Universality of the least singular value for the sum of random matrices

Ziliang Che†  Patrick Lopatto‡

Abstract
We consider the least singular value of $M = R^*XT + U^*YV$, 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.

Keywords: random matrix theory; sparse; universality; singular value.
MSC2020 subject classifications: 60B20.
Submitted to EJP on August 13, 2019, final version accepted on March 12, 2021.

Contents

1 Introduction 2
2 Overview and main result 3
  2.1 Overview 3
  2.2 Model 4
  2.3 Main result 6
3 Definition of dynamics 6
  3.1 Unitary Brownian motion 6
  3.2 Canceling mesoscopic drift 7

*Z.C. is partially supported by NSF grant DMS-1607871. P.L. is partially supported by the NSF Graduate Research Fellowship Program under Grant DGE-1144152.
†Harvard University, United States of America. E-mail: zche@math.harvard.edu
‡Harvard University, United States of America. E-mail: lopatto@math.harvard.edu
Universality of the least singular value

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

\[ M = R^\dagger X T + U^\dagger Y V, \]  

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

4 Local law

4.1 Concentration of Green’s functions .................................. 8
4.2 Invariant identities ...................................................... 10
4.3 Asymptotic equations .................................................. 10
4.4 Weak law ............................................................... 14

5 Eigenvector estimates .............................................. 17
5.1 Global eigenvector bounds ........................................... 17
5.2 Deterministic estimates .............................................. 18
5.3 Bulk eigenvector bounds ............................................ 20

6 Well posedness of dynamics .................................... 22

7 Analysis of SDEs ..................................................... 24
7.1 Interpolating process ................................................ 25
7.2 Conclusion ............................................................ 26

A Derivation of dynamics ........................................... 28
A.1 Calculation .......................................................... 28

B Real case .............................................................. 31

C Preliminary estimates .............................................. 32
C.1 Diagonal matrices .................................................... 32
C.2 Unitary flow ........................................................ 33
C.3 Sufficient conditions for positive density ...................... 33

References .............................................................. 34
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,
Universality of the least singular value

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

$$M = R^t X T + U^t Y V,$$  \hspace{1cm} (2.1)

where $X = \text{diag}(x_1, \ldots, x_N)$ and $Y = \text{diag}(y_1, \ldots, y_N)$ 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

$$0 \leq x_i \leq C_0, \quad 0 \leq y_i \leq C_0$$  \hspace{1cm} (2.2)

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

$$\mu_X = \frac{1}{N} \sum_{1 \leq i \leq N} \delta_{x_i}, \quad \mu_Y = \frac{1}{N} \sum_{1 \leq i \leq N} \delta_{y_i}.$$  \hspace{1cm} (2.3)

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

$$x_{-i} = -x_i, \quad y_{-i} = -y_i.$$  \hspace{1cm} (2.4)

We denote the symmetrized version of the empirical measures of $X$ and $Y$ by

$$\mu_X^{\text{sym}} = \frac{1}{2N} \sum_{1 \leq |i| \leq N} \delta_{x_i}, \quad \mu_Y^{\text{sym}} = \frac{1}{2N} \sum_{1 \leq |i| \leq N} \delta_{y_i}.$$  \hspace{1cm} (2.5)

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

$$m_X(z) = \int_{\mathbb{R}} \frac{\mu_X^{\text{sym}}(dx)}{x - z}, \quad m_Y(z) = \int_{\mathbb{R}} \frac{\mu_Y^{\text{sym}}(dy)}{y - z}.$$  \hspace{1cm} (2.6)

We assume:

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

$$\sup_{E \in \mathbb{R}, \eta \geq N^{-1+a}} |m_Y(E + i\eta)| \leq C_a.$$  \hspace{1cm} (2.7)

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

$^1$For a general Borel measure $\nu$ we define $\nu^{\text{sym}}(A) = \frac{1}{2} [\nu(A) + \nu(-A)]$ for any Borel set $A \subset \mathbb{R}$.

$^2$Observe this implies $\mu_X^{\text{sym}} \rightarrow \mu_1^{\text{sym}}$ and $\mu_Y^{\text{sym}} \rightarrow \mu_2^{\text{sym}}$ weakly.
Universality of the least singular value

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

4. The Stieltjes transform of the measure $\mu_X$ converges to that of $\mu_1$ with polynomial speed, in the sense that there exists a constant $c_X > 0$ such that

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

for $\eta \geq N^{-c_X}$.

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

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

where $y^*_k$ is the $k$-th $N$-quantile of $\mu_2$ defined by

$$
y^*_k = \inf \left\{ s : \int_{-\infty}^{s} \mu_2(dy) = \frac{k}{N} \right\}.
$$

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

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

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

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

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 spectral edges of $\mu_2$. 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 $\mu_2$ behave sub-linearly, for example like the edges of the semicircle distribution $\rho_{sc}(x) = \frac{1}{\pi \sqrt{4-x^2}} dx$. 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 $\mu_1$ 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 $\mu_1^{\text{sym}}$ and $\mu_2^{\text{sym}}$ have positive density at zero, or $\mu_1^{\text{sym}} = \mu_2^{\text{sym}}$.

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.

---

3We recall the definition of free convolution in Subsection 4.4

---
We therefore find by the Lévy criterion that we introduce the index set

\[ \text{with variance } \frac{1}{c^2} \text{ defined in Section 2.2.} \]

For all \( r \geq 0 \), we have

\[ P(N \lambda_1(M_N) \leq r) = 1 - e^{-r^2} + O(N^{-c}), \quad (2.13) \]

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 \( (\sqrt{2} \text{Im } B(t), \sqrt{2} \text{Re } B(t)) \) 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 \( (\sqrt{2} \text{Im } B_{ij}, \sqrt{2} \text{Re } B_{ij})_{i < j}, \ (B_{ii})_{1 \leq i \leq N} \) are a collection of independent real standard Brownian motions, and \( B_{ij} = B_{ji} \).

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

\[ \mathcal{I}_a = \{(i, j) : |y_i - y_j| \geq N^{-1 + a}\} \quad (3.1) \]

and let \( \mathcal{I}_a \) be the set of pairs \((i, j)\) not in \( \mathcal{I}_a \).

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

\[ dU(t) = i dW_1(t) U(t) - \frac{1}{2} A U(t) dt, \quad dV(t) = i dW_2 V(t) - \frac{1}{2} A V(t) dt \quad (3.2) \]

Here \( dW_1, dW_2 \), and \( A \) are defined as follows. Let \( \tilde{W}_1 \) and \( \tilde{W}_2 \) 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

\[ (W_1)_{ij} = \frac{1}{\sqrt{2N}} \left( \frac{1}{|y_i - y_j|} (\tilde{W}_1)_{ij} + \frac{1}{y_i + y_j} (\tilde{W}_2)_{ij} \right), \quad (3.3) \]

\[ (W_2)_{ij} = \frac{1}{\sqrt{2N}} \left( \frac{1}{|y_i - y_j|} (\tilde{W}_1)_{ij} - \frac{1}{y_i + y_j} (\tilde{W}_2)_{ij} \right). \quad (3.4) \]

The diagonal matrix \( A \) in (3.2) is given by

\[ A_{ii} = \frac{1}{2N} \sum_{j:(i,j) \in \mathcal{I}_a} \left( \frac{1}{(y_i - y_j)^2} + \frac{1}{(y_i + y_j)^2} \right). \quad (3.5) \]

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

\[ (idW_1 Y - iY dW_2)_{ij} = \frac{i \mathbb{1}_{(i,j) \in \mathcal{I}_a}}{\sqrt{2N}} \left( \text{sgn}(y_j - y_i) (d(\tilde{W}_1))_{ij} + (d(\tilde{W}_2))_{ij} \right). \quad (3.6) \]

We therefore find by the Lévy criterion that \( \sqrt{N} ((idW_1 Y - iY dW_2)_{ij})_{(i,j) \in \mathcal{I}_a} \) is a family of independent standard complex Brownian motions. In particular, there is no longer a
Universality of the least singular value

Hermitian symmetry. We write

\[(i dW_1 Y - i Y dW_2)_{ij} = \frac{1_{(i,j) \in I_s}}{\sqrt{N}} d\tilde{B}_{ij}, \tag{3.7}\]

where \((d\tilde{B}_{ij})_{1 \leq i,j \leq N}\) is a family of independent standard complex Brownian motions. We choose \(\lambda\) 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

\[dM(t) = \frac{1}{\sqrt{N}} U^*(t)(1_{(i,j) \in I_s} d\tilde{B}_{ij}) V(t) + U^*(t)\hat{A}V(t) dt, \tag{3.8}\]

where \(\hat{A}\) is a diagonal matrix whose entries are given by

\[\hat{A}_{ii} = \frac{1}{2N} \sum_{j,(i,j) \in I_s} \left( \frac{1}{y_j - y_i} - \frac{1}{y_i + y_j} \right). \tag{3.9}\]

### 3.2 Canceling mesoscopic drift

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

\[\hat{M}(t) = M(t) + (\tau - t)U^*(0)\hat{A}V(0). \tag{3.10}\]

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

\[d\hat{M} = \frac{1}{\sqrt{N}} U^*(t) (1_{(i,j) \in I_s} d\tilde{B}_{ij}) V(t) + \left( U^*(t)\hat{A}V(t) - U^*(0)\hat{A}V(0) \right) dt \tag{3.11}\]

\[= \frac{1}{\sqrt{N}} d\tilde{B} + \left( U^*(t)\hat{A}V(t) - U^*(0)\hat{A}V(0) \right) dt - \frac{1}{\sqrt{N}} U^*(t) (1_{(i,j) \in I_s} d\tilde{B}_{ij}) V(t). \tag{3.12}\]

