SPECTRUM AND PSEUDOSPECTRUM FOR QUADRATIC POLYNOMIALS IN GINIBRE MATRICES

Nicholas A. Cook, Alice Guionnet, Jonathan Husson

SPECTRUM AND PSEUDOSPECTRUM FOR QUADRATIC POLYNOMIALS IN GINIBRE MATRICES

NICHOLAS COOK*, ALICE GUIONNET†, AND JONATHAN HUSSON

ABSTRACT. For a fixed quadratic polynomial $\mathfrak{p}$ in n non-commuting variables, and n independent $N \times N$ complex Ginibre matrices $X_1^N, \ldots, X_n^N$, we establish the convergence of the empirical measure of the eigenvalues of $P^N = \mathfrak{p}(X_1^N, \ldots, X_n^N)$ to the Brown measure of $\mathfrak{p}$ evaluated at n freely independent circular elements $c_1, \ldots, c_n$ in a non-commutative probability space. As in previous works on non-normal random matrices, a key step is to obtain quantitative control on the pseudospectrum of P^N. Via a linearization trick of Haagerup–Thorbjørnsen for lifting non-commutative polynomials to tensors, we obtain this as a consequence of a lower tail estimate for the smallest singular value of patterned block matrices with strongly dependent entries. This reduces to establishing anti-concentration for determinants of random walks in a matrix space of bounded dimension, for which we encounter novel structural obstacles of an algebro-geometric nature.

RÉSUMÉ. Pour un polynôme quadratique $\mathfrak{p}$ en n variables non-commutatives, et n matrices $N \times N$ de Ginibre complexes $X_1^N, \ldots, X_n^N$, nous établissons la convergence de la mesure empirique des valeurs propres de $P^N = \mathfrak{p}(X_1^N, \ldots, X_n^N)$ vers la mesure de Brown de $\mathfrak{p}$ évaluée en n éléments circulaires librement indépendants $c_1, ..., c_n$ dans un espace de probabilité non-commutatif. Comme dans de précédents travaux portant sur des matrices aléatoires non normales, une étape clé est d'obtenir un contrôle qualitatif sur le pseudo-spectre de P^N. Par une méthode de linéarisation due à Haagerup et Thorbjørnsen qui permet de relever les polynômes non-commutatifs en des tenseurs, nous obtenons ce contrôle comme conséquence d'une estimée de la queue de la loi de la plus petite valeur singulière d'une matrice dotée d'une structure par blocs avec des coefficients fortement dépendants. Cela nous ramène à établir l'anti-concentration pour les déterminants de marches aléatoires dans un espace de matrices de dimension bornée, pour lesquelles nous rencontrons de nouveaux obstacles de nature algèbrico-géométriques.

1. INTRODUCTION

1.1. **Background.** Recall that for an $N \times N$ complex matrix A with complex eigenvalues $\lambda_1(A), \ldots, \lambda_N(A)$ (not necessarily distinct), the *empirical spectral distribution (ESD)* is the probability measure

$$\mu_A := \frac{1}{N} \sum_{j=1}^{N} \delta_{\lambda_j(A)} \,. \tag{1.1}$$

For an ensemble of random $N \times N$ matrices $(A^N)_{N \geq 1}$, a central problem in random matrix theory is to establish a law of large numbers for the ESDs – that is, to prove that (in the vague topology) μ_{A^N} converges in probability to some deterministic probability measure μ. The seminal works of Wigner [57] and Marchenko–Pastur [35] addressed this problem for matrices with i.i.d. entries above the diagonal (Wigner ensembles) and Gram matrices for i.i.d. rectangular matrices using the moment and Stieltjes transform methods.

Date: December 2, 2021.

2010 *Mathematics Subject Classification.* 60B20, 46L54.

Key words and phrases. Random matrix, small-ball estimate, Brown measure, non-Hermitian matrix, smallest singular value.

*Partially supported by NSF grant DMS-1606310.

†Partially supported by Labex MILYON/ANR-10-LABX-0070 .

For non-Hermitian *i.i.d. ensembles* X^N, having i.i.d. entries and no symmetry assumption, the problem was only addressed at a comparable level of generality much more recently in [50, 28, 49], where it was shown that $\mu_{\frac{1}{\sqrt{N}}X^N}$ converges to Girko's circular law μ_{circ}, the uniform measure on the unit disk. The moment and Stieltjes transform methods were insufficient to establish the circular law due to the instability of the spectrum of non-normal matrices; these obstacles were overcome using tools from additive combinatorics and geometric functional analysis. We refer to the survey [14] for further background.

The circular law was recently generalized to the inhomogeneous case to allow the variance of the entries depends on the site [19, 1] and further to the case of correlated entries provided their correlations decay fast enough [2]. In [40] it was extended to random block matrices, with each block an i.i.d. matrix, and entries in separate blocks uncorrelated. Girko's *elliptic law* for matrices with correlation between opposite pairs of entries ξ_{ij}, ξ_{ji}, interpolating between the Wigner and i.i.d. ensembles, was established in [39]. The case of matrices invariant under left multiplication by Haar unitary matrices led to the single ring theorem proven in [29]; see also [23] for a non-rigorous derivation. The sum of i.i.d. Haar unitary, orthogonal and permutation matrices were treated in [9, 8].

Random matrices have played an important role in the theory of operator algebras since the seminal work of Voiculescu [56]. In these applications one is interested not in a single random matrix X^N but rather a collection $X_1^N, \ldots, X_n^N$ of a bounded number of independent random matrices, and the algebra they generate in the large N limit. For self-adjoint polynomials in several independent Wigner matrices, Voiculescu showed that the theory of free probability gives powerful tools to understand the limit of ESDs for a non-commutative polynomial in those matrices [56]. Indeed, if $W_1^N, ..., W_n^N$ are independent $N \times N$ Wigner matrices and $\mathfrak{p}$ is a non-commutative self-adjoint polynomial, then the ESD of $\mathfrak{p}(W_1^N, ..., W_n^N)$ converges towards the spectral distribution of $\mathfrak{p}(s_1, ..., s_n)$, where $s_1, ..., s_n$ are n freely independent semi-circular elements of a von Neumann algebra.

The problem is more involved in the case of a non-self-adjoint polynomial since the convergence in $*$-moments does not yield the convergence of the ESDs. In this case, the analogue of the spectral distribution for a non-normal element of a von Neumann algebra is the Brown measure [15], and the result one can expect for usual matrix models is the convergence of the empirical measure toward the Brown measure of the $*$-moments limit. Even the computation of the candidate Brown measure limit is a non-trivial task: there has been recent progress by Speicher, Mai, Belinschi and Sniady, who found an algorithm that gives such Brown measures using linearization techniques and subordination results [11, 12]; see also [13] for the computation of some specific Brown measures.

The convergence of ESDs of polynomials in independent matrices was so far tackled only in specific cases. The product of independent Ginibre matrices was studied by F. Götze, A. Naumov and A. Tikhomirov [52, 28] and S. O'Rourke and A. Soshnikov [43], as well as the sum of such products [32]. For products of Girko's elliptic random matrices see [42]. Yet, there are no general results for the convergence of the ESDs of non-self-adjoint polynomials in independent non-normal random matrices.

Part of the challenge is to develop an approach that can deal with all polynomials in a unified way. This difficulty has also been encountered in recent work on the local law for self-adjoint polynomials [22], which covers quadratic polynomials and symmetrized products of independent Wigner matrices; the specific case of the anti-commutator $W_1W_2 + W_2W_1$ of Wigner matrices was the subject of the earlier work of Anderson [3].

Here we develop a general approach for non-self-adjoint quadratic polynomials. Modulo a key reduction step, many of the essential ideas can be illustrated with the case of the anti-commutator $X_1X_2 + X_2X_1$ in i.i.d. matrices, as we do in Section 2. After lifting the matrix polynomials to

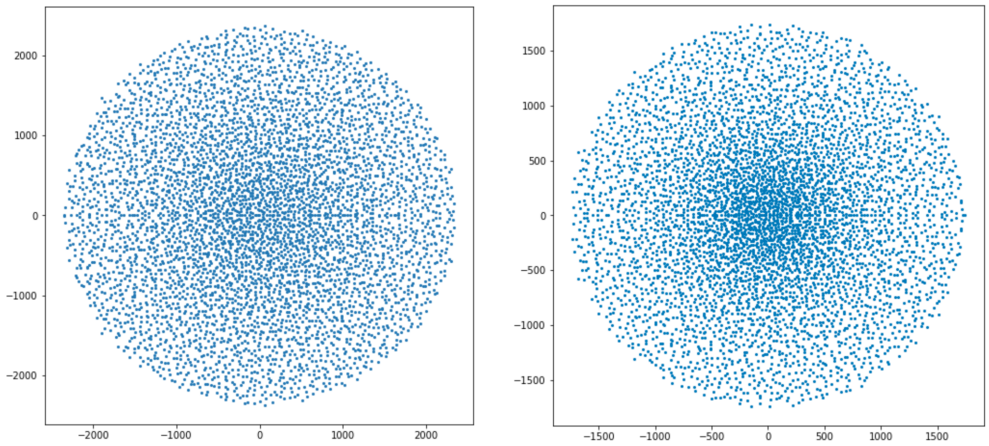

FIGURE 1. Simulated spectra of $X_1X_2+X_2X_1$ (left) and $X_1X_2-0.3X_2X_3+0.1X_3X_1$ (right) for independent 5000×5000 matrices X_1, X_2, X_3 with entries independently and uniformly drawn from $[-1, 1]$.

flattened random tensors of higher dimension, we encounter a new layer to the problem of quantitative invertibility of random matrices. The Ginibre hypothesis allows us to relatively efficiently get to what we feel are the core novel difficulties for non-self-adjoint polynomials. Generalizing to other distributions, particularly discrete, would involve extra considerations of a purely technical nature (many of which are by now standard); in order to keep the paper of reasonable length we do not pursue such extensions here. As with the local law for self-adjoint polynomials, the main outstanding challenge is to generalize to higher degree – we hope that some of the techniques developed here will be useful for this problem.

1.2. **Spectral convergence and the pseudospectrum.** Interestingly, the qualitative problem of establishing convergence of spectral measures to the Brown measure is intimately related to quantitative (finite-N) questions of interest in numerical analysis. The key difficulty for non-normal matrices lies in the instability of their eigenvalues: it is well known that for certain matrices even a tiny perturbation of a single entry can drastically change the spectrum.

One might expect however that *random* non-normal matrices have a more stable spectrum. A long line of works beginning with [47] has shown that in many cases, for a sequence of non-normal matrices converging in star moments, it is sufficient to perturb by a small random matrix (of vanishing norm) in order to "regularize" the ESDs and guarantee convergence to the Brown measure – see [47, 30, 58, 10, 55] and references therein. This regularizing effect has been exploited for applications in numerical analysis in the recent works [5, 6]. Yet, these results are based on independence of the entries of random matrices and there are, as quoted above, no general results when the entries start to be strongly coupled, for instance for the commutator of two independent Ginibre matrices.

We briefly sketch how quantitative measures of spectral instability arise in the study of limiting spectral distributions; a more formal discussion is deferred to Section 1.5. The starting point is to note that the empirical spectral distribution (ESD) μ_A of an $N\times N$ matrix A can be recovered from the Laplacian of the log-modulus of the characteristic polynomial: recalling (1.1), we have

$$\mu_A = \frac{1}{2\pi N}\sum_{i=1}^{N}\Delta_z \log|\lambda_i(A)-z| = \frac{1}{2\pi N}\Delta_z \log|\det(A-z)|.$$

On the other hand, the log-modulus of the characteristic polynomial can also be expressed

$$\frac{1}{N}\log|\det(A-z)| = \frac{1}{2N}\log\det[(A-z)(A-z)^*] = \frac{1}{2}\int_0^\infty \log x\, d\mu_{(A-z)(A-z)^*}(x)$$

and so we have the identity

$$\mu_A = \frac{1}{4\pi}\Delta_z \int_0^\infty \log x\, d\mu_{(A-z)(A-z)^*}(x) \tag{1.2}$$

expressing the ESD of a (possibly non-normal) matrix A in terms of the ESDs of the collection of Hermitian matrices $(A-z)(A-z)^*$ with $z \in \mathbb{C}$.

Now for a sequence A^N of $N \times N$ matrices converging in $*$-moments to an element a of a von Neumann algebra one has convergence of the ESDs $\mu_{(A^N-z)(A^N-z)^*}$ to the spectral measure $\mu_{(a-z)(a-z)^*}$, and so one might hope to have convergence of the ESDs μ_{A^N} to the measure μ_a obtained by substituting a for A on the right hand side of (1.2). This is not true in general: we have already mentioned that convergence in $*$-moments does not guarantee convergence of the spectral distributions, and indeed one notes that the hoped-for identity fails if there is escape of mass of $\mu_{(A^N-z)(A^N-z)^*}$ to the singularities of the logarithm at 0 and $+\infty$. However, we may take (1.2) as a reasonable guess, and indeed this amounts to the definition of the Brown measure ν_a (reviewed in Section 1.5 below).

We thus see that in order to access the limiting measure via (1.2) it is crucial to control the largest and smallest eigenvalues of $(A^N-z)(A^N-z)^*$. For random matrices it turns out that the more delicate task is to bound the smallest eigenvalue from below. Recall that the ε-pseudospectrum of a square matrix A is the set

$$\Lambda_\varepsilon(A) = \{z \in \mathbb{C} : \sigma_{\min}(A-z) \le \varepsilon\} \tag{1.3}$$

where $\sigma_{\min}(A) = \sqrt{\lambda_{\min}(AA^*)}$ denotes the smallest singular value of a matrix A. Alternatively, $\Lambda_\varepsilon(A)$ is the union of the spectra of $A+E$ over all perturbations E of spectral norm at most ε (we refer to [53, Chapter 2] for the demonstration of this equivalence). The pseudospectrum is an important object in numerical analysis that quantifies the stability of the spectrum under small perturbations. We refer to the textbook [53] for further background. To prove convergence of the ESDs μ_{A^N} one must show that $\Lambda_\varepsilon(A^N)$ has measure $o(1)$ for $\varepsilon = \varepsilon(N) \to 0$ not too fast (any polynomial order will be sufficient).

We point out that the problem of bounding $\sup_{z\in\mathbb{C}} \mathbb{P}\{\sigma_{\min}(A-z) \le \varepsilon\}$ for a random $N \times N$ matrix A generalizes the well-studied *anti-concentration problem* for scalar random variables, which is the case $N = 1$. Moreover, scalar concentration inequalities have played a fundamental role in the study of invertibility of random matrices.

To establish uniform integrability of the logarithm in (1.2) also requires some control on the k-th smallest singular value of $A^N - z$ for moderately small values of k (in the range $[N^{-c}, \delta N]$ for small fixed $c, \delta > 0$). For a polynomial P^N in Ginibre matrices this can be done thanks to local laws which can be found in this context in [31]. Hence, in the present work the main issue is to obtain control on the pseudospectrum of P^N. For this, the dependency structure of the polynomial presents significant obstacles that were not present in work on the circular law. As this structure is different for each polynomial, it is a challenge to develop an approach to deal with all polynomials in a unified way. For the quadratic case, the singular value decomposition of the associated quadratic form can be used to organize the dependencies among blocks in the associated tensor matrix in a certain common pattern (see Corollary 4.2). We comment further on the novel challenges for the proof after stating our results below.

1.3. **Main results.** In the sequel we write $\mathbb{C}\langle x_1, \ldots, x_n\rangle$ for the set of polynomials with complex coefficients in n non-commuting indeterminates $x_1, \ldots x_n$. We recall the following definition of the complex Ginibre ensemble:

Definition 1.1. A random matrix $X = X^N$ is an $N \times N$ complex Ginibre matrix if the families $(\sqrt{N}\Re(X_{i,j}))_{i,j\in[N]}$ and $(\sqrt{N}\Im(X_{i,j}))_{i,j\in[N]}$ are independent i.i.d. families of random variables of law $\mathcal{N}(0, 1/2)$.

Our first main result establishes convergence of the empirical spectral distribution of $\mathfrak{p}(X_1^N, \dots, X_n^N)$ for any quadratic non-commutative polynomial $\mathfrak{p}$.

Theorem 1.2 (Convergence to the Brown measure)**.** *Let $n \in \mathbb{N}$ and $\mathfrak{p} \in \mathbb{C}\langle x_1, \dots, x_n\rangle$ be a non-commutative polynomial of degree two. For each $N \in \mathbb{N}$ let $X_1^N, \dots, X_n^N$ be n independent $N \times N$ complex Ginibre matrices and set $P^N = \mathfrak{p}(X_1^N, \dots, X_n^N)$. Then*

$$\mu_{P^N} \to \nu_{\mathfrak{p}(c_1,\dots,c_n)}$$

weakly in probability, where $c_1, \dots, c_n$ are $*$-free circular elements of a W^*-probability space and $\nu_{\mathfrak{p}(c_1,\dots,c_n)}$ is the Brown measure of $\mathfrak{p}(c_1, \dots, c_n)$ (defined in Section 1.5 below).

As we described above, the main step for proving Theorem 1.2 is to control the pseudospectrum of P^N, which is accomplished in our second main result. Whereas Theorem 1.2 was stated for a sequence of matrices P^N of growing size, the following is a non-asymptotic result for matrices of any fixed size.

Theorem 1.3 (Control on the pseudospectrum)**.** *Let $N, n \in \mathbb{N}$ and $\mathfrak{p} \in \mathbb{C}\langle x_1, \dots, x_n\rangle$ be a non-commutative polynomial of degree two. There is an absolute constant $C_0 > 0$ and $C(\mathfrak{p}), c(\mathfrak{p}) > 0$ depending only on $\mathfrak{p}$ such that the following holds. Let $X_1, \dots, X_n$ be n independent $N \times N$ complex Ginibre matrices and set $P = \mathfrak{p}(X_1, \dots, X_n)$. For any $z \in \mathbb{C}$ and any $\varepsilon > 0$,*

$$\mathbb{P}\big\{\sigma_{\min}(P - z) \le \varepsilon\big\} \le C(N^{C_0}\varepsilon^c + e^{-N}).$$

Remark 1.4. Our proof shows one can take $C_0 = 13/3$ but we have not tried to optimize this constant. We obtain the dependence $c(\mathfrak{p}) = 1/(n_0 + 1)$, where n_0 is the rank of the quadratic form associated to the homogeneous degree-two part of $\mathfrak{p}$ (see Lemma 4.1). The constant C depends on $\mathfrak{p}$ only through the rank parameter n_0 and the size of the coefficients, quantified by the norm of the matrix $(s_0, \dots, s_{n_0})$ with columns s_k as in Corollary 4.2. The dependence on these two parameters is polynomial, but we have not tracked the precise order.

From the Fubini–Tonelli theorem and Markov's inequality, we obtain the following corollary on the density of the pseudospectrum (recall (1.3)).

Corollary 1.5. *With hypotheses in Theorem 1.3, for any Borel set $\Omega \subset \mathbb{C}$ and any $\varepsilon > 0$ we have*

$$\mathbb{E}\,\mathrm{Leb}(\Lambda_\varepsilon(P) \cap \Omega) \le C(N^{C_0}\varepsilon^c + e^{-N})\,\mathrm{Leb}(\Omega)$$

where Leb *denotes the Lebesgue measure on $\mathbb{C}$. In particular, for any $K > 0$ there exists $C(K, \mathfrak{p}) > 0$ such that for all N sufficiently large depending on $\mathfrak{p}$, for any fixed Borel set Ω of finite measure we have that with probability at least $1 - N^{-100}$,*

$$\mathrm{Leb}(\Lambda_{N^{-C}}(P) \cap \Omega) \le N^{-K}\,\mathrm{Leb}(\Omega).$$

1.4. **A matrix Littlewood–Offord problem.** Previous work on the invertibility and pseudospectrum of matrices with independent entries proceeds by reduction to an anti-concentration problem for scalar random walks. For given deterministic $w_0 \in \mathbb{C}$, $u = (u_1, \dots, u_N) \in \mathbb{C}^N$ and i.i.d. copies $\xi_1, \dots, \xi_N$ of a random variable $\xi \in \mathbb{C}$, the problem is to bound the probability that the scalar random variable

$$w = w_0 + \sum_{j=1}^N u_j\xi_j \tag{1.4}$$

lies in a small neighborhood of the origin. When ξ is Bernoulli this is the classic Littlewood–Offord problem [33]. Strong small-ball estimates for w depending on arithmetic structure among the scalars u_j were obtained in [51, 50, 44]. (For analogous results on the invertibility of symmetric Wigner matrices – first obtained in [20, 38, 54], with some notable recent improvements in [24, 17, 16] – the problem reduces to anti-concentration for quadratic forms, but these works make use of a bilinear reduction in order to apply the linear Littlewood–Offord theory.) Sharp extensions to the vector setting, with $w, w_0 \in \mathbb{R}^d$, $u \in \mathbb{R}^{d\times N}$ and ξ scalar, were obtained in [26, 45]. Extensions to scalar polynomials of higher degree have been obtained in [40, 36, 25]. The problem has also been studied in the setting of finite abelian groups; for further background we refer to the survey [41].

Our approach to Theorem 1.3 leads naturally to the following matrix generalization of the scalar anti-concentration problem. Indeed, our first step (see Section 2.2 for further discussion) is to apply the so-called linearization trick incepted by Haagerup and Thorbjørnsen in [31] to reduce to the problem of estimating the smallest singular value of a larger $Nd \times Nd$ block matrix $\boldsymbol{L}$ with dependent entries, for some bounded d determined by $\mathfrak{p}$, with coefficients given by linear combinations of the Ginibre matrices. The matrix $\boldsymbol{L}$ can be viewed as an $N \times N$ matrix whose $d \times d$ entries $L_{i,j}$ are independent random matrices, which themselves have correlated entries. In an analogous way to arguments for i.i.d. matrices we are then led to the following matrix generalization of the scalar anti-concentration problem.

Matrix Littlewood–Offord problem: *For given deterministic $W_0, U_1, \dots, U_N \in \mathbb{C}^{d\times d}$ and i.i.d. copies L_j of a $d \times d$ random matrix L, bound the probability that the matrix random walk*

$$W = W_0 + \sum_{j=1}^{N} U_j^* L_j \tag{1.5}$$

is close to the variety of singular matrices $\mathcal{S}_d = \{A \in \mathbb{C}^{d\times d} : \det A = 0\}$.

For Theorem 1.3 we specifically seek a bound of the form $N^C\varepsilon^c + e^{-N}$ on the probability that W is within distance ε of $\mathcal{S}_d$ (in the Hilbert–Schmidt norm, say), for some $C, c^{-1} = O_{d,L}(1)$ (depending only on d and the distribution of L), but it would be interesting to determine the optimal exponents.

As for the scalar problem there are obvious conditions under which anti-concentration is poor or completely fails. For instance, when $W_0 = 0$, one of the U_j is singular and the rest are zero, then $\mathbb{P}(W \in \mathcal{S}_d) = 1$. We also have $\mathbb{P}(W \in \mathcal{S}_d) = 1$ in the "simple" case that $L = \xi\, \mathrm{I}_d$ is a random scalar multiple of the identity and all of the matrices $W_0, U_1, \dots, U_N$ lie in a linear subspace contained in $\mathcal{S}_d$. The problem of classifying such subspaces was raised by Lovász in [34] and as far as we are aware there has been little progress since then. In our study of quadratic polynomials in random matrices we can avoid this case as L will have a very different structure, but we note that the case $L = \xi\, \mathrm{I}_d$ arises naturally when considering higher-degree monomials. Here, we have the opposite problem that L itself is singular with probability 1, so that in bounding the probability that W is close to $\mathcal{S}_d$ we must make use of the geometry of the sequence $U_1, \dots, U_N$.

