RESEARCH ARTICLE

Wegner estimate and upper bound on the eigenvalue condition number of non-Hermitian random matrices

László Erdős | **Hong Chang Ji**

Institute of Science and Technology Austria, Klosterneuburg, Austria

Correspondence
Hong Chang Ji, Institute of Science and Technology Austria, Am Campus 1, 3400 Klosterneuburg, Austria.
Email: Hongchang.Ji@ist.ac.at

Abstract

We consider $N \times N$ non-Hermitian random matrices of the form $X + A$, where A is a general deterministic matrix and $\sqrt{N}X$ consists of independent entries with zero mean, unit variance, and bounded densities. For this ensemble, we prove (i) a Wegner estimate, that is, that the local density of eigenvalues is bounded by $N^{1+o(1)}$ and (ii) that the expected condition number of any bulk eigenvalue is bounded by $N^{1+o(1)}$; both results are optimal up to the factor $N^{o(1)}$. The latter result complements the very recent matching lower bound obtained by Cipolloni et al. and improves the N-dependence of the upper bounds by Banks et al. and Jain et al. Our main ingredient, a near-optimal lower tail estimate for the small singular values of $X + A - z$, is of independent interest.

Mathematics Subject Classification (MSC) 2020
60B20, 15A12, 15B52

1 | INTRODUCTION

1.1 | Setup

We consider large $N \times N$ random matrices with independent entries and without any symmetry constraint, so that they are typically not normal and have complex eigenvalues. More precisely,

our random matrix is of the form $X + A$, where A is a general deterministic "data" matrix and X is a random "noise matrix" consisting of independent complex or real entries with continuous distributions; the ensemble is often called *"Information-plus-Noise"* model, see for example [12, 21, 22, 26]. The main objects of this paper are the three fundamental spectral quantities of $X + A$ determining its stability properties, namely, eigenvalues, diagonal *eigenvector overlaps* (also known as *eigenvalue condition number*), and small singular values. We will typically use the scaling where both $\|A\|$ and $\|X\|$ are bounded, independently of N, however our results on the singular values will be uniform in A, hence they also cover the especially interesting small noise regime after a trivial rescaling.

We next explain our main motivations to study these quantities, along with their definitions. The fundamental difficulty in studying a non-normal matrix A, as opposed to a Hermitian one, is the instability of its spectrum; even a tiny perturbation to A may result in a very large change in its eigenvalues. To make this precise, we assume that A has a simple spectrum, and write its spectral decomposition

$$A = \sum_{i=1}^{N} \sigma_i \boldsymbol{r}_i \boldsymbol{l}_i^{\mathsf{T}}, \qquad A\boldsymbol{r}_i = \sigma_i \boldsymbol{r}_i, \qquad \boldsymbol{l}_i^{\mathsf{T}} A = \sigma_i \boldsymbol{l}_i^{\mathsf{T}}, \tag{1.1}$$

where $\sigma_i = \sigma_i(A)$'s are the eigenvalues of A and $\{\boldsymbol{l}_i = \boldsymbol{l}_i(A), \boldsymbol{r}_i = \boldsymbol{r}_i(A) : 1 \leq i \leq N\}$ is a bi-orthogonal family (so that $\boldsymbol{l}_i^{\mathsf{T}} \boldsymbol{r}_j = \delta_{ij}$) in $\mathbb{C}^N$ consisting of left and right eigenvectors of A. Under this normalization, the *diagonal eigenvector overlaps* are defined as the scale-invariant quantity

$$\mathcal{O}_{ii}(A) := \|\boldsymbol{l}_i(A)\|^2 \|\boldsymbol{r}_i(A)\|^2. \tag{1.2}$$

With this definition, we have the well-known variational identity (see e.g. [8])

$$\sqrt{\mathcal{O}_{ii}(A)} = \lim_{t \to 0} \sup \left\{ \left| \frac{\sigma_i(A + tE) - \sigma_i(A)}{t} \right| : E \in \mathbb{C}^{N \times N}, \|E\| = 1 \right\}. \tag{1.3}$$

This formula shows that the diagonal overlap quantifies the instability of σ_i against the worst perturbation, therefore $\sqrt{\mathcal{O}_{ii}}$ is also called the *eigenvalue condition number*. However, it is practically impossible to control $\mathcal{O}_{ii}(A)$ for a general deterministic A: One can easily construct a bi-orthogonal family with arbitrarily large $\|\boldsymbol{l}_1\|^2 \|\boldsymbol{r}_1\|^2$, and thus construct a matrix A with arbitrarily large overlap $\mathcal{O}_{11}(A)$ but bounded $\sigma_1(A)$.

Besides quantifying the instability σ_i, another interesting feature of the overlap $\mathcal{O}_{ii}(A)$ is that it connects σ_i to the shifted singular values of A. This follows from another standard variational identity;

$$\sqrt{\mathcal{O}_{ii}(A)} = \lim_{z \to \sigma_i} \frac{|\sigma_i(A) - z|}{\lambda_1(A - z)}, \tag{1.4}$$

where $\lambda_1(A - z) \leq \cdots \leq \lambda_N(A - z)$ stand for the singular values of $A - z$ for $z \in \mathbb{C}$. Note that singular values, as (square roots of) the eigenvalues of the *Hermitian* matrix $(A - z)(A - z)^*$, are much more stable under perturbations. Thus (1.4) shows that the overlap exactly quantifies the magnification of the instability of σ_i in terms of $\lambda_1(A - z)$ for z near σ_i.

While the overlap is an uncontrolled object in the worst case, numerical algorithms, whose computational cost or accuracy should deteriorate with $\mathcal{O}_{ii}$, still perform well in practice; see, for example [43] and references therein concerning Gaussian elimination. This is often explained by a mechanism called *smoothed analysis* capitalizing on the fact that even for the worst A, the instability of its spectrum is typically regularized by a noise, that is, a random perturbation, for example in [6–8, 35, 43]. This motivates the investigation of three quantities $\sigma_i(A + X)$, $\mathcal{O}_{ii}(A + X)$, and $\lambda_k(X + A)$ for small k, where A is a general deterministic matrix and X is a random matrix with independent entries.

Interestingly, overlaps also appear in a completely unrelated context, namely, in the Dyson-type stochastic eigenvalue dynamics. Consider the matrix Brownian motion $A_t = A + B_t$ where the entries of $B_t \in \mathbb{C}^{N \times N}$ are independent standard complex Brownian motions and A has a simple spectrum. The eigenvalues $\sigma_i(A_t)$ of A_t are martingales with brackets given by

$$\mathrm{d}\langle \sigma_i(A_t), \overline{\sigma_j(A_t)} \rangle_t = \boldsymbol{l}_j(A_t)^* \boldsymbol{l}_i(A_t) \boldsymbol{r}_j(A_t)^* \boldsymbol{r}_i(A_t) \mathrm{d}t, \qquad \langle \sigma_i(A_t), \sigma_j(A_t) \rangle_t = 0; \qquad (1.5)$$

see [15, Appendix A]. Taking $i = j$, the diagonal overlap $\mathcal{O}_{ii}(A_t)$ is exactly the time derivative of the quadratic variation of $\sigma_i(A_t)$. In particular, (1.5) shows that the dynamics of $\sigma_i(A_t)$ involves eigenvectors, hence is not autonomous in contrast to the standard Hermitian Dyson Brownian motion [23].

Finally, we point out yet another context of our results. The noise matrix is rescaled by a factor $1/\sqrt{N}$ to keep its norm of order one as N grows, that is, the entries of $\sqrt{N}X$ remain on constant scale. With this scaling, the celebrated *circular law* [5, 31, 46] states that if the entries of $\sqrt{N}X$ are centered i.i.d. random variables with unit variance, then $\sigma_i(X)$'s are asymptotically uniformly distributed on the unit disk with density of order N (we consider unnormalized densities throughout the paper unless stated otherwise). Extension of the circular law for $X + A$ was proved in [46, Corollary 1.17] as well if A has a limiting $*$-distribution a, in which case the limit is the Brown measure of free sum $x + a$ where x is a free circular element. Recently this Brown measure was studied in detail in [53], where it was proved that there is an open set $\mathcal{B} \subset \mathbb{C}$ so that this measure is supported on $\overline{\mathcal{B}}$ and has a strictly positive real analytic density in $\mathcal{B}$. Consequently, the number of eigenvalues $\sigma_i(X + A)$ in a complex domain $\mathcal{D} \subset \mathcal{B}$ is typically comparable to $N|\mathcal{D}|$ where $|\mathcal{D}|$ is the area of $\mathcal{D}$; see Remark 2.2 for details. We refer to [53, Section 1.1] for a historical exposition on the eigenvalue density of $X + A$.

In the special case of the *Ginibre ensemble*, that is, when the entries of X are i.i.d. Gaussian and $A = 0$, the density, or one-point function, of the eigenvalues can be computed explicitly [30] and it is essentially given by $\rho_N(z) = \frac{N}{\pi}$ for $|z| < 1$. For general entry distribution of X, the *local circular law* [16] (see also [3, 4]), asserts that $\sum_i f(\sigma_i)$ is well approximated by $\frac{N}{\pi} \int f$ as long as f lives on a scale much larger than $N^{-1/2}$, the typical eigenvalue spacing. Thus the density of eigenvalues is well understood on *mesoscopic* scales, but it is a highly non-trivial statement that the density remains absolutely continuous and even constant on arbitrary small scales (in particular this obviously requires that the distribution of the matrix elements has a continuous component). Motivated by the theory of random Schrödinger operators, in the Hermitian setup an upper bound on the eigenvalue density is called *Wegner estimate* [51]. Its optimal form has been established in [27, 36] for a large class of Hermitian random matrices (Wigner matrices). Prior to the current work, no Wegner-type result below the mesoscopic scale $N^{-1/2+\varepsilon}$ has been known for non-Hermitian i.i.d. random matrices besides the Ginibre case and its elliptic generalizations

where explicit formulas using planar orthogonal polynomials and contour integral methods are available, see the recent comprehensive paper [2] and references therein. We finally remark that optimal upper bounds on the local k-point correlation functions (in particular Wegner estimates for $k = 1$) for two-dimensional Coulomb gases with general potential were established in [48] using non-explicit conditioning methods.

Eigenvalues and eigenvector overlaps of $X + A$ are genuinely *non-Hermitian* objects which are hard to access directly. Instead, one typically works with the singular values and singular vectors of $X + A$, which are accessible via *Hermitian* methods, however, the relation between them is subtle. On the level of the spectrum, Girko's formula (see 1.23 below) provides an explicit connection, but it requires to control the lower tail of the small singular values of $X + A - z$ for all $z \in \mathbb{C}$ simultaneously. For eigenvectors the situation is much more delicate; there is no direct formula relating eigenvectors of $X + A$ and singular vectors of $X + A - z$ apart from the special case when z is exactly an eigenvalue, a conditioning that is technically very hard to handle. Our main results on eigenvector overlaps and Wegner estimate for eigenvalues both heavily rely on very accurate lower tail estimates on the lowest singular values of $X + A - z$ and we develop a new method to estimate them. We remark that singular values and singular vectors of the Information-plus-Noise model $X + A$ have also been extensively studied in statistics, but with a strong focus on the large singular values. Our current main interest is the opposite regime, so we refrain from reviewing the extensive statistics literature on the subject.

1.2 | Our results

The standing assumption on X for all our results is that the entries of $\sqrt{N}X$ have continuous distributions with bounded densities. This guarantees the main mechanism behind the key Wegner estimates. For certain results, we may additionally assume that the entries of $\sqrt{N}X$ are centered, have variance one and have all moments finite. The matrix A can be completely general, for some results we only assume that A has bounded norm. Before formally stating our main theorems with precise conditions in Section 2 we informally summarize their content and compare them with previous results.

The first main result, given precisely in Theorem 2.3 later, concerns an upper bound for the averaged density of the eigenvalues σ_i's. Specifically, we prove that

$$\mathbb{E}\#\{i \,:\, \sigma_i(X + A) \in \mathcal{D}\} \lesssim \begin{cases} N^{o(1)}(N|\mathcal{D}|)^{1-o(1)} & \text{for complex } X, \\ N^{o(1)}\dfrac{(N|\mathcal{D}|)^{1-o(1)}}{\min_{z \in \mathcal{D}} \lambda_1(\mathfrak{I}[A - z])} & \text{for real } X, \end{cases} \tag{1.6}$$

for any ball $\mathcal{D}$ *of arbitrarily small size* in the bulk spectrum (see 2.2 for its definition), where $\mathfrak{I}[A - z]$ is the matrix with entrywise imaginary part of $A - z$. This result is a weak *Wegner estimate*: Up to the $o(1)$-powers it almost shows the absolute continuity and boundedness of the normalized eigenvalue density. Note that in the real case the estimate necessarily deteriorates near the real axis; this is because real random matrices tend to have many real eigenvalues (e.g., the real Ginibre matrices have approximately $\sqrt{N}$ real eigenvalues [25]) and once eigenvalues concentrate on a one-dimensional submanifold, their *two-dimensional* density naturally has a singular component.

The second main results (Theorem 2.7–2.9 later) concern an upper bound on the diagonal overlap $\mathcal{O}_{ii}$ in expectation sense. Since in the Ginibre case it is known $\mathcal{O}_{ii}$ has a fat tail [15], hence controlling second or higher moments is impossible. In Theorem 2.7 we prove that

$$\mathbb{E}\left[\mathbb{1}_{\Xi_\mathcal{D}} \sum_{i:\sigma_i(X+A)\in\mathcal{D}} \mathcal{O}_{ii}(X+A)\right] \lesssim \begin{cases} |\log N|^2 \cdot N(N|\mathcal{D}|) & \text{for complex } X, \\ |\log N|^2 \cdot N \dfrac{N|\mathcal{D}|}{\min_{z\in\mathcal{D}} \lambda_1(\mathfrak{I}[A-z])} & \text{for real } X, \end{cases} \tag{1.7}$$

where $\mathcal{D} \subset \mathbb{C}$ can be any polynomially small (in N) ball and $\Xi_\mathcal{D}$ is a very high-probability event. In Theorem 2.9, under somewhat stronger assumptions, we remove the exceptional set, that is, the factor of $\Xi_\mathcal{D}$ from (1.7). Theorem 2.9 also covers overlaps for *real eigenvalues* when both X and A are real, and shows that

$$\mathbb{E}\left[\sum_{i:\sigma_i(X+A)\in\mathcal{I}} \sqrt{\mathcal{O}_{ii}(X+A)}\right] \lesssim N^\delta \cdot \sqrt{N}\left(\sqrt{N}|\mathcal{I}|\right), \tag{1.8}$$

where $\mathcal{I} \subset \mathbb{R}$ can be any interval in the bulk spectrum. Up to expectation and the $\log N$ (or N^δ) factors, both of (1.7) and (1.8) roughly imply $\mathcal{O}_{ii} = O(N)$: The first factors N and $\sqrt{N}$ indicate the sizes of $\mathcal{O}_{ii}$ and $\sqrt{\mathcal{O}_{ii}}$, while $(N|\mathcal{D}|)$ and $\left(\sqrt{N}|\mathcal{I}|\right)$ stand for the expected number of eigenvalues in the given domains. Our estimates also naturally provide an upper bound on the *off-diagonal* overlaps $\mathcal{O}_{ij} := (l_i, l_j)(r_i, r_j)$ as well by a trivial Schwarz inequality, $|\mathcal{O}_{ij}| \le (\mathcal{O}_{ii}\mathcal{O}_{jj})^{1/2}$.

Finally, in Theorems 2.10–2.13, we prove almost optimal lower tail bounds on the singular values of $X + A$. To be precise, in Theorem 2.10 we prove that, *uniformly over all $s > 0$*,

$$\mathbb{P}[N\lambda_k(X+A) \le s] \lesssim \begin{cases} s^{2k^2}(|\log s| + \log N)^k & \text{for complex } X, \text{ fixed } k \ge 1, \\ s^{k^2}(|\log s| + \log N)^k & \text{for real } X, \text{ fixed } k \ge 2, \end{cases} \tag{1.9}$$

where $\lambda_k(X+A)$ denotes the k-th smallest singular value of $X + A$. In Theorem 2.11 we remove the $\log s$ factor to exhibit the optimal s-dependence in (1.9) assuming more restrictive conditions on X, and extend the result for real X to $k = 1$ when A is also real. The exponents $2k^2$ and k^2 exactly express the correct strength of *level repulsion* among singular values. Finally, in Theorem 2.13, we prove that for *real X*, a genuinely *complex shift A* can improve (1.9) for $k = 1$ in the sense that

$$\mathbb{P}[N\lambda_1(X+A) \le s] \lesssim s^2\left(1 + \frac{|\log s| + \log N}{\lambda_1(\mathfrak{I}[A])}\right) \tag{1.10}$$

uniformly over all $s > 0$. The second power of s shows that the tail of the lowest singular value of $X + A$ follows the behavior of the *complex* ensemble even though X is *real*. As a consequence, for a genuinely complex data matrix A, the same regularization effect can be achieved with a real noise matrix as with a complex one.

In the next sections we collect previously known results on σ_i, $\mathcal{O}_{ii}$, and λ_k, and explain how our results (1.6)–(1.10) improve upon them. The situation is somewhat different for general A matrix and for the important special case when $A = -z$ scalar matrix for some $z \in \mathbb{C}$. Most previous results have only been established for the real or complex Ginibre ensemble, that is, when

the entries of $\sqrt{N}X$ are i.i.d. real or complex standard Gaussians. Results valid for general i.i.d. matrices will be collected separately.

1.3 | Previous results I: X = Ginibre, A = scalar

Since adding a scalar matrix A affects $\sigma_i(X + A)$ and $\mathcal{O}_{ii}(X + A)$ merely as a shift, for these quantities we may assume $A = 0$.

The density of eigenvalues σ_i was explicitly computed for complex and real Ginibre ensembles respectively in [30] and [24]. In particular, results therein imply that uniformly over $z \in \mathbb{C}$

$$\rho_{\mathbb{C}}(z) \sim \mathbb{1}(|z| \le 1) + o(1), \qquad \rho_{\mathbb{C}\setminus\mathbb{R}}(z) \sim \mathbb{1}(|z| \le 1) \min\left(\sqrt{N}|\operatorname{Im} z|, 1\right) + o(1), \qquad (1.11)$$

where $\rho_{\mathbb{C}}$ and $\rho_{\mathbb{C}\setminus\mathbb{R}}$ denote the normalized densities of all and non-real eigenvalues of complex and real Ginibre ensembles, respectively. Taking a domain $\mathcal{D}$ away from the real axis, our result (1.6) generalizes the upper bound (1.11) to an almost boundedness of $\rho_{\mathbb{C}}$ and $\rho_{\mathbb{C}\setminus\mathbb{R}}$ even for general (bounded) matrix A and general i.i.d. distribution for X. See Remark 2.6 for more detailed comparison in the real case. For the real eigenvalues of the real Ginibre ensemble X we mention that [25, Corollaries 4.5 and 5.2] proved

$$E_N := |\{i : \sigma_i(X) \in \mathbb{R}\}| \sim \sqrt{N}, \qquad \rho_{\mathbb{R}}(x) := \frac{1}{E_N} \frac{\mathrm{d}}{\mathrm{d}x} \mathbb{E}|\{i : \sigma_i(X) \le x\}| \sim \mathbb{1}(x \in [-1, 1]) + o(1).$$
$$(1.12)$$

While (1.12) is not directly related to (1.6), it shows that the factor of $\left(\sqrt{N}|\mathcal{I}|\right)$ in (1.8) indeed accurately describes the number of real eigenvalues $\sigma_i(X + A)$ in a real interval $\mathcal{I}$.

Likewise, the overlap $\mathcal{O}_{ii}$ has also been studied thoroughly for Ginibre ensembles. In [50] it was proved for the complex Ginibre ensemble that, uniformly over $|z| < 1$,

$$\mathbb{E}[\mathcal{O}_{ii}(X)|\sigma_i = z] = N(1 - |z|^2) + o(1), \qquad (1.13)$$

with a sub-exponential error in N, where the expectation is conditioned on the event that $\sigma_i = z$. Furthermore, it was recently proved in [15, 29] that $\mathcal{O}_{ii}$ has a heavy tail even after proper rescaling: Uniformly over $|z| < 1$ and conditionally on the event $\sigma_i = z$,

$$\frac{\mathcal{O}_{ii}(X)}{N(1 - |z|^2)} \implies \begin{cases} \gamma_2^{-1} & \text{if } X \text{ is complex Ginibre ensemble [15, Theorem 1.1],} \\ \gamma_1^{-1} & \text{if } X \text{ is real Ginibre ensemble and } z \in \mathbb{R} \text{ [29, Theorem 2.1],} \end{cases} \qquad (1.14)$$

where γ_1 and γ_2 are gamma distributed random variables with parameters 1 and 2. This implies that $\mathcal{O}_{ii}$ has a heavy tail without first and second moment in the real and complex case, respectively. In particular, the probabilistic L^1-control in (1.7) is practically the strongest possible form of an upper bound for $\mathcal{O}_{ii}$, and the same applies to (1.8). We also mention that for real Ginibre ensemble X and $|\mathcal{D}| \gtrsim 1/N$, the following nearly optimal upper bound corresponding to the second

estimate of (1.7) was proved in [18, Theorem 2.5]:

$$\mathbb{E}\mathbb{1}_{\Xi_D} \sum_{\sigma_i(X) \in D} \mathcal{O}_{ii}(X) \lesssim (\log N) \cdot N(N|\mathcal{D}|) \tag{1.15}$$

for a high-probability event Ξ_D. Comparing (1.13)–(1.15) with (1.7)–(1.8), we find that our results are optimal modulo the factors of $\log N$ (or N^δ) and the denominator in the second estimate of (1.7). The main novelty is that we are not restricted to Ginibre ensembles and we allow a general matrix A.

Next, we recall previous results on the singular values when X is real or complex Ginibre ensemble. For singular values a scalar shift $A = -z$ does matter. For the special $A = 0$ case, the following optimal result was proved in [44]:

$$\mathbb{P}[N\lambda_k(X) \le s] \sim \begin{cases} s^{2k^2} & \text{if } X \text{ is complex Ginibre,} \\ s^{k^2} & \text{if } X \text{ is real Ginibre.} \end{cases} \tag{1.16}$$

The exponents reflect the strong level repulsion among the first k singular values indicating an underlying Vandermonde determinant structure for the joint density function.

For the more general $A = -z \ne 0$ case; when X is complex Ginibre one can explicitly compute $\mathbb{P}[\lambda_k(X - z) \le s]$ for any fixed k, in fact the singular values form a determinantal point process [10] and thus the analogue of (1.16) for the complex case holds. No explicit formula is available when X is a real Ginibre ensemble but via supersymmetric methods [18] proved an almost optimal counterpart of (1.10) for the lowest singular value. See Remarks 2.12 and 2.14 for more details on $\lambda_k(X + A)$ when X is Ginibre and $A = -z$.

1.4 | Previous results II: $X = $ Ginibre, A general

As we already mentioned, Wegner-type estimates for σ_i's such as (1.6) have not been considered for general A before.

As far as $\mathcal{O}_{ii}$ and λ_k for general A are concerned, all upper bounds on $\mathcal{O}_{ii}$ listed below are obtained using (variants of) the inequalities

$$\pi\mathbb{E} \sum_{i:\sigma_i \in D} \mathcal{O}_{ii}(X + A) \le N^2 \liminf_{s \to 0} \int_D \frac{\mathbb{P}[N\lambda_1(X + A - z) \le s]}{s^2} d^2z,$$

$$2\mathbb{E} \sum_{i:\sigma_i \in \mathcal{I}} \sqrt{\mathcal{O}_{ii}(X + A)} \le N \liminf_{s \to 0} \int_{\mathcal{I}} \frac{\mathbb{P}[N\lambda_1(X + A - x) \le s]}{s} dx, \tag{1.17}$$

proved respectively in [15, Section 3.6] and [8, Lemma 6.2] for the condition numbers of the complex and real eigenvalues. Notice that it is essential to have a tail bound on λ_1 *uniformly* for all small s; in particular the prominent tail bounds on $N\lambda_1$ that are valid only down to some scale $s \gtrsim N^{-a}$ with some exponent $a > 0$, for example [14, Section 4.4], [45, Theorem 2.1], [20, Theorem 1.18], or bounds with a tiny additive factor (to account for the possible discrete distribution of X which we exclude in our setup), for example [41, Theorem 1.2] would not be useful in (1.17).

The inequalities (1.17) translate estimates such as (1.9) and (1.10) into upper bounds for $\mathcal{O}_{ii}$, and in particular one can simply replace A by $A - z$ and plug these estimates into (1.17) if not for the

logarithmic corrections. In light of (1.17), in what follows, we list previous results on $\lambda_1(A + X)$ together with their direct consequences for $\mathcal{O}_{ii}$, even if the resulting upper bounds for $\mathcal{O}_{ii}$ were not formulated explicitly.

The first result concerning $\mathcal{O}_{ii}(X + A)$ for general A appeared in [7], where the authors proved for real and complex Ginibre ensemble that

$$\mathbb{P}[N\lambda_1(X + A) \leq s] \leq s^2, \quad \mathbb{E} \sum_{i:\sigma_i(X+A)\in\mathcal{D}} \mathcal{O}_{ii}(X + A) \leq \frac{N}{\pi}(N|\mathcal{D}|), \quad \text{for complex } X, \quad (1.18)$$

$$\mathbb{P}[N\lambda_1(X + A) \leq s] \leq s, \quad \mathbb{E} \sum_{i:\sigma_i(X+A)\in\mathcal{I}} \sqrt{\mathcal{O}_{ii}(X + A)} \leq \frac{\sqrt{N}}{2}\left(\sqrt{N}|\mathcal{I}|\right), \quad \text{for real } X \text{ and } A. \tag{1.19}$$

A weaker version of (1.19) was proved earlier in [43] with an additional factor of 2.35. Comparing these estimates with (1.13)–(1.15) shows that (1.18)–(1.19) are optimal.

1.5 | Previous results III: X, A general

Beyond Gaussian ensembles, (1.18) and (1.19) have been generalized to any X such that the entries of $\sqrt{N}X$ have bounded densities, as in the current paper. Namely, for such general X and A, it is known that

$$\mathbb{P}[N\lambda_1(X + A) \leq s] \lesssim Ns^2, \quad \mathbb{E} \sum_{i:\sigma_i(X+A)\in\mathcal{D}} \mathcal{O}_{ii}(X + A) \lesssim N^2(N|\mathcal{D}|), \quad \text{for complex } X, \quad (1.20)$$

$$\mathbb{P}[N\lambda_1(X + A) \leq s] \lesssim \sqrt{N}s, \quad \mathbb{E} \sum_{i:\sigma_i(X+A)\in\mathcal{I}} \sqrt{\mathcal{O}_{ii}(X + A)} \lesssim N\left(\sqrt{N}|\mathcal{I}|\right), \quad \text{for real } X \text{ and } A, \tag{1.21}$$

where the first result (1.20) was essentially[1] proved in [35, Lemma 3.4] and the second result (1.21) is due to [49, Corollary 1.4]. Prior to [49], a similar result in the real case appeared in [37, Corollary 1.8], which had weaker bound but covered λ_k beyond $k = 1$. In the same setting with general real X, the overlap $\mathcal{O}_{ii}$ at a genuinely complex eigenvalue σ_i was considered simultaneously in [35] and [8] improving (1.21) away from the real axis. Specifically, [35, Proposition 4.2] proved for real X and A and $z \in \mathbb{C}$ that

$$\mathbb{P}[N\lambda_1(X + A - z) \leq s] \lesssim \frac{N^{3/2}s^2}{|\operatorname{Im} z|}, \quad \mathbb{E} \sum_{i:\sigma_i(X+A)\in\mathcal{D}} \mathcal{O}_{ii}(X + A) \lesssim \frac{N^{5/2}(N|\mathcal{D}|)}{\min_{z\in\mathcal{D}} |\operatorname{Im} z|}. \tag{1.22}$$

A similar result was obtained in [8, Theorem 1.5] with larger powers of N on the right-hand sides, but this result also covers λ_k for $k > 1$. Notice the additional powers of N in (1.20)–(1.22) compared to the optimal results (1.18)–(1.19) in the Ginibre case. These are largely due to the geometric methods used in [35], namely, "invertibility via distance" introduced in [49]. In contrast, in (1.7)–(1.8) we obtain upper bounds on the overlaps and in (1.9)–(1.10) for the singular values with a nearly optimal N-dependence using purely analytic methods.

---

[1] Estimates therein carry an additional factor of N compared to (1.20), due to an apparent typo in the proof of [35, Lemma 3.4]; the factor n^2 in the rightmost side of the last inequality in [35, p.3018] should be n.

Finally, we remark that a matching *lower* bound for $\mathcal{O}_{ii}(X + A)$ with general A and i.i.d. X was recently obtained in [19, Theorem 2.4], which is optimal up to a factor of $N^{o(1)}$. Lower bounds require very different methods than upper bounds and they are much more robust, in particular $\mathcal{O}_{ii}(X + A) \geq N^{1-o(1)}$ can be proven with very high probability, while the overlap has a heavy upper tail. The main method of [19] is a very precise multi-resolvent local law on the Hermitization of $X + A$ closely related to the *Eigenstate Thermalization Hypothesis* and *Quantum Unique Ergodicity* for Hermitian random matrices. No regularity on the density of X was necessary in [19].

Summarizing, the current paper substantially generalizes many previous results on upper bounds on overlaps and lower tail bounds on singular values for matrices of the form $X + A$. We can deal with very general distributions of X and general matrices A without sacrificing much of the optimality.

The key new approach is a Wegner type estimate for the singular values; an idea that has been exploited earlier in the Hermitian (Wigner) setup [27] but has never been used for non-Hermitian problems. In the main body of this paper we develop the necessary tools to establish a Wegner estimate for the Hermitization of $X + A$, then we derive all our estimates from this input. Before entering the precise details, in the next subsection we sketch the main ideas, especially we indicate how Hermitian Wegner type estimates for singular values are used to estimate non-Hermitian eigenvalue density and overlaps.

1.6 | Outline of the proofs

The common ingredient for our main non-Hermitian results (1.6)–(1.8) is the lower tail bound (1.9) on the first few singular values $\lambda_k(X + A)$ or $\lambda_k(X + A - z)$. First we explain how we use this bound for our main results and then we comment on its proof.