Here \(\hat{B}\) is a matrix of standard complex Brownian motions. We show in Section 7 that the second term, when integrated from 0 to \(\tau\), 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

\[
\begin{bmatrix}
0 & \hat{M}_s \\
\hat{M}_s & 0
\end{bmatrix}
\]

governed by the following system of SDEs:

\[d\lambda_i = \frac{1}{\sqrt{2N}} dB_i + \frac{1}{2N} \sum_{j \neq i} \frac{1 - \gamma_{ij}}{\lambda_i - \lambda_j} dt + R_i, \tag{3.13}\]

where

\[R_i = \text{Re} \left\langle j_{i*}, \left( U^*(t)\hat{A}V(t) - U^*(0)\hat{A}V(0) \right) \right\rangle_{k_i} \tag{3.14}\]

\[\gamma_{ij} = \frac{1}{2} \sum_{(a,b) \in I_s} \left| w_i(a) \right|^2 \left| z_j(b) \right|^2 + \sum_{(a,b) \in I_s} \left| w_j(a) \right|^2 \left| z_i(b) \right|^2, \quad w_i = U j_i, \quad z_i = V k_i. \tag{3.15}\]
Universality of the least singular value

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 \( \hat{M} = JSK^* \) 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 \( \hat{M}(t) \).

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 \( a, b \) such that
\[
0 < b < \frac{a}{100} < 10^{-4},
\]
and let \( \tau = N^{-1+b} \) denote the short time we study. Any constant \( C \) without further specification is a universal constant that may depend on \( a \) 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
\[
\mathcal{H}(t) = \begin{bmatrix} 0 & Z \\ Z^\dagger & 0 \end{bmatrix}
\]
for
\[
Z = UR^\dagger XTV^\dagger + Y + (\tau - t)UU(0)^\dagger \hat{A}V(0)V^\dagger
\]
and \( \mathcal{G} = (\mathcal{H} - z)^{-1} \). Note that the \( 2N \) eigenvalues of \( \mathcal{H}(t) \) are exactly the \( N \) singular values of \( \hat{M}(t) \) 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 \( \hat{M}(t) \) multiplied by \( U \) on the left and \( V^\dagger \) on the right.) We also define
\[
\hat{\mathcal{H}}(t) = \begin{bmatrix} 0 & UR^\dagger XTV^\dagger + Y + (\tau - t)\hat{A} \\ (UR^\dagger XTV^\dagger + Y + (\tau - t)\hat{A})^\dagger & 0 \end{bmatrix},
\]
and \( \hat{\mathcal{G}}(t) = (\hat{\mathcal{H}}(t) - z)^{-1} \). We will begin by studying \( \hat{\mathcal{H}}(t) \), since the lack of randomness in the term involving \( \hat{A} \) 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:
\[
\|A\|_{\text{HS}} = \sqrt{\text{Tr}(AA^*)}.
\]
We also recall the equivalent characterization of this norm in terms of the sum of the squares of the matrix entries:
\[
\|A\|_{\text{HS}}^2 = \sum_{i,j} |A_{ij}|^2.
\]

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

**Proposition 4.1.** Let \( g: U(N) \to \mathbb{C} \) be a Lipschitz function with Lipschitz constant \( L \) in the Hilbert–Schmidt norm:
\[
|g(X) - g(Y)| \leq L\|X - Y\|_{\text{HS}}, \quad \forall X, Y \in U(N).
\]


Universality of the least singular value

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

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

The above proposition can be applied to the Green’s function $G$, which is a smooth function of $U$, $V$, and $z$. In particular, the Lipschitz constant of $G$ 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 $1 \leq i \leq 2N$,

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

The same bounds hold for $\tilde{G}$.

**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

$$\text{Im} G_{ii} = \text{Im}(e_i, Ge_i) = \frac{1}{2i} \langle e_i, (G - G^*)e_i \rangle. \quad (4.10)$$

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

$$\text{Im} G_{ii} = \langle Ge_i, Ge_i \rangle \eta = \|Ge_i\|^2 \eta. \quad (4.11)$$

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

$$\text{Im} G_{ii} = \|Ge_i\|^2 \eta \geq \frac{\eta}{(C + |z|)^2}. \quad (4.12)$$

This proves the first inequality. The second inequality follows from the spectral theorem. □

**Proposition 4.3.** Fix $z = E + i\eta \in \mathbb{C}^+$. Then, for any $1 \leq i, j \leq 2N$, $G_{ij}$ 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 $\tilde{G}_{ij}$.

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

$$\tilde{G} - G = G \begin{pmatrix} 0 & (U - \tilde{U})R^1XTV^1 \dagger \\ (U - \tilde{U})R^1XTV^1 \dagger \\ 0 \end{pmatrix} \tilde{G}. \quad (4.13)$$

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

$$\|\tilde{G} - G\|_{\text{HS}} \leq 2\|\tilde{G}\|_{\text{HS}} \|\tilde{G}\|_{\text{HS}} \|U - \tilde{U}\|_{\text{HS}}. \quad (4.14)$$

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

$$\|\tilde{G} - G\|_{\text{HS}} \leq \frac{2C}{\eta^2} \|U - \tilde{U}\|_{\text{HS}}, \quad (4.15)$$

and the conclusion follows. □
4.2 Invariant identities

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

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

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

For brevity we set $\hat{\Upsilon} = \Upsilon + (\tau - t)\hat{A}$, and we define

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

so that $\hat{H} = \mathcal{H}_1 + \mathcal{H}_2$. 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

$$E \left[ \hat{G}\mathcal{H}_1 E_{ij} \hat{G} \right] = E \left[ \hat{G}E_{ij} \mathcal{H}_1 \hat{G} \right]. \quad (4.18)$$

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

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

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

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

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

$$\partial_\zeta G(\zeta, z)|_{\zeta = 0} = \hat{G} \begin{bmatrix} 0 \\ - (UR^*XTV^\dagger)^\dagger E_{ij} \end{bmatrix} E_{ij} E_{ij}^\dagger \hat{G}. \quad (4.21)$$

Note that the distribution of $Q(\zeta)U$ is invariant with respect to $\zeta$, so $E[\partial_\zeta G(\zeta, z)] = 0$.

Therefore, the above equality yields

$$E \left[ \hat{G} \begin{bmatrix} 0 \\ - (UR^*XTV^\dagger)^\dagger E_{ij} \end{bmatrix} E_{ij} E_{ij}^\dagger \hat{G} \right] = 0. \quad (4.22)$$

This can also be rearranged as

$$E \left[ \hat{G}E_{ij}\mathcal{H}_1 \hat{G} \right] = E \left[ \hat{G}E_{ij}E_{ij}\hat{G} \right]. \quad (4.23)$$

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. \qed

4.3 Asymptotic equations

Let $m_N(z)$ be the Stieltjes transform of $\hat{H}$. 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 $X = O(Y)$ if there are $c_1, c_2, c_3, N_0 > 0$ that do not depend on $N$ such that for $N > N_0$,

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

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

$$\mathbb{P}[\exists \alpha \in A : |X(\alpha, N)| \geq c_1 Y(\alpha, N)] \leq \exp (-c_2 N^{c_3}). \quad (4.25)$$
Universality of the least singular value

Define

\[ w_X = -\frac{\text{Tr}(\hat{\mathcal{G}}_1)}{\text{Tr} \hat{\mathcal{G}}}, \quad w_{\mathcal{F}} = -\frac{\text{Tr}(\hat{\mathcal{G}}_2)}{\text{Tr} \hat{\mathcal{G}}}, \tag{4.26} \]

which are holomorphic functions on \( C^+ \).

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

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

uniformly for \( 1 \leq i, k \leq 2N \).

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

\[ (\hat{\mathcal{G}}_1)_{kl} \hat{\mathcal{G}}_{ji} = \hat{\mathcal{G}}_{ki}(\mathcal{G}_1 \hat{\mathcal{G}})_{ji} + O\left( \eta^{-3} N^{-\frac{1}{2}+a} \right). \tag{4.28} \]

We used here our assumptions on \( X \) to conclude that \( \|\mathcal{H}_1\| \leq C \).

