Universality of the least singular value for the sum of random matrices[*]

Ziliang Che[†]   Patrick Lopatto[‡]

Abstract

We consider the least singular value of $M = R^* X T + U^* Y V$, where R, T, U, V are independent Haar-distributed unitary matrices and X, Y are deterministic diagonal matrices. Under weak conditions on X and Y, we show that the limiting distribution of the least singular value of M, suitably rescaled, is the same as the limiting distribution for the least singular value of a matrix of i.i.d. Gaussian random variables. Our proof is based on the dynamical method used by Che and Landon to study the local spectral statistics of sums of Hermitian matrices.

Contents

1 Introduction

The problem of effectively bounding the least singular value of a random matrix with independent entries has received tremendous attention from mathematicians and computer scientists [17, 35, 36, 62, 66–70, 72, 74–77]. Recent investigations have also focused on matrices with dependent entries [2, 27, 38, 52, 57, 63, 78] and random tensors [1, 10, 25, 79].

A closely related question is to determine the limiting distribution of the least singular value, suitably rescaled, as the size of the matrix tends to infinity. For square matrices with independent entries, it is known that this distribution does not depend on the entry distributions and is equal to the one obtained from a matrix of i.i.d. Gaussian random variables (which may be computed exactly). This phenomenon is known as *universality* of the least singular value and was proved for entry distributions with mean zero and variance one in [73] using ideas from the method of property testing in the study of algorithms. In [34], universality of the least singular value for square matrices was studied from a dynamical viewpoint and shown to hold for matrices whose entries may be sparse, weakly correlated, and have unequal variances. We also note that the case of genuinely rectangular matrices was taken up in [51].

In this work we prove universality of the least singular value for the matrix

$$M = R^\dagger X T + U^\dagger Y V, \tag{1.1}$$

where X and Y are deterministic matrices and R, T, U, V are independently sampled from the Haar measure on the unitary group. The model M is a natural interpretation of the notion of the sum of two generic random matrices and exhibits strong correlations

between its entries, unlike the matrices studied in [34, 73]. It was previously studied in [16], where its spectrum was controlled on scale $N^{-1+\varepsilon}$.

The Hermitian version of this model, $H = V^*XV + U^*YU$, has attracted significant interest. The weak convergence of the empirical distribution was obtained first in [80] and later shown in [26, 37, 65, 71] using alternative techniques. Convergence was then investigated on scales decaying in N in [55, 56] and established on the optimal scale $N^{-1+\varepsilon}$ through the series of works [12–15]. The latter results were used to show universality of local spectral statistics in the bulk of the spectrum [33]. Very recently, [6] provided a detailed analysis of the edge of the spectrum.

Our work is based on the dynamical approach to random matrix theory developed in the last decade. Based on resolvent estimates and a precise analysis of the short-time behavior of Dyson Brownian motion [28, 39, 44, 58, 59], it has succeeded in its original goal of establishing the universality of local spectral statistics for Wigner matrices [29, 41, 45–50, 61], and has since been applied to investigate universality for numerous other random matrix models. These include random graph models [3, 18–20, 42, 43, 54], general Wigner-type matrices [8, 9], band matrices [30–32, 82], and matrices with few moments [4, 5].

Our proof follows closely the method used in [33] to show universality for the Hermitian model. Two primary inputs in that work were a carefully chosen flow $U(t)$ on the unitary group which leaves the eigenvalue distribution of H unchanged but produces a system of SDEs for the eigenvalue process closely resembling Dyson Brownian motion, and a weak local law throughout the spectrum (including the spectral edges) which was used as an *a priori* input to study the flow of the eigenvalues. We show how similar inputs may be obtained for the model M through a slightly more involved analysis, which proceeds by transforming the problem from one about the singular values of a $N \times N$ non-Hermitian matrix to the eigenvalues of a $2N \times 2N$ Hermitian matrix. The resulting eigenvalue flow is not a Dyson Brownian motion, but instead similar to a symmetrized version studied in [34], and the short-time relaxation result for the symmetrized flow in that work is a crucial input here.

Compared to [33], we derive the weak local law in a slightly different way, involving a general stability analysis of the system of equations that define the free convolution of two measures. While the essential technical content is unchanged, this somewhat streamlines the proofs. Further, we use [16] to prove a strong law at small energies, paralleling the use of [12–15] in [33] to establish a strong law in the bulk of the spectrum. We also comment on an interesting difference between the real and complex cases which does not arise in the Hermitian model.

2 Overview and main result

2.1 Overview

In this section we define the model under consideration and state our main result. The main technical input is Theorem 7.6 about short-time universality for the singular values of the model as it undergoes a time-dependent perturbation. Its proof occupies the bulk of this work. In Section 3 we define this perturbation and the associated stochastic differential equations governing the evolution of the singular values. In Sections 4 and 5 we prove various estimates necessary to study the short-time behavior of these SDEs. Their well posedness and the fact that they represent the claimed singular value evolution are proved in Section 6. In Section 7 we compare the SDEs for the singular values to a symmetrized Dyson Brownian Motion flow. The short-time behavior of this flow was studied in [34], and by combining our comparison with the main result of that work, we achieve a proof of Theorem 7.6. As corollary we deduce our main result,

Theorem 2.1. Appendix A contains a computation using Itô's formula that is required for Section 6.

For concreteness, we focus on a model where deterministic initial data is conjugated by unitary matrices. It is natural to also consider the analogous model with conjugation by orthogonal matrices. The SDEs for the evolution of the singular values in the second case lack a certain influential repulsion term compared to the first, and as a result the least singular value displays qualitatively different behavior. (This distinction was already noted in [34] for a different ensemble). Fortunately, our methods suffice to treat this case too. The difference between the behavior of the least singular value in the real and complex models and the necessary modifications to the proof are discussed in Appendix B.

Finally, Appendix C contains some preliminary estimates required for our analysis.

2.2 Model

Define

$$M = R^\dagger X T + U^\dagger Y V, \tag{2.1}$$

where $X = \mathrm{diag}(x_1, \ldots, x_N)$ and $Y = \mathrm{diag}(y_1, \ldots, y_N)$ are deterministic diagonal matrices and R, T, U, V are independent and distributed according to the Haar measure on the unitary group $U(N)$. We suppose that

$$0 \le x_i \le C_0, \quad 0 \le y_i \le C_0 \tag{2.2}$$

for some constant C_0 independent of N. Denote the empirical measures of X and Y by

$$\mu_X = \frac{1}{N} \sum_{1 \le i \le N} \delta_{x_i}, \quad \mu_Y = \frac{1}{N} \sum_{1 \le i \le N} \delta_{y_i}. \tag{2.3}$$

For integers $1 \le i \le N$, we define

$$x_{-i} = -x_i, \quad y_{-i} = -y_i. \tag{2.4}$$

We denote the symmetrized version of the empirical measures of X and Y by[1]

$$\mu_X^{\mathrm{sym}} = \frac{1}{2N} \sum_{1 \le |i| \le N} \delta_{x_i}, \quad \mu_Y^{\mathrm{sym}} = \frac{1}{2N} \sum_{1 \le |i| \le N} \delta_{y_i}. \tag{2.5}$$

Let $\mathbb{C}^+ = \{z : \mathrm{Im}\, z > 0\}$ denote the complex upper half plane. For $z \in \mathbb{C}^+$, define the Stietjes transforms

$$m_X(z) = \int_{\mathbb{R}} \frac{\mu_X^{\mathrm{sym}}(\mathrm{d}x)}{x - z}, \quad m_Y(z) = \int_{\mathbb{R}} \frac{\mu_Y^{\mathrm{sym}}(\mathrm{d}y)}{y - z}. \tag{2.6}$$

We assume:

1. For any $a > 0$, there is a constant $C_a > 0$ independent of N such that

$$\sup_{E \in \mathbb{R},\, \eta \ge N^{-1+a}} |m_Y(E + i\eta)| \le C_a. \tag{2.7}$$

2. There are compactly supported probability measures μ_1, μ_2 such that $\mu_X \to \mu_1$ and $\mu_Y \to \mu_2$ weakly, and at least one of μ_1, μ_2 has a bounded Stieltjes transform.[2]

---

[1] For a general Borel measure ν we define $\nu^{\mathrm{sym}}(A) = \frac{1}{2}[\mu(A) + \mu(-A)]$ for any Borel set $A \subset \mathbb{R}$.
[2] Observe this implies $\mu_X^{\mathrm{sym}} \to \mu_1^{\mathrm{sym}}$ and $\mu_Y^{\mathrm{sym}} \to \mu_2^{\mathrm{sym}}$ weakly.

3. Neither of $\mu_1^{\mathrm{sym}}, \mu_2^{\mathrm{sym}}$ is a single point mass, and at least one is supported at more than two points.

4. The Stieltjes transform of the measure μ_X converges to that of μ_1 with polynomial speed, in the sense that there exists a constant $\mathfrak{c}_X > 0$ such that

$$\left| \frac{1}{N} \sum_{i=1}^{N} \frac{1}{x_i - E - i\eta} - \int \frac{d\mu_1(x)}{x - E - i\eta} \right| \leq N^{-\mathfrak{c}_X} \tag{2.8}$$

for $\eta \geq N^{-\mathfrak{c}_X}$.

5. The particle y_k is close to the deterministic location y_k^* in the sense that for any $c > 0$,

$$\sup_{1 \leq k \leq N} |y_k - y_k^*| \leq N^{-1+c}, \tag{2.9}$$

where y_k^* is the k-th N-quantile of μ_2 defined by

$$y_k^* = \inf \left\{ s \colon \int_{-\infty}^{s} \mu_2(dy) = \frac{k}{N} \right\}. \tag{2.10}$$

6. The measure μ_2 has a continuous density and there are constants $\mathfrak{c}, \delta_0 > 0$ such that for any $x \in \mathrm{supp}\,\mu_2$ and $0 \leq h \leq \delta_0$,

$$\mu_2([x-h, x+h]) \geq h^{2-\mathfrak{c}}. \tag{2.11}$$

7. The free convolution[3] $\mu_1^{\mathrm{sym}} \boxplus \mu_2^{\mathrm{sym}}$ has a density $\rho(x)$ in a neighborhood of zero such that

$$c \leq \rho(x) \leq C, \quad \rho(0) = \frac{1}{\pi} \tag{2.12}$$

for some constants $C, c > 0$ and all $x \in [-c, c]$.

The first assumption is to prevent the y_i from accumulating around any point $E \in \mathbb{R}$. This is illustrated by Proposition C.1. The second and third are required to apply [16, Theorem 4.4] to control the Green functions of (a modification of) M, as is done in Section 5. The remaining assumptions are required in Subsection 5.2.

Assumption (5) requires that Y obey a strict rigidity condition. However, it can often be relaxed in practice, for instance near the the spectral edges of μ_2. In particular, Y can be taken to be the spectrum of a Wigner matrix, or more generally a matrix of general Wigner-type, as in [9]. We refer the reader to the remark in [33, Subsection 2.1] and [9, Corollary 2.11] for more on this point.

The condition (6) is technical and says, roughly, that the spectral edges of μ_2 behave sub-linearly, for example like the edges of the semicircle distribution $\rho_{\mathrm{sc}}(x) = (2\pi)^{-1}\sqrt{4 - x^2}\,dx$. While this is a strong condition, it is true for a broad class of spectral distributions arising in random matrix theory, including those coming from a matrix of general Wigner-type [9, Theorem 4.1]; see [7, Theorem 2.6] or [8, Theorem 2.6] for more.

The assumption (7) is difficult to check in general. For example, the case where μ_1 is a point mass was considered in [16, Theorem 2.2], whose proof is quite technical. In Appendix C we prove two simple sufficient conditions for (7): both μ_1^{sym} and μ_2^{sym} have positive density at zero, or $\mu_1^{\mathrm{sym}} = \mu_2^{\mathrm{sym}}$.

The second equation in (2.12) is necessary to make the scale of the smallest singular value match that of the analogous Gaussian ensemble. We include it for technical convenience, but it could be trivially removed by an appropriate rescaling of X and Y. We use it in Section 7.

---

[3]We recall the definition of free convolution in Subsection 4.4

2.3 Main result

The following is our main result. It is proved at the end of Section 7.

Theorem 2.1. Let $\lambda_1(M_N)$ be the least singular value of the random matrix ensemble (2.1) defined in Section 2.2. For all $r \geq 0$, we have

$$\mathbb{P}(N\lambda_1(M_N) \leq r) = 1 - e^{-r^2} + O(N^{-c}), \tag{2.13}$$

where $c > 0$ is an absolute constant uniform in r.

3 Definition of dynamics

3.1 Unitary Brownian motion

We use the following definitions. Recall a standard complex Gaussian random variable is such that its real and imaginary parts are independent mean zero normal distributions with variance $1/2$.

Definition 3.1. A complex-valued stochastic process $B(t)$ is called a *standard complex Brownian motion* if $(\sqrt{2}\operatorname{Im}B(t), \sqrt{2}\operatorname{Re}B(t))$ are independent real standard Brownian motions.

Definition 3.2. A complex-valued stochastic process $(B_{ij}(t))_{1 \leq i,j \leq N}$ is called a *Hermitian Brownian motion* if $(\sqrt{2}\operatorname{Im}B_{ij}, \sqrt{2}\operatorname{Re}B_{ij})_{i<j}$, $(B_{ii})_{1 \leq i \leq N}$ are a collection of independent real standard Brownian motions, and $\overline{B_{ij}} = B_{ji}$.

The following construction parallels [33, Section 2.3.1]. Given a parameter $\mathfrak{a} \in (0,1)$ we introduce the index set

$$\mathcal{I}_\mathfrak{a} = \{(i,j)\colon |y_i - y_j| \geq N^{-1+\mathfrak{a}}\} \tag{3.1}$$

and let $\mathcal{I}_\mathfrak{a}^c$ be the set of pairs (i,j) not in $\mathcal{I}_\mathfrak{a}$.

We let $U(0) := U$ and $V(0) := V$ evolve according to the following equations.

$$\mathrm{d}U(t) = i\mathrm{d}W_1 U(t) - \frac{1}{2}AU(t)\,\mathrm{d}t, \quad \mathrm{d}V(t) = i\mathrm{d}W_2 V(t) - \frac{1}{2}AV(t)\,\mathrm{d}t \tag{3.2}$$

Here $\mathrm{d}W_1, \mathrm{d}W_2$, and A are defined as follows. Let $\widetilde{W}_1$ and $\widetilde{W}_2$ be independent Hermitian Brownian motions in the sense of Definition 3.2. For $1 \leq i,j \leq N$, define the matrix processes W_1 and W_2 entrywise by

$$(W_1)_{ij} = \frac{\mathbb{1}_{(i,j)\in\mathcal{I}_\mathfrak{a}}}{\sqrt{2N}}\left(\frac{1}{|y_i - y_j|}(\widetilde{W}_1)_{ij} + \frac{1}{y_i + y_j}(\widetilde{W}_2)_{ij}\right), \tag{3.3}$$

$$(W_2)_{ij} = \frac{\mathbb{1}_{(i,j)\in\mathcal{I}_\mathfrak{a}}}{\sqrt{2N}}\left(\frac{1}{|y_i - y_j|}(\widetilde{W}_1)_{ij} - \frac{1}{y_i + y_j}(\widetilde{W}_2)_{ij}\right). \tag{3.4}$$

The diagonal matrix A in (3.2) is given by

$$A_{ii} = \frac{1}{2N}\sum_{j:(i,j)\in\mathcal{I}_\mathfrak{a}}\left(\frac{1}{(y_i - y_j)^2} + \frac{1}{(y_i + y_j)^2}\right). \tag{3.5}$$

Let us explain why these choices are made. With this definition of W_1 and W_2, we see that

$$(i\mathrm{d}W_1 Y - iY\mathrm{d}W_2)_{ij} = \frac{i\mathbb{1}_{(i,j)\in\mathcal{I}_\mathfrak{a}}}{\sqrt{2N}}\left(\operatorname{sgn}(y_j - y_i)(\mathrm{d}\widetilde{W}_1)_{ij} + (\mathrm{d}\widetilde{W}_2)_{ij}\right). \tag{3.6}$$

We therefore find by the Lévy criterion that $\sqrt{N}\left((i\mathrm{d}W_1 Y - iY\mathrm{d}W_2)_{ij}\right)_{(i,j)\in\mathcal{I}_\mathfrak{a}}$ is a family of independent standard complex Brownian motions. In particular, there is no longer a

Hermitian symmetry. We write

$$(idW_1 Y - iY dW_2)_{ij} = \frac{\mathbb{1}_{(i,j)\in\mathcal{I}_\mathfrak{a}}}{\sqrt{N}} d\widetilde{B}_{ij}, \qquad (3.7)$$

where $(d\widetilde{B}_{ij})_{1\le i,j\le N}$ is a family of independent standard complex Brownian motions. We choose A so that the solutions $U(t)$ and $V(t)$ stay on the unitary group. One can verify this by differentiating $(UU^*)(t)$ using Itô's formula and the above definitions to see that $d(UU^*)(t)$ is constant.

Having defined $U(t)$ and $V(t)$, we can differentiate $M(t)$ (defined in the obvious way using $U(t)$ and $V(t)$) and use (3.7) to see

$$dM(t) = \frac{1}{\sqrt{N}} U^*(t)(\mathbb{1}_{(i,j)\in\mathcal{I}_\mathfrak{a}} d\widetilde{B}_{ij})V(t) + U^*(t)\widehat{A}V(t)\, dt, \qquad (3.8)$$

where $\widehat{A}$ is a diagonal matrix whose entries are given by

$$\widehat{A}_{ii} = \frac{1}{2N} \sum_{j:(i,j)\in\mathcal{I}_\mathfrak{a}} \left(\frac{1}{y_j - y_i} - \frac{1}{y_i + y_j} \right). \qquad (3.9)$$

3.2 Canceling mesoscopic drift

Let $\tau = N^{-1+\mathfrak{b}}$ for a parameter $\mathfrak{b} > 0$ that will be chosen in the next section. It is hard to use (3.8) as written because after time τ, the contribution from the second term will be order $N^{-1+\mathfrak{b}}$, which is larger than the order of the microscopic statistic we are interested in. Therefore we introduce an auxiliary matrix

$$\widehat{M}(t) = M(t) + (\tau - t)U^*(0)\widehat{A}V(0). \qquad (3.10)$$

The process $\widehat{M}(t)$ has the property that $M(\tau) = \widehat{M}(\tau)$ and

$$d\widehat{M} = \frac{1}{\sqrt{N}} U^*(t) \left(\mathbb{1}_{(i,j)\in\mathcal{I}_\mathfrak{a}} d\widetilde{B}_{ij} \right) V(t) + \left(U^*(t)\widehat{A}V(t) - U^*(0)\widehat{A}V(0) \right) dt \qquad (3.11)$$

$$= \frac{1}{\sqrt{N}} d\widehat{B} + \left(U^*(t)\widehat{A}V(t) - U^*(0)\widehat{A}V(0) \right) dt - \frac{1}{\sqrt{N}} U^*(t) \left(\mathbb{1}_{(i,j)\in\mathcal{I}_\mathfrak{a}^c} d\widetilde{B}_{ij} \right) V(t). \qquad (3.12)$$

Here $\widehat{B}$ is a matrix of standard complex Brownian motions. We show in Section 7 that the second term, when integrated from 0 to τ, is $o(N^{-1})$. This is small enough not to disturb the microscopic scale $O(N^{-1})$.