For the proof of (1.6) (Theorem 2.3), we use Girko's Hermitization formula [31, 46]

$$\sum_{i=1}^{N} f(\sigma_i(X + A)) = -\frac{1}{4\pi} \int_{\mathbb{C}} \Delta f(z) \int_0^{\infty} \mathrm{Tr} \frac{\eta}{(X + A - z)(X + A - z)^* + \eta^2} \mathrm{d}\eta \mathrm{d}^2 z. \quad (1.23)$$

Girko's formula translates the eigenvalue statistics of $X + A$ into singular value statistics of $X + A - z$, and has been proved to be an effective tool for studying the spectrum of non-Hermitian random matrices; see for example [3, 5, 17, 31–33, 47]. To control the left-hand side of (1.6) we take f in (1.23) to be a smoothed indicator $\mathbb{1}_D$. To estimate the right hand side of (1.23) from above, we first perform two integrations by parts with respect to z. Then, for an upper bound, besides the lower bounds on the singular values of $X + A - z$, we also need *upper* bounds on certain scalar products of their singular vectors uniformly in z – this information is provided by the thermalization result [19, Theorem 2.2]. We remark that exactly the same information was used in [19, Theorem 2.4] to prove the high probability lower bounds for $\mathcal{O}_{ii}$, so morally an *upper* bound on the r.h.s. of (1.23) requires *lower* bound on the singular values and *lower* bound on the overlaps $\mathcal{O}_{ii}$.

As for the *upper* bounds on the overlap (1.7)–(1.8) (Theorem 2.7) we relate the smallest singular value $\lambda_1(X + A - z)$ to the overlap $\mathcal{O}_{ii}(X + A)$ via the contour integral

$$-\frac{1}{2\pi \mathrm{i}} \oint_{|z-w|=r} \frac{1}{X + A - w} \mathrm{d}w = \sum_{i:|\sigma_i(X+A)-z|<r} \boldsymbol{r}_i(X + A)\boldsymbol{l}_i(X + A)^*. \quad (1.24)$$

Using Weyl's inequality between products of eigenvalues and singular values and the estimates for the *second* smallest singular value $\lambda_2(X + A - z)$ we can take $r \sim N^{-C}$ so small that there is at most one σ_i within distance r from z. For such an r, the overlap $\mathcal{O}_{ii}$ is exactly the squared norm of the right-hand side of (1.24), thus an upper bound on it can be obtained by the square of the left-hand side, which in turn can be bounded from above via a lower tail bound on $\lambda_1(X + A - z)$. We need to tile the spectrum into disjoint tiny boxes of linear size r and carefully control the error terms to obtain a linear bound in the area $|\mathcal{D}|$ after summing them up.

Now we explain the proof of the optimal lower tail estimates (1.9)–(1.10) for the lowest singular values $\lambda_k(X + A)$. These are proven in Theorems 2.10–2.13 using Wegner type arguments. The key point is that the regularity of the randomness in X forces the singular values to *spread*; they cannot stick to a particular value (like zero) as X is sampled from a continuous probability space. While this mechanism is quite transparent, the essential difficulty is to get the optimal N-powers in the estimate. Playing with the randomness of a *single* matrix element x_{ij} would not have a sufficiently strong effect on λ_1 since x_{ij} has size only $N^{-1/2}$, while playing with *all* matrix elements simultaneously would not be manageable due to the correlations of their effects. We operate with the randomness of a few rows of X at the same time; this already has the desired effect on λ_1 and is still technically feasible.

Part of these arguments is inspired by [27] that proved Wegner estimates for eigenvalues of Hermitian random matrices. The lower tail of $\lambda_1(X + A)$ is estimated by the trace of the resolvent of $(X + A)(X + A)^*$ outside its spectrum at a negative spectral parameter $-\eta^2$ with $\eta \lesssim 1/N$. Then we express the i-th diagonal entry of the resolvent as a quadratic form of the i-th row vector of $(X + A)$ with a random weight involving the minor of $(X + A)$ with the i-th row deleted. We work in the probability space of this row vector which is independent of the weight and prove effective anti-concentration bounds. Higher singular values $\lambda_k(X + A)$ require an inductive procedure to remove k rows. Since the actual proofs are rather technical, we present a more detailed outline in Section 5. See also [1, 28] for other instances of smoothed analysis in the Hermitian context.

While our proofs of Theorems 2.10–2.13 start with a similar idea as [27], the proofs in [27] actually assumed much stronger conditions as they were studying the entire bulk spectrum. This resulted in a non-vanishing Hermitian part of the resolvent whose control required an additional integration by parts in the probability space forcing much stronger regularity assumptions. In our case, the resolvent is always skew-Hermitian (see 3.22) since we work at the *hard edge* of $(X + A)(X + A)^*$. We also point out that [27] assumed certain decay for the Fourier transforms of entry-distribution, which was later weakened in [36, Appendix A] by using Brascamp–Lieb inequality. Instead, here we use a more refined input from [42] whose proof contains the same arguments as in [36], eventually requiring only a minimal regularity condition (bounded density) for the entry-distribution.

As a final remark, we point out that the improved bound (1.10) for the least singular value $\lambda_1(X + A)$ for real X and complex A is more difficult to handle than that for complex X, while their lower tails have the same s^2-decay by (1.9) and (1.10). Along the proof of Theorem 2.13, we find that the singular vectors of $(X + A)$ affect the lower tail of the singular values; we need to show that singular vectors are genuinely complex if A is complex. Hence the difference between real and complex *non-Hermitian X* is fundamentally harder to capture than that between real symmetric and complex *Hermitian* random matrices which do not involve eigenvectors.

1.7 | Organization

In Section 2, we rigorously define our model and state the main results. Sections 3 and 4 are devoted to the proofs of Theorems 2.3 and 2.7 concerning eigenvalues and overlaps of $A + X$, respectively. We prove the singular value estimates in the remaining sections; Sections 5–7 contain proofs of Theorems 2.11, 2.10, and 2.13, in that order.

1.8 | Notations

For $x, y \in \mathbb{R}$, we denote $[\![x, y]\!] = [x, y] \cap \mathbb{Z}$ and $[\![x]\!] = [1, x] \cap \mathbb{Z}$. For a square matrix $A \in \mathbb{C}^{d \times d}$, we write $\langle A \rangle = \frac{1}{d} \operatorname{Tr} A$ for its normalized trace and denote its Hermitian, skew-Hermitian, real, and imaginary parts by

$$\operatorname{Re}[A] = \frac{A + A^*}{2}, \quad \operatorname{Im}[A] = \frac{A - A^*}{2i}, \quad \mathfrak{R}[A] = \frac{A + \overline{A}}{2}, \quad \mathfrak{I}[A] = \frac{A - \overline{A}}{2i}. \tag{1.25}$$

For a matrix B of any size we denote by $\|B\|$ its operator norm. For each $i \in \mathbb{N}$, we denote $\boldsymbol{e}_i$ by the i-th coordinate vector, whose dimension can vary by lines.

For $p \in \mathbb{N}$ and integrable functions f and g respectively on $\mathbb{R}^p$ and $\mathbb{C}^p$, we denote their Lebesgue integrals by

$$\int_{\mathbb{R}^p} f(\boldsymbol{x}) \mathrm{d}^p \boldsymbol{x}, \qquad \int_{\mathbb{C}^p} g(\boldsymbol{z}) \mathrm{d}^{2p} \boldsymbol{z}.$$

For a random variable x and $q \geq 1$, we write $\|x\|_q$ to denote $(\mathbb{E}|x|^q)^{1/q}$.

We denote $\mathbb{C}_+ := \{z \in \mathbb{C} : \operatorname{Im} z > 0\}$ and $\mathbb{D} := \{|z| < 1\} \subset \mathbb{C}$. For a Borel set $\mathcal{D} \subset \mathbb{C}$, we define $|\mathcal{D}|$ to be its Lebesgue measure. The letter N always denotes the dimension of our matrix, we consider the large N regime and we use the standard asymptotic relation $\lesssim$ and $\sim$ with respect to N. For example, for non-negative functions f and g of N we write $f \lesssim g$ when there is a constant $C > 0$ such that $f(N) \leq Cg(N)$ for all N, and write $f \sim g$ when $f \lesssim g$ and $g \lesssim f$ are both true. We use the letters c and C to denote positive constants whose exact value may vary from line to line.

2 | MAIN RESULTS

We consider matrices over the real or complex field, denoted commonly by $\mathbb{K} = \mathbb{R}$ or $\mathbb{C}$.

Definition 2.1. An $(N \times N)$ real ($\mathbb{K} = \mathbb{R}$) or complex ($\mathbb{K} = \mathbb{C}$) random matrix X is called *regular* if the following hold true for a constant $\mathfrak{b} > 0$:

(i)   The collection $\{\operatorname{Re} X_{ij}, \operatorname{Im} X_{ij} : i, j \in [\![N]\!]\}$ is independent:

(ii.$\mathbb{R}$)   When $\mathbb{K} = \mathbb{R}$, the random variables $\sqrt{N} X_{ij}$ have densities bounded by $\mathfrak{b}$:

(ii.$\mathbb{C}$)   When $\mathbb{K} = \mathbb{C}$, the random variables $\sqrt{N} \operatorname{Re} X_{ij}$ and $\sqrt{N} \operatorname{Im} X_{ij}$ have densities bounded by $\mathfrak{b}$.

A regular matrix X is called a *regular i.i.d. matrix* if the following hold in addition to (i)–(ii).

(iii) The entries are identically distributed, that is, $X_{11} \overset{d}{=} X_{ij}$ for all $i, j \in [\![N]\!]$:

(iv) $\mathbb{E}X = 0$ and $\mathbb{E}XX^* = I$: If $\mathbb{K} = \mathbb{C}$ we also assume $\mathbb{E}XX^\mathsf{T} = 0$:

(v) For all $p \in \mathbb{N}$, $\mathfrak{m}_p := \mathbb{E}|\sqrt{N}X_{11}|^p$ is finite. In this case we define $\mathbf{m} := (\mathfrak{m}_p)_{p \in \mathbb{N}} \in \mathbb{R}^\mathbb{N}$.

Note that we do not impose any assumption on the decay of the entry distribution of a *regular* matrix, but only on that of a *regular i.i.d.* matrix. For example, if X_{ij}'s are i.i.d. Cauchy-distributed random variables, that is,

$$\frac{\mathrm{d}}{\mathrm{d}x}\mathbb{P}\big[\sqrt{N}X_{11} \le x\big] = \frac{1}{\pi}\frac{1}{x^2 + 1},$$

then X is a regular matrix but not a regular i.i.d. matrix. Also note that the shifted matrix $X + A$ remains regular with the same $\mathfrak{b} > 0$ for any $A \in \mathbb{K}^{N \times N}$ when X is regular.

For a given deterministic matrix $A \in \mathbb{C}^{N \times N}$, we denote the (complex) eigenvalues of $X + A$ by $\{\sigma_i \equiv \sigma_i(X + A) : i \in [\![N]\!]\}$ and their normalized empirical distribution by

$$\rho \equiv \rho_{X+A}^{(N)} := \frac{1}{N}\sum_{i=1}^{N}\delta_{\sigma_i(X+A)}. \tag{2.1}$$

The eigenvalues of $X + A$ do not have a canonical ordering, but we still consider them indexed by $[\![N]\!]$ to simplify notations. Nonetheless all our arguments are insensitive to this ad hoc ordering. Finally, for each $z \in \mathbb{C}$ and $r > 0$, we define the number of eigenvalues in a ball of radius r about z;

$$\mathcal{N}_{z,r} := |\{i \in [\![N]\!] : |\sigma_i - z| \le r\}|.$$

Our results on eigenvalues and overlaps are often restricted to the indices i for which σ_i is well inside the limiting spectrum of $X + A$. We quantify the "bulk" of the non-Hermitian spectrum of $(X + A)$ in terms of the singular value spectrum of $(X + A - z)$. Namely, for $\tau > 0$ we define

$$\mathcal{B}_\tau := \left\{z \in \mathbb{C} : \frac{\mathrm{d}}{\mathrm{d}x}(\mu_{\mathrm{sc}} \boxplus \mu_{|A-z|}^{\mathrm{symm}})(0) > \tau\right\} \subset \mathbb{C}, \tag{2.2}$$

where μ_{sc} is the usual semi-circle distribution, $\mu_{|A-z|}^{\mathrm{symm}}$ is the symmetrization of the singular value distribution of $A - z$, and $\boxplus$ denotes the free additive convolution. Using the standard defining equation for the Stieltjes transform of $\mu_{\mathrm{sc}} \boxplus \mu_{|A-z|}^{\mathrm{symm}}$ (see, e.g., [38] or [19, Eq. (2.4)]), $\mathcal{B}_\tau$ may be written in terms of the singular value distribution of $A - z$ as (recall from Section 1.8 that $\langle \cdot \rangle$ is the normalized trace)

$$\mathcal{B}_\tau = \left\{z \in \mathbb{C} : \left\langle\frac{1}{(A - z)(A - z)^* + \tau^2}\right\rangle > 1\right\}. \tag{2.3}$$

Note that $\mu_{\mathrm{sc}} \boxplus \mu_{|A-z|}^{\mathrm{symm}}$ is exactly the symmetrization of the limiting singular value distribution of $(X + A - z)$, see [11, 34]. Thus $z \in \mathcal{B}_\tau$ for some fixed $\tau > 0$ guarantees that z is well inside the limiting spectrum of $(X + A)$. We will make this relation more precise in the following Remark 2.2 that may be skipped at the first reading as it is not used in the rest of the paper.

Remark 2.2 (Characterization of the bulk). For technical convenience we defined the "bulk" spectrum of $X + A$ in terms of $\mathcal{B}_\tau$. Here we show that the domains $\mathcal{B}_\tau$ for small τ indeed coincide with the traditional concept of "bulk spectrum" of ρ in the following sense:

(i) The sets $\mathcal{B}_\tau$ roughly cover the support of ρ in the limit $\tau \to 0$;

(ii) for any disk $\mathcal{D} \subset \mathcal{B}_\tau$ on macroscopic scale ($|\mathcal{D}| \sim 1$) we have

$$|\{i \in [\![N]\!] : \sigma_i \in \mathcal{D}\}| \sim N|\mathcal{D}|. \tag{2.4}$$

To make these two statements a bit more precise, consider the Brown measure ρ_{a+x} of the sum $a + x$ in a non-commutative probability space where $a \equiv a_N$ has the same $*$-distribution as A and x is a circular element $*$-free from a. By a standard corollary of the proof of [19, Theorem 2.6][2], we can prove that if A is norm-bounded, then the empirical eigenvalue distribution $\rho = \rho^{(N)}$ from (2.1) is close to ρ_{a+x} in a weak sense, namely, for any fixed, smooth, compactly supported test function f on $\mathbb{C}$,

$$\int_\mathbb{C} f(z)\mathrm{d}(\rho^{(N)} - \rho_{a+x})(z) \to 0 \qquad \text{as } N \to \infty \tag{2.5}$$

with high probability. On the other hand, define the open set $\mathcal{B}_0 \subset \mathbb{C}$ by

$$\mathcal{B}_0 := \left\{ z \in \mathbb{C} : \left\langle \frac{1}{(A-z)(A-z)^*} \right\rangle > 1 \right\} = \bigcup_{\tau > 0} \mathcal{B}_\tau, \tag{2.6}$$

where the last equality is due to (2.3). For this open set [53, Theorem B] implies that the measure ρ_{a+x} has no atom, is supported on $\overline{\mathcal{B}_0}$, and is absolutely continuous on $\mathcal{B}_0$ with a strictly positive real analytic density therein. Together with (2.5) this implies (i), that is, that $\mathrm{supp}(\rho_{a+x}) = \overline{\mathcal{B}_0}$ covers the bulk of ρ, that is, the regime where ρ is typically positive.

Furthermore, [53, Eq. (4.2)] (first proved for normal A in [13, Theorem 1.4]) gives an implicit formula for the density of ρ_{a+x}, which implies the following quantitative estimate on $\mathcal{B}_\tau$:

$$\frac{\tau^2}{(\|A-z\|^2 + \tau^2)^2} \leq \pi \frac{\mathrm{d}\rho_{a+x}}{\mathrm{d}m}(z) \leq \|A-z\| \left(\frac{\|A-z\|^2 + \tau^2}{\tau^2} \right)^2 + \frac{1}{\tau^2}, \qquad \forall z \in \mathcal{B}_\tau, \tag{2.7}$$

where m denotes the Lebesgue measure on $\mathbb{C}$. Taking f in (2.5) to be a smoothed indicator function $\mathbb{1}_D$ and combining with (2.7) yields the second statement (ii) and (2.4).

2.1 | Eigenvalues and eigenvector overlaps of $X + A$

In this section, we consider $X + A$ where A is a general deterministic matrix with $\|A\| = O(1)$ and X is a regular i.i.d. matrix.

The first result is our Wegner-type estimate for the eigenvalues in the bulk: it asserts that the expected density of σ_i's, or equivalently, the one point correlation function, is almost bounded in the bulk.

---

[2] This result gives the optimal bulk local law for the Hermitization of $X + A$, see also (3.4). If we restrict $w \in \mathrm{i}(\eta, \infty)$ in [19, Theorem 2.6] with a small enough but fixed η, their result easily extends to all norm-bounded A, beyond those with 0 in the bulk spectrum of $|X + A - z|$. From this local law, a standard argument using Girko's formula (1.23) implies (2.5), the corresponding analogue of the macroscopic "circular law". The standard additional ingredient on the tail of the lowest singular value is given as usual by the regularity condition on X.

Theorem 2.3. *Fix $\gamma \in (0,1)$, $\mathfrak{K} > 0$, and (small) $\delta, \tau > 0$. Let X be a real or complex regular i.i.d. matrix, and $A \in \mathbb{C}^{N \times N}$ be deterministic with $\|A\| \leq \mathfrak{K}$. Then there exists a constant $C \equiv C(\gamma, \delta, \tau, \mathfrak{b}, \mathfrak{m}, \mathfrak{K}) > 0$ such that the following hold for all $N \in \mathbb{N}$:*

(i) If X is complex, then for all $r \in [0, N^{-1/2}]$ and $z \in \mathcal{B}_\tau$ we have

$$\mathbf{P}\big[\mathcal{N}_{z,r} \geq 1\big] \leq \mathbb{E}\mathcal{N}_{z,r} \leq C N^\delta (Nr^2)^{1-\gamma}. \tag{2.8}$$

(ii) If X is real, then for all $r \in [0, N^{-1/2}]$, $z \in \mathcal{B}_\tau$, and $\lambda_1(\mathfrak{I}[A - z]) > 2r$ we have

$$\mathbf{P}\big[\mathcal{N}_{z,r} \geq 1\big] \leq \mathbb{E}\mathcal{N}_{z,r} \leq C N^\delta \frac{(Nr^2)^{1-\gamma}}{\lambda_1(\mathfrak{I}[A - z])}, \tag{2.9}$$

where we recall that $\lambda_1(\mathfrak{I}[B])$ denotes the smallest singular value of $\mathfrak{I}B = (B - \overline{B})/2i$ for $B \in \mathbb{C}^{N \times N}$.

Note that the quotient $\mathbb{E}\mathcal{N}_{z,r}/(N\pi r^2)$ tends to the averaged density of states in the limit $r \to 0$, hence Theorem 2.3 proves the Wegner estimate up to an exponent $\gamma > 0$. As a corollary, we show that $\langle (X + A - z)^{-1} \rangle$, the normalized trace of the resolvent, is essentially bounded in (probabilistic) $L^{2-o(1)}$. This is a truly random effect since z lies in the support of the limiting density of states, hence $\langle (X + A - z)^{-1} \rangle$ may be unbounded, but its $(2 - o(1))$-th moment is finite.

Corollary 2.4. *Fix $\mathfrak{K}, \tau > 0$, $\delta_1 \in (0,1]$ and let $\delta > 0$ be sufficiently small. Let X be a real or complex regular i.i.d. matrix and $A \in \mathbb{C}^{N \times N}$ with $\|A\| \leq \mathfrak{K}$.*

(i) If X is complex, then there exists a constant $C \equiv C(\delta_1, \tau, \mathfrak{b}, \mathfrak{m}, \mathfrak{K}) > 0$ such that the following hold for all $z \in \mathcal{B}_\tau$ and $N \in \mathbb{N}$:

$$\mathbb{E}|\langle (X + A - z)^{-1} \rangle|^{2-\delta_1} \leq C. \tag{2.10}$$

(ii) If X is real, then there exists a constant $C \equiv C(\delta_1, \delta, \tau, \mathfrak{b}, \mathfrak{m}, \mathfrak{K}) > 0$ such that the following hold for all $z \in \mathcal{B}_\tau$ and $N \in \mathbb{N}$:

$$\mathbb{E}|\langle (X + A - z)^{-1} \rangle|^{2-\delta_1} \leq C\left(1 + \frac{N^{-\delta_1/2+\delta}}{y}\left(1 \wedge (Ny)^{\delta_1/2-\delta}\right)\right) \tag{2.11}$$

where we abbreviated $y := \lambda_1(\mathfrak{I}[A - z])$.

A similar result as (2.10) appeared in [40, Lemma 2.2], with an extra factor of N^2 on the right-hand side.

Notice that we take $\delta_1 \in (0,1]$ in Corollary 2.4, instead of $(0,2]$. While we use this choice in the proof (see 3.35), Corollary 2.4 trivially implies an upper bound for $\mathbb{E}|\langle X + A - z \rangle|^{2-\delta_1}$ even when $\delta_1 \in (1,2]$ by Hölder's inequality. In particular, (2.10) is true for all $\delta_1 \in (0,2]$, and the same holds true in the real case if $\lambda_1(\mathfrak{I}[A - z])$ is of order one.

Remark 2.5. In light of (1.11) for the Ginibre ensemble, we believe that Theorem 2.3 remains true for $\delta = 0 = \gamma$ even in the general case. In our current Theorem 2.3 the factor N^δ is due to the error in the local laws for the singular values of $X + A - z$, and the exponent $\gamma > 0$ is the cost for neglecting the correlation between singular values and vectors. We defer the details to Remark 3.5

after the proof of Theorem 2.3. Also note that $\delta_1 > 0$ in Corollary 2.4 is necessary since the contribution from a single eigenvalue to the second moment of $\langle (X + A - z)^{-1} \rangle$, that is, $N^{-2} \mathbb{E} |\sigma_i - z|^{-2}$, is already logarithmically divergent.

Remark 2.6. When $A = 0$ and X is real Ginibre ensemble, (1.11) and (1.12) imply that $\mathbb{E} \mathcal{N}_{z,r}$ remains bounded as long as $|\operatorname{Im} z| \gtrsim N^{-1/2}$. Taking $A = 0$ in our results, (2.9) and (2.11), they do not feature this sharp $|\operatorname{Im} z|$-dependence due to the factor $|\operatorname{Im} z|^{-1}$ in (2.9). This factor originates from Theorem 2.13, where we prove a lower tail estimate for the smallest singular value of $(X + A - z)$ that also carries the same $|\operatorname{Im} z|^{-1}$ factor. In Remark 2.14 following Theorem 2.13, we explain more details on its source and a possibly optimal $|\operatorname{Im} z|$-dependence.

Recall that when the spectrum of $(X + A)$ is simple we have the spectral decomposition

$$X + A = \sum_i \sigma_i \boldsymbol{r}_i \boldsymbol{l}_i^*, \qquad \boldsymbol{l}_i^* \boldsymbol{r}_j = \delta_{ij} \tag{2.12}$$

and that the overlaps between eigenvectors $\boldsymbol{l}_i$ and $\boldsymbol{r}_j$ are defined as

$$\mathcal{O}_{ij} := \boldsymbol{l}_i^* \boldsymbol{l}_j \boldsymbol{r}_j^* \boldsymbol{r}_i. \tag{2.13}$$

Before presenting our results on the diagonal overlaps $\mathcal{O}_{ii}$, we briefly pause to show that $X + A$ has simple spectrum with probability 1 when X is regular. To see this, notice that $X + A$ has a repeated eigenvalue if and only if the characteristic polynomial and its derivative has a common root. This is in turn equivalent to the resultant of these two polynomials being identically zero. Thus $X + A$ is not simple when its entries satisfy a polynomial equation, so that X is in an algebraic submanifold of $\mathbb{K}^{N \times N}$ with positive codimension. Since this manifold has Lebesgue measure zero and X has a density, X falls into this set with probability zero.

In the next two theorems, we prove upper bounds for the expectation of $\mathcal{O}_{ii}$. The first result holds anywhere in the spectrum but only on a set with very high probability. In the second result we remove this exceptional set in the bulk.

Theorem 2.7. *Fix $\mathfrak{K}_0 > 0$ and (large) $D, K > 0$. Let X be a real or complex regular i.i.d. matrix and $A \in \mathbb{C}^{N \times N}$ with $\|A\| \le N^{\mathfrak{K}_0}$. Then there exists a constant $C \equiv C(\mathfrak{b}, \mathfrak{m}, \mathfrak{K}_0, D, K) > 0$ such that the following hold true:*

(i) If X is complex, for any square $\mathcal{D} \subset \mathbb{C}$ with $|\mathcal{D}| \ge N^{-2K}$ we have

$$\mathbb{E} \mathbb{1}_{\Xi_\mathcal{D}} \sum_{i : \sigma_i \in \mathcal{D}} \mathcal{O}_{ii} \le CN (\log N)^2 (N|\mathcal{D}|), \tag{2.14}$$

for some event $\Xi_\mathcal{D}$ with $\mathbb{P}[\Xi_\mathcal{D}^c] \le N^{-D}$.

(ii) If X is real, for any square $\mathcal{D} \subset \mathbb{C}$ with $|\mathcal{D}| \ge N^{-2K}$ and $\min_{z \in \mathcal{D}} \lambda_1(\mathfrak{I}[A - z]) \ge N^{-K}$ we have

$$\mathbb{E} \mathbb{1}_{\Xi_\mathcal{D}} \sum_{i : \sigma_i \in \mathcal{D}} \mathcal{O}_{ii} \le CN (\log N)^2 \frac{(1 + \|A\|^2)(N|\mathcal{D}|)}{\min_{z \in \mathcal{D}} \lambda_1(\mathfrak{I}[A - z])}, \tag{2.15}$$

for some event $\Xi_\mathcal{D}$ with $\mathbb{P}[\Xi_\mathcal{D}^c] \le N^{-D}$.

In particular, if X is complex, for each fixed $0 < \epsilon' < \epsilon$ and $K > 0$ we have

$$\mathbb{P}\left[\sum_{i:\sigma_i\in D}\mathcal{O}_{ii}\geq N^{1+\epsilon}N|\mathcal{D}|\right]\leq N^{-\epsilon'} \tag{2.16}$$

uniformly over squares $\mathcal{D}\subset\mathbb{C}$ with $|\mathcal{D}|>N^{-2K}$. The same holds true for real X for all $\mathcal{D}\subset\mathbb{C}$ with $|\mathcal{D}|>N^{-2K}$ as long as $\min_{z\in D}\lambda_1(\mathfrak{J}[A-z])\sim 1$.

Remark 2.8. Notice that in Theorem 2.7 we only assume $\|A\|$ is at most polynomially large in N, allowing A to be much larger than X. A closely related canonical model is $\gamma A + X$, with $\|A\| = O(1)$, where $\gamma > 0$ is a large control parameter. In fact, controlling the effect of large γ on the condition number $\mathcal{O}_{ii}$ has been a central task in quantitative linear algebra, for example, in [7, 8, 35] after the trivial rescaling $\gamma A + X = \gamma(A + \gamma^{-1}X)$ and viewing $\gamma^{-1}X$ as a small random perturbation of the deterministic A. Since the constant $C > 0$ in Theorem 2.7 depends only on $\log\|A\|$ via $\mathfrak{K}_0$, our result can be used for such a purpose. In fact, if we replace A with γA in Theorem 2.7 (i) and (ii), we recover the same γ-dependence for $\mathcal{O}_{ii}$ as in respectively [7, Theorem 1.5] and [8, Proposition 6.4] modulo $\log\gamma$ factors.

We also remark that our proof of Theorem 2.7 easily extends the result to regular matrices beyond *regular i.i.d.* matrices (i.e., without conditions (iii)–(iv) in Definition 2.1). In this case we need the additional assumptions that $\mathcal{D}$ is contained in a bounded set and, if X is real, that $\mathbb{E}\|X\|^4 = O(1)$.

Since (2.14) and (2.15) are true for all fixed $K > 0$ and additive in $\mathcal{D}$, we may easily deduce the same result for many other domains that can be approximately tiled by small squares. Indeed, in Appendix A we prove that Theorem 2.7 holds true not only for square domains but also for any Borel measurable domain $\mathcal{D}$ satisfying a natural geometric regularity condition of the form $|\mathcal{D} + [-N^{-K}, N^{-K}]^2| \leq C|\mathcal{D}|$ for some constants $C, K > 0$. Here $+$ refers to the Minkowski sum of $\mathcal{D}$ and a small square. In particular, we may take $\mathcal{D}$ to be a disk or a "polynomially thin" rectangle.

In the next theorem, we remove the probabilistic factor $\mathbb{1}_{\Xi_D}$ from (2.14) assuming that $\mathcal{D}$ is in the bulk $\mathcal{B}_\tau$, and extend the result to $\mathcal{O}_{ii}$'s at real eigenvalues when A and X are both real.

Theorem 2.9. *Fix $\mathfrak{K} > 0$ and (small) $\delta, \tau > 0$. Let X be a real or complex regular i.i.d. matrix and $A \in \mathbb{C}^{N\times N}$ with $\|A\| \leq \mathfrak{K}$. Then there exists a constant $C \equiv C(\delta, \tau, \mathfrak{b}, \mathfrak{m}, \mathfrak{K})$ such that the following hold:*

(i) If X is complex, for any Borel set $\mathcal{D} \subset \mathcal{B}_\tau$ we have

$$\mathbb{E}\sum_{i:\sigma_i\in D}\mathcal{O}_{ii}\leq CN^{1+\delta}(N|\mathcal{D}|). \tag{2.17}$$

(ii) If X and A are real, for any Borel set $\mathcal{I} \subset \mathcal{B}_\tau \cap \mathbb{R}$ we have

$$\mathbb{E}\sum_{i:\sigma_i\in I}\sqrt{\mathcal{O}_{ii}}\leq CN^{1/2+\delta}\left(\sqrt{N}|\mathcal{I}|\right), \tag{2.18}$$

where $|\mathcal{I}|$ is the Lebesgue measure of $\mathcal{I}$ in $\mathbb{R}$.

The most relevant choice for $\mathcal{D}$ in both of Theorems 2.7 and 2.9 is a square of area $N^{-1+\epsilon}$ in the bulk, so that typically there is at least one eigenvalue in $\mathcal{D}$ with high probability. More precisely, since the local law in [19, Theorem 2.6] covers all mesoscopic scales, we can extend the convergence in (2.5) to test functions f on mesoscopic scales supported in $\mathcal{B}_\tau$; see for example [3, Section 5.2] for a completely analogous proof in a slightly different setup. In particular (2.4) extends to squares $\mathcal{D} \subset \mathcal{B}_\tau$ with $|\mathcal{D}| = N^{-1+\epsilon}$, so that there is at least one eigenvalue in $\mathcal{D}$ with high probability. For such sets $\mathcal{D}$, Theorem 2.9 and (1.13) are on the same footing; our result shows that the expectation of $\mathcal{O}_{ii}$ is bounded by $N^{1+\epsilon}$.

