LOCAL INHOMOGENEOUS CIRCULAR LAW[1]

BY JOHANNES ALT, LÁSZLÓ ERDŐS AND TORBEN KRÜGER

Institute of Science and Technology Austria

We consider large random matrices X with centered, independent entries, which have comparable but not necessarily identical variances. Girko's circular law asserts that the spectrum is supported in a disk and in case of identical variances, the limiting density is uniform. In this special case, the *local circular law* by Bourgade et al. [*Probab. Theory Related Fields* **159** (2014) 545–595; *Probab. Theory Related Fields* **159** (2014) 619–660] shows that the empirical density converges even locally on scales slightly above the typical eigenvalue spacing. In the general case, the limiting density is typically inhomogeneous and it is obtained via solving a system of deterministic equations. Our main result is the local *inhomogeneous* circular law in the bulk spectrum on the optimal scale for a general variance profile of the entries of X.

CONTENTS

1. Introduction. The density of eigenvalues of large random matrices typically converges to a deterministic limit as the dimension n of the matrix tends to infinity. In the Hermitian case, the best known examples are the Wigner semicircle law for Wigner ensembles and the Marchenko–Pastur law for sample covariance

matrices. In both cases, the spectrum is real, and these laws state that the empirical eigenvalue distribution converges to an explicit density on the real line.

The spectra of non-Hermitian random matrices concentrate on a domain of the complex plane. The most prominent case is the *circular law*, asserting that for an $n \times n$ matrix X with independent, identically distributed entries, satisfying $\mathbb{E}x_{ij} = 0$, $\mathbb{E}|x_{ij}|^2 = n^{-1}$, the empirical density converges to the uniform distribution on the unit disk $\{z : |z| < 1\} \subset \mathbb{C}$. Despite the apparent similarity in the statements, it is considerably harder to analyze non-Hermitian random matrices than their Hermitian counterparts since eigenvalues of non-Hermitian matrices may respond very drastically to small perturbations. This instability is one reason why the universality of local eigenvalue statistics in the bulk spectrum, exactly on the scale of the eigenvalue spacing, is not yet established for X with independent (even for i.i.d.) entries, while the corresponding statement for Hermitian Wigner matrices, known as the Wigner–Dyson–Mehta universality conjecture, has been proven recently; see [16] for an overview.

The circular law for i.i.d. entries has a long history; we refer to the extensive review [11]. The complex Gaussian case (Ginibre ensemble) has been settled in the 1960s by Mehta using explicit computations. Girko in [20] found a key formula to relate linear statistics of eigenvalues of X to eigenvalues of the family of Hermitian matrices $(X - z)^*(X - z)$ where $z \in \mathbb{C}$ is a complex parameter. Technical difficulties still remained until Bai [9] presented a complete proof under two additional assumptions requiring higher moments and bounded density for the single entry distribution. After a series of further partial results [22, 26, 29], the circular law for i.i.d. entries under the optimal condition, assuming only the existence of the second moment, was established by Tao and Vu [30].

Another line of research focused on the local version of the circular law with constant variances, $\mathbb{E}|x_{ij}|^2 = n^{-1}$, which asserts that the local density of eigenvalues is still uniform on scales $n^{-1/2+\epsilon}$, that is, slightly above the typical spacing between neighboring eigenvalues. The optimal result was achieved in Bourgade, Yau and Yin [12, 13] and Yin [33] both inside the unit disk ("bulk regime") and at the edge $|z| = 1$. If the first three moments match those of a standard complex Gaussian, then a similar result has also been obtained by Tao and Vu in [31]. In [31], this result was used to prove the universality of local eigenvalue statistics under the assumption that the first four moments match those of a complex Gaussian. While there is no proof of universality for general distributions without moment matching conditions yet, similar to the development in the Hermitian case, the local law is expected to be one of the key ingredients of such a proof in the future.

In this paper, we study non-Hermitian matrices X with a general matrix of variances $S = (s_{ij})$, that is, we assume that x_{ij} are centered, independent, but $s_{ij} := \mathbb{E}|x_{ij}|^2$ may depend nontrivially on the indices i, j. We show that the eigenvalue density is close to a deterministic density σ on the smallest possible scale. As a direct application, our local law implies that the spectral radius $\rho(X)$ of X is

arbitrarily close to $\sqrt{\rho(S)}$, where $\rho(S)$ is the spectral radius of S. More precisely, we prove that for every $\varepsilon > 0$

$$\sqrt{\rho(S)} - \varepsilon \leq \rho(X) \leq \sqrt{\rho(S)} + \varepsilon$$

with a very high probability as n tends to infinity. The fact that the spectral radius of X becomes essentially deterministic is the key mathematical mechanism behind the sharp "transition to chaos" in a commonly studied mean field model of dynamical neural networks [28]. This transition is described by the stability/instability of the system of ordinary differential equations

$$\dot{q}_i(t) = q_i(t) - \lambda \sum_{j=1}^{n} x_{ij} q_j(t)$$

for $i = 1, \ldots, n$ as λ varies. Moreover, the number of unstable modes close to the critical value of the parameter λ is determined by the behaviour of σ at the spectral edge which we also analyze. Such systems have originally been studied under the assumption that the coefficients x_{ij} are independent and identically distributed [25]. More recently, however, it was argued [5, 6] that for more realistic applications in neuroscience one should allow x_{ij} to have varying distributions with an arbitrary variance profile S.

   After Girko's Hermitization, understanding the spectrum of X reduces to analyzing the spectrum of the family of Hermitian matrices

$$(1.1) \qquad \boldsymbol{H}^z := \begin{pmatrix} 0 & X - z\mathbb{1} \\ X^* - \bar{z}\mathbb{1} & 0 \end{pmatrix}$$

of double dimension, where $z \in \mathbb{C}$. The Stieltjes transform of the spectral density of $\boldsymbol{H}^z$ at any spectral parameter w in the upper half plane $\mathbb{H} := \{w \in \mathbb{C} : \operatorname{Im} w > 0\}$ is approximated via the solution of a system of $2n$ nonlinear equations, written concisely as

$$
\begin{aligned}
(1.2) \qquad -\frac{1}{m_1} &= w + Sm_2 - \frac{|z|^2}{w + S^t m_1}, \\
-\frac{1}{m_2} &= w + S^t m_1 - \frac{|z|^2}{w + Sm_2},
\end{aligned}
$$

where $m_a = m_a^z(w) \in \mathbb{H}^n$, $a = 1, 2$ are n-vectors with each component in the upper half plane. The normalized trace of the resolvent, $\frac{1}{2n}\operatorname{trace}(\boldsymbol{H}^z - w)^{-1}$, is approximately equal to $\frac{1}{n}\sum_j [m_1^z(w)]_j = \frac{1}{n}\sum_j [m_2^z(w)]_j$ in the $n \to \infty$ limit. The spectral density of $\boldsymbol{H}^z$ at any $E \in \mathbb{R}$ is then given by setting $w = E + i\eta$ and taking the limit $\eta \to 0+$ for the imaginary part of these averages. In fact, for Girko's formula it is sufficient to study the resolvent only along the positive imaginary axis $w \in i\mathbb{R}_+$. Heuristically, equations (1.2) arise from second-order perturbation theory and in physics they are commonly called *Dyson equations*. Their analogues for general

Hermitian ensembles with independent or weakly dependent entries play an essential role in random matrix theory. They have been systematically studied by Girko, for example, equation (1.2) in the current random matrix context appears as the *canonical equation of type K_{25}* in Theorem 25.1 in [21]. In particular, under the condition that all s_{ij} variances are comparable, that is, $c/n \leq s_{ij} \leq C/n$ with some positive constants c, C, Girko identifies the limiting density. From his formulas it is clear that this density is rotationally symmetric. He also presents a proof for the weak convergence of the empirical eigenvalue distribution but the argument was considered incomplete. This deficiency can be resolved in a similar manner as for the circular law assuming a bounded density of the single entry distribution using the argument from Section 4.4 of [11]. In a recent preprint [14], Cook et al. substantially relax the condition on the uniform bound $s_{ij} \geq c/n$ by replacing it with a concept of *robust irreducibility*. Moreover, relying on the bound by Cook [15] on the smallest singular value of X, they also remove any condition on the regularity of the single entry distribution and prove weak convergence on the global scale.

The matrix $\boldsymbol{H}^z$ may be viewed as the sum of a *Wigner-type matrix* [1] with centered, independent (up to Hermitian symmetry) entries and a deterministic matrix whose two off-diagonal blocks are $-z\mathbb{1}$ and $-\bar{z}\mathbb{1}$, respectively. Disregarding these z terms for the moment, (1.2) has the structure of the *quadratic vector equations* that were extensively studied in [2, 4]. Including the z-terms, $\boldsymbol{H}^z$ at first sight seems to be a special case of the random matrix ensembles with nonzero expectations analyzed in [3] and (1.2) is the diagonal part of the corresponding *Matrix Dyson Equation* (MDE). In [3], an optimal local law was proved for such ensembles. However, the large zero blocks in the diagonal prevent us from applying these results to $\boldsymbol{H}^z$ or even to $\boldsymbol{H}^{z=0}$. In fact, the flatness condition **A**1 in [3] [see (3.1) later] prohibit such large zero diagonal blocks. These conditions are essential for the proofs in [3] since they ensure the stability of the corresponding Dyson equation against *any* small perturbation. In this case, there is only one potentially unstable direction that is associated to a certain Perron–Frobenius eigenvector, and this direction is regularized by the positivity of the density of states at least in the bulk regime of the spectrum.

If the flatness condition **A**1 is not satisfied, then the MDE can possess further unstable directions. In particular, in our setup, the MDE is not stable in the previously described strong sense; there is at least one additional unstable direction which cannot be regularized by the positivity of the density of states. Owing to the specific structure of $\boldsymbol{H}^z$, the *matrix* Dyson equation decouples and its diagonal parts satisfy a closed system of *vector* equations (1.2). Compared to the MDE, the reduced vector equations (1.2) are rather cubic than quadratic in nature. For this reduced system, however, we can show that there is only one further unstable direction, at least when S is entrywise bounded from below by some c/n. The system is not stable against an arbitrary perturbation, but for the perturbation arising

in the random matrix problem we reveal a key cancellation in the leading contribution to the unstable direction. Armed with this new insight, we will perform a detailed stability analysis of (1.2).

This delicate stability analysis is the key ingredient for the proof of our main result, the optimal local law for X with an optimal speed of convergence as $n \to \infty$. In this paper, we consider the bulk regime, that is, spectral parameter z inside the disk with boundary $|z|^2 = \rho(S)$, where $\rho(S)$ is the spectral radius of S. We defer the analysis of the edge of the spectrum of X to later works.

In the special case $z = 0$, we thoroughly studied the system of equations (1.2) even for the case when S is a rectangular matrix in [7]; the main motivation was to prove the local law for random *Gram matrices*, that is, matrices of the form XX^*. Note that in [7] we needed to tackle a much simpler quadratic system since taking $z = 0$ in (1.2) removes the most complicated nonlinearity.

Finally, we list two related recent results. Local circular law on the optimal scale in the bulk has been proven in [32] for ensembles of the form TX, where T is a deterministic $N \times M$ matrix and X is a random $M \times N$ matrix with independent, centered entries whose variances are constant and have vanishing third moments. The structure of the product matrix TX is very different from our matrices that could be viewed as the Hadamard (entrywise) product of the matrix $(s_{ij}^{1/2})$ and a random matrix with identical variances. The approach of [32] is also very different from ours: it relies on first assuming that X is Gaussian and using its invariance to reduce the problem to the case when T^*T is diagonal. Then the corresponding Dyson equations are much simpler, in fact, they consist of only two scalar equations and they are characterized by a vector of parameters (of the singular values of T) instead of an entire matrix of parameters S. The vanishing third moment condition in [32] is necessary to compare the general distribution with the Gaussian case via a moment matching argument. We also mention the recent proof of the local *single ring theorem* on optimal scale in the bulk [10]. This concerns another prominent non-Hermitian random matrix ensemble that consists of matrices of the form $U\Sigma V$, where U, V are two independent Haar distributed unitaries and Σ is deterministic (may be assumed to be diagonal). The spectrum lies in a ring about the origin and the limiting density can be computed via free convolution [23].

Notation. For vectors $v, w \in \mathbb{C}^l$, we write their componentwise product as $vw = (v_i w_i)_{i=1}^l$. If $f : U \to \mathbb{C}$ is a function on $U \subset \mathbb{C}$, then we define $f(v) \in \mathbb{C}^l$ for $v \in U^l$ to be the vector with components $f(v)_i = f(v_i)$ for $i = 1, \ldots, l$. We will in particular apply this notation with $f(z) = 1/z$ for $z \in \mathbb{C} \setminus \{0\}$. We say that a vector $v \in \mathbb{C}^l$ is positive, $v > 0$, if $v_i > 0$ for all $i = 1, \ldots, l$. Similarly, the notation $v \leq w$ means $v_i \leq w_i$ for all $i = 1, \ldots, l$. For vectors $v, w \in \mathbb{C}^l$, we define $\langle w \rangle = l^{-1} \sum_{i=1}^l w_i$, $\langle v, w \rangle = l^{-1} \sum_{i=1}^l \overline{v_i} w_i$, $\|w\|_2^2 = l^{-1} \sum_{i=1}^l |w_i|^2$ and $\|w\|_\infty = \max_{i=1,\ldots,l} |w_i|$, $\|v\|_1 := \langle |v| \rangle$. Note that $\langle w \rangle = \langle 1, w \rangle$, where we used the convention that 1 also denotes the vector $(1, \ldots, 1) \in \mathbb{C}^l$. In general, we use

the notation that if a scalar α appears in a vector-valued relation, then it denotes the constant vector $(\alpha, \ldots, \alpha)$. In most cases, we will work in n or $2n$ dimensional spaces. Vectors in $\mathbb{C}^{2n}$ will usually be denoted by boldface symbols like $\boldsymbol{v}$, $\boldsymbol{u}$ or $\boldsymbol{y}$. Correspondingly, capitalized boldface symbols denote matrices in $\mathbb{C}^{2n \times 2n}$, for example, $\boldsymbol{R}$. We use the symbol $\mathbb{1}$ for the identity matrix in $\mathbb{C}^{l \times l}$, where the dimension $l = n$ or $l = 2n$ is understood from the context. For a matrix $A \in \mathbb{C}^{l \times l}$, we use the short notation $\|A\|_\infty := \|A\|_{\infty \to \infty}$ and $\|A\|_2 := \|A\|_{2 \to 2}$ if the domain and the target are equipped with the same norm whereas we use $\|A\|_{2 \to \infty}$ to denote the matrix norm of A when it is understood as a map $(\mathbb{C}^l, \|\cdot\|_2) \to (\mathbb{C}^l, \|\cdot\|_\infty)$. We define the normalized trace of an $l \times l$ matrix $B = (b_{ij})_{i,j=1}^l \in \mathbb{C}^{l \times l}$ as

$$(1.3) \qquad \operatorname{Tr} B := \frac{1}{l} \sum_{j=1}^l b_{jj}.$$

For a vector $y \in \mathbb{C}^l$, we write $\operatorname{diag} y$ or $\operatorname{diag}(y)$ for the diagonal $l \times l$ matrix with y on its diagonal, that is, this matrix acts on any vector $x \in \mathbb{C}^l$ as

$$(1.4) \qquad \operatorname{diag}(y)x = yx.$$

We write $\mathrm{d}^2 z$ for indicating integration with respect to the Lebesgue measure on $\mathbb{C}$. For $a \in \mathbb{C}$ and $\varepsilon > 0$, the open disk in the complex plane centered at a with radius ε is denoted by $D(a, \varepsilon) := \{b \in \mathbb{C} | |a - b| < \varepsilon\}$. Furthermore, we denote the characteristic function of some event A by $\chi(A)$, the positive real numbers by $\mathbb{R}_+ := (0, \infty)$ and the nonnegative real numbers by $\mathbb{R}_0^+ := [0, \infty)$.

2. Main results. Let X be a random $n \times n$ matrix with centered entries, $\mathbb{E}x_{ij} = 0$, and $s_{ij} := \mathbb{E}|x_{ij}|^2$ the corresponding variances. We introduce its variance matrix $S := (s_{ij})_{i,j=1}^n$.

ASSUMPTIONS.

(A) *The variance matrix S is* flat, *that is, there are $0 < s_* < s^*$ such that*

$$(2.1) \qquad \frac{s_*}{n} \le s_{ij} \le \frac{s^*}{n}$$

for all $i, j = 1, \ldots, n$.

(B) *All entries of X have bounded moments in the sense that there are $\mu_m > 0$ for $m \in \mathbb{N}$ such that*

$$(2.2) \qquad \mathbb{E}|x_{ij}|^m \le \mu_m n^{-m/2}$$

for all $i, j = 1, \ldots, n$.

(C) *Each entry of $\sqrt{n}X$ has a density, that is, there are probability densities $f_{ij} : \mathbb{C} \to [0, \infty)$ such that*

$$\mathbb{P}(\sqrt{n}x_{ij} \in B) = \int_B f_{ij}(z) \, \mathrm{d}^2 z$$

for all $i, j = 1, \ldots, n$ and $B \subset \mathbb{C}$ a Borel set. There are $\alpha, \beta > 0$ such that $f_{ij} \in L^{1+\alpha}(\mathbb{C})$ and

$$(2.3) \qquad\qquad \|f_{ij}\|_{1+\alpha} \le n^{\beta}$$

for all $i, j = 1, \ldots, n$.

In the following, we will assume that s_*, s^*, α, β and the sequence $(\mu_m)_m$ are fixed constants which we will call *model parameters*. The constants in all our estimates will depend on the model parameters without further notice.

REMARK 2.1.   The Assumption (C) is used in our proof solely for controlling the smallest singular value of $X - z\mathbb{1}$ with very high probability uniformly for $z \in D(0, \tau^*)$ with some fixed $\tau^* > 0$ in Proposition 5.7. All our other results do not make use of Assumption (C). Provided a version of Proposition 5.7 that tracks the z-dependence can effectively be obtained without (C), our main result, the local inhomogeneous circular law in Theorem 2.5, will hold true solely assuming (A) and (B). For example, a very high probability estimate uniform in z in a statement similar to Corollary 1.22 of [15] would be sufficient.

The density of states of X will be expressed in terms of v_1^{τ} and v_2^{τ}, which are the positive solutions of the following two coupled vector equations:

$$(2.4a) \qquad\qquad \frac{1}{v_1^{\tau}} = \eta + Sv_2^{\tau} + \frac{\tau}{\eta + S^t v_1^{\tau}},$$

$$(2.4b) \qquad\qquad \frac{1}{v_2^{\tau}} = \eta + S^t v_1^{\tau} + \frac{\tau}{\eta + S v_2^{\tau}},$$

for all $\eta \in \mathbb{R}_+$ and $\tau \in \mathbb{R}_0^+$. Here, $v_1^{\tau}, v_2^{\tau} \in \mathbb{R}_+^n$ and recall that the algebraic operations are understood componentwise, for example, $(1/v)_i = 1/v_i$ for the ith component of the vector $1/v$. The system (2.4) is a special case of (1.2) with $w = i\eta$, $\tau = |z|^2$ and $v_a = \operatorname{Im} m_a$ for $a = 1, 2$. The existence and uniqueness of solutions to equations of the type (2.4) are considered standard knowledge in the literature [21]. The equations can be viewed as a special case of the matrix Dyson equation for which existence and uniqueness was proven in [24]. We explain this connection in more detail in the Appendix where we give the proof of Lemma 2.2 for the convenience of the reader.

LEMMA 2.2 (Existence and uniqueness).   *For every $\tau \in \mathbb{R}_0^+$, there exist two uniquely determined functions $v_1^{\tau} : \mathbb{R}_+ \to \mathbb{R}_+^n$, $v_2^{\tau} : \mathbb{R}_+ \to \mathbb{R}_+^n$, which satisfy (2.4).*

We denote the spectral radius of S by $\rho(S)$, that is,

$$\rho(S) := \max|\operatorname{Spec}(S)|.$$

Now, we define the density of states of X through the solution to (2.4).

DEFINITION 2.3 (Density of states of X). Let v_1^τ and v_2^τ be the unique positive solutions of (2.4). The *density of states* $\sigma : \mathbb{C} \to \mathbb{R}$ of X is defined through

$$(2.5) \qquad \sigma(z) := -\frac{1}{2\pi} \int_0^\infty \Delta_z \langle v_1^\tau(\eta)|_{\tau=|z|^2} \rangle \, d\eta$$

for $|z|^2 < \rho(S)$ and $\sigma(z) := 0$ for $|z|^2 \geq \rho(S)$. The right-hand side of (2.5) is well defined by part (i) of the following proposition.

In the following proposition, we present some key properties of the density of states σ of X. Some of them have previously been known [14, 21]. For an alternative representation of σ, see (4.8) later.

PROPOSITION 2.4 (Properties of σ). *Let v_1^τ and v_2^τ be the unique positive solutions of* (2.4). *Then*:

(i) *The function $\mathbb{R}_+ \times \mathbb{C} \to \mathbb{R}_+^{2n}, (\eta, z) \mapsto (v_1^\tau(\eta), v_2^\tau(\eta))|_{\tau=|z|^2}$ is infinitely often differentiable and $\eta \mapsto \Delta_z \langle v_1^\tau(\eta)|_{\tau=|z|^2} \rangle$ is integrable on $\mathbb{R}_+$ for each $z \in D(0, \sqrt{\rho(S)})$.*

(ii) *The function σ, defined in* (2.5), *is a rotationally symmetric probability density on $\mathbb{C}$.*

(iii) *The restriction $\sigma|_{D(0, \sqrt{\rho(S)})}$ is infinitely often differentiable such that for every $\varepsilon > 0$ each derivative is bounded on $D(0, \sqrt{\rho(S)} - \varepsilon)$ uniformly in n. Moreover, there exist constants $c_1 > c_2 > 0$, which depend only on s_* and s^*, such that*

$$(2.6) \qquad c_1 \geq \sigma(z) \geq c_2$$

for all $z \in D(0, \sqrt{\rho(S)})$. In particular, the support of σ is the closed disk of radius $\sqrt{\rho(S)}$ around zero. In fact, the jump height $\lim \sigma(z)$ as $|z| \uparrow \sqrt{\rho(S)}$ can be computed explicitly (see Remark 4.2).

The next theorem, the main result of the present article, states that the eigenvalue distribution of X, with a very high probability, can be approximated by σ on the mesoscopic scales n^{-a} for any $a \in (0, 1/2)$. Note that $n^{-1/2}$ is the typical eigenvalue spacing so our result holds down to the optimal local scale. To study the local scale, we shift and rescale the test functions as follows. Let $f \in C_0^2(\mathbb{C})$. For $z_0 \in \mathbb{C}$ and $a > 0$, we define

$$f_{z_0,a} : \mathbb{C} \to \mathbb{C}, \qquad f_{z_0,a}(z) := n^{2a} f(n^a(z - z_0)).$$

We denote the eigenvalues of X by $\sigma_1, \ldots, \sigma_n$.

THEOREM 2.5 (Local inhomogeneous circular law). *Let X be a random matrix which has independent centered entries and satisfies* (A), (B) *and* (C). *Furthermore, let $a \in (0, 1/2], \varphi > 0, \tau_* > 0$ and σ defined as in* (2.5):

(i) (*Bulk spectrum.*) *For every $\varepsilon > 0$, $D > 0$, there is a positive constant $C_{\varepsilon,D}$ such that*

$$(2.7) \qquad \mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n} f_{z_0,a}(\sigma_i) - \int_{\mathbb{C}} f_{z_0,a}(z)\sigma(z)\,\mathrm{d}^2 z\right| \geq \frac{\|\Delta f\|_{L^1}}{n^{1-2a-\varepsilon}}\right) \leq \frac{C_{\varepsilon,D}}{n^D}$$

holds true for all $n \in \mathbb{N}$, for every $z_0 \in \mathbb{C}$ satisfying $|z_0|^2 \leq \rho(S) - \tau_*$ and for every $f \in C_0^2(\mathbb{C})$ satisfying $\mathrm{supp}\, f \subset D(0,\varphi)$. *The point z_0 and the function f may depend on n.*

(ii) (*Away from the spectrum.*) *For every $D > 0$, there exists a positive constant C_D such that*

$$(2.8) \qquad \mathbb{P}\big(\exists i \in \{1,\ldots,n\} | |\sigma_i|^2 \geq \rho(S) + \tau_*\big) \leq \frac{C_D}{n^D}$$

holds true for all $n \in \mathbb{N}$.

In addition to the model parameters, the constant $C_{\varepsilon,D}$ in (2.7) *depends only on a, φ and τ_* (apart from ε and D) and the constant C_D in* (2.8) *only on τ_* (apart from D).*

The key technical input for the proof of Theorem 2.5 is the local law for $\boldsymbol{H}^z$ (see Theorem 5.2). In Figure 1 below, we illustrate how the empirical spectral measure of X converges to σ for an example with a nontrivial variance profile S. We now state a simple corollary of the local law for $\boldsymbol{H}^z$ on the complete delocalization of the bulk eigenvectors of X.

COROLLARY 2.6 (Eigenvector delocalization).   *Let $\tau_* > 0$. For all $\varepsilon > 0$ and $D > 0$, there is a positive constant $C_{\varepsilon,D}$ such that*

$$(2.9) \qquad \mathbb{P}\big(\|y\|_\infty \geq n^{-1/2+\varepsilon}\big) \leq \frac{C_{\varepsilon,D}}{n^D}$$

holds true for all $n \in \mathbb{N}$ and for all eigenvectors $y \in \mathbb{C}^n$ of X, normalized as $\sum_{i=1}^{n} |y_i|^2 = 1$, corresponding to an eigenvalue $\sigma \in \mathrm{Spec}\, X$ with $|\sigma|^2 \leq \rho(S) - \tau_*$. The constant $C_{\varepsilon,D}$ in (2.9) *depends only on τ_* and the model parameters (in addition to ε and D).*

The proof of Corollary 2.6 will be given after the statement of Theorem 5.2. We remark that eigenvector delocalization for random matrices with independent entries was first proved by Rudelson and Vershynin in [27].

2.1. *Short outline of the proof.*   We start with the Hermitization trick due to Girko, which expresses $\sum_{i=1}^{n} f_{z_0,a}(\sigma_i)$ in terms of an integral of the log-determinant of $X - z\mathbb{1}$ for any $z \in \mathbb{C}$. Furthermore, the log-determinant of $X - z\mathbb{1}$ can be rewritten as the log-determinant of a Hermitian matrix $\boldsymbol{H}^z$.

