Spectrum of Markov generators on sparse random graphs

Charles Bordenave, Pietro Caputo, Djalil Chafai

SPECTRUM OF MARKOV GENERATORS ON SPARSE RANDOM GRAPHS

CHARLES BORDENAVE, PIETRO CAPUTO, AND DJALIL CHAFAÏ

Abstract. We investigate the spectrum of the infinitesimal generator of the continuous time random walk on a randomly weighted oriented graph. This is the non-Hermitian random $n \times n$ matrix L defined by $L_{jk} = X_{jk}$ if $k \neq j$ and $L_{jj} = -\sum_{k \neq j} L_{jk}$, where $(X_{jk})_{j \neq k}$ are i.i.d. random weights. Under mild assumptions on the law of the weights, we establish convergence as $n \to \infty$ of the empirical spectral distribution of L after centering and rescaling. In particular, our assumptions include sparse random graphs such as the oriented Erdős-Rényi graph where each edge is present independently with probability $p(n) \to 0$ as long as $np(n) \gg (\log(n))^6$. The limiting distribution is characterized as an additive Gaussian deformation of the standard circular law. In free probability terms, this coincides with the Brown measure of the free sum of the circular element and a normal operator with Gaussian spectral measure. The density of the limiting distribution is analyzed using a subordination formula. Furthermore, we study the convergence of the invariant measure of L to the uniform distribution and establish estimates on the extremal eigenvalues of L.

Contents

1. Introduction 2
1.1. Convergence results 4
1.2. Limiting distribution 4
1.3. Extremal eigenvalues and the invariant measure 5
1.4. Comments and remarks 6
2. Convergence of singular values: Proof of theorem 1.1 7
2.1. Concentration of singular values measure 7
2.2. Centralization and truncation 7
2.3. Tightness 9
2.4. Linearization 9
2.5. Approximate fixed point equation 10
2.6. Uniqueness of the fixed point equation 12
3. Convergence of eigenvalues: Proof of theorem 1.2 12
3.1. Strategy of proof 12
3.2. Large singular values 13
3.3. Smallest singular value 13
3.4. Moderately small singular values 20
3.5. Proof of theorem 3.2 21
4. Limiting distribution: Proof of theorems 1.3 and 1.4 21
4.1. Brown measure 21
4.2. Proof of theorem 1.3 22
4.3. Quaternionic resolvent 22
4.4. Proof of theorem 1.4 24
5. Extremal eigenvalues: Proof of theorems 1.5 and 1.6 26
5.1. Proof of theorem 1.5 26
5.2. Proof of theorem 1.6 27
6. Invariant measure: Proof of theorem 1.7 27
Appendix A. Concentration function and small ball probabilities 28
Acknowledgments 32
References 32

Date: Preprint – February 2012 – Compiled February 3, 2012.
2000 *Mathematics Subject Classification.* 05C80 (05C81 15B52 46L54 47A10 60B20).
Key words and phrases. Random graphs; Random matrices; Free probability; Combinatorics; Spectral Analysis.

1. Introduction

For each integer $n \geq 1$ let $X = (X_{ij})_{1 \leq i,j \leq n}$ be the random matrix whose entries are i.i.d. copies of a complex valued random variable $\mathbf{x}$ with variance σ^2. The circular law theorem (see e.g. the survey papers [40, 12]) asserts that the empirical spectral distribution of X - after centering and rescaling by $\sigma\sqrt{n}$ - converges weakly to the uniform distribution on the unit disc of $\mathbb{C}$. The sparse regime is obtained by allowing the law of $\mathbf{x}$ to depend on n with a variance satisfying $\sigma^2(n) \to 0$ as $n \to \infty$. As an example, if $\mathbf{x} = \epsilon(n)$ is a Bernoulli random variable with parameter $p(n)$, then X is the adjacency matrix of the oriented Erdős-Rényi random graph where each edge is present independently with probability $p(n)$. In this example $\sigma^2(n) = p(n)(1-p(n)) \sim p(n)$ when $p(n) \to 0$. It is expected that the circular law continues to hold in the sparse regime, as long as $n\sigma^2(n) \to \infty$. Results in this direction have been recently established in [37, 22, 32], where the convergence is proved under some extra assumptions including that $n^{1-\varepsilon}\sigma^2 \to \infty$ for some $\varepsilon > 0$.

In this paper, we consider instead random matrices of the form

$$L = X - D \tag{1.1}$$

where X is a matrix with i.i.d. entries as above, and D is the diagonal matrix obtained from the row sums of X, i.e. for $i = 1, \ldots, n$,

$$D_{ii} = \sum_{k=1}^{n} X_{ik}.$$

If X is interpreted as the adjacency matrix of a weighted oriented graph, then L is the associated Laplacian matrix, with zero row sums. In particular, if the weights X_{ij} take values in $[0, \infty)$, then L is the infinitesimal generator of the continuous time random walk on that graph, and properties of the spectrum of L can be used to study its long-time behavior. Clearly, L has non independent entries but independent rows. A related model is obtained by considering the stochastic matrix $P = D^{-1}X$. The circular law for the latter model in the non sparse regime has been studied in [9]. Here, we investigate the behavior of the spectrum of L both in the non sparse and the sparse regime. We are mostly concerned with three issues:

(1) convergence of the empirical spectral distribution;
(2) properties of the limiting distribution;
(3) invariant measure and extremal eigenvalues of L.

As in the case of the circular law, the main challenge in establishing point (1) is the estimate on the smallest singular value of deterministic shifts of L. This is carried out by combining the method of Rudelson and Vershynin [35], and Götze and Tikhomirov [22], together with some new arguments needed to handle the non independence of the entries of L and the possibility of zero singular values - for instance, L itself is not invertible since all its rows sum to zero. As for point (2) the analysis of the resolvent is combined with free probabilistic arguments to characterize the limiting distribution as the Brown measure of the free sum of the circular element and an unbounded normal operator with Gaussian spectral measure. Further properties of this distribution are obtained by using a subordination formula. This result can be interpreted as the asymptotic independence of X and D. The Hermitian counterpart of these facts has been discussed by Bryc, Dembo and Jiang in [14], who showed that if X is an i.i.d. Hermitian matrix, then the limiting spectral distribution of L - after centering and rescaling - is the free convolution of the semi-circle law with a Gaussian law. Finally, in point (3) we estimate the total variation distance between the invariant measure of L and the uniform distribution. This analysis is based on perturbative arguments similar to those recently used to analyze bounded rank perturbations of matrices with i.i.d. entries [2, 36, 4, 39]. Further perturbative reasoning is used to give bounds on the spectral radius and on the spectral gap of L.

Before stating our main results, we introduce the notation to be used. If A is an $n \times n$ matrix, we denote by $\lambda_1(A), \ldots, \lambda_n(A)$ its eigenvalues, i.e. the roots in $\mathbb{C}$ of its characteristic polynomial. We label them in such a way that $|\lambda_1(A)| \geq \cdots \geq |\lambda_n(A)|$. We denote by $s_1(A), \ldots, s_n(A)$ the singular values of A, i.e. the eigenvalues of the Hermitian positive semidefinite matrix $|A| = \sqrt{A^*A}$, labeled so that $s_1(A) \geq \cdots \geq s_n(A) \geq 0$. The *operator norm* of A is $\|A\| = s_1(A)$ while the *spectral radius*

is $|\lambda_1(A)|$. We define the discrete probability measures

$$\mu_A = \frac{1}{n}\sum_{k=1}^{n} \delta_{\lambda_k(A)} \quad \text{and} \quad \nu_A = \mu_{|A|} = \frac{1}{n}\sum_{k=1}^{n} \delta_{s_k(A)}.$$

We denote "$\rightsquigarrow$" the weak convergence of measures against bounded continuous functions on $\mathbb{R}$ or on $\mathbb{C}$. If μ and (μ_n) are random finite measures on $\mathbb{R}$ or on $\mathbb{C}$, we say that $\mu_n \rightsquigarrow \mu$ *in probability* when for every bounded continuous function f and for every $\varepsilon > 0$, $\lim_{n\to\infty} \mathbb{P}(\left|\int f\, d\mu_n - \int f\, d\mu\right| > \varepsilon) = 0$. In this paper, all the random variables are defined on a common probability space $(\Omega, \mathcal{A}, \mathbb{P})$. For each integer $n \geq 1$, let $\mathbf{x} = \mathbf{x}(n)$ denote a complex valued random variable with law $\mathcal{L} = \mathcal{L}(n)$ possibly dependent on n, with variance

$$\sigma^2(n) = \mathrm{Var}(\mathbf{x}) = \mathbb{E}(|\mathbf{x} - \mathbb{E}\mathbf{x}|^2) = \mathrm{Var}(\mathfrak{Re}\,\mathbf{x}) + \mathrm{Var}(\mathfrak{Im}\,\mathbf{x}),$$

and mean $m(n) = \mathbb{E}\mathbf{x}$. Throughout the paper, it is always assumed that $\mathbf{x}$ satisfies

$$\lim_{n\to\infty} n\sigma^2(n) = \infty, \quad \sup_n \sigma^2(n) < \infty, \tag{1.2}$$

and the Lindeberg type condition:

$$\text{For all } \varepsilon > 0, \quad \lim_{n\to\infty} \sigma(n)^{-2}\mathbb{E}\big[|\mathbf{x} - m(n)|^2 \mathbf{1}_{\{|\mathbf{x}-m(n)|^2 \geq \varepsilon n\sigma^2(n)\}}\big] = 0; \tag{1.3}$$

It is also assumed that the normalized covariance matrix converges:

$$\lim_{n\to\infty} \sigma(n)^{-2}\mathrm{Cov}(\mathfrak{Re}(\mathbf{x}), \mathfrak{Im}(\mathbf{x})) = K \tag{1.4}$$

for some 2×2 matrix K. This allows the real and imaginary parts of $\mathbf{x}$ to be correlated. Note that the matrix K has unit trace by construction.

The two main examples we have in mind are:

A) non sparse case: $\mathbf{x}$ has law $\mathcal{L}$ independent of n with finite positive variance;
B) sparse case:

$$\mathbf{x} = \epsilon(n)\mathbf{y}, \tag{1.5}$$

with $\mathbf{y}$ a bounded random variable with law independent of n and $\epsilon(n)$ an independent Bernoulli variable with parameter $p(n)$ satisfying $p(n) \to 0$ and $np(n) \to \infty$.

These cases are referred to as model A and model B in the sequel. It is immediate to check that either case satisfies the assumptions (1.2), (1.3) and (1.4). We will sometimes omit the script n in $\mathbf{x}(n)$, $\mathcal{L}(n)$, $m(n)$, $\sigma(n)$, etc. The matrix L is defined by (1.1), and we consider the rescaled matrix

$$M = \frac{L + nmI}{\sigma\sqrt{n}} = \frac{X}{\sigma\sqrt{n}} - \frac{D - nmI}{\sigma\sqrt{n}}. \tag{1.6}$$

By the central limit theorem, the distribution of $(D_{ii} - nm)/(\sigma\sqrt{n})$ converges to the Gaussian law with mean 0 and covariance K. Combined with the circular law for $X/(\sigma\sqrt{n})$, this suggests the interpretation of the spectral distribution of M, in the limit $n \to \infty$, as an additive Gaussian deformation of the circular law.

Define $\mathbb{C}_+ = \{z \in \mathbb{C} : \mathfrak{Im}(z) > 0\}$. If ν is a probability measure on $\mathbb{R}$ then its Cauchy-Stieltjes transform is the analytic function $S_\nu : \mathbb{C}_+ \to \mathbb{C}_+$ given for any $z \in \mathbb{C}_+$ by

$$S_\nu(z) = \int \frac{1}{t - z}\, d\nu(t). \tag{1.7}$$

We denote by $\check{\nu}$ the symmetrization of ν, defined for any Borel set A of $\mathbb{R}$ by

$$\check{\nu}(A) = \frac{\nu(A) + \nu(-A)}{2}. \tag{1.8}$$

If ν is supported in $\mathbb{R}_+$ then ν is characterized by its symmetrization $\check{\nu}$. In the sequel, G is a Gaussian random variable on $\mathbb{R}^2 \cong \mathbb{C}$ with law $\mathcal{N}(0, K)$ i.e. mean 0 and covariance matrix K. This law has a Lebesgue density on $\mathbb{R}^2$ if and only if K is invertible, given by $(2\pi\sqrt{\det(K)})^{-1}\exp(-\frac{1}{2}\langle K^{-1}\cdot, \cdot\rangle)$.

1.1. **Convergence results.** We begin with the singular values of shifts of the matrix M, a useful proxy to the eigenvalues.

Theorem 1.1 (Singular values)**.** *For every $z \in \mathbb{C}$, there exists a probability measure ν_z on $\mathbb{R}_+$ which depends only on z and K such that with probability one,*

$$\nu_{M-zI} \underset{n\to\infty}{\rightsquigarrow} \nu_z.$$

Moreover, the limiting law ν_z is characterized as follows: $\check{\nu}_z$ is the unique symmetric probability measure on $\mathbb{R}$ with Cauchy-Stieltjes transform satisfying, for every $\eta \in \mathbb{C}_+$,

$$S_{\check{\nu}_z}(\eta) = \mathbb{E}\bigg(\frac{S_{\check{\nu}_z}(\eta) + \eta}{|G - z|^2 - (\eta + S_{\check{\nu}_z}(\eta))^2}\bigg). \tag{1.9}$$

The next result concerns the eigenvalues of M. On top of our running assumptions (1.2), (1.3) and (1.4), here we need to assume further:

(i) variance growth:

$$\lim_{n\to\infty} \frac{n\sigma^2(n)}{(\log(n))^6} = +\infty; \tag{1.10}$$

(ii) tightness conditions:

$$\lim_{t\to\infty} \sup_{n\geq 1} \frac{\mathbb{E}\Big[|\mathbf{x}|^2 \mathbf{1}_{\{\frac{1}{t} < |\mathbf{x}| < t\}}\Big]}{\mathbb{E}[|\mathbf{x}|^2]} = 1 \quad \text{and} \quad \sup_{n\geq 1} \frac{\mathbb{E}\big[|\mathbf{x}|^2\big]}{\sigma^2(n)} < \infty; \tag{1.11}$$

(iii) the set Λ of accumulation points of $(\sqrt{n}\, m(n)/\sigma(n))_{n\geq 1}$ has zero Lebesgue measure in $\mathbb{C}$.

It is not hard to check that assumptions (i),(ii),(iii) are all satisfied by model A and model B, provided in B we require that $p(n) \gg (\log(n))^6/n$.

Theorem 1.2 (Eigenvalues)**.** *Assume that (i),(ii),(iii) above hold. Let μ be the probability measure on $\mathbb{C}$ defined by*

$$\mu = -\frac{1}{2\pi}\Delta U \quad \textit{with} \quad U(z) := -\int_0^\infty \log(t)\, d\nu_z(t),$$

where the Laplacian $\Delta = \partial_x^2 + \partial_y^2$ is taken in the sense of Schwartz-Sobolev distributions in the space $\mathcal{D}'(\mathbb{R}^2)$, and where ν_z is as in theorem 1.1. Then, in probability,

$$\mu_M \underset{n\to\infty}{\rightsquigarrow} \mu.$$

1.2. **Limiting distribution.** The limiting distribution in theorem 1.2 is independent of the mean m of the law $\mathcal{L}$. This is rather natural since shifting the entries produces a deterministic rank one perturbation. As in other known circumstances, a rank one additive perturbation produces essentially a single outlier, and therefore does not affect the limiting spectral distribution, see e.g. [2, 36, 16, 39]. To obtain further properties of the limiting distribution, we turn to free probability.

A non-commutative probability space is a pair $(\mathcal{M}, \tau)$ where $\mathcal{M}$ is a von Neumann algebra of bounded operators and τ is a normal, faithful, tracial state on $\mathcal{M}$. In the present work, we need to deal with possibly unbounded operators in order to interpret $\sigma^{-1}n^{-1/2}(D - nmI)$ when $n \to \infty$. Following Brown [13] and Haagerup and Schultz [26], one can associate to every element a in a class $\bar{\mathcal{M}} \supset \mathcal{M}$ of closed densely defined operators affiliated with $\mathcal{M}$, a probability measure μ_a on $\mathbb{C}$, called the Brown spectral measure. For normal operators the Brown measure coincides with the usual spectral measure. The notion of $\star$-free operators still makes sense in $(\bar{\mathcal{M}}, \tau)$ even if the elements of $\bar{\mathcal{M}}$ are not necessarily bounded. We refer to section 4.1 below for precise definitions in our setting and to [26] for a complete treatment. We use the standard notation $|a| = \sqrt{a^*a}$ for the square root of the non negative self-adjoint operator a^*a.

Theorem 1.3 (Free probability interpretation of the limiting laws)**.** *Let c and g be $\star$-free operators in $(\bar{\mathcal{M}}, \tau)$, with c circular, and g normal with spectral measure equal to $\mathcal{N}(0, K)$. Then, if ν_z and μ are as in theorems 1.1-1.2, we have*

$$\nu_z = \mu_{|c+g-z|} \quad \textit{and} \quad \mu = \mu_{c+g}.$$

Having identified the limit law μ, we obtain some additional information on it.

Theorem 1.4 (Properties of the limiting measure)**.** *Let c and g be as in theorem 1.3. The support of the Brown measure μ_{c+g} of $c+g$ is given by*

$$\mathrm{supp}(\mu_{c+g}) = \left\{ z \in \mathbb{C} : \mathbb{E}\left(\frac{1}{|G-z|^2} \right) \geq 1 \right\}.$$

There exists a unique function $f : \mathrm{supp}(\mu_{c+g}) \to [0,1]$ such that for all $z \in \mathrm{supp}(\mu_{c+g})$,

$$\mathbb{E}\left[\frac{1}{|G-z|^2 + f(z)^2} \right] = 1.$$

Moreover, f is C^∞ in the interior of $\mathrm{supp}(\mu_{c+g})$*, and letting $\Phi(w,z) := (|w-z|^2 + f(z)^2)^{-2}$, the probability measure μ_{c+g} is absolutely continuous with density given by*

$$z \mapsto \frac{1}{\pi} f(z)^2 \mathbb{E}[\Phi(G,z)] + \frac{1}{\pi} \frac{\left| \mathbb{E}[(G-z)\Phi(G,z)] \right|^2}{\mathbb{E}[\Phi(G,z)]}. \tag{1.12}$$

It can be seen that μ is rotationally invariant when K is a multiple of the identity, while this is not the case if $\mathrm{supp}(\mathcal{L}) \subset \mathbb{R}$, in which case $K_{22} = K_{12} = K_{21} = 0$ (in this case G does not have a density on $\mathbb{C}$ since K is not invertible). Figures 1.1-1.2 provide numerical simulations illustrating this phenomenon in two special cases. Note also that the support of μ_{c+g} is not bounded since it contains the support of $\mathcal{N}(0,K)$. Thus, $\mathrm{supp}(\mu_{c+g}) = \mathbb{C}$ if K is invertible. If K is not invertible, it can be checked that the boundary of $\mathrm{supp}(\mu_{c+g})$ is

$$\left\{ z \in \mathbb{C} : \mathbb{E}\left(\frac{1}{|G-z|^2} \right) = 1 \right\}$$

On this set, $f(z) = 0$, but from (1.12), we see that the density does not vanish there. This phenomenon, not unusual for Brown measures, occurs for the circular law and more generally for R-diagonal operators, see Haagerup and Larsen [25].

The formula (1.12) is slightly more explicit than the formulas given in Biane and Lehner [7, Section 5]. Equation (1.12) will be obtained via a subordination formula for the circular element (forthcoming proposition 4.3) in the spirit of the works of Biane [6] or Voiculescu [42]. This subordination formula can also be used to compute more general Brown measures of the form μ_{a+c} with a, c $\star$-free and c circular.

1.3. **Extremal eigenvalues and the invariant measure.** Theorem 1.2 suggests that the bulk of the spectrum of L is concentrated around the value $-mn$ in a two dimensional window of width $\sigma\sqrt{n}$. Actually, it is possible to localize more precisely the support of the spectrum, by controlling the extremal eigenvalues of L. Recall that L has always the trivial eigenvalue 0. Theorem 1.5 below describes the positions of the remaining eigenvalues. For simplicity, we restrict our analysis to the Markovian case in which $\mathrm{supp}(\mathcal{L}) \subset [0,\infty)$ and to either model A or B. Analogous statements hold however in the general case. Note that we have here $m > 0$ and $K_{11} = 1$ while $K_{22} = K_{12} = K_{21} = 0$ (in particular, K is not invertible). We define for convenience the centered random matrices

$$\underline{X} = X - \mathbb{E}X, \quad \underline{D} = D - \mathbb{E}D, \quad \underline{L} = L - \mathbb{E}L = \underline{X} - \underline{D}. \tag{1.13}$$

If J stands for the $n \times n$ matrix with all entries equal to 1, then we have

$$\mathbb{E}L = L - \underline{L} = mJ - mnI.$$

Theorem 1.5 (Spectral support for model A)**.** *Assume model A, that* $\mathrm{supp}(\mathcal{L}) \subset \mathbb{R}_+$ *and that* $\mathbb{E}|\mathbf{x}|^4 < \infty$*. Then with probability one, for $n \gg 1$, every eigenvalue $\underline{\lambda}$ of $\underline{L}$ satisfies*

$$|\mathfrak{Re}\underline{\lambda}| \leq \sigma\sqrt{2n\log(n)}\,(1+o(1)) \quad \textit{and} \quad |\mathfrak{Im}\underline{\lambda}| \leq \sigma\sqrt{n}(2+o(1)). \tag{1.14}$$

Moreover, with probability one, for $n \gg 1$, every eigenvalue $\lambda \neq 0$ of L satisfies

$$|\mathfrak{Re}\lambda + mn| \leq \sigma\sqrt{2n\log(n)}\,(1+o(1)) \quad \textit{and} \quad |\mathfrak{Im}\lambda| \leq \sigma\sqrt{n}(2+o(1)). \tag{1.15}$$

To interpret the above result, recall that Yin and Bai [44, theorem 2] prove that, in model A, if $\mathbb{E}|\mathbf{x}|^4 < \infty$ then the operator norm of $\underline{X}$ is $\sigma\sqrt{n}\,(2+o(1))$. On the other hand, from the central limit theorem one expects that the operator norm and the spectral radius of the diagonal (thus normal) matrix $\underline{D}$ are of order $\sigma\sqrt{2n\log(n)}\,(1+o(1))$ (as for maximum of i.i.d. Gaussian random

variables). Note that if one defines a *spectral gap* κ of the Markov generator L as the minimum of $|\mathfrak{Re}\lambda|$ for $\lambda \neq 0$ in the spectrum of L, then by theorem 1.5 one has a.s.

$$\kappa \geq mn - \sigma\sqrt{2n\log(n)}\,(1+o(1)). \tag{1.16}$$

In theorem 1.5, we have restricted our attention to model A to be in position to use Yin and Bai [44]. Beyond model A, their proof cannot be extended to laws which do not satisfy the assumption $\mathbb{E}(|\mathbf{x}-m|^4) = \mathcal{O}(\sigma^4)$. The latter will typically not hold when $\sigma(n)$ goes to 0. For example, in model B, one has $\sigma^2(n) \sim p(n)\mathbb{E}|\mathbf{y}|^2$ and $\mathbb{E}(|\mathbf{x}-m|^4) \sim p(n)\mathbb{E}|\mathbf{y}|^4$. In this situation, we have the following result.

