Circular law for the sum of random permutation matrices

Anirban Basak[*]   Nicholas Cook[†]   Ofer Zeitouni[‡]

Abstract

Let $P_n^1, \ldots, P_n^d$ be $n \times n$ permutation matrices drawn independently and uniformly at random, and set $S_n^d := \sum_{\ell=1}^d P_n^\ell$. We show that if $\log^{12} n / (\log \log n)^4 \leq d = O(n)$, then the empirical spectral distribution of $S_n^d / \sqrt{d}$ converges weakly to the circular law in probability as $n \to \infty$.

Keywords: random permutation; logarithmic potential; singular values; Stieltjes transform; local laws.
AMS MSC 2010: Primary 15B52, Secondary 15A18; 60B10; 60B20.
Submitted to EJP on May 29, 2017, final version accepted on March 27, 2018.

1 Introduction

For an $n \times n$ matrix M_n let $\lambda_1(M_n), \lambda_2(M_n), \ldots, \lambda_n(M_n)$ be its eigenvalues. We define the *empirical spectral distribution* (ESD) of M_n as follows:

$$L_{M_n} := \frac{1}{n} \sum_{i=1}^n \delta_{\lambda_i(M_n)}.$$

For a sequence of random probability measures $\{\mu_n\}_{n \in \mathbb{N}}$, supported on the complex plane, we say that μ_n converges weakly to a limiting probability measure μ, in probability, if for every bounded continuous function $f : \mathbb{C} \mapsto \mathbb{R}$,

$$\int f d\mu_n - \int f d\mu \to 0 \qquad \text{as } n \to \infty, \tag{1.1}$$

in probability. If (1.1) holds almost surely we say that μ_n converges weakly to μ, almost surely.

[*]Department of Mathematics, Weizmann Institute of Science, Israel. E-mail: anirban.basak@weizmann.ac.il. Research partially supported by grant 147/15 from the Israel Science Foundation.
[†]Department of Mathematics, University of California Los Angeles, USA. E-mail: nickcook@math.ucla.edu. Research partially supported by NSF postdoctoral fellowship DMS-1606310.
[‡]Department of Mathematics, Weizmann Institute of Science, Israel. E-mail: ofer.zeitouni@weizmann.ac.il. Research partially supported by grant 147/15 from the Israel Science Foundation.

We are concerned in this paper with the ESD of certain random, non-normal matrices, defined as follows. For a positive integer n, let π_n^i, $i = 1, 2, \ldots$ denote i.i.d. permutations, distributed uniformly on the symmetric group $\mathbb{S}_n$. Let P_n^i denote the associated permutation matrices, i.e., $P_n^\ell(i, j) := \mathbb{I}(\pi_n^\ell(i) = j)$ for $\ell \in [d]$, $i, j \in [n]$ where for any integer m we denote $[m] := \{1, 2, \ldots, m\}$. For d an integer, define S_n^d as

$$S_n^d(i, j) := \sum_{\ell=1}^d P_n^\ell(i, j) = \sum_{\ell=1}^d \mathbb{I}(\pi_n^\ell(i) = j). \tag{1.2}$$

Note that S_n^d can be viewed as the adjacency matrix of a d-regular directed multigraph.

For two sequences of positive reals $\{a_n\}$ and $\{b_n\}$ we say that $a_n = O(b_n)$ (or $a_n = o(b_n)$) if for some universal constant C, $\limsup_{n \to \infty} a_n / b_n \leq C$ (respectively, $= 0$). We say that $a_n = \omega(b_n)$ if $b_n = o(a_n)$. The main result of this paper is the following theorem.

Theorem 1.1. If $\log^{12} n / (\log \log n)^4 \leq d = O(n)$ then the ESD of $S_n^d / \sqrt{d}$ converges weakly to the uniform distribution on the unit disk in the complex plane, in probability, as $n \to \infty$.

We refer to this result as the weak circular law for sums of permutations.

Remark 1.2. One expects the conclusion of Theorem 1.1 to hold almost surely. However, the estimate on the smallest singular value of $S_n^d / \sqrt{d} - zI$ contained in Theorem 2.1 below is not sharp enough to allow for the application of the Borel–Cantelli lemma. On the other hand, other estimates in the paper, and in particular the concentration inequalities and the estimates on moderately small singular values, see Section 2 for definitions, are not an obstacle to the application of Borel–Cantelli.

Remark 1.3. Theorem 1.1 is established for $d \geq \log^{12} n / (\log \log n)^4$. One expects its conclusion to hold as soon as $d = \omega(1)$. Obvious obstacles to proving this by our methods are that the minimal singular value estimate, Theorem 2.1 below, requires $d = \omega(\log^8 n)$ to be useful, and our loop equations main theorem, Theorem 2.6, is only effective when d grows like a power of $\log n$. Proving Theorem 1.1 for $d = \omega(1)$ remains a major challenge and seems to require new ideas. It is possible that one could use the methods of [29] to relax the assumptions in Theorem 2.1 to $d = \omega(1)$.

1.1 Background: ESD's for non-normal matrices

The study of the ESD for random Hermitian matrices can be traced back to Wigner [42, 43] who showed that the ESD's of $n \times n$ Hermitian matrices with i.i.d. centered entries of variance $1/n$ (upper diagonal) satisfying appropriate moment bounds (e.g., Gaussian) converge to the semicircle distribution. The conditions on finiteness of moments were removed in subsequent work, see e.g. [5, 34] and the references therein. We refer to the texts [30, 21, 39, 3, 5] for further background and a historical perspective.

Wigner's proof employed the method of moments: one notes that the moments of the semicircle law determine it, and then one computes by combinatorial means the expectation (and variance) of the trace of powers of the matrix. This method (as well as related methods based on evaluating the Stieltjes transform of the ESD) fails for non-normal matrices since moments do not determine the ESD.

An analogue of Wigner's semicircle law in the non-normal regime is the following *circular law* theorem:

Circular law. Let M_n be an $n \times n$ matrix with i.i.d. entries of zero mean and unit variance. Then the ESD of $M_n / \sqrt{n}$ converges to the uniform distribution on the unit disk on the complex plane.

The circular law was posed as a conjecture based on numerical evidence in the 1950's. For the case that the entries have a complex Gaussian distribution it can be derived from Ginibre's explicit formula for the joint density function of the eigenvalues [23, 30]. The case of real Gaussian entries, where a similar formula is available, was settled by Edelman [18]. For the general case when there is no such formula, the problem remained open for a very long time. An approach to the problem, which eventually played an important role in the resolution of the conjecture, was suggested by Girko in the 1980's [24], but mathematically it contained significant gaps. The first non-Gaussian case (assuming existence of density for the entries) was rigorously treated by Bai [4], and after a series of partial results (see [12] and the references therein), the circular law conjecture was established in its full generality in the seminal work of Tao and Vu [41].

Theorem 1.4 (Circular law for i.i.d. entries [41, Theorem 1.10]). Let M_n be an $n \times n$ random matrix whose entries are i.i.d. copies of a fixed (independent of n) complex random variable x with zero mean and unit variance. Then the ESD of $\frac{1}{\sqrt{n}} M_n$ converges weakly to the uniform distribution on the unit disk on the complex plane, both in probability and in the almost sure sense.

A remarkable feature of Theorem 1.4 is its *universality*: the asymptotic behavior of the ESD is insensitive to the specific details of the entry distributions as long as they are i.i.d. and have zero mean and unit variance. It also extends to the sparse set-up. Namely consider a matrix of i.i.d. entries where each entry is the product of a zero mean and unit variance random variable, and an independent Bernoulli(p) random variable. From the two concurrent works of Götze and Tikhomirov [25] and Tao and Vu [40] it follows that if p decays polynomially in n, i.e. $p \geq n^{\varepsilon-1}$ for some $\varepsilon > 0$, then the limit is still the circular law. Later Wood [44] relaxed the moment assumptions of the entries. A recent article by Basak and Rudelson [7] shows that the same limit continues to hold when p decays at a poly-logarithmic rate. In all these works the entries of the matrix still enjoys the independence and this feature plays a key role in the proofs. In [11], Bordenave, Caputo and Chafaï studied random Markov generators where one puts i.i.d. entries on the off-diagonal positions and sets each diagonal to be the negative of the corresponding row-sum, showing that the limit law is a *free additive convolution* of the circular law and a Gaussian random variable. Their result covers sparse ensembles, including the Markov generator for a directed Erdős–Rényi graph with edge probability $p(n) = \omega(n^{-1} \log^6 n)$.

Circular laws for matrices with less independence between entries were subsequently proved in [10], [1], [33], [2], and [32]. In particular, in [32] Nguyen showed that the ESD of a uniformly chosen random *doubly stochastic matrices* converges weakly to the circular law. Since the adjacency matrix of a random d-regular directed graph (digraph) is a random doubly stochastic matrix, one is naturally led to the question of establishing the limits of the ESD for such matrices. This was addressed in recent work of the second author [16], where it was shown that the circular law holds for adjacency matrices of random regular digraphs assuming a poly-log(n) lower bound on the degree.

A completely different story emerges when one replaces the Ginibre matrices by other models whose distribution is invariant under the action of some large group (note that Ginibre matrices are indeed invariant under right or left multipliction by unitary matrices). The study of such invariant models was initiated by Feinberg and Zee [20], who evaluated non-rigorously the limit of the ESD for such matrices and showed various properties of the limit, e.g. that it is supported on a single ring in the complex plane. By using a variant of Girko's method adapted to the unitary group, this was put on a rigorous basis by Guionnet, Krishnapur and Zeitouni [26], who evaluated the limit of the ESD for a matrix of the form UD where D is diagonal satisfying some assumptions and U is a random Haar-distributed unitary, and showed that it coincides with the Brown

measure of the associated limiting operators (an improved version appears in [37]). Building on this and closer to the topic of this paper, Basak and Dembo [6] showed that the ESDs of the sum $\hat{U}_n^d$ of d i.i.d. Haar distributed Unitary/Orthogonal matrices converge to a probability measure μ_d whose density with respect to Lebesgue measure is given by

$$f_d(z) := \frac{1}{\pi} \frac{d^2(d-1)}{(d^2 - |z|^2)^2} \mathbb{I}(|z| \le \sqrt{d}), \tag{1.3}$$

which coincides with the Brown measure of a sum of d free Haar unitaries. Note that from this one easily concludes the existence of a sequence $d = d(n)$ so that the ESD of $\hat{U}_n^{d(n)}/\sqrt{d(n)}$ converges to the circular law.

We finally get to our model: it sits at the intersection of sparse models of regular directed (multi-)graphs and the sum of unitaries treated in [6]. Indeed, from the point of view of the latter we replace unitary matrices which are Haar-distributed on the full unitary group by unitaries which are Haar-distributed on the subgroup of permutation matrices. In this case a formal application of Girko's method leads one to expect convergence to μ_d (if d is fixed, see e.g. [12]) or to the circular law when $d = \omega(1)$ (after rescaling by $\sqrt{d}$). The goal of this paper is to establish that the latter indeed holds, at least when d does not grow too slowly or too rapidly.

Remark 1.5. Our methods are not sharp enough to handle the case of d constant, both for the reasons mentioned in Remark 1.3 and the fact that the loop equations for fixed d are much more complicated. See however the recent work [8] for progress in this direction for random d-regular graphs of sufficiently large fixed degree.

We end this section by pointing out that for *fixed* d, the random regular digraph model considered in [16] is contiguous with the sum of permutations model conditioned to have no parallel edges (i.e. with the matrix conditioned to have no entries larger than 1, an event which occurs with positive probability) [31, 27]. However, we are unaware of any quantitative contiguity results that allow d to grow with n. Given such a result (allowing d to grow faster than $\log^{12} n$) it could be possible to deduce the main result of [16] from Theorem 1.1, for some range of d; however, this would require a quantitative version of Theorem 1.1 with failure probability smaller than the probability for the sum of permutations to yield a 0/1 matrix, which is of order $\exp(-cd^2)$.

1.2 Outline of the paper

In Section 2 we provide a brief outline of the proof techniques of Theorem 1.1. We begin Section 2 by a short description of Girko's method, which in a nutshell consists of focusing attention on the *logarithmic potential* of the ESD of $S_n^d/\sqrt{d}$. This is done by analyzing the Hermitian matrix $T_n(z) = (z - S_n^d/\sqrt{d})^*(z - S_n^d/\sqrt{d})$ with $z \in \mathbb{C}$ (hereafter, for any $n \times n$ matrix B_n and $z \in \mathbb{C}$, for brevity, we often write $z - B_n$ to denote $zI_n - B_n$). To implement Girko's method one requires good control on the smallest singular value of $T_n(z)$ as well as on its *smallish* singular values. The required control on the smallest singular value is derived in Theorem 2.2 and an outline of its proof can be found in Section 2.2. The desired control on the smallish singular values is obtained in Theorem 2.6 by controlling the difference of the *Stieltjes transform* of the ESD of $T_n^{1/2}(z)$ at the finite n level and at the putative limit, all the way up to (almost) the real line. An outline of the proof of Theorem 2.6 is given in Section 2.3.

For Theorem 2.2, to control the smallest singular value of a matrix A_n we need to control the infimum of $\|A_n u\|_2$ over all u in the unit sphere. To this end, we break the sphere into the set of "flat" vectors and its complement, where a vector is said to be flat if it is close in ℓ_2 norm to a vector with a large number of equal components (for a

precise formulation see Definition 2.4). The infimum over flat vectors is taken care of in Section 3 and the infimum over the remaining vectors is treated in Section 4.

Section 5 and Section 6 are devoted to control certain traces of polynomials in S_n^d and to derive concentration inequalities for Lipschitz functions of sum of permutations, respectively. We then turn to the control on the Stieltjes transform of the ESD of $T_n^{1/2}(z)$. In Section 7.1 we show that the Stieltjes transform satisfies an (approximate) fixed point equation, first in expectation and then, using the concentration results of Section 6, also with high probability. In Section 7.2 we then finish the proof of Theorem 2.6 using the stability of the fixed point equation, apriori lower bound on Stieltjes transform of the ESD of $T_n^{1/2}(z)$ far away from the real line, and a bootstrap argument.

Finally in Section 8 combining Theorem 2.2, Theorem 2.6, and using a *replacement principle* (see Lemma 8.1) we finish the proof of Theorem 1.1.

1.3 Notational conventions

We write $\mathbb{C}^J$ for the subspace of vectors in $\mathbb{C}^n$ supported on $J \subset [n]$, and write $\mathbb{B}^J, \mathbb{S}^J$ for the closed Euclidean unit ball and sphere in this subspace. If $J = [n]$, we write $\mathbb{B}^n, \mathbb{S}^{n-1}$ for brevity. Given $v \in \mathbb{C}^n$ and $J \subset [n]$, v_J denotes the projection of v to $\mathbb{C}^J$. $\mathbf{1} = \mathbf{1}_n$ denotes the n-dimensional vector with all components equal to one, and consequently $\mathbf{1}_J$ denotes the vector with jth component equal to 1 for $j \in J$ and 0 otherwise. For $x, y \in \mathbb{R}$ we sometimes write $x \wedge y$ to mean $\min(x, y)$.

2 Preliminaries and proof outline

2.1 Proof overview

In this section we provide an outline of the proof of Theorem 1.1. As we go along we introduce necessary definitions and notation.

The standard technique to analyze the asymptotics of the ESD of a non-normal matrix is Girko's method [24]. The basis of this method is the following identity which is a consequence of Green's theorem: for any polynomial $P(z) = \prod_{i=1}^n (z - \lambda_i)$ and any test function $\psi \in C_c^2(\mathbb{C})$,

$$\sum_{j=1}^n \psi(\lambda_j) = \frac{1}{2\pi} \int_{\mathbb{C}} \Delta\psi(z) \log|P(z)| d\mathfrak{m}(z),$$

where $\mathfrak{m}$ is the Lebesgue measure on $\mathbb{C}$ and Δ denotes the two-dimensional Laplacian. Applying this identity with the characteristic polynomial $P(\cdot)$ of a matrix M_n yields

$$\int_{\mathbb{C}} \psi(z) dL_{M_n}(z) = \frac{1}{2\pi n} \int_{\mathbb{C}} \Delta\psi(z) \log|\det(zI_n - M_n)| d\mathfrak{m}(z) \tag{2.1}$$

$$= \frac{1}{4\pi n} \int_{\mathbb{C}} \Delta\psi(z) \log\det[(zI_n - M_n)(zI_n - M_n)^*] d\mathfrak{m}(z).$$

Next, associate with any n-dimensional non-Hermitian matrix M_n and every $z \in \mathbb{C}$ the $2n$-dimensional Hermitian matrix

$$\boldsymbol{M}_n^z := \begin{bmatrix} 0 & (zI_n - M_n) \\ (zI_n - M_n)^* & 0 \end{bmatrix}. \tag{2.2}$$

The eigenvalues of $\boldsymbol{M}_n^z$ are merely ± 1 times the singular values of $zI_n - M_n$. Therefore, denoting by ν_n^z the ESD of $\boldsymbol{M}_n^z$, we have that

$$\frac{1}{n} \log\det[(zI_n - M_n)(zI_n - M_n)^*] = \frac{1}{n} \log|\det \boldsymbol{M}_n^z| = 2\langle \text{Log}, \nu_n^z \rangle,$$

where for any probability measure μ on $\mathbb{R}$, $\langle \text{Log}, \mu \rangle := \int_{\mathbb{R}} \log|x| d\mu(x)$. Therefore we have the following key identity

$$\int_{\mathbb{C}} \psi(z) dL_{M_n}(z) = \frac{1}{2\pi} \int_{\mathbb{C}} \Delta\psi(z) \langle \text{Log}, \nu_n^z \rangle d\mathfrak{m}(z). \tag{2.3}$$

The utility of Eqn. (2.3) lies in the following general recipe for proving convergence of L_{M_n} of a given family of non-Hermitian random matrices $\{M_n\}$:

Step 1: Show that for (Lebesgue almost) every $z \in \mathbb{C}$, as $n \to \infty$, the measures ν_n^z converge weakly, in probability, to some measure ν^z.

Step 2: Justify that $\langle \text{Log}, \nu_n^z \rangle \to \langle \text{Log}, \nu^z \rangle$ in probability.

Step 3: A uniform integrability argument allows one to convert the z-a.e. convergence of $\langle \text{Log}, \nu_n^z \rangle$ to the convergence of $\int_{\mathbb{C}} \Delta\psi(z) \langle \text{Log}, \nu_n^z \rangle d\mathfrak{m}(z)$, for a suitable collection $\mathcal{S} \subseteq C_c^2(\mathbb{C})$ of (smooth) test functions ψ. Consequently, it then follows from (2.3) that for each fixed, non-random $\psi \in \mathcal{S}$,

$$\int_{\mathbb{C}} \psi(z) dL_{M_n}(z) \to \frac{1}{2\pi} \int_{\mathbb{C}} \Delta\psi(z) \langle \text{Log}, \nu^z \rangle d\mathfrak{m}(z), \tag{2.4}$$

in probability.

Step 4: Upon checking that $f(z) := \langle \text{Log}, \nu^z \rangle$ is smooth enough to justify the integration by parts, one has that for each fixed, non-random $\psi \in \mathcal{S}$,

$$\int_{\mathbb{C}} \psi(z) dL_{M_n}(z) \to \frac{1}{2\pi} \int_{\mathbb{C}} \psi(z) \Delta f(z) d\mathfrak{m}(z), \tag{2.5}$$

in probability. For $\mathcal{S}$ large enough, this implies the weak convergence of the ESDs L_{M_n} to a limit which has the density $\frac{1}{2\pi} \Delta f$ with respect to Lebesgue measure on $\mathbb{C}$, in probability.

To prove Theorem 1.1 our plan is to establish **Steps 1–4** for $M_n = S_n^d/\sqrt{d}$. As has been the case for other models of random matrices, **Step 2** is the most challenging part. Since ν_z is the ESD of a Hermitian matrix one can use tools such as the method of moments or the Stieltjes transform to deduce **Step 1**. However $\log(\cdot)$ being unbounded both near zero and infinity the conclusion of **Step 1** is not enough to establish **Step 2**. One needs additional control on the large as well as small singular values of $S_n^d/\sqrt{d} - z$. To this end, we first note that the limit of the ESD of $S_n^d/\sqrt{d}$, the circular law, is compactly supported. Therefore one can actually check that establishing **Steps 1–4** for z in a large ball in the complex plane is enough to complete the proof of Theorem 1.1.

Next note that each row-sum and column-sum of S_n^d is d and hence the maximal singular value of $S_n^d/\sqrt{d} - z$ is $O(\sqrt{d})$ for any z in a large ball. One can also easily show that the trace of $S_n^d(S_n^d)^*/nd$ is bounded with high probability (see Section 5), which can be used to show that ν_n^z integrates x^2, and hence $\log(x)$, near infinity.

Most of this paper is devoted to obtaining bounds on the small singular values of $S_n^d/\sqrt{d} - z$. First, one needs to have a lower bound on the smallest singular value. This is derived in Theorem 2.1. The idea behind the proof of Theorem 2.1 is outlined in Section 2.2.

Next we need to show that there are not too many singular values near zero. Equivalently, we need to show that the total mass of a small interval I around zero under the ESD of M_n^z is not too large. That mass can be estimated by obtaining bounds on the Stieltjes transform of the ESD at a distance from the real line which is comeasurate with the length of I (for example, see Lemma 8.3). In Section 2.3 we provide an outline on how to achieve the desired bounds on the Stieltjes transform of M_n^z (see Theorem 2.6).

2.2 Control on the smallest singular value

The following result provides the required lower bound on the smallest singular value of $\frac{1}{\sqrt{d}}S_n^d - z$.

Theorem 2.1. Fix any $R > 0$ and let $z \in B_{\mathbb{C}}(0, R) := \{z' \in \mathbb{C} : |z'| \leq R\}$. Assume $1 \leq d \leq n^{100}$. There exists $C_{2.1} < \infty$ depending only on R and an absolute constant $\overline{C}_{2.1} > 0$ such that

$$\mathbb{P}\left\{s_n\left(\frac{1}{\sqrt{d}}S_n^d - zI_n\right) \leq n^{-\overline{C}_{2.1}\log_{d\wedge n} n}\right\} \leq C_{2.1}\frac{\log^4 n}{\sqrt{d \wedge n}}, \tag{2.6}$$

where $s_n(\cdot)$ denotes the smallest singular value.

We deduce Theorem 2.1 from the following more general result. First we introduce some notation. For an $n \times n$ matrix M_n we write

$$\|M_n\|_{\langle \mathbf{1}\rangle^{\perp}} := \sup_{u \in \mathbb{S}^{n-1}\cap\langle\mathbf{1}\rangle^{\perp}} \|M_n u\|_2, \tag{2.7}$$

where we recall $\mathbb{S}^{n-1} := \{u \in \mathbb{C}^n : \|u\|_2 = 1\}$ and $\|\cdot\|_2$ denotes ℓ_2 norm.

Theorem 2.2. Fix an arbitrary $\gamma_0 \geq 1$. Let $1 \leq d \leq n^{\gamma_0}$, and let Z_n be a deterministic $n \times n$ matrix such that $\|Z_n\|_{\langle\mathbf{1}\rangle^{\perp}} \leq n^{\gamma_0}$ and $Z_n\mathbf{1} = \zeta\mathbf{1}$, $Z_n^*\mathbf{1} = \overline{\zeta}\mathbf{1}$ for some $\zeta \in \mathbb{C}$. There exists $C_{2.2} < \infty$ depending only on γ_0 and an absolute constant $\overline{C}_{2.2} < \infty$ such that

$$\mathbb{P}\left\{s_n(S_n^d + Z_n) < n^{-\overline{C}_{2.2}\gamma_0\log_{d\wedge n} n} \wedge |d + \zeta|\right\} \leq C_{2.2}\frac{\log^4 n}{\sqrt{d \wedge n}}. \tag{2.8}$$

By taking $Z_n = \sqrt{d}zI_n$ we immediately deduce Theorem 2.1 from Theorem 2.2.

Remark 2.3. In the proof of Theorem 2.2 it will be convenient to assume $d \leq n$. We now show how to reduce to this case (in fact we could reduce assuming $d \leq c_0 n$ for any fixed constant $c_0 > 0$). Suppose $d > n$, and let

$$Z_n' = Z_n + S_n^d - S_n^n.$$

Condition on $\pi_n^{n+1}, \ldots, \pi_n^d$ to fix Z_n'. Then we have

- $Z_n'\mathbf{1} = (\zeta + d - n)\mathbf{1} =: \zeta'\mathbf{1}$,
- $(Z_n')^*\mathbf{1} = \overline{\zeta}'\mathbf{1}$,
- $|n + \zeta'| = |d + \zeta|$,
- $\|Z_n'\|_{\langle\mathbf{1}\rangle^{\perp}} \leq \|Z_n\|_{\langle\mathbf{1}\rangle^{\perp}} + \|S_n^d - S_n^n\| \leq n^{\gamma_0} + d \leq 2n^{\gamma_0}$.

Thus, after modifying γ_0 slightly, we see that it is enough to prove Theorem 2.2 under the additional assumption that $d \leq n$.

On a high level, the proof of Theorem 2.2 follows the general strategy of the recent work [16] of the second author, which establishes a similar result with S_n^d replaced by a uniform random 0–1 matrix constrained to have all row and column sums equal to d. We now motivate some of the main ideas of this strategy.

From the definition of the smallest singular value we have

$$s_n(S_n^d + Z_n) = \inf_{u \in \mathbb{S}^{n-1}} \left\|(S_n^d + Z_n)u\right\|_2. \tag{2.9}$$

We note that $\mathbf{1}$ is an eigenvector of $(S_n^d + Z_n)^*(S_n^d + Z_n)$ with eigenvalue $|d + \zeta|^2$. A short argument then shows that to obtain (2.8) it suffices to control the infimum of $\|(S_n^d + Z_n)u\|_2$ for $u \in \mathbb{S}^{n-1} \cap \langle \mathbf{1} \rangle^\perp$. Denoting the rows of $S_n + Z_n$ by $R_1, \ldots, R_n$, we have

$$(S_n^d + Z_n)u = (R_1 \cdot u, \ldots, R_n \cdot u).$$

Thus, for a fixed vector $u \in \mathbb{S}^{n-1} \cap \langle \mathbf{1} \rangle^\perp$, the task of controlling the probability that $(S_n^d + Z_n)u$ concentrates near the origin will involve bounding the probability that the scalar random variables $R_i \cdot u$ concentrate near zero.

First we briefly review the argument from [36] for the case where S_n^d is replaced by a matrix X_n with i.i.d. centered entries ξ_{ij} of unit variance. In this case we have

$$R_i \cdot u = w + \sum_{j=1}^n \xi_{ij} u_j, \tag{2.10}$$

$w \in \mathbb{C}$ is a deterministic quantity involving the entries of u and Z_n. Then we can bound $\mathbb{P}(|R_i \cdot u| \le t)$ for small $t > 0$ using standard anti-concentration estimates. For instance, we have the following Berry–Esséen-type bound (see Lemma 4.8): for fixed nonzero $v \in \mathbb{C}^n$ and any $r \ge 0$,

$$\sup_{z \in \mathbb{C}} \mathbb{P}\left(\left| z + \sum_{j=1}^n \xi_j v_j \right| \le r \right) = O\left(\frac{r + \|v\|_\infty}{\|v\|_2} \right). \tag{2.11}$$

For this bound to be effective when applied to u, we need u to be "spread" in the sense that there is a set $J \subset [n]$ with $|J| \ge cn$ such that $|u_j| \sim 1/\sqrt{n}$ for all $j \in J$. After conditioning on the variables ξ_{ij} with $j \notin J$, (2.11) gives

$$\mathbb{P}(|R_i \cdot u| \le t) = O\left(t + \frac{1}{\sqrt{n}} \right). \tag{2.12}$$

This motivates partitioning the sphere into *compressible* and *incompressible* vectors, which we now define. Denote $\mathrm{supp}(v) := \{ j \in [n] : v_j \ne 0 \}$, and for $m \in [n]$ define the set of m-*sparse vectors*

$$\mathrm{Sparse}(m) := \left\{ v \in \mathbb{C}^n : |\,\mathrm{supp}(v)| \le m \right\}. \tag{2.13}$$

For $m \in [n]$ and $\rho > 0$, the set of (m, ρ)-*compressible* unit vectors is defined to be the ρ-neighborhood of the set of m-sparse vectors in the sphere:

$$\mathrm{Comp}(m, \rho) := \mathbb{S}^{n-1} \cap \left(\mathrm{Sparse}(m) + \rho \mathbb{B}^n \right).$$

For $m \ge cn$ and ρ of constant order, one can show that *incompressible* vectors $u \in \mathbb{S}^{n-1} \setminus \mathrm{Comp}(m, \rho)$ are spread in the above sense, *i.e.* $|u_j| \sim 1/\sqrt{n}$ for $\ge c'n$ elements $j \in [n]$ for some constant $c' > 0$. Thus, (2.12) is effective for incompressible vectors. While we only have a crude anti-concentration bound for compressible vectors, the bound can be *tensorized* to show $\mathbb{P}(\|Mu\|_2 \le c\sqrt{n}) \le e^{-cn}$ for any fixed compressible vector u. Then, from the fact that $\mathrm{Comp}(m, \rho)$ has low metric entropy (i.e. it can be covered by a relatively small number of small balls) one can apply the union bound over a suitable net to show $\inf_{u \in \mathrm{Comp}(c_1 n, c_2)} \|(X_n + Z_n)u\|_2 \ge c'\sqrt{n}$ with high probability if c_1, c_2 are sufficiently small constants.