(a) Eigenvalue locations

(b) Variance profile S
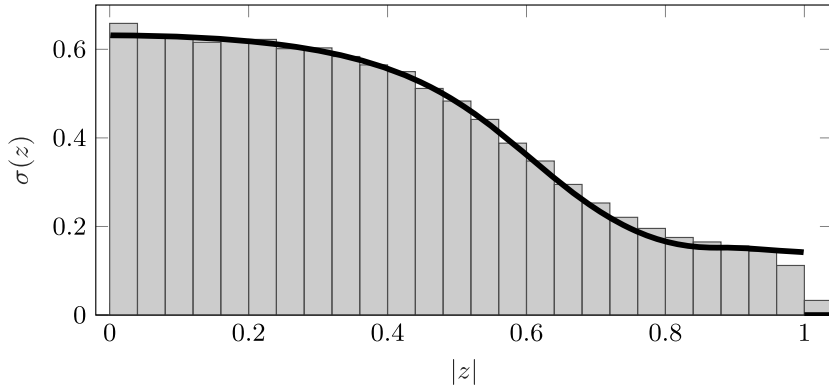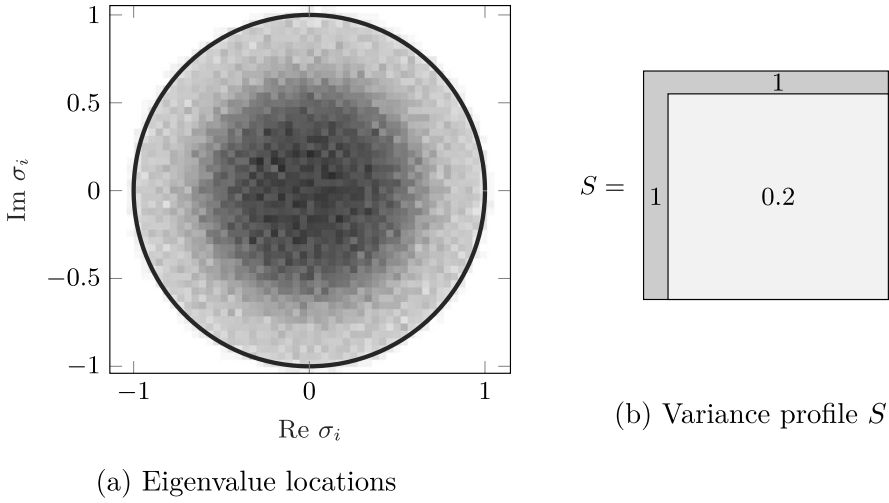
(c) Eigenvalue histogram and density of states σ

FIG. 1. *These figures were obtained by sampling* 100 *matrices of size* 1000×1000 *with centered complex Gaussian entries and the variance profile S. Figure* (a) *shows the eigenvalue density for the variance profile S given in Figure* (b) *(we rescaled S such that $\rho(S) = 1$). The eigenvalue density is rotationally invariant and almost all eigenvalues are contained in the disk of radius 1 around zero. Moreover, the eigenvalue density is considerably higher around 0. Figure* (c) *compares the histogram of the eigenvalue with the density of states σ obtained from* (2.4) *and* (2.5).

Using the log-transform of the empirical spectral measure of X, we obtain

$$(2.10) \qquad \frac{1}{n} \sum_{i=1}^{n} f_{z_0,a}(\sigma_i) = \frac{1}{2\pi n} \int_{\mathbb{C}} \Delta f_{z_0,a}(z) \log|\det(X - z\mathbb{1})| \, \mathrm{d}^2 z.$$

To express the log-determinant of $X - z\mathbb{1}$ in terms of a Hermitian matrix, we introduce the $2n \times 2n$ matrix

$$(2.11) \qquad \boldsymbol{H}^z := \begin{pmatrix} 0 & X - z\mathbb{1} \\ X^* - \bar{z}\mathbb{1} & 0 \end{pmatrix}$$

for all $z \in \mathbb{C}$. Note that the eigenvalues of $\boldsymbol{H}^z$ come in opposite pairs and we denote them by $\lambda_{2n} \le \cdots \le \lambda_{n+1} \le 0 \le \lambda_n \le \cdots \le \lambda_1$ with $\lambda_i = -\lambda_{2n+1-i}$ for $i = 1, \ldots, 2n$. We remark that the moduli of these real numbers are the singular values of $X - z\mathbb{1}$. The Stieltjes transform of its empirical spectral measure is denoted by m^z, that is,

$$(2.12) \qquad m^z(w) = \frac{1}{2n} \sum_{i=1}^{2n} \frac{1}{\lambda_i(z) - w}$$

for $w \in \mathbb{C}$ satisfying $\operatorname{Im} w > 0$. It will turn out that on the imaginary axis $\operatorname{Im} m^z(i\eta)$ is very well approximated by $\langle v_1^\tau(\eta) \rangle = \langle v_2^\tau(\eta) \rangle$, where $\tau = |z|^2$ and (v_1^τ, v_2^τ) is the solution of (2.4). This fact is commonly called a *local law* for $\boldsymbol{H}^z$. With this notation, we have the following relation between the determinant of $X - z\mathbb{1}$ and the determinant of $\boldsymbol{H}^z$:

$$(2.13) \qquad \log|\det(X - z\mathbb{1})| = \frac{1}{2} \log|\det \boldsymbol{H}^z|.$$

We write the log-determinant in terms of the Stieltjes transform (this formula was used by Tao and Vu [31] in a similar context)

$$(2.14) \qquad \log|\det \boldsymbol{H}^z| = \log|\det(\boldsymbol{H}^z - iT\mathbb{1})| - 2n \int_0^T \operatorname{Im} m^z(i\eta) \, d\eta,$$

for any $T > 0$. Combining (2.5), (2.10), (2.13) and (2.14) as well as subtracting $1/(1+\eta)$ freely and using integration by parts, we obtain

$$\frac{1}{n} \sum_{i=1}^n f_{z_0,a}(\sigma_i) - \int_{\mathbb{C}} f_{z_0,a}(z)\sigma(z) \, d^2z$$

$$(2.15) \qquad \begin{aligned} &= \frac{1}{4\pi n} \int_{\mathbb{C}} \Delta f_{z_0,a}(z) \log|\det(\boldsymbol{H}^z - iT\mathbb{1})| \, d^2z \\ &\quad - \frac{1}{2\pi} \int_{\mathbb{C}} \Delta f_{z_0,a}(z) \int_0^T \left[\operatorname{Im} m^z(i\eta) - \langle v_1^\tau(\eta)|_{\tau=|z|^2} \rangle \right] d\eta \, d^2z \\ &\quad + \frac{1}{2\pi} \int_{\mathbb{C}} \Delta f_{z_0,a}(z) \int_T^\infty \left(\langle v_1^\tau(\eta)|_{\tau=|z|^2} \rangle - \frac{1}{\eta + 1} \right) d\eta \, d^2z. \end{aligned}$$

The task is then to prove that each of the terms on the right-hand side of (2.15) is dominated by $n^{-1+2a} \|\Delta f\|_1$ with very high probability. The parameter T will be chosen to be a large power of n, so that the first and the third term will easily

satisfy this bound. Estimating the second term on the right-hand side of (2.15) is much more involved and we focus only on this term in this outline.

We split its $d\eta$-integral into two parts. For $\eta \leq n^{-1+\varepsilon}$, $\varepsilon \in (0, 1/2)$, the integral is controlled by an estimate on the smallest singular value of $X - z\mathbb{1}$. This is the only step in our proof which uses assumption (C), that is, the entries of X have bounded densities (2.3).

For $\eta \geq n^{-1+\varepsilon}$, we use a local law for $\boldsymbol{H}^z$, that is, an optimal pointwise estimate (up to negligible n^ε-factors) on

$$(2.16) \qquad \operatorname{Im} m^z(i\eta) - \langle v_1^\tau(\eta)|_{\tau=|z|^2}\rangle,$$

uniformly in η and z (see Theorem 5.2 for the precise formulation). Note that a local law for $\boldsymbol{H}^z$ is needed only at spectral parameters on the imaginary axis. This will simplify the proof of the local law we need in this paper.

The proof of the local law is based on a stability estimate of (2.4). To write these equations in a more concise form, we introduce the $2n \times 2n$ matrices:

$$(2.17) \qquad \boldsymbol{S}_o = \begin{pmatrix} 0 & S \\ S^t & 0 \end{pmatrix}, \qquad \boldsymbol{S}_d = \begin{pmatrix} S^t & 0 \\ 0 & S \end{pmatrix}.$$

With this notation, the system of equations (2.4) can be written as

$$(2.18) \qquad i\boldsymbol{v} + \left(i\eta + \boldsymbol{S}_o i\boldsymbol{v} - \frac{\tau}{i\eta + \boldsymbol{S}_d i\boldsymbol{v}}\right)^{-1} = 0,$$

where we introduced $\boldsymbol{v} := (v_1, v_2) \in \mathbb{R}^{2n}$.

Let $\boldsymbol{G}^z(\eta) := (\boldsymbol{H}^z - i\eta\mathbb{1})^{-1}$, $\eta > 0$, be the resolvent of $\boldsymbol{H}^z$ at spectral parameter $i\eta$. We will prove that its diagonal $\boldsymbol{g}(\eta) = (\langle \boldsymbol{e}_i, \boldsymbol{G}^z(\eta)\boldsymbol{e}_i\rangle)_{i=1}^{2n}$, where $\boldsymbol{e}_i$ denotes the ith standard basis vector in $\mathbb{C}^{2n}$, satisfies a perturbed version of (2.18),

$$(2.19) \qquad \boldsymbol{g} + \left(i\eta + \boldsymbol{S}_o \boldsymbol{g} - \frac{\tau}{i\eta + \boldsymbol{S}_d \boldsymbol{g}}\right)^{-1} = \boldsymbol{d},$$

with $\tau = |z|^2$ and a small random error term $\boldsymbol{d}$. As $m^z(i\eta) = \langle \boldsymbol{g}(\eta) \rangle$ [cf. (2.12)] obtaining a local law, that is, an optimal pointwise estimate on (2.16), reduces to a stability problem for the *Dyson equation* (2.18).

Computing the difference of (2.19) and (2.18), we obtain

$$(2.20) \qquad \boldsymbol{L}(\boldsymbol{g} - i\boldsymbol{v}) = \boldsymbol{r}$$

for some error vector $\boldsymbol{r} = O(\|\boldsymbol{d}\|)$ [for the precise definition we refer to (3.24) below] and with the matrix $\boldsymbol{L}$ defined through its action on $\boldsymbol{y} \in \mathbb{C}^{2n}$ via

$$(2.21) \qquad \boldsymbol{L}\boldsymbol{y} := \boldsymbol{y} + \boldsymbol{v}^2(\boldsymbol{S}_o \boldsymbol{y}) - \tau \frac{\boldsymbol{v}^2}{(\eta + \boldsymbol{S}_d \boldsymbol{v})^2}(\boldsymbol{S}_d \boldsymbol{y}).$$

Therefore, a bound on $\boldsymbol{g} - i\boldsymbol{v}$ uniformly for $\eta \geq n^{-1+\varepsilon}$ requires a uniform bound on the inverse of $\boldsymbol{L}$ down to this local spectral scale.

In fact, the mere invertibility of L even for η bounded away from zero is a nontrivial fact that is not easily seen from (2.21). In Section 3, we will factorize L into the form

$$L = V^{-1}(\mathbb{1} - TF)V$$

for some invertible matrix V and self-adjoint matrices T and F with the properties $\|T\|_2 = 1$ and $\|F\|_2 \leq 1 - c\eta$ for some $c > 0$. In particular, this representation shows the a priori bound $\|L^{-1}\|_2 \leq C\eta^{-1}$ for some $C > 0$. The blow-up in the norm of L^{-1} is potentially caused by the two extremal eigendirections f_+ and f_- of F, which satisfy

$$F f_\pm = \pm\|F\|_2 f_\pm.$$

However, it turns out that the positivity of the solutions v_1, v_2 of (2.4) implies that $\|T f_+\|_2$ is strictly smaller than 1, so that $\|(\mathbb{1} - TF)f_+\|_2 \geq c\|f_+\|_2$ for some constant $c > 0$. In this sense, the solution of the Dyson equation regularizes the potentially unstable direction f_+.

In contrast, the other instability caused by f_- persists since we will find that $(\mathbb{1} - TF)f_- = O(\eta)$. This problem can only be resolved by exploiting an extra cancellation that originates from the special structure of the random matrix H^z. The leading contribution of the random error $r = O(\|d\|)$ from (2.20) pointing in the unstable direction happens to vanish with a remaining subleading term of order $\eta\|d\|$. The extra η-factor cancels the η^{-1}-divergence of $\|L^{-1}\|_2$ and allows us to invert the stability operator L in (2.20).

From this analysis, we conclude $\|g - iv\| \leq C\|d\|$. This result allows us to follow the general arguments developed in [3] for verifying the optimal local law for H^z. These steps are presented only briefly in Section 5.

3. Dyson equation for the inhomogeneous circular law. As explained in Section 2.1 a main ingredient in the proof of Theorem 2.5 is the local law for the self-adjoint random matrix H^z with noncentered independent entries above the diagonal. In [3], such a local law was proven for a large class of self-adjoint random matrices with noncentered entries and general short range correlations. For any fixed $z \in \mathbb{C}$, the matrix H^z satisfies the assumptions made for the class of random matrices covered in [3] with one crucial exception: H^z is not *flat* (cf. (2.28) in [3]), that is, for any constant $c > 0$, the inequality

$$(3.1) \qquad \frac{1}{n}\mathbb{E}|\langle a, (H - \mathbb{E}H)b\rangle|^2 \geq c\|a\|_2^2\|b\|_2^2,$$

is not satisfied for $H = H^z$ and vectors a, b that both have support either in $\{1, \ldots, n\}$ or $\{n + 1, \ldots, 2n\}$. Nevertheless, we will show that the conclusion of Theorem 2.9 of [3] remains true for spectral parameters $i\eta$ on the imaginary axis,

namely that the resolvent $G^z(\eta) := (H^z - i\eta\mathbb{1})^{-1}$ approaches the solution $M^z(\eta)$ of the MDE

$$(3.2) \qquad -M^z(\eta)^{-1} = i\eta\mathbb{1} - A^z + \mathcal{S}[M^z(\eta)], \qquad \eta > 0,$$

as $n \to \infty$. In fact, the solution of (3.2) is unique under the constraint that the imaginary part $\operatorname{Im} M := (M - M^*)/(2i)$ is positive definite [24]. The data $A^z \in \mathbb{C}^{2n \times 2n}$ and $\mathcal{S}: \mathbb{C}^{2n \times 2n} \to \mathbb{C}^{2n \times 2n}$ determining (3.2) are given in terms of the first and second moments of the entries of H^z,

$$(3.3) \quad A^z := \mathbb{E}H^z = \begin{pmatrix} 0 & -z \\ -\bar{z} & 0 \end{pmatrix}, \qquad \mathcal{S}[W] := \begin{pmatrix} \operatorname{diag}(Sw_2) & 0 \\ 0 & \operatorname{diag}(S^t w_1) \end{pmatrix},$$

for an arbitrary $2n \times 2n$ matrix $W = (w_{ij})_{i,j=1}^{2n}$ with

$$(3.4) \qquad W = \begin{pmatrix} W_{11} & W_{12} \\ W_{21} & W_{22} \end{pmatrix}, \qquad w_1 := (w_{ii})_{i=1}^{n}, \qquad w_2 := (w_{ii})_{i=n+1}^{2n}.$$

In the following, we will not keep the z-dependence in our notation and just write M, A and G instead of M^z, A^z and G^z. A simple calculation (cf. the proof of Lemma 2.2 in the Appendix) shows that $M: \mathbb{R}_+ \to \mathbb{C}^{2n \times 2n}$ is given by

$$(3.5) \qquad M^z(\eta) := \begin{pmatrix} i\operatorname{diag}(v_1^\tau(\eta)) & -z\operatorname{diag}(u^\tau(\eta)) \\ -\bar{z}\operatorname{diag}(u^\tau(\eta)) & i\operatorname{diag}(v_2^\tau(\eta)) \end{pmatrix},$$

where $z \in \mathbb{C}$, $\tau = |z|^2$, (v_1^τ, v_2^τ) is the solution of (2.4) and $u^\tau := v_1^\tau/(\eta + S^t v_1^\tau)$. In this section, we will therefore analyze the solution and the stability of (2.4).

3.1. *Analysis of the Dyson equation* (2.4). Combining the equations in (2.4), recalling $v = (v_1, v_2)$ and the definitions of S_o and S_d in (2.17), we obtain

$$(3.6) \qquad \frac{1}{v} = \eta + S_o v + \frac{\tau}{\eta + S_d v}$$

for $\eta > 0$ and $\tau \in \mathbb{R}_0^+$, where $v: \mathbb{R}_+ \to \mathbb{R}_+^{2n}$. This equation is equivalent to (2.18). The τ-dependence of v, v_1 and v_2 will mostly be suppressed but sometimes we view $v = v^\tau(\eta)$ as a function of both parameters.

Equation (3.6) has an obvious scaling invariance when S is rescaled to λS for $\lambda > 0$. If $v^\tau(\eta)$ is the positive solution of (3.6), then $v_\lambda^\tau(\eta) := \lambda^{-1/2} v^{\tau\lambda^{-1}}(\eta\lambda^{-1/2})$ is the positive solution of

$$(3.7) \qquad \frac{1}{v_\lambda} = \eta + \lambda S_o v_\lambda + \frac{\tau}{\eta + \lambda S_d v_\lambda}.$$

Therefore, without loss of generality, we may assume that the spectral radius of S is one,

$$\rho(S) = 1,$$

in the remainder of the paper.

The following proposition, the first main result of this section, collects some basic estimates on the solution $\boldsymbol{v}$ of (3.6). For the whole section, we fix $\tau_* > 0$ and $\tau^* > \tau_* + 1$ and except for Proposition 3.2, we exclude the small interval $[1 - \tau_*, 1 + \tau_*]$ from our analysis of $\boldsymbol{v}^\tau$. Because of the definition of σ in (2.5)—recall $\tau = |z|^2$ in the definition—we will talk about inside and outside regimes for $\tau \in [0, 1 - \tau_*]$ and $\tau \in [1 + \tau_*, \tau^*]$, respectively.

Recalling s_* and s^* from (2.1) we make the following convention in order to suppress irrelevant constants from the notation.

CONVENTION 3.1.   For nonnegative scalars or vectors f and g, we will use the notation $f \lesssim g$ if there is a constant $c > 0$, depending only on τ_*, τ^*, s_* and s^* such that $f \leq cg$ and $f \sim g$ if $f \lesssim g$ and $f \gtrsim g$ both hold true. If f, g and h are scalars or vectors and $h \geq 0$ such that $|f - g| \lesssim h$, then we write $f = g + O(h)$. Moreover, we define

$$\mathcal{P} := \{\tau_*, \tau^*, s_*, s^*\}$$

because many constants in the following will depend only on $\mathcal{P}$.

PROPOSITION 3.2.   *The solution $\boldsymbol{v}^\tau$ of (3.6) satisfies*

$$(3.8) \qquad \langle v_1^\tau(\eta) \rangle = \langle v_2^\tau(\eta) \rangle$$

for all $\eta > 0$ and $\tau \in \mathbb{R}_0^+$ as well as the following estimates:

   (i) (*Large η*) *Uniformly, for $\eta \geq 1$ and $\tau \in [0, \tau^*]$, we have*

$$(3.9) \qquad \boldsymbol{v}^\tau(\eta) \sim \eta^{-1}.$$

   (ii) (*Inside regime*) *Uniformly, for $\eta \leq 1$ and $\tau \in [0, 1]$, we have*

$$(3.10) \qquad \boldsymbol{v}^\tau(\eta) \sim \eta^{1/3} + (1 - \tau)^{1/2}.$$

   (iii) (*Outside regime*) *Uniformly, for $\eta \leq 1$ and $\tau \in [1, \tau^*]$, we have*

$$(3.11) \qquad \boldsymbol{v}^\tau(\eta) \sim \frac{\eta}{\tau - 1 + \eta^{2/3}}.$$

PROOF.   We start with proving (3.8). By multiplying (2.4a) by $(\eta + S^t v_1)$ and (2.4b) by $(\eta + S v_2)$ and realizing that both right-hand sides agree, we obtain

$$(3.12) \qquad \frac{v_1}{\eta + S^t v_1} = \frac{v_2}{\eta + S v_2}.$$

From (3.12), we also get

$$0 = \eta(v_1 - v_2) + v_1 S v_2 - v_2 S^t v_1.$$

We take the average on both sides, use $\langle v_1 S v_2 \rangle = \langle v_1, S v_2 \rangle = \langle v_2 S^t v_1 \rangle$ and divide by $\eta > 0$ to infer (3.8).

From (2.1), we immediately deduce the following auxiliary bounds:

$$(3.13) \qquad \langle v_1 \rangle \lesssim S^t v_1 \lesssim \langle v_1 \rangle, \qquad \langle v_2 \rangle \lesssim S v_2 \lesssim \langle v_2 \rangle.$$

We start with establishing $v \sim \langle v \rangle$. Since the entries of S are strictly positive and $\rho(S) = 1$ there is a unique vector $p \in \mathbb{R}_+^n$, which has strictly positive entries such that

$$(3.14) \qquad Sp = p, \qquad \langle p \rangle = 1, \qquad p \sim 1$$

by the Perron–Frobenius theorem and (2.1). We multiply (2.4a) by v_1 as well as $\eta + S^t v_1$ and obtain $\eta + S^t v_1 = v_1(\eta + Sv_2)(\eta + S^t v_1) + \tau v_1$. Taking the scalar product with p and using $\langle p \rangle = 1$ and $\rho(S) = 1$ yield

$$(3.15) \qquad \eta + \langle pv_1 \rangle = \langle pv_1(\eta + S^t v_1)(\eta + Sv_2) \rangle + \tau \langle pv_1 \rangle.$$

Therefore, (3.13), $\langle v_1 \rangle = \langle v_2 \rangle = \langle v \rangle$ by (3.8) and (3.14) imply

$$(3.16) \qquad \eta + \langle v \rangle \sim \left[(\eta + \langle v \rangle)^2 + \tau \right] \langle v \rangle.$$

We use (3.13) in (2.4a) and (2.4b) to conclude

$$(3.17) \qquad v \sim \frac{1}{\eta + \langle v \rangle + \frac{\tau}{\eta + \langle v \rangle}} = \frac{\eta + \langle v \rangle}{(\eta + \langle v \rangle)^2 + \tau} \sim \langle v \rangle,$$

where we applied (3.16) in the last step. Hence, it suffices to prove all estimates (3.9), (3.10) and (3.11) for v replaced by $\langle v \rangle$ only.

We start with an auxiliary upper bound on $\langle v \rangle$. By multiplying (3.6) with v, we get $1 = \eta v + vS_o v + \tau v/(\eta + S_d v) \geq vS_o v$. Hence, $1 \geq \langle v_1 S v_2 \rangle \gtrsim \langle v_1 \rangle \langle v_2 \rangle = \langle v \rangle^2$, where we used (3.13) in the second step and (3.8) in the last step.

Next, we show (3.9). Clearly, (3.6) implies $v \leq \eta^{-1}$. Moreover, as $\tau \leq \tau^*$ and $\eta \geq 1 \gtrsim \langle v \rangle$ we find $\eta \lesssim \eta^2 \langle v \rangle$ from (3.16). This gives the lower bound on v in (3.9) when combined with (3.17).

We note that (3.16) immediately implies $\langle v \rangle \gtrsim \eta$ for $\eta \leq 1$. Now, we show (3.10). For $\tau \in [0, 1]$, we bring the term $\tau \langle pv_1 \rangle$ to the left-hand side in (3.15) and use $v_1 \sim v_2 \sim \langle v \rangle$ and (3.13) as well as $\langle v \rangle \gtrsim \eta$ to obtain

$$(3.18) \qquad \eta + (1 - \tau) \langle v \rangle \sim \langle v \rangle^3.$$

From (3.18), it is an elementary exercise to conclude (3.10) for $\eta \leq 1$.

Similarly, for $1 \leq \tau \leq \tau^*$, we bring $\langle pv_1 \rangle$ to the right-hand side of (3.15), use $\langle v \rangle \gtrsim \eta$ for $\eta \leq 1$ and conclude

$$(3.19) \qquad \eta \sim \langle v \rangle^3 + (\tau - 1) \langle v \rangle.$$

As before, it is easy to conclude (3.11) from (3.19). We leave this to the reader. This completes the proof of Proposition 3.2. $\square$

Our next goal is a stability result for (3.6) in the regime $\tau \in [0, 1 - \tau_*] \cup [1 + \tau_*, \tau^*]$. In the following proposition, the second main result of this section, we

prove that $i\boldsymbol{v}(\eta)$ well approximates $\boldsymbol{g}(\eta)$ for all $\eta > 0$ if $\boldsymbol{g}$ satisfies (2.19) and as long as $\boldsymbol{d}$ is small. However, we will need an additional assumption on $\boldsymbol{g} = (g_1, g_2)$, namely that $\langle g_1 \rangle = \langle g_2 \rangle$ [see (3.20) below]. Note that this is imposed on the solution $\boldsymbol{g}$ of (2.19) and not directly on the perturbation $\boldsymbol{d}$. Nevertheless, in our applications, the constraint (3.20) will be automatically satisfied owing to the specific block structure of the matrix $\boldsymbol{H}^z$ from (2.11).

PROPOSITION 3.3 (Stability). *Suppose that some functions* $\boldsymbol{d} \colon \mathbb{R}_+ \to \mathbb{C}^{2n}$ *and* $\boldsymbol{g} = (g_1, g_2) \colon \mathbb{R}_+ \to \mathbb{H}^{2n}$ *satisfy* (2.19) *and*

$$(3.20) \qquad\qquad \langle g_1(\eta) \rangle = \langle g_2(\eta) \rangle$$

for all $\eta > 0$. *There is a number* $\lambda_* \gtrsim 1$, *depending only on* $\mathcal{P}$, *such that*

$$(3.21) \qquad \big\| \boldsymbol{g}(\eta) - i\boldsymbol{v}(\eta) \big\|_\infty \cdot \chi \big(\big\| \boldsymbol{g}(\eta) - i\boldsymbol{v}(\eta) \big\|_\infty \leq \lambda_* \big) \lesssim \big\| \boldsymbol{d}(w) \big\|_\infty$$

uniformly for $\eta > 0$ *and* $\tau \in [0, 1 - \tau_*] \cup [1 + \tau_*, \tau^*]$.