Theorem 1.6 (Spectral support for model B)**.** *Assume model B, that* $\mathrm{supp}(\mathcal{L}) \subset \mathbb{R}_+$ *has bounded diameter, and that*

$$\lim_{n\to\infty} \frac{n\sigma^2}{\log(n)} = +\infty; \tag{1.17}$$

Then with probability one, for $n \gg 1$*, every eigenvalue* $\underline{\lambda}$ *of* $\underline{L}$ *satisfies*

$$|\mathfrak{Re}\underline{\lambda}| \leq (2+o(1))\sigma\sqrt{n\log(n)} + \mathcal{O}\Big(\sigma^{\frac{1}{2}} n^{\frac{1}{4}}\log(n)\Big) \tag{1.18}$$
$$\text{and} \qquad |\mathfrak{Im}\underline{\lambda}| \leq (2+o(1))\sigma\sqrt{n} + \mathcal{O}\Big(\sigma^{\frac{1}{2}} n^{\frac{1}{4}}\log(n)\Big).$$

Moreover, with probability one, for $n \gg 1$*, every eigenvalue* $\lambda \neq 0$ *of* L *satisfies*

$$|\mathfrak{Re}\lambda + mn| \leq (2+o(1))\sigma\sqrt{n\log(n)} + \mathcal{O}\Big(\sigma^{\frac{1}{2}} n^{\frac{1}{4}}\log(n)\Big) \tag{1.19}$$
$$\text{and} \qquad |\mathfrak{Im}\lambda| \leq (2+o(1))\sigma\sqrt{n} + \mathcal{O}\Big(\sigma^{\frac{1}{2}} n^{\frac{1}{4}}\log(n)\Big).$$

Note that $\sigma^{\frac{1}{2}} n^{\frac{1}{4}}\log(n) = o(\sigma\sqrt{n\log(n)})$ whenever (1.17) is strenghtened to $n\sigma^2 \gg (\log(n))^2$. The term $\sigma^{\frac{1}{2}} n^{\frac{1}{4}}\log(n)$ comes in our proof from an estimate of Vu [43] on the norm of sparse matrices with independent bounded entries.

We turn to the properties of the invariant measure of L. If $\mathrm{supp}(\mathcal{L}) \subset \mathbb{R}_+$ and L is irreducible, then from the Perron-Frobenius theorem, the kernel of L has dimension 1 and there is a unique vector $\Pi \in (0,1)^n$ such that $L^\top \Pi = 0$ and $\sum_{i=1}^n \Pi_i = 1$. The vector Π is the invariant measure of the Markov process with infinitesimal generator L.

Theorem 1.7 (Invariant measure)**.** *Assume that either the assumptions of theorems 1.5 or 1.6 hold. Then, a.s. for* $n \gg 1$*, the Markov generator* L *is irreducible and*

$$\|\Pi - U_n\|_{\mathrm{TV}} = \mathcal{O}\left(\frac{\sigma}{m}\sqrt{\frac{\log(n)}{n}}\right) + \mathcal{O}\left(\frac{\sqrt{\sigma}}{m}\frac{\log(n)}{n^{3/4}}\right),$$

where $U_n = \frac{1}{n}(1,\dots,1)^\top$ *is the uniform probability distribution on the finite set* $\{1,\dots,n\}$ *and* $\|Q\|_{\mathrm{TV}} := \frac{1}{2}\sum_{i=1}^n |Q_i|$ *is the total variation norm.*

1.4. **Comments and remarks.** We conclude the introduction with a list of comments and open questions.

1.4.1. *Interpolation.* A first observation is that all our results for the matrix $L = X - D$ can be extended with minor modifications to the case of the matrix $L^{(\delta)} = X - \delta D$, where $\delta \in \mathbb{R}$ is independent of n, provided the law $\mathcal{N}(0,K)$ characterizing our limiting spectral distributions is replaced by $\mathcal{N}(0,\delta^2 K)$. This gives back the circular law for $\delta = 0$.

1.4.2. *Almost sure convergence.* One expects that the convergence in theorem 1.2 holds almost surely and not simply in probability. This weaker convergence comes from our poor control of the smallest singular value of the random matrix $M - z$. In the special case when the law $\mathcal{L}(n)$ of the entries has a bounded density growing at most polynomially in n, then arguing as in [9], it is possible to prove that the convergence in theorem 1.2 holds almost surely.

1.4.3. *Sparsity.* It is natural to conjecture that theorem 1.2 continues to hold even if (1.10) is replaced by the weaker condition (1.2). However, this is a difficult open problem even in the simpler case of the circular law. To our knowledge, our assumption (1.10) improves over previous works [22, 32], where the variance is assumed to satisfy $n^{1-\varepsilon}\sigma^2 \to \infty$ for some $\varepsilon > 0$. Assumption (1.10) is crucial for the control of the smallest singular value of M. We believe that with some extra effort the power 6 could be reduced. However, some power of $\log(n)$ is certainly needed for the arguments used here. It is worthy of note that theorem 1.1 holds under the minimal assumption (1.2).

1.4.4. *Heavy tails.* A different model for random Markov generators is obtained when the law $\mathcal{L}$ of $\mathbf{x}$ has heavy tails, with e.g. infinite first moment. In this context, we refer e.g. to [11, 12] for the spectral analysis of non-Hermitian matrices with i.i.d. entries, and to [10] for the case of reversible Markov transition matrices. It is natural to expect that, in contrast with the cases considered here, there is no asymptotic independence of the matrices X and D in the heavy tailed case.

1.4.5. *Spectral edge and spectral gap.* Concerning theorem 1.5, it seems natural to conjecture the asymptotic behavior $\kappa = mn - \sigma\sqrt{2n\log(n)}\,(1 + o(1))$ for the spectral gap (1.16), but we do not have a proof of the corresponding upper bound. In the same spirit, in the setting of theorem 1.5 or theorem 1.6 we believe that with probability one, with $\underline{L} = L - \mathbb{E}L$,

$$\lim_{n\to\infty} \frac{s_1(\underline{L})}{\sigma\sqrt{2n\log(n)}} = \lim_{n\to\infty} \frac{|\lambda_1(\underline{L})|}{\sigma\sqrt{2n\log(n)}} = 1,$$

which contrasts with the behavior of $\underline{X}$ for which $s_1/|\lambda_1| \to 2$ as $n \to \infty$ under a finite fourth moment assumption [44].

The rest of the article is structured as follows. Sections 2 and 3 provide the proof of theorem 1.1 and of theorem 1.2 respectively. Section 4 is devoted to the proof of theorem 1.3 and of theorem 1.4. Section 5 gives the proof of theorems 1.5 and 1.6, section 6 contains a proof theorem 1.7. Finally, an appendix collects some facts on concentration function and small probabilities.

2. Convergence of singular values: Proof of theorem 1.1

We adapt the strategy of proof described in [12, section 4.5].

2.1. **Concentration of singular values measure.** A standard separability and density argument shows that it is sufficient to prove that for any compactly supported $\mathcal{C}^\infty(\mathbb{R})$ function f, a.s.

$$\int f\, d\nu_{M-z} - \int f\, \nu_z \to 0.$$

Since the matrix $M - z$ has independent rows, we can rely on the concentration of measure phenomenon for matrices with independent rows, see [11], [12], or [24]. In particular, using [12, lemma 4.18] and the Borel-Cantelli lemma, we obtain that for any compactly supported continuous function f, a.s.

$$\int f\, d\nu_{M-z} - \mathbb{E}\int f\, d\nu_{M-z} \to 0.$$

In other words, in order to prove the first part of theorem 1.1, it is sufficient to prove the convergence to ν_z of the averaged measure $\mathbb{E}\nu_{M-z}$.

2.2. **Centralization and truncation.** We now prove that it is sufficient to prove the convergence for centered entries with bounded support. We first notice that

$$\mathbb{E}M = \sigma^{-1}n^{-1/2}mJ$$

has rank one, where J stands for the matrix with all entries equal to 1. Hence, writing $\underline{M} = M - \mathbb{E}M$, from standard perturbation inequalities (see e.g. [29, Th. 3.3.16]):

$$\left|\int f\, d\nu_{\underline{M}-zI} - \int f\, d\nu_{M-zI}\right| \leq \|f\|_{\mathrm{BV}} \frac{\mathrm{rank}(M - \underline{M})}{n} \leq \frac{\|f\|_{\mathrm{BV}}}{n}, \tag{2.1}$$

where $\|f\|_{\mathrm{BV}} = \int |f'(t)|\, dt$ denotes the bounded variation norm of f. In particular, it is sufficient to prove the convergence of $\mathbb{E}\nu_{\underline{M}-zI}$ to ν_z. Recall the definition (1.13) of the centered matrices $\underline{X}, \underline{D}$. Define

$$X'_{ij} = \underline{X}_{ij} \mathbf{1}_{\{|\underline{X}_{ij}| \le \varepsilon\sigma\sqrt{n}\}},$$

where $\varepsilon = \varepsilon_n$ is a sequence going to 0 such that

$$\lim_{n\to\infty} \mathbb{E}\left[\frac{|X_{ij}-m|^2}{\sigma^2} \mathbf{1}_{\{|X_{ij}-m|^2 \ge \varepsilon^2\sigma^2 n\}} \right] = 0. \tag{2.2}$$

(its existence is guaranteed by assumption (1.3)). Then, let $X' = (X'_{ij})_{1\le i,j\le n}$, $L' = X' - \underline{D}$ and $M' = L'/(\sigma\sqrt{n})$. From Hoffman-Wielandt inequality, we have

$$\begin{aligned} \frac{1}{n}\sum_{k=1}^{n} |s_k(M'-zI) - s_k(\underline{M}-zI)|^2 &\le \frac{1}{n} \sum_{1\le i,j\le n} |\underline{M}_{ij} - M'_{ij}|^2 \\ &= \frac{1}{n^2\sigma^2} \sum_{1\le i,j\le n} |\underline{X}_{ij}|^2 \mathbf{1}_{\{|\underline{X}_{ij}| > \varepsilon\sigma\sqrt{n}\}} \end{aligned}$$

Then by (2.2), we deduce that

$$\mathbb{E}\, \frac{1}{n}\sum_{k=1}^{n} |s_k(M'-zI) - s_k(\underline{M}-zI)|^2 \to 0.$$

The left hand side above is the square of the expected Wasserstein W_2 coupling distance between $\nu_{M'-zI}$ and $\nu_{\underline{M}-zI}$. Since the convergence in W_2 distance implies weak convergence, we deduce that it is sufficient to prove the convergence of $\mathbb{E}\nu_{M'-zI}$ to ν_z. We then center the entries of X', and set $\widetilde{L} = L' - m'J$ and $\widetilde{M} = \widetilde{L}/(\sigma\sqrt{n})$, where $m' = \mathbb{E}X'_{11}$. As in (2.1), we find

$$\left| \int f\, d\nu_{\widetilde{M}-zI} - \int f\, d\nu_{M'-zI} \right| \le \|f\|_{\mathrm{BV}} \frac{\mathrm{rank}(\widetilde{M}-M')}{n} \le \frac{\|f\|_{\mathrm{BV}}}{n}.$$

Finally, consider the generator associated to $X'' := X' - \mathbb{E}X'$. Namely, define the matrices $L'' = X'' - D''$ and $M'' = L''/(\sigma\sqrt{n})$, where D'' is the diagonal matrix with

$$D''_{ii} = \sum_j (X'')_{ij} = \underline{D}_{ii} - \sum_j \underline{X}_{ij} \mathbf{1}_{\{|\underline{X}_{ij}| > \varepsilon\sigma\sqrt{n}\}} - nm'.$$

Using $m' = -\mathbb{E}\underline{X}_{11}\mathbf{1}_{\{|\underline{X}_{ij}| > \varepsilon\sigma\sqrt{n}\}}$ and again the Hoffman-Wielandt inequality, one has

$$\begin{aligned} \frac{1}{n}\sum_{k=1}^{n} \left| s_k(\widetilde{M}-zI) - s_k(M''-zI) \right|^2 &\le \frac{1}{n^2\sigma^2} \sum_{1\le i\le n} |D''_{ii} - \underline{D}_{ii}|^2 \\ &\le \frac{1}{n^2\sigma^2} \sum_{1\le i\le n} \left| \sum_{1\le j\le n} \left(\underline{X}_{ij} \mathbf{1}_{\{|\underline{X}_{ij}| > \varepsilon\sigma\sqrt{n}\}} - \mathbb{E}\underline{X}_{ij} \mathbf{1}_{\{|\underline{X}_{ij}| > \varepsilon\sigma\sqrt{n}\}} \right) \right|^2. \end{aligned}$$

The expectation of the above expression equals

$$\frac{1}{n^2\sigma^2} \sum_{1\le i,j\le n} \mathbb{E} \left| \underline{X}_{ij} \mathbf{1}_{\{|\underline{X}_{ij}| > \varepsilon\sigma\sqrt{n}\}} - \mathbb{E}\underline{X}_{ij} \mathbf{1}_{\{|\underline{X}_{ij}| > \varepsilon\sigma\sqrt{n}\}} \right|^2,$$

which tends to 0 by (2.2).

In summary, for the remainder of the proof, we will assume without loss of generality that the law of X_{ij} satisfies

$$\mathbb{E}X_{11} = 0\,, \quad \mathbb{P}(|X_{11}| \ge \kappa(n)) = 0 \quad \text{and} \quad \mathbb{E}|X_{11}|^2 = \sigma'(n)^2, \tag{2.3}$$

where

$$\kappa(n) = o(\sigma(n)\sqrt{n}) \quad \text{and} \quad \sigma'(n) = \sigma(n)(1+o(1)).$$

2.3. **Tightness.** Let us check that $\mathbb{E}\nu_{M-zI}$ is tight. Recall an instance of the Weyl inequality: for all A, B in $\mathcal{M}_n(\mathbb{C})$, for all i:

$$|s_i(A) - s_i(B)| \leq s_1(A - B).$$

Consequently,

$$\int s^2 \, d\nu_{M-zI}(s) \leq \int s^2 \, d\nu_M(s) + |z|^2.$$

It is thus sufficient to check that $\mathbb{E}\int s^2 \, d\nu_M(s)$ is uniformly bounded. However,

$$\begin{aligned}\int s^2 \, d\nu_M(s) &= \frac{1}{n} \sum_{1 \leq i,j \leq n} |M_{ij}|^2 \\ &= \frac{1}{n^2\sigma^2} \sum_{1 \leq i \neq j \leq n} |X_{ij}|^2 + \frac{1}{n^2\sigma^2} \sum_{1 \leq i \leq n} \left| \sum_{1 \leq j \leq n, j \neq i} X_{ij} \right|^2.\end{aligned}$$

The conclusion follows by taking expectation and using (2.3).

2.4. **Linearization.** We use a common linearization technique. With the notation from (1.7) and (1.8) one has the identity of the Cauchy-Stieltjes transform, for $\eta \in \mathbb{C}_+$,

$$S_{\check{\nu}_{M-zI}}(\eta) = \frac{1}{2n} \mathrm{Tr}\left[(H(z) - \eta I_{2n})^{-1}\right], \tag{2.4}$$

where I_{2n} is the $2n \times 2n$ identity matrix and $H(z)$ is the $2n \times 2n$ hermitian matrix

$$H(z) := \begin{pmatrix} 0 & M - z \\ (M-z)^* & 0 \end{pmatrix},$$

with eigenvalues $\{\pm s_i(M - z),\, i = 1, \ldots, n\}$. Define $\mathbb{H}_+ \subset \mathcal{M}_2(\mathbb{C})$ as

$$\mathbb{H}_+ := \left\{ \begin{pmatrix} \eta & z \\ \bar{z} & \eta \end{pmatrix}, z \in \mathbb{C}, \eta \in \mathbb{C}_+ \right\}. \tag{2.5}$$

For $q \in \mathbb{H}_+$, with

$$q(z, \eta) := \begin{pmatrix} \eta & z \\ \bar{z} & \eta \end{pmatrix},$$

let $q(z,\eta) \otimes I_n$ denote the $2n \times 2n$ matrix obtained by repeating n times along the diagonal the 2×2 block q. Through a permutation of the entries, the matrix $H(z) - \eta I_{2n}$ is equivalent to the matrix

$$B - q(z, \eta) \otimes I_n, \tag{2.6}$$

where B is obtained from the 2×2 blocks B_{ij}, $1 \leq i, j \leq n$:

$$B_{ij} := \begin{pmatrix} 0 & M_{ij} \\ \bar{M}_{ji} & 0 \end{pmatrix}.$$

If $B(z) := B - q(z, 0) \otimes I_n$, then $B(z) \in \mathcal{M}_n(\mathcal{M}_2(\mathbb{C})) \simeq \mathcal{M}_{2n}(\mathbb{C})$ is Hermitian, and its resolvent is denoted by

$$R(q) = (B(z) - \eta I_{2n})^{-1} = (B - q(z, \eta) \otimes I_n)^{-1}. \tag{2.7}$$

Then $R(q) \in \mathcal{M}_n(\mathcal{M}_2(\mathbb{C}))$ and, by (2.4), we deduce that

$$S_{\check{\nu}_{M-zI}}(\eta) = \frac{1}{2n} \mathrm{Tr} R(q).$$

We set

$$R(q)_{kk} = \begin{pmatrix} a_k(q) & b_k(q) \\ c_k(q) & d_k(q) \end{pmatrix} \in \mathcal{M}_2(\mathbb{C}).$$

As in [12, lemma 4.19], it is not hard to check that

$$a(q) := \frac{1}{n} \sum_{k=1}^n a_k(q) = \frac{1}{n} \sum_{k=1}^n d_k(q) \quad \text{and} \quad b(q) := \frac{1}{n} \sum_{k=1}^n b_k(q) = \frac{1}{n} \sum_{k=1}^n \bar{c}_k(q), \tag{2.8}$$

It follows that

$$S_{\check{\nu}_{M-zI}}(\eta) = a(q). \tag{2.9}$$

Hence, in order to prove that $\mathbb{E}\nu_{M-z}$ converges, it is sufficient to prove that $\mathbb{E}a(q)$ converges to, say, $\alpha(q)$, for all $q \in \mathbb{H}_+$. By tightness, $\alpha(q)$ will necessarily be the Cauchy-Stieltjes transform

of a symmetric measure. (Indeed, since $\check{\nu}_{M-zI}$ is tight and symmetric, any accumulation point of $\check{\nu}_{M-zI}$ will be a symmetric probability measure. Also, recall that the weak convergence of a sequence of probability measures on $\mathbb{R}$, $(\nu_n)_{n\geq 1}$ to ν, is equivalent to the convergence for all $\eta \in \mathbb{C}_+$ of $S_{\nu_n}(\eta)$ to $S_\nu(\eta)$).

2.5. **Approximate fixed point equation.** We use a resolvent method to deduce an approximate fixed point equation satisfied by $a(q)$. The Schur block inversion formula states that if A is a $k \times k$ matrix then for every partition $\{1,\dots,k\} = I \cup I^c$,

$$(A^{-1})_{I,I} = (A_{I,I} - A_{I,I^c}(A_{I^c,I^c})^{-1}A_{I^c,I})^{-1}.$$

Applied to $k = 2n$, $A = B(z) - \eta I_{2n}$, $q = q(z,\eta)$, it gives

$$R(q)_{nn} = \left(\begin{pmatrix} 0 & M_{nn} \\ \bar{M}_{nn} & 0 \end{pmatrix} - q - Q^* \widetilde{R}(q) Q \right)^{-1}, \tag{2.10}$$

where $Q \in \mathcal{M}_{n-1,1}(\mathcal{M}_2(\mathbb{C}))$, is the $2(n-1) \times 2$ matrix given by the blocks

$$Q_i = \begin{pmatrix} 0 & M_{ni} \\ \bar{M}_{in} & 0 \end{pmatrix} = \frac{1}{\sigma\sqrt{n}} \begin{pmatrix} 0 & X_{ni} \\ \bar{X}_{in} & 0 \end{pmatrix}$$

for $i = 1,\dots,n-1$, and $\widetilde{B} = (B_{ij})_{1\leq i,j\leq n-1}$, $\widetilde{B}(z) = \widetilde{B} - q(z,0) \otimes I_{n-1}$,

$$\widetilde{R}(q) = (\widetilde{B} - q \otimes I_{n-1})^{-1} = (\widetilde{B}(z) - \eta I_{2(n-1)})^{-1}$$

is the resolvent of a minor.