In an unrelated context, we remark that Theorem 2.9(i) can also be used to give an upper bound on the diffusivity of the complex Dyson-type eigenvalue dynamics defined via (1.5). This matches the analogous lower bound given in [19, Eq. (2.14)].

2.2 | Small singular values of shifted regular matrices

As mentioned in the introduction, our proofs of Theorems 2.3–2.9 translate the problems to understanding the singular values of $X + A - z$ for any shift parameter $z \in \mathbb{C}$. In this section we present results on the singular values of $X + A$, where X is a general regular matrix in Definition 2.1 and $A \in \mathbb{C}^{N\times N}$ is a deterministic matrix. By replacing A with $A - z$, we use the results in this section as inputs for those in Section 2.1. These results are of independent interest, so we list them in this section separately.

We stress that all results in this section except Theorem 2.11 will be uniform in A, in particular no norm bound on A is assumed unlike in the previous Section 2.1. This means that by a simple rescaling $\gamma X + A = \gamma(X + A/\gamma), \gamma > 0$, these results also hold for the singular values of the matrix $\gamma X + A$ with a rescaled noise, as often presented in numerical applications, for example [7, 8, 35].

We denote the ordered singular values of $X + A$ by

$$0 \leq \lambda_1 \leq \cdots \leq \lambda_N. \tag{2.19}$$

As eigenvalues of a Hermitian random matrix, λ_k's are subject to *level repulsion* that in turn forces an especially small lower tail probability for them. The next theorems contain such results. Their optimality will be discussed afterwards.

Theorem 2.10 (Regular matrices). *Let $k \in \mathbb{N}$ be fixed, $A \in \mathbb{C}^{N\times N}$, and X be a regular matrix. Then for the singular values (2.19) of $(X + A)$ we have the following:*

(i) *[Complex case] If X is complex, there exists a constant $C \equiv C(k, \mathfrak{b}) > 0$ such that for all $N \geq k \vee 2$ and $s \in [0, 1]$*

$$\mathbb{P}[N\lambda_k \leq s] \leq C(|\log s| + \log N)^k s^{2k^2}. \tag{2.20}$$

(ii) *[Real case] If X is real, there exists a constant $C \equiv C(k, \mathfrak{b}) > 0$ such that for all $2 \leq k \leq N$ and $s \in [0, 1]$*

$$\mathbb{P}[N\lambda_k \leq s] \leq C(|\log s| + \log N)^k s^{k^2}. \tag{2.21}$$

Notice that the result for real X, as formulated in (2.21), requires $k \geq 2$. The origin of this restriction will be explained in Section 6, along the proof of Proposition 6.1. Nevertheless, we still have results that are optimal in s for $k = 1$ and real X, see the following Theorems 2.11 and 2.13, respectively for real A and genuinely complex A.

Note that the constant C in Theorem 2.10 does not depend on A. Also note that the only assumption on X is the regularity, and in particular some moments[3] of the entries of X may be infinite. Although Theorem 2.10 has such robustness, it is really relevant when the (non-Hermitian) spectrum of $X + A$ indeed reaches zero, otherwise typically even the lowest singular value λ_1 is far away from zero and its $1/N$ scaling indicated by (2.20)–(2.21) is completely off. In particular, recalling the definition of $\mathcal{B}_\tau$ in (2.2), this is the case for $X + A - z$ if X is a regular i.i.d. matrix and $z \in \mathcal{B}_\tau$. To be precise, denoting the singular values of $X + A - z$ by

$$0 \leq \lambda_1^z \leq \cdots \leq \lambda_N^z, \tag{2.22}$$

we remove the $|\log s|$ factors from Theorem 2.10 under this setting.

Theorem 2.11 (Regular i.i.d. matrices). *Fix $k \in \mathbb{N}$ and $\delta, \tau, \mathfrak{K} > 0$. Let X be a* regular i.i.d. *matrix and $A \in \mathbb{C}^{N \times N}$ with $\|A\| \leq \mathfrak{K}$. Then we have the following for the singular values (2.22) of $X + A - z$.*

(i) *[Complex case] If X is complex, there exist constants $C_1 \equiv C_1(k, \mathfrak{b})$ and $C_2 \equiv C_2(k, \tau, \delta, \mathfrak{m}, \mathfrak{K})$ such that for all $N \geq 3k$, $s \in [0, 1]$, and $z \in \mathcal{B}_\tau$ we have*

$$\mathbb{P}[N\lambda_k^z \leq s] \leq C_1 s^{2k^2} \mathbb{E}\left[(1 + N\lambda_{3k}^z)^{2k^2}\right] \leq C_1 C_2 N^\delta s^{2k^2}. \tag{2.23}$$

(ii) *[Real case] If X and A are both real, there exist constants $C_1 \equiv C_1(k, \mathfrak{b})$ and $C_2 \equiv C_2(k, \tau, \delta, \mathfrak{m}, \mathfrak{K})$ such that for all $N \geq 3k$, $s \in [0, 1]$, and $z \in \mathcal{B}_\tau \cap \mathbb{R}$ we have*

$$\mathbb{P}[N\lambda_k^z \leq s] \leq C_1 s^{k^2} \mathbb{E}\left[(1 + N\lambda_{3k}^z)^{k^2}\right] \leq C_1 C_2 N^\delta s^{k^2}. \tag{2.24}$$

As easily seen from (2.23) and (2.24), the only suboptimal factor N^δ comes from the estimate for λ_{3k}^z. This is due to local law (see Lemma 3.1 for its statement), which carries a small power of N. Note that for real X, Theorem 2.11 (ii) applies also to $k = 1$, but requires A to be real.

Remark 2.12. For complex Gaussian X and general A, λ_k's are exactly the square roots of eigenvalues of deformed Laguerre unitary ensemble, whose joint density was computed in [10, Proposition 3.1]. In particular one can easily deduce that the joint density of unordered eigenvalues $(x_1, \ldots, x_N)$ of $(X - z)(X - z)^*$ is proportional (ignoring N-factors) to

$$V(\boldsymbol{x}) \exp\left(-N^2 |z|^2 - N \sum_{i \in [\![N]\!]} x_i\right) \det\left(x_j^{i-1} F^{(i-1)}(N^2 |z|^2 x_j)\right)_{i,j \in [\![N]\!]}, \tag{2.25}$$

---

[3] Still one should think of the second moment of $\sqrt{N} X_{ij}$ to be finite, for otherwise $\lambda_k \sim N^{-1}$ might not be true; see [9] for the case when $h_{ij}(x) \sim 1/|x|^{1+\alpha}$ with $\alpha \in (0, 2)$. Our result still holds for such matrices, but our $N\lambda_k$ scaling may not be adequate.

where $V(\boldsymbol{x}) := \prod_{i<j}(x_i - x_j)$ is the Vandermonde determinant, and F is a real analytic function directly related to the zeroth modified Bessel function I_0. After some algebra, one can easily see that the second determinant in (2.25) is $O(V(\boldsymbol{x}))$ up to an N-dependent factor. Thus we obtain that (ignoring N-factors)

$$\mathbb{P}[\lambda_k(X - z) \le s] = \mathbb{P}[|\{x_j \,:\, x_j \le s^2\}| \ge k] \lesssim (s^2)^k \cdot (s^2)^{k(k-1)} = s^{2k^2}, \tag{2.26}$$

where the first factor comes from the volume of $[0, s^2]^k$ and the second from $V(\boldsymbol{x})^2$. In particular (2.25) as well as (1.16) show that Theorem 2.11 is optimal for general k as far as the s-power is concerned in the small s regime.

In the next result, we show that if the entrywise imaginary part $\Im A$ of the shift matrix A is non-singular, Theorem 2.10 (ii) extends to $k = 1$ with quadratic decay s^2; in terms of Theorem 2.11, the linear s-dependence in (2.24) for $k = 1$ can be improved to s^2.

Theorem 2.13 (Real case, complex shift, improved). *Let X be a* real *regular matrix, $A \in \mathbb{C}^{N \times N}$, and λ_1 be the smallest singular value of $(X + A)$. Then there exists a constant $C \equiv C(\mathfrak{b}) > 0$ such that the following holds for all $N \ge 4$ and $s \in [0, 1]$;*

$$\mathbb{P}[N\lambda_1 \le s] \le Cs^2\left(1 + \frac{(\mathbb{E}\|X\|^4)^{1/2} + \|A\|^2}{\lambda_1(\Im A)}(|\log s| + \log N)\right). \tag{2.27}$$

Consequently, if X is a real regular i.i.d. *matrix, A is real, and $\mathfrak{K} > 0$, then there exists a constant $C \equiv C(\mathfrak{b}, \mathfrak{m}, \mathfrak{K})$ such that the following holds whenever $\|A\| \le \mathfrak{K}$, $z \in \mathbb{C}$, $N \ge 4$, and $s \in [0, 1]$;*

$$\mathbb{P}[N\lambda_1^z \le s] \le Cs^2\left(\frac{1 + |z|^2}{|\operatorname{Im} z|}(|\log s| + \log N)\right). \tag{2.28}$$

Remark 2.14. In particular when $|\operatorname{Im} z| \sim 1$, (2.28) shows that $\mathbb{P}[\lambda_1^z \le s] = O(s^2)$ up to a logarithmic factor, similarly to the complex case. The same improvement due to (genuinely) complex shift was previously proved for real Ginibre ensemble X and $A = 0$ in [18, Theorem 2.1]; for z in the bulk, the result therein reads

$$\mathbb{P}[N\lambda_1^z \le s] \lesssim (1 + |\log s|)s^2 + e^{-(\sqrt{N}|\operatorname{Im} z|)^2}\left(s \wedge \frac{s^2}{\sqrt{N}|\operatorname{Im} z|}\right). \tag{2.29}$$

From (2.29) it is clear that the critical scale of $|\operatorname{Im} z|$ is $N^{-1/2}$, and that the right-hand side is roughly of the same size as s^2 or s depending on whether $|\operatorname{Im} z|$ exceeds that scale or not. Our result fails to capture this scale of transition due to the suboptimality of Lemma 7.1, where we estimate to what extent are the singular vectors of $(X - z)$ genuinely complex when z is. We refer to Remark 7.3 for more details.

3 | WEGNER-TYPE ESTIMATE FOR X: PROOF OF THEOREM 2.3

Throughout the rest of the paper, we fix parameters τ, $\mathfrak{b}$, $\mathfrak{m}$, and $\mathfrak{K}$; since they affect Theorems 2.3–2.13 only implicitly via the constant C, considering them to be fixed does no harm to the proof.

For each $A \in \mathbb{C}^{N \times N}$ and $z \in \mathbb{C}$, we define $H_z \equiv H_{z,A} \in \mathbb{C}^{2N \times 2N} \cong \mathbb{C}^{N \times N} \otimes \mathbb{C}^{2 \times 2}$ as

$$H_z := \begin{pmatrix} 0 & X + A - z \\ (X + A - z)^* & 0 \end{pmatrix}, \tag{3.1}$$

which is also known as the *Hermitization* of $X + A$. As in (3.1), we often omit the dependence on A for simplicity. We may write the spectral decomposition of H_z as

$$H_z = \sum_{|i| \in [\![N]\!]} \lambda_i^z \mathbf{w}_i^z (\mathbf{w}_i^z)^* = \sum_{|i| \in [\![N]\!]} \lambda_i^z \begin{pmatrix} \boldsymbol{u}_i^z \\ \boldsymbol{v}_i^z \end{pmatrix} \begin{pmatrix} \boldsymbol{u}_i^z \\ \boldsymbol{v}_i^z \end{pmatrix}^*, \qquad \|\mathbf{w}_i^z\|^2 = \|\boldsymbol{u}_i^z\|^2 + \|\boldsymbol{v}_i^z\|^2 = 1, \tag{3.2}$$

where the eigenvalues $\{\lambda_i^z : |i| \in [\![N]\!]\}$ are increasingly ordered and we decomposed the eigenvectors $\mathbf{w}_i^z \in \mathbb{C}^{2N}$ into two parts with $\boldsymbol{u}_i^z, \boldsymbol{v}_i^z \in \mathbb{C}^N$. Note that λ_i^z and $\boldsymbol{u}_i^z, \boldsymbol{v}_i^z$ for $i \geq 1$ are exactly the singular values and left and right singular vectors of $(X + A - z)$, respectively. We denote the resolvent of H_z by

$$G_z(w) := (H_z - w)^{-1}$$

for each $w \in \mathbb{C}_+$. To simplify the notation, we further omit the dependence on z to write, for example, $G \equiv G_z$ and $\lambda_i \equiv \lambda_i^z$ when there is no confusion.

Since H_z has only block off-diagonal component, it is clear that $\lambda_{-i} = -\lambda_i$ and $\langle G(\mathrm{i}\eta) \rangle = \mathrm{i}\langle \mathrm{Im}\, G(\mathrm{i}\eta) \rangle$ hold true. Again due to the block structure of H_z, we may take $(\boldsymbol{u}_i, \boldsymbol{v}_i)$ that satisfies $\boldsymbol{u}_{-i} = \boldsymbol{u}_i$ and $\boldsymbol{v}_{-i} = -\boldsymbol{v}_i$, which in turn implies $\|\boldsymbol{u}_i\|^2 = \|\boldsymbol{v}_i\|^2 = 1/2$ whenever $\lambda_i \neq 0$. Since the event $[\lambda_1 > 0]$ has probability 1, we will work on this event in what follows.

For $w \in \mathbb{C}_+$, it is known that $G_z(w)$ concentrates around a deterministic block constant matrix $M_z(w) \in (\mathbb{C}^{N \times N})^{2 \times 2}$; such results are commonly called *local laws*. The matrix $M_z(w) \equiv M$ is the unique solution of the matrix Dyson equation

$$-M^{-1} = -\begin{pmatrix} -w & A - z \\ (A - z)^* & -w \end{pmatrix} + \langle M \rangle, \tag{3.3}$$

with the side condition $\mathrm{Im}\, M = (M - M^*)/(2\mathrm{i}) \geq 0$. Note that the *self consistent density of states (scDos)* corresponding to H_z, given by $\rho^z(x) = \pi^{-1} \langle \mathrm{Im}\, M(x + \mathrm{i}0) \rangle$, is exactly the density of the measure $\mu_{\mathrm{sc}} \boxplus \mu_{|A-z|}^{\mathrm{symm}}$ that we used to define the bulk in (2.2).

More concretely, we have the following local law for H_z from [19, Theorem 2.6]:

Lemma 3.1 [19, Theorem 2.6]. *Let $\tau, \mathfrak{K} > 0$ be fixed. Then for each (small) $\epsilon, \xi > 0$ and (large) $D > 0$, there exists $N_0 \in \mathbb{N}$ such that the following holds uniformly over $N \geq N_0$, $\|A\| \leq \mathfrak{K}$, $z \in \mathcal{B}_\tau$, and $\eta \in [N^{-1+\epsilon}, 1]$;*

$$\mathbf{P}\left[|\langle G_z(\mathrm{i}\eta) - M_z(\mathrm{i}\eta) \rangle| \geq \frac{N^\xi}{N\eta} \right] \leq N^{-D}. \tag{3.4}$$

Consequently, for any fixed (small) $\epsilon, \xi > 0$ and (large) $D > 0$, the following holds uniformly over $|z| \leq \tau$;

$$\mathbb{P}[\Xi_z(\epsilon, \xi)] \geq 1 - N^{-D}, \qquad \Xi_z(\epsilon, \xi) := \bigcap_{\eta \in [N^{-1+\epsilon}, 1]} \left[|\{i \in [\![N]\!] : \lambda_i^z \leq \eta\}| \leq N^{1+\xi}\eta \right]. \tag{3.5}$$

The second result (3.5) is a direct consequence of the local law (3.4) via the inequality

$$|\{i \in [\![N]\!] \,:\, \lambda_i^z \leq \eta\}| \leq 2N\eta \langle \operatorname{Im} G_z(i\eta) \rangle$$

together with the fact that $\|M\| \lesssim 1$ from [19, Appendix A].

Next, we present the two main technical inputs for our proof of Theorem 2.3, (i) an upper bound for $\operatorname{Im}\langle G_z(i\eta)\rangle$ with small $\eta > 0$ and (ii) a high-moment bound for overlaps between the left and right eigenvectors $\boldsymbol{u}_i$ and $\boldsymbol{v}_i$.

Lemma 3.2. *Let X be a real or complex regular i.i.d. matrix and $A \in \mathbb{C}^{N \times N}$ with $\|A\| \leq \mathfrak{K}$. Then for each fixed $\delta, \tau, \xi > 0$ and $\epsilon \in [0, 1]$, there exists a constant $C > 0$ such that the following holds uniformly over all $z \in \mathcal{B}_\tau$ and $\eta \in (0, 1)$;*

$$\mathbb{E}|\langle G_z(i\eta)\rangle|^{2+\epsilon} \leq \begin{cases} 1 + CN^\delta(N\eta)^{-\epsilon-\xi}(|\log \eta| + \log N) & \text{if } X \text{ is complex,} \\[2mm] 1 + CN^\delta(N\eta)^{-\epsilon-\xi}\dfrac{|\log \eta| + \log N}{|\operatorname{Im} z|} & \text{if } X \text{ is real.} \end{cases} \tag{3.6}$$

Proposition 3.3 [19, Theorem 2.2]. *Let X and A be as in Lemma 3.2, and fix $\delta, \tau > 0$ and $p \in \mathbb{N}$. Then there exists a constant $c \in (0, 1)$ such that the following holds uniformly over $z \in \mathcal{B}_\tau$;*

$$N^p \mathbb{E} \sup_{i,j \in [\![cN]\!]} |\langle \boldsymbol{u}_j^z, \boldsymbol{v}_i^z \rangle - \delta_{i,j} q_i(z)|^{2p} \lesssim N^\delta, \tag{3.7}$$

where $\{q_i(z) \,:\, i \in [\![cN]\!]\}$ is the collection of deterministic functions of z defined by

$$q_i(z) := \frac{\langle \operatorname{Im}[M(\gamma_i)]F \rangle}{\langle \operatorname{Im} M(\gamma_i) \rangle}, \qquad F := \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \in (\mathbb{C}^{N \times N})^{2 \times 2} \tag{3.8}$$

and γ_i is the i-th N-quantile of the scDOS of H_z, that is,

$$\gamma_i = \inf\left\{ x \in \mathbb{R} \,:\, \int_0^x \rho^z(t)\mathrm{d}t \geq \frac{i}{2N} \right\}. \tag{3.9}$$

Furthermore, there exists a constant $C > 0$ such that

$$|q_i(z)| \leq C\frac{i}{N}, \qquad \forall i \in [\![cN]\!], \forall z \in \mathcal{B}_\tau. \tag{3.10}$$

The proof of Lemma 3.2 is postponed to the end of this section. It heavily relies on the tail estimate of the lowest singular value λ_1^z from Theorem 2.10(i) for the complex case and Theorem 2.13 for the real case. These theorems are proven separately in Sections 5 and 6. Proposition 3.3 is a direct consequence of the thermalization result in [19, Theorem 2.2] which computes any quadratic form $(\boldsymbol{u}_i^z, B\boldsymbol{v}_j^z)$ of the bulk singular vectors of $X + A - z$ in terms of a leading deterministic quantity up to an error or order $N^{-1/2}$. The upper bound for $q_i(z)$ in (3.10) follows by combining Lemma A.1 (b) and (c) therein, which shows Lipschitz continuity of M (in operator norm) and $\operatorname{Im}[M(i\eta)]F \equiv 0$ for $\eta \in \mathbb{R}$, respectively.

Proof of Theorem 2.3. We only prove the complex case, and the real case follows from the exact same proof. Fix $z_0 \in \mathbb{D}$ with $|z_0| \leq 1 - \tau$, and let $f_0 \in C_c^\infty(\mathbb{C})$ be a fixed cut-off function with $\mathbb{1}_\mathbb{D} \leq$

$f_0 \leq \mathbb{1}_{2\mathbb{D}}$ and $f := f_0(\frac{\cdot - z_0}{r})$. Since $\mathbb{1}(|\cdot - z_0| \leq r) \leq f$, Girko's formula (1.23) and $\mathrm{Im}\langle G_z(i\eta)\rangle = -\mathrm{i}\langle G_z(i\eta)\rangle$ gives

$$
\mathbb{E}\mathcal{N}_{z_0,r} \leq N\mathbb{E} \int_{\mathbb{C}} f(z)\mathrm{d}\rho(z) = -\frac{N}{2\pi} \int_{\mathbb{C}} \Delta f(z) \int_0^T \mathbb{E}\,\mathrm{Im}\langle G_z(i\eta)\rangle \mathrm{d}\eta \mathrm{d}^2 z \tag{3.11}
$$

$$
+ \frac{1}{4\pi} \int_{\mathbb{C}} \Delta f(z)\mathbb{E} \log |\det(H_z - \mathrm{i}T)|\mathrm{d}^2 z
$$

$$
= \frac{N}{2\pi} \int_{\mathbb{C}} \Delta f(z)\left(\int_0^{\eta_0} + \int_{\eta_0}^{\eta_1} + \int_{\eta_1}^T\right)\mathrm{i}\mathbb{E}\langle G_z(i\eta)\rangle \mathrm{d}\eta \mathrm{d}^2 z \tag{3.11}
$$

$$
+ \frac{1}{4\pi} \int_{\mathbb{C}} \Delta f(z)\mathbb{E} \log |\det(H_z - \mathrm{i}T)|\mathrm{d}^2 z
$$

$$
= : I_0 + I_1 + I_T + I_\infty, \tag{3.11}
$$

where we defined

$$
\eta_0 := N^{-1} \cdot (Nr^2)^D, \qquad \eta_1 := 1, \qquad T := N^D(Nr^2)^{-D} \tag{3.12}
$$

for a fixed (large) constant $D > 0$. Since the spectral parameters in $G_z(i\eta)$ are always z and $i\eta$, we abbreviate $G \equiv G_z(i\eta)$ throughout the rest of the proof.

The bound for the first term I_0 follows by taking $\epsilon = 0$ and small enough $\xi > 0$ in Lemma 3.2;

$$
|I_0| \leq \frac{N}{2\pi} \int_{\mathbb{C}} |\Delta f(z)| \int_0^{\eta_0} \|\langle G\rangle\|_2 \mathrm{d}\eta \mathrm{d}^2 z
$$
$$
\lesssim N^{\delta/2} \int_0^{N\eta_0} t^{-\xi/2}\sqrt{|\log t|}\mathrm{d}t \lesssim N^{\delta/2}(N\eta_0)^{1-\xi} \lesssim N^{\delta/2}(Nr^2)^{D/2}, \tag{3.13}
$$

where we used $\|\Delta f\|_{L^1} \lesssim 1$.

To handle I_1 and I_T in (3.11), we use integration by parts, that is, for any $f \in C_c^\infty(\mathbb{C})$ and $0 < A < B < \infty$

$$
\mathbb{E} \int_{\mathbb{C}} \int_A^B \Delta f(z)\langle G\rangle \mathrm{d}\eta \mathrm{d}^2 z = \int_A^B \int_{\mathbb{C}} f(z)\mathbb{E}\langle \Delta_z G_z(i\eta)\rangle \mathrm{d}^2 z \mathrm{d}\eta. \tag{3.14}
$$

The derivative $\Delta_z G$ can be calculated directly; for $F \in \mathbb{C}^{2N \times 2N}$ defined in (3.8),

$$
\frac{1}{4}\Delta_z G = \partial_z \overline{\partial}_z G = -\partial_z GFG = GF^*GFG + GFGF^*G. \tag{3.15}
$$

Now we can estimate I_T in (3.11). Applying the trivial deterministic bound $\|G\| \leq \eta^{-1}$ gives

$$
\mathbb{E}|\langle GFGF^*G + GF^*GFG\rangle| \leq 2\|F\|^2 \mathbb{E}\|G\|^3 \leq 2\eta^{-3},
$$

which yields

$$
|I_T| \lesssim Nr^2 \int_1^T \eta^{-3}\mathrm{d}\eta \leq Nr^2. \tag{3.16}
$$

The estimate for $\mathbb{E}|\langle\Delta G\rangle|$ when $\eta < 1$ is much trickier than $\eta > 1$. We first state the result here as a lemma, use it to conclude Theorem 2.3, and then prove the lemma.

Lemma 3.4. *Let X be a complex regular i.i.d. matrix and $\epsilon, \delta, \tau > 0$ be fixed. Then the following holds uniformly over $|z| \leq 1 - \tau$ and $\eta \in (0, 1)$;*

$$\mathbb{E}|\langle\Delta_z G_z(i\eta)\rangle| \lesssim \frac{N^{1+\delta}}{N\eta}(N\eta \wedge 1)^{-\epsilon}. \tag{3.17}$$

Given Lemma 3.4, we substitute (3.17) into (3.14) with $A = \eta_0$ and $B = \eta_1$ so that

$$|I_1| \leq N \int_{\eta_0}^{\eta_1} \int_{\mathbb{C}} |f(z)||\mathbb{E}\langle\Delta G\rangle|d^2zd\eta \lesssim N^{\delta/2}Nr^2 \int_{\eta_0}^{\eta_1} \frac{(N\eta \wedge 1)^{-\epsilon}}{N\eta}(Nd\eta)$$

$$\lesssim N^{\delta/2}(Nr^2)((N\eta_0)^{-\epsilon} + \log N) = N^\delta(Nr^2)^{1-D\epsilon} \leq N^{2\delta}(Nr^2)^{(1-\gamma)}, \tag{3.18}$$

where in the last step we took $\epsilon < \gamma/D$.

Finally, the last term I_∞ is estimated using $\log(1 + x^2) \leq x^2$ and $\mathbb{E}\langle H_z^2\rangle \lesssim 1$ as

$$|I_\infty| = \frac{1}{2}\left|\int_{\mathbb{C}} \Delta f(z)\mathbb{E}\,\mathrm{Tr}\log(1 + T^{-2}H_z^2)d^2z\right| \leq \frac{N}{T^2}\int_{\mathbb{C}} |\Delta f(z)|\mathbb{E}\langle H_z^2\rangle d^2z \lesssim T^{-2}N \lesssim Nr^2. \tag{3.19}$$

Combining (3.13), (3.16), (3.18), and (3.19), we have proved

$$\mathbb{E}\mathcal{N}_{z_0,r} \lesssim N^{1+2\delta}(Nr^2)^{1-\gamma}. \tag{3.20}$$

This completes the proof of Theorem 2.3 modulo Lemma 3.4. $\qquad\square$

Proof of Lemma 3.4. We start with expressing the trace of the right-hand side of (3.15) in terms of the eigenvalues and eigenvectors of H_z. First we write

$$\langle GF^*GFG\rangle = -\mathrm{i}\frac{\eta}{2}\left\langle \frac{1}{\eta^2 + (X + A - z)^*(X + A - z)} \frac{\eta^2 - (X + A - z)(X + A - z)^*}{(\eta^2 + (X + A - z)(X + A - z)^*)^2}\right\rangle, \tag{3.21}$$

and the second term of (3.15), $\langle GF^*GFG\rangle$, is given by the same quantity as (3.21) with roles of X and X^* interchanged. Here we used the Schur complement form of G, that is,

$$G = G(\mathrm{i}\eta) = \begin{pmatrix} \dfrac{\mathrm{i}\eta}{\eta^2 + (X + A - z)(X + A - z)^*} & \dfrac{1}{\eta^2 + (X + A - z)(X + A - z)^*}(X + A - z) \\ (X + A - z)^*\dfrac{1}{\eta^2 + (X + A - z)(X + A - z)^*} & \dfrac{\mathrm{i}\eta}{\eta^2 + (X + A - z)^*(X + A - z)} \end{pmatrix}. \tag{3.22}$$

Recalling the definition of $\boldsymbol{u}_i$ and $\boldsymbol{v}_i$ and (3.21), we may further write

$$
\begin{aligned}
|\langle GF^*GFG \rangle| &\leq \frac{1}{2N} \sum_{i,j \in \llbracket N \rrbracket} \frac{\eta}{\eta^2 + \lambda_i^2} \frac{|\eta^2 - \lambda_j^2|}{(\eta^2 + \lambda_j^2)^2} |\langle \boldsymbol{u}_j, \boldsymbol{v}_i \rangle|^2 \\
&\leq \frac{1}{2N\eta} \left(\sum_{i,j \in \llbracket cN \rrbracket} + \sum_{i \in \llbracket cN, N \rrbracket, j \in \llbracket N \rrbracket} + \sum_{i \in \llbracket N \rrbracket, j \in \llbracket cN, N \rrbracket} \right) \frac{\eta}{\eta^2 + \lambda_i^2} \frac{\eta}{\eta^2 + \lambda_j^2} |\langle \boldsymbol{u}_j, \boldsymbol{v}_i \rangle|^2 \\
&\leq \frac{1}{2N\eta} \sum_{i,j \in \llbracket cN \rrbracket} \frac{\eta^2}{(\eta^2 + \lambda_i^2)(\eta^2 + \lambda_j^2)} |\langle \boldsymbol{u}_j, \boldsymbol{v}_i \rangle|^2 + \frac{1}{\eta^2 + \lambda_{\lceil cN \rceil}^2} \frac{1}{N} \sum_i \frac{\eta}{\eta^2 + \lambda_i^2},
\end{aligned}
\tag{3.23}
$$

where the constant $c > 0$ is from Proposition 3.3 and in the last inequality we used

$$
\sum_{i \in \llbracket N \rrbracket, j \in \llbracket cN \rrbracket} \frac{\eta}{\eta^2 + \lambda_i^2} \frac{\eta}{\eta^2 + \lambda_j^2} |\langle \boldsymbol{u}_j, \boldsymbol{v}_i \rangle|^2 \leq \frac{\eta}{\eta^2 + \lambda_{\lceil cN \rceil}^2} \sum_{i \in \llbracket N \rrbracket} \frac{\eta}{\eta^2 + \lambda_i^2} \left(\sum_{j \in \llbracket N \rrbracket} |\langle \boldsymbol{u}_j, \boldsymbol{v}_i \rangle|^2 \right),
$$

and that $\boldsymbol{u}_j$'s are orthogonal to perform the j-summation. For the first term on the rightmost side of (3.23), we recall the deterministic function $q_i(z)$ from Proposition 3.3 and write

$$
\begin{aligned}
&\frac{1}{2N\eta} \sum_{i,j \in \llbracket cN \rrbracket} \frac{\eta^2}{(\eta^2 + \lambda_i^2)(\eta^2 + \lambda_j^2)} |\langle \boldsymbol{u}_j, \boldsymbol{v}_i \rangle|^2 \\
&\lesssim \frac{1}{\eta} \langle \operatorname{Im} G \rangle^2 \left(N \sup_{i,j \in \llbracket cN \rrbracket} |\langle \boldsymbol{u}_j, \boldsymbol{v}_i \rangle - \delta_{ij} q_i(z)|^2 \right) + \frac{1}{N\eta} \sum_{i \in \llbracket cN \rrbracket} \frac{\eta^2}{(\eta^2 + \lambda_i^2)^2} \frac{i^2}{N^2}.
\end{aligned}
\tag{3.24}
$$