Moreover, there is a matrix-valued function $\boldsymbol{R} \colon \mathbb{R}_+ \to \mathbb{C}^{2n \times 2n}$, *depending only on* τ *and* S *and satisfying* $\| \boldsymbol{R}(\eta) \|_\infty \lesssim 1$, *such that*

$$(3.22) \qquad \begin{aligned} &\big| \langle \boldsymbol{y}, \boldsymbol{g}(\eta) - i\boldsymbol{v}(\eta) \rangle \big| \cdot \chi \big(\big\| \boldsymbol{g}(\eta) - i\boldsymbol{v}(\eta) \big\|_\infty \leq \lambda_* \big) \\ &\qquad \lesssim \| \boldsymbol{y} \|_\infty \big\| \boldsymbol{d}(\eta) \big\|_\infty^2 + \big| \langle \boldsymbol{R}(\eta) \boldsymbol{y}, \boldsymbol{d}(\eta) \rangle \big| \end{aligned}$$

uniformly for all $\boldsymbol{y} \in \mathbb{C}^{2n}$, $\eta > 0$ *and* $\tau \in [0, 1 - \tau_*] \cup [1 + \tau_*, \tau^*]$.

The proof of this result is based on deriving a quadratic equation for the difference $\boldsymbol{h} := \boldsymbol{g} - i\boldsymbol{v}$ and establishing a quantitative estimate on $\boldsymbol{h}$ in terms of the perturbation $\boldsymbol{d}$. Computing the difference of (2.19) and (2.18), we obtain an equation for $\boldsymbol{g} - i\boldsymbol{v}$. A straightforward calculation yields

$$(3.23) \qquad\qquad \boldsymbol{L}\boldsymbol{h} = \boldsymbol{r}, \qquad \text{for } \boldsymbol{h} = \boldsymbol{g} - i\boldsymbol{v},$$

where we used $\boldsymbol{L}$ defined in (2.21) and introduced the vector $\boldsymbol{r}$ through

$$(3.24) \qquad \boldsymbol{r} := \boldsymbol{d} + i\boldsymbol{v}(\boldsymbol{h} - \boldsymbol{d}) S_o \boldsymbol{h} - \tau \boldsymbol{u} \bigg[\frac{\boldsymbol{d} - \boldsymbol{g}}{i\eta + S_d \boldsymbol{g}} + \boldsymbol{u} \bigg] S_d \boldsymbol{h}.$$

The vector $\boldsymbol{u}$ in (3.24) is defined through

$$(3.25) \qquad u := \frac{v_1}{\eta + S^t v_1} = \frac{v_2}{\eta + S v_2}, \qquad \boldsymbol{u} := (u, u) = \frac{\boldsymbol{v}}{\eta + S_d \boldsymbol{v}}$$

which is consistent by (3.12).

Notice that all terms on the right-hand side of (3.24) are either second order in $\boldsymbol{h}$ or they are of order $\boldsymbol{d}$, so (3.23) is the linearization of (2.19) around (2.18).

In the following estimates, we need a bound on $\boldsymbol{u}$ as well. Indeed, Proposition 3.2 yields

$$(3.26) \qquad\qquad \boldsymbol{u} = \frac{\boldsymbol{v}}{\eta + S_d \boldsymbol{v}} \sim \frac{1}{1 + \eta^2}$$

uniformly for $\eta > 0$ and $\tau \in [0, \tau^*]$.

To shorten the upcoming relations, we introduce the vector

$$\widetilde{\boldsymbol{v}} := (v_2, v_1)$$

and the matrices $\boldsymbol{T}$, $\boldsymbol{F}$ and $\boldsymbol{V}$ defined by their action on a vector $\boldsymbol{y} = (y_1, y_2)$, $y_1, y_2 \in \mathbb{C}^n$ as follows:

$$(3.27a) \qquad \boldsymbol{T}\boldsymbol{y} := \frac{1}{\boldsymbol{u}}\begin{pmatrix} -v_1 v_2 y_1 + \tau u^2 y_2 \\ \tau u^2 y_1 - v_1 v_2 y_2 \end{pmatrix},$$

$$(3.27b) \qquad \boldsymbol{F}\boldsymbol{y} := \sqrt{\frac{\boldsymbol{v}\boldsymbol{u}}{\widetilde{\boldsymbol{v}}}}\,\boldsymbol{S}_o\!\left(\sqrt{\frac{\boldsymbol{v}\boldsymbol{u}}{\widetilde{\boldsymbol{v}}}}\,\boldsymbol{y}\right),$$

$$(3.27c) \qquad \boldsymbol{V}\boldsymbol{y} := \sqrt{\frac{\widetilde{\boldsymbol{v}}}{\boldsymbol{u}\boldsymbol{v}}}\,\boldsymbol{y}.$$

All these matrices are functions of η and τ. They provide a crucial factorization of the stability operator $\boldsymbol{L}$; indeed, a simple calculation shows that

$$(3.28) \qquad \boldsymbol{L} = \boldsymbol{V}^{-1}(\mathbb{1} - \boldsymbol{T}\boldsymbol{F})\boldsymbol{V}.$$

This factorization reveals many properties of $\boldsymbol{L}$ which are difficult to observe directly. Owing to (3.23), the stability analysis of (3.6) requires a control on the invertibility of the matrix $\boldsymbol{L}$. The matrices $\boldsymbol{V}$ and $\boldsymbol{V}^{-1}$ are harmless. A good understanding of the spectral decompositions of the simpler matrices $\boldsymbol{F}$ and $\boldsymbol{T}$ will then yield that $\boldsymbol{L}$ has only one direction, in which its inverse is not bounded. We remark that the factorization (3.28) is the diagonal part of the one used in the stability analysis of the matrix Dyson equation in [3].

Because of (3.28), we can study the stability of

$$(3.29) \qquad (\mathbb{1} - \boldsymbol{T}\boldsymbol{F})(\boldsymbol{V}\boldsymbol{h}) = \boldsymbol{V}\boldsymbol{r}$$

instead of (3.23). From Proposition 3.2 and (3.26), we conclude that

$$(3.30) \qquad \|\boldsymbol{V}\|_\infty \|\boldsymbol{V}^{-1}\|_\infty \lesssim 1$$

uniformly for all $\eta > 0$ and $\tau \in [0, 1 - \tau_*] \cup [1 + \tau_*, \tau^*]$. Hence, it suffices to control the invertibility of $\mathbb{1} - \boldsymbol{T}\boldsymbol{F}$.

For later usage, we derive two relations for $\boldsymbol{u}$. From (3.25), recalling $\widetilde{\boldsymbol{v}} = (v_2, v_1)$, we immediately get

$$(3.31) \qquad \frac{\widetilde{\boldsymbol{v}}}{\boldsymbol{u}} = \eta + \boldsymbol{S}_o \boldsymbol{v}.$$

We multiply (3.6) by $\boldsymbol{v}\boldsymbol{u}$ and use (3.31) to obtain

$$(3.32) \qquad \boldsymbol{u} = \boldsymbol{v}\widetilde{\boldsymbol{v}} + \tau u^2, \qquad 1 = \frac{\boldsymbol{v}\widetilde{\boldsymbol{v}}}{\boldsymbol{u}} + \tau \boldsymbol{u}.$$

The next lemma collects some properties of $\boldsymbol{F}$. For this formulation, we introduce

$$\boldsymbol{e}_- := (1, -1) \in \mathbb{C}^{2n}.$$

LEMMA 3.4 (Spectral properties of $\boldsymbol{F}$).   *The eigenspace of $\boldsymbol{F}$ corresponding to its largest eigenvalue $\|\boldsymbol{F}\|_2$ is one dimensional. It is spanned by a unique positive normalized eigenvector $\boldsymbol{f}_+$, that is, $\boldsymbol{F}\boldsymbol{f}_+ = \|\boldsymbol{F}\|_2\boldsymbol{f}_+$ and $\|\boldsymbol{f}_+\|_2 = 1$. For every $\eta > 0$, the norm of $\boldsymbol{F}$ is given by*

$$(3.33) \qquad \|\boldsymbol{F}\|_2 = 1 - \eta \frac{\langle \boldsymbol{f}_+ \sqrt{\boldsymbol{v}/(\eta + S_o\boldsymbol{v})}\rangle}{\langle \boldsymbol{f}_+ \sqrt{\boldsymbol{v}(\eta + S_o\boldsymbol{v})}\rangle}.$$

Defining $\boldsymbol{f}_- := \boldsymbol{f}_+ \boldsymbol{e}_-$, we have

$$(3.34) \qquad \boldsymbol{F}\boldsymbol{f}_- = -\|\boldsymbol{F}\|_2\boldsymbol{f}_-.$$

(i) (*Inside regime*) *The following estimates hold true uniformly for $\tau \in [0, 1 - \tau_*]$. We have*

$$(3.35) \qquad 1 - \|\boldsymbol{F}\|_2 \sim \eta$$

uniformly for $\eta \in (0, 1]$. Furthermore, uniformly for $\eta \geq 1$, we have

$$(3.36) \qquad 1 - \|\boldsymbol{F}\|_2 \sim 1.$$

   Moreover, uniformly for $\eta \in (0, 1]$, $\boldsymbol{f}_+$ satisfies

$$(3.37) \qquad \boldsymbol{f}_+ \sim 1$$

and there is $\varepsilon \sim 1$ such that

$$(3.38) \qquad \|\boldsymbol{F}\boldsymbol{x}\|_2 \leq (1 - \varepsilon)\|\boldsymbol{x}\|_2$$

for all $\boldsymbol{x} \in \mathbb{C}^{2n}$ satisfying $\boldsymbol{x} \perp \boldsymbol{f}_+$ and $\boldsymbol{x} \perp \boldsymbol{f}_-$.
   (ii) (*Outside regime*) *Uniformly for all $\eta > 0$ and $\tau \in [1 + \tau_*, \tau^*]$, we have*

$$(3.39) \qquad 1 - \|\boldsymbol{F}\|_2 \sim 1.$$

PROOF.   The statements about the eigenspace corresponding to $\|\boldsymbol{F}\|_2$ and $\boldsymbol{f}_+$ follow from Lemma 3.3 in [7].

For the proof of (3.33), we multiply (3.6) by $\boldsymbol{v}$ and take the scalar product of the resulting relation with $\boldsymbol{f}_+ \sqrt{\boldsymbol{u}/(\boldsymbol{v}\widetilde{\boldsymbol{v}})}$. Using $S_o = S_o^t$ and the definition of $\boldsymbol{F}$ in (3.27b), we get

$$\left\langle \boldsymbol{f}_+ \sqrt{\frac{\boldsymbol{u}}{\boldsymbol{v}\widetilde{\boldsymbol{v}}}}, \boldsymbol{v}S_o\boldsymbol{v}\right\rangle = \left\langle S_o\left(\boldsymbol{f}_+\sqrt{\frac{\boldsymbol{v}\boldsymbol{u}}{\widetilde{\boldsymbol{v}}}}\right), \boldsymbol{v}\right\rangle = \|\boldsymbol{F}\|_2\left\langle \boldsymbol{f}_+, \sqrt{\frac{\boldsymbol{v}\widetilde{\boldsymbol{v}}}{\boldsymbol{u}}}\right\rangle.$$

This yields

$$\|\boldsymbol{F}\|_2\left\langle \boldsymbol{f}_+, \sqrt{\frac{\boldsymbol{v}\widetilde{\boldsymbol{v}}}{\boldsymbol{u}}}\right\rangle = \left\langle \boldsymbol{f}_+\sqrt{\frac{\boldsymbol{u}}{\boldsymbol{v}\widetilde{\boldsymbol{v}}}}, 1 - \tau\boldsymbol{u}\right\rangle - \eta\left\langle \boldsymbol{f}_+\sqrt{\frac{\boldsymbol{u}}{\boldsymbol{v}\widetilde{\boldsymbol{v}}}}, \boldsymbol{v}\right\rangle.$$

We conclude (3.33) from applying (3.32) and (3.31) to the last relation.

Since $\boldsymbol{F}$ from (3.27b) has the form

$$\boldsymbol{F} = \begin{pmatrix} 0 & F \\ F^t & 0 \end{pmatrix},$$

for some $F \in \mathbb{C}^{n \times n}$ we have $\boldsymbol{F}(\boldsymbol{e}_- \boldsymbol{y}) = -\boldsymbol{e}_-(\boldsymbol{F}\boldsymbol{y})$ for all $\boldsymbol{y} \in \mathbb{C}^{2n}$. Thus, we get (3.34) from $\boldsymbol{F}\boldsymbol{f}_+ = \|\boldsymbol{F}\|_2 \boldsymbol{f}_+$.

In the regime $\tau \in [0, 1 - \tau_*]$ and $\eta \in (0, 1]$, we have uniform lower and upper bounds on $\boldsymbol{v}$ from Proposition 3.2. Therefore, the estimates in (3.37) and (3.38) follow from Lemma 3.3 in [7]. Combining (3.37), (3.33) and Proposition 3.2 yields (3.35). In the large η regime, that is, for $\eta \geq 1$, since $\boldsymbol{v} \sim \eta^{-1}$ by Proposition 3.2 we obtain

$$(3.40) \qquad \frac{\boldsymbol{v}}{\eta + S_o \boldsymbol{v}} \sim \eta^{-2}, \qquad \boldsymbol{v}(\eta + S_o \boldsymbol{v}) \sim 1.$$

Hence, as $\boldsymbol{f}_+ > 0$ we conclude

$$(3.41) \qquad \frac{\langle \boldsymbol{f}_+ \sqrt{\boldsymbol{v}/(\eta + S_o \boldsymbol{v})} \rangle}{\langle \boldsymbol{f}_+ \sqrt{\boldsymbol{v}(\eta + S_o \boldsymbol{v})} \rangle} \sim \frac{\langle \boldsymbol{f}_+ \rangle}{\langle \boldsymbol{f}_+ \rangle} \frac{1}{\eta} = \frac{1}{\eta},$$

uniformly for all $\eta \geq 1$. This shows that (3.36) holds true for all $\eta \geq 1$ and $\tau \in [0, 1 - \tau_*]$.

We now turn to the proof of (ii). If $\tau \in [1 + \tau_*, \tau^*]$, then $\boldsymbol{v} \sim \eta$ by (3.11) for $\eta \leq 1$ and, therefore,

$$\frac{\boldsymbol{v}}{\eta + S_o \boldsymbol{v}} \sim 1, \qquad \boldsymbol{v}(\eta + S_o \boldsymbol{v}) \sim \eta^2.$$

As $\boldsymbol{f}_+ > 0$, we thus have

$$(3.42) \qquad \eta \frac{\langle \boldsymbol{f}_+ \sqrt{\boldsymbol{v}/(\eta + S_o \boldsymbol{v})} \rangle}{\langle \boldsymbol{f}_+ \sqrt{\boldsymbol{v}(\eta + S_o \boldsymbol{v})} \rangle} \sim \frac{\langle \boldsymbol{f}_+ \rangle}{\langle \boldsymbol{f}_+ \rangle} = 1.$$

For $\eta \geq 1$, we argue as in (3.40) and (3.41) and arrive at the same conclusion (3.42). Thus, because of (3.33) the estimate (3.39) holds true for all $\eta > 0$ and $\tau \in [1 + \tau_*, \tau^*]$. $\quad \square$

Next, we give an approximation for the eigenvector $\boldsymbol{f}_-$ belonging to the isolated single eigenvalue $-\|\boldsymbol{F}\|_2$ of $\boldsymbol{F}$ by constructing an approximate eigenvector. For $\eta \leq 1$ and $\tau \in [0, 1 - \tau_*]$, we define

$$(3.43) \qquad \boldsymbol{a} := \frac{\boldsymbol{e}_-(V\boldsymbol{v})}{\|V\boldsymbol{v}\|_2}$$

which is normalized as $\|\boldsymbol{e}_-(V\boldsymbol{v})\|_2 = \|V\boldsymbol{v}\|_2$. We compute

$$(3.44) \qquad \boldsymbol{F}(V\boldsymbol{v}) = \sqrt{\frac{\boldsymbol{u}}{\boldsymbol{v}\widetilde{\boldsymbol{v}}}} \boldsymbol{v}(S_o \boldsymbol{v}) = \sqrt{\frac{\boldsymbol{v}\widetilde{\boldsymbol{v}}}{\boldsymbol{u}}} - \eta \boldsymbol{v} \sqrt{\frac{\boldsymbol{u}}{\boldsymbol{v}\widetilde{\boldsymbol{v}}}} = \|\boldsymbol{F}\|_2 V\boldsymbol{v} + O(\eta).$$

Here, we used $\boldsymbol{v}\boldsymbol{S}_o\boldsymbol{v} = -\eta\boldsymbol{v} + \boldsymbol{v}\widetilde{\boldsymbol{v}}/\boldsymbol{u}$ by (3.31). For estimating the $O(\eta)$ term we applied (3.10), (3.26) and (3.35) since $\tau \in [0, 1 - \tau_*]$ and $\eta \leq 1$. Using the block structure of $\boldsymbol{F}$ as in the proof of (3.34), we obtain

$$(3.45) \qquad \boldsymbol{F}\big(\boldsymbol{e}_-(\boldsymbol{V}\boldsymbol{v})\big) = -\|\boldsymbol{F}\|_2\boldsymbol{e}_-(\boldsymbol{V}\boldsymbol{v}) + O(\eta).$$

The following lemma states that $\boldsymbol{a}$ approximates the nondegenerate eigenvector $\boldsymbol{f}_-$.

LEMMA 3.5. *The eigenvector* $\boldsymbol{f}_-$ *can be approximated by* $\boldsymbol{a}$ *in the* ℓ^∞*-norm, that is,*

$$(3.46) \qquad \|\boldsymbol{f}_- - \boldsymbol{a}\|_\infty = O(\eta)$$

uniformly for $\eta \leq 1$ *and* $\tau \in [0, 1 - \tau_*]$.

This lemma is proved in Appendix B. In the following lemma, we show some properties of $\boldsymbol{T}$.

LEMMA 3.6 (Spectral properties of $\boldsymbol{T}$). *The symmetric operator* $\boldsymbol{T}$, *defined in* (3.27a), *satisfies*:

  (i) $\|\boldsymbol{T}\|_2 = 1$, $\|\boldsymbol{T}\|_\infty = 1$.
  (ii) *The spectrum of* $\boldsymbol{T}$ *is given by*

$$\mathrm{Spec}(\boldsymbol{T}) = \{-1\} \cup \left\{\tau\boldsymbol{u}_i - \frac{(\boldsymbol{v}\widetilde{\boldsymbol{v}})_i}{\boldsymbol{u}_i}\Big| i = 1, \ldots, n\right\}.$$

  (iii) *For all* $\eta > 0$, *we have* $\boldsymbol{T}(\tau = 0) = -\mathbb{1}$ *and if* $\tau > 0$, *then the eigenspace of* $\boldsymbol{T}$ *corresponding to the eigenvalue* -1 *is* n*-fold degenerate and given by*

$$(3.47) \qquad \mathrm{Eig}(-1, \boldsymbol{T}) = \big\{(y, -y) | y \in \mathbb{C}^n\big\}.$$

  (iv) *The spectrum of* $\boldsymbol{T}$ *is strictly away from one, that is, there is* $\varepsilon > 0$, *depending only on* $\mathcal{P}$, *such that*

$$(3.48) \qquad \mathrm{Spec}(\boldsymbol{T}) \subset [-1, 1 - \varepsilon]$$

uniformly for $\tau \in [0, 1 - \tau_*]$ *and* $\eta \in (0, 1]$.

PROOF. The second relation in (3.32) implies $\|\boldsymbol{T}\|_\infty = 1$ and $\boldsymbol{T}(\tau = 0) = -\mathbb{1}$. Moreover, it yields that all vectors of the form $(y, -y)$ for $y \in \mathbb{C}^n$ are contained in $\mathrm{Eig}(-1, \boldsymbol{T})$. We define the vector $\boldsymbol{y}^{(j)} \in \mathbb{C}^{2n}$ by $\boldsymbol{y}^{(j)} := (\delta_{i,j} + \delta_{i,j+n})_{i=1}^{2n}$ and observe that

$$\boldsymbol{T}\boldsymbol{y}^{(j)} = \left(\tau\boldsymbol{u}_j - \frac{(\boldsymbol{v}\widetilde{\boldsymbol{v}})_j}{\boldsymbol{u}_j}\right)\boldsymbol{y}^{(j)}$$

for $j = 1, \ldots, n$. Counting dimensions implies that we have found all eigenvalues, hence (ii) follows. For $\tau > 0$, we have $\tau\boldsymbol{u}_j - (\boldsymbol{v}\widetilde{\boldsymbol{v}})_j/\boldsymbol{u}_j = 2\tau\boldsymbol{u}_j - 1 > -1$ by (3.32)

and $u_j > 0$ for all $j = 1, \ldots, n$. This yields the missing inclusion in (3.47). Since T is a symmetric operator, $\|T\|_2 = 1$ follows from (ii) and $|\tau u - v\widetilde{v}/u| \leq 1$ by (3.32).

For the proof of (iv), we remark that there is $\varepsilon > 0$, depending only on $\mathcal{P}$, such that $2v\widetilde{v}/u \geq \varepsilon$ for all $\eta \in (0, 1]$ and $\tau \in [0, 1 - \tau_*]$ by (3.10) and (3.26). Thus,

$$\tau u - \frac{v\widetilde{v}}{u} = 1 - 2\frac{v\widetilde{v}}{u} \leq 1 - \varepsilon$$

by (3.32). This concludes the proof of the lemma. $\square$

Now we are ready to give a proof of Proposition 3.3 based on inverting $\mathbb{1} - TF$.

PROOF OF PROPOSITION 3.3. We recall that $h = g - iv$. Throughout the proof, we will omit arguments, but we keep in mind that g, d, h and v depend on η and τ. The proof will be given in three steps.

The first step is to control $\|r\|_\infty$ from (3.24) in terms of $\|h\|_\infty^2$ and $\|d\|_\infty$, that is, to show

(3.49)
$$\|r\|_\infty \chi(\|h\|_\infty \leq 1) \lesssim \|h\|_\infty^2 + \|d\|_\infty.$$

Inverting $V^{-1}(\mathbb{1} - TF)V$ in (3.29), controlling the norm of the inverse and choosing $\lambda_* \leq 1$ small enough, we will conclude Proposition 3.3 from (3.49). For any $\eta_* \in (0, 1]$, depending only on $\mathcal{P}$, this argument will be done in the second step for $\tau \in [0, 1 - \tau_*] \cup [1 + \tau_*, \tau^*]$ and $\eta \geq \eta_*$ as well as for $\tau \in [1 + \tau_*, \tau^*]$ and $\eta \in (0, \eta_*]$. In the third step, we consider the most interesting regime $\tau \in [0, 1 - \tau_*]$ and $\eta \leq \eta_*$ for a sufficiently small η_*, depending on $\mathcal{P}$ only. In this regime, we will use an extra cancellation for the contribution of r in the unstable direction of L.

Step 1: For all $\eta > 0$ and $\tau \in [0, 1 - \tau_*] \cup [1 + \tau_*, \tau^*]$, (3.49) holds true.

From (2.19), we obtain

$$\tau\frac{g - d}{i\eta + S_d g} = 1 + (i\eta + S_o g)(g - d).$$

We start from (3.24), use the previous relation, $\tau u = 1 + iv(i\eta + S_o iv)$ by (3.6) and $\widetilde{v} = (v_2, v_1) = u(\eta + S_o v)$ by (3.32) and get

$$r = d + iv(h - d)S_o h - u[iv(i\eta + S_o iv) - (g - d)(i\eta + S_o g)]S_d h$$

$$= d + iv(h - d)S_o h + u[h(i\eta + S_o iv) + gS_o h]S_d h$$

(3.50)
$$\quad - du(i\eta + S_o g)S_d h$$

$$= ivhS_o h + i\widetilde{v}hS_d h + ugS_o hS_d h$$

$$\quad + d - ivdS_o h - du(i\eta + S_o g)S_d h.$$

Notice that the first three terms are quadratic in $\boldsymbol{h}$ (the linear terms dropped out), while the last three terms are controlled by $\boldsymbol{d}$. Now, we show that all other factors are bounded, and hence irrelevant whenever $\|\boldsymbol{g} - i\boldsymbol{v}\|_\infty \le \lambda_*$ for $\eta > 0$ and $\tau \in [0, 1 - \tau_*] \cup [1 + \tau_*, \tau^*]$. In this case, we conclude $\|\boldsymbol{g}\|_\infty \lesssim 1$ uniformly for all $\eta > 0$ and $\tau \in [0, 1 - \tau_*] \cup [1 + \tau_*, \tau^*]$ by (3.9) and (3.10) from Proposition 3.2. Therefore, starting from (3.50) and using $\|\boldsymbol{v}\|_\infty \lesssim 1$ by (3.9) and (3.10), and $\|\boldsymbol{u}\|_\infty \lesssim 1$ by (3.26), we obtain (3.49).

Step 2: For any $\eta_* \in (0, 1]$, there exists $\lambda_* \gtrsim 1$, depending only on $\mathcal{P}$ and η_*, such that (3.21) holds true for $\eta \ge \eta_*$ and $\tau \in [0, 1 - \tau_*] \cup [1 + \tau_*, \tau^*]$ as well as for $\eta \in (0, \eta_*)$ and $\tau \in [1 + \tau_*, \tau^*]$. Moreover, with this choice of λ_*, (3.22) holds true in these (η, τ) parameter regimes as well.

Within Step 2, we redefine the comparison relation to depend both on $\mathcal{P}$ and η_*. Later in Step 3 we will choose an appropriate η_* depending only on $\mathcal{P}$, so eventually the comparison relations for our choice will depend only on $\mathcal{P}$.

We are now working in the regime, where $\eta \ge \eta_*$ and $\tau \in [0, 1 - \tau_*] \cup [1 + \tau_*, \tau^*]$ or $\eta \in (0, \eta_*)$ and $\tau \in [1 + \tau_*, \tau^*]$. In this case, to prove (3.21), we invert $\boldsymbol{L} = \boldsymbol{V}^{-1}(\mathbb{1} - \boldsymbol{T}\boldsymbol{F})\boldsymbol{V}$ [cf. (2.21)] in $\boldsymbol{L}\boldsymbol{h} = \boldsymbol{r}$, bound $\|\boldsymbol{L}^{-1}\|_\infty \lesssim 1$, which is proved below, and conclude

$$\|\boldsymbol{h}\|_\infty \chi(\|\boldsymbol{h}\|_\infty \le 1) \lesssim \|\boldsymbol{h}\|_\infty^2 + \|\boldsymbol{d}\|_\infty$$

from (3.49) for $\eta \ge \eta_*$ and $\tau \in [0, 1 - \tau_*] \cup [1 + \tau_*, \tau^*]$ as well as for $\eta \in (0, \eta_*)$ and $\tau \in [0, 1 - \tau_*]$. This means that there are $\Psi_1, \Psi_2 \sim 1$ such that

$$\|\boldsymbol{h}\|_\infty \chi(\|\boldsymbol{h}\|_\infty \le 1) \le \Psi_1 \|\boldsymbol{h}\|_\infty^2 + \Psi_2 \|\boldsymbol{d}\|_\infty.$$

Choosing $\lambda_* := \min\{1, (2\Psi_1)^{-1}\}$ this yields

$$\|\boldsymbol{h}\|_\infty \chi(\|\boldsymbol{h}\|_\infty \le \lambda_*) \le 2\Psi_2 \|\boldsymbol{d}\|_\infty.$$

Thus, we are left with controlling $\|\boldsymbol{L}^{-1}\|_\infty$, that is, proving $\|\boldsymbol{L}^{-1}\|_\infty \lesssim 1$.