Define the matrix $M' \in \mathcal{M}_{n-1}(\mathbb{C})$ by

$$M'_{ij} = \frac{X_{ij}}{\sigma\sqrt{n}} - \delta_{i,j} \sum_{1\leq k\leq n-1} \frac{X_{ik}}{\sigma\sqrt{n}},$$

for $i,j = 1,\dots,n-1$. Let R' and B' in $\mathcal{M}_{n-1}(\mathcal{M}_2(\mathbb{C}))$ be the matrices obtained as in (2.6) and (2.7) with M replaced by M'. From the resolvent formula and the bounds $\|R'\|, \|\widetilde{R}\| \leq (\mathfrak{Im}(\eta))^{-1}$:

$$\left\|\widetilde{R} - R'\right\| = \left\|R'(\widetilde{B} - B')\widetilde{R}\right\| \leq \frac{1}{\sigma\sqrt{n}\mathfrak{Im}(\eta)^2} \left\| \mathrm{diag}\left(\begin{pmatrix} 0 & X_{ni} \\ \bar{X}_{in} & 0 \end{pmatrix} \right)_{1\leq i\leq n-1} \right\|.$$

Hence using (2.3), we deduce that

$$\|\widetilde{R} - R'\| = o(1). \tag{2.11}$$

(Here and below, $o(1)$ depends on $q(z,\eta)$ through $\mathfrak{Im}(\eta)$). Since $\|Q^*SQ\| \leq \|S\|\|Q^*Q\|$ and $\mathbb{E}\|Q^*Q\| = \mathcal{O}(1)$, with $S = \widetilde{R} - R'$ we obtain

$$R_{nn} = -\left(- \begin{pmatrix} 0 & M_{nn} \\ \bar{M}_{nn} & 0 \end{pmatrix} + q + Q^* R' Q + \varepsilon_1 \right)^{-1}.$$

with ε_1 a 2×2 matrix satisfying $\mathbb{E}\|\varepsilon_1\| = o(1)$.

We denote by $\mathcal{F}_{n-1}$ the σ-algebra spanned by the variables $(X_{ij})_{1\leq i,j\leq n-1}$. Then R' is $\mathcal{F}_{n-1}$-measurable and is independent of Q. If $\mathbb{E}_n[\,\cdot\,] := \mathbb{E}[\,\cdot\,|\mathcal{F}_{n-1}]$, we get, using (2.3) and (2.8)

$$\begin{aligned} \mathbb{E}_n[Q^* R' Q] &= \sum_{1\leq k,\ell\leq n-1} \mathbb{E}_n[Q_k^* R'_{k\ell} Q_\ell] = \frac{\sigma'^2}{\sigma^2 n} \sum_{k=1}^{n-1} \begin{pmatrix} a'_k & 0 \\ 0 & d'_k \end{pmatrix} \\ &= \frac{\sigma'^2}{\sigma^2 n} \sum_{k=1}^{n-1} \begin{pmatrix} a'_k & 0 \\ 0 & a'_k \end{pmatrix} = \frac{1}{n} \sum_{k=1}^{n-1} \begin{pmatrix} \widetilde{a}_k & 0 \\ 0 & \widetilde{a}_k \end{pmatrix} + \varepsilon_2, \end{aligned}$$

where

$$R'_{kk} = \begin{pmatrix} a'_k & b'_k \\ c'_k & d'_k \end{pmatrix} \quad \text{and} \quad \widetilde{R}_{kk} = \begin{pmatrix} \widetilde{a}_k & \widetilde{b}_k \\ \widetilde{c}_k & \widetilde{d}_k \end{pmatrix},$$

and ε_2 is a 2×2 matrix. Using (2.11), we have the bound

$$\left| \frac{1}{n} \sum_{k=1}^{n-1} \widetilde{a}_k - \frac{1}{n} \sum_{k=1}^{n-1} a'_k \right| = \left| \frac{1}{2n} \mathrm{Tr}(\widetilde{R}) - \frac{1}{2n} \mathrm{Tr}(R') \right| \leq \left\|\widetilde{R} - R'\right\| = o(1)$$

We deduce that $\|\varepsilon_2\| = o(1)$. Similarly, recall that $\widetilde{B}(z)$ is a minor of $B(z)$. We may thus use the interlacing inequality (2.1) for the function $f = (\cdot - \eta)^{-1}$. We find

$$\left|\sum_{k=1}^{n-1} \widetilde{a}_k - \sum_{k=1}^{n} a_k\right| \leq 2\int_{\mathbb{R}} \frac{1}{|x-\eta|^2} dx = O\Big(\frac{1}{\mathfrak{Im}(\eta)}\Big).$$

In summary, we have checked that

$$\mathbb{E}_n\Big[Q^*\widetilde{R}Q\Big] = \begin{pmatrix} a & 0 \\ 0 & a \end{pmatrix} + \varepsilon_3,$$

where $a = a(q)$ is as in (2.9), and ε_3 satisfies $\|\varepsilon_3\| = o(1)$. Moreover, we define

$$\varepsilon_4 := \mathbb{E}_n\Big[\Big(Q^*\widetilde{R}Q - \mathbb{E}_n\Big[Q^*\widetilde{R}Q\Big]\Big)^*\Big(Q^*\widetilde{R}Q - \mathbb{E}_n\Big[Q^*\widetilde{R}Q\Big]\Big)\Big].$$

Since $\|\widetilde{R}\| \leq \mathfrak{Im}(\eta)^{-1}$, we have

$$\|\widetilde{R}_{ii}^*\widetilde{R}_{ii}\| \leq \mathfrak{Im}(\eta)^{-2} \quad \text{and} \quad \mathrm{Tr}\Big(\sum_{i,j}\widetilde{R}_{ij}^*\widetilde{R}_{ji}\Big) = \mathrm{Tr}(\widetilde{R}^*\widetilde{R}) \leq 2n\mathfrak{Im}(\eta)^{-2}.$$

Also, by (2.3)

$$\mathbb{E}|X_{ij}^2 - {\sigma'}^2|^2 \leq 2\kappa^2{\sigma'}^2.$$

Then, an elementary computation gives

$$\|\varepsilon_4\| \leq \mathrm{Tr}(\varepsilon_4) = \mathcal{O}\left(\frac{\kappa^2{\sigma'}^2}{n\mathfrak{Im}(\eta)^2\sigma^4}\right) = o(1).$$

Moreover, $a(q)$ is close to its expectation. More precisely, from [12, lemma 4.21],

$$\mathbb{E}|a(q) - \mathbb{E}a(q)|^2 = \mathcal{O}\Big(\frac{1}{n\mathfrak{Im}(\eta)^2}\Big) = o(1).$$

We recall finally that the central limit theorem with Lindeberg condition implies that

$$M_{nn} = -\frac{1}{\sigma\sqrt{n}}\sum_{i=1}^{n-1} X_{ni}$$

converges weakly to G with distribution $\mathcal{N}(0,K)$. From Skorokhod's representation theorem, we may assume that this convergence holds almost surely. Then, we have proved that the 2×2 matrix

$$A = -\begin{pmatrix} 0 & M_{nn} \\ \bar{M}_{nn} & 0 \end{pmatrix} + \begin{pmatrix} 0 & G \\ \bar{G} & 0 \end{pmatrix} + Q^*\widetilde{R}Q - \mathbb{E}\begin{pmatrix} a & 0 \\ 0 & a \end{pmatrix}$$

has a norm which converges to 0 in probability as $n \to \infty$. On the other hand, from (2.10),

$$R_{nn}A = -R_{nn}\left(q + \mathbb{E}\begin{pmatrix} a & 0 \\ 0 & a \end{pmatrix} - \begin{pmatrix} 0 & G \\ \bar{G} & 0 \end{pmatrix}\right) - I_2 \tag{2.12}$$

Since the norms of $\left(q + \mathbb{E}\begin{pmatrix} a & 0 \\ 0 & a \end{pmatrix} - \begin{pmatrix} 0 & G \\ \bar{G} & 0 \end{pmatrix}\right)^{-1}$ and R_{nn} are at most $\mathfrak{Im}(\eta)^{-1}$, we get

$$\mathbb{E}R_{nn} = -\mathbb{E}\left(q + \mathbb{E}\begin{pmatrix} a & 0 \\ 0 & a \end{pmatrix} - \begin{pmatrix} 0 & G \\ \bar{G} & 0 \end{pmatrix}\right)^{-1} + \varepsilon$$

with $\|\varepsilon\| = o(1)$. Using exchangeability, we get that the functions in (2.8) satisfy

$$\mathbb{E}\begin{pmatrix} a & b \\ \bar{b} & a \end{pmatrix} = -\mathbb{E}\left(q + \mathbb{E}\begin{pmatrix} a & 0 \\ 0 & a \end{pmatrix} - \begin{pmatrix} 0 & G \\ \bar{G} & 0 \end{pmatrix}\right)^{-1} + \varepsilon.$$

2.6. **Uniqueness of the fixed point equation.** From what precedes, any accumulation point of $\mathbb{E}\begin{pmatrix} a & b \\ \bar{b} & a \end{pmatrix}$ is solution of the fixed point equation

$$\begin{pmatrix} \alpha & \beta \\ \bar{\beta} & \alpha \end{pmatrix} = -\mathbb{E}\left(q + \begin{pmatrix} \alpha & 0 \\ 0 & \alpha \end{pmatrix} - \begin{pmatrix} 0 & G \\ \bar{G} & 0 \end{pmatrix} \right)^{-1}. \tag{2.13}$$

with $\alpha = \alpha(q) \in \mathbb{C}_+$. Therefore, for $q = q(z,\eta) \in \mathbb{H}_+$:

$$\alpha = \mathbb{E}\frac{\alpha + \eta}{|G - z|^2 - (\alpha + \eta)^2}.$$

The above identity is precisely the fixed point equation satisfied by $S_{\check{\nu}_z}(\eta)$ given in theorem 1.1. Hence, to conclude the proof of theorem 1.1, it is sufficient to prove that there is a unique symmetric measure whose Cauchy-Stieltjes transform is solution of this fixed point equation. We know from (2.9) and Montel's theorem that $\eta \in \mathbb{C}_+ \mapsto \alpha(q(z,\eta)) \in \mathbb{C}_+$ is analytic for every fixed $z \in \mathbb{C}$. In particular, it is sufficient to check that there is a unique solution in $\mathbb{C}_+$ for $\eta = it$, for a fixed $t > 0$. If $h(z,t) = \mathfrak{Im}(\alpha(q))$, we find

$$h = \mathbb{E}\,\frac{h + t}{|G - z|^2 + (h + t)^2}.$$

Thus, $h \neq 0$ and

$$1 = \mathbb{E}\,\frac{1 + th^{-1}}{|G - z|^2 + (h + t)^2}.$$

The right hand side in a decreasing function in h on $(0,\infty)$ with limits equal to $+\infty$ and 0 at $h \to 0$ and $h \to \infty$. Thus, there is a unique solution $h > 0$ of the above equation. The proof of theorem 1.1 is over.

3. Convergence of eigenvalues: Proof of theorem 1.2

3.1. **Strategy of proof.** In order to prove theorem 1.2, we will use the Hermitization method; see e.g. [12, lemma 4.3] for the proof of the next lemma.

Lemma 3.1 (Hermitization)**.** *Let $(A_n)_{n\geq 1}$ be a sequence of complex random matrices where A_n is $n \times n$ for every $n \geq 1$. Suppose that there exists a family $(\nu_z)_{z\in\mathbb{C}}$ of (non-random) probability measures on $\mathbb{R}_+$ such that for a.a. $z \in \mathbb{C}$,*

(i) $\nu_{A_n - z}$ tends weakly in probability to ν_z;
(ii) $\log(\cdot)$ is uniformly integrable in probability for $(\nu_{A_n - z})_{n\geq 1}$.

Then, in probability, μ_{A_n} converges weakly to the probability measure μ defined by

$$\mu = \frac{1}{2\pi}\Delta \int_0^\infty \log(t)\, d\nu_z(t).$$

Applied to our matrix $M(n) = \sigma^{-1}\sqrt{n}(L/n + mI)$, the validity of (i) follows from theorem 1.1 (convergence a.s. implies convergence in probability). The proof of (ii) is performed in the remainder of this section using ideas developed by Tao and Vu in [38] and by the authors in [9], together with an analysis of the smallest singular value which follows closely the work of Götze and Tikhomirov [22]. We are going to prove the following theorem.

Theorem 3.2 (Uniform integrability)**.** *Under the assumptions of theorem 1.2, there exists an increasing function $J : [0,\infty) \to [0,\infty)$ with $J(t)/t \to \infty$ as $t \to \infty$ such that for all $z \in \mathbb{C}\backslash\Lambda$,*

$$\lim_{t\to\infty} \limsup_{n\to\infty} \mathbb{P}\left(\int_0^\infty J(|\log(s)|)\, d\nu_{M(n)-z}(s) > t \right) = 0.$$

From the de La Vallée Poussin criterion for uniform integrability, theorem 3.2 implies point (ii) above, i.e. the uniform integrability in probability of $\log(\cdot)$ for $(\nu_{M(n)-z})_{n\geq 1}$, see [12]. Therefore, it implies theorem 1.2. The proof of theorem 3.2 is divided into three steps corresponding to the control of the large singular values, of the moderately small singular values and of the smallest singular value of $M(n) - z$.

3.2. **Large singular values.**

Lemma 3.3. *For all $z \in \mathbb{C}$, there exists $C > 0$ such that, for $n \geq 1$, $t > 0$,*

$$\mathbb{P}\Big(\int x^2 \, d\nu_{M(n)-z}(x) > t \Big) \leq C\, t^{-1}.$$

In particular, for any $u > 0$,

$$\mathbb{P}(\|M(n) - z\| \geq u) \leq C\, u^{-2} n. \tag{3.1}$$

Proof. As in Section 2.3, we find

$$\begin{aligned} \int x^2 \, d\nu_{M-z}(x) &\leq \int x^2 \, d\nu_M(x) + |z|^2 \\ &= \frac{1}{\sigma^2 n^2} \sum_{i \neq j} |X_{ij}|^2 + \frac{1}{\sigma^2 n^2} \sum_i \Big| \sum_{j \neq i} (X_{ij} - m) \Big|^2 + |z|^2. \end{aligned}$$

Since $\mathbb{E}|X_{ij}|^2 = \mathcal{O}(\sigma^2)$ by (1.11), taking expectation, we find

$$\mathbb{E} \int x^2 \, d\nu_{M-z}(x) = \mathcal{O}(1).$$

It remains to apply Markov's inequality. The bound (3.1) now follows from $\|M(n) - z\|^2 \leq n \int x^2 \, d\nu_{M(n)-z}(x)$. □

3.3. **Smallest singular value.** A crucial step towards the proof of theorem 3.2 is a lower bound, in probability, on the smallest singular value $s_n(M - z)$. Here, our main result is a quasi polynomial lower bound.

Proposition 3.4 (Smallest singular value)**.** *Under the assumptions of theorem 1.2, for all $z \in \mathbb{C} \backslash \Lambda$,*

$$\lim_{n \to \infty} \mathbb{P}(s_n(M - z) \geq e^{-(\log(n))^2}) = 1. \tag{3.2}$$

The proof of Proposition 3.4 follows closely the strategy developed in [22], in turn inspired by the works [35, 31]. However, some crucial modifications are needed due to the extra terms coming form the diagonal matrix D. If one assumes that $\sigma^2(n) \geq n^{-1/4+\varepsilon}$ for some $\varepsilon > 0$, then the quasi polynomial $e^{-(\log(n))^2}$ can be replaced by n^{-4} in (3.2), see Section 3.3.4 below. Before we start the proof of Proposition 3.4, we collect some important preliminaries.

3.3.1. *Distance of a random vector to a subspace.* The following lemma is valid under the sole assumption (1.11). For the proof we adapt the argument of [38, Proposition 5.1], see also [12, Appendix A], but some extra care is needed to handle the sparse regime $\sigma^2 \to 0$. Remark that condition (1.11) implies that there exists $a > 0$ such that

$$\inf_{T \geq a} \inf_{n \geq 1} \sigma^{-2} \mathrm{Var}(\mathbf{x}^{(T)}) > 0, \tag{3.3}$$

where $\mathbf{x}^{(T)}$ is the random variable $\mathbf{x}$ conditioned on $|\mathbf{x}| \leq T$.

Lemma 3.5. *Let $R := (X_{i1}, \ldots, X_{in})$ denote a row of the matrix X and assume* (1.11)*. Let $\psi : \mathbb{N} \to \mathbb{N}$ be such that $\psi(n) \to \infty$, and $\psi(n) < n$. There exists $\varepsilon > 0$ such that for any subspace H of $\mathbb{C}^n$ with $1 \leq \dim(H) \leq n - \psi(n)$, one has*

$$\mathbb{P}\Big(\mathrm{dist}(R, H) \leq \varepsilon \sigma \sqrt{n - \dim(H)} \Big) \leq e^{-\varepsilon \sigma^2 \psi(n)} + e^{-\varepsilon \psi(n)^2 / n}. \tag{3.4}$$

Proof. As in [38, Proposition 5.1] we can assume that the random variables X_{ij} are centered, since this amounts to replace H with H', the linear span of H and a deterministic one-dimensional space, satisfying $\dim(H') \leq n - \psi(n) + 1$. Next, we truncate X_{ij}: Fix $T > 0$ and use Chebyshev's inequality to bound $\mathbb{P}(|X_{1j}| > T) \leq T^{-2}\sigma^2$. Let E_n denote the event that $\sum_{j=1}^n \mathbf{1}(|X_{1j}| > T)) \geq c\psi(n)$, for some $c > 0$ to be chosen later. We take T such that $T^2 \geq 4n\sigma^2/(c\,\psi(n))$. Then, from Hoeffding's inequality one has,

$$\mathbb{P}(E_n) \leq e^{-c^2 \psi(n)/n}. \tag{3.5}$$

Thus, in proving (3.4) we can now assume that the complementary event E_n^c occurs. Set $T = \max\{2\sigma\sqrt{n/(c\psi(n))}, a\}$, where $a > 0$ is the parameter in (3.3). Conditioning on the set I such that $|X_{1j}| \leq T$ iff $j \in I$, $|I| \geq n - c\,\psi(n)$, and conditioning on the values of X_{1j} for $j \in I^c$,

one can reduce the problem to estimating $\mathrm{dist}(R, H')$ by $\mathrm{dist}(\tilde{R}, H'')$ where $\tilde{R}$ is the vector $\tilde{X}_{1j}$, $j = 1, \dots, |I|$, made of i.i.d. copies of the centered variable $\mathbf{x}^{(T)} - \mathbb{E}\mathbf{x}^{(T)}$, and H'' has dimension at most $\dim(H') + |I^c| + 1 \le n - \psi(n) + 2 + c\,\psi(n) \le n - (1-2c)\psi(n)$. A simple computation yields

$$\mathbb{E}\Big[\mathrm{dist}(\tilde{R}, H'')^2\Big] = \tilde{\sigma}^2(n - \dim(H''))\,, \qquad \tilde{\sigma}^2 = \mathrm{Var}(\mathbf{x}^{(T)}).$$

Using (3.3), one has $\tilde{\sigma}^2 \ge c_1\sigma^2$ for some constant $c_1 > 0$. As in [38] we may now invoke Talagrand's concentration inequality for Lipschitz convex functions of bounded independent variables. Using $n - \dim(H'') \ge (1-2c)\psi(n)$, if c is sufficiently small this implies that for some $\varepsilon > 0$ one has

$$\mathbb{P}\Big(\mathrm{dist}(\tilde{R}, H'') \le \varepsilon\sigma\sqrt{n - \dim(H)}\Big) \le \exp(-\varepsilon\sigma^2 T^{-2}\psi(n)). \tag{3.6}$$

The expression above is then an upper bound for the probability of the event

$$E_n^c \cap \big\{\mathrm{dist}(R, H) \le \varepsilon\sigma\sqrt{n - \dim(H)}\big\}.$$

Therefore if T is bounded one has the upper bound $e^{-\varepsilon'\sigma^2\psi(n)}$, for some $\varepsilon' > 0$, while if $T \to \infty$, then $\sigma^2 T^{-2} = c\psi(n)/(4n)$ and one has the upper bound $e^{-\varepsilon'\psi(n)^2/n}$. This ends the proof of (3.4). □

3.3.2. *Compressible and incompressible vectors.* For $\delta \in (0,1)$, define the set of *sparse vectors*

$$\mathrm{Sparse}(\delta) := \{x \in \mathbb{C}^n : \ |\mathrm{supp}(x)| \le \delta n\}$$

where $\mathrm{supp}(x) = \{i : x_i \ne 0\}$ and $|\mathrm{supp}(x)|$ is its cardinality. Given $\rho \in (0,1)$, consider the partition of the unit sphere $\mathbb{S}^{n-1}$ into a set of *compressible vectors* and the complementary set of *incompressible vectors* as follows:

$$\mathrm{Comp}(\delta, \rho) := \{x \in \mathbb{S}^{n-1} : \mathrm{dist}(x, \mathrm{Sparse}(\delta)) \le \rho\}$$
$$\mathrm{Incomp}(\delta, \rho) := \mathbb{S}^{n-1} \setminus \mathrm{Comp}(\delta, \rho).$$

For any matrix $A \in \mathcal{M}_n(\mathbb{C})$:

$$s_n(A) = \min_{x \in \mathbb{S}^{n-1}} \|Ax\|_2 = \min\left(\min_{x \in \mathrm{Comp}(\delta,\rho)} \|Ax\|_2, \min_{x \in \mathrm{Incomp}(\delta,\rho)} \|Ax\|_2\right). \tag{3.7}$$

We will apply (3.7) to $A = Y^\top$, the transpose of the matrix

$$Y = \sigma\sqrt{n}(M(n) - z), \tag{3.8}$$

and then use the obvious identities $s_i(Y^\top) = s_i(Y) = \sigma\sqrt{n}\, s_i(M - z)$, $i = 1, \dots, n$.