Formally applying Itô's formula (see Appendix A for details) suggests that the evolution of the eigenvalues of $\begin{bmatrix} 0 & \widehat{M} \\ \widehat{M}^* & 0 \end{bmatrix}$ is governed by the following system of SDEs:

$$d\lambda_i = \frac{1}{\sqrt{2N}} dB_i + \frac{1}{2N} \sum_{j\ne i} \frac{1 - \gamma_{ij}}{\lambda_i - \lambda_j}\, dt + R_i, \qquad (3.13)$$

where

$$R_i = \operatorname{Re} \left\langle j_i, \left(U^*(t)\widehat{A}V(t) - U^*(0)\widehat{A}V(0) \right) k_i \right\rangle dt$$

$$+ \frac{1}{\sqrt{N}} \operatorname{Re} \left\langle j_i, \left(U^*(t) \left(\mathbb{1}_{(i,j)\in\mathcal{I}_\mathfrak{a}} d\widetilde{B}_{ij} \right) V(t) \right) k_i \right\rangle, \quad (3.14)$$

$$\gamma_{ij} = \frac{1}{2} \sum_{(a,b)\in\mathcal{I}_\mathfrak{a}^c} |w_i(a)|^2 |z_j(b)|^2 + \sum_{(a,b)\in\mathcal{I}_\mathfrak{a}^c} |w_j(a)|^2 |z_i(b)|^2, \quad w_i = U j_i, \quad z_i = V k_i, \quad (3.15)$$

and for $i < 0$ we set $R_i = -R_{-i}$ and $\gamma_{ij} = -\gamma_{-i,j}$. Here j_i and k_i are the columns of the matrices J and K in the singular value decomposition $\widehat{M} = JSK^*$ with S diagonal.

In Section 6, we justify this formal calculation, proving that the SDE (3.13) is well-posed and its solution is the eigenvalue process for $\widehat{M}(t)$.

4 Local law

In this section we prove a local law that is used in the next section to obtain global control on the quantities $w_\alpha(i)$, $z_\alpha(i)$, and $\gamma_{\alpha\beta}$. We fix constants $\mathfrak{a}, \mathfrak{b}$ such that

$$0 < \mathfrak{b} < \frac{\mathfrak{a}}{100} < 10^{-4}, \tag{4.1}$$

and let $\tau = N^{-1+\mathfrak{b}}$ denote the short time we study. Any constant C without further specification is a universal constant that may depend on $\mathfrak{a}$ but not on N. It may change from line to line, but only finitely many times, so that it remains finite. The norm $\|\cdot\|$ on matrices denotes the operator norm as an operator $\ell^2 \to \ell^2$.

We now write U for $U(t)$ and V for $V(t)$, and define

$$\mathcal{H}(t) = \begin{bmatrix} 0 & Z \\ Z^\dagger & 0 \end{bmatrix} \tag{4.2}$$

for

$$Z = UR^\dagger XTV^\dagger + Y + (\tau - t)UU(0)^\dagger \widehat{A}V(0)V^\dagger \tag{4.3}$$

and $\mathcal{G} = (\mathcal{H} - z)^{-1}$. Note that the $2N$ eigenvalues of $\mathcal{H}(t)$ are exactly the N singular values of $\widehat{M}(t)$ for $t \in [0, \tau]$, where each singular value appears with positive and negative sign. (The off-diagonal blocks of $\mathcal{H}$ come from the matrix $\widehat{M}(t)$ multiplied by U on the left and $V^\dagger$ on the right.) We also define

$$\widehat{\mathcal{H}}(t) = \begin{bmatrix} 0 & UR^\dagger XTV^\dagger + Y + (\tau - t)\widehat{A} \\ \left(UR^\dagger XTV^\dagger + Y + (\tau - t)\widehat{A}\right)^\dagger & 0 \end{bmatrix}, \tag{4.4}$$

and $\widehat{\mathcal{G}}(t) = (\widehat{\mathcal{H}}(t) - z)^{-1}$. We will begin by studying $\widehat{\mathcal{H}}(t)$, since the lack of randomness in the term involving $\widehat{A}$ makes it more tractable. We then relate it to $\mathcal{H}(t)$ in the proof of Corollary 4.13.

4.1 Concentration of Green's functions

The main probabilistic tool in this section is the following concentration result about the Haar measure on the unitary group $U(N)$. We use the following notation for the Hilbert–Schmidt norm of matrices:

$$\|A\|_{\mathrm{HS}} = \sqrt{\mathrm{Tr}(AA^*)}. \tag{4.5}$$

We also recall the equivalent characterization of this norm in terms of the sum of the squares of the matrix entries:

$$\|A\|_{\mathrm{HS}}^2 = \sum_{i,j} |A_{ij}|^2. \tag{4.6}$$

The next proposition follows from a theorem by Gromov; see [11, Corollary 4.4.28].

Proposition 4.1. Let $g\colon U(N) \to \mathbb{C}$ be a Lipschitz function with Lipschitz constant L in the Hilbert–Schmidt norm:

$$|g(X) - g(Y)| \le L\|X - Y\|_{\mathrm{HS}}, \quad \forall X, Y \in U(N). \tag{4.7}$$

Let $\mathbb{P}$ be the normalized Haar measure on the unitary group $U(N)$ and $\mathbb{E}$ be the corresponding expectation. Then there is a constant $c > 0$ not depending on N such that

$$\mathbb{P}[|g - \mathbb{E}g| \geq \delta L N^{-\frac{1}{2}}] \leq \exp\left(-c\delta^2\right), \quad \forall \delta \geq 0. \tag{4.8}$$

The above proposition can be applied to the Green's function $\mathcal{G}$, which is a smooth function of U, V, and z. In particular, the Lipschitz constant of $\mathcal{G}$ with respect to the variable U or V can be bounded using the imaginary part of z, as illustrated by the following propositions.

Proposition 4.2. For any $z = E + i\eta \in \mathbb{C}_+$ and $1 \leq i \leq 2N$,

$$\operatorname{Im} \mathcal{G}_{ii} \geq \frac{\eta}{(C + |z|)^2}, \quad \|\mathcal{G}\| \leq \eta^{-1}. \tag{4.9}$$

The same bounds hold for $\widehat{\mathcal{G}}$.

Proof. Let $e_i \in \mathbb{C}^N$ be the unit vector with 1 on the i-th coordinate and 0 otherwise. Then we have

$$\operatorname{Im} \mathcal{G}_{ii} = \operatorname{Im}\langle e_i, \mathcal{G}e_i \rangle = \frac{1}{2i}\langle e_i, (\mathcal{G} - \mathcal{G}^\dagger)e_i \rangle. \tag{4.10}$$

Note that $\mathcal{G} - \mathcal{G}^\dagger = \mathcal{G}^*(z - \bar{z})\mathcal{G} = 2i\eta\mathcal{G}^*\mathcal{G}$. Therefore,

$$\operatorname{Im} \mathcal{G}_{ii} = \langle \mathcal{G}e_i, \mathcal{G}e_i \rangle\eta = \|\mathcal{G}e_i\|^2\eta. \tag{4.11}$$

By the definition of $\mathcal{G}$ we have $\|e_i\| \leq \|\mathcal{H} - z\|\|\mathcal{G}e_i\|$. Note that our assumptions and the bound on $\widehat{A}$ given in (C.10) imply that $\|\mathcal{H} - z\| \leq C + |z|$. Then

$$\operatorname{Im} \mathcal{G}_{ii} = \|\mathcal{G}e_i\|^2\eta \geq \frac{\eta}{(C + |z|)^2}. \tag{4.12}$$

This proves the first inequality. The second inequality follows from the spectral theorem. $\square$

Proposition 4.3. Fix $z = E + i\eta \in \mathbb{C}^+$. Then, for any $1 \leq i, j \leq 2N$, $\mathcal{G}_{ij}$ is a Lipschitz function of U (or V) in the Hilbert–Schmidt norm with Lipschitz constant bounded by $C\eta^{-2}$ for some $C > 0$ independent of N. The same holds for $\widehat{\mathcal{G}}_{ij}$.

Proof. Let $\widetilde{\mathcal{G}}$ be the Green's function $\mathcal{G}$ after replacing U with $\widetilde{U} \in U(N)$. Then the resolvent identity yields

$$\widetilde{\mathcal{G}} - \mathcal{G} = \mathcal{G}\begin{bmatrix} 0 & (U - \widetilde{U})R^\dagger XTV^\dagger \\ \left((U - \widetilde{U})R^\dagger XTV^\dagger\right)^\dagger & 0 \end{bmatrix}\widetilde{\mathcal{G}}. \tag{4.13}$$

Therefore, using the general inequalities $\|AB\|_{\mathrm{HS}} \leq \|A\|_{\mathrm{HS}}\|B\|$ and $\|AB\|_{\mathrm{HS}} \leq \|A\|\|B\|_{\mathrm{HS}}$,

$$\|\widetilde{\mathcal{G}} - \mathcal{G}\|_{\mathrm{HS}} \leq 2\|\mathcal{G}\|\|\widetilde{\mathcal{G}}\|\|X\|\|U - \widetilde{U}\|_{\mathrm{HS}}. \tag{4.14}$$

By Proposition 4.2, $\|\mathcal{G}\| \leq \eta^{-1}$, and similarly the spectral theorem yields $\|\widetilde{\mathcal{G}}\| \leq \eta^{-1}$. By assumption (2.2) we have $\|X\| \leq C$. Hence (4.14) implies that

$$\|\mathcal{G} - \widetilde{\mathcal{G}}\|_{\mathrm{HS}} \leq \frac{2C}{\eta^2}\|U - \widetilde{U}\|_{\mathrm{HS}}, \tag{4.15}$$

and the conclusion follows. $\square$

4.2 Invariant identities

Let E_{ij} be the matrix whose (i, j)-th entry is 1 and all the other entries are 0:

$$(E_{ij})_{kl} = \delta_{ik}\delta_{jl}. \tag{4.16}$$

The matrix E_{ij} will be either N by N or $2N$ by $2N$, depending on the context.

For brevity we set $\overline{Y} = Y + (\tau - t)\widehat{A}$, and we define

$$\mathcal{H}_1 = \begin{bmatrix} 0 & UR^*XTV^\dagger \\ (UR^*XTV^\dagger)^* & 0 \end{bmatrix}, \quad \mathcal{H}_2 = \begin{bmatrix} 0 & \overline{Y} \\ \overline{Y}^* & 0 \end{bmatrix}, \tag{4.17}$$

so that $\widehat{\mathcal{H}} = \mathcal{H}_1 + \mathcal{H}_2$. We require the following lemma.

Lemma 4.4. For $1 \leq i, j \leq N$ or $N + 1 \leq i, j \leq 2N$, we have

$$\mathbb{E}\left[\widehat{\mathcal{G}}\mathcal{H}_1 E_{ij}\widehat{\mathcal{G}}\right] = \mathbb{E}\left[\widehat{\mathcal{G}}E_{ij}\mathcal{H}_1\widehat{\mathcal{G}}\right]. \tag{4.18}$$

Proof. For any $\zeta \in \mathbb{C}$, define an $N \times N$ unitary matrix $Q(\zeta)$ by

$$Q(\zeta) = \exp(\zeta E_{ij} - \bar{\zeta}E_{ji}). \tag{4.19}$$

The derivatives of $Q(\zeta)$ and $Q^*(\zeta)$ with respect to ζ at $\zeta = 0$ are

$$\partial_\zeta Q(\zeta)|_{\zeta=0} = E_{ij}, \quad \partial_\zeta Q^*(\zeta)|_{\zeta=0} = -E_{ij}. \tag{4.20}$$

Let $M(\zeta) = Q(\zeta)UR^*XTV^* + \overline{Y}$, let $\mathcal{H}(\zeta)$ be its symmetrization, and let $G(\zeta, z)$ be the Green's function of $\mathcal{H}(\zeta)$. By definition, $M(0) = M$, $\mathcal{H}(0) = \widehat{\mathcal{H}}$, and $\mathcal{G}(0, z) = \widehat{\mathcal{G}}(z)$. We differentiate $G(\zeta, z)$ and evaluate the derivative at $\zeta = 0$ to obtain

$$\partial_\zeta G(\zeta, z)|_{\zeta=0} = \widehat{\mathcal{G}}\begin{bmatrix} 0 & E_{ij}UR^\dagger XTV^\dagger \\ -(UR^\dagger XTV^\dagger)^\dagger E_{ij} & 0 \end{bmatrix}\widehat{\mathcal{G}}. \tag{4.21}$$

Note that the distribution of $Q(\zeta)U$ is invariant with respect to ζ, so $\mathbb{E}[\partial_\zeta G(\zeta, z)] = 0$. Therefore, the above equality yields

$$\mathbb{E}\left[\widehat{\mathcal{G}}\begin{bmatrix} 0 & E_{ij}UR^\dagger XTV^\dagger \\ -(UR^\dagger XTV^\dagger)^\dagger E_{ij} & 0 \end{bmatrix}\widehat{\mathcal{G}}\right] = 0. \tag{4.22}$$

This can also be rearranged as

$$\mathbb{E}\left[\widehat{\mathcal{G}}E_{i,j}\mathcal{H}_1\widehat{\mathcal{G}}\right] = \mathbb{E}\left[\widehat{\mathcal{G}}\mathcal{H}_1 E_{i,j}\widehat{\mathcal{G}}\right]. \tag{4.23}$$

This proves the case where $1 \leq i, j \leq N$. The other case follows from a similar argument after multiplying $Q(\zeta)$ on the right. $\square$

4.3 Asymptotic equations

Let $m_N(z)$ be the Sieltjes transform of $\widehat{\mathcal{H}}$. In this subsection we provide a system of equations that $m_N(z)$ satisfies asymptotically. We require the following high probability notation.

Definition 4.5. Given two sequences of random variables (X_N) and (Y_N), we write $X = \mathcal{O}(Y)$ if there are $c_1, c_2, c_3, N_0 > 0$ that do not depend on N such that for $N > N_0$,

$$\mathbb{P}[|X_N| \geq c_1 Y_N] \leq \exp(-c_2 N^{c_3}). \tag{4.24}$$

In general, for some index set $\mathcal{A}$ (possibly N-dependent) and families of random variables $(X(\alpha, N))$ and $(Y(\alpha, N))$ with parameters $\alpha \in \mathcal{A}$ and $N \in \mathbb{N}$, we say that $X = \mathcal{O}(Y)$ uniformly in α, if there are constants $c_1, c_2, c_3, N_0 > 0$ that do not depend on N such that for $N > N_0$,

$$\mathbb{P}[\exists \alpha \in \mathcal{A} : |X(\alpha, N)| \geq c_1 Y(\alpha, N)] \leq \exp(-c_2 N^{c_3}). \tag{4.25}$$

Define

$$w_X = -\frac{\mathrm{Tr}(\widehat{\mathcal{G}}\mathcal{H}_1)}{\mathrm{Tr}\,\widehat{\mathcal{G}}}, \quad w_{\overline{Y}} = -\frac{\mathrm{Tr}(\widehat{\mathcal{G}}\mathcal{H}_2)}{\mathrm{Tr}\,\widehat{\mathcal{G}}}, \tag{4.26}$$

which are holomorphic functions on $\mathbb{C}^+$.

Lemma 4.6. Fix $a > 0$. For any $z = E + i\eta \in \mathbb{C}^+$ with $|z| \le \log N$ and $\eta > N^{-1/10+a}$, we have

$$(\widehat{\mathcal{G}}\mathcal{H}_1)_{ki} = \widehat{\mathcal{G}}_{ki}\left(-w_X + \mathcal{O}\left(\eta^{-5}N^{-\frac{1}{2}+a}\right)\right) + \mathcal{O}\left(\eta^{-4}N^{-\frac{1}{2}+a}\right), \tag{4.27}$$

uniformly for $1 \le i, k \le 2N$.

Proof. By Lemma 4.4 combined with Proposition 4.1 and Proposition 4.3, for $1 \le i, j \le N$ or $N + 1 \le i, j \le 2N$, and any $a > 0$, $0 \le k, l \le 2N$, we have, using the general inequality $\|AB\|_{\mathrm{HS}} \le \|A\|\|B\|_{\mathrm{HS}}$ and the symmetry of the matrices in question,

$$(\widehat{\mathcal{G}}\mathcal{H}_1)_{ki}\widehat{\mathcal{G}}_{jl} = \widehat{\mathcal{G}}_{ki}(\mathcal{H}_1\widehat{\mathcal{G}})_{jl} + \mathcal{O}\left(\eta^{-3}N^{-\frac{1}{2}+a}\right). \tag{4.28}$$

We used here our assumptions on X to conclude that $\|\mathcal{H}_1\| \le C$.

Take $i, j \in [1, N]$ and let $l = j$, then sum over j:

$$(\widehat{\mathcal{G}}\mathcal{H}_1)_{ki}\frac{1}{N}\sum_{j=1}^{N}\widehat{\mathcal{G}}_{jj} = \widehat{\mathcal{G}}_{ki}\frac{1}{N}\sum_{j=1}^{N}(\mathcal{H}_1\widehat{\mathcal{G}})_{jj} + \mathcal{O}\left(\eta^{-3}N^{-\frac{1}{2}+a}\right), \forall 1 \le i \le N, 1 \le k \le 2N. \tag{4.29}$$

Note that $UR^*XTV^\dagger$ has the same distribution as $(UR^*XTV^\dagger)^*$, so $\begin{bmatrix} 0 & I \\ I & 0 \end{bmatrix} \widehat{\mathcal{G}} \begin{bmatrix} 0 & I \\ I & 0 \end{bmatrix}$ has the same probability distribution as $\widehat{\mathcal{G}}$. It follows that $\widehat{\mathcal{G}}_{ii}$ has the same distribution as $\widehat{\mathcal{G}}_{n+i,n+i}$ for $1 \le i \le N$. Therefore,

$$\mathbb{E}\left[\frac{1}{N}\sum_{j=1}^{N}\widehat{\mathcal{G}}_{jj}\right] = \mathbb{E}\left[\frac{1}{2N}\mathrm{Tr}\,\widehat{\mathcal{G}}\right]. \tag{4.30}$$

Then Proposition 4.1 and Proposition 4.3 imply that the above identity holds without expectation up to a small error term:

$$\frac{1}{N}\sum_{j=1}^{N}\widehat{\mathcal{G}}_{jj} = \frac{1}{2N}\mathrm{Tr}\,\widehat{\mathcal{G}} + \mathcal{O}\left(\eta^{-2}N^{-\frac{1}{2}+a}\right). \tag{4.31}$$

Similarly, we have

$$\frac{1}{N}\sum_{j=1}^{N}(\mathcal{H}_1\widehat{\mathcal{G}})_{jj} = \frac{1}{2N}\mathrm{Tr}(\mathcal{H}_1\widehat{\mathcal{G}}) + \mathcal{O}\left(\eta^{-2}N^{-\frac{1}{2}+a}\right). \tag{4.32}$$

Recall the definition (4.26). Take the quotient of the above two equations and use Proposition 4.2 to get

$$\frac{\sum_{j=1}^{N}(\mathcal{H}_1\widehat{\mathcal{G}})_{jj}}{\sum_{j=1}^{N}\widehat{\mathcal{G}}_{jj}} = -w_X + \mathcal{O}\left(\eta^{-5}N^{-\frac{1}{2}+a}\right). \tag{4.33}$$

Here we used the assumed lower bound $\eta > N^{-1/10+a}$ to ensure the error in the denominator is small. We also used Hölder's inequality for Schatten norms and the second bound in (4.2) (which bounds the largest – in absolute value – eigenvalue of $\widehat{\mathcal{G}}$) to conclude that

$$\left|\mathrm{Tr}(\mathcal{H}_1\widehat{\mathcal{G}})\right| \le \|\mathcal{H}_1\|\|\widehat{\mathcal{G}}\|_1 \le CN\eta^{-1}. \tag{4.34}$$

We used that the Schatten p-norm with $p = \infty$ is the operator norm.