In the regime $\eta \ge \eta_*$ and $\tau \in [0, 1 - \tau_*] \cup [1 + \tau_*, \tau^*]$, we have $\boldsymbol{v} \sim 1/\eta$ by Proposition 3.2 and $\boldsymbol{u} \sim 1/\eta^2$ by (3.26). Hence, $\boldsymbol{V} \sim \eta$ and $\boldsymbol{V}^{-1} \sim 1/\eta$. Therefore, $\|\boldsymbol{V}\|_\infty \|\boldsymbol{V}^{-1}\|_\infty \lesssim 1$ and due to $\|\boldsymbol{L}^{-1}\|_\infty \lesssim \|\boldsymbol{V}^{-1}\|_\infty \|(\mathbb{1} - \boldsymbol{T}\boldsymbol{F})^{-1}\|_\infty \|\boldsymbol{V}\|_\infty$, it suffices to show $\|(\mathbb{1} - \boldsymbol{T}\boldsymbol{F})^{-1}\|_\infty \lesssim 1$. Basic facts on the operator $\mathbb{1} - \boldsymbol{T}\boldsymbol{F}$ are collected in Lemma B.1 in the Appendix. In particular, because of (B.9), the ℓ^∞ bound follows from $\|(\mathbb{1} - \boldsymbol{T}\boldsymbol{F})^{-1}\|_2 \lesssim 1$. Using (3.35), (3.36) and (3.39), we get that $1 - \|\boldsymbol{F}\|_2 \sim 1$ for all $\eta \ge \eta_*$ and $\tau \in [0, 1 - \tau_*] \cup [1 + \tau_*, \tau^*]$. Hence, $1 - \|\boldsymbol{T}\boldsymbol{F}\|_2 \sim 1$ by Lemma 3.6(i), so the bound $\|(\mathbb{1} - \boldsymbol{T}\boldsymbol{F})^{-1}\|_2 \lesssim 1$ immediately follows. This proves (3.21) for $\eta \ge \eta_*$ and $\tau \in [0, 1 - \tau_*] \cup [1 + \tau_*, \tau^*]$.

For $\eta \le \eta_*$ and $\tau \in [1 + \tau_*, \tau^*]$, we have $\boldsymbol{v} \sim \eta$ by (3.11), $\boldsymbol{u} \sim 1$ by (3.26). Thus, $\boldsymbol{V} \sim 1$, $\boldsymbol{V}^{-1} \sim 1$ as well as $\|\boldsymbol{V}\|_\infty \|\boldsymbol{V}^{-1}\|_\infty \lesssim 1$. As above it is enough to show $\|(\mathbb{1} - \boldsymbol{T}\boldsymbol{F})^{-1}\|_2 \lesssim 1$. By Lemma 3.6(i) and (3.39), $1 - \|\boldsymbol{T}\boldsymbol{F}\|_2 \sim 1$ which again leads to $\|(\mathbb{1} - \boldsymbol{T}\boldsymbol{F})^{-1}\|_2 \lesssim 1$. We conclude (3.21) for $\eta \le \eta_*$ and $\tau \in [1 + \tau_*, \tau^*]$.

Next, we verify (3.22) in these two regimes. Using $\boldsymbol{h} \cdot \chi(\|\boldsymbol{h}\|_\infty \leq \lambda_*) = O(\|\boldsymbol{d}\|_\infty)$ by (3.21), $\boldsymbol{v} \lesssim 1$ and $\boldsymbol{u} \lesssim 1$, we see that with the exception of $\boldsymbol{d}$, all terms in (3.50) are second order in $\boldsymbol{d}$. Therefore,

$$(3.51) \qquad \boldsymbol{r} \cdot \chi(\|\boldsymbol{h}\|_\infty \leq \lambda_*) = \boldsymbol{d} \cdot \chi(\|\boldsymbol{h}\|_\infty \leq \lambda_*) + O(\|\boldsymbol{d}\|_\infty^2)$$

uniformly for $\eta \geq \eta_*$ and $\tau \in [0, 1 - \tau_*] \cup [1 + \tau_*, \tau^*]$ as well as for $\eta \in (0, \eta_*]$ and $\tau \in [1 + \tau_*, \tau^*]$.

We start from $\boldsymbol{L}\boldsymbol{h} = \boldsymbol{r}$ and compute

$$(3.52) \qquad \langle \boldsymbol{y}, \boldsymbol{h} \rangle = \langle (\boldsymbol{L}^{-1})^* \boldsymbol{y}, \boldsymbol{r} \rangle = \langle \boldsymbol{R}\boldsymbol{y}, \boldsymbol{d} \rangle + \langle (\boldsymbol{L}^{-1})^* \boldsymbol{y}, \boldsymbol{r} - \boldsymbol{d} \rangle.$$

Here, we defined the operator $\boldsymbol{R} = \boldsymbol{R}(\eta)$ on $\mathbb{C}^{2n}$ in the last step through its action on any $\boldsymbol{x} \in \mathbb{C}^{2n}$ via

$$(3.53) \qquad \boldsymbol{R}\boldsymbol{x} := (\boldsymbol{L}^{-1})^* \boldsymbol{x} = \boldsymbol{V}^{-1}(\mathbb{1} - \boldsymbol{F}\boldsymbol{T})^{-1} \boldsymbol{V}\boldsymbol{x}.$$

Now, we establish that $\|(\boldsymbol{L}^{-1})^*\|_\infty \lesssim 1$ in the two regimes considered in Step 2. From this, we conclude that $\|\boldsymbol{R}\|_\infty \lesssim 1$ and that the last term in (3.52) when multiplied by $\chi(\|\boldsymbol{h}\|_\infty \leq \lambda_*)$ is bounded by $\lesssim \|\boldsymbol{y}\|_\infty \|\boldsymbol{d}\|_\infty^2$ because of (3.51). By Lemma 3.6(i), (3.35), (3.36) and (3.39), we have $1 - \|\boldsymbol{F}\boldsymbol{T}\|_2 \sim 1$. Thus, $\|(\mathbb{1} - \boldsymbol{F}\boldsymbol{T})^{-1}\|_2 \lesssim 1$ and hence $\|(\mathbb{1} - \boldsymbol{F}\boldsymbol{T})^{-1}\|_\infty \lesssim 1$ by Lemma B.1(ii). As $\|\boldsymbol{V}\|_\infty \|\boldsymbol{V}^{-1}\|_\infty \lesssim 1$ we get $\|(\boldsymbol{L}^{-1})^*\|_\infty \lesssim 1$. Therefore, we conclude that (3.22) holds true uniformly for $\eta \geq \eta_*$ and $\tau \in [0, 1 - \tau_*] \cup [1 + \tau_*, \tau^*]$ as well as for $\eta \in (0, \eta_*]$ and $\tau \in [1 + \tau_*, \tau^*]$. Thus, we have proved the proposition for these combinations of η and τ.

Finally, we prove the proposition in the most interesting regime, $\tau \in [0, 1 - \tau_*]$ and for small η:

Step 3: There exists $\eta_* > 0$, depending only on $\mathcal{P}$, and $\lambda_* \gtrsim 1$ such that (3.21) holds true for $\eta \in (0, \eta_*]$ and $\tau \in [0, 1 - \tau_*]$. Moreover, with this choice of λ_*, (3.22) holds true for $\eta \in (0, \eta_*]$ and $\tau \in [0, 1 - \tau_*]$.

The crucial step for proving (3.21) and (3.22) was the order one bound on $\|(\mathbb{1} - \boldsymbol{T}\boldsymbol{F})^{-1}\|_2$. However, in the regime $\tau \in [0, 1 - \tau_*]$ and small η such bound is not available since $(\mathbb{1} - \boldsymbol{T}\boldsymbol{F})\boldsymbol{f}_- = O(\eta)$ which can be deduced from (3.62) below. The simple bound

$$(3.54) \qquad \|(\mathbb{1} - \boldsymbol{T}\boldsymbol{F})^{-1}\|_2 \lesssim \eta^{-1}$$

which is a consequence of (3.35) and $\|\boldsymbol{T}\|_2 = 1$ is not strong enough. In order to control $\|(\mathbb{1} - \boldsymbol{T}\boldsymbol{F})^{-1}\boldsymbol{V}\boldsymbol{r}\|_2$ we will need to use a special property of the vector $\boldsymbol{V}\boldsymbol{r}$, namely that it is almost orthogonal to $\boldsymbol{f}_-$. This mechanism is formulated in the following *contraction-inversion lemma*, which is proved in Appendix B. It is closely related to the rotation-inversion lemmas—Lemma 5.8 in [4] and Lemma 3.6 in [7]—which control the invertibility of $\mathbb{1} - \boldsymbol{U}\boldsymbol{F}$, where $\boldsymbol{U}$ is a unitary operator and $\boldsymbol{F}$ is symmetric.

LEMMA 3.7 (Contraction–inversion lemma).   *Let $\varepsilon, \eta, c_1, c_2, c_3 > 0$ satisfying $\eta \le \varepsilon c_1 / (2c_2^2)$ and $A, B \in \mathbb{C}^{2n \times 2n}$ be two Hermitian matrices such that*

$$(3.55) \qquad \|A\|_2 \le 1, \qquad \|B\|_2 \le 1 - c_1 \eta.$$

Suppose that there are ℓ^2-normalized vectors $b_\pm \in \mathbb{C}^{2n}$ satisfying

$$(3.56) \quad Bb_+ = \|B\|_2 b_+, \qquad Bb_- = -\|B\|_2 b_-, \qquad \|Bx\|_2 \le (1 - \varepsilon)\|x\|_2$$

for all $x \in \mathbb{C}^{2n}$ such that $x \perp \operatorname{span}\{b_+, b_-\}$.

Furthermore, assume that

$$(3.57) \qquad \langle b_+, Ab_+ \rangle \le 1 - \varepsilon, \qquad \big\|(\mathbb{1} + A)b_-\big\|_2 \le c_2 \eta.$$

Then there is a constant $C > 0$, depending only on c_1, c_2, c_3 and ε, such that for each $p \in \mathbb{C}^{2n}$ satisfying

$$(3.58) \qquad \big|\langle b_-, p \rangle\big| \le c_3 \eta \|p\|_2,$$

it holds true that

$$(3.59) \qquad \big\|(\mathbb{1} - AB)^{-1} p\big\|_2 \le C\|p\|_2.$$

We will apply this lemma with the choices $A = T$, $B = F$, $b_\pm = f_\pm$ and $p = Vr$. The resulting bound on $\|(\mathbb{1} - TF)^{-1} Vr\|_2$ will be lifted to a bound on $\|(\mathbb{1} - TF)^{-1} Vr\|_\infty$ by (B.9). All estimates in the remainder of this proof will hold true uniformly for $\tau \in [0, 1 - \tau_*]$. However, we will not stress this fact for each estimate. Moreover, the estimates will be uniform for $\eta \in (0, \eta_*]$. The threshold $\eta_* \le 1$ will be chosen later such that it depends on $\mathcal{P}$ only and the assumptions of Lemma 3.7 are fulfilled. We now start checking the assumptions of Lemma 3.7.

By Proposition 3.2, there is $\Phi_1 \sim 1$ such that

$$(3.60) \qquad \Phi_1^{-1} \le v \le \Phi_1$$

for all $\eta \in (0, 1]$. We recall from (3.35) that there is a constant $c_1 \sim 1$ such that $\|F\|_2 \le 1 - c_1 \eta$ for all $\eta \in (0, 1]$. Recalling the definition of a from (3.43), we conclude from (3.46) the existence of $\Phi_2 \sim 1$ such that

$$(3.61) \qquad \|f_- - a\|_2 \le \|f_- - a\|_\infty \le \Phi_2 \eta$$

for all $\eta \in (0, 1]$. Here, we used that $\|y\|_2 \le \|y\|_\infty$ for all $y \in \mathbb{C}^{2n}$ due to the normalization of the ℓ^2 norm.

Since the first and the second n-component of the vector Vv are the same we have $Ta = -a$ by (3.43) and Lemma 3.6(iii). Hence,

$$(3.62) \qquad \|f_- + Tf_-\|_2 \le \|f_- - a\|_2 + \|T\|_2 \|f_- - a\|_2 \le 2\Phi_2 \eta$$

by $\|T\|_2 = 1$ and (3.61).

Due to (3.38), there exists $\varepsilon \sim 1$ such that

$$\|Fx\|_2 \le (1 - \varepsilon)\|x\|_2$$

for all $x \in \mathbb{C}^{2n}$ such that $x \perp f_+$ and $x \perp f_-$ and for all $\eta \in (0, 1]$. As T is Hermitian, we can also assume by (3.48) that

$$\langle f_+, T f_+ \rangle \leq 1 - \varepsilon$$

for all $\eta \in (0, 1]$ by possibly reducing ε but keeping $\varepsilon \gtrsim 1$.

So far we checked the conditions (3.55)–(3.57), it remains to verify (3.58) with the choice $p = Vr$. Assuming that $\langle a, Vr \rangle = 0$, we deduce from (3.61) that

$$(3.63) \qquad |\langle f_-, Vr \rangle| \leq |\langle a, Vr \rangle| + \|f_- - a\|_2 \|Vr\|_2 \leq \Phi_2 \eta \|Vr\|_2.$$

This is the estimate required in (3.58). Hence, it suffices to show that Vr is perpendicular to a, that is,

$$(3.64) \qquad \langle e_-(Vv), Vr \rangle = \langle e_-(V^2 v), Lh \rangle = \left\langle L^*\left(e_- \frac{\widetilde{v}}{u}\right), h \right\rangle = 0,$$

where we used the symmetry of V, that V is diagonal and (3.23) in the first equality, and the notation $\widetilde{v} = (v_2, v_1)$.

Using (3.31) and the n-component relations of the second identity in (3.32), we compute

$$(3.65) \qquad L^*\left(e_- \frac{\widetilde{v}}{u}\right) = \begin{pmatrix} \eta + Sv_2 - S\left(v_2\left(\dfrac{v_1 v_2}{u} + \tau u\right)\right) \\ -\eta - S^t v_1 + S^t\left(v_1\left(\dfrac{v_1 v_2}{u} + \tau u\right)\right) \end{pmatrix} = \eta e_-.$$

Since $\langle e_- g \rangle = \langle e_- v \rangle = 0$ by (3.20) and (3.8), respectively, this proves (3.64) and, therefore, (3.63) as well. Thus, we checked all conditions of Lemma 3.7.

By possibly reducing η_* but keeping $\eta_* \gtrsim 1$, we can assume that $\eta_* \leq \varepsilon c_1/(8\Phi_2^2)$. Now, we can apply Lemma 3.7 with ε, c_1, $c_2 = 2\Phi_2$, $c_3 = \Phi_2$ for any $\eta \in (0, \eta_*]$. Thus, applying (3.59) in Lemma 3.7 to (3.29), we obtain $\|Vh\|_2 \lesssim \|Vr\|_2$, and hence $\|Vh\|_\infty \lesssim \|Vr\|_\infty$ because of (B.9). Therefore, for any $\lambda_* > 0$, depending only on $\mathcal{P}$, we have

$$\|h\|_\infty \chi(\|h\|_\infty \leq \lambda_*) \lesssim \|V^{-1}\|_\infty \|Vr\|_\infty \chi(\|h\|_\infty \leq \lambda_*) \lesssim \|h\|_\infty^2 + \|d\|_\infty$$

uniformly for $\eta \in (0, \eta_*]$ and $\tau \in [0, 1 - \tau_*]$. Here, we used (3.30) and (3.49) in the second step. Choosing $\lambda_* > 0$ small enough as before, we conclude (3.21) for $\eta \in (0, \eta_*]$ and $\tau \in [0, 1 - \tau_*]$. Since $\eta_* > 0$ depends only on $\mathcal{P}$, and η_* was arbitrary in the proof of Step 2 we proved (3.21) for all $\eta > 0$ and $\tau \in [0, 1 - \tau_*] \cup [1 + \tau_*, \tau^*]$.

In order to prove (3.22), we remark that because of (3.21) and (3.50) the estimate (3.51) holds true for $\eta \in (0, \eta_*]$ and $\tau \in [0, 1 - \tau_*]$ as well. Due to the instability (3.54) of $(\mathbb{1} - TF)^{-1}$ and, correspondingly, of its adjoint, the definition of R in (3.53) will not yield an operator satisfying $\|R\|_\infty \lesssim 1$ in this regime. Therefore, we again employ that the inverse of $\mathbb{1} - TF$ is bounded on the subspace orthogonal to f_- and the blow-up in the direction of f_- is compensated by the smallness of $\langle f_-, Vr \rangle$ following from $\langle a, Vr \rangle = 0$ and $\|f_- - a\|_\infty = O(\eta)$ by (3.46).

Let Q be the orthogonal projection onto the subspace $f_-^\perp$, that is, $Qx := x - \langle f_-, x \rangle f_-$ for all $x \in \mathbb{C}^{2n}$. Recalling the definition of a in (3.43), we now define the operator $R = R(\eta)$ on $\mathbb{C}^{2n}$ as follows:

$$(3.66) \quad Rx := V\big((\mathbb{1} - TF)^{-1}Q\big)^* V^{-1}x - \langle V^{-1}(\mathbb{1} - TF)^{-1}f_-, x \rangle V(f_- - a)$$

for every $x \in \mathbb{C}^{2n}$. Note that this R is different from the one given in (3.53) that is used in the other parameter regimes. Now, we estimate $\|Rx\|_\infty$. For the first term, we use the bound (B.11) whose assumptions we check first. The first condition, $\|(\mathbb{1} - TF)^{-1}Q\|_2 \lesssim 1$, in (B.10) follows from (3.59) as (3.58) with $p = Qx$ is trivially satisfied, and hence $\|(\mathbb{1} - TF)^{-1}Qx\|_2 \lesssim \|Qx\|_2 \lesssim \|x\|_2$. The second condition in (B.10) is met by (3.35) and the third condition is exactly (3.62). Using $\|f_-\|_\infty \lesssim 1$ from (3.37), (B.11) and (3.30), we conclude that the first term in (3.66) is $\lesssim \|x\|_\infty$. In the second term, we use the trivial bound

$$(3.67) \qquad\qquad \|(\mathbb{1} - TF)^{-1}\|_\infty \lesssim \eta^{-1}$$

which is a consequence of the corresponding bound on $\|(1 - TF)^{-1}\|_2$ in (3.54) and (B.9). The potential blow-up in (3.67) for small η is compensated by the estimate $\|f_- - a\|_\infty = O(\eta)$ from (3.46). Altogether this yields $\|R(\eta)\|_\infty \lesssim 1$ for all $\eta \in (0, \eta_*]$.

From the definition of R and $h = V^{-1}(\mathbb{1} - TF)^{-1}Vr$, we obtain

$$\langle y, h \rangle = \langle V^{-1}y, (\mathbb{1} - TF)^{-1}QV(r - d) \rangle$$
$$(3.68) \qquad\qquad + \langle y, V^{-1}(\mathbb{1} - TF)^{-1}f_- \rangle \langle f_- - a, V(r - d) \rangle + \langle Ry, d \rangle.$$

Notice that we first inserted $\mathbb{1} = Q + |f_-\rangle\langle f_-|$ before Vr, then we inserted the vector a in the second term for free by using $\langle a, Vr \rangle = 0$ from (3.64). This brought in the factor $f_- - a \sim O(\eta)$ that compensates the $(\mathbb{1} - TF)^{-1}$ on the unstable subspace parallel to f_-. Finally, we subtracted the term d to r freely and we defined the operator R exactly to compensate for it. The reason for this counter term d is the formula (3.51) showing that $r - d$ is one order better in d than r. Thus, the first two terms in the right-hand side of (3.68) are bounded by $\|d\|_\infty^2 \|y\|_\infty$. The compensating term, $\langle Ry, d \rangle$ remains first order in d but only in weak sense, tested against the vector Ry, and not in norm sense. This is the essential improvement of (3.22) over (3.21). Recalling now $h = g - iv$, the identity (3.68) together with the bounds we just explained concludes the proof of Proposition 3.3. $\quad\square$

4. Proof of Proposition 2.4. As in the previous section, we assume without loss of generality that $\rho(S) = 1$. See the remark about (3.7).

For $\tau_* > 0$ and $\tau^* > \tau_* + 1$, we define

$$(4.1) \qquad \mathbb{D}_< := \{z \in \mathbb{C} | |z|^2 \leq 1 - \tau_*\}, \qquad \mathbb{D}_> := \{z \in \mathbb{C} | 1 + \tau_* \leq |z|^2 \leq \tau^*\}.$$

Via $\tau = |z|^2$ these sets correspond to the regimes $[0, 1 - \tau_*]$ and $[1 + \tau_*, \tau^*]$ in the previous section.

PROOF OF PROPOSITION 2.4. Since the defining equations in (2.4) are smooth functions of η, τ and $(v_i)_{i=1,\dots,2n}$ and the operator L is invertible for $\eta > 0$ the implicit function theorem implies that the function $v \colon \mathbb{R}_+ \times \mathbb{R}_0^+ \to \mathbb{R}_+^{2n}$ is smooth. Therefore, the function $\mathbb{R}_+ \times \mathbb{C} \to \mathbb{R}_+^{2n}$, $(\eta, z) \mapsto v^\tau(\eta)|_{\tau = |z|^2}$ is also smooth.

For $\alpha = (\alpha_1, \alpha_2) \in \mathbb{N}^2$, we define

$$\partial^\alpha v := \partial_\eta^{\alpha_1} \partial_\tau^{\alpha_2} v.$$

For fixed $\tau_* > 0$ and $\tau^* > \tau_* + 1$, we first prove that for all $\alpha \in \mathbb{N}^2$, we have

(4.2) $$\|\partial^\alpha v\|_\infty \lesssim 1$$

uniformly for all $\eta > 0$ and $\tau \in [0, 1 - \tau_*] \cup [1 + \tau_*, \tau^*]$.

Differentiating (2.4) with respect to η and τ, respectively, yields

(4.3) $$L(\partial_\eta v) = -v^2 + \tau u^2, \qquad L(\partial_\tau v) = -uv.$$

By further differentiating with respect to η and τ, we iteratively obtain that for any multi-index $\alpha \in \mathbb{N}^2$

(4.4) $$L\partial^\alpha v = r_\alpha,$$

where r_α only depends on η, τ and $\partial^\beta v$ for $\beta \in \mathbb{N}^2$, $|\beta| = \beta_1 + \beta_2 < |\alpha|$. In fact, for all $\alpha \in \mathbb{N}^2$, we have

(4.5a) $$L\big(\partial^{\alpha + e_1} v\big) = \partial^\alpha\big(-v^2 + \tau u^2\big) - \sum_{\nu \leq \alpha, \nu \neq (0,0)} \binom{\alpha}{\nu} \big(\partial^\nu L\big)\big(\partial^{\alpha - \nu + e_1} v\big),$$

(4.5b) $$L\big(\partial^{\alpha + e_2} v\big) = \partial^\alpha\big(-vu\big) - \sum_{\nu \leq \alpha, \nu \neq (0,0)} \binom{\alpha}{\nu} \big(\partial^\nu L\big)\big(\partial^{\alpha - \nu + e_2} v\big).$$

As an example, we compute $L(\partial_\tau^2 v)$. Thus, the second relation in (4.3) yields

(4.6) $$L\partial_\tau^2 v = \frac{2}{v}(\partial_\tau v)^2 + 2u^2 S_d \partial_\tau v - \frac{2\tau u^3}{v}(S_d \partial_\tau v)^2.$$

By induction on $|\alpha| = \alpha_1 + \alpha_2$, we prove $\|r_\alpha\|_\infty \lesssim 1$ and $\|\partial^\alpha v\|_\infty \lesssim 1$ simultaneously. From (4.5), we conclude that $r_{\alpha + e_1}$ and $r_{\alpha + e_2}$ are bounded in ℓ^∞-norm if $\|\partial^\nu v\|_\infty \lesssim 1$ for all $\nu \leq \alpha$ as the first term on the right-hand side of (4.5a) and (4.5b), respectively, and $\partial^\nu L$ for all $\nu \leq \alpha$ are bounded. In order to conclude that $\partial^{\alpha + e_1} v$ and $\partial^{\alpha + e_2} v$ are bounded, it suffices to prove that $\|\partial^\alpha v\|_\infty \lesssim \|r_\alpha\|_\infty$ by controlling L^{-1} in (4.4).

As in the proof of Proposition 3.3 the norm of L^{-1} is bounded, $\|L^{-1}\|_\infty \lesssim 1$, for $\tau \in [1 + \tau_*, \tau^*]$ or $\tau \in [0, 1 - \tau_*]$ and large η as well as $\tau \in [0, 1 - \tau_*]$ and small η separately. We thus focus on the most interesting regime where $\tau \in [0, 1 - \tau_*]$ and small η. As for the proof of Proposition 3.3, we apply Lemma 3.7 in this regime. We only check the condition (3.58) here since the others are established

in the same way as in the proof of Proposition 3.3. Recall the definition of $\boldsymbol{a}$ in (3.43). Using $\langle \boldsymbol{e}_- \partial^\alpha \boldsymbol{v} \rangle = 0$ from (3.8) for all $\alpha \in \mathbb{N}^2$, we obtain

$$\langle \boldsymbol{a}, \boldsymbol{V} \boldsymbol{r}_\alpha \rangle = \langle \boldsymbol{L}^*(\boldsymbol{e}_- \boldsymbol{V}^2 \boldsymbol{v}), \partial^\alpha \boldsymbol{v} \rangle = \langle \eta \boldsymbol{e}_-, \partial^\alpha \boldsymbol{v} \rangle = 0$$

for all $\alpha \in \mathbb{N}^2$. Here, we used $\boldsymbol{L}^*(\boldsymbol{e}_- \boldsymbol{V}^2 \boldsymbol{v}) = \eta \boldsymbol{e}_-$ which is shown in (3.65) in the proof of Proposition 3.3. This concludes the proof of (4.2).