Next, we recall two lemmas from [35]; see also [12, Appendix A].

Lemma 3.6. *Let $x \in \mathrm{Incomp}(\delta, \rho)$. There exists a subset $\pi \subset \{1, \dots, n\}$ such that $|\pi| \ge \delta n/2$, and for all $i \in \pi$,*

$$\frac{\rho}{\sqrt{n}} \le |x_i| \le \sqrt{\frac{2}{\delta n}}.$$

Lemma 3.7. *Let $A \in \mathcal{M}_n(\mathbb{C})$ be any random matrix and let W_k denote its k-th column. For $1 \le k \le n$, let $H_k = \mathrm{span}\{W_j,\ j \ne k\}$. Then, for any $t \ge 0$,*

$$\mathbb{P}\left(\min_{x \in \mathrm{Incomp}(\delta,\rho)} \|Ax\|_2 \le \frac{t\rho}{\sqrt{n}}\right) \le \frac{2}{\delta n}\sum_{k=1}^{n} \mathbb{P}(\mathrm{dist}(W_k, H_k) \le t).$$

3.3.3. *Small ball probabilities and related estimates.* We turn to some crucial estimates. We assume that the hypothesis of theorem 1.2 hold.

Lemma 3.8. *Let $R := (X_{i1}, \dots, X_{in})$ denote a row of the matrix X. There exists $C > 0$, such that for any $t \ge 0$, any $\delta, \rho \in (0,1)$, and $x \in \mathrm{Incomp}(\delta, \rho)$, and any $w \in \mathbb{C}$,*

$$\mathbb{P}(|\langle R, x\rangle - w)| \le t) \le \frac{C}{\sigma\sqrt{\delta n}}\left(\frac{t\sqrt{n}}{\rho} + 1\right). \tag{3.9}$$

Proof. Note that the left hand side is bounded above by $p_x(t)$, where p_x is the concentration function defined in (A.1). By lemma 3.6, we can assume that for all $1 \le i \le \delta n/2$, $|x_i| \ge \rho/\sqrt{n}$. It suffices to invoke theorem A.1 with n replaced by $\lfloor \delta n/2 \rfloor$, with x replaced by $y = x\sqrt{n}/\rho$ and use the identity: $p_{\gamma y}(\gamma t) = p_y(t)$ for all $\gamma > 0$. □

Lemma 3.9. *Let $s(n) = n^\kappa$ for some $\kappa > 0$. There exists $\varepsilon > 0$ such that, if*

$$\delta = \frac{\sigma^2(n)}{\log(n)}, \quad \rho = \frac{\varepsilon\sigma(n)}{s(n)\sqrt{\delta}}, \tag{3.10}$$

then for all n large enough:

$$\mathbb{P}\bigg(\min_{x\in \mathrm{Comp}(\delta,\rho)} \big\|Y^\top x\big\|_2 \le \frac{\varepsilon\sigma(n)}{\sqrt{\delta}} \;;\; s_1(Y) \le s(n)\bigg) \le \exp(-\varepsilon n\sigma^2(n)). \tag{3.11}$$

Proof. If $A \in \mathcal{M}_n(\mathbb{C})$ and $y \in \mathbb{C}^n$ is such that $\mathrm{supp}(y) \subset \pi \subset \{1,\dots,n\}$, then

$$\|Ay\|_2 \ge \|y\|_2 s_n(A_{|\pi}),$$

where $A_{|\pi}$ is the $n \times |\pi|$ matrix formed by the columns of A selected by π. Therefore,

$$\min_{x\in \mathrm{Comp}(\delta,\rho)} \|Ax\|_2 \ge (1-\rho) \min_{\pi:\, |\pi| = \lfloor \delta n\rfloor} s_n(A_{|\pi}) - \rho s_1(A). \tag{3.12}$$

On the other hand, for any $x \in \mathbb{C}^{|\pi|}$,

$$\begin{aligned}\big\|A_{|\pi}x\big\|_2^2 &= \Big\|\sum_{i\in\pi} x_i W_i\Big\|_2^2 \\ &\ge \max_{i\in\pi} |x_i|^2 \mathrm{dist}^2(W_i, H_i) \\ &\ge \min_{i\in\pi} \mathrm{dist}^2(W_i,H_i)\, \frac{1}{|\pi|}\sum_{j\in\pi}|x_j|^2,\end{aligned}$$

where W_i is the i-th column of A and $H_i := \mathrm{span}\{W_j : j \in \pi, j \ne i\}$. In particular,

$$s_n(A_{|\pi}) \ge \frac{1}{\sqrt{|\pi|}} \min_{i\in\pi} \mathrm{dist}(W_i, H_i). \tag{3.13}$$

Next, we apply (3.13) to $A = Y^\top$. We have $W_i = R_i^\top + (-D_{ii} + mn - z\sqrt{n})e_i$ where R_i is the i-th row of X. Therefore,

$$\mathrm{dist}(W_i, H_i) \ge \mathrm{dist}(R_i, \mathrm{span}\{H_i, e_i\}) = \mathrm{dist}(R_i, H_i')$$

where $H_i' = \mathrm{span}\{H_i, e_i\} = \mathrm{span}\{R_j, e_i : j \in \pi, j \ne i\}$. H_i' has dimension at most $n\delta + 1$ and is independent of R_i. By lemma 3.5, with e.g. $\psi(n) = n - 2\delta n \ge n/2$, one has that, for some $\varepsilon > 0$

$$\mathbb{P}\bigg(\min_{i\in\pi} \mathrm{dist}(W_i, H_i) \le \varepsilon\sigma\sqrt{n}\bigg) \le n\exp(-\varepsilon\sigma^2 n)$$

for all n large enough. From (3.13), for $|\pi| \le \delta n$:

$$\mathbb{P}\bigg(s_n((Y^\top)_{|\pi}) \le \frac{\varepsilon\sigma}{\sqrt{\delta}}\bigg) \le n\exp(-\varepsilon\sigma^2 n).$$

Therefore, using the union bound and $1 - \rho \ge 3/4$, we deduce from (3.12)

$$\mathbb{P}\bigg(\min_{x\in\mathrm{Comp}} \big\|Y^\top x\big\|_2 \le \frac{\varepsilon\sigma}{2\sqrt{\delta}} \;;\; s_1(Y) \le s(n)\bigg) \le \binom{n}{\lfloor \delta n\rfloor} n e^{-\varepsilon\sigma^2 n} = e^{n(h(\delta)(1+o(1)) - \varepsilon\sigma^2)},$$

with $h(\delta) := -\delta\log\delta - (1-\delta)\log(1-\delta)$. As $n \to \infty$, $h(\delta) = -\delta\log\delta(1+o(1))$ and using $n\sigma^2 \gg \log(n)$ one has $-\delta\log\delta \le \varepsilon\sigma^2/2$ for all n large enough. Therefore, (3.11) follows by adjusting the value of ε. □

Lemma 3.10. *Let W_k, H_k be as in lemma 3.7, with $A = Y^\top$. Let $\delta, \rho, s(n)$ be as in lemma 3.9. There exists $C > 0$ such that, for all $1 \le k \le n$ and $t \ge 0$,*

$$\mathbb{P}(\mathrm{dist}(W_k, H_k) \le t\,;\, s_1(Y) \le s(n)) \le \frac{C}{\sigma\sqrt{\delta n}}\bigg(\frac{t\, s(n)}{\rho\,\sigma|\alpha|} + 1\bigg), \tag{3.14}$$

where $\alpha = m\sqrt{n}/\sigma - z$.

Note that the z dependence, represented by the coefficient α in the above estimate, cannot be completely avoided since if $z = m\sqrt{n}/\sigma$ then $s_n(Y) = 0$.

The proof of lemma 3.10 requires a couple of intermediate steps. Fix k and, conditional on H_k, consider a unit vector ζ orthogonal to H_k. Since W_k is independent of H_k, the random vector ζ can be assumed to be independent of W_k. Clearly,

$$\operatorname{dist}(W_k, H_k) \geq |\langle \zeta, W_k \rangle|. \tag{3.15}$$

Define $\phi = (1, \dots, 1)$ and $\Phi = \operatorname{span}\{\phi\} = \{\lambda\phi,\, \lambda \in \mathbb{C}\}$.

Lemma 3.11. *Let $n \geq 2$. The unit vector ζ orthogonal to H_k satisfies*

$$\operatorname{dist}(\zeta, \Phi) \geq \frac{|\alpha|\sigma\sqrt{n}}{2s_1(Y)}, \tag{3.16}$$

where $\alpha = m\sqrt{n}/\sigma - z$. In particular, if $s_1(Y) \leq s$, then one has, for any $\lambda \in \mathbb{C}$,

$$\|\zeta - \lambda\phi\| \geq \frac{|\alpha|\sigma\sqrt{n}}{2s}.$$

Proof. Set $\hat{\phi} = \frac{1}{\sqrt{n}}\phi$. Since $\|\zeta\| = \|\hat{\phi}\| = 1$, we have

$$\operatorname{dist}(\zeta, \Phi) = \operatorname{dist}(\hat{\phi}, \operatorname{span}\{\zeta\}) = \operatorname{dist}(\hat{\phi}, \operatorname{span}\{\bar{\zeta}\}).$$

(the last identity follows from $\hat{\phi} \in \mathbb{R}^n$). Let $B \in \mathcal{M}_n(\mathbb{C})$ be the matrix obtained from Y by replacing the k-th row with the zero vector. Then, by construction $B\bar{\zeta} = 0$. Hence, $\bar{\zeta} \in \ker B$ and

$$\operatorname{dist}(\zeta, \Phi) \geq \operatorname{dist}(\hat{\phi}, \ker B).$$

Observe that $\hat{\phi} = au + bv$, where $a, b \in \mathbb{C}$, and some unit vectors $v \in \ker B$ and $u \in (\ker B)^{\perp}$. Then $|a| = \operatorname{dist}(\hat{\phi}, \ker B)$ can be bounded as follows. Note that Y satisfies $Y\phi = \alpha\sigma\sqrt{n}\phi$ and therefore $aBu = B\hat{\phi} = \alpha\sigma\sqrt{n}\hat{\phi} - \alpha\sigma e_k$. Consequently, one has

$$s_1(B)|a| \geq \|aBu\| = \|B\hat{\phi}\| = |\alpha|\sigma\sqrt{n-1} \geq |\alpha|\sigma\sqrt{n}/2$$

This implies (3.16) since $s_1(B) \leq s_1(Y)$. □

Lemma 3.12. *Let $\delta, \rho, s(n)$ be as in lemma 3.9. There exists $\varepsilon > 0$ such that, for all n large enough:*

$$\mathbb{P}(\exists \lambda \in \mathbb{C} : \eta(\lambda) \in \operatorname{Comp}(\delta, \rho)\,; s_1(Y) \leq s(n)) \leq \exp(-\varepsilon\sigma^2 n),$$

where for $\lambda \in \mathbb{C}$, $\eta(\lambda) := (\zeta - \lambda\phi)/\|\zeta - \lambda\phi\|$.

Proof. Let B be as in the proof of lemma 3.11. Thus $B\bar{\zeta} = 0$. Note that by lemma 3.11, $\eta(\lambda)$ is well defined as soon as $\alpha \neq 0$, and then $\|\eta(\lambda)\| = 1$. If there exists $\lambda \in \mathbb{C}$ such that $\eta(\lambda) \in \operatorname{Comp}(\delta, \rho)$, then $B(\bar{\eta}(\lambda) + \lambda'\phi) = 0$ for $\lambda' = \bar{\lambda}/\|\zeta - \lambda\phi\| \in \mathbb{C}$. Therefore

$$\min_{x \in \operatorname{Comp}(\delta,\rho),\, \lambda \in \mathbb{C}} \|B(x + \lambda\phi)\| = 0.$$

Note that if $\Phi_k := \operatorname{span}\{\phi, e_k\}$, then $Y\phi \in \Phi, B\phi \in \Phi_k$ and $Bx - Yx \in \Phi_k$ for all vectors x. Thus

$$\min_{x \in \operatorname{Comp}(\delta,\rho),\, v \in \Phi_k} \|Yx + v\|_2 = 0.$$

The above equation can be rewritten as

$$\min_{x \in \operatorname{Comp}(\delta,\rho)} \|\Pi Y x\|_2 = 0, \tag{3.17}$$

where Π is the orthogonal projection on the orthogonal complement of Φ_k. On the other hand, one has

$$\|\Pi Y x\|_2 \geq \|x\|_2 s_n([\Pi Y]_{|\pi}),$$

if π is the support of x. As in (3.13) one has

$$s_n([\Pi Y]_{|\pi}) \geq \frac{1}{\sqrt{|\pi|}} \min_{i \in \pi} \operatorname{dist}(\tilde{W}_i, \tilde{H}_i),$$

where $\tilde{W}_i$ is the i-th column of ΠY and $\tilde{H}_i := \operatorname{span}\{\tilde{W}_j : j \in \pi, j \neq i\}$. Note that

$$\operatorname{dist}(\tilde{W}_i, \tilde{H}_i) \geq \operatorname{dist}(C_i, H_i''),$$

where $H_i'' = \mathrm{span}\{C_j, e_\ell, \Phi_k : j, \ell \in \pi, j \neq i\}$ and C_i is the i-th column of X. Then, since H_i'' has dimension at most $1 + 2n\delta$ and is independent of C_i, by lemma 3.5, with e.g. $\psi(n) = n - 3\delta n$,

$$\mathbb{P}\bigg(s_n([\Pi Y]_{|\pi}) \leq \frac{\varepsilon\sigma}{\sqrt{\delta}} \bigg) \leq n \exp(-\varepsilon\sigma^2 n),$$

for some $\varepsilon > 0$. In particular, as in the proof of (3.11), one sees that the probability of the event in (3.17) intersected with the event $s_1(Y) \leq s(n)$ is bounded by $\exp(-\varepsilon\sigma^2 n)$, for some new $\varepsilon > 0$ and all n large enough. This proves the lemma. □

Let us now go back to (3.15). Observe that

$$\langle \zeta, W_k \rangle = \sum_{i=1}^n (\bar{\zeta}_i - \bar{\zeta}_k) X_{ki} + \langle \zeta, v \rangle = \langle \zeta - \zeta_k \phi, R_k \rangle + \langle \zeta, v \rangle,$$

where v is a deterministic vector, and we use the notation R_k for the k-th row of X. With the notation of lemma 3.11 and lemma 3.12, on the event $s_1(Y) \leq s$ we can write

$$\mathrm{dist}(W_k, H_k) \geq \frac{|\alpha|\sigma\sqrt{n}}{2s} \, |\langle \eta, R_k \rangle + w_k|, \tag{3.18}$$

where $\eta = \eta(\lambda)$, at $\lambda = \zeta_k$, and w_k depends only on ζ and therefore is independent of R_k. By lemma 3.12 and using $\exp(-\varepsilon\sigma^2 n) \ll (\sigma\sqrt{\delta n})^{-1}$, in order to prove (3.14), it is sufficient to invoke lemma 3.8 above. This ends the proof of lemma 3.10.

3.3.4. *Proof of Proposition 3.4.* Take δ, ρ and $s(n)$ as in lemma 3.9. From (3.14) and lemma 3.7 we find, for all $t \geq 0$,

$$\mathbb{P}\bigg(\min_{x \in \mathrm{Incomp}(\delta,\rho)} \big\| Y^\top x \big\|_2 \leq \frac{\rho^2 t}{\sqrt{n}} \ ; \ s_1(Y) \leq s \bigg) \leq \frac{C}{\sigma\sqrt{\delta^3 n}} \bigg(\frac{ts}{|\alpha|\sigma} + 1 \bigg).$$

Using our choice of $\rho, \delta, s(n)$, we obtain for some new constant $C > 0$, for all $t \geq 0$,

$$\mathbb{P}\bigg(\min_{x \in \mathrm{Incomp}(\delta,\rho)} \big\| Y^\top x \big\|_2 \leq t \ ; \ s_1(Y) \leq n^\kappa \bigg) \leq \frac{C(\log(n))^{3/2}}{\sigma^4\sqrt{n}} \bigg(\frac{t n^{3\kappa}\sqrt{n}}{|\alpha|\sigma\log(n)} + 1 \bigg). \tag{3.19}$$

From (3.11) we know that

$$\mathbb{P}\bigg(\min_{x \in \mathrm{Comp}(\delta,\rho)} \big\| Y^\top x \big\|_2 \leq \varepsilon\sqrt{\log(n)} \ ; \ s_1(Y) \leq n^\kappa \bigg) \leq \exp(-\varepsilon n \sigma^2(n)). \tag{3.20}$$

Since $s_1(Y) = \sigma\sqrt{n}\|M(n) - z\|$, by (3.1) one has

$$\mathbb{P}\big(\, s_1(Y) > s \big) \leq C n^2 s^{-2}. \tag{3.21}$$

Suppose that

$$\sigma^2 \geq n^{-1/4+\varepsilon}, \tag{3.22}$$

for some $\varepsilon > 0$. Choosing e.g. $\kappa = 1 + \gamma$, $s = n^{1+\gamma}$, with $\gamma = 0.1$, and $t = n^{-4}$, then the above expressions and (3.7) imply that

$$\lim_{n\to\infty} \mathbb{P}\big(\, s_n(Y) > n^{-4} \big) = 0. \tag{3.23}$$

In particular, this proves Proposition 3.4 under the mild sparsity assumption (3.22).

In the general case we cannot count on (3.22), and we only assume $n\sigma^2 \gg (\log(n))^6$. In particular, while the bound (3.20) is still meaningful, the bound (3.19) becomes useless, even at $t = 0$, if e.g. $\sigma^2 \leq n^{-1/4}$. To deal with this problem we use a further partition of the set $\mathrm{Incomp}(\delta, \rho)$ inspired by the method of Göetze and Tikhomirov [22]. More precisely, for fixed $\varepsilon > 0, \kappa > 1$, define

$$\begin{aligned} \delta_1 = \frac{\sigma^2(n)}{\log(n)}, \quad \rho_1 = \frac{\varepsilon\sigma(n)}{n^\kappa\sqrt{\delta_1}}, \\ \delta_\ell = \delta_{\ell-1}\log(n), \quad \rho_\ell = \frac{\varepsilon\sqrt{\delta_{\ell-1}}}{n^\kappa}\,\rho_{\ell-1}, \quad \ell = 2, \dots, N, \end{aligned} \tag{3.24}$$

where $N = N(n)$ is defined as the smallest integer $k \geq 2$ such that $(\log n)^{k+1}\sigma^2(n) \geq 1$. Note that δ_1, ρ_1 are the choice of values of δ, ρ from lemma 3.8. Define further

$$\delta_{N+1} = \frac{1}{(\log(n))^2}\,, \quad \text{and} \quad \rho_{N+1} = \frac{\varepsilon\sqrt{\delta_N}\rho_N}{n^\kappa}\,. \tag{3.25}$$

It is immediate to check that this defines an increasing sequence $\delta_\ell = (\log(n))^{\ell-2}\sigma^2(n)$ and a decreasing sequence ρ_ℓ, $\ell = 1, \ldots, N+1$, such that $\sigma^2(n)(\log(n))^{-1} \leq \delta_\ell \leq (\log n)^{-2}$, $\rho_{N+1} \leq \rho_\ell \leq \varepsilon(\log(n))^{1/2}n^{-\kappa}$, with $N(n) = \mathcal{O}(\log(n)/(\log\log(n)))$, and

$$\rho_{N+1} \geq \exp\left(-\eta(\log(n))^2\right), \tag{3.26}$$

for any $\eta > 0$, for n large enough.

As in our previous argument, cf. (3.23), the value of κ is not essential as long as $\kappa > 1$, and it may be fixed for the rest of this proof as e.g. $\kappa = 2$. Next, using the sequences (δ_ℓ, ρ_ℓ), define

$$\begin{aligned} C_0 &= \mathrm{Comp}(\delta_1, \rho_1)\,, \quad C_{N+1} = \mathrm{Incomp}(\delta_{N+1}, \rho_{N+1})\,, \\ C_\ell &= \mathrm{Comp}(\delta_{\ell+1}, \rho_{\ell+1}) \cap \mathrm{Incomp}(\delta_\ell, \rho_\ell)\,, \quad 1 \leq \ell \leq N. \end{aligned} \tag{3.27}$$

Note that these sets form a partition $\mathbb{S}^{n-1} = \cup_{\ell=0}^{N+1} C_\ell$ and therefore, for any $t \geq 0$:

$$\mathbb{P}(s_n(Y) \leq t) \leq \sum_{\ell=0}^{N+1} \mathbb{P}\Big(\min_{x\in C_\ell} \big\|Y^\top x\big\|_2 \leq t\Big). \tag{3.28}$$

Thus, it will be sufficient to show that

$$\lim_{n\to\infty} \sum_{\ell=0}^{N+1} \mathbb{P}\Big(\min_{x\in C_\ell} \big\|Y^\top x\big\|_2 \leq e^{-(\log(n))^2}\Big) = 0. \tag{3.29}$$

For the term C_0, since δ_1, ρ_1 coincide with the choice of lemma 3.9, one can use the estimate (3.20) and (3.21). For $\ell \geq 1$ we need a refinement of the argument used for (3.19) and (3.20).