Plugging in (3.24) to (3.23) and using the same set of inequalities to the second term of (3.15), we find that

$$
\begin{aligned}
|\langle \Delta_z G \rangle| &\lesssim \frac{1}{\eta} \langle \operatorname{Im} G \rangle^2 \left(N \sup_{i,j \in \llbracket cN \rrbracket} |\langle \boldsymbol{u}_j, \boldsymbol{v}_i \rangle - \delta_{ij} q_i(z)|^2 \right) \\
&\quad + \frac{1}{N\eta} \sum_{i \in \llbracket cN \rrbracket} \frac{\eta^2}{(\eta^2 + \lambda_i^2)^2} \frac{i^2}{N^2} + \frac{1}{\eta^2 + \lambda_{\lceil cN \rceil}^2} \langle \operatorname{Im} G \rangle.
\end{aligned}
\tag{3.25}
$$

Next, we estimate the expectation of the right-hand side of (3.25) for $\eta \in (0, 1)$. For the first term in (3.25), we apply Hölder's inequality and $\langle G \rangle = i \operatorname{Im} \langle G \rangle$ to get for all $\eta \in (0, 1)$ that

$$
\begin{aligned}
&\frac{1}{\eta} \mathbb{E} \langle \operatorname{Im} G \rangle^2 \left(N \sup_{i,j \in \llbracket cN \rrbracket} |\langle \boldsymbol{u}_j, \boldsymbol{v}_i \rangle - \delta_{ij} q_i(z)|^2 \right) \\
&\qquad\qquad\qquad\qquad\qquad \leq \frac{1}{\eta} \| \langle G \rangle^2 \|_{1 + c'\varepsilon} \left\| \sup_{i,j \in \llbracket cN \rrbracket} N |\langle \boldsymbol{u}_j, \boldsymbol{v}_i \rangle - \delta_{ij} q_i(z)|^2 \right\|_{1 + 1/(c'\varepsilon)},
\end{aligned}
\tag{3.26}
$$

where $c' > 0$ is a small constant, say $c' = 1/100$. Then we use Lemma 3.2 and Proposition 3.3 with a union bound, respectively, to the first and second factors on the right-hand side of (3.26) to get

$$\| |\langle G \rangle|^2 \|_{1+c'\epsilon} = \| \langle G \rangle \|^2_{2(1+c'\epsilon)} \lesssim \left(1 + N^{\delta/3}(N\eta)^{-2c'\epsilon}(|\log \eta| + \log N) \right)^{1/(1+c'\epsilon)} \lesssim N^{\delta/2}(N\eta \wedge 1)^{-2c'\epsilon},$$

$$\| N \sup_{i,j} |\langle \boldsymbol{u}_j, \boldsymbol{v}_i \rangle - \delta_{ij} q_i(z)|^2 \|_{1+1/(c'\epsilon)} \leq \left(\mathbb{E} \sup_{i,j \in [\![cN]\!]} (N|\langle \boldsymbol{u}_j, \boldsymbol{v}_i \rangle - \delta_{ij} q_i(z)|^2)^{(1+1/(c'\epsilon))} \right)^{c'\epsilon} \lesssim N^{\delta/2}.$$

Thus we conclude

$$\frac{1}{\eta} \mathbb{E} \langle \operatorname{Im} G \rangle^2 \left(N \sup_{i,j \in [\![cN]\!]} |\langle \boldsymbol{u}_j, \boldsymbol{v}_i \rangle|^2 \right) \lesssim \frac{1}{\eta} N^\delta (1 + (N\eta)^{-2c'\epsilon}) = \frac{N^{1+\delta}}{N\eta}(N\eta \wedge 1)^{-2c'\epsilon}. \tag{3.27}$$

To handle the expectation of the second term in (3.25), recall the definition of the event $\Xi_z(\epsilon, \xi)$ from (3.5). On the event $\Xi := \Xi_z(c'\delta/2, c'\delta/2)$ we have

$$\lambda_i \geq N^{-c'\delta/2} \frac{i}{N} \qquad \forall i \in [\![N^{c'\delta}, N]\!], \tag{3.28}$$

which follows from (3.5) by putting in $\eta = N^{-c'\delta/2}(i/N)$. We divide the second term of (3.25) as

$$\sum_{i \in [\![cN]\!]} \frac{\eta^2}{(\eta^2 + \lambda_i^2)^2} \frac{i^2}{N^2} \leq \left(\sum_{i \in [\![N^{c'\delta}]\!]} + \mathbb{1}_{\Xi^c} \sum_{i \in [\![cN]\!]} + \mathbb{1}_{\Xi} \sum_{i \in [\![N^{c'\delta}, cN]\!]} \right) \frac{\eta^2}{(\eta^2 + \lambda_i^2)^2} \frac{i^2}{N^2}$$

$$\leq (N^{2c'\delta} + N^2 \mathbb{1}_{\Xi^c})\langle \operatorname{Im} G \rangle^2 + N^{2c'\delta} \sum_{i \in [\![N]\!]} \frac{\eta^2 \cdot (N^{-1}i)^2}{(\eta^2 + (N^{-1}i)^2)^2} \lesssim (N^{2c'\delta} + N^2 \mathbb{1}_{\Xi^c})\langle \operatorname{Im} G \rangle^2 + N^{1+2c'\delta}\eta.$$

Then we follow the same argument as in (3.26) to conclude for any fixed $D > 0$ that

$$\frac{1}{N\eta} \mathbb{E} \sum_{i \in [\![cN]\!]} \frac{\eta^2}{(\eta^2 + \lambda_i^2)^2} \frac{i^2}{N^2} \leq \frac{1}{N\eta} \left\| N^{2c'\delta} + N^2 \mathbb{1}_{\Xi^c} \right\|_{1+1/(c'\epsilon)} \| \langle G \rangle^2 \|_{1+c'\epsilon} + \frac{N^{2c'\delta}}{\eta}$$

$$\lesssim \frac{N^{2c'\delta}}{\eta} + \frac{N^{-D+\delta/2}}{\eta}(N\eta \wedge 1)^{-2c'\epsilon} \leq \frac{N^{1+\delta}}{N\eta}(N\eta \wedge 1)^{-\epsilon}. \tag{3.29}$$

For the expectation of the last term of (3.25), we again use the same event $\Xi = \Xi_z(c'\delta/2, c'\delta/2)$ but we only require that $\lambda_{\lceil cN \rceil} \geq cN^{-c'\delta/2}$ on the event. Then we write

$$\mathbb{E} \mathbb{1}_{\Xi} \frac{1}{\eta^2 + \lambda_{\lceil cN \rceil}^2} \langle \operatorname{Im} G \rangle \lesssim N^{c'\delta} \mathbb{E} \langle \operatorname{Im} G \rangle \leq N^{c'\delta} \| \langle G \rangle \|_2 \lesssim N^\delta (1 + (N\eta)^{-\epsilon}) \leq \frac{N^{1+\delta}}{N\eta}(N\eta \wedge 1)^{-\epsilon}. \tag{3.30}$$

The complementary event Ξ^c can be dealt with the same reasoning as in (3.29). We thus conclude

$$\mathbb{E} |\langle \Delta_z G \rangle| \lesssim N^{1+\delta}(N\eta \wedge 1)^{-1-\epsilon} \tag{3.31}$$

as desired. $\qquad \square$

Remark 3.5. As easily seen from the proof, the factor of N^δ in Theorem 2.3 is due to the same factors in Lemma 3.2 and Proposition 3.3. In the proof of Lemma 3.2, we use the local laws for

H_z, Lemma 3.1, to show that λ_1^z determines the size of $\langle G_z \rangle$ up to a factor of N^δ. If we can show that the random variable $|\{i : \lambda_i^z \leq K/N\}|$ has finite high-moment for any fixed K, then we may use this as an alternative input and remove N^δ from Lemma 3.2. Likewise, Proposition 3.3 uses the two-resolvent local laws in [19, Theorem 4.4] that is designed for mesoscopic scales. If h has a better decay rate, we expect Proposition 3.3 to be true without the factor of N^δ on the right-hand side of (3.7), which would reduce N^δ in Theorem 2.3 to a logarithmic correction.

On the other hand, changing the power $(1 - \gamma)$ to the optimal first power of Nr^2 seems harder. It is due to a Hölder inequality in (3.26), whose purpose is to separate the singular values and the overlaps since we do not know how to estimate their joint distribution effectively.

Proof of Corollary 2.4. We first prove the complex case and later show how to modify the proof for the real case. Fix $z \in \mathcal{B}_\tau$. We decompose the special test function $w \mapsto 1/w$ into three parts by inserting cutoff functions as follows.

$$\frac{1}{w} = \frac{1}{w}(1 - f_0(w/\kappa)) + \frac{1}{w}f_0(N^{1/2}w) + \frac{1}{w}f_0(w/\kappa)(1 - f_0(N^{1/2}w)) =: f_1(w) + f_2(w) + f_3(w),$$
(3.32)

where $\kappa \geq N^{-1/2}$ is such that[4] $D(z, 2\kappa) \subset \mathcal{B}_{\tau/2}$ and the cutoff function f_0 was defined in the proof of Theorem 2.3. Thus we may write

$$\mathbb{E}|\langle (X - z)^{-1} \rangle|^{2-\delta_1} = \mathbb{E}\left| \int_{\mathbb{C}} (f_1(w - z) + f_2(w - z) + f_3(w - z))d\rho(w) \right|^{2-\delta_1}$$
(3.33)

where ρ is the spectral distribution of X defined in (2.1).

The first integral is the easiest as $|f_1(w)| \leq \kappa^{-1}$ for all $w \in \mathbb{C}$, so that

$$\mathbb{E}\left| \int_{\mathbb{C}} f_1(w - z)d\rho(w) \right|^{2-\delta_1} \leq \kappa^{-2+\delta_1}.$$
(3.34)

For the second integral, we use Jensen's inequality (recall $\delta_1 \in (0, 1]$) and supp $f_0 \subset D(0, 2)$ to write

$$\left| \int_{\mathbb{C}} f_2(w - z)d\rho(w) \right|^{2-\delta_1} \leq \rho(D(z, 2N^{-1/2}))^{1-\delta_1} \int_{\mathbb{C}} \frac{1}{|w - z|^{2-\delta_1}} \mathbb{1}_{D(z,2N^{-1/2})}(w)d\rho(w).$$
(3.35)

We then consider the following dyadic decomposition of $D(z, 2N^{-1/2})$ according to $|w - z|$;

$$D(z, 2N^{-1/2}) = \bigcup_{k=0}^{\infty}(D(z, 2^{-k+1}N^{-1/2}) \setminus D(z, 2^{-k}N^{-1/2})) =: \bigcup_{k=0}^{\infty} D_k.$$
(3.36)

Applying Theorem 2.3 (i) to each domain D_k with some $\delta < \delta_1/2$ and $\gamma < \delta_1/2$, we have

$$\mathbb{E}\int_{D_k} \frac{1}{|w - z|^{2-\delta_1}}d\rho(w) \leq (2^k N^{1/2})^{2-\delta_1}\mathbb{E}\int_{D_k}d\rho(w) \leq N^{-1}(2^k N^{1/2})^{2-\delta_1}\mathbb{E}\mathcal{N}_{z,N^{-1/2}2^{-k+1}}$$

$$\leq CN^{\delta-\delta_1/2}2^{k(2-\delta_1)-2(k-1)(1-\gamma)} \leq CN^{\delta-\delta_1/2}2^{k(2\gamma-\delta_1)}.$$
(3.37)

---

[4] If N is so small that such a κ does not exists, we pick κ with $D(z, 2\kappa) \subset \mathcal{B}_{\tau/2}$ and write $f_2 \equiv 0$, $f_3(w) = f_0(w/\kappa)/\kappa$.

Therefore, summing over k and using $\rho(D(z, 2N^{-1/2})) \leq \rho(\mathbb{C}) = 1$, we obtain

$$\mathbb{E}\left|\int_{\mathbb{C}} f_2(w - z)\mathrm{d}\rho(w)\right|^{2-\delta_1} \leq \sum_{k=0}^{\infty} \mathbb{E}\int_{D_k} \frac{1}{|w - z|^{2-\delta_1}}\mathrm{d}\rho(w) \leq CN^{\delta-\delta_1/2} \leq C. \tag{3.38}$$

The last integral in (3.33) can be dealt with the local law for $X + A$. This is the natural extension of the local circular law to $X + A$ and it asserts that for all (possibly N-dependent) C_c^2 test functions g supported in $\mathcal{B}_{\tau/2}$ and fixed positive ϵ and D we have

$$\mathbb{P}\left[\left|\int_{\mathbb{C}} g(w)\mathrm{d}\rho(w) - \int_{\mathbb{C}} g(w)\mathrm{d}\rho_{a+x}(w)\right| \geq \frac{N^\epsilon}{N}\|\Delta g\|_{L^1}\right] \lesssim N^{-D}. \tag{3.39}$$

We omit its proof as it is standard, given the optimal local law [19, Theorem 2.6] as an input, following the analogous argument in the proof of [3, Theorem 2.5]. We apply (3.39) to $g(w) = f_3(w - z)$, which is supported on $\mathcal{B}_{\tau/2}$ by definition. We denote the "good" event in (3.39) for this choice of g by Ξ_ϵ, that is,

$$\Xi_\epsilon := \left[\left|\int_{\mathbb{C}} f_3(w - z)\mathrm{d}\rho(w) - \int_{\mathbb{C}} f_3(w - z)\mathrm{d}\rho_{a+x}(w)\right| \leq \frac{N^\epsilon}{N}\|\Delta f_3\|_{L^1}\right], \tag{3.40}$$

so that $\mathbb{P}[\Xi_\epsilon^c] \lesssim N^{-D}$ for all fixed $D > 0$. Now notice that

$$\|\Delta f_3\|_{L^1} \lesssim N^{1/2}. \tag{3.41}$$

On the other hand, since ρ_{a+x} has a bounded density in $\mathcal{B}_\tau$ by (2.7), we have

$$\left|\int_{\mathbb{C}} f_3(w - z)\mathrm{d}\rho_{a+x}\right| \lesssim \int_{D(0,\kappa)} \frac{1}{|z|}\mathrm{d}^2 z \lesssim 1. \tag{3.42}$$

Hence on the event Ξ_ϵ we have

$$\mathbb{E}\mathbb{1}_{\Xi_\epsilon}\left|\int_{\mathbb{C}} f_3(w - z)\mathrm{d}\rho_{a+x}(w)\right|^{2-\delta_1} \leq 1 + \left(\frac{N^\epsilon}{N^{1/2}}\right)^{2-\delta_1} \lesssim 1. \tag{3.43}$$

On the complementary event Ξ_ϵ^c, we recall $|f_3(w)| \leq N^{1/2}$ from the definition of f_3, which gives

$$\mathbb{E}\mathbb{1}_{\Xi_\epsilon^c}\left|\int_{\mathbb{C}} f_3(w - z)\mathrm{d}\rho(w)\right|^{2-\delta_1} \leq \mathbb{P}[\Xi_\epsilon^c]N^{1-\delta_1/2} \lesssim N^{1-D}. \tag{3.44}$$

Combining (3.43) and (3.44) we arrive at

$$\mathbb{E}\left|\int_{\mathbb{C}} f_3(w - z)\mathrm{d}\rho(w)\right|^{2-\delta_1} \lesssim 1. \tag{3.45}$$

Plugging in (3.34), (3.38), and (3.45) to (3.33) concludes the proof of Corollary 2.4 in the complex case.

Finally, we show how to modify the proof for real X. The estimates for f_1 and f_3 are completely analogous to the complex case, and their contributions are both bounded by a constant. Henceforth we focus on estimating the contribution of f_2. We consider the cases $y = \lambda_1(\Im[A - z]) \geq N^{-1}$ and $y \leq N^{-1}$ separately.

When $y \geq N^{-1}$, we define $\widetilde{y} := \min(N^{-1/2}, y/4)$ and further divide f_2 into two parts,

$$f_2(w) = \frac{1}{w} f_0(\widetilde{y}^{-1} w) + \frac{1}{w} f_0(N^{1/2} w)(1 - f_0(\widetilde{y}^{-1} w)) =: f_{21}(w) + f_{22}(w). \tag{3.46}$$

The contribution of f_{21} can be estimated in the exact same fashion as f_2 for the complex case, using a dyadic decomposition. The only difference is that we use radii $2^{-k+1}\widetilde{y}$ instead of $2^{-k+1}N^{-1/2}$, and apply Theorem 2.3 (ii) in place of (i). As a result we obtain

$$\mathbb{E}\left|\int_{\mathbb{C}} f_{21}(w-z)\mathrm{d}\rho(w)\right|^{2-\delta_1} \leq C\frac{N^{\delta-\delta_1/2}}{y}(N\widetilde{y}^2)^{\delta_1/2-\gamma} \leq C\frac{N^{\delta-\delta_1/2}}{y}. \tag{3.47}$$

For f_{22}, we use the crude bound $|f_{22}(w)| \leq 2\widetilde{y}^{-1}f_0(N^{1/2}(w))$ to get

$$\mathbb{E}\left|\int_{\mathbb{C}} f_{22}(w-z)\mathrm{d}\rho(w)\right|^{2-\delta_1} \leq C\widetilde{y}^{\delta_1-2}\mathbb{E}\int_{\mathbb{C}} f_0(N^{1/2}(w-z))\mathrm{d}\rho(w) \leq CN^{-1+\delta}\widetilde{y}^{\delta_1-2}$$

$$\leq CN^{\delta-\delta_1/2} + CN^{-1+\delta}y^{\delta_1-2}. \tag{3.48}$$

where in the first line we replaced $\mathrm{d}\rho(w)$ with the bounded density $\mathrm{d}\rho_{a+x}(w)$ by applying (3.39) to $g(w) = f_0(N^{1/2}(w-z))$, using that $\|\Delta g\|_{L^1} = \|\Delta f_0\|_{L^1} = O(1)$ for this choice of g. Adding (3.47) and (3.48) proves

$$\mathbb{E}\left|\int_{\mathbb{C}} f_2(w-z)\mathrm{d}\rho(w)\right|^{2-\delta_1} \leq C + C\frac{N^\delta}{y}\left(N^{-\delta_1/2} + \frac{1}{Ny^{1-\delta_1}}\right) \leq C + \frac{N^{-\delta_1/2+\delta}}{y}, \tag{3.49}$$

where we used $y \geq N^{-1}$ in the last inequality.

Now it only remains to consider the case $y \leq 1/N$ and prove

$$\mathbb{E}\left|\int_{\mathbb{C}} f_2(w-z)\mathrm{d}\rho(w)\right|^{2-\delta_1} \leq C\frac{N^\delta}{y^{1-\delta_1/2+\delta}}. \tag{3.50}$$

Here we use that $|f_2(w-z)| \leq f_0(N^{1/2}(w-z))/\lambda_1^z$, which gives

$$\mathbb{E}\left|\int_{\mathbb{C}} f_2(w-z)\mathrm{d}\rho(w)\right|^{2-\delta_1} \leq \mathbb{E}(N\lambda_1^z)^{\delta_1-2}\left|N\int_{\mathbb{C}} f_0(N^{1/2}(w-z))\mathrm{d}\rho(w)\right|^{2-\delta_1}. \tag{3.51}$$

Then we apply Hölder inequality in (3.51) to obtain for all fixed $\epsilon'' \in (0, \delta_1)$ that

$$\mathbb{E}\left|\int_{\mathbb{C}} f_2(w-z)\mathrm{d}\rho(w)\right|^{2-\delta_1} \lesssim N^{\delta/2}(\mathbb{E}(N\lambda_1^z)^{-2+\epsilon''})^{(2-\delta_1)/(2-\epsilon'')}, \tag{3.52}$$

where we used local laws to the second factor in (3.51) to bound its high moment by $N^{\delta/2}$, as in (3.48). Now by (2.27) in Theorem 2.13 and an integration by parts we can compute the right-hand

side of (3.52) as

$$\mathbb{E}(N\lambda_1^z)^{-2+\epsilon''} = (2 - \epsilon'') \int_0^\infty t^{-3+\epsilon''} \mathbb{P}[N\lambda_1^z \le t] dt \lesssim 1 + \frac{N^{\delta/2}}{y} \int_0^1 t^{-1+\epsilon''} \log t \, dt \lesssim \frac{N^{\delta/2}}{y}.$$

Thus we have

$$\mathbb{E}\left| \int_{\mathbb{C}} f_2(w - z) d\rho(w) \right|^{2-\delta_1} \lesssim N^\delta y^{-(2-\delta_1)/(2-\epsilon'')},$$

and choosing suitable $\epsilon'' \in (0, \delta_1)$ concludes (3.50). This completes the proof of Corollary 2.4. $\qquad\square$

Proof of Lemma 3.2. The only input for the proof, other than Lemma 3.1, is Theorem 2.10.(i) for the complex case and Theorem 2.13 for the real case. We restrict ourselves to the complex case, since the real case follows directly by replacing the input accordingly.

First, we assume $\eta \le N^{-1+\delta/2}$ without loss of generality, since for $\eta \ge N^{-1+\delta/2}$ the result is a direct consequence of Lemma 3.1 and $\|M\| \lesssim 1$ from [19, Appendix A]. Then we separate the contribution of the singular values λ_i^z below $N^{-1+\delta/2}$ from the rest;

$$\mathbb{E}|\langle G \rangle|^{2+\epsilon} \lesssim \mathbb{E} n_0^{2+\epsilon} \left(\frac{N\eta}{(N\lambda_1^z)^2 + (N\eta)^2} \right)^{2+\epsilon} + \mathbb{E} \left(\frac{1}{N} \sum_{i : \lambda_i^z > N^{-1+\delta/2}} \frac{\eta}{(\lambda_i^z)^2 + \eta^2} \right)^{2+\epsilon}, \tag{3.53}$$

where we denoted by n_0 the number of singular values below $N^{-1+\delta/2}$;

$$n_0 := |\{i \in [\![N]\!] : \lambda_i^z \le N^{-1+\delta/2}\}|. \tag{3.54}$$

By (3.5), the random variable n_0 satisfies $\|n_0\|_p \lesssim N^\delta$ for any fixed $p \in \mathbb{N}$. Thus applying Hölder's inequality to the first term of (3.53) gives that for each $\xi > 0$

$$\mathbb{E}|\langle G \rangle|^{2+\epsilon} \lesssim N^\delta \left\| \frac{N\eta}{(N\lambda_1^z)^2 + (N\eta)^2} \right\|_{2+\epsilon+\xi}^{2+\epsilon} + \mathbb{E}\langle \mathrm{Im}\, G_z(iN^{-1+\delta/2}) \rangle^{2+\epsilon}, \tag{3.55}$$

where we estimated the second term of (3.53) using that $\eta \mapsto \eta/(x^2 + \eta^2)$ is increasing in $\eta \in [0, |x|]$. Denoting $F_1(\cdot) = \mathbb{P}[N\lambda_1^z \le \cdot]$, the first term of (3.55) is estimated using (2.28) as

$$\mathbb{E}\left(\frac{N\eta}{(N\lambda_1^z)^2 + (N\eta)^2} \right)^{2+\epsilon+\xi} \lesssim \int_0^\infty \int_x^\infty \frac{(N\eta)^{2+\epsilon+\xi} \cdot t}{(t^2 + (N\eta)^2)^{3+\epsilon+\xi}} dt \, dF_1(x)$$

$$= \int_0^\infty \frac{(N\eta)^{2+\epsilon+\xi} \cdot t}{(t^2 + (N\eta)^2)^{3+\epsilon+\xi}} F_1(t) dt \lesssim \int_0^\infty \frac{(N\eta)^{2+\epsilon+\xi} \cdot t^3}{(t^2 + (N\eta)^2)^{3+\epsilon+\xi}} (|\log t| + \log N) dt \tag{3.56}$$

$$= (N\eta)^{-\epsilon-\xi} \int_0^\infty \frac{s^3}{(s^2 + 1)^{3+\epsilon+\xi}} (|\log(N\eta s)| + \log N) ds \lesssim (N\eta)^{-\epsilon-\xi} (|\log \eta| + \log N),$$

where in the first inequality we used the identity

$$\frac{1}{(x^2 + y^2)^p} = 2p \int_x^\infty \frac{t}{(t^2 + y^2)^{p+1}} dt, \qquad p, x, y > 0. \tag{3.57}$$

For the second expectation in (3.55), we directly use (3.4) and $\|M\| \lesssim 1$ from [19, Appendix A] to get

$$\mathbb{E}\langle \operatorname{Im} G_z(iN^{-1+\delta/2})\rangle^{2+\epsilon} \lesssim 1 + \frac{1}{N^{\delta/2}} \lesssim 1. \tag{3.58}$$

Substituting (3.56) and (3.58) into (3.55) completes the proof of Lemma 3.2 for the complex case. □

4 | UPPER BOUND FOR THE OVERLAP: PROOF OF THEOREMS 2.7 AND 2.9

We start with the proof of Theorem 2.9 for it motivates that of Theorem 2.7 while being much shorter.

Proof of Theorem 2.9. Recall the inequalities in (1.17). In fact, they follow from more general, elementary deterministic inequalities (see e.g., [7, Lemma 3.2] for a proof)

$$\begin{aligned}
\pi \sum_{i:\sigma_i \in \mathcal{D}} \mathcal{O}_{ii} &\leq \liminf_{\epsilon \to 0} \frac{\left|\{z \in \mathcal{D} : \lambda_1^z \leq \epsilon\right|}{\epsilon^2}, \\
2 \sum_{i:\sigma_i \in \mathcal{I}} \sqrt{\mathcal{O}_{ii}} &\leq \liminf_{\epsilon \to 0} \frac{\left|\{x \in \mathcal{I} : \lambda_1^x \leq \epsilon\}\right|}{\epsilon},
\end{aligned} \tag{4.1}$$

that hold true for any matrix with simple spectrum and any Borel sets $\mathcal{D} \subset \mathbb{C}$ and $\mathcal{I} \subset \mathbb{R}$. Simply taking the expectation of these two inequalities and using Fatou's lemma gives

$$\begin{aligned}
\mathbb{E} \sum_{i:\sigma_i \in \mathcal{D}} \mathcal{O}_{ii} &\leq \liminf_{\epsilon \to 0} \frac{1}{\pi\epsilon^2} \mathbb{E}|\{z \in \mathcal{D} : \lambda_1^z \leq \epsilon| = \frac{N^2}{\pi} \liminf_{\epsilon \to 0} \int_{\mathcal{D}} \frac{\mathbb{P}[N\lambda_1^z \leq N\epsilon]}{(N\epsilon)^2} \mathrm{d}^2 z, \\
\mathbb{E} \sum_{i:\sigma_i \in \mathcal{I}} \sqrt{\mathcal{O}_{ii}} &\leq \liminf_{\epsilon \to 0} \frac{1}{2\epsilon} \mathbb{E}|\{x \in \mathcal{I} : \lambda_1^x \leq \epsilon| = \frac{N}{2} \liminf_{\epsilon \to 0} \int_{\mathcal{I}} \frac{\mathbb{P}[N\lambda_1^x \leq N\epsilon]}{N\epsilon} \mathrm{d}x.
\end{aligned} \tag{4.2}$$

Then plugging in the results of Theorem 2.11 into (4.2) proves Theorem 2.9. □

One can immediately see that it is vital to have the exact rate $\mathbb{P}[N\lambda_1 \leq s] \lesssim s^2$ or s for the proof above. Any suboptimal factor in s, even $|\log s|$ as in Theorems 2.10 and 2.13, completely ruins the proof, due to the fact that the limit $\epsilon \to 0$ in (4.1) is not quantitative. Inspecting the original proof of (4.1) in [7], one finds that in order for an inequality like (4.1) to be true, ϵ has to be smaller than, say, (i) minimal gap between eigenvalues in $\mathcal{D}$ and (ii) $1/\lambda_1^z$. Roughly, our proof of Theorem 2.7 shows that a suitable deterministic choice of ϵ is smaller than both of (i) and (ii) with high probability, and quantifies the limit in (4.1) via contour integrals.

Proof of Theorem 2.7. First of all, notice that (2.16) follows immediately from (2.14) after an application of Markov's inequality, so we will focus on (2.14) and (2.15).

Next we prove that it suffices to assume $\mathcal{D} \subset [-4 - \|A\|, 4 + \|A\|]^2$. Suppose that we have the result for squares in $[-4 - \|A\|, 4 + \|A\|]^2$, and take a general square $\mathcal{D}$ with $|\mathcal{D}| \geq N^{-2K}$. Define

$$
\mathcal{D}_0 := \begin{cases} \emptyset & \text{if } \mathcal{D} \cap (-3 - \|A\|, 3 + \|A\|)^2 = \emptyset, \\ \mathcal{D} \cap [-4 - \|A\|, 4 + \|A\|]^2 & \text{otherwise.} \end{cases}
$$

Then $\mathcal{D}_0$ is either empty or a rectangle whose side lengths are at least N^{-K}. Since (2.14) and (2.15) are additive with respect to the domain, the conclusion is true for $\mathcal{D}_0$ with an exceptional event $\Xi_{\mathcal{D}_0}$ such that $\mathbb{P}[\Xi_{\mathcal{D}_0}^c] \leq N^{-D}$. We then consider the event $\Xi_0 := [\|X + A\| \leq 3 + \|A\|]$, which satisfies $\mathbb{P}[\Xi_0^c] \lesssim N^{-D}$ for all fixed $D > 0$ by, for example [39, Theorem 3.1]. Defining $\Xi_{\mathcal{D}} = \Xi_{\mathcal{D}_0} \cap \Xi_0$ we have $\mathbb{P}[\Xi_{\mathcal{D}}^c] \lesssim N^{-D}$. Since there is no eigenvalue in $[-3 - \|A\|, 3 + \|A\|]^c \supset \mathcal{D} \setminus \mathcal{D}_0$ on the event Ξ_0, (2.14) and (2.15) hold true for the general square $\mathcal{D}$.

Thus from now on we assume $\mathcal{D} \subset [-4 - \|A\|, 4 + \|A\|]^2$. For a parameter $r < \sqrt{|\mathcal{D}|}$ to be chosen later, we define $\mathcal{C}_r$ to be a partition of $\mathcal{D}$ consisting of solid, open squares with side length r so that $|\mathcal{C}_r| \sim |\mathcal{D}| r^{-2}$ where $|\mathcal{C}_r|$ denotes the number of elements in the partition. Now for this partition, we define the events

$$
\Xi_1 := \bigcap_{C \in \mathcal{C}_r} [|\lambda_1^z| > \mathfrak{a}, \forall z \in \partial C], \qquad \Xi_2 := \bigcap_{C \in \mathcal{C}_r} [|\{i : \sigma_i \in C\}| \leq 1], \quad \text{and} \quad \Xi := \Xi_1 \cap \Xi_2
$$

(4.3)

where $\mathfrak{a} > 0$ is an N-dependent parameter that will also be chosen later. Under these notations, we estimate $\mathbb{P}[\Xi^c]$ and the expectation in (2.14) and (2.15) with the following lemma.