Next, we show the integrability of $\Delta_z \langle v_1^\tau |_{\tau=|z|^2} \rangle$ as a function of η for $z \in \mathbb{D}_<$ for fixed $\tau_* > 0$. Note that $\langle v_1^\tau \rangle = \langle \boldsymbol{v}^\tau \rangle$ by (3.8). Using

$$\Delta_z (\boldsymbol{v}^\tau |_{\tau=|z|^2}) = 4(\tau \partial_\tau^2 \boldsymbol{v}^\tau + \partial_\tau \boldsymbol{v}^\tau)|_{\tau=|z|^2}$$

together with (4.3) and (4.6), we obtain

$$(4.7) \quad \boldsymbol{L}\Delta_z(\boldsymbol{v}^\tau |_{\tau=|z|^2}) = 4\left(\frac{2\tau}{\boldsymbol{v}}(\partial_\tau \boldsymbol{v})^2 + 2\tau \boldsymbol{u}^2 \boldsymbol{S}_d \partial_\tau \boldsymbol{v} - \frac{2\tau^2 \boldsymbol{u}^3}{\boldsymbol{v}}(\boldsymbol{S}_d \partial_\tau \boldsymbol{v})^2 - \boldsymbol{u}\boldsymbol{v} \right).$$

From (3.9), (3.10) and (3.26), we conclude that $\boldsymbol{u}\boldsymbol{v} \sim (1 + \eta^3)^{-1}$, and hence $|\partial_\tau \boldsymbol{v}| \lesssim (1 + \eta^3)^{-1}$ uniformly for $z \in \mathbb{D}_<$ since $\|\partial^\alpha \boldsymbol{v}\|_\infty \lesssim \|\boldsymbol{r}_\alpha\|_\infty$. Therefore, the right-hand side of (4.7) is of order $(1 + \eta^3)^{-1}$ for $z \in \mathbb{D}_<$, and hence using the control on $\boldsymbol{L}^{-1}$ as before, we conclude that $|\Delta_z(\boldsymbol{v}^\tau |_{\tau=|z|^2})| \lesssim (1 + \eta^3)^{-1}$ uniformly for $\eta > 0$. Thus, $\Delta_z \langle v_1^\tau |_{\tau=|z|^2} \rangle = \Delta_z \langle \boldsymbol{v}^\tau |_{\tau=|z|^2} \rangle$ as a function of η is integrable on $\mathbb{R}_+$ and the integral is a continuous function of $z \in \mathbb{D}_<$. As $\tau_* > 0$ was arbitrary, this concludes the proof of part (i) of Proposition 2.4 and shows that σ is a rotationally invariant function on $\mathbb{C}$, which is continuous on $D(0, 1)$.

Now, we establish that for $\tau < 1$, the derivative of the average of $\boldsymbol{u}$ with respect to τ gives an alternative representation of the density of states as follows:

$$(4.8) \quad \sigma(z) = \frac{1}{\pi} \partial_\tau \left(\tau \langle \boldsymbol{u}_0 \rangle \right) \Big|_{\tau=|z|^2} = -\frac{2}{\pi} \langle \boldsymbol{S}_o \boldsymbol{v}_0, \partial_\tau \boldsymbol{v}_0 \rangle \Big|_{\tau=|z|^2},$$

where $\boldsymbol{u}_0 := \lim_{\eta \downarrow 0} \boldsymbol{u}(\eta)$ and $\boldsymbol{v}_0 := \lim_{\eta \downarrow 0} \boldsymbol{v}(\eta)$. The first relation in (4.8) will be proved below and the second one follows immediately using $\tau \boldsymbol{u}_0 = 1 - \boldsymbol{v}_0 \boldsymbol{S}_o \boldsymbol{v}_0$ by (3.6) and (3.25) for $\eta \downarrow 0$, as well as $\boldsymbol{S}_o^t = \boldsymbol{S}_o$.

We first give a heuristic derivation of the first equality in (4.8) (see, e.g., Section 4.6 of [11]). Writing the resolvent $\boldsymbol{G}^z$ of $\boldsymbol{H}^z$ as

$$\boldsymbol{G}^z = \begin{pmatrix} G_{11} & G_{12} \\ G_{21} & G_{22} \end{pmatrix},$$

we obtain

$$\operatorname{Tr} G_{12} = -\partial_{\bar{z}} \operatorname{Tr} \log((X - z)(X^* - \bar{z}) + \eta^2) = -\frac{2}{n} \partial_{\bar{z}} \log|\det(\boldsymbol{H}^z - i\eta)|$$

for the normalized trace of G_{12} [see (1.3)]. Since $\Delta_z = 4\partial_z \partial_{\bar{z}}$, taking the ∂_z-derivative of the previous identity, we obtain

$$(4.9) \quad \frac{1}{2n} \Delta_z \log|\det(\boldsymbol{H}^z - i\eta)| = -\partial_z \operatorname{Tr} G_{12}.$$

Using (2.5), (2.14) and $\operatorname{Im} m^z \approx \langle v_1^\tau |_{\tau=|z|^2} \rangle$, the left-hand side of (4.9) is approximately $\pi \sigma(z)$ after taking the $\eta \downarrow 0$ limit. On the other hand, $\boldsymbol{G}^z$ converges to $\boldsymbol{M}^z$ for $n \to \infty$. Thus, by (3.5) the right-hand side of (4.9) can be approximated by $\partial_z(z \langle u^\tau |_{\tau=|z|^2}(\eta) \rangle)$. Therefore, taking $\eta \downarrow 0$, we conclude

$$\pi \sigma(z) \approx \partial_z z \langle u_0^\tau |_{\tau=|z|^2} \rangle = (\partial_\tau \tau \langle u_0^\tau \rangle)|_{\tau=|z|^2}.$$

In fact, this approximation holds not only in the $n \to \infty$ limit but it is an identity for any fixed n. This completes the heuristic argument for (4.8).

We now turn to the rigorous proof of the first relation in (4.8). In fact, for $\tau < 1$, we prove the following integrated version:

$$\text{(4.10)} \qquad \int_{|z'|^2 \le \tau} \sigma(z') \, d^2 z' = \tau \langle \boldsymbol{u}_0^\tau \rangle.$$

Since σ is a continuous function on $D(0, 1)$ differentiating (4.10) with respect to τ immediately yields (4.8).

In order to justify the existence of the limits of $\boldsymbol{v}$ and $\boldsymbol{u}$ for $\eta \downarrow 0$ and the computations in the proof of (4.10), we remark that by (4.2), $(\eta, z) \mapsto \boldsymbol{v}^\tau(\eta)|_{\tau=|z|^2}$ can be uniquely extended to a positive C^∞ function on $[0, \infty) \times D(0, 1)$. In the following, $\boldsymbol{v}$ and $\boldsymbol{v}_0^\tau := \boldsymbol{v}^\tau |_{\eta=0}$ denote this function and its restriction to $\{0\} \times [0, 1)$, respectively. In particular, the restriction $\boldsymbol{v}_0^\tau |_{\tau=|z|^2}$ is a smooth function on $D(0, 1)$ which satisfies

$$\text{(4.11)} \qquad \frac{1}{\boldsymbol{v}_0^\tau} = S_o \boldsymbol{v}_0^\tau + \frac{\tau}{S_d \boldsymbol{v}_0^\tau}$$

with $\tau = |z|^2$. Moreover, derivatives of $\boldsymbol{v}$ in η and τ and limits in η and τ for $\tau < 1$ can be freely interchanged.

For the proof of (4.10), we use integration by parts to obtain

$$\text{(4.12)} \qquad \int_{|z'|^2 \le \tau} \sigma(z') \, d^2 z' = -2\tau \int_0^\infty \partial_\tau \langle \boldsymbol{v} \rangle \, d\eta = -\tau \int_0^\infty \partial_\tau (\langle \boldsymbol{v} \rangle + \langle \widetilde{\boldsymbol{v}} \rangle) \, d\eta.$$

We recall $\widetilde{\boldsymbol{v}} = (v_2, v_1)$ and get

$$\boldsymbol{v} = \frac{\eta + S_d \boldsymbol{v}}{(\eta + S_d \boldsymbol{v})(\eta + S_o \boldsymbol{v}) + \tau}, \qquad \widetilde{\boldsymbol{v}} = \frac{\eta + S_o \boldsymbol{v}}{(\eta + S_d \boldsymbol{v})(\eta + S_o \boldsymbol{v}) + \tau}$$

from (3.6). This implies the identity

$$\partial_\eta \log((\eta + S_d \boldsymbol{v})(\eta + S_o \boldsymbol{v}) + \tau) = \boldsymbol{v} + \widetilde{\boldsymbol{v}} + \widetilde{\boldsymbol{v}} S_d \partial_\eta \boldsymbol{v} + \boldsymbol{v} S_o \partial_\eta \boldsymbol{v}.$$

Using

$$\langle \widetilde{\boldsymbol{v}} S_d \partial_\eta \boldsymbol{v} \rangle + \langle \boldsymbol{v} S_o \partial_\eta \boldsymbol{v} \rangle = \langle \boldsymbol{v} S_o \partial_\eta \boldsymbol{v} \rangle + \langle \boldsymbol{v} S_o \partial_\eta \boldsymbol{v} \rangle = \partial_\eta \langle \boldsymbol{v} S_o \boldsymbol{v} \rangle$$

and recalling $\boldsymbol{v}_0 := \lim_{\eta \downarrow 0} \boldsymbol{v}(\eta)$, we find for (4.12) the expression

$$\text{(4.13)} \qquad \int_0^\infty \partial_\tau (\langle \boldsymbol{v} \rangle + \langle \widetilde{\boldsymbol{v}} \rangle) \, d\eta = -\langle \partial_\tau \log((S_d \boldsymbol{v}_0)(S_o \boldsymbol{v}_0) + \tau) \rangle + \partial_\tau \langle \boldsymbol{v}_0 S_o \boldsymbol{v}_0 \rangle.$$

Hence, using $\langle \widetilde{\boldsymbol{v}}_0 \boldsymbol{S}_d \partial_\tau \boldsymbol{v}_0 \rangle = \langle \boldsymbol{v} \boldsymbol{S}_o \partial_\tau \boldsymbol{v}_0 \rangle$, we obtain

$$\langle \partial_\tau \log((\boldsymbol{S}_d \boldsymbol{v}_0)(\boldsymbol{S}_o \boldsymbol{v}_0) + \tau) \rangle = \langle \boldsymbol{u} \rangle + \partial_\tau \langle \boldsymbol{v}_0 \boldsymbol{S}_o \boldsymbol{v}_0 \rangle$$

and conclude (4.10) from (4.13). The formula (4.10) was also obtained in [14] with a different method.

We prove (iii) before (ii). As $\boldsymbol{v}_0$ is infinitely often differentiable in τ and $\tau = |z|^2$, we conclude from (4.8) that σ is infinitely often differentiable in z. The following lemma shows (2.6) which completes the proof of part (iii).

LEMMA 4.1 (Positivity and boundedness of σ).   *Uniformly for $z \in D(0, 1)$, we have*

(4.14) $$\sigma(z) \sim 1,$$

where $\sim$ only depends on s_* and s^*.

PROOF.   We will compute the derivative in (4.8) and prove the estimate (4.14) first for $z \in \mathbb{D}_<$ and arbitrary $\tau_* > 0$ depending only on s_* and s^*. Then we show that there is $\tau_* > 0$ depending only on s_* and s^* such that (4.14) holds true for $z \in D(0, 1) \setminus \mathbb{D}_<$.

In this proof, we write $\mathscr{D}(y) := \text{diag}(y)$ for $y \in \mathbb{C}^l$ for brevity. Furthermore, we introduce the $2n \times 2n$ matrix

$$\boldsymbol{E} := \begin{pmatrix} \mathbb{1} & \mathbb{1} \\ \mathbb{1} & \mathbb{1} \end{pmatrix}.$$

In the following, $\boldsymbol{v}$ and all related quantities will be evaluated at $\tau = |z|^2$. We start the proof from (4.8), recall $\boldsymbol{L} = \boldsymbol{V}^{-1}(\mathbb{1} - \boldsymbol{T}\boldsymbol{F})\boldsymbol{V}$ and use the second relation in (4.3) as well as (3.31) to obtain

(4.15)
$$\begin{aligned}
\sigma(z) &= -\frac{2}{\pi} \langle \boldsymbol{S}_o \boldsymbol{v}_0, \partial_\tau \boldsymbol{v}_0 \rangle \\
&= \lim_{\eta \downarrow 0} \frac{2}{\pi} \left\langle \boldsymbol{V}^{-1} \frac{\widetilde{\boldsymbol{v}}}{\boldsymbol{u}}, (\mathbb{1} - \boldsymbol{T}\boldsymbol{F})^{-1} \boldsymbol{V}(\boldsymbol{v}\boldsymbol{u}) \right\rangle \\
&= \lim_{\eta \downarrow 0} \frac{2}{\pi} \left\langle \sqrt{\boldsymbol{v}\widetilde{\boldsymbol{v}}}, \frac{1}{\sqrt{\boldsymbol{u}}} (\mathbb{1} - \boldsymbol{T}\boldsymbol{F})^{-1} \sqrt{\boldsymbol{u}} \sqrt{\boldsymbol{v}\widetilde{\boldsymbol{v}}} \right\rangle \\
&= \lim_{\eta \downarrow 0} \frac{2}{\pi} \langle \sqrt{\boldsymbol{v}\widetilde{\boldsymbol{v}}}, (\mathbb{1} - \mathscr{D}(\boldsymbol{u}^{-1/2}) \boldsymbol{T}\boldsymbol{F} \mathscr{D}(\boldsymbol{u}^{1/2}))^{-1} \sqrt{\boldsymbol{v}\widetilde{\boldsymbol{v}}} \rangle.
\end{aligned}$$

Note that the inverses of $\mathbb{1} - \boldsymbol{T}\boldsymbol{F}$ and $\mathbb{1} - \tau \mathscr{D}(\boldsymbol{u}^{-1/2}) \boldsymbol{T}\boldsymbol{F} \mathscr{D}(\boldsymbol{u}^{1/2})$ exist by Lemma 3.6 and Lemma 3.4 as $\eta > 0$ and $\tau < 1$.

Due to (3.27a) and (3.32), we have $T = -\mathbb{1} + \tau u E$, which implies

(4.16)
$$
\begin{aligned}
\mathbb{1} &- \mathscr{D}(u^{-1/2}) T F \mathscr{D}(u^{1/2}) \\
&= \mathbb{1} + \mathscr{D}(u^{-1/2}) F \mathscr{D}(u^{1/2}) - \tau \mathscr{D}(u^{1/2}) E F \mathscr{D}(u^{1/2}) \\
&= \left(\mathbb{1} - \tau \mathscr{D}(u^{1/2}) E F (\mathbb{1} + F)^{-1} \mathscr{D}(u^{1/2})\right) \\
&\quad \times \left(\mathbb{1} + \mathscr{D}(u^{-1/2}) F \mathscr{D}(u^{1/2})\right).
\end{aligned}
$$

From (3.33) and (3.44), we deduce $\sqrt{u} F \sqrt{v\tilde{v}/u} = \sqrt{v\tilde{v}} + O(\eta)$. Hence, due to (4.16), (4.15) yields

(4.17) $\qquad \sigma(z) = \lim_{\eta \downarrow 0} \dfrac{1}{\pi} \langle \sqrt{v\tilde{v}}, (\mathbb{1} - \tau \mathscr{D}(u^{1/2}) E F (\mathbb{1} + F)^{-1} \mathscr{D}(u^{1/2}))^{-1} \sqrt{v\tilde{v}} \rangle.$

Defining the matrix $F \in \mathbb{C}^{n \times n}$ through $F y = \sqrt{v_1 u / v_2} S \sqrt{v_2 u / v_1} y$ for $y \in \mathbb{C}^n$, we obtain

(4.18)
$$
F = \begin{pmatrix} 0 & F \\ F^t & 0 \end{pmatrix},
$$
$$
(\mathbb{1} + F)^{-1} = \begin{pmatrix} (\mathbb{1} - F F^t)^{-1} & -(\mathbb{1} - F F^t)^{-1} F \\ -F^t (\mathbb{1} - F F^t)^{-1} & (\mathbb{1} - F^t F)^{-1} \end{pmatrix}.
$$

Furthermore, we introduce the $n \times n$ matrix A by

$$
A := 2 \cdot \mathbb{1} + (F^t - \mathbb{1})(\mathbb{1} - F F^t)^{-1} + (F - \mathbb{1})(\mathbb{1} - F^t F)^{-1}.
$$

From the computation,

$$
E F (\mathbb{1} + F)^{-1} = \begin{pmatrix} \mathbb{1} + (F^t - \mathbb{1})(\mathbb{1} - F F^t)^{-1} & \mathbb{1} + (F - \mathbb{1})(\mathbb{1} - F^t F)^{-1} \\ \mathbb{1} + (F^t - \mathbb{1})(\mathbb{1} - F F^t)^{-1} & \mathbb{1} + (F - \mathbb{1})(\mathbb{1} - F^t F)^{-1} \end{pmatrix},
$$

we conclude that

(4.19)
$$
\begin{aligned}
&(\mathbb{1} - \tau \mathscr{D}(u^{1/2}) E F (\mathbb{1} + F)^{-1} \mathscr{D}(u^{1/2}))^{-1} \begin{pmatrix} x \\ x \end{pmatrix} \\
&\qquad = \begin{pmatrix} (\mathbb{1} - \tau \mathscr{D}(u^{1/2}) A \mathscr{D}(u^{1/2}))^{-1} x \\ (\mathbb{1} - \tau \mathscr{D}(u^{1/2}) A \mathscr{D}(u^{1/2}))^{-1} x \end{pmatrix}
\end{aligned}
$$

for all $x \in \mathbb{C}^n$. Before applying this relation to (4.17), we show that $\mathbb{1} - \tau \mathscr{D}(u^{1/2}) A \mathscr{D}(u^{1/2})$ is invertible for $\tau < 1$. The relations in (4.18) yield

(4.20) $\qquad \langle x, A x \rangle = 2 \|x\|_2^2 - 2 \left\langle \begin{pmatrix} x \\ x \end{pmatrix}, (\mathbb{1} + F)^{-1} \begin{pmatrix} x \\ x \end{pmatrix} \right\rangle$

for all $x \in \mathbb{C}^n$ and $\eta > 0$. In particular, since $\|F\|_2 \leq 1$ by (3.33) we conclude $A \leq \mathbb{1}$. Hence, $\tau u = 1 - v_1 v_2 / u < 1$ for $\tau < 1$ by (3.32) implies that

$\mathbb{1} - \tau \mathscr{D}(u^{1/2}) A \mathscr{D}(u^{1/2})$ is invertible for $\tau < 1$. Thus, we apply (4.19) to (4.17) and obtain for $z \in D(0, 1)$

$$(4.21) \qquad \sigma(z) = \frac{2}{\pi} \lim_{\eta \downarrow 0} \langle \sqrt{v_1 v_2}, (\mathbb{1} - \tau \mathscr{D}(u^{1/2}) A \mathscr{D}(u^{1/2}))^{-1} \sqrt{v_1 v_2} \rangle.$$

Let $\tau_* > 0$ depend only on s_* and s^*. From (3.10) and (4.2), we conclude that $|\sigma| \lesssim 1$ uniformly for $z \in \mathbb{D}_<$ because of (4.8). This proves the upper bound in (4.14) for $z \in \mathbb{D}_<$.

For the proof of the lower bound, we infer some further properties of A and $\mathbb{1} - \tau \mathscr{D}(u^{1/2}) A \mathscr{D}(u^{1/2})$, respectively, from information about $\boldsymbol{F}$ via (4.20). In the following, we use versions of Proposition 3.2, (3.26) and Lemma 3.4 extended to the limiting case $\eta = 0+$. Recalling $\boldsymbol{v}_0 = \lim_{\eta \downarrow 0} \boldsymbol{v}$, these results are a simple consequence of the uniform convergence $\partial^\alpha \boldsymbol{v} \to \partial^\alpha \boldsymbol{v}_0$ for $\eta \downarrow 0$ and all $\alpha \in \mathbb{N}^2$ by (4.2).

Since $\boldsymbol{f}_- = (\sqrt{v_1 v_2 / u}, -\sqrt{v_1 v_2 / u}) + O(\eta)$ by (3.45) there are $\eta_*, \varepsilon \sim 1$ by Lemma 3.4 such that $\mathrm{Spec}(\boldsymbol{F}|_W) \subset [-1 + \varepsilon, 1]$ on the subspace $W := \{(x, x) | x \in \mathbb{C}^n\} \subset \mathbb{C}^{2n}$ as $\boldsymbol{f}_- \perp W$ uniformly for all $\eta \in [0, \eta_*]$. Therefore, for $\|x\|_2 = 1$, the right-hand side of (4.20) is contained in $[2(\varepsilon - 1)/\varepsilon, 1]$. Since $(F^t(\mathbb{1} - FF^t)^{-1})^t = F(\mathbb{1} - F^t F)^{-1}$, the matrix A is real symmetric, and hence the spectrum of A is contained in $[2(\varepsilon - 1)/\varepsilon, 1]$ for all $\eta \in [0, \eta_*]$ as well.

The real symmetric matrix A has a positive and a negative part, that is, there are positive matrices A_+ and A_- such that $A = A_+ - A_-$. Hence, we have

$$(4.22) \qquad \begin{aligned} &\mathbb{1} - \tau \mathscr{D}(u^{1/2}) A \mathscr{D}(u^{1/2}) \\ &= \mathbb{1} - \tau \mathscr{D}(u^{1/2}) A_+ \mathscr{D}(u^{1/2}) + \tau \mathscr{D}(u^{1/2}) A_- \mathscr{D}(u^{1/2}). \end{aligned}$$

The above statements about (4.20) yield $\mathrm{Spec}\, A_+ \subset [0, 1]$ and $\mathrm{Spec}\, A_- \subset [0, 2(1 - \varepsilon)/\varepsilon]$. As $0 \leq u\tau$ we conclude from (4.22) that the spectrum of $\mathbb{1} - \tau \mathscr{D}(u^{1/2}) A \mathscr{D}(u^{1/2})$ is contained in $(0, 2/\varepsilon]$ for all $\eta \in [0, \eta_*]$. Therefore, using (4.21), we obtain

$$\sigma(z) = \frac{2}{\pi} \lim_{\eta \downarrow 0} \langle \sqrt{v_1 v_2}, (\mathbb{1} - \tau \mathscr{D}(u^{1/2}) A \mathscr{D}(u^{1/2}))^{-1} \sqrt{v_1 v_2} \rangle \geq \frac{\varepsilon}{\pi} \langle \boldsymbol{v}_0 \tilde{\boldsymbol{v}}_0 \rangle \gtrsim 1$$

uniformly for all $z \in \mathbb{D}_<$. Here, we used (3.10) in the last step. This shows (4.14) for $z \in \mathbb{D}_<$ for any $\tau_* > 0$ depending only on s_* and s^*.

We now show that there is $\tau_* > 0$ depending only on s_* and s^* such that (4.14) holds true for $z \in D(0, 1) \setminus \mathbb{D}_<$. This is proved by tracking the blowup of $(\mathbb{1} - \tau \mathscr{D}(u^{1/2}) A \mathscr{D}(u^{1/2}))^{-1}$ in $1 - \tau$ for $\tau \uparrow 1$ in (4.21) and establishing a compensation through $v_1 \sim v_2 \sim (1 - \tau)^{1/2}$ due to (3.10). This yields the upper and lower bound in (4.14). Since $\mathbb{1} - \tau \mathscr{D}(u^{1/2}) A \mathscr{D}(u^{1/2})$ in (4.21) is also invertible for $\eta = 0$ we may directly set $\eta = 0$ in the following argument.

We multiply the first component of the first relation in (3.32) by τ and solve for τu to obtain

$$\tau u = \frac{1}{2}(1 + \sqrt{1 - 4\tau v_1 v_2}) = 1 - \tau v_1 v_2 + O((1 - \tau)^2).$$

Therefore, using $v_1 \sim v_2 \sim (1-\tau)^{1/2}$, we have

$$\tau \mathscr{D}(u^{1/2})A\mathscr{D}(u^{1/2}) = A - \frac{\tau}{2}\big(\mathscr{D}(v_1v_2)A + A\mathscr{D}(v_1v_2)\big) + O\big((1-\tau)^2\big).$$

Moreover, from (4.20) we conclude that $Aa = a$ for $a := \sqrt{v_1v_2/u}/\|\sqrt{v_1v_2/u}\|_2$. Here, we also used (3.44) and (3.33) with $\eta = 0$.

Thus, the smallest eigenvalue of the positive operator $\mathbb{1} - \tau\mathscr{D}(u^{1/2})A\mathscr{D}(u^{1/2})$ satisfies

$$\lambda_{\min}\big(\mathbb{1} - \tau\mathscr{D}(u^{1/2})A\mathscr{D}(u^{1/2})\big) = \lambda_{\min}(\mathbb{1} - A) + \tau\langle a^2 v_1 v_2\rangle + O\big((1-\tau)^2\big)$$
$$= \tau\langle a^2 v_1 v_2\rangle + O\big((1-\tau)^2\big).$$

Here, we used multiple times that $Aa = a$. Therefore, as A is symmetric we conclude from (4.21) that

$$\sigma(z) = \frac{2}{\pi}\langle\sqrt{v_1v_2}, \big(\mathbb{1} - \tau\mathscr{D}(u^{1/2})A\mathscr{D}(u^{1/2})\big)^{-1}\sqrt{v_1v_2}\rangle$$
$$\geq \frac{\langle a, \sqrt{v_1v_2}\rangle^2}{\tau\langle a^2 v_1 v_2\rangle} + O(1-\tau).$$

Since $a \sim 1$ and $v_1 \sim v_2 \sim (1-\tau)^{1/2}$, there is $\tau_* \sim 1$ such that the lower bound in (4.14) holds true for $z \in D(0,1) \setminus \mathbb{D}_<$. Starting from (4.21), we similarly obtain

$$\sigma(z) \leq \frac{\langle v_1 v_2\rangle}{\tau\langle a^2 v_1 v_2\rangle} + O(1-\tau).$$

Using the positivity of a, $v_1 \sim v_2 \sim (1-\tau)^{1/2}$ and possibly shrinking $\tau_* \sim 1$ the upper bound in (4.14) for $z \in D(0,1) \setminus \mathbb{D}_<$ follows. This concludes the proof of Lemma 4.1. $\square$

As $\sigma(z) = 0$ for $|z| \geq 1$, we conclude from (2.6) that σ is nonnegative on $\mathbb{C}$. We use (4.10) to compute the total mass of the measure on $\mathbb{C}$ defined by σ. Clearly, $\boldsymbol{u}_0 = \boldsymbol{v}_0/S_d\boldsymbol{v}_0$ and using (4.11) and (4.10), we obtain

$$\lim_{\tau\uparrow 1}\int_{|z'|^2 \leq \tau}\sigma(z')\,\mathrm{d}^2 z' = 1 - \lim_{\tau\uparrow 1}\langle\boldsymbol{v}_0 S_o \boldsymbol{v}_0\rangle = 1.$$

Here, we used that $\lim_{\tau\uparrow 1}\boldsymbol{v}_0 = 0$ by (3.10). Hence, as $\sigma(z) = 0$ for $|z| \geq 1$ it defines a probability density on $\mathbb{C}$, which concludes the proof of Proposition 2.4.
$\square$

REMARK 4.2 (Jump height).   In fact, it is possible to compute the jump height of the density of states σ at the edge $\tau = |z|^2 = 1$. Let s_1 and s_2 be two eigenvectors of S^t and S, respectively, associated to the eigenvalue 1, that is, $S^t s_1 = s_1$ and $Ss_2 = s_2$. Note that s_1 and s_2 are unique up to multiplication by a scalar.