After obtaining uniform control on $\|(X_n + Z_n)u\|_2$ for $u \in \mathrm{Comp}(c_1 n, c_2)$, an averaging argument shows that in order to obtain an estimate of the form

$$\mathbb{P}\big(s_n(X_n + Z_n) \le t/\sqrt{n} \big) = O(t) + o(1),$$

it suffices to get a bound of the form $\mathbb{P}(|R_i \cdot u| \le t) = O(t) + o(1)$ for an arbitrary *fixed* row R_i and $u \in \mathbb{S}^{n-1} \backslash \mathrm{Comp}(c_1 n, c_2)$. But this now follows from (2.12). See [36] for the detailed presentation of this argument.

The distribution of S_n^d necessitates a somewhat modified approach, and in particular a different notion of structure than compressibility. In order to make use of the anti-concentration estimate (2.11) we will consider *pairs* of rows R_{i_1}, R_{i_2}. For each $\ell \in [d]$, conditioning on the remaining $n-2$ rows of P_n^ℓ fixes $\pi_n^\ell(\{i_1, i_2\})$. It follows that the i_1-st row of P_n^ℓ is e_j where j is drawn uniformly from $\pi_n^\ell(\{i_1, i_2\})$, and e_k denotes the k-th standard basis vector. Since the matrices $\{P_n^\ell\}_{\ell \in [d]}$ are independent, it is then possible to express

$$R_{i_1} \cdot u = w + \sum_{\ell=1}^{d} \xi_\ell (u_{\pi_n^\ell(i_1)} - u_{\pi_n^\ell(i_2)}) \tag{2.14}$$

where $\{\xi_\ell\}_{\ell \in [d]}$ are i.i.d. Rademacher variables and $w \in \mathbb{C}$ is some quantity that is deterministic under conditioning on the rows $[n] \backslash \{i_1, i_2\}$ of all of the matrices $\{P_n^\ell\}_{\ell \in [d]}$. By the discussion under (2.10), we can then get a bound on $\mathbb{P}(|R_{i_1} \cdot u| \le t)$ for small $t > 0$ via the Berry–Esséen-type bound (2.11), which will be effective when the vector of *differences* $(u_{\pi_n^\ell(i_1)} - u_{\pi_n^\ell(i_2)})_{\ell \in [d]}$ is spread. This motivates the following:

Definition 2.4. For $m \in [n]$ and $\rho \in (0, 1)$, define the set of (m, ρ)-*flat vectors*

$$\mathrm{Flat}(m, \rho) := \mathbb{S}^{n-1} \cap \left(\rho \mathbb{B}^n + \bigcup_{\lambda \in \mathbb{C}} \left(\lambda \mathbf{1} + \mathrm{Sparse}(m) \right) \right)$$

$$= \left\{ u \in \mathbb{S}^{n-1} : \exists v \in \mathrm{Sparse}(m),\, \lambda \in \mathbb{C} \text{ with } \| u - v - \lambda \mathbf{1} \|_2 \le \rho \right\} \tag{2.15}$$

(where the set $\mathrm{Sparse}(m)$ was defined in (2.13)). We denote the mean-zero flat vectors by

$$\mathrm{Flat}_0(m, \rho) := \mathrm{Flat}(m, \rho) \cap \langle \mathbf{1} \rangle^\perp. \tag{2.16}$$

For non-integral $x \ge 0$ we will sometimes abuse notation and write $\mathrm{Sparse}(x)$, $\mathrm{Flat}(x, \rho)$, etc. to mean $\mathrm{Sparse}(\lfloor x \rfloor)$, $\mathrm{Flat}(\lfloor x \rfloor, \rho)$.

Our first task is get a lower bound on $\inf_{u \in \mathrm{Flat}_0(m, \rho)} \|(S_n^d + Z_n) u\|_2$ holding with high probability for a suitable choice of m, ρ, which we obtain in Proposition 2.5 below. For a parameter $K \ge 1$ define the *boundedness event*

$$\mathcal{B}(K) := \left\{ \|S_n^d + Z_n\|_{\langle \mathbf{1} \rangle^\perp} \le K \sqrt{d} \right\} \tag{2.17}$$

(recall our notation (2.7)). We will eventually take $K = n^{\gamma_0}$ for an arbitrary fixed $\gamma_0 \ge 1$ (cf. Section 4.4). For $m \in [n]$ and $\rho \in (0, 1)$ (possibly depending on n), define the event

$$\mathcal{E}_K(m, \rho) := \mathcal{B}(K) \cap \left\{ \exists u \in \mathrm{Flat}_0(m, \rho) : \|(S_n^d + Z_n) u\|_2 \le \rho K \sqrt{d} \right\}. \tag{2.18}$$

Proposition 2.5 (Invertibility over flat vectors)**.** There exist absolute constants $C_{2.5}$, $c_{2.5}, \bar{c}_{2.5} > 0$ such that the following holds. Let $\gamma \ge 1$ and $1 \le K \le n^\gamma$. Assume $\log^3 n \le d \le n$. Then

$$\mathbb{P} \left\{ \mathcal{E}_K \left(\frac{c_{2.5} n}{\gamma \log^2 n},\, n^{-C_{2.5} \gamma \log_d n} \right) \right\} \le e^{-\bar{c}_{2.5} d} \tag{2.19}$$

for all n sufficiently large depending on γ.

Section 3 is devoted to the proof of Proposition 2.5, and we defer discussion of the proof ideas to that section.

The remainder of the proof of Theorem 2.2 is given in Section 4. Having obtained control on flat vectors, our aim will then be to reduce the problem to obtaining an anti-concentration estimate on $R_{i_1} \cdot u$, which we express as in (2.14), for a fixed row R_{i_1} and fixed $u \in \mathbb{S}^{n-1} \cap \langle \mathbf{1} \rangle^{\perp} \cap \mathrm{Flat}(m, \rho)^c$. (Actually we will consider dot products of the form $(R_{i_1} - R_{i_2}) \cdot u$, but these can also be expressed in the form (2.14).) As in the i.i.d. setting discussed above this can be accomplished by an averaging argument, but the argument here is more delicate due to the dependencies among the entries of S_n^d. We adapt an approach used in [29] for the invertibility problem for random regular digraphs. The vector u must be chosen to be almost-orthogonal to the span of rows $\{R_i : i \notin \{i_1, i_2\}\}$, and we want to ensure that the differences $u_{\pi_n^\ell(i_1)} - u_{\pi_n^\ell(i_2)}$ are large for a large number of $\ell \in [d]$. If the indices $\pi_n^\ell(i_1), \pi_n^\ell(i_2)$ were independent of u then it would be relatively easy to show that because u is non-flat, a random choice of i_1, i_2 will give us a large number of differences, on average. However, since both u and $\pi_n^\ell(i_1), \pi_n^\ell(i_2)$ are fixed by conditioning on $\{\pi_n^\ell(i) : i \in [n] \setminus \{i_1, i_2\}, \ell \in [d]\}$ the argument requires some care. See Lemma 4.4 for the details.

Having reduced to consideration of a random walk of the form (2.14) with a large number of large differences $u_{\pi_n^\ell(i_1)} - u_{\pi_n^\ell(i_2)}$, we can conclude using the Berry–Esséen-type bound (2.11); this is done in Lemma 4.6. In Section 4.4 we combine all of these elements to complete the proof of Theorem 2.2.

2.3 Control on the Stieltjes transform

We begin this section by fixing some notation. Denote $\mathbb{C}^+ := \{\xi \in \mathbb{C} : \mathrm{Im}\, \xi > 0\}$. Fixing any $z \in B_{\mathbb{C}}(0, R)$, for some $R > 0$, and $\xi \in \mathbb{C}^+$ we define the Green function as follows:

$$G(S_n^d) := G(S_n^d, \xi, z) := \left(\xi - \left(z - \frac{S_n^d}{\sqrt{d}} \right) \left(z - \frac{S_n^d}{\sqrt{d}} \right)^* \right)^{-1}.$$

Instead of working with the Green function $G_n(\cdot)$, we will see that it will be easier to work with its symmetrized version

$$\widetilde{G}(S_n^d) := \widetilde{G}(S_n^d, \xi, z) := \left[\xi - \begin{bmatrix} 0 & \left(z - \frac{S_n^d}{\sqrt{d}} \right) \\ \left(z - \frac{S_n^d}{\sqrt{d}} \right)^* & 0 \end{bmatrix} \right]^{-1}. \tag{2.20}$$

We next define the Stieltjes transform of the ESD of $(z - S_n^d/\sqrt{d})(z - S_n^d/\sqrt{d})^*$ and its symmetrized version,

$$m_n(\xi) := m_n(\xi, z) := \frac{1}{n} \mathrm{Tr}\, G(S_n^d, \xi, z), \quad \widetilde{m}_n(\xi) := \frac{1}{2n} \mathrm{Tr}\, \widetilde{G}(S_n^d, \xi, z).$$

Recall that the eigenvalues of the matrix

$$\boldsymbol{S}_n^{d,z} := \begin{bmatrix} 0 & \left(z - \frac{S_n^d}{\sqrt{d}} \right) \\ \left(z - \frac{S_n^d}{\sqrt{d}} \right)^* & 0 \end{bmatrix} \tag{2.21}$$

are $\pm s_i(z - S_n^d/\sqrt{d})$ where $s_i(z - S_n^d/\sqrt{d})$ are the singular values of $z - S_n^d/\sqrt{d}$. Therefore, $\widetilde{m}_n(\xi)$ is the Stieltjes transform of the symmetrized version of the empirical measure of the singular values of $z - S_n^d/\sqrt{d}$, and one has

$$\widetilde{m}_n(\xi) = \xi m_n(\xi^2). \tag{2.22}$$

Our goal is to show that $\widetilde{m}$ converges to a limit $\widetilde{m}_\infty$ which is the Stieltjes transform of a probability measure on $\mathbb{R}$ and satisfies the equation

$$\widetilde{m}_\infty(\xi)(\xi - \widetilde{m}_\infty(\xi))^2 + \widetilde{m}_\infty(\xi)(1 - |z|^2) - \xi = 0. \tag{2.23}$$

As explained above, we need a bit more: we need to control the difference $|\widetilde{m}_n(\xi) - \widetilde{m}_\infty(\xi)|$ for all $\xi \in \mathbb{C}^+$ such that $\operatorname{Im} \xi \geq (\log^2 n)^{-1}$. The proof of Theorem 1.1 only requires such control for ξ purely imaginary. This is achieved in Theorem 2.6 below.

Theorem 2.6. Fix any sufficiently small $\varepsilon > 0$ and $z \in B_{\mathbb{C}}(0, 1 - \varepsilon)$. Take any sequence of reals $\{\varpi_n\}_{n \in \mathbb{N}}$ such that $\varpi_n \to \infty$ as $n \to \infty$. Then there exist a constant $\widetilde{C}_{2.6}$, depending only on ε, absolute constants $c_{2.6}, \bar{C}_{2.6}, C_{2.6}$, and an event Ω_n with

$$\mathbb{P}(\Omega_n^c) \leq C_{2.6} \exp(-c_{2.6}(\log n)^2) + \exp(-c_{2.6}d),$$

such that for all large n, on the event Ω_n we have

$$|\widetilde{m}_n(\xi) - \widetilde{m}_\infty(\xi)| \leq \widetilde{C}_{2.6} \max\left\{\frac{1}{d^{1/2}}, \frac{\log n}{n^{1/4}}\right\} (\operatorname{Im} \xi)^{-3}$$

for all $\xi \in \mathscr{S}_{\varepsilon, \varpi}$ where

$$\mathscr{S}_{\varepsilon, \varpi} := \left\{\xi = i\eta : \eta \in (0, \bar{C}_{2.6}], \ \eta^3 \min\{d^{1/2}, n^{1/4}(\log n)^{-1}\} \geq \varpi_n\right\}.$$

Remark 2.7. In Theorem 2.6 we treat the case when ξ is purely imaginary, which simplifies some of the computations. One can use a similar idea as in the proof of Theorem 2.6 to control the difference of $\widetilde{m}_n(\xi)$ and $\widetilde{m}_\infty(\xi)$ for all $\xi \in \mathbb{C}^+$ when $\operatorname{Im} \xi \geq (\log n)^{-C}$ for some $C > 0$. The key is to establish stability of the equation (2.23) for all $\xi \in \mathbb{C}^+$. Since the proof of Theorem 1.1 does not require such control we do not attempt it here.

The key to the proof of Theorem 2.6 is to establish that $\widetilde{m}_n(\xi)$ satisfies an approximate version of the equation (2.23). That is we need to show that $\widetilde{P}(\widetilde{m}_n(\xi)) \approx 0$ where $\widetilde{P}(m) := m(\xi - m)^2 + m(\xi)(1 - |z|^2) - \xi$. To show this, it is easier to work with $\widehat{m}_n(\xi)$, the Stieltjes transform of the symmetrized version of the empirical measure of the singular values of $z - \widehat{S}_n^d/\sqrt{d}$ where the entries of $\widehat{S}_n^d$ are now centered. Then concentration bounds for Lipschitz functions of permutations under the Hamming metric also allow us only to consider $\widetilde{P}(\mathbb{E}\widehat{m}_n(\xi))$.

To show that $\widetilde{P}(\mathbb{E}\widehat{m}_n(\xi)) \approx 0$ we start with a function related to $\widetilde{G}(\widehat{S}_n^d)$, where $\widetilde{G}(\widehat{S}_n^d)$ is defined by replacing $\widetilde{S}_n^d$ with $\widehat{S}_n^d$ in (2.20). Then we use the resolvent identity and the fact that $\{P_n^\ell\}$ are independent to identify the dominant and negligible terms. This yields an approximate equation involving $\mathbb{E}\widehat{m}_n(\xi)$ and an auxiliary variable. To remove the auxiliary variable we derive another approximate equation.

However, this alone does not yield Theorem 2.6. Because $\widetilde{P}(\cdot)$ is cubic polynomial, bounds on $\widetilde{P}(\cdot)$ do not translate to bounds on $|\widetilde{m}_n(\xi) - \widetilde{m}_\infty(\xi)|$. Moreover, the bound on $\widetilde{P}(\widetilde{m}_n(\xi))$ depends implicitly on an bound on $\widetilde{m}_n(\xi)$ (see Lemma 7.1). To overcome this difficulty, in Lemma 7.6 we show that if $\widetilde{m}_n(\xi)$ if bounded below then a bound on $\widetilde{P}(\cdot)$ can be translated to a bound on the difference between $\widetilde{m}_n(\xi)$ and $\widetilde{m}_\infty(\xi)$. On other hand, we can easily show that the desired bounds on $\widetilde{m}_n(\xi)$ hold when ξ is far away from the real line. This gives Theorem 2.6 when ξ away from the real line.

To propagate the above bound for all $\xi \in \mathscr{S}_{\varepsilon, \varpi}$ we use a *bootstrap* argument. In the random matrix literature the bootstrap argument has already been used on many occasions to prove local law for different random matrix ensembles. Specifically, Erdős, Schlein, and Yau [19] used it to prove the local semicircle law for Wigner matrices down to the optimal scale. Subsequently it was generalized to prove local laws for other ensembles of random matrices (see [9] and references therein).

To carry out the above scheme for $\xi \in \mathbb{C}^+$ such that $\operatorname{Im} \xi$ is small we note that by Lipschitz continuity and the boundedness property of $\widetilde{m}_\infty(\xi)$ derived in Lemma 7.8, the bounds on $\widetilde{m}_n(\xi)$ translates to a bound on the same with ξ replaced by ξ', whenever $|\operatorname{Im} \xi - \operatorname{Im} \xi'|$ is small. These bounds on $\widetilde{m}_n(\xi')$ together with Lemma 7.1 yield the desired bound on $|\widetilde{m}_n(\xi') - \widetilde{m}_\infty(\xi')|$. Repeating this scheme we obtain the desired result for all $\xi \in \mathscr{S}_{\varepsilon,\varpi}$.

We note that in the work [16] on the spectrum of the adjacency matrix $A_{n,d}$ for a random d-regular digraph, a completely different argument is used to obtain quantitative control on the Stieltjes transforms $g_{\xi,z}(A_{n,d}) = \frac{1}{n}\operatorname{Tr} G(A_{n,d}, \xi, z)$. There the approach is by comparison, first replacing $A_{n,d}$ with an i.i.d. 0–1 Bernoulli matrix $B_{n,p}$ with entries of mean $p = d/n$, and then replacing $B_{n,p}$ with a suitably rescaled real Ginibre matrix G_n (for which the desired bounds are known to hold), showing that $g_{\xi,z}$ changes by a negligible amount at each replacement. The comparison between $g_{\xi,z}(B_{n,p})$ and $g_{\xi,z}(G_n)$ is done using the standard Lindeberg swapping argument, whose use in random matrix theory goes back to Chatterjee [14]. The comparison of $g_{\xi,z}(A_{n,d})$ with $g_{\xi,z}(B_{n,p})$ is done by conditioning, basically showing that $g_{\xi,z}(B_{n,p})$ concentrates near its expected value with failure probability smaller than the probability that $B_{n,p}$ lies in $\mathcal{A}_{n,d}$, the set of adjacency matrices for d-regular digraphs. Since $A_{n,d}$ is uniform in $\mathcal{A}_{n,d}$, obtaining a lower bound for the latter probability amounts to the enumerative problem of estimating the cardinality of $\mathcal{A}_{n,d}$, which can be solved with known techniques. It is possible that this comparison approach could be adapted to the current setup, first replacing S_n^d with a discrete i.i.d. matrix M_n^d having i.i.d. Poisson entries, and then replacing M_n^d with a Gaussian matrix. However, as S_n^d is not drawn uniformly from a set of matrices the first step would not reduce to an enumeration problem as it did for $A_{n,d}$, and hence this step appears more challenging. Instead we would need a coupling between S_n^d and M_n^d, together with a lower bound on the probability that they are close in an appropriate norm. It is likely that a proof along these lines, even if doable, would be somewhat lengthier than the approach taken in the present article.

3 Invertibility over flat vectors

In this section we prove Proposition 2.5. Throughout this section and Section 4 we let S_n^d and Z_n be as in the statement of Theorem 2.2, except that some lemmas and propositions are stated under additional assumptions on the range of d. (Recall from Remark 2.3 that we are free to assume $d \le n$; also note that Theorem 2.2 trivially holds for $d \le \log^8 n$.)

The general approach is similar to the proof in [16], and indeed we make use of two lemmas from that work (Lemma 3.5 and Lemma 3.6). However, the differences between the distribution of S_n^d and the adjacency matrix of a uniform random regular digraph $A_{n,d}$ cause the proof here to differ on most of the particulars. We have attempted to structure the proof in roughly the same way as in [16], and use Lemma 3.1 to encapsulate the parts of the proof which are most different from that work. On a technical level, the proof here is somewhat simpler as the joint independence of the permutations π_n^ℓ allows us to avoid the difficult coupling constructions of [16], as well as the use of heavy-powered graph discrepancy results.

3.1 Anti-concentration for the image of a fixed vector

To lighten notation we will drop subscripts n from $\boldsymbol{\pi}_n, \pi_n^\ell$ in this section.

We begin by obtaining lower tail bounds for the norm of $(S_n^d + Z_n)u$ for a fixed vector $u \in \mathbb{S}^{n-1}$.

Lemma 3.1 (Image of a fixed vector)**.** There exist absolute constants $c_{3.1}, \bar{c}_{3.1} > 0$ such

that the following holds. Let $d \geq 1$, and let $u \in \mathbb{C}^n$ be such that there are disjoint sets $J_1, J_2 \subset [n]$ and $\rho > 0$, with $|J_1| = |J_2| = m$, such that

$$|u_{j_1} - u_{j_2}| \geq \frac{\rho}{\sqrt{n}} \qquad \forall j_1 \in J_1, \ j_2 \in J_2. \tag{3.1}$$

Then

$$\mathbb{P}\left(\|(S_n^d + Z_n)u\|_2 \leq c_{3.1}\rho \min\left\{ \sqrt{\frac{md}{n}}, 1 \right\} \right) \leq \exp\left(-\bar{c}_{3.1} \min(md, n) \right). \tag{3.2}$$

Remark 3.2. We note that (3.2) is essentially optimal when md is small compared with n, at least for the case $Z_n = 0$ (and we are aiming for estimates that are uniform in Z_n). Indeed, $\|S_n^d u\|_2^2 = \sum_{i=1}^n |R_i \cdot u|^2$, where R_i is the ith row of S_n^d. When $md = o(n)$ the number of "good" rows R_i whose support overlaps the support of u will be roughly md on average (in fact it concentrates near md, as will be shown in the proof). For each good row R_i we will have $\mathbb{E}|R_i \cdot u|^2 \approx 1/n$, since the overlap of supports is of order 1 on average, and coordinates u_j are typically of size $1/\sqrt{n}$. This means we should expect $\|S_n^d u\|_2^2 \approx md/n$, and (3.2) gives a lower bound at this scale. However, the bound is suboptimal when $m \approx n$, in which case there will be roughly $\asymp n$ good rows with overlaps of order d, which suggests $\mathbb{E}\|S_n^d u\|_2^2 \approx d$ in this case. Thus, we expect (3.2) to hold with $\min(\sqrt{md/n}, 1)$ replaced with $\sqrt{md/n}$. The proof could be extended to give such a bound by exploiting the randomness of all d permutations within each row (in the proof we only use one permutation per row) but such a refinement is not necessary for our purposes.

The above lemma is a quick consequence of Lemma 3.3 below. First we need some notation. We write $J := J_1 \cup J_2$, and for each $k \in [d]$ we set $I_k := (\pi^k)^{-1}(J) = I_k^1 \cup I_k^2$ with $I_k^a := (\pi^k)^{-1}(J_a)$ for $a = 1, 2$. Note that I_k^1, I_k^2 are disjoint sets of size m. We also denote

$$\boldsymbol{\pi}^{<k} := (\pi^\ell)_{\ell \in [k-1]}, \qquad \boldsymbol{\pi}^{>k} := (\pi^\ell)_{\ell \in [k+1,d]}, \qquad \boldsymbol{\pi}^{(k)} := (\pi^\ell)_{\ell \in [d] \setminus \{k\}} \tag{3.3}$$

and

$$U_{<k} := \bigcup_{\ell \in [k-1]} I_\ell, \qquad U_{>k} := \bigcup_{\ell \in [k+1,d]} I_\ell, \qquad U_{(k)} := \bigcup_{\ell \in [d] \setminus \{k\}} I_\ell \tag{3.4}$$

(with $U_{<1} = U_{>d} := \varnothing$). We further write $\pi^{\leq k} := \pi^{<k+1}$, $U_{\geq k} := U_{>k-1}$, etc. For fixed $u \in \mathbb{C}^n$ and for $k \in [d]$ let

$$W_k(u) := \sum_{i \in U_{\leq k}} |R_i \cdot u|^2, \qquad X_k(u) := \exp\left(-\frac{n}{\rho^2} W_k(u) \right).$$

Lemma 3.3. Let J_1, J_2, u, and m be as in Lemma 3.1. There are absolute constants $c_0, c_1 > 0$ such that for any $d_0 \leq \min(d, c_0 n/m)$,

$$\mathbb{E}X_{d_0}(u) \leq e^{-c_1 m d_0}. \tag{3.5}$$

Proof of Lemma 3.1. Let $c_0, c_1 > 0$ be as in Lemma 3.3 and let $d_0 = \lfloor \min(d, c_0 n/m) \rfloor$. For any $c_2 > 0$ we have

$$\mathbb{P}\left(\|(S_n^d + Z_n)u\|_2 \leq c_2 \rho \sqrt{\frac{md_0}{n}} \right) = \mathbb{P}\left(\sum_{i=1}^n |R_i \cdot u|^2 \leq c_2^2 \rho^2 \frac{md_0}{n} \right)$$

$$\leq \mathbb{P}\left(W_{d_0}(u) \leq c_2^2 \rho^2 \frac{md_0}{n} \right).$$

Using the pointwise bound $\mathbb{I}_{[0,\infty)}(x) \le \exp((n/\rho^2)x)$ followed by Lemma 3.3, we can bound the last expression by

$$e^{c_2^2 m d_0} \mathbb{E} X_{d_0}(u) \le \exp((c_2^2 - c_1)m d_0).$$

Taking $c_2 = c_1^{1/2}/2$, the claim follows. $\qquad\square$

Proof of Lemma 3.3. Fix u as in the statement of the lemma. To lighten notation we will drop the dependence on u from $X_k(u), W_k(u)$ and write X_k, W_k.

First we note that for any $\ell \in [d]$, $j_1 \in J_1$, and $j_2 \in J_2$, if $i \in (\pi_n^\ell)^{-1}(\{j_1, j_2\})$, then we have

$$\mathbb{P}\left\{ |R_i \cdot u| \le \frac{\rho}{4\sqrt{n}} \;\middle|\; \boldsymbol{\pi}^{(\ell)}, ((\pi^\ell)^{-1}(j))_{j \notin \{j_1, j_2\}} \right\} \le \frac{1}{2}. \tag{3.6}$$

Indeed, fixing $((\pi^\ell)^{-1}(j))_{j \notin \{j_1, j_2\}}$ we see that for any $i \in (\pi^\ell)^{-1}(\{j_1, j_2\})$, either $\pi^\ell(i) = j_1$ or j_2 with equal probability. Thus, under the conditioning in (3.6) we have

$$|R_i \cdot u| = \Delta_i + u_{j_1} \qquad \text{or} \qquad \Delta_i + u_{j_2},$$

with equal probability, where Δ_i is some non-random quantity depending on $\boldsymbol{\pi}^{(\ell)}$, $((\pi^\ell)^{-1}(j))_{j \notin \{j_1, j_2\}}$. Using the assumption (3.1) and the triangle inequality we immediately deduce (3.6). Now using (3.6),

$$\mathbb{E}\left(\exp\left(-\frac{n}{\rho^2} |R_i \cdot u|^2 \right) \;\middle|\; \boldsymbol{\pi}^{(\ell)}, ((\pi^\ell)^{-1}(j))_{j \notin \{j_1, j_2\}} \right)$$
$$= \int_0^1 \mathbb{P}\left\{ e^{-\frac{n}{\rho^2}|R_i \cdot u|^2} \ge s \;\middle|\; \boldsymbol{\pi}^{(\ell)}, ((\pi^\ell)^{-1}(j))_{j \notin \{j_1, j_2\}} \right\} ds$$
$$\le \frac{1}{2}(1 - e^{-1/16}) + e^{-1/16} =: 1 - q. \tag{3.7}$$

Now we establish the claim for the case $m = 1$, in which case $J = \{j_1, j_2\}$, and for each $k \in [d]$ we set $I_k = \{i_k^1, i_k^2\}$. The sets I_k are i.i.d. uniform random subsets of $[n]$ of size 2. Let $d_1 = c_3 d$ for some $c_3 > 0$ to be determined later. For $2 \le k \le d_1$, say that k is "bad" if $I_k \cap U_{<k} \ne \varnothing$, and let $\mathcal{B}_k$ be the event that k is bad. Then for each $2 \le k \le d_1$,

$$\mathbb{E}\mathbb{I}(\mathcal{B}_k) \le 2d_1 \cdot n \Big/ \binom{n}{2} = O(c_3) \tag{3.8}$$

(recall our assumption $d \le n$ from Remark 2.3). Thus we have that $B := \sum_{k=2}^{d_1} \mathbb{I}(\mathcal{B}_k)$ is stochastically dominated by a sum of i.i.d. indicator variables with expectation $O(c_3)$. From the Chernoff bound it thus follows that

$$\mathbb{P}(B > d_1/2) \le e^{-cd}, \tag{3.9}$$

taking c_3 sufficiently small. Let us denote the complement of this event by $\mathcal{G}$. On $\mathcal{G}$, there exists a set $G \subset [d_1]$ with $|G| \ge d_1/2$ such that the sets $\{I_k\}_{k \in G}$ are pairwise disjoint. We take G to be the largest such set (in the event of a tie we pick one in some measurable fashion). We have

$$\mathbb{E}\left[X_{d_1} \;\middle|\; G, (\pi^k)_{k \notin G}, (I_k)_{k \in G} \right] \mathbb{I}(\mathcal{G})$$
$$\le \mathbb{E}\left[\exp\left(-\frac{n}{\rho^2} \sum_{k \in G} |R_{i_k^1} \cdot u|^2 \right) \;\middle|\; G, (\pi^k)_{k \notin G}, (I_k)_{k \in G} \right] \mathbb{I}(\mathcal{G})$$
$$= \prod_{k \in G} \mathbb{E}\left[\exp\left(-\frac{n}{\rho^2} |R_{i_k^1} \cdot u|^2 \right) \;\middle|\; G, (\pi^k)_{k \notin G}, (I_k)_{k \in G} \right] \mathbb{I}(\mathcal{G}) \le (1 - q)^{|G|} \mathbb{I}(\mathcal{G}).$$

Thus, since $d_1 \le d_0$,

$$\mathbb{E}X_{d_0} \le \mathbb{E}X_{d_1} \le \mathbb{P}(\mathcal{G}^c) + \mathbb{E}X_{d_1}\mathbb{I}(\mathcal{G}) \le e^{-cd} + (1-q)^{d_1/2} \le e^{-c'd}$$

for some constant $c' > 0$, establishing the lemma for the case $m = 1$.