Lemma 3.13. *There exists a constant $\varepsilon > 0$ such that if δ_ℓ, ρ_ℓ are defined by (3.24), then for all $\ell = 1, \ldots, N$:*

$$\mathbb{P}\Big(\min_{x\in C_\ell} \|Y^\top x\|_2 \leq \varepsilon\rho_\ell\sqrt{\delta_\ell}\,;\; s_1(Y) \leq n^\kappa;\; s_1(X) \leq n^\kappa\Big) \leq \exp\big(-\varepsilon n(1 \wedge \sigma^2\delta_\ell n)\big). \tag{3.30}$$

Proof. Let $x \in \mathrm{Incomp}(\delta_\ell, \rho_\ell)$ and assume that $\mathrm{supp}(x) \subset \pi$ with $|\pi| = \delta_{\ell+1}n$. Let $W_i = R_i^\top + (-D_{ii} + mn - z\sqrt{n})e_i$ be the i-th column of $Y^\top$ and $H = \mathrm{span}(e_i, i \in \pi)$, and write

$$\begin{aligned} \|Y^\top x\|_2 = \|\sum_{i\in\pi} W_i x_i\|_2 &\geq \mathrm{dist}\Big(\sum_{i\in\pi} W_i x_i, H\Big) \\ &= \mathrm{dist}\Big(\sum_{i\in\pi} R_i x_i, H\Big) = \mathrm{dist}\big(X^\top x, H\big). \end{aligned} \tag{3.31}$$

Next, we want to apply corollary A.4. Taking $s = n^\kappa$, from (3.26) one has that $\log(\frac{s}{\rho_\ell^2\delta_\ell}) \ll (\log(n))^2$. Therefore, using $n\sigma^2(n) \gg (\log(n))^3$, $\delta_{\ell+1} = (\log(n))^{\ell-1}\sigma^2(n)$, one has

$$\delta_{\ell+1} \leq \frac{\eta(1 \wedge \sigma^2\delta_\ell n)}{\log(\frac{s}{\rho_\ell^2\delta_\ell})}, \tag{3.32}$$

for any $\eta > 0$, and all n large enough. Hence, if c_1 is as in corollary A.4, then

$$\mathbb{P}\Big(\big\|Y^\top x\big\|_2 \leq c_1\rho_\ell\sqrt{\delta_\ell}\,;\; s_1(X) \leq n^\kappa\Big) \leq \exp\big(-c_1 n(1 \wedge \sigma^2\delta_\ell n)\big).$$

If $\eta \in (0,1)$ and $B_H = H \cap \mathrm{Incomp}(\delta_\ell, \rho_\ell)$, there exists an η-net N of B_H, of cardinality $(3/\eta)^k$, $k = \delta_{\ell+1}n$, such that

$$\min_{x:\mathrm{dist}(x,B_H)\leq\eta} \|Y^\top x\|_2 \geq \min_{x\in N} \|Y^\top x\|_2 - \eta s_1(Y).$$

Hence, if we take $\eta = \rho_{\ell+1} = \frac{\varepsilon\rho_\ell\sqrt{\delta_\ell}}{n^\kappa}$, from the union bound, and then using (3.32):

$$\mathbb{P}\Big(\min_{x:\mathrm{dist}(x,B_H)\le\rho_{\ell+1}} \big\|Y^\top x\big\|_2 \le (c_1-\varepsilon)\rho_\ell\sqrt{\delta_\ell}\,;\, s_1(Y)\le n^\kappa;\, s_1(X)\le n^\kappa\Big)$$
$$\le \exp\left(\delta_{\ell+1}n\log\left(\frac{6n^\kappa}{\rho_\ell\sqrt{\delta_\ell}}\right) - c_1 n(1\wedge\sigma^2\delta_\ell n)\right) \le \exp\Big(-\frac{c_1}{2}n(1\wedge\sigma^2\delta_\ell n)\Big).$$

Finally, summing over all choices of $|\pi| = \delta_{\ell+1}n$, one finds

$$\mathbb{P}\Big(\min_{x\in C_\ell} \|Y^\top x\|_2 \le (c_1-\varepsilon)\rho_\ell\sqrt{\delta_\ell}\,;\, s_1(Y)\le n^\kappa;\, s_1(X)\le n^\kappa\Big)$$
$$\le \exp\Big(nh(\delta_{\ell+1})(1+o(1)) - \frac{c_1}{2}n(1\wedge\sigma^2\delta_\ell n)\Big),$$

with $h(\delta) = -\delta\log(\delta) - (1-\delta)\log(1-\delta)$. Since $\delta_{\ell+1}\log\delta_{\ell+1} \ll (1\wedge\sigma^2\delta_\ell n)$, the above expression is bounded by $e^{-\frac{c_1}{3}n(1\wedge\sigma^2\delta_\ell n)}$ for all n large enough. The conclusion follows by choosing e.g. $\varepsilon = \frac{c_1}{3}$. □

Let us now conclude the proof of Proposition 3.4. Observe that by (3.21) we may assume that $s_1(Y)\le n^\kappa$. Moreover, the same argument proving (3.21) proves the same bound for $s_1(X)$. Thus, one may assume that $s_1(X)\le n^\kappa$ as well at the price of adding a vanishing term to (3.29). Using (3.20) (for the case $\ell = 0$) and (3.30) (for the case $\ell = 1,\dots,N$), together with the simple bounds $\varepsilon\rho_\ell\sqrt{\delta_\ell} \gg e^{-(\log(n))^2}$, $n(1\wedge\sigma^2\delta_\ell n)\gg(\log n)^2$, one has

$$\lim_{n\to\infty}\sum_{\ell=0}^{N}\mathbb{P}\Big(\min_{x\in C_\ell}\|Y^\top x\|_2 \le e^{-(\log(n))^2}\Big) = 0. \tag{3.33}$$

Thus, to end the proof of (3.29), it remains to prove

$$\mathbb{P}\Big(\min_{x\in C_{N+1}}\|Y^\top x\|_2 \le e^{-(\log(n))^2}\Big)\to 0. \tag{3.34}$$

To prove (3.34), observe that lemma 3.7 and lemma 3.11, as in (3.18), imply that for all $t\ge 0$:

$$\mathbb{P}\Big(\min_{x\in C_{N+1}}\|Y^\top x\|_2 \le \frac{t\rho_{N+1}}{\sqrt{n}}\Big) \le \frac{2}{\delta_{N+1}n}\sum_{k=1}^{n}\mathbb{P}\Big(|\langle\eta^{(k)},R_k\rangle + w_k| \le \frac{2ts_1(Y)}{|\alpha|\sigma\sqrt{n}}\Big),$$

where $w_k\in\mathbb{C}$ and $\eta^{(k)}\in\mathbb{S}^{n-1}$ denote suitable random variables independent of R_k, the k-th row of X. Thanks to (3.21), one can safely assume that $s_1(Y)\le n^\kappa$. By exchangeability it is enough to consider the first row R of X, and the associated random variables η, w. Using $\delta_{N+1} = (\log(n))^{-2}$, taking $t = e^{-(\log(n))^2}\sqrt{n}/\rho_{N+1}$, and using $2n^\kappa e^{-(\log(n))^2}/(|\alpha|\sigma\rho_{N+1}) \le e^{-\frac12(\log(n))^2}/|\alpha|$, for n large, it is then sufficient to prove

$$\lim_{n\to\infty}(\log(n))^2\,\mathbb{P}\Big(|\langle\eta,R\rangle + w| \le |\alpha|^{-1}e^{-\frac12(\log(n))^2}\Big) = 0. \tag{3.35}$$

By conditioning on the event $\eta\in\mathrm{Incomp}(\delta_{N+1},\rho_{N+1})$, lemma 3.8 implies that

$$\mathbb{P}\Big(|\langle\eta,R\rangle + w| \le |\alpha|^{-1}e^{-\frac12(\log(n))^2}; \eta\in\mathrm{Incomp}(\delta_{N+1},\rho_{N+1})\Big)$$
$$\le \frac{C}{\sigma\sqrt{\delta_{N+1}n}}\left(\frac{e^{-\frac12(\log(n))^2}\sqrt{n}}{|\alpha|\rho_{N+1}} + 1\right) \le \frac{2C\log(n)}{\sigma\sqrt{n}}, \tag{3.36}$$

where the last bound holds for all $z\in\mathbb{C}\setminus\Lambda$, for n sufficiently large, so that α is bounded away from 0. Since by assumption (1.10) we have $\sigma\sqrt{n}\gg(\log(n))^3$, this proves (3.35), provided that

$$\lim_{n\to\infty}(\log(n))^2\,\mathbb{P}(\eta\in\mathrm{Comp}(\delta_{N+1},\rho_{N+1})) = 0. \tag{3.37}$$

As in the proof of lemma 3.12, cf. (3.17), $\eta\in\mathrm{Comp}(\delta_{N+1},\rho_{N+1})$ implies

$$\min_{x\in\mathrm{Comp}(\delta_{N+1},\rho_{N+1})}\|\Pi Y x\|_2 = 0,$$

where Π is the orthogonal projection on $(\mathrm{span}\{\phi, e_1\})^\perp$. Since $\mathrm{Comp}(\delta_{N+1}, \rho_{N+1}) = \cup_{\ell=0}^N C_\ell$, (3.37) may be reduced to the estimate

$$\lim_{n\to\infty} (\log(n))^2 \sum_{\ell=0}^{N} \mathbb{P}\Big(\min_{x\in C_\ell} \|\Pi Y x\|_2 = 0 \Big) = 0. \tag{3.38}$$

To prove (3.38), one repeats the argument in the proof of lemma 3.13. More precisely, (3.31) is now replaced by

$$\|\Pi Y x\|_2 \geq \mathrm{dist}(Xx, H')$$

where $H' = \mathrm{span}\{H, \phi, e_1\}$. Since H' has dimension at most $\delta_{\ell+1} n + 2$, the same arguments apply here. As in the proof of (3.33), this implies (3.38). This concludes the proof of Proposition 3.4.

3.4. **Moderately small singular values.**

Lemma 3.14 (Moderately small singular values)**.** *Assume* (1.10) *and* (1.11)*. Let* $u(n) = n/[(\log n)^5]$*. There exists* c_0 *such that for any* $z \in \mathbb{C}$*, a.s. for* $n \gg 1$

$$s_{n-i}(M(n) - z) \geq c_0 \frac{i}{n}, \qquad u(n) \leq i \leq n-1,$$

Proof. We follow the original proof of Tao and Vu [38] for the circular law. To lighten the notations, we denote by $s_1 \geq \cdots \geq s_n$ the singular values of $M - zI$. We fix $u(n) \leq i \leq n-1$, and consider the matrix Y' formed by the first $m := n - \lceil i/2 \rceil$ rows of $\sigma\sqrt{n}(M - z)$. Let $s'_1 \geq \cdots \geq s'_m$ be the singular values of Y'. By the Cauchy-Poincaré interlacing, we get

$$\sigma^{-1} n^{-1/2} s'_{n-i} \leq s_{n-i}$$

(see e.g. [29, corollary 3.1.3]). Next, by [38, lemma A4], we have

$$s_1'^{-2} + \cdots + s_{n-\lceil i/2 \rceil}'^{-2} = \mathrm{dist}_1^{-2} + \cdots + \mathrm{dist}_{n-\lceil i/2 \rceil}^{-2},$$

where $\mathrm{dist}_j := \mathrm{dist}(R'_j, H'_j)$ is the distance from the j^{th} row R'_j of the matrix Y' to H'_j, the subspace spanned by all other rows of Y'. In particular, we have

$$\frac{i}{2n} s_{n-i}^{-2} \leq \frac{i\sigma^2}{2} s_{n-i}'^{-2} \leq \sigma^2 \sum_{j=n-\lceil i \rceil}^{n-\lceil i/2 \rceil} s_j'^{-2} \leq \sigma^2 \sum_{j=1}^{n-\lceil i/2 \rceil} \mathrm{dist}_j^{-2}. \tag{3.39}$$

Now, we note that

$$\mathrm{dist}_j = \mathrm{dist}(R'_j, H'_j) \geq \mathrm{dist}(R_j, H_j),$$

where $H_j = \mathrm{span}\{H'_j, e_j\}$ and R_j is the j^{th} row of X. Now, H_j is independent of R_j and $\dim(H_j) \leq n - \frac{i}{2} + 1 \leq n - \frac{1}{4} u(n)$. We may use lemma 3.5 with the choice $\psi(n) = \frac{1}{4} u(n)$. By assumption (1.10) one has

$$\min\{\sigma^2(n)\psi(n), \psi(n)^2/n\} \gg \log(n). \tag{3.40}$$

By the union bound, this implies

$$\sum_{n\geq 1} \mathbb{P}\left(\bigcup_{i=u(n)}^{n-1} \bigcup_{j=1}^{n-\lceil i/2 \rceil} \left\{ \mathrm{dist}_j \leq \frac{\sigma\sqrt{i}}{2\sqrt{2}} \right\} \right) < \infty. \tag{3.41}$$

Consequently, by the first Borel-Cantelli lemma, we obtain that a.s. for $n \gg 1$, all $u(n) \leq i \leq n-1$, and all $1 \leq j \leq n - \lceil i/2 \rceil$,

$$\mathrm{dist}_j \geq \frac{\sigma\sqrt{i}}{2\sqrt{2}} \geq \frac{\sigma\sqrt{i}}{4}$$

Finally, (3.39) gives $s_{n-i}^2 \geq (i^2)/(32n^2)$, i.e. the desired result with $c_0 := 1/(4\sqrt{2})$. □

3.5. **Proof of theorem 3.2.** Let us choose $J(t) = t^2$. By lemma 3.3, it is sufficient to prove that

$$\lim_{t\to\infty} \limsup_{n\to\infty} \mathbb{P}\left(\int_0^1 J(|\log s|)\, d\nu_{M(n)-z(s)} > t \right) = 0.$$

We shall actually prove that if n_* is the last i such that $s_{n-i}(M-z) \leq 1$, then there exists $C > 0$ such that

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{1}{n} \sum_{i=0}^{n_*} J(|\log s_{n-i}(M-z)|) \leq C \right) = 1. \tag{3.42}$$

With the notation, of lemma 3.14, let F_n be the event, that $s_n(M-z) \geq e^{-(\log(n))^2}$ and that for all $u(n) \leq i \leq n-1$, $s_{n-i}(M-z) \geq c_0\, i/n$. Then by Proposition 3.4 and lemma 3.14, F_n has probability tending to 1. Also, if F_n holds, writing s_{n-i} for $s_{n-i}(M-z)$ one has

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^{n_*} J(|\log s_{n-i}|) &\leq \frac{1}{n} \sum_{i=1}^{u(n)} J((\log(n))^2) + \frac{1}{n} \sum_{i=1}^{n} J(|\log(c_0\, i/n)|) \\ &= \frac{u(n)(\log(n))^4}{n} + \frac{1}{n} \sum_{i=1}^{n} (\log(n/i))^2. \end{aligned}$$

This last expression is uniformly bounded since $u(n) = n/[(\log(n))^5]$ and the sum is approximated by a finite integral. This concludes the proof of (3.42).

4. Limiting distribution: Proof of theorems 1.3 and 1.4

4.1. **Brown measure.** In this paragraph, we recall classical notions of operator algebra. Consider the pair $(\mathcal{M}, \tau)$, where $\mathcal{M}$ is a von Neumann algebra and τ is a normal, faithful, tracial state on $\mathcal{M}$. For $a \in \mathcal{M}$, set $|a| = \sqrt{a^* a}$. For a self-adjoint element $a \in \mathcal{M}$, we denote by μ_a the spectral measure of a, that is the unique probability measure on the real line satisfying, for any $z \in \mathbb{C}_+$,

$$\tau((a-z)^{-1}) = \int \frac{d\mu_a(t)}{t-z} = S_{\mu_a}(z).$$

The Brown measure [13] of $a \in \mathcal{M}$ is the probability measure μ_a on $\mathbb{C}$, which satisfies for almost all $z \in \mathbb{C}$,

$$\int \log|z-\lambda|\, d\mu_a(\lambda) = \int \log(t)\, d\mu_{|a-z|}(t)$$

In distribution, it is given by the formula

$$\mu_a = \frac{1}{2\pi} \Delta \int \log(t)\, d\mu_{|a-z|}(t). \tag{4.1}$$

Our notation is consistent: firstly, if a is self-adjoint, then the Brown measure coincides with the spectral measure; secondly, if $\mathcal{M} = \mathcal{M}_n(\mathbb{C})$ and $\tau = \frac{1}{n}\mathrm{Tr}$ is the normalized trace on $\mathcal{M}_n(\mathbb{C})$, then the Brown measure of A is simply equal to $\mu_A = \frac{1}{n}\sum_{i=1}^n \delta_{\lambda_i(A)}$.

The $\star$-distribution of $a \in \mathcal{M}$ is the collection of all its $\star$-moments $\tau(a^{\varepsilon_1} a^{\varepsilon_2} \cdots a^{\varepsilon_n})$ where a^{ε_i} is either a or a^*. The element $c \in \mathcal{M}$ is circular if it has the $\star$-distribution of $(s_1 + i s_2)/\sqrt{2}$ where s_1 and s_2 are free semi-circular variables. We refer to Voiculescu, Dykema and Nica [41] for a complete treatment of free non-commutative variables.

As explained in Haagerup and Schultz [26], it is possible to extend these notions to unbounded operators. Let $\bar{\mathcal{M}}$ be the set of closed, densely defined operators a affiliated with $\mathcal{M}$ satisfying

$$\int \log(1+t)\, d\mu_{|a|}(t) < +\infty.$$

In particular the normal operator g in theorem 1.3 is an element of $\bar{\mathcal{M}}$. Also, note that if $a \in \bar{\mathcal{M}}$ and $z \in \mathbb{C}$, then $a - z \in \bar{\mathcal{M}}$. For all $a \in \bar{\mathcal{M}}$, Haagerup and Schultz check that it is possible to define the Brown measure by (4.1).

4.2. **Proof of theorem 1.3.** From theorem 1.2, μ is given by the formula, in distribution,

$$\mu = \frac{1}{2\pi}\Delta \int \log(t)\, d\nu_z(t).$$

Hence in view of (4.1), the statement of theorem 1.3 will follow once we prove that for all $z \in \mathbb{C}$,

$$\nu_z = \mu_{|c+g-z|}.$$

To prove the latter identity, assume that $\mathbf{x}$ has distribution $\mathcal{N}(0, K)$, i.e. $(X_{ij})_{1\leq i,j\leq n}$ are i.i.d. centered Gaussian variable with covariance K, and $\sigma = 1$. Let $(G_i)_{i\geq 1}$ be an independent sequence of i.i.d. Gaussian random variables, $G_i \sim \mathcal{N}(0, K)$. We define the diagonal matrix $D' = \mathrm{diag}(G_1, \ldots, G_n)$, which is independent of X and set $M' = \frac{1}{\sqrt{n}}X - D'$. The proof of theorem 1.1 shows that ν_{M-z}, $\nu_{M'-z}$ both converge a.s. to ν_z. However, it is a consequence of Capitaine and Casalis [15, proposition 5.1] or Anderson, Guionnet and Zeitouni [1, theorem 5.4.5], that $\mathbb{E}\nu_{M'-z}$ converges weakly to $\mu_{|c+g-z|}$ ([1, theorem 5.4.5] is stated for Wigner matrices but the result can be lifted to our case, see [1, exercice 5.4.14]).

4.3. **Quaternionic resolvent.** Here we give another characterization of the Brown measure μ_{c+g}. We use the same linearization procedure as in section 2.4 to develop a quaternionic resolvent approach for the Brown measure. This approach was introduced in the mathematical physics literature [20, 23, 34] for the analysis of non-hermitian random matrices; see also [11] and [12, §4.6]. As above, we consider the operator algebra $(\bar{\mathcal{M}}, \tau)$ associated to the von Neumann algebra $(\mathcal{M}, \tau)$. If elements of $\mathcal{M}$ act on a Hilbert space H, we define the Hilbert space $H_2 = H \times \mathbb{Z}/2\mathbb{Z}$ and for $x = (y, \varepsilon) \in H_2$, we set $\hat{x} = (y, \varepsilon + 1)$. In particular, this transform is an involution $\hat{\hat{x}} = x$. There is the direct sum decomposition $H_2 = H_0 \oplus H_1$ with $H_\varepsilon = \{x = (y, \varepsilon) : y \in H\}$. An operator b acting on H_2 has the 2×2 representation

$$b = \begin{pmatrix} b_{00} & b_{01} \\ b_{10} & b_{11} \end{pmatrix} \tag{4.2}$$

where b_{ij} are operators on H. That is, if $x = (y_0, 0) + (y_1, 1) \in H_2$, then $bx = (z_0, 0) + (z_1, 1)$, where $z_0 = b_{00}y_0 + b_{01}y_1$, and $z_1 = b_{10}y_0 + b_{11}y_1$. We define the linear map τ_2 on operators acting on H_2, with values in $\mathcal{M}_2(\mathbb{C})$, through the formula,

$$\tau_2(b) = \begin{pmatrix} \tau(b_{00}) & \tau(b_{01}) \\ \tau(b_{10}) & \tau(b_{11}) \end{pmatrix}. \tag{4.3}$$

Given $a \in \bar{\mathcal{M}}$ we define the operator

$$\mathrm{bip}(a) = \begin{pmatrix} 0 & a \\ a^* & 0 \end{pmatrix}$$

The operator $\mathrm{bip}(a)$ is self-adjoint. It will be called the *bipartization* of a.