With this notation, expanding $\mathbf{v}^\tau$ for $\tau \leq 1$ around $\tau = 1$ yields

$$v_1 = \sqrt{1-\tau}\Big(\frac{\langle s_1 s_2 \rangle \langle s_2 \rangle}{\langle s_1^2 s_2^2 \rangle \langle s_1 \rangle}\Big)^{1/2} s_1 + O\big((1-\tau)^{3/2}\big),$$

$$v_2 = \sqrt{1-\tau}\Big(\frac{\langle s_1 s_2 \rangle \langle s_1 \rangle}{\langle s_1^2 s_2^2 \rangle \langle s_2 \rangle}\Big)^{1/2} s_2 + O\big((1-\tau)^{3/2}\big).$$

Therefore, solving (3.32) for τu and expanding in $1-\tau$, we obtain that σ has a jump of height

$$\lim_{|z|^2 \uparrow 1} \sigma(z) = \frac{1}{\pi} \lim_{\tau \uparrow 1} \partial_\tau \big(\tau \langle \mathbf{u}_0 \rangle\big) = \frac{1}{\pi} \frac{\langle s_1 s_2 \rangle^2}{\langle s_1^2 s_2^2 \rangle}.$$

5. Local law. We begin this section with a notion for high probability estimates.

DEFINITION 5.1 (Stochastic domination).   Let $C \colon \mathbb{R}_+^2 \to \mathbb{R}_+$ be a given function which depends only on a, φ, τ_*, τ^* and the model parameters. If $\Phi = (\Phi^{(n)})_n$ and $\Psi = (\Psi^{(n)})_n$ are two sequences of nonnegative random variables, then we will say that Φ is *stochastically dominated* by Ψ, $\Phi \prec \Psi$, if for all $\varepsilon > 0$ and $D > 0$ we have

$$\mathbb{P}\big(\Phi^{(n)} \geq n^\varepsilon \Psi^{(n)}\big) \leq \frac{C(\varepsilon, D)}{n^D}$$

for all $n \in \mathbb{N}$.

As a trivial consequence of $\mathbb{E} x_{ij} = 0$, (2.1) and (2.2) we remark that

(5.1) $$|x_{ij}| \prec n^{-1/2}.$$

5.1. *Local law for $\mathbf{H}^z$.*   Let (v_1^τ, v_2^τ) be the positive solution of (2.4) and u^τ defined as in (3.25). In the whole section, we will always evaluate v_1^τ, v_2^τ and u^τ at $\tau = |z|^2$ and mostly suppress the dependence on τ and $|z|^2$, respectively, in our notation. Recall that $\mathbf{M}^z$ is defined in (3.5). Note that although v_1, v_2 and u are rotationally invariant in $z \in \mathbb{C}$, the dependence of $\mathbf{M}^z$ on z is not rotationally symmetric.

For the following theorem, we remark that the sets $\mathbb{D}_<$ and $\mathbb{D}_>$ were introduced in (4.1).

THEOREM 5.2 (Local law for $\mathbf{H}^z$).   *Let X satisfy* (A) *and* (B) *and let $\mathbf{G} = \mathbf{G}^z$ be the resolvent of $\mathbf{H}^z$ as defined in (2.11). For fixed $\varepsilon \in (0, 1/2)$, the entrywise*

local law

$$(5.2) \quad \|\boldsymbol{G}^z(\eta) - \boldsymbol{M}^z(\eta)\|_{\max} \prec \begin{cases} \dfrac{1}{\sqrt{n\eta}} & \text{for } z \in \mathbb{D}_<, \eta \in \left[n^{-1+\varepsilon}, 1\right], \\[2ex] \dfrac{1}{\sqrt{n}} + \dfrac{1}{n\eta} & \text{for } z \in \mathbb{D}_>, \eta \in \left[n^{-1+\varepsilon}, 1\right], \\[2ex] \dfrac{1}{\sqrt{n}\eta^2} & \text{for } z \in \mathbb{D}_< \cup \mathbb{D}_>, \eta \in [1, \infty), \end{cases}$$

holds true. In particular,

$$(5.3) \quad \|\boldsymbol{g}(\eta) - \mathrm{i}\boldsymbol{v}(\eta)\|_\infty \prec \begin{cases} \dfrac{1}{\sqrt{n\eta}} & \text{for } z \in \mathbb{D}_<, \eta \in \left[n^{-1+\varepsilon}, 1\right], \\[2ex] \dfrac{1}{\sqrt{n}} + \dfrac{1}{n\eta} & \text{for } z \in \mathbb{D}_>, \eta \in \left[n^{-1+\varepsilon}, 1\right], \\[2ex] \dfrac{1}{\sqrt{n}\eta^2} & \text{for } z \in \mathbb{D}_< \cup \mathbb{D}_>, \eta \in [1, \infty), \end{cases}$$

where $\boldsymbol{g} = (\langle \boldsymbol{e}_i, \boldsymbol{G}\boldsymbol{e}_i \rangle)_{i=1}^{2n}$ denotes the vector of diagonal entries of the resolvent $\boldsymbol{G}^z$.

For a nonrandom vector $\boldsymbol{y} \in \mathbb{C}^{2n}$ with $\|\boldsymbol{y}\|_\infty \le 1$, we have

$$(5.4) \quad |\langle \boldsymbol{y}, \boldsymbol{g}(\eta) - \mathrm{i}\boldsymbol{v}(\eta) \rangle| \prec \begin{cases} \dfrac{1}{n\eta} & \text{for } z \in \mathbb{D}_<, \eta \in \left[n^{-1+\varepsilon}, 1\right], \\[2ex] \dfrac{1}{n} + \dfrac{1}{(n\eta)^2} & \text{for } z \in \mathbb{D}_>, \eta \in \left[n^{-1+\varepsilon}, 1\right], \\[2ex] \dfrac{1}{n\eta^2} & \text{for } z \in \mathbb{D}_< \cup \mathbb{D}_>, \eta \in [1, \infty). \end{cases}$$

As an easy consequence, we can now prove Corollary 2.6.

PROOF OF COROLLARY 2.6. Let $y \in \mathbb{C}^n$ be an eigenvector of X corresponding to the eigenvalue $\sigma \in \operatorname{Spec} X$ with $|\sigma|^2 \le \rho(S) - \tau_*$. Then the $2n$-vector $(0, y)$ is contained in the kernel of $\boldsymbol{H}^\sigma$. Therefore, (2.9) is an easy consequence of (5.3) (Compare with the proof of Corollary 1.14 in [1]). $\square$

We recall our normalization of the trace, $\operatorname{Tr} \mathbb{1} = 1$, from (1.3).

PROOF OF THEOREM 5.2. Recall from the beginning of Section 3 how our problem can be cast into the setup of [3]. In the regime $z \in \mathbb{D}_<$, we follow the structure of the proof of Theorem 2.9 in [3] and in the regime $z \in \mathbb{D}_>$ the proof of Proposition 7.1 in [3] until the end of Step 1. In fact, the arguments from these proofs can be taken over directly with three important adjustments. The flatness assumption (3.1) is used heavily in [3] in order to establish bounds (Theorem 2.5 in [3]) on the deterministic limit of the resolvent and for establishing the stability

of the matrix Dyson equation, cf. (5.5) below (Theorem 2.6 in [3]). Since this assumption is violated in our setup we present appropriately adjusted versions of these theorems (Proposition 3.2 and Proposition 3.3 in [3]). We will also take over the proof of the fluctuation averaging result (Proposition 5.5 below) for $\boldsymbol{H}^z$ from [3] since the flatness did not play a role in that proof at all. Note that the η^{-2}-decay in the spectral parameter regime $\eta \geq 1$ was not covered in [3]. But this decay simply follows by using the bounds $\|\boldsymbol{M}^z(\eta)\|_{\max} + \|\boldsymbol{G}^z(\eta)\|_{\max} \leq \frac{2}{\eta}$ instead of just $\|\boldsymbol{M}^z(\eta)\|_{\max} + \|\boldsymbol{G}^z(\eta)\|_{\max} \leq C$ along the proof.

As in [3], we choose a pseudo-metric d on $\{1, \ldots, 2n\}$. Here, this pseudo-metric is particularly simple,

$$d(i, j) := \begin{cases} 0 & \text{if } i = j \text{ or } i = j + n \text{ or } j = i + n, \\ \infty & \text{otherwise,} \end{cases} \qquad i, j = 1, \ldots, 2n.$$

With this choice of d, the matrix $\boldsymbol{H}^z$ satisfies all assumptions in [3] apart from the flatness.

We will now show that as in [3] the resolvent $\boldsymbol{G}^z$ satisfies the *perturbed matrix Dyson equation*:

$$(5.5) \qquad -\mathbb{1} = \left(\mathrm{i}\eta\mathbb{1} - \boldsymbol{A}^z + \widetilde{\mathcal{S}}[\boldsymbol{G}^z(\eta)]\right)\boldsymbol{G}^z(\eta) + \boldsymbol{D}^z(\eta).$$

Here, $\boldsymbol{A}^z$ is given by (3.3),

$$(5.6) \qquad \boldsymbol{D}(\eta) := \boldsymbol{D}^z(\eta) := -\left(\widetilde{\mathcal{S}}[\boldsymbol{G}^z(\eta)] + \boldsymbol{H}^z - \boldsymbol{A}^z\right)\boldsymbol{G}^z(\eta),$$

is a random error matrix and $\widetilde{\mathcal{S}}$ is a slight modification of the operator $\mathcal{S}$ defined in (3.3),

$$(5.7) \qquad \widetilde{\mathcal{S}}[\boldsymbol{W}] := \mathbb{E}(\boldsymbol{H}^z - \boldsymbol{A}^z)\boldsymbol{W}(\boldsymbol{H}^z - \boldsymbol{A}^z) = \begin{pmatrix} \mathrm{diag}(Sw_2) & T \odot W_{21}^t \\ T^* \odot W_{12}^t & \mathrm{diag}(S^t w_1) \end{pmatrix}.$$

Here, $\odot$ denotes the Hadamard product, that is, for matrices $A = (a_{ij})_{i,j=1}^l$ and $B = (b_{ij})_{i,j=1}^l$, we define their Hadamard product through $(A \odot B)_{ij} := a_{ij}b_{ij}$ for $i, j = 1, \ldots, l$. Moreover, we used the conventions from (3.4) for $\boldsymbol{W}$ and introduced the matrix $T \in \mathbb{C}^{n \times n}$ with entries

$$t_{ij} := \mathbb{E}x_{ij}^2.$$

Note that in contrast to [3] the matrix $\boldsymbol{M}$ solves (3.2), which is given in terms of the operator $\mathcal{S}$ and not $\widetilde{\mathcal{S}}$ (we remark that $\widetilde{\mathcal{S}}$ was denoted by $\mathcal{S}$ in [3]). As we will see below, this will not affect the proof, since the entries of the matrix T are of order N^{-1}, and thus the off-diagonal terms in (5.7) of $\widetilde{\mathcal{S}}$ are negligible.

We will see that $\boldsymbol{D} = \boldsymbol{D}^z$ is small in the entrywise maximum norm

$$\|\boldsymbol{W}\|_{\max} := \max_{i,j=1}^{2n} |w_{ij}|,$$

$W = (w_{ij})_{i,j=1}^{2n}$, and use the stability of (5.5) to show that $G(\eta) = G^z(\eta)$ approaches $M(\eta) = M^z(\eta)$ defined in (3.5) as $n \to \infty$, that is, we will show that

$$(5.8) \qquad \Lambda(\eta) := \| G(\eta) - M(\eta) \|_{\max},$$

converges to zero. For simplicity, we will only consider the most difficult regime $z \in \mathbb{D}_<$ and $\eta \le 1$ inside the spectrum. The cases $z \in \mathbb{D}_>$ and $\eta \ge 1$ are similar but simpler and left to the reader. In a more general setup, they have recently been addressed in [8]. We simply follow the proof in Section 3 of [3] line by line until the flatness assumption is used. This happens for the first time inside the proof of Lemma 3.3. We therefore replace this lemma by the following modification.

LEMMA 5.3.   *Let $z \in \mathbb{D}_<$. Then*

$$\| D(\eta) \|_{\max} \prec \frac{1}{\sqrt{n}}, \qquad \eta \ge 1.$$

Furthermore, we have

$$(5.9) \qquad \| D(\eta) \|_{\max} \chi \big(\Lambda(\eta) \le n^{-\varepsilon} \big) \prec \frac{1}{\sqrt{n\eta}}, \qquad \eta \in [n^{-1+\varepsilon}, 1].$$

To show Lemma 5.3, we follow the proof of its analog, Lemma 3.3 in [3], where the flatness assumption as well as the assumptions that the spectral parameter is in the bulk of the spectrum (formulated as $\rho(\zeta) \ge \delta$ in [3]) are used only implicitly through the upper bound on M (Theorem 2.5 in [3]). However, the conclusion of this theorem clearly still holds in our setup because M has the 2×2-diagonal structure (3.5) and the vectors v_1, v_2 and u are bounded by Proposition 3.2 and (3.26).

We continue following the arguments of Section 3 of [3] using our Lemma 5.3 above instead of Lemma 3.3 there. The next step that uses the flatness assumption is the stability of the MDE (Theorem 2.6 in [3]) which shows that the bound (5.9) also implies

$$\Lambda(\eta) \chi \big(\Lambda(\eta) \le n^{-\varepsilon} \big) \prec \frac{1}{\sqrt{n\eta}}.$$

In our setup, this stability result is replaced by the following lemma whose proof is postponed until the end of the proof of Theorem 5.2.

LEMMA 5.4 (MDE stability).   *Suppose that some functions $D_{ab}, G_{ab} : \mathbb{R}_+ \to \mathbb{C}^{n \times n}$ for $a, b = 1, 2$ satisfy (5.5) with*

$$(5.10) \qquad D := \begin{pmatrix} D_{11} & D_{12} \\ D_{21} & D_{22} \end{pmatrix}, \qquad G := \begin{pmatrix} G_{11} & G_{12} \\ G_{21} & G_{22} \end{pmatrix},$$

and the additional constraints

(5.11) $\operatorname{Tr} G_{11} = \operatorname{Tr} G_{22}, \qquad \operatorname{Im} \boldsymbol{G} = \dfrac{1}{2i}(\boldsymbol{G} - \boldsymbol{G}^*) \qquad$ *is positive definite.*

There is a constant $\lambda_* \gtrsim 1$, depending only on $\mathcal{P}$, such that

(5.12) $\|\boldsymbol{G} - \boldsymbol{M}\|_{\max} \chi \lesssim \|\boldsymbol{D}\|_{\max} + \dfrac{1}{n}, \qquad \chi := \chi(\|\boldsymbol{G} - \boldsymbol{M}\|_{\max} \le \lambda_*),$

uniformly for all $z \in \mathbb{D}_< \cup \mathbb{D}_>$, where $\boldsymbol{M}(\eta) = \boldsymbol{M}^z(\eta)$ is defined in (3.5).

   Furthermore, there exist eight matrix valued functions $R_{ab}^{(k)} : \mathbb{R}_+ \to \mathbb{C}^{n \times n}$ with $a, b, k = 1, 2$, depending only on z and S, and satisfying $\|R_{ab}^{(k)}\|_\infty \lesssim 1$, such that

$$\big|\operatorname{Tr}\big[\operatorname{diag}(\boldsymbol{y})(\boldsymbol{G} - \boldsymbol{M})\big]\big| \chi \lesssim \max_{a,b,k=1,2} \big|\operatorname{Tr}\big[\operatorname{diag}\big(R_{ab}^{(k)} y_k\big) D_{ab}\big]\big|$$

(5.13)
$$+ \|\boldsymbol{y}\|_\infty \Big(\dfrac{1}{n} + \|\boldsymbol{D}\|_{\max}^2\Big),$$

uniformly for all $z \in \mathbb{D}_< \cup \mathbb{D}_>$ and $\boldsymbol{y} = (y_1, y_2) \in \mathbb{C}^{2n}$.

   The important difference between Theorem 2.6 in [3] and Lemma 5.4 above is the additional assumption (5.11) imposed on the solution of the perturbed MDE. This assumption is satisfied for the resolvent of the matrix $\boldsymbol{H}^z$ because of the 2×2-block structure (2.11). In fact, with the block decomposition for $\boldsymbol{G}$ as in (5.10), we have

$$G_{11}(\eta) = \dfrac{i\eta\mathbb{1}}{(X - z\mathbb{1})(X - z\mathbb{1})^* + \eta^2\mathbb{1}},$$

$$G_{22}(w) = \dfrac{i\eta\mathbb{1}}{(X - z\mathbb{1})^*(X - z\mathbb{1}) + \eta^2\mathbb{1}}.$$

Using Lemma 5.4 in the remainder of the proof of the entrywise local law in Section 3 of [3] completes the proof of (5.2).

   To see (5.4), we use the fluctuation averaging mechanism, which was first established for generalized Wigner matrices with Bernoulli entries in [17]. The following proposition is stated and proven as Proposition 3.4 in [3]. Since the flatness condition was not used in its proof at all, we simply take it over.

   PROPOSITION 5.5 (Fluctuation averaging). *Let $z \in \mathbb{D}_< \cup \mathbb{D}_>$, $\varepsilon \in (0, 1/2)$, $\eta \ge n^{-1}$ and Ψ a nonrandom control parameter such that $n^{-1/2} \le \Psi \le n^{-\varepsilon}$. Suppose the local law holds true in the form*

$$\big\|\boldsymbol{G}(\eta) - \boldsymbol{M}(\eta)\big\|_{\max} \prec \Psi.$$

Then for any nonrandom vector $y \in \mathbb{C}^n$ with $\|y\|_\infty \le 1$ we have

$$\max_{a,b=1,2} \big|\operatorname{Tr}\big[\operatorname{diag}(y) D_{ab}\big]\big| \prec \Psi^2,$$

where $D_{ab} \in \mathbb{C}^{n \times n}$, $a, b = 1, 2$, are the blocks of the error matrix

$$\boldsymbol{D}(\eta) = \begin{pmatrix} D_{11} & D_{12} \\ D_{21} & D_{22} \end{pmatrix},$$

which was defined in (5.6).

Using this proposition, the averaged local law (5.4) follows from (5.2) and (5.13). This completes the proof of Theorem 5.2. $\square$

PROOF OF LEMMA 5.4. We write (5.5) in the 2×2-block structure

$$\begin{pmatrix} \mathrm{diag}(\mathrm{i}\eta + Sg_2) & z\mathbb{1} \\ \bar{z}\mathbb{1} & \mathrm{diag}(\mathrm{i}\eta + S^t g_1) \end{pmatrix} \begin{pmatrix} G_{11} & G_{12} \\ G_{21} & G_{22} \end{pmatrix}$$

$$(5.14) \qquad = -\begin{pmatrix} \mathbb{1} & 0 \\ 0 & \mathbb{1} \end{pmatrix}$$

$$- \begin{pmatrix} D_{11} + (T \odot G_{21}^t)G_{21} & D_{12} + (T \odot G_{21}^t)G_{22} \\ D_{21} + (T^* \odot G_{12}^t)G_{11} & D_{22} + (T^* \odot G_{12}^t)G_{22} \end{pmatrix},$$

where we introduced $\boldsymbol{g} = (g_1, g_2) \in \mathbb{C}^{2n}$, the vector of the diagonal elements of $\boldsymbol{G}$.

We restrict the following calculation to the regime where $\|\boldsymbol{G}(\eta) - \boldsymbol{M}(\eta)\|_{\max} \leq \lambda_*$ for some sufficiently small λ_* in accordance with the characteristic function on the left-hand side of (5.12). In particular,

$$(5.15) \qquad \qquad \|\boldsymbol{g}(\eta) - \mathrm{i}\boldsymbol{v}(\eta)\|_\infty \leq \lambda_*.$$

Since by (2.4) and (3.5) the identity

$$\begin{pmatrix} \mathrm{i}\,\mathrm{diag}(\eta + Sv_2(\eta)) & z\mathbb{1} \\ \bar{z}\mathbb{1} & \mathrm{i}\,\mathrm{diag}(\eta + S^t v_1(\eta)) \end{pmatrix}^{-1} = -\boldsymbol{M}(\eta),$$

holds we infer from the smallness of $\|\boldsymbol{g} - \mathrm{i}\boldsymbol{v}\|_{\max}$ that the inverse of the first matrix factor on the left-hand side of (5.14) is bounded and satisfies

$$(5.16) \qquad \left\| \begin{pmatrix} \mathrm{diag}(\mathrm{i}\eta + Sg_2) & z\mathbb{1} \\ \bar{z}\mathbb{1} & \mathrm{diag}(\mathrm{i}\eta + S^t g_1) \end{pmatrix}^{-1} + \boldsymbol{M} \right\|_{\max} \lesssim \|\boldsymbol{g} - \mathrm{i}\boldsymbol{v}\|_{\max}.$$

Using this in (5.14) yields

$$(5.17) \quad \begin{aligned} \boldsymbol{G} + &\begin{pmatrix} \mathrm{diag}(\mathrm{i}\eta + Sg_2) & z\mathbb{1} \\ \bar{z}\mathbb{1} & \mathrm{diag}(\mathrm{i}\eta + S^t g_1) \end{pmatrix}^{-1} \\ &= \boldsymbol{M}\boldsymbol{D} + O\left(\|\boldsymbol{g} - \boldsymbol{v}\|_{\max}\|\boldsymbol{D}\|_{\max} + \|\boldsymbol{G} - \boldsymbol{M}\|_{\max}^2 + \frac{1}{n} \right), \end{aligned}$$

where we applied the simple estimate

$$\big\| (T \odot G_{ab}^t) G_{cd} \big\|_{\max}$$

$$(5.18) \qquad \lesssim \|G - M\|_{\max}^2 + \frac{1}{n} \|G - M\|_{\max} \|M\|_{\max} + \frac{1}{n} \|M\|_{\max}^2$$

$$\lesssim \|G - M\|_{\max}^2 + \frac{1}{n},$$

which follows from

$$\|T\|_{\max} \lesssim \frac{1}{n}.$$

Thus the diagonal elements g of G satisfy (2.19) with an error term d that is given by

$$(5.19) \qquad d = ((MD)_{ii})_{i=1}^{2n} + O\left(\|G - M\|_{\max}^2 + \frac{1}{n} \right).$$

Here, we used $\|D\|_{\max} \lesssim \|G - M\|_{\max}$, which follows directly from (5.5) and (3.2). With (3.21) and (3.22) in Proposition 3.3, the stability result on (2.19), we conclude that

$$(5.20) \qquad \|g - i v\|_{\infty} \lesssim \|D\|_{\max} + \|G - M\|_{\max}^2 + \frac{1}{n},$$

and that

$$(5.21) \quad |\langle y, g - i v \rangle| \lesssim |\mathrm{Tr}[\mathrm{diag}(R y) M D]| + \|D\|_{\max}^2 + \|G - M\|_{\max}^2 + \frac{1}{n},$$

for some bounded $R \in \mathbb{C}^{2n \times 2n}$ and any $y \in \mathbb{C}^{2n}$ with $\|y\|_{\infty} \leq 1$, respectively. Combining (5.16) with (5.17) and (5.20) yields

$$\|G - M\|_{\max} \lesssim \|D\|_{\max} + \|G - M\|_{\max}^2 + \frac{1}{n}.$$

By choosing λ_* sufficiently small, we may absorb the quadratic term of the difference $G - M$ on the right-hand side into the left-hand side and (5.12) follows. Using (5.12) in (5.21) to estimate the term $\|G - M\|_{\max}^2$ proves (5.13). $\square$

We use a standard argument to conclude from (5.4) the following statement about the number of eigenvalues $\lambda_i(z)$ of H^z in a small interval centered at zero.

LEMMA 5.6.   *Let $\varepsilon > 0$. Then*

$$(5.22) \qquad \#\{i : |\lambda_i(z)| \leq \eta\} \prec n\eta,$$

uniformly for all $\eta \geq n^{-1+\varepsilon}$ and $z \in \mathbb{D}_<$.
    Furthermore, we have

$$(5.23) \qquad \sup_{z \in \mathbb{D}_>} \frac{1}{|\lambda_i(z)|} \prec n^{1/2}.$$

PROOF. For the proof of (5.22), we realize that (5.2) implies a uniform bound on the resolvent elements up to the spectral scale $\eta \geq n^{-1+\varepsilon}$. Thus we have

$$\frac{\#\Sigma_\eta}{2\eta} \leq \sum_{i \in \Sigma_\eta} \frac{\eta}{\eta^2 + \lambda_i(z)^2} \leq 2n \operatorname{Im} \operatorname{Tr} \boldsymbol{G}^z(\eta) \prec n,$$

where $\Sigma_\eta := \{i : |\lambda_i(z)| \leq \eta\}$. Here, we used the normalization of the trace (1.3).