For the smallest singular value problem, the matrices L_j are taken from an $Nd \times d$ column of $\boldsymbol{L}$, and the sequence of matrices $U_1, \dots, U_N$ are random and depending on the remaining, independent columns of the matrix. The main technical challenge for the proof of Theorem 1.3 is thus to rule out the case that the sequence $U_1, \dots, U_N$ has special structure. The structural events here, already present in the Gaussian case, are algebro-geometric in nature, orthogonal to arithmetic issues for the discrete scalar problem. Controlling these events requires delicate partitionings and coverings of the Stiefel manifold of orthonormal d-frames in $\mathbb{C}^{Nd}$ based on the structure of the polynomial $\mathfrak{p}$ (indeed, the dependency structure and support of the step variable L is determined by $\mathfrak{p}$).

In order to focus on this new layer to the problem of quantitative invertibility of random matrices, we consider here the case of Gaussian entries, which allows us to get to the heart of the issue in

a more transparent way; an extension to general, possibly discrete distributions would involve a combination of algebro-geometric and arithmetic structural considerations. We illustrate these new issues in Section 2 using the example of the anti-commutator $P = X_1X_2 + X_2X_1$.

1.5. **The Brown measure limit.** In this subsection we define the Brown measure appearing as the limit in Theorem 1.2. For further background on free probability we refer to [4, Chapter 5] and [37, Chapter 11].

Recall that a (tracial) W^*-probability space $(\mathcal{A}, \tau)$ is a von Neumann algebra $\mathcal{A}$ equipped with a tracial faithful normal state τ. According to Voiculescu's central result [56], the non-commutative distribution of $(X_1^N, \dots, X_n^N)$ converges in $*$-moments towards the non-commutative distribution of n free circular elements $(c_1, \dots, c_n)$ in $\mathcal{A}$ in the sense that for any polynomial $\mathfrak{p} \in \mathbb{C}\langle x_1, \dots, x_n, x_1^*, \dots, x_n^*\rangle$

$$\lim_{N\to\infty} \frac{1}{N}\mathrm{Tr}\left(\mathfrak{p}(X_1^N, \dots, X_n^N, (X_1^N)^*, \dots, (X_n^N)^*)\right) = \tau\left(\mathfrak{p}(c_1, \dots, c_n, c_1^*, \dots, c_n^*)\right) \qquad a.s.$$

The right hand side is a linear map on the set of polynomials in $\mathbb{C}\langle x_1, \dots, x_n, x_1^*, \dots, x_n^*\rangle$ which is uniquely defined by its value on monomials given, for any choices of $i_1, \dots, i_k \in [1, n]$ and any $\varepsilon_1, \dots, \varepsilon_k \in \{1, *\}$, by setting that $\tau\left(c_{i_1}^{\varepsilon_1} \cdots c_{i_k}^{\varepsilon_k}\right)$ is the number of non-crossing pair partitions of $\{i_1, \dots, i_k\}$ so that each block $b = (b_1, b_2)$ is such that $i_{b_1} = i_{b_2}$ and $(\varepsilon_{b_1}, \varepsilon_{b_2}) = (1, *)$ or $(*, 1)$. By density of polynomial functions in the set of continuous functions, and the fact that the Ginibre matrices are bounded with high probability, we see that the latter implies that for any bounded continuous function f

$$\lim_{N\to\infty} \frac{1}{N}\mathrm{Tr}\left(f\left(\mathfrak{p}(X_1^N, \dots, X_n^N, (X_1^N)^*, \dots, (X_n^N)^*)\right)\right) = \tau\left(f(\mathfrak{p}(c_1, \dots, c_n, c_1^*, \dots, c_n^*))\right) \quad a.s.$$

If $\mathfrak{p}$ is self-adjoint, this guarantees the convergence of the empirical spectral distribution of $\mathfrak{p}\left(X_1^N, \dots, X_n^N, (X_1^N)^*, \dots, (X_n^N)^*\right)$ towards the distribution of $\mathfrak{p}(c_1, \dots, c_n, c_1^*, \dots, c_n^*)$ which is specified by its moments as given above. If $\mathfrak{p}$ is non-self-adjoint, there is no such simple relation between the eigenvalues and moments. Girko's idea [27] to overcome this difficulty is based on Green's formula which states that for any twice continuously and bounded function ψ, any complex numbers $\lambda_i, 1 \le i \le N$,

$$\frac{1}{N}\sum_{i=1}^N \psi(\lambda_i) = \frac{1}{2\pi}\int \Delta\psi(z)\frac{1}{N}\sum_{i=1}^N \log|z - \lambda_i| dz \tag{1.6}$$

where dz denotes the Lebesgue measure on $\mathbb{C}$. Taking the $\lambda_i, 1 \le i \le N$ to be the eigenvalues of $Q := \mathfrak{p}\left(X_1^N, \dots, X_n^N, (X_1^N)^*, \dots, (X_n^N)^*\right)$, we deduce

$$\frac{1}{N}\sum_{i=1}^N \psi(\lambda_i) = \frac{1}{4\pi}\int \Delta\psi(z)\frac{1}{N}\mathrm{Tr}\left(\log|z - Q|^2\right) dz\,.$$

Noting that $|z - Q|^2$ is a self-adjoint polynomial in $X_1^N, \dots, X_n^N, (X_1^N)^*, \dots, (X_n^N)^*$, and neglecting the singularity and unboundedness of the logarithm, we are hence prompted to conjecture that the empirical measure of the eigenvalues of Q converges towards its Brown measure given for any twice continuously differentiable function ψ on $\mathbb{C}$ by

$$\int_{\mathbb{C}} \psi(\lambda) d\nu_{\mathfrak{p}(c_1,\dots,c_n,c_1^*,\dots,c_n^*)}(\lambda) := \frac{1}{4\pi}\int \Delta\psi(z)\tau\left(\log|z - \mathfrak{p}(c_1, \dots, c_n, c_1^*, \cdots, c_n^*)|^2\right) dz.$$

Note that the right hand side makes sense as soon as $\Delta\psi$ is nonnegative, taking at worst the value $-\infty$, since the circular elements $c_1, \dots, c_n$ are bounded.

1.6. **Organization of the paper.** The proof of Theorem 1.3 occupies the bulk of the paper; the deduction of Theorem 1.2 from Theorem 1.3 following the by-now standard Hermitization procedure is deferred to Section 8. In Section 2 we give an informal overview of the proof ideas for Theorem 1.3 using the example of the anti-commutator $P = X_1X_2 + X_2X_1$ in two Ginibre matrices. The formal proofs of Theorems 1.3 and 1.2 occupy the remainder of the paper. Section 3 summarizes our notational conventions and records a few standard facts. The proof of Theorem 1.3 occupies Sections 4–7.

2. Proof outline

The proof of Theorem 1.3 proceeds in four steps, carried out in Sections 4–7. Here we illustrate the main ideas for the special case of the *anti-commutator* $P = X_1X_2 + X_2X_1$ of independent Ginibre matrices X_1, X_2 (the same arguments apply with trivial modifications to the commutator $X_1X_2 - X_2X_1$). Note that Anderson [3] showed the convergence of the empirical measure of the eigenvalues of the anti-commutator $X_1X_2 + X_2X_1$ in two Wigner matrices, which is self-adjoint. For our conventions on asymptotic notation we refer to Section 3 below; in the simple setting of the anti-commutator we are able to avoid the more burdensome notation for block matrices described there that is useful for treating general polynomials.

2.1. **A word on the problem for a single i.i.d. matrix.** Previous works on the invertibility and pseudospectrum of a single i.i.d. matrix X proceed by reduction to the task of bounding the probability that some column of X, say the jth column, which we will denote by $\mathrm{col}_j(X)$, is close to the span $\mathsf{V}_{(j)}$ of the remaining $N-1$ columns. The independence of the columns allows one to condition on the remaining columns and focus on controlling the distance of a random vector to a fixed subspace. When the entries are i.i.d. Gaussian, by rotational invariance one may take the fixed space to be the span of the standard basis vectors $e_2, \dots, e_N$, and the distance is simply the magnitude of the first coordinate of $\mathrm{col}_j(X)$. The necessary control then follows from the boundedness of the Gaussian density.

For non-Gaussian entries, and in particular for discrete distributions, the problem is more complicated as the small ball probability for the distance depends on the position of the random hyperplane spanned by the remaining columns – in particular on arithmetic structure in a normal vector to the hyperplane. Such issues will not arise in the present work and we refer the interested reader to [44]. By taking the entries of our matrices to be Gaussian we can focus on the novel structural pathologies that can arise for algebraic reasons related to the form of the polynomial $\mathfrak{p}$.

2.2. **Step 1: Linearization.** When considering distances of columns to the span of remaining columns for the anti-commutator $P = X_1X_2 + X_2X_1$ in Ginibre matrices, we immediately encounter the problem of complete lack of independence of the entries. However, we can retain some independence by using the linearization trick of Haagerup and Thorbjørnsen [31]: one verifies with the Schur complement formula that $(P-z)^{-1}$ is the top-left $N \times N$ block of $(\boldsymbol{L}^z)^{-1}$, where $\boldsymbol{L}^z$ is the $3N \times 3N$ block matrix

$$\boldsymbol{L}^z = \begin{pmatrix} -z & X_1 & X_2 \\ X_2 & -\mathrm{I} & 0 \\ X_1 & 0 & -\mathrm{I} \end{pmatrix}. \tag{2.1}$$

In particular we have the deterministic bound

$$\|(P-z)^{-1}\|_{\mathrm{op}} \le \|(\boldsymbol{L}^z)^{-1}\|_{\mathrm{op}} \tag{2.2}$$

and so it suffices to show

$$\mathbb{P}\{\sigma_{\min}(\boldsymbol{L}^z) \le \varepsilon\} \lesssim N^C\varepsilon^c + e^{-N}. \tag{2.3}$$

The advantage of this new problem is that $\boldsymbol{L}^z$ can be viewed as an $N \times N$ matrix with entries that are independent 3×3 random matrices:

$$L^z_{i,j} = \begin{pmatrix} -z\delta_{i,j} & X_1(i,j) & X_2(i,j) \\ X_2(i,j) & -\delta_{i,j} & 0 \\ X_1(i,j) & 0 & -\delta_{i,j} \end{pmatrix}, \qquad i,j \in [N]. \tag{2.4}$$

(To avoid ambiguity in the proof we will actually write $\widetilde{\boldsymbol{L}}^z$ for the element of $\mathbb{M}_N(\mathbb{M}_3(\mathbb{C}))$ associated to P, whereas we write $\boldsymbol{L}^z$ for the associated element of $\mathbb{M}_3(\mathbb{M}_N(\mathbb{C}))$ as represented in (2.1), but we avoid this notation here.)

2.3. **Step 2: Reduction to a random matrix of bounded dimension.** Now we write $\widehat{L}^z_j$ for 3×3 matrix obtained by projecting the three columns of $\mathrm{col}_j(\boldsymbol{L}^z) =: (L^z_{i,j})_{i=1}^N \in \mathbb{M}_3(\mathbb{C})^N \cong \mathbb{M}_{3N,3}(\mathbb{C})$ to the orthogonal complement of the span $\mathsf{V}^z_{(j)}$ of the remaining $3N-3$ columns. Specifically, we let $\boldsymbol{U}_j \in \mathbb{M}_{3N,3}(\mathbb{C})$ have columns that are orthonormal in $(\mathsf{V}^z_{(j)})^\perp$, and put

$$\widehat{L}^z_j = \boldsymbol{U}_j^* \,\mathrm{col}_j(\boldsymbol{L}^z) = \sum_{i=1}^N U^*_{i,j} L^z_{i,j} \tag{2.5}$$

where $U_{i,j} \in \mathbb{M}_3(\mathbb{C})$ are the corresponding 3×3 blocks of $\boldsymbol{U}_j$. (We remark that, *a posteriori*, with high probability $(\mathsf{V}^z_{(j)})^\perp$ has dimension 3 and the columns of $\boldsymbol{U}_j$ are in fact a basis.) The key is that $\mathrm{col}_j(\boldsymbol{L}^z)$ is independent of $\mathsf{V}^z_{(j)}$, and hence $\boldsymbol{U}_j$ can be chosen independently of $\mathrm{col}_j(\boldsymbol{L}^z)$.

Now on the event that $\sigma_{\min}(\boldsymbol{L}^z) \le \varepsilon$, there exists a unit vector $\boldsymbol{v} = (v_j)_{j=1}^N \in (\mathbb{C}^3)^N$ such that

$$\varepsilon \ge \|\boldsymbol{L}^z \boldsymbol{v}\|_2 = \Big\| \sum_{j=1}^N \mathrm{col}_j(\boldsymbol{L}^z) v_j \Big\|_2.$$

Since $\boldsymbol{v}$ has unit norm there must be some $j_0 \in [N]$ such that $\|v_{j_0}\|_2 \ge N^{-1/2}$. Projecting to the orthocomplement of $V_{(j_0)}$ we obtain

$$\varepsilon \ge \Big\| \sum_{j=1}^N \boldsymbol{U}^*_{j_0} \,\mathrm{col}_j(\boldsymbol{L}^z) v_j \Big\|_2 = \|\boldsymbol{U}^*_{j_0} \,\mathrm{col}_{j_0}(\boldsymbol{L}^z) v_{j_0}\|_2 = \|\widehat{L}^z_{j_0} v_{j_0}\|_2 \ge N^{-1/2} \sigma_{\min}(\widehat{L}^z_{j_0}).$$

After applying a union bound to fix j_0 (and using symmetry to assume $j_0 = 1$) we reduce to showing

$$\mathbb{P}\{\sigma_{\min}(\widehat{L}^z_1) \le \varepsilon\} \lesssim N^C \varepsilon^c + e^{-N} \tag{2.6}$$

for an adjusted constant C.

2.4. **Step 3: Gaussian polynomial anti-concentration.** It will be convenient to instead bound the lower tail of $|\det(\widehat{L}^z_1)| = \prod_{k=1}^3 \sigma_k(\widehat{L}^z_1)$ rather than the smallest singular value. Since $\widehat{L}^z_1$ has bounded norm $O(1)$ with probability $1 - O(e^{-N})$ (a property inherited from $\boldsymbol{L}^z$, which in turn inherits this from X_1 and X_2 via the triangle inequality), we have $|\det(\widehat{L}^z_1)| \lesssim \sigma_{\min}(\widehat{L}^z_1)$, so it suffices to show

$$\mathbb{P}\{|\det(\widehat{L}^z_1)| \le \varepsilon\} \lesssim N^C \varepsilon^c + e^{-N}. \tag{2.7}$$

We condition on an arbitrary realization of $\boldsymbol{U}_1$ (recall it is independent of $\mathrm{col}_1(\boldsymbol{L}^z)$). From (2.5) we can view the random matrix $\widehat{L}^z_1$ as a random walk in $\mathbb{M}_3(\mathbb{C})$ with independent steps $U^*_{i,1} L^z_{i,1}$, and our aim is to control the probability that this walk lands in a small neighborhood of the codimension-one variety of singular matrices.

For a single i.i.d. matrix X the analogous problem of bounding $\mathbb{P}\{|u^* \,\mathrm{col}_1(X)| \le \varepsilon\}$ is easily handled when the entries $X(i,j)$ have bounded density, as one can condition on all components but some j_0 for which u_{j_0} is not too small. In the present setting we cannot reduce so easily to consideration of a single step of the walk. Indeed, (2.7) fails to hold for general $\boldsymbol{U}_1$: consider for

instance the case all of the $U_{i,1}$ are zero except for a single $U_{i_0,1}$ with $i_0 \geq 2$, in which case $\det(\widehat{L}_1^z) = \det(U_{i_0,1})\det(L_{i_0,1}) \equiv 0$. On the other hand, it seems unlikely that the random orthonormal set $\boldsymbol{U}_1$ in $(\mathsf{V}_{(1)}^z)^\perp$ would concentrate on a single submatrix $U_{i_0,1}$.

We highlight in particular a key challenge for establishing (2.7): one notes from (2.4) that the random matrices $L_{i,1}^z$ are themselves singular for $i \geq 2$. Thus, the invertibility of $\widehat{L}_1^z$ cannot rely on the randomness of the $L_{i,1}^z$ alone, but must also come from the geometry of the sequence $(U_{1,1}, \ldots, U_{N,1})$.

We proceed to identify sufficient structural conditions on the matrix $\boldsymbol{U}_1$ in order to have (2.7). Let us denote the columns of $U_{i,1}^*$ by $v_i^0, v_i^1, v_i^2 \in \mathbb{C}^3$. For compactness of notation we write $\xi_i^a = X_a(i,1)$, $a = 1, 2$, for the entries of X_1, X_2 that enter in $\mathrm{col}_1(\boldsymbol{L}^z)$. For simplicity of exposition we ignore here the deterministic shift in $L_{1,1}^z$ and focus on the random walk

$$W = \sum_{i=1}^N U_{i,1}^* L_{i,1}\,, \qquad L_{i,1} = \begin{pmatrix} 0 & \xi_i^1 & \xi_i^2 \\ \xi_i^2 & 0 & 0 \\ \xi_i^1 & 0 & 0 \end{pmatrix}.$$

For a single step of the matrix random walk we have

$$U_{i,1}^* L_{i,1} = \big(\xi_i^2 v_i^1 + \xi_i^1 v_i^2\,,\ \xi_i^1 v_i^0\,,\ \xi_i^2 v_i^0\big).$$

Expanding $\det(W)$ using multilinearity of the determinant, we have

$$p(\xi^1, \xi^2) = \det(W) = \sum_{i_0, i_1, i_2 \in [N]} \big\{\xi_{i_0}^2 \xi_{i_1}^1 \xi_{i_2}^2 \det(v_{i_0}^1, v_{i_1}^0, v_{i_2}^0) + \xi_{i_0}^1 \xi_{i_1}^1 \xi_{i_2}^2 \det(v_{i_0}^2, v_{i_1}^0, v_{i_2}^0)\big\}$$

where we view the determinant as a degree-3 polynomial p in the i.i.d. complex Gaussian variables $\xi^1 = (\xi_i^1)_{i\in[N]}, \xi^2 = (\xi_i^2)_{i\in[N]}$. Now a simple consequence of the Carbery–Wright inequality (cf. Lemma 6.2) gives an anticoncentration bound of the form

$$\mathbb{P}\{|p(\xi^1, \xi^2)| \leq \varepsilon\} \lesssim N^{O(1)} \varepsilon^{1/3} \tag{2.8}$$

as soon as we can find a monomial of p with coefficient of size at least $N^{-O(1)}$. It will be sufficient to focus on the coefficients of the $2N^2$ monomials $\xi_i^1(\xi_j^2)^2$ and $(\xi_i^1)^2\xi_j^2$ for $1 \leq i, j \leq N$, which are

$$\Delta_{i,j}^1 := \det(v_j^1, v_i^0, v_j^0)\,, \qquad \Delta_{i,j}^2 := \det(v_i^2, v_i^0, v_j^0). \tag{2.9}$$

2.5. **Step 4: Ruling out structured bases.** We hence obtain (2.7) as soon as either $|\Delta_{i,j}^1| = N^{-O(1)}$ or $|\Delta_{i,j}^2| = N^{-O(1)}$ for some $i, j \in [N]$. We thus say that the set $\boldsymbol{U}_1$ of orthonormal columns in $(\mathsf{V}_{(1)}^z)^\perp$ is *structured* if

$$|\Delta_{i,j}^1|,\, |\Delta_{i,j}^2| \leq N^{-\gamma} \qquad \forall\, i, j \in [N] \tag{2.10}$$

for some $\gamma > 0$ to be taken sufficiently large, and aim to show that, in the randomness of the remaining columns $\mathrm{col}_j(\boldsymbol{L}^z)$, $2 \leq j \leq N$, $\boldsymbol{U}_1$ is unstructured except with probability $1 - O(e^{-N})$.

Previous works on the invertibility of i.i.d. matrices have controlled the event that a set of columns has a structured normal vector using net arguments, and we do the same here. However, the notion of structure here is quite different from the arithmetic structure encountered in those works, as it involves an orthonormal set rather than a single normal vector, and involves relations between a set of tuples (in this case triples) of coordinates that is determined by the polynomial $\mathfrak{p}$. Indeed, one notes that the pattern of indices in (2.9) results from the form of the linearization (2.1), which will be different for other polynomials (and in fact there are multiple linearizations one can consider for any given polynomial).

To bound the probability that $\boldsymbol{U}_1$ satisfies (2.10) we aim to construct an ε-net $\mathcal{N}$ for the set of all possible structured orthonormal bases $\boldsymbol{U} = (U_1, \ldots, U_N) \in \mathbb{M}_3(\mathbb{C})^N \cong \mathbb{M}_{3N,3}(\mathbb{C})$. Since $\boldsymbol{U}_1^* \mathrm{col}_j(\boldsymbol{L}^z) = 0_{3\times 3}$ for every $2 \leq j \leq N$, then if $\boldsymbol{U}_1$ is approximated within ε (in the Hilbert–Schmidt metric) by an element $\widehat{\boldsymbol{U}}$ of the net, we have $\|\widehat{\boldsymbol{U}}^* \mathrm{col}_j(\boldsymbol{L}^z)\|_{\mathrm{HS}} = O(\varepsilon)$ for all $2 \leq j \leq N$

(here we use that $\boldsymbol{L}^z$ has bounded operator norm with high probability). Viewing the 3×3 random matrix $\widehat{\boldsymbol{U}}^* \operatorname{col}_j(\boldsymbol{L}^z)$ as a random vector in $\mathbb{C}^9$, after a bit of algebra one sees

$$\|\widehat{\boldsymbol{U}}^* \operatorname{col}_j(\boldsymbol{L}^z)\|_{\mathrm{HS}} = \|w_0 + \boldsymbol{W}^*\boldsymbol{\xi}\|_2$$

for a deterministic vector $w_0 \in \mathbb{C}^9$ depending on $\widehat{\boldsymbol{U}}$ and z, where

$$\boldsymbol{\xi} = (X_1(1,j),\dots,X_1(N,j),X_2(1,j),\dots,X_2(N,j)) \in \mathbb{C}^{2N}$$

is a Gaussian vector and

$$\boldsymbol{W} = \begin{pmatrix} \widehat{U}^2 & \widehat{U}^0 & 0 \\ \widehat{U}^1 & 0 & \widehat{U}^0 \end{pmatrix} \in \mathbb{M}_{2N,9}(\mathbb{C}) \tag{2.11}$$

with $\widehat{U}^0, \widehat{U}^1, \widehat{U}^2 \in \mathbb{M}_{N,3}(\mathbb{C})$ denoting the blocks of $\widehat{\boldsymbol{U}}$ under the partitioning of coordinates as in (2.1). Thus, the small ball probability for the matrix random walk $\widehat{\boldsymbol{U}}^* \operatorname{col}_j(\boldsymbol{L}^z) \in \mathbb{M}_3(\mathbb{C})$ is controlled by that of the Gaussian vector $\boldsymbol{W}^*\boldsymbol{\xi} \in \mathbb{C}^9$, which in turn is determined by the effective rank r of $\boldsymbol{W}$ (one can take r to be the number of singular values of $\boldsymbol{W}$ that exceed some small fixed threshold).