Now we go back to (4.29), plugging in the above equation to see

$$(\widehat{\mathcal{G}}\mathcal{H}_1)_{ki} = \widehat{\mathcal{G}}_{ki}\left(-w_X + \mathcal{O}\left(\eta^{-5}N^{-\frac{1}{2}+a}\right)\right) + \mathcal{O}\left(\eta^{-4}N^{-\frac{1}{2}+a}\right), \quad \forall 1 \le i \le N, 1 \le k \le 2N, \tag{4.35}$$

where the extra power of η^{-1} in the second error term comes from Proposition 4.2 after taking the quotient. This proves the conclusion for $1 \le i \le N$. Similarly, we can take $i \in [N+1, 2N]$ to obtain the same identity for $N+1 \le i \le 2N, 1 \le k \le 2N$. This proves the conclusion for fixed i and k. The constants in the $\mathcal{O}$ notation are uniform in i and k. $\qquad\square$

Before proceeding to the next lemma, we prove the following technical proposition.

Proposition 4.7. Let A, B, R be square matrices of the same dimension such that $\|AR\| \le \delta < 1/2$ and

$$A(B + R) = I. \tag{4.36}$$

Then

$$A = B^{-1} + O(\delta \|A\|). \tag{4.37}$$

Here $O(\cdot)$ is in the sense of operator norm.

Proof. We immediately have $B = A^{-1}(I - AR)$. Hence

$$B^{-1} = (I - AR)^{-1}A = \left(\sum_{k=0}^{\infty}(AR)^k\right)A. \tag{4.38}$$

By the assumption that $\|AR\| < 1/2$, we have $\left\|\sum_{k=0}^{\infty}(AR)^k\right\| \le 2$. Hence,

$$\left\|B^{-1}\right\| \le 2\|A\|. \tag{4.39}$$

On the other hand, $A(B + R) = I$ implies

$$A = B^{-1} - ARB^{-1}. \tag{4.40}$$

Note that $\left\|ARB^{-1}\right\| \le \delta\left\|B^{-1}\right\| \le 2\delta\|A\|$. This gives

$$A = B^{-1} + O(\delta\|A\|). \tag{4.41}$$

$\qquad\square$

Now we are ready to prove that the three holomorphic functions m_N, w_X, and $w_{\overline{Y}}$ approximately satisfy a system of equations for $z \in \mathbb{C}^+$. We recall that m_X was defined in (2.6) as the Stieltjes transform of μ_X^{sym}, and we let $m_{\overline{Y}}$ denote the Stieltjes transform of $\mu_{\overline{Y}}^{\text{sym}}$.

Lemma 4.8. Fix $a > 0$. For any $z \in \mathbb{C}^+$ with $|z| \le \log N$ and $\eta > N^{-1/10+a}$, we have

$$\begin{cases} m_N(z) = m_X(z + w_{\overline{Y}}) + \mathcal{O}\left(\eta^{-6}N^{-\frac{1}{2}+a}\right), \\[2mm] m_N(z) = m_{\overline{Y}}(z + w_X) + \mathcal{O}\left(\eta^{-6}N^{-\frac{1}{2}+a}\right), \\[2mm] \dfrac{1}{m_N(z)} = -z - w_X - w_{\overline{Y}}, \end{cases} \tag{4.42}$$

and for $1 \le k \le N$,

$$\begin{aligned} \widehat{\mathcal{G}}_{kk} &= \frac{1}{2}\left(\frac{1}{\overline{Y}_k - z - w_X} + \frac{1}{-\overline{Y}_k - z - w_X}\right) + \mathcal{O}\left(\eta^{-4}N^{-\frac{1}{2}+a}\right), \\ \widehat{\mathcal{G}}_{N+k,N+k} &= \frac{1}{2}\left(\frac{1}{\overline{Y}_k - z - w_X} + \frac{1}{-\overline{Y}_k - z - w_X}\right) + \mathcal{O}\left(\eta^{-4}N^{-\frac{1}{2}+a}\right). \end{aligned} \tag{4.43}$$

Proof. We start with the following identity, which is equivalent to the definition of $\widehat{\mathcal{G}}$.

$$\widehat{\mathcal{G}} \begin{bmatrix} 0 & \overline{Y} \\ \overline{Y} & 0 \end{bmatrix} + \widehat{\mathcal{G}}\mathcal{H}_1 - z\widehat{\mathcal{G}} = I \tag{4.44}$$

Taking the (k,k)-th entry of each of the four blocks, we have

$$\overline{Y}_k \begin{bmatrix} \widehat{\mathcal{G}}_{k,N+k} & \mathcal{G}_{kk} \\ \widehat{\mathcal{G}}_{N+k,N+k} & \widehat{\mathcal{G}}_{N+k,k} \end{bmatrix} + \begin{bmatrix} (\widehat{\mathcal{G}}\mathcal{H}_1)_{kk} & (\widehat{\mathcal{G}}\mathcal{H}_1)_{k,N+k} \\ (\widehat{\mathcal{G}}\mathcal{H}_1)_{N+k,k} & (\widehat{\mathcal{G}}\mathcal{H}_1)_{N+k,N+k} \end{bmatrix}$$
$$- z \begin{bmatrix} \widehat{\mathcal{G}}_{kk} & \widehat{\mathcal{G}}_{k,N+k} \\ \widehat{\mathcal{G}}_{N+k,k} & \widehat{\mathcal{G}}_{N+k,N+k} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad (4.45)$$

or equivalently,

$$\begin{bmatrix} \widehat{\mathcal{G}}_{kk} & \widehat{\mathcal{G}}_{k,N+k} \\ \widehat{\mathcal{G}}_{N+k,k} & \widehat{\mathcal{G}}_{N+k,N+k} \end{bmatrix} \begin{bmatrix} -z & \overline{Y}_k \\ \overline{Y}_k & -z \end{bmatrix} + \begin{bmatrix} (\widehat{\mathcal{G}}\mathcal{H}_1)_{kk} & (\widehat{\mathcal{G}}\mathcal{H}_1)_{k,N+k} \\ (\widehat{\mathcal{G}}\mathcal{H}_1)_{N+k,k} & (\widehat{\mathcal{G}}\mathcal{H}_1)_{N+k,N+k} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}. \quad (4.46)$$

We apply Corollary 4.6 to the matrix involving $\widehat{\mathcal{G}}\mathcal{H}_1$ to get

$$\begin{bmatrix} \widehat{\mathcal{G}}_{kk} & \widehat{\mathcal{G}}_{k,N+k} \\ \widehat{\mathcal{G}}_{N+k,k} & \widehat{\mathcal{G}}_{N+k,N+k} \end{bmatrix} \left(\begin{bmatrix} -z - w_X & \overline{Y}_k \\ \overline{Y}_k & -z - w_X \end{bmatrix} + \mathcal{O}\left(\eta^{-5}N^{-\frac{1}{2}+a}\right) \right)$$
$$+ \mathcal{O}\left(\eta^{-4}N^{-\frac{1}{2}+a}\right) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad (4.47)$$

where the $\mathcal{O}$ notation is used entrywise. Using Proposition 4.7 and the fact that $|\widehat{\mathcal{G}}_{ij}| \leq \eta^{-1}$ for all $1 \leq i, j \leq 2N$, we have

$$\begin{bmatrix} \widehat{\mathcal{G}}_{kk} & \widehat{\mathcal{G}}_{k,N+k} \\ \widehat{\mathcal{G}}_{N+k,k} & \widehat{\mathcal{G}}_{N+k,N+k} \end{bmatrix} = \begin{bmatrix} -z - w_X & \overline{Y}_k \\ \overline{Y}_k & -z - w_X \end{bmatrix}^{-1} \left(I + \mathcal{O}\left(\eta^{-6}N^{-\frac{1}{2}+a}\right) \right). \tag{4.48}$$

This can be written explicitly as

$$\begin{bmatrix} \widehat{\mathcal{G}}_{kk} & \widehat{\mathcal{G}}_{k,N+k} \\ \widehat{\mathcal{G}}_{N+k,k} & \mathcal{G}_{N+k,N+k} \end{bmatrix}$$
$$= \begin{bmatrix} \frac{1}{2}\left(\frac{1}{\overline{Y}_k - z - w_X} + \frac{1}{-\overline{Y}_k - z - w_X} \right) & * \\ * & \frac{1}{2}\left(\frac{1}{\overline{Y}_k - z - w_X} + \frac{1}{-\overline{Y}_k - z - w_X} \right) \end{bmatrix} + \mathcal{O}\left(\eta^{-6}N^{-\frac{1}{2}+a}\right), \tag{4.49}$$

where we omit the off-diagonal terms and write them as $*$. We sum over the diagonal terms to get

$$\frac{1}{2N} \operatorname{Tr} \widehat{\mathcal{G}} = m_{\overline{Y}}(z + w_X) + \mathcal{O}\left(\eta^{-4}N^{-\frac{1}{2}+a}\right). \tag{4.50}$$

This proves the second equation. To prove the second equation, one replaces $\overline{Y}$ with $RU\overline{Y}V^\dagger T^*$ and $UR^*XTV^\dagger$ with X in the definition of $\mathcal{H}$, then repeats the entire argument word for word. Note that this replacement does not change the definition of $w_{\overline{Y}}$, w_X, and m_N, since the trace of a matrix does not change under unitary conjugation.

Finally, the third equation follows from

$$\operatorname{Tr}(\widehat{\mathcal{G}}\mathcal{H}_1) + \operatorname{Tr}(\widehat{\mathcal{G}}\mathcal{H}_2) - z \operatorname{Tr}\widehat{\mathcal{G}} = 2N, \tag{4.51}$$

which follows from the definition of $\widehat{\mathcal{G}}$. $\qquad \square$

4.4 Weak law

In this subsection we prove a weak law for the $\mathcal{G}_{ii}$. We use the term *weak* because the result is only valid in the regime $\operatorname{Im} z \geq N^{-c}$ for some small constant c. Thus, it is only slightly stronger than the weak convergence of the corresponding measure μ_N. Nevertheless, the weak law provides necessary bounds for the eigenvectors of $\mathcal{H}$.

We deal with equation (4.42) in a general setting. Let m_α, m_β be the Stieltjes transforms of probability measures μ_α, μ_β. Consider the following deterministic equations for fixed $z \in \mathbb{C}^+$.

$$\begin{cases} m = m_\alpha(z + w_\alpha) \\ m = m_\beta(z + w_\beta) \\ \dfrac{1}{m} = -z - w_\alpha - w_\beta \end{cases} \tag{4.52}$$

Observe that equation (4.42) is a special case of the above equations plus some error terms. The existence and uniqueness of a solution to this system is known. We call the measure $\mu_\alpha \boxplus \mu_\beta$ given by the following proposition the *free convolution* of μ_α and μ_β.

Proposition 4.9. Given two probability measures μ_α and μ_β on $\mathbb{R}$, there exists unique analytic functions $w_\alpha, w_\beta, m \colon \mathbb{C}^+ \to \mathbb{C}^+$ satisfying (4.52), where m is the Stieltjes transform of a probability measure we denote $\mu_\alpha \boxplus \mu_\beta$. Further, suppose $\mu_\alpha \boxplus \mu_\beta$ has a density on an interval $I \subset \mathbb{R}$ that is bounded away from zero. If μ_α, μ_β are compactly supported, and none of them is a point mass, then w_α, w_β extend continuously to I; in particular, $|w_\alpha| \vee |w_\beta|$ is uniformly bounded on compact subsets of $\mathbb{C}^+ \cup I$. If in addition, $\mu_\alpha(\{a\}) + \mu_\beta(\{b\}) < 1$ for all $a, b \in \mathbb{R}$, then $\mu_\alpha \boxplus \mu_\beta$ has a continuous density on $\mathbb{R}$.

Proof. For existence and uniqueness see [24, Theorem 4.1]. For the continuous extension see [21, Remark 2.4] or [22, Theorem 3.3]. For the continuous density claim see [23, Corollary 8].

We remark that the referenced works permit the continuous boundary extensions of w_α and w_β to take the value ∞, but our hypothesis on the density of $\mu_\alpha \boxplus \mu_\beta$ on I rules this out. Indeed, [21, Theorem 2.3] shows that m also extends continuously to I (again with values that may be ∞), and this hypothesis shows that $\operatorname{Im} m(a) > c$ uniformly for $a \in I$ and some constant $c > 0$. The uniform boundedness claim then follows from the last equation of (4.52). $\qquad\square$

We need the stability of the solution to (4.52) under perturbation. To investigate this, it is convenient to write the equation in a more symmetric form. In fact, the above equation can be rephrased in terms of w_α and w_β only. Define

$$\hat{m}_\alpha(\zeta) = -\zeta - \frac{1}{m_\alpha(\zeta)}, \quad \hat{m}_\beta(\zeta) = -\zeta - \frac{1}{m_\beta(\zeta)}. \tag{4.53}$$

Proposition 2.2 in [64] says that $\hat{m}_\alpha$ and $\hat{m}_\beta$ are Stieltjes transforms of Borel measures $\hat{\mu}_\alpha$ and $\hat{\mu}_\beta$ on $\mathbb{R}$, whose total masses are $\sigma_\alpha^2 := \int t^2 \mu_\alpha(\mathrm{d}t)$ and $\sigma_\beta^2 := \int t^2 \mu_\beta(\mathrm{d}t)$ respectively (which are $\leq C$ by assumption (2.2)).

After eliminating m, equation (4.52) becomes

$$\begin{cases} w_\alpha = \hat{m}_\beta(z + w_\beta) \\ w_\beta = \hat{m}_\alpha(z + w_\alpha). \end{cases} \tag{4.54}$$

Define a function $\Phi \colon (\mathbb{C}^+)^2 \to (\mathbb{C}^+)^2$ by

$$\Phi(\zeta_1, \zeta_2) = (\zeta_1 - \hat{m}_\beta(z + \zeta_2), \zeta_2 - \hat{m}_\alpha(z + \zeta_1)). \tag{4.55}$$

Note Φ depends on a choice of $z \in \mathbb{C}^+$, but we omit this in the notation. The equation (4.54) is equivalent to

$$\Phi(w_\alpha, w_\beta) = (0, 0). \tag{4.56}$$

To show stability, we want to show that the solution to the following perturbed version is close to the solution of (4.56), when the perturbation (r_1, r_2) is small enough:

$$\Phi(w'_\alpha, w'_\beta) = (r_1, r_2). \tag{4.57}$$

It suffices to prove that the matrix of first derivatives $D\Phi$ is non-degenerate at (w_α, w_β), i.e., $(D\Phi)^{-1}$ is bounded.

Proposition 4.10. For all $z \in \mathbb{C}^+$,

$$\left\| (D\Phi(w_\alpha, w_\beta))^{-1} \right\| \leq C(1 \vee \eta^{-4}). \tag{4.58}$$

Proof. Note that

$$D\Phi(\zeta_1, \zeta_2) = \begin{bmatrix} 1 & -\int_\mathbb{R} \frac{\hat\mu_\beta(\mathrm{d}x)}{(x-z-\zeta_2)^2} \\ -\int_\mathbb{R} \frac{\hat\mu_\alpha(\mathrm{d}x)}{(x-z-\zeta_1)^2} & 1 \end{bmatrix}. \tag{4.59}$$

To simplify notation, we define

$$p = \int_\mathbb{R} \frac{\hat\mu_\beta(\mathrm{d}x)}{(x-z-w_\beta)^2}, \quad q = \int_\mathbb{R} \frac{\hat\mu_\alpha(\mathrm{d}x)}{(x-z-w_\alpha)^2}, \tag{4.60}$$

and

$$\tilde p = \int_\mathbb{R} \frac{\hat\mu_\beta(\mathrm{d}x)}{|x-z-w_\beta|^2}, \quad \tilde q = \int_\mathbb{R} \frac{\hat\mu_\alpha(\mathrm{d}x)}{|x-z-w_\alpha|^2}. \tag{4.61}$$

Note $|p| \leq \tilde p \leq \eta^{-2}, |q| \leq \tilde q \leq \eta^{-2}$ and

$$D\Phi(w_\alpha, w_\beta) = \begin{bmatrix} 1 & -p \\ -q & 1 \end{bmatrix}. \tag{4.62}$$

Therefore we can take the inverse of $D\Phi$ explicitly:

$$(D\Phi(w_\alpha, w_\beta))^{-1} = \frac{1}{1-pq} \begin{bmatrix} 1 & p \\ q & 1 \end{bmatrix}. \tag{4.63}$$

Hence by the elementary bound $\|A\| \leq \|A\|_{\mathrm{HS}}$ and the triangle inequality,

$$\left\| (D\Phi(w_\alpha, w_\beta))^{-1} \right\| \leq \frac{\sqrt{2+\tilde p^2+\tilde q^2}}{1-\tilde p\tilde q} \leq \sqrt{2}\frac{1+\eta^{-2}}{1-\tilde p\tilde q}. \tag{4.64}$$

Taking the imaginary part of equation (4.54), we have

$$\begin{cases} \operatorname{Im} w_\alpha = (\eta + \operatorname{Im} w_\beta)\tilde p \\ \operatorname{Im} w_\beta = (\eta + \operatorname{Im} w_\alpha)\tilde q. \end{cases} \tag{4.65}$$

Therefore,

$$\tilde p\tilde q = \frac{(\operatorname{Im} w_\alpha)(\operatorname{Im} w_\beta)}{(\eta + \operatorname{Im} w_\alpha)(\eta + \operatorname{Im} w_\beta)} < 1. \tag{4.66}$$

A simple calculation yields $\frac{1}{1-\tilde p\tilde q} \leq (1 + \frac{\operatorname{Im} w_\alpha \wedge \operatorname{Im} w_\beta}{\eta})$. Therefore,

$$\left\| (D\Phi(w_\alpha, w_\beta))^{-1} \right\| \leq \sqrt{2} \left(1 + \frac{\operatorname{Im} w_\alpha \wedge \operatorname{Im} w_\beta}{\eta} \right) (1 + \eta^{-2}). \tag{4.67}$$