Now assume $m \ge 2$. In fact we are now free to assume $m \ge C_0$ for some absolute constant $C_0 > 0$ to be specified later. Indeed, for $m \le C_0$ we can simply pass to singleton subsets of J_1, J_2 and apply the case $m = 1$ (adjusting the constant c_1).

We next show that for any fixed $k \in [d]$,

$$\mathbb{E}\left[\exp\left(-\frac{n}{\rho^2} \sum_{i \in I_k \backslash U_{(k)}} |R_i \cdot u|^2 \right) \, \middle| \, \boldsymbol{\pi}^{(k)}, (\pi^k(i))_{i \in U_{(k)}} \right]$$

$$\le \mathbb{E}\left[(1-q)^{M_k} \, \middle| \, \boldsymbol{\pi}^{(k)}, (\pi^k(i))_{i \in U_{(k)}} \right], \qquad (3.10)$$

where

$$M_k := \min(|I_k^1 \backslash U_{(k)}|, |I_k^2 \backslash U_{(k)}|). \qquad (3.11)$$

Note that the expectation in (3.10) is only taken over part of the randomness of the permutation π^k. The idea for the proof is that after some further conditioning we can reduce to using only the randomness of π^k on M_k pairwise disjoint sets $T_1, \ldots, T_{M_k} \subset I_k \backslash U_{(k)}$ of size two, and the action of π^k on these sets can be realized as the application of M_k independent transpositions. Thus, we can extract a subsequence of M_k rows R_{i_j} that are jointly independent under the conditioning, and apply the bound (3.7) to each one.

We turn to the details. Fix $k \in [d]$ and write $\hat{I}_k^a := I_k^a \backslash U_{(k)}$ for $a = 1, 2$. For given $m_0 \in \mathbb{N}$ and $U \subset [n]$ let $\mathcal{T}(m_0, U)$ be the collection of all sequences $\mathbf{T} := (T_j)_{j=1}^{m_0}$ of pairwise disjoint 2-sets $T_j := \{i_j^1, i_j^2\} \subset [n] \backslash U$. Given $\mathbf{T} \in \mathcal{T}(m_0, U)$, define the set

$$\$_{\mathbf{T}} := \bigcap_{j \in [m_0]} \{\pi \in \$_n : |T_j \cap (\pi^{-1}(J_1))\backslash U| = |T_j \cap (\pi^{-1}(J_2))\backslash U| = 1\}.$$

(Since $\pi^{-1}(J_1), \pi^{-1}(J_2)$ are disjoint, this is the event that they bisect each of the sets T_j for $1 \le j \le m_0$.) Conditional on $\boldsymbol{\pi}^{(k)}$ and M_k, for any $\mathbf{T} \in \mathcal{T}(M_k, U_{(k)})$,

$$\mathbb{E}\left[\exp\left(-\frac{n}{\rho^2} \sum_{i \in I_k \backslash U_{(k)}} |R_i \cdot u|^2 \right) \, \middle| \, \boldsymbol{\pi}^{(k)}, (\pi^k(i))_{i \in U_{(k)}}, M_k \right] \mathbb{I}(\pi^k \in \$_{\mathbf{T}})$$

$$\le \mathbb{E}\left[\exp\left(-\frac{n}{\rho^2} \sum_{j=1}^{M_k} |R_{i_j^1} \cdot u|^2 \right) \, \middle| \, \boldsymbol{\pi}^{(k)}, (\pi^k(i))_{i \in U_{(k)}}, M_k \right] \mathbb{I}(\pi^k \in \$_{\mathbf{T}})$$

$$= \prod_{j=1}^{M_k} \mathbb{E}\left[\exp\left(-\frac{n}{\rho^2} |R_{i_j^1} \cdot u|^2 \right) \, \middle| \, \boldsymbol{\pi}^{(k)}, (\pi^k(i))_{i \in U_{(k)}}, M_k \right] \mathbb{I}(\pi^k \in \$_{\mathbf{T}})$$

$$\le (1-q)^{M_k}\mathbb{I}(\pi^k \in \$_{\mathbf{T}}),$$

where in the penultimate line we noted that under the conditioning and restriction to $\pi^k \in \$_{\mathbf{T}}$ the pairs of rows $\{(R_{i_j^1}, R_{i_j^2})\}_{j=1}^{M_k}$ are jointly independent, and in the last line we applied (3.7). Now letting $\mathcal{T}' := \mathcal{T}'(M_k, U_{(k)}) \subset \mathcal{T}(M_k, U_{(k)})$ be a sub-collection such that $\{\$_{\mathbf{T}}\}_{\mathbf{T} \in \mathcal{T}'}$ partitions the range of π^k under the conditioning on $\boldsymbol{\pi}^{(k)}, (\pi^k(i))_{i \in U_{(k)}}, M_k$, we

have

$$\mathbb{E}\left[\exp\left(-\frac{n}{\rho^2}\sum_{i\in I_k\setminus U_{(k)}}|R_i\cdot u|^2\right)\;\middle|\;\boldsymbol{\pi}^{(k)},(\pi^k(i))_{i\in U_{(k)}},M_k\right]$$
$$\leq\sum_{\mathbf{T}\in\mathcal{T}'}(1-q)^{M_k}\mathbb{I}(\pi^k\in\$_{\mathbf{T}})\leq(1-q)^{M_k}.$$

Undoing the conditioning on M_k yields (3.10) as desired.

Define the decreasing sequence of sigma algebras

$$\mathcal{F}_k:=\langle\pi^{>k},(\pi^\ell(i))_{\ell\leq k,i\in U_{>k}}\rangle,\quad k=1,\dots,d-1 \tag{3.12}$$

and set $\mathcal{F}_d$ to be the trivial sigma algebra. In words, conditioning on $\mathcal{F}_k$ fixes the permutations $\pi_{k+1},\dots,\pi_d$, along with the values $\pi^\ell(i)$ for $\ell\leq k$ and all i in the preimages of $J=J_1\cup J_2$ under $\pi_{k+1},\dots,\pi_d$. Note that $\mathcal{F}_1\supset\cdots\supset\mathcal{F}_d$. Note also that for any $k\in[d]$ the random variable M_k defined in (3.11) is $\mathcal{F}_{k-1}$-measurable. Indeed, conditioning on $\mathcal{F}_{k-1}$ fixes $I_k^1,I_k^2,U_{\geq k}$, and $\pi^\ell(i)$ for all $\ell\leq k-1$ and $i\in I_k\subset U_{\geq k}$, which in turn determine

$$I_k^a\cap U_{<k}=\bigcup_{\ell\leq k-1}\{i\in I_k^a:\pi^\ell(i)\in J\},\qquad a=1,2,$$

so $I_k^a\cap U_{<k}^c\cap U_{>k}^c=I_k^a\setminus U_{(k)}$ are fixed as well for $a=1,2$.

From (3.10),

$$\mathbb{E}[X_1|\mathcal{F}_1]=\mathbb{E}\left[\exp\left(-\frac{n}{\rho^2}\sum_{i\in I_1}|R_i\cdot u|^2\right)\;\middle|\;\mathcal{F}_1\right]$$
$$\leq\mathbb{E}\left[\exp\left(-\frac{n}{\rho^2}\sum_{i\in I_1\setminus U_{(1)}}|R_i\cdot u|^2\right)\;\middle|\;\mathcal{F}_1\right]\leq\mathbb{E}\left[(1-q)^{M_1}\;\middle|\;\mathcal{F}_1\right]. \tag{3.13}$$

Now letting $2\leq k\leq d_0$, we have

$$\mathbb{E}[X_k|\mathcal{F}_k]=\mathbb{E}\left[X_{k-1}\mathbb{E}\left[\exp\left(-\frac{n}{\rho^2}\sum_{i\in I_k\setminus U_{(k)}}|R_i\cdot u|^2\right)\;\middle|\;\boldsymbol{\pi}^{(k)},(\pi^k(i))_{i\in U_{(k)}}\right]\;\middle|\;\mathcal{F}_k\right]$$
$$\leq\mathbb{E}\left[X_{k-1}\mathbb{E}\left[(1-q)^{M_k}\;\middle|\;\boldsymbol{\pi}^{(k)},(\pi^k(i))_{i\in U_{(k)}}\right]\;\middle|\;\mathcal{F}_k\right]$$
$$=\mathbb{E}\left[X_{k-1}(1-q)^{M_k}\;\middle|\;\mathcal{F}_k\right]=\mathbb{E}\left[\mathbb{E}[X_{k-1}|\mathcal{F}_{k-1}](1-q)^{M_k}\;\middle|\;\mathcal{F}_k\right],$$

where the penultimate equality follows upon noting that

$$\mathcal{F}_k\subset\langle\boldsymbol{\pi}^{(k)},(\pi^k(i))_{i\in U_{(k)}}\rangle$$

and applying the tower property of the conditional expectation, and in the last step we have used that M_k is $\mathcal{F}_{k-1}$-measurable. Iterating this bound over $2\leq k\leq d_0$ and combining with (3.13) we obtain

$$\mathbb{E}[X_{d_0}|\mathcal{F}_{d_0}]\leq\mathbb{E}\left[\prod_{\ell=1}^{d_0}(1-q)^{M_\ell}\;\middle|\;\mathcal{F}_{d_0}\right]\leq\mathbb{E}\left[\prod_{\ell=1}^{d_0}\left((1-q)^{m/2}+\mathbb{I}\left(M_\ell<\frac{m}{2}\right)\right)\;\middle|\;\mathcal{F}_{d_0}\right].$$

Thus,

$$\mathbb{E}X_{d_0}\leq\mathbb{E}\prod_{\ell=1}^{d_0}\left((1-q)^{m/2}+\mathbb{I}\left(M_\ell<\frac{m}{2}\right)\right)=\sum_{k=0}^{d_0}(1-q)^{\frac{1}{2}m(d_0-k)}\sum_{L\in\binom{[d_0]}{k}}\mathbb{P}\left(\mathcal{E}_L\right) \tag{3.14}$$

where

$$\mathcal{E}_L := \bigcap_{\ell \in L} \left\{ M_\ell < \frac{m}{2} \right\}.$$

Next we will show that for any $L \subset [d_0]$,

$$\mathbb{P}(\mathcal{E}_L) \leq e^{-cm|L|}, \tag{3.15}$$

for some absolute constant $c > 0$. Assuming (3.15), we have from (3.14) that

$$\mathbb{E}X_{d_0} \leq \sum_{k=0}^{d_0} \binom{d_0}{k} (1-q)^{\frac{1}{2}m(d_0-k)} e^{-cmk} = ((1-q)^{m/2} + e^{-cm})^{d_0} \leq e^{-c'md_0},$$

where the last inequality is obtained by taking the constant $C_0 > 0$ sufficiently large and thus $m \geq C_0$. This yields (3.5) and hence Lemma 3.3.

It only remains to establish (3.15). Since the variables M_ℓ are exchangeable we may take $L = [k]$. As $I_\ell = I_\ell^1 \cup I_\ell^2$ with $|I_\ell^1| = |I_\ell^2| = m$, on $\mathcal{E}_L$ we have $|I_\ell \backslash U_{(\ell)}| < 3m/2$ for each $\ell \leq k$. Hence,

$$|U_{\leq k}| = \left| \bigcup_{\ell \leq k} I_\ell \right| = \sum_{\ell \leq k} |I_\ell \backslash U_{<\ell}| < \frac{3}{2}mk.$$

On the other hand,

$$|U_{<\ell}| \leq |U_{\leq k}| \leq 2mk \leq 2md_0 \leq 2c_0 n,$$

and since the sets I_ℓ are independent and uniformly distributed over $\binom{[n]}{2m}$, we have

$$\mathbb{E}|I_\ell \backslash U_{<\ell}| \geq 2m(1 - 2c_0) \geq 1.9m$$

for each $\ell \leq k$, where we took the constant c_0 sufficiently small. Hence,

$$\mathbb{E}|U_{\leq k}| = \sum_{\ell=1}^{k} \mathbb{E}|I_\ell \backslash U_{<\ell}| \geq 1.9mk.$$

We have thus shown

$$\mathbb{P}(\mathcal{E}_L) \leq \mathbb{P}(|U_{\leq k}| < 0.99\mathbb{E}|U_{\leq k}|).$$

The latter probability can be shown to be at most e^{-cmk} by an argument using stochastic domination and the Chernoff bound similar to what was done in (3.8)–(3.9). This gives (3.15) and hence the claim. $\square$

3.2 Weak control on flat vectors

In this subsection we establish the following, which already implies Proposition 2.5 when $d \geq n/\log n$, but is weaker for smaller values of d. Recall the events $\mathcal{E}_K(m, \rho)$ from (2.18).

Lemma 3.4 (Invertibility over flat vectors, weak version)**.** There are absolute constants $c_{3.4}$, $\overline{c}_{3.4}$, $c'_{3.4} > 0$ such that the following holds. Let $\gamma \geq 1$ and $1 \leq K \leq n^\gamma$. Assume $d \geq 1$. Then for any $1 \leq m_0 \leq c'_{3.4}d/\gamma \log n$,

$$\mathbb{P}\left\{ \mathcal{E}_K\left(m_0, \frac{c_{3.4}}{K\sqrt{m_0}} \right) \right\} \leq e^{-\overline{c}_{3.4}d}. \tag{3.16}$$

We will need the following lemma from [16].

Lemma 3.5 (Metric entropy for flat vectors, cf. [16, Lemma 3.3])**.** Let $1 \leq m \leq n/10$ and $\rho \in (0, 1)$. There exists $\Sigma_0 := \Sigma_0(m, \rho) \subset \text{Flat}_0(m, \rho)$ such that Σ_0 is a ρ-net for $\text{Flat}_0(m, \rho)$ and $|\Sigma_0| \leq \left(\frac{C_{3.5}n}{m\rho^2} \right)^m$ for some absolute constant $C_{3.5} > 0$.

Proof of Lemma 3.4. Our plan is to use Lemma 3.1 first to obtain a bound on $\|(S_n^d + Z_n)u\|_2$ for any arbitrary but fixed $u \in \mathrm{Flat}(m_0, \rho_0)$, where $\rho_0 := c/K\sqrt{m_0}$ for some c to be determined determined during the course of the proof. Then using Lemma 3.5 we claim that the metric entropy of $\mathrm{Flat}(m_0, \rho_0)$ is small enough to allow us to take a union bound.

In order to apply Lemma 3.1 we need to find disjoint sets J_1 and J_2 such that $|u_{j_1} - u_{j_2}|$ is large for every $j_1 \in J_1$ and $j_2 \in J_2$. To this end, consider an arbitrary vector $u \in \mathrm{Flat}_0(m_0, \rho_0)$. By definition, there exists $\lambda \in \mathbb{C}$, $v \in \mathrm{Sparse}(m_0)$ and $w \in \rho_0 B_{\mathbb{C}}(0, 1)$ such that $u = v + \frac{\lambda}{\sqrt{n}} \mathbf{1} + w$. First we claim that

$$\|v + w\|_2 \geq 1/2. \tag{3.17}$$

Indeed, by the triangle inequality,

$$|\lambda| = \left\| \frac{\lambda}{\sqrt{n}} \mathbf{1} \right\|_2 \geq \|u\|_2 - \|v + w\|_2 = 1 - \|v + w\|_2. \tag{3.18}$$

On the other hand by the assumption $u \in \mathbb{S}^{n-1} \cap \langle \mathbf{1} \rangle^{\perp}$ and applying the Cauchy–Schwarz inequality we get

$$|\lambda|\sqrt{n} = \left| \sum_{j=1}^{n} (v_j + w_j) \right| \leq \|v + w\|_2 \sqrt{n}$$

and so

$$|\lambda| \leq \|v + w\|_2.$$

Combined with (3.18) this gives (3.17).

Let $J \subset [n]$ with $|J| = m_0$ such that $\mathrm{supp}(v) \subset J$. Shrinking ρ_0, if necessary, from (3.17) we obtain

$$\frac{1}{8} \leq \frac{1}{4} - \|w\|_2^2 \leq \sum_{j \in J} |v_j + w_j|^2 \leq m_0 \max_{j \in J} \left| u_j - \frac{\lambda}{\sqrt{n}} \right|^2.$$

It follows that there exists $j_1 \in J$ with

$$\left| u_{j_1} - \frac{\lambda}{\sqrt{n}} \right| \geq \frac{1}{2\sqrt{2m_0}}. \tag{3.19}$$

On the other hand, since $\sum_{j \in J^c} |w_j|^2 \leq \|w\|_2^2 \leq \rho_0^2$ it follows from the pigeonhole principle that there exists $j_2 \in J^c$ such that

$$\left| u_{j_2} - \frac{\lambda}{\sqrt{n}} \right| = |w_{j_2}| \leq \frac{\rho_0}{\sqrt{n - m_0}} \leq \frac{1}{5\sqrt{m_0}},$$

where we have used the fact that $m_0 = o(n)$ and the definition of ρ_0. Now using the triangle inequality we have

$$|u_{j_1} - u_{j_2}| \geq \frac{1}{4\sqrt{m_0}}. \tag{3.20}$$

To complete the proof of the lemma we then apply Lemma 3.1 with $J_1 = \{j_1\}$, $J_2 = \{j_2\}$, $m = 1$ and $\rho = \frac{1}{4}\sqrt{n/m_0}$. Recalling that $u \in \mathrm{Flat}_0(m_0, \rho_0)$ was abitrary, we conclude the bound

$$\sup_{u \in \mathrm{Flat}_0(m_0, \rho_0)} \mathbb{P}\left\{ \|(S_n^d + Z_n)u\|_2 \leq \frac{c_{3.1}}{4}\sqrt{\frac{d}{m_0}} \right\} \leq e^{-\bar{c}_{3.1} d}, \tag{3.21}$$

where we also use the fact that $d \leq n$.

Now by Lemma 3.5 we may fix a ρ_0-net $\Sigma_0(m_0, \rho_0) \subset \mathrm{Flat}_0(m_0, \rho_0)$ for $\mathrm{Flat}_0(m_0, \rho_0)$ of cardinality at most $(C_{3.5}n/m_0\rho_0^2)^{m_0}$. On the event $\mathcal{E}_K(m_0, \rho_0)$ we have $\|(S_n^d + Z_n)v\|_2 \leq \rho_0 K\sqrt{d}$ for some $v \in \mathrm{Flat}_0(m_0, \rho_0)$. Letting $u \in \Sigma_0(m_0, \rho_0)$ such that $\|u - v\|_2 \leq \rho_0$, by the triangle inequality we have

$$\|(S_n^d + Z_n)u\|_2 \leq \|(S_n^d + Z_n)v\|_2 + \|(S_n^d + Z_n)(u-v)\|_2 \leq \rho_0 K\sqrt{d} + \rho_0\|S_n^d + Z_n\|_{\langle \mathbf{1}\rangle^\perp} \leq 2\rho_0 K\sqrt{d},$$

where in the last step we have used the fact that $\mathcal{E}_K(m_0, \rho_0) \subset \mathcal{B}(K)$. Thus, by the union bound,

$$\mathbb{P}\{\mathcal{E}_K(m_0, \rho_0)\} \leq \sum_{u \in \Sigma_0(m_0, \rho_0)} \mathbb{P}\left\{\|(S_n^d + Z_n)u\|_2 \leq 2\rho_0 K\sqrt{d}\right\}.$$

We choose $c_{3.4}$ such that $c_{3.4} \leq c_{3.1}/2$ and hence $2\rho_0 K\sqrt{d} \leq c_{3.1}\sqrt{\frac{d}{m_0}}$. Therefore, by (3.21),

$$\begin{aligned}
\mathbb{P}\{\mathcal{E}_K(m_0, \rho_0)\} &\leq |\Sigma_0(m_0, \rho_0)|e^{-\overline{c}_{3.1}d} \leq \left(\frac{C_{3.5}n}{m_0\rho_0^2}\right)^{m_0} e^{-\overline{c}_{3.1}d} \\
&= \left(\frac{C_{3.5}nK^2}{c_{3.4}^2}\right)^{m_0} e^{-\overline{c}_{3.1}d} \\
&\leq \exp\left((1 + 2\gamma)m_0\log n + m_0\log(C_{3.5}/c_{3.4}^2) - \overline{c}_{3.1}d\right) \\
&\leq \exp\left(-\frac{\overline{c}_{3.1}}{3}d\right),
\end{aligned}$$

where in the last step we choose $c_{3.4}'$ suffciently small. The proof of the lemma thus completes. $\qquad\square$

3.3 Proof of Proposition 2.5

In this subsection we upgrade the weak control on flat vectors obtained in Lemma 3.4 to obtain Proposition 2.5 by iterative application of Lemma 3.7 below. The idea is that once we have shown $S_n^d + Z_n$ is well-invertible over $\mathrm{Flat}(m_0, \rho_0)$ for some small value of $m_0 \in [n]$ we can exploit the improved anti-concentration properties of vectors in $\mathbb{S}^{n-1}\backslash\mathrm{Flat}(m_0, \rho_0)$. (Here and in the sequel, by saying that a matrix A is well-invertible over a subspace B we mean that with high probability a good lower bound on $\|Au\|_2$ holds for all $u \in B$.) This allows us to beat the increased metric entropy cost for $\mathrm{Flat}(m_1, \rho_1)$ for some $m_1 > m_0$ that exceeds m_0 by a factor (essentially) d, and some $\rho_1 > 0$ somewhat smaller than ρ_0. We can iterate this roughly $\log_d n$ times to obtain control on $\mathrm{Flat}(m, \rho)$ with m essentially size n (up to log corrections). A similar iterative approach was used in the sparse i.i.d. setup in [25] (with the sets $\mathrm{Flat}(m_0, \rho_0)$ replaced by sets of vectors lying close to m_0-sparse vectors).

For deducing the improved anti-concentration properties as we increment the parameter m we will need the following lemma from [16].

Lemma 3.6 (Locating a bimodal component, cf. [16, Lemma 3.5]). *Let* $u \in \mathbb{S}^{n-1}\backslash\mathrm{Flat}(m^\star, \rho)$. *There exist disjoint sets* $J_1, J_2 \subset [n]$ *such that* $|J_1| \geq m^\star$, $|J_2| \geq c_{3.6}(n - m^\star)$ *and*

$$|u_{j_1} - u_{j_2}| \geq \frac{\rho}{4\sqrt{n}} \qquad \forall\, j_1 \in J_1,\, j_2 \in J_2, \tag{3.22}$$

where $c_{3.6} > 0$ *is some absolute constant.*

Lemma 3.7 (Incrementing control on flat vectors). *There exists absolute constants* $c_{3.7}, c_{3.7}', \overline{c}_{3.7} > 0$ *such that the following holds. Let* $\gamma \geq 1$ *and* $1 \leq K \leq n^\gamma$. *Assume* $1 \leq d \leq n$. *Let*

$$e^{-\gamma\log^2 n} \leq \rho^\star < 1 \qquad \text{and} \qquad 1 \leq m^\star \leq \min\left(\frac{1}{d}, \frac{c_{3.6}}{1 + c_{3.6}}\right)n \tag{3.23}$$

and let m', ρ' satisfy

$$m^\star < m' \le \left(\frac{c'_{3.7} d}{\gamma \log^2 n} \right) m^\star, \qquad 0 < \rho' \le \left(\frac{c_{3.7} \sqrt{m^\star d}}{Kn} \right) \rho^\star. \qquad (3.24)$$

Then

$$\mathbb{P} \left\{ \mathcal{E}_K(m', \rho') \backslash \mathcal{E}_K(m^\star, \rho^\star) \right\} \le \exp\left(-\overline{c}_{3.7} m^\star d \right). \qquad (3.25)$$

Proof. Let $m^\star, m', \rho^\star, \rho'$ be as in the statement of the lemma (note that the lemma holds vacuously for $d \le \log^2 n$ by the assumptions (3.24)). Since the event $\mathcal{E}_K(m, \rho)$ is monotone in the parameters m, ρ, we may and will assume that the upper bounds (3.24) hold with equality.

First we will argue

$$\sup_{u \in \mathrm{Flat}_0(m', \rho') \backslash \mathrm{Flat}_0(m^\star, \rho^\star)} \mathbb{P} \left\{ \|(S_n^d + Z_n) u\|_2 \le \frac{c_{3.1} \rho^\star}{4} \sqrt{\frac{m^\star d}{n}} \right\} \le \exp\left(-\overline{c}_{3.1} m^\star d \right). \quad (3.26)$$

Indeed, consider an arbitrary fixed element $u \in \mathrm{Flat}_0(m', \rho') \backslash \mathrm{Flat}_0(m^\star, \rho^\star)$. Note that

$$\mathrm{Flat}_0(m', \rho') \backslash \mathrm{Flat}_0(m^\star, \rho^\star) = \mathbb{S}^{n-1} \cap \langle \mathbf{1} \rangle^\perp \cap \mathrm{Flat}(m', \rho') \cap \mathrm{Flat}(m^\star, \rho^\star)^c$$
$$\subset \mathbb{S}^{n-1} \backslash \mathrm{Flat}(m^\star, \rho^\star).$$

By the assumed upper bound on $m^\star$ we can apply Lemma 3.6 to obtain disjoint sets $J_1, J_2 \subset [n]$ with $|J_1| \ge m^\star$, $|J_2| \ge c_{3.6}(n - m^\star) \ge m^\star$, such that

$$|u_{j_1} - u_{j_2}| \ge \frac{\rho^\star}{4\sqrt{n}} \qquad \forall j_1 \in J_1, \ j_2 \in J_2. \qquad (3.27)$$

By deleting elements from J_1 and J_2 we may assume $|J_1| = |J_2| = m^\star$. Now we apply Lemma 3.1 to obtain

$$\mathbb{P} \left\{ \|(S_n^d + Z_n) u\|_2 \le \frac{c_{3.1} \rho^\star}{4} \sqrt{\frac{m^\star d}{n}} \right\} \le \exp\left(-\overline{c}_{3.1} m^* d \right)$$

where we have used the fact that $m^\star d \le n$. Since u was arbitrary, (3.26) follows.