Note that the conditions (2.10) are geometric in nature, as they imply that a large collection of triples of the rows are not in general position, but lie in some hyperplane in $\mathbb{C}^3$. This is what allows for the construction of an efficient net for the set Struct of structured bases $\boldsymbol{U}$. Supposing we can find a net of size $|\mathcal{N}| = O(1/\varepsilon^2)^{dN}$, where $0 \le d \le 9$ is the effective metric dimension of Struct (which is a subset of the $9N$-dimensional complex vector space $\mathbb{M}_{3N,3}(\mathbb{C})$, so that this estimate holds trivially for $d = 9$) we can then bound the probability that $\boldsymbol{U}_1$ is structured by

$$\sum_{\widehat{\boldsymbol{U}}\in\mathcal{N}} \mathbb{P}\{\|\widehat{\boldsymbol{U}}^* \operatorname{col}_j(\boldsymbol{L}^z)\|_{\mathrm{HS}} \le C\varepsilon\} \le O(1/\varepsilon^2)^{dN} \cdot O(\varepsilon^2)^{r(N-1)} = O(\varepsilon^2)^{2(r-d)N-2r}.$$

We thus hope to show that the effective rank r of matrices $\boldsymbol{W}$ as in (2.11) is strictly larger than the effective metric dimension d of Struct. Note that the former is determined by relations between the columns of $\widehat{\boldsymbol{U}}$, while the latter is determined by relations between the rows.

In fact one cannot show the effective rank r is strictly larger than the effective dimension d uniformly for all structured $\boldsymbol{U}$ – we will instead need to stratify the set Struct according to the effective rank and use the geometric relations (2.10) between rows together with the rank constraint to construct an efficient net on each stratum. We defer further explanation of this step to Section 7.

3. Notation and preliminaries

3.1. **Asymptotic notation.** We use C, C_0, c, c', etc. to denote positive, finite constants. If the constants depend on parameters $p, q, \dots$ we indicate this by writing $C = C(p, q, \dots)$. If no dependence on parameters is given then the constants are understood to be absolute.

For $g \in \mathbb{R}_+$ we write $O(g)$ to stand for a quantity $f \in \mathbb{C}$ such that $|f| \le Cg$ for some absolute constant $C \in (0, +\infty)$. For $f, g \in \mathbb{R}_+$, $f \lesssim g$ and $g \gtrsim f$ mean that $f = O(g)$. For a parameter (or list of parameters) q, we write $f = O_q(g)$ to mean $|f| \le Cg$ for some $C = C(q) \in (0, +\infty)$, and similarly for $f \lesssim_q g, f \gtrsim_q g$.

3.2. **Matrices.** For integers $n \ge m \ge 1$ we use the common abbreviations $[m, n] := \{m, \dots, n\}$ and $[n] := [1, n]$. We write $|I|$ for the cardinality of a finite set I, and for $n < |I|$ we denote by $\binom{I}{n}$ the set of subsets of I with cardinality n.

We write $\mathbb{M}_{n,m}(\mathcal{X})$ for the set of $n\times m$ matrices with entries in a set $\mathcal{X}$, and abbreviate $\mathbb{M}_n(\mathcal{X}) := \mathbb{M}_{n,n}(\mathcal{X})$. The $n \times n$ identity matrix is denoted I_n, with the subscript sometimes omitted if it is clear from the context. For $a \in \mathbb{C}$ we often write a for $a\,\mathrm{I}_n$ when there can be no confusion.

We frequently work with block matrices, and the following more general matrix notation facilitates referral to their entries and submatrices of various dimensions. For finite indexing sets S, T

we write $\mathbb{M}_S^T(\mathcal{X})$ for the set of $|S| \times |T|$ matrices with entries in $\mathcal{X}$ and rows and columns indexed by S and T, respectively. Thus, $\mathbb{M}_{n,m}(\mathcal{X}) = \mathbb{M}_{[n]}^{[m]}(\mathcal{X})$. We write $\mathbb{M}_S(\mathcal{X}) := \mathbb{M}_S^S(\mathcal{X})$ when there can be no confusion (we avoid this when S is itself a product set). For $A \in \mathbb{M}_S^T(\mathcal{X})$ and $(s,t) \in S \times T$ we write $A(s,t)$ for the (s,t) entry of A. A block matrix $\boldsymbol{A}$ that is an $n \times n'$ array of blocks $A^{k,\ell} \in \mathbb{M}_{N,N'}(\mathcal{X})$ is viewed as an element of

$$\mathbb{M}_{[n]\times[N]}^{[n']\times[N']}(\mathcal{X})$$

and is naturally associated with an element of $\mathbb{M}_{n,n'}(\mathbb{M}_{N,N'}(\mathcal{X}))$; we abusively write $\boldsymbol{A}$ for both. Thus,

$$\boldsymbol{A}(k,\ell) = A^{k,\ell}, \qquad \boldsymbol{A}((k,i),(\ell,j)) = A^{k,\ell}(i,j) \tag{3.1}$$

for $(k,\ell) \in [n] \times [n']$ and $(i,j) \in [N] \times [N']$. We will sometimes index the rows by $I \times J$ for some other sets $I, J \subset \mathbb{Z}_{\geq 0}$ of size n, N, respectively, and similarly for the columns. It will often be convenient to view a block matrix "inside-out" – that is, $\boldsymbol{A}$ is naturally associated to an element

$$\widetilde{\boldsymbol{A}} \in \mathbb{M}_{[N]\times[n]}^{[N']\times[n']}(\mathcal{X}) \cong \mathbb{M}_{N,N'}(\mathbb{M}_{n,n'}(\mathcal{X})) \tag{3.2}$$

with $\widetilde{\boldsymbol{A}}((i,k),(j,\ell)) = \boldsymbol{A}((k,i),(\ell,j))$.

Typically, one of the pairs (N, N') will contain a "large" dimension (usually $(N,N') \in \{(N,N),(N,N-1),(N,1)\}$, with N the size of the Ginibre matrices in Theorems 1.2 and 1.3) while the other pair (n,n') remains bounded (being related to parameters of the fixed polynomial $\mathfrak{p}$). In this case we always use boldface for the block matrix $\boldsymbol{A}$, with its large submatrices $A^{k,\ell}$ indexed with superscripts and its small submatrices indexed with subscripts:

$$\widetilde{\boldsymbol{A}}(i,j) = A_{i,j}\,, \qquad (i,j) \in [N] \times [N']. \tag{3.3}$$

Thus

$$\boldsymbol{A}((k,i),(\ell,j)) = A^{k,\ell}(i,j) = A_{i,j}(k,\ell) = \widetilde{\boldsymbol{A}}((i,k),(j,\ell)). \tag{3.4}$$

We always use indices i, j, i_0, i_1, etc. for the large dimensions and k, ℓ, etc. for the small dimensions.

For $A \in \mathbb{M}_S^T(\mathcal{X})$, we denote its rows and columns

$$\mathrm{row}_s(A) = (A(s,t))_{t\in T} \in \mathcal{X}^T, \qquad \mathrm{col}_t(A) = (A(s,t))_{s\in S} \in \mathcal{X}^S.$$

We sometimes manipulate these as row and column vectors (note that the entries may be matrices). Thus, for $\boldsymbol{A} = (A^{k,\ell}) \in \mathbb{M}_{n,n'}(\mathbb{M}_{N,N'}(\mathbb{C}))$ with $\widetilde{\boldsymbol{A}} = (A_{i,j}) \in \mathbb{M}_{N,N'}(\mathbb{M}_{n,n'}(\mathbb{C}))$ we have that

$$\mathrm{col}_\ell(\boldsymbol{A}) = (A^{1,\ell}, \dots, A^{n,\ell}) \in \mathbb{M}_{N,N'}(\mathbb{C})^n, \qquad \ell \in [n']$$
$$\mathrm{col}_j(\widetilde{\boldsymbol{A}}) = (A_{1,j}, \dots, A_{N,j}) \in \mathbb{M}_{n,n'}(\mathbb{C})^N, \qquad \ell \in [N']$$

are vectors of matrices, whereas

$$\mathrm{col}_{(\ell,j)}(\boldsymbol{A}) = (\boldsymbol{A}((k,i),(\ell,j)))_{(k,i)\in[n]\times[N]} \in \mathbb{C}^{[n]\times[N]}, \qquad (\ell,j) \in [n'] \times [N']$$
$$\mathrm{col}_{(j,\ell)}(\widetilde{\boldsymbol{A}}) = (\boldsymbol{A}((k,i),(\ell,j)))_{(i,k)\in[N]\times[n]} \in \mathbb{C}^{[N]\times[n]}, \qquad (j,\ell) \in [N'] \times [n']$$

are arrays of scalars, which we view as vectors with block coordinate structure.

For $A \in \mathbb{M}_S^T(\mathcal{A})$ and $B \in \mathbb{M}_T^U(\mathcal{A})$ with $\mathcal{A}$ a $*$-algebra we define the product

$$AB = \big(\mathrm{row}_s(A) \cdot \mathrm{col}_u(B)\big)_{s\in S, u\in U} = \Big(\sum_{t\in T} A(s,t)B(t,u)\Big)_{s\in S,u\in U} \in \mathbb{M}_S^U(\mathcal{A})$$

and the conjugate transpose

$$A^* = \big(A(t,s)^*\big)_{t\in T, s\in S} \in \mathbb{M}_T^S(\mathcal{A})$$

in the usual way. For $A \in \mathbb{M}_S^T(\mathcal{A})$ and $B \in \mathbb{M}_U^V(\mathcal{A})$ we denote the tensor product $A \otimes B \in \mathbb{M}_{S\times U}^{T\times V}(\mathcal{A})$ with entries

$$(A \otimes B)((s,u),(t,v)) = A(s,t)B(u,v).$$

3.3. **Norms and singular values.** For finite indexing sets S, T we equip $\mathbb{C}^T$ with the Euclidean inner product $\langle \cdot, \cdot \rangle$ and ℓ_2-norm $\| \cdot \|_2$, and $\mathbb{M}_S^T(\mathbb{C})$ with the Hilbert–Schmidt norm

$$\|A\|_{\mathrm{HS}} = \Big(\sum_{s \in S, t \in T} |A(s,t)|^2 \Big)^{1/2}.$$

We let $\mathbb{B}^T \subset \mathbb{C}^T$ and $\mathbb{B}_S^T \subset \mathbb{M}_S^T(\mathbb{C})$ denote the closed unit balls under the Euclidean ℓ_2 and Hilbert–Schmidt norms, respectively. We denote the boundary of $\mathbb{B}^T$ by $\mathbb{S}^T$. For $T = [n]$ we write $\mathbb{B}^n$, $\mathbb{S}^{n-1}$. We sometimes write $\mathrm{dist}(u, v) := \|u - v\|_2$ for the Euclidean metric on $\mathbb{C}^T$. We also equip $\mathbb{M}_S^T(\mathbb{C})$ with the $\ell_2(T) \to \ell_2(S)$ operator norm

$$\|A\|_{\mathrm{op}} = \sup_{v \in \mathbb{S}^T} \|Av\|_2 .$$

For $A \in \mathbb{M}_S^T(\mathbb{C})$ we label its singular values in non-increasing order:

$$\|A\|_{\mathrm{op}} = \sigma_1(A) \geq \cdots \geq \sigma_{|S| \vee |T|}(A) \geq 0 ,$$

where the last $|S| \vee |T| - |S| \wedge |T|$ singular values are trivially zero, and denote by

$$\sigma_{\min}(A) := \sigma_{|S| \wedge |T|}(A) \geq 0$$

the smallest nontrivial singular value (which may be zero). In particular, for $|S| = |T|$ we have that A is invertible if and only if $\sigma_{\min}(A) > 0$. For a square block matrix $\boldsymbol{A} \in \mathbb{M}_{[n] \times [N]}^{[n] \times [N]}(\mathbb{C})$ we have $\sigma_l(\boldsymbol{A}) = \sigma_l(\widetilde{\boldsymbol{A}})$ for all $1 \leq l \leq nN$, as the singular values are invariant under relabeling of the rows and columns.

For subsets $E_1, \ldots, E_m$ of a vectors space we write $\mathrm{Span}(E_1, \ldots, E_m) := \mathrm{Span}(E_1 \cup \cdots E_m)$ for the linear span of their union. With slight abuse we allow some of the E_k to be single points, understood to denote the singleton sets $\{E_k\}$. For $A \in \mathbb{M}_S^T(\mathbb{C})$, by $\mathrm{Span}(A)$ we mean the linear span of its columns in $\mathbb{C}^S$. For vectors $v_j \in \mathbb{C}^S$, $j \in J$ we sometimes write $\langle v_j : j \in J \rangle = \mathrm{Span}(\{v_j\}_{j \in J})$ and $\langle v_j \rangle = \mathrm{Span}(v_j)$. $v_j \wedge v_k$ denotes the wedge product of v_j and v_k.

3.4. **Basic facts.** Recall that an *ε-net* for a subset E of a metric space $(\mathcal{X}, d)$ is a finite set $\Sigma \subset E$ such that

$$\sup_{x \in E} \min_{y \in \Sigma} d(x, y) < \varepsilon.$$

In the sequel all nets will be with respect to the appropriate Euclidean metric (which for matrices coincides with the Hilbert–Schmidt norm). The following is standard:

Lemma 3.1. *There is an absolute constant $C > 0$ such that for any subset E of the closed ball of radius R in $\mathbb{C}^d$ with Euclidean metric and any $\varepsilon \in (0, R)$, there is an ε-net for E of cardinality at most $(CR/\varepsilon)^{2d}$.*

The following two lemmas provide control on the norm and the norm of the inverse of the individual Ginibre inputs $X_1, \ldots, X_n$ for the polynomial P.

Lemma 3.2. *There is a constant $C_0 < \infty$ such that the following holds. For any $N \geq 1$ and X an $N \times N$ Ginibre matrix,*

$$\mathbb{P}\{\|X\|_{\mathrm{op}} > C_0\} \lesssim e^{-N}.$$

Proof. This holds more generally for matrices with independent complex uniformly-sub-Gaussian entries; see for instance [44] (one reduces to the real Ginibre case with the triangle inequality). □

Lemma 3.3. *Let $N \geq 1$ and X an $N \times N$ Ginibre matrix. For any deterministic $M \in \mathbb{M}_N(\mathbb{C})$ and any $\varepsilon > 0$*

$$\mathbb{P}\{\sigma_{\min}(X + M) \leq \varepsilon\} \lesssim N^3 \varepsilon^2 ,$$

where we emphasize that (following our previously stated convention) the implied constant is absolute and in particular is independent of the shift M.

We include the short proof below for completeness. We remark in passing that Edelman obtained the sharp bound of $N^2\varepsilon^2$ (with no constant factor loss) on the left hand side for the unshifted case $M = 0$ [21]. The sharpening to $O(N^2\varepsilon^2)$ for the shifted case can be found in [5] (cf. Lemma 3.3 there), extending the analogous bound of $O(N\varepsilon)$ for real shifts of real Ginibre matrices from [48].

Proof. Write $A = X + M$. On the event that $\sigma_{\min}(A) \leq \varepsilon$, there exists $u \in \mathbb{S}^{N-1}$ such that $\|Au\|_2 \leq \varepsilon$. Moreover, since u must have a coordinate i such that $|u_i| \geq 1/\sqrt{N}$, we obtain after projecting the vector Au to the orthocomplement of $A_{-i} := \mathrm{Span}\{\mathrm{col}_j(A) : j \neq i\}$ that

$$\mathrm{dist}(\mathrm{col}_i(A), A_{-i}) \leq \varepsilon\sqrt{N}.$$

By rotational invariance of the distribution of $\mathrm{col}_i(X)$, the probability of the above event is equal to

$$\mathbb{P}\{\mathrm{dist}(\mathrm{col}_i(X) + y, \mathrm{Span}\{e_2, \ldots, e_N\}) \leq \varepsilon\sqrt{N}\}$$

for a vector y that depends only on M and A_{-i}, and hence is independent of $\mathrm{col}_i(X)$. Conditioning on the columns $\mathrm{col}_j(X)$ with $j \neq i$, the above is equal to

$$\mathbb{P}\{|X(1,i) + y(1)| \leq \varepsilon\sqrt{N}\} \lesssim \varepsilon^2 N^2,$$

where we used that $\sqrt{N} \cdot X(1,i) \sim \mathcal{N}_{\mathbb{C}}(0,1)$ has bounded density on $\mathbb{C}$. Taking a union bound over the possible choices of i yields the claim. □

4. Linearization

Recall that $\mathbb{C}\langle \boldsymbol{x} \rangle = \mathbb{C}\langle x_1, \ldots, x_n \rangle$ is the set of polynomials with complex coefficients in n non-commuting indeterminates $x_1, \ldots, x_n$. We express $\mathfrak{p} \in \mathbb{C}\langle \boldsymbol{x} \rangle$ of degree two as

$$\mathfrak{p}(\boldsymbol{x}) = \sum_{\ell,m=1}^{n} a_{\ell,m} x_\ell x_m + \sum_{\ell=1}^{n} b_\ell x_\ell + a_0. \tag{4.1}$$

Lemma 4.1 (Linearization)**.** *Let $\mathcal{A}$ be a unital $*$-algebra over $\mathbb{C}$, let $x_1, \ldots, x_n \in \mathcal{A}$, and let $\mathfrak{p} \in \mathbb{C}\langle \boldsymbol{x} \rangle$ have degree two. Let $n_0 \in [n]$ be the rank of the matrix $A = (a_{\ell,m})$ as in* (4.1)*. There exist vectors $r_1, \ldots, r_{n_0}, s_0, s_1, \ldots, s_{n_0} \in \mathbb{C}^n$ such that $r_1, \ldots, r_{n_0}$ are orthonormal and $s_1, \ldots, s_{n_0}$ are nonzero and orthogonal (note we leave out s_0), and the following holds. Define the linear mapping*

$$\mathcal{L} = \mathcal{L}_{\mathfrak{p}} : \mathcal{A}^n \to \mathbb{M}_{[0,n_0]}(\mathcal{A}), \qquad \mathcal{L}(\boldsymbol{x}) := \begin{pmatrix} \langle s_0, \boldsymbol{x} \rangle & \langle r_1, \boldsymbol{x} \rangle & \cdots & \langle r_{n_0}, \boldsymbol{x} \rangle \\ \langle s_1, \boldsymbol{x} \rangle & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \langle s_{n_0}, \boldsymbol{x} \rangle & 0 & \cdots & 0 \end{pmatrix} \tag{4.2}$$

and for $z \in \mathbb{C}$ set

$$\mathcal{L}^z = \mathcal{L}^z_{\mathfrak{p}} : \mathcal{A}^n \to \mathbb{M}_{[0,n_0]}(\mathcal{A}), \qquad \mathcal{L}^z(\boldsymbol{x}) := \mathcal{L}(\boldsymbol{x}) + K^z \otimes 1_{\mathcal{A}} \tag{4.3}$$

where

$$K^z = K^z_{\mathfrak{p}} := \begin{pmatrix} a_0 - z & 0 \\ 0 & -\mathrm{I}_{n_0} \end{pmatrix} \in \mathbb{M}_{[0,n_0]}(\mathbb{C})\,. \tag{4.4}$$

We have that for any $z \in \mathbb{C}$,

$$(\mathfrak{p} - z)^{-1} = ((\mathcal{L}^z(\boldsymbol{x}))^{-1})_{1,1}. \tag{4.5}$$

For the proof we recall the Schur complement formula: for $M \in \mathbb{M}_{m+n}(\mathcal{A})$ written in block form as

$$M = \begin{pmatrix} M_{1,1} & M_{1,2} \\ M_{2,1} & M_{2,2} \end{pmatrix}, \quad M_{1,1} \in \mathbb{M}_m(\mathcal{A}),\ M_{2,2} \in \mathbb{M}_n(\mathcal{A})$$

with $M_{2,2}$ invertible, we have that the top-left $m \times m$ block of M^{-1} is given by

$$(M^{-1})_{1,1} = \big(M_{1,1} - M_{1,2}(M_{2,2})^{-1}M_{2,1}\big)^{-1}.$$

Proof. For $s_0 = (s_{0,1}, \dots, s_{0,n})$ we take $s_{0,\ell} = \overline{b_\ell}$. Write $\mathfrak{p}_{(2)}(\boldsymbol{x}) = \sum_{\ell,m=1}^n a_{\ell,m} x_\ell x_m$ for the homogeneous degree-2 part of $\mathfrak{p}(\boldsymbol{x})$. Let $A = U\Sigma V^*$ be the singular value decomposition for A, with $U = (u_{\ell,k}), V = (v_{m,k}) \in \mathbb{M}_{n,n_0}(\mathbb{C})$, and $\Sigma = \operatorname{diag}(\sigma_1, \dots, \sigma_{n_0})$ with $\sigma_1 \geq \cdots \geq \sigma_{n_0} > 0$. For $1 \leq k \leq n_0$ and $1 \leq \ell \leq n$, we set $r_{k,\ell} := \overline{u_{\ell,k}}$ and $s_{k,\ell} = \sigma_k v_{\ell,k}$, so that

$$\begin{aligned}
\mathfrak{p}_{(2)}(\boldsymbol{x}) &= \sum_{k=1}^{n_0} \sigma_k \sum_{\ell=1}^n u_{\ell,k} x_\ell \sum_{m=1}^n \overline{v_{m,k}} x_m \\
&= \sum_{k=1}^{n_0} \Big(\sum_{\ell=1}^n \overline{r_{k,\ell}} x_\ell \Big) \Big(\sum_{m=1}^n \overline{s_{k,m}} x_m \Big) \\
&= \sum_{k=1}^n \langle r_k, \boldsymbol{x} \rangle \langle s_k, \boldsymbol{x} \rangle.
\end{aligned}$$

With $\mathcal{L}^z(\boldsymbol{x})$ as in (4.3), by the Schur complement formula we have

$$((\mathcal{L}^z(\boldsymbol{x}))^{-1})_{1,1} = \Big(\langle s_0, \boldsymbol{x} \rangle + a_0 - z + \sum_{k=1}^{n_0} \langle r_k, \boldsymbol{x} \rangle \langle s_k, \boldsymbol{x} \rangle \Big)^{-1} = (\mathfrak{p}(\boldsymbol{x}) - z)^{-1}$$

as desired. □

Corollary 4.2. *Let $\boldsymbol{X} = (X_1, \dots, X_n)$, and P be as in Theorem 1.3, with $\mathfrak{p}$ as in* (4.1)*. With n_0 the rank of $A = (a_{\ell,m})$, there exist vectors $s_0, s_1, \dots, s_{n_0} \in \mathbb{C}^n$ depending only on $\mathfrak{p}$, with $s_1, \dots, s_{n_0}$ nonzero and mutually orthogonal, such that the following holds. Let*

$$\boldsymbol{L}^z := \boldsymbol{L} + K^z \otimes \mathrm{I}_N \tag{4.6}$$

with K^z as in (4.4) *and*

$$\boldsymbol{L} := \begin{pmatrix} Y_0 & X_1 & \cdots & X_{n_0} \\ Y_1 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ Y_{n_0} & 0 & \cdots & 0 \end{pmatrix}, \tag{4.7}$$

where

$$Y_k := \sum_{\ell=1}^n \overline{s_{k,\ell}} X_\ell, \qquad k \in [0, n_0]. \tag{4.8}$$

For all $\varepsilon > 0$ and $z \in \mathbb{C}$,

$$\mathbb{P}\{\sigma_{\min}(P - z) \leq \varepsilon\} \leq \mathbb{P}\{\sigma_{\min}(\boldsymbol{L}^z) \leq \varepsilon\}. \tag{4.9}$$

Proof. We apply Lemma 4.1 to get a linearized matrix $\mathcal{L}^z(\boldsymbol{X})$ as in (4.3), with

$$(P - z)^{-1} = ((\mathcal{L}^z(\boldsymbol{X}))^{-1})_{1,1}$$

where the right hand side is the $N \times N$ top-left block of $(\mathcal{L}^z(\boldsymbol{X}))^{-1}$. In particular,

$$\|(P - z)^{-1}\|_{\mathrm{op}} = \|((\mathcal{L}^z(\boldsymbol{X}))^{-1})_{1,1}\|_{\mathrm{op}} \leq \|(\mathcal{L}^z(\boldsymbol{X}))^{-1}\|_{\mathrm{op}}. \tag{4.10}$$

Now extend $r_1, \dots, r_{n_0}$ to an orthonormal basis $\mathbb{C}^n$ and let R be the unitary matrix with rows $r_1, \dots, r_n$. Replacing the N^2 independent complex standard Gaussian vectors $\boldsymbol{X}(i,j) = (X_1(i,j), \dots, X_n(i,j))$ with $R^*\boldsymbol{X}(i,j)$ and $s_0, \dots, s_{n_0}$ with $s_0R^*, \dots, s_{n_0}R^*$ we obtain the matrix (4.6). The claim now follows from (4.10) and the invariance of the complex standard Gaussian measure on $\mathbb{C}^n$ under unitary transformations. □

It will be convenient later in the proof to assume $n_0 \geq 2$, so we now dispense with the case that $n_0 = 1$.