By equation (4.54) and the fact that $\hat\mu_\alpha$ and $\hat\mu_\beta$ have total masses $\leq C$, we have $\operatorname{Im} w_\alpha \wedge \operatorname{Im} w_\beta \leq C\eta^{-1}$. Therefore

$$\left\| (D\Phi(w_\alpha, w_\beta))^{-1} \right\| \leq C(1 + \eta^{-2})^2 \leq C(1 \vee \eta^{-4}). \tag{4.68}$$

$\square$

Proposition 4.11. Fix $z = E + i\eta \in \mathbb{C}^+$. Let $w_\alpha, w_\beta \in \mathbb{C}^+$ solve (4.54). Let $w_1, w_2, r \in \mathbb{C}^+$ be such that

$$\Phi(w_1, w_2) = (r_1, r_2) = r. \tag{4.69}$$

Define $\delta w = (w_1 - w_\alpha, w_2 - w_\beta)$. There exists $c > 0$ such that if $\|\delta w\|_2 \leq c(\eta^3 \wedge \eta^7)$, then

$$\|\delta w\|_2 \leq C(1 \vee \eta^{-4})\|r\|_2. \tag{4.70}$$

Proof. Using (4.59), it is straightforward to see that for fixed $z = E + i\eta$,

$$\|D\Phi(\zeta_1, \zeta_2)\|_\infty \leq C(1 \vee \eta^{-2}), \quad \|D^2\Phi(\zeta_1, \zeta_2)\|_\infty \leq C\eta^{-3}, \tag{4.71}$$

for any $(\zeta_1, \zeta_2) \in (\mathbb{C}^+)^2$. By Taylor expansion at (w_α, w_β), we have

$$\|r - D\Phi(w_\alpha, w_\beta)\delta w\|_2 \leq \sup_{(\zeta_1', \zeta_2') \in (\mathbb{C}^+)^2} 4\|D^2\Phi(\zeta_1', \zeta_2')\|_\infty \|\delta w\|_\infty^2 \leq \frac{C}{\eta^3}\|\delta w\|_2^2. \tag{4.72}$$

Using (4.72) and the bound (4.58), we have

$$\|(D\Phi(w_\alpha, w_\beta))^{-1}r - \delta w\| \leq \|(D\Phi(w_\alpha, w_\beta))^{-1}\|\|r - D\Phi(w_\alpha, w_\beta)\delta w\|_2 \leq C\eta^{-3}(1 \vee \eta^{-4})\|\delta w\|_2^2. \tag{4.73}$$

Hence,

$$\begin{aligned}
\|\delta w\|_2 &\leq \|(D\Phi(w_\alpha, w_\beta))^{-1}r\|_2 + \|(D\Phi(w_\alpha, w_\beta))^{-1}r - \delta w\|_2 \\
&\leq C(1 \vee \eta^{-4})\|r\|_2 + C\eta^{-3}(1 \vee \eta^{-4})\|\delta w\|_2^2. \quad (4.74)
\end{aligned}$$

Using the condition that $\|\delta w\| \leq c(\eta^3 \wedge \eta^7)$ and choosing $c > 0$ small enough, the second term on the right side can be absorbed into the left side. Thus,

$$\|\delta w\|_2 \leq C(1 \vee \eta^{-4})\|r\|_2. \tag{4.75}$$
$$\square$$

Let $K > 0$ be a constant smaller than the constant c in the assumption (2.12). Define the spectral domain

$$\Sigma = \left\{ z = E + i\eta \colon \eta \in [N^{-1/100}, 1], E \in [-2K, 2K] \right\}, \tag{4.76}$$

and let $\tilde{w}_X$ and $\tilde{w}_{\overline{Y}}$ solve

$$\begin{cases} \tilde{w}_X = \hat{m}_X(z + \tilde{w}_{\overline{Y}}) \\ \tilde{w}_{\overline{Y}} = \hat{m}_{\overline{Y}}(z + \tilde{w}_X). \end{cases} \tag{4.77}$$

Corollary 4.12. For any $z \in \Sigma$ and $t \in [0, \tau]$,

$$|w_X - \tilde{w}_X| \vee |w_{\overline{Y}} - \tilde{w}_{\overline{Y}}| = \mathcal{O}\left(\eta^{-12}N^{-1/3}\right). \tag{4.78}$$

Proof. We restrict the following claims to $z \in \Sigma$. Multiplying the first and third equations of (4.42) gives

$$1 = \left(m_X(z + w_{\overline{Y}}) + \mathcal{O}\left(\eta^{-6}N^{-\frac{1}{2}+a}\right)\right)(-z - w_X - w_{\overline{Y}}). \tag{4.79}$$

By Proposition 4.2, $\operatorname{Im} m_N(z) \geq c\eta$, so

$$|-z - w_X - w_{\overline{Y}}| = \frac{1}{|m_N(z)|} \leq C\eta^{-1}, \tag{4.80}$$

$$1 = m_X(z + w_{\overline{Y}})(-z - w_X - w_{\overline{Y}}) + \mathcal{O}\left(\eta^{-7}N^{-\frac{1}{2}+a}\right). \tag{4.81}$$

Similarly, using the first equation of (4.42) (with a small enough, say $a = 1/6$) and Proposition 4.2 we have

$$\frac{1}{|m_X(z + w_{\overline{Y}})|} \leq C\eta^{-1}. \tag{4.82}$$

Dividing (4.81) by $m_X(z + w_{\overline{Y}})$ and rearranging using $\hat{m}_X(z + w_{\overline{Y}}) = -z - w_{\overline{Y}} - \frac{1}{m_X(z+w_{\overline{Y}})}$ yields

$$\hat{m}_X(z + w_{\overline{Y}}) = w_X + \mathcal{O}\left(\eta^{-8} N^{-\frac{1}{2}+a}\right). \tag{4.83}$$

Analogously,

$$\hat{m}_Y(z + w_X) = w_{\overline{Y}} + \mathcal{O}\left(\eta^{-8} N^{-\frac{1}{2}+a}\right). \tag{4.84}$$

The claim now follows from choosing $a = 1/6$ and applying Proposition 4.11. $\qquad \square$

Corollary 4.13. There exists a constant $c(\mathfrak{b}) > 0$ such that with probability at least $1 - e^{-cN^c}$,

$$\sup_{(z,t)\in\Sigma\times[0,\tau]} \left| \mathcal{G}_{ii} - \frac{1}{2}\left(\frac{1}{\overline{Y}_i - z - \tilde{w}_X} + \frac{1}{-\overline{Y}_i - z - \tilde{w}_X}\right)\right| \leq C\eta^{-14} N^{-1/3}, \tag{4.85}$$

where the i in $\overline{Y}_i$ is taken modulo N.

Proof. By Corollary 4.12, $|w_X - \tilde{w}_X| = \mathcal{O}\left(\eta^{-12} N^{-1/3}\right)$, and by Lemma 4.8,

$$\widehat{\mathcal{G}}_{ii} = \frac{1}{2}\left(\frac{1}{\overline{Y}_i - z - w_X} + \frac{1}{-\overline{Y}_i - z - w_X}\right) + \mathcal{O}\left(\eta^{-6} N^{-\frac{1}{2}+a}\right). \tag{4.86}$$

Because

$$\left|\frac{1}{\overline{Y}_i - z - w_X} - \frac{1}{\overline{Y}_i - z - \tilde{w}_X}\right| \leq \frac{|w_X - \tilde{w}_X|}{\eta^2}, \tag{4.87}$$

we obtain

$$\widehat{\mathcal{G}}_{ii} = \frac{1}{2}\left(\frac{1}{\overline{Y}_i - z - \tilde{w}_X} + \frac{1}{-\overline{Y}_i - z - \tilde{w}_X}\right) + \mathcal{O}\left(\eta^{-14} N^{-1/3}\right). \tag{4.88}$$

This statement for fixed z and t may be upgraded to the desired statement uniform over $\Sigma \times [0, \tau]$ by a standard stochastic continuity argument, as indicated in the proof of [33, Theorem 3.16]. Observe that we use Lemma C.3 in place of [33, Theorem 3.3] in that argument. Finally, the estimate may be transferred from $\widehat{\mathcal{G}}$ to $\mathcal{G}$ by using the resolvent expansion and Ward identity, again as in the proof of [33, Theorem 3.16]. $\quad \square$

5 Eigenvector estimates

5.1 Global eigenvector bounds

Theorem 5.1. There are constants $p, c(\mathfrak{b}) > 0$ such that the following holds. With probability $1 - e^{-N^c}$,

$$\sup_{0\leq t\leq\tau} \max_{1\leq\alpha,i\leq N} |w_\alpha(i)|^2 + |z_\alpha(i)|^2 \leq N^{-1/p}, \quad \sup_{0\leq t\leq\tau} \max_{\beta\neq\alpha} \gamma_{\alpha\beta} \leq N^{-2/p+\mathfrak{a}}. \tag{5.1}$$

Proof. Recall the eigenvectors of $\mathcal{H}$ are of the form (w_α, z_α) and $(-w_\alpha, z_\alpha)$, corresponding to eigenvalues λ_α and $-\lambda_\alpha$ respectively. Recall that Σ was defined in (4.76). Then for $(z,t) \in \Sigma \times [0, \tau]$ and $1 \leq \alpha \leq N$,

$$\operatorname{Im}\mathcal{G}_{\alpha\alpha}(z) = \eta\sum_{i=1}^N \frac{|w_\alpha(i)|^2}{|\lambda_i - E|^2 + \eta^2} + \eta\sum_{i=1}^N \frac{|z_\alpha(i)|^2}{|-\lambda_i - E|^2 + \eta^2}. \tag{5.2}$$

Setting $E = \lambda_i$, we obtain

$$|w_\alpha(i)|^2 \leq \eta \operatorname{Im} \mathcal{G}_{\alpha\alpha}(\lambda_i + i\eta). \tag{5.3}$$

Denote $z_i = \lambda_i + i\eta$. By the local law for $\mathcal{G}_{\alpha\alpha}$, Corollary 4.13, it suffices to lower bound

$$|-z_i - \overline{Y}_\alpha - \operatorname{Re}\tilde{w}_X| \vee \operatorname{Im}\tilde{w}_X. \tag{5.4}$$

and

$$|-z_i + \overline{Y}_\alpha - \operatorname{Re}\tilde{w}_X| \vee \operatorname{Im}\tilde{w}_X. \tag{5.5}$$

By Corollary 5.5, proved in the next subsection,

$$|-z_i + \overline{Y}_\alpha - \operatorname{Re}\tilde{w}_X| \vee \operatorname{Im}\tilde{w}_X \geq \eta^{1-\mathfrak{c}/4}, \quad |-z_i - \overline{Y}_\alpha - \operatorname{Re}\tilde{w}_X| \vee \operatorname{Im}\tilde{w}_X \geq \eta^{1-\mathfrak{c}/4}. \tag{5.6}$$

Therefore, using Corollary 4.13, we have

$$|w_\alpha(i)|^2 \leq \eta^{\mathfrak{c}/4} + C\eta^{-11}N^{-1/3} \tag{5.7}$$

with exponentially high probability. Taking $\eta = N^{-1/q}$ with q large finishes the proof of the bound on $|w_\alpha(i)|^2$.

The bound on $|z_\alpha(i)|^2$ is analogous. Given the bound on $|w_\alpha(i)|^2 + |z_\alpha(i)|^2$, the bound for $\gamma_{\alpha\beta}$ follows from the definition of $\gamma_{\alpha\beta}$ after taking p large enough. $\qquad\square$

5.2 Deterministic estimates

Let m_1, m_2 be the Stieltjes transforms of $\mu_1^{\mathrm{sym}}, \mu_2^{\mathrm{sym}}$, and let w_1, w_2 be the solution to the system

$$\begin{cases} w_1 = \hat{m}_1(z + w_2) \\ w_2 = \hat{m}_2(z + w_1). \end{cases} \tag{5.8}$$

The proof of the following proposition is the same as [33, Proposition 3.9].

Proposition 5.2. There exists $p > 0$ such that if $\operatorname{Im} z \geq N^{-1/p}$, then

$$|m_1(z) - m_X(z)| \leq N^{-1/p}, \quad |m_2(z) - m_{\overline{Y}}(z)| \leq N^{-1/p} \tag{5.9}$$

for all z.

Given $p > 10$, define the spectral domain

$$\Sigma_p = \left\{ z = E + i\eta \colon \eta \in [N^{-1/p^2}, 1], E \in [-2K, 2K] \right\}, \tag{5.10}$$

Corollary 5.3. There exists a universal constant $p > 0$ such that for $z \in \Sigma_p$,

$$|w_1 - \tilde{w}_X| \vee |w_2 - \tilde{w}_{\overline{Y}}| \leq N^{-1/p}\eta^{-p}, \quad |\tilde{w}_X| \vee |\tilde{w}_{\overline{Y}}| \leq C. \tag{5.11}$$

Proof. This follows from Proposition 4.11 with

$$\delta w = \left(\frac{1}{m_1(z + \tilde{w}_{\overline{Y}})} - \frac{1}{m_X(z + \tilde{w}_{\overline{Y}})}, \frac{1}{m_2(z + \tilde{w}_X)} - \frac{1}{m_{\overline{Y}}(z + \tilde{w}_X)} \right). \tag{5.12}$$

We indicate how to bound the first coordinate; the second is analogous. From Proposition 5.2,

$$|\delta w_1| \leq \frac{N^{-1/p}}{m_1(z + \tilde{w}_{\overline{Y}})m_X(z + \tilde{w}_{\overline{Y}})}. \tag{5.13}$$

For any Stieltjes transform $m(z)$ of a measure μ,

$$\operatorname{Im} m(z) \geq \frac{\operatorname{Im} z}{(|z| + \sup_{x \in \operatorname{supp}\mu} |x|)^2}. \tag{5.14}$$

Recall from Proposition 4.9 that $\operatorname{Im} \tilde{w}_X \wedge \operatorname{Im} \tilde{w}_Y > 0$, and $|\hat{m}_X(z)| \leq C\eta^{-1}$ since $\hat{\mu}_X$ has finite mass. Taking p large and using $|\tilde{w}_X| \vee |\tilde{w}_{\overline{Y}}| \leq C\eta^{-1}$, which follows from (4.77) and the preceding comment, this shows

$$|\delta w_1| \leq C N^{-1/p} \eta^{-6} \leq C\eta. \tag{5.15}$$

In the last inequality we used the hypothesis that $\eta \geq N^{-1/p^2}$.

Finally, the second bound follows from the first and Proposition 4.9. $\qquad\square$

Define $\tilde{m}(z)$ to be the Stieltjes transform of $\mu_X^{\mathrm{sym}} \boxplus \mu_{\overline{Y}}^{\mathrm{sym}}$. We recall this means that $w_X, w_{\overline{Y}}, m_X, m_{\overline{Y}}, \tilde{m}$ satisfy (4.52).

Corollary 5.4. Under the same assumptions as Corollary 5.3,

$$\tilde{m}(z) \leq C. \tag{5.16}$$

Proof. As noted in Section 2.2, we assume that either μ_1 or μ_2 has a bounded Stieltjes transform. By Proposition 5.2, for $\eta \geq N^{-1/p}$,

$$|m_X(z + \tilde{w}_{\overline{Y}})| \wedge |m_Y(z + \tilde{w}_X)| \leq |m_1(z + \tilde{w}_{\overline{Y}})| \wedge |m_2(z + \tilde{w}_X)| + 2N^{-1/p} \leq C. \tag{5.17}$$

Using the definition of $\tilde{m}$ (recall (4.52)), this completes the proof. $\qquad\square$

The following corollary is essentially contained in the proof of [33, Theorem 3.15]. We include it for completeness.

Corollary 5.5. There exists $N_0 > 0$ such that for $z \in \Sigma_p$ and $N \geq N_0$,

$$|-z + \overline{Y}_\alpha + \operatorname{Re} \tilde{w}_X| \vee \operatorname{Im} \tilde{w}_X \geq \eta^{1-\mathfrak{c}/4}, \quad |-z - \overline{Y}_\alpha + \operatorname{Re} \tilde{w}_X| \vee \operatorname{Im} \tilde{w}_X \geq \eta^{1-\mathfrak{c}/4}. \tag{5.18}$$

Proof. We first show

$$|-z + \overline{Y}_\alpha + \operatorname{Re} \tilde{w}_X| \vee \operatorname{Im} \tilde{w}_X \geq \eta^{1-\mathfrak{c}/4}. \tag{5.19}$$

We suppose, for the sake of contradiction, that $|-z + \overline{Y}_\alpha + \operatorname{Re} \tilde{w}_X| \vee \operatorname{Im} \tilde{w}_X < \eta^{1-\mathfrak{c}/4}$. Recall

$$\tilde{w}_X = -z - \tilde{w}_{\overline{Y}} - \frac{1}{m_X(z + \tilde{w}_{\overline{Y}})}. \tag{5.20}$$

Taking the imaginary part and using Corollary 5.4 together with the definition (4.52) shows

$$\operatorname{Im} \tilde{w}_X + \operatorname{Im} \tilde{w}_{\overline{Y}} = -\eta + \frac{\operatorname{Im} m_X(z + \tilde{w}_{\overline{Y}})}{|m_X(z + \tilde{w}_{\overline{Y}})|^2} \geq -\eta + c \operatorname{Im} m_X(z + \tilde{w}_{\overline{Y}}) \tag{5.21}$$

for some $c > 0$. Set $I(\alpha, \eta) = \left|\left\{\beta : \left|\overline{Y}_\beta - \overline{Y}_\alpha\right| \leq \eta\right\}\right|$. By definition,

$$\begin{aligned}
\operatorname{Im} m_X(z + \tilde{w}_{\overline{Y}}) &= \frac{1}{2N} \sum_{\beta=-N}^{N} \frac{\eta + \operatorname{Im} \tilde{w}_{\overline{Y}}}{|-z + \overline{Y}_\beta + \operatorname{Re} \tilde{w}_X|^2 + |\eta + \operatorname{Im} \tilde{w}_X|^2} \\
&\geq \frac{1}{2N} \sum_{|\overline{Y}_\beta - \overline{Y}_\alpha| \leq \eta} \frac{\eta + \operatorname{Im} \tilde{w}_{\overline{Y}}}{|-z + \overline{Y}_\beta + \operatorname{Re} \tilde{w}_X|^2 + |\eta + \operatorname{Im} \tilde{w}_X|^2} \\
&\geq \frac{I(\alpha, \eta)}{2N} \frac{\eta}{4\eta^{2-\mathfrak{c}/2}} = \frac{I(\alpha, \eta)\eta^{\mathfrak{c}/2-1}}{8N}.
\end{aligned} \tag{5.22}$$

Using model assumptions (5) and (6), applying (C.10), and recalling $\tau = N^{-1+\mathfrak{b}}$, we find

$$I(\alpha, \eta) \geq |\{\beta : |y_\beta - y_\alpha| \leq 2\eta/3\}| \geq \frac{N}{2} \mu_2([y_\alpha - \eta/2, y_\alpha + \eta/2]) \geq cN\eta^{2-\mathfrak{c}}. \tag{5.23}$$

This implies $\operatorname{Im} m_X(z + \tilde{w}_{\overline{Y}}) \geq c\eta^{1-\mathfrak{c}/2}$, from which we conclude using the above work that

$$\operatorname{Im} \tilde{w}_X + \operatorname{Im} \tilde{w}_{\overline{Y}} \geq 2\eta^{1-\mathfrak{c}/3} \tag{5.24}$$

when N is large enough. By the assumption that $\operatorname{Im} \tilde{w}_X < \eta^{1-\mathfrak{c}/4}$, this implies $\operatorname{Im} \tilde{w}_{\overline{Y}} \geq \eta^{1-\mathfrak{c}/3}$. From $\tilde{w}_X = \hat{m}_X(z + \tilde{w}_{\overline{Y}})$ and Corollary 5.3, we obtain

$$\operatorname{Im} \tilde{w}_X = \int_{\mathbb{R}} \frac{(\eta + \operatorname{Im} \tilde{w}_{\overline{Y}})\hat{\mu}_X(dx)}{|x - z + \tilde{w}_{\overline{Y}}|^2} \geq \eta^{1-\mathfrak{c}/3} \int_{\mathbb{R}} \frac{\hat{\mu}_X(dx)}{|x - z + \tilde{w}_{\overline{Y}}|^2} \tag{5.25}$$

$$\geq \eta^{1-\mathfrak{c}/3} \int_{\mathbb{R}} \frac{\hat{\mu}_X(dx)}{c(x^2 + 1)} \geq c\eta^{1-\mathfrak{c}/3} \operatorname{Im} \hat{m}_X(i). \tag{5.26}$$

By Proposition 5.2, $\operatorname{Im} \hat{m}_X(i) \geq c$, which implies $\operatorname{Im} \tilde{w}_X \geq \eta^{1-\mathfrak{c}/4}$, contradicting our assumption (for large enough N).