Before proving (5.23), we first establish that

$$(5.24) \qquad \frac{1}{|\lambda_i(z)|} \prec n^{1/2},$$

uniformly for $z \in \mathbb{D}_>$. We use (5.4) and $\langle \boldsymbol{v}(\eta) \rangle \sim \eta$ to estimate

$$(5.25) \qquad \frac{\eta}{\eta^2 + \lambda_i(z)^2} \leq 2n \operatorname{Im} \operatorname{Tr} \boldsymbol{G}^z(\eta) \prec n\eta + \frac{1}{n\eta^2},$$

with the choice $\eta := n^{-1/2-\varepsilon}$ for any $\varepsilon > 0$. This immediately implies $|\lambda_i(z)|^{-1} \prec n^{1/2+\varepsilon}$, hence (5.24). For the stronger bound (5.23), we use that $z \mapsto \operatorname{Im} \operatorname{Tr} \boldsymbol{G}^z(\eta)$ is a Lipschitz continuous function (with a Lipschitz constant $C\eta^{-2}$ uniformly in z) and that $\mathbb{D}_>$ is compact, so the second bound in (5.25) holds even after taking the supremum over $z \in \mathbb{D}_>$. Thus

$$\sup_{z \in \mathbb{D}_>} \frac{\eta}{\eta^2 + \lambda_i(z)^2} \leq 2n \sup_{z \in \mathbb{D}_>} \operatorname{Im} \operatorname{Tr} \boldsymbol{G}^z(\eta) \prec n\eta + \frac{1}{n\eta^2}$$

holds for $\eta := n^{-1/2-\varepsilon}$. From the last inequality, we easily conclude (5.23). $\square$

5.2. Local inhomogeneous circular law.

We start with an estimate on the smallest singular value of $X - z\mathbb{1}$, which will be used to control the dη-integral in the second term on the right-hand side of (2.15) for $\eta \leq n^{-1+\varepsilon}$. Notice that Proposition 5.7 is the only result in our proof of Theorem 2.5, which requires the entries of X to have a bounded density.

Adapting the proof of [11], Lemma 4.12, with the bounded density assumption to our setting, we obtain the following proposition.

PROPOSITION 5.7 (Smallest singular value of $X - z\mathbb{1}$). *Under condition* (2.3), *there is a constant* C, *depending only on* α, *such that*

$$(5.26) \qquad \mathbb{P}\left(\min_{i=1}^{2n} |\lambda_i(z)| \leq \frac{u}{n}\right) \leq C u^{2\alpha/(1+\alpha)} n^{\beta+1}$$

for all $u > 0$ *and* $z \in \mathbb{C}$.

PROOF. We follow the proof in [11] and explain the differences. Let $R_1, \ldots, R_n$ denote the rows of $\sqrt{n} X - z\mathbb{1}$. Proceeding as in [11] but using our normalization conventions, we are left with estimating

$$\mathbb{P}\left(n|\langle R_i, y \rangle| \leq \frac{u}{\sqrt{n}}\right)$$

uniformly for u and for arbitrary $y \in \mathbb{C}^n$ satisfying $\|y\|_2 = 1/\sqrt{n}$. We choose $j \in \{1, \dots, n\}$ such that $|y_j| \geq 1/\sqrt{n}$ and compute the conditional probability

$$\mathbb{P}_{ij} := \mathbb{P}\left(n|\langle R_i, y\rangle| \leq \frac{u}{\sqrt{n}} \Big| x_{i1}, \dots, \widehat{x_{ij}}, \dots, x_{in}\right)$$

$$= \int_{\mathbb{C}} \chi\left(\left|\frac{a}{y_j} + w\right| \leq \frac{u}{y_j\sqrt{n}}\right) f_{ij}(w)\, d^2w,$$

where a is independent of x_{ij}. Using (2.3) and $|y_j| \geq 1/\sqrt{n}$, we get

$$|\mathbb{P}_{ij}| \leq \left|\pi \frac{u}{y_j\sqrt{n}}\right|^{2\alpha/(1+\alpha)} \|f_{ij}\|_{1+\alpha} \leq (\pi u)^{2\alpha/(1+\alpha)} n^\beta.$$

Thus, $\mathbb{P}(n|\langle R_i, y\rangle| \leq u/\sqrt{n}) \leq (\pi u)^{2\alpha/(1+\alpha)} n^\beta$, which concludes the proof of (5.26) as in [11]. $\square$

For the following proof of Theorem 2.5, we recall that, without loss of generality, we are assuming that $\rho(S) = 1$ which can be obtained by a simple rescaling of X. Moreover, from (4.1), for $\tau_* > 0$ and $\tau^* > 1 + \tau_*$, we recall the notation

$$\mathbb{D}_< := \{z \in \mathbb{C}| |z|^2 \leq 1 - \tau_*\}, \qquad \mathbb{D}_> := \{z \in \mathbb{C}| 1 + \tau_* \leq |z|^2 \leq \tau^*\}.$$

PROOF OF THEOREM 2.5.   We start with the proof of part (i) of Theorem 2.5. We will estimate each term on the right-hand side of (2.15). Let $z_0 \in \mathbb{D}_<$. We suppress the τ dependence of v_1 in this proof but it will always be evaluated at $\tau = |z|^2$.

As $\operatorname{supp} f \subset D_\varphi(0)$, $a > 0$ and $z_0 \in \mathbb{D}_<$ we can assume that the integration domains of the d^2z integrals in (2.15) are $\mathbb{D}_<$ instead of $\mathbb{C}$. Hence, it suffices to prove every bound along the proof of (i) uniformly for $z \in \mathbb{D}_<$.

To begin, we estimate the first term in (2.15). Since

$$\log\left|\det(H^z - iT\mathbb{1})\right| = 2n\log T + \sum_{j=1}^n \log\left(1 + \frac{\lambda_j^2}{T^2}\right)$$

and the integral of $\Delta f_{z_0,a}$ over $\mathbb{C}$ vanishes as $f \in C_0^2(\mathbb{C})$, we obtain

(5.27)

$$\left|\frac{1}{4\pi n}\int_{\mathbb{C}} \Delta f_{z_0,a}(z) \log\left|\det(H^z - iT\mathbb{1})\right| d^2z\right|$$

$$\leq \frac{1}{2\pi}\int_{\mathbb{C}} |\Delta f_{z_0,a}(z)| \frac{\operatorname{Tr}((H^z)^2)}{T^2}\, d^2z.$$

Here, we used $\log(1 + x) \leq x$ for $x \geq 0$. Furthermore, if $|z| \leq 1$, then we have

(5.28) $\operatorname{Tr}((H^z)^2) = \frac{1}{n}\sum_{i,j=1}^n (x_{ij} - z\delta_{ij})(\overline{x_{ij}} - \bar{z}\delta_{ij}) \leq \frac{2}{n}\sum_{i,j=1}^n |x_{ij}|^2 + 2|z|^2 \prec 1,$

where we applied (1.3) in the first and (5.1) in the last step. Therefore, choosing $T := n^{100}$, we conclude from (5.27) and (5.28) that the first term in (2.15) is stochastically dominated by $n^{-1+2a}\|\Delta f\|_1$.

To control the second term on right-hand side of (2.15), we define

$$(5.29) \qquad I(z) := \int_0^T \left| \operatorname{Im} m^z(i\eta) - \langle v_1(\eta) \rangle \right| \mathrm{d}\eta$$

for $z \in \mathbb{D}_<$. We will conclude below the following lemma.

LEMMA 5.8. *For every $\delta > 0$ and $p \in \mathbb{N}$, there is a positive constant C, depending only on δ and p in addition to the model parameters and τ_*, such that*

$$(5.30) \qquad \sup_{z \in \mathbb{D}_<} \mathbb{E} I(z)^p \leq C \frac{n^{\delta p}}{n^p}.$$

We now show that this moment bound on $I(z)$ will yield that the second term in (2.15) is $\prec n^{-1+2a}\|\Delta f\|_1$. Indeed, for every $p \in \mathbb{N}$ and $\delta > 0$, using Hölder's inequality, we estimate

$$\mathbb{E}\left| \int_{\mathbb{C}} \Delta f_{z_0,a}(z) \int_0^T \left[\operatorname{Im} m^z(i\eta) - \langle v_1(\eta) \rangle \right] \mathrm{d}\eta\, \mathrm{d}^2 z \right|^p$$

$$(5.31) \qquad \leq \int_{\mathbb{C}} \cdots \int_{\mathbb{C}} \prod_{i=1}^p |\Delta f_{z_0,a}(z_i)| \prod_{i=1}^p \left(\mathbb{E} I(z_i)^p \right)^{1/p} \mathrm{d}^2 z_1 \cdots \mathrm{d}^2 z_p$$

$$\leq C \|\Delta f\|_1^p \frac{n^{\delta p + 2ap}}{n^p}.$$

Applying Chebyshev's inequality to (5.31) and using that $\delta > 0$ and p were arbitrary, we get

$$\left| \int_{\mathbb{C}} \Delta f_{z_0,a}(z) \int_0^T \operatorname{Im} m^z(i\eta) - \langle v_1(\eta) \rangle \mathrm{d}\eta\, \mathrm{d}^2 z \right| \prec n^{-1+2a}\|\Delta f\|_1.$$

Hence, the bound on the second term on the right-hand side of (2.15) follows once we have proven (5.30).

For the third term in (2.15), notice that the integrand is bounded by $C\eta^{-2}$ so it is bounded by $n^{2a}T^{-1}\|\Delta f\|_1$. This concludes the proof of (i) of Theorem 2.5 up to the proof of Lemma 5.8, which is given below.

We now turn to the proof of (ii). We will use an interpolation between the random matrix X and an independent Ginibre matrix $\widehat{X}$ together with the well-known result that a Ginibre matrix does not have any eigenvalues $|\lambda| \geq 1 + \tau_*$ with very high probability. With the help of (5.23), we will control the number of eigenvalues outside of the disk of radius $1 + \tau^*$ along the flow. We fix $\tau^* > 1 + \tau_*$.

Let $(\widehat{x}_{ij})^n_{i,j=1}$ be independent centered complex Gaussians of variance n^{-1}, that is, $\mathbb{E}\widehat{x}_{ij} = 0$ and $\mathbb{E}|\widehat{x}_{ij}|^2 = n^{-1}$. We set $\widehat{X} := (\widehat{x}_{ij})^n_{i,j=1}$, that is, $\widehat{X}$ is a Ginibre matrix. We denote the eigenvalues of $\widehat{X}$ by $\widehat{\sigma}_1, \ldots, \widehat{\sigma}_n$.

For $t \in [0, 1]$, we denote the spectral radius of the matrix $tS + (1 - t)E$ by $\rho_t := \rho(tS + (1-t)E)$, where E is the $n \times n$ matrix with entries $e_{ij} := 1/n$, $E = (e_{ij})^n_{i,j=1}$. Furthermore, we define

$$X^t := \rho_t^{-1/2}(tX + (1-t)\widehat{X}), \qquad \boldsymbol{H}^{z,t} := \begin{pmatrix} 0 & X^t - z\mathbb{1} \\ (X^t - z\mathbb{1})^* & 0 \end{pmatrix}$$

for $t \in [0, 1]$. The eigenvalues of X^t and $\boldsymbol{H}^{z,t}$ are denoted by σ_i^t and $\lambda_k^t(z)$, respectively, for $i = 1, \ldots, n$ and $k = 1, \ldots, 2n$. The one parameter family $t \mapsto X^t$ interpolates between X and $\widehat{X}$ by keeping the spectral radius of the variance matrix at constant one.

Note that $\|(X^t - z)^{-1}\|_2 = \max_{k=1}^{2n} |\lambda_k^t(z)|^{-1}$. We can apply Lemma 5.6 to the matrices X^t for any t to get

$$\sup_{z \in \mathbb{D}_>} \|(X^t - z)^{-1}\|_2 \prec n^{1/2}$$

uniformly in t from (5.23). In fact, the estimate can be strengthened to

$$(5.32) \qquad \sup_{t \in [0,1]} \sup_{z \in \mathbb{D}_>} \|(X^t - z)^{-1}\|_2 \prec n^{1/2}$$

exactly in the same way as (5.24) was strengthened to (5.23), we only need to observe that the two parameter family $(z, t) \mapsto \operatorname{Im} \operatorname{Tr} \boldsymbol{G}^{z,t}(\eta)$ is Lipschitz continuous in both variables, where $\boldsymbol{G}^{z,t}$ denotes the resolvent of $\boldsymbol{H}^{z,t}$.

Let γ be the circle in $\mathbb{C}$ centered at zero with radius $1 + \tau_*$. For $t \in [0, 1]$, we have

$$N(t) := \#\{i | |\sigma_i^t| \leq 1 + \tau_*\} = \frac{n}{2\pi \mathrm{i}} \int_\gamma \operatorname{Tr}((X^t - z)^{-1}) \, \mathrm{d}z,$$

where $\operatorname{Tr}: \mathbb{C}^{n \times n} \to \mathbb{C}$ denotes the normalized trace, that is, $\operatorname{Tr} \mathbb{1} = 1$. Due to (5.32) $N(t)$ is a continuous function of t. Thus, $N(t)$ is constant as a continuous integer valued function.

Using Corollary 2.3 of [19], we obtain that $\#\{k | |\widehat{\sigma}_k| \geq \tau^*\} = 0$ with very high probability. Furthermore, $\#\{k | \widehat{\sigma}_k \in \mathbb{D}_>\} = 0$ with very high probability by (5.32). Thus,

$$N(1) = N(0) = n - \#\{k | \widehat{\sigma}_k \in \mathbb{D}_>\} - \#\{k | |\widehat{\sigma}_k| \geq \tau^*\} = n$$

with very high probability which concludes the proof of (ii), and hence of Theorem 2.5. $\quad \square$

REMARK 5.9.   In the above proof, we showed that $\|\boldsymbol{H}^z\| \le C$ with very high probability via an interpolation argument using the norm-boundedness of a Ginibre matrix and the local law for the entire interpolating family. Robust upper bounds on the norm of random matrices are typically proven by a simple moment method. Such approach also applies here. For example, one may follow the proof of Lemma 7.2 in [18], and estimate every moment $\mathbb{E}|x_{ij}|^k$ by its maximum over all i, j. The final constant estimating $\|\boldsymbol{H}^z\|$ will not be optimal due to these crude bounds, but it will still only depend on s^* and μ_m from (2.1), (2.2). This argument is very robust, in particular it does not use hermiticity.

PROOF OF LEMMA 5.8.   To show (5.30), we use the following estimate, which converts a bound in $\prec$ into a moment bound. For every nonnegative random variable satisfying $Y \prec 1/n$ and $Y \le n^c$ for some $c > 0$, the pth moment is bounded by

$$(5.33) \qquad \mathbb{E}Y^p \le \mathbb{E}Y^p \chi\big(Y \le n^{\delta-1}\big) + \big(\mathbb{E}Y^{2p}\big)^{1/2}\big(\mathbb{P}(Y > n^{\delta-1})\big)^{1/2} \le C\frac{n^{p\delta}}{n^p},$$

for all $p \in \mathbb{N}$, $\delta > 0$ and for some $C > 0$, depending on c, p and δ.

As a first step in the proof of (5.30), we choose $\varepsilon \in (0, 1/2)$, split the $\mathrm{d}\eta$ integral in the definition of $I(z)$, (5.29), and consider the regimes $\eta \le n^{-1+\varepsilon}$ and $\eta \ge n^{-1+\varepsilon}$, separately. For $\eta \le n^{-1+\varepsilon}$, we compute

$$\int_0^{n^{-1+\varepsilon}} \operatorname{Im} m^z(i\eta)\, \mathrm{d}\eta = \frac{1}{2n} \sum_{i=1}^n \log\bigg(1 + \frac{n^{-2+2\varepsilon}}{\lambda_i^2}\bigg).$$

We recall that $\lambda_1, \ldots, \lambda_{2n}$ are the eigenvalues of $\boldsymbol{H}^z$. Therefore, (5.29) yields

$$
\begin{aligned}
\int_0^T \big[\operatorname{Im} m^z(i\eta) - \langle v_1(\eta)\rangle\big]\, \mathrm{d}\eta = {}& \frac{1}{n} \sum_{|\lambda_i| < n^{-l}} \log\bigg(1 + \frac{n^{-2+2\varepsilon}}{\lambda_i^2}\bigg) \\
& + \frac{1}{n} \sum_{|\lambda_i| \ge n^{-l}} \log\bigg(1 + \frac{n^{-2+2\varepsilon}}{\lambda_i^2}\bigg) \\
& - \int_0^{n^{-1+\varepsilon}} \langle v_1(\eta)\rangle\, \mathrm{d}\eta \\
& + \int_{n^{-1+\varepsilon}}^1 \big[\operatorname{Im} m^z(i\eta) - \langle v_1(\eta)\rangle\big]\, \mathrm{d}\eta \\
& + \int_1^T \big[\operatorname{Im} m^z(i\eta) - \langle v_1(\eta)\rangle\big]\, \mathrm{d}\eta.
\end{aligned}
$$

(5.34)

Here, $l \in \mathbb{N}$ is a large fixed integer to be chosen later.

We will estimate each of the terms on the right-hand side of (5.34) individually. For the first term in (5.34), we compute

$$\mathbb{E}\bigg(\frac{1}{n}\sum_{|\lambda_i|\leq n^{-l}}\log\bigg(1+\frac{n^{-2+2\varepsilon}}{\lambda_i^2}\bigg)\bigg)^p \leq \mathbb{E}\bigg[\log^p\bigg(1+\frac{n^{-2+2\varepsilon}}{\lambda_n^2}\bigg)\chi\big(\lambda_n\leq n^{-l}\big)\bigg]$$

$$\leq C\mathbb{E}\big[|\log\lambda_n|^p\chi\big(\lambda_n\leq n^{-l}\big)\big]$$

for some constant $C > 0$ independent of n. We compute the expectation directly

$$\mathbb{E}\big[|\log\lambda_n|^p\chi\big(\lambda_n\leq n^{-l}\big)\big] = p\int_{l\log n}^{\infty}\mathbb{P}\big(\lambda_n\leq e^{-t}\big)t^{p-1}\,\mathrm{d}t$$

$$\leq Cn^{\beta+1+2\alpha/(1+\alpha)}\int_{l\log n}^{\infty}t^{p-1}e^{-2\alpha t/(1+\alpha)}\,\mathrm{d}t.$$

Here, we applied (5.26) in Proposition 5.7 with $u = e^{-t}n$. Choosing l large enough, depending on α, β and p, we obtain that the right-hand side is smaller than n^{-p}. This shows the bound (5.30) for the first term in (5.34).

To estimate the second term on the right-hand side of (5.34), we decompose the sum into three regimes, $n^{-l} \leq |\lambda_i| < n^{-1+\varepsilon}$, $n^{-1+\varepsilon} \leq |\lambda_i| < n^{-1/2}$ and $n^{-1/2} \leq |\lambda_i|$.

For the first regime, we use (5.22) with $\eta = n^{-1+\varepsilon}$ and $\log(1 + n^{-2+2\varepsilon+l}) \leq C\log n$ to get

$$(5.35)\quad \frac{1}{n}\sum_{|\lambda_i|\in[n^{-l},n^{-1+\varepsilon}]}\log\bigg(1+\frac{n^{-2+2\varepsilon}}{\lambda_i^2}\bigg) \leq \frac{C\log n}{n}\#\{i:|\lambda_i|\leq n^{-1+\varepsilon}\} \prec \frac{n^{\varepsilon}}{n}.$$

As this sum is clearly polynomially bounded in n, we can apply (5.33) to conclude that the first regime of the second term in (5.34) fulfills the moment bound in (5.30).

For the intermediate regime, due to the symmetry $\mathrm{Spec}(\boldsymbol{H}^z) = -\mathrm{Spec}(\boldsymbol{H}^z)$, we only consider the positive eigenvalues. We decompose the interval $[n^{-1+\varepsilon}, n^{-1/2}]$ into dyadic intervals of the form $[\eta_k, \eta_{k+1}]$, where $\eta_k := 2^k n^{-1+\varepsilon}$. Thus, we obtain

$$(5.36)\quad \begin{aligned}\frac{1}{n}&\sum_{|\lambda_i|\in[n^{-1+\varepsilon},n^{-1/2}]}\log\bigg(1+\frac{n^{-2+2\varepsilon}}{\lambda_i^2}\bigg)\\ &\leq \frac{2}{n}\sum_{k=0}^{N}\sum_{\lambda_i\in[\eta_k,\eta_{k+1}]}\log\bigg(1+\frac{n^{-2+2\varepsilon}}{\lambda_i^2}\bigg)\prec\frac{n^{\varepsilon}}{n},\end{aligned}$$

where we introduced $N = O(\log n)$ in the first step. Moreover, we used the monotonicity of the logarithm, $\log(1 + x) \leq x$ in the last step and the following consequence of (5.22):

$$\#\{i:\lambda_i\in[\eta_k,\eta_{k+1}]\} \leq \#\{i:|\lambda_i|\leq\eta_{k+1}\} \prec n^{\varepsilon}2^{k+1}.$$

The left-hand side of (5.36) is trivially bounded by $\log 2$. Therefore, applying (5.33) to the left-hand side of (5.36), we conclude that it satisfies the moment estimate in (5.30).

For estimating the second term in (5.34) in the third regime, employing $|\lambda_i| \geq n^{-1/2}$ and $\log(1 + x) \leq x$, we obtain

$$(5.37) \quad \frac{1}{n} \sum_{|\lambda_i| \geq n^{-1/2}} \log\left(1 + \frac{n^{-2+2\varepsilon}}{\lambda_i^2}\right) \leq \frac{1}{n} \sum_{|\lambda_i| \geq n^{-1/2}} \log(1 + n^{-1+2\varepsilon}) \leq \frac{2n^{2\varepsilon}}{n}.$$

Here, we used that $\boldsymbol{H}^z$ has $2n$ eigenvalues (counted with multiplicities). This deterministic bound and (5.33) imply that the moments of this sum are bounded by the right-hand side in (5.30).

Combining the estimates in these three regimes, (5.35), (5.36) and (5.37), we conclude that the second term in (5.34) satisfies the moment bound in (5.30).

We now estimate the third term on the right-hand side of (5.34). Since $\boldsymbol{v} \sim 1$ for $z \in \mathbb{D}_<$ and $\eta \leq 1$ by (3.10), the pth power of the third term is immediately bounded by the right-hand side of (5.30).

To bound the fourth and fifth term in (5.34), we note that $\operatorname{Im} m^z(\mathrm{i}\eta) = \langle \boldsymbol{g}(\eta) \rangle$ for $\eta > 0$. Recalling the choice $T = n^{100}$, we obtain

$$\int_{n^{-1+\varepsilon}}^1 \left|\operatorname{Im} m^z(\mathrm{i}\eta) - \langle v_1(\eta) \rangle\right| \mathrm{d}\eta \prec \frac{n^\varepsilon}{n}, \qquad \int_1^T \left|\operatorname{Im} m^z(\mathrm{i}\eta) - \langle v_1(\eta) \rangle\right| \mathrm{d}\eta \prec \frac{1}{n}$$

from the first and third regime in (5.4) with $\boldsymbol{y} = 1$. As the integrands are bounded by n^2 trivially (5.33) yields that the moments of the fourth and fifth term in (5.34) are bounded by the right-hand side in (5.30).

Since $\varepsilon \in (0, 1/2)$ was arbitrary, this concludes the proof of (5.30). $\square$

APPENDIX A: EXISTENCE AND UNIQUENESS

In this section, we prove Lemma 2.2.

The existence and uniqueness of the solution of (2.4) will be a consequence of the existence and uniqueness of the matrix Dyson equation

$$(A.1) \qquad -\boldsymbol{M}^{-1}(\eta) = \mathrm{i}\eta\mathbb{1} - \boldsymbol{A} + \mathcal{S}[\boldsymbol{M}(\eta)].$$

Note that $\boldsymbol{A} \in \mathbb{C}^{2n \times 2n}$ and $\mathcal{S} \colon \mathbb{C}^{2n \times 2n} \to \mathbb{C}^{2n \times 2n}$ were defined in (3.3).

The matrix Dyson equation, (A.1), has a unique solution under the constraint that the imaginary part

$$\operatorname{Im} \boldsymbol{M} := \frac{1}{2\mathrm{i}}(\boldsymbol{M} - \boldsymbol{M}^*)$$

is positive definite. This was established in [24]. In the context of random matrices, (A.1) was studied in [3].

In the following proof, for vectors $a, b, c, d \in \mathbb{C}^n$, we will denote the $2n \times 2n$ matrix having diagonal matrices with diagonals a, b, c, d on its top-left, top-right, lower-left and lower-right $n \times n$ blocks, respectively, by

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} := \begin{pmatrix} \operatorname{diag} a & \operatorname{diag} b \\ \operatorname{diag} c & \operatorname{diag} d \end{pmatrix} \in \mathbb{C}^{2n \times 2n}.$$

PROOF OF LEMMA 2.2.   We show that there is a bijection between the solutions of (A.1) with positive definite imaginary part $\operatorname{Im} M$ and the positive solutions of (3.6).

We remark that (A.1) implies that there are functions $a, b, c, d \colon \mathbb{R}_+ \to \mathbb{C}^n$ such that for all $\eta > 0$ we have

$$(A.2) \qquad\qquad M(\eta) = \begin{pmatrix} a(\eta) & b(\eta) \\ c(\eta) & d(\eta) \end{pmatrix}.$$

First, we show that $\operatorname{Im} \operatorname{diag} M$ is a positive solution of (3.6) if M satisfies (A.1) and $\operatorname{Im} M$ is positive definite. Due to (A.2), multiplying (A.1) by M yields that (A.1) is equivalent to

$$(A.3) \qquad \begin{aligned} -1 &= \mathrm{i}\eta a + aSd + b\bar{z}, & 0 &= \mathrm{i}\eta b + za + bS^t a, \\ -1 &= \mathrm{i}\eta d + dS^t a + zc, & 0 &= \mathrm{i}\eta c + \bar{z}d + cSd. \end{aligned}$$

Solving the second and fourth relations in (A.3) for b and c, we obtain

$$(A.4) \qquad\qquad b = -\frac{za}{\mathrm{i}\eta + S^t a}, \qquad c = -\frac{\bar{z}d}{\mathrm{i}\eta + Sd}.$$

Plugging the first relation in (A.4) into the first relation in (A.3) and the second relation in (A.4) into the third relation in (A.3) and dividing the results by a and d, respectively, implies

$$-\frac{1}{a} = \mathrm{i}\eta + Sd - \frac{|z|^2}{\mathrm{i}\eta + S^t a}, \qquad -\frac{1}{d} = \mathrm{i}\eta + S^t a - \frac{|z|^2}{\mathrm{i}\eta + Sd}.$$

Therefore, if a and d are purely imaginary then $(\operatorname{Im} a, \operatorname{Im} d) = -\mathrm{i}(a, d)$ will fulfill (3.6).