Lemma 4.3. *Theorem 1.3 holds for the case that $A = (a_{\ell,m})$ as in (4.1) has rank one.*

Proof. With a unitary change of basis we can reduce to a matrix of the form

$$P = X_1(\alpha_1 X_1 + \alpha_2 X_2) + \sum_{k=1}^{3} \beta_k X_k + \gamma.$$

If $\beta_3 \neq 0$ then, conditioning on X_1, X_2, we have from Lemma 3.3 that

$$\mathbb{P}\{\sigma_{\min}(P) \leq \varepsilon\} = \mathbb{P}\{\sigma_{\min}(X_3 + M) \leq \varepsilon/|\beta_3|\} \lesssim (\varepsilon/|\beta_3|)^2 N^3 = O_{\mathfrak{p}}(N^3\varepsilon^2),$$

where M is a deterministic shift depending on X_1, X_2. Having obtained the claim in this case, we henceforth assume that $\beta_3 = 0$. Now write

$$P = (\alpha_2 X_1 + \beta_2)X_2 + (\alpha_1 X_1^2 + \beta_1 X_1 + \gamma).$$

If α_2 or β_2 is nonzero then, conditioning on X_1, we have

$$\mathbb{P}\{\sigma_{\min}(P) \leq \varepsilon\} \leq \mathbb{P}\{\sigma_{\min}(\alpha_2 X_1 + \beta_2) \leq \delta\} + \mathbb{P}\{\sigma_{\min}(X_2 + M') \leq \varepsilon/\delta\}$$

where $M' = (\alpha_2 X_1 + \beta_2)^{-1}(\alpha_1 X_1^2 + \beta_1 X_1 + \gamma)$ is a deterministic shift depending on X_1 (note that M' is well defined off a null event since $\alpha_2 X_1 + \beta_2$ is almost-surely invertible). Taking $\delta = \min\{\sqrt{\varepsilon}, |\beta_2|/2\}$, the claim again follows from Lemma 3.3.

Finally, assuming $\beta_3 = \beta_2 = \alpha_2 = 0$, we have

$$P = \alpha_1 X_1^2 + \beta_1 X_1 + \gamma.$$

If $\alpha_1 = 0$ we can conclude along the same lines as in previous cases. Otherwise we can factorize $P = \alpha_1(X_1 - a_+)(X_1 - a_-)$ for some $a_\pm \in \mathbb{C}$. Then since $\sigma_{\min}((X_1 - a_+)(X_1 - a_-)) \geq \sigma_{\min}(X_1 - a_+)\sigma_{\min}(X_1 - a_-)$ we have

$$\mathbb{P}\{\sigma_{\min}(P) \leq \varepsilon\} \leq \mathbb{P}\{\sigma_{\min}(X_1 - a_+) \leq (\varepsilon/|\alpha_1|)^{1/2}\} + \mathbb{P}\{\sigma_{\min}(X_1 - a_-) \leq (\varepsilon/|\alpha_1|)^{1/2}\} \lesssim_{\mathfrak{p}} \varepsilon N^3$$

and the lemma is proved. □

5. Reduction to a bounded-dimensional test projection

In view of Corollary 4.2, we henceforth write

$$\mathcal{L}^z(\boldsymbol{x}) = \mathcal{L}(\boldsymbol{x}) + K^z \otimes 1_{\mathcal{A}}, \quad \text{with} \quad \mathcal{L}(\boldsymbol{x}) := \begin{pmatrix} \langle s_0, \boldsymbol{x}\rangle & x_1 & \cdots & x_{n_0} \\ \langle s_1, \boldsymbol{x}\rangle & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \langle s_{n_0}, \boldsymbol{x}\rangle & 0 & \cdots & 0 \end{pmatrix} \tag{5.1}$$

(i.e. taking r_k to be the k-th standard basis vector in $\mathbb{C}^n$ for each $k \in [n_0]$), so that $\boldsymbol{L} = \mathcal{L}(\boldsymbol{X}), \boldsymbol{L}^z = \mathcal{L}^z(\boldsymbol{X})$ are as in Corollary 4.2. Our aim is to establish a lower tail bound for $\sigma_{\min}(\boldsymbol{L}^z)$.

Recall our notational conventions from Section 3.2. We view $\boldsymbol{L}, \boldsymbol{L}^z$ as elements of $\mathbb{M}_{[0,n_0]\times[N]}^{[0,n_0]\times[N]}(\mathbb{C}) \cong \mathbb{M}_{[0,n_0]}(\mathbb{M}_N(\mathbb{C}))$ (note that the first row and column have index 0). Thus, $\boldsymbol{L}$ has entries

$$\boldsymbol{L}((k,i),(l,j)) = L^{k,l}(i,j) = L_{i,j}(k,l) = \widetilde{\boldsymbol{L}}((i,k),(j,l)), \quad k,l \in [0,n_0],\ i,j \in [N]$$

with

$$L_{i,j} = \mathcal{L}(X_1(i,j), \ldots, X_n(i,j)) \in \mathbb{M}_{[0,n_0]}(\mathbb{C}),$$

and similarly $\boldsymbol{L}^z((k,i),(l,j)) = L^z_{i,j}(k,l)$, with

$$L^z_{i,j} = \mathcal{L}^z(X_1(i,j), \ldots, X_n(i,j)) = L_{i,j} + K^z\delta_{i,j}.$$

Note that as an element of $\mathbb{M}_N(\mathbb{M}_{[0,n_0]}(\mathbb{C}))$, $\widetilde{\boldsymbol{L}}^z = (L^z_{i,j})_{i,j\in[N]}$ has independent entries, and $\widetilde{\boldsymbol{L}} = (L_{i,j})_{i,j\in[N]}$ has entries that are i.i.d. and centered; however, we stress that the $(n_0+1)^2$ scalar entries of $L_{i,j}$ in general are *not* independent.

For $1 \le j \le N$ we write $\boldsymbol{L}^z_{(j)} \in \mathbb{M}_{N,N-1}(\mathbb{M}_{[0,n_0]}(\mathbb{C}))$ (resp. $\boldsymbol{L}_{(j)}$) for the submatrix of $\boldsymbol{L}^z$ (resp. $\boldsymbol{L}$) obtained by removing the jth column from each $N \times N$ submatrix. We let $\mathsf{V}^z_{(j)} \subset \mathbb{C}^{[0,n_0]\times[N]}$ denote the span of the columns of $\boldsymbol{L}^z_{(j)}$, viewed as an element of $\mathbb{M}^{[0,n_0]\times[N-1]}_{[0,n_0]\times[N]}(\mathbb{C})$, that is

$$\mathsf{V}^z_{(j)} := \mathrm{Span}\{\mathrm{col}_{(l,j')}(\boldsymbol{L}^z) : j' \in [N]\setminus\{j\},\, l \in [0,n_0]\}. \tag{5.2}$$

For each j, conditional on these $(N-1)(n_0+1)$ columns, we draw a matrix $\boldsymbol{U}_j \in \mathbb{M}^{[0,n_0]}_{[0,n_0]\times[N]}(\mathbb{C})$ with columns that are orthonormal in $(\mathsf{V}^z_{(j)})^\perp$, with $\boldsymbol{U}_j$ independent of $\mathrm{col}_j(\widetilde{\boldsymbol{L}}^z) = (L^z_{i,j})_{i=1}^N$. (To be more precise, we can for instance draw N i.i.d. Haar unitaries $H_j \in \mathcal{U}(n_0+1)$, independent of $(X_1,\dots,X_n)$, and fixing arbitrary matrices $\boldsymbol{V}_j$ with n_0+1 orthonormal columns in $\mathsf{V}^z_{(j)}$, chosen measurably with respect to the sigma algebra generated by $\{L_{i,j'}\}_{i\in[N],j'\in[N]\setminus\{j\}}$, we set $\boldsymbol{U}_j = \boldsymbol{V}_j H_j$.) Recalling our notational conventions from Section 3.2, we have that $\boldsymbol{U}_j$ is naturally associated to $\widetilde{\boldsymbol{U}}_j \in \mathbb{M}_{[0,n_0]}(\mathbb{C})^N$, an $N\times 1$ matrix of $(n_0+1)\times(n_0+1)$ blocks, which we denote by $U_{1,j},\dots,U_{N,j}$.

In the remainder of this section we establish the following:

Lemma 5.1 (Reduction to invertibility of a small test projection)**.** *With notation as above, for $j \in [N]$ denote*

$$\widehat{L}^z_j := \widetilde{\boldsymbol{U}}_j^* \,\mathrm{col}_j(\widetilde{\boldsymbol{L}}^z) = \sum_{i=1}^N U^*_{i,j} L^z_{i,j} \in \mathbb{M}_{[0,n_0]}(\mathbb{C}). \tag{5.3}$$

Then for any $\varepsilon > 0$,

$$\mathbb{P}\{\sigma_{\min}(\boldsymbol{L}^z) \le \varepsilon\} \le N\,\mathbb{P}\big\{\,|\det(\widehat{L}^z_1)| \le \varepsilon(|z|+C_{\mathfrak{p}})^{n_0}\sqrt{N}\,\big\} + e^{-N}$$

for some $C_{\mathfrak{p}} < \infty$ depending only on $\mathfrak{p}$.

Proof. On the event that $\sigma_{\min}(\boldsymbol{L}^z) \le \varepsilon$ we have that there exists a unit vector $\boldsymbol{v} = (v_1,\dots,v_N) \in \mathbb{C}^{[N]\times[0,n_0]}$ with each $v_j \in \mathbb{C}^{[0,n_0]}$ such that

$$\varepsilon \ge \|\boldsymbol{L}^z\boldsymbol{v}\|_2 = \Big\| \sum_{j=1}^N \mathrm{col}_j(\widetilde{\boldsymbol{L}}^z) v_j \Big\|_2.$$

Since $\boldsymbol{v}$ is a unit vector we must have $\|v_{j_0}\|_2 \ge 1/\sqrt{N}$ for some $j_0 \in [N]$. Since the norm of $\boldsymbol{L}^z\boldsymbol{v}$ can only decrease under projection to the column span of $\boldsymbol{U}_{j_0}$, we have

$$\varepsilon \ge \big\|\widetilde{\boldsymbol{U}}^*_{j_0}\,\mathrm{col}_j(\widetilde{\boldsymbol{L}}^z)v_j\big\|_2 = \|\widehat{L}^z_{j_0} v_{j_0}\|_2 \ge \frac{1}{\sqrt{N}}\sigma_{\min}(\widehat{L}^z_{j_0}).$$

Now note that

$$\|\widehat{L}^z_{j_0}\|_{\mathrm{op}} \le \|\boldsymbol{L}^z\|_{\mathrm{op}} \le \|\boldsymbol{L}\|_{\mathrm{op}} + |z| + O_{\mathfrak{p}}(1).$$

From Lemma 3.2 and the triangle inequality,

$$\mathbb{P}\{\|\boldsymbol{L}\|_{\mathrm{op}} > B\} \le e^{-N}$$

for some $B = O_{\mathfrak{p}}(1)$. Thus, by the union bound we have for some $B' = O_{\mathfrak{p}}(1)$,

$$\mathbb{P}\{\sigma_{\min}(\boldsymbol{L}^z) \le \varepsilon\} \le e^{-N} + \sum_{j=1}^{N} \mathbb{P}\{\sigma_{\min}(\widehat{L}_j^z) \le \varepsilon\sqrt{N},\ \|\widehat{L}_j^z\|_{\mathrm{op}} \le B' + |z|\}$$
$$\le e^{-N} + \sum_{j=1}^{N} \mathbb{P}\{|\det(\widehat{L}_j^z)| \le (B' + |z|)^{n_0}\varepsilon\sqrt{N}\}.$$

Since the distribution of $\widetilde{\boldsymbol{L}}^z = (L_{i,j}^z)_{i,j\in[N]}$ is invariant under simultaneous permutation of the N row and N column indices, the claim follows. □

6. Anti-concentration for matrix random walks

From Lemma 5.1 we see that it suffices to control the lower tail for the determinant of the (n_0+1)-dimensional matrix $\widehat{L}_1^z$ defined in (5.3). The advantage of this perspective is that for a fixed realization of $\boldsymbol{U}_1$ (which we recall is independent of $\mathrm{col}_1(\widetilde{\boldsymbol{L}}^z)$), the summands $U_{i,1}^* L_{i,1}^z$ are N independent random matrices. We can thus view $\widehat{L}_1^z$ as a random walk in $\mathbb{M}_{[0,n_0]}(\mathbb{C})$.

In this section we consider an arbitrary *fixed* (deterministic) matrix

$$\boldsymbol{U} \in \mathbb{M}_{[0,n_0]\times[N]}^{[0,n_0]}(\mathbb{C}) \cong \mathbb{M}_N(\mathbb{C})^{[0,n_0]}$$

with orthonormal columns in $\mathbb{C}^{[0,n_0]\times[N]}$, identified with the sequence of its block submatrices as $\boldsymbol{U} = (U^0, \dots, U^{n_0}) \in \mathbb{M}_N(\mathbb{C})^{[0,n_0]}$ and with the sequence of its square submatrices denoted $\widetilde{\boldsymbol{U}} = (U_1, \dots, U_N) \in \mathbb{M}_{[0,n_0]}(\mathbb{C})^N$. Thus, denoting the rows of U_i by $(v_i^k)^*$, $k \in [0, n_0]$, we have that $(v_i^k)^*$ is the ith row of U^k.

Let $(g_i^k)_{i\in[N],k\in[n]}, (h_i^k)_{i\in[N],k\in[n]} \in \mathbb{C}^{[N]\times[n]}$ be independent arrays of Nn i.i.d. standard real Gaussians and denote $g_i = (g_i^1, \dots, g_i^n)$, $g^k = (g_i^k)_{i\in[N]}$, and similarly for h_i, h^k. We set $\xi_i^k = (2N)^{-1/2}(g_i^k + \mathrm{i}h_i^k)$, $\xi_i = (\xi_i^1, \dots, \xi_i^n)$, $\xi^k = (2N)^{-1/2}(g^k + \mathrm{i}h^k)$. Then $\widehat{L}_1^z$ from (5.3) is identically distributed to the $(n_0+1)\times(n_0+1)$ random matrix

$$W = M + \sum_{i=1}^{N} U_i^* L_i \tag{6.1}$$

where

$$M = M(\boldsymbol{U}, z, a_0) := U_1^* K^z \in \mathbb{M}_{[0,n_0]}(\mathbb{C}) \tag{6.2}$$

is deterministic and the square matrices

$$L_i := \mathcal{L}(\xi_i^1, \dots, \xi_i^n) = \begin{pmatrix} \langle s_0, \xi_i\rangle & \xi_i^1 & \cdots & \xi_i^{n_0} \\ \langle s_1, \xi_i\rangle & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \langle s_{n_0}, \xi_i\rangle & 0 & \cdots & 0 \end{pmatrix} \in \mathbb{M}_{[0,n_0]}(\mathbb{C}) \tag{6.3}$$

are i.i.d., with $a_0, s_0, \dots, s_{n_0}$ as in Corollary 4.2.

From Lemma 5.1 we see that in order to prove Theorem 1.3, it suffices to prove an anti-concentration estimate for $\det(W)$, a degree-(n_0+1) polynomial in the Nn Gaussian variables g_i^k, of the form

$$\mathbb{P}\{|\det(W)| \le \varepsilon\} \le N^C \varepsilon^c. \tag{6.4}$$

In this section we identify sufficient structural conditions on the matrix $\boldsymbol{U}$ in order to have (6.4). In Section 7 we will show that such conditions hold with high probability for the matrix $\boldsymbol{U}_1$.

For $1 \le \ell \le n$ let

$$Q^\ell = Q^\ell(\boldsymbol{U}) = \sum_{k=0}^{n_0} s_{k,\ell} U^k \in \mathbb{M}_{[N]}^{[0,n_0]}(\mathbb{C}) \tag{6.5}$$

and for $i \in [N]$ denote the ith row of Q^ℓ by $(w_i^\ell)^*$; thus,

$$w_i^\ell = \sum_{k=0}^{n_0} \overline{s_{k,\ell}} v_i^k. \tag{6.6}$$

For $\ell \in [n]$ and $i_0, i_1, \dots, i_{n_0} \in [N]$ denote

$$\Delta^\ell_{i_0,i_1,\dots,i_{n_0}} = \Delta^\ell_{i_0,i_1,\dots,i_{n_0}}(\boldsymbol{U}) := \det(w^\ell_{i_0}, v^0_{i_1}, \dots, v^0_{i_{n_0}}). \tag{6.7}$$

For a parameter $\delta > 0$ we define sets of *structured matrix-columns*:

$$\mathrm{Struct}_1(\delta) := \bigcap_{i_0,i_1,\dots,i_{n_0} \in [N]} \bigcap_{\ell \in [n_0+1,n]} \left\{ \boldsymbol{U} \in \mathbb{M}^{[0,n_0]}_{[0,n_0]\times[N]}(\mathbb{C}) : |\Delta^\ell_{i_0,i_1,\dots,i_{n_0}}| < \delta \right\}, \tag{6.8}$$

$$\mathrm{Struct}_2(\delta) := \bigcap_{i_1,\dots,i_{n_0} \in [N]} \bigcap_{\ell \in [n_0]} \left\{ \boldsymbol{U} \in \mathbb{M}^{[0,n_0]}_{[0,n_0]\times[N]}(\mathbb{C}) : |\Delta^\ell_{i_\ell,i_1,\dots,i_{n_0}}| < \delta \right\}, \tag{6.9}$$

$$\mathrm{Struct}(\delta) := \mathrm{Struct}_1(\delta) \cap \mathrm{Struct}_2(\delta). \tag{6.10}$$

Lemma 6.1 (Anti-concentration for the determinant of a matrix random walk)**.** *Let* $\boldsymbol{U} = (U_1, \dots, U_N) \in \mathbb{M}^{[0,n_0]}_{[0,n_0]\times[N]}(\mathbb{C})$ *and suppose* $\boldsymbol{U} \notin \mathrm{Struct}(\delta)$ *for some* $\delta > 0$*. Then for any* $\varepsilon > 0$*,*

$$\sup_{M \in \mathbb{M}_{[0,n_0]}(\mathbb{C})} \mathbb{P}\left\{ \left| \det\left(M + \sum_{i=1}^N U_i^* L_i \right) \right| \le \varepsilon \right\} \lesssim_{n_0} N^{1/2} (\varepsilon/\delta)^{1/(n_0+1)}. \tag{6.11}$$

To establish the lemma we make use of the following easy corollary of the Carbery–Wright inequality for anti-concentration of polynomials in Gaussian variables.

Lemma 6.2. *Let* $g_1, \cdots, g_n$ *be i.i.d. standard gaussian variables and let* p *be a real-valued degree-*d *polynomial. Let* $a \neq 0$ *be the coefficient of one of the degree-*d *monomials of* p*. Then for any* $\varepsilon > 0$*,*

$$\sup_{t \in \mathbb{R}} \mathbb{P}\left(|p(g_1, \dots, g_n) - t| \le \varepsilon \right) \lesssim_d (\varepsilon/|a|)^{1/d}.$$

Proof. By rescaling p and ε by a we may assume $a = 1$. From the Carbery–Wright inequality [18] we have

$$\sup_{t \in \mathbb{R}} \mathbb{P}\left(|p(g_1, \dots, g_n) - t| \le \varepsilon\sigma \right) \lesssim_d \varepsilon^{1/d}$$

where $\sigma^2 = \mathbb{V}\mathrm{ar}(p(g_1, \dots, g_n))$ is the variance of p under the product Gaussian measure. It hence suffices to verify that

$$\mathbb{V}\mathrm{ar}(p(g_1, \dots, g_n)) \gtrsim_d 1.$$

One readily notes the above indeed holds by, for instance, expanding p in the orthonormal basis of Hermite polynomials and noting our assumption implies that one of the coefficients among the highest degree terms in the expansion must be of size $\gtrsim_d 1$. The claim follows. □

Proof of Lemma 6.1. Write

$$D = \det \widehat{L}, \qquad \widehat{L} := M + \sum_{i=1}^N U_i^* L_i. \tag{6.12}$$

Denoting the columns of M by $y^0, \dots, y^{n_0}$, we have

$$D = \det\left(y^0 + \sum_{i_0=1}^{N} \sum_{k=0}^{n_0} \langle s_k, \xi_{i_0} \rangle v_{i_0}^k \, , \, y^1 + \sum_{i_1=1}^{N} \xi_{i_1}^1 v_{i_1}^0, \dots, y^{n_0} + \sum_{i_{n_0}=1}^{N} \xi_{i_{n_0}}^{n_0} v_{i_{n_0}}^0 \right)$$
$$= \det\left(y^0 + \sum_{i_0=1}^{N} \sum_{\ell=1}^{n} \xi_{i_0}^{\ell} w_{i_0}^{\ell} \, , \, y^1 + \sum_{i_1=1}^{N} \xi_{i_1}^1 v_{i_1}^0, \dots, y^{n_0} + \sum_{i_{n_0}=1}^{N} \xi_{i_{n_0}}^{n_0} v_{i_{n_0}}^0 \right). \tag{6.13}$$

Now we condition on the variables $\boldsymbol{h} = (h_i^k)_{i \in [N], k \in [n]}$, the imaginary parts of $\sqrt{2N}\boldsymbol{\xi}$, and view D as a degree-(n_0+1) polynomial in the nN i.i.d. standard real Gaussian variables $\boldsymbol{g} = (g_i^k)$, the real parts of $\sqrt{2N}\boldsymbol{\xi}$. Write $\widehat{D}$ for the homogeneous degree-(n_0+1) part of this polynomial. We have

$$\widehat{D} = (2N)^{-(n_0+1)/2} \sum_{\ell=1}^{n} \sum_{i_0, \dots, i_{n_0}=1}^{N} g_{i_0}^{\ell} g_{i_1}^1 \cdots g_{i_{n_0}}^{n_0} \cdot \det(w_{i_0}^{\ell}, v_{i_1}^0, \dots, v_{i_{n_0}}^0) \tag{6.14}$$

$$= (2N)^{-(n_0+1)/2} \sum_{\ell=1}^{n} \sum_{i_0, \dots, i_{n_0}=1}^{N} g_{i_0}^{\ell} g_{i_1}^1 \cdots g_{i_{n_0}}^{n_0} \cdot \Delta_{i_0, \dots, i_{n_0}}^{\ell}. \tag{6.15}$$