Finally, an analogous argument shows $|-z - \overline{Y}_\alpha + \operatorname{Re} \tilde{w}_X| \vee \operatorname{Im} \tilde{w}_X \geq \eta^{1-\mathfrak{c}/4}$, and this completes the proof. $\square$

5.3 Bulk eigenvector bounds

Theorem 5.6. Let I be the interval in assumption (7). Fix $\nu > 0$ and set

$$\mathcal{D}_I = \left\{ z = E + i\eta \colon E \in I, N^{-1+\nu} \leq \eta \leq 1 \right\}. \tag{5.27}$$

Then

$$\inf_{z \in \mathcal{D}_I} \operatorname{Im} \tilde{w}_X \geq c, \tag{5.28}$$

and it holds with overwhelming probability that

$$\sup_{0 \leq t \leq \tau} \sup_{z \in \mathcal{D}_I} \max_{1 \leq i \leq 2N} \left| \mathcal{G}_{ii}(z, t) - \frac{z + \tilde{w}_X}{(y_i + (\tau - t)\widehat{A}_{ii})^2 - (z + \tilde{w}_X)^2} \right| \leq \frac{N^\nu}{\sqrt{N\eta}}, \tag{5.29}$$

where the indices in y_i and $\widehat{A}_{ii}$ are taken modulo N, and we require $N \geq N_0$ for some N_0 depending on ν.

Proof. By assumption, the empirical measure of $UR^\dagger XTV^\dagger$ converges to μ_1 weakly. Using $\tau = o(1)$ and (C.10), $(\tau - t)\widehat{A}$ is negligible and $Y + (\tau - t)\widehat{A}$ converges to μ_2 weakly. Fix a small $\sigma > 0$. Then by Theorem 4.4 of [16], for any fixed t,

$$\sup_{z \in \mathcal{D}_I} \max_{1 \leq i \leq N} \left| \widehat{\mathcal{G}}_{ii}(z, t) - \frac{z + \tilde{w}_X}{(y_i + (\tau - t)\widehat{A}_{ii})^2 - (z + \tilde{w}_X)^2} \right| \leq \frac{N^\sigma}{\sqrt{N\eta}} \tag{5.30}$$

with overwhelming probability for sufficiently large N not depending on t. Further, by Lemma A.2 of [16], there exists $c > 0$ such that $\inf_{z \in \mathcal{D}_I} \operatorname{Im} \tilde{w}_X \geq c$ for large enough N, independent of t. This implies the desired claim for $\widehat{\mathcal{G}}(z, t)$ at any fixed t. Observe there is an implicit dependence of $\tilde{w}_X$ on t.

This estimate may then be transferred to $\mathcal{G}_{ii}$, using the resolvent identity, and made uniform in t, using a standard stochastic continuity argument, as in [33, Theorem 3.16]. This completes the proof. $\square$

Corollary 5.7. Let I be the interval in assumption (7). Then for any $\nu > 0$, there exists $N_0(\nu) > 0$ such that the following estimates hold with overwhelming probability for $N \geq N_0$:

$$\sup_{0 \leq t \leq \tau} \max_{\lambda_\alpha \in I} \max_{1 \leq i \leq N} |w_\alpha(i)| + |z_\alpha(i)| \leq \frac{N^\nu}{\sqrt{N}}, \quad \sup_{0 \leq t \leq \tau} \max_{\lambda_\alpha \in I} \max_{\beta \neq \alpha} \gamma_{\alpha\beta} \leq \frac{N^{\mathfrak{a}+\nu}}{N}. \tag{5.31}$$

Proof. The proof is the same as [33, Corollary 3.17] (using both conclusions of Theorem 5.6) after observing that the eigenvectors of $\mathcal{H}$ are of the form (w_α, z_α) and $(-w_\alpha, z_\alpha)$. □

Define $m_3(z)$ as the Stieltjes transform of $\mu_1^{\mathrm{sym}} \boxplus \mu_2^{\mathrm{sym}}$. The next proof follows [33, Theorem 3.14].

Theorem 5.8. Let I and $\mathcal{D}_I$ be as in Theorem 5.6. There exist constants $q > 0$ and $N_0 > 0$ such that

$$\sup_{z \in \mathcal{D}_I} |w_1 - \tilde{w}_X| \vee |w_2 - \tilde{w}_{\overline{Y}}| \le N^{-1/q}, \quad \sup_{z \in \mathcal{D}_I} |m_3(z) - \tilde{m}(z)| \le N^{-1/q} \quad (5.32)$$

for $N \ge N_0$.

Proof. Let $q > 0$ be a constant to be determined later. We define $\Sigma_1 \subset \mathcal{D}_I$ by

$$\Sigma_1 = \left\{ z \in \mathcal{D}_I : |w_1 - \tilde{w}_X| \vee |w_2 - \tilde{w}_{\overline{Y}}| \le N^{-1/q} \right\}. \quad (5.33)$$

By Corollary 5.3, Σ_1 is nonempty. Because the functions involved in its definition are continuous, it is closed. Therefore to show $\Sigma_1 = \mathcal{D}_I$, it suffices to show Σ_1 is open in $\mathcal{D}_I$.

Because the density of $\mu_1^{\mathrm{sym}} \boxplus \mu_2^{\mathrm{sym}}$ is bounded above and below by positive constants, there exists $c > 0$ such that $c^{-1} \ge \operatorname{Im} m_3 \ge c$. By taking imaginary parts in (5.8) and the last equation of (4.52), using (5.14), and recalling that $|w_1| \vee |w_2|$ is bounded on compact subset of $\mathbb{C}^+ \cup \mathbb{R}$ by Proposition 4.9, we have

$$\operatorname{Im} w_2 \ge c \operatorname{Im} w_1, \quad \operatorname{Im} w_1 \ge c \operatorname{Im} w_2, \quad \operatorname{Im} w_1 + \operatorname{Im} w_2 \ge c > 0, \quad (5.34)$$

which implies $\operatorname{Im} w_1 \wedge \operatorname{Im} w_2 \ge c > 0$.[4] These lower bounds permit the use of Proposition 5.2 to conclude that on Σ_1,

$$|m_1(z + \tilde{w}_{\overline{Y}}) - m_X(z + \tilde{w}_{\overline{Y}})| \vee |m_2(z + \tilde{w}_X) - m_{\overline{Y}}(z + \tilde{w}_X)| \le 2N^{-1/p}. \quad (5.35)$$

Therefore, using the definition (4.53) and the lower bound $\operatorname{Im} m_X(z) \wedge \operatorname{Im} m_{\overline{Y}}(z) \ge c\eta$ for some $c > 0$ (which follows from the definition of the Stieltjes transform as the trace of a Green's function, as in Proposition 4.2), we have

$$|\hat{m}_1(z + \tilde{w}_{\overline{Y}}) - \hat{m}_X(z + \tilde{w}_{\overline{Y}})| \vee |\hat{m}_2(z + \tilde{w}_X) - \hat{m}_{\overline{Y}}(z + \tilde{w}_X)| \le CN^{-1/p}. \quad (5.36)$$

We now claim that on $\mathcal{D}_I$ the stability of the system of equations (5.8) is improved, so that the operator Φ from (4.55) satisfies

$$\left\| (D\Phi(w_1, w_2))^{-1} \right\| \le C. \quad (5.37)$$

To see this, one can reinspect the proof of Proposition 4.10 using the bound

$$\operatorname{Im} w_1 \wedge \operatorname{Im} w_2 \ge c > 0, \quad (5.38)$$

which implies $\tilde{p} \vee \tilde{q} \le C$, and the bound

$$\sup_{\operatorname{Im} z \ge 0} |pq| < \sup_{\operatorname{Im} z \ge 0} |\tilde{p}\tilde{q}| \le 1, \quad (5.39)$$

which holds because $\hat{\mu}_2$ is not a point mass and $\operatorname{Im} w_1 \wedge \operatorname{Im} w_2 \ge c > 0$. Because p, q are continuous, we find $|1 - pq| \ge c > 0$ on I. Repeating (4.63) and (4.64) with these improved bounds proves the claim.

---

[4]We also need $\eta = O(N^{-c})$ for some $c > 0$ for the third inequality in (5.34), but the first inequality of (5.32) for the complementary regime has already been proved in Corollary 5.3.

Similar reasoning gives $\|D\Phi(\zeta_1, \zeta_2)\|_\infty < C$, $\|D^2\Phi(\zeta_1, \zeta_2)\|_\infty \le C$ on Σ_1. We can therefore repeat the reasoning of the proof of Proposition 4.11 to show that for any $z \in \Sigma_1$, $|w_1 - \tilde{w}_X| \vee |w_2 - \tilde{w}_{\overline{Y}}| \le CN^{-1/p}$. The remainder term r in that proof is now bounded using (5.36). Therefore, there is a neighborhood of z such that $|w_1 - \tilde{w}_X| \vee |w_2 - \tilde{w}_{\overline{Y}}| \le N^{-1/q}$ when $q > p$. This shows that $\Sigma_1 = \mathcal{D}_I$.

Finally, on $\mathcal{D}_I$ we have using (4.52), the lower bound on $\operatorname{Im} w_2$, and Proposition 5.2,

$$|m_3(z) - \tilde{m}(z)| \le |m_X(z + \tilde{w}_{\overline{Y}}) - m_1(z + \tilde{w}_{\overline{Y}})| + |m_1(z + \tilde{w}_{\overline{Y}}) - m_1(z + w_2)| \le N^{-1/q}. \quad (5.40)$$

To bound the second term in the sum, we used $\operatorname{Im}(z + \tilde{w}_{\overline{Y}}) \wedge \operatorname{Im}(z + w_2) \ge c$ and the fact that $|\partial_z m_1(z)| \le C\eta^{-2}$. This completes the proof. $\qquad \square$

Corollary 5.9. Let I and $\mathcal{D}_I$ be as in Theorem 5.6. There exist constants $p > 0$ and N_0 such that, with overwhelming probability,

$$\sup_{z \in \mathcal{D}_I} \left| \frac{1}{N} \operatorname{Tr}\left(\frac{1}{\mathcal{H}(0) - z} \right) - m_3(z) \right| \le N^{-1/p} + \frac{N^\nu}{\sqrt{N\eta}} \quad (5.41)$$

for $N \ge N_0$.

Proof. The claim follows from Theorem 5.6, the identity $\tilde{m}(z) = m_{\overline{Y}}(z + \tilde{w}_X)$ from (4.52), and the second inequality in (5.32). Since $\overline{Y}$ is diagonal, $m_{\overline{Y}}(z + \tilde{w}_X)$ is sum of terms identical to the fractions in Theorem 5.6, because the latter are

$$\frac{z + \tilde{w}_X}{(y_i + (\tau - t)\widehat{A}_{ii})^2 - (z + \tilde{w}_X)^2} = \frac{1}{2}\left(\frac{1}{-z - \tilde{w}_X + \overline{Y}_i} + \frac{1}{-z - \tilde{w}_X - \overline{Y}_i} \right). \quad (5.42)$$

$\qquad \square$

6 Well posedness of dynamics

To show the well posedness of (3.13), it is important to ensure that the drift term, which depends on the the inverses of the eigenvalue spacings, does not become too singular. We guarantee this by adding a small perturbation to the diagonal matrix X defined in (2.1). Let

$$X' = \operatorname{diag}(x_1, \ldots, x_N) + e^{-N}Q, \quad (6.1)$$

where Q is an a $N \times N$ matrix of i.i.d. standard complex Gaussians. We first note that because the perturbation is exponentially small, it does not affect our desired conclusion. The proof is trivial and hence omitted. For the rest of this work, we use the redefined version of $\widehat{M}$ with X' and may not explicitly indicate this.

Lemma 6.1. If Theorem 2.1 holds when X' replaces X in definition (2.1), then Theorem 2.1 holds.

We now prove the desired eigenvalue repulsion estimates. The proof of the following lemma is similar to [33, Proposition 2.3]. For completeness we provide some details in the current context.

Lemma 6.2. Let P be a $N \times N$ matrix of complex numbers, and let Q be a $N \times N$ matrix of i.i.d. standard complex Gaussians. Define the $2N \times 2N$ matrix $\widetilde{P}$ by

$$\widetilde{P} = \begin{bmatrix} 0 & P + e^{-N}Q \\ P^* + e^{-N}Q^* & 0 \end{bmatrix} \quad (6.2)$$

Let $\gamma_1 \le \cdots \le \gamma_N$ be the eigenvalues of P and $\alpha_1 \le \cdots \le \alpha_N$ be the positive eigenvalues of $\widetilde{P}$. Let $\alpha_{-i} = -\alpha_i$ denote the corresponding negative eigenvalues. Then the α_i are

almost surely distinct, and we have the following estimates for every $\delta \in (0,1)$:

$$\mathbb{E} \sum_{i \neq j} \frac{1}{|\alpha_i - \alpha_j|} < c_N \psi(N, P), \quad \mathbb{P}[\min_{i \neq j} |\alpha_i - \alpha_j| \leq \delta] \leq c_N \psi(N, P) \delta^2, \tag{6.3}$$

where c_N is an N-dependent constant and

$$\psi(N, P) = \exp\left(e^{2N} \sum_{k,l} |P_{kl}|^2\right). \tag{6.4}$$

Finally, we have

$$\mathbb{P}[\max_{1 \leq k \leq N} |\alpha_k - \gamma_k| \geq e^{-N/2}] \leq e^{-e^{N/2}}. \tag{6.5}$$

Proof. Recall that P has a singular value decomposition $P = USV^*$, where S is diagonal and U and V are unitary. Therefore, after conjugating by the unitary block matrix $\frac{1}{\sqrt{2}} \begin{bmatrix} U & -U \\ V & V \end{bmatrix}$, which leaves invariant the eigenvalues and the distribution of Q, we may suppose P is real and diagonal.

Define the index set corresponding to the off-diagonal blocks by

$$\mathcal{J} = \{(i,j) \colon 1 \leq i, j \leq 2N, \ i \leq N < j \text{ or } j \leq N < i\}. \tag{6.6}$$

Let $\mathcal{H}_N$ be set of $2N \times 2N$ Hermitian matrices with zeros in the indices $\mathcal{J}^c$ (the diagonal $N \times N$ blocks). We parameterize $\mathcal{H}_N$ by the coordinates $(w_{ij}) \in \mathbb{R}^{2N \times 2N}$, where $w_{ij} = 0$ if $(i,j) \in \mathcal{J}^c$ and $h_{ij} = w_{ij} + iw_{ji}$ for $j > i$ otherwise. This space is naturally equipped with the Lebesgue measure for $\mathbb{R}^{2N^2}$.

Set $\sigma_N = e^{-N}$ and write the density for $\widetilde{P}$ as

$$p_{\widetilde{P}}(w) = \frac{1}{Z_N} \exp\left(-\frac{1}{2\sigma_N^2} \sum_{(i,j) \in \mathcal{J}} |w_{ij} - \delta_{j,i+N} P_{ii}|^2\right), \tag{6.7}$$

where we use that P is real, so only the w_{ij} representing the real parts of the diagonals of the off-diagonal blocks are shifted. Note the normalization constant Z_N does not depend on P.

In the eigenvalue–eigenvector coordinates,[5] we have

$$p_{\widetilde{P}}(\lambda, u, v) =$$
$$\frac{1}{Z_N} \exp\left(-\frac{1}{2\sigma_N^2}\left(2 \sum_{k=1}^{N} \lambda_k^2 + \sum_{k=1}^{N} P_{kk}^2 - 2 \sum_{k,\ell \leq N} \lambda_\ell \operatorname{Re}(u_{k\ell} v_{\ell k}^*) P_{kk}\right)\right) \prod_{i \neq j} (\lambda_i - \lambda_j)^2 g(u, v), \tag{6.8}$$

where we used

$$w_{k,k+N} = \sum_{\ell=1}^{N} \lambda_\ell \operatorname{Re}(u_{k\ell} v_{\ell k}^*) \tag{6.9}$$

from the singular value decomposition for the upper-right block of the (w_{ij}) matrix. Here $g(u,v)$ is an integrable function on the compact subdomain of $\mathbb{C}^{N(N-1/2)} \times \mathbb{C}^{N(N-1/2)}$ where the map $(u,v) \to (U(u), V(v))$ taking the strictly upper triangular part of a matrix to the full Hermitian matrix is well-defined.

---

[5]The technical details of this reparameterization are similar to [33, Proposition 2.3] and therefore omitted.

Using the trivial bound of 1 on the eigenvector entries and the AM–GM inequality, we obtain

$$2 \sum_{k,\ell \leq N} \lambda_\ell \operatorname{Re}(u_{k\ell} v_{\ell k}^*) P_{kk} \leq \sum_{k=1}^N \lambda_k^2 + \sum_{k=1}^N P_{kk}^2. \tag{6.10}$$

This implies

$$p_{\widetilde{P}}(\lambda, u, v) \leq \frac{1}{Z_N} \exp\left(-\frac{1}{2\sigma_N^2} \sum_{k=1}^N \lambda_k^2\right) \prod_{i \neq j} (\lambda_i - \lambda_j)^2 g(u, v). \tag{6.11}$$

Then integrating out the $g(u, v)$ term and integrating again to compute $\mathbb{E} \sum_{i \neq j} |\alpha_i - \alpha_j|^{-1}$, we obtain the first bound (where we use (4.5) and (4.6) to simplify the sum of the eigenvalues squared). The final inequality follows as in [33, Proposition 2.3]. □

With this estimate, the following well posedness theorem is proved nearly identically to [33, Theorem 5.2]. For any $\mathfrak{t} > 0$ we define the filtration

$$(\mathcal{F}_t)_{0 \leq t \leq \mathfrak{t}} = (\sigma(\widehat{M}(0), (B_s)_{0 \leq s \leq t}))_{0 \leq t \leq \mathfrak{t}}, \tag{6.12}$$

where B_s is the multi-dimensional Brownian motion driving (3.13).