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

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

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

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

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

\[ \frac{1}{N} \sum_{j=1}^{N} \hat{\mathcal{G}}_{jj} = \frac{1}{2N} \text{Tr} \hat{\mathcal{G}} + O\left( \eta^{-2} N^{-\frac{1}{2}+a} \right). \tag{4.31} \]

Similarly, we have

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

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

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

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 \( \hat{\mathcal{G}} \)) to conclude that

\[ |\text{Tr}(\mathcal{H}_1 \hat{\mathcal{G}})| \leq \|\mathcal{H}_1\| \|\hat{\mathcal{G}}\|_1 \leq CN\eta^{-1}. \tag{4.34} \]
Universality of the least singular value

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
\[
(\hat{G}H_1)_{ki} = \hat{G}_{ki} \left( -w_X + O\left( \eta^{-5}N^{-\frac{1}{2}+a}\right) \right) + O\left( \eta^{-4}N^{-\frac{1}{2}+a}\right), \quad \forall 1 \leq i \leq N, 1 \leq k \leq 2N,
\] (4.35)
where the extra power of $\eta^{-1}$ in the second error term comes from Proposition 4.2 after taking the quotient. This proves the conclusion for $1 \leq i \leq N$. Similarly, we can take $i \in [N+1, 2N]$ to obtain the same identity for $N + 1 \leq i \leq 2N, 1 \leq k \leq 2N$. This proves the conclusion for fixed $i$ and $k$. The constants in the $O$ notation are uniform in $i$ and $k$. \hfill \Box

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 $\|AR\| \leq \delta < 1/2$ and
\[
A(B + R) = I.
\] (4.36)

Then
\[
A = B^{-1} + O(\delta \|A\|).
\] (4.37)

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

**Proof.** We immediately have $B = A^{-1}(I - AR)$. Hence
\[
B^{-1} = (I - AR)^{-1}A = \left( \sum_{k=0}^{\infty} (AR)^k \right) A.
\] (4.38)

By the assumption that $\|AR\| < 1/2$, we have $\|\sum_{k=0}^{\infty} (AR)^k\| \leq 2$. Hence,
\[
\|B^{-1}\| \leq 2 \|A\|.
\] (4.39)

On the other hand, $A(B + R) = I$ implies
\[
A = B^{-1} - ARB^{-1}.
\] (4.40)

Note that $\|ARB^{-1}\| \leq \delta \|B^{-1}\| \leq 2\delta \|A\|$. This gives
\[
A = B^{-1} + O(\delta \|A\|).
\] (4.41) \hfill \Box

Now we are ready to prove that the three holomorphic functions $m_N, w_X, \text{ and } w_T$ 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 $\mu_X^{\text{sym}}$, and we let $w_T$ denote the Stieltjes transform of $\mu_T^{\text{sym}}$.

**Lemma 4.8.** Fix $a > 0$. For any $z \in \mathbb{C}^+$ with $|z| \leq \log N$ and $\eta > N^{-1/10+a}$, we have
\[
\begin{align*}
\left\{ \begin{array}{l}
m_N(z) = m_X(z + w_T) + O\left( \eta^{-6}N^{-\frac{1}{2}+a}\right), \\
m_N(z) = m_T(z + w_X) + O\left( \eta^{-6}N^{-\frac{1}{2}+a}\right), \\
\frac{1}{m_N(z)} = -z - w_X - w_T,
\end{array} \right. \tag{4.42}
\end{align*}
\]
and for $1 \leq k \leq N$,
\[
\begin{align*}
\hat{G}_{kk} &= \frac{1}{2} \left( \frac{1}{\overline{Y}_k - z - w_X} + \frac{1}{-\overline{Y}_k - z - w_X} \right) + O\left( \eta^{-4}N^{-\frac{1}{2}+a}\right), \\
\hat{G}_{N+k,N+k} &= \frac{1}{2} \left( \frac{1}{\overline{Y}_k - z - w_X} + \frac{1}{-\overline{Y}_k - z - w_X} \right) + O\left( \eta^{-4}N^{-\frac{1}{2}+a}\right). \tag{4.43}
\end{align*}
\]
Universality of the least singular value

Proof. We start with the following identity, which is equivalent to the definition of \( \hat{G} \).

\[
\hat{G} \begin{bmatrix} 0 & Y \\ Y & 0 \end{bmatrix} + \hat{G}H_1 - z\hat{G} = I \tag{4.44}
\]

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

\[
Y_k \begin{bmatrix} \hat{G}_{k,k} & \hat{G}_{k,N+k} \\ \hat{G}_{N+k,k} & \hat{G}_{N+k,N+k} \end{bmatrix} + \begin{bmatrix} (\hat{G}H_1)_{kk} & (\hat{G}H_1)_{k,N+k} \\ (\hat{G}H_1)_{N+k,k} & (\hat{G}H_1)_{N+k,N+k} \end{bmatrix}
- z \begin{bmatrix} \hat{G}_{kk} & \hat{G}_{k,N+k} \\ \hat{G}_{N+k,k} & \hat{G}_{N+k,N+k} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \tag{4.45}
\]

or equivalently,

\[
\begin{bmatrix} \hat{G}_{k,k} & \hat{G}_{k,N+k} \\ \hat{G}_{N+k,k} & \hat{G}_{N+k,N+k} \end{bmatrix} \begin{bmatrix} -z & Y_k \\ Y_k & -z \end{bmatrix} + \begin{bmatrix} (\hat{G}H_1)_{kk} & (\hat{G}H_1)_{k,N+k} \\ (\hat{G}H_1)_{N+k,k} & (\hat{G}H_1)_{N+k,N+k} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}. \tag{4.46}
\]

We apply Corollary 4.6 to the matrix involving \( \hat{G}H_1 \) to get

\[
\begin{bmatrix} \hat{G}_{k,k} & \hat{G}_{k,N+k} \\ \hat{G}_{N+k,k} & \hat{G}_{N+k,N+k} \end{bmatrix} \begin{bmatrix} -z - w_X & Y_k \\ Y_k & -z - w_X \end{bmatrix} + \mathcal{O}(\eta^{-5}N^{-\frac{1}{2} + a})
+ \mathcal{O}(\eta^{-4}N^{-\frac{1}{2} + a}) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}. \tag{4.47}
\]

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

\[
\begin{bmatrix} \hat{G}_{k,k} & \hat{G}_{k,N+k} \\ \hat{G}_{N+k,k} & \hat{G}_{N+k,N+k} \end{bmatrix} = \begin{bmatrix} -z - w_X & Y_k \\ Y_k & -z - w_X \end{bmatrix}^{-1} \left( I + \mathcal{O}(\eta^{-6}N^{-\frac{1}{2} + a}) \right). \tag{4.48}
\]

This can be written explicitly as

\[
\begin{bmatrix} \hat{G}_{k,k} & \hat{G}_{k,N+k} \\ \hat{G}_{N+k,k} & \hat{G}_{N+k,N+k} \end{bmatrix} = \begin{bmatrix} \frac{1}{2} \left( \frac{1}{Y_k - z - w_X} + \frac{1}{Y_k - z - w_X} \right) \\ \frac{1}{2} \left( \frac{1}{Y_k - z - w_X} + \frac{1}{Y_k - z - w_X} \right) \end{bmatrix} + \mathcal{O}(\eta^{-6}N^{-\frac{1}{2} + a}), \tag{4.49}
\]

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

\[
\frac{1}{2N} \text{Tr} \hat{G} = m_{\mathcal{F}}(z + w_X) + \mathcal{O}(\eta^{-4}N^{-\frac{1}{2} + a}). \tag{4.50}
\]

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

Finally, the third equation follows from

\[
\text{Tr}(\hat{G}H_1) + \text{Tr}(\hat{G}H_2) - z \text{Tr} \hat{G} = 2N, \tag{4.51}
\]

which follows from the definition of \( \hat{G} \). \( \Box \)
4.4 Weak law

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

We deal with equation (4.42) in a general setting. Let \( m_\alpha, m_\beta \) be the Stieltjes transforms of probability measures \( \mu_\alpha, \mu_\beta \). Consider the following deterministic equations for fixed \( z \in \mathbb{C}^+ \).

\[
\begin{aligned}
 m &= m_\alpha(z + w_\alpha) \\
 m &= m_\beta(z + w_\beta) \\
 \frac{1}{m} &= -z - w_\alpha - w_\beta
\end{aligned}
\]  

(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 \( \mu_\alpha \oplus \mu_\beta \) given by the following proposition the free convolution of \( \mu_\alpha \) and \( \mu_\beta \).

**Proposition 4.9.** Given two probability measures \( \mu_\alpha \) and \( \mu_\beta \) on \( \mathbb{R} \), there exists unique analytic functions \( w_\alpha, w_\beta, m: \mathbb{C}^+ \rightarrow \mathbb{C}^+ \) satisfying (4.52), where \( m \) is the Stieltjes transform of a probability measure we denote \( \mu_\alpha \oplus \mu_\beta \). Further, suppose \( \mu_\alpha \oplus \mu_\beta \) has a density on an interval \( I \subset \mathbb{R} \) that is bounded away from zero. If \( \mu_\alpha, \mu_\beta \) are compactly supported, and none of them is a point mass, then \( w_\alpha, w_\beta \) 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 \( \mu_\alpha \oplus \mu_\beta \) 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 \( w_\alpha \) and \( w_\beta \) to take the value \( \infty \), but our hypothesis on the density of \( \mu_\alpha \oplus \mu_\beta \) on \( I \) rules this out. Indeed, [21, Theorem 2.3] shows that \( m \) also extends continuously to \( I \) (again with values that may be \( \infty \)), and this hypothesis shows that \( \Im m(a) > c \) uniformly for \( a \in I \) and some constant \( c > 0 \). The uniform boundedness claim then follows from the last equation of (4.52).

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 \( w_\alpha \) and \( w_\beta \) only. Define

\[
\hat{m}_\alpha(\zeta) = -\zeta - \frac{1}{m_\alpha(\zeta)}, \quad \hat{m}_\beta(\zeta) = -\zeta - \frac{1}{m_\beta(\zeta)}.
\]  

(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 \( \sigma^2_\alpha := \int t^2 \mu_\alpha(\mathbb{R}) \) and \( \sigma^2_\beta := \int t^2 \mu_\beta(\mathbb{R}) \) respectively (which are \( \leq C \) by assumption (2.2)).