As in the proof of Lemma 3.4 we conclude by application of the union bound. Indeed, using Lemma 3.5 we fix a ρ'-net $\Sigma'_0 \subset \mathrm{Flat}_0(m', \rho')$ for $\mathrm{Flat}_0(m', \rho')$ with $|\Sigma'_0| \le (C_{3.5} n / m' \rho'^2)^{m'}$. By similar reasoning as in the proof of Lemma 3.4, on the event $\mathcal{E}_K(m', \rho')$, there exists $u \in \Sigma'_0$ such that $\|(S_n^d + Z_n) u\|_2 \le 2\rho' K \sqrt{d}$. Since $d \le n$, choosing $c_{3.7}$ sufficiently small we also have that $2\rho' K \sqrt{d} \le (c_{3.1} \rho^\star / 4) \sqrt{m^\star d / n}$. Therefore, applying the union bound and (3.26) we deduce,

$$\mathbb{P} \left\{ \mathcal{E}_K(m', \rho') \backslash \mathcal{E}_K(m^\star, \rho^\star) \right\} \le \sum_{u \in \Sigma'_0} \mathbb{P} \left(\mathcal{E}_K(m^\star, \rho^\star)^c \cap \left\{ \|(S_n^d + Z_n) u\|_2 \le 2\rho' K \sqrt{d} \right\} \right)$$
$$\le \sum_{u \in \Sigma'_0 \backslash \mathrm{Flat}_0(m^\star, \rho^\star)} \mathbb{P} \left(\|(S_n^d + Z_n) u\|_2 \le 2\rho' K \sqrt{d} \right)$$
$$\le \left(\frac{C_{3.5} n}{m' \rho'^2} \right)^{m'} \exp\left(-\overline{c}_{3.1} m^\star d \right)$$
$$\le \exp\left(m' \left(\log(n^3 K^2) + 2 \log \frac{1}{\rho^\star} + \log\left(\frac{C_{3.5}}{c'^2_{3.7} c^2_{3.7}} \right) \right) - \overline{c}_{3.1} m^\star d \right).$$

Since $K \le n^\gamma$, and $\rho^\star$ and m' satisfies (3.23) and (3.24) respectively we further obtain that

$$\mathbb{P} \left\{ \mathcal{E}_K(m', \rho') \backslash \mathcal{E}_K(m^\star, \rho^\star) \right\} \le \exp\left(3 c'_{3.7} m^\star d - \overline{c}_{3.1} m^\star d \right).$$

Now we choose $c'_{3.7}$ sufficiently small to complete the proof of the lemma. $\square$

Proof of Proposition 2.5. We may and will assume throughout that n is sufficiently large depending on γ. In the sequel, we will frequently apply the observation that the events $\mathcal{E}_K(m, \rho)$ are monotone increasing in the parameters m and ρ.

For $k \geq 0$, set

$$m_k := \left(\frac{c_{2.5} d}{\gamma \log^2 n} \right)^k, \qquad \tilde{\rho}_k := n^{-10\gamma k}, \tag{3.28}$$

where $c_{2.5} := c'_{3.4} \wedge c'_{3.7}$, and denote

$$\mathcal{E}_k := \mathcal{E}_K(m_k, \tilde{\rho}_k).$$

Note that m_k is an increasing sequence by our assumption $d \geq \log^3 n$. From Lemma 3.4 and monotonicity of $\mathcal{E}_K(m, \cdot)$, we have

$$\mathbb{P}(\mathcal{E}_1) \leq e^{-\bar{c}_{3.4} d}. \tag{3.29}$$

Let $k^* \geq 0$ be such that

$$\frac{n}{d} \in [m_{k^*}, m_{k^*+1}). \tag{3.30}$$

From the definitions of k^* and m_k and using the fact that $d \geq \log^3 n$ we see that

$$k^* \leq \frac{C \log n}{\log d} \tag{3.31}$$

for a sufficiently large constant $C > 0$. By monotonicity of $\mathcal{E}_K(\cdot, \rho)$,

$$\mathcal{E}_K \left(\frac{n}{d}, \tilde{\rho}_{k^*+1} \right) \subset \mathcal{E}_{k^*+1}.$$

Thus, applying the union bound,

$$\mathbb{P} \left\{ \mathcal{E}_K \left(\frac{c_{2.5} n}{\gamma \log^2 n}, \tilde{\rho}_{k^*+2} \right) \right\} \leq \mathbb{P} \left(\mathcal{E}_K \left(\frac{c_{2.5} n}{\gamma \log^2 n}, \tilde{\rho}_{k^*+2} \right) \backslash \mathcal{E}_K \left(\frac{n}{d}, \tilde{\rho}_{k^*+1} \right) \right)$$
$$+ \mathbb{P}(\mathcal{E}_1) + \sum_{k=1}^{k^*} \mathbb{P}(\mathcal{E}_{k+1} \backslash \mathcal{E}_k) \tag{3.32}$$

where we interpret the last sum as zero if $k^* = 0$. From (3.31) we have

$$\tilde{\rho}_{k^*+1} = n^{-10(k^*+2)\gamma} \geq \exp\left(-\gamma \log^2 n \right),$$

for n sufficiently large. Thus, we can apply Lemma 3.7 with $m^\star = n/d$ and $\rho^\star = \tilde{\rho}_{k^*+1}$ to bound

$$\mathbb{P} \left\{ \mathcal{E}_K \left(\frac{c_{2.5} n}{\gamma \log^2 n}, \tilde{\rho}_{k^*+2} \right) \backslash \mathcal{E}_K \left(\frac{n}{d}, \tilde{\rho}_{k^*+1} \right) \right\}$$
$$\leq \mathbb{P} \left\{ \mathcal{E}_K \left(\frac{n}{d} \times \frac{c_{2.5} d}{\gamma \log^2 n}, \tilde{\rho}_{k^*+1} \times \frac{c_{3.7} \sqrt{(n/d) \times d}}{Kn} \right) \backslash \mathcal{E}_K \left(\frac{n}{d}, \tilde{\rho}_{k^*+1} \right) \right\}$$
$$\leq e^{-\bar{c}_{3.7} n} \leq e^{-\bar{c}_{3.7} d}. \tag{3.33}$$

For the case that $k^* \geq 1$, since

$$\frac{m_{k+1}}{m_k} \leq \frac{c'_{3.7} d}{\gamma \log^2 n}, \qquad \frac{\tilde{\rho}_{k+1}}{\tilde{\rho}_k} = n^{-10\gamma} \leq \frac{c_{3.7} \sqrt{m_k d}}{Kn}$$

we may similarly apply Lemma 3.7 to deduce

$$\mathbb{P}(\mathcal{E}_{k+1} \backslash \mathcal{E}_k) \leq e^{-\bar{c}_{3.7} d}, \tag{3.34}$$

for each $1 \leq k \leq k^*$. Combining (3.29) and (3.33)–(3.34), from (3.32) and our assumption $d \geq \log^3 n$ we conclude

$$\mathbb{P}\left\{ \mathcal{E}_K\left(\frac{c_{2.5} n}{\gamma \log^2 n}, \tilde{\rho}_{k^*+2} \right) \right\} \leq 4k^* e^{-cd} \leq e^{-cd/2},$$

where c is a sufficiently small positive constant. From (3.31) we have $\tilde{\rho}_{k^*+2} \geq n^{-C'\gamma \log_d n}$ for a sufficiently large constant $C' > 0$. This completes the proof of the proposition. $\quad\square$

4 Invertibility over non-flat vectors

Having shown that $S_n^d + Z_n$ is well-invertible over vectors in $\mathrm{Flat}_0(m, \rho)$ with m essentially of size n (up to log factors), it remains to control the infimum of $\|(S_n^d + Z_n)u\|_2$ over the non-flat vectors $u \in \mathbb{S}^{n-1} \cap \langle \mathbf{1} \rangle^{\perp} \cap \mathrm{Flat}(m, \rho)^c$. The metric entropy of non-flat vectors is too large to take union bounds, so a different approach must be used for reducing to consideration of $(S_n^d + Z_n)u$ for a fixed vector u. We follow [36] by using an averaging argument, which in the setting of i.i.d. matrices reduces the problem to consideration of a dot product $R_i \cdot u$ for a single row vector R_i and a unit vector u that is orthogonal to the span of the remaining rows (and hence may be treated as fixed).

In the present setting, in order to use random transpositions we must consider a fixed pair of rows R_{i_1}, R_{i_2} and the dot product $(R_{i_1} - R_{i_2}) \cdot u$. Here u is a unit vector that is (almost) orthogonal to the remaining $n - 2$ vectors as well as $R_{i_1} + R_{i_2}$. The lack of independence between the rows makes the argument considerably more delicate than in [38]. In particular, the vectors R_{i_1}, R_{i_2} and u all depend on the rows $\{R_i : i \neq i_1, i_2\}$, and we want to avoid the event that, after conditioning on these $n - 2$ rows, the vector u is not flat on the supports of R_{i_1} and R_{i_2}. To overcome this we will adapt an argument of Litvak et al. that was used to bound the singularity probability for adjacency matrices of random regular digraphs [29]. Specifically, we define "good overlap events" $\mathcal{O}_{i_1, i_2}$ on which we may select an appropriate (almost-) normal vector u that has "high variation" on the supports of R_{i_1}, R_{i_2}, see Definition 4.3. In Lemma 4.4 we show that, if we restrict to the events that

1. $S_n^d + Z_n$ is well-invertible over flat vectors, and

2. S_n^d has *no holes* in the sense that the nonzero entries are uniformly distributed in all sufficiently large submatrices,

then the events $\mathcal{O}_{i_1, i_2}$ hold for a constant proportion of pairs $i_1, i_2 \in [n]$. Event (1) holds with high probability by Proposition 2.5, while the no-holes property (2) is shown to hold with high probability in Section 4.1. We can then restrict to $\mathcal{O}_{i_1, i_2}$ for some fixed i_1, i_2 by an averaging argument, at which point we can control the dot product $(R_{i_1} - R_{i_2}) \cdot u$ using a Berry–Esséen-type bound. As with the previous section, the arguments are similar to those in the work [16] for random regular digraphs, but differ in many particulars due to the different nature of the distribution of S_n^d.

4.1 The no-holes property

In the graph theory literature, a graph is said to enjoy a *discrepancy property* if the number of edges between all sufficiently large pairs of vertex sets U, V is roughly $\delta |U||V|$, where δ is the density of the graph. In terms of the adjacency matrix this says that all sufficiently large submatrices have roughly the same density. We will need a one-sided version of this property, called the *no-holes property*, to hold for S_n^d with high probability – namely, that all sufficiently large submatrices have density at least half of the expected value. In fact, we will need this property to hold for all matrices

$\{S_n^T : T \subset [d]\}$ obtained by summing only the permutation matrices P_n^ℓ with $\ell \in T$. (Note that S_n^T can be interpreted as the adjacency matrix for a random regular directed multigraph.)

For $L \subset [d]$ and $I, J \subset [n]$, write

$$e_L(I, J) := \sum_{\ell \in L} \sum_{i \in I} \mathbb{I}(\pi_n^\ell(i) \in J). \tag{4.1}$$

Since the permutations π_n^ℓ have uniform distribution, by linearity of the expectation,

$$\mathbb{E} e_L(I, J) = \frac{1}{n} |L||I||J|. \tag{4.2}$$

For $k_0 \in [d]$, $n_0 \in [n]$ we define the *no-holes event*

$$\mathcal{D}(k_0, n_0) := \bigcap_{\substack{L \subset [d]: \\ |L| \geq k_0}} \bigcap_{\substack{I, J \subset [n]: \\ |I|, |J| \geq n_0}} \left\{ e_L(I, J) \geq \frac{|L||I||J|}{2n} \right\}. \tag{4.3}$$

(This event actually only enforces a one-sided discrepancy property.)

Lemma 4.1 (No-holes property). Assume $1 \leq d \leq 10n$. If $k_0 n_0^2 \geq C_{4.1} n^2$ for a sufficiently large absolute constant $C_{4.1} > 0$, then

$$\mathbb{P}\left(\mathcal{D}(k_0, n_0)\right) \geq 1 - e^{-n}. \tag{4.4}$$

Proof. The proof follows from a result of [17] upon taking the union bound. Indeed, from [17, Theorem 1.13] we have that for any fixed $L \subset [d], I, J \subset [n]$,

$$\mathbb{P}\left(e_L(I, J) \leq \frac{|L||I||J|}{2n}\right) \leq 2 \exp\left(-\frac{1}{10n}|L||I||J|\right). \tag{4.5}$$

Combining this with the union bound,

$$\mathbb{P}(\mathcal{D}(k_0, n_0)^c) = \mathbb{P}\left\{\exists L \subset [d], I, J \subset [n] : |L| \geq k_0, |I|, |J| \geq n_0, \ e_L(I, J) \leq \frac{|L||I||J|}{2n}\right\}$$
$$\leq 2^{d+1} 4^n \exp\left(-\frac{k_0 n_0^2}{10n}\right).$$

Since $d \leq n$ the result immediately follows. $\qquad\square$

Remark 4.2. It is interesting to note that the dual property that S_n^d has no *dense* patches with high probability was a crucial ingredient in the work of Kahn–Szemerédi [22] on the mirror problem of proving an upper tail bound for the *second largest* singular value of S_n^d (i.e. the operator norm of the centered matrix $S_n^d - \frac{d}{n} \mathbf{1} \mathbf{1}^*$).

4.2 Good overlap via an averaging argument

In this and the next subsection we make use of the following notation: for distinct $i_1, i_2 \in [n]$ we denote

$$\mathcal{F}(i_1, i_2) := \left\langle \{\pi_n^\ell(i) : i \neq i_1, i_2\} \right\rangle \tag{4.6}$$

that is, the sigma algebra of events generated by all but the i_1-st and i_2-nd rows of each permutation matrix P_n^ℓ, $\ell \in [d]$.

Definition 4.3 (Good overlap events). For $i_1, i_2 \in [n]$ distinct, $\rho, t > 0$ and $k \geq 1$, we define the *good overlap event* $\mathcal{O}_{i_1,i_2}(k, \rho, t)$ to be the event that there exist $u \in \mathbb{S}^{n-1} \cap \langle \mathbf{1} \rangle^\perp$ and $L \subset [d]$ with $|L| \geq k$ such that the following properties hold:

(a) $\left| u_{\pi_n^\ell(i_1)} - u_{\pi_n^\ell(i_2)} \right| \geq \frac{\rho}{\sqrt{n}}$ for all $\ell \in L$,

(b) $\left\| (S_n^d + Z_n)^{(i_1, i_2)} u \right\|_2 \leq \frac{t}{\sqrt{n}}$, and

(c) $\left| (R_{i_1} + R_{i_2}) \cdot u \right| \leq \frac{2t}{\sqrt{n}}$.

Here $(S_n^d + Z_n)^{(i_1, i_2)}$ denotes the matrix obtained by removing rows i_1, i_2 from $S_n^d + Z_n$. We note that the event $\mathcal{O}_{i_1, i_2}(k, \rho, t)$ is $\mathcal{F}(i_1, i_2)$-measurable. Indeed, conditioning on $\mathcal{F}(i_1, i_2)$ fixes the $(S_n^d + Z_n)^{(i_1, i_2)}$ as well as the pairs $\{\pi_n^\ell(i_1), \pi_n^\ell(i_2)\}_{\ell \in [d]}$, and the latter determine the vector $R_{i_1} + R_{i_2}$ and the differences $\{|u_{\pi_n^\ell(i_1)} - u_{\pi_n^\ell(i_2)}|\}_{\ell \in [d]}$.

For each pair of distinct indices $i_1, i_2 \in [n]$ we choose an $\mathcal{F}(i_1, i_2)$-measurable random vector $u^{(i_1, i_2)} \in \mathbb{S}^{n-1} \cap \langle \mathbf{1} \rangle^\perp$ and an $\mathcal{F}(i_1, i_2)$-measurable random set $L(i_1, i_2) \subset [d]$ which, on the event $\mathcal{O}_{i_1, i_2}(k, \rho, t)$, satisfy the stated properties (a)–(c) for u, L; off this event we define $u^{(i_1, i_2)}$ and $L(i_1, i_2)$ arbitrarily (but in an $\mathcal{F}(i_1, i_2)$-measurable way).

For $m \geq 1$ and $\rho, t > 0$ we define the "good" event that $(S_n^d + Z_n)$ is well-invertible over mean-zero flat vectors:

$$\mathcal{G}(m, \rho, t) := \left\{ \forall u, v \in \mathrm{Flat}_0(m, \rho), \ \min(\|(S_n^d + Z_n)u\|_2, \|(S_n^d + Z_n)^* v\|_2) > \frac{t}{\sqrt{n}} \right\}. \quad (4.7)$$

Lemma 4.4 (Good overlap on average). Assume $d \geq 1$ and let $1 \leq m \leq \frac{c_{3.6}}{1 + c_{3.6}} n$. For all $\rho > 0$ and $0 < t \leq |d + \zeta| \sqrt{n}$,

$$\mathbb{P}\left(\left\{ s_n(S_n^d + Z_n) < \frac{t}{\sqrt{n}} \right\} \cap \mathcal{G}(m, \rho, t) \cap \mathcal{D}\left(\frac{c_{4.4} m d}{n}, \frac{m}{4} \right) \right)$$

$$\leq \frac{2}{mn} \sum_{i_1, i_2 = 1}^{n} \mathbb{P}\left(\mathcal{O}_{i_1, i_2}\left(\frac{c_{4.4} m d}{n}, \frac{\rho}{4}, t \right) \cap \left\{ |(R_{i_1} - R_{i_2}) \cdot u^{(i_1, i_2)}| \leq \frac{8t}{\rho} \right\} \right) \quad (4.8)$$

for some absolute constant $c_{4.4} > 0$.

Remark 4.5. The condition $t \leq |d + \zeta| \sqrt{n}$ is needed in order to bypass the possibility that $\mathbf{1}$ is an approximate minimal singular eigenvector of $S_n^d + Z_n$. This can be best seen if one chooses $\zeta = -d$.

Proof of Lemma 4.4. Suppose the event on the left hand side of (4.8) holds. Let $u, v \in \mathbb{S}^{n-1}$ be the respective eigenvectors of $(S_n^d + Z_n)^*(S_n^d + Z_n)$, $(S_n^d + Z_n)(S_n^d + Z_n)^*$ with eigenvalue $(s_n(S_n^d + Z_n))^2$. By our assumptions on Z_n we have that $\mathbf{1}$ is also an eigenvector of these matrices with eigenvalue $|d + \zeta|^2$. Then since

$$s_n(S_n^d + Z_n) < \frac{t}{\sqrt{n}} \leq |d + \zeta|$$

by assumption, it follows that u and $\mathbf{1}$ are associated to distinct eigenvalues of $(S_n^d + Z_n)^*(S_n^d + Z_n)$ and hence $u \perp \mathbf{1}$; we similarly have that $v \perp \mathbf{1}$. We have thus located vectors $u, v \in \mathbb{S}^{n-1} \cap \langle \mathbf{1} \rangle^\perp$ such that

$$\|(S_n^d + Z_n)u\|_2, \ \|(S_n^d + Z_n)^* v\|_2 \leq \frac{t}{\sqrt{n}}. \quad (4.9)$$

Furthermore, by the restriction to $\mathcal{G}(m, \rho, t)$ we have that $u, v \in \mathbb{S}^{n-1} \cap \langle \mathbf{1} \rangle^\perp \cap \mathrm{Flat}(m, \rho)^c$.

In the first stage of the proof, we show that there is a large number of "good" pairs $(i_1, i_2) \in [n]^2$ such that (1) $|v_{i_1} - v_{i_2}|$ is reasonably large, and (2) $|u_{\pi_n^\ell(i_1)} - u_{\pi_n^\ell(i_2)}|$ is reasonably large for a large number of $\ell \in [d]$.

We begin with (2), counting pairs (i_1, i_2) that are "good" with respect to u. Since $u \in \mathbb{S}^{n-1} \backslash \operatorname{Flat}(m, \rho)$, by Lemma 3.6 there exist disjoint sets $J_1, J_2 \subset [n]$ with $|J_1| = m$ and

$$|J_2| \geq c_{3.6}(n - m) \geq \frac{c_{3.6}}{1 + c_{3.6}} n \geq m \tag{4.10}$$

such that

$$|u_{j_1} - u_{j_2}| \geq \frac{\rho}{4\sqrt{n}} \quad \forall j_1 \in J_1, \; j_2 \in J_2. \tag{4.11}$$

For $i \in [n]$ and $\alpha \in \{1, 2\}$, write

$$L_\alpha(i) := \{\ell \in [d] : \pi_n^\ell(i) \in J_\alpha\}.$$

Fixing $c_{4.4} < \frac{c_{3.6}}{4(1 + c_{3.6})}$, define

$$\mathcal{I}(u) := \left\{ (i_1, i_2) \in [n]^2 : |L_1(i_1) \cap L_2(i_2)| > \frac{c_{4.4} dm}{n} \right\}. \tag{4.12}$$

We will use our restriction to the no-holes event $\mathcal{D}(c_{4.4} md/n, m/4)$ to show that $\mathcal{I}(u)$ is large. First, let

$$I_1 := \left\{ i \in [n] : |L_1(i)| \geq \frac{dm}{2n} \right\}.$$

We claim

$$|I_1| > n - \frac{m}{4}. \tag{4.13}$$

Indeed, suppose $|I_1^c| \geq m/4$. By our restriction to $\mathcal{D}(c_{4.4} md/n, m/4)$ and the fact that $|J_1| = m > m/4$, we have

$$\frac{d|I_1^c|m}{2n} \leq e_{[d]}(I_1^c, J_1) = \sum_{i \in I_1^c} |L_1(i)| < \frac{dm|I_1^c|}{2n},$$

a contradiction. Hence, (4.13) holds. Now for $i_1 \in [n]$ let

$$I_2(i_1) := \left\{ i \in [n] : |L_1(i_1) \cap L_2(i)| \geq \frac{c_{4.4} dm}{n} \right\}.$$

We claim that for any $i_1 \in I_1$,

$$|I_2(i_1)| > n - \frac{m}{4}. \tag{4.14}$$

Indeed, suppose towards a contradiction that $|I_2(i_1)^c| \geq m/4$ for some $i_1 \in I_1$. From (4.10) we have $|J_2| \geq m$, so by our restriction to $\mathcal{D}(c_{4.4} md/n, m/4)$,

$$\frac{|L_1(i_1)||I_2(i_1)^c||J_2|}{2n} \leq e_{L_1(i_1)}(I_2(i_1)^c, J_2) = \sum_{i \in I_2(i_1)^c} |L_1(i_1) \cap L_2(i)| < |I_2(i_1)^c| \frac{c_{4.4} dm}{n},$$

which rearranges to

$$|L_1(i_1)||J_2| < 2c_{4.4} dm.$$

Since $|J_2| \geq \frac{c_{3.6}}{1 + c_{3.6}} n$ and $c_{4.4} < \frac{c_{3.6}}{4(1 + c_{3.6})}$, we have $|L_1(i_1)| < dm/2n$, which contradicts the fact that $i_1 \in I_1$. This establishes (4.14). From (4.13) and (4.14) it follows that

$$|\mathcal{I}(u)| \geq |\{(i_1, i_2) : i_1 \in I_1, i_2 \in I_2(i_1)\}| > \left(n - \frac{m}{4} \right)^2 > n^2 - \frac{mn}{2}. \tag{4.15}$$

Now we count pairs that are "good" with respect to v. For $i_1 \in [n]$ write

$$J_v(i_1) := \left\{ i \in [n] : |v_{i_1} - v_i| \geq \frac{\rho}{\sqrt{n}} \right\}.$$

Since $v \in \mathbb{S}^{n-1} \setminus \mathrm{Flat}(m, \rho)$ we must have that $|J_v(i_1)| > m$ for any $i_1 \in [n]$. Indeed, suppose $|J_v(i_1)| \leq m$ for some $i_1 \in [n]$. Denoting $w := (v - v_{i_1} \mathbf{1})_{J_v(i_1)}$ (for any vector v' and $J' \in [n]$ we write $v'_{J'}$ to denote the projection of the vector v' onto coordinates indexed by J'), we have

$$\|v - v_{i_1} \mathbf{1} - w\|_2 = \|(v - v_{i_1} \mathbf{1})_{J_v(i_1)^c}\|_2 < \rho.$$

But since $w \in \mathrm{Sparse}(m)$ this contradicts the assumption that $v \notin \mathrm{Flat}(m, \rho)$. Thus, putting

$$\tilde{\mathcal{I}}(v) := \left\{ (i_1, i_2) \in [n]^2 : |v_{i_1} - v_{i_2}| \geq \frac{\rho}{\sqrt{n}} \right\}$$

we have $|\tilde{\mathcal{I}}(v)| = \sum_{i_1 \in [m]} |J_v(i_1)| \geq nm$. Set

$$\mathcal{I}'(u, v) := \mathcal{I}(u) \cap \tilde{\mathcal{I}}(v).$$

Using the bound (4.15) we have

$$|\mathcal{I}'(u, v)| \geq |\tilde{\mathcal{I}}(v)| - |\tilde{\mathcal{I}}(v) \setminus \mathcal{I}(u)| \geq |\tilde{\mathcal{I}}(v)| - |\mathcal{I}(u)^c| \geq mn - \frac{mn}{2} = \frac{mn}{2}. \qquad (4.16)$$

Now we show that $\mathcal{O}_{i_1, i_2}(c_{4.4} md/n, \rho/4, t)$ holds for all $(i_1, i_2) \in \mathcal{I}'(u, v)$ (in fact it holds for all $(i_1, i_2) \in \mathcal{I}(u)$). Indeed, the vector u and the set $L = L_1(i_1) \cap L_2(i_2)$ witness the conditions (a)–(c) from Definition 4.3, as we now demonstrate. The condition that $|L| \geq c_{4.4} md/n$ follows from the definition of $\mathcal{I}(u)$. The condition (a) follows from (4.11) and the definitions of $L_1(i_1), L_2(i_2)$. Finally, (b) and (c) follow easily from (4.9) and the triangle inequality:

$$\|(S_n^d + Z_n)^{(i_1, i_2)} u\|_2 \leq \|(S_n^d + Z_n) u\|_2 \leq \frac{t}{\sqrt{n}},$$

$$|(R_{i_1} + R_{i_2}) \cdot u| \leq |R_{i_1} \cdot u| + |R_{i_2} \cdot u| \leq 2\|(S_n^d + Z_n) u\|_2 \leq \frac{2t}{\sqrt{n}}.$$

A key point here is that while u and $L = L_1(i_1) \cap L_2(i_2)$ witness that the event $\mathcal{O}_{i_1, i_2}$ $(c_{4.4} md/n, \rho/4, t)$ holds, we cannot take these to be $u^{(i_1, i_2)}$ and $L(i_1, i_2)$, respectively, as u and L are not themselves measurable with respect to $\mathcal{F}(i_1, i_2)$.