Recall the definition (2.5) of $\mathbb{H}_+$ and $q = q(z, \eta) \in \mathbb{H}_+$ We define the *quaternionic transform* of a as the 2×2 matrix

$$\Gamma_a(q) := \tau_2((\mathrm{bip}(a) - q \otimes I_H)^{-1}).$$

Here $q \otimes I_H$ is the operator on H_2 defined by (4.2) with $b_{00} = b_{11} = \eta I$ and $b_{01} = zI$, $b_{10} = \bar{z}I$, with I the identity operator on H. Note that $(\mathrm{bip}(a) - q \otimes I_H)^{-1}$ is the usual resolvent at η of the self-adjoint operator $b(z) := \mathrm{bip}(a) - q(z, 0) \otimes I_H$. Hence $(\mathrm{bip}(a) - q \otimes I_H)^{-1}$ inherits the usual properties of resolvent operators (analyticity in $\eta \in \mathbb{C}_+$, bounded norm). For a proof of the next lemma, see [12, lemma 4.19]. We use the notation $\partial = \frac{1}{2}(\partial_x - i\partial_y)$, for the derivative at $z = x + iy \in \mathbb{C}$.

Lemma 4.1 (Properties of the quaternionic transform)**.** *For all $q = q(z, \eta) \in \mathbb{H}_+$,*

$$\Gamma_a(q) = \begin{pmatrix} \alpha(q) & \beta(q) \\ \bar{\beta}(q) & \alpha(q) \end{pmatrix} \in \mathbb{H}_+,$$

with

$$\alpha(q) = S_{\check{\mu}_{|a-z|}}(\eta),$$

and, in distribution,

$$\mu_a = -\frac{1}{\pi} \lim_{t\downarrow 0} \partial\beta(q(z, it)).$$

Recall that a sequence of matrices $(A_n)_{n\geq 1}$ is said to converge in $\star$-moments to $a \in \mathcal{M}$ if for any integer k and $(\varepsilon_i)_{1\leq i\leq k} \in \{1, *\}^k$,

$$\lim_{n\to\infty} \frac{1}{n}\mathrm{Tr} A_n^{\varepsilon_1} A_n^{\varepsilon_2} \cdots A_n^{\varepsilon_k} = \tau(a^{\varepsilon_1} a^{\varepsilon_2} \cdots a^{\varepsilon_k}).$$

Also, if $(A_n)_{n\geq 1}$ is a sequence of random matrices, $(A_n)_{n\geq 1}$ converges in expected $\star$-moments to $a \in \mathcal{M}$ if the above convergence holds in expectation.

Lemma 4.2 (Continuity of the quaternionic transform)**.** *If (A_n) is a sequence of matrices converging in $\star$-moments to $a \in \mathcal{M}$ then for all $q \in \mathbb{H}_+$, $\Gamma_{A_n}(q)$ converges to $\Gamma_a(q)$. If (A_n) is a sequence of random matrices converging in expected $\star$-moments to $a \in \mathcal{M}$ then $\mathbb{E}\Gamma_{A_n}(q)$ converges to $\Gamma_a(q)$.*

Proof. For ease of notation, let $a_z = a - z$. Then, in matrix form,

$$(\mathrm{bip}(a) - q \otimes I_H)^{-1} = \begin{pmatrix} -\eta & a_z \\ a_z^* & -\eta \end{pmatrix}^{-1} = -\begin{pmatrix} \eta(\eta^2 - a_z a_z^*)^{-1} & a_z(\eta^2 - a_z^* a_z)^{-1} \\ a_z^*(\eta^2 - a_z a_z^*)^{-1} & \eta(\eta^2 - a_z^* a_z)^{-1} \end{pmatrix}.$$

Hence

$$\Gamma_a(q) = -\begin{pmatrix} \eta\tau(\eta^2 - a_z a_z^*)^{-1} & \tau\big(a_z(\eta^2 - a_z^* a_z)^{-1}\big) \\ \tau\big(a_z^*(\eta^2 - a_z a_z^*)^{-1}\big) & \eta\tau(\eta^2 - a_z^* a_z)^{-1} \end{pmatrix}.$$

For a replaced by A_n and τ by $\frac{1}{n}\mathrm{Tr}$, we may expand in series the terms of the above expression. Each term of the series is a $\star$-moment of $(A_n - z)$ and it converges by assumption to the $\star$-moment in $a - z$. Since $|a|$ is bounded, for $|\eta|$ large enough, the series is absolutely convergent. We thus obtain the convergence of $\Gamma_{A_n}(q)$ for $|\eta|$ large enough. Finally, we may extend by analyticity to all $\eta \in \mathbb{C}_+$. □

If $\Gamma = \begin{pmatrix} \Gamma_{11} & \Gamma_{12} \\ \Gamma_{21} & \Gamma_{22} \end{pmatrix}$ is a 2×2 matrix, we define $\mathrm{diag}(\Gamma) = \begin{pmatrix} \Gamma_{11} & 0 \\ 0 & \Gamma_{22} \end{pmatrix}$.

Proposition 4.3 (Subordination formula)**.** *If c and a are $\star$-free operators in $(\bar{\mathcal{M}}, \tau)$ with c circular then for all $q = q(z, \eta) \in \mathbb{H}_+$,*

$$\Gamma_{c+a}(q) = \Gamma_{|a-z|}(q(0, \eta) + \mathrm{diag}(\Gamma_{c+a}(q))). \tag{4.4}$$

Moreover, if a is normal then

$$\Gamma_{c+a}(q) = \Gamma_a(q + \mathrm{diag}(\Gamma_{c+a}(q))).$$

A version of proposition 4.3, in the language of random matrices, was obtained by Rogers [34, theorem 2]. However, our formulas coincide only for normal operators. This subordination formula is also reminiscent of the subordination formula in Biane [6, proposition 2] on the free sum with a semi-circular. Our argument is indirect and relies on random matrices.

Proof of proposition 4.3. Let u_0, u_1 be unitary operators on H. We consider the unitary operator u in H_2 defined in matrix form by

$$u = \begin{pmatrix} u_0 & 0 \\ 0 & u_1 \end{pmatrix}.$$

From (4.3), for any operator b,

$$\tau_2(u^* b u) = \tau_2(b).$$

Now, for ease of notation, let $a_z = a - z$. There exist unitary operators $u_0 = u_0(z), u_1 = u_1(z)$ such that $u_0^* a_z u_1 = |a_z|$. Then, with u as above,

$$\mathrm{bip}(a_z) = \begin{pmatrix} 0 & a_z \\ a_z^* & 0 \end{pmatrix} = u^* \begin{pmatrix} 0 & |a_z| \\ |a_z| & 0 \end{pmatrix} u.$$

Since $u_i c$, $u_i^* c$ and c have the same $\star$-distribution, $u\mathrm{bip}(c)u^*$ and $\mathrm{bip}(c)$ have also the same $\star$-distribution, we obtain from what precedes that

$$\Gamma_{c+a}(q(z, \eta)) = \tau_2((\mathrm{bip}(c) + \mathrm{bip}(a_z) - q(0, \eta))^{-1}) = \Gamma_{c+|a_z|}(q(0, \eta)).$$

As a consequence, in order to prove proposition 4.3, it is sufficient to prove the statement for a normal operator. Now, if a is normal, let μ_a denote its spectral measure.

We first assume that $a \in \mathcal{M}$. Consider a diagonal matrix A of size n with i.i.d. diagonal entries with distribution μ_a and Y an independent complex Ginibre matrix of size n (i.e. $(Y_{ij})_{1\leq i,j\leq n}$ is

an array of i.i.d. $\mathcal{N}(0, I_2/2)$ random variables). The proof of theorem 1.1, cf. (2.13), shows that $\mathbb{E}\Gamma_{A+Y/\sqrt{n}}$ converges to the function Γ which satisfies

$$\Gamma(q) = \Gamma_a(q + \mathrm{diag}(\Gamma(q))).$$

On the other hand, it is known that $A + Y/\sqrt{n}$ converges in expected $\star$-moments to the free sum $a + c$, see [15, proposition 5.1] or [1, theorem 5.4.5]. It thus remains to invoke lemma 4.2. This completes the proof of lemma 4.3 when $a \in \mathcal{M}$.

In the general case, let $a \in \bar{\mathcal{M}}$ be a normal operator. From the spectral theorem, (see e.g. [17, §X.4]), there is a resolution of the identity E (i.e. a projection valued probability measure) such that

$$a = \int_{\mathbb{C}} \lambda dE(\lambda),$$

and $D(a) = \{\psi \in H : \int |\lambda|^2 d\langle \psi, E(\lambda)\psi\rangle < \infty\}$. For n integer, define

$$a_n = \int_{\{\lambda : |\lambda| \le n\}} \lambda dE(\lambda).$$

By construction, as $n \to \infty$, for any $\psi \in D(a)$,

$$\|a_n\psi - a\psi\|_2^2 = \int_{\{\lambda : |\lambda| > n\}} |\lambda|^2 d\langle \psi, E(\lambda)\psi\rangle \to 0.$$

That is, a_n converges in strong sense toward a. Hence the sequences of operators, $\mathrm{bip}(a_n) - q(z,0) \otimes I_H$ and $\mathrm{bip}(a_n + c) - q(z,0) \otimes I_H$ also converge in the strong sense to $\mathrm{bip}(a) - q(z,0) \otimes I_H$ and $\mathrm{bip}(a + c) - q(z,0) \otimes I_H$, respectively. In particular, for any $q \in \mathbb{H}_+$,

$$\Gamma_{a_n}(q) \to \Gamma_a(q) \quad \text{and} \quad \Gamma_{a_n+c}(q) \to \Gamma_{a+c}(q). \tag{4.5}$$

(see e.g. [33, theorem VIII.25(a)]).

Moreover, by construction a_n is a bounded operator and, from what precedes, Γ_{a_n+c} satisfies the fixed point equation

$$\Gamma(q) = \Gamma_{a_n}(q + \mathrm{diag}(\Gamma(q))). \tag{4.6}$$

We note finally that

$$\|\Gamma_a(q) - \Gamma_a(q')\| \le C\|q - q'\|, \tag{4.7}$$

where for $q = q(z,\eta)$, $q' = q(z',\eta')$, $C = \min(\Im\mathfrak{m}(\eta), \Im\mathfrak{m}(\eta'))^{-2}$. Indeed, by the resolvent identity:

$$(b - q)^{-1} - (b - q')^{-1} = (b - q)^{-1}((q' - q) \otimes I_H)(b - q')^{-1}.$$

Hence

$$\|\Gamma_a(q) - \Gamma_a(q')\| \le \left\|(b - q)^{-1}\right\| \|q - q'\| \left\|(b - q')^{-1}\right\| \le C\|q - q'\|.$$

The conclusion follows from (4.5),(4.6) and (4.7). □

Remark 4.4 (Uniqueness of the solution to the fixed point equation)**.** *Note that* (4.4) *characterizes completely the quaternionic transform of $c + a$. Indeed, in section 2.6, we have proved that there exists a unique map $\Gamma : \mathbb{H}_+ \to \mathbb{H}_+$ which satisfies* (4.4) *for all $q \in \mathbb{H}_+$ and such that, with $\alpha(q) = \Gamma(q)_{11}$, for all $z \in \mathbb{C}$, $\eta \mapsto \alpha(q(z,\eta))$ is analytic on $\mathbb{C}_+$ and is the Cauchy-Stieltjes transform of a symmetric measure on $\mathbb{R}$. Indeed, in* (2.13)*, replace $z - G$ by a random variable G_z with law $\mu_{|a-z|}$ to obtain* (4.4)*.*

4.4. **Proof of theorem 1.4.** Set

$$\Gamma_{c+a}(q) = \begin{pmatrix} \alpha(q) & \beta(q) \\ \bar{\beta}(q) & \alpha(q) \end{pmatrix}.$$

By proposition 4.3, Γ_{c+h} satisfies the fixed point equation

$$\begin{aligned} \begin{pmatrix} \alpha & \beta \\ \bar{\beta} & \alpha \end{pmatrix} &= \mathbb{E}\left(\begin{pmatrix} 0 & G \\ \bar{G} & 0 \end{pmatrix} - q - \begin{pmatrix} \alpha & 0 \\ 0 & \alpha \end{pmatrix} \right)^{-1} \\ &= \mathbb{E}\frac{1}{|G - z|^2 - (\alpha + \eta)^2} \begin{pmatrix} \alpha + \eta & G - z \\ \bar{G} - \bar{z} & \alpha + \eta \end{pmatrix}, \end{aligned} \tag{4.8}$$

where G has law $\mathcal{N}(0, K)$ and $q = q(z,\eta)$. For ease of notation, we set

$$G_z = z - G$$

We also define

$$\Sigma = \left\{ z \in \mathbb{C} : \mathbb{E}\frac{1}{|G-z|^2} > 1 \right\},$$

and its closure $\bar{\Sigma} = \left\{ z \in \mathbb{C} : \mathbb{E}\frac{1}{|G-z|^2} \geq 1 \right\}$. As in section 2.6, for $\eta = it$, we find $\alpha = ih(z,t) \in i\mathbb{R}_+$ and

$$1 = \mathbb{E}\frac{1 + th^{-1}}{|G_z|^2 + (h+t)^2}.$$

In particular, if $0 < t \leq 1$, then

$$1 \leq \mathbb{E}\frac{1 + h^{-1}}{|G_z|^2 + h^2}.$$

When h goes to infinity, the right hand side goes to 0 (uniformly in z). Hence there exists $c > 0$, such that for all $z \in \mathbb{C}$ and $0 < t \leq 1$, $h(z,t) \leq c$. Similarly,

$$1 \geq \mathbb{E}\frac{1}{|G_z|^2 + (h+t)^2}.$$

Thus for all $z \in \Sigma$ there exists $c_z > 0$ depending continuously on z such that $h(z,t) + t \geq c_z$.

We now let $t \downarrow 0$. From what precedes, if $z \in \Sigma$, any accumulation point, say $f(z)$, of $h(z,t)$ satisfies $f(z) \in [c_z, c]$ and

$$1 = \mathbb{E}\frac{1}{|G_z|^2 + f(z)^2}. \tag{4.9}$$

The function $\varphi_z : x \mapsto \mathbb{E}(|G_z|^2 + x^2)^{-1}$ is decreasing, for $x \geq 0$. Hence, for all $z \in \Sigma$, there exists a unique value $f(z)$ which satisfies (4.9). Moreover, for any $0 < \varepsilon < 1$ and $z \in \mathbb{C}$, the map φ_z is C^∞ on $[\varepsilon, \varepsilon^{-1}]$, while for any $x > \varepsilon$, the map $\psi_x : z \mapsto \mathbb{E}(|G_z|^2 + x^2)^{-1}$ is C^∞ on $\mathbb{C}$. Then, the implicit function theorem implies that $z \mapsto f(z)$ is C^∞ on Σ.

Now, take $z \notin \Sigma$, we recall that

$$h(t,z) = \mathbb{E}\frac{h(t,z) + t}{|G_z|^2 + (h+t)^2},$$

and for $0 < t \leq 1$, $0 \leq h(z,t) \leq c$. Hence, letting $t \downarrow 0$, any accumulation point $f(z)$ of $h(z,t)$ satisfies $f(z) \in [0,c]$ and

$$f(z) = \mathbb{E}\frac{f(z)}{|G_z|^2 + f(z)^2}.$$

If $f(z) \neq 0$ then (4.9) would hold true. However, this would contradict the assumption $z \notin \Sigma$. Therefore, for all $z \notin \Sigma$, we have

$$f(z) = 0.$$

By (4.8), it follows that

$$\beta(z) := \lim_{t \downarrow 0} \beta(q(z, it)) = -\mathbb{E}\frac{G_z}{|G_z|^2 + f(z)^2}$$

By lemma 4.1, the Brown measure of $c + g$ is equal in distribution to

$$\mu_{c+g} = -\frac{1}{\pi}\partial\beta(z) = \frac{1}{\pi}\partial\mathbb{E}\frac{G_z}{|G_z|^2 + f(z)^2}.$$

Now, if $z \notin \bar{\Sigma}$, then $f(z)$ is 0 in a neighborhood of z. Hence, $-\beta(z) = \mathbb{E}(\bar{G} - \bar{z})^{-1}$. Since $\partial\bar{z} = 0$, $\partial\beta(z) = 0$, and we deduce that the density of μ_{c+g} is 0 on $(\bar{\Sigma})^c$.

Assume now that $z \in \Sigma$. We find that μ_{c+g} has a density given by $1/\pi$ times

$$-\partial\beta(z) = \mathbb{E}\frac{1}{|G_z|^2 + f(z)^2} - \mathbb{E}\frac{|G_z|^2}{(|G_z|^2 + f(z)^2)^2} - 2f(z)f'(z)\mathbb{E}\frac{G_z}{(|G_z|^2 + f(z)^2)^2},$$

where we use $\partial G_z = 1$, $\partial|G_z|^2 = \bar{G}_z$. Here $f'(z) = \partial f(z)$. Using (4.9), the first term on the right hand side is equal to 1 and

$$0 = \mathbb{E}\frac{\bar{G}_z}{(|G_z|^2 + f(z)^2)^2} + 2f(z)f'(z)\mathbb{E}\frac{1}{(|G_z|^2 + f(z)^2)^2}.$$

Hence,

$$-\partial\beta(z) = 1 - \mathbb{E}\frac{|G_z|^2}{(|G_z|^2+f(z)^2)^2} + \frac{\left|\mathbb{E}\frac{G_z}{(|G_z|^2+f(z)^2)^2}\right|^2}{\mathbb{E}\frac{1}{(|G_z|^2+f(z)^2)^2}}$$
$$= \mathbb{E}\frac{f(z)^2}{(|G_z|^2+f(z)^2)^2} + \frac{\left|\mathbb{E}\frac{G_z}{(|G_z|^2+f(z)^2)^2}\right|^2}{\mathbb{E}\frac{1}{(|G_z|^2+f(z)^2)^2}}.$$

From what precedes, if $z \in \Sigma$, $f(z) > 0$. We have thus proved that the density of μ_{c+g} is positive on Σ, the interior of $\bar{\Sigma}$, and given by $1/\pi$ times the above expression, while on $\mathbb{C}\backslash\bar{\Sigma}$ the density is 0. In particular, the support of μ_{c+g} is $\bar{\Sigma}$. This concludes the proof of theorem 1.4.

5. Extremal eigenvalues: Proof of theorems 1.5 and 1.6

5.1. **Proof of theorem 1.5.** The next lemma allows us to control the spectral norm of the diagonal matrix $\underline{D}$ defined in (1.13). The proof uses a refined central limit theorem together with estimates for the maximum of i.i.d. standard Gaussian random variables. We refer to [14, theorem 1.5] for a proof.

Lemma 5.1 (Gaussianization)**.** *Under the assumptions of theorem 1.5, almost surely*

$$\max_{1\le i\le n} |\underline{D}_{ii}| = \sigma\sqrt{2n\log(n)}\,(1+o(1)).$$

Next, we observe that from the Bauer-Fike theorem [5, theorem 25.1] one has that the eigenvalues of $\underline{L} = \underline{X} - \underline{D}$ are all contained in the subset of $\mathbb{C}$ defined by

$$\bigcup_{i=1}^{n} B(-\underline{D}_{ii}, s_1(\underline{X})),$$

where $B(z,t)$ stands for the Euclidean closed ball (actually a disk) around z with radius t, and $s_1(\underline{X})$ is the largest singular value of $\underline{X}$. Since $s_1(\underline{X}) = 2\sigma\sqrt{n}\,(1+o(1))$ (by [3, theorem 2]), using lemma 5.1 one finds that all eigenvalues $\underline{\lambda}$ of $\underline{L}$ must satisfy (1.14).

We turn to the proof of (1.15). We observe that from the Bauer-Fike theorem all eigenvalues of L must be contained in the subset of $\mathbb{C}$ defined by

$$\bigcup_{i=1}^{n} B(\zeta_i, s_1(\underline{X})),$$

where $\zeta_1,\dots,\zeta_n$ are the (real) eigenvalues of $-\underline{D} + mJ - mnI$. The eigenvalues of $mJ - mnI$ are easily seen to be $z_1 = 0$ and $z_2 = \cdots = z_n = -mn$. Thus, the bound $s_1(\underline{X}) = \sigma\sqrt{n}(2+o(1))$ proves statement (1.15) on $\mathfrak{Im}(\lambda)$. For the statement on $\mathfrak{Re}(\lambda)$, we first notice that

$$-mn - \sigma\sqrt{2n\log(n)}\,(1+o(1)) \le -mn - \|\underline{L}\| \le \mathfrak{Re}(\lambda) \le 0.$$

This proves one side of the bound on $|\mathfrak{Re}(\lambda)+mn|$ and, for the other side, we can assume without loss of generality that

$$\overline{\lim_{n\to\infty}}\ \frac{\sigma}{m}\sqrt{\frac{2\log(n)}{n}} < 1. \tag{5.1}$$

Moreover one can arrange the labeling in such a way that $\max_{1\le j\le n}|\zeta_j - z_j| = \mathcal{O}(\max_i |\underline{D}_{ii}|)$, so that one has

$$|\zeta_1| = \mathcal{O}(\max_{1\le i\le n}|\underline{D}_{ii}|) \quad\text{and}\quad \max_{2\le j\le n}|\zeta_j + mn| = \mathcal{O}(\max_{1\le i\le n}|\underline{D}_{ii}|).$$

If n is large enough, by lemma 5.1, the bound $s_1(\underline{X}) = \sigma\sqrt{n}(2+o(1))$ and assumption (5.1), we see that

$$B(\zeta_1, s_1(\underline{X})) \cap B(\zeta_j, s_1(\underline{X})) = \varnothing,$$

for all $2 \le j \le n$. Thus, a continuity argument [28, proof of Gershgorin's theorem 6.1.1] implies that apart from $\lambda = 0$, which belongs to $B(\zeta_1, s_1(\underline{X}))$, all other eigenvalues of L belong to $\cup_{j=2}^{n} B(\zeta_j, s_1(\underline{X}))$. Using the bound $s_1(\underline{X}) = 2\sigma\sqrt{n}\,(1+o(1))$ and lemma 5.1 we see that any $\lambda \neq 0$ in the spectrum of L must satisfy (1.15).