Lemma 4.1. *Let $\mathcal{D} \subset \mathbb{C}$ be a square and define $\Xi_1, \Xi_2,$ and Ξ as in (4.3).*

- *If X is complex, we have*

$$
\mathbb{E}\mathbb{1}_{\Xi} \sum_{i : \sigma_i \in \mathcal{D}} \mathcal{O}_{ii} \lesssim N(N|\mathcal{D}|)|\log \mathfrak{a}|^2,
$$

(4.4)

$$
\mathbb{P}[\Xi_1^c] \lesssim N^2 r^{-1} \mathfrak{a} |\log \mathfrak{a}| |\mathcal{D}|,
$$

(4.5)

$$
\mathbb{P}[\Xi_2^c] \lesssim N^{16/5} r^{6/5} |\log r|^2 |\mathcal{D}|.
$$

(4.6)

uniformly over $0 < \mathfrak{a} \leq r \leq \min((4N)^{-1}, \sqrt{|\mathcal{D}|})$.

- *If X is real, we have*

$$
\mathbb{E}\mathbb{1}_{\Xi} \sum_{i : \sigma_i \in \mathcal{D}} \mathcal{O}_{ii} \lesssim N \frac{(1 + \|A\|^2)N|\mathcal{D}|}{y} |\log \mathfrak{a}|^2,
$$

(4.7)

$$
\mathbb{P}[\Xi_1^c] \lesssim N^2 r^{-1} \mathfrak{a} \frac{|\log \mathfrak{a}|}{y} |\mathcal{D}|,
$$

(4.8)

$$
\mathbb{P}[\Xi_2^c] \lesssim N^{8/3} r^{2/3} y^{-2/3} (|\log r| + |\log y|)^2 |\mathcal{D}|.
$$

(4.9)

uniformly over $0 < \mathfrak{a} \leq r \leq \min\left\{(4N)^{-1}, \sqrt{|\mathcal{D}|}, y/2\right\}$, where we denoted $y := \min_{z \in \mathcal{D}} \lambda_1(\mathfrak{I}[A - z])$.

We postpone the proof of Lemma 4.1 and first deduce the complex case, (2.14), from the lemma. Writing $\mathfrak{a} = N^{-k_1}$ and $r = N^{-k_2}$ with $k_1, k_2 > 1$, we will show that we can choose constant parameters k_1 and k_2 so that (2.14) is true and the right-hand sides of (4.5), (4.6) are $O(N^{-D})$.

First of all, taking $\mathfrak{a} = N^{-k_1}$ in (4.4) with constant k_1 immediately proves (2.14). Then we plug in $\mathfrak{a} = N^{-k_1}$ and $r = N^{-k_2}$ into (4.5) and (4.6), so that

$$\mathbb{P}[\Xi_1^c] \lesssim N^{k_2 - k_1 + 2 + 2\aleph_0} \log N \quad \text{and} \quad \mathbb{P}[\Xi_2^c] \lesssim N^{\frac{16 - 6k_2}{5} + 2\aleph_0} (\log N)^2, \tag{4.10}$$

where we used $|\mathcal{D}| \leq (N^{\aleph_0} + 4)^2$. Therefore choosing $k_2 = K + D + 2\aleph_0 + 100$ and $k_1 = k_2 + D + 2\aleph_0 + 100$ proves $P[\Xi^c] \lesssim N^{-D}$. We omit the proof for the real case since it is completely analogous except that we use $N^{-K} \leq y \leq 2N^{\aleph_0} + 4$. This finishes the proof of Theorem 2.7 modulo Lemma 4.1. □

Proof of Lemma 4.1. We first prove the complex case, and start with the proof of (4.4). The major problem is that we cannot pull the sum over $\{i : \sigma_i \in \mathcal{D}\}$ out of the expectation, since the set of indices itself is random. To circumvent this, we use the partition $\mathcal{C}_r$ and the fact that each $C \in \mathcal{C}_r$ contains at most one eigenvalue on the event $\Xi \subset \Xi_2$, hence

$$\mathbb{E}\mathbb{1}_\Xi \sum_{i : \sigma_i \in \mathcal{D}} \mathcal{O}_{ii} = \mathbb{E}\mathbb{1}_\Xi \sum_{C \in \mathcal{C}_r} \sum_{i : \sigma_i \in C} \mathcal{O}_{ii} = \mathbb{E}\mathbb{1}_\Xi \sum_{C \in \mathcal{C}_r} \mathbb{1}(\exists \sigma_i \in C) \mathcal{O}_{ii} = \sum_{C \in \mathcal{C}_r} \mathbb{E}\mathbb{1}_\Xi \mathbb{1}(\exists \sigma_i \in C) \mathcal{O}_{ii}. \tag{4.11}$$

Note that the deterministic sum over $C \in \mathcal{C}_r$ in effect replaces the sum over i, and in the last equality of (4.11) we interchanged it with the expectation. Furthermore, for each $C \in \mathcal{C}_r$ we have

$$\frac{1}{2\pi i} \oint_{\partial C} \frac{1}{X + A - z} dz = \frac{1}{2\pi i} \oint_{\partial C} \sum_i \frac{dz}{\sigma_i - z} r_i l_i^* = -\sum_{i : \sigma_i \in C} r_i l_i^*, \tag{4.12}$$

which is valid as $\mathbb{P}[\exists \sigma_i \in \partial C] = 0$ and $X + A$ is simple almost surely. Recalling the definition of $\mathcal{O}_{ij}$ from (2.13), this in turn implies

$$\frac{1}{4\pi^2} \mathbb{1}_\Xi \left\| \oint_{\partial C} \frac{1}{X + A - z} dz \right\|^2 = \mathbb{1}_\Xi \left\| \sum_{i : \sigma_i \in C} r_i l_i^* \right\|^2 = \mathbb{1}_\Xi \mathbb{1}(\exists \sigma_i \in C) \mathcal{O}_{ii}, \tag{4.13}$$

where we used that there is at most one eigenvalue in each C on the event Ξ. Since the equality in (4.13) is true for every $C \in \mathcal{C}_r$, we take the expectation and plug it into (4.11) to obtain

$$\mathbb{E}\mathbb{1}_\Xi \sum_{i : \sigma_i \in \mathcal{D}} \mathcal{O}_{ii} \leq \frac{1}{4\pi^2} \sum_{C \in \mathcal{C}_r} \mathbb{E}\mathbb{1}_\Xi \left\| \oint_{\partial C} \frac{1}{(X + A - z)} dz \right\|^2. \tag{4.14}$$

We next estimate the contour integral for each C. First, we use $\|Y\|^2 = \|YY^*\|$ to write

$$\mathbb{E}\mathbb{1}_\Xi \left\| \oint_{\partial C} \frac{1}{(X+A-z)} \mathrm{d}z \right\|^2 = \mathbb{E}\mathbb{1}_\Xi \left\| \oint_{\partial C} \oint_{\partial C} \frac{1}{(X+A-z)} \frac{1}{(X+A-w)^*} \mathrm{d}z \mathrm{d}\overline{w} \right\|$$

$$\leq \mathbb{E}\mathbb{1}_\Xi \int_{\partial C} \int_{\partial C} \left\| \frac{1}{(X+A-z)} \frac{1}{(X+A-w)^*} \right\| |\mathrm{d}z||\mathrm{d}w|$$

$$= \int_{\partial C} \int_{\partial C} \mathbb{E}\mathbb{1}_\Xi \left\| \frac{1}{(X+A-z)} \frac{1}{(X+A-w)^*} \right\| |\mathrm{d}z||\mathrm{d}w| \qquad (4.15)$$

$$\leq 16r^2 \sup_{z,w \in \partial C} \mathbb{E}\mathbb{1}_{\Xi_1 \cap \Xi_2} \left\| \frac{1}{X+A-z} \frac{1}{(X+A-w)^*} \right\| \leq 16r^2 \sup_{z \in \partial C} \mathbb{E}\mathbb{1}_{\Xi_1} \left\| \frac{1}{X+A-z} \right\|^2,$$

where in last line we used Cauchy-Schwarz inequality and also dropped $\mathbb{1}_{\Xi_2}$. Since $\lambda_1^z \geq \mathfrak{a}$ on the event Ξ_1, for all $i \in [\![N]\!]$ and $z \in \partial C$ we have from Theorem 2.10 that

$$\mathbb{E}\mathbb{1}_{\Xi_1} \left\| \frac{1}{X+A-z} \right\|^2 = N^2 \mathbb{E}\mathbb{1}_{\Xi_1} \frac{1}{(N\lambda_1^z)^2} = 2N^2 \int_{N\mathfrak{a}}^{\infty} \frac{1}{s^3} \mathbb{P}[N\lambda_1^z \leq s]\mathrm{d}s$$

$$\leq N^2 + CN^2 \int_{N\mathfrak{a}}^{1} \frac{1}{s}(|\log s| + \log N)\mathrm{d}s \leq N^2 + CN^2|\log(N\mathfrak{a})|(\log N + |\log(N\mathfrak{a})|) \leq CN^2|\log\mathfrak{a}|^2,$$

$$(4.16)$$

where we used $\mathfrak{a} \leq (4N)^{-1}$ in the second line. Combining (4.14)–(4.16), we conclude that

$$\mathbb{E}\mathbb{1}_\Xi \sum_{i:\sigma_i \in D} \mathcal{O}_{ii} \lesssim |\mathcal{C}_r| r^2 N^2 |\log\mathfrak{a}|^2 \sim N^2 |\mathcal{D}| |\log\mathfrak{a}|^2. \qquad (4.17)$$

This completes the proof of (4.4).

Next, we prove the second estimate (4.5). We take a regular $\mathfrak{a}$-grid $\mathcal{L}$ of points in $\bigcup_{C \in \mathcal{C}_r} \partial C$ so that

$$\max_{C \in \mathcal{C}_r} \max_{z \in \partial C} \min_{w \in \mathcal{L}} |z - w| \leq \mathfrak{a} \quad \text{and} \quad |\mathcal{L}| \sim |\mathcal{C}_r| r\mathfrak{a}^{-1} \sim |\mathcal{D}| r^{-1}\mathfrak{a}^{-1}. \qquad (4.18)$$

We here remark that it is crucial to have the first negative power $\mathfrak{a}^{-1}$ on the right-hand side, not the second; this comes from the fact that $\mathcal{L}$ is a grid on ∂C, not C. Then, on the event

$$\bigcap_{w \in \mathcal{L}} [\lambda_1^w \geq 2\mathfrak{a}], \qquad (4.19)$$

for any $z \in \bigcup_{C \in \mathcal{C}_r} \partial C$ we can find a point $w \in \mathcal{L}$ with $|z - w| \leq \mathfrak{a}$ so that

$$\lambda_1^z \geq \lambda_1^w - \|H_z - H_w\| = \lambda_1^w - |z - w| \geq \mathfrak{a}. \qquad (4.20)$$

Thus, by Theorem 2.10(i) with $k = 1$ and $\mathfrak{a} \leq r \leq (4N)^{-1}$, we obtain

$$\mathbb{P}[\Xi_1^c] = \mathbb{P}\left[\bigcup_{C \in \mathcal{C}_r} [\exists z \in \partial C \text{ such that } \lambda_i^z \leq \mathfrak{a}] \right] \leq \mathbb{P}\left[\bigcup_{w \in \mathcal{L}} [\lambda_1^w \leq 2\mathfrak{a}] \right]$$

$$\lesssim |\mathcal{L}|(N\mathfrak{a})^2(\log N + |\log\mathfrak{a}|) \lesssim N^2|\mathcal{D}|r^{-1}\mathfrak{a}|\log\mathfrak{a}|. \qquad (4.21)$$

This proves (4.5).

We finally prove (4.6). By a union bound we trivially have

$$\mathbb{P}[\Xi_2^c] = \mathbb{P}\left[\bigcup_{C \in \mathcal{C}_r} [|\{i : \sigma_i \in C\}| \geq 2]\right] \leq |\mathcal{C}_r| \max_{C \in \mathcal{C}_r} \mathbb{P}[|\{i : \sigma_i \in C\}| \geq 2]. \tag{4.22}$$

To bound the probability on the right-hand side, we fix a square $C \in \mathcal{C}_r$, take z_C to be the center of C, and label the eigenvalues so that $|\sigma_i - z_C|$ increases in i. Then we have

$$\mathbb{P}[|\{i : \sigma_i \in C\}| \geq 2] \leq \mathbb{P}[|\sigma_2 - z_C| \leq 2r] \leq \mathbb{P}[|\sigma_1 - z_C| \cdot |\sigma_2 - z_C| \leq 4r^2] \leq \mathbb{P}[\lambda_1^{z_C} \lambda_2^{z_C} \leq 4r^2], \tag{4.23}$$

where the last inequality is due to Weyl;

$$\prod_{i=1}^{k} |\sigma_i - z_C| \geq \prod_{i=1}^{k} \lambda_i^{z_C}, \qquad \forall k \in [\![N]\!]. \tag{4.24}$$

Now for a threshold $\kappa > 0$ to be optimized later, by Theorem 2.10.(i) with $k = 1, 2$ we have

$$\begin{aligned}
\mathbb{P}[\lambda_1^{z_C} \lambda_2^{z_C} \leq 4r^2] &= \mathbb{P}[\lambda_1^{z_C} \lambda_2^{z_C} \leq 4r^2, \lambda_2^{z_C} \leq \kappa] + \mathbb{P}[\lambda_1^{z_C} \lambda_2^{z_C} \leq 4r^2, \lambda_2^{z_C} \geq \kappa] \\
&\leq \mathbb{P}[\lambda_2^{z_C} \leq \kappa] + \mathbb{P}[\lambda_1^{z_C} \leq 4r^2/\kappa] \\
&\lesssim \left((N\kappa)^8 + N^2\kappa^{-2}r^4\right) \cdot (\log N + |\log \kappa| + |\log r|)^2.
\end{aligned} \tag{4.25}$$

Note that here we used that $s \in [0, 1]$ in Theorem 2.10 can be extended to $s \in [0, \infty)$ by increasing C slightly. Taking the optimal choice $\kappa = N^{-3/5} r^{2/5}$, we have $\mathbb{P}[\lambda_1^{z_C} \lambda_2^{z_C} \leq 4r^2] \lesssim (Nr)^{16/5} |\log r|^2$. Therefore we conclude from (4.22) that

$$\mathbb{P}[\Xi_2^c] \lesssim |\mathcal{C}_r|(Nr)^{16/5}|\log r|^2 \sim N^{16/5}|\mathcal{D}|r^{6/5}|\log r|^2, \tag{4.26}$$

completing the proof of Lemma 4.1 in the complex case.

We next show how to modify the proof for the real case. Firstly in (4.16), we use Theorem 2.13 instead of Theorem 2.10 to get

$$\mathbb{E}\mathbb{1}_{\Xi_1}\left\|\frac{1}{X + A - z}\right\|^2 \lesssim N^2|\log \mathfrak{a}|^2 \frac{\|A\|^2 + 1}{\lambda_1(\mathrm{Im}[A - z])},$$

which leads to (4.7) via the same argument as in the complex case.

Secondly in (4.21), we again use the same replacement to prove (4.8);

$$\mathbb{P}[\Xi_1^c] \leq \mathbb{P}\left[\bigcup_{w \in \mathcal{L}} [\lambda_1^w \leq 2\mathfrak{a}]\right] \lesssim |\mathcal{L}|(N\mathfrak{a})^2 \frac{(|\log \mathfrak{a}| + \log N)}{y} \lesssim N^2 r^{-1} \mathfrak{a} \frac{|\log \mathfrak{a}|}{y}|\mathcal{D}|. \tag{4.27}$$

Lastly in (4.25), we use Theorems 2.10 (ii) and 2.13 to get

$$\mathbb{P}[\lambda_1^{z_C} \lambda_2^{z_C} \leq 4r^2] \lesssim \left((N\kappa)^4 + \frac{1}{y}N^2\kappa^{-2}r^4\right)(\log N + |\log \kappa| + |\log r|)^2. \tag{4.28}$$

After optimizing κ we obtain

$$\mathbb{P}[\lambda_1^{z_C} \lambda_2^{z_C} \le 4r^2] \lesssim \frac{(Nr)^{8/3}}{y^{2/3}} (\log N + |\log r| + |\log y|)^2, \tag{4.29}$$

which gives

$$\mathbb{P}[\Xi_2^c] \lesssim y^{-2/3} N^{8/3} |\mathcal{D}| r^{2/3} (|\log r| + |\log y|)^2. \tag{4.30}$$

This concludes the proof of Lemma 4.1. $\qquad\qquad\qquad\qquad\qquad\square$

5 | PROOF OF THEOREM 2.11

In the rest of the paper we prove results in Section 2.2. We start with the proof of Theorem 2.11, since it best represents the common core ideas.

Recall that all of Theorems 2.10–2.13 concern singular values of $X + A$, where X is a regular matrix and A is a deterministic shift sometimes with additional restrictions. We denote the Hermitization of $X + A$ by H and the resolvent $(H - w)^{-1}$ by $G(w)$ as in (3.1). The spectral parameter of G is always taken to be $w = i\eta$, where

$$\eta \equiv \eta(s) := \frac{s}{N} \tag{5.1}$$

with the same parameter $s > 0$ in Theorems 2.10–2.13. Henceforth we often abbreviate $G \equiv G(i\eta)$. As seen below, all arguments and inequalities for the complex case is also used for the real case with some changes to exponents. In order to unify the presentation, we introduce the exponent β defined by

$$\beta := \begin{cases} 1 & \text{if } X \text{ is real}, \\ 2 & \text{if } X \text{ is complex}. \end{cases} \tag{5.2}$$

Before commencing the proofs, we introduce a few notations and observations regarding minors of matrices, that are used throughout the rest of the paper. Firstly, for each subset[5] $\mathcal{I} \subset [\![N]\!]$ we define the matrix $J^{(\mathcal{I})} \in \mathbb{C}^{([\![N]\!] \setminus \mathcal{I}) \times [\![N]\!]}$ by

$$(J^{(\mathcal{I})})_{ij} := \delta_{ij} \mathbb{1}(j \notin \mathcal{I}) \tag{5.3}$$

Note that $J^{(\mathcal{I})}$ acts as a $\mathcal{I}$-shift operator; for each $A \in \mathbb{C}^{[\![N]\!] \times [\![N]\!]}$, taking product $J^{(\mathcal{I})} A \in \mathbb{C}^{([\![N]\!] \setminus \mathcal{I}) \times [\![N]\!]}$ removes the j-th rows of A for $j \in \mathcal{I}$. Secondly, we write $H^{(\mathcal{I})}$ and $G^{(\mathcal{I})}$ for the Hermitization of $J^{(\mathcal{I})}(X + A) \in \mathbb{C}^{([\![N]\!] \setminus \mathcal{I}) \times [\![N]\!]}$ and its resolvent, that is,

$$H^{(\mathcal{I})} := \begin{pmatrix} 0 & J^{(\mathcal{I})}(X+A) \\ (X+A)^*(J^{(\mathcal{I})})^* & 0 \end{pmatrix} \in \mathbb{C}^{(([\![N]\!] \setminus \mathcal{I}) \cup [\![N+1,2N]\!])^2},$$

$$G^{(\mathcal{I})} \equiv G^{(\mathcal{I})}(i\eta) := (H^{(\mathcal{I})} - i\eta)^{-1}. \tag{5.4}$$

---

[5] The index set $\mathcal{I}$ is unrelated to the domain $\mathcal{I} \subset \mathbb{R}$ in Theorem 2.9; $\mathcal{I}$ always denote an index set in what follows.

Note that $H^{(\mathcal{I})}$ is obtained from $H^{(\emptyset)} = H$ by removing the j-th rows and columns for $j \in \mathcal{I}$. Thirdly, we write the spectral decomposition of $H^{(\mathcal{I})}$ as follows;

$$H^{(\mathcal{I})} = \sum_{i \in [\![-(N-|\mathcal{I}|),-1]\!] \cup [\![1,N]\!]} \lambda_i^{(\mathcal{I})} \begin{pmatrix} \boldsymbol{u}_i^{(\mathcal{I})} \\ \boldsymbol{v}_i^{(\mathcal{I})} \end{pmatrix} \begin{pmatrix} \boldsymbol{u}_i^{(\mathcal{I})} \\ \boldsymbol{v}_i^{(\mathcal{I})} \end{pmatrix}^*, \qquad \boldsymbol{u}_i^{(\mathcal{I})} \in \mathbb{C}^{N-|\mathcal{I}|}, \, \boldsymbol{v}_i^{(\mathcal{I})} \in \mathbb{C}^N \tag{5.5}$$

where the eigenvalues are ordered increasingly. Note that the superscript $(\emptyset)$ is superfluous, for example $\lambda_i^{(\emptyset)} = \lambda_i$.

Now we list the key observations under these notations. Almost surely, the matrix $H^{(\mathcal{I})}$ has rank $2(N - |\mathcal{I}|)$, and by the block structure of $H^{(\mathcal{I})}$ we have

$$0 = \lambda_1^{(\mathcal{I})} = \cdots = \lambda_{|\mathcal{I}|}^{(\mathcal{I})} < \lambda_{|\mathcal{I}|+1}^{(\mathcal{I})} < \cdots < \lambda_N^{(\mathcal{I})} \quad \text{and} \quad \lambda_{-i}^{(\mathcal{I})} = -\lambda_{i+|\mathcal{I}|}^{(\mathcal{I})} \quad \forall i \in [\![N-|\mathcal{I}|]\!], \tag{5.6}$$

where the strict inequalities are due to the fact that $H^{(\mathcal{I})}$ has simple spectrum. Furthermore, for any $\mathcal{I} \subset [\![N]\!]$ and $i \in [\![N]\!] \setminus \mathcal{I}$, since $H^{(\mathcal{I} \cup \{i\})}$ is a principal minor of $H^{(\mathcal{I})}$, Cauchy interlacing theorem implies that

$$\lambda_i^{(\mathcal{I} \cup \{i\})} \leq \lambda_i^{(\mathcal{I})} \leq \lambda_{i+1}^{(\mathcal{I} \cup \{i\})} \qquad \forall i \in [\![N]\!] \setminus \mathcal{I}. \tag{5.7}$$

For the eigenvectors, we have

$$\boldsymbol{v}_{-i}^{(\mathcal{I})} = \boldsymbol{v}_{i+|\mathcal{I}|}^{(\mathcal{I})}, \quad \boldsymbol{u}_{-i}^{(\mathcal{I})} = -\boldsymbol{u}_{i+|\mathcal{I}|}^{(\mathcal{I})}, \quad \text{and} \quad \|\boldsymbol{v}_i^{(\mathcal{I})}\|^2 = \frac{1}{2} = \|\boldsymbol{u}_i^{(\mathcal{I})}\|^2 \quad \text{for } i \in [\![N-|\mathcal{I}|]\!],$$

$$\boldsymbol{u}_i^{(\mathcal{I})} = 0 \quad \text{and} \quad \|\boldsymbol{v}_i^{(\mathcal{I})}\|^2 = 1 \qquad\qquad \text{for } i \in [\![|\mathcal{I}|]\!]. \tag{5.8}$$

In particular, we can write the spectral decompositions

$$(X + A)^* \left(J^{(\mathcal{I})}\right)^* J^{(\mathcal{I})}(X + A) = \sum_{i \in [\![|\mathcal{I}|]\!]} 0 \cdot \boldsymbol{v}_i^{(\mathcal{I})} \left(\boldsymbol{v}_i^{(\mathcal{I})}\right)^* + \sum_{i \in [\![|\mathcal{I}|+1,N]\!]} \left(\lambda_i^{(\mathcal{I})}\right)^2 \left(\sqrt{2}\boldsymbol{v}_i^{(\mathcal{I})}\right) \left(\sqrt{2}\boldsymbol{v}_i^{(\mathcal{I})}\right)^*,$$

$$J^{(\mathcal{I})}(X + A)(X + A)^* (J^{(\mathcal{I})})^* = \sum_{i \in [\![|\mathcal{I}|+1,N]\!]} \left(\lambda_i^{(\mathcal{I})}\right)^2 \left(\sqrt{2}\boldsymbol{u}_i^{(\mathcal{I})}\right) \left(\sqrt{2}\boldsymbol{u}_i^{(\mathcal{I})}\right)^*. \tag{5.9}$$

Finally, since the entries of X are independent, the minor $H^{(\mathcal{I})}$ and the j-th rows of X for $j \in \mathcal{I}$ are independent. To be precise, we define the σ-algebras and the corresponding conditional expectations

$$\mathcal{F}^{(\mathcal{I})} := \sigma(\{X_{ab} : a \in [\![N]\!] \setminus \mathcal{I}, b \in [\![N]\!]\}), \qquad \mathbb{E}^{(\mathcal{I})}[\cdot] := \mathbb{E}[\cdot | \mathcal{F}^{(\mathcal{I})}]. \tag{5.10}$$

Then all of $H^{(\mathcal{I})}$, $G^{(\mathcal{I})}$, and $(\lambda_i^{(\mathcal{I})}, \boldsymbol{u}_i^{(\mathcal{I})}, \boldsymbol{v}_i^{(\mathcal{I})})_{i \in [\![N]\!]}$ are measurable with respect to $\mathcal{F}^{(\mathcal{I})}$ and thus independent of the $\mathcal{I}$-rows $\{X_{ab} : a \in \mathcal{I}, b \in [\![N]\!]\}$ of X.

Proofs of Theorems 2.10 an 2.11 both involve an induction over minors $H^{(\mathcal{I})}$ as $|\mathcal{I}|$ runs through $[\![k]\!]$, with the same k as in their statements. To make this rigorous, we define for each $\mathcal{I} \subset [\![N]\!]$ with $|\mathcal{I}| \leq k$ the events[6]

$$\Xi_k := [\lambda_k \leq \eta], \qquad \Xi_k^{(\mathcal{I})} := \left[\lambda_k^{(\mathcal{I})} \leq \eta\right], \tag{5.11}$$

---

[6] The events Ξ_k's here are completely unrelated to those in (4.3).

where we recall that $\eta = s/N$. Thus the goal of these two theorems is to bound $\mathbb{P}[\Xi_k]$ by $(N\eta)^{\beta k^2}$ with the exponent β from (5.2). Note that $\Xi_k^{(I)} \in \mathcal{F}^{(I)}$ and that, due to $\lambda_k^{(I\cup\{i\})} \leq \lambda_k^{(I)}$ from (5.7),

$$\Xi_k \equiv \Xi_k^{(\emptyset)} \subset \Xi_k^{(I_1)} \subset \cdots \subset \Xi_k^{(I_k)} \tag{5.12}$$

holds true for any increasing sequence of index sets $(I_j)_{j\in[\![k]\!]}$. Also note that if $|I_k| = k$, then the last event $\Xi_k^{(I_k)}$ has probability one since $\lambda_k^{(I_k)} \equiv 0$.

As we will see shortly, Theorems 2.10 and 2.11 both follow from an induction over the chain (5.12) for sequences $(I_j)_{j\in[\![k]\!]}$ with $|I_j| = j$, gaining a factor of $(N\eta)^{\beta k}$ at each step. To be precise, we prove

$$\mathbb{E}[\boldsymbol{\mathcal{X}}_I \mathbb{1}_{\Xi_k^{(I)}}] \leq (N\eta)^{\beta k} \frac{1}{N - |I|} \sum_{i\in[\![N]\!]\setminus I} \mathbb{E}[\boldsymbol{\mathcal{X}}_{I\cup\{i\}} \mathbb{1}_{\Xi_k^{(I\cup\{i\})}}], \quad \forall \eta \in [0, N^{-1}], \ \forall I \subset [\![N]\!], |I| \leq k-1, \tag{5.13}$$

for a collection of positive random variables $(\boldsymbol{\mathcal{X}}_I)_{I\subset[\![N]\!]}$ such that $\boldsymbol{\mathcal{X}}_I \in \mathcal{F}^{(I)}$ and $\boldsymbol{\mathcal{X}}_\emptyset \equiv 1$. Once we have (5.13), then it immediately follows that

$$\mathbb{P}[\Xi_k] = \mathbb{E}[\boldsymbol{\mathcal{X}}_\emptyset \mathbb{1}_{\Xi_k^{(0)}}] \leq (N\eta)^{\beta k} \frac{1}{N} \sum_{i\in[\![N]\!]} \mathbb{E}[\boldsymbol{\mathcal{X}}_{\{i\}} \mathbb{1}_{\Xi_k^{(\{i\})}}] \leq (N\eta)^{2\beta k} \frac{2}{N(N-1)} \sum_{i_1 \neq i_2 \in [\![N]\!]} \mathbb{E}\boldsymbol{\mathcal{X}}_{\{i_1, i_2\}} \mathbb{1}_{\Xi_k^{(\{i_1, i_2\})}} \leq \cdots$$

$$\leq (N\eta)^{\beta k^2} \frac{(N-k)! k!}{N!} \sum_{I\subset[\![N]\!], |I|=k} \mathbb{E}[\boldsymbol{\mathcal{X}}_I \mathbb{1}_{\Xi_k^{(I)}}] \leq (N\eta)^{\beta k^2} \max_{I\subset[\![N]\!], |I|=k} \mathbb{E}[\boldsymbol{\mathcal{X}}_I], \tag{5.14}$$

so that the only remaining thing is to estimate $\mathbb{E}[\boldsymbol{\mathcal{X}}_I]$ for $|I| = k$, which becomes the correction. The exact form of $\boldsymbol{\mathcal{X}}_I$ in (5.13) varies depending on the goal, and in particular $\boldsymbol{\mathcal{X}}_I$'s are deterministic in the proof of Theorem 2.10.

Although all discussions above are written in terms of a general index set I for the sake of rigor, in the actual proof we often choose $I = [\![j]\!] \subset [\![k]\!]$ for clarity. To further simplify the notation, we write the superscript (j) instead of $([\![j]\!])$ for an integer j, for example $\lambda^{(j)} \equiv \lambda^{([\![j]\!])}$, $\Xi_k^{(0)} \equiv \Xi_k^{([\![0]\!])} = \Xi_k^{(\emptyset)}$, et cetera.

We now present the two inputs for the proof of Theorem 2.11, whose proofs are postponed to the end of this section. Note that the first input, Proposition 5.1, applies to a general regular matrix X but assumes that A is real when X is.