Now it remains to show that occurrence of all the events on the left hand side of (4.8) implies also the occurrence of the event $\{|(R_{i_1} - R_{i_2}) \cdot u^{(i_1, i_2)}| \leq \frac{8t}{\rho}\}$ for all $(i_1, i_2) \in \mathcal{I}'(u, v)$. By several applications of the Cauchy–Schwarz inequality and the fact that $\mathcal{O}_{i_1, i_2}(c_{4.4} md/n, \rho/4, t)$ holds, we have

$$\frac{t}{\sqrt{n}} \geq \|v^*(S_n^d + Z_n)\|_2 \geq |v^*(S_n^d + Z_n) u^{(i_1, i_2)}|$$

$$= \left| \sum_{i=1}^n \overline{v}_i R_i \cdot u^{(i_1, i_2)} \right|$$

$$\geq |(\overline{v}_{i_1} R_{i_1} + \overline{v}_{i_2} R_{i_2}) \cdot u^{(i_1, i_2)}| - \|(S_n^d + Z_n)^{(i_1, i_2)} u^{(i_1, i_2)}\|_2$$

$$\geq |(\overline{v}_{i_1} R_{i_1} + \overline{v}_{i_2} R_{i_2}) \cdot u^{(i_1, i_2)}| - \frac{t}{\sqrt{n}},$$

which implies that $|(\overline{v}_{i_1} R_{i_1} + \overline{v}_{i_2} R_{i_2}) \cdot u^{(i_1, i_2)}| \leq \frac{2t}{\sqrt{n}}$. Using the triangle inequality, recalling

the definition of $\tilde{\mathcal{I}}(v)$, and using the fact that $\max_i |v_i| \leq \|v\|_2 = 1$, we further obtain

$$
\begin{aligned}
&\left| (\overline{v}_{i_1} R_{i_1} + \overline{v}_{i_2} R_{i_2}) \cdot u^{(i_1,i_2)} \right| \\
&= \frac{1}{2} \left| (\overline{v}_{i_1} + \overline{v}_{i_2})(R_{i_1} + R_{i_2}) \cdot u^{(i_1,i_2)} + (\overline{v}_{i_1} - \overline{v}_{i_2})(R_{i_1} - R_{i_2}) \cdot u^{(i_1,i_2)} \right| \\
&\geq \frac{1}{2} \left| (\overline{v}_{i_1} - \overline{v}_{i_2})(R_{i_1} - R_{i_2}) \cdot u^{(i_1,i_2)} \right| - \left| (R_{i_1} + R_{i_2}) \cdot u^{(i_1,i_2)} \right| \\
&\geq \frac{1}{2} \left| (\overline{v}_{i_1} - \overline{v}_{i_2})(R_{i_1} - R_{i_2}) \cdot u^{(i_1,i_2)} \right| - \frac{2t}{\sqrt{n}} \\
&\geq \frac{\rho}{2\sqrt{n}} \left| (R_{i_1} - R_{i_2}) \cdot u^{(i_1,i_2)} \right| - \frac{2t}{\sqrt{n}},
\end{aligned}
$$

where in the second-to-last inequality we have used the property (c) of the event $\mathcal{O}_{i_1,i_2}(c_{4.4}md/n, \rho/4, t)$. Combining and rearranging we have

$$
\left| (R_{i_1} - R_{i_2}) \cdot u^{(i_1,i_2)} \right| \leq \frac{8t}{\rho}.
$$

We have thus shown that on the event $\mathcal{E} := \left\{ s_n(S_n^d + Z_n) \leq \frac{t}{\sqrt{n}} \right\} \cap \mathcal{G}(m, \rho, t) \cap \mathcal{D}\left(\frac{c_{4.4}md}{n}, \frac{m}{4} \right)$, the event $\mathcal{E}(i_1, i_2) := \mathcal{O}_{i_1,i_2}\left(\frac{cmd}{n}, \frac{\rho}{2}, t \right) \cap \left\{ \left| (R_{i_1} - R_{i_2}) \cdot u^{(i_1,i_2)} \right| \leq \frac{8t}{\rho} \right\}$ holds for at least $mn/2$ values of $(i_1, i_2) \in [n]^2$. By double counting,

$$
\sum_{i_1,i_2=1}^{n} \mathbb{I}_{\mathcal{E}(i_1,i_2)} \geq \frac{mn}{2} \mathbb{I}_{\mathcal{E}}.
$$

Taking expectations on both sides and rearranging yields the desired bound. $\qquad\square$

4.3 Anti-concentration for random walks

In the previous section we essentially reduced our task to obtaining an anti-concentration estimate for the random variable $(R_{i_1} - R_{i_2}) \cdot u^{(i_1,i_2)}$ for a fixed pair of distinct indices $i_1, i_2 \in [n]$. We accomplish this in the following lemma (recall our notation (4.6)).

Lemma 4.6 (Anti-concentration for row-pair random walk). Let $i_1, i_2 \in [n]$ be distinct, and suppose $\mathcal{O}_{i_1,i_2}(k, \rho, t)$ holds for some $k \geq 1$, $\rho, t > 0$. Then for all $r \geq 0$,

$$
\mathbb{P}\left\{ \left| (R_{i_1} - R_{i_2}) \cdot u^{(i_1,i_2)} \right| \leq r \;\Big|\; \mathcal{F}(i_1, i_2) \right\} \leq C_{4.6}\left(1 + \frac{r\sqrt{n}}{\rho} \right) \left(\frac{\log(n/\rho)}{k} \right)^{1/2} \tag{4.17}
$$

for some absolute constant $C_{4.6}$.

Remark 4.7. In the proof we will only use the lower bound $|L(i_1, i_2)| \geq k$ and property (a) for $u^{(i_1,i_2)}$ and $L(i_1, i_2)$ from Definition 4.3, which is why the bound is independent of the parameter t.

We will need the following standard anti-concentration bound of Berry–Esséen-type; see for instance [15, Lemma 2.7] (the condition there of κ-controlled second moment is easily verified to hold with $\kappa = 1$ for a Rademacher variable).

Lemma 4.8 (Berry–Esséen-type small-ball inequality). Let $v \in \mathbb{C}^n$ be a fixed nonzero vector and let $\xi_1, \ldots, \xi_n$ be independent Rademacher variables. There exists an absolute constant $C_{4.8}$ such that for any $r \geq 0$,

$$
\sup_{z \in \mathbb{C}} \mathbb{P}\left(\left| z + \sum_{j=1}^{n} \xi_j v_j \right| \leq r \right) \leq C_{4.8}\left(\frac{r + \|v\|_\infty}{\|v\|_2} \right).
$$

Proof of Lemma 4.6. By symmetry we may take $(i_1, i_2) = (1, 2)$. Condition on a realization of $\{\pi^\ell(i) : i \notin \{1, 2\}, 1 \le \ell \le d\}$ such that $\mathcal{O}_{i_1, i_2}(k, \rho, t)$ holds. This fixes the vector $u^{(1,2)}$ and the set $L(1, 2) \subset [d]$. For ease of notation we write $u = u^{(1,2)}$ and $L = L(1, 2)$ for the remainder of the proof. Let $r \ge 0$. Our aim is to show

$$\mathbb{P}\left(\left| (R_1 - R_2) \cdot u \right| \le r \mid \mathcal{F}(1, 2) \right) \le C \left(1 + \frac{r\sqrt{n}}{\rho} \right) \left(\frac{\log(n/\rho)}{k} \right)^{1/2} \qquad (4.18)$$

for some sufficiently large constant C. Let $\xi_1, \ldots, \xi_d$ be i.i.d. Rademacher variables, independent of all other variables, and for each $\ell \in [d]$ put

$$\widetilde{\pi}_n^\ell := \pi_n^\ell \circ \tau_{(1,2)}^{\frac{1}{2}(\xi_\ell + 1)}$$

where we recall $\tau_{(i_1, i_2)}$ denotes the transposition that switches i_1, i_2, and we interpret $\tau_{(i_1, i_2)}^1 = \tau_{(i_1, i_2)}$, $\tau_{(i_1, i_2)}^0 = \mathrm{Id}$. Now let $\widetilde{S}_n^d$ be as in (1.2) but with each π_n^ℓ replaced by $\widetilde{\pi}_n^\ell$. By the Haar distribution of $\pi_n^1, \ldots, \pi_n^d$ and their independence from the Rademacher variables $\xi_1, \ldots, \xi_d$, we have that $\widetilde{S}_n^d \overset{d}{=} S_n^d$, even under conditioning on $\mathcal{F}(i_1, i_2)$. Moreover, it is clear from the construction that $\widetilde{\pi}_n^\ell(i) = \pi_n^\ell(i)$ for all $3 \le i \le n$ and $1 \le \ell \le d$, so that $\widetilde{S}_n^d$ agrees with S_n^d on the third through n-th rows. We denote the first two rows of $\widetilde{S}_n^d$ by $\widetilde{R}_1$ and $\widetilde{R}_2$. By replacing S_n^d with $\widetilde{S}_n^d$ in (4.18), it now suffices to show

$$\mathbb{P}\left(\left| (\widetilde{R}_1 - \widetilde{R}_2) \cdot u \right| \le r \mid \mathcal{F}(1, 2) \right) \le C \left(1 + \frac{r\sqrt{n}}{\rho} \right) \left(\frac{\log(n/\rho)}{k} \right)^{1/2}. \qquad (4.19)$$

Turning to prove (4.19) we note

$$
\begin{aligned}
(\widetilde{R}_1 - \widetilde{R}_2) \cdot u &= \sum_{\ell \in [d]} u_{\widetilde{\pi}_n^\ell(1)} - u_{\widetilde{\pi}_n^\ell(2)} \\
&= \sum_{\ell \in [d]} \left(u_{\pi^\ell(1)} - u_{\pi^\ell(2)} \right) \mathbb{I}(\xi_\ell = -1) - \left(u_{\pi^\ell(1)} - u_{\pi^\ell(2)} \right) \mathbb{I}(\xi_\ell = +1) \\
&= \sum_{\ell \in [d]} \xi_\ell \partial_\ell(u), \qquad (4.20)
\end{aligned}
$$

where

$$\partial_\ell(u) := u_{\pi^\ell(2)} - u_{\pi^\ell(1)}. \qquad (4.21)$$

For $j \ge -1$ let

$$L^{(j)} := \left\{ \ell \in L : 2^{-(j+1)} < |\partial_\ell(u)| \le 2^{-j} \right\}.$$

By condition (a) in Definition 4.3 we have that $|\partial_\ell(u)| \ge \rho/\sqrt{n}$ for all $\ell \in L$. Therefore,

$$L \subset \bigcup_{j=-1}^{\log_2(\sqrt{n}/\rho)} L^{(j)}.$$

Since $|L| = k$ by the pigeonhole principle there must exists some $j_\star$ such that

$$|L^{(j_\star)}| \ge \frac{k}{2 \log_2(\sqrt{n}/\rho)}.$$

Set

$$v := \left(\partial_\ell(u) \mathbb{I}(\ell \in L^{(j_\star)}) \right)_{\ell \in [d]} \in \mathbb{C}^d.$$

For all $\ell \in L^{(j)}$ we have $|v_\ell| \ge \rho/\sqrt{n}$ and so

$$\|v\|_2 \ge \frac{\rho}{\sqrt{n}} |L^{(j_\star)}|^{1/2} \ge \frac{\rho}{\sqrt{n}} \left(\frac{k}{2 \log_2(\sqrt{n}/\rho)} \right)^{1/2}. \qquad (4.22)$$

Moreover, since the components of v vary by at most a factor of 2 on $L^{(j_\star)}$ we also have $|v_\ell| \geq \|v\|_\infty/2$ for all $\ell \in L^{(j_\star)}$. Therefore

$$\|v\|_\infty \leq \frac{2\|v\|_2}{|L^{(j)}|^{1/2}} \leq \left(\frac{8\log_2(\sqrt{n}/\rho)}{k}\right)^{1/2} \|v\|_2. \tag{4.23}$$

Conditioning on $\{\pi^\ell : \ell \in [d]\}$ and applying Lemma 4.8, we have

$$\sup_{z \in \mathbb{C}} \mathbb{P}_{L^{(j_\star)}}\left(\left|z + \sum_{\ell \in L^{(j_\star)}} \xi_\ell \partial_\ell(u)\right| \leq r\right) = \sup_{z \in \mathbb{C}} \mathbb{P}_{L^{(j_\star)}}\left(\left|z + \sum_{\ell=1}^{d} \xi_\ell v_\ell\right| \leq r\right)$$

$$\leq C_{4.8}\left(\frac{r}{\|v\|_2} + \frac{\|v\|_\infty}{\|v\|_2}\right)$$

$$\leq C_{4.8}\left(1 + \frac{r\sqrt{n}}{\rho}\right)\left(\frac{8\log_2(\sqrt{n}/\rho)}{k}\right)^{1/2},$$

where $\mathbb{P}_{L^{(j_\star)}}$ denotes the law of $\{\xi_\ell\}_{\ell \in L^{(j_\star)}}$. Applying this bound to the expression (4.20) (after conditioning on $\{\xi_\ell : \ell \notin L^{(j)}\}$ and absorbing the resulting deterministic summands into the scalar z), we obtain (4.19) as desired. $\qquad\square$

4.4 Proof of Theorem 2.2

Now we combine the results of this section and Section 3 to complete the proof of Theorem 2.2. Fix $\gamma_0 \geq 1$ and let $\Gamma_0 = \overline{C}_{2.2}\gamma_0 \log_d n$ with $\overline{C}_{2.2}$ an absolute constant to be chosen sufficiently large. We may and will assume that n is sufficiently large depending on γ_0. By Remark 2.3 we may assume

$$\log^8 n \leq d \leq n \tag{4.24}$$

(the desired bound holds trivially for smaller values of d). Recall the boundedness event $\mathcal{B}(K)$ from (2.17). From our hypotheses and the fact that $\|S_n^d\|_{\langle \mathbf{1}\rangle^\perp} \leq \|S_n^d\| = d$ we have

$$\|S_n^d + Z_n\|_{\langle \mathbf{1}\rangle^\perp} \leq \|S_n^d\|_{\langle \mathbf{1}\rangle^\perp} + \|Z_n\|_{\langle \mathbf{1}\rangle^\perp} \leq d + n^{\gamma_0} \leq 2n^{\gamma_0} \leq n^{\gamma_0}\sqrt{d}.$$

Thus the event $\mathcal{B}(n^{\gamma_0})$ holds.

Set

$$m = \frac{c_{2.5}n}{\gamma_0 \log^2 n}, \qquad \rho = n^{-C_{2.5}\gamma_0 \log_d n}, \qquad t = \sqrt{n}(n^{-\Gamma_0} \wedge |d + \zeta|). \tag{4.25}$$

Now using Lemma 4.4 we have

$$\mathbb{P}\left(s_n(S_n^d + Z_n) < n^{-\Gamma_0} \wedge |d + \zeta|\right) \leq \mathbb{P}\left(\mathcal{G}(m,\rho,t)^c\right) + \mathbb{P}\left(\mathcal{D}\left(\frac{c_{4.4}md}{n}, \frac{m}{4}\right)^c\right)$$

$$+ \frac{2}{mn}\sum_{i_1,i_2=1}^{n}\mathbb{P}\left\{\mathcal{O}_{i_1,i_2}\left(\frac{c_{4.4}md}{n}, \frac{\rho}{4}, t\right) \cap \left\{\left|(R_{i_1} - R_{i_2})\cdot u^{(i_1,i_2)}\right| \leq \frac{8t}{\rho}\right\}\right\}. \tag{4.26}$$

Taking $\overline{C}_{2.2} \geq C_{2.5} + 1$ we have $t \leq \rho$. Then by Proposition 2.5 we see that

$$\mathbb{P}\left(\mathcal{G}(m,\rho,t)^c\right) \leq e^{-\overline{c}_{2.5}d}. \tag{4.27}$$

Using Lemma 4.1 and the lower bound in (4.24) (here we only need $d = \omega(\log^6 n)$), we see that

$$\mathbb{P}\left(\mathcal{D}\left(\frac{c_{4.4}md}{n}, \frac{m}{4}\right)^c\right) \leq e^{-n}. \tag{4.28}$$

Next applying Lemma 4.6 yields that the third term in (4.26) is bounded by

$$\frac{2}{mn} \times n^2 \times C_{4.6} \left(1 + \frac{128t\sqrt{n}}{\rho^2}\right) \left(\log n + \log \frac{4}{\rho}\right)^{1/2} \sqrt{\frac{n}{c_{4.4}md}}$$

$$\leq \frac{256C_{4.6}}{\sqrt{c_{4.4}}} \frac{1}{\sqrt{d}} \left(\frac{n}{m}\right)^{3/2} \left(1 + n^{-\Gamma_0 + 1 + 2C_{2.5}\gamma_0 \log_d n}\right) (2 + C_{2.5}\gamma_0 \log_d n)^{1/2} (\log n)^{1/2}$$

$$\leq \frac{C_{\gamma_0} \log^4 n}{\sqrt{d}} \left(1 + n^{-\Gamma_0 + 1 + 2C_{2.5}\gamma_0 \log_d n}\right) \tag{4.29}$$

for some constant C_{γ_0} depending only on γ_0. Taking $\overline{C}_{2.2} \geq 3C_{2.5}$ and combining (4.26)–(4.29) we conclude

$$\mathbb{P}\left(s_n(S_n^d + Z_n) < n^{-\Gamma_0} \wedge |d + \zeta|\right) \leq e^{-\bar{c}_{2.5}d} + e^{-n} + \frac{2C_{\gamma_0} \log^4 n}{\sqrt{d}}$$

$$\leq \frac{3C_{\gamma_0} \log^4 n}{\sqrt{d}}. \tag{4.30}$$

The proof of Theorem 2.2 is now complete.

5 Control on traces

In this short section, we derive simple estimates on traces for permutation matrices and for $S_n^d(S_n^d)^*$. We begin with the following simple estimate. Let π_n be a random, uniformly chosen permutation on $[n]$, and let P_n denote the corresponding permutation matrix.

Lemma 5.1. With notation as above,

$$\mathbb{P}(\operatorname{Tr} P_n \geq k) \leq \frac{1}{k!}, \quad k \geq 1. \tag{5.1}$$

Proof. Let N_ℓ denote the number of cycles of length ℓ in π_n. Note that $\operatorname{Tr} P_n = N_1$. Thus, the event $\{\operatorname{Tr} P_n \geq k\}$ is the union of the events that k particular indices are fixed points in the permutation π_n and therefore

$$\mathbb{P}(\operatorname{Tr} P_n \geq k) = \mathbb{P}(N_1 \geq k) \leq \binom{n}{k} \frac{1}{n \cdot (n-1) \cdots (n-k+1)} = \frac{1}{k!}. \tag{5.2}$$
$\square$

Let now S_n^d be as in (1.2). We have the following lemma.

Lemma 5.2. With notation as above, there exists absolute constants $c_{5.2}$, $C'_{5.2}$, and $C_{5.2}$ so that

$$\mathbb{P}(\operatorname{Tr} S_n^d(S_n^d)^* \geq nd + xd^2) \leq de^{-d(x-e)}, \quad x \geq e, \tag{5.3}$$

for any $d \geq C_{5.2}$. In particular, there exists an absolute constant $\overline{C}_{5.2}$ so that

$$\mathbb{E}\operatorname{Tr}(S_n^d(S_n^d)^*)^2 \leq 2nd^2 + \overline{C}_{5.2}d^4. \tag{5.4}$$

Proof. Note that

$$S_n^d(S_n^d)^* = dI_n + \sum_{i \neq j = 1}^{d} P_n^i(P_n^j)^*. \tag{5.5}$$

Therefore, using that $P_n^i(P_n^j)^*$ with $i \neq j$ is distributed like P_n, and that for fixed i they are independent of each other, we get from (5.1) that

$$\mathbb{P}(\operatorname{Tr} S_n^d(S_n^d)^* \geq nd + xd^2) \leq \mathbb{P}\left(\sum_{i \neq j=1}^{d} \operatorname{Tr} P_n^i(P_n^j)^* \geq xd^2\right) \leq d\,\mathbb{P}\left(\sum_{j=2}^{d} \operatorname{Tr} P_n^i(P_n^1)^* \geq xd\right)$$

$$\leq d\,\mathbb{P}\left(\sum_{i=1}^{d} \operatorname{Tr} P_n^i \geq xd\right). \tag{5.6}$$

From (5.1) we have that $\mathbb{E}(e^{\operatorname{Tr} P_n^i}) \leq e^e$, and therefore, by independence and Markov's inequality,

$$\mathbb{P}\left(\sum_{i=1}^{d} \operatorname{Tr} P_n^i \geq xd\right) \leq e^{-xd}e^{ed} = e^{-(x-e)d}. \tag{5.7}$$

Substituting in (5.6) we obtain that

$$\mathbb{P}(\operatorname{Tr} S_n^d(S_n^d)^* \geq nd + xd^2) \leq de^{-(x-e)d},$$

which completes the proof. $\qquad\square$

Note that Lemma 5.2 together with (5.1) imply that with $Q_n = (z - S_n^d/\sqrt{d})(z - S_n^d/\sqrt{d})^*$,

$$\mathbb{P}(\operatorname{Tr} Q_n > ((|z|^2 + 1)n + 2|z|\sqrt{d}x + dx)) \leq de^{-c'dx}, \tag{5.8}$$

for some absolute constant c', and d and x sufficiently large. Indeed,

$$\operatorname{Tr} Q_n \leq |z|^2 n + \frac{1}{d}\operatorname{Tr} S_n^d(S_n^d)^* + 2|z|\frac{1}{\sqrt{d}}\operatorname{Tr} S_n^d,$$

and the conclusion follows by a union bound and the estimates in (5.3) and (5.7).

6 Concentration for resolvent sub-traces

In this section we derive concentration bounds on the traces of the diagonal and the off-diagonal blocks of the resolvent $\widetilde{G}(S_n^d)$. To prove Theorem 2.6 we will need to consider the resolvent of S_n^d shifted by some deterministic matrices. Hence, we introduce the following notation. Let $M_n := M$ be a deterministic matrix of size $n \times n$. Fix $\xi \in \mathbb{C}\backslash\mathbb{R}$, $z \in \mathbb{C}$ and define

$$F^M(\xi) =: \begin{bmatrix} F_{11}^M(\xi) & F_{12}^M(\xi) \\ F_{21}^M(\xi) & F_{22}^M(\xi) \end{bmatrix} =: \widetilde{G}^M(S_n^d, \xi, z)$$

$$:= \left[\xi I - \begin{bmatrix} 0 & \left(z - \frac{S_n^d}{\sqrt{d}} + M\right) \\ \left(z - \frac{S_n^d}{\sqrt{d}} + M\right)^* & 0 \end{bmatrix}\right]^{-1}.$$

Theorem 6.1. Fix $z \in B_{\mathbb{C}}(0, R)$ and $\xi \in \mathbb{C}\backslash\mathbb{R}$ such that $|\operatorname{Im}\xi| \leq C_0$ for some $C_0 \geq 1$. Let $M_n := M$ of size $n \times n$ be an $n \times n$ deterministic matrix with $\|M\| \leq C_0$. Then, for $i, j = 1, 2$ and $u \geq 0$ we have

$$\mathbb{P}\left(\left|\frac{1}{n}\operatorname{Tr} F_{ij}^M(\xi) - \mathbb{E}\left[\frac{1}{n}\operatorname{Tr} F_{ij}^M(\xi)\right]\right| \geq u\right) \leq 4\exp\left(-c_{6.1}n(\operatorname{Im}\xi)^4 u^2\right)$$

for some constant $c_{6.1} > 0$, depending only on C_0.

The following is an immediate corollary of Theorem 6.1.

Corollary 6.2. With notation as in Theorem 6.1, there exists an n_0 so that if $\operatorname{Im}\xi > n^{-1/16}$ and $n > n_0$ then, for $i, j = 1, 2$,

$$\mathbb{E}\left[\left(\frac{1}{n}\operatorname{Tr} F_{ij}^M(\xi) - \mathbb{E}\left[\frac{1}{n}\operatorname{Tr} F_{ij}^M(\xi)\right]\right)^2\right] \leq \frac{1}{n^{3/4}}. \tag{6.1}$$

We first prove Corollary 6.2 using Theorem 6.1. The proof of Theorem 6.1 follows that.

Proof of Corollary 6.2. Let $Z := |\frac{1}{n}\operatorname{Tr} F_{ij}(\xi) - \mathbb{E}\left[\frac{1}{n}\operatorname{Tr} F_{ij}(\xi)\right]|$. Substituting $u = x/n^{1/4}$ in Theorem 6.1 gives that for $x > 0$ we have

$$\mathbb{P}(Z > u) \leq 4\exp\left(-c_{6.1}x^2 n^{1/4}\right).$$

This completes the proof upon using integration by parts. $\qquad\square$

We next establish Theorem 6.1, using a standard martingale approach. Specifically, we will apply a consequence of Azuma's inequality from [28] that is conveniently phrased for our setting. This will reduce the task to bounding the change in $n^{-1}\operatorname{Tr} F_{ij}^M(\xi)$ under the application of a transposition to one of the permutations π_n^ℓ.

Define the Hamming distance between two permutations $\pi, \sigma \in \mathbb{S}_n$ as follows:

$$d_H(\pi, \sigma) := |\{i \in [n] : \pi(i) \neq \sigma(i)\}|. \tag{6.2}$$

We extend to a Hamming metric on product space $\mathbb{S}_n^d$ in the natural way: for two sequences $\boldsymbol{\pi} = (\pi^\ell)_{\ell \in [d]}$, $\boldsymbol{\sigma} = (\sigma^\ell)_{\ell \in [d]}$, set

$$d_H(\boldsymbol{\pi}, \boldsymbol{\sigma}) := \sum_{\ell=1}^d d_H(\pi^\ell, \sigma^\ell). \tag{6.3}$$

Lemma 6.3 (Concentration for Hamming-Lipschitz functions)**.** Let $f : \mathbb{S}_n^d \to \mathbb{C}$ be an L-Lipschitz function with respect to the Hamming metric (6.3), and let $\boldsymbol{\pi} = (\pi^\ell)_{\ell \in [d]}$ be a uniform random element of $\mathbb{S}_n^d$. Then, for any $u \geq 0$,

$$\mathbb{P}(|f(\boldsymbol{\pi}) - \mathbb{E}f(\boldsymbol{\pi})| \geq u) \leq 4\exp\left(-\frac{u^2}{8ndL^2}\right). \tag{6.4}$$

Proof. First we note that it is enough to prove that (6.4) holds for 1-Lipschitz function. Next, splitting $f(\boldsymbol{\pi})$ into real and imaginary parts and applying the pigeonhole principle and the union bound, it suffices to show that for f a real-valued 1-Lipschitz function on $\mathbb{S}_n^d$,

$$\mathbb{P}(f(\boldsymbol{\pi}) - \mathbb{E}f(\boldsymbol{\pi}) \geq u) \leq \exp\left(-\frac{u^2}{8nd}\right). \tag{6.5}$$

By Chebycheff's inequality, (6.5) would follow if, for any $\lambda > 0$,

$$\mathbb{E}\left(e^{\lambda(f(\boldsymbol{\pi}) - \mathbb{E}f(\boldsymbol{\pi}))}\right) \leq e^{2nd\lambda^2}. \tag{6.6}$$

For $d = 1$, the inequality (6.6) follows as in the proof of [28, Corollary 4.3], using that in Lemma 4.1 there, one actually controls the Laplace transform and not just the probabilities. To prove the case of general d, we use tensorization. For an arbitrary 1-Lipschitz function $f : \mathbb{S}_n^d \mapsto \mathbb{R}$ and for $i \in [d]$, denote

$$h_i(\boldsymbol{\pi}^{<i+1}) := \mathbb{E}[f|\boldsymbol{\pi}^{<i+1}] - \mathbb{E}[f|\boldsymbol{\pi}^{<i}],$$

where we recall that $\boldsymbol{\pi}^{<k} := (\pi^\ell)_{\ell \in [k-1]}$. For any fixed $i \in [d]$ and $\boldsymbol{\pi}^{<i}$, the function h_i viewed as a function of π_i is a 1-Lipschitz function with respect to the Hamming metric while $\mathbb{E}_i[h_i] = 0$, where $\mathbb{E}_i$ denotes the expectation with respect to π^i. Therefore, applying the $d = 1$ case of (6.6) we obtain, for any $i \in [d]$,

$$\mathbb{E}\left[\exp(\lambda h_i(\boldsymbol{\pi}^{<i+1})) | \boldsymbol{\pi}^{<i}\right] \le \exp(2\lambda^2 n).$$

Since $f - \mathbb{E}f = \sum_{i=1}^d h_i$ and h_i are measurable with respect to $\boldsymbol{\pi}^{<i+1}$, iterating the above bound gives (6.6). $\square$

Lemma 6.3 reduces our task to showing the normalized traces of $F_{ij}(\xi)$ are L-Hamming-Lipschitz for an appropriate L. For this task we will make use of the following:

Lemma 6.4 (Resolvent identity)**.** Let A and B be two Hermitian matrices, and let $\xi \in \mathbb{C} \backslash \mathbb{R}$. Then

$$(\xi - A)^{-1} - (\xi - B)^{-1} = (\xi - A)^{-1}(A - B)(\xi - B)^{-1} = (\xi - B)^{-1}(A - B)(\xi - A)^{-1}.$$

More generally for any two invertible matrices C and D, we have

$$C^{-1} - D^{-1} = C^{-1}(D - C)D^{-1}. \tag{6.7}$$

Proof of Theorem 6.1. Fix $i, j \in \{1, 2\}$, $\ell \in [d]$ and set $H_n(\xi) := \frac{1}{n}\mathrm{Tr}\, F_{ij}^M(\xi)$. As mentioned above we need to show that $H_n(\cdot)$ is an L-Lipschitz function of $\boldsymbol{\pi} = (\pi_n^1, \dots, \pi_n^d)$ with respect to the Hamming distance (6.3) for an appropriate value of L. By the triangle inequality it suffices to show it is L-Lipschitz as a function of π_n^ℓ with respect to the Hamming distance (6.2) on $\mathbb{S}_n$, for arbitrary fixed $\ell \in [d]$.