5.2. **Proof of theorem 1.6.** Since $\|\underline{L}\| \leq \|\underline{X}\| + \|\underline{D}\|$, we may bound separately the norms of $\|\underline{X}\|$ and $\|\underline{D}\|$. We have

$$\|\underline{D}\| = \max_{1\leq i\leq n} \left| \sum_{j=1}^{n} \underline{X}_{ij} \right|.$$

By assumption there exists $a > 0$ such that with probability one, $|\underline{X}_{ij}| \leq a$. From Bennett's inequality, for any $1 \leq i \leq n$, $t > 0$:

$$\mathbb{P}\left(\left| \sum_{j=1}^{n} \underline{X}_{ij} \right| \geq t\sigma\sqrt{n} \right) \leq 2\exp\left(-\frac{\sigma^2 n}{a^2} h\Big(\frac{at}{\sigma\sqrt{n}} \Big) \right),$$

where $h(s) = (1+s)\log(1+s) - s \sim s^2/2$ as s goes to 0. We choose $t = \sqrt{2c\log(n)}$, we find by (1.17)

$$\mathbb{P}\left(\left| \sum_{j=1}^{n} \underline{X}_{ij} \right| \geq \sigma\sqrt{2cn\log(n)} \right) \leq n^{-c(1+o(1))}.$$

In particular, if $c = 2(1+\varepsilon)^2$, from the union bound, for $n \gg 1$,

$$\mathbb{P}\left(\max_{1\leq i\leq n} \left| \sum_{j=1}^{n} \underline{X}_{ij} \right| \geq 2(1+\varepsilon)\sigma\sqrt{n\log(n)} \right) \leq n^{-1+\varepsilon}.$$

Hence, from Borel-Cantelli lemma we get a.s. for $n \gg 1$,

$$\|\underline{D}\| \leq (2+o(1))\sigma\sqrt{n\log(n)}.$$

We now turn to the bound on $\|\underline{X}\|$. This is a much more delicate matter. Fortunately, we may use a result by Vu [43, theorem 1.4], which extends Fűredi and Komlós [21]. It asserts that a.s. for $n \gg 1$,

$$\|\underline{X}\| = (2+o(1))\sigma\sqrt{n} + c\sigma^{\frac{1}{2}} n^{\frac{1}{4}} \log(n).$$

This proves (1.18). To prove (1.19), we may follow the proof of theorem 1.5.

6. Invariant measure: Proof of theorem 1.7

We start by proving the matrix L is irreducible. Consider the graph G on $\{1,\dots,n\}$ whose adjacency matrix is $A = (A_{ij})_{1\leq i,j\leq n}$ with $A_{ii} = 0$ and $A_{ij} = \mathbf{1}_{\{X_{ij}\neq 0\}}$ for $i \neq j$. Note that G is an oriented Erdős-Rényi random graph where each edge is present independently with probability $p := \mathbb{P}(X_{ij} \neq 0)$. To prove irreducibility of L, we shall prove that the oriented graph G is connected. Hence, by a fundamental result of Erdős and Rényi [19], see also e.g. [8, theorem 7.3], [30, Corollary 3.31], it is sufficient to check that

$$\varliminf_{n\to\infty} \frac{pn}{\log(n)} > 1. \tag{6.1}$$

Note that the cited references deal with non-oriented Erdős-Rényi graphs. This does not change much, see the discussion [18, p.2]. For either model A or model B, the statement (6.1) follows from assumption (1.10).

We may now turn to the analysis of the invariant measure Π. We will rely on a method that has already been successfully used in random matrix models with finite rank perturbations, for example in [4, 39]. We write

$$L = \underline{L} + m\phi\phi^* - mnI,$$

with $\phi = (1,\dots,1)^\top$. Then, if z is not an eigenvalue of $\underline{L}$, we have

$$\begin{aligned} \det(L - zI + mnI) &= \det(\underline{L} + m\phi\phi^* - zI) \\ &= \det(I + m(\underline{L} - zI)^{-1}\phi\phi^*)\det(\underline{L} - zI) \\ &= \big(1 + m\phi^*(\underline{L} - zI)^{-1}\phi\big)\det(\underline{L} - zI), \end{aligned}$$

where at the last line we have used the well known Sylvester determinant theorem: for all $A \in \mathcal{M}_{n,p}(\mathbb{C})$, $B \in \mathcal{M}_{p,n}(\mathbb{C})$,

$$\det(I_n + AB) = \det(I_p + BA).$$

In particular, $z - mn$ will be an eigenvalue of L if and only if

$$f(z) := 1 + m\phi^*(\underline{L} - zI)^{-1}\phi = 0.$$

Further, still if z is not an eigenvalue of $\underline{L}$,

$$\begin{aligned}\phi^*(\underline{L} - zI)^{-1}(L - zI + mnI) &= \phi^*(\underline{L} - zI)^{-1}(\underline{L} - zI + m\phi\phi^*)\\ &= f(z)\phi^*.\end{aligned}$$

We deduce that if $z - mn$ is an eigenvalue of L then $\phi^*(\underline{L} - zI)^{-1}$ is a left eigenvector of L with eigenvalue $z - mn$. Now 0 is an eigenvalue of L and, by theorems 1.5-1.6, a.s. for $n \gg 1$, mn is not an eigenvalue of $\underline{L}$. Hence from what precedes, for $1 \le i \le n$,

$$\Pi_i = \frac{u_i}{\sum_{k=1}^n u_k},$$

where

$$u^* = \phi^*\left(I - \frac{\underline{L}}{mn}\right)^{-1}.$$

From the resolvent identity,

$$\left(I - \frac{\underline{L}}{mn}\right)^{-1} - I = \frac{\underline{L}}{mn}\left(I - \frac{\underline{L}}{mn}\right)^{-1}.$$

Thus, using theorems 1.5-1.6, and $\|\phi\| = \sqrt{n}$,

$$\|\phi - u\| \le \|\phi\|\left\|\frac{\underline{L}}{mn}\right\|\left\|\left(I - \frac{\underline{L}}{mn}\right)^{-1}\right\| = \mathcal{O}\Big(\frac{\sigma}{m}\sqrt{\log(n)}\Big) + \mathcal{O}\left(\frac{\sqrt{\sigma}}{m}\frac{\log(n)}{n^{1/4}}\right).$$

In other words, writing $u_i = 1 + \varepsilon_i$, we find

$$\sum_{k=1}^n |\varepsilon_k| \le \sqrt{n}\left(\sum_{k=1}^n \varepsilon_i^2\right)^{\frac{1}{2}} = \mathcal{O}\Big(\frac{\sigma}{m}\sqrt{n\log(n)}\Big) + \mathcal{O}\left(\frac{\sqrt{\sigma}}{m}n^{1/4}\log(n)\right).$$

Consequently, uniformly in $i = 1, \dots, n$:

$$\Pi_i = \frac{1 + \varepsilon_i}{n + \sum_{k=1}^n \varepsilon_k} = \frac{1}{n} + \frac{\varepsilon_i}{n} + \mathcal{O}\left(\frac{\sigma}{m}\sqrt{\frac{\log(n)}{n^3}}\right) + \mathcal{O}\left(\frac{\sqrt{\sigma}}{m}\frac{\log(n)}{n^{7/4}}\right).$$

Therefore,

$$\begin{aligned}\sum_{i=1}^n \left\|\Pi_i - \frac{1}{n}\right\| &\le \frac{1}{n}\sum_{i=1}^n \left(|\varepsilon_i| + \mathcal{O}\left(\frac{\sigma}{m}\sqrt{\frac{\log(n)}{n}}\right) + \mathcal{O}\left(\frac{\sqrt{\sigma}}{m}\frac{\log(n)}{n^{7/4}}\right)\right)\\ &= \mathcal{O}\left(\frac{\sigma}{m}\sqrt{\frac{\log(n)}{n}}\right) + \mathcal{O}\left(\frac{\sqrt{\sigma}}{m}\frac{\log(n)}{n^{3/4}}\right).\end{aligned}$$

Appendix A. Concentration function and small ball probabilities

For $x \in \mathbb{C}^n$ and $t \ge 0$, define the concentration function as

$$p_x(t) = \max_{w\in\mathbb{C}} \mathbb{P}\left(\left|\sum_{i=1}^n X_i x_i - w\right| \le t\right), \tag{A.1}$$

where $\{X_i,\, i = 1, \dots, n\}$ are i.i.d. copies of the complex valued random variable $\mathbf{x}$ with law $\mathcal{L}(n)$. Throughout this section we assume that $\mathbf{x}$ satisfies (1.11) and $n\sigma^2 \ge 1$ only. This implies the existence of constants $a, b > 0$ such that for all $n \ge 1$, letting $\mathbf{x}^{(a)}$ denote the random variable $\mathbf{x}$ conditioned on $|\mathbf{x}| \le a$:

$$\mathbb{P}(|\mathbf{x}| \le a) \ge b\,, \quad \widetilde{\sigma}^2 = \mathrm{Var}\Big(\mathbf{x}^{(a)}\Big) \ge b\sigma^2, \tag{A.2}$$

$$\mathbb{P}(a^{-1} \le |\mathbf{x}| \le a) \ge b\sigma^2. \tag{A.3}$$

We start with a consequence of Halász [27].

Theorem A.1 (Halász concentration bound)**.** *Let $1 \le m \le n$. There exists a constant $C > 0$ independent of (m, n) such that if $|x_i| \ge 1$, $1 \le i \le m$, then for all $t \ge 0$,*

$$p_x(t) \le \frac{C}{\sigma\sqrt{m}}(t+1).$$

Proof. For any constant $c > 0$, e^{-cm} is smaller than $\frac{1}{\sigma\sqrt{m}}$ for all $m \gg 1$. Hence, from Hoeffding inequality, it is sufficient to prove the statement conditioned on $\{|X_1| \le a, \dots, |X_k| \le a\}$ with $k = bm/2$, if $b > 0$ is as in (A.2). At the price of replacing m by $bm/2$ and σ by $\widetilde{\sigma}$, one can replace $\mathbf{x}$ by $\mathbf{x}^{(a)}$ from the start. For ease of notation, we will simply assume that the variables are bounded by a.

We first assume that all X_i and x_i are real valued. It is sufficient to prove the statement for $t \ge 1$. Following Rudelson and Vershynin in [35], set $Y = |X_1 - X_2|/2$. For any $t > 0$,

$$\mathbb{P}(|Y| \ge t) \ge \mathbb{P}(|X_1| \ge 3t)\mathbb{P}(|X_2| \le t).$$

Using (A.3), (1.11) and Markov inequality, we find, for some constant $c_0 > 0$,

$$\mathbb{P}(|Y| \ge 1/(3a)) \ge b\sigma^2(1 - \mathbb{E}|X_2|^2a^2) \ge b\sigma^2(1 - c_0\sigma^2a^2).$$

In particular, if $c_0\sigma^2a^2 < 1/2$ then there exists a constant $c > 0$ such that

$$\mathbb{P}(|Y| \ge c) \ge c\sigma^2.$$

On the other hand, if $c_0\sigma^2a^2 \ge 1/2$, from (A.3), we may use Paley-Zygmund inequality and deduce easily that the above inequality also holds (for some new constant c).

Let $\hat{Y}$ the law of the variable Y conditioned on $|Y| \ge c$. It follows that if f is a non-negative measurable function,

$$\mathbb{E}f(Y) \ge c\sigma^2\mathbb{E}f(\hat{Y}). \tag{A.4}$$

From Esseen inequality, cf. [35, lemma 4.2], for some universal constant $C > 0$,

$$p_x(t) \le C\int_0^{2\pi} \Big|\mathbb{E}e^{i\frac{\theta}{t}\sum_{k=1}^m X_k x_k}\Big|\, d\theta = C\,t\int_0^{2\pi t^{-1}} \prod_{k=1}^m \big|\mathbb{E}e^{i\theta X_k x_k}\big|\, d\theta.$$

Now, we use the bound $|z| \le \exp(-\frac{1}{2}(1-|z|^2))$ valid for all z, and the identity,

$$\big|\mathbb{E}e^{i\theta X_k x_k}\big|^2 = \mathbb{E}\cos(2\theta Y x_k),$$

to write

$$\prod_{k=1}^n \big|\mathbb{E}e^{i\theta X_k x_k}\big| \le \prod_{k=1}^m \exp\big(-\frac{1}{2}(1 - \mathbb{E}\cos(2\theta Y x_k)\big) = \prod_{k=1}^m \exp\big(-\mathbb{E}\sin^2(\theta Y x_k)\big).$$

Therefore,

$$p_x(t) \le C\,t\int_0^{2\pi t^{-1}} \exp\Big(-\sum_{k=1}^m \mathbb{E}\sin^2(\theta x_k Y)\Big)d\theta.$$

This implies, using (A.4) and $|x_k| \ge 1$,

$$p_x(t) \le C\,t\sup_{\lambda \ge c}\int_0^{2\pi t^{-1}} \exp\Big(-c\,m\sigma^2\sin^2(\theta\lambda)\Big)d\theta.$$

By the change of variable $\theta' = \theta\lambda$, using $t^{-1} \le 1$, one has

$$p_x(t) \le C\,t\sup_{\lambda \ge c}\frac{1}{\lambda}\int_0^{2\pi\lambda} e^{-c\,m\sigma^2\sin^2(\theta)}d\theta \le \frac{C'\,t}{\sigma\sqrt{m}}. \tag{A.5}$$

The right hand side follows easily by decomposing the integral along the periods of $\sin(\theta)$.

It remains to prove the statement for complex random variables. It is a consequence of the real case. Indeed, we may assume for example that at least $m/2$ of the x_i's satisfy $\mathfrak{Re}(x_i) \ge 1/\sqrt{2}$ (otherwise, this is satisfied by the imaginary part). We then notice that the function p_x does not change if we rotate $\mathbf{x}$ into $e^{iu}\mathbf{x}$ with $u \in [0, 2\pi]$. But, we may argue as in [37, lemma 2.4] : there exists u such that for all $z \in \mathbb{C}$,

$$\mathrm{Var}\Big(\mathfrak{Re}(e^{iu}\mathbf{x}z)\Big) \ge \mathfrak{Re}(z)^2\frac{\sigma^2}{2}.$$

Finally, we note that

$$p_x(t) \le \max_{w\in\mathbb{C}} \mathbb{P}\Big(\Big|\mathfrak{Re}\Big(\sum_{k=1}^m e^{iu} X_k x_k - w\Big)\Big| \le t\Big) = \max_{w\in\mathbb{R}} \mathbb{P}\Big(\Big|\sum_{k=1}^m \mathfrak{Re}(e^{iu} X_k x_k) - w\Big| \le t\Big).$$

Therefore, the case of complex random variables follows from the case of real random variable. □

If $\sigma\sqrt{m}$ is not large, the bound in theorem A.1 becomes useless. In this case, we may however give a weak bound on the concentration function.

Lemma A.2. *Let $1 \le m \le n$. There exists a constant $c_0 > 0$ independent of (m, n) such that if x satisfies $|x_i| \ge 1$, for $1 \le i \le m$, then*

$$p_x(c_0) \le 1 - c_0(m\sigma^2 \wedge 1)\,. \tag{A.6}$$

Proof. Let us start with some comments. First from (A.3), (1.11) and Markov inequality,

$$\mathbb{P}(|X_1| \ge 1/(2a)) \le 4\mathbb{E}|X_1|^2 a^2 \le C\sigma^2 a^2 \quad \text{and} \quad \mathbb{P}(|X_1| \ge 1/a) \ge b\sigma^2.$$

Hence, if $|x_1| \ge 1$, for $n = 1$, we find $p_{x_1}(1/(2a)) \le 1 - c_0(\sigma^2 \wedge 1)$. Since the concentration function of the sum is bounded by the concentration function of one of its summands, we get (A.6) for $m = 1$.

Also, from theorem A.1, if $\sigma\sqrt{m} \ge 4C$, then $p_x(1) \le 1/2$. Hence, it suffices to prove the statement for $\sigma^2 m \le 16C^2$. Then, simple manipulations show that, at the price of reducing the constant c_0, it suffices to prove the statement of the lemma with the extra assumptions $\sigma^2 m \le \varepsilon$ and $m \ge 1/\varepsilon$ for some $\varepsilon > 0$ arbitrarily small (but independent of m, n).

We will first give a proof in the specific case when $\mathbf{x}$ is a Bernoulli random variable with parameter p. We will generalize the argument afterward. We fix $t = 1/4$ and define $p_k = p_{(x_1,\dots,x_k)}(t)$. We will assume that $pm \le 1/4$ and $m \ge 4$. We are going to show that $p_m \le 1 - mp/4$. Let $k \ge 2$ and assume that $p_{k-1} \ge 1 - kp$. Let $S = \sum_{i=1}^{k-1} X_i x_i$ and u_* be such that $p_{k-1} = \mathbb{P}(|S - u_*| \le t)$. We write, for $u \in \mathbb{C}$,

$$\mathbb{P}(|S + x_k X_k - u| \le t) = (1-p)\mathbb{P}(|S - u| \le t) + p\mathbb{P}(|S - u + x_k| \le t)$$

Now, since $|x_k| \ge 1$, u and $u - x_k$ cannot be both at distance less than $t = 1/4$ from u_*. Using $\mathbb{P}(|S - u_*| > t) \le kp$ and $\mathbb{P}(|S - u| \le t) \le p_{k-1}$, we deduce

$$\begin{aligned} p_k &\le \max\big((1-p)p_{k-1} + kp^2, (1-p)kp + pp_{k-1}\big) \\ &\le \max\left(p_{k-1} - p(1-2kp), p(k+1-2kp)\right) \\ &\le p_{k-1} - p/2, \end{aligned}$$

provided that $pk \le 1/4$. Now the argument goes as follows, if there exists $\lfloor m/2\rfloor \le k < m$ such that $p_k \le 1 - kp$, then since $p_m \le p_k$, we find $p_m \le 1 - \lfloor m/2\rfloor p \le 1 - mp/4$. Otherwise, for all $\lfloor m/2\rfloor \le k < m$, $p_k \ge 1 - kp$, and we apply recursively the above argument. We find $p_m \le p_{\lfloor m/2\rfloor} - (m - \lfloor m/2\rfloor)p/2 \le 1 - mp/4$. This concludes the proof when $\mathbf{x}$ is Bernoulli random variable.

In the general case, we take $t = 1/(4a)$, with $a > 0$ as in (A.2). We have $p = b\sigma^2 \le \mathbb{P}(|X_1| \ge 1/a)$, and, from Markov inequality, $1 - q = 1 - c\sigma^2 \le \mathbb{P}(|X_1| \le 1/(4a))$. Setting $p_k = p_{(x_1,\dots,x_k)}(t)$, the above argument gives $p_m \le 1 - mp/4$ as soon as $\sigma^2 m$ is small enough. Indeed, as above, let $k \ge 2$ and assume that $p_{k-1} \ge 1 - kp$. Note that, with the above notation, if $|X_k - X_k'| \ge 3/(4a)$, then $u + x_k X_k'$ and $u + x_k X_k$ cannot be both at distance less than $t = a/4$ from u_*. In particular, from

$$\begin{aligned} \mathbb{P}(|S + x_k X_k - u| \le t) \;\le\; & (q-p)p_{k-1} \\ & + (1-q)\mathbb{P}(|S - u + x_k X_k| \le t \mid |X_k| \le 1/(4a)) \\ & + p\mathbb{P}(|S - u + x_k X_k| \le t \mid |X_k| \ge 1/a), \end{aligned}$$

we get

$$\begin{aligned} p_k &\le \max\big((q-p)p_{k-1} + (1-q)p_{k-1} + kp^2, (q-p)p_{k-1} + (1-q)kp + pp_{k-1}\big) \\ &\le \max\left(p_{k-1} - p(1-2kp), p(k + \frac{q}{p} - 2kq)\right) \\ &\le p_{k-1} - p/2, \end{aligned}$$

provided that $\sigma^2 k$ is small enough. The rest of the argument is identical. □

The following corollary is a version of [22, lemma 4.6] (our variable $\mathbf{x}$ satisfies however more general statistical assumptions than in [22]).

Corollary A.3. *There exists $c > 0$ such that if $x \in \mathrm{Incomp}(\delta, \rho)$, with $\delta, \rho \in (0,1)$, $u \in \mathbb{C}^n$, then*

$$\mathbb{P}(\|Xx - u\|_2 \le c\,\rho\sqrt{\delta}) \le e^{-c\,n(1\wedge\sigma^2\delta n)}. \tag{A.7}$$

Proof. We have

$$\|Xx - u\|_2 = \sqrt{\sum_{i=1}^n S_i^2} \quad \text{with} \quad S_i = \left|\sum_{k=1}^n X_{ik}x_k - u_i\right|.$$

Thanks to lemma 3.6 one may assume that $|x_i| \ge \rho/\sqrt{n}$ for all $1 \le i \le \delta n/2$. From theorem A.1, one obtains

$$\mathbb{P}\left(S_i \le \frac{\rho t}{\sqrt{n}}\right) \le p_x(\rho t/\sqrt{n}) \le \frac{C\,(t+1)}{\sigma\sqrt{\delta n}},$$

for some constant $C > 0$, and all $t \ge 0$. Assume first that $\sigma\sqrt{\delta n} \ge 4C$. Then, taking $t = 1$, we get

$$\mathbb{P}\left(S_i \le \frac{\rho}{\sqrt{n}}\right) \le \frac{1}{2}.$$

The S_i's are independent random variables. By Hoeffding's inequality,

$$\mathbb{P}(\|Xx - u\|_2 \le \rho/2) \le \mathbb{P}\left(\sum_{i=1}^n \mathbf{1}_{\{S_i \ge \frac{\rho}{\sqrt{n}}\}} \le n/4\right) \le e^{-n/8}.$$

This implies (A.7).