Proposition 5.1. *Let $k, N \in \mathbb{N}$ with $N \geq 3k$, $\mathbb{K} = \mathbb{R}$ or $\mathbb{C}$, $X \in \mathbb{K}^{N\times N}$ be a regular matrix, and $A \in \mathbb{K}^{N\times N}$ be deterministic. Then there exists a constant $C \equiv C(k, \mathfrak{b})$ such that*

$$\mathbb{E}\left(1 + N\lambda_m^{(I)}\right)^n \mathbb{1}_{\Xi_k^{(I)}} \leq C(N\eta)^{\beta k} \frac{1}{N - |I|} \sum_{i\in[\![N]\!]\setminus I} \mathbb{E}\left(1 + N\lambda_{m+1}^{(I\cup\{i\})}\right)^{n+\beta k} \mathbb{1}_{\Xi_k^{(I\cup\{i\})}} \quad \forall \eta \in [0, N^{-1}] \tag{5.15}$$

holds for any $I \subset [\![N]\!]$ with $|I| \leq k-1$, $m \in [\![2k, N-1]\!]$, and $n \geq 0$.

Lemma 5.2. *Let X be a regular i.i.d. matrix, $A \in \mathbb{C}^{N\times N}$ with $\|A\| \leq \mathfrak{K}$, and for each $z \in \mathbb{C}$ let $\lambda_1^z \leq \cdots \leq \lambda_N^z$ be singular values of $X + A - z$. For each fixed $\delta, \tau > 0$, there exists a constant $C_2 \equiv C_2(k, \delta, \tau, \mathfrak{m}, \mathfrak{K})$ such that*

$$\mathbb{E}[(1 + N\lambda_{3k}^z)^{2k^2}] \leq C_2 N^\delta \tag{5.16}$$

holds for all $N \geq 3k$ and $|z| \leq 1 - \tau$.

We finally prove Theorem 2.11 using these inputs.

Proof of Theorem 2.11. Recall the definitions of η and β respectively from (5.1) and (5.2). Define

$$\mathfrak{X}_{\mathcal{I}} := C^{|\mathcal{I}|} \left(1 + N \lambda_{2k+|\mathcal{I}|}^{(\mathcal{I})} \right)^{\beta |\mathcal{I}| k} \tag{5.17}$$

where C is the constant from Proposition 5.1. Then, since we assumed $z \in \mathbb{R}$ in the real case, we may apply (5.15) with A replaced by $A - z$. This immediately proves (5.13) and hence, by (5.14) we have

$$\mathbb{P}[N \lambda_k^z \leq s] \leq s^{\beta k^2} \max_{\mathcal{I} \subset [\![N]\!], |\mathcal{I}| = k} \mathbb{E}[\mathfrak{X}_{\mathcal{I}}] \leq C^k s^{\beta k^2} \mathbb{E}\left[(1 + N \lambda_{3k}^z)^{\beta k^2} \right], \tag{5.18}$$

where we used $\lambda_{3k}^{(\mathcal{I})} \leq \lambda_{3k}$ from (5.7) in the last inequality. In particular, by taking $C_1 = C^k$, this proves the first inequalities in (2.23) and (2.24).

Finally, since

$$\mathbb{E}(1 + N \lambda_{3k})^{k^2} \leq \left(\mathbb{E}(1 + N \lambda_{3k})^{2k^2} \right)^{1/2},$$

substituting (5.16) into (5.18) concludes the proof of Theorem 2.11. $\qquad \square$

Proof of Proposition 5.1. First of all, we prove that one may take $\mathcal{I} = [\![j - 1]\!]$ for some $j \in [\![k]\!]$ without loss of generality. To this end, we take a general $\mathcal{I} \subset [\![N]\!]$ with $|\mathcal{I}| \leq k - 1$ and assume that Proposition 5.1 is valid for $[\![|\mathcal{I}|]\!]$. Consider a permutation matrix $P_{\mathcal{I}}$ that maps $\mathcal{I}$ into $[\![|\mathcal{I}|]\!]$, so that $J^{(|\mathcal{I}|)} P_{\mathcal{I}} = J^{(\mathcal{I})}$. Then it is easy to see that $P_{\mathcal{I}} X$ is regular with the same density bound $\mathfrak{b}$ if and only if X is, so that we may apply Proposition 5.1 to $P_{\mathcal{I}}(X + A)$. Noticing that the constant C_1 in (5.15) depends only on $(k, \mathfrak{b})$ and that

$$\lambda_m^{(|\mathcal{I}|)}(P_{\mathcal{I}}(X + A)) = \lambda_m^{(\mathcal{I})}, \tag{5.19}$$

where the left-hand side stands for the m-th smallest singular value of $J^{(|\mathcal{I}|)} P_{\mathcal{I}}(X + A)$, the result for general $\mathcal{I}$ immediately follows.

Therefore in what follows we take $\mathcal{I} = [\![j - 1]\!]$ for some $j \in [\![k]\!]$, and all superscripts $(\mathcal{I})$ are replaced by $(j - 1)$. Next, we write

$$\text{Im} \sum_{i \in [\![j, N]\!]} G_{ii}^{(j-1)}(i\eta) = \text{Im} \, \text{Tr} \frac{i\eta}{J^{(j-1)}(X + A)(X + A)^* J^{(j-1)*} + \eta^2} = \sum_{i \in [\![j, N]\!]} \frac{\eta}{|\lambda_i^{(j-1)}|^2 + \eta^2}, \tag{5.20}$$

where we recall that $H^{(j-1)}$ and $G^{(j-1)}$ are indexed by $[\![j, 2N]\!]^2$. Recalling the definition of $\Xi_k^{(j-1)}$ from (5.11), we have

$$\mathbb{1}_{\Xi_k^{(j-1)}} \text{Im} \sum_{i \in [\![j, N]\!]} G_{ii}^{(j-1)} \geq \mathbb{1}_{\Xi_k^{(j-1)}} \sum_{i=j}^{k} \frac{\eta}{|\lambda_i^{(j-1)}|^2 + \eta^2} \geq \frac{1}{2\eta}(k - j + 1) \mathbb{1}_{\Xi_k^{(j-1)}} \geq \frac{1}{2\eta} \mathbb{1}_{\Xi_k^{(j-1)}}, \tag{5.21}$$

and hence, raising it to the (βk)-th power (recall that $\beta = 1$ or 2 depending on real or complex X),

$$\mathbb{1}_{\Xi_k^{(j-1)}} \leq (2N\eta)^{\beta k} \left(\frac{1}{N-j+1} \sum_{i \in [\![j,N]\!]} \operatorname{Im} G_{ii}^{(j-1)} \right)^{\beta k} \mathbb{1}_{\Xi_k^{(j-1)}}. \tag{5.22}$$

Multiplying both sides of (5.22) by the factor $(1 + N\lambda_m^{(j-1)})^n$ and using Jensen's inequality, we have

$$\mathbb{E}\mathbb{1}_{\Xi_k^{(j-1)}}\left(1 + N\lambda_m^{(j-1)}\right)^n \leq \frac{1}{N-j+1}(2N\eta)^{\beta k}\mathbb{E}\mathbb{1}_{\Xi_k^{(j-1)}}\left(1 + N\lambda_m^{(j-1)}\right)^n \sum_{i \in [\![j,N]\!]} |G_{ii}^{(j-1)}|^{\beta k}. \tag{5.23}$$

Then, for each $i \in [\![j,N]\!]$ we use that $\lambda_m^{(j-1)} \leq \lambda_{m+1}^{([\![j-1]\!]\cup\{i\})}$ from (5.7) and $\Xi_k^{(j-1)} \subset \Xi_k^{([\![j-1]\!]\cup\{i\})}$ to estimate the i-th summand in the right-hand side of (5.23) as

$$\mathbb{E}\mathbb{1}_{\Xi_k^{(j-1)}}\left(1 + N\lambda_m^{(j-1)}\right)^n |G_{ii}^{(j-1)}|^{\beta k} \leq \mathbb{E}\mathbb{1}_{\Xi_k^{([\![j-1]\!]\cup\{i\})}}\left(1 + N\lambda_{m+1}^{([\![j-1]\!]\cup\{i\})}\right)^n \mathbb{E}^{([\![j-1]\!]\cup\{i\})}|G_{ii}^{(j-1)}|^{\beta k}, \tag{5.24}$$

where we recall the definition of $\mathbb{E}^{(I)}$ from (5.10) and that $\lambda_{m+1}^{(I)}, \Xi_k^{(I)}$ are measurable with respect to $\mathcal{F}^{(I)}$. Substituting (5.24) into (5.23) and then comparing with the final goal (5.15), we find that it suffices to prove for all $i \in [\![j,N]\!]$ that

$$\mathbb{1}_{\Xi_k^{([\![j-1]\!]\cup\{i\})}}\mathbb{E}^{([\![j-1]\!]\cup\{i\})}|G_{ii}^{(j-1)}|^{\beta k} \leq C\mathbb{1}_{\Xi_k^{([\![j-1]\!]\cup\{i\})}}\left(1 + N\lambda_{2k+1}^{([\![j-1]\!]\cup\{i\})}\right)^{\beta k}, \tag{5.25}$$

for a constant C depending only on k, $\mathfrak{b}$. At this point we may further assume $i = j$ without loss of generality, using the same argument as in (5.19) involving permutation matrices. Finally, recalling the assumption $m \geq 2k$, it only remains to prove

$$\mathbb{1}_{\Xi_k^{(j)}}\mathbb{E}^{(j)}|G_{jj}^{(j-1)}|^{\beta k} \leq C\mathbb{1}_{\Xi_k^{(j)}}\left(1 + N\lambda_{2k+1}^{(j)}\right)^{\beta k}. \tag{5.26}$$

Next, we decouple the j-th row of X from $G_{jj}^{(j-1)}$. Applying Schur complement formula with respect to the (j,j)-th entry of $H^{(j-1)}$ gives

$$\frac{1}{G_{jj}^{(j-1)}} = -\mathrm{i}\eta - \sum_{k,\ell \in [\![j,2N]\!]} H_{jk}^{(j-1)}G_{k\ell}^{(j)}H_{\ell j}^{(j-1)} = -\mathrm{i}\eta - \sum_{k,\ell \in [\![N]\!]} (X+A)_{jk}G_{k+N,\ell+N}^{(j)}(\overline{X}+\overline{A})_{j\ell}. \tag{5.27}$$

On the other hand, another application of Schur's complement with respect to the bottom right $(N \times N)$ block of $(H^{(j)} - \mathrm{i}\eta)$ gives

$$G_{k+N,\ell+N}^{(j)} = \mathrm{i}\eta\left((X+A)^*J^{(j)*}J^{(j)}(X+A) + \eta^2\right)_{k\ell}^{-1}, \qquad k,\ell \in [\![N]\!], \tag{5.28}$$

so that (5.9) gives

$$\frac{1}{G_{jj}^{(j-1)}} = -\mathrm{i}\eta - \mathrm{i}\eta \sum_{i=1}^N \frac{(2 - \mathbb{1}(i \leq j))}{(\lambda_i^{(j)})^2 + \eta^2}|e_j^*(X+A)v_i^{(j)}|^2 = -\mathrm{i}\eta - \mathrm{i}\sum_{i=1}^N c_i^{(j)}|w_i^{(j)}|^2, \tag{5.29}$$

where we defined (recall that $\|\boldsymbol{v}_i^{(j)}\|^2$ is equal to 1 for $i \leq j$ and $1/2$ for $i > j$)

$$c_i^{(j)} := \frac{N\eta}{(N\lambda_i^{(j)})^2 + (N\eta)^2}, \qquad \boldsymbol{w}^{(j)} := \left(w_1^{(j)}, \ldots, w_N^{(j)}\right) := U^{(j)}\left(\boldsymbol{b}^{(j)} + \boldsymbol{a}^{(j)}\right), \qquad \boldsymbol{b}^{(j)} := \sqrt{N}X^*\boldsymbol{e}_j,$$
(5.30)

$$\boldsymbol{a}^{(j)} := \sqrt{N}A^*\boldsymbol{e}_j, \qquad U^{(j)} := \left(\boldsymbol{v}_1^{(j)} \quad \cdots \quad \boldsymbol{v}_j^{(j)} \quad \sqrt{2}\boldsymbol{v}_{j+1}^{(j)} \quad \cdots \quad \sqrt{2}\boldsymbol{v}_N^{(j)}\right)^*.$$

While (5.29) is an identity, we do not need all summands on the right-hand side but only that

$$|G_{ii}^{(j-1)}| = \left(\eta + \sum_{i=1}^{N} c_i^{(j)}|w_i^{(j)}|^2\right)^{-1} \leq \left(\sum_{i \in [\![2k+1]\!]} c_i^{(j)}|w_i^{(j)}|^2\right)^{-1}.$$
(5.31)

Note that $c_i^{(j)}$ and $U^{(j)}$ are measurable with respect to $\mathcal{F}^{(j)}$ defined in (5.10) and that $\boldsymbol{b}^{(j)}$ is exactly the j-th row of X with a deterministic shift, hence $c_i^{(j)}$ and $U^{(j)}$ are independent of $\boldsymbol{b}^{(j)}$. Also the matrix $(U^{(j)})^*$, and hence $U^{(j)}$, is unitary due to (5.9). Here we emphasize that, on the event $\Xi_k^{(j)} = [\lambda_k^{(j)} \leq \eta]$, sizes of $c_i^{(j)}$'s are roughly

$$\frac{1}{2N\eta} \leq c_i^{(j)} \leq \frac{1}{N\eta} \quad \text{for } i \leq k, \quad \text{and} \quad c_i^{(j)} \sim (N\eta) \quad \text{for } i > k$$
(5.32)

since $N\lambda_i^{(j)} \sim 1$ for $i > k$. Notice that the first inequality in (5.32) is not a heuristic, but deterministically true on $\Xi_k^{(j)}$.

Substituting (5.31) into the left-hand side of (5.26), we find that it suffices to prove

$$\mathbb{1}_{\Xi_k^{(j)}}\mathbb{E}^{(j)}\left(\sum_{i \in [\![2k+1]\!]} c_i^{(j)}|w_i^{(j)}|^2\right)^{-\beta k} \leq C\mathbb{1}_{\Xi_k^{(j)}}\left(1 + N\lambda_{2k+1}^{(j)}\right)^{\beta k}$$
(5.33)

for a constant C depending only on k and $\mathfrak{b}$. The inequality (5.33) follows from the following technical lemma, whose proof is presented after completing that of Theorem 2.11.

Lemma 5.3. *Let $k, N \in \mathbb{N}$ with $2k + 1 \leq N$, $\mathbb{K} = \mathbb{R}$ or $\mathbb{C}$, $\mathfrak{b} > 0$, and $\boldsymbol{b} \in \mathbb{K}^N$ be a random vector with independent components. Assume that each component of $\boldsymbol{b}$ (Re $\boldsymbol{b}$ and Im $\boldsymbol{b}$, resp.) has a density bounded by $\mathfrak{b}$ if $\mathbb{K} = \mathbb{R}$ (if $\mathbb{K} = \mathbb{C}$, resp.). Then there exists a constant $\mathfrak{C}_1 \equiv \mathfrak{C}_1(k, \mathfrak{b}) > 0$ such that the following holds for any positive sequence $(c_i)_{i \in [\![N]\!]}$ and a unitary matrix $U \in \mathbb{K}^{N \times N}$;*

$$\mathbb{E}\left(\sum_{i \in [\![2k+1]\!]} c_i|\boldsymbol{e}_i^*U\boldsymbol{b}|^2\right)^{-\beta k} \leq 1 + \mathfrak{C}_1 \prod_{i=1}^{k} c_i^{-\beta/2} \cdot \prod_{i=k+1}^{2k+1} c_i^{-\beta k/(2(k+1))}.$$
(5.34)

Notice that Lemma 5.3 assumes that U is real orthogonal when $\mathbb{K} = \mathbb{R}$. Assuming Lemma 5.3 is valid, we complete the proof of Theorem 2.11. We aim at applying Lemma 5.3 with the choices $(\boldsymbol{b}, c_i, U) = (\boldsymbol{b}^{(j)} + \boldsymbol{a}^{(j)}, c_i^{(j)}, U^{(j)})$. First of all, recalling that A is real when X is, the matrix $(X + A)^* \in \mathbb{K}^{N \times N}$ is regular so that each component of the vector $(\boldsymbol{b}^{(j)} + \boldsymbol{a}^{(j)}) \in \mathbb{K}^N$ has a density bounded by $\mathfrak{b}$. Secondly, again due to $A \in \mathbb{K}^{N \times N}$, the singular vectors $(\boldsymbol{v}_i^{(j)})_{i \in [\![N]\!]}$ are in $\mathbb{K}^N$ so that $U^{(j)} \in \mathbb{K}^{N \times N}$. Lastly, since the constant $\mathfrak{C}_1$ in (5.34) is uniform over all (c_i) and U, we may as well take them to be random as long as they are independent of $\boldsymbol{b}$. More precisely, for any σ-algebra $\mathcal{F}$

independent of $\boldsymbol{b}$ such that $\sigma((c_i), U) \subset \mathcal{F}$, we may also replace $\mathbb{E}$ on the left-hand side of (5.34) with the conditional expectation $\mathbb{E}[\cdot|\mathcal{F}]$. In particular for $\boldsymbol{b} = \boldsymbol{b}^{(j)}$ we may replace $\mathbb{E}$ with $\mathbb{E}^{(j)}$.

Thus we may apply Lemma 5.3 to the left-hand side of (5.33), so that

$$\mathbb{E}^{(j)}\left(\sum_{i\in[\![2k+1]\!]} c_i^{(j)}|w_i^{(j)}|^2\right)^{-\beta k} \leq 1 + \mathfrak{C}_1 \prod_{i=1}^{k}\left(c_i^{(j)}\right)^{-\beta/2} \prod_{i=k+1}^{2k+1}\left(c_i^{(j)}\right)^{-\beta k/(2(k+1))}$$

$$\leq 1 + \mathfrak{C}_1\left(c_k^{(j)}c_{2k+1}^{(j)}\right)^{-\beta k/2}, \tag{5.35}$$

in the second line we used that $c_i^{(j)}$ is decreasing in $i \in [\![N]\!]$. Then recalling $\Xi_k^{(j)} = [\lambda_k^{(j)} \leq \eta]$ we have

$$\mathbb{1}_{\Xi_k^{(j)}} \frac{1}{c_k^{(j)}c_{2k+1}^{(j)}} = \mathbb{1}_{\Xi_k^{(j)}}\left((N\eta)^2 + \left(N\lambda_{2k+1}^{(j)}\right)^2\right) \cdot \frac{\left(N\lambda_k^{(j)}\right)^2 + (N\eta)^2}{(N\eta)^2} \leq 2\mathbb{1}_{\Xi_k^{(j)}}\left(1 + (N\lambda_{2k+1}^{(j)})^2\right) \tag{5.36}$$

where we used $N\eta = s \leq 1$. Therefore we conclude

$$\mathbb{1}_{\Xi_k^{(j)}}\mathbb{E}^{(j)}\left(\sum_{i\in[\![N]\!]} c_i^{(j)}|w_i^{(j)}|^2\right)^{-\beta k} \leq \mathbb{1}_{\Xi_k^{(j)}}\left(1 + \mathfrak{C}_1\left(2 + 2(N\lambda_{2k+1}^{(j)})^2\right)^{\beta k/2}\right) \leq C'\mathbb{1}_{\Xi_k^{(j)}}\left(1 + N\lambda_{2k+1}^{(j)}\right)^{\beta k},$$

by taking a suitable constant C' that still depends only on $(k, \mathfrak{b})$. This finishes the proof of (5.33), concluding that of Proposition 5.1. $\square$

Proof of Lemma 5.3. We define probability measures μ and μ_p on $\mathbb{K}^N$ and $\mathbb{K}^p$ to be the laws of $U\boldsymbol{b}$ and its first p coordinates. More specifically, for any suitable test functions $f : \mathbb{K}^N \to \mathbb{C}$ and $f_p : \mathbb{K}^p \to \mathbb{C}$ we define

$$\int_{\mathbb{K}^N} f\mathrm{d}\mu := \int_{\mathbb{K}^N} f(U\boldsymbol{b})h(\boldsymbol{b})\mathrm{d}^{\beta N}\boldsymbol{b}, \qquad \int_{\mathbb{K}^p} f_p\mathrm{d}\mu_p := \int_{\mathbb{K}^N} f_p(w_1, \dots, w_p)\mathrm{d}\mu(\boldsymbol{w}) \tag{5.37}$$

where we abbreviated $h(\boldsymbol{b}) := \prod_{i=1}^{N} h_i(b_i)$. With μ_{2k+1}, we may rewrite the left-hand side of (5.34) as

$$\mathbb{E}\left(\sum_{i\in[\![2k+1]\!]} c_i|\boldsymbol{e}_i^*U\boldsymbol{b}|^2\right)^{-\beta k} = \int_{\mathbb{K}^{2k+1}}\left(\sum_{i=1}^{2k+1} c_i|w_i|^2\right)^{-\beta k}\mathrm{d}\mu_{2k+1}(\boldsymbol{w}). \tag{5.38}$$

For a threshold $\kappa > 0$ that will be optimized afterwards, we divide the integration in (5.38) by inserting the following factor;

$$1 = \mathbb{1}_{S_1^c} + \mathbb{1}_{S_1 \cap S_2^c} + \mathbb{1}_{S_1 \cap S_2}, \tag{5.39}$$

$$S_1 := \left\{\boldsymbol{w} \in \mathbb{K}^{2k+1} : \sum_{i=1}^{k} c_i|w_i|^2 \leq 1\right\}, \qquad S_2 := \left\{\boldsymbol{w} \in \mathbb{K}^{2k+1} : \sum_{i=k+1}^{2k+1} c_i|w_i|^2 \leq \kappa^2\right\}.$$

The integral in (5.38) corresponding to the first regime S_1^c is simply bounded by 1 as μ_{2k+1} is a probability measure and the integrand is bounded by one. For the second domain $S_1 \cap S_2^c$, denoting

the L^∞ norm of the Lebesgue density of a measure ν by $\|\nu\|_\infty$, we have

$$\int_{\mathbb{K}^{2k+1}} \frac{\mathbb{1}_{S_1 \cap S_2^c}(\boldsymbol{w})}{\left(\sum_{i=1}^{2k+1} c_i |w_i|^2\right)^{\beta k}} \mathrm{d}\mu_{2k+1}(\boldsymbol{w}) \le \int_{\mathbb{K}^k} \frac{\mathbb{1}\left(\sum_{i=1}^k c_i |w_i|^2 \le 1\right)}{\left(\sum_{i=1}^k c_i |w_i|^2 + \kappa^2\right)^{\beta k}} \mathrm{d}\mu_k(\boldsymbol{w})$$

$$\le \|\mu_k\|_\infty \prod_{i=1}^k c_i^{-\beta/2} \int_{\mathbb{K}^k} \frac{\mathbb{1}(\|\boldsymbol{x}\| \le 1)}{(\|\boldsymbol{x}\|^2 + \kappa^2)^{\beta k}} \mathrm{d}^{\beta k} \boldsymbol{x} \le C \|\mu_k\|_\infty \prod_{i=1}^k c_i^{-\beta/2} \int_0^1 \frac{s^{\beta k - 1}}{(s^2 + \kappa^2)^{\beta k}} \mathrm{d}s \quad (5.40)$$

$$\le C \|\mu_k\|_\infty \kappa^{-\beta k} \prod_{i=1}^k c_i^{-\beta/2},$$

where we used the change of variables $x_i = \sqrt{c_i} w_i$ in the second inequality. Here the constant $C > 0$ is a number that depends solely on β and k. Likewise, for the integral over the third domain $S_1 \cap S_2$ we get

$$\int_{\mathbb{K}^{2k+1}} \frac{\mathbb{1}_{S_1 \cap S_2}(\boldsymbol{w})}{\left(\sum_{i=1}^{2k+1} c_i |w_i|^2\right)^{\beta k}} \mathrm{d}\mu_{2k+1}(\boldsymbol{w})$$

$$\le C \|\mu_{2k+1}\|_\infty \prod_{i=1}^{2k+1} c_i^{-\beta/2} \int_{\mathbb{K}^{k+1}} \int_{\mathbb{K}^k} \frac{\mathbb{1}(\|\boldsymbol{x}\| \le 1)\mathbb{1}(\|\boldsymbol{y}\| \le \kappa)}{(\|\boldsymbol{x}\|^2 + \|\boldsymbol{y}\|^2)^{\beta k}} \mathrm{d}^{\beta k}\boldsymbol{x}\,\mathrm{d}^{\beta(k+1)}\boldsymbol{y} \quad (5.41)$$

$$\le C \|\mu_{2k+1}\|_\infty \prod_{i=1}^{2k+1} c_i^{-\beta/2} \int_0^\kappa \int_0^1 \frac{s^{\beta k - 1} t^{\beta(k+1)-1}}{(s^2 + t^2)^{\beta k}} \mathrm{d}s\mathrm{d}t \le C \|\mu_{2k+1}\|_\infty \kappa^\beta \prod_{i=1}^{2k+1} c_i^{-\beta/2}.$$

The fact that $\|\mu_p\|_\infty \le C(p, \mathfrak{b})$ for any $p \le N$ is a direct consequence of the following lemma.

Lemma 5.4 [42, Theorem 1.1]. *Let $\widetilde{p} \le \widetilde{N}$ be positive integers, $\mathfrak{b} > 0$, and $\widetilde{\boldsymbol{b}} = (\widetilde{b}_i)_{i \in [\![N]\!]} \in \mathbb{R}^{\widetilde{N}}$ be a random vector with independent components. If the densities of $\widetilde{b}_i$ are bounded by $\mathfrak{b}$, then the density of $\widetilde{P}\widetilde{\boldsymbol{b}}$ on $\widetilde{P}\mathbb{R}^N$ is bounded by $(\mathfrak{C}_0\mathfrak{b})^{\widetilde{p}}$ for any orthogonal projection $\widetilde{P}$ on $\mathbb{R}^N$ with $\mathrm{rank}\,\widetilde{P} = \widetilde{p}$ where $\mathfrak{C}_0 > 0$ is an absolute constant.*

In the real case, we apply Lemma 5.4 with $\widetilde{p} = p$, $\widetilde{N} = N$, $\widetilde{\boldsymbol{b}} = \boldsymbol{b}$, and

$$\widetilde{P} = U^* \begin{pmatrix} I_p & O \\ O & O \end{pmatrix} U, \quad (5.42)$$

to find that the density of $\widetilde{P}\widetilde{\boldsymbol{b}}$ is bounded on $\widetilde{P}\mathbb{R}^N$. Since μ_p is the law of $U\widetilde{P}\widetilde{\boldsymbol{b}}$ on $U\widetilde{P}\mathbb{R}^N = (\mathbb{R}^p, \mathbf{0})$, we immediately have $\|\mu_p\|_\infty \le (\mathfrak{C}_0\mathfrak{b})^p$. To apply Lemma 5.4 in the complex case, we employ the following notations: For a matrix $A \in \mathbb{C}^{d_1 \times d_2}$, we define

$$\mathcal{J}[A] := \begin{pmatrix} \mathfrak{R}[A] & -\mathfrak{I}[A] \\ \mathfrak{I}[A] & \mathfrak{R}[A] \end{pmatrix} \in \mathbb{R}^{2d_1 \times 2d_2}. \quad (5.43)$$

where $\mathfrak{R}[A]$ and $\mathfrak{I}[A]$ are the entrywise real and imaginary parts of A defined in (1.25). With a slight abuse of notation, when $\boldsymbol{v} \in \mathbb{C}^d$ is a vector, we define $\mathcal{J}[\boldsymbol{v}] := (\mathrm{Re}[\boldsymbol{v}]^\intercal, \mathrm{Im}[\boldsymbol{v}]^\intercal)^\intercal \in \mathbb{R}^{2d}$. Note that $\mathcal{J}$ is an $\mathbb{R}$-linear algebra homomorphism in the sense that, for all $A \in \mathbb{C}^{d_1 \times d_2}, B \in$

$\mathbb{C}^{d_2 \times d_3}$, and $\boldsymbol{v} \in \mathbb{C}^{d_2}$,

$$\mathcal{J}[AB] = \mathcal{J}[A]\mathcal{J}[B], \qquad \mathcal{J}[A\boldsymbol{v}] = \mathcal{J}[A]\mathcal{J}[\boldsymbol{v}], \qquad \mathcal{J}[A^*] = \mathcal{J}[A]^\mathsf{T}. \tag{5.44}$$

Now we apply Lemma 5.4 with the choices $\widetilde{p} = 2p$, $\widetilde{N} = 2N$, $\widetilde{\boldsymbol{b}} = \mathcal{J}[\boldsymbol{b}]$, and

$$\widetilde{P} = \mathcal{J}[U^*]\mathcal{J}\left[\begin{pmatrix} I_p & O \\ O & O \end{pmatrix}\right]\mathcal{J}[U]. \tag{5.45}$$

Noticing that $\mathcal{J}[U] \in \mathbb{R}^{2N \times 2N}$ is a real orthogonal matrix, we immediately find $\|\mu_p\|_\infty \le (\mathfrak{C}_0\mathfrak{b})^{2p}$. Combining (5.40), (5.41), $\|\mu_p\|_\infty \le (\mathfrak{C}_0\mathfrak{b})^{\beta p}$, and optimizing for κ gives

$$\mathbb{E}\left(\sum_{i=1}^{2k+1} c_i|\boldsymbol{e}_i^*U\boldsymbol{b}|^2\right)^{-\beta k} \le 1 + \mathfrak{C}_1\left(\prod_{i=1}^{k} c_i^{-\beta/2}\right)\left(\kappa^{-\beta k} + \kappa^\beta \prod_{i=k+1}^{2k+1} c_i^{-\beta/2}\right)$$

$$\le 1 + \mathfrak{C}_1 \prod_{i=1}^{k} c_i^{-\beta/2} \prod_{i=k+1}^{2k+1} c_i^{-k\beta/(2(k+1))},$$

for some constant $\mathfrak{C}_1 \equiv \mathfrak{C}_1(k, \mathfrak{b})$. This completes the proof of Lemma 5.3. $\qquad\square$

Proof of Lemma 5.2. Recall that the goal is to prove

$$\mathbb{E}[(1 + N\lambda_{3k})^{2k^2}] \le C(k, \delta, \tau, \mathfrak{m})N^\delta. \tag{5.46}$$

Observe that we may assume that N is sufficiently large, as long as the threshold depends on the same parameters as C. Recall the definition of Ξ_z from (3.5). We apply Lemma 3.1 with the choices $\epsilon = \xi = \delta/(100k^2)$ and $D = 100k^2$, so that $[\Xi_z(\epsilon, \xi)^c] \le N^{-D}$ holds for all $|z| \le 1 - \tau$. Then, on the event $\Xi_z(\epsilon, \xi)$ we have

$$N\lambda_{3k}^z \le N\lambda_{\lfloor N^{\epsilon+\xi}\rfloor}^z \le N^{1+\epsilon}\eta \le N^{1+2\epsilon},$$

hence

$$\mathbb{E}\mathbb{1}_{\Xi_z(\epsilon,\xi)}(1 + N\lambda_{3k}^z)^{2k^2} \le N^{5k^2\epsilon}. \tag{5.47}$$

On the complementary event $\Xi_z(\epsilon, \xi)^c$ we use Cauchy-Schwarz and $\lambda_N^z = \|A + X - z\| \le \|X\| + |z| + \|\mathfrak{K}\|$ to get

$$\mathbb{E}\mathbb{1}_{\Xi_z(\epsilon,\xi)^c}(1 + N\lambda_{3k}^z)^{2k^2} \le \mathbb{P}[\Xi_z(\epsilon,\xi)^c]^{1/2}\mathbb{E}[(1 + N\lambda_N^z)^{4k^2}]^{1/2} \le N^{2k^2-D/2}(1 + \|\|X\|\|_{4k^2})^{2k^2}. \tag{5.48}$$

Then we finally use

$$\mathbb{E}\left[\|X\|^{4k^2}\right] \le \mathbb{E}(\mathrm{Tr}\,XX^*)^{2k^2} \le N^{2k^2}\mathfrak{m}_{4k^2} \tag{5.49}$$

so that

$$\mathbb{E}\mathbb{1}_{\Xi_z(\epsilon,\xi)^c}(1 + N\lambda_{3k}^z)^{2k^2} \le N^{3k^2-D/2}\mathfrak{m}_{4k^2}^{1/2}. \tag{5.50}$$

Substituting the definitions of ϵ, ξ, D into (5.47) and (5.50) proves (5.16), concluding the proof of Lemma 5.2. $\qquad\square$

6 | PROOF OF THEOREM 2.10

The proof of Theorem 2.10 follows a similar strategy as in Section 5 for the proof of Theorem 2.11 but with some non-trivial modifications that will be summarized in Remark 6.4. The key step again is to prove (5.13) but this time with the deterministic choice $\mathfrak{X}_\mathcal{I} = C^{|\mathcal{I}|}|\log \eta|^{|\mathcal{I}|}$; we state this in the next proposition. Recall the definition of β from (5.2).