In order to prove that a and d are purely imaginary, we define

$$\widetilde{M} := \begin{pmatrix} \widetilde{a}(\eta) & \widetilde{b}(\eta) \\ \widetilde{c}(\eta) & \widetilde{d}(\eta) \end{pmatrix} := \begin{pmatrix} -\bar{a} & \frac{z}{\bar{z}}\bar{b} \\ \frac{\bar{z}}{z}\bar{c} & -\bar{d} \end{pmatrix}.$$

The goal is to conclude $M = \widetilde{M}$, and hence $a = -\bar{a}$ and $d = -\bar{d}$, from the uniqueness of the solution of (A.1) with positive definite imaginary part. Since the relations (A.3) are fulfilled if a, b, c, d are replaced by $\widetilde{a}, \widetilde{b}, \widetilde{c}, \widetilde{d}$, respectively, $\widetilde{M}$ satisfies (A.1). For $j = 1, \ldots, n$, we define the 2×2 matrices

$$M_j := \begin{pmatrix} a_j & b_j \\ c_j & d_j \end{pmatrix}, \qquad \widetilde{M}_j := \begin{pmatrix} \widetilde{a}_j & \widetilde{b}_j \\ \widetilde{c}_j & \widetilde{d}_j \end{pmatrix}.$$

Note that $\operatorname{Im} \boldsymbol{M}$ is positive definite if and only if $\operatorname{Im} M_j$ is positive definite for all $j = 1, \ldots, n$. Similarly, the positive definiteness of $\operatorname{Im} \widetilde{\boldsymbol{M}}$ is equivalent to the positive definiteness of $\operatorname{Im} \widetilde{M}_j$ for all $j = 1, \ldots, n$. We have

$$
\operatorname{Im} M_j = \begin{pmatrix} \operatorname{Im} a_j & \frac{1}{2\mathrm{i}}(b_j - \bar{c}_j) \\ \frac{1}{2\mathrm{i}}(c_j - \bar{b}_j) & \operatorname{Im} d_j \end{pmatrix},
$$

$$
\operatorname{Im} \widetilde{M}_j := \begin{pmatrix} \operatorname{Im} a_j & \frac{z}{2\mathrm{i}\bar{z}}(\bar{b}_j - c_j) \\ \frac{\bar{z}}{2\mathrm{i}z}(\bar{c}_j - b_j) & \operatorname{Im} d_j \end{pmatrix}.
$$

As $\operatorname{Tr} \operatorname{Im} \widetilde{M}_j = \operatorname{Tr} \operatorname{Im} M_j$ and $\det \operatorname{Im} \widetilde{M}_j = \det \operatorname{Im} M_j$ for all $j = 1, \ldots, n$ we get that $\widetilde{\boldsymbol{M}}$ is a solution of (A.1) with positive definite imaginary part $\operatorname{Im} \widetilde{\boldsymbol{M}}$. Thus, the uniqueness of the solution of (A.1) implies $\boldsymbol{M} = \widetilde{\boldsymbol{M}}$ as well as $a = -\bar{a}$ and $d = -\bar{d}$.

Moreover, since

$$
\operatorname{Im} \boldsymbol{M} = \begin{pmatrix} \operatorname{Im} a & (b - \bar{c})/(2\mathrm{i}) \\ (c - \bar{b})/(2\mathrm{i}) & \operatorname{Im} d \end{pmatrix}
$$

is positive definite we have that $\operatorname{Im} a > 0$ and $\operatorname{Im} d > 0$. Hence, $(\operatorname{Im} a, \operatorname{Im} d)$ is a positive solution of (3.6).

Conversely, let $\boldsymbol{v} = (v_1, v_2) \in \mathbb{C}^{2n}$ be a solution of (3.6) satisfying $\boldsymbol{v} > 0$ and u be defined as in (3.25). Because of (3.25), we obtain that $\boldsymbol{M} = \boldsymbol{M}^z$, defined as in (3.5), is a solution of (A.1). To conclude that $\operatorname{Im} \boldsymbol{M}$ is positive definite, it suffices to show that $\det \operatorname{Im} M_j > 0$ for all $j = 1, \ldots, n$ with

$$
M_j := \begin{pmatrix} \mathrm{i}(v_1)_j & -z u_j \\ -\bar{z} u_j & \mathrm{i}(v_2)_j \end{pmatrix}
$$

as $\operatorname{Tr} \operatorname{Im} M_j = (v_1)_j + (v_2)_j > 0$. Since $z u_j - \overline{\bar{z} u_j} = 0$ for all $j = 1, \ldots, n$ by (3.25) we obtain

$$
\det \operatorname{Im} M_j = (v_1)_j (v_2)_j - \frac{1}{4} |z u_j - \overline{\bar{z} u_j}|^2 = (v_1)_j (v_2)_j > 0.
$$

Therefore, there is a bijection between the solutions of (A.1) with positive definite imaginary part and the positive solutions of (3.6). Appealing to the existence and uniqueness of (A.1) proved in [24] concludes the proof of Lemma 2.2.   □

APPENDIX B: CONTRACTION-INVERSION LEMMA

PROOF OF LEMMA 3.7.   The bounds (3.55) imply that $\mathbb{1} - \boldsymbol{A}\boldsymbol{B}$ is invertible and

$$
\big\| (\mathbb{1} - \boldsymbol{A}\boldsymbol{B})^{-1} \big\|_2 \le \frac{1}{c_1 \eta}.
$$

The main point of this lemma is to show that $(\mathbb{1} - \boldsymbol{AB})^{-1}\boldsymbol{p}$ can be bounded independently of η for $\boldsymbol{p}$ satisfying (3.58). We introduce $\boldsymbol{h} := (\mathbb{1} - \boldsymbol{AB})^{-1}\boldsymbol{p}$. Thus, (3.59) is equivalent to $\|\boldsymbol{h}\|_2 \leq C\|\boldsymbol{p}\|_2$ for some $C > 0$, which depends only on c_1, c_2, c_3 and ε. Without loss of generality, we may assume that $\|\boldsymbol{h}\|_2 = 1$. We decompose

$$(B.1) \qquad\qquad \boldsymbol{h} = \alpha \boldsymbol{b}_- + \beta \boldsymbol{b}_+ + \gamma \boldsymbol{x},$$

where $\alpha = \langle \boldsymbol{b}_-, \boldsymbol{h} \rangle$, $\beta = \langle \boldsymbol{b}_+, \boldsymbol{h} \rangle$ and $\boldsymbol{x} \perp \boldsymbol{b}_\pm$ satisfying $\|\boldsymbol{x}\|_2 = 1$, thus $|\alpha|^2 + |\beta|^2 + |\gamma|^2 = 1$. Since $\boldsymbol{B} = \boldsymbol{B}^*$, we have $\boldsymbol{b}_+ \perp \boldsymbol{b}_-$ and $\boldsymbol{Bx} \perp \boldsymbol{b}_\pm$. Hence, we obtain

$$\|\boldsymbol{ABh}\|_2^2 \leq \|\boldsymbol{Bh}\|_2^2 \leq |\alpha|^2 \|\boldsymbol{B}\|_2^2 + |\beta|^2 \|\boldsymbol{B}\|_2^2 + |\gamma|^2 \|\boldsymbol{Bx}\|_2^2 \leq 1 - \varepsilon + \varepsilon\big(|\alpha|^2 + |\beta|^2\big),$$

where we used $\|\boldsymbol{A}\|_2 \leq 1$, $\|\boldsymbol{B}\|_2 \leq 1$ and $\|\boldsymbol{Bx}\|_2 \leq 1 - \varepsilon$ in the last step. Therefore, if $|\alpha|^2 + |\beta|^2 \leq 1 - \delta$ for some $\delta > 0$ to be determined later, then $\|\boldsymbol{ABh}\|_2 \leq \sqrt{1 - \varepsilon\delta}\|\boldsymbol{h}\|_2 \leq (1 - \varepsilon\delta/2)\|\boldsymbol{h}\|_2$, and thus

$$(B.2) \qquad\qquad 1 = \|\boldsymbol{h}\|_2 \leq \frac{2}{\varepsilon\delta}\|\boldsymbol{p}\|_2.$$

For the rest of the proof, we assume that $|\alpha|^2 + |\beta|^2 \geq 1 - \delta$. In the regime, where $|\alpha|$ is relatively large, we compute $\langle \boldsymbol{b}_-, (\mathbb{1} - \boldsymbol{AB})\boldsymbol{h} \rangle$, capitalize on the positivity of $\langle \boldsymbol{b}_-, (\mathbb{1} - \boldsymbol{AB})\boldsymbol{b}_- \rangle$ and treat all other terms as errors. In the opposite regime, where $|\beta|$ is relatively large, we use the positivity of $\langle \boldsymbol{b}_+, (\mathbb{1} - \boldsymbol{AB})\boldsymbol{b}_+ \rangle$.

Using (B.1), we compute

$$\begin{aligned}
\langle \boldsymbol{b}_-, \boldsymbol{p} \rangle &= \langle \boldsymbol{b}_-, (\mathbb{1} - \boldsymbol{AB})\boldsymbol{h} \rangle \\
&= \alpha\big(1 + \|\boldsymbol{B}\|_2 \langle \boldsymbol{b}_-, \boldsymbol{Ab}_- \rangle\big) - \beta\|\boldsymbol{B}\|_2 \langle \boldsymbol{b}_-, \boldsymbol{Ab}_+ \rangle - \gamma \langle \boldsymbol{b}_-, \boldsymbol{ABx} \rangle.
\end{aligned}$$

From $\|\boldsymbol{A}\|_2 \leq 1$, the hermiticity of $\boldsymbol{A}$, $\langle \boldsymbol{b}_-, \boldsymbol{Bx} \rangle = 0$, (3.57) and (3.56), we deduce

$$\big|\langle \boldsymbol{b}_-, \boldsymbol{Ab}_- \rangle\big| \leq 1, \qquad \big|\langle \boldsymbol{b}_-, \boldsymbol{Ab}_+ \rangle\big| = \big|\langle \boldsymbol{b}_- + \boldsymbol{Ab}_-, \boldsymbol{b}_+ \rangle\big| \leq c_2 \eta,$$

$$\big|\langle \boldsymbol{b}_-, \boldsymbol{ABx} \rangle\big| = \big|\langle \boldsymbol{b}_- + \boldsymbol{Ab}_-, \boldsymbol{Bx} \rangle\big| \leq c_2 \eta (1 - \varepsilon).$$

Employing these estimates, $\|\boldsymbol{B}\|_2 \leq 1 - c_1\eta$ and (3.58), together with $|\gamma|^2 \leq \delta$, we obtain

$$(B.3) \qquad\qquad c_3\|\boldsymbol{p}\|_2 \geq |\alpha|c_1 - |\beta|c_2 - \sqrt{\delta}c_2(1 - \varepsilon)$$

after dividing through by $\eta > 0$. If $|\alpha|c_1 \geq c_2|\beta| + \sqrt{\delta}c_2(1 - \varepsilon) + \delta\varepsilon c_3/2$ then we obtain (B.2).

Therefore, it suffices to show (B.2) in the regime

$$(B.4) \qquad |\gamma|^2 \leq \delta, \qquad |\alpha|c_1 \leq c_2|\beta| + \sqrt{\delta}c_2(1 - \varepsilon) + \delta\varepsilon c_3/2.$$

For this regime, we use (B.1) and obtain

$$\begin{aligned}
\langle \boldsymbol{b}_+, \boldsymbol{p} \rangle &= \langle \boldsymbol{b}_+, (\mathbb{1} - \boldsymbol{AB})\boldsymbol{h} \rangle \\
(B.5) \qquad &= \beta\big(1 - \|\boldsymbol{B}\|_2 \langle \boldsymbol{b}_+, \boldsymbol{Ab}_+ \rangle\big) - \alpha\|\boldsymbol{B}\|_2 \langle \boldsymbol{b}_+, \boldsymbol{Ab}_- \rangle \\
&\quad - \gamma \langle \boldsymbol{b}_+, \boldsymbol{ABx} \rangle.
\end{aligned}$$

We employ (3.56), (3.57), the hermiticity of A and $\langle b_-, b_+ \rangle = 0$ to obtain

(B.6)
$$\langle b_+, A b_+ \rangle \leq 1 - \varepsilon, \qquad |\langle b_+, A B x \rangle| \leq 1 - \varepsilon,$$
$$|\langle b_+, A b_- \rangle| = |\langle b_+, b_- + A b_- \rangle| \leq c_2 \eta.$$

Applying (B.6) to (B.5) yields

(B.7)
$$\|p\|_2 \geq |\langle b_+, p \rangle|$$
$$\geq |\beta|\varepsilon - |\alpha|c_2\eta - |\gamma|(1 - \varepsilon)$$
$$\geq |\beta|\varepsilon - |\alpha|\frac{\varepsilon c_1}{2c_2} - \sqrt{\delta}(1 - \varepsilon),$$

where we used the assumption $\eta \leq \varepsilon c_1 / 2c_2^2$. Since $|\alpha|c_1/c_2 \leq |\beta| + O(\sqrt{\delta})$ from (B.4), we obtain that $\|p\|_2 \geq |\beta|\varepsilon/3$ for any $\delta \leq \delta_0(c_1, c_2, c_3, \varepsilon)$ sufficiently small. Furthermore, $|\alpha|^2 + |\beta|^2 \geq 1 - \delta$ and the fact that $|\beta|$ is large compared with $|\alpha|$ in the sense (B.4) guarantee that $|\beta|^2 \geq \frac{1}{3}[1 + (c_2/c_1)^2]^{-1}$, if δ is sufficiently small. In particular, $\|p\|_2 \geq \varepsilon\delta/2$ can be achieved with a small δ, that is, (B.2) holds true in the regime (B.4) as well. This concludes the proof of Lemma 3.7. □

LEMMA B.1.   (i) *Uniformly for $z \in \mathbb{D}_< \cup \mathbb{D}_>$ and $\eta > 0$, we have*

(B.8) $\|F\|_{2\to\infty} \lesssim 1, \qquad \|TF\|_{2\to\infty} \lesssim 1, \qquad \|FT\|_{2\to\infty} \lesssim 1.$

(ii) *If $w \notin \mathrm{Spec}(TF) \cup \{0\}$ and $\|(w\mathbb{1} - TF)^{-1}y\|_2 \lesssim \|y\|_2$ for some $y \in \mathbb{C}^{2n}$ then*

(B.9)
$$\left\|(w\mathbb{1} - TF)^{-1}y\right\|_\infty \lesssim \frac{1}{|w|}\|y\|_\infty.$$

A similar statement holds true for $(\bar{w}\mathbb{1} - FT)^{-1} = [(w\mathbb{1} - TF)^{-1}]^*$.

(iii) *For every $\eta_* > 0$, depending only on τ_* and the model parameters, such that*

(B.10)
$$\left\|(\mathbb{1} - TF)^{-1}Q\right\|_2 \lesssim 1,$$
$$1 - \|F\|_2 \gtrsim \eta,$$
$$\|f_- + Tf_-\|_2 \lesssim \eta,$$
$$\|f_-\|_\infty \lesssim 1$$

uniformly for all $\eta \in (0, \eta_*]$ and $z \in \mathbb{D}_<$, we have

(B.11)
$$\left\|((\mathbb{1} - TF)^{-1}Q)^*\right\|_\infty \lesssim 1$$

uniformly for $\eta \in (0, \eta_*]$ and $z \in \mathbb{D}_<$. Here, Q denotes the orthogonal projection onto the subspace $f_-^\perp$, that is, $Qy := y - \langle f_-, y \rangle f_-$ for every $y \in \mathbb{C}^{2n}$.

The estimate (B.9) is proved similarly as (5.28) in [2].

PROOF OF LEMMA B.1.   As $\|S_o\|_{2\to\infty} \lesssim 1$ by (2.1), we obtain from Proposition 3.2, and (3.26)

$$\|F\|_{2\to\infty} \le \|V^{-1}\|_\infty \|S_o\|_{2\to\infty} \|V^{-1}\|_2 = \left\|\frac{u\,v}{\widetilde{v}}\right\|_\infty \|S_o\|_{2\to\infty} \lesssim 1$$

uniformly for all $\eta > 0$ and $z \in \mathbb{D}_< \cup \mathbb{D}_>$. This proves the first estimate in (B.8). From Lemma 3.6(i), we conclude the second and the third estimate in (B.8).

We set $x := (w\mathbb{1} - T\,F)^{-1}y$. By assumption there is $C \sim 1$ such that

$$\|x\|_2 \le C\|y\|_2 \le C\|y\|_\infty.$$

Moreover, since $wx = T\,F\,x + y$ we obtain from the previous estimate

$$|w|\|x\|_\infty \le \|T\,F\,x\|_\infty + \|y\|_\infty \le (\|T\,F\|_{2\to\infty}C + 1)\|y\|_\infty.$$

Using the second estimate in (B.8), this concludes the proof of (B.9). The statement about $(\bar{w}\mathbb{1} - F\,T)^{-1}$ follows in the same way using the third estimate in (B.8) instead of the second.

For the proof of (B.11), we remark that the first condition in (B.10) implies that

(B.12) $$\left\|\left((\mathbb{1} - T\,F)^{-1}Q\right)^*\right\|_2 = \left\|(\mathbb{1} - T\,F)^{-1}Q\right\|_2 \lesssim 1.$$

The second assumption in (B.10) yields

(B.13) $$\left\|(\mathbb{1} - T\,F)^{-1}\right\|_2 \lesssim \eta^{-1}.$$

Take $y \in \mathbb{C}^{2n}$ arbitrary. We get

$$[T,\,Q]y = \langle T\,f_- + f_-,\,y\rangle f_- - \langle f_-,\,y\rangle(T\,f_- + f_-),$$

where $[T,\,Q] = T\,Q - Q\,T$ denotes the commutator of T and Q. Therefore,

(B.14) $$\left\|[T,\,Q]\right\|_2 \le 2\|f_- + T\,f_-\|_2 \lesssim \eta$$

by the third condition in (B.10). We set $x := Q(\mathbb{1} - F\,T)^{-1}y$ and compute

$$x = F\,T\,x + Q\,y - F[T,\,Q](\mathbb{1} - F\,T)^{-1}y,$$

where we commuted $\mathbb{1} - F\,T$ and Q and used that F and Q commute. Hence, using $\|x\|_2 \lesssim \|y\|_2 \lesssim \|y\|_\infty$ by (B.12), $\|Q\|_\infty \le 1 + \|f_-\|_\infty$, (B.14) and (B.13), we obtain

$$\|x\|_\infty \lesssim (\|F\,T\|_{2\to\infty} + 1 + \|f_-\|_\infty + \|F\|_{2\to\infty})\|y\|_\infty \lesssim \|y\|_\infty.$$

Here, we used the fourth assumption in (B.10) and (B.8). Notice that the η^{-1} factor from the trivial estimate (B.13) was compensated by the smallness of the commutator $[T,\,Q]$ which was a consequence of the third assumption in (B.10). Since $x = ((\mathbb{1} - T\,F)^{-1}Q)^*y$ this concludes the proof of (B.11). $\square$

PROOF OF LEMMA 3.5. We first prove that

(B.15) $$\|\boldsymbol{f}_- - \boldsymbol{a}\|_2 = O(\eta)$$

uniformly for $\eta \le 1$ and $\tau \in [0, 1 - \tau_*]$. To that end, we introduce the auxiliary operator:

$$\boldsymbol{A} := \|\boldsymbol{F}\|_2 \mathbb{1} + \boldsymbol{F}.$$

Therefore, we obtain from $\boldsymbol{F}\boldsymbol{f}_- = -\|\boldsymbol{F}\|_2 \boldsymbol{f}_-$ and (3.45)

$$\boldsymbol{A}\boldsymbol{f}_- = 0, \qquad \boldsymbol{A}\boldsymbol{a} = O(\eta).$$

Let $\boldsymbol{Q}$ be the orthogonal projection onto the subspace $\boldsymbol{f}_-^{\perp}$ orthogonal to $\boldsymbol{f}_-$, that is, $\boldsymbol{Q}\boldsymbol{y} := \boldsymbol{y} - \langle \boldsymbol{f}_-, \boldsymbol{y} \rangle \boldsymbol{f}_-$ for $\boldsymbol{y} \in \mathbb{C}^{2n}$. We then obtain $\boldsymbol{A}\boldsymbol{Q}\boldsymbol{a} = O(\eta)$ which implies $\boldsymbol{Q}\boldsymbol{a} = O(\eta)$ as $\boldsymbol{A}$ is invertible on $\boldsymbol{f}_-^{\perp}$ and $\|(\boldsymbol{A}|_{\boldsymbol{f}_-^{\perp}})^{-1}\|_2 \sim 1$ by (3.38). We infer (B.15).

For the proof of (3.46), we follow the proof of (B.11), replace $\boldsymbol{T}$ by $-\mathbb{1}$ and use Lemma 3.4(i) instead of the second and fourth condition in (B.10). □

REFERENCES

[1] AJANKI, O., ERDŐS, L. and KRÜGER, T. (2016). Universality for general Wigner-type matrices. *Probab. Theory Related Fields*. To appear. DOI:10.1007/s00440-016-0740-2.

[2] AJANKI, O., ERDŐS, L. and KRÜGER, T. (2015). Quadratic vector equations on complex upper half-plane. Available at arXiv:1506.05095v4.

[3] AJANKI, O., ERDŐS, L. and KRÜGER, T. (2016). Stability of the matrix Dyson equation and random matrices with correlations. Available at arXiv:1604.08188.

[4] AJANKI, O., ERDŐS, L. and KRÜGER, T. (2017). Singularities of solutions to quadratic vector equations on the complex upper half-plane. *Comm. Pure Appl. Math.* **70** 1672–1705. MR3684307

[5] ALJADEFF, J., RENFREW, D. and STERN, M. (2015). Eigenvalues of block structured asymmetric random matrices. *J. Math. Phys.* **56** 103502. MR3403052

[6] ALJADEFF, J., STERN, M. and SHARPEE, T. (2015). Transition to chaos in random networks with cell-type-specific connectivity. *Phys. Rev. Lett.* **114** 088101.

[7] ALT, J., ERDŐS, L. and KRÜGER, T. (2017). Local law for random Gram matrices. *Electron. J. Probab.* **22** paper no. 25, 41 pp. MR3622895

[8] ALT, J., ERDŐS, L. KRÜGER, T. and NEMISH, Y. (2017). Location of the spectrum of Kronecker random matrices. Available at arXiv:1706.08343.

[9] BAI, Z. D. (1997). Circular law. *Ann. Probab.* **25** 494–529. MR1428519

[10] BAO, Z., ERDŐS, L. and SCHNELLI, K. (2016). Local single ring theorem on optimal scale. Available at arXiv:1612.05920.

[11] BORDENAVE, C. and CHAFAÏ, D. (2012). Around the circular law. *Probab. Surv.* **9** 1–89. MR2908617

[12] BOURGADE, P., YAU, H.-T. and YIN, J. (2014). Local circular law for random matrices. *Probab. Theory Related Fields* **159** 545–595. MR3230002

[13] BOURGADE, P., YAU, H.-T. and YIN, J. (2014). The local circular law II: The edge case. *Probab. Theory Related Fields* **159** 619–660. MR3230004

[14] COOK, N., HACHEM, W., NAJIM, J. and RENFREW, D. (2016). Limiting spectral distribution for non-Hermitian random matrices with a variance profile. Available at arXiv:1612.04428.

[15] COOK, N. A. (2016). Lower bounds for the smallest singular value of structured random matrices. Available at arXiv:1608.07347v3.

[16] ERDŐS, L. and YAU, H.-T. (2012). Universality of local spectral statistics of random matrices. *Bull. Amer. Math. Soc.* (*N.S.*) **49** 377–414. MR2917064

[17] ERDŐS, L., YAU, H.-T. and YIN, J. (2011). Universality for generalized Wigner matrices with Bernoulli distribution. *J. Comb.* **2** 15–81. MR2847916

[18] ERDŐS, L., YAU, H.-T. and YIN, J. (2012). Bulk universality for generalized Wigner matrices. *Probab. Theory Related Fields* **154** 341–407. MR2981427

[19] FEY, A., VAN DER HOFSTAD, R. and KLOK, M. J. (2008). Large deviations for eigenvalues of sample covariance matrices, with applications to mobile communication systems. *Adv. in Appl. Probab.* **40** 1048–1071. MR2488532

[20] GIRKO, V. L. (1984). The circular law. *Teor. Veroyatn. Primen.* **29** 669–679. MR0773436

[21] GIRKO, V. L. (2001). *Theory of Stochastic Canonical Equations*: *Vol. I and II. Mathematics and Its Applications* **535**. Kluwer Academic, Dordrecht. MR1887676

[22] GÖTZE, F. and TIKHOMIROV, A. (2010). The circular law for random matrices. *Ann. Probab.* **38** 1444–1491. MR2663633

[23] GUIONNET, A., KRISHNAPUR, M. and ZEITOUNI, O. (2011). The single ring theorem. *Ann. of Math.* (2) **174** 1189–1217. MR2831116

[24] HELTON, J. W., RASHIDI FAR, R. and SPEICHER, R. (2007). Operator-valued semicircular elements: Solving a quadratic matrix equation with positivity constraints. *Int. Math. Res. Not. IMRN* **2007**. MR2376207

[25] MAY, R. M. (1972). Will a large complex system be stable? *Nature* **238** 413–414.

[26] PAN, G. and ZHOU, W. (2010). Circular law, extreme singular values and potential theory. *J. Multivariate Anal.* **101** 645–656. MR2575411

[27] RUDELSON, M. and VERSHYNIN, R. (2015). Delocalization of eigenvectors of random matrices with independent entries. *Duke Math. J.* **164** 2507–2538. MR3405592

[28] SOMPOLINSKY, H., CRISANTI, A. and SOMMERS, H.-J. (1988). Chaos in random neural networks. *Phys. Rev. Lett.* **61** 259–262. MR0949871

[29] TAO, T. and VU, V. (2008). Random matrices: The circular law. *Commun. Contemp. Math.* **10** 261–307. MR2409368

[30] TAO, T. and VU, V. (2010). Random matrices: Universality of ESDs and the circular law. *Ann. Probab.* **38** 2023–2065. MR2722794

[31] TAO, T. and VU, V. (2015). Random matrices: Universality of local spectral statistics of non-Hermitian matrices. *Ann. Probab.* **43** 782–874. MR3306005

[32] XI, H., YANG, F. and YIN, J. (2017). Local circular law for the product of a deterministic matrix with a random matrix. *Electron. J. Probab.* **22** paper no. 60, 77 pp.

[33] YIN, J. (2014). The local circular law III: General case. *Probab. Theory Related Fields* **160** 679–732. MR3278919

INSTITUTE OF SCIENCE AND TECHNOLOGY AUSTRIA
AM CAMPUS 1
A-3400 KLOSTERNEUBURG
AUSTRIA
E-MAIL: johannes.alt@ist.ac.at
          lerdos@ist.ac.at
          torben.krueger@ist.ac.at