In particular, for $\ell \le n_0$, the coefficient of the monomial $g_{i_\ell}^{\ell} g_{i_1}^1 \cdots g_{i_{n_0}}^{n_0}$ is

$$2 \cdot (2N)^{-(n_0+1)/2} \Delta_{i_\ell, i_1, \dots, i_{n_0}}^{\ell} \tag{6.16}$$

and for $\ell > n_0$, the coefficient of the monomial $g_{i_0}^{\ell} g_{i_1}^1 \cdots g_{i_{n_0}}^{n_0}$ is

$$(2N)^{-(n_0+1)/2} \Delta_{i_0, i_1, \dots, i_{n_0}}^{\ell}. \tag{6.17}$$

Since $\boldsymbol{U} \notin \mathrm{Struct}(\delta)$, at least one of these coefficients has modulus at least $(2N)^{-(n_0+1)/2}\delta$. Suppose this holds for a coefficient (6.16) for some $\ell \le n_0$ and $i_1, \dots, i_{n_0} \in [N]$. Then either the real or imaginary part is of size at least $2^{-1/2}(2N)^{-(n_0+1)/2}\delta$. Supposing further that this holds for the real part, then for the real Gaussian polynomial $\Re D$ there is a coefficient of a maximal degree monomial of size at least $2^{-1/2}(2N)^{-(n_0+1)/2}\delta$. Applying Lemma 6.2 to $\Re D$, we have

$$\mathbb{P}\{|D| \le \varepsilon\} \le \mathbb{P}\{|\Re D| \le \varepsilon\} \lesssim_{n_0} N^{1/2} (\varepsilon/\delta)^{1/(n_0+1)}$$

yielding (6.11) in this case. For the case that the imaginary part of this coefficient is of size at least $2^{-1/2}(2N)^{-(n_0+1)/2}\delta$ we argue similarly with the real Gaussian polynomial $\Im D$, and we repeat the same reasoning for the case that a coefficient as in (6.17) is large. □

7. Ruling out structured bases

Recall from (5.2) that $\mathsf{V}_{(1)}^z \subset \mathbb{C}^{[0,n_0]\times[N]}$ denotes the span of the $(n_0+1)(N-1)$ columns of $\boldsymbol{L}_{(1)}^z$, with the latter viewed as an element of $\mathbb{M}_{[0,n_0]\times[N]}^{[0,n_0]\times[N-1]}(\mathbb{C})$, and that the columns of $\boldsymbol{U}_1$ are a random orthonormal set in $(\mathsf{V}_{(1)}^z)^{\perp}$. (In view of Theorem 1.3 we have, *a posteriori*, that with high probability $\boldsymbol{L}_{(1)}^z$ is full rank and the columns of $\boldsymbol{U}_1$ in fact comprise a basis for $(\mathsf{V}_{(1)}^z)^{\perp}$.) The aim of this section is to establish the following:

Proposition 7.1 (Structured bases are rare)**.** *Assume $n_0 \ge 2$. Then, we have*

$$\mathbb{P}\{\boldsymbol{U}_1 \in \mathrm{Struct}(N^{-\frac{1}{2}n_0 - 10})\} \lesssim_{\mathfrak{p},z} e^{-N} \tag{7.1}$$

where the sets $\mathrm{Struct}(\delta)$ *were defined in* (6.10)*.*

We now conclude the proof of Theorem 1.3 assuming the above proposition. From Lemma 4.3 we may assume $n_0 \geq 2$. From Lemma 3.2, the triangle inequality and sub-multiplicativity of the operator norm we have $\|P\|_{\mathrm{op}} = O_{\mathfrak{p}}(1)$. In particular we may assume $|z| = O_{\mathfrak{p}}(1)$ since otherwise we obtain the claim by simply lower bounding $\sigma_{\min}(P - z) \geq |z| - \|P\|_{\mathrm{op}}$. Now by Corollary 4.2,

$$\mathbb{P}\{\sigma_{\min}(P - z) \leq \varepsilon\} \leq \mathbb{P}\{\sigma_{\min}(\boldsymbol{L}^z) \leq \varepsilon\}$$

and from Lemma 5.1 the latter is in turn bounded by

$$N\mathbb{P}\big\{ \, |\det(\widehat{L}_1^z)| \leq \varepsilon(|z| + C_{\mathfrak{p}})^{n_0}\sqrt{N} \, \big\} + e^{-N}.$$

Finally, from Proposition 7.1, (6.1) and Lemma 6.1,

$$\begin{aligned}\mathbb{P}\big\{ \, |\det(\widehat{L}_1^z)| \leq \varepsilon(|z| + C_{\mathfrak{p}})^{n_0}\sqrt{N} \, \big\} &\lesssim_{\mathfrak{p}} e^{-N} + N^{\frac{1}{2}(1+(n_0+21)/(n_0+1))}(|z| + C_{\mathfrak{p}})^{n_0/(n_0+1)}\varepsilon^{1/(n_0+1)} \\ &\lesssim_{\mathfrak{p}} e^{-N} + N^{13/3}\varepsilon^{1/(n_0+1)},\end{aligned}$$

where in the second line we bounded the exponent of N by its maximum value at $n_0 = 2$. The claim follows by combining all of these estimates. □

7.1. **High-level proof of Proposition 7.1.** Our approach is to cut the set $\mathrm{Struct}(\delta)$ (for sufficiently small $\delta = N^{-O_{\mathfrak{p}}(1)}$) into several pieces, and to bound the event that $\boldsymbol{U}_1$ lies in each piece by taking union bounds over nets. (Recall the definition of a net from Section 3.4.) Once we have approximated $\boldsymbol{U}_1$ by a some fixed element $\boldsymbol{U}$ of a net, our task is then to bound the probability that the columns of $\boldsymbol{U}$ are nearly contained in $(\mathsf{V}_{(1)}^z)^{\perp}$, i.e. that $\widetilde{\boldsymbol{U}}^* \operatorname{col}_j(\widetilde{\boldsymbol{L}}^z) \approx 0$ (the zero matrix in $\mathbb{M}_{[0,n_0]}(\mathbb{C})$) for each $2 \leq j \leq N$. We have

$$\begin{aligned}\widetilde{\boldsymbol{U}}^* \operatorname{col}_j(\widetilde{\boldsymbol{L}}^z) = \sum_{i=1}^N U_i^* L_{i,j}^z &= U_j^* K^z + \sum_{i=1}^N U_i^* L_{i,j} \\ &=: M_j(\boldsymbol{U}, z) + \mathrm{Walk}_j(\boldsymbol{U}) \in \mathbb{M}_{[0,n_0]}(\mathbb{C}) \end{aligned} \tag{7.2}$$

where the matrices M_j are deterministic shifts and $\mathrm{Walk}_j(\boldsymbol{U})$ are i.i.d. copies of the centered random walk

$$\mathrm{Walk}(\boldsymbol{U}) = \sum_{i=1}^N U_i^* L_i \tag{7.3}$$

with

$$L_i = \mathcal{L}(\xi_i^1, \dots, \xi_i^n) = \frac{1}{\sqrt{N}}\mathcal{L}(\zeta_i^1, \dots, \zeta_i^n)$$

as in (6.3), where $\zeta_i^k = \frac{1}{\sqrt{2}}(g_i^k + \mathrm{i}h_i^k)$ are i.i.d. standard complex Gaussians. Our task is thus reduced to proving an anti-concentration bound for $\mathrm{Walk}(\boldsymbol{U})$. Note that whereas in Section 6 we were concerned with anti-concentration for the determinant of $W = M_1 + \mathrm{Walk}(\boldsymbol{U})$, here we need to show the matrix $\mathrm{Walk}(\boldsymbol{U})$ is anti-concentrated as a random element of the vector space $\mathbb{M}_{[0,n_0]}(\mathbb{C})$.

Anti-concentration for $\mathrm{Walk}(\boldsymbol{U})$ is most transparent when viewing the matrix as a Gaussian vector in $\mathbb{C}^{(n_0+1)^2} \cong \mathbb{C}^{[0,n_0]^2}$. Indeed, we note from (6.3) that L_i is a Gaussian linear combination $\frac{1}{\sqrt{N}}\sum_{\ell=1}^n \zeta_i^\ell M_\ell$ of matrices M_ℓ, where M_ℓ^* has zeroth row $(s_0(\ell), \dots, s_{n_0}(\ell))$, ℓth row e_ℓ (the ℓth standard basis vector) and all other rows equal to zero. Thus

$$\mathrm{Walk}(\boldsymbol{U}) = \frac{1}{\sqrt{N}} \sum_{(\ell,i)\in[n]\times[N]} \zeta_i^\ell U_i^* M_\ell$$

Recalling (6.6), the matrix $U_i^* M_\ell$ has zeroth column w_i^ℓ, ℓth column v_i^ℓ, and all other columns equal to zero. We can hence identify the matrix (7.3) with the Gaussian vector

$$\frac{1}{\sqrt{N}}\boldsymbol{W}^*\boldsymbol{\zeta} \in \mathbb{C}^{[0,n_0]^2} \tag{7.4}$$

where $\boldsymbol{\zeta} = (\zeta_i^k)_{(k,i)\in[n]\times[N]}$, and the *walk matrix* $\boldsymbol{W} = \boldsymbol{W}(\boldsymbol{U})$ is defined

$$\boldsymbol{W}(\boldsymbol{U}) := \begin{pmatrix} Q^1 & U^0 & 0 & \cdots & 0 \\ Q^2 & 0 & U^0 & \cdots & 0 \\ \vdots & \vdots & & \ddots & \vdots \\ Q^{n_0} & 0 & \cdots & 0 & U^0 \\ Q^{n_0+1} & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ Q^n & 0 & \cdots & 0 & 0 \end{pmatrix} =: (\, \boldsymbol{Q}(\boldsymbol{U}) \; \boldsymbol{R}(U^0) \,) \in \mathbb{M}_{[n]\times[N]}^{[0,n_0]^2}(\mathbb{C}), \tag{7.5}$$

recalling the matrices Q^ℓ from (6.5). (See (2.11) in the proof outline for the form of the walk matrix in the case of the anti-commutator polynomial $\mathfrak{p} = x_1x_2 + x_2x_1$.) From (7.4) we see that the probability the Gaussian vector Walk($\boldsymbol{U}$) lands in a Hilbert–Schmidt ball of radius ε scales like $O(\varepsilon^2)^{\operatorname{rank}(\boldsymbol{W})}$ for small ε (cf. Lemma 7.7). In particular, from the form of (7.5) we see that a lower bound $\sigma_{\min}(U^0) \gtrsim 1$ already guarantees a small ball probability of order $O(\varepsilon^2)^{n_0(n_0+1)}$. The next lemma shows we may assume such a bound holds. For a parameter $\alpha > 0$, we define the set of "good" matrices $\boldsymbol{U}$ having a well-conditioned zeroth block:

$$\mathcal{A}(\alpha) := \{\, \boldsymbol{U} \in \mathbb{M}_{[0,n_0]\times[N]}^{[0,n_0]}(\mathbb{C}) : \; \sigma_{\min}(U^0) \geq \alpha \,\}. \tag{7.6}$$

Recall that $\boldsymbol{L}_{(1)}$ is obtained by removing the first column from each of the $n_0 + 1$ blocks of N columns of $\boldsymbol{L}$, and that $\boldsymbol{U}_1$ depends on $\boldsymbol{L}$ only through $\boldsymbol{L}_{(1)}$.

Lemma 7.2. *There is constant $c_{\mathfrak{p}} > 0$ depending only on $\mathfrak{p}$ such that for any $B \geq 1$ and $z \in \mathbb{C}$,*

$$\mathbb{P}\big\{\, \boldsymbol{U}_1 \notin \mathcal{A}(\alpha) \wedge \|\boldsymbol{L}_{(1)}\|_{\mathrm{op}} \leq B \,\big\} \lesssim_{\mathfrak{p}} e^{-N} \tag{7.7}$$

for any $\alpha < c_{\mathfrak{p}}/(B^2 + |z|)$.

We prove this lemma in Section 7.2.

By a standard tensorization argument (cf. Lemma 7.8), we obtain from the anti-concentration bound for Walk($\boldsymbol{U}$) the following anti-concentration estimate for $\boldsymbol{U}^*\boldsymbol{L}_{(1)}^z$. In other words, we bound the probability that the columns of a fixed matrix $\boldsymbol{U}$ are almost contained in $\mathsf{V}_{(1)}^z$. The bounds assume U^0 is well conditioned, and depend on how many columns of $\boldsymbol{Q}$ are in general position with respect to each other and to the columns of $\boldsymbol{R}$. We use the following notation: For given $\boldsymbol{U}$, denoting the columns of $\boldsymbol{Q}$ by $\boldsymbol{q}_k$, $k \in [0, n_0]$, we write $\boldsymbol{W}_I$ for the submatrix of $\boldsymbol{W}$ formed by the columns of $\boldsymbol{R}$ together with the columns $\{\boldsymbol{q}_k\}_{k\in I}$ of $\boldsymbol{Q}$ (in the natural order, say, though we note the order of columns will not be important). We sometimes write $\boldsymbol{W}_\emptyset := \boldsymbol{R}$. The column span of $\boldsymbol{W}_I$ is denoted W_I, that is:

$$\mathsf{W}_I := \operatorname{Span}(\mathsf{W}_\emptyset, \boldsymbol{q}_{k_1}, \ldots, \boldsymbol{q}_{k_r}) \qquad \text{for } I = \{k_1, \ldots, k_r\} \subset [0, n_0]. \tag{7.8}$$

Lemma 7.3. *Let $z \in \mathbb{C}$ and $\alpha > 0$, and fix some arbitrary $\boldsymbol{U} \in \mathcal{A}(\alpha)$. Then for any $\varepsilon > 0$ and $1 \leq j \leq N$,*

$$\mathbb{P}\{\|\boldsymbol{U}^*\boldsymbol{L}_{(1)}^z\|_{\mathrm{HS}} \leq \varepsilon\} = O_{\mathfrak{p},\alpha}(\varepsilon^2)^{n_0(n_0+1)(N-1)}. \tag{7.9}$$

Furthermore, if additionally it holds that for some $\beta > 0$, $1 \leq r \leq n_0 + 1$ and $k_1, \ldots, k_r \in [0, n_0]$ distinct,

$$\operatorname{dist}(\boldsymbol{q}_{k_i}, \mathsf{W}_{\{k_1,\ldots,k_{i-1}\}}) \geq \beta \qquad \forall\, 1 \leq i \leq r, \tag{7.10}$$

(interpreting the left hand side as $\operatorname{dist}(\boldsymbol{q}_{k_1}, \mathsf{W}_\emptyset)$ when $i = 1$), then for any $\varepsilon > 0$ and $1 \leq j \leq N$,

$$\mathbb{P}\{\|\boldsymbol{U}^*\boldsymbol{L}_{(1)}^z\|_{\mathrm{HS}} \leq \varepsilon\} = O_{\mathfrak{p},\alpha,\beta}(\varepsilon^2)^{[n_0(n_0+1)+r](N-1)}. \tag{7.11}$$

We defer the proof to Section 7.3.

We get the best anti-concentration estimate in (7.11) when $r = n_0 + 1$, i.e. when the full walk matrix $\boldsymbol{W}(\boldsymbol{U})$ is well conditioned. The following lemma shows we may assume this is the case. For $\beta > 0$ we define the set

$$\mathcal{B}(\beta) := \bigcup_{k\in[0,n_0]} \{\, \boldsymbol{U} \in \mathbb{M}^{[0,n_0]}_{[0,n_0]\times[N]}(\mathbb{C}) : \operatorname{dist}(\boldsymbol{q}_k, \mathsf{W}_{[0,n_0]\setminus\{k\}}) < \beta \,\} \tag{7.12}$$

Lemma 7.4 (Walks usually have full rank)**.** *For any $\alpha > 0$ and $B \geq 1$ there exists $\beta_\star(\mathfrak{p}, \alpha, B) > 0$ such that for any $\beta \in (0, \beta_\star]$ and $z \in \mathbb{C}$,*

$$\mathbb{P}\{\, \boldsymbol{U}_1 \in \mathcal{B}(\beta) \cap \mathcal{A}(\alpha) \ \wedge\ \|\boldsymbol{L}^z\|_{\mathrm{op}} \leq B \,\} \lesssim_{\mathfrak{p}} e^{-N}.$$

We prove this lemma in Section 7.4. While the lemma allows us to apply (7.11) with $r = n_0 + 1$, we will need to apply this estimate with smaller values of r in the proof.

In view of Lemmas 7.2 and 7.4 it only remains to bound the probability that $\boldsymbol{U}_1$ lies in the set

$$\mathcal{E} = \mathcal{E}(\delta, \alpha, \beta) := \mathrm{Struct}(\delta) \cap \mathcal{A}(\alpha) \cap \mathcal{B}(\beta)^c. \tag{7.13}$$

We will bound the probability that $\boldsymbol{U}_1$ lies in $\mathcal{E}$ by combining the the anti-concentration estimate (7.11) (taking $r = n_0 + 1$) with a union bound over a suitable net, provided by the following:

Lemma 7.5. *For any $\alpha > 0$, $\rho_\star \in (0, 1)$ and $0 < \delta \leq \rho_\star^2 N^{-\frac{n_0}{2}-2}$, $\mathcal{E}$ has a (Hilbert–Schmidt) $\rho_\star$-net $\mathcal{N} \subset \mathcal{E}$ of size*

$$|\mathcal{N}| = O_{\mathfrak{p},\alpha}(1)^N N^{(2n_0+3)N} O(1/\rho_\star^2)^{({n_0}^2+\frac{3}{2}n_0+\frac{3}{2})N}. \tag{7.14}$$

Remark 7.6. The proof in fact yields a net for any subset of $\mathrm{Struct}(\delta) \cap \mathcal{A}(\alpha)$, which is why the bound (7.14) is independent of β.

The proof of Lemma 7.5 is deferred to Section 7.5. We now conclude the proof of Proposition 7.1 assuming the above lemmas. Let $\delta > 0$ to be chosen sufficiently small in the course of the proof. From Lemma 3.2 and the triangle inequality we have that $\|\boldsymbol{L}\|_{\mathrm{op}} = O_{\mathfrak{p}}(1)$ with probability $1 - O_{\mathfrak{p}}(e^{-N})$, and hence $\|\boldsymbol{L}_{(1)}\|_{\mathrm{op}}, \|\boldsymbol{L}^z\|_{\mathrm{op}}, \|\boldsymbol{L}^z_{(1)}\|_{\mathrm{op}} = O_{\mathfrak{p},z}(1)$ with probability $1 - O_{\mathfrak{p}}(e^{-N})$. Applying the union bound, for $\alpha, \beta > 0$ and some $B = O_{\mathfrak{p},z}(1)$ we have

$$\begin{aligned}\mathbb{P}\{\boldsymbol{U}_1 \in \mathrm{Struct}(\delta)\} \leq\ & \mathbb{P}\{\boldsymbol{U}_1 \in \mathcal{E}(\delta, \alpha, \beta) \ \wedge\ \|\boldsymbol{L}^z_{(1)}\|_{\mathrm{op}} \leq B\} \\ &+ \mathbb{P}\{\boldsymbol{U}_1 \notin \mathcal{A}(\alpha) \ \wedge\ \|\boldsymbol{L}_{(1)}\|_{\mathrm{op}} \leq B\} \\ &+ \mathbb{P}\{\boldsymbol{U}_1 \in \mathcal{A}(\alpha) \cap \mathcal{B}(\beta) \ \wedge\ \|\boldsymbol{L}^z\|_{\mathrm{op}} \leq B\} + O_{\mathfrak{p}}(e^{-N}).\end{aligned}$$

For this choice of B we now choose $\alpha \gtrsim_{\mathfrak{p},z} 1$ satisfying the constraint in Lemma 7.2, followed by $\beta \gtrsim_{\mathfrak{p},z} 1$ satisfying the constraint in Lemma 7.4. From those lemmas we then have

$$\mathbb{P}\{\boldsymbol{U}_1 \in \mathrm{Struct}(\delta)\} \leq \mathbb{P}\{\boldsymbol{U}_1 \in \mathcal{E}(\delta, \alpha, \beta) \ \wedge\ \|\boldsymbol{L}^z\|_{\mathrm{op}} \leq B\} + O_{\mathfrak{p}}(e^{-N}). \tag{7.15}$$

Let $\rho_\star > 0$ to be chosen later and let $\mathcal{N}$ be as in Lemma 7.5 (assuming δ is sufficiently small depending on the choice of $\rho_\star$). By approximating any realization of $\boldsymbol{U}_1 \in \mathcal{E}$ by some $\boldsymbol{U} \in \mathcal{N}$ with $\|\boldsymbol{U}_1 - \boldsymbol{U}\|_{\mathrm{HS}} \leq \rho_\star$ and using that

$$\|(\boldsymbol{U}_1 - \boldsymbol{U})^* \boldsymbol{L}^z_{(1)}\|_{\mathrm{HS}} \leq \|\boldsymbol{L}^z\|_{\mathrm{op}} \|\boldsymbol{U}_1 - \boldsymbol{U}\|_{\mathrm{HS}}$$

we have, since $\boldsymbol{U}_1^* \boldsymbol{L}^z_{(1)}$ vanishes everywhere by definition,

$$\begin{aligned}\mathbb{P}\{\boldsymbol{U}_1 \in \mathcal{E} \ \wedge\ \|\boldsymbol{L}^z\|_{\mathrm{op}} \leq B\} &\leq \mathbb{P}\{\exists \boldsymbol{U} \in \mathcal{E} : \boldsymbol{U}^* \boldsymbol{L}^z_{(1)} = 0 \ \wedge\ \|\boldsymbol{L}^z\|_{\mathrm{op}} \leq B\} \\ &\leq \mathbb{P}\{\exists \boldsymbol{U} \in \mathcal{N} : \|\boldsymbol{U}^* \boldsymbol{L}^z_{(1)}\|_{\mathrm{HS}} \leq B\rho_\star\}. \end{aligned} \tag{7.16}$$

Now applying the union bound and (7.11) with $r = n_0 + 1$, using that $\mathcal{N} \subset \mathcal{E} \subset \mathcal{B}(\beta)^c$, we have

$$\begin{aligned}\mathbb{P}\{\boldsymbol{U}_1 \in \mathcal{E} \ \wedge \ \|\boldsymbol{L}^z\|_{\mathrm{op}} \le B\} &\le |\mathcal{N}| O_{\mathfrak{p},z}(\rho_\star^2)^{(n_0+1)^2(N-1)} \\ &\le O_{\mathfrak{p},z}(1)^N N^{(2n_0+3)N} \rho_\star^{2[(n_0+1)^2 - {n_0}^2 - \frac{3}{2}n_0 - \frac{3}{2}]N - O_{\mathfrak{p}}(1)} \\ &= O_{\mathfrak{p},z}(1)^N N^{(2n_0+3)N} \rho_\star^{(n_0-1)N - O_{\mathfrak{p}}(1)}.\end{aligned} \tag{7.17}$$

Recalling our assumption that $n_0 \ge 2$, one verifies the last expression is at most

$$O_{\mathfrak{p},z}(1)^N (c')^{N - O_{\mathfrak{p}}(1)} \lesssim_{\mathfrak{p}} e^{-N}$$

if we take

$$\rho_\star = c' N^{-(2n_0+3)/(n_0-1)}$$

for a sufficiently small constant $c'(\mathfrak{p}, z) > 0$. Now one verifies that the condition of Lemma 7.5 is satisfied for all $N \ge N_0$ for a sufficiently large constant $N_0(\mathfrak{p}, z) > 0$ when $\delta = N^{-\frac{n_0}{2} - 10}$. This completes the proof of Proposition 7.1 assuming Lemmas 7.2, 7.3, 7.4 and 7.5. □

7.2. **Proof of Lemma 7.2: Reduction to bases with a well-conditioned zeroth block.** With notation as in Corollary 4.2, we express $\boldsymbol{L}^z_{(1)}$ in block form as follows:

$$\boldsymbol{L}^z_{(1)} = \left(\begin{array}{c|c|c|c|c} \begin{matrix}\mathrm{row}_1(Y_0') \\ Y_0'' + (a_0 - z)\,\mathrm{I}_{N-1}\end{matrix} & \begin{matrix}\mathrm{row}_1(X_1') \\ X_1''\end{matrix} & \begin{matrix}\mathrm{row}_1(X_2') \\ X_2''\end{matrix} & \cdots & \begin{matrix}\mathrm{row}_1(X_{n_0}') \\ X_{n_0}''\end{matrix} \\ \hline \begin{matrix}\mathrm{row}_1(Y_1') \\ Y_1''\end{matrix} & \begin{matrix}0_{1,N-1} \\ -\,\mathrm{I}_{N-1}\end{matrix} & 0 & \cdots & 0 \\ \hline \begin{matrix}\mathrm{row}_1(Y_2') \\ Y_2''\end{matrix} & 0 & \begin{matrix}0_{1,N-1} \\ -\,\mathrm{I}_{N-1}\end{matrix} & & \vdots \\ \hline \vdots & \vdots & & \ddots & 0 \\ \hline \begin{matrix}\mathrm{row}_1(Y_{n_0}') \\ Y_{n_0}''\end{matrix} & 0 & \cdots & 0 & \begin{matrix}0_{1,N-1} \\ -\,\mathrm{I}_{N-1}\end{matrix} \end{array}\right).$$

Here, Y_k' (resp. X_k') is the matrix obtained by removing the first column from Y_k (resp. X_k), while Y_k'' (resp. X_k'') is the $N-1 \times N-1$ matrix obtained by removing the first row and column from Y_k (resp. X_k). $0_{1,N-1}$ is the row vector of $N-1$ zeros.