After eliminating \( m \), equation (4.52) becomes

\[
\begin{aligned}
 w_\alpha &= \hat{m}_\beta(z + w_\beta) \\
 w_\beta &= \hat{m}_\alpha(z + w_\alpha).
\end{aligned}
\]  

(4.54)

Define a function \( \Phi: (\mathbb{C}^+)^2 \rightarrow (\mathbb{C}^+)^2 \) by

\[
\Phi(\zeta_1, \zeta_2) = (\zeta_1 - \hat{m}_\beta(z + \zeta_2), \zeta_2 - \hat{m}_\alpha(z + \zeta_1)).
\]  

(4.55)
Universality of the least singular value

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

$$\Phi(w_\alpha, w_\beta) = (0, 0).$$

(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:

$$\Phi(w_\alpha', w_\beta') = (r_1, r_2).$$

(4.57)

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

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

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

(4.58)

**Proof.** Note that

$$D\Phi(z_1, z_2) = \left[ \begin{array}{c} 1 \\ -q \end{array} \right].$$

(4.59)

To simplify notation, we define

$$p = \int_R \frac{\tilde{\mu}_\beta(dx)}{(x - z - w_\beta)^2}, \quad q = \int_R \frac{\tilde{\mu}_\alpha(dx)}{(x - z - w_\alpha)^2}.$$  

(4.60)

and

$$\tilde{p} = \int_R \frac{\tilde{\mu}_\beta(dx)}{|x - z - w_\beta|^2}, \quad \tilde{q} = \int_R \frac{\tilde{\mu}_\alpha(dx)}{|x - z - w_\alpha|^2}.$$  

(4.61)

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

$$D\Phi(w_\alpha, w_\beta) = \left[ \begin{array}{c} 1 \\ -q \end{array} \right].$$

(4.62)

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

$$(D\Phi(w_\alpha, w_\beta))^{-1} = \frac{1}{1 - pq} \left[ \begin{array}{c} 1 \\ p \end{array} \right].$$

(4.63)

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

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

(4.64)

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

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

(4.65)

Therefore,

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

(4.66)

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

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

(4.67)

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

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

(4.68)
Universality of the least singular value

**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
\[
\Phi(w_1, w_2) = (r_1, r_2) = r.
\]
Define \( \delta w = (w_1 - w_\alpha, w_2 - w_\beta) \). There exists \( c > 0 \) such that if \( \|\delta w\|_2 \leq c(\eta^3 \land \eta^7) \), then
\[
\|\delta w\|_2 \leq C(1 \lor \eta^{-4})\|r\|_2.
\]

**Proof.** Using (4.59), it is straightforward to see that for fixed \( z = E + i\eta \),
\[
\|D\Phi(\zeta_1, \zeta_2)\|_\infty \leq C(1 \lor \eta^{-2}), \quad \|D^2\Phi(\zeta_1, \zeta_2)\|_\infty \leq C\eta^{-3},
\]
for any \( (\zeta_1, \zeta_2) \in (\mathbb{C}^+)^2 \). By Taylor expansion at \( (w_\alpha, w_\beta) \), we have
\[
\|r - D\Phi(w_\alpha, w_\beta)\delta w\|_2 \leq \sup_{(\zeta_1', \zeta_2') \in (\mathbb{C}^+)^2} 4\|D^2\Phi(\zeta_1', \zeta_2')\|_\infty \|\delta w\|_2 \leq \frac{C}{\eta^3}\|\delta w\|_2^2.
\]
Using (4.72) and the bound (4.58), we have
\[
\|(D\Phi(w_\alpha, w_\beta))^{-1}r - \delta w\| \leq \|(D\Phi(w_\alpha, w_\beta))^{-1}\|\|r - D\Phi(w_\alpha, w_\beta)\delta w\|_2 \leq C\eta^{-3}(1 \lor \eta^{-4})\|\delta w\|_2^2.
\]
Hence,
\[
\|\delta w\|_2 \leq \|(D\Phi(w_\alpha, w_\beta))^{-1}\|_2 + \|(D\Phi(w_\alpha, w_\beta))^{-1}r - \delta w\|_2 \leq C(1 \lor \eta^{-4})\|r\|_2 + C\eta^{-3}(1 \lor \eta^{-4})\|\delta w\|_2^2.
\]
Using the condition that \( \|\delta w\| \leq c(\eta^3 \land \eta^7) \) and choosing \( c > 0 \) small enough, the second term on the right side can be absorbed into the left side. Thus,
\[
\|\delta w\|_2 \leq C(1 \lor \eta^{-4})\|r\|_2.
\]

Let \( K > 0 \) be a constant smaller than the constant \( c \) in the assumption (2.12). Define the spectral domain
\[
\Sigma = \left\{ z = E + i\eta: \eta \in [N^{-1/100}, 1], E \in [-2K, 2K] \right\},
\]
and let \( \tilde{\omega}_X \) and \( \tilde{\omega}_\Gamma \) solve
\[
\begin{cases}
\tilde{\omega}_X = \tilde{m}_X(z + \tilde{\omega}_\Gamma) \\
\tilde{\omega}_\Gamma = \tilde{m}_\Gamma(z + \tilde{\omega}_X).
\end{cases}
\]

**Corollary 4.12.** For any \( z \in \Sigma \) and \( t \in [0, \tau] \),
\[
|w_X - \tilde{\omega}_X| \lor |w_\Gamma - \tilde{\omega}_\Gamma| = \mathcal{O}\left(\eta^{-12}N^{-1/3}\right).
\]

**Proof.** We restrict the following claims to \( z \in \Sigma \). Multiplying the first and third equations of (4.42) gives
\[
1 = \left( m_X(z + w_\Gamma) + \mathcal{O}\left(\eta^{-6}N^{-\frac{1}{2}+\alpha}\right) \right)(-z - w_X - w_\Gamma).
\]
By Proposition 4.2, \( \text{Im } m_N(z) \geq c\eta, \) so
\[
| -z - w_X - w_\Gamma | \leq \frac{1}{|m_N(z)|} \leq C\eta^{-1},
\]
\[
1 = m_X(z + w_\Gamma)(-z - w_X - w_\Gamma) + \mathcal{O}\left(\eta^{-7}N^{-\frac{1}{2}+\alpha}\right).
\]
Universality of the least singular value

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

\[
\frac{1}{|m_X(z + w)\rangle} \leq C\eta^{-1}.
\]  

(4.82)

Dividing (4.81) by \(m_X(z + w)\rangle\) and rearranging yields

\[
\hat{m}_X(z + w)\rangle = w_X + O(\eta^{-8}N^{-\frac{1}{2}+a}).
\]  

(4.83)

Analogously,

\[
\hat{m}_Y(z + w)\rangle = w_T + O(\eta^{-8}N^{-\frac{1}{2}+a}).
\]  

(4.84)

The claim now follows from choosing \(a = 1/6\) and applying Proposition 4.11.

\[\square\]

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

\[
\sup_{(z,t) \in \Sigma \times [0,\tau]} \left| G_{ii} - \frac{1}{2} \left( \frac{1}{\nu_{i} - z - w_{X}} + \frac{1}{-\nu_{i} - z - w_{X}} \right) \right| \leq C\eta^{-14}N^{-1/3},
\]  

(4.85)

where the \(i\) in \(\nu\) is taken modulo \(N\).

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

\[
\hat{G}_{ii} = \frac{1}{2} \left( \frac{1}{\nu_{i} - z - w_{X}} + \frac{1}{-\nu_{i} - z - w_{X}} \right) + O(\eta^{-6}N^{-\frac{1}{2}+a}).
\]  

(4.86)

Because

\[
\left| \frac{1}{\nu_{i} - z - w_{X}} - \frac{1}{-\nu_{i} - z - w_{X}} \right| \leq \frac{|w_X - \bar{w}_X|}{\eta^2},
\]  

(4.87)

we obtain

\[
\hat{G}_{ii} = \frac{1}{2} \left( \frac{1}{\nu_{i} - z - w_{X}} + \frac{1}{-\nu_{i} - z - w_{X}} \right) + O(\eta^{-14}N^{-1/3}).
\]  

(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 \(\hat{G}\) to \(\hat{G}\) by using the resolvent expansion and Ward identity, again as in the proof of [33, Theorem 3.16]. \[\square\]

**5 Eigenvector estimates**

**5.1 Global eigenvector bounds**

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

\[
\sup_{0 \leq t \leq \tau} \max_{1 \leq \eta \leq N} \left| w_{\alpha}(i) \right|^2 + \left| z_{\alpha}(i) \right|^2 \leq N^{-1/p}, \quad \sup_{0 \leq t \leq \tau} \max_{\beta \neq \alpha} \gamma_{\alpha\beta} \leq N^{-2/p + a}.
\]  

(5.1)

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

\[
\text{Im} G_{\alpha\alpha}(z) = \eta \sum_{i=1}^{N} \left| w_{\alpha}(i) \right|^2 \left| \lambda_{i} - E \right|^{2} + \eta \sum_{i=1}^{N} \left| z_{\alpha}(i) \right|^2 \left| -\lambda_{i} - E \right|^{2} + \eta^{2}.
\]  

(5.2)
Universality of the least singular value

Setting $E = \lambda_i$, we obtain
\[ |w_\alpha(i)|^2 \leq \eta \Im \mathcal{G}_{\alpha\alpha}(\lambda_i + i\eta). \tag{5.3} \]
Denote $z_i = \lambda_i + i\eta$. By the local law for $\mathcal{G}_{\alpha\alpha}$, Corollary 4.13, it suffices to lower bound
\[ | - z_i - Y_\alpha | \Re \tilde{w}_X \vee \Im \tilde{w}_X. \tag{5.4} \]
and
\[ | - z_i + Y_\alpha | \Re \tilde{w}_X \vee \Im \tilde{w}_X. \tag{5.5} \]
By Corollary 5.5, proved in the next subsection,
\[ | - z_i + Y_\alpha | \Re \tilde{w}_X \vee \Im \tilde{w}_X \geq \eta^{1-c/4}, \quad | - z_i - Y_\alpha | \Re \tilde{w}_X \vee \Im \tilde{w}_X \geq \eta^{1-c/4}. \tag{5.6} \]
Therefore, using Corollary 4.13, we have
\[ |w_\alpha(i)|^2 \leq \eta^{c/4} + C\eta^{-11} N^{-1/3} \tag{5.7} \]
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. \hfill \Box

5.2 Deterministic estimates

Let $m_1, m_2$ be the Stieltjes transforms of $\mu_1^{\text{sym}}, \mu_2^{\text{sym}}$, 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 $\Im z \geq N^{-1/p}$, then
\[ |m_1(z) - m_X(z)| \leq N^{-1/p}, \quad |m_2(z) - m_T(z)| \leq N^{-1/p} \tag{5.9} \]
for all $z$.