Proposition 6.1. *Let* $k, N \in \mathbb{N}$ *with* $2 \leq k \leq N$, X *be an* $(N \times N)$ *regular complex or real matrix, and* $A \in \mathbb{C}^{N \times N}$. *Then there exists a constant* $C \equiv C(k, \mathfrak{b}) > 0$ *such that*

$$\mathbb{P}\left[\Xi_k^{(\mathcal{I})}\right] \leq C|\log \eta|(N\eta)^{\beta k}\frac{1}{N - |\mathcal{I}|}\sum_{i \in [\![N]\!] \setminus \mathcal{I}} \mathbb{P}\left[\Xi_k^{(\mathcal{I} \cup \{i\})}\right] \qquad \forall \eta \in [0, N^{-1}], \tag{6.1}$$

holds for all $\mathcal{I} \subset [\![N]\!]$ *with* $|\mathcal{I}| \leq k - 1$. *If* X *is complex, the same result also holds for* $k = 1$.

Note that Proposition 6.1, in contrast to Proposition 5.1, does not assume that A is real even if X is. As we will see below, complex A poses additional technicalities in its proof when X is real. Given Proposition 6.1, Theorem 2.10 follows directly by substituting $\eta = s/N$ and $\mathfrak{X}_\mathcal{I} = C^{|\mathcal{I}|}|\log \eta|^{|\mathcal{I}|}$ into (5.14) where C is the constant from (6.1). Thus we move on to the proof of Proposition 6.1.

Proof of Proposition 6.1. We make two modifications to the proof Proposition 5.1 in order to prove (6.1). Firstly in (5.22), we raise (5.21) to the $(\beta k/2)$-th power instead of (βk), and we do not smuggle in the factor $(1 + N\lambda_m^{(j-1)})^n$ in the following steps. Notice here that we need to take $k \geq 2$ in the real case, so that we may apply Jensen's inequality in (5.23) with the power $k/2$. Secondly in (5.31), we keep the first term η and only use the first k coordinates of $\mathbf{w}^{(j)}$ instead of the first $(2k + 1)$. After these modifications, it suffices to prove

$$\mathbb{1}_{\Xi_k^{(j)}}\mathbb{E}^{(j)}\left(\eta + \sum_{i \in [\![k]\!]} c_i^{(j)}|w_i^{(j)}|^2\right)^{-\beta k/2} \leq C(N\eta)^{\beta k/2}|\log \eta| \tag{6.2}$$

for a constant $C \equiv C(k, \mathfrak{b}) > 0$.

Recall from (5.32) that $c_i^{(j)} \geq 1/(2N\eta)$ for all $i \in [\![k]\!]$ on the event $\Xi_k^{(j)}$. Thus we have

$$\mathbb{1}_{\Xi_k^{(j)}}\mathbb{E}^{(j)}\left(\eta + \sum_{i \in [\![N]\!]} c_i^{(j)}|w_i^{(j)}|^2\right)^{-\beta k/2} \leq C(N\eta)^{\beta k/2}\mathbb{1}_{\Xi_k^{(j)}}\mathbb{E}^{(j)}\left(N\eta^2 + \sum_{i \in [\![k]\!]} |w_i^{(j)}|^2\right)^{-\beta k/2}, \tag{6.3}$$

where $C > 0$ depends only on k. Therefore (6.2) follows once we prove

$$\mathbb{E}^{(j)}\left(N\eta^2 + \sum_{i \in [\![k]\!]} |w_i^{(j)}|^2\right)^{-\beta k/2} \leq C|\log \eta| \tag{6.4}$$

for a constant C depending only on k and $\mathfrak{b}$. We state this estimate as the next lemma. To simplify the presentation we define $P_k : \mathbb{C}^N \to \mathbb{C}^k$ to be the projection onto the first k components, that is,

$$P_k := \begin{pmatrix} I_k & | & O \end{pmatrix} \in \mathbb{C}^{(k \times N)}. \tag{6.5}$$

Lemma 6.2. *Let $k, N \in \mathbb{N}$ with $k \leq N$, $\mathbb{K} = \mathbb{R}$ or $\mathbb{C}$, and $\boldsymbol{b} \in \mathbb{K}^N$ be a random vector with independent components satisfying the same assumptions as in Lemma 5.3. Then there exists a constant $\mathfrak{C}_2 \equiv \mathfrak{C}_2(k, \mathfrak{b}) > 0$ such that the following holds for any $c_0 \in (0, 1/2)$, $\boldsymbol{a} \in \mathbb{C}^N$, and a unitary matrix $U \in \mathbb{C}^{N \times N}$:*

$$\mathbb{E}\left(c_0 + \|P_k U(\boldsymbol{b} + \boldsymbol{a})\|^2\right)^{-\beta k/2} \leq \mathfrak{C}_2(1 + |\log c_0|). \tag{6.6}$$

Note that when $\mathbb{K} = \mathbb{R}$, the vector $PU(\boldsymbol{b} + \boldsymbol{a})$ in (6.6) involves both real and complex matrices (or vectors) in contrast to (5.34). Hereafter, any binary operation concerning both real and complex matrices treats $\mathbb{R}$ as canonically embedded in $\mathbb{C}$.

As in Section 5, we postpone the proof of Lemma 6.2 and first use this lemma to conclude Proposition 6.1. We choose $c_0 = N\eta^2$, $U = U^{(j)}$, $\boldsymbol{b} = \sqrt{N} X^* \boldsymbol{e}_j$, and $\boldsymbol{a} = \sqrt{N} A^* \boldsymbol{e}_j$. With these choices we have

$$P_k U(\boldsymbol{b} + \boldsymbol{a}) = \sqrt{N} P_k U^{(j)}(X + A)^* \boldsymbol{e}_j = P U^{(j)} \boldsymbol{b}^{(j)} = P \boldsymbol{w}^{(j)}, \tag{6.7}$$

so that the left-hand side of (6.6) is exactly that of (6.4). Following similar arguments as in the proof of Proposition 6.1, these choices satisfy all assumptions of Lemma 6.2 so that

$$\mathbb{E}^{(j)}\left(N\eta^2 + \sum_{i \in [\![k]\!]} |w_i^{(j)}|^2\right)^{-\beta k/2} \leq \mathfrak{C}_2(1 + |\log(N\eta^2)|) \leq 10\mathfrak{C}_2 |\log \eta|, \tag{6.8}$$

where we used $\eta \in [0, N^{-1}]$. This proves (6.4), concluding the proof of Proposition 6.1. $\qquad\square$

Next we will present the proof of Lemma 6.2. Recall that in Lemma 6.2 the unitary matrix U is not real orthogonal even if X is real, in contrast to Lemma 5.3. Hence the proof of (6.4) for the real case does not simply follow from the complex case just after some changes in the exponents. Nonetheless, we still expect that the vector $P_k U \boldsymbol{b}$ carries at least half the degrees of freedom compared to the complex case, regardless of U. As before, these will be used to regularize the potentially singular expectation in (6.6) even if c_0 is very small. In other words, we will construct a "projection" $R : \mathbb{C}^k \to \mathbb{R}^k$ so that $RP_k U \boldsymbol{b} \in \mathbb{R}^k$ has a continuous distribution. The next lemma is used to construct such an R.

Lemma 6.3. *Let U be an $N \times N$ complex unitary matrix, $k \in [\![N]\!]$, P_k be given by (6.5), and define*

$$Q := \mathcal{J}[P_k] \mathcal{J}[U] \begin{pmatrix} I_N \\ O \end{pmatrix} \in \mathbb{R}^{2k \times N}. \tag{6.9}$$

Then the singular values $m_1 \leq \cdots \leq m_{2k} = \|Q\|$ of Q satisfy

$$m_{2k} \leq 1, \qquad m_{k+1} \geq \frac{1}{\sqrt{k+1}}. \tag{6.10}$$

The proof of Lemma 6.3 is postponed to the end of this section, and we proceed to prove Lemma 6.2.

Proof of Lemma 6.2. We first prove the complex case, $\beta = 2$. One can easily see that the vector $\widetilde{\boldsymbol{b}} := \boldsymbol{b} + \boldsymbol{a}$ has independent components whose densities are bounded by $\mathfrak{b}$, exactly when $\boldsymbol{b}$ does. Since $\|PU(\boldsymbol{b} + \boldsymbol{a})\| = \|PU\widetilde{\boldsymbol{b}}\|$, we may replace $PU(\boldsymbol{b} + \boldsymbol{a})$ by $PU\boldsymbol{b}$ in the left-hand side of (6.6), that is, it suffices to prove

$$\mathbb{E}(c_0 + \|PU\boldsymbol{b}\|^2)^{-k} \leq \mathfrak{C}_2(1 + |\log c_0|). \tag{6.11}$$

Note that this argument using a complex shift $\boldsymbol{a}$ only applies to the case $\mathbb{K} = \mathbb{C}$, and later for the real case we will need to use Lemma 6.3 instead.

Now we apply similar arguments as in Lemma 5.3; first, define μ_k to be the distribution of $PU\boldsymbol{b}$; for a test function $f : \mathbb{C}^k \to \mathbb{C}$ we write

$$\int_{\mathbb{C}^N} f \mathrm{d}\mu := \int_{\mathbb{C}^N} f(PU\boldsymbol{b})h(\boldsymbol{b})\mathrm{d}^{2N}\boldsymbol{b}$$

where $h(\boldsymbol{b})$ was defined in (5.37). Then we have, for a constant $C \equiv C(k) > 0$,

$$\mathbb{E}\left(c_0 + \|PU\boldsymbol{b}\|^2\right)^{-k} = \int_{\mathbb{C}^k}\left(c_0 + \|\boldsymbol{w}\|^2\right)^{-k}\mathrm{d}\mu_k(\boldsymbol{w}) \leq 1 + \int_{\mathbb{C}^k}\frac{\mathbb{1}(\|\boldsymbol{w}\| \leq 1)}{(c_0 + \|\boldsymbol{w}\|^2)^k}\mathrm{d}\mu_k(\boldsymbol{w})$$

$$\leq 1 + \|\mu_k\|_\infty \int_{\mathbb{C}^k}\frac{\mathbb{1}(\|\boldsymbol{w}\| \leq 1)}{(c_0 + \|\boldsymbol{w}\|^2)^k}\mathrm{d}^{2k}\boldsymbol{w} \leq 1 + C\|\mu_k\|_\infty \int_0^1 \frac{s^{2k-1}}{(c_0 + s^2)^k}\mathrm{d}s, \tag{6.12}$$

where we used in the first inequality that the integral over $\|\boldsymbol{w}\| > 1$ is less than 1. Recall that this μ_k is exactly the same as that in the proof of Lemma 5.3, so that $\|\mu_k\|_\infty \leq C(k, \mathfrak{b})$. Since the last integral in (6.12) is bounded by $|\log c_0|$, therefore we have

$$\mathbb{E}\left(c_0 + \|PU\boldsymbol{b}\|^2\right)^{-k} \leq \mathfrak{C}_2(1 + |\log c_0|), \tag{6.13}$$

for some constant $\mathfrak{C}_2 \equiv \mathfrak{C}_2(k, \mathfrak{b}) > 0$. This proves (6.11), thus concludes Lemma 6.2 for the complex case.

Now we move on to the real case. We first notice that

$$\mathcal{J}[P_k U\boldsymbol{b}] = \mathcal{J}[P_k U]\mathcal{J}[\boldsymbol{b}] = \mathcal{J}[P_k]\mathcal{J}[U]\begin{pmatrix} I_N \\ O \end{pmatrix}\boldsymbol{b} = Q\boldsymbol{b} \in \mathbb{R}^{2k}, \tag{6.14}$$

where Q was defined in (6.9). Consider the singular value decomposition

$$Q = \sum_{i=1}^{2k} m_i \boldsymbol{x}_i \boldsymbol{y}_i^\mathsf{T}, \qquad \boldsymbol{x}_i \in \mathbb{R}^{2k}, \boldsymbol{y}_i \in \mathbb{R}^N, \tag{6.15}$$

and define

$$R := \sum_{i=k+1}^{2k} \boldsymbol{y}_i \boldsymbol{x}_i^\mathsf{T}, \quad \widetilde{P} := \sum_{i=k+1}^{2k} \boldsymbol{y}_i \boldsymbol{y}_i^\mathsf{T}, \quad V := \sum_{i=1}^{k} \boldsymbol{e}_i \boldsymbol{y}_{i+k}^\mathsf{T} \in \mathbb{R}^{k\times N}, \quad D := \mathrm{diag}(m_{k+1}, \dots, m_{2k}). \tag{6.16}$$

Note that $\widetilde{P}$ is an orthogonal projection on $\mathbb{R}^N$, V is a unitary map between $\mathbb{R}^k$ and $\mathrm{span}(\boldsymbol{y}_{k+1}, \dots \boldsymbol{y}_{2k})$, and that

$$VRQ = \sum_{i=k+1}^{2k} m_i \boldsymbol{e}_i \boldsymbol{y}_i^\intercal = DV\widetilde{P}. \tag{6.17}$$

Hence, using $\|VR\| \leq 1$ and $\|\mathcal{J}[\boldsymbol{v}]\| = \|\boldsymbol{v}\|$ for any complex vector $\boldsymbol{v}$, we obtain

$$\|P_k U(\boldsymbol{b} + \boldsymbol{a})\| = \|\mathcal{J}[P_k U(\boldsymbol{b} + \boldsymbol{a})]\| \geq \|VR\mathcal{J}[P_k U(\boldsymbol{b} + \boldsymbol{a})]\|$$
$$= \|DV(\widetilde{P}\boldsymbol{b} + V^\intercal D^{-1} VR\mathcal{J}[P_k U\boldsymbol{a}])\| =: \|DV(\widetilde{P}\boldsymbol{b} + \widetilde{\boldsymbol{a}})\|, \tag{6.18}$$

where we abbreviated

$$\widetilde{\boldsymbol{a}} = V^\intercal D^{-1} VR\mathcal{J}[P_k U\boldsymbol{a}] \in \mathrm{span}(\boldsymbol{y}_{k+1}, \dots, \boldsymbol{y}_{2k}). \tag{6.19}$$

Thus we have that

$$(c_0 + \|PU(\boldsymbol{b} + \boldsymbol{a})\|^2)^{-k/2} \leq (c_0 + \|DV(\widetilde{P}\boldsymbol{b} + \widetilde{\boldsymbol{a}})\|^2)^{-k/2}. \tag{6.20}$$

Next, we define ν_k to be the distribution (on $\mathbb{R}^k$) of $V(\widetilde{P}\boldsymbol{b} + \widetilde{\boldsymbol{a}})$, that is,

$$\int_{\mathbb{R}^k} f(\boldsymbol{x}) \mathrm{d}\nu_k(\boldsymbol{x}) = \int_{\mathbb{R}^N} f(V(\widetilde{P}\boldsymbol{b} + \widetilde{\boldsymbol{a}})) h(\boldsymbol{b}) \mathrm{d}^N \boldsymbol{b} \tag{6.21}$$

where $h(\boldsymbol{b})$ was defined in (5.37). Then we may follow the same lines as in (6.12) to write

$$\mathbb{E}(c_0 + \|PU(\boldsymbol{b} + \boldsymbol{a})\|^2)^{-k/2} \leq \mathbb{E}(c_0 + \|DV(\widetilde{P}\boldsymbol{b} + \widetilde{\boldsymbol{a}})\|^2)^{-k/2}$$
$$= \int_{\mathbb{R}^k} (c_0 + \|D\boldsymbol{x}\|^2)^{-k/2} \mathrm{d}\nu_k(\boldsymbol{x}) \leq 1 + \frac{\|\nu_k\|_\infty}{(k+1)^{k/2}} \int_0^\infty \frac{s^{k-1}}{(c_0 + s^2)^{k/2}} \mathrm{d}s \tag{6.22}$$

where we used in the last inequality that $\det D \geq m_{k+1}^k \geq (k+1)^{k/2}$.

Since the last integral on the right-hand side of (6.22) is comparable to $|\log c_0|$, it suffices to prove $\|\nu_k\|_\infty \leq C(k, \mathfrak{b})$ in order to conclude (6.6) for the real case. To prove this, we apply Lemma 5.4 with the choices $\widetilde{\boldsymbol{b}} = \boldsymbol{b}$ and $\widetilde{P}$ in (6.16). As a result, the densities of $\widetilde{P}\boldsymbol{b}$ and hence $(\widetilde{P}\boldsymbol{b} + \widetilde{\boldsymbol{a}})$ on $\mathrm{span}(\boldsymbol{y}_{k+1}, \dots, \boldsymbol{y}_{2k})$ are both bounded by $C(k, \mathfrak{b})$. Since V is an isometry between $\mathrm{span}(\boldsymbol{y}_{k+1}, \dots, \boldsymbol{y}_{2k})$ and $\mathbb{R}^k$, we immediately find $\|\nu_k\|_\infty \leq C(k, \mathfrak{b})$. This completes the proof of Lemma 6.2 modulo Lemma 6.3. □

Proof of Lemma 6.3. First of all, $m_{2k} = \|Q\| \leq 1$ follows directly from

$$\|Q\boldsymbol{v}\| = \|\mathcal{J}[P_k U\boldsymbol{v}]\| = \|P_k U\boldsymbol{v}\| \leq \|\boldsymbol{v}\|, \qquad \forall \boldsymbol{v} \in \mathbb{R}^N. \tag{6.23}$$

To prove $m_{k+1} \geq (k+1)^{-1/2}$, we use the definition of $\mathcal{J}$ to write the matrix Q in (6.9) as

$$Q = \begin{pmatrix} P_k \mathfrak{R}[U] \\ -P_k \mathfrak{I}[U] \end{pmatrix} \in \mathbb{R}^{2k \times N}, \tag{6.24}$$

so that, using the same singular decomposition as in (6.15),

$$QQ^\intercal = \begin{pmatrix} P_k \mathfrak{R}[U] \\ -P_k \mathfrak{I}[U] \end{pmatrix} \begin{pmatrix} \mathfrak{R}[U]^\intercal P^\intercal & -\mathfrak{I}[U]^\intercal P^\intercal \end{pmatrix} = \sum_{i=1}^{2k} m_i^2 \boldsymbol{x}_i \boldsymbol{x}_i^\intercal. \tag{6.25}$$

Then from $\mathfrak{R}[UU^*] = \mathfrak{R}[I] = I$ we have

$$\sum_{i=1}^{2k} m_i^2 = \mathrm{Tr} \begin{pmatrix} P\mathfrak{R}[U] \\ -P\mathfrak{I}[U] \end{pmatrix} \begin{pmatrix} \mathfrak{R}[U]^\intercal P^\intercal & -\mathfrak{I}[U]^\intercal P^\intercal \end{pmatrix}$$

$$= \mathrm{Tr}\, P(\mathfrak{R}[U]\mathfrak{R}[U]^\intercal + \mathfrak{I}[U]\mathfrak{I}[U]^\intercal)P^\intercal = \mathrm{Tr}\, P\mathfrak{R}[UU^*]P^\intercal = \mathrm{Tr}\, PP^\intercal = k. \tag{6.26}$$

Hence for each $\ell \in [\![2k]\!]$ we have

$$k = \sum_{i=1}^{2k} m_i^2 \le \ell m_\ell^2 + (2k - \ell)m_{2k}^2 \le \ell m_\ell^2 + (2k - \ell), \tag{6.27}$$

where we used in the last inequality that $m_{2k} \le 1$. Thus for all $\ell \ge k$ we have

$$m_\ell \ge \sqrt{\frac{\ell - k}{\ell}}, \qquad \text{in particular} \quad m_{k+1} \ge \frac{1}{\sqrt{k+1}}. \tag{6.28}$$

This completes the proof of Lemma 6.3. $\qquad\square$

Remark 6.4. As one can see from the proofs, the difference between Theorem 2.10 and 2.11 is rooted in that between (5.33) and (6.2). Comparing (6.2) to (5.33), we find that the left-hand sides concern different inverse powers of the same random variable $\sum_i c_i^{(j)}|w_i^{(j)}|^2$ modulo irrelevant summands. On the other hand, the right-hand side of (6.2) has a factor of $(N\eta)^{\beta k/2}$, which is small, whereas that of (5.33) is (roughly) $O(1)$.

Here we briefly explain how these two different estimates for the inverse powers of the same random variable arise. In fact, when $N\eta \ll 1$, (6.2) better represents the natural size of this random variable. To see this, we recall (5.32) and write for general $\boldsymbol{w} \in \mathbb{C}^N$ that, on the event $\Xi_k^{(j)}$,

$$\frac{1}{2N\eta} \sum_{i=1}^{k} |w_i|^2 \le \sum_{i=1}^{k} c_i^{(j)}|w_i|^2 \le \sum_{i\in[\![N]\!]} c_i|w_i|^2 \le \frac{1}{N\eta} \sum_{i=1}^{k} |w_i|^2 + \frac{1}{N} \sum_{i>k} \frac{\eta}{(\lambda_i^{(j)})^2 + \eta^2} |w_i|^2. \tag{6.29}$$

Typically we take w_i to be $O(1)$ random variables with continuous joint distribution, and hence the second term on the right-hand side of (6.29) is roughly of the same size as $\langle G(i\eta) \rangle$ which is $O(1)$ (see Lemma 3.2). Thus essentially the first k summands determine the size of $\sum_i c_i^{(j)}|w_i^{(j)}|^2$, which is in turn comparable to $(N\eta)^{-1}\|P\boldsymbol{w}\|^2$ recalling that the relevant regime is $N\eta \ll 1$.

Therefore, in effect, one can only use these βk degrees of freedom (from k variables in $\mathbb{K}$) upon estimating the m-th negative moment to regularize the potential singularity. This essentially leads to integrals of the form

$$\int_{\mathbb{K}^k} \frac{\mathbb{1}(\|\boldsymbol{w}\| \le 1)}{\|\boldsymbol{w}\|^{2m}} \mathrm{d}^{\beta k}\boldsymbol{w}, \tag{6.30}$$

which is finite only if $m < \beta k/2$. When m is exactly $\beta k/2$ this integral has logarithmic singularity, which is responsible for the logarithmic corrections in Theorem 2.10. On the other hand, for any $m > \beta k/2$, the *a priori* uncontrolled singular values $\{\lambda_i^{(j)} : i > k\}$ affect the m-th negative moment of $\sum_i c_i^{(j)} |w_i^{(j)}|^2$ via $c_i^{(j)}$. The main point of Theorem 2.11 is that $\lambda_i^{(j)}$'s, and hence $c_i^{(j)}$'s, for $i > k$ are controlled if replace A with $A - z$ for z in the bulk spectrum and thus effectively contribute to regularizing the integral.

7 | PROOF OF THEOREM 2.13

Now we consider the singular values of real regular matrices, but shifted by genuinely complex matrix $A \in \mathbb{C}^{N \times N}$. More precisely, we show that (2.21) extends to $k = 1$ with the improved rate s^2 if $\Im[A]$ is strictly positive definite.

Recall, from the proofs of Theorems 2.10 and 2.11 for the complex case with $k = 1$, that the scale s^2 is due to the fact that the random variable $w_1^{(1)} = (v_1^{(1)})^* b$ is genuinely complex; more specifically, that the distribution μ_1 of $w^{(1)}$ on $\mathbb{C}$ has a bounded density. Here we prove the same statement for the real case when the shift A is genuinely complex.

Before presenting the proof, we briefly explain which parts of the proof of $\|\mu_1\|_\infty = O(1)$ have to be modified compared to those in Sections 5 and 6. If A is real, then $v_1^{(1)}$, the null vector of $J^1(X + A)$, is also real and thus $w_1^{(1)}$ is real, so that its density μ_1 is singular if viewed as a density on $\mathbb{C}$. If the shift A is complex, then $b = \sqrt{N}(X + A)^* e_1$ becomes complex but only due to an irrelevant shift. Hence the regularity of μ_1 should come from the fact that $v_1^{(1)}$ is genuinely complex. Furthermore, whatever estimate we obtain for $\|\mu_1\|_\infty$, it should deteriorate as A tends to a real matrix; we have already seen in the proof of Theorem 2.11 that $w_1^{(1)}$ is real when A is.

The next lemma summarizes the technical input we need for the proof of Theorem 2.13. It proves that $v_1^{(1)}$ is a genuinely complex vector, quantitatively with an explicit dependence on $\Im A = (A - \overline{A})/(2\mathrm{i})$.

Lemma 7.1. *Let $Y, B \in \mathbb{R}^{N \times N}$ and $v \in \mathbb{C}^N$ be a unit null vector of $J^{(1)}(Y + \mathrm{i}B)$. Then we have*

$$\inf_{\theta \in [0, 2\pi]} \| \mathrm{Re}[e^{\mathrm{i}\theta} v] \|^2 \geq \frac{1}{5} \lambda_1(B)^2 \frac{\|J^{(1)} Y w\|^2}{(\|J^{(1)} Y\| + \|B\|)^4}, \tag{7.1}$$

where $\lambda_1(B)$ is the smallest singular value of B and $w \in \mathbb{R}^N$ is the unit null vector of $J^{(1)} B$.

Note that zero is an eigenvalue of $B^*(J^{(1)})^* J^{(1)} B$ and that, by Cauchy interlacing theorem, its multiplicity is exactly one provided $\lambda_1(B) > 0$. Hence w is uniquely determined if $\lambda_1(B) > 0$. We postpone the proof of Lemma 7.1 to the end of this section and move on to that of Theorem 2.13.

Proof of Theorem 2.13. In order to prove (2.27), we follow the proof of Theorem 2.10 in the complex case (hence $\beta = 2$) for $j = k = 1$. One can easily find that all the arguments until Lemma 6.2 remain intact, which is replaced by the following lemma;

Lemma 7.2. *Let $b = (b_i)_{i \in [\![N]\!]} \in \mathbb{R}^N$ be a random vector with independent components whose densities are bounded by $\mathfrak{b} > 0$. Then there exists a constant $\mathfrak{C}_3 \equiv \mathfrak{C}_3(\mathfrak{b}) > 0$ such that the following holds*

for all $c_0 \in (0, 1/2)$ and deterministic vectors $\boldsymbol{v}, \boldsymbol{a} \in \mathbb{C}^N$ with $\|\boldsymbol{v}\| = 1$;

$$\mathbb{E}(c_0 + |\boldsymbol{v}^*(\boldsymbol{b} + \boldsymbol{a})|^2)^{-1} \leq \mathfrak{C}_3 \left(1 + \left(\min_{\theta \in [0,2\pi]} \| \operatorname{Re}[e^{i\theta} \boldsymbol{v}]\| \right)^{-1} |\log c_0| \right). \tag{7.2}$$

The proof of Lemma 7.2 is presented after completing that of Theorem 2.13, and we proceed assuming its validity. If we replace Lemma 6.2 by 7.2, we obtain

$$\mathbb{P}[\Xi_1] \leq (N\eta)^2 \max_{i \in [\![N]\!]} \mathbb{E} \left(N\eta^2 + |w_1^{(\{i\})}|^2 \right)^{-1}$$

$$\leq 10 \mathfrak{C}_3 (N\eta)^2 \max_{i \in [\![N]\!]} \left(|\log \eta| \mathbb{E} \left(\min_{\theta \in [0,2\pi]} \| \operatorname{Re} \left[e^{i\theta} \boldsymbol{v}_1^{(\{i\})} \right] \| \right)^{-1} \right), \tag{7.3}$$

so that it only remains to estimate the expectation on the right-hand side. Also, as before, we take $i = 1$ without loss of generality since the final result is uniform in $i \in [\![N]\!]$. To this end, we apply Lemma 7.1 with the choices $Y = X + \mathfrak{R}A$ and $B = \mathfrak{I}A$. As a result, we obtain

$$\mathbb{E} \left(\min_{\theta \in [0,2\pi]} \| \operatorname{Re} \left[e^{i\theta} \boldsymbol{v}^{(1)} \right] \| \right)^{-1} \leq \frac{C}{\lambda_1(\mathfrak{I}A)} \mathbb{E} \frac{(\|J^{(1)}(X + \mathfrak{R}A)\| + \|\mathfrak{I}A\|)^2}{\|J^{(1)}(X + \mathfrak{R}A)\boldsymbol{w}\|}$$

$$\leq \frac{C}{\lambda_1(\mathfrak{I}A)} \left\| \|J^{(1)}(X + \mathfrak{R}A)\boldsymbol{w}\|^{-1} \right\|_2 \left(\left\| \|X\|^2 \right\|_2 + \|A\|^2 \right) \tag{7.4}$$

for a numeric constant $C > 0$, where $\boldsymbol{w}$ is the unit null vector of $J^{(1)}\mathfrak{I}A$.