Theorem 6.3. For any $\mathfrak{t} > 0$, the singular values $\lambda_i(t)$ of $\widehat{M}(t)$ and their negatives $-\lambda_i(t) = \lambda_{-i}(t)$ are the unique strong solution to the equation (3.13) on $[0, \mathfrak{t}]$ such that

- $\lambda(t)$ is adapted to the filtration $(\mathcal{F}_t)_{0 \leq t \leq \mathfrak{t}}$, and

- $\mathbb{P}[\lambda_{-N}(t) < \cdots < \lambda_{-1}(t) < \lambda_1(t) < \lambda_2(t) < \cdots < \lambda_N(t),\ \text{for almost all } t \in [0, \mathfrak{t}]] = 1$.

7 Analysis of SDEs

The system of SDEs for the evolution of the singular values of $\widehat{M}$ is

$$d\lambda_i = \frac{1}{\sqrt{2N}} dB_i + \frac{1}{2N} \sum_{j \neq i} \frac{1 - \gamma_{ij}}{\lambda_i - \lambda_j}\, dt + R_i, \tag{7.1}$$

for $1 \leq |i| \leq N$, where

$$R_i = \operatorname{Re}\left\langle j_i, \left(U^*(t)\widehat{A}V(t) - U^*(0)\widehat{A}V(0)\right) k_i \right\rangle dt$$
$$+ \frac{1}{\sqrt{N}} \operatorname{Re}\left\langle j_i, \left(U^*(t)\left(\mathbb{1}_{(i,j) \in \mathcal{I}_\mathfrak{a}} d\widetilde{B}_{ij}\right) V(t)\right) k_i \right\rangle. \tag{7.2}$$

for $i \geq 1$. We recall that with λ_i and λ_{-i} are coupled as discussed above so that $\lambda_i = -\lambda_{-i}$ (and the remainder terms and the γ_{ij} are coupled in the same way). We use the redefinition noted in Lemma 6.1, so that our well posedness result Theorem 6.3 applies.

Our plan is to study this system for times $0 \leq t \leq \tau$ with $\tau = N^{-1+\mathfrak{b}}$ and compare it to the process defined by

$$d\mu_i = \frac{1}{\sqrt{2N}} dB_i + \frac{1}{2N} \sum_{j \neq i} \frac{1}{\mu_i - \mu_j}\, dt, \qquad \mu_i(0) = \lambda_i(0), \tag{7.3}$$

which we treat using the methods of [34]. We follow closely the strategy in [33], commenting on the minor differences in the current setting.

7.1 Interpolating process

Fix a constant c_4 satisfying $0 < c_4 < \mathfrak{c}$. For $0 \leq \alpha \leq 1$ we define the interpolating process $z_i(t, \alpha)$ by the SDE

$$dz_i(t, \alpha) = \frac{1}{\sqrt{2N(1 + N^{-c_4})}} dB_i + \frac{1}{2N} \sum_{j \neq i} \frac{1 - \alpha \hat{\gamma}_{ij}}{z_i(t, \alpha) - z_j(t, \alpha)} dt, \quad z_i(0, \alpha) = \lambda_i(0), \quad (7.4)$$

with

$$\hat{\gamma}_{ij} = \gamma_{ij} \wedge N^{-\mathfrak{c}}. \quad (7.5)$$

The well posedness of (7.4) follows from the same method used to prove the well posedness of the Hermitian analogue [33, (4.13)]; see for example [33, Proposition 5.4]. The necessity of the N^{-c_4} term is technical and explained following [33, (4.15)]. Its presence means that the $z_i(t, 0)$ and $z_i(t, 1)$ are not exactly the same as the $\mu_i(t)$ and $\lambda_i(t)$. However, the difference is negligible [33, Lemma 4.2].

Define

$$m_0(z) = \frac{1}{2N} \sum_{1 \leq |i| \leq N} \frac{1}{\lambda_i(0) - z}, \quad (7.6)$$

and let $m_t(z)$ be the free convolution of m_0 with the semicircle law at time t (see [60] for details):

$$m_t(z) = m_0(z + t m_t(z)), \quad \lim_{|z| \to \infty} m_t(z) = 0. \quad (7.7)$$

Let $I = [-c, c]$ be the interval from Theorem 5.6 on which $\mu_1^{\mathrm{sym}} \boxplus \mu_2^{\mathrm{sym}}$ has a positive density bounded above and away from zero. Let $\gamma_i^{\boxplus}$ be the i-th classical eigenvalue location (the i-th N-quantile) for the measure $\mu_1^{\mathrm{sym}} \boxplus \mu_2^{\mathrm{sym}}$, and define the index set $\mathcal{J}$ by

$$\mathcal{J} = \{i \colon \gamma_i^{\boxplus} \in I\}. \quad (7.8)$$

From Corollary 5.9 we can deduce by standard arguments (cf. [46, Chapter 11]) that there exists $c > 0$ such that

$$|\gamma_i^{\boxplus} - \gamma_i(0)| \leq N^{-c} \quad (7.9)$$

for $i \in \mathcal{J}$ with overwhelming probability.

The function $m_t(z)$ is the Stieltjes transform of some probability density $\rho_t(E)$. Let the classical eigenvalue locations of the free convolution ρ_t be $\{\gamma_i(t)\}_{|i|=1}^N$. Note that by the same reasoning given in [33, Section 4.4], that for any $\nu > 0$ and $i, j \in \mathcal{J}$ with $|i - j| \geq N^\nu$,

$$c \frac{|i - j|}{N} \leq |\gamma_i(t) - \gamma_j(t)| \leq C \frac{|i - j|}{N}. \quad (7.10)$$

The following rigidity lemmas hold. They are straightforward adaptations of the proofs of Theorem 3.1 and Corollary 3.2 of [53], and the discussion in [33, Section 4.5]. The main difference is that our Brownian motions are coupled in pairs, $B_i = -B_{-i}$. However, this does not affect the bound on the Brownian motion terms in equation (3.33) of [53] in the proof the deformed law, so the same method applies here. Observe our global eigenvector bounds from Corollary 5.7 are used to prove the second lemma.

Lemma 7.1. For any $\nu \geq 0$, it holds with overwhelming probability that

$$\sup_{0 \leq t \leq \tau} \sup_{i \in \mathcal{J}} |z_i(t, 0) - \gamma_i(t)| \leq \frac{N^\nu}{N}. \quad (7.11)$$

Lemma 7.2. With overwhelming probability for $i \in \mathcal{J}$,

$$|z_i(t, \alpha) - \gamma_i(t)| \leq \frac{N^{5\mathfrak{a}}}{N}. \quad (7.12)$$

7.2 Conclusion

The remaining stochastic analysis, including a short-range approximation and use of the energy method, is virtually identical to the argument given in [33], and we obtain the following coupling.

Proposition 7.3. Fix $\kappa > 0$. Suppose that $\mathfrak{b} < \mathfrak{a}/100$ and $\mathfrak{a} < \mathfrak{c}/10$. For every time t such that $0 \leq t \leq \tau$, we have with overwhelming probability for every index $i \in \mathcal{J}$ that

$$|\lambda_i(t) - \mu_i(t)| \leq \frac{1}{N}\left(N^{-\mathfrak{c}/5} + N^{-5\mathfrak{b}} + N^{-1/4}\right). \tag{7.13}$$

The following proposition is essentially [34, Theorem 3.2].[6] Compared to that reference, a certain repulsion term is present in the dynamics we study here (cf. Appendix B), but the proof is nearly identical (and in fact strictly easier) in our case.

We first recall the setup from that reference. Fix $\delta_1 > 0$ and let g and G be N-dependent parameters such that

$$N^{-1+\delta_1} \leq g \leq N^{-\delta_1}, \quad G \leq N^{-\delta_1}. \tag{7.14}$$

Let V be a deterministic matrix and let $B_t = \{B_{ij}(t)\}_{1 \leq i,j \leq N}$ be a matrix of i.i.d. standard complex Brownian motions. Define

$$M_t = V + \frac{1}{\sqrt{N}}B_t, \quad H_t = \begin{bmatrix} 0 & M_t \\ M_t^\dagger & 0 \end{bmatrix}. \tag{7.15}$$

Let $\{s_i(t)\}_{i=-N}^N$ (omitting the zero index) be the eigenvalues of H_t. We set

$$m_V(z) = \frac{1}{2N}\sum_{i=-N}^N \frac{1}{s_i(0) - z}, \tag{7.16}$$

where again $i = 0$ is omitted in the sum.

For the next definition, we recall that $m_3(z)$ was defined as the Stieltjes transform of $\mu_1^{\text{sym}} \boxplus \mu_2^{\text{sym}}$.

Definition 7.4. With g and G as above, we say V is (g, G)-regular with respect to m_3 if there exists $c > 0$ such that

$$|\text{Im}\, m_V(E + i\eta) - m_3(z)| \leq N^{-c} \tag{7.17}$$

for $z = E + i\eta$ with $|E| \leq G$ and $\eta \in [g, 10]$, for large enough N, and if there exists a constant C_V such that $|v_i| \leq N^{C_V}$ for all v_i.

Let W be a random matrix whose entries are i.i.d. complex normal variables of variance N^{-1}, and let $\tilde{B}_t = \{\tilde{B}_{ij}(t)\}_{1 \leq i,j \leq N}$ be a matrix of i.i.d. standard complex Brownian motions. Define $W_t = W + N^{-1/2}\tilde{B}_t$. Recall $\{s_i(t)\}_{i=1}^N$ are the singular values of M_t, and let $\{r_i(t)\}_{i=1}^N$ be the singular values of W_t.

Proposition 7.5. Fix $\sigma > 0$, and let V be a deterministic matrix that is (g, G)-regular with respect to m_3. Let M_t, W_t, $\{s_i(t)\}$, and $\{r_i(t)\}$ be defined as above. Then there exists a coupling of the processes $\{s_i(t)\}$ and $\{r_i(t)\}$ such that the following holds. Given parameters $0 < \omega_1 < \omega_0$ and times $t_0 = N^{-1+\omega_0}$, $t_1 = N^{-1+\omega_1}$, with the restrictions that

$$gN^\sigma \leq t_0 \leq N^{-\sigma}G^2, \quad 2\omega_1 < \omega_0, \tag{7.18}$$

---

[6]It is likely that the techniques in the recent work [28] could be used provide a shorter proof than the one given in [34], but since the result is already established we do not take this up here.

there exist $C, \omega, \delta > 0$ such that

$$|s_i(t_a) - r_i(t_a)| < CN^{-1-\delta} \tag{7.19}$$

with overwhelming probability for $i < N^\omega$ and $t_a = t_0 + t_1$. Here C, ω, δ are constants that depend only on $\delta_1, \sigma, \omega_0, \omega_1$, and the constants used to verify Definition 7.4 for V.

Remark. Unfortunately, the statement of [34, Theorem 3.2] omits a necessary hypothesis used in its proof. We have corrected this in the statement of Proposition 7.5, and we now explain the changes in detail.

The definition of (g, G)-regularity used in [34, Theorem 3.2] is weaker than Definition 7.4, and merely requires that $\operatorname{Im} m_V$ be bounded above and below. Here, we impose the stronger condition (7.17). Together with assumption (2.12), (7.17) ensures that that $\hat{\rho}_{t_0}(0)$, the value at 0 of the density corresponding to the Stieltjes transform of the free convolution of the data V with t_0 times the semicircle law, is close to $1/\pi$, the value at 0 of the density of the semicircle law. The latter law governs the density of the singular values of the reference Gaussian ensemble, and this matching of densities is necessary to place the particles V on the same scale as that ensemble and permit the coupling at time t_0 between the $s_i(t)$ and $r_i(t)$ used in the proof of [34, Theorem 3.2]. This condition on $\hat{\rho}_{t_0}(0)$ is tacitly assumed in the proof of [34, Theorem 3.2] but missing from its statement.

To prove Proposition 7.5, one may follow the proof of [34, Theorem 3.2] to obtain (7.19) up to a scaling of the particles s_i by $\pi \hat{\rho}_{t_0}(0)$. Using (7.17), it can be shown that this scaling is $1 + O(N^{-c})$, and we obtain the claimed result. The details of the latter argument can be found in the proof of [33, Theorem 2.4]; see the discussion starting above (4.110).

The hypotheses of Proposition 7.5 are verified with overwhelming probability for the singular values of $\widehat{M}(0)$ by Corollary 5.9. Combining Proposition 7.5 with Proposition 7.3, we obtain short-time relaxation of the singular value dynamics.

Theorem 7.6. Fix $\sigma > 0$, $\kappa > 0$, suppose that $\mathfrak{b} < \mathfrak{a}/100$ and $\mathfrak{a} < \mathfrak{c}/10$, and retain the definitions of Proposition 7.3. Then there exists a coupling of the processes $\{\lambda_i(t)\}$ and $\{r_i(t)\}$ and a constant $N_0(\sigma, \kappa, \mathfrak{a}, \mathfrak{b}, \mathfrak{c})$ such that

$$|\lambda_i(t_a) - r_i(t_a)| < N^{-1-\delta} \tag{7.20}$$

with overwhelming probability for $i < N^\omega$ and $N \geq N_0$.

We are now positioned to prove our main theorem.

Proof of Theorem 2.1. Setting $t_a = \tau$ in Theorem 7.6, we have $|\lambda_1(\tau) - r_1(\tau)| < N^{-1-\delta}$ where λ_1 is the least singular value of $\widetilde{M}(\tau) = M(\tau)$ and $r_1(\tau)$ is the least singular value of a matrix with distribution $\sqrt{1 + \tau} W$, where W is a matrix of i.i.d. standard complex Gaussians. Note that $M(\tau)$ has the same law as M, so if $\lambda_1(M_N)$ is the least singular value of M_N, we have

$$|\lambda_1(M_N) - r_1(\tau)| < N^{-1-\delta}. \tag{7.21}$$

The distribution of the least singular value of a Gaussian matrix is known explicitly. For W and any $r \geq 0$ [40],

$$\mathbb{P}(N\lambda_1(W) \leq r) = 1 - e^{-r^2}. \tag{7.22}$$

Therefore,

$$\mathbb{P}(N\lambda_1(\sqrt{1 + \tau} W) \leq \sqrt{1 + \tau} r) = 1 - e^{-r^2}. \tag{7.23}$$

We now show the $\sqrt{1 + \tau}$ factor is negligible, so that we may compare $\lambda_1(M_N)$ directly

to $\lambda_1(W)$. We compute, using $1 - e^{-x} \leq x$,

$$\left| \mathbb{P}(N\lambda_1(W) \leq r) - \mathbb{P}(N\lambda_1(\sqrt{1+\tau}W) \leq r) \right| = \left| e^{-r^2/(1+\tau)} - e^{-r^2} \right|$$

$$= e^{-r^2/(1+\tau)} \left| 1 - \exp\left(-r^2 \left(\frac{\tau}{1+\tau} \right) \right) \right| \leq \left[e^{-\frac{r^2}{1+\tau}} \frac{r^2}{1+\tau} \right] \cdot \tau \leq C\tau = O(N^{-c}). \quad (7.24)$$

By (7.21),

$$\mathbb{P}(Nr_1(\tau) \leq r - N^{-\delta}) \leq \mathbb{P}(N\lambda_1(M_N) \leq r) \leq \mathbb{P}(Nr_1(\tau) \leq r + N^{-\delta}) \quad (7.25)$$

We deduce

$$\mathbb{P}(N\lambda_1(W) \leq r - N^{-\delta}) - CN^{-c}$$
$$\leq \mathbb{P}(N\lambda_1(M_N) \leq r) \leq \mathbb{P}(N\lambda_1(W) \leq r + N^{-\delta}) + CN^{-c}. \quad (7.26)$$

By (7.22), $N\lambda_1(W)$ has a bounded density, so the $N^{-\delta}$ terms in the above may be removed with $O(N^{-c})$ error, and we conclude that

$$\left| \mathbb{P}(N\lambda_1(M_N) \leq r) - \mathbb{P}(N\lambda_1(W) \leq r) \right| = O(N^{-c}) \quad (7.27)$$

as desired. $\qquad\square$

A Derivation of dynamics

The following is a formal calculation that ignores the technical issue of possible eigenvalue collisions. It is used in Section 6, where this issue is dealt with rigorously.