To this end, we define

$$\widetilde{F}^M(\xi) := F^M(\xi, z, \widetilde{S}_n^d(\ell)), \text{ where } \widetilde{S}_n^d(\ell) := \sum_{k \in [d] \backslash \{\ell\}} \pi_n^k + \widetilde{\pi}_n^\ell,$$

and $\widetilde{\pi}_n^\ell$ is some fixed but arbitrary permutation over $[n]$. We similarly define $\widetilde{F}_{ij}^M(\xi)$ and $\widetilde{H}_n(\xi)$. Now using the resolvent identity we note that

$$F^M(\xi) - \widetilde{F}^M(\xi) = \frac{1}{\sqrt{d}} F^M(\xi) \left(\Delta_n^\ell + (\Delta_n^\ell)^*\right) \widetilde{F}^M(\xi),$$

where

$$\Delta_n^\ell := \begin{bmatrix} 0 & (\widetilde{\pi}_n^\ell - \pi_n^\ell) \\ 0 & 0 \end{bmatrix}.$$

Therefore,

$$H_n(\xi) - \widetilde{H}_n(\xi) = \frac{1}{n\sqrt{d}}\mathrm{Tr}\left[\begin{pmatrix} E_i^\mathsf{T} \\ \mathbf{0}^\mathsf{T} \end{pmatrix} F^M(\xi) \left(\Delta_n^\ell + (\Delta_n^\ell)^*\right) \widetilde{F}^M(\xi)(E_j \quad \mathbf{0})\right], \tag{6.8}$$

where

$$E_1 := \begin{pmatrix} I_n \\ 0_n \end{pmatrix}, \quad E_2 := \begin{pmatrix} 0_n \\ I \end{pmatrix}, \quad \mathbf{0} := \begin{pmatrix} 0_n \\ 0_n \end{pmatrix},$$

and 0_n is the $n \times n$ matrix of zeros. To simplify (6.8) further, we note that the $(k, n+k')$-th entry of Δ_n^ℓ is non-zero for some $k, k' \in [n]$, if and only if $\pi_n^\ell(k) \ne \widetilde{\pi}_n^\ell(k)$ and one of $\pi_n^\ell(k)$ and $\widetilde{\pi}_n^\ell(k)$ equals k'. Hence, using the triangle inequality and recalling the definition of $d_H(\cdot, \cdot)$, it follows that $|H_n(\xi) - \widetilde{H}_n(\xi)|$ is bounded by the sum of $4d_H(\pi_n^\ell, \widetilde{\pi}_n^\ell)$ terms of the form

$$\frac{1}{n\sqrt{d}}\left|\mathrm{Tr}\left[\begin{pmatrix} E_i^\mathsf{T} \\ \mathbf{0}^\mathsf{T} \end{pmatrix} F^M(\xi)e_k e_{k'}^\mathsf{T} \widetilde{F}^M(\xi)(E_j \quad \mathbf{0})\right]\right|, \tag{6.9}$$

for some $k, k' \in [2n]$. Here e_m denotes the canonical basis vector which has one in the m-th position. Since $|\mathrm{Im}\,\xi|, \|M\| \leq C_0$ we have the operator norm bounds

$$\|F^M(\xi)\|, \|\widetilde{F}^M(\xi)\| \leq \|M\| + |\mathrm{Im}\,\xi|^{-1} \leq 2C_0^2 |\mathrm{Im}\,\xi|^{-1}$$

As $\|E_i\| = 1$ for $i = 1, 2$, we have

$$
\left| \mathrm{Tr} \left[\begin{pmatrix} E_i^\mathsf{T} \\ \mathbf{0}^\mathsf{T} \end{pmatrix} F^M(\xi) e_k e_{k'}^\mathsf{T} \widetilde{F}^M(\xi) (E_j \quad \mathbf{0}) \right] \right| = \left| e_{k'}^\mathsf{T} \widetilde{F}^M(\xi) (E_j \quad \mathbf{0}) \begin{pmatrix} E_i^\mathsf{T} \\ \mathbf{0}^\mathsf{T} \end{pmatrix} F^M(\xi) e_k \right|
$$
$$
\leq \left\| \widetilde{F}^M(\xi) (E_j \quad \mathbf{0}) \begin{pmatrix} E_i^\mathsf{T} \\ \mathbf{0}^\mathsf{T} \end{pmatrix} F^M(\xi) \right\|
$$
$$
\leq \frac{4C_0^4}{(\mathrm{Im}\,\xi)^2}. \tag{6.10}
$$

Now combining (6.9)-(6.10) and (6.8), we obtain

$$|H_n(\xi) - \widetilde{H}_n(\xi)| \leq \frac{16 C_0^4 d_H(\pi_n^\ell, \widetilde{\pi}_n^\ell)}{n\sqrt{d}(\mathrm{Im}\,\xi)^2}. \tag{6.11}$$

This shows that we can apply Lemma 6.3 with $f(\boldsymbol{\pi}) = H_n(\xi)$ and $L = 16 C_0^4 / n\sqrt{d}(\mathrm{Im}\,\xi)^2$, and the result follows. $\qquad \square$

7 Proof of the local law

In this section we prove Theorem 2.6. The proof consists of two key components. First we derive an approximate fixed point equation for $\widetilde{m}_n(\xi)$, the Stieltjes transform of the symmetrized version of the empirical measure of the singular values of $z - S_n^d/\sqrt{d}$. Since the fixed point equation is an equation of degree three it is not apriori immediate that $\widetilde{m}_n(\xi)$ is close to the correct solution of the fixed point equation. To tackle this, we need certain properties of the roots of that cubic equation. We also need to employ a bootstrap argument to quantify the difference between $\widetilde{m}_n(\xi)$ and its limit $\widetilde{m}_\infty(\xi)$ when $\mathrm{Im}\,\xi$ approaches zero.

7.1 Derivation of the approximate fixed point equation

The main technical result of this section is the following lemma.

Lemma 7.1 (Loop equation). Fix $\xi \in \mathbb{C}^+$ such that $(\log n)^{-2} \leq \mathrm{Im}\,\xi \leq C_0$ for some $C_0 > 0$. Fix $z \in B_{\mathbb{C}}(0, R)$ for some $R < \infty$. Then, there exists an event $\Omega_n(\xi)$ with

$$\mathbb{P}(\Omega_n(\xi)^c) \leq \exp\left(-c_{7.1}(\log n)^2\right)$$

such that on $\Omega_n(\xi)$ we have

$$
\left| \widetilde{m}_n(\xi)(\widetilde{m}_n(\xi) - \xi)^2 + (1 - |z|^2)\widetilde{m}_n(\xi) - \xi \right|
$$
$$
\leq C_{7.1} \max\{d^{-1/2}, n^{-1/4} \log n\}(\mathrm{Im}\,\xi)^{-3}(1 + |\mathbb{E}\widetilde{m}_n(\xi)|), \tag{7.1}
$$

where $c_{7.1}$ is an absolute constant and $C_{7.1}$ depends only on C_0 and R.

Since we have concentration bounds in Theorem 6.1, as we will see below, it will be enough to show that inequality (7.1) holds for $\mathbb{E}\widetilde{m}_n(\xi)$. To show the same, it will be convenient to consider the Stieltjes transform of symmetrized version of the empirical measure of the singular values of $z - \widehat{S}_n^d/\sqrt{d}$, where $\widehat{S}_n^d$ is now centered. For ease of writing, let us denote $\mathsf{S}_n^d := \sum_{\ell=1}^d \mathsf{P}_n^\ell$, where for $\ell \in [d]$,

$$
\mathsf{P}_n^\ell := \begin{bmatrix} 0 & (P_n^\ell - \mathbb{E}P_n^\ell) \\ (P_n^\ell - \mathbb{E}P_n^\ell)^* & 0 \end{bmatrix},
$$

and $\{P_n^\ell\}$ are i.i.d. uniformly distributed permutation matrices. Define the resolvent as

$$\widehat{G}(S_n^d) := \widehat{G}(S_n^d, \xi, z) := \left[\xi I_{2n} - \begin{bmatrix} 0 & zI_n \\ \bar{z}I_n & 0 \end{bmatrix} + \frac{\mathsf{S}_n^d}{\sqrt{d}}\right]^{-1}$$

and denote $\widehat{m}_n(\xi) := \frac{1}{2n}\mathrm{Tr}\,\widehat{G}(S_n^d)$.

Lemma 7.2 (Loop equation for the sum of centered permutation matrices)**.** Fix $\xi \in \mathbb{C}^+$ such that $n^{-1/16} \leq \mathrm{Im}\,\xi \leq C_0$ for some $C_0 > 0$. Fix $z \in B_{\mathbb{C}}(0, R)$ for some $R < \infty$. Then, there exists a constant $C_{7.2}$, depending on C_0 and R, such that

$$\left|\mathbb{E}\widehat{m}_n(\xi)(\mathbb{E}\widehat{m}_n(\xi) - \xi)^2 + (1 - |z|^2)\mathbb{E}\widehat{m}_n(\xi) - \xi\right| \leq C_{7.2}d^{-1/2}(\mathrm{Im}\,\xi)^{-3}(1 + |\mathbb{E}\widehat{m}_n(\xi)|). \quad (7.2)$$

Recalling the definition of $\widetilde{G}(S_n^d)$ (see (2.20)) we observe that $\widetilde{m}_n(\xi)$ and $\widehat{m}_n(\xi)$ are the normalized traces of the resolvent of two Hermitian matrices differed by a finite rank perturbation. Therefore, one can use the following result to bound the difference between $\widetilde{m}_n(\xi)$ and $\widehat{m}_n(\xi)$. Its proof is a simple application of Cauchy's interlacing inequality. We include it for completeness.

Lemma 7.3. Let A_i, $i = 1, 2$, be two $n \times n$ Hermitian matrices such that $\mathrm{rank}(A_1 - A_2) \leq C_1$ for some absolute constant C_1. For $i = 1, 2$, and $\xi \in \mathbb{C}\backslash\mathbb{R}$, let $m_n^{A_i}(\xi)$ denote the Stieltjes transform of the empirical law of the eigenvalues of A_i. That is,

$$m_n^{A_i}(\xi) := \int \frac{1}{\xi - x}dL_{A_n^i}(x), \qquad i = 1, 2.$$

Then

$$|m_n^{A_1}(\xi) - m_n^{A_2}(\xi)| \leq \frac{C_1\pi}{n|\mathrm{Im}\,\xi|}.$$

Proof. Since $\frac{1}{\xi - x} = \int_{-\infty}^{x} \frac{1}{(t-\xi)^2}dt$ we observe that

$$m_n^{A_i}(\xi) = \int_{-\infty}^{\infty} \int_{-\infty}^{x} \frac{1}{(t-\xi)^2}dt dL_{A_n^i}(x) = \frac{1}{n}\int_{-\infty}^{\infty} \frac{\mathfrak{n}_i(-\infty, t]}{(\xi - t)^2},$$

for $i = 1, 2$, where $\mathfrak{n}_i(-\infty, t]$ denotes the number of eigenvalues of A_i in the interval $(-\infty, t]$. As $\mathrm{rank}(A_1 - A_2) \leq C_1$, by Cauchy's interlacing inequality it also follows that $|\mathfrak{n}_1(-\infty, t] - \mathfrak{n}_2(-\infty, t]| \leq C_1$. Therefore,

$$\left|m_n^{A_1}(\xi) - m_n^{A_2}(\xi)\right| \leq \frac{C_1}{n}\int \frac{1}{|\xi - t|^2}dt = \frac{C_1}{n}\int \frac{1}{(\mathrm{Im}\,\xi)^2 + t^2}dt = \frac{C_1\pi}{n|\mathrm{Im}\,\xi|}. \qquad \square$$

Equipped with Lemma 7.3 and assuming Lemma 7.2 we now prove Lemma 7.1.

Proof of Lemma 7.1. Using Lemma 7.3 and the trivial bounds $|\widehat{m}_n(\xi)|, |\widetilde{m}_n(\xi)| \leq 1/\mathrm{Im}\,\xi$ we obtain that $|\widetilde{P}(\mathbb{E}\widehat{m}_n(\xi)) - \widetilde{P}(\mathbb{E}\widetilde{m}_n(\xi))| = O(n^{-1/2}(\mathrm{Im}\,\xi)^{-3})$. Therefore, Lemma 7.2 implies that

$$\widetilde{P}(\mathbb{E}\widetilde{m}_n(\xi)) = O\left(d^{-1/2}(\mathrm{Im}\,\xi)^{-3}(1 + |\mathbb{E}\widehat{m}_n(\xi)|)\right) = O\left(d^{-1/2}(\mathrm{Im}\,\xi)^{-3}(1 + |\mathbb{E}\widetilde{m}_n(\xi)|)\right),$$
$$(7.3)$$

where we have used Lemma 7.3 again and the fact that $n\mathrm{Im}\,\xi \geq 1$. It remains to show that

$$\left|\widetilde{P}(\widetilde{m}_n(\xi)) - \widetilde{P}(\mathbb{E}\widetilde{m}_n(\xi))\right| = O\left(\frac{\log n}{n^{1/4}(\mathrm{Im}\,\xi)^3}\right), \quad (7.4)$$

with high probability. This will complete the proof of the lemma.

To this end, applying Theorem 6.1, seting $M = 0$ there, using the trivial bound $|\widetilde{m}_n(\xi)| \le 1/\mathrm{Im}\,\xi$ again, and the triangle inequality we obtain that

$$\mathbb{P}\left(|\widetilde{m}_n^3(\xi) - (\mathbb{E}\widetilde{m}_n(\xi))^3| \ge \frac{3\log n}{n^{1/2}(\mathrm{Im}\,\xi)^4}\right) \le \mathbb{P}\left(|\widetilde{m}_n(\xi) - \mathbb{E}\widetilde{m}_n(\xi)| \ge \frac{\log n}{n^{1/2}(\mathrm{Im}\,\xi)^2}\right)$$

$$\le 4e^{-c_{6.1}(\log n)^2} \qquad (7.5)$$

and

$$\mathbb{P}\left(|\widetilde{m}_n^2(\xi) - (\mathbb{E}\widetilde{m}_n(\xi))^2| \ge \frac{2\log n}{n^{1/2}(\mathrm{Im}\,\xi)^3}\right) \le 4e^{-c_{6.1}(\log n)^2}. \qquad (7.6)$$

Since $n^{-1/16} \le \mathrm{Im}\,\xi \le C_0$ we also have that

$$\frac{\log n}{n^{1/2}(\mathrm{Im}\,\xi)^3} = O\left(\frac{\log n}{n^{1/2}(\mathrm{Im}\,\xi)^4}\right) = O\left(\frac{\log n}{n^{1/4}(\mathrm{Im}\,\xi)^3}\right),$$

yielding (7.4). The desired probability bounds (7.5)-(7.6). The proof of the lemma now completes. □

Now it remains to prove Lemma 7.2. As we will see below, to prove the same we will first derive an approximate fixed point equation involving $\mathbb{E}\widehat{m}_n(\xi)$ and an auxiliary variable $\mathbb{E}\hat{\nu}_n(\xi)$ where

$$\widehat{\nu}_n(\xi) := \frac{1}{n}\sum_{i=1}^n (\widehat{G}(S_n^d))_{i,n+i}.$$

Then an additional equation will be derived to eliminate $\mathbb{E}\nu_n(\xi)$ from the first equation. To obtain these two equations we will need to consider the expectation of the entries of product of matrices that are functions of centered permutation matrices. Hence, it will be useful to introduce the following notation. For ease of writing, for any permutation π_n uniformly distributed on $\mathbb{S}_n$, we denote

$$\mathsf{P} := \mathsf{P}_n := \begin{bmatrix} 0 & (P_n - \mathbb{E}P_n) \\ (P_n - \mathbb{E}P_n)^* & 0 \end{bmatrix}, \qquad \text{where} \qquad P_n(i,j) := \mathbb{I}(\pi_n(i) = j). \quad (7.7)$$

Equipped with the above notation we have the following lemma.

Lemma 7.4. Let $M := M_n$ be a $2n \times 2n$ deterministic matrix. Then

(i)

$$\mathbb{E}\left[(\mathsf{P}M\mathsf{P}M)_{i,i}\right] = M_{i,i}\left[\frac{1}{n}\sum_{j=1}^n M_{n+j,n+j}\right] + O\left(\frac{\|M\|^2}{\sqrt{n}}\right),$$

(ii)

$$\mathbb{E}\left[(\mathsf{P}M\mathsf{P}M)_{n+i,n+i}\right] = M_{n+i,n+i}\left[\frac{1}{n}\sum_{j=1}^n M_{j,j}\right] + O\left(\frac{\|M\|^2}{\sqrt{n}}\right),$$

(iii)

$$\mathbb{E}\left[(\mathsf{P}M\mathsf{P}M)_{n+i,i}\right] = M_{n+i,i}\left[\frac{1}{n}\sum_{j=1}^n M_{j,j}\right] + O\left(\frac{\|M\|^2}{\sqrt{n}}\right),$$

(iv)

$$\mathbb{E}\left[(\mathsf{P}M\mathsf{P}M)_{i,n+i}\right] = M_{i,n+i}\left[\frac{1}{n}\sum_{j=1}^n M_{n+j,n+j}\right] + O\left(\frac{\|M\|^2}{\sqrt{n}}\right).$$

Proof. Recalling (7.7), we make the following observations:

$$\mathbb{E}[\mathsf{P}_{i,n+j}\mathsf{P}_{n+k,\ell}] = \begin{cases} \frac{1}{n}(1-\frac{1}{n}) & \text{if } i=\ell, j=k \\ -\frac{1}{n^2} & \text{if } i\neq\ell, j=k, \text{or } i=\ell, j\neq k \\ \frac{1}{n^2(n-1)} & \text{if } i\neq\ell, j\neq k \end{cases} \tag{7.8}$$

and

$$\mathbb{E}[\mathsf{P}_{i,n+j}\mathsf{P}_{k,n+\ell}] = \begin{cases} \frac{1}{n}(1-\frac{1}{n}) & \text{if } i=k, j=\ell \\ -\frac{1}{n^2} & \text{if } i\neq k, j=\ell, \text{or } i=k, j\neq\ell \\ \frac{1}{n^2(n-1)} & \text{if } i\neq k, j\neq\ell \end{cases}. \tag{7.9}$$

Since the diagonal blocks of P are zero it follows that

$$\mathbb{E}[(\mathsf{P}M\mathsf{P}M)_{i,i}] = \mathbb{E}\left[\sum_{j,k,\ell=1}^{n} \mathsf{P}_{i,n+j}M_{n+j,n+k}\mathsf{P}_{n+k,\ell}M_{\ell,i} \right]$$

$$+ \mathbb{E}\left[\sum_{j,k,\ell=1}^{n} \mathsf{P}_{i,n+j}M_{n+j,k}\mathsf{P}_{k,n+\ell}M_{n+\ell,i} \right]$$

$$=: \text{Term I} + \text{Term II}. \tag{7.10}$$

Using (7.8) we have

$$\text{Term I} = M_{i,i}\left[\frac{1}{n}\sum_{j=1}^{n}M_{n+j,n+j} \right] \cdot \left(1-\frac{1}{n}\right)$$

$$- \frac{1}{n^2}\left[\sum_{\substack{j,\ell=1 \\ \ell\neq i}}^{n} M_{n+j,n+j}M_{\ell,i} + M_{i,i}\sum_{\substack{j,k=1 \\ j\neq k}}^{n} M_{n+j,n+k} \right]$$

$$+ \frac{1}{n^2(n-1)}\sum_{\substack{j,k,\ell=1 \\ j\neq k, i\neq\ell}}^{n} M_{n+j,n+k}M_{\ell,i}.$$

Since

$$\max_{\ell=1}^{2n}|M_{\ell,i}| \leq \sum_{\ell=1}^{2n}|M_{\ell,i}| \leq \sqrt{2n\sum_{\ell=1}^{2n}|M_{\ell,i}|^2} \leq \sqrt{2n}\|M\|, \quad \text{for all } i\in[2n], \tag{7.11}$$

we deduce from above that

$$\text{Term I} = M_{i,i}\left[\frac{1}{n}\sum_{j=1}^{n}M_{n+j,n+j} \right] + O\left(\frac{\|M\|^2}{\sqrt{n}} \right). \tag{7.12}$$

Using (7.9) and a similar argument as above we also deduce that

$$\text{Term II} = \frac{1}{n}\sum_{j=1}^{n}M_{n+j,i}^2 + O\left(\frac{\|M\|^2}{\sqrt{n}} \right) = O\left(\frac{\|M\|^2}{\sqrt{n}} \right), \tag{7.13}$$

where the last step follows from (7.11). Thus, the part (i) of the lemma now follows upon plugging the bounds (7.12)-(7.13) in (7.10). To prove (iii) we apply (7.8)-(7.9), (7.11),

and Cauchy–Schwarz inequality to deduce that

$$\mathbb{E}\left[(\mathsf{P}M\mathsf{P}M)_{n+i,i}\right] = M_{n+i,i}\left[\frac{1}{n}\sum_{j=1}^{n}M_{j,j}\right] + \frac{1}{n}\sum_{j=1}^{n}M_{j,n+i}M_{n+j,i} + O\left(\frac{\|M\|^2}{\sqrt{n}}\right)$$

$$= M_{i,n+i}\left[\frac{1}{n}\sum_{j=1}^{n}M_{n+j,n+j}\right] + O\left(\frac{\|M\|^2}{\sqrt{n}}\right).$$

This yields part (iii). The proofs of parts (ii) and (iv) follow from a similar argument as above and hence omitted. $\qquad\square$

We will apply Lemma by setting $\mathsf{P} = \mathsf{P}_n^\ell$ for some $\ell \in [d]$ and M will be functions of $\{\mathsf{P}_n^j\}$. For $\ell \in [d]$ denote

$$\widehat{G}^{(\ell)}(S_n^d) := \left[\xi - \begin{bmatrix} 0 & z \\ \bar{z} & 0 \end{bmatrix} + \frac{1}{\sqrt{d}}\mathsf{S}_n^{d,(\ell)}\right]^{-1},$$

where $\mathsf{S}_n^{d,(\ell)} := \sum_{j\neq\ell}\mathsf{P}_n^j$, and

$$\widehat{\nu}_n^{(\ell)}(\xi) := \frac{1}{n}\sum_{i=1}^{n}(\widehat{G}^{(\ell)}(S_n^d))_{i,n+i}.$$

Recall the following result regarding the inverse of a block matrix.

Lemma 7.5 (Inverse of a block matrix)**.**

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} = \begin{bmatrix} (A - BD^{-1}C)^{-1} & -A^{-1}B(D - CA^{-1}B)^{-1} \\ -D^{-1}C(A - BD^{-1}C)^{-1} & (D - CA^{-1}B)^{-1} \end{bmatrix}.$$

Applying Lemma 7.5 wth $A = D = \xi I$, where $\xi \in \mathbb{C}\backslash\mathbb{R}$, and $C = B^*$ we obtain that

$$\frac{1}{2}\mathrm{Tr}\left(\begin{bmatrix} \xi & B \\ B^* & \xi \end{bmatrix}^{-1}\right) = \xi\mathrm{Tr}\left\{(\xi^2 - BB^*)^{-1}\right\} = \xi\mathrm{Tr}\left\{(\xi^2 - B^*B)^{-1}\right\}.$$

This, in particular, implies that for every $\ell \in [d]$,

$$\frac{1}{n}\sum_{i=1}^{n}(\widehat{G}^{(\ell)}(S_n^d))_{n+i,n+i} = \frac{1}{n}\sum_{j=1}^{n}(\widehat{G}^{(\ell)}(S_n^d))_{j,j} = \frac{1}{2n}\mathrm{Tr}\,\widehat{G}^{(\ell)}(S_n^d) =: \widehat{m}_n^{(\ell)}(\xi). \qquad (7.14)$$

We now prove Lemma 7.2.