It remains to consider the case $\sigma\sqrt{\delta n} < 4C$. Here, we may use lemma A.2 with $m = \delta n/2$. Therefore, there exists $c_0 > 0$, such that

$$\mathbb{P}\left(S_i \le \frac{\rho c_0}{\sqrt{n}}\right) \le p_x\left(\frac{\rho c_0}{\sqrt{n}}\right) \le 1 - c_0\sigma^2\delta n. \tag{A.8}$$

Once (A.8) is available, we can conclude the proof as follows. From Bennett's inequality, there exists $c > 0$ such that if ε_i are i.i.d. Bernoulli random variables with parameter q, $0 < q < 1/2$,

$$\mathbb{P}\left(\sum_{i=1}^n \varepsilon_i \le \frac{nq}{4}\right) \le \mathbb{P}\left(\sum_{i=1}^n \mathbb{E}\varepsilon_i - \varepsilon_i \ge \frac{3nq}{4}\right) \le e^{-cnq}.$$

From (A.8), we can take $\varepsilon_i = \mathbf{1}_{\{S_i > \frac{c_0\rho}{\sqrt{n}}\}}$, with $1/2 \ge q \ge c_0 n\sigma^2\delta$. Then, using $n\sigma^2 \ge 1$, one has that $\|Xx - u\|_2 \le c\rho\sqrt{\delta}$ implies $\sum_{i=1}^n \varepsilon_i \le \frac{nq}{4}$, for some $c > 0$. It follows that for some new $c > 0$:

$$\mathbb{P}\Big(\|Xx - u\|_2 \le c\rho\sqrt{\delta}\Big) \le e^{-cn^2\sigma^2\delta}.$$

This ends the proof. □

Corollary A.3 implies a probabilistic bound on the distance of Xx to a vector space.

Corollary A.4. *There exists $c_1 > 0$ such that if $x \in \mathrm{Incomp}(\delta, \rho)$, with $\delta, \rho \in (0,1)$, $s \ge 1$, and H is a deterministic vector space such that*

$$\dim(H) \le \frac{c_1\, n(1 \wedge \sigma^2\delta n)}{\log(\frac{s}{\rho^2\delta})}, \tag{A.9}$$

then

$$\mathbb{P}\Big(\mathrm{dist}(Xx, H) \le c_1\,\rho\sqrt{\delta}\,;\, s_1(X) \le s\Big) \le e^{-c_1\,n(1\wedge\sigma^2\delta n)}.$$

Proof. By definition,

$$\mathrm{dist}(Xx, H) = \min_{u\in H} \|Xx - u\|_2.$$

On the event $s_1(X) \le s$, if $\|u\|_2 \ge 2s$ then $\|Xx - u\|_2 \ge s \ge 1 \ge \rho\sqrt{\delta}$. Hence, we may restrict our attention to

$$\min_{u \in B_H(s)} \|Xx - u\|_2,$$

where $B_H(s) = \{u \in H; \|u\|_2 \leq 2s\}$. Now, if $\varepsilon < 1$ and $k = \dim(H)$, there exists an ε-net N of $B_H(s)$ of cardinality $(5s/\varepsilon)^k$, with

$$\min_{u \in B_H(s)} \|Xx - u\|_2 \geq \min_{u \in N} \|Xx - u\|_2 - \varepsilon.$$

Let $c > 0$ and take $\varepsilon = \frac{c\rho\sqrt{\delta}}{2}$. From the union bound,

$$\mathbb{P}\left(\operatorname{dist}(Xx, H) \leq \frac{c\rho\sqrt{\delta}}{2} \,;\, s_1(X) \leq s \right) \leq \left(\frac{10s}{c\rho\sqrt{\delta}} \right)^k \max_{u \in \mathbb{C}^n} \mathbb{P}\Big(\|Xx - u\|_2 \leq c\rho\sqrt{\delta} \Big).$$

Then if c is chosen as in corollary A.3, we find

$$\mathbb{P}\left(\operatorname{dist}(Xx, H) \leq \frac{c\rho\sqrt{\delta}}{2} \,;\, s_1(X) \leq s \right) \leq e^{k \log\left(\frac{10s}{c\rho\sqrt{\delta}}\right) - cn(1 \wedge \sigma^2\delta n)}.$$

In particular if $k = \dim(H)$ satisfies (A.9), then the conclusion follows by taking c_1 sufficiently small. □

Acknowledgments

The idea of studying this model came from animated discussions with Laurent Miclo in Marne-la-Vallée. We are grateful to Philippe Biane for helpful suggestions regarding the algebraic literature. This work was supported by the GDRE GREFI-MEFI CNRS-INdAM. Support of the European Research Council through the "Advanced Grant" PTRELSS 228032 is also acknowledged.

References

[1] G. W. Anderson, A. Guionnet and O. Zeitouni – *An introduction to random matrices*, Cambridge Studies in Advanced Mathematics, vol. 118, Cambridge University Press, Cambridge, 2010.
[2] A. L. Andrew – "Eigenvalues and singular values of certain random matrices", *J. Comput. Appl. Math.* **30** (1990), no. 2, p. 165–171.
[3] Z. D. Bai and Y. Q. Yin – "Limit of the smallest eigenvalue of a large-dimensional sample covariance matrix", *Ann. Probab.* **21** (1993), no. 3, p. 1275–1294.
[4] F. Benaych-Georges and R. R. Nadakuditi – "The eigenvalues and eigenvectors of finite, low rank perturbations of large random matrices", *Adv. Math.* **227** (2011), no. 1, p. 494–521.
[5] R. Bhatia – *Perturbation bounds for matrix eigenvalues*, Classics in Applied Mathematics, vol. 53, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2007, Reprint of the 1987 original.
[6] P. Biane – "On the free convolution with a semi-circular distribution", *Indiana Univ. Math. J.* **46** (1997), no. 3, p. 705–718.
[7] P. Biane and F. Lehner – "Computation of some examples of Brown's spectral measure in free probability", *Colloq. Math.* **90** (2001), no. 2, p. 181–211.
[8] B. Bollobás – *Random graphs*, second ed., Cambridge Studies in Advanced Mathematics, vol. 73, Cambridge University Press, Cambridge, 2001.
[9] C. Bordenave, P. Caputo and D. Chafaï – "Circular Law Theorem for Random Markov Matrices", preprint arXiv:0808.1502 [math.PR] to appear in Probability Theory and Related Fields, 2011.
[10] ———, "Spectrum of large random reversible Markov chains: Heavy-tailed weigths on the complete graph", *Annals of Probability* **39** (2011), no. 4, p. 1544–1590.
[11] ———, "Spectrum of non-Hermitian heavy tailed random matrices", *Communications in Mathematical Physics* **307** (2011), no. 2, p. 513–560.
[12] C. Bordenave and D. Chafaï – "Around the circular law", *Probab. Surveys* **9** (2012), no. 0, p. 1–89.
[13] L. G. Brown – "Lidskiĭ's theorem in the type II case", Geometric methods in operator algebras (Kyoto, 1983), Pitman Res. Notes Math. Ser., vol. 123, Longman Sci. Tech., Harlow, 1986, p. 1–35.
[14] W. Bryc, A. Dembo and T. Jiang – "Spectral measure of large random Hankel, Markov and Toeplitz matrices", *Ann. Probab.* **34** (2006), no. 1, p. 1–38.
[15] M. Capitaine and M. Casalis – "Asymptotic freeness by generalized moments for Gaussian and Wishart matrices. Application to beta random matrices", *Indiana Univ. Math. J.* **53** (2004), no. 2, p. 397–431.
[16] D. Chafaï – "Circular law for noncentral random matrices", *Journal of Theoretical Probability* **23** (2010), no. 4, p. 945–950.
[17] J. B. Conway – *A course in functional analysis*, second ed., Graduate Texts in Mathematics, vol. 96, Springer-Verlag, New York, 1990.
[18] C. Cooper and A. Frieze – "Stationary distribution and cover time of random walks on random digraphs", preprint arXiv:1103.4317v1 [math.CO], 2011.
[19] P. Erdős and A. Rényi – "On random graphs. I", *Publ. Math. Debrecen* **6** (1959), p. 290–297.
[20] J. Feinberg and A. Zee – "Non-Hermitian Random Matrix Theory: Method of Hermitian Reduction", *Nucl. Phys. B* (1997), no. 3, p. 579–608.

[21] Z. Füredi and J. Komlós – "The eigenvalues of random symmetric matrices", *Combinatorica* **1** (1981), no. 3, p. 233–241.
[22] F. Götze and A. Tikhomirov – "The circular law for random matrices", *Ann. Probab.* **38** (2010), no. 4, p. 1444–1491.
[23] E. Gudowska-Nowak, A. Jarosz, M. Nowak and G. Pappe – "Towards non-Hermitian random Lévy matrices", *Acta Physica Polonica B* **38** (2007), no. 13, p. 4089–4104.
[24] A. Guntuboyina and H. Leeb – "Concentration of the spectral measure of large Wishart matrices with dependent entries", *Electron. Commun. Probab.* **14** (2009), p. 334–342.
[25] U. Haagerup and F. Larsen – "Brown's spectral distribution measure for R-diagonal elements in finite von Neumann algebras", *J. Funct. Anal.* **176** (2000), no. 2, p. 331–367.
[26] U. Haagerup and H. Schultz – "Brown measures of unbounded operators affiliated with a finite von Neumann algebra", *Math. Scand.* **100** (2007), no. 2, p. 209–263.
[27] G. Halász – "Estimates for the concentration function of combinatorial number theory and probability", *Period. Math. Hungar.* **8** (1977), no. 3-4, p. 197–211.
[28] R. A. Horn and C. R. Johnson – *Matrix analysis*, Cambridge University Press, Cambridge, 1990, Corrected reprint of the 1985 original.
[29] ———, *Topics in matrix analysis*, Cambridge University Press, Cambridge, 1994, Corrected reprint of the 1991 original.
[30] S. Janson, T. Łuczak and A. Rucinski – *Random graphs*, Wiley-Interscience Series in Discrete Mathematics and Optimization, Wiley-Interscience, New York, 2000.
[31] A. E. Litvak, A. Pajor, M. Rudelson and N. Tomczak-Jaegermann – "Smallest singular value of random matrices and geometry of random polytopes", *Adv. Math.* **195** (2005), no. 2, p. 491–523.
[32] P. Matchett Wood – "Universality and the circular law for sparse random matrices", preprint , 2010.
[33] M. Reed and B. Simon – *Methods of modern mathematical physics. I*, second ed., Academic Press Inc. [Harcourt Brace Jovanovich Publishers], New York, 1980, Functional analysis.
[34] T. Rogers – "Universal sum and product rules for random matrices", *J. Math. Phys.* **51** (2010), no. 9, p. 093304, 15.
[35] M. Rudelson and R. Vershynin – "The Littlewood-Offord problem and invertibility of random matrices", *Adv. Math.* **218** (2008), no. 2, p. 600–633.
[36] J. W. Silverstein – "The spectral radii and norms of large-dimensional non-central random matrices", *Comm. Statist. Stochastic Models* **10** (1994), no. 3, p. 525–532.
[37] T. Tao and V. Vu – "Random matrices: the circular law", *Commun. Contemp. Math.* **10** (2008), no. 2, p. 261–307.
[38] ———, "Random matrices: universality of ESDs and the circular law", *Ann. Probab.* **38** (2010), no. 5, p. 2023–2065, With an appendix by Manjunath Krishnapur.
[39] T. Tao – "Outliers in the spectrum of iid matrices with bounded rank perturbations", preprint arXiv:1012.4818 [math.PR] to appear in Probability Theory and Related Fields, 2011.
[40] T. Tao and V. Vu – "From the Littlewood-Offord problem to the circular law: universality of the spectral distribution of random matrices", *Bull. Amer. Math. Soc. (N.S.)* **46** (2009), no. 3, p. 377–396.
[41] D. V. Voiculescu, K. J. Dykema and A. Nica – *Free random variables*, CRM Monograph Series, vol. 1, American Mathematical Society, Providence, RI, 1992, A noncommutative probability approach to free products with applications to random matrices, operator algebras and harmonic analysis on free groups.
[42] D. Voiculescu – "The coalgebra of the free difference quotient and free probability", *Internat. Math. Res. Notices* (2000), no. 2, p. 79–106.
[43] V. H. Vu – "Spectral norm of random matrices", *Combinatorica* **27** (2007), no. 6, p. 721–736.
[44] Y. Q. Yin, Z. D. Bai and P. R. Krishnaiah – "On the limit of the largest eigenvalue of the large-dimensional sample covariance matrix", *Probab. Theory Related Fields* **78** (1988), no. 4, p. 509–521.

(Ch. Bordenave) Université de Toulouse, UMR 5219 CNRS Institut de Mathématiques de Toulouse, 118 route de Narbonne, F-31062 Toulouse, France
E-mail address: charles.bordenave(at)math.univ-toulouse.fr
URL: http://www.math.univ-toulouse.fr/~bordenave/

(P. Caputo) Dipartimento di Matematica, Università Roma Tre, Largo San Murialdo 1, 00146 Roma, Italy
E-mail address: caputo(at)mat.uniroma3.it
URL: http://www.mat.uniroma3.it/users/caputo/

(D. Chafaï) Université Paris-Est Marne-la-Vallée, UMR 8050 CNRS, 5 boulevard Descartes, F-77454 Champs-sur-Marne, France
E-mail address: djalil(at)chafai.net
URL: http://djalil.chafai.net/

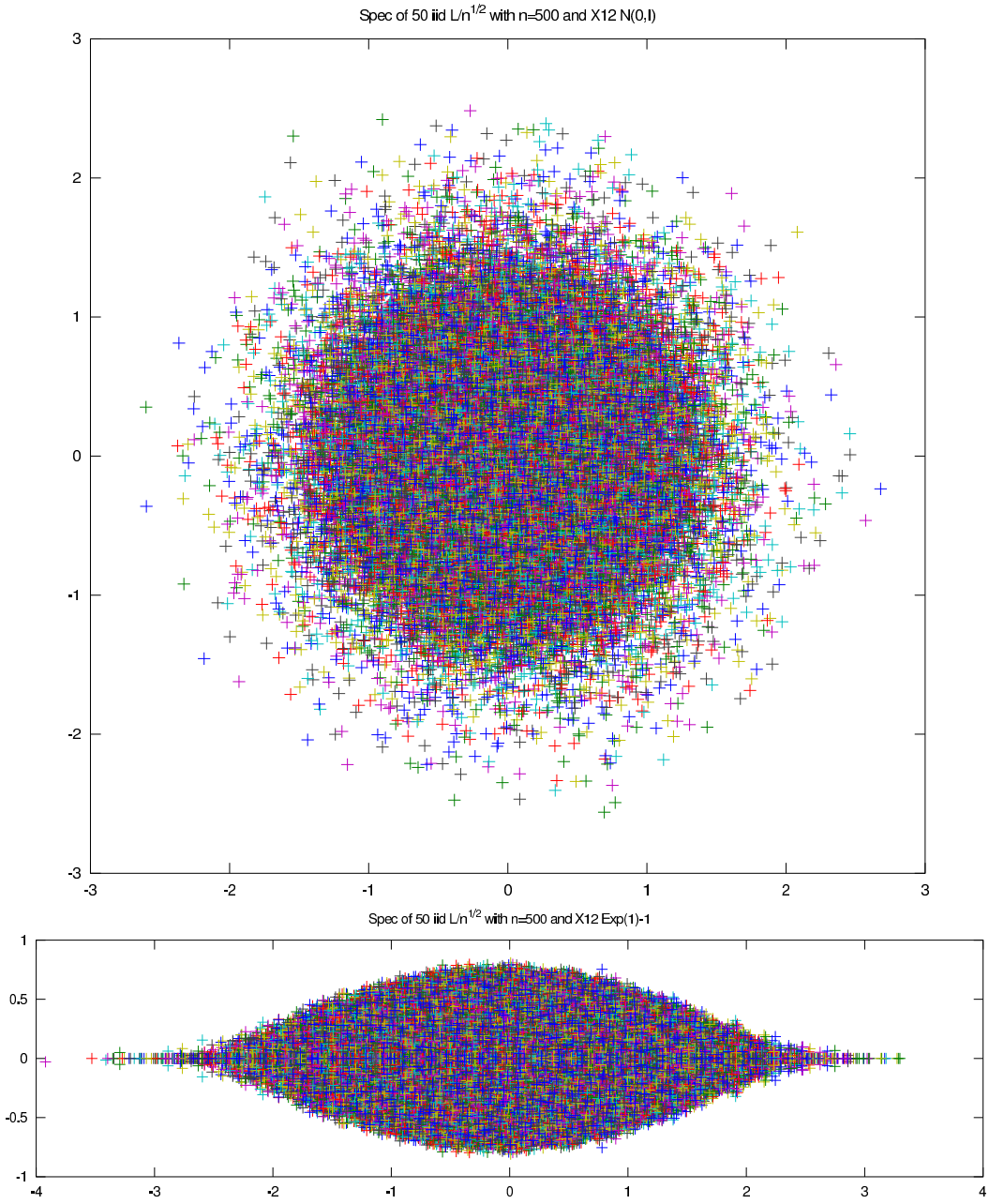

FIGURE 1.1. The bottom graphic shows a simulation of 50 i.i.d. copies of the spectrum of $n^{-1/2}L$ with $n = 500$ when $\mathcal{L}$ is the exponential law on $\mathbb{R}$ of parameter 1 shifted by -1 ($m = 0$, $K_{11} = 1$, and $K_{12} = K_{21} = K_{22} = 0$). The top graphics shows a simulation of 50 i.i.d. copies of the spectrum of $n^{-1/2}L$ with $n = 500$ when $\mathcal{L}$ is the Gaussian law on $\mathbb{C}$ of mean 0 and covariance $K = \frac{1}{2}I_2$. These simulations and graphics were produced with the free software GNU Octave provided by Debian GNU/Linux.

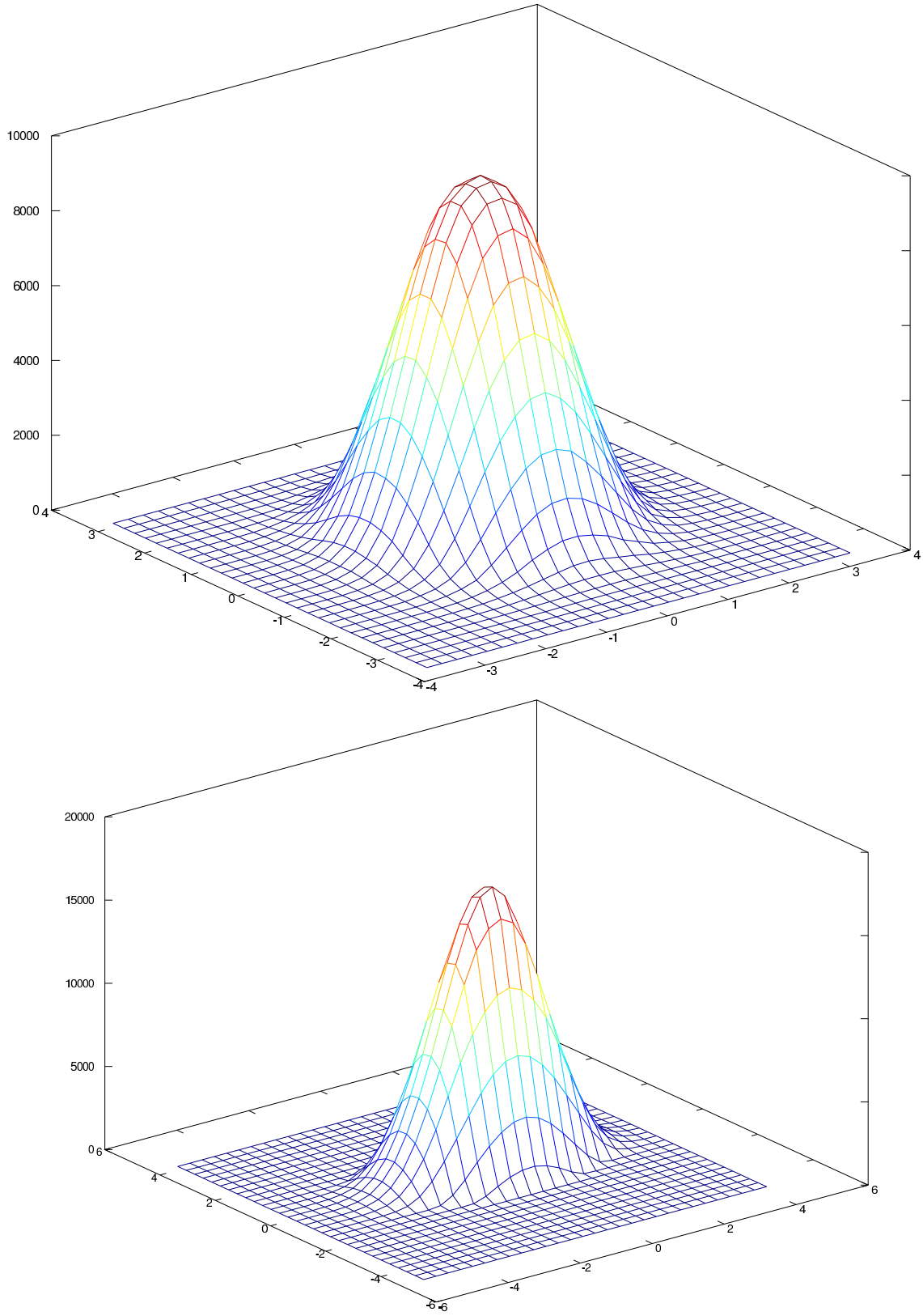

FIGURE 1.2. Kernel estimators of the densities for the simulations in figure 1.1.