A.1 Calculation

With $\widehat{M}$ as above, we define the $2N \times 2N$ block matrix

$$X = \begin{bmatrix} 0 & \widehat{M} \\ \widehat{M}^\dagger & 0 \end{bmatrix} = \begin{bmatrix} 0 & M(t) + (\tau - t)U^*(0)\widehat{A}V(0) \\ M(t)^* + (\tau - t)V(0)^*\widehat{A}U(0) & 0 \end{bmatrix}. \quad (A.1)$$

Observe that the eigenvalues of X are the singular values of $\widehat{M}$ and their negatives. Let $\widehat{M} = JSK^*$ be the singular value decomposition of $\widehat{M}$. Then a matrix of normalized eigenvectors for X is

$$H = \frac{1}{\sqrt{2}} \begin{bmatrix} J & -J \\ K & K \end{bmatrix}. \quad (A.2)$$

We follow the approach of [46, Chapter 12] to compute the dynamics of the eigenvalues of X. Denote the eigenvalues of X by λ_α with corresponding eigenvectors u_α. For the elements x_{ij} of X that are not identically zero, we have

$$\frac{\partial \lambda_\alpha}{\partial x_{ij}} = u_\alpha^*(i)u_\alpha(j), \quad \frac{\partial u_\alpha(i)}{\partial x_{kl}} = \sum_{\beta \neq \alpha} \frac{u_\beta^*(k)u_\alpha(l)}{\lambda_\alpha - \lambda_\beta} u_\beta(i), \quad (A.3)$$

and by the chain rule,

$$\frac{\partial^2 \lambda_\alpha}{\partial x_{kl} \partial x_{ij}} = \sum_{\beta \neq \alpha} \frac{1}{\lambda_\alpha - \lambda_\beta} \left[u_\alpha^*(k)u_\beta(l)u_\beta^*(i)u_\alpha(j) + u_\beta^*(k)u_\alpha(l)u_\alpha^*(i)u_\beta(j) \right]. \quad (A.4)$$

Itô's formula gives

$$d\lambda_\alpha = \sum_{i,j} \frac{\partial \lambda_\alpha}{\partial x_{ij}} dx_{ij} + \frac{1}{2} \sum_{i,j,k,l} \frac{\partial^2 \lambda_\alpha}{\partial x_{kl} \partial x_{ij}} (dx_{ij})(dx_{kl}). \quad (A.5)$$

The first term is, for $\alpha \leq N$,

$$\sum_{i,j} \frac{\partial \lambda_\alpha}{\partial x_{ij}} dx_{ij} = \sum_{i,j} u_\alpha^*(i) u_\alpha(j) dx_{ij} = u_\alpha^*(dX) u_\alpha \tag{A.6}$$

$$= \frac{1}{2} j_\alpha^* \left(\frac{1}{\sqrt{N}} d\widehat{B} + (U^* \widehat{A} V - U^*(0)\widehat{A}V(0)) \, \mathrm{d}t - \frac{1}{\sqrt{N}} U^*(\mathbb{1}_{(i,j)\in\mathcal{I}_{\mathfrak{a}}^c} \, \mathrm{d}\widetilde{B}_{ij})V \right) k_\alpha \tag{A.7}$$

$$+ \frac{1}{2} k_\alpha^* \left(\frac{1}{\sqrt{N}} d\widehat{B} + (U^* \widehat{A} V - U(0)\widehat{A}V(0)^*) \, \mathrm{d}t - \frac{1}{\sqrt{N}} U^*(\mathbb{1}_{(i,j)\in\mathcal{I}_{\mathfrak{a}}^c} \, \mathrm{d}\widetilde{B}_{ij})V \right)^* j_\alpha. \tag{A.8}$$

We see that

$$\frac{1}{2\sqrt{N}} \left(j_\alpha^* d\widehat{B} k_\alpha + k_\alpha^* d\widehat{B}^* j_\alpha \right) = \frac{1}{\sqrt{2N}} dB_\alpha, \tag{A.9}$$

where $\{dB_\alpha\}_{\alpha=1}^N$ is a set of independent standard real Brownian motions. The independence follows from an explicit computation, noting that $(d\widehat{B}_{ij})(d\widehat{B}_{kl}) = \delta_{il}\delta_{jk}$. The remaining terms are

$$\mathrm{Re}\langle j_\alpha, (U\widehat{A}V^* - U(0)\widehat{A}V(0)^*)k_\alpha\rangle \, \mathrm{d}t + \frac{1}{\sqrt{N}} \mathrm{Re}\langle j_\alpha, (U(\mathbb{1}_{(i,j)\in\mathcal{I}_{\mathfrak{a}}} \, \mathrm{d}\widetilde{B}_{ij})V^*)k_\alpha\rangle. \tag{A.10}$$

The second term is

$$\frac{1}{2} \sum_{i,j,k,l} \frac{\partial^2 \lambda_\alpha}{\partial x_{kl}\partial x_{ij}}(dx_{ij})(dx_{kl})$$

$$= \frac{1}{2} \sum_{i,j,k,l} \sum_{\beta \neq \alpha} \frac{1}{\lambda_\alpha - \lambda_\beta} \left[u_\alpha^*(k)u_\beta(l)u_\beta^*(i)u_\alpha(j) + u_\beta^*(k)u_\alpha(l)u_\alpha^*(i)u_\beta(j) \right] dx_{ij} dx_{kl}. \tag{A.11}$$

The first contribution is

$$\frac{1}{2} \sum_{i,j,k,l} \sum_{\beta \neq \alpha} \frac{1}{\lambda_\alpha - \lambda_\beta} \left[u_\alpha^*(k)u_\beta(l)u_\beta^*(i)u_\alpha(j) + u_\beta^*(k)u_\alpha(l)u_\alpha^*(i)u_\beta(j) \right] \frac{1}{N}(d\widehat{B}_{ij}')(d\widehat{B}_{kl}'), \tag{A.12}$$

where

$$d\widehat{B}' = \begin{bmatrix} 0 & d\widehat{B} \\ d\widehat{B}^* & 0 \end{bmatrix}. \tag{A.13}$$

This vanishes unless $i = l$, $j = k$, and exactly one of i or j is greater than N, due to the covariation factor. Summing over i and j, we obtain the norm of the first or last half of each u_α, that is $\|j_\alpha\|_2/2$ or $\|k_\alpha\|_2/2$, both of which are $1/2$. We then recover the drift term

$$\frac{1}{2N} \sum_{\beta \neq \alpha} \frac{dt}{\lambda_\alpha - \lambda_\beta}. \tag{A.14}$$

The remaining contribution is

$$-\frac{1}{2N} \sum_{i,j,k,l} \sum_{\beta \neq \alpha} \frac{1}{\lambda_\alpha - \lambda_\beta} \left[(u_\alpha^*(k)R_{kl}'u_\beta(l))(u_\beta^*(i)R_{ij}'u_\alpha(j)) + (u_\beta^*(k)R_{kl}'u_\alpha(l))(u_\alpha^*(i)R_{ij}'u_\beta(j)) \right]. \tag{A.15}$$

where

$$R' = \begin{bmatrix} 0 & U^*(\mathbb{1}_{(i,j)\in\mathcal{I}_{\mathfrak{a}}^c} \, \mathrm{d}\widetilde{B}_{ij})V \\ (U^*(\mathbb{1}_{(i,j)\in\mathcal{I}_{\mathfrak{a}}^c} \, \mathrm{d}\widetilde{B}_{ij})V)^* & 0 \end{bmatrix}. \tag{A.16}$$

We perform the sum on i and j first. We have

$$\sum_{i,j} u_\alpha^*(i) R'_{ij} u_\beta(j) = \frac{1}{2}\left(j_\alpha^* U^*(\mathbb{1}_{(i,j)\in\mathcal{I}_a^c}\,\mathrm{d}\widetilde{B}_{ij})Vk_\beta + k_\alpha^* V^*(\mathbb{1}_{(i,j)\in\mathcal{I}_a^c}\,\mathrm{d}\widetilde{B}_{ij})^*Uj_\beta\right). \quad \text{(A.17)}$$

Define the column vectors

$$w_\alpha = Uj_\alpha, \quad z_\alpha = Vk_\alpha, \quad \text{(A.18)}$$

and set $R = (\mathbb{1}_{(i,j)\in\mathcal{I}_a^c}\,\mathrm{d}\widetilde{B}_{ij})$. Then since the quadratic variation of a standard complex Brownian motion is zero, and the elements of R are independent,

$$(w_\alpha^* R z_\beta + z_\alpha^* R^* w_\beta)(w_\beta^* R z_\alpha + z_\beta^* R^* w_\alpha) = w_\alpha^* R z_\beta z_\beta^* R^* w_\alpha + z_\alpha^* R^* w_\beta w_\beta^* R z_\alpha \quad \text{(A.19)}$$

$$= \sum_{(i,j)\in\mathcal{I}_a^c} |w_\alpha(i)|^2|z_\beta(j)|^2 + \sum_{(i,j)\in\mathcal{I}_a^c} |w_\beta(i)|^2|z_\alpha(j)|^2 \quad \text{(A.20)}$$

$$:= 2\gamma_{\alpha\beta}. \quad \text{(A.21)}$$

Then (A.15) becomes, remembering the factors of $1/2$ from (A.17) and using $\gamma_{\alpha\beta} = \gamma_{\beta\alpha}$,

$$-\frac{1}{2N}\sum_{\beta\neq\alpha} \frac{\gamma_{\alpha\beta}\,dt}{\lambda_\alpha - \lambda_\beta}. \quad \text{(A.22)}$$

We obtain the following SDE for the eigenvalues of X, which are the singular values of $\widehat{M}$ and their negatives. We label the positive eigenvalues by $\{\lambda_i\}_{i=1}^N$ and the negative eigenvalues by $\{\lambda_i\}_{i=-1}^{-N}$, where we have set $\lambda_{-i} = -\lambda_i$. The same convention holds for the Brownian motions B_i: there are $2N$ of them, and the ones with positive indices are coupled to those with negative indices by $B_{-i} = -B_i$. The final SDE is, for $i > 0$:

$$d\lambda_i = \frac{1}{\sqrt{2N}}dB_i + \frac{1}{2N}\sum_{j\neq i} \frac{1 - \gamma_{ij}}{\lambda_i - \lambda_j}\,dt + R_i, \quad \text{(A.23)}$$

where

$$R_i = \mathrm{Re}\langle j_i, (U^*\widehat{A}V - U^*(0)\widehat{A}V(0))k_i\rangle\,\mathrm{d}t + \frac{1}{\sqrt{N}}\mathrm{Re}\langle j_i, (U^*(\mathbb{1}_{(i,j)\in\mathcal{I}_a}\,\mathrm{d}\widetilde{B}_{ij})V)k_i\rangle. \quad \text{(A.24)}$$

For $i < 0$ one can check that $R_i = -R_{-i}$ and $\gamma_{ij} = -\gamma_{-i,j}$.

In preparation for the next section, we note that when we conjugate the initial data by orthogonal instead of unitary matrices, the matrix (A.1) is a function of N^2 real variables instead of $2N^2$ real variables (N^2 complex entries). In this case we obtain (see, for example, [46, Chapter 12]) that

$$\frac{\partial\lambda_\alpha}{\partial x_{ij}} = 2u_\alpha(i)u_\alpha(j), \quad \frac{\partial u_\alpha(i)}{\partial x_{kl}} = \sum_{\beta\neq\alpha} \frac{u_\beta(k)u_\alpha(l) + u_\beta(l)u_\alpha(k)}{\lambda_\alpha - \lambda_\beta}u_\beta(i), \quad \text{(A.25)}$$

where we view X as a function of N^2 real variables x_{ij} with $1 \le i \le N$ and $N \le j \le 2N$. Using the representation (A.2), we find

$$u_{-\alpha}(k)u_\alpha(l) + u_\beta(l)u_\alpha(k) = u_{-\alpha}(k)u_\alpha(l) + u_{-\alpha}(l)u_\alpha(k) \quad \text{(A.26)}$$

$$= -u_\alpha(k)u_\alpha(l) + u_\alpha(l)u_\alpha(k) \quad \text{(A.27)}$$

$$= 0. \quad \text{(A.28)}$$

Therefore, we see that the sum in the drift component now omits the term with $\beta = -\alpha$, and there is no repulsion between λ_α and $\lambda_{-\alpha}$. The rest of the derivation is completed as before.
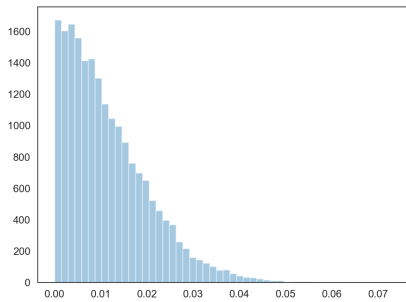
Figure 1: Simulated distribution of the least singular value of the real model, with the elements of X and Y chosen uniformly from $[0,1]$, matrix size $N = 200$, and 2×10^4 samples.
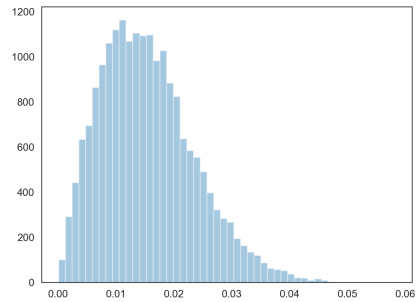

Figure 2: Simulated distribution of the least singular value of the complex model, with the elements of X and Y chosen uniformly from $[0,1]$, matrix size $N = 200$, and 2×10^4 samples.

B Real case

We now consider the real analogue of the model of Section 2.2, where the initial data is conjugated by orthogonal matrices. Precisely, in this section we consider the matrix ensemble

$$M = R^{\dagger} X T + U^{\dagger} Y V, \tag{B.1}$$

where $X = \mathrm{diag}(x_1, \ldots, x_N)$ and $Y = \mathrm{diag}(y_1, \ldots, y_N)$ are deterministic diagonal matrices and R, T, U, V are independent and distributed according to the Haar measure on the orthogonal group $O(N)$. We retain the hypothesis (2.2) and the assumptions labeled (1) through (7) on X and Y given in Section 2.2.

The least singular value in the real case displays qualitatively different behavior than its counterpart in the complex case, as indicated by the accompanying simulation results. The density for λ_1 vanishes at zero in the complex model, but remains positive in the real model. The singular value distribution in the real case is said to have a *hard edge* at zero.

This phenomenon may be understood dynamically. As discussed in Appendix A, the drift term in the complex case has the repulsion component

$$\frac{1}{2N} \sum_{j \neq i} \frac{1 - \gamma_{ij}}{\lambda_i - \lambda_j} \, dt, \tag{B.2}$$

while the same computation in the real case yields the repulsion term

$$\frac{1}{2N} \sum_{j \neq i, -i} \frac{1 - \gamma_{ij}}{\lambda_i - \lambda_j} \, dt, \tag{B.3}$$

with the interaction between λ_i and λ_{-i} removed. For λ_1, this means there is no force from λ_{-i} pushing it away from the origin, resulting in the hard edge.

The model (B.1) can be handled by the same method used for (2.1). The definition of the matrix dynamics in Section 3 is the same except for obvious changes, such as the use of orthogonal matrices and real symmetric Brownian motions. This leads to virtually the same singular value dynamics as in Appendix A, with the important exception of the interaction term noted above. The estimates of Section C are also essentially unchanged. An inspection of the proofs referenced in Section 6 and Section 7 shows that they

still apply to the dynamics in the real case. An important point is that the short-time universality result Proposition 7.5 still holds without the regularizing force from λ_{-1}; this was the original form of the result stated in [34]. Finally, for the exact form of the distribution of the least singular value for the Gaussian matrix, we use the form with quantitative error given in [73, Theorem 1.3].

We obtain the following analogue of Theorem 2.1 for the real model.

Theorem B.1. Let $\lambda_1(M_N)$ be the least singular value of the random matrix ensemble (B.1). For all $r \geq 0$, we have

$$\mathbb{P}(N\lambda_1(M_N) \leq r) = 1 - e^{-r^2/2 - r} + O(N^{-c}) \tag{B.4}$$

where $c > 0$ is an absolute constant uniform in r.

C Preliminary estimates

C.1 Diagonal matrices

Proposition C.1. For any $a > 0$ and $E \in \mathbb{R}$, we have

$$\frac{1}{2N} \sum_{|y_i - E| \geq N^{-1+a}} \frac{1}{|y_i - E|^2} \leq 2C_a N^{1-a}, \tag{C.1}$$

$$\frac{1}{2N} \sum_{|y_i - E| \geq N^{-1+a}} \frac{1}{|y_i - E|} \leq 2C_a \log N + 4, \tag{C.2}$$

where the sums are taken over indices i such that $1 \leq |i| \leq N$.

Proof. Let $\eta = N^{-1+a}$. Note that

$$\frac{1}{2N} \sum_{|y_i - E| \geq \eta} \frac{\eta}{|y_i - E|^2} \leq \frac{1}{N} \operatorname{Im} \sum_{|i|=1}^{N} \frac{1}{y_i - (E + i\eta)} \leq 2C_a. \tag{C.3}$$

Divide both sides by η to obtain

$$\frac{1}{2N} \sum_{|y_i - E| \geq \eta} \frac{1}{|y_i - E|^2} \leq \frac{2C_a}{\eta} = 2C_a N^{1-a}. \tag{C.4}$$

This proves the first inequality in the proposition. For the second inequality, note that for $x > \eta$, we have

$$\frac{1}{x} \leq \frac{2}{x+1} + \int_\eta^1 \frac{2\,dt}{(x+t)^2} \leq 2 + \int_\eta^1 \frac{2\,dt}{x^2 + t^2}. \tag{C.5}$$

Taking $x = |y_i - E|$ and summing over i such that $|y_i - E| \geq \eta$, we have

$$\frac{1}{2N} \sum_{|y_i - E| \geq \eta} \frac{1}{|y_i - E|} \leq 4 + \frac{1}{2N} \int_\eta^1 \sum_{|i|=1}^{N} \frac{4\,dt}{|y_i - (E + it)|^2}. \tag{C.6}$$

Using $\frac{1}{|y_i - (E+it)|^2} = \frac{1}{t} \operatorname{Im}\left(\frac{1}{y_i - (E+it)}\right)$ and the hypothesized bound (2.7) on $m_Y(E + i\eta)$, we have

$$\frac{1}{2N} \sum_{|y_i - E| \geq \eta} \frac{1}{|y_i - E|} \leq 4 + C \int_\eta^1 \frac{dt}{t} \leq 4 + C\,|\log \eta|. \tag{C.7}$$

Here $C = 2\sup_{\eta \leq t \leq 1} |m_Y(E + it)|$, which is bounded by $2C_a$ according to the maximum principle for holomorphic functions. Recall that $\eta = N^{-1+a}$, so we have

$$\frac{1}{2N} \sum_{|y_i - E| \geq \eta} \frac{1}{|y_i - E|} \leq 2C_a \log N + 4. \tag{C.8}$$

$\square$

Recalling definition (3.1) and using Proposition C.1 with $E = y_i$, we immediately have a bound for A:

$$\|A\| = \max_{1 \leq i \leq N} |A_{ii}| \leq CN^{1-\mathfrak{a}}. \tag{C.9}$$

Similarly we obtain

$$\|\widehat{A}\| = \max_{1 \leq i \leq N} |\widehat{A}_{ii}| \leq C(1 + \log N). \tag{C.10}$$

C.2 Unitary flow

The proofs of the following lemmas are essentially identical to those of [33, Theorems 3.1 and 3.3].[7]

Lemma C.2. For $\mathfrak{a}, \mathfrak{b}, U$, as above,

$$\mathbb{P}[\sup_{0 \leq t \leq \tau} \|U(t) - I\| \geq N^{-10\mathfrak{b}}] \leq \exp\left(-N^{10\mathfrak{b}}\right), \tag{C.11}$$

and the same estimate holds for V.

For any $t_0 \leq \tau$ define $\widehat{U}(t_0) = U(t)U(t_0)^*$.

Lemma C.3. For N large enough the following holds. For any $0 \leq t_0 \leq t \leq \tau$, $|t - t_0| \leq 1/N$,

$$\mathbb{P}\left[\sup_{t_0 \leq s \leq t} \|\widehat{U}(s) - \widehat{U}(t_0)\| \geq (N(t - t_0))^{1/4}\right] \leq \exp\left(-N^{\mathfrak{a}/3}\right). \tag{C.12}$$

Also, for any $0 \leq t_0 \leq t \leq \tau$, $|t - t_0| \leq r \leq 1/N$,

$$\mathbb{P}\left[\sup_{t_0 \leq s \leq t} \|\widehat{U}(s) - \widehat{U}(t_0)\| \geq r^{9/20}\right] \leq \exp\left(-c_N r^{-1/10}\right). \tag{C.13}$$

where $c_N > 0$ depends on N.

C.3 Sufficient conditions for positive density

The next lemma follows from the argument in [15, Lemma 3.2]. We provide the reasoning again here for completeness.

Lemma C.4. Let μ_α, μ_β be probability measures with density functions ρ_α, ρ_β that are symmetric about zero and are strictly positive on $[-r_0, r_0]$ for some $r_0 > 0$. Then $\mu_\alpha \boxplus \mu_\beta$ has a density, and that density is bounded above and away from zero in a neighborhood of zero.