Proof of Lemma 7.2. From the identity $\widehat{G}(S_n^d)^{-1}\widehat{G}(S_n^d) = I_{2n}$ we obtain

$$\begin{bmatrix} \xi & -z \\ -\bar{z} & \xi \end{bmatrix}\widehat{G}(S_n^d) = I_{2n} - \frac{\mathsf{S}_n^d}{\sqrt{d}}\cdot\widehat{G}(S_n^d). \qquad (7.15)$$

Applying the resolvent identity ((6.7) in Lemma 6.4) twice we further obtain that for any $\ell \in [d]$,

$$\widehat{G}(S_n^d) = \widehat{G}^{(\ell)}(S_n^d) - \widehat{G}^{(\ell)}(S_n^d)\cdot\frac{\mathsf{P}_n^\ell}{\sqrt{d}}\cdot\widehat{G}^{(\ell)}(S_n^d) + \widehat{G}^{(\ell)}(S_n^d)\cdot\frac{\mathsf{P}_n^\ell}{\sqrt{d}}\cdot\widehat{G}^{(\ell)}(S_n^d)\cdot\frac{\mathsf{P}_n^\ell}{\sqrt{d}}\cdot\widehat{G}(S_n^d).$$

Fixing $\ell \in [d]$ we use the above identity to expand $\widehat{G}(S_n^d)$, which we then plug in (7.15). Therefore, now summing over $\ell \in [d]$, from (7.15) we deduce that

$$
\begin{bmatrix} \xi & -z \\ -\bar{z} & \xi \end{bmatrix} \widehat{G}(S_n^d) = I_{2n} - \sum_{\ell=1}^{d} \frac{\mathsf{P}_n^\ell}{\sqrt{d}} \cdot \widehat{G}^{(\ell)}(S_n^d) + \sum_{\ell=1}^{d} \frac{\mathsf{P}_n^\ell}{\sqrt{d}} \cdot \widehat{G}^{(\ell)}(S_n^d) \cdot \frac{\mathsf{P}_n^\ell}{\sqrt{d}} \cdot \widehat{G}^{(\ell)}(S_n^d)
$$
$$
- \sum_{\ell=1}^{d} \frac{\mathsf{P}_n^\ell}{\sqrt{d}} \cdot \widehat{G}^{(\ell)}(S_n^d) \cdot \frac{\mathsf{P}_n^\ell}{\sqrt{d}} \cdot \widehat{G}^{(\ell)}(S_n^d) \cdot \frac{\mathsf{P}_n^\ell}{\sqrt{d}} \cdot \widehat{G}(S_n^d). \tag{7.16}
$$

Next we need to simplify (7.16). To this end, for every $\ell \in [d]$, let $\mathbb{E}_\ell(\cdot)$ denote the expectation with respect to the randomness of P_n^ℓ and $\mathbb{E}_{-\ell}(\cdot)$ denote the expectation with respect to the randomness of $\{P_n^j\}_{j \neq \ell}$. Since $\{P_n^\ell\}$ are independent we have that P_n^ℓ and $\widehat{G}^{(\ell)}(S_n^d)$ are independent for every $\ell \in [d]$, which in particular implies that

$$
\mathbb{E}[\widehat{G}^{(\ell)}(S_n^d)\mathsf{P}_n^\ell] = \mathbb{E}_{-\ell}\left[\mathbb{E}_\ell(\widehat{G}^{(\ell)}(S_n^d)\mathsf{P}_n^\ell)\right] = 0, \qquad \ell \in [d], \tag{7.17}
$$

where we have used the fact that the entries of P_n^ℓ are centered. Applying Lemma 7.4 we also note that

$$
\frac{1}{n} \sum_{i=1}^{n} \sum_{\ell=1}^{d} \mathbb{E}\left[\left\{ \frac{\mathsf{P}_n^\ell}{\sqrt{d}} \cdot \widehat{G}^{(\ell)}(S_n^d) \cdot \frac{\mathsf{P}_n^\ell}{\sqrt{d}} \cdot \widehat{G}^{(\ell)}(S_n^d) \right\}_{n+i,n+i} \right]
$$
$$
= \frac{1}{d} \sum_{\ell=1}^{d} \mathbb{E}_{-\ell}\left[\left\{ \frac{1}{n} \sum_{i=1}^{n} (\widehat{G}^{(\ell)}(S_n^d))_{n+i,n+i} \right\} \cdot \left\{ \frac{1}{n} \sum_{j=1}^{n} (\widehat{G}^{(\ell)}(S_n^d))_{j,j} \right\} \right]
$$
$$
= \frac{1}{d} \sum_{\ell=1}^{d} \mathbb{E}_{-\ell}\left[\widehat{m}_n^{(\ell)}(\xi)^2 \right] + O(n^{-1/2}(\mathrm{Im}\,\xi)^{-2}), \tag{7.18}
$$

where the last step follows from (7.14) and the operator norm bound $\|\widehat{G}^{(\ell)}(S_n^d)\| \leq 1/\mathrm{Im}\,\xi$. Therefore, considering the $(n+i, n+i)$-th entry of the both sides of (7.16), taking an average over $i \in [n]$, followed by taking an expectation over the randomness of $\{\mathsf{P}_n^\ell\}$, upon using (7.17), we obtain

$$
-\bar{z}\mathbb{E}\widehat{\nu}_n(\xi) + \xi\mathbb{E}\widehat{m}_n(\xi) = 1 + \frac{1}{d} \sum_{\ell=1}^{d} \mathbb{E}_{-\ell}\left[\widehat{m}_n^{(\ell)}(\xi)^2 \right] - \mathrm{Term}\,\mathrm{E}_1 + O(n^{-1/2}(\mathrm{Im}\,\xi)^{-2}), \tag{7.19}
$$

where

$$
\mathrm{Term}\,\mathrm{E}_1 := \sum_{\ell=1}^{d} \mathbb{E}\left[\left(\frac{\mathsf{P}_n^\ell}{\sqrt{d}} \cdot \widehat{G}^{(\ell)}(S_n^d) \cdot \frac{\mathsf{P}_n^\ell}{\sqrt{d}} \cdot \widehat{G}^{(\ell)}(S_n^d) \cdot \frac{\mathsf{P}_n^\ell}{\sqrt{d}} \cdot \widehat{G}(S_n^d) \right)_{n+i,n+i} \right].
$$

Using the resolvent identity once again we observe that for any $\ell \in [d]$,

$$
\|\widehat{G}^{(\ell)}(S_n^d) - \widehat{G}(S_n^d)\| \leq \frac{1}{\sqrt{d}} \|\widehat{G}(S_n^d)\| \cdot \|P_n^\ell - \mathbb{E}P_n^\ell\| \cdot \|\widehat{G}^{(\ell)}(S_n^d)\| \leq 2d^{-1/2}(\mathrm{Im}\,\xi)^{-2}, \tag{7.20}
$$

where the last inequality follows from the facts that $\|\widehat{G}(S_n^d)\|, \|\widehat{G}^{(\ell)}(S_n^d)\| \leq (\mathrm{Im}\,\xi)^{-1}$ and $\|P_n^\ell - \mathbb{E}P_n^\ell\| \leq 2$. Thus $\mathrm{Term}\,\mathrm{E}_1 = O(d^{-1/2}(\mathrm{Im}\,\xi)^{-3})$, which in particular implies that the first term in the RHS of (7.19) is the dominant term. Using (7.20) we also note that $|\widehat{m}_n(\xi)^2 - \widehat{m}^{(\ell)}(\xi)^2| \leq 4d^{-1/2}(\mathrm{Im}\,\xi)^{-3}$. Hence, from (7.19), upon using the facts that $d = O(n)$ and $\mathrm{Im}\,\xi \leq C_0$, we deduce

$$
-\bar{z}\mathbb{E}\widehat{\nu}_n(\xi) + \xi\mathbb{E}\widehat{m}_n(\xi) = 1 + \frac{1}{d} \sum_{\ell=1}^{d} \mathbb{E}_{-\ell}\left[\widehat{m}_n^{(\ell)}(\xi)^2 \right] + O(d^{-1/2}(\mathrm{Im}\,\xi)^{-3}) \tag{7.21}
$$
$$
= 1 + \mathbb{E}\left[\widehat{m}_n(\xi)^2 \right] + O(d^{-1/2}(\mathrm{Im}\,\xi)^{-3})
$$
$$
= 1 + (\mathbb{E}\widehat{m}_n(\xi))^2 + O(d^{-1/2}(\mathrm{Im}\,\xi)^{-3}),
$$

where the last step follows from Corollary 6.2 upon taking

$$
M = \begin{bmatrix} 0 & \frac{\mathbb{E}S_n^d}{\sqrt{d}} \\ \frac{\mathbb{E}(S_n^d)^*}{\sqrt{d}} & 0 \end{bmatrix} = \begin{bmatrix} 0 & \frac{\sqrt{d}}{n}\mathbf{1}\mathbf{1}^\mathsf{T} \\ \frac{\sqrt{d}}{n}\mathbf{1}\mathbf{1}^\mathsf{T} & 0 \end{bmatrix}
$$

(recall that $\mathbf{1}$ is the n-dimensional vector consisting of all ones) and observing that $d^{1/2}(\mathrm{Im}\,\xi)^3 = O(n^{1/2}) = o(n^{3/4})$.

Note that (7.21) involves $\mathbb{E}\widehat{\nu}_n(\xi)$. To derive the desired approximate fixed point equation for $\mathbb{E}\widehat{m}_n(\xi)$ one needs eliminate $\mathbb{E}\widehat{\nu}_n(\xi)$ from (7.21). To this end, consider the $(i, n+i)$-th entry of the both sides of (7.16), take an average over $i \in [n]$, and proceed similarly as in the steps leading to (7.19) to deduce that

$$
\xi\mathbb{E}\widehat{\nu}_n(\xi) - z\mathbb{E}\widehat{m}_n(\xi) = \frac{1}{d}\sum_{\ell=1}^{d}\mathbb{E}_{-\ell}\left[\widehat{\nu}_n^{(\ell)}(\xi)\widehat{m}_n^{(\ell)}(\xi)\right] + O(n^{-1/2}(\mathrm{Im}\,\xi)^{-2}) - \mathrm{Term}\,\mathrm{E}_2, \quad (7.22)
$$

where

$$
\begin{aligned}
\mathrm{Term}\,\mathrm{E}_2 &:= \sum_{\ell=1}^{d}\mathbb{E}\left[\left(\frac{\mathsf{P}_n^\ell}{\sqrt{d}}\cdot\widehat{G}^{(\ell)}(S_n^d)\cdot\frac{\mathsf{P}_n^\ell}{\sqrt{d}}\cdot\widehat{G}^{(\ell)}(S_n^d)\cdot\frac{\mathsf{P}_n^\ell}{\sqrt{d}}\cdot\widehat{G}(S_n^d)\right)_{i,n+i}\right] \\
&= O(d^{-1/2}(\mathrm{Im}\,\xi)^{-3})
\end{aligned}
$$

and the last step follows from the operator norm bounds $\|\widehat{G}(S_n^d)\|, \|\widehat{G}^{(\ell)}(S_n^d)\| \le (\mathrm{Im}\,\xi)^{-1}$ and $\|P_n^\ell - \mathbb{E}P_n^\ell\| \le 2$. Using (7.20) and the resolvent identity we also have that

$$
|\widehat{\nu}_n^{(\ell)}(\xi)\widehat{m}_n^{(\ell)}(\xi) - \widehat{\nu}_n(\xi)\widehat{m}_n(\xi)| = O(d^{-1/2}(\mathrm{Im}\,\xi)^{-3}).
$$

On the other hand an application of Corollary 6.2 and Cauchy-Schwarz inequality yield that

$$
|\mathbb{E}(\widehat{m}_n(\xi)\widehat{\nu}_n(\xi)) - \mathbb{E}\widehat{m}_n(\xi)\mathbb{E}\widehat{\nu}_n(\xi)| \le n^{-3/4}.
$$

Therefore, the approximate equation (7.22) simplifies to

$$
(\mathbb{E}\widehat{m}_n(\xi) - \xi)\mathbb{E}\widehat{\nu}_n(\xi) = -z\mathbb{E}\widehat{m}_n(\xi) + O(d^{-1/2}(\mathrm{Im}\,\xi)^{-3}). \tag{7.23}
$$

Finally multiplying both sides of (7.21) by $(\mathbb{E}\widehat{m}_n(\xi) - \xi)$, using (7.23), and recalling that $\mathrm{Im}\,\xi \le C_0$, $|z| \le R$, we arrive at (7.2). This completes the proof of the lemma. $\qquad\square$

7.2 Proof of Theorem 2.6

In Section 7.1 we have shown that for $\xi \in \mathbb{C}^+$ with $(\log n)^{-2} \le \mathrm{Im}\,\xi \le C_0$ we have $\widetilde{P}(\widetilde{m}_n(\xi)) = o(1)$, with high probability, where

$$
\widetilde{P}(m) := \widetilde{P}(m,\xi) := \widetilde{P}(m,\xi,z) = m(m-\xi)^2 + (1 - |z|^2)m - \xi.
$$

Since $\widetilde{P}(\cdot)$ it is not evident from Lemma 7.1 that $|\widetilde{m}_n(\xi) - \widetilde{m}_\infty(\xi)| = o(1)$, where $\widetilde{m}_\infty(\xi)$ is the Stieltjes transform of the desired limit. Hence, it requires some additional properties of the roots of the equation $\widetilde{P}(m,\xi) = 0$.

During the proof of Theorem 1.1 we will see that it is enough to show that $|\widetilde{m}_n(\xi) - \widetilde{m}_\infty(\xi)| = o(1)$ holds for ξ purely imaginary, that is $\xi = \mathrm{i}\eta$ for some $\eta > 0$. On the other hand, for any symmetric probability measure μ on $\mathbb{R}$ (i.e. $\mu((a,b)) = \mu(((-b,-a))$ for any $0 < a < b \le \infty$), denoting $m_\mu(\xi)$ to be its Stieltjes transform, we have

$$
m_\mu(\mathrm{i}\eta) = \int\frac{1}{\mathrm{i}\eta - x}d\mu(x) = -\int\frac{x + \mathrm{i}\eta}{x^2 + \eta^2}d\mu(x) = -\mathrm{i}\eta\int\frac{1}{x^2 + \eta^2}d\mu(x).
$$

This means that $m_\mu(i\eta) = -ix$ for some $x > 0$. Therefore

$$\widetilde{P}(m_\mu(i\eta), i\eta) = ix(x + \eta)^2 - i(1 - |z|^2)x - i\eta. \qquad (7.24)$$

Thus for any symmetric probability measure μ on $\mathbb{R}$, the map $\eta \mapsto \widetilde{P}(m_\mu(i\eta, i\eta))$ is essentially a cubic polynomial over the reals. Since $\widetilde{m}_n(\xi)$ and $\widetilde{m}_\infty(\xi)$ are both Stieltjes transforms of symmetric probability measures and we need to control their differences only when ξ is purely imaginary, it is enough to derive properties of the roots of the equation

$$Q(x) := Q(x, \eta) := Q(x, \delta, \eta) := x(x + \eta)^2 - \delta x - \eta.$$

where $\delta, \eta > 0$.

Lemma 7.6 (Stability of the fixed point equation). Fix any $\delta, \eta > 0$. Then the following properties hold for the cubic equation $Q(x) = 0$.

  (i) There exists a unique positive root $x_\star$ of the equation $Q(x) = 0$.

  (ii) For any $c_0 > 0$,
$$\inf_{x \geq c_0} \frac{|Q(x)|}{|x - x_\star|} \geq c_0^2.$$

Proof. Since $Q(0) = -\eta < 0$ and $\lim_{x \to \infty} Q(x) = \infty$, the number of roots of the equation $Q(x) = 0$ in the interval $(0, \infty)$ is either one or three. If the number of positive roots of the equation $Q(x) = 0$ is three, then the Rolle's theorem implies that there exists $x_0 \in (0, \infty)$ such that $Q''(x_0) = 0$ which is clearly a contradiction, as we note that $Q''(x) = 3x^2 + 4\eta > 0$ for all $x \in \mathbb{R}$. Thus there exists a unique $x_\star \in (0, \infty)$ such that $Q(x_\star) = 0$. Turning to prove the second part of the lemma we note that

$$\begin{aligned}
Q(x) = Q(x) - Q(x_\star) &= (x - x_\star)((x + \eta)^2 + x_\star(x + x_\star + 2\eta) - \delta) \\
&= (x - x_\star)(x^2 + 2x\eta + xx_\star + x_\star^2 + \eta^2 + 2x_\star\eta - \delta) \\
&= (x - x_\star)\left(x^2 + 2x\eta + xx_\star + \frac{\eta}{x_\star}\right) \qquad (7.25)
\end{aligned}$$

where the last equality follows from the fact that $Q(x_\star) = 0$. Since $x, x_\star, \eta > 0$, we have that

$$x^2 + 2x\eta + xx_\star + \frac{\eta}{x_\star} \geq c_0^2,$$

for all $x \geq c_0$. This completes the proof of the lemma. $\qquad \square$

    Recalling (7.24) we see that for any symmetric probability measure μ, $\widetilde{P}(m_\mu(i\eta, i\eta)) = iQ(x, \eta)$ where $m_\mu(i\eta) = -ix$. Therefore, Lemma 7.6(i) implies that there is a unique symmetric probability measure $\widetilde{\mu}_\infty$ such that its Stieltjes transform $\widetilde{m}_\infty(\xi)$ satisfies the fixed point equation $\widetilde{P}(m) = 0$. The second part of Lemma 7.6 ensures that

$$|\widetilde{P}(\widetilde{m}_n(i\eta))| \geq |\widetilde{m}_n(i\eta)|^2 |\widetilde{m}_n(i\eta) - \widetilde{m}_\infty(i\eta)| \qquad (7.26)$$

for all $\eta > 0$ and in particular

$$|\widetilde{P}(\widetilde{m}_n(i\eta))| \geq c_0^2 |\widetilde{m}_n(i\eta) - \widetilde{m}_\infty(i\eta)|, \qquad (7.27)$$

provided $|\widetilde{m}_n(i\eta)| \geq c_0$. The inequalities (7.26)-(7.27) will be crucially used to derive a bound on the difference between $\widetilde{m}_n(\xi)$ and $\widetilde{m}_\infty(\xi)$ from a bound on $|\widetilde{P}(\widetilde{m}_n(\xi))|$. However, these inequalities need apriori lower bound on $|\widetilde{m}_n(\xi)|$. Hence, to initiate the bootstrap argument we need to show that such lower bounds hold when ξ far away from the real line.

Lemma 7.7 (Preliminary lower bound). Fix $z \in \mathbb{C}$ such that $|z| \leq 1$. For any positive constant C denote

$$\mathbb{H}_C := \{\xi \in \mathbb{C} : \text{Im}\,\xi \geq C, \xi \in B_{\mathbb{C}}(0, 2C)\}.$$

Then there exist a sufficiently large absolute constant $\bar{C}_{7.7}$ and a sufficiently small absolute constant $c'_{7.7}$ such that the following holds: There exists a set $\Omega_{7.7,n}$ such that for every $\xi \in \mathbb{H}_{\bar{C}_{7.7}}$, we have $|\widetilde{m}_n(\xi)| \geq 1/10$ on the event $\Omega_{7.7,n}$, where

$$\mathbb{P}(\Omega^c_{7.7,n}) \leq \exp(-c'_{7.7}d).$$

Proof. We set

$$\Omega^c_{7.7,n} := \left\{ \frac{1}{n}\text{Tr}\,(z - \hat{S}^d_n)^*(z - \hat{S}^d_n) > C \right\},$$

where C is chosen to be sufficiently large and for brevity we write $\hat{S}^d_n := S^d_n/\sqrt{d}$. Recalling that $d = O(n)$ and $|z| \leq 1$ it follows from (5.8) that for C large,

$$\mathbb{P}\left(\frac{1}{n}\text{Tr}\,(z - \hat{S}^d_n)^*(z - \hat{S}^d_n) > C \right) \leq d\exp(-c'd),$$

for some absolute constant c' establishing the desired assertion on the probability bound of $\Omega^c_{7.7,n}$.

Now note that

$$-\text{Im}\,\widetilde{m}_n(\xi) = \frac{1}{2n}\sum_{i=1}^{2n} \frac{\text{Im}\,\xi}{|\xi - \lambda_i|^2} \geq \frac{\text{Im}\,\xi}{\frac{1}{2n}\sum_{i=1}^{2n}|\xi - \lambda_i|^2} \geq \frac{\text{Im}\,\xi}{2|\xi|^2 + \frac{1}{n}\sum_{i=1}^{2n}\lambda_i^2},$$

where $\{\lambda_i\}_{i=1}^{2n}$ are the eigenvalues of

$$\begin{bmatrix} 0 & (z - \hat{S}^d_n) \\ (z - \hat{S}^d_n)^* & 0 \end{bmatrix}.$$

Thus,

$$-\text{Im}\,\widetilde{m}_n(\xi) \geq \frac{\bar{C}_{7.7}}{8\bar{C}^2_{7.7} + \frac{2}{n}\text{Tr}\,(z - \hat{S}^d_n)^*(z - \hat{S}^d_n)}.$$

The desired lower bound on $\widetilde{m}_n(\xi)$ on the event $\Omega_{7.7,n}$ now follows upon setting $\bar{C}_{7.7} = C$. □

When $\text{Im}\,\xi$ is close to zero we cannot use Lemma 7.7. In that case, the desired bound $|\widetilde{m}_n(\xi)|$ can be obtained by showing that it is close to $\widetilde{m}_\infty(\xi)$ and then obtaining bounds on $|\widetilde{m}_\infty(\xi)|$ which we derive in the lemma below.

Lemma 7.8 (Properties of $\widetilde{m}_\infty$). Fix any $\varepsilon > 0$ and let $z \in B_{\mathbb{C}}(0, 1 - \varepsilon)$. Fix any $\xi \in \mathbb{C}^+$ such that $|\xi| \leq \varepsilon^{-1}$. Then there exist $\varepsilon_0 > 0$ such that for any $\varepsilon < \varepsilon_0$ there exists constants $c_{7.8}$ and $C_{7.8}$, depending only on ε, such that $c_{7.8} \leq |\widetilde{m}_\infty(\xi)| \leq C_{7.8}$.

Proof. The proof of this lemma follows from [13, Lemma 4.3]. There they analyzed properties of the solution $m_c(\xi)$ of the cubic equation

$$m(1 + m)^2\xi + (1 - |z|^2)m + 1 = 0,$$

which has nonnegative imaginary part for all $\xi \in \mathbb{C}$. In [4] it was shown that for any $\xi \in \mathbb{C}^+$, $-m_c(\xi)$ is the Stieltjes transform of the limiting distribution of the empirical measure of the singular values of $z - A_n/\sqrt{n}$ where A_n is an $n \times n$ matrix of i.i.d. entries with certain moment assumptions on its entries. Note that the limiting measure is the

same in our set-up. Therefore $m_\infty(\xi) = -m_c(\xi)$ on $\mathbb{C}^+$. Since $\widetilde{m}_\infty(\xi) = \xi m_\infty(\xi^2)$, we use the relation between $m_\infty(\xi)$ and $m_c(\xi)$ to extract the properties of $\widetilde{m}_\infty(\xi)$.

From [13, Eqn. (4.9)] we note that

$$c|\xi|^{-1} \le |m_\infty(\xi^2)| \le C|\xi|^{-1}, \tag{7.28}$$

whenever $\mathrm{Im}\,(\xi^2) > 0$, for some constants c and C depending only on ε. When $\mathrm{Im}\,(\xi^2) < 0$ then we note that $m_\infty(\xi^2) = \overline{m_\infty(\bar\xi^2)} = \overline{m_c(\bar\xi^2)}$ and therefore (7.28) also holds for all ξ such that $\mathrm{Im}\,(\xi^2) < 0$. Multiplying both sides of (7.28) by $|\xi|$ and using the relation between $\widetilde{m}_\infty(\cdot)$ and $m_\infty(\cdot)$ we establish the desired conclusion for $\widetilde{m}_\infty(\cdot)$ for all ξ such that $\mathrm{Re}\,\xi \ne 0$. We extend our conclusion for all ξ such that $\mathrm{Re}\,\xi = 0$ by continuity of $\widetilde{m}_\infty(\cdot)$ on $\mathbb{C}^+$. $\qquad\square$

Equipped with all ingredients we are now ready to prove Theorem 2.6.

Proof of Theorem 2.6. Recall that

$$\mathscr{S}_{\varepsilon,\varpi} := \left\{ \xi = \mathrm{i}\eta : 0 < \eta \le \bar{C}_{2.6}, \min\{d^{1/2}, n^{1/4}(\log n)^{-1}\}\eta^3 \ge \varpi_n \right\},$$

where we set $\bar{C}_{2.6} = 2\bar{C}_{7.7}$. We need to show that $\widetilde{m}_n(\xi)$ is close to $\widetilde{m}_\infty(\xi)$ uniformly for all $\xi \in \mathscr{S}_{\varepsilon,\varpi}$. Consider a decreasing sequence of positive reals $\{\eta_i\}_{i=0}^N$ such that $\eta_0 = \bar{C}_{2.6}$, $1/(2n) < \eta_i - \eta_{i+1} < 1/n$ and $\eta_N \in \mathscr{S}_{\varepsilon,\varpi}$. Note that $N = O(n)$. Denote

$$\Upsilon_n(\xi) := 3C_{7.1}\max\{d^{-1/2}, n^{-1/4}\log n\}(\mathrm{Im}\,\xi)^{-3}(1 + 4C_{7.8}) \tag{7.29}$$

and set

$$c_0 = \frac{1}{4}\min\{c_{7.8}, 1/10\}. \tag{7.30}$$

Note that $\Upsilon_n(\xi) = o(1)$ for all $\xi \in \mathscr{S}_{\varepsilon,\varpi}$. Now applying Lemma 7.1 we see that on the event $\Omega_n(\xi_0)$ we have

$$\widetilde{P}(\widetilde{m}_n(\xi_0)) = O\left(\max\{d^{-1/2}, n^{-1/4}\log n\}\right),$$

as $|\mathbb{E}\widetilde{m}_n(\xi_0)| \le 1/\eta_0 < 1$. From (7.26) we have that

$$|\widetilde{P}(\widetilde{m}_n(\xi_0)) \ge |\widetilde{m}_n(\xi_0)|^2|\widetilde{m}_n(\xi_0) - \widetilde{m}_\infty(\xi_0)|.$$

This together with Lemma 7.7 further implies that on the event $\Omega_{7.7,n} \cap \Omega_n(\xi_0)$ we have

$$|\widetilde{m}_n(\xi_0) - \widetilde{m}_\infty(\xi_0)| = O\left(\max\{d^{-1/2}, n^{-1/4}\log n\}\right) = o(1).$$

Therefore, Lemma 7.8 and the triangle inequality yields

$$2c_0 \le |\widetilde{m}_n(\xi_0)| \le 2C_{7.8} \tag{7.31}$$

on the event $\Omega_{7.7,n} \cap \Omega_n(\xi_0)$, for all large n. Note that we also have

$$|\mathbb{E}[\widetilde{m}_n(\xi_0)]| \le 2C_{7.8} + (\mathrm{Im}\,\xi_0)^{-1}\mathbb{P}(\Omega_{7.7,n}^c \cup \Omega_n(\xi_0)^c) \le 3C_{7.8}, \tag{7.32}$$

for all large n, where we use the fact that $\mathrm{Im}\,\xi_0 > \mathrm{Im}\,\xi_N \ge (\log n)^{-2}$.

Now we are ready to carry out the bootstrap argument. Indeed, applying Lemma 7.1 again and using the inequality $|\widetilde{m}_n(\xi) - \widetilde{m}_n(\xi')| \leq |\xi - \xi'|/((\operatorname{Im}\xi) \cdot (\operatorname{Im}\xi'))$ we deduce that

$$
\begin{aligned}
&|\widetilde{P}(\widetilde{m}_n(\xi))| \\
&\leq |\widetilde{P}(\widetilde{m}_n(\xi_0))| + |\widetilde{P}(\widetilde{m}_n(\xi)) - \widetilde{P}(\widetilde{m}_n(\xi_0))| \\
&\leq C_{7.1} \max\{d^{-1/2}, n^{-1/2}(\log n)^3\}(\operatorname{Im}\xi_0)^{-3}(1 + |\mathbb{E}\widetilde{m}_n(\xi_0)|) + O\left(\frac{(\log n)^8}{n}\right) \\
&\leq 2C_{7.1} \max\{d^{-1/2}, n^{-1/2}(\log n)^3\}(\operatorname{Im}\xi_0)^{-3}(1 + |\mathbb{E}\widetilde{m}_n(\xi_0)|) \\
&\leq 2C_{7.1} \max\{d^{-1/2}, n^{-1/2}(\log n)^3\}(\operatorname{Im}\xi)^{-3}\left(1 + O\left(\frac{(\log n)^8}{n}\right)\right)(1 + |\mathbb{E}\widetilde{m}_n(\xi_0)|) \\
&\leq \Upsilon_n(\xi),
\end{aligned}
\tag{7.33}
$$

for all $\xi = i\eta$ with $\eta \in [\eta_1, \eta_0]$, on the event $\Omega_{7.7,n} \cap \Omega_n(\xi_0)$, where in the last step we have used (7.32). On other hand, from (7.31) and the inequality $|\widetilde{m}_n(\xi) - \widetilde{m}_n(\xi')| \leq |\xi - \xi'|/((\operatorname{Im}\xi) \cdot (\operatorname{Im}\xi'))$ we obtain $|\widetilde{m}_n(\xi)| \geq c_0$ for all $\xi = i\eta$ with $\eta \in [\eta_1, \eta_0]$, on the event $\Omega_{7.7,n} \cap \Omega_n(\xi_0)$. This together with (7.27) implies that

$$
|\widetilde{m}_n(\xi) - \widetilde{m}_\infty(\xi)| \leq c_0^{-2}\Upsilon_n(\xi)
\tag{7.34}
$$

for all $\xi = i\eta$ with $\eta \in [\eta_1, \eta_0]$, on the event $\Omega_{7.7,n} \cap \Omega_n(\xi_0)$.

We complete the proof by induction. Indeed, we denote $\Omega_j := \cap_{i=0}^{j-1}\Omega_n(\xi_i) \cap \Omega_{7.7,n}$. By the induction hypothesis we assume that (7.34) holds for all $\xi = i\eta$ with $\eta \in [\eta_k, \eta_0]$ on the event Ω_k. To finish the proof we need to show that (7.34) continue to hold for all $\xi = i\eta$ with $\eta \in [\eta_{k+1}, \eta_k]$ on the event Ω_{k+1}.

First we note that using Lemma 7.8 and proceeding similarly as in (7.32) we obtain $|\mathbb{E}\widetilde{m}_n(\xi_k)| \leq 3C_{7.8}$. Therefore, arguing similarly as in (7.33) we deduce that the conclusion of (7.33) continue to hold for all $\xi = i\eta$ with $\eta \in [\eta_{k+1}, \eta_k]$ on the event Ω_{k+1}. Using this we also get that (7.34) holds for all $\xi = i\eta$ with $\eta \in [\eta_{k+1}, \eta_k]$ on the event Ω_{k+1}.

Thus by induction we have shown that for all $\xi = i\eta$ with $\eta \in [\eta_N, \eta_0]$ the inequality (7.34) holds on the event Ω_N. Since,

$$
\mathbb{P}(\Omega_N^c) \leq \mathbb{P}(\Omega_{7.7,n}^c) + \sum_{j=0}^{N-1}\mathbb{P}(\Omega_n(\xi_j)^c),
$$

the proof of the theorem now completes from the probability bounds obtained in Lemma 7.1 and Lemma 7.7, and using the fact $N = O(n)$. This finishes the proof. $\qquad \square$

8 Proof of Theorem 1.1

Recall that the key step in Girko's method is the integrability of $\log(\cdot)$ with respect to the empirical distribution of the singular values of $S_n^d/\sqrt{d} - zI_n$ for Lebesgue almost every $z \in \mathbb{C}$. From Theorem 2.2 we have quantitive bounds on the smallest singular value of $S_n^d/\sqrt{d} - zI_n$. The conclusion of Theorem 2.6 will show that there are not too many singular values in small intervals near zero. However, we note that Theorem 2.6 holds only for $z \in B_{\mathbb{C}}(0, 1 - \varepsilon)$, where $\varepsilon > 0$ is arbitrary but fixed. So the steps of Girko's method, as stated in Section 2, cannot be carried out. To overcome this difficulty we use the *replacement principle*, already present e.g. in the works of Tao and Vu, see in particular [41, Theorem 2.1]. However, their proof requires control on the small singular values for Lebesgue almost every $z \in \mathbb{C}$. Below we adapt their proof to obtain a version of the replacement principle, which is suited to our purpose. Before stating the result we introduce more definitions.