Let

$$\widehat{Y} = \begin{pmatrix} \mathrm{row}_1(Y_1') \\ \vdots \\ \mathrm{row}_1(Y_{n_0}') \end{pmatrix} \in \mathbb{M}_{n_0, N-1}(\mathbb{C}). \tag{7.18}$$

We first show that

$$\mathbb{P}\{\sigma_{\min}(\widehat{Y}) < c_{\mathfrak{p}}'\} \lesssim_{\mathfrak{p}} e^{-N} \tag{7.19}$$

if $c_{\mathfrak{p}}' > 0$ is sufficiently small. On the event that $\sigma_{\min}(\widehat{Y}) < c_{\mathfrak{p}}'$, there exists $x \in \mathbb{S}^{n_0 - 1}$ such that

$$c_{\mathfrak{p}}' > \|x^* \widehat{Y}\|_2 = \Big\| \sum_{k \in [n_0]} x(k) \sum_{\ell \in [n]} \overline{s_{k,\ell}}\, \mathrm{row}_1(X_\ell') \Big\|_2 = \|y(1)\,\mathrm{row}_1(X_1') + \cdots + y(n)\,\mathrm{row}_1(X_n')\|_2$$

where $y := \sum_{k=1}^{n_0} x(k) \overline{s_k}$. Since $s_1, \ldots, s_{n_0}$ are mutually orthogonal and nonzero it follows that $\|y\|_2 \gtrsim_{\mathfrak{p}} 1$, and in particular y must have a coordinate $y(\ell_0)$ of size $|y(\ell_0)| \gtrsim_{\mathfrak{p}} 1$. By projecting $\mathrm{row}_1(X_{\ell_0}')$ to the orthocomplement of $\widetilde{\mathsf{W}}_{(\ell_0)} = \mathrm{Span}\{\mathrm{row}_1(X_\ell')\}_{\ell \ne \ell_0}$ we see that $\mathrm{dist}(\mathrm{row}_1(X_{\ell_0}'), \widetilde{\mathsf{W}}_{(\ell_0)}) \lesssim_{\mathfrak{p}} c_{\mathfrak{p}}'$. Taking a union bound to fix ℓ_0 and conditioning on the rows $\mathrm{row}_1(X_\ell')$ with $\ell \ne \ell_0$, we are left with bounding the probability that the vector $\mathrm{row}_1(X_{\ell_0}')$ is within distance $O_{\mathfrak{p}}(c_{\mathfrak{p}}')$ of a fixed subspace of $\mathbb{C}^{N-1}$ of dimension $n-1$. Since X_ℓ has i.i.d. complex Gaussian entries of variance $1/N$,

by rotational invariance of the Gaussian measure we may assume $\widetilde{\mathsf{W}}_{(\ell_0)}$ is the coordinate subspace spanned by $e_2, \dots, e_n$, and we have reduced to bounding the probability of the event that

$$\sum_{j=n+1}^{N} |X_{\ell_0}(1,j)|^2 = O_{\mathfrak{p}}(c'_{\mathfrak{p}}).$$

Now we can take $c'_{\mathfrak{p}}$ sufficiently small to make the right hand side smaller than any fixed constant, so that this event has probability at most e^{-N} as soon as $N \geq 10n$, say, and (7.19) follows.

We now fix for the remainder of the proof an arbitrary realization of $\boldsymbol{L}_{(1)}$ such that $\|\boldsymbol{L}_{(1)}\|_{\mathrm{op}} \leq B$ and $\sigma_{\min}(\widehat{Y}) \geq c'_{\mathfrak{p}}$. This fixes the subspace $(\mathsf{V}^z_{(1)})^{\perp}$. We claim that

$$\left\{ \boldsymbol{U} = (\boldsymbol{u}_0, \dots, \boldsymbol{u}_{n_0}) : \boldsymbol{u}_0, \boldsymbol{u}_1, \dots, \boldsymbol{u}_{n_0} \text{ orthonormal in } (\mathsf{V}^z_{(1)})^{\perp} \right\} \cap \mathcal{A}(\alpha)^c = \emptyset \tag{7.20}$$

for $\alpha < c_{\mathfrak{p}}/(B^2 + |z|)$. From this the lemma clearly follows.

Turning to prove (7.20), towards a contradiction we let $\boldsymbol{U}$ be an arbitrary element of the left hand side of (7.20). Then there exists a unit vector $\boldsymbol{u} \in (\mathsf{V}^z_{(1)})^{\perp}$ such that $\|u^0\|_2 < \alpha$ (recall that $u^0, \dots, u^{n_0}$ denote the blocks of N coordinates of $\boldsymbol{u}$). For each k we write $u^k = (u^k(1), \check{u}^k)$, i.e. $\check{u}^k = (u^k(2), \dots, u^k(N))$. Fixing such a $\boldsymbol{u}$, we note that $\boldsymbol{u} \in (\mathsf{V}^z_{(1)})^{\perp}$ is equivalent to the statement that $\boldsymbol{u}^*$ is a left null vector for $\boldsymbol{L}^z_{(1)}$. In particular, for each $1 \leq k \leq n_0$ we have

$$\overline{u^0(1)}\, \mathrm{row}_1(X'_k) + (\check{u}^0)^* X''_k = (\check{u}^k)^*. \tag{7.21}$$

Since $\|\boldsymbol{L}_{(1)}\|_{\mathrm{op}} \leq B$ and $\|u^0\|_2 < \alpha$ this implies

$$\|\check{u}^k\|_2 \leq 2B/\alpha, \quad \forall\, 1 \leq k \leq n_0. \tag{7.22}$$

Again from the fact that $\boldsymbol{u}^*$ is a left null vector for $\boldsymbol{L}^z_{(1)}$, we have

$$\sum_{k=1}^{n_0} \overline{u^k(1)}\, \mathrm{row}_1(Y'_k) = -\overline{u^0(1)}\, \mathrm{row}_1(Y'_0) - (\check{u}^0)^*(Y''_0 + (a_0 - z)\, \mathrm{I}_{N-1}) - \sum_{k=1}^{n_0} (\check{u}^k)^* Y''_k$$

for each $1 \leq k \leq n_0$. Writing $\check{u}_1 = (u^1(1), \dots, u^{n_0}(1))$, the left hand side has norm at least $\sigma_{\min}(\widehat{Y})\|\check{u}_1\|_2 \geq c'_{\mathfrak{p}}\|\check{u}_1\|_2$. Since $\boldsymbol{L}_{(1)}$, and hence any of its submatrices, has norm at most B, together with (7.22) and $\|u^0\|_2 < \alpha$ this implies by the triangle inequality that the right hand side is $O_{\mathfrak{p}}(\alpha(B^2 + |z|))$. Putting these bounds together we have

$$\|\check{u}_1\|_2 \lesssim_{\mathfrak{p}} (B^2 + |z|)\alpha.$$

Combined with (7.22) and $\|u^0\|_2 < \alpha$ we conclude

$$1 = \|\boldsymbol{u}\|_2^2 = \|u^0\|_2^2 + \|\check{u}_1\|_2^2 + \sum_{k=1}^{n_0} \|\check{u}^k\|_2^2 \leq \alpha^2(1 + O_{\mathfrak{p}}(B^2 + |z|)^2)$$

and we obtain a contradiction for $\alpha < c_{\mathfrak{p}}/(B^2 + |z|)$ if $c_{\mathfrak{p}}$ is taken sufficiently small. This implies the the left hand side of (7.20) is empty for such a choice of α, and completes the proof.

7.3. **Proof of Lemma 7.3: Anti-concentration for matrix random walks.**

Lemma 7.7. *Let $M \in \mathbb{M}_{d,d'}(\mathbb{C})$ and let $Z = (\zeta_1, \dots, \zeta_{d'})$ be a vector of i.i.d. standard complex Gaussians. For $I \subseteq [d]$ let $M_I \in \mathbb{M}_{|I|,d'}(\mathbb{C})$ be the submatrix of M formed by the rows indexed by I. If M_I has full row rank for some $I \subseteq [d]$ of size d_0, then*

$$\sup_{w \in \mathbb{C}^d} \mathbb{P}(\|MZ - w\|_2 \leq \varepsilon) \leq \frac{O(\varepsilon^2)^{d_0}}{\det(M_I M_I^*)}\,. \tag{7.23}$$

Proof. By projecting MZ to the coordinate subspace $\mathbb{C}^I$ we may assume $I = [d]$. Let M have singular value decomposition $U\Sigma V^*$ with U and V square unitary matrices of respective dimensions d and d', and with $\Sigma = \mathrm{diag}(\sigma_1(M), \dots, \sigma_d(M))$. We have

$$\begin{aligned}\sup_{w\in\mathbb{C}^d} \mathbb{P}(\|MZ - w\|_2 \le \varepsilon) &= \sup_{w\in\mathbb{C}^d} \mathbb{P}(\|\Sigma V^* Z - U^* w\|_2 \le \varepsilon) \\ &= \sup_{w\in\mathbb{C}^d} \mathbb{P}(\|\Sigma V^* Z - w\|_2 \le \varepsilon) \\ &= \sup_{w\in\mathbb{C}^d} \mathbb{P}(\|\Sigma Z - w\|_2 \le \varepsilon)\end{aligned}$$

where we used the invariance of the Euclidean norm and the distribution of Z under unitary transformations of $\mathbb{C}^d$ and $\mathbb{C}^{d'}$, respectively. Now we have

$$\mathbb{P}(\|\Sigma Z - w\|_2 \le \varepsilon) = \mathbb{P}\Big(\sum_{j=1}^d |\sigma_j(M)\zeta_j - w_j|^2 \le \varepsilon^2\Big) \le \prod_{j=1}^d \mathbb{P}(|\sigma_j(M)\zeta_j - w_j| \le \varepsilon)$$

and the latter quantity is $O(\varepsilon)^{2d}/\prod_{j=1}^d \sigma_j(M)^2$ by the boundedness of the standard Gaussian density. The claim follows. □

Lemma 7.8 (Tensorization of anti-concentration (cf. [44]))**.** *Suppose that $\xi_1, \dots, \xi_m$ are independent non-negative random variables such that, for some $\varepsilon_1, \varepsilon_2 > 0$, $\mathbb{P}\{\xi_i \le \varepsilon_1\} \le \varepsilon_2$ for all $1 \le i \le m$. Then*

$$\mathbb{P}\Big\{ \sum_{i=1}^m \xi_i^2 \le \varepsilon_1^2 m \Big\} = O(\varepsilon_2)^m.$$

We now prove Lemma 7.3. Since $\boldsymbol{U} \in \mathcal{A}(\alpha)$ we have $\sigma_{\min}(\boldsymbol{R}) = \sigma_{\min}(U^0) \ge \alpha$, and hence

$$\det(\boldsymbol{R}^*\boldsymbol{R}) \gtrsim_{\mathfrak{p},\alpha} 1. \tag{7.24}$$

Letting $\varepsilon > 0$ be arbitrary, from (7.4) and Lemma 7.7 with $M = \boldsymbol{W}^*$, $M_I = \boldsymbol{R}^*$ we obtain

$$\sup_{M\in\mathbb{M}_{[0,n_0]}(\mathbb{C})} \mathbb{P}\big\{ \|M + \mathrm{Walk}(\boldsymbol{U})\|_{\mathrm{HS}} \le \varepsilon N^{-1/2} \big\} = O_{\mathfrak{p},\alpha}(\varepsilon^2)^{n_0(n_0+1)}.$$

In particular, from (7.2) we have that for each $2 \le j \le N$,

$$\mathbb{P}\big\{ \|\widetilde{\boldsymbol{U}}^* \mathrm{col}_j(\widetilde{\boldsymbol{L}}^z)\|_{\mathrm{HS}} \le \varepsilon N^{-1/2} \big\} = O_{\mathfrak{p},\alpha}(\varepsilon^2)^{n_0(n_0+1)}.$$

Since the columns $\mathrm{col}_j(\widetilde{\boldsymbol{L}}^z)$ are independent, from Lemma 7.8 we obtain

$$\mathbb{P}\{ \|\boldsymbol{U}^* \boldsymbol{L}^z_{(1)}\|_{\mathrm{HS}} \le \varepsilon \} = \mathbb{P}\Big\{ \sum_{j=1}^N \|\widetilde{\boldsymbol{U}}^* \mathrm{col}_j(\widetilde{\boldsymbol{L}}^z)\|_{\mathrm{HS}}^2 \le \varepsilon^2 \Big\} = O_{\mathfrak{p},\alpha}(\varepsilon^2)^{n_0(n_0+1)(N-1)}$$

which is (7.9).

For (7.11), under the assumptions $\sigma_{\min}(U^0) \ge \alpha$ and (7.10), with $I = \{k_1, \dots, k_r\}$ we have

$$\det(\boldsymbol{W}_I^* \boldsymbol{W}_I) \ge \det(\boldsymbol{R}^*\boldsymbol{R}) \cdot \beta^{2r} \gtrsim_{\mathfrak{p},\alpha,\beta} 1,$$

(recalling the notation around (7.8) and using the fact that

$$\det(\boldsymbol{W}^*_{I_{i+1}} \boldsymbol{W}_{I_{i+1}}) = |\operatorname{dist}(\boldsymbol{q}_{k_{i+1}}, \mathsf{W}_{I_i})|^2 \det(\boldsymbol{W}^*_{I_i} \boldsymbol{W}_{I_i})$$

for $I_i = \{k_1, ..., k_i\}$) and the proof concludes by following the same lines as we did for (7.9). □

7.4. **Proof of Lemma 7.4: Reduction to matrix walks of full rank.** We will argue iteratively, incrementing the rank parameter r in (7.10) from 0 to n_0. For brevity, in this section we denote the dilated Hilbert–Schmidt ball

$$\mathbb{B}_\star := 2\sqrt{n_0+1}\cdot \mathbb{B}^{[0,n_0]}_{[0,n_0]\times[N]}$$

(recall our notation $\mathbb{B}_S^T \subset \mathbb{M}_S^T(\mathbb{C})$ for the closed unit Hilbert–Schmidt ball). Note that $\boldsymbol{U}_1 \in \mathbb{B}_\star$ almost surely since its n_0+1 columns are unit vectors. For $\beta_1 > 0$ let

$$\mathcal{B}_\emptyset(\beta_1) := \big\{\, \boldsymbol{U} \in \mathbb{B}_\star : \mathrm{dist}(\boldsymbol{q}_k, \mathsf{W}_\emptyset) < \beta_1 \ \forall k \in [0,n_0] \,\big\} \tag{7.25}$$

and for $\beta_0 \geq \beta_1 > 0$, $1 \leq r \leq n_0$ and $k_1, \ldots, k_r \in [0,n_0]$ distinct,

$$\mathcal{B}_{(k_1,\ldots,k_r)}(\beta_0) := \big\{\, \boldsymbol{U} \in \mathbb{B}_\star : \mathrm{dist}(\boldsymbol{q}_{k_i}, \mathsf{W}_{\{k_1,\ldots,k_{i-1}\}}) \geq \beta_0 \quad \forall\, 1 \leq i \leq r \,\big\} \tag{7.26}$$

$$\widetilde{\mathcal{B}}_{(k_1,\ldots,k_r)}(\beta_0,\beta_1) := \mathcal{B}_{(k_1,\ldots,k_r)}(\beta_0) \setminus \bigcup_{k_{r+1}\notin\{k_1,\ldots,k_r\}} \mathcal{B}_{(k_1,\ldots,k_{r+1})}(\beta_1). \tag{7.27}$$

(interpreting $\mathsf{W}_{\{k_1,\ldots,k_{i-1}\}}$ as $\mathsf{W}_\emptyset$ for $i=1$).

Claim 7.9. *For any $\alpha \in (0, \frac{1}{2}]$ there exists $c_0(\mathfrak{p},\alpha) > 0$ such that for any $\beta_1 \in (0, c_0]$, there is a set $\mathcal{N}_\emptyset(\beta_1) \subset \mathbb{B}_\star \cap \mathcal{A}(\alpha)$ with*

$$|\mathcal{N}_\emptyset(\beta_1)| \leq O_{\mathfrak{p}}(1/\beta_1^2)^{(n_0+1)N} \tag{7.28}$$

and such that for any $\boldsymbol{U} \in \mathcal{B}_\emptyset(\beta_1) \cap \mathcal{A}(\alpha)$ there exists $\widehat{\boldsymbol{U}} \in \mathcal{N}_\emptyset(\beta_1)$ with

$$\|\boldsymbol{U} - \widehat{\boldsymbol{U}}\|_{\mathrm{HS}} \lesssim_{\mathfrak{p},\alpha} \beta_1. \tag{7.29}$$

Claim 7.10. *For any $\alpha, \beta_0 \in (0, \frac{1}{2}]$ there exists $c_1(\mathfrak{p},\alpha,\beta_0) > 0$ such that for any $\beta_1 \in (0, c_1]$, $1 \leq r \leq n_0$ and distinct $k_1, \ldots, k_r \in [0,n_0]$, there is a set $\mathcal{N}_{(k_1,\ldots,k_r)}(\beta_0,\beta_1) \subset \mathcal{B}_{(k_1,\ldots,k_r)}(\beta_0/2) \cap \mathcal{A}(\alpha)$ with*

$$|\mathcal{N}_{(k_1,\ldots,k_r)}(\beta_0,\beta_1)| \leq O_{\mathfrak{p}}(1/\beta_1^2)^{((r+1)n_0+1)N} \tag{7.30}$$

and such that for any $\boldsymbol{U} \in \widetilde{\mathcal{B}}_{(k_1,\ldots,k_r)}(\beta_0,\beta_1) \cap \mathcal{A}(\alpha)$ there exists $\widehat{\boldsymbol{U}} \in \mathcal{N}_{(k_1,\ldots,k_r)}(\beta_0,\beta_1)$ with

$$\|\boldsymbol{U} - \widehat{\boldsymbol{U}}\|_{\mathrm{HS}} \lesssim_{\mathfrak{p},\alpha,\beta_0} \beta_1. \tag{7.31}$$

We now conclude the proof of Lemma 7.4 on the above claims. By the same lines we used in the proof of Proposition 7.1 (cf. (7.16)–(7.17)) we have from Claim 7.9 and (7.9) that for $\beta_0 \leq c_0(\mathfrak{p},\alpha)$,

$$\begin{aligned}\mathbb{P}\{\boldsymbol{U}_1 \in \mathcal{B}_\emptyset(\beta_0) \cap \mathcal{A}(\alpha) \ \wedge\ \|\boldsymbol{L}^z\|_{\mathrm{op}} \leq B\} &\leq |\mathcal{N}_\emptyset(\beta_0)| O_{\mathfrak{p},\alpha}(B^2\beta_0^2)^{n_0(n_0+1)(N-1)} \\ &\leq O_{\mathfrak{p},\alpha,B}(1)^N O(\beta_0^2)^{[n_0(n_0+1)-(n_0+1)]N - O_{\mathfrak{p}}(1)}.\end{aligned}$$

Since $n_0(n_0+1)-(n_0+1) = {n_0}^2 - 1 \geq 1$ (recall our assumption that $n_0 \geq 2$), it follows that there exists $\beta_0 = \beta_0(\mathfrak{p},\alpha,B) > 0$ such that

$$\mathbb{P}\{\boldsymbol{U}_1 \in \mathcal{B}_\emptyset(\beta_0) \cap \mathcal{A}(\alpha) \ \wedge\ \|\boldsymbol{L}^z\|_{\mathrm{op}} \leq B\} \lesssim_{\mathfrak{p}} e^{-N}. \tag{7.32}$$

Similarly, from Claim 7.10 and (7.10), for any $\gamma_0 \in (0, \frac{1}{2}]$ and $\gamma_1 \leq c_1(\mathfrak{p},\alpha,\gamma_0)$, any $1 \leq r \leq n_0$ and distinct $k_1, \ldots, k_r \in [0,n_0]$,

$$\begin{aligned}&\mathbb{P}\{\boldsymbol{U}_1 \in \widetilde{\mathcal{B}}_{(k_1,\ldots,k_r)}(\gamma_0,\gamma_1) \cap \mathcal{A}(\alpha) \ \wedge\ \|\boldsymbol{L}^z\|_{\mathrm{op}} \leq B\} \\ &\qquad \leq |\mathcal{N}_{(k_1,\ldots,k_r)}(\gamma_0,\gamma_1)| O_{\mathfrak{p},\alpha,\gamma_0}(B^2\gamma_1^2)^{[n_0(n_0+1)+r](N-1)} \\ &\qquad \leq O_{\mathfrak{p},\alpha,B,\gamma_0}(1)^N O(\gamma_1^2)^{[n_0(n_0+1)+r-(r+1)n_0-1]N - O_{\mathfrak{p}}(1)}.\end{aligned}$$

Since $n_0(n_0+1)+r-(r+1)n_0-1 = n_0(n_0-r)+r-1 \geq 1$ for all n_0, r with $1 \leq r \leq n_0$ and $n_0 \geq 2$, it follows that there exists $\gamma_1(\mathfrak{p},\alpha,B,\gamma_0) \in (0,\gamma_0]$ such that

$$\mathbb{P}\{\boldsymbol{U}_1 \in \widetilde{\mathcal{B}}_{(k_1,\ldots,k_r)}(\gamma_0,\gamma_1) \cap \mathcal{A}(\alpha) \ \wedge\ \|\boldsymbol{L}^z\|_{\mathrm{op}} \leq B\} \lesssim_{\mathfrak{p}} e^{-N}. \tag{7.33}$$

Now we recursively obtain a sequence $\beta_0 \geq \beta_1 \geq \cdots \geq \beta_{n_0} =: \beta_\star > 0$ with $\beta_0 = \beta_0(\mathfrak{p}, \alpha, B)$ as in (7.32), and $\beta_k = \gamma_1(\mathfrak{p}, \alpha, B, \beta_{k-1})$ for $1 \leq k \leq n_0$. We cover

$$\mathcal{B}(\beta_\star) \cap \mathbb{B}_\star \subseteq \mathcal{B}_\emptyset(\beta_0) \cup \bigcup_{\substack{k_1,\ldots,k_{n_0} \in [0,n_0] \\ \text{distinct}}} \widetilde{\mathcal{B}}_{(k_1)}(\beta_0, \beta_1) \cup \widetilde{\mathcal{B}}_{(k_1,k_2)}(\beta_1, \beta_2) \cup \cdots \cup \widetilde{\mathcal{B}}_{(k_1,\ldots,k_{n_0})}(\beta_{n_0-1}, \beta_{n_0}).$$

Applying the union bound followed by (7.32) and (7.33) with $(\gamma_0, \gamma_1) = (\beta_{k-1}, \beta_k)$ for $1 \leq k \leq n_0$, we obtain

$$\mathbb{P}\{\, \boldsymbol{U}_1 \in \mathcal{B}(\beta) \cap \mathcal{A}(\alpha) \ \wedge\ \|\boldsymbol{L}^z\|_{\mathrm{op}} \leq B \,\} \lesssim_{\mathfrak{p}} e^{-N}$$

for any $0 < \beta \leq \beta_\star$. □

It remains to prove Claims 7.9 and 7.10. The proof of Claim 7.9 is essentially contained in that of Claim 7.10 – we give the proof of the latter and describe at the end which steps can be skipped to establish the former.