We next handle the first factor on the right-hand side of (7.4). We claim that $\mathbb{E}\|J^{(1)}(X + \mathfrak{R}A)\boldsymbol{w}\|^{-r} \leq C$ as long as $N - 1 > r$. To this end, we estimate the lower tail of $\|J^{(1)}(X + \operatorname{Re}A)\boldsymbol{w}\|$ by its Laplace transform;

$$\mathbb{P} \left[\|J^{(1)}(X + \mathfrak{R}A)\boldsymbol{w}\|^2 \leq t^2 \right] = \mathbb{P} \left[-\frac{1}{t^2} \sum_{j=2}^{N} \left(\boldsymbol{e}_j^\top(\sqrt{N}(X + \mathfrak{R}A))\boldsymbol{w} \right)^2 \geq -N \right]$$

$$\leq e^N \prod_{j=2}^{N} \mathbb{E} \exp \left(-\frac{1}{t^2} \left(\boldsymbol{e}_j^\top(\sqrt{N}(X + \mathfrak{R}A))\boldsymbol{w} \right)^2 \right). \tag{7.5}$$

Then for each $j \in [\![2, N]\!]$, since $\|\boldsymbol{w}\| = 1$, the random variable $\boldsymbol{e}_j^\top(\sqrt{N}X + \mathfrak{R}A)\boldsymbol{w}$ has a density bounded by $\mathfrak{C}_0\mathfrak{b}$ by Lemma 5.4. This gives

$$\mathbb{P}[\|J^{(1)}(X + \mathfrak{R}A)\boldsymbol{w}\| \leq t] \leq (C_0 t)^{N-1}, \tag{7.6}$$

for a constant C_0 depending only on $\mathfrak{b}$, which in turn implies

$$\mathbb{E}\|J^{(1)}(X + \mathfrak{R}A)\boldsymbol{e}_1\|^{-r} \leq C_0^r + \mathbb{E}\|J^{(1)}(X + \mathfrak{R}A)\boldsymbol{e}_1\|^{-r} \mathbb{1}_{\|J^{(1)}(X+\mathfrak{R}A)\boldsymbol{e}_1\| \leq C_0^{-1}}$$

$$= C_0^r + \int_{C_0^r}^{\infty} \mathbb{P}[\|J^{(1)}(X + \mathfrak{R}A)\boldsymbol{e}_1\| \leq x^{-1/r}]\mathrm{d}x \leq C_0^r + C_0^r \int_1^{\infty} x^{-(N-1)/r}\mathrm{d}x. \tag{7.7}$$

Taking $r = 2$ and recalling $N \geq 4 > r + 1$ we have

$$\left\| \|J^{(1)}(X + \Re A)e_1\|^{-1} \right\|_2 \leq C(\mathfrak{b}). \tag{7.8}$$

Substituting this into (7.4) and then (7.3) proves (2.27). Given (2.27), (2.28) follows directly from $\mathbb{E}\|X\|^4 \leq C(\mathfrak{m})$, which is a classical result that can be proved following [52]. $\qquad \square$

Proof of Lemma 7.2. We reuse the same notations as in the proof of Lemma 6.2 for the real case. Namely, for a suitable unitary matrix U we may write $\boldsymbol{v} = U^* P_1^*$ with P_1 defined in (6.5), so that

$$\begin{pmatrix} \mathrm{Re}[\boldsymbol{v}^*\boldsymbol{b}] \\ \mathrm{Im}[\boldsymbol{v}^*\boldsymbol{b}] \end{pmatrix} = \mathcal{J}[P_1 U \boldsymbol{b}] = Q\boldsymbol{b}, \tag{7.9}$$

with the same Q as in Lemma 6.3 for $k = 1$. Again we write the same singular value decomposition of Q as in (6.15), but here we replace the matrices in (6.16) by the following, that use both m_1 and m_2:

$$R' := \sum_{i=1,2} \boldsymbol{y}_i \boldsymbol{x}_i^\intercal, \quad \widetilde{P}' := \sum_{i=1,2} \boldsymbol{y}_i \boldsymbol{y}_i^\intercal, \quad V' := \sum_{i=1,2} \boldsymbol{e}_i \boldsymbol{y}_i^\intercal, \quad D' := \mathrm{diag}(m_1, m_2). \tag{7.10}$$

Now we may follow similar lines as in the proof of Lemma 6.2: We define $\widetilde{\boldsymbol{a}}'$ by the exact same formula as in (6.19), but with matrices in (6.16) replaced by those in (7.10) so that $\widetilde{\boldsymbol{a}}' \in \mathrm{span}(\boldsymbol{y}_1, \boldsymbol{y}_2)$. Then, denoting the distribution of $V'(\widetilde{P}'\boldsymbol{b} + \widetilde{\boldsymbol{a}}')$ by ν_1', we have

$$\mathbb{E}(c_0 + |\boldsymbol{v}^*(\boldsymbol{b} + \boldsymbol{a})|^2)^{-1} = \mathbb{E}(c_0 + \|P_1 U(\boldsymbol{b} + \boldsymbol{a})\|^2)^{-1} \leq \mathbb{E}(c_0 + \|D' V'(\widetilde{P}'\boldsymbol{b} + \widetilde{\boldsymbol{a}}')\|^2)^{-1}$$

$$= \int_{\mathbb{R}^2} (c_0 + \|D'\boldsymbol{x}\|^2)^{-1} \mathrm{d}\nu_1'(\boldsymbol{x}) \leq 1 + C \frac{\|\nu_1'\|_\infty}{\det D'} |\log c_0|. \tag{7.11}$$

By the exact same argument as in the proof of Lemma 6.2 we have $\|\nu_1'\|_\infty \leq C(\mathfrak{b})$, thus it only remains to estimate $(\det D')^{-1} = (m_1 m_2)^{-1}$.

Applying Lemma 6.3 we immediately have $m_1 \geq 1/\sqrt{2}$, hence it suffices to estimate m_2 from below. Writing out the definition of Q in this case, we have

$$Q = \begin{pmatrix} \mathrm{Re}[\boldsymbol{v}]^\intercal \\ -\mathrm{Im}[\boldsymbol{v}]^\intercal \end{pmatrix}. \tag{7.12}$$

Then we immediately find

$$m_2 = \min_{\boldsymbol{s} \in \mathbb{R}^2, \|\boldsymbol{s}\|=1} \|Q^\intercal \boldsymbol{s}\| = \min_{\theta \in [0,2\pi]} \left\| (\mathrm{Re}[\boldsymbol{v}] \quad -\mathrm{Im}[\boldsymbol{v}]) \begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix} \right\| = \min_{\theta \in [0,2\pi]} \| \mathrm{Re}[e^{\mathrm{i}\theta}\boldsymbol{v}] \|. \tag{7.13}$$

We thus obtain

$$\det D = m_1 m_2 \geq \sqrt{2} \min_{\theta \in [0,2\pi]} \| \mathrm{Re}[e^{\mathrm{i}\theta}\boldsymbol{v}] \|, \tag{7.14}$$

and plugging this into (7.11) concludes the proof of Lemma 7.2. $\qquad \square$

Proof of Lemma 7.1. Throughout the proof, we abbreviate $J := J^{(1)}$ and $\lambda := \lambda_1(B)$ for simplicity. Since $\boldsymbol{v}$ is a null vector of $J(Y + iB)$, we get from taking the real part of $e^{i\theta}J(Y + iB)\boldsymbol{v} = 0$ that

$$0 = JX\,\mathrm{Re}[e^{i\theta}\boldsymbol{v}] - JB\,\mathrm{Im}[e^{i\theta}\boldsymbol{v}] = JX\,\mathrm{Re}[e^{i\theta}\boldsymbol{v}] - JBP_{\boldsymbol{w}^\perp}\,\mathrm{Im}[e^{i\theta}\boldsymbol{v}] \tag{7.15}$$

for any $\theta \in \mathbb{R}$, where $P_{\boldsymbol{w}^\perp} \in \mathbb{R}^{N \times N}$ is the projection onto the orthogonal complement of $\boldsymbol{w}$. Then, writing the smallest non-zero singular value of JB by $\lambda_1(JB)$, (7.15) implies

$$\lambda^2(\|\,\mathrm{Im}[e^{i\theta}\boldsymbol{v}]\|^2 - |\,\mathrm{Im}[e^{i\theta}\boldsymbol{w}^\mathsf{T}\boldsymbol{v}]|^2) = \lambda^2\|P_{\boldsymbol{w}^\perp}\,\mathrm{Im}[e^{i\theta}\boldsymbol{v}]\|^2 \leq \lambda_1(JB)\|P_{\boldsymbol{w}^\perp}\,\mathrm{Im}[e^{i\theta}\boldsymbol{v}]\|^2$$
$$\leq \|JBP_{\boldsymbol{w}^\perp}\,\mathrm{Im}[e^{i\theta}\boldsymbol{v}]\|^2 \leq \|JY\|^2\|\,\mathrm{Re}[e^{i\theta}\boldsymbol{v}]\|^2, \tag{7.16}$$

where we used $\lambda \leq \lambda_1(JB)$ from Cauchy interlacing theorem in the first inequality. This in turn gives

$$\lambda^2(1 - |\boldsymbol{w}^*\boldsymbol{v}|^2) \leq (\|JY\|^2 + \lambda^2)\|\,\mathrm{Re}[e^{i\theta}\boldsymbol{v}]\|^2 \leq (\|JY\| + \|B\|)^2\|\,\mathrm{Re}[e^{i\theta}\boldsymbol{v}]\|^2. \tag{7.17}$$

Comparing (7.1) with (7.17) and noticing the first factor on the right-hand side of (7.17), since θ was arbitrary, in order to prove (7.1) it only remains to show

$$1 - |\boldsymbol{w}^*\boldsymbol{v}|^2 \geq \frac{1}{5}\frac{\|JY\boldsymbol{w}\|^2}{(\|JY\| + \|B\|)^2}. \tag{7.18}$$

Next, we prove (7.18). First of all, since $\boldsymbol{v}$ is a null vector of $J(Y + iB)$, we have the following inequality of positive semi-definite matrices for all $\widetilde{\eta} > 0$;

$$\boldsymbol{v}\boldsymbol{v}^* \leq \frac{\widetilde{\eta}^2}{\widetilde{\eta}^2 + (Y + iB)^*J^*J(Y + iB)}. \tag{7.19}$$

We then take $\widetilde{U} \in \mathbb{R}^{N \times N}$ to be a real orthogonal matrix with $\widetilde{U}\boldsymbol{e}_1 = \boldsymbol{w}$, so that for all $\widetilde{\eta} > 0$

$$|\boldsymbol{w}^*\boldsymbol{v}|^2 = |\boldsymbol{e}_1^*\widetilde{U}^*\boldsymbol{v}|^2 \leq \boldsymbol{e}_1^*\frac{\widetilde{\eta}^2}{\widetilde{\eta}^2 + \widetilde{U}^*(Y + iB)^*J^*J(Y + iB)\widetilde{U}}\boldsymbol{e}_1 = \widetilde{\eta}\,\mathrm{Im}\,\big[\widetilde{G}^{(1)}(i\widetilde{\eta})\big]_{N+1,N+1} \tag{7.20}$$

where in the last equality we used (5.28) and defined

$$\widetilde{G}^{(1)}(i\widetilde{\eta}) := \big(\widetilde{H}^{(1)} - i\widetilde{\eta}\big)^{-1}, \qquad \widetilde{H}^{(1)} := \begin{pmatrix} 0 & J(Y + iB)\widetilde{U} \\ \widetilde{U}^*(Y + iB)^*J^* & 0 \end{pmatrix} \in \mathbb{C}^{[\![2,2N]\!] \times [\![2,2N]\!]}. \tag{7.21}$$

On the other hand, we use Schur complement formula to get

$$-\frac{1}{[\widetilde{G}^{(1)}(i\widetilde{\eta})]_{N+1,N+1}} = i\widetilde{\eta} + \sum_{i,j \in [\![2,N]\!] \cup [\![N+2,2N]\!]} \widetilde{H}^{(1)}_{N+1,i}\big[(\widetilde{H}^{(1,1)} - i\widetilde{\eta})^{-1}\big]_{ij}\widetilde{H}^{(1)}_{j,N+1}, \tag{7.22}$$

where $\widetilde{H}^{(1,1)}$ is the matrix obtain from $\widetilde{H}^{(1)}$ by deleting the $(N + 1)$-th rows and columns. Equivalently, $\widetilde{H}^{(1,1)}$ is the Hermitization of $J(Y + iB)\widetilde{U}J^* \in \mathbb{C}^{[\![2,N]\!]^2}$, that is,

$$\widetilde{H}^{(1,1)} := \begin{pmatrix} 0 & J(Y + iB)\widetilde{U}J^* \\ J\widetilde{U}^*(Y + iB)^*J^* & 0 \end{pmatrix} \in \mathbb{C}^{([\![2,N]\!] \cup [\![N+2,2N]\!])^2}. \tag{7.23}$$

Thus, by the fact that $(\widetilde{G}^{(1)}(i\widetilde{\eta}))_{N+1,N+1} \in i\mathbb{R}$, the definition of $\widetilde{H}^{(1)}$, and another application of Schur complement formula, we have

$$\frac{1}{\widetilde{\eta}\,\mathrm{Im}\left[\widetilde{G}^{(1)}(i\widetilde{\eta})\right]_{N+1,N+1}} = 1 + \frac{1}{\widetilde{\eta}}\boldsymbol{e}_1^*\widetilde{U}^*(Y+iB)^*J^*\left(\frac{\widetilde{\eta}}{J(Y+iB)\widetilde{U}J^*J\widetilde{U}^*(Y+iB)^*J^*+\widetilde{\eta}^2}\right)$$

$$J(Y+iB)\widetilde{U}\boldsymbol{e}_1 \geq 1 + \frac{\|J(Y+iB)\boldsymbol{w}\|^2}{\|J(Y+iB)\widetilde{U}J^*\|^2+\widetilde{\eta}^2} \geq 1 + \frac{\|JY\boldsymbol{w}\|^2}{\|J(Y+iB)\|^2+\widetilde{\eta}^2},$$

(7.24)

where in the second line we used $\widetilde{U}\boldsymbol{e}_1 = \boldsymbol{w}$, $JB\boldsymbol{w} = 0$, and the positive semi-definiteness of

$$\frac{1}{ZZ^*+1} - \frac{1}{\|Z\|^2+1} \geq 0 \tag{7.25}$$

which is true for any matrix Z. We then combine (7.20) and (7.24) to get

$$1 - |\boldsymbol{e}_1^*\boldsymbol{v}|^2 \geq 1 - \widetilde{\eta}\,\mathrm{Im}[\widetilde{G}^{(1)}(i\widetilde{\eta})]_{N+1,N+1} \geq 1 - \left(1 + \frac{\|JY\boldsymbol{w}\|^2}{\|J(Y+iB)\|^2+\widetilde{\eta}^2}\right)^{-1}. \tag{7.26}$$

Finally, since $\widetilde{\eta}$ can be arbitrary, we take the limit $\widetilde{\eta} \searrow 0$ in (7.26) to obtain

$$1 - |\boldsymbol{e}_1^*\boldsymbol{v}|^2 \geq \frac{\|JY\boldsymbol{w}\|^2}{\|JY\boldsymbol{w}\|^2 + \|J(Y+iB)\|^2} \geq \frac{1}{5}\frac{\|JY\boldsymbol{w}\|^2}{(\|JY\|+\|B\|)^2}. \tag{7.27}$$

This completes the proof of Lemma 7.1. $\qquad\square$

Remark 7.3. As pointed out in Remark 2.14, the suboptimality in (2.28) is due to that of Lemma 7.1. In particular the inequality (7.16) is far from being optimal: For Gaussian X and $A = -z$, so that $Y = X - \mathrm{Re}\,z$ and $B = -\mathrm{Im}\,z$, numerical experiments show that $\mathrm{Re}[\boldsymbol{v}]$ and $\mathrm{Im}[\boldsymbol{v}]$ have almost equal size when $|\mathrm{Im}\,z| \gtrsim N^{-1/2}$. In this case, we believe that the typical size of the right-hand side of (7.1) is $(1 \wedge N|\mathrm{Im}\,z|^2)$ up to a positive random variable of size $O(1)$. Nonetheless, we do not know whether the first negative moment of this random variable is finite which is crucial for our proof; see (7.4)

ACKNOWLEDGMENTS

REFERENCES

1. M. Aizenman, R. Peled, J. Schenker, M. Shamis, and S. Sodin, *Matrix regularizing effects of Gaussian perturbations*, Commun. Contemp. Math. **19** (2017), no. 3, 1750028.
2. G. Akemann, M. Duits, and L. D. Molag, *The elliptic Ginibre ensemble: A unifying approach to local and global statistics for higher dimensions*, J. Math. Phys. **64** (2023), no. 2, 023503.
3. J. Alt, L. Erdős, and T. Krüger, *Local inhomogeneous circular law*, Ann. Appl. Probab. **28** (2018), no. 1, 148–203.
4. J. Alt and T. Krüger, *Inhomogeneous circular law for correlated matrices*, J. Funct. Anal. **281** (2021), no. 7, 109120.
5. Z. D. Bai, *Circular law*, Ann. Probab. **25** (1997), no. 1, 494–529.
6. J. Banks, J. Garza-Vargas, A. Kulkarni, and N. Srivastava, *Pseudospectral shattering, the sign function, and diagonalization in nearly matrix multiplication time*, 2020 IEEE 61st Annual Symposium on Foundations of Computer Science, IEEE Computer Soc., Los Alamitos, CA, 2020, pp. 529–540.

7. J. Banks, A. Kulkarni, S. Mukherjee, and N. Srivastava, *Gaussian regularization of the pseudospectrum and Davies' conjecture*, Comm. Pure Appl. Math. **74** (2021), no. 10, 2114–2131.

8. J. Banks, J. G. Vargas, A. Kulkarni, and N. Srivastava, *Overlaps, eigenvalue gaps, and pseudospectrum under real Ginibre and absolutely continuous perturbations*, arXiv preprint arXiv:2005.08930, 2020. To appear in Ann. Inst. Henri Poincaré Probab. Stat.

9. S. Belinschi, A. Dembo, and A. Guionnet, *Spectral measure of heavy tailed band and covariance random matrices*, Comm. Math. Phys. **289** (2009), no. 3, 1023–1055.

10. G. Ben Arous and S. Péché, *Universality of local eigenvalue statistics for some sample covariance matrices*, Comm. Pure Appl. Math. **58** (2005), no. 10, 1316–1357.

11. F. Benaych-Georges, *Rectangular random matrices, related convolution*, Probab. Theory Relat. Fields **144** (2009), no. 3–4, 471–515.

12. C. Bordenave and M. Capitaine, *Outlier eigenvalues for deformed i.i.d. random matrices*, Comm. Pure Appl. Math. **69** (2016), no. 11, 2131–2194.

13. C. Bordenave, P. Caputo, and D. Chafaï, *Spectrum of Markov generators on sparse random graphs*, Comm. Pure Appl. Math. **67** (2014), no. 4, 621–669.

14. C. Bordenave and D. Chafaï, *Around the circular law*, Probab. Surv. **9** (2012), 1–89.

15. P. Bourgade and G. Dubach, *The distribution of overlaps between eigenvectors of Ginibre matrices*, Probab. Theory Relat. Fields **177** (2020), no. 1–2, 397–464.

16. P. Bourgade, H.-T. Yau, and J. Yin, *Local circular law for random matrices*, Probab. Theory Relat. Fields **159** (2014), no. 3-4, 545–595.

17. G. Cipolloni, L. Erdős, and D. Schröder, *Edge universality for non-Hermitian random matrices*, Probab. Theory Relat. Fields **179** (2021), no. 1–2, 1–28.

18. G. Cipolloni, L. Erdős, and D. Schröder, *On the condition number of the shifted real Ginibre ensemble*, SIAM J. Matrix Anal. Appl. **43** (2022), no. 3, 1469–1487.

19. G. Cipolloni, L. Erdős, J. Henheik, and D. Schröder, *Optimal lower bound on eigenvector overlaps for non-Hermitian random matrices*, arXiv preprint arXiv:2301.03549, 2023.

20. N. Cook, *Lower bounds for the smallest singular value of structured random matrices*, Ann. Probab. **46** (2018), no. 6, 3442–3500.

21. R. B. Dozier and J. W. Silverstein, *Analysis of the limiting spectral distribution of large dimensional information-plus-noise type matrices*, J. Multivariate Anal. **98** (2007), no. 6, 1099–1122.

22. R. B. Dozier and J. W. Silverstein, *On the empirical distribution of eigenvalues of large dimensional information-plus-noise-type matrices*, J. Multivariate Anal. **98** (2007), no. 4, 678–694.

23. F. J. Dyson, *A Brownian-motion model for the eigenvalues of a random matrix*, J. Mathematical Phys. **3** (1962), 1191–1198.

24. A. Edelman, *The probability that a random real Gaussian matrix has k real eigenvalues, related distributions, and the circular law*, J. Multivariate Anal. **60** (1997), no. 2, 203–232.

25. A. Edelman, E. Kostlan, and M. Shub, *How many eigenvalues of a random matrix are real?* J. Amer. Math. Soc. **7** (1994), no. 1, 247–267.

26. N. El Karoui, *On information plus noise kernel random matrices*, Ann. Statist. **38** (2010), no. 5, 3191–3216.

27. L. Erdős, B. Schlein, and H.-T. Yau, *Wegner estimate and level repulsion for Wigner random matrices*, Int. Math. Res. Not. IMRN (2010), no. 3, 436–479.

28. B. Farrell and R. Vershynin, *Smoothed analysis of symmetric random matrices with continuous distributions*, Proc. Amer. Math. Soc. **144** (2016), no. 5, 2257–2261.

29. Y. V. Fyodorov, *On statistics of bi-orthogonal eigenvectors in real and complex Ginibre ensembles: Combining partial Schur decomposition with supersymmetry*, Comm. Math. Phys. **363** (2018), no. 2, 579–603.

30. J. Ginibre, *Statistical ensembles of complex, quaternion, and real matrices*, J. Mathematical Phys. **6** (1965), 440–449.

31. V. L. Girko, *The circular law*, Teor. Veroyatnost. i Primenen. **29** (1984), no. 4, 669–679.

32. F. Götze and A. Tikhomirov, *The circular law for random matrices*, Ann. Probab. **38** (2010), no. 4, 1444–1491.

33. A. Guionnet, M. Krishnapur, and O. Zeitouni, *The single ring theorem*, Ann. Math. (2) **174** (2011), no. 2, 1189–1217.

34. U. Haagerup and F. Larsen, *Brown's spectral distribution measure for R-diagonal elements in finite von Neumann algebras*, J. Funct. Anal. **176** (2000), no. 2, 331–367.

35. V. Jain, A. Sah, and M. Sawhney, *On the real Davies' conjecture*, Ann. Probab. **49** (2021), no. 6, 3011–3031.

36. A. Maltsev and B. Schlein, *Average density of states of Hermitian Wigner matrices*, Adv. Math. **228** (2011), no. 5, 2797–2836.

37. H. H. Nguyen, *Random matrices: Overcrowding estimates for the spectrum*, J. Funct. Anal. **275** (2018), no. 8, 2197–2224.

38. L. A. Pastur, *The spectrum of random matrices*, Teoret. Mat. Fiz. **10** (1972), no. 1, 102–112.

39. N. S. Pillai and J. Yin, *Universality of covariance matrices*, Ann. Appl. Probab. **24** (2014), no. 3, 935–1001.

40. B. Rider and J. W. Silverstein, *Gaussian fluctuations for non-Hermitian random matrix ensembles*, Ann. Probab. **34** (2006), no. 6, 2118–2143.

41. M. Rudelson and R. Vershynin, *The Littlewood-Offord problem and invertibility of random matrices*, Adv. Math. **218** (2008), no. 2, 600–633.

42. M. Rudelson and R. Vershynin, *Small ball probabilities for linear images of high-dimensional distributions*, Int. Math. Res. Not. (2015), no. 19, 9594–9617.

43. A. Sankar, D. A. Spielman, and S.-H. Teng, *Smoothed analysis of the condition numbers and growth factors of matrices*, SIAM J. Matrix Anal. Appl. **28** (2006), no. 2, 446–476.

44. S. J. Szarek, *Condition numbers of random matrices*, J. Complexity **7** (1991), no. 2, 131–149.

45. T. Tao and V. Vu, *Random matrices: The circular law*, Commun. Contemp. Math. **10** (2008), no. 2, 261–307.

46. T. Tao and V. Vu, *Random matrices: Universality of ESDs and the circular law*, Ann. Probab. **38** (2010), no. 5, 2023–2065. With an appendix by Manjunath Krishnapur.

47. T. Tao and V. Vu, *Random matrices: Universality of local spectral statistics of non-Hermitian matrices*, Ann. Probab. **43** (2015), no. 2, 782–874.

48. E. Thoma, *Overcrowding and separation estimates for the Coulomb gas*, Comm. Pure Appl. Math. Early View article: https://doi.org/10.1002/cpa.22188, 2023.

49. K. Tikhomirov, *Invertibility via distance for noncentered random matrices with continuous distributions*, Random Structures Algorithms **57** (2020), no. 2, 526–562.

50. M. Walters and S. Starr, *A note on mixed matrix moments for the complex Ginibre ensemble*, J. Math. Phys. **56** (2015), no. 1, 013301.

51. F. Wegner, *Bounds on the density of states in disordered systems*, Z. Phys. B **44** (1981), no. 1–2, 9–15.

52. Y. Q. Yin, Z. D. Bai, and P. R. Krishnaiah, *On the limit of the largest eigenvalue of the large-dimensional sample covariance matrix*, Probab. Theory Relat. Fields **78** (1988), no. 4, 509–521.

53. P. Zhong, *Brown measure of the sum of an elliptic operator and a free random variable in a finite von Neumann algebra*, arXiv preprint arXiv:2108.09844, 2021.

APPENDIX A: GENERALIZED DOMAIN FOR THEOREM 2.7

The goal of this section is to prove that we may replace the square $\mathcal{D}$ in Theorem 2.7 by a general Borel set $\mathcal{D}_0$, instead of a square, under the condition that

$$|\mathcal{D}_0 + \left[-N^{-K}, N^{-K}\right]^2| \le C|\mathcal{D}_0| \tag{A.1}$$

for some constants $C, K > 0$, where we recall that $|\mathcal{D}|$ is the Lebesgue measure of any planar domain $|\mathcal{D}|$.

For simplicity, we write $x := N^{-K}$ and $S := [-x, x]^2$. First, note that it suffices to prove that Theorem 2.7 is true with the choice $\mathcal{D} = \mathcal{A} + S$ for any Borel set $\mathcal{A} \subset \mathbb{C}$. Suppose this has been done, then for a given $\mathcal{D}_0$ satisfying (A.1), we define $\Xi_{\mathcal{D}_0} := \Xi_{\mathcal{D}_0 + S}$ and write

$$\mathbb{E}\mathbb{1}_{\Xi_{\mathcal{D}_0}} \sum_{i : \sigma_i \in \mathcal{D}_0} \mathcal{O}_{ii} \le \mathbb{E}\mathbb{1}_{\Xi_{\mathcal{D}_0 + S}} \sum_{i : \sigma_i \in \mathcal{D}_0 + S} \mathcal{O}_{ii} \lesssim N^{1+\xi}(N|\mathcal{D}_0 + S|) \lesssim N^{1+\xi}(N|\mathcal{D}_0|).$$

This proves Theorem 2.7 for $\mathcal{D} = \mathcal{D}_0$.

In order to prove the result for $\mathcal{D} = \mathcal{A} + S$, we prove that this Minkowski sum can be covered by non-intersecting copies of S, so that the total area is comparable to $|\mathcal{A} + S|$.

Lemma A.1. *For any Borel measurable $\mathcal{A} \subset \mathbb{C}$, the discrete set $\mathcal{L} := (2x\mathbb{Z})^2 \cap (\mathcal{A} + 2S)$ satisfies*

$$\mathcal{A} + S \subset \mathcal{L} + S \quad and \quad |\mathcal{L} + S| \leq 3|\mathcal{A} + S|. \tag{A.2}$$

Proof. Define $|z|_\infty = |\operatorname{Re} z| \vee |\operatorname{Im} z|$ for $z \in \mathbb{C}$. To prove the first assertion, take a point $z \in \mathcal{A} + S$. Then we have a point $w \in (2x\mathbb{Z})^2$ and $z_0 \in \mathcal{A}$ such that $|z - w|_\infty \leq x$ and $|z_0 - z|_\infty \leq x$. Thus we immediately have $|w - z_0|_\infty \leq 2x$ so that $w \in (2x\mathbb{Z})^2 \cap (z_0 + 2S) \subset \mathcal{L}$. Thus $z \in w + S \subset \mathcal{L} + S$.

The second assertion follows from $\mathcal{L} \subset \mathcal{A} + 2S$ since

$$|\mathcal{L} + S| \leq |\mathcal{A} + 2S + S| = |\mathcal{A} + 3S| \leq 3|\mathcal{A} + S|.$$

This finishes the proof. $\qquad\square$

Next, we prove that Theorem 2.7 is true for $\mathcal{D} = \mathcal{A} + S$ with any Borel set $\mathcal{A}$. Let $\mathcal{L}_0 := \mathcal{L} \cap [-2, 2]^2$ where $\mathcal{L}$ is the discrete set given by Lemma A.1. Apply Theorem 2.7 for each square $z + S$ with $z \in \mathcal{L}_0$, so that $\mathbb{P}[\Xi^c_{z+S}] \leq N^{-2K-D-1}$ and (2.14) holds true with the choice $\mathcal{D} = z + S$.

Define $\Xi_{\mathcal{A}+S} := \bigcap_{z \in \mathcal{L}_0} \Xi_{z+S} \cap \Xi_0$ where Ξ_0 is the event $[\|X\| \leq 3]$. Then we have $\mathbb{P}[\Xi^c_{\mathcal{A}+S}] \leq N^{-D}$ from $|\mathcal{L}_0| = O(N^{2K})$ and $\mathbb{P}[\Xi^c_0] \leq N^{-D-1}$. Note that on the event Ξ_0 there is no eigenvalue in the rightmost side, hence all, of

$$(\mathcal{A} + S) \setminus (\mathcal{L}_0 + S) \subset (\mathcal{L} \setminus \mathcal{L}_0) + S = (\mathcal{L} \cap ([-3, 3]^2)^c) + S,$$

where we used the first inequality of (A.2). Then it follows that

$$\mathbb{E}\mathbb{1}_{\Xi_{\mathcal{A}+S}} \sum_{i:\sigma_i \in \mathcal{A}+S} \mathcal{O}_{ii} \leq \sum_{z \in \mathcal{L}_0} \mathbb{E}\mathbb{1}_{\Xi_{z+S}} \sum_{i:\sigma_i \in (z+S)} \mathcal{O}_{ii} \lesssim N^{1+\xi}$$

$$(N \sum_{z \in \mathcal{L}_0} |z + S|) = N^{1+\xi}(N|\mathcal{L}_0 + S|) \leq 3N^{1+\xi}(N|\mathcal{D}|),$$

where in the last step we used the second inequality of (A.1). This proves that we may take $\mathcal{D} = \mathcal{A} + S$, and hence $\mathcal{D}_0$ satisfying (A.1), in Theorem 2.7.