127 − 204, 2011.
[18] C. A. TRACY AND H. WIDOM, Level spacing distributions and the Airy kernel, Comm. Math. Phys., Vol. 159, No. 1, 151 − 174, 1994.
[19] E. P. WIGNER, On the Distribution of the Roots of Certain Symmetric Matrices, Annals of Mathematics, Second Series, Vol. 67, No. 2. 325 − 327, 1958.