Given $p > 10$, define the spectral domain
\[ \Sigma_p = \left\{ z = E + i\eta : \eta \in [N^{-1/p^2}, 1], E \in [-2K, 2K] \right\}, \tag{5.10} \]

Corollary 5.3. There exists a universal constant $p > 0$ such that for $z \in \Sigma_p$,
\[ |w_1 - \tilde{w}_X| \vee |w_2 - \tilde{w}_T| \leq N^{-1/p} \eta^{-p}, \quad |\tilde{w}_X| \vee |\tilde{w}_T| \leq C. \tag{5.11} \]

Proof. This follows from Proposition 4.11 with
\[ \delta w = \left( \frac{1}{m_1(z + \tilde{w}_T)} - \frac{1}{m_X(z + \tilde{w}_T)} \right) \right\) \tag{5.12} \]
We indicate how to bound the first coordinate; the second is analogous. From Proposition 5.2,
\[ |\delta w_1| \leq \frac{N^{-1/p}}{m_1(z + \tilde{w}_T) m_X(z + \tilde{w}_T)}. \tag{5.13} \]
For any Stieltjes transform $m(z)$ of a measure $\mu$,
\[ \Im m(z) \geq \frac{\Im z}{(|z| + \sup_{x \in \supp \mu} |x|)^2} \tag{5.14} \]
Recall from Proposition 4.9 that $\text{Im} \tilde{w}_X \land \text{Im} \tilde{w}_Y > 0$, and $|\hat{m}_X(z)| \leq C\eta^{-1}$ since $\hat{m}_X$ has finite mass. Taking $p$ large and using $|\tilde{w}_X| \lor |\tilde{w}_Y| \leq C\eta^{-1}$, which follows from (4.77) and the preceding comment, this shows

$$|\delta w_1| \leq CN^{-1/p}\eta^{-6} \leq C\eta.$$  
(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. □

Define $\tilde{m}(z)$ to be the Stieltjes transform of $\mu_X^\text{sym} \boxplus \mu_Y^\text{sym}$. We recall this means that $w_X, w_Y, m_X, m_Y, \tilde{m}$ satisfy (4.52).

**Corollary 5.4.** Under the same assumptions as Corollary 5.3,

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

**Proof.** As noted in Section 2.2, we assume that either $\mu_1$ or $\mu_2$ has a bounded Stieltjes transform. By Proposition 5.2, for $\eta \geq N^{-1/p}$,

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

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

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 $N \geq N_0$,

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

**Proof.** We first show

$$|z - \overline{Y}_\alpha + \text{Re} \tilde{w}_X| \lor \text{Im} \tilde{w}_X \geq \eta^{1-\epsilon/4}. \tag{5.19}$$

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

Recall

$$\tilde{w}_X = -z - \tilde{w}_Y - \frac{1}{m_X(z + \tilde{w}_X)}.$$  
(5.20)

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

$$\text{Im} \tilde{w}_X + \text{Im} \tilde{w}_Y = -\eta + \frac{\text{Im} m_X(z + \tilde{w}_Y)}{|m_X(z + \tilde{w}_Y)|^2} \geq -\eta + c \text{Im} m_X(z + \tilde{w}_Y)$$  
(5.21)

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

$$\text{Im} m_X(z + \tilde{w}_Y) = \frac{1}{2N} \sum_{\beta = -N}^{N} \frac{\eta + \text{Im} \tilde{w}_Y}{|z + \overline{Y}_\beta + \text{Re} \tilde{w}_X|^2 + |\eta + \text{Im} \tilde{w}_X|^2} \geq \frac{\eta + \text{Im} \tilde{w}_Y}{2N} \sum_{|\overline{Y}_\beta - \overline{Y}_\alpha| \leq \eta} \frac{1}{|z + \overline{Y}_\beta + \text{Re} \tilde{w}_X|^2 + |\eta + \text{Im} \tilde{w}_X|^2} \geq \frac{I(\alpha, \eta)}{2N} \frac{\eta}{4\eta^{2-\epsilon/2}} \geq \frac{I(\alpha, \eta)}{8N}. \tag{5.22}$$

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

$$I(\alpha, \eta) \geq |\{\beta : |y_\beta - y_\alpha| \leq 2\eta/3\}| \geq \frac{N}{2} \mu_2([y_\alpha - \eta/2, y_\alpha + \eta/2]) \geq cN\eta^{2-\epsilon}. \tag{5.23}$$
Universality of the least singular value

This implies \( \text{Im} m_X(z + \tilde{\omega}_\tau) \geq c\eta^{1-t^2} \), from which we conclude using the above work that

\[
\text{Im} \tilde{\omega}_X + \text{Im} \tilde{\omega}_\tau \geq 2\eta^{1-t^3} \tag{5.24}
\]

when \( N \) is large enough. By the assumption that \( \text{Im} \tilde{\omega}_X \leq \eta^{1-t^4} \), this implies \( \text{Im} \tilde{\omega}_\tau \geq \eta^{1-t^3} \). From \( \tilde{\omega}_X = \hat{m}_X(z + \tilde{\omega}_\tau) \) and Corollary 5.3, we obtain

\[
\text{Im} \tilde{\omega}_X = \int_\mathbb{R} \frac{(\eta + \text{Im} \tilde{\omega}_\tau)\mu_X(dx)}{|x-z+\tilde{\omega}_\tau|^2} \geq \eta^{1-t^3} \int_\mathbb{R} \frac{\hat{\mu}_X(dx)}{|x-z+\tilde{\omega}_\tau|^2} \tag{5.25}
\]

\[
\geq \eta^{1-t^3} \int_\mathbb{R} \frac{\hat{\mu}_X(dx)}{c(x^2+1)} \geq c\eta^{1-t^3} \text{Im} \hat{m}_X(i). \tag{5.26}
\]

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

Finally, an analogous argument shows \( | - z - Y_\alpha + \text{Re} \tilde{\omega}_X | \lor \text{Im} \tilde{\omega}_X \geq \eta^{1-t^4} \), 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

\[
D_I = \{ z = E + i\eta; E \in I, N^{-1+\nu} \leq \eta \leq 1 \}. \tag{5.27}
\]

Then

\[
\inf_{z \in D_I} \text{Im} \tilde{\omega}_X \geq c, \tag{5.28}
\]

and it holds with overwhelming probability that

\[
\sup_{0 \leq t \leq \tau} \max_{z \in D_I, 1 \leq i \leq 2N} \left| G_{ii}(z,t) - \frac{z + \tilde{\omega}_X}{(y_i + (\tau - t)\hat{A}_{ii})^2 - (z + \tilde{\omega}_X)^2} \right| \leq \frac{N^\nu}{\sqrt{N\eta}}, \tag{5.29}
\]

where the indices in \( y_i \) and \( \hat{A}_{ii} \) are taken modulo \( N \), and we require \( N \geq N_0 \) for some \( N_0 \) depending on \( \nu \).

**Proof.** By assumption, the empirical measure of \( U R^t X T V^t \) converges to \( \mu_1 \) weakly. Using \( \tau = o(1) \) and (C.10), \( (\tau - t)\hat{A} \) is negligible and \( Y + (\tau - t)\hat{A} \) converges to \( \mu_2 \) weakly. Fix a small \( \sigma > 0 \). Then by Theorem 4.4 of [16], for any fixed \( t \),

\[
\sup_{z \in D_I} \max_{1 \leq i \leq N} \left| G_{ii}(z,t) - \frac{z + \tilde{\omega}_X}{(y_i + (\tau - t)\hat{A}_{ii})^2 - (z + \tilde{\omega}_X)^2} \right| \leq \frac{N^\sigma}{\sqrt{N\eta}} \tag{5.30}
\]

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

This estimate may then be transferred to \( G_{ii} \), 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 \( N \geq N_0 \):

\[
\sup_{0 \leq t \leq \tau} \max_{\lambda_\alpha \in I} |w_\alpha(i)| + |z_\alpha(i)| \leq \frac{N^\nu}{\sqrt{N}}, \quad \sup_{0 \leq t \leq \tau} \max_{\lambda_\alpha \in I} \max_{\beta \neq \alpha} \gamma_{\alpha\beta} \leq \frac{N^{a+\nu}}{N}. \tag{5.31}
\]
Universality of the least singular value

**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 $(w_\alpha, z_\alpha)$ and $(-w_\alpha, z_\alpha)$.

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

**Theorem 5.8.** Let $I$ and $D_I$ be as in Theorem 5.6. There exist constants $q > 0$ and $N_0 > 0$ such that

$$\sup_{z \in D_I} |w_1 - \tilde{w}_X| \vee |w_2 - \tilde{w}_X| \leq N^{-1/q}, \quad \sup_{z \in D_I} |m_3(z) - \tilde{m}(z)| \leq N^{-1/q}$$

(5.32)

for $N \geq N_0$.