Proof. According to [23, Corollary 8], $\mu_\alpha \boxplus \mu_\beta$ has a bounded density. It remains to show it is bounded away from zero. By Proposition 4.9, the corresponding subordination functions w_α, w_β extend continuously to 0 with values in $\mathbb{C}^+ \cup \mathbb{R} \cup \{\infty\}$. By the equations defining the free convolution, it suffices to show these limits are not infinite to show that the density $\mu_\alpha \boxplus \mu_\beta$ is bounded below in a neighborhood of 0.

We proceed by contradiction. Fix $r < r_0/2$ and define

$$\mathcal{E} = \left\{z \in \mathbb{C}^+ \cup \mathbb{R} \colon |z| \leq r\right\}. \tag{C.14}$$

Let $L > r_0$ and $M > 10$ be large parameters to be fixed later. We first suppose that there exists $z \in \mathcal{E}$ such that $|w_\alpha(z)| > LM$ and $|w_\beta(z)| > L$. The defining equations for the free convolution give

$$(w_\alpha + w_\beta - z)^{-1} = \int_{\mathbb{R}} \frac{d\mu_\beta(x)}{w_\alpha - x} = w_\alpha^{-1} + O(w_\alpha^{-2}), \tag{C.15}$$

---

[7]We recall that equation (3.7) in this reference is derived by applying the formula $\frac{d}{dt}e^{\theta X(t)} = \int_0^\theta e^{\alpha X(t)} \frac{dx(t)}{dt} e^{(\theta-\alpha)X(t)} dt$, which holds for any one-parameter matrix subgroup $X(t)$ [81], to compute the derivative of the matrix exponential with respect to each matrix entry, in conjunction with Itô's formula.

where the O notation is with respect to the limit $L \to \infty$. The above equation gives

$$\frac{w_\beta}{w_\alpha} = O(w_\alpha^{-1}). \tag{C.16}$$

This contradicts $L/|w_\alpha| \leq |w_\beta/w_\alpha|$ (which holds by our assumptions on w_α, w_β) for L sufficiently large.

We next suppose $|w_\alpha(z)| > LM$ and $|w_\beta(z)| \leq L$, and find from the definition of free convolution that for $z \in \mathcal{E}$ and M sufficiently large,

$$\frac{1}{|m_\alpha(w_\beta)|} = |w_\alpha + w_\beta - z| \geq \frac{ML}{2}. \tag{C.17}$$

By symmetry of μ_α and μ_β we know that w_β is imaginary for z on the imaginary line $\{i\eta \mid \eta \in \mathbb{R}\}$. But $m_\alpha(z)$ has no zeros on the imaginary line, as ρ_α is positive near 0. So it is bounded away from zero in $z \in \mathcal{E}$. For M large we reach a contradiction. This completes the proof. $\qquad \square$

In the case $\mu_\alpha = \mu_\beta$, only the first part of the previous argument is required.

Lemma C.5. Let μ_α be a symmetric probability measure, not necessarily absolutely continuous, supported at more than 2 points. Then $\mu_\alpha \boxplus \mu_\alpha$ has a density, and that density is bounded above and away from zero in a neighborhood of zero.

References

[1] Emmanuel Abbe, Amir Shpilka, and Avi Wigderson, *Reed–Muller codes for random erasures and errors*, IEEE Transactions on Information Theory **61** (2015), no. 10, 5229–5252. MR-3400278

[2] Radoslaw Adamczak, Olivier Guédon, Alexander Litvak, Alain Pajor, and Nicole Tomczak-Jaegermann, *Smallest singular value of random matrices with independent columns*, Comptes Rendus Mathematique **346** (2008), no. 15-16, 853–856. MR-2441920

[3] Ben Adlam and Ziliang Che, *Spectral statistics of sparse random graphs with a general degree distribution*, Preprint arXiv:1509.03368 (2015).

[4] Amol Aggarwal, *Bulk universality for generalized Wigner matrices with few moments*, Probability Theory and Related Fields **173** (2019), no. 1-2, 375–432. MR-3916110

[5] Amol Aggarwal, Patrick Lopatto, and Horng-Tzer Yau, *GOE statistics for Lévy matrices*, arXiv preprint arXiv:1806.07363 (2018).

[6] Andrew Ahn, *Airy point process via supersymmetric lifts*, arXiv preprint arXiv:2009.06839 (2020).

[7] Oskari Ajanki, László Erdős, and Torben Krüger, *Quadratic vector equations on complex upper half-plane*, Preprint arXiv:1506.05095 (2015). MR-4031100

[8] Oskari Ajanki, László Erdős, and Torben Krüger, *Singularities of solutions to quadratic vector equations on the complex upper half plane*, Communications on Pure and Applied Mathematics **70** (2017), no. 9. MR-3684307

[9] Oskari Ajanki, László Erdős, and Torben Krüger, *Universality for general Wigner-type matrices*, Probability Theory and Related Fields **169** (2017), no. 3-4, 667–727. MR-3719056

[10] Nima Anari, Constantinos Daskalakis, Wolfgang Maass, Christos Papadimitriou, Amin Saberi, and Santosh Vempala, *Smoothed analysis of discrete tensor decomposition and assemblies of neurons*, Advances in Neural Information Processing Systems, 2018, pp. 10857–10867.

[11] Greg W Anderson, Alice Guionnet, and Ofer Zeitouni, *An Introduction to Random Matrices*, Cambridge University Press, 2010. MR-2760897

[12] Zhigang Bao, László Erdős, and Kevin Schnelli, *Local stability of the free additive convolution*, Journal of Functional Analysis **271** (2016), no. 3, 672–719. MR-3506962

[13] Zhigang Bao, László Erdős, and Kevin Schnelli, *Convergence rate for spectral distribution of addition of random matrices*, Advances in Mathematics **319** (2017), 251–291. MR-3695875

[14] Zhigang Bao, László Erdős, and Kevin Schnelli, *Local law of addition of random matrices on optimal scale*, Communications in Mathematical Physics **349** (2017), no. 3, 947–990. MR-3602820

[15] Zhigang Bao, László Erdős, and Kevin Schnelli, *Spectral rigidity for addition of random matrices at the regular edge*, arXiv preprint arXiv:1708.01597 (2017). MR-4102163

[16] Zhigang Bao, László Erdős, and Kevin Schnelli, *Local single ring theorem on optimal scale*, The Annals of Probability **47** (2019), no. 3, 1270–1334. MR-3945747

[17] Anirban Basak and Mark Rudelson, *Invertibility of sparse non-Hermitian matrices*, Advances in Mathematics **310** (2017), 426–483. MR-3620692

[18] Roland Bauerschmidt, Jiaoyang Huang, Antti Knowles, and Horng-Tzer Yau, *Bulk eigenvalue statistics for random regular graphs*, The Annals of Probability **45** (2017), no. 6A, 3626–3663. MR-3729611

[19] Roland Bauerschmidt, Jiaoyang Huang, and Horng-Tzer Yau, *Local Kesten–McKay law for random regular graphs*, Communications in Mathematical Physics (2016), 1–114. MR-3962004

[20] Roland Bauerschmidt, Antti Knowles, and Horng-Tzer Yau, *Local semicircle law for random regular graphs*, Comm. Pure Appl. Math. **70** (2017), 1898–1960. MR-3688032

[21] Serban Teodor Belinschi, *A note on regularity for free convolutions*, Ann. Inst. Henri Pointcarè Probab. Stat. **42** (2006), no. 5, 635–648. MR-2259979

[22] Serban Teodor Belinschi, *The Lebesgue decomposition of the free additive convolution of two probability distributions*, Probability Theory and Related Fields **142** (2008), no. 1-2, 125–150. MR-2413268

[23] Serban Teodor Belinschi, l^∞-*boundedness of density for free additive convolutions*, Rev. Roumaine Math. Pures Appl. **59** (2014), no. 2, 173–184. MR-3299499

[24] Serban Teodor Belinschi and Hari Bercovici, *A new approach to subordination results in free probability*, J. Anal. Math. **101** (2007), no. 1, 357–365. MR-2346550

[25] Aditya Bhaskara, Moses Charikar, Ankur Moitra, and Aravindan Vijayaraghavan, *Smoothed analysis of tensor decompositions*, Proceedings of the forty-sixth annual ACM symposium on Theory of computing, ACM, 2014, pp. 594–603.

[26] Philippe Biane, *Representations of symmetric groups and free probability*, Advances in Mathematics **138** (1998), no. 1, 126–181. MR-1644993

[27] Arup Bose and Walid Hachem, *Smallest singular value and limit eigenvalue distribution of a class of non-Hermitian random matrices with statistical application*, Journal of Multivariate Analysis (2020), 104623. MR-4082540

[28] Paul Bourgade, *Extreme gaps between eigenvalues of Wigner matrices*, arXiv preprint arXiv:1812.10376 (2018).

[29] Paul Bourgade, L Erdős, H-T Yau, and Jun Yin, *Fixed energy universality for generalized Wigner matrices*, Comm. Pure Appl. Math. (2015). MR-3541852

[30] Paul Bourgade, László Erdős, Horng-Tzer Yau, and Jun Yin, *Universality for a class of random band matrices*, Advances in Theoretical and Mathematical Physics **21** (2017), no. 3, 739–800. MR-3695802

[31] Paul Bourgade, Fan Yang, H-T Yau, and Jun Yin, *Random band matrices in the delocalized phase, II: Generalized resolvent estimates*, Journal of Statistical Physics (2019), 1–33. MR-3934695

[32] Paul Bourgade, Horng-Tzer Yau, and Jun Yin, *Random band matrices in the delocalized phase, I: Quantum unique ergodicity and universality*, Communications on Pure and Applied Mathematics **73** (2020), no. 7, 1526–1596. MR-4156609

[33] Ziliang Che and Benjamin Landon, *Local spectral statistics of the addition of random matrices*, Probability Theory and Related Fields **175** (2019), no. 1-2, 579–654. MR-4009717

[34] Ziliang Che and Patrick Lopatto, *Universality of the least singular value for sparse random matrices*, Electronic Journal of Probability **24** (2019). MR-3916329

[35] Giorgio Cipolloni, László Erdős, and Dominik Schröder, *Optimal lower bound on the least singular value of the shifted Ginibre ensemble*, Probability and Mathematical Physics **1** (2020), no. 1, 101–146.

[36] Giorgio Cipolloni, László Erdős, and Dominik Schröder, *Edge universality for non-hermitian random matrices*, Probability Theory and Related Fields **179** (2021), no. 1, 1–28. MR-4221653

[37] Benoît Collins, *Moments and cumulants of polynomial random variables on unitary groups, the Itzykson–Zuber integral, and free probability*, International Mathematics Research Notices **2003** (2003), no. 17, 953–982. MR-1959915

[38] Nicholas Cook, *Lower bounds for the smallest singular value of structured random matrices*, The Annals of Probability **46** (2018), no. 6, 3442–3500. MR-3857860

[39] Freeman J Dyson, *A Brownian-motion model for the eigenvalues of a random matrix*, J. Math. Phys. **3** (1962), no. 6, 1191–1198. MR-0148397

[40] A. Edelman, *Eigenvalues and condition numbers of random matrices*, SIAM J. Matrix Anal. SIAM J. Matrix Anal. Appl. **9** (1988), 543–560. MR-0964668

[41] Erdős, S. Péché, J. A. Ramirez, and B. Schlein, *Bulk universality for Wigner matrices*, Comm. Pure Appl. Math. **63** (2010), no. 7, 895–925. MR-2662426

[42] L. Erdős, A. Knowles, H.-T. Yau, and J. Yin, *Spectral statistics of Erdős–Rényi graphs II: eigenvalue spacing and the extreme eigenvalues*, Comm. Math. Phys. **314** (2012), no. 3, 587–640. MR-2964770

[43] L. Erdős, A. Knowles, H.-T. Yau, and J. Yin, *Spectral statistics of Erdős–Rényi graphs I: local semicircle law*, Ann. Probab. **41** (2013), no. 3B, 2279–2375. MR-3098073

[44] L. Erdős and H.-T. Yau, *Gap universality of generalized Wigner and β-ensembles*, J. Eur. Math. **17** (2015), no. 8, 1927–2036. MR-3372074

[45] L. Erdős, H.-T. Yau, and B. Schlein, *Universality of random matrices and local relaxation flow*, Invent. Math. **185** (2011), no. 1, 75–119. MR-2810797

[46] László Erdős and Horng-Tzer Yau, *Dynamical approach to random matrix theory*, Courant Lecture Notes in Mathematics **28** (2017). MR-3699468

[47] L. Erdős, H.-T. Yau, and J. Yin, *Bulk universality for generalized Wigner matrices*, Probab. Theory Related Fields **154** (2012), no. 1-2, 341–407. MR-2981427

[48] László Erdös, José A Ramírez, Benjamin Schlein, Terence Tao, Vu Van, and Horng-Tzer Yau, *Bulk universality for Wigner Hermitian matrices with subexponential decay*, Mathematical Research Letters **17** (2010), no. 4. MR-2661171

[49] László Erdős, Benjamin Schlein, and Horng-Tzer Yau, *Universality of random matrices and local relaxation flow*, Invent. Math. **185** (2011), no. 1, 75–119. MR-2810797

[50] László Erdős, Benjamin Schlein, Horng-Tzer Yau, and Jun Yin, *The local relaxation flow approach to universality of the local statistics for random matrices*, Annales de l'I.H.P. Probabilités et statistiques **48** (2012), no. 1, 1–46. MR-2919197

[51] Ohad Feldheim and Sasha Sodin, *A universality result for the smallest eigenvalues of certain sample covariance matrices*, Geometric And Functional Analysis **20** (2010), no. 1, 88–123. MR-2647136

[52] Friedrich Götze, Alexey Naumov, and Alexander Tikhomirov, *On minimal singular values of random matrices with correlated entries*, Random Matrices: Theory and Applications **4** (2015), no. 02, 1550006. MR-3356884

[53] Jiaoyang Huang and Benjamin Landon, *Local law and mesoscopic fluctuations of Dyson Brownian motion for general β and potential*, arXiv preprint arXiv:1612.06306 (2016).

[54] Jiaoyang Huang, Benjamin Landon, and Horng-Tzer Yau, *Bulk universality of sparse random matrices*, J. Math. Phys. **56** (2015), no. 12, 123301. MR-3429490

[55] Vladislav Kargin, *A concentration inequality and a local law for the sum of two random matrices*, Probability Theory and Related Fields **154** (2012), no. 3-4, 677–702. MR-3000559

[56] Vladislav Kargin, *Subordination for the sum of two random matrices*, The Annals of Probability **43** (2015), no. 4, 2119–2150. MR-3353823

[57] Phil Kopel, Sean O'Rourke, and Van Vu, *Random matrix products: Universality and least singular values*, Annals of Probability **48** (2020), no. 3, 1372–1410. MR-4112718

[58] Benjamin Landon, Philippe Sosoe, and Horng-Tzer Yau, *Fixed energy universality of Dyson Brownian motion*, Advances in Mathematics **346** (2019), 1137–1332. MR-3914908

[59] Benjamin Landon and Horng-Tzer Yau, *Convergence of local statistics of Dyson Brownian motion*, Comm. Math. Phys. **355** (2017), 949–1000. MR-3687212

[60] Ji Oon Lee, Kevin Schnelli, Ben Stetler, and Horng-Tzer Yau, *Bulk universality for deformed Wigner matrices*, Ann. Probab. **44** (2016), no. 3, 2349–2425. MR-3502606

[61] Ji Oon Lee and Jun Yin, *A necessary and sufficient condition for edge universality of Wigner matrices*, Duke Mathematical Journal **163** (2014), no. 1, 117–173. MR-3161313

[62] Alexander Litvak and Omar Rivasplata, *Smallest singular value of sparse random matrices*, Studia Mathematica **212** (2011), no. 3. MR-3009072

[63] Galyna Livshyts, *The smallest singular value of heavy-tailed not necessarily i.i.d. random matrices via random rounding*, arXiv preprint arXiv:1811.07038 (2018).

[64] Hans Maassen, *Addition of freely independent random variables*, J. Funct. Anal. **106** (1992), 409–438. MR-1165862

[65] L Pastur and V Vasilchuk, *On the law of addition of random matrices*, Communications in Mathematical Physics **214** (2000), no. 2, 249–286. MR-1796022

[66] Elizaveta Rebrova and Konstantin Tikhomirov, *Coverings of random ellipsoids, and invertibility of matrices with i.i.d. heavy-tailed entries*, Israel Journal of Mathematics **227** (2018), no. 2, 507–544. MR-3846333

[67] Mark Rudelson, *Invertibility of random matrices: norm of the inverse*, Annals of Mathematics (2008), 575–600. MR-2434885

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

[69] Mark Rudelson and Roman Vershynin, *Smallest singular value of a random rectangular matrix*, Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences **62** (2009), no. 12, 1707–1739. MR-2569075

[70] Mark Rudelson and Roman Vershynin, *Non-asymptotic theory of random matrices: extreme singular values*, Proceedings of the International Congress of Mathematicians 2010, World Scientific, 2010, pp. 1576–1602. MR-2827856

[71] Roland Speicher, *Free convolution and the random sum of matrices*, Publications of the Research Institute for Mathematical Sciences **29** (1993), no. 5, 731–744. MR-1245015

[72] D.A. Spielman and S.H. Teng, *Smoothed analysis of algorithms*, Proceedings of the International Congress of Mathematicians 2002, Higher Ed. Press, 2002, pp. 1576–1602. MR-1989210

[73] T. Tao and V. Vu, *Random matrices: the distribution of the smallest singular value*, Geometric and Functional Analysis **20** (2010), no. 1, 260–297. MR-2647142

[74] Terence Tao and Van H Vu, *Inverse Littlewood–Offord theorems and the condition number of random discrete matrices*, Annals of Mathematics (2009), 595–632. MR-2480613

[75] Kateryna Tatarko, *An upper bound on the smallest singular value of a square random matrix*, Journal of Complexity **48** (2018), 119–128. MR-3828841

[76] Konstantin Tikhomirov, *The limit of the smallest singular value of random matrices with i.i.d. entries*, Advances in Mathematics **284** (2015), 1–20. MR-3391069

[77] Konstantin Tikhomirov, *The smallest singular value of random rectangular matrices with no moment assumptions on entries*, Israel Journal of Mathematics **212** (2016), no. 1, 289–314. MR-3504328

[78] Konstantin Tikhomirov, *Invertibility via distance for noncentered random matrices with continuous distributions*, Random Structures & Algorithms (2020). MR-4129729

[79] Roman Vershynin, *Concentration inequalities for random tensors*, Bernoulii **26** (2020), no. 4, 3139–3162. MR-4140540

[80] Dan Voiculescu, *Limit laws for random matrices and free products*, Inventiones mathematicae **104** (1991), no. 1, 201–220. MR-1094052

[81] R.M. Wilcox, *Exponential operators and parameter differentiation in quantum physics*, Journal of Mathematical Physics **8** (1967), no. 4, 962–982. MR-0234689

[82] Fan Yang and Jun Yin, *Random band matrices in the delocalized phase, III: Averaging fluctuations*, Probability Theory and Related Fields **179** (2021), no. 1, 451–540. MR-4221663