If $\{X_n\}$ is a sequence of random variables, we say that X_n is bounded in probability if we have

$$\lim_{K \to \infty} \liminf_{n \to \infty} \mathbb{P}(|X_n| \leq K) = 1.$$

Next for a matrix B_n, we denote $\|B_n\|_2$ to be its Frobenius norm, i.e. $\|B_n\|_2 := \sqrt{\mathrm{Tr}\left(B_n^* B_n\right)}$. Now we are ready to state the result on replacement principle.

Lemma 8.1 (Replacement lemma)**.** Let $B_n^{(1)}$ and $B_n^{(2)}$ are two sequences of $n \times n$ random matrices, such that
(i) The expression

$$\frac{1}{n} \left\| B_n^{(1)} \right\|_2^2 + \frac{1}{n} \left\| B_n^{(2)} \right\|_2^2, \tag{8.1}$$

is bounded in probability
and
(ii) For Lebesgue almost all $z \in \mathbb{D} \subset B_{\mathbb{C}}(0, R) \subset \mathbb{C}$, for some domain $\mathbb{D}$ and some R finite,

$$\frac{1}{n} \log |\det(B_n^{(1)} - zI_n)| - \frac{1}{n} \log |\det(B_n^{(2)} - zI_n)| \to 0,$$

in probability.
    Then for every $f \in C_c^2(\mathbb{C})$ supported on $\mathbb{D}$,

$$\int f(z) dL_{B_n^{(1)}}(z) - \int f(z) dL_{B_n^{(2)}}(z) \to 0,$$

in probability.

Since Theorem 2.6 holds for all $z \in B_{\mathbb{C}}(0, 1 - \varepsilon)$, for every $\varepsilon > 0$, we can set $\mathbb{D}_\varepsilon := B_{\mathbb{C}}(0, 1 - \varepsilon)$ and apply Lemma 8.1 to conclude that $\int f dL_{S_n^d/\sqrt{d}} \to \frac{1}{2\pi} \int f d\mathfrak{m}$ for all smooth functions f supported on $\mathbb{D}_\varepsilon$, where we recall $\mathfrak{m}(\cdot)$ is the Lebesgue measure on $\mathbb{C}$. Since $\varepsilon > 0$ is arbitrary and the circular law is supported on $B_{\mathbb{C}}(0, 1)$, the above is enough to conclude the weak convergence of $L_{S_n^d/\sqrt{d}}$ (for more details see the proof of Theorem 1.1).

We now turn our attention to the proof of Lemma 8.1. A key tool is the following dominated convergence theorem.

Lemma 8.2. ([41, Lemma 3.1]) Let $(\mathcal{X}, \mu)$ be a finite measure space. For each integer $n \geq 1$, let $f_n : \mathcal{X} \to \mathbb{R}$ be random functions which are jointly measurable with respect to $\mathcal{X}$ and the underlying probability space. Assume that
(i) There exists $\delta > 0$ such that $\int_{\mathcal{X}} |f_n(x)|^{1+\delta} d\mu(x)$ is bounded in probability.
(ii) For μ-almost every $x \in \mathcal{X}$, $f_n(x)$ converges to zero in probability.
Then $\int_{\mathcal{X}} f_n(x) d\mu(x)$ converges to zero in probability.

With the help of Lemma 8.2, one can check that the proof of Lemma 8.1 actually follows from an easy adaptation of the alternative proof of [41, Theorem 2.1] sketched in [41, Section 3.6]. We provide a short proof for completeness.

Proof of Lemma 8.1. From (2.1), it follows that for any $f \in C_c^2(\mathbb{C})$,

$$\int f(z) dL_{B_n^{(1)}}(z) - \int f(z) dL_{B_n^{(2)}}(z) \tag{8.2}$$

$$= \frac{1}{2\pi n} \int \Delta f(z) \Big(\log |\det(B_n^{(1)} - zI_n)| - \log |\det(B_n^{(2)} - zI_n)| \Big) d\mathfrak{m}(z).$$

Set $\mathcal{X} := \mathbb{D}$,

$$f_n(z) := \frac{1}{2\pi n} \Delta f(z) \Big(\log |\det(B_n^{(1)} - zI_n)| - \log |\det(B_n^{(2)} - zI_n)| \Big),$$

and μ to be the Lebesgue measure on $\mathbb{D}$ in Lemma 8.2. We see that with these choices the assumption (ii) of Lemma 8.2 is satisfied. To prove assumption (i) of Lemma 8.2 note that, for any $\lambda \in \mathbb{C}$,

$$\int |\Delta f(z)|^2 \Big(\log |\lambda - z| \Big)^2 d\mathfrak{m}(z)$$

$$\leq \int_{z \in B_\mathbb{C}(\lambda,1)} |\Delta f(z)|^2 \Big(\log |\lambda - z| \Big)^2 d\mathfrak{m}(z) + \int_{z \notin B_\mathbb{C}(\lambda,1)} |\Delta f(z)|^2 2(|\lambda|^2 + |z|^2) d\mathfrak{m}(z)$$

$$\leq C(1 + |\lambda|^2),$$

for some positive finite constant C depending on f. Here the last step follows from the fact that $f \in C_c^2(\mathbb{C})$. Therefore using Cauchy–Schwarz inequality, denoting $\lambda_i^{B_n^{(j)}}, i = 1, 2, \ldots, n$, to be the eigenvalues of $B_n^{(j)}$, for $j = 1, 2$, we have that

$$\int_\mathcal{X} |f_n(z)|^2 d\mathfrak{m}(z) \leq C' \left(1 + \frac{1}{n} \sum_{i=1}^n |\lambda_i^{B_n^{(1)}}|^2 + \frac{1}{n} \sum_{i=1}^n |\lambda_i^{B_n^{(2)}}|^2 \right),$$

for some another positive finite constant C'. Finally, using assumption (i) of Lemma 8.1, and Weyl's comparison inequality for second moment (cf. [41, Lemma A.2]), we see that the assumption (i) of Lemma 8.2 is satisfied. Thus, recalling (8.2), the proof now completes upon applying Lemma 8.2. $\square$

Now we are almost ready to complete the proof of Theorem 1.1. Recall that we earlier mentioned that the control on the Stieltjes transform derived in Theorem 2.6 provides us necessary estimates on the number of singular values near zero. Indeed, the following lemma does that job.

Lemma 8.3. ([26, Lemma 15]) Let μ be a probability measure on $\mathbb{R}$. Then for any real $y > 0$,

$$\mu\Big((-y,y)\Big) \leq 2y|\mathrm{Im}\, G_\mu(iy)|.$$

We now proceed to the proof of Theorem 1.1. The idea behind the proof is the following. From Theorem 2.1 we have that $s_n(S_n^d/\sqrt{d} - z)$ is not very small with large probability. Therefore we can exclude a small region near zero while computing $\langle \mathrm{Log}, \nu_n^z \rangle$ where we recall ν_n^z is be the ESD of $\boldsymbol{S}_n^{d,z}$ and $\boldsymbol{S}_n^{d,z}$ was defined in (2.21). Then we use Theorem 2.6 to show that the integration of $\log(|\cdot|)$ around zero, with respect to the probability measure ν_n^z, is negligible. Using Theorem 2.6 we also deduce that $\{\nu_n^z\}$ converges weakly, which therefore in combination with the last observation yields **Step 2** of Girko's method. Then applying the replacement lemma we finish the proof. Below we make this idea precise.

Proof of Theorem 1.1. Fix $\varepsilon > 0$ and $z \in \mathbb{D}_\varepsilon := B_\mathbb{C}(0, 1-\varepsilon)$. Denote $c_n := e^{-\overline{C}_{2.1}(\log n)^2/\log d}$ and let

$$\Omega'_n := \left\{ s_n \left(\frac{S_n^d}{\sqrt{d}} - z \right) \geq c_n \right\}.$$

Fixing any $\tau > 0$, on the event Ω'_n, we see that

$$\int_{-\tau}^\tau |\log(|x|)| d\nu_n^z(x) = 2 \int_0^\tau |\log(x)| d\nu_n^z(x)$$

$$= 2 \int_{c_n}^\tau |\log(x)| d\nu_n^z(x)$$

$$= 2 \int_{c_n}^{\tau_n} |\log(x)| d\nu_n^z(dx) + 2 \int_{\tau_n}^\tau |\log(x)| d\nu_n^z(dx), \qquad (8.3)$$

where we set $\tau_n := \frac{(\log\log n)^{5/6}}{(\log n)^2}$.

Using Theorem 2.6, Lemma 8.3, Lemma 7.8, and the triangle inequality we obtain

$$\int_{c_n}^{\tau_n} |\log(x)| d\nu_n^z(dx) \leq |\log c_n| \times \nu_n^z((0, \tau_n))$$

$$\leq 2|\log c_n| \cdot \tau_n |\widetilde{m}_n(\mathrm{i}\tau_n)|,$$

$$\leq 2|\log c_n| \cdot \tau_n \left(|\widetilde{m}_\infty(\mathrm{i}\tau_n)| + \widetilde{C}_{2.6} \tau_n^{-3} \cdot \max\left\{\frac{1}{d^{1/2}}, \frac{\log n}{n^{1/4}}\right\} \right)$$

$$\leq 4C_{7.8}|\log c_n| \cdot \tau_n = o(1), \tag{8.4}$$

on the event $\Omega_n \cap \Omega_n'$ (recall the definition of Ω_n from the statement of Theorem 2.6), where we used the fact $d \geq \frac{(\log n)^{12}}{(\log\log n)^4}$.

Next using integration by parts it is easy to check that for any probability measure μ on $\mathbb{R}$ and $0 \leq a_1 < a_2 < 1$,

$$\int_{a_1}^{a_2} |\log(x)| d\mu(x) \leq |\log(a_2)|\mu((0, a_2)) + \int_{a_1}^{a_2} \frac{\mu((0, t))}{t} dt. \tag{8.5}$$

Therefore arguing similarly as above and using (8.5) we further deduce

$$\int_{\tau_n}^{\tau} |\log(x)| d\nu_n^z(x) \leq |\log(\tau)|\nu_n^z((0, \tau)) + \int_{\tau_n}^{\tau} \frac{\nu_n^z((0, t))}{t} dt$$

$$\leq \tau|\log(\tau)||\widetilde{m}_n(\mathrm{i}\tau)| + \int_{\tau_n}^{\tau} |\widetilde{m}_n(\mathrm{i}t)| dt$$

$$\leq \tau|\log(\tau)||\widetilde{m}_\infty(\mathrm{i}\tau)| + \int_{\tau_n}^{\tau} |\widetilde{m}_\infty(\mathrm{i}t)| dt$$

$$+ 2\widetilde{C}_{2.6}\tau|\log(\tau)|\tau_n^{-3} \cdot \max\left\{\frac{1}{d^{1/2}}, \frac{\log n}{n^{1/4}}\right\}$$

$$\leq 2C_{7.8}\tau|\log(\tau)| + 2\widetilde{C}_{2.6}\tau|\log(\tau)|\tau_n^{-3} \cdot \max\left\{\frac{1}{d^{1/2}}, \frac{\log n}{n^{1/4}}\right\} \tag{8.6}$$

on the event $\Omega_n \cap \Omega_n'$. Hence, combining (8.4)-(8.6) from (8.3) we see that for any given $\delta > 0$ there exists a $\tau_\delta := \tau(\delta)$, with the property $\lim_{\delta \to 0} \tau_\delta = 0$, such that

$$\limsup_{n \to \infty} \mathbb{P}\left(\int_{-\tau_\delta}^{\tau_\delta} |\log|x|| d\nu_n^z(x) \geq \delta\right) \leq \limsup_{n \to \infty} \mathbb{P}\left(\left\{\int_{-\tau_\delta}^{\tau_\delta} |\log|x|| d\nu_n^z(x) \geq \delta\right\} \cap \Omega_n \cap \Omega_n'\right) = 0. \tag{8.7}$$

We next recall that Theorem 2.6 also implies that, for any $\delta' > 0$,

$$\lim_{n \to \infty} \mathbb{P}\left(\sup_{\xi = \mathrm{i}\eta: \bar{C}_{2.6}/2 \leq \eta \leq \bar{C}_{2.6}} |\widetilde{m}_n(\xi) - \widetilde{m}_\infty(\xi)| > \delta'\right) = 0.$$

This in particular implies that ν_n^z converges weakly to ν_∞^z, in probability (for example, apply Montel's theorem in conjunction with [3, Theorem 2.4.4(c)]), where ν_∞^z is the probability measure corresponding to the Stieltjes transform $\widetilde{m}_\infty(\xi)$. Therefore

$$\int_{(-R, -\tau_\delta) \cup (\tau_\delta, R)} |\log|x|| d\nu_n^z(x) \to \int_{(-R, -\tau_\delta) \cup (\tau_\delta, R)} |\log|x|| d\nu_\infty^z(x) \qquad \text{in probability}, \tag{8.8}$$

for any R positive. Recall that for $z \in \mathbb{D}_\varepsilon$ the support of ν_∞^z is contained in $[-7, 7]$. On the other hand, using that $\log|x|/|x|$ is decreasing for $|x| > e$, we have that

$$\mathbb{E}\int_{(-R, R)^c} |\log|x|| d\nu_n^z(x) \leq \frac{\log R}{R}\mathbb{E}\int |x| d\nu_n^z(x) \leq C\frac{\log R}{R},$$

where C is an absolute constant, and (5.3) was used in the last inequality. Therefore, choosing $R_\delta := R(\delta)$ sufficiently large we obtain from Markov's inequality that

$$\lim_{\delta \to 0} \limsup_{n \to \infty} \mathbb{P}\left(\left|\int_{(-R_\delta, R_\delta)^c} |\log|x||d\nu_n^z(x) - \int_{(-R_\delta, R_\delta)^c} |\log|x||d\nu_\infty^z(x)\right| > \delta\right) = 0. \quad (8.9)$$

From Lemma 7.8, using Lemma 8.3 and (8.5) one can also check that

$$\int_{-\tau_\delta}^{\tau_\delta} |\log|x||d\nu_\infty^z(x) \le 4C_{7.8}\tau_\delta|\log \tau_\delta|. \quad (8.10)$$

Since $\delta > 0$ is arbitrary and $\tau_\delta \to 0$ as $\delta \to 0$, combining (8.7)-(8.10) we deduce that

$$\langle \text{Log}, \nu_n^z \rangle \to \langle \text{Log}, \nu_\infty^z \rangle, \text{ in probability.} \quad (8.11)$$

Now the remainder of the proof is completed using Lemma 8.1. Indeed, consider A_n the $n \times n$ matrix with i.i.d. centered Gaussian entries with variance one. It is well-known that, for Lebesgue almost all z,

$$\frac{1}{n} \log|\det(A_n/\sqrt{n} - zI_n)| \to \langle \text{Log}, \nu_\infty^z \rangle, \text{ almost surely.} \quad (8.12)$$

For example, one can obtain a proof of (8.12) using [12, Lemma 4.11, Lemma 4.12], [13, Theorem 3.4], and [35, Lemma 3.3].

Thus setting $\mathbb{D} = \mathbb{D}_\varepsilon$, $B_n^{(1)} = S_n^d/\sqrt{d}$, and $B_n^{(2)} = A_n/\sqrt{n}$ in Lemma 8.1 we see that assumption (ii) there is satisfied. The assumption (i) of Lemma 8.1 follows from (5.3). Hence, using Lemma 8.1 and the circular law for i.i.d. complex Gaussian matrices (which follows from e.g. [4], but essentially goes back to Ginibre [23]), we obtain that for every $\varepsilon > 0$ and every $f_\varepsilon \in C_c^2(\mathbb{C})$, supported on $\mathbb{D}_\varepsilon$,

$$\int f_\varepsilon(z)d\mu_n(z) \to \frac{1}{\pi}\int f_\varepsilon(z)d\mathfrak{m}(z), \text{ in probability,} \quad (8.13)$$

where for brevity we denote $\mu_n := L_{S_n^d/\sqrt{d}}$. To finish the proof it now remains to show that one can extend the convergence of (8.13) to all $f \in C_c^2(\mathbb{C})$. That is we need to show that for any $\delta > 0$ and $f \in C_c^2(\mathbb{C})$

$$\mathbb{P}\left(\left|\int f(z)d\mu_n(z) - \frac{1}{\pi}\int_{B_\mathbb{C}(0,1)} f(z)d\mathfrak{m}(z)\right| \ge \delta\right) \to 0 \quad \text{as} \quad n \to \infty. \quad (8.14)$$

To this end, for any $\varepsilon > 0$ define a function $i_\varepsilon \in C_c^2(\mathbb{C})$ such that i_ε is supported on $\mathbb{D}_\varepsilon$, $i_\varepsilon \equiv 1$ on $\mathbb{D}_{2\varepsilon}$ and $i_\varepsilon \in [0,1]$ on $\mathbb{D}_\varepsilon \backslash \mathbb{D}_{2\varepsilon}$. Next fix ε such that $M(1 - (1-2\varepsilon)^2) \le \delta/8$ where $M := \sup_x |f(x)|$. Denote $f_\varepsilon := fi_\varepsilon$ and $\bar{f}_\varepsilon := f - f_\varepsilon$. Applying (8.13) for the function i_ε and the triangle inequality we note that

$$\mathbb{P}\left(\left|\int \bar{f}_\varepsilon(z)d\mu_n(z)\right| \ge \delta/4\right) \le \mathbb{P}\left(\left|\int (1 - i_\varepsilon(z))d\mu_n(z)\right| \ge \frac{\delta}{4M}\right)$$
$$\le \mathbb{P}\left(\left|\int i_\varepsilon(z)d\mu_n(z) - \frac{1}{\pi}\int i_\varepsilon(z)d\mathfrak{m}(z)\right| \ge \frac{\delta}{8M}\right) \to 0, \quad (8.15)$$

as $n \to \infty$, where we have used the fact that

$$\left|\frac{1}{\pi}\int_{B_\mathbb{C}(0,1)} (1 - i_\varepsilon(z))d\mathfrak{m}(z)\right| \le \frac{1}{\pi}\int_{B_\mathbb{C}(0,1)\backslash \mathbb{D}_{2\varepsilon}} d\mathfrak{m}(z) \le \frac{\delta}{8M}, \quad (8.16)$$

by our choice of ε. Therefore combining (8.13), (8.15)-(8.16) and the triangle inequality we deduce

$$\mathbb{P}\left(\left|\int f(z)d\mu_n(z) - \frac{1}{\pi}\int f(z)d\mathfrak{m}(z)\right| \geq \delta\right)$$
$$\leq \mathbb{P}\left(\left|\int f_\varepsilon(z)d\mu_n(z) - \frac{1}{\pi}\int f_\varepsilon(z)d\mathfrak{m}(z)\right| \geq \delta/2\right) + \mathbb{P}\left(\left|\int \bar{f}_\varepsilon(z)d\mu_n(z)\right| \geq \delta/4\right) \to 0,$$

as $n \to \infty$. This completes the proof of the theorem. $\qquad\square$

References

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

[2] Radosław Adamczak, Djalil Chafaï, and Paweł Wolff, *Circular law for random matrices with exchangeable entries*, Random Structures Algorithms **48** (2016), no. 3, 454–479. MR-3481269

[3] Greg W. Anderson, Alice Guionnet, and Ofer Zeitouni, *An introduction to random matrices*, Cambridge Studies in Advanced Mathematics, vol. 118, Cambridge University Press, Cambridge, 2010. MR-2760897

[4] Zhidong Bai, *Circular law*, Ann. Probab. **25** (1997), no. 1, 494–529. MR-1428519

[5] Zhidong Bai and Jack W. Silverstein, *Spectral analysis of large dimensional random matrices*, second ed., Springer Series in Statistics, Springer, New York, 2010. MR-2567175

[6] Anirban Basak and Amir Dembo, *Limiting spectral distribution of sums of unitary and orthogonal matrices*, Electron. Commun. Probab. **18** (2013), no. 69, 19. MR-3091727

[7] Anirban Basak and Mark Rudelson, *The circular law for sparse non-hermitian matrices*, preprint (2017). arXiv:1707.03675

[8] Roland Bauerschmidt, Jiaoyang Huang, and Horng-Tzer Yau, *Local Kesten–Mckay law for random regular graphs*, preprint (2016). arXiv:1609.09052

[9] Florent Benaych-Georges and Antti Knowles, *Lectures on the local semicircle law for Wigner matrices*, To appear in SMF volume Panoramas et Synthéses (2016).

[10] Charles Bordenave, Pietro Caputo, and Djalil Chafaï, *Circular law theorem for random Markov matrices*, Probab. Theory Related Fields **152** (2012), no. 3-4, 751–779. MR-2892961

[11] Charles Bordenave, Pietro Caputo, and Djalil Chafaï, *Spectrum of Markov generators on sparse random graphs*, Comm. Pure Appl. Math. **67** (2014), no. 4, 621–669. MR-3168123

[12] Charles Bordenave and Djalil Chafaï, *Around the circular law*, Probab. Surv. **9** (2012), 1–89. MR-2908617

[13] Paul Bourgade, Horng-Tzer Yau, and Jun Yin, *Local circular law for random matrices*, Probab. Theory Related Fields **159** (2014), no. 3-4, 545–595. MR-3230002

[14] Sourav Chatterjee, *A generalization of the Lindeberg principle*, Ann. Probab. **34** (2006), no. 6, 2061–2076. MR-2294976

[15] Nicholas A. Cook, *Lower bounds for the smallest singular value of structured random matrices*, Ann. Probab., to appear (2018). arXiv:1608.07347

[16] Nicholas A. Cook, *The circular law for random regular digraphs*, preprint (2017). arXiv:1703.05839

[17] Nicholas A. Cook, *Discrepancy properties for random regular digraphs*, Random Structures Algorithms **50** (2017), no. 1, 23–58. MR-3583025

[18] Alan Edelman, *Eigenvalues and condition numbers of random matrices*, SIAM J. Matrix Anal. Appl. **9** (1988), no. 4, 543–560. MR-964668

[19] László Erdős, Benjamin Schlein, and Horng-Tzer Yau, *Semicircle law on short scales and delocalization of eigenvectors for Wigner random matrices*, Ann. Probab. **37** (2009), no. 3, 815–852. MR-2537522

[20] Joshua Feinberg and A. Zee, *Non-Hermitian random matrix theory: method of Hermitian reduction*, Nuclear Phys. B **504** (1997), no. 3, 579–608. MR-1488584

[21] Peter J. Forrester, *Log-gases and random matrices*, London Mathematical Society Monographs Series, vol. 34, Princeton University Press, Princeton, NJ, 2010. MR-2641363

[22] Joel Friedman, Jeff Kahn, and Endre Szemerédi, *On the second eigenvalue of random regular graphs*, Proceedings of the twenty-first annual ACM symposium on Theory of computing (New York, NY, USA), STOC, 1989, pp. 587–598.

[23] Jean Ginibre, *Statistical ensembles of complex, quaternion, and real matrices*, Journal of Mathematical Physics **6** (1965), 440–449. MR-0173726

[24] Vyacheslav L. Girko, *The circular law*, Teor. Veroyatnost. i Primenen. **29** (1984), no. 4, 669–679. MR-773436

[25] Friedrich Götze and Alexander Tikhomirov, *The circular law for random matrices*, Ann. Probab. **38** (2010), no. 4, 1444–1491. MR-2663633

[26] Alice Guionnet, Manjunath Krishnapur, and Ofer Zeitouni, *The single ring theorem*, Ann. of Math. (2) **174** (2011), no. 2, 1189–1217. MR-2831116

[27] Svante Janson, *Random regular graphs: asymptotic distributions and contiguity*, Combin. Probab. Comput. **4** (1995), no. 4, 369–405. MR-1377557

[28] Michel Ledoux, *The concentration of measure phenomenon*, Mathematical Surveys and Monographs, vol. 89, American Mathematical Society, Providence, RI, 2001. MR-1849347

[29] Alexander E. Litvak, Anna Lytova, Konstantin Tikhomirov, Nicole Tomczak-Jaegermann, and Pierre Youssef, *Adjacency matrices of random digraphs: singularity and anti-concentration*, Journal of Mathematical Analysis and Applications **445** (2017), no. 2, 1447–1491. MR-3545253

[30] Madan L. Mehta, *Random matrices and the statistical theory of energy levels*, Academic Press, New York-London, 1967. MR-0220494

[31] Michael S. O. Molloy, Hanna D. Robalewska, Robert W. Robinson, and Nicholas C. Wormald, 1-*factorizations of random regular graphs*, Random Structures Algorithms **10** (1997), no. 3, 305–321. MR-1606218

[32] Hoi H. Nguyen, *Random doubly stochastic matrices: the circular law*, Ann. Probab. **42** (2014), no. 3, 1161–1196. MR-3189068

[33] Hoi H. Nguyen and Van H. Vu, *Circular law for random discrete matrices of given row sum*, J. Comb. **4** (2013), no. 1, 1–30. MR-3064040

[34] Leonid A. Pastur, *The spectrum of random matrices*, Teoret. Mat. Fiz. **10** (1972), no. 1, 102–112. MR-0475502

[35] Mark Rudelson, *Invertibility of random matrices: norm of the inverse*, Ann. of Math. (2) **168** (2008), no. 2, 575–600. MR-2434885

[36] Mark Rudelson and Roman Vershynin, *The Littlewood-Offord problem and invertibility of random matrices*, Advances in Mathematics **218** (2008), no. 2, 600–633. MR-2407948

[37] Mark Rudelson and Roman Vershynin, *Invertibility of random matrices: unitary and orthogonal perturbations*, J. Amer. Math. Soc. **27** (2014), no. 2, 293–338. MR-3164983

[38] Mark Rudelson and Ofer Zeitouni, *Singular values of Gaussian matrices and permanent estimators*, Random Structures & Algorithms **48** (2016), no. 1, 183–212. MR-3432577

[39] Terence Tao, *Topics in random matrix theory*, Graduate Studies in Mathematics, vol. 132, American Mathematical Society, Providence, RI, 2012. MR-2906465

[40] Terence Tao and Van Vu, *Random matrices: the circular law*, Commun. Contemp. Math. **10** (2008), no. 2, 261–307. MR-2409368

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

[42] Eugene P. Wigner, *Characteristic vectors of bordered matrices with infinite dimensions*, Ann. of Math. (2) **62** (1955), 548–564. MR-0077805

[43] Eugene P. Wigner, *On the distribution of the roots of certain symmetric matrices*, Ann. of Math. (2) **67** (1958), 325–327. MR-0095527

[44] Philip Matchett Wood, *Universality and the circular law for sparse random matrices*, Ann. Appl. Probab. **22** (2012), no. 3, 1266–1300. MR-2977992

Electronic Journal of Probability
Electronic Communications in Probability

Advantages of publishing in EJP-ECP

- Very high standards

- Free for authors, free for readers

- Quick publication (no backlog)

- Secure publication (LOCKSS[1])

- Easy interface (EJMS[2])

Economical model of EJP-ECP

- Non profit, sponsored by IMS[3], BS[4] , ProjectEuclid[5]

- Purely electronic

Help keep the journal free and vigorous

- Donate to the IMS open access fund[6] (click here to donate!)

- Submit your best articles to EJP-ECP

- Choose EJP-ECP over for-profit journals

---

[1]LOCKSS: Lots of Copies Keep Stuff Safe http://www.lockss.org/
[2]EJMS: Electronic Journal Management System http://www.vtex.lt/en/ejms.html
[3]IMS: Institute of Mathematical Statistics http://www.imstat.org/
[4]BS: Bernoulli Society http://www.bernoulli-society.org/
[5]Project Euclid: https://projecteuclid.org/
[6]IMS Open Access Fund: http://www.imstat.org/publications/open.htm