**Proof.** Let $q > 0$ be a constant to be determined later. We define $\Sigma_1 \subset D_I$ by

$$\Sigma_1 = \left\{ z \in D_I : |w_1 - \tilde{w}_X| \vee |w_2 - \tilde{w}_X| \leq N^{-1/q} \right\}.$$  \hspace{1cm} (5.33)

By Corollary 5.3, $\Sigma_1$ is nonempty. Because the functions involved in its definition are continuous, it is closed. Therefore to show $\Sigma_1 = D_I$, it suffices to show $\Sigma_1$ is open in $D_I$.

Because the density of $\mu_1^{\text{sym}} \oplus \mu_2^{\text{sym}}$ is bounded above and below by positive constants, there exists $c > 0$ such that $c^{-1} \geq \text{Im} m_3 \geq c$. 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

$$\text{Im} w_2 \geq c \text{Im} w_1, \quad \text{Im} w_1 \geq c \text{Im} w_2, \quad \text{Im} w_1 + \text{Im} w_2 \geq c > 0,$$

(5.34)

which implies $\text{Im} w_1 \wedge \text{Im} w_2 \geq c > 0$.\footnote{We also need $\eta = O(N^{-\epsilon})$ for some $\epsilon > 0$ 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.}

These lower bounds permit the use of Proposition 5.2 to conclude that on $\Sigma_1$,

$$|m_1(z + \tilde{w}_X) - m_X(z + \tilde{w}_X)| \vee |m_2(z + \tilde{w}_X) - m_{\tilde{w}_X}(z + \tilde{w}_X)| \leq 2N^{-1/p}.$$  \hspace{1cm} (5.35)

Therefore, using the definition (4.53) and the lower bound $\text{Im} m_3(z) \wedge \text{Im} m_{\tilde{w}_X}(z) \geq c\eta$ 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

$$|\tilde{m}_1(z + \tilde{w}_X) - m_X(z + \tilde{w}_X)| \vee |\tilde{m}_2(z + \tilde{w}_X) - m_{\tilde{w}_X}(z + \tilde{w}_X)| \leq CN^{-1/p}.$$  \hspace{1cm} (5.36)

We now claim that on $D_I$ the stability of the system of equations (5.8) is improved, so that the operator $\Phi$ from (4.55) satisfies

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

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

$$\text{Im} w_1 \wedge \text{Im} w_2 \geq c > 0,$$

(5.38)

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

$$\sup_{\text{Im} z \geq 0} |pq| < \sup_{\text{Im} z \geq 0} |\tilde{p}\tilde{q}| \leq 1,$$

(5.39)

which holds because $\tilde{m}_2$ is not a point mass and $\text{Im} w_1 \wedge \text{Im} w_2 \geq c > 0$. Because $p, q$ are continuous, we find $|1 - pq| \geq c > 0$ on $I$. Repeating (4.63) and (4.64) with these improved bounds proves the claim.
Similar reasoning gives \( \|D\Phi(\zeta_1, \zeta_2)\|_\infty < C, \|D^2\Phi(\zeta_1, \zeta_2)\|_\infty \leq C \) on \( \Sigma_1 \). We can therefore repeat the reasoning of the proof of Proposition 4.11 to show that for any \( z \in \Sigma_1 \),

\[ |w_1 - \tilde{w}_X| \vee |w_2 - \tilde{w}_Y| \leq CN^{-1/p}. \]

The remainder term \( r \) in that proof is now bounded using (5.36). Therefore, there is a neighborhood of \( z \) such that \( |w_1 - \tilde{w}_X| \vee |w_2 - \tilde{w}_Y| \leq N^{-1/q} \) 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 \( \text{Im} w_2 \), and Proposition 5.2,

\[ |m_3(z) - \tilde{m}(z)| \leq |m_1(z + \tilde{w}_Y) - m_1(z + \tilde{w}_X)| + |m_1(z + \tilde{w}_Y) - m_1(z + w_2)| \leq N^{-1/q}. \]

To bound the second term in the sum, we used \( \text{Im}(z + \tilde{w}_Y) \wedge \text{Im}(z + w_2) \geq c \) and the fact that \( |\partial_1 m_1(z)| \leq C \eta^{-2} \). This completes the proof.

**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,

\[ \sup_{z \in \mathcal{D}_I} \left| \frac{1}{N} \text{Tr} \left( \frac{1}{H(0) - z} \right) - m_3(z) \right| \leq N^{-1/p} + \frac{N^\nu}{\sqrt{\nu N \eta}} \]

(5.41)

for \( N \geq N_0 \).

**Proof.** The claim follows from Theorem 5.6, the identity \( \tilde{m}(z) = m_1(z + \tilde{w}_X) \) from (4.52), and the second inequality in (5.32). Since \( \tilde{Y} \) is diagonal, \( m_1(z + \tilde{w}_X) \) is sum of terms identical to the fractions in Theorem 5.6, because the latter are

\[ \frac{z + \tilde{w}_X}{(y_i + (\tau - t)\tilde{A}_{ii})^2 - (z + \tilde{w}_X)^2} = \frac{1}{2} \left( \frac{1}{-z - \tilde{w}_X + \tilde{Y}_i} + \frac{1}{-z - \tilde{w}_X - \tilde{Y}_i} \right). \]

(5.42)

\[ \square \]

**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

\[ X' = \text{diag}(x_1, \ldots, x_N) + e^{-N}Q, \]

(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 \( \tilde{M} \) 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 \( \tilde{P} \) by

\[ \tilde{P} = \begin{bmatrix} 0 & P + e^{-N}Q \\ P^* + e^{-N}Q^* & 0 \end{bmatrix} \]

(6.2)

Let \( \gamma_1 \leq \cdots \leq \gamma_N \) be the eigenvalues of \( P \) and \( \alpha_1 \leq \cdots \leq \alpha_N \) be the positive eigenvalues of \( \tilde{P} \). Let \( \alpha_{-i} = -\alpha_i \) denote the corresponding negative eigenvalues. Then the \( \alpha_i \) are
Universality of the least singular value

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

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

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

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

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

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

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 \( \tilde{P} \) as

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

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

\[
p_{\tilde{P}}(\lambda, \mathbf{u}, \mathbf{v}) = \frac{1}{Z_N} \exp \left( -\frac{1}{2\sigma_N^2} \left( 2 \sum_{k=1}^{N} \lambda^2_k + \sum_{k=1}^{N} P_{kk}^2 - 2 \sum_{k \leq N} \lambda_k \Re(u_k v_k^*) P_{kk} \right) \right) \prod_{i \neq j} (\lambda_i - \lambda_j)^2 g(u, v), \tag{6.8}
\]

where we used

\[
w_{k,k+N} = \sum_{\ell=1}^{N} \lambda_{\ell} \Re(u_k v_{k+\ell}^*). \tag{6.9}
\]

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.
Universality of the least singular value

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

\[ 2 \sum_{k,i \leq N} \lambda_i \text{Re}(u_k \ell_i u_{i,k}^*) P_{kk} \leq \sum_{k=1}^{N} \lambda_k^2 + \sum_{k=1}^{N} P_{kk}^2. \]  \hspace{1cm} (6.10)

This implies

\[ p_F(\lambda, u, v) \leq \frac{1}{Z_N} \exp \left( -\frac{1}{2\sigma_N^2} \sum_{k=1}^{N} \lambda_k^2 \right) \prod_{i \neq j} (\lambda_i - \lambda_j)^2 g(u, v). \] \hspace{1cm} (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]. \( \square \)

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

\[ (\mathcal{F}_t)_{0 \leq t \leq t} = (\sigma(\tilde{M}(0), (B_s)_{0 \leq s \leq t}))_{0 \leq t \leq t}, \] \hspace{1cm} (6.12)

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

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

- \( \lambda(t) \) is adapted to the filtration \( (\mathcal{F}_t)_{0 \leq t \leq t} \), and
- \( \mathbb{P} |\lambda_{-N}(t) < \cdots < \lambda_{-1}(t) < \lambda_1(t) < \lambda_2(t) < \cdots < \lambda_N(t) |, \) for almost all \( t \in [0, t] \) = 1.