We will use the following elementary lemma.

Lemma 7.11 (Stability of distances to subspaces)**.** *Let $1 \leq k \leq m$. Let $b, a_1, \ldots, a_k, b', a'_1, \ldots, a'_k \in \mathbb{C}^m$ and $A = (a_1, \ldots, a_k)$. Suppose $\sigma_{\min}(A) \geq \alpha > 0$, $\|b - b'\| \leq \varepsilon$, and $\|a_i - a'_i\|_2 \leq \varepsilon$ for each $i \in [k]$. If $\varepsilon \leq \frac{1}{2}k^{-1/2}\alpha$, then*

$$\big|\operatorname{dist}(b, \operatorname{Span}(a_1, \ldots, a_k)) - \operatorname{dist}(b', \operatorname{Span}(a'_1, \ldots, a'_k)\big| \lesssim (1 + \|b\|_2)k\varepsilon/\alpha.$$

Proof. From the triangle inequality it suffices to show

$$\big|\operatorname{dist}(b, \operatorname{Span}(a_1, \ldots, a_k)) - \operatorname{dist}(b, \operatorname{Span}(a'_1, \ldots, a'_k)\big| \lesssim k\varepsilon\|b\|_2/\alpha. \tag{7.34}$$

Set $A^{(0)} = A$ and for each $1 \leq j \leq k$ let $A^{(j)}$ be obtained by replacing the first j columns of A with $a'_1, \ldots, a'_j$, so that $A^{(k)} = A'$. By expanding the left hand side of (7.34) as a telescoping sum and applying the triangle inequality, we see it suffices to show

$$|\operatorname{dist}(b, \operatorname{Span}(A^{(j)})) - \operatorname{dist}(b, \operatorname{Span}(A^{(j-1)}))| \leq \frac{4\|b\|_2\varepsilon}{\sigma_{\min}(A)} \tag{7.35}$$

for each $1 \leq j \leq k$.

Fix such a j, and let V be the span of all columns of $A^{(j)}$ but the jth one. Let $\tilde{b}, \tilde{a}_j, \tilde{a}'_j$ denote the projections of b, a_j, a'_j to $V^\perp$. Then the left hand side above is

$$\begin{aligned}
|\operatorname{dist}(\tilde{b}, \langle \tilde{a}_k \rangle) - \operatorname{dist}(\tilde{b}, \langle \tilde{a}'_k \rangle)| &= \left| \frac{\|\tilde{b} \wedge \tilde{a}_j\|_2}{\|\tilde{a}_j\|_2} - \frac{\|\tilde{b} \wedge \tilde{a}'_j\|_2}{\|\tilde{a}'_j\|_2} \right| \\
&\leq \frac{1}{\|\tilde{a}_j\|_2} \big| \|\tilde{b} \wedge \tilde{a}_j\|_2 - \|\tilde{b} \wedge \tilde{a}'_j\|_2 \big| + \left| \frac{\|\tilde{a}'_j\|_2}{\|\tilde{a}_j\|_2} - 1 \right| \frac{\|\tilde{b} \wedge \tilde{a}'_j\|_2}{\|\tilde{a}'_j\|_2} \\
&\leq \frac{1}{\|\tilde{a}_j\|_2} \|\tilde{b}\|_2 \varepsilon + \frac{\varepsilon}{\|\tilde{a}_j\|_2} \|\tilde{b}\|_2 \leq \frac{2\varepsilon\|b\|_2}{\|\tilde{a}_j\|_2}.
\end{aligned}$$

Noting that $\|\tilde{a}_j\|_2 = \operatorname{dist}(a_j, V)$, we see it suffices to show

$$\operatorname{dist}(a_j, V) \geq \sigma_{\min}(A)/2. \tag{7.36}$$

Now for some $c_1, \ldots, c_{j-1}, c_{j+1}, \ldots, c_k \in \mathbb{C}$ we have

$$\operatorname{dist}(a_j, V) = \Big\| a_j - \sum_{i \neq j} c_i \operatorname{col}_i(A^{(j)}) \Big\|_2 \geq \sigma_{\min}(A^{(j)}) \Big(1 + \sum_{i \neq j} |c_i|^2\Big)^{1/2} \geq \sigma_{\min}(A^{(j)}).$$

Now since $\|A^{(j)} - A\|_{\mathrm{op}} \le \|A^{(j)} - A\|_{\mathrm{HS}} \le \varepsilon\sqrt{k}$ for all $1 \le j \le k$,

$$\sigma_{\min}(A^{(j)}) \ge \sigma_{\min}(A) - \|A^{(j)} - A\|_{\mathrm{op}} \ge \sigma_{\min}(A) - \varepsilon\sqrt{k}$$

and (7.36) follows from the previous two displays and our assumption on ε. □

Proof of Claim 7.10. The key property of $\boldsymbol{W} = \boldsymbol{W}(\boldsymbol{U})$ is that the columns of $\boldsymbol{Q}$ are determined by those of $\boldsymbol{U}$. Indeed, let $S \in \mathbb{M}_{[n]}^{[0,n_0]}(\mathbb{C})$ have columns $s_0, \dots, s_{n_0}$ (cf. (6.3)) and let $\widetilde{S} \in \mathbb{M}_{[0,n]}^{[0,n_0]}(\mathbb{C})$ be the result of adding the standard basis vector $e_1 = (1, 0, \dots, 0) \in \mathbb{C}^n$ as zeroth row. We have

$$\begin{pmatrix} U^0 \\ \boldsymbol{Q} \end{pmatrix} = (\widetilde{S} \otimes \mathrm{I}_N)\boldsymbol{U}\,, \qquad \text{in particular } \ \boldsymbol{Q} = (S \otimes \mathrm{I}_N)\boldsymbol{U}. \tag{7.37}$$

Since $s_1, \dots, s_{n_0}$ are linearly independent, it follows that $\widetilde{S}$ has full column rank, so

$$\boldsymbol{U} = (\widetilde{S}^{+} \otimes \mathrm{I}_N) \begin{pmatrix} U^0 \\ \boldsymbol{Q} \end{pmatrix} \tag{7.38}$$

where $\widetilde{S}^{+} = (\widetilde{S}^{*}\widetilde{S})^{-1}\widetilde{S}^{*}$ is the Moore–Penrose pseudoinverse of $\widetilde{S}$.

For $U^0 = (u_0^0, \dots, u_{n_0}^0) \in \mathbb{M}_{[N]}^{[0,n_0]}(\mathbb{C})$ and $\boldsymbol{u}_1', \dots, \boldsymbol{u}_r' \in \mathbb{C}^{[n_0]\times[N]}$ with $r \le n_0$, we write $\mathsf{W}^{(r)}(U^0, \boldsymbol{u}_1', \dots, \boldsymbol{u}_r')$ for the span of the columns of $\boldsymbol{R}(U^0)$ together with $(S \otimes \mathrm{I}_N)\boldsymbol{u}_1, \dots, (S \otimes \mathrm{I}_N)\boldsymbol{u}_r$, where

$$\boldsymbol{u}_k = \begin{pmatrix} u_k^0 \\ \boldsymbol{u}_k' \end{pmatrix}.$$

In particular, for $I = \{k_1, \dots, k_r\} \subset [0, n_0]$ we have

$$\mathsf{W}_I(\boldsymbol{U}) = \mathsf{W}^{(r)}(U^0, \boldsymbol{u}_{k_1}', \dots, \boldsymbol{u}_{k_r}'). \tag{7.39}$$

We introduce this notation to make it clear that this subspace is fixed by the partial data $U^0, \boldsymbol{u}_{k_1}', \dots, \boldsymbol{u}_{k_r}'$, which will be crucial for the net construction.

For any $\beta \in (0, 1/2]$ we let $\Sigma_0(\beta, \alpha)$ be a β-net for the set

$$\left\{ U \in 2\sqrt{n_0 + 1} \cdot \mathbb{B}_{[N]}^{[0,n_0]} : \sigma_{\min}(U) \ge \alpha \right\}$$

and let $\Sigma_1(\beta)$ be a β-net for $\mathbb{B}^{[n_0]\times[N]}$ (the unit ball in $\mathbb{C}^{[n_0]\times[N]}$). (Both nets are taken with respect to the Euclidean metric.) For each $\widehat{U}^0 \in \Sigma_0(\beta, \alpha)$, $\widehat{\boldsymbol{u}}_1', \dots, \widehat{\boldsymbol{u}}_r' \in \Sigma_1(\beta)$ and any $R \ge 1$, we let $\Sigma^{(r)}_{\widehat{U}^0, \widehat{\boldsymbol{u}}_1', \dots, \widehat{\boldsymbol{u}}_r'}(\beta, R)$ be a (Euclidean) β-net for the ball of radius R in $\mathsf{W}^{(r)}(\widehat{U}^0, \widehat{\boldsymbol{u}}_1', \dots, \widehat{\boldsymbol{u}}_r')$. For $r = 0$ we let $\Sigma^{(0)}_{\widehat{U}^0}(\beta, R)$ be a β-net for the ball of radius R in $\mathsf{W}^{(0)}(\widehat{U}^0)$. By Lemma 3.1 we may choose these nets so that

$$|\Sigma_0(\beta, \alpha)| \le O_{\mathfrak{p}}(1/\beta^2)^{(n_0+1)N}, \tag{7.40}$$

$$|\Sigma_1(\beta)| \le O(1/\beta^2)^{n_0 N}, \tag{7.41}$$

$$|\Sigma^{(0)}_{\widehat{U}^0}(\beta, R)|,\ |\Sigma^{(r)}_{\widehat{U}^0, \widehat{\boldsymbol{u}}_1', \dots, \widehat{\boldsymbol{u}}_r'}(\beta, R)| \le O(R^2/\beta^2)^{O_{\mathfrak{p}}(1)}. \tag{7.42}$$

Now let $1 \le r \le n_0$ and fix distinct $k_1, \dots, k_r \in [0, n_0]$. In the sequel we write $I_i = \{k_1, \dots, k_i\}$. Let $\mathcal{N}$ be the set of $\widehat{\boldsymbol{U}} \in \mathbb{B}_\star$ such that

$$\widehat{U}^0 \in \Sigma_0(\beta_1, \alpha), \qquad \widehat{\boldsymbol{u}}_k' \in \Sigma_1(\beta_1) \quad \forall\, k \in I_r \tag{7.43}$$

and

$$\widehat{\boldsymbol{u}}_k' \in \left\{ (S' \otimes \mathrm{I}_N) \begin{pmatrix} \widehat{u}_k^0 \\ \widehat{\boldsymbol{q}}_k \end{pmatrix} \,:\, \widehat{\boldsymbol{q}}_k \in \Sigma_{\widehat{U}^0, (\widehat{\boldsymbol{u}}_k' : k \in I_r)}(\beta_1, R) \right\} \qquad \forall\, k \notin I_r, \tag{7.44}$$

with $R = O_{\mathfrak{p}}(1)$ to be taken sufficiently large, and where S' is the matrix $\widetilde{S}^{+}$ with its first row removed (see (7.38)).

Fix an arbitrary $\boldsymbol{U} \in \widetilde{\mathcal{B}}_{(k_1,\dots,k_r)}(\beta_0,\beta_1) \cap \mathcal{A}(\alpha)$. We claim there exists $\widehat{\boldsymbol{U}} \in \mathcal{N}$ such that (7.31) holds and $\widehat{\boldsymbol{U}} \in \mathcal{B}_{(k_1,\dots,k_r)}(\beta_0/2)$. From this the claim will follow by taking $\mathcal{N}_{(k_1,\dots,k_r)}(\beta_0,\beta_1) = \mathcal{N} \cap \mathcal{B}_{(k_1,\dots,k_r)}(\beta_0/2)$, noting that $\widehat{\boldsymbol{U}} \in \mathcal{A}(\alpha)$ by (7.43), and that the bound (7.30) follows from (7.40)–(7.42).

We first fix the following submatrices of $\widehat{\boldsymbol{U}}$: take $\widehat{U}^0 \in \Sigma_0(\beta_1,\alpha)$ and for each $k \in I_r$ take $\widehat{\boldsymbol{u}}'_k \in \Sigma_1(\beta_1)$ such that

$$\|U^0 - \widehat{U}^0\|_2 \le \beta_1 \quad \text{and} \quad \|\boldsymbol{u}'_k - \widehat{\boldsymbol{u}}'_k\|_2 \le \beta_1 \quad \forall k \in I_r. \tag{7.45}$$

By the triangle inequality, it only remains to choose the vectors $\widehat{\boldsymbol{u}}'_k$ for $k \notin I_r$ so that

$$\|\boldsymbol{u}'_k - \widehat{\boldsymbol{u}}'_k\|_2 \lesssim_{\mathfrak{p},\alpha,\beta_0} \beta_1 \tag{7.46}$$

and so that the resulting matrix $\widehat{\boldsymbol{U}}$ lies in $\mathcal{B}_{(k_1,\dots,k_r)}(\beta_0/2)$.

For arbitrary $k \in [0,n_0]$, since $\boldsymbol{q}_k = (S \otimes \mathrm{I}_N)\boldsymbol{u}_k$ and $\|\boldsymbol{u}_k\|_2 \le 2\sqrt{n_0+1}$, it follows that

$$\|\boldsymbol{q}_k\|_2 \le 2\sqrt{n_0+1}\|S\|_{\mathrm{op}} \lesssim_{\mathfrak{p}} 1.$$

For $k \in I_r$, with

$$\widehat{\boldsymbol{q}}_k := (S \otimes \mathrm{I}_N)\widehat{\boldsymbol{u}}_k, \qquad \widehat{\boldsymbol{u}}_k = \begin{pmatrix} \widehat{u}_k^0 \\ \widehat{\boldsymbol{u}}'_k \end{pmatrix},$$

we similarly have

$$\|\boldsymbol{q}_k - \widehat{\boldsymbol{q}}_k\|_2 \lesssim_{\mathfrak{p}} \beta_1 \quad \forall\, k \in I_r. \tag{7.47}$$

Since $\boldsymbol{U} \in \widetilde{\mathcal{B}}_{(k_1,\dots,k_r)}(\beta_0,\beta_1)$,

$$\operatorname{dist}(\boldsymbol{q}_{k_i}, \mathsf{W}_{I_{i-1}}) = \operatorname{dist}(\boldsymbol{q}_{k_i}, \mathsf{W}^{(i-1)}(U^0, \boldsymbol{u}'_{k_1}, \dots, \boldsymbol{u}'_{k_{i-1}})) \ge \beta_0 \qquad \forall\, 1 \le i \le r \tag{7.48}$$

and for any $k \notin I_r$,

$$\operatorname{dist}(\boldsymbol{q}_k, \mathsf{W}_{I_r}) = \operatorname{dist}(\boldsymbol{q}_k, \mathsf{W}^{(r)}(U^0, \boldsymbol{u}'_{k_1}, \dots, \boldsymbol{u}'_{k_r})) \le \beta_1 \tag{7.49}$$

(recalling (7.39)). Since $U^0 \in \mathcal{A}(\alpha)$ we have $\sigma_{\min}(\boldsymbol{R}) = \sigma_{\min}(U^0) \ge \alpha$. Together with (7.48) this implies

$$\sigma_{\min}(\boldsymbol{W}_{I_{i-1}}) \gtrsim_{\alpha,\beta_0} 1 \quad \forall\, 1 \le i \le r$$

(where $\boldsymbol{W}_{I_0} := \boldsymbol{R}$). This together with (7.45), Lemma 7.11 and (7.48), and assuming $c_1(\mathfrak{p},\alpha,\beta_0)$ is sufficiently small, implies

$$\operatorname{dist}(\widehat{\boldsymbol{q}}_{k_i}, \mathsf{W}^{(i-1)}(\widehat{U}^0, \widehat{\boldsymbol{u}}'_{k_1}, \dots, \widehat{\boldsymbol{u}}'_{k_{i-1}})) \ge \beta_0 - O_{\mathfrak{p},\alpha,\beta_0}(\beta_1) \qquad \forall\, 1 \le i \le r. \tag{7.50}$$

We similarly have

$$\operatorname{dist}(\boldsymbol{q}_k, \mathsf{W}^{(r)}(\widehat{U}^0, \widehat{\boldsymbol{u}}'_{k_1}, \dots, \widehat{\boldsymbol{u}}'_{k_r})) \lesssim_{\mathfrak{p},\alpha,\beta_0} \beta_1 \qquad \forall\, k \notin I_r. \tag{7.51}$$

By the triangle inequality, taking $R = O_{\mathfrak{p}}(1)$ sufficiently large, there exist $\widehat{\boldsymbol{q}}_k \in \Sigma_{\widehat{U}^0,\widehat{\boldsymbol{u}}'_{k_1},\dots,\widehat{\boldsymbol{u}}'_{k_r}}(\beta_1,R)$ such that

$$\|\boldsymbol{q}_k - \widehat{\boldsymbol{q}}_k\|_2 \lesssim_{\mathfrak{p},\alpha,\beta_0} \beta_1 \qquad \forall\, k \notin I_r.$$

Finally, setting $\widehat{\boldsymbol{u}}'_k = (S' \otimes \mathrm{I}_N)\widehat{\boldsymbol{q}}_k$, we have for each $k \notin I_r$ that

$$\|\boldsymbol{u}'_k - \widehat{\boldsymbol{u}}'_k\|_2 = \left\|(S' \otimes \mathrm{I}_N)\begin{pmatrix} u_k^0 - \widehat{u}_k^0 \\ \boldsymbol{q}_k(\boldsymbol{U}) - \widehat{\boldsymbol{q}}_k \end{pmatrix}\right\|_2 \lesssim_{\mathfrak{p}} \left\|\begin{pmatrix} u_k^0 - \widehat{u}_k^0 \\ \boldsymbol{q}_k(\boldsymbol{U}) - \widehat{\boldsymbol{q}}_k \end{pmatrix}\right\|_2 \lesssim_{\mathfrak{p},\alpha,\beta_0} \beta_1$$

giving (7.46) as desired. That $\widehat{\boldsymbol{U}} \in \mathcal{B}_{(k_1,\dots,k_r)}(\beta_0/2)$ follows from (7.50) and taking the constant $c_1(\mathfrak{p},\alpha,\beta_0)$ sufficiently small.

The proof of Claim 7.9 follows similar lines. The net $\mathcal{N}_\emptyset(\beta_1)$ is taken as in (7.43)–(7.44) with $I_r = \emptyset$ (so no columns $\widehat{\boldsymbol{u}}'_k$ are fixed at this initial stage). The estimates (7.49) and (7.51) are obtained by the same lines, with the subspaces $\mathsf{W}_\emptyset = \mathsf{W}^{(0)}(U^0)$ and $\mathsf{W}^{(0)}(\widehat{U}^0)$, while we skip (7.48) and (7.50). □

7.5. **Proof of Lemma 7.5: Constructing a net for structured matrices.** We first record two elementary lemmas.

Lemma 7.12. *Let $A \in \mathbb{M}_m(\mathbb{C})$ with $|\det A| \geq \kappa$ and columns $x_1, \dots, x_m \in \mathbb{B}^m$. Suppose that for some $v \in \mathbb{C}^m$ we have*

$$|\det((v, A_J))| < \delta \qquad \forall\, J \in \binom{[m]}{m-1} \tag{7.52}$$

where we write A_J for the $m \times |J|$ submatrix with columns $\{x_j : j \in J\}$. Then $\|v\|_2 < m\delta/\kappa$.

Proof. We expand $v = \sum_{i=1}^m b_i x_i$. Applying (7.52) with $J = [m] \setminus \{j\}$ gives

$$\delta > |\det((v, A_J)| = |b_j||\det A| \geq \kappa |b_j| \qquad \forall\, j \in [m],$$

so $\|b\|_2 < \sqrt{m}\delta/\kappa$. From our assumptions we have $\|A\|_{\mathrm{op}} \leq \|A\|_{\mathrm{HS}} \leq \sqrt{m}$, and so

$$\|v\|_2 = \|Ab\|_2 \leq \sqrt{m}\|b\|_2 < m\delta/\kappa$$

as desired. □

Lemma 7.12 will be used in conjunction with the following, which locates a well-conditioned basis of rows in the matrix U^0. Recall that $v_i^0 \in \mathbb{C}^{[0,n_0]}$, $i \in [N]$ denote the rows of U^0.

Lemma 7.13. *For any $\boldsymbol{U} \in \mathcal{A}(\alpha)$, there exists $I^* = \{i_0, \dots, i_{n_0}\} \subset [N]$ such that*

$$\|v_{i_0}^0 \wedge \cdots \wedge v_{i_k}^0\|_2 \geq (\alpha/\sqrt{N})^{k+1} \tag{7.53}$$

for each $0 \leq k \leq n_0$.

Proof. We iteratively construct I^* as follows: first, there exists $i_0 \in [N]$ such that $\|v_{i_0}^0\|_2 \geq \alpha/\sqrt{N}$, as otherwise $\sigma_{\min}(U^0) \leq \|U^0\|_{\mathrm{op}} \leq \|U^0\|_{\mathrm{HS}} < \alpha$. Now for $1 \leq k \leq n_0 - 1$, having picked $i_0, \dots, i_k$, there exists $i_{k+1} \in [N]$ such that $\mathrm{dist}(v_{i_{k+1}}^0, \langle v_{i_0}^0, \dots, v_{i_k}^0\rangle) \geq \alpha/\sqrt{N}$, as otherwise, for any unit vector $u \in \langle v_{i_0}^0, \dots, v_{i_k}^0\rangle^\perp$ we would have

$$\|U^0 u\|_2^2 = \sum_{j \in [N] \setminus \{i_0, \dots, i_k\}} |\langle v_j^0, u\rangle|^2 < \alpha^2,$$

which contradicts $\sigma_{\min}(U^0) \geq \alpha$. (7.53) now follows from the base-times-height formula for the norm of the wedge product. □

Fix an arbitrary $k_0 \in [n_0]$. Our proof of Lemma 7.5 is divided into two cases depending on $\mathfrak{p}$:

(A) $s_{k_0,\ell_0} \neq 0$ for some $\ell_0 \in [n_0 + 1, n]$;

(B) $s_{k_0,\ell_0} \neq 0$ for some $\ell_0 \in [n_0]$.