7 Analysis of SDEs

The system of SDEs for the evolution of the singular values of \( \tilde{M} \) is

\[ d\lambda_i = \frac{1}{\sqrt{2N}} dB_i + \frac{1}{2N} \sum_{j \neq i} \frac{1}{\lambda_i - \lambda_j} \ dt + R_i, \] \hspace{1cm} (7.1)

for \( 1 \leq |i| \leq N \), where

\[ R_i = \text{Re} \left( \check{\lambda}_i \left( U^*(t) \tilde{A} V(t) - U^*(0) \tilde{A} V(0) \right) k_i \right) \ dt \]
\[ + \frac{1}{\sqrt{N}} \text{Re} \left( \check{\lambda}_i \left( U^*(t) \left( 1_{(i,j) \in I^*} dB_{ij} \right) V(t) \right) k_i \right). \] \hspace{1cm} (7.2)

for \( i \geq 1 \). We recall that with \( \lambda_i \) and \( \lambda_{-i} \) are coupled as discussed above so that \( \lambda_i = -\lambda_{-i} \) (and the remainder terms and the \( \gamma_{ij} \) 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 \( \tau = N^{-1+b} \) and compare it to the process defined by

\[ d\mu_i = \frac{1}{\sqrt{2N}} dB_i + \frac{1}{2N} \sum_{j \neq i} \frac{1}{\mu_i - \mu_j} \ dt, \quad \mu_i(0) = \lambda_i(0), \] \hspace{1cm} (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 $0 < c_4 < c$. For $0 \leq \alpha \leq 1$ we define the interpolating process $z_i(t, \alpha)$ by the SDE
\[
 dz_i(t, \alpha) = \frac{1}{\sqrt{2N(1 + N^{-c_4})}} dB_i + \frac{1}{2N} \sum_{j \neq i} \frac{1 - \alpha \gamma_{ij}}{z_i(t, \alpha) - z_j(t, \alpha)} \, dt, \quad z_i(0, \alpha) = \lambda_i(0),
\] (7.4)
with
\[
 \gamma_{ij} = \gamma_{ij} \land N^{-c}. \tag{7.5}
\]
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
\[
 m_0(z) = \frac{1}{2N} \sum_{1 \leq |i| \leq N} \frac{1}{\lambda_i(0) - z}, \tag{7.6}
\]
and let $m_t(z)$ be the free convolution of $m_0$ with the semicircle law at time $t$ (see [60] for details):
\[
 m_t(z) = m_0(z + tm_t(z)), \quad \lim_{|z| \to \infty} m_t(z) = 0. \tag{7.7}
\]

Let $I = [-c, c]$ be the interval from Theorem 5.6 on which $\mu_1^{\text{sym}} \boxplus \mu_2^{\text{sym}}$ has a positive density bounded above and away from zero. Let $\gamma_i^\nu$ be the $i$-th classical eigenvalue location (the $i$-th $N$-quantile) of the measure $\mu_1^{\text{sym}} \boxplus \mu_2^{\text{sym}}$, and define the index set $\mathcal{J}$ by
\[
 \mathcal{J} = \{ i : \gamma_i^\nu \in I \}. \tag{7.8}
\]

From Corollary 5.9 we can deduce by standard arguments (cf. [46, Chapter 11]) that there exists $c > 0$ such that
\[
 |\gamma_i^\nu - \gamma_i(0)| \leq N^{-c} \tag{7.9}
\]
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 $\rho_t$ 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 $|i - j| \geq N^\nu$,
\[
 e^{|i - j|/N} \leq |\gamma_i(t) - \gamma_j(t)| \leq C |i - j|/N. \tag{7.10}
\]

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
\[
 \sup_{0 \leq t \leq T} \sup_{i \in \mathcal{J}} |z_i(t, 0) - \gamma_i(t)| \leq N^\nu / N. \tag{7.11}
\]

**Lemma 7.2.** With overwhelming probability for $i \in \mathcal{J}$,
\[
 |z_i(t, \alpha) - \gamma_i(t)| \leq N^{5a} / N. \tag{7.12}
\]
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 $b < a/100$ and $a < c/10$. For every time $t$ such that $0 \leq t \leq \tau$, we have with overwhelming probability for every index $i \in J$ that

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

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

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

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

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

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

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

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

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

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 $B_t = \{B_{ij}(t)\}_{1 \leq i,j \leq N}$ be a matrix of i.i.d. standard complex Brownian motions. Define $W_t = W + N^{-1/2} B_t$. 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 $t_0 = N^{-1+\omega_1}, t_1 = N^{-1+\omega_1}$, with the restrictions that

$$g N^\sigma \leq t_0 \leq N^{-\sigma} G^2, \quad 2\omega_1 < \omega_0, \quad (7.18)$$

\(^6\)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
\[ |s_i(t_a) - r_i(t_a)| < CN^{-1-\delta} \quad (7.19) \]
with overwhelming probability for $i < N^\omega$ and $t_a = t_0 + t_1$. Here $C, \omega, \delta$ are constants that depend only on $\delta_1$, $\sigma$, $\omega$, 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 $\text{Im} \, m_V$ 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 $\hat{M}(0)$ 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 $b < a/100$ and $a < \epsilon/10$, 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 $N_0(\sigma, \kappa, a, b, \epsilon)$ such that
\[ |\lambda_i(t_a) - r_i(t_a)| < N^{-1-\delta} \quad (7.20) \]
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 $\lambda_i$ is the least singular value of $\hat{M}(\tau) = M(\tau)$ 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
\[ |\lambda_1(M_N) - r_1(\tau)| < N^{-1-\delta}. \quad (7.21) \]

The distribution of the least singular value of a Gaussian matrix is known explicitly. For $W$ and any $r \geq 0$ [40],
\[ P(N\lambda_1(W) \leq r) = 1 - e^{-r^2}. \quad (7.22) \]
Therefore,
\[ P(N\lambda_1(\sqrt{1 + \tau W}) \leq \sqrt{1 + \tau r}) = 1 - e^{-r^2}. \quad (7.23) \]
We now show the $\sqrt{1 + \tau}$ factor is negligible, so that we may compare $\lambda_1(M_N)$ directly
Universality of the least singular value

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

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

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

By (7.21),

$$P(N_{r_1}(\tau) \leq r - N^{-\delta}) \leq P(N_{\lambda_1}(M_N) \leq r) \leq P(N_{r_1}(\tau) \leq r + N^{-\delta}) \quad (7.25)$$

We deduce

$$P(N_{\lambda_1}(W) \leq r - N^{-\delta}) \leq C N^{-c}$$

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

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

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

as desired.

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

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

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

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

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

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

and by the chain rule,

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

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 $\alpha \leq N$,
\[
\sum_{i,j} \frac{\partial \lambda_\alpha}{\partial x_{ij}} dx_{ij} = \sum_{i,j} u_\alpha^*(i) u_\alpha(j) dx_{ij} = u_\alpha^*(dX) u_\alpha
\]  
(A.6)
\[
= \frac{1}{2} j_{\alpha}^* \left( \frac{1}{\sqrt{N}} d\tilde{B} + \left( U^* \tilde{A}V - U^*(0) \tilde{A}V(0) \right) dt - \frac{1}{\sqrt{N}} U^*(1_{(i,j)\in I_2^*} d\tilde{B}_{ij}) V \right) \kappa_\alpha
\]
(A.7)
\[
+ \frac{1}{2} k_{\alpha}^* \left( \frac{1}{\sqrt{N}} d\tilde{B} + \left( U^* \tilde{A}V - U(0) \tilde{A}V(0) \right) dt - \frac{1}{\sqrt{N}} U^*(1_{(i,j)\in I_2^*} d\tilde{B}_{ij}) V \right)^* j_{\alpha}.
\]
(A.8)

We see that
\[
\frac{1}{2\sqrt{N}} \left( j_{\alpha}^* d\tilde{B} k_{\alpha} + k_{\alpha}^* d\tilde{B}^* j_{\alpha} \right) = \frac{1}{\sqrt{2N}} d\tilde{B}_{\alpha},
\]
where \( \{d\tilde{B}_{\alpha}\}_{\alpha=1}^N \) is a set of independent standard real Brownian motions. The independence follows from an explicit computation, noting that \( (d\tilde{B}_{ij})(d\tilde{B}_{kl}) = \delta_{il}\delta_{jk} \). The remaining terms are
\[
\text{Re} \langle j_{\alpha}, (U \tilde{A}V^* - U(0) \tilde{A}V(0)) k_{\alpha} \rangle dt + \frac{1}{\sqrt{N}} \text{Re} \langle j_{\alpha}, (U(1_{(i,j)\in I_2} d\tilde{B}_{ij}) V^*) k_{\alpha} \rangle.
\]
(A.10)

The second term is
\[
\frac{1}{2} \sum_{i,j,k,l} \frac{\partial^2 \lambda_\alpha}{\partial x_{ki} \partial x_{lj}} (dx_{ij})(dx_{kl})
\]
\[
= \frac{1}{2} \sum_{i,j,k,l} \sum_{\beta \neq \alpha} \frac{1}{\lambda_\alpha - \lambda_\beta} \left[ u_\alpha^*(k) u_\beta^*(l) u_\beta^*(i) u_\alpha(j) + u_\beta^*(k) u_\alpha^*(l) u_\alpha^*(i) u_\beta(j) \right] dx_{ij} dx_{kl}.
\]
(A.11)

The first contribution is
\[
\frac{1}{2} \sum_{i,j,k,l} \sum_{\beta \neq \alpha} \frac{1}{\lambda_\alpha - \lambda_\beta} \left[ u_\alpha^*(k) u_\beta^*(l) u_\beta^*(i) u_\alpha(j) + u_\beta^*(k) u_\alpha^*(l) u_\alpha^*(i) u_\beta(j) \right] \frac{1}{N} (d\tilde{B}_{ij})(d\tilde{B}_{kl}),
\]

where
\[
d\tilde{B}' = \begin{bmatrix} 0 & d\tilde{B} \\ d\tilde{B}^* & 0 \end{bmatrix}.
\]
(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 \( u_\alpha \), that is \( \|j_{\alpha}\|_2^2 \) or \( \|k_{\alpha}\|_2^2 \), both of which are \( 1/2 \). We then recover the drift term
\[
\frac{1}{2N} \sum_{\beta \neq \alpha} \frac{dt}{\lambda_\alpha - \lambda_\beta}.
\]
(A.14)