Since the vector $s_{k_0} \in \mathbb{C}^n$ has nonzero norm at least one of these cases must hold. Our construction of $\mathcal{N}$ will be different for each case.

In what follows, for parameters $R, \rho > 0$ we let $\Sigma_0(R, \rho)$ be a $\rho/\sqrt{N}$-net for the ball of radius R in $\mathbb{C}^{[0,n_0]}$, and for $J \subseteq [N]$, we let $\Sigma_J(R, \rho) \subset \mathbb{M}_{[N]}^{[0,n_0]}(\mathbb{C})$ be a ρ-net for the Hilbert–Schmidt ball of radius R on the subspace of matrices supported on the rows indexed by J. (We take $\Sigma_\emptyset(R, \rho)$ to consist of the zero matrix.) As before, all nets are with respect to the appropriate Euclidean metric. By Lemma 3.1 we may take

$$|\Sigma_0(R, \rho)| = O_{\mathfrak{p}}(R^2 N/\rho^2)^{n_0+1}, \qquad |\Sigma_J(R, \rho)| = O(R^2/\rho^2)^{(n_0+1)|J|}. \tag{7.54}$$

Proof of Lemma 7.5 under Case A. Let $\mathcal{N}'$ be the set of matrices $\boldsymbol{U} = (U^0, \dots, U^{n_0}) \in \mathbb{M}_{[0,n_0]\times[N]}^{[0,n_0]}(\mathbb{C})$ such that $U^k \in \Sigma_{[N]}(\sqrt{n_0 + 1}, \rho_\star)$ for each $k \in [0, n_0] \setminus \{k_0\}$, and U^{k_0} has rows given by

$$v_j^{k_0} = -\frac{1}{s_{k_0,\ell_0}} \sum_{k \in [0,n_0] \setminus \{k_0\}} s_{k,\ell_0} v_j^k, \qquad j \in [N], \tag{7.55}$$

where v_j^k is the jth row of U^k. Now for each $\boldsymbol{U}' \in \mathcal{N}'$ take an element $\boldsymbol{U}'' \in \mathcal{E}$ within a distance $2\,\mathrm{dist}(\boldsymbol{U}', \mathcal{E})$ of $\boldsymbol{U}'$, and let $\mathcal{N} \subset \mathcal{E}$ be the set of all $\boldsymbol{U}''$ obtained in this way. By construction we have

$$|\mathcal{N}| \le |\mathcal{N}'| \le |\Sigma_{[N]}(\sqrt{n_0+1}, \rho_\star)|^{n_0} = O_{\mathfrak{p}}(1/\rho_\star^2)^{n_0(n_0+1)N}. \tag{7.56}$$

We claim that for any $\boldsymbol{U} \in \mathcal{E}$ there exists $\widehat{\boldsymbol{U}} \in \mathcal{N}'$ with

$$\|\boldsymbol{U} - \widehat{\boldsymbol{U}}\|_{\mathrm{HS}} \lesssim_{\mathfrak{p},\alpha} \rho_\star. \tag{7.57}$$

The lemma (under Case A) clearly follows from this and the triangle inequality, and replacing δ above with $c(\mathfrak{p},\alpha)\delta$ for a sufficiently small constant $c(\mathfrak{p},\alpha) > 0$.

To show (7.57), for arbitrary $\boldsymbol{U} \in \mathcal{E}$, we take $\widehat{\boldsymbol{U}} \in \mathcal{N}'$ to be an element such that $\|U^k - \widehat{U}^k\|_{\mathrm{HS}} < \rho_\star$ for each $k \ne k_0$. (Recall that $\widehat{U}^{k_0}$ is determined by (7.55) once U^k for $k \ne k_0$ are fixed.) It only remains to show

$$\|U^{k_0} - \widehat{U}^{k_0}\|_{\mathrm{HS}} \lesssim_{\mathfrak{p},\alpha} \rho_\star. \tag{7.58}$$

Let $I^* = \{i_0, \dots, i_{n_0}\} \subset [N]$ satisfying (7.53) for $\boldsymbol{U}$. Since $\boldsymbol{U} \in \mathrm{Struct}(\delta) \subset \mathrm{Struct}_1(\delta)$ we have

$$|\Delta_{j,I}| := |\det(w_j, v_I^0)| < \delta \qquad \forall\, j \in [N],\ \forall\, I \in \binom{I^*}{n_0}$$

where we write $v_I^0 := (v_j^0 : j \in I)$, and for ease of notation we write $w_j := w_j^{\ell_0}$ (ℓ_0 being fixed at this stage). From (7.53) (with $k = n_0$) and Lemma 7.12 it follows that

$$\|w_j\|_2 < \delta(n_0+1)(\sqrt{N}/\alpha)^{n_0+1} \qquad \forall\, j \in [N].$$

Now for each $j \in [N]$, writing $\widehat{v}_j^{k_0}$ for the jth row of $\widehat{U}^{k_0}$, from (6.6), the triangle and AM-GM inequalities, and the above inequality, we have

$$\begin{aligned}\|v_j^{k_0} - \widehat{v}_j^{k_0}\|_2 &= \frac{1}{|s_{k_0,\ell_0}|^2}\Big\|w_j - \sum_{k \in [0,n_0]\setminus\{k_0\}} s_{k,\ell_0}(v_j^k - \widehat{v}_j^k)\Big\|^2 \\ &\lesssim_{\mathfrak{p},\alpha} \delta^2 N^{n_0+1} + \sum_{k \in [0,n_0]\setminus\{k_0\}} \|v_j^k - \widehat{v}_j^k\|_2^2.\end{aligned}$$

Summing over j and applying the bounds $\|U^k - \widehat{U}^k\|_{\mathrm{HS}} < \rho_\star$ for $k \ne k_0$ and our assumed bound on δ, we obtain (7.58) as desired. □

Proof of Lemma 7.5 under Case B. Let Λ_0 be a $\rho_\star/\sqrt{N}$-mesh for the interval $[-R, R]$, for some $R = O_{\mathfrak{p}}(1)$ to be taken sufficiently large. We take $\mathcal{N}''$ be the set of matrices $\widehat{\boldsymbol{U}} = (\widehat{U}^0, \dots, \widehat{U}^{n_0}) \in \mathbb{M}_{[0,n_0]\times[N]}^{[0,n_0]}(\mathbb{C})$ such that

$$\widehat{U}^k \in \Sigma_{[N]}(\sqrt{n_0+1}, \rho_\star) \qquad \forall\, k \in [n_0] \setminus \{k_0\}, \tag{7.59}$$

and for some (possibly empty) $J \subseteq [N]$,

$$\widehat{U}^0_{J^c} \in \Sigma_{J^c}(\rho_\star^{1/2}, \rho_\star) \quad \text{and} \quad \widehat{U}^{k_0}_{J^c} \in \Sigma_{J^c}(\sqrt{n_0+1}, \rho_\star) \tag{7.60}$$

while for each $j \in J$,

$$\widehat{v}_j^0 \in \Sigma_0(1, \rho_1), \qquad \rho_1 := \rho_\star^{3/2}/N \tag{7.61}$$

and

$$\widehat{v}_j^{k_0} = \frac{1}{s_{k_0,\ell_0}}\Big(\widehat{w}_j - \sum_{k \in [0,n_0]\setminus\{k_0\}} s_{k,\ell_0}\widehat{v}_j^k\Big) \tag{7.62}$$

for some $\widehat{w}_j \in \Lambda_0 \cdot \langle \widehat{v}_j^0 \rangle$. Here we write $J^c := [N] \setminus J$, and $\widehat{U}_I^k$ for the matrix obtained from $\widehat{U}^k$ by zeroing out the rows with indices in I^c.

Constructing $\mathcal{N} \subset \mathcal{E}$ from $\mathcal{N}''$ analogously to how $\mathcal{N}$ was obtained from $\mathcal{N}'$ in the proof for Case A, we have by construction that

$$\begin{aligned}
|\mathcal{N}| &\le |\mathcal{N}''| \\
&\le |\Sigma_{[N]}(\sqrt{n_0+1},\rho_\star)|^{n_0-1} \sum_{J\subseteq[N]} |\Sigma_{J^c}(\rho_\star^{1/2},\rho_\star)||\Sigma_{J^c}(\sqrt{n_0+1},\rho_\star)||\Sigma_0(1,\rho_1)|^{|J|}|\Lambda_0|^{|J|} \\
&= O_{\mathfrak{p}}(1/\rho_\star^2)^{(n_0-1)(n_0+1)N} \sum_{J\subseteq[N]} O_{\mathfrak{p}}(1/\rho_\star)^{(n_0+1)(N-|J|)} O_{\mathfrak{p}}(1/\rho_\star^2)^{(n_0+1)(N-|J|)} \\
&\qquad\qquad \times O_{\mathfrak{p}}(1/\rho_1^2)^{(n_0+1)|J|} O_{\mathfrak{p}}(N/\rho_\star^2)^{|J|} \\
&= O_{\mathfrak{p}}(1)^N O(1/\rho_\star^2)^{(n_0+1/2)(n_0+1)N} \sum_{J\subseteq[N]} O(N)^{(2n_0+3)|J|} O(1/\rho_\star^2)^{|J|} \\
&= O_{\mathfrak{p}}(N)^{(2n_0+3)N} O(1/\rho_\star^2)^{({n_0}^2+\frac{3}{2}n_0+\frac{3}{2})N}
\end{aligned}$$

where in the last line we simply bounded $|J| \le N$ and absorbed the harmless factor 2^N in the implied constant.

As in the proof for Case A, we will be done if we can show that for every $\boldsymbol{U} \in \mathcal{E}$ there exists $\widehat{\boldsymbol{U}} \in \mathcal{N}''$ such that (7.57) holds. Fixing now an arbitrary $\boldsymbol{U} \in \mathcal{E}$, for each $k \in [n_0] \setminus \{k_0\}$ we let $\widehat{U}^k$ be as in (7.59) with $\|U^k - \widehat{U}^k\|_{\mathrm{HS}} < \rho_\star$. It only remains to pick $\widehat{U}^0, \widehat{U}^{k_0}$ as in (7.60)–(7.62) for some choice of $J = J(\boldsymbol{U}) \subseteq [N]$ with

$$\|U^k - \widehat{U}^k\|_{\mathrm{HS}} \lesssim_{\mathfrak{p},\alpha} \rho_\star\,, \quad k \in \{0, k_0\}. \tag{7.63}$$

We take J to be the set of large rows of U^0:

$$J = J(U^0) = \{j \in [N] : \|v_j^0\|_2 \ge \rho_\star^{1/2}/\sqrt{N}\}.$$

We have $\|U^0_{J^c}\|_{\mathrm{HS}} < \rho_\star^{1/2}$, and so there exists $\widehat{U}^0_{J^c}$ as in (7.60) with $\|U^0_{J^c} - \widehat{U}^0_{J^c}\|_{\mathrm{HS}} \le \rho_\star$. Completing $\widehat{U}^0$ with rows $\widehat{v}_j^0$ as in (7.61) such that $\|v_j^0 - \widehat{v}_j^0\|_2 \le \rho_1$ for each $j \in J$, we obtain (7.63) for $k = 0$.

Since $\|U^{k_0}\|_{\mathrm{HS}} \le \sqrt{n_0+1}$ we can take $\widehat{U}^{k_0}_{J^c}$ as in (7.60) with $\|U^{k_0}_{J^c} - \widehat{U}^{k_0}_{J^c}\|_{\mathrm{HS}} \le \rho_\star$. It only remains to show there exist $\widehat{v}_j^{k_0}, j \in J$ as in (7.62) such that

$$\sum_{j\in J} \|v_j^{k_0} - \widehat{v}_j^{k_0}\|_2^2 \lesssim_{\mathfrak{p},\alpha} \rho_\star^2. \tag{7.64}$$

Turning to this task, let $I^* = \{i_0, \dots, i_{n_0}\} \subset [N]$ satisfying (7.53) for $\boldsymbol{U}$. For ease of writing we set $m := n_0 + 1$ and denote $x_k := v^0_{i_{k-1}}$ for $1 \le k \le m$. We also use the shorthand $x_{\wedge I} := x_{k_1} \wedge \cdots \wedge x_{k_\ell}$ for $I = \{k_1, \dots, k_\ell\} \subseteq [m]$ with $k_1 < \cdots < k_m$. From (7.53) we have

$$|x_1 \wedge \cdots \wedge x_m| \ge (\alpha/\sqrt{N})^m =: \kappa. \tag{7.65}$$

Consider an arbitrary $j \in J$. From Lemma 7.12, and the bounds (7.65) and $\|v_j^0\|_2 \ge \rho_\star^{1/2}/\sqrt{N}$ it follows that there exists $\widehat{I} = \widehat{I}_j \subset [m]$ of size $m-1$ such that

$$|v_j^0 \wedge x_{\wedge \widehat{I}}| \ge \frac{\kappa \rho_\star^{1/2}}{m\sqrt{N}}. \tag{7.66}$$

Since $\boldsymbol{U} \in \mathrm{Struct}(\delta) \subset \mathrm{Struct}_2(\delta)$ we have that for every $I \subset [m]$ of size $m-2$,

$$|\widetilde{\Delta}_{j,I}| := |\det(w_j, v_j^0, (x_k)_{k\in I})| = |w_j \wedge v_j^0 \wedge x_{\wedge I}| < \delta.$$

Letting $\widetilde{w}_j, \widetilde{x}_1, \dots, \widetilde{x}_m$ denote the projections of $w_j, x_1, \dots, x_m$ to $\langle v_j^0\rangle^\perp$, we have in particular that

$$\delta > |w_j \wedge v_j^0 \wedge x_{\wedge I}| = \|v_j^0\|_2 \|\widetilde{w}_j \wedge \widetilde{x}_{\wedge I}\|_2 \ge \frac{\rho_\star^{1/2}}{\sqrt{N}} \|\widetilde{w}_j \wedge \widetilde{x}_{\wedge I}\|_2 \quad \forall\, I \in \binom{\widehat{I}}{m-2}. \tag{7.67}$$

On the other hand, since $\|v_j^0\|_2 \le 1$, we have from (7.66) that

$$\|x_{\wedge \widehat{I}}\|_2 \ge |v_j^0 \wedge x_{\wedge \widehat{I}}| \ge \frac{\kappa \rho_\star^{1/2}}{m\sqrt{N}}.$$

Together with (7.67) and Lemma 7.12 (identifying the subspace $\langle v_j^0\rangle^\perp$ with $\mathbb{C}^{m-1}$) this implies

$$\operatorname{dist}(w_j, \langle v_j^0\rangle) = \|\widetilde{w}_j\|_2 < \frac{m(m-1)\delta N}{\kappa\rho_\star} \lesssim_{\mathfrak{p},\alpha} \delta N^{(n_0+3)/2}/\rho_\star \lesssim \rho_\star/\sqrt{N}$$

using our assumption on $\rho_\star$ and recalling $m = n_0 + 1 = O_{\mathfrak{p}}(1)$. Now from the estimates

$$\|w_j\|_2 = O_{\mathfrak{p}}(1)\,, \qquad \|v_j^0\|_2 \ge \frac{\rho_\star^{1/2}}{\sqrt{N}}\,, \qquad \|v_j^0 - \widehat{v}_j^0\|_2 \le \rho_1 = \rho_\star^{3/2}/N,$$

the identity

$$\operatorname{dist}(w_j, \langle v_j^0\rangle) = \frac{\|w_j \wedge v_j^0\|_2}{\|v_j^0\|_2}$$

and multilinearity of the wedge product, we have

$$\operatorname{dist}(w_j, \langle \widehat{v}_j^0\rangle) = \operatorname{dist}(w_j, \langle v_j^0\rangle) + O_{\mathfrak{p},\alpha}(\rho_\star/\sqrt{N}) \lesssim_{\mathfrak{p},\alpha} \rho_\star/\sqrt{N}.$$

Since $\|w_j\|_2 = O_{\mathfrak{p}}(1)$, by taking R sufficiently large we have that for every $j \in J$ there exists $\widehat{w}_j \in \Lambda_0 \cdot \widehat{v}_j^0$ such that

$$\|w_j - \widehat{w}_j\|_2 \lesssim_{\mathfrak{p},\alpha} \rho_\star/\sqrt{N}. \tag{7.68}$$

Taking $\widehat{v}_j^{k_0}$ as in (7.62) for each $j \in J$ (having by now fixed all vectors on the right hand side), we have by (6.6) and the triangle and AM-GM inequalities

$$\|v_j^{k_0} - \widehat{v}_j^{k_0}\|_2^2 \lesssim_{\mathfrak{p}} \|w_j - \widehat{w}_j\|_2^2 + \sum_{k \in [n_0]\setminus\{k_0\}} \|v_j^k - \widehat{v}_j^k\|_2^2.$$

Substituting (7.68), summing over $j \in J$, and applying the bounds $\|U^k - \widehat{U}^k\|_{\mathrm{HS}} \lesssim_{\mathfrak{p}} \rho_\star$ for $k \ne k_0$ yields (7.64) and completes the proof. □

Remark 7.14. Let us explain why we cannot take any linearization for $\mathfrak{p}$. Consider a linearization of the form

$$\boldsymbol{L}^z = \begin{pmatrix} -z + T & R_1 & \dots & R_{n_0} \\ Y_1 & -\mathrm{I} & \dots & 0 \\ \vdots & & \ddots & \\ Y_{n_0} & \cdots & & -\mathrm{I} \end{pmatrix}$$

(compare (4.6)) where the R_k, Y_ℓ and T are linear forms in the matrices $X_1, \dots, X_n$ with respective coefficient vectors $r_k, s_\ell, t \in \mathbb{C}^n$. If the family $\{r_1, ..., r_{n_0}\}$ is not of full rank, there exists a non-null vector $x \in \mathbb{C}^{n_0}$ such that $\sum_{i=1}^{n_0} x_i r_i = 0$, then if $\tilde{x} = (0, x_1, ..., x_{n_0})$, the matrices L_i as in (6.3) (with $\langle r_k, \boldsymbol{\xi}_i\rangle$ in place of ξ_i^k) all satisfy $L_i\tilde{x} = 0$. When we look for the coefficients of the terms of order $n_0 + 1$ of $\det(M + \sum_{i=1}^N U_i^* L_i)$, we find that these are the same as in $\det(\sum_{i=1}^N U_i^* L_i)$, and hence they are zero. On the other hand, if $\{s_1, ..., s_{n_0}\}$ is not of full rank, we then have a non-null vector $x \in \mathbb{C}^{n_0}$ such that $\sum_{k=1}^{n_0} x_k Y_k = 0$ and so $\sum x_k \operatorname{row}_1(Y_k') = 0$. Then considering the vector u with first block $u^0 = 0$ and $u^k(1) = x_k$ for $k = 1, ..., n_0$, $u^k(i) = 0$ for all k and all $i \in [2, N]$, we have that u is a non-null element of $(\mathsf{V}_{(1)}^z)^\perp$, so that $\sigma_{min}(U^0) = 0$. And so we cannot have our lower bound on the effective rank $\boldsymbol{R}(U^0)$ (as in (7.5)) and our anti-concentration bound does not beat the cardinality of the net. The problem of finding a "nice" linearization for $\mathfrak{p}$ is also why the question of polynomials of degrees higher than 2 is more difficult. For instance, it is not clear that a homogeneous polynomial of degree 3 admits a linearization of the form $\mathfrak{p} = \sum_{i=1}^{n_0} R_i S_i T_i$ where the families $\{R_i\}_{1\le i\le n_0}$ and $\{T_i\}_{1\le i\le n_0}$ are of full rank.

8. Proof of Theorem 1.2

The proof of Theorem 1.2 follows Girko's idea [27] based on Green's formula (1.6). For $z \in \mathbb{C}$, we set

$$\nu_N^z := \mu_{(z\,\mathrm{I}_N - P^N)(z\,\mathrm{I}_N - P^N)^*},$$

where we recall that $P^N = \mathfrak{p}(X_1^N, \dots, X_n^N)$, as well as our notation (1.1). Applying the above formula to the eigenvalues of P^N, we deduce that

$$\int_{\mathbb{C}} \psi(z) d\mu_{P^N}(z) = \frac{1}{4\pi} \int_{\mathbb{C}} \Delta\psi(z) \int_0^\infty \log(x) d\nu_N^z(x) dz\,. \tag{8.1}$$

The proof of the convergence of the right hand side is broken into the following steps.

(1) For all $z \in \mathbb{C}$, we show that ν_N^z converges weakly almost surely to some identifiable probability measure ν^z on $\mathbb{R}_+$.

(2) Using our lower bound in Theorem 1.3, we show that for almost every $z \in \mathbb{C}$, $\int_{\mathbb{R}_+} \log(x) d\nu_N^z(x)$ converges to $\int_{\mathbb{R}_+} \log(x) d\nu^z(x)$ in probability.

(3) We show that, in probability, $z \mapsto \int_{\mathbb{R}_+} \log(x) d\nu_N^z(x)$ converges in L^1 to $z \mapsto \int_{\mathbb{R}_+} \log(x) d\nu^z(x)$, and therefore μ_{P^N} converges in distribution to a limit that we identify as $\nu_{\mathfrak{p}(c_1,\dots,c_n)}$.

Proof of (1): Because $(z\,\mathrm{I}_N - P^N)(z\,\mathrm{I}_N - P^N)^*$ is a self-adjoint polynomial in independent Ginibre matrices and their adjoints, the first point is a direct consequence of Voiculescu's theorem [56], see [4] for a review. In particular we have that ν^z is the distribution of $|z - \mathfrak{p}(c_1, \dots, c_n)|^2$ in the sense of $*$-moments, and from the boundedness of the circular elements it follows that ν^z is compactly supported for every fixed $z \in \mathbb{C}$.

Proof of (2): To prove the second point, we need to deal with the fact that the logarithm is unbounded. To this end, first observe that by Lemma 3.2, there is a constant $B > 0$ such that

$$\limsup_{N\to\infty} \|P^N\| \le B \qquad a.s. \tag{8.2}$$

Therefore, fixing $\epsilon > 0$, the first point implies that for any smooth nonnegative function χ_ϵ which vanishes on $[0, \epsilon/2]$ and equals one on $[\epsilon, \infty)$,

$$\lim_{N\to\infty} \int_0^B \chi_\epsilon(x) \log(x) d\nu_N^z(x) = \int_0^B \chi_\epsilon(x) \log(x) d\nu^z(x) \qquad a.s. \tag{8.3}$$

The main point is therefore to show that $\int_0^\epsilon \log(x) d\nu_N^z(x)$ is negligible. We will show that for every $\delta, \delta' > 0$ there is $\epsilon \in (0,1)$ such that

$$\limsup_{N\to\infty} \mathbb{P}\Big\{ \Big| \int_0^\epsilon \log x d\nu_N^z(x) \Big| \ge \delta \Big\} \le \delta'.$$

Denoting $\mathcal{G}_N = \{\|P^N\| \le B, \sigma_{\min}(z - P^N) \ge N^{-\beta}\}$, by Theorem 1.3 and (8.2), we can choose β large enough so that $\mathbb{P}(\mathcal{G}_N)$ goes to one. Hence, it suffices to show that

$$\lim_{\epsilon\to 0} \limsup_{N\to\infty} \mathbb{E}\Big(\mathbb{1}