The remaining contribution is
\[
- \frac{1}{2N} \sum_{i,j,k,l} \sum_{\beta \neq \alpha} \frac{1}{\lambda_\alpha - \lambda_\beta} \left[ (u_\alpha^*(k) R'_{kl} u_\beta(l))(u_\beta^*(i) R'_{ij} u_\alpha(j)) + (u_\beta^*(k) R'_{kl} u_\alpha(l))(u_\alpha^*(i) R'_{ij} u_\beta(j)) \right].
\]
(A.15)

where
\[
R' = \begin{bmatrix} 0 & U^*(1_{(i,j)\in I_2^*} d\tilde{B}_{ij}) V \end{bmatrix}^*.
\]
(A.16)
Universality of the least singular value

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

$$\sum_{i,j} u^*_\alpha(i) R^*_{ij} u_\beta(j) = \frac{1}{2} \left( j^*_\alpha U^*(1_{(i,j)\in T^2} d\tilde{B}_{ij}) V k_\beta + k^*_\alpha V^*(1_{(i,j)\in T^2} d\tilde{B}_{ij})^* U j_\beta \right). \tag{A.17}$$

Define the column vectors

$$w_\alpha = U j_\alpha, \quad z_\alpha = V k_\beta, \tag{A.18}$$

and set $R = (1_{(i,j)\in T^2} d\tilde{B}_{ij})$. Then since the quadratic variation of a standard complex Brownian motion is zero, and the elements of $R$ are independent,

$$(w^*_\alpha R z_\beta + z^*_\alpha R^* w_\beta)(w^*_\alpha R z_\alpha + z^*_\alpha R^* w_\alpha) = w^*_\alpha R z_\beta z^*_\alpha R^* w_\alpha + z^*_\alpha R^* w_\beta R z_\alpha = \sum_{(i,j)\in T^2} |w_\alpha(i)|^2 |z_\beta(j)|^2 + \sum_{(i,j)\in T^2} |w_\beta(i)|^2 |z_\alpha(j)|^2$$

$$:= 2\gamma_{\alpha\beta}. \tag{A.19}$$

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

$$-\frac{1}{2N} \sum_{\beta \neq \alpha} \gamma_{\alpha\beta} \frac{dt}{\lambda_\alpha - \lambda_\beta}. \tag{A.20}$$

We obtain the following SDE for the eigenvalues of $X$, which are the singular values of $\tilde{M}$ and their negatives. We label the positive eigenvalues by $\{\lambda_i\}_{i=1}^N$ and the negative eigenvalues by $\{\lambda_i\}_{i=1}^{N-1}$, 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 $i > 0$:

$$d\lambda_i = \frac{1}{\sqrt{2N}} dB_i + \frac{1}{2N} \sum_{\beta \neq \alpha} \frac{1 - \gamma_{ij}}{\lambda_i - \lambda_j} dt + R_i, \tag{A.21}$$

where

$$R_i = \text{Re}(j_i, (U^*\tilde{A} V - U^*(0)\tilde{A} V(0))k_i) dt + \frac{1}{\sqrt{N}} \text{Re}(j_i, (U^*(1_{(i,j)\in T^2} d\tilde{B}_{ij}) V)k_i). \tag{A.22}$$

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

$$\frac{\partial \lambda_\alpha}{\partial x_{ij}} = 2u_\alpha(i)u_\alpha(j), \quad \frac{\partial u_\alpha(i)}{\partial x_{kl}} = \sum_{\beta \neq \alpha} \frac{u_\beta(k)u_\alpha(l) + u_\beta(l)u_\alpha(k)}{\lambda_\alpha - \lambda_\beta} u_\beta(i), \tag{A.23}$$

where we view $X$ as a function of $N^2$ real variables $x_{ij}$ with $1 \leq i \leq N$ and $N \leq j \leq 2N$.

Using the representation (A.2), we find

$$u_{-\alpha}(k)u_\alpha(l) + u_\beta(l)u_\alpha(k) = u_{-\alpha}(k)u_\alpha(l) + u_{-\alpha}(l)u_\alpha(k)$$

$$= -u_\alpha(k)u_\alpha(l) + u_\alpha(l)u_\alpha(k)$$

$$= 0. \tag{A.24}$$

Therefore, we see that the sum in the drift component now omits the term with $\beta = -\alpha$, and there is no repulsion between $\lambda_\alpha$ and $-\lambda_\alpha$. The rest of the derivation is completed as before.
Universality of the least singular value

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 $2 \times 10^4$ 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 $2 \times 10^4$ 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

$$ M = R^\dagger X T + U^\dagger Y V, \quad (B.1) $$

where $X = \text{diag}(x_1, \ldots, x_N)$ and $Y = \text{diag}(y_1, \ldots, y_N)$ 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 $\lambda_1$ 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

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

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

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

with the interaction between $\lambda_i$ and $\lambda_{-i}$ removed. For $\lambda_1$, this means there is no force from $\lambda_{-1}$ 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
Universality of the least singular value

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 \( \lambda_1 \); 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

\[
P(N \lambda_1(M_N) \leq r) = 1 - e^{-r^2/2 - r} + O(N^{-c})
\]

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

\[
\frac{1}{2N} \sum_{|y_i - E| \geq N^{1+a}} \frac{1}{|y_i - E|^2} \leq 2C_a N^{1-a},
\]

\[
\frac{1}{2N} \sum_{|y_i - E| \geq N^{1+a}} \frac{1}{|y_i - E|} \leq 2C_a \log N + 4,
\]

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

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

\[
\frac{1}{2N} \sum_{|y_i - E| \geq \eta} \frac{\eta}{|y_i - E|^2} \leq \frac{1}{N} \Im \sum_{|i|=1}^{N} \frac{1}{y_i - (E + i\eta)} \leq 2C_a.
\]

Divide both sides by \( \eta \) to obtain

\[
\frac{1}{2N} \sum_{|y_i - E| \geq \eta} \frac{1}{|y_i - E|^2} \leq \frac{2C_a}{\eta} = 2C_a N^{1-a}.
\]

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

\[
\frac{1}{x} \leq \frac{2}{x + 1} + \int_{\eta}^{1} \frac{2 dt}{(x + t)^2} \leq 2 + \int_{\eta}^{1} \frac{2 dt}{x^2 + t^2}.
\]

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

\[
\frac{1}{2N} \sum_{|y_i - E| \geq \eta} \frac{1}{|y_i - E|} \leq 4 + \frac{1}{2N} \int_{\eta}^{1} \frac{N \cdot 4 dt}{|y_i - (E + it)|^2}.
\]

Using \( \frac{1}{|y_i - (E + it)|^2} = \frac{1}{4} \Im \left( \frac{1}{y_i - (E + it)} \right) \) and the hypothesized bound (2.7) on \( m_Y(E + i\eta) \), we have

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

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

\[
\frac{1}{2N} \sum_{|y_i - E| \geq \eta} \frac{1}{|y_i - E|} \leq 2C_a \log N + 4.
\]
Universality of the least singular value

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

\[
\| A \| = \max_{1 \leq i \leq N} |A_{ii}| \leq CN^{1-s}.
\]

Similarly we obtain

\[
\| \hat{A} \| = \max_{1 \leq i \leq N} |\hat{A}_{ii}| \leq C(1 + \log N).
\]

C.2 Unitary flow

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

**Lemma C.2.** For \( a, b, U \), as above,

\[
P( \sup_{0 \leq t \leq \tau} \| U(t) - I \| \geq N^{-10b} ) \leq \exp \left( -N^{10b} \right),
\]

and the same estimate holds for \( V \).

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

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

\[
P( \sup_{0 \leq s \leq t} \| \hat{U}(s) - \hat{U}(t_0) \| \geq (N(t - t_0))^{1/4} \) \leq \exp \left( -N^{6/5} \right).
\]

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

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

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 \( \mu_\alpha, \mu_\beta \) be probability measures with density functions \( \rho_\alpha, \rho_\beta \) 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 \( w_\alpha, w_\beta \) extend continuously to 0 with values in \( C^+ \cup \mathbb{R} \cup \{\infty\} \). 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

\[
\mathcal{E} = \{ z \in C^+ \cup \mathbb{R} : |z| \leq r \}.
\]

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

\[
(w_\alpha + w_\beta - z)^{-1} = \int_{\mathbb{R}} \frac{d\mu_\beta(x)}{w_\alpha - x} = w_\alpha^{-1} + O(w_\alpha^{-2}),
\]

We recall that equation (3.7) in this reference is derived by applying the formula \( \frac{d}{dt} e^{X(t)} = \int_0^\theta e^{\alpha X(t)} e^{X(t)} e^{(\theta - \alpha)X(t)} dt \), which holds for any one-parameter matrix subgroup \( X(t) \) [B1], to compute the derivative of the matrix exponential with respect to each matrix entry, in conjunction with Itô’s formula.
Universality of the least singular value

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

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

This contradicts $L/|w_\alpha| \leq |w_\beta/w_\alpha|$ (which holds by our assumptions on $w_\alpha, w_\beta$) 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 $z \in E$ and $M$ sufficiently large,

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

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

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

**Lemma C.5.** Let $\mu_\alpha$ 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**


Universality of the least singular value


Universality of the least singular value


Universality of the least singular value


EJP **26** (2021), paper 40. https://www.imstat.org/ejp

Page 37/38
Universality of the least singular value


**Acknowledgments.** The authors thank Benjamin Landon for comments on a preliminary draft of this paper. They also thank the anonymous referees for their detailed comments, which substantially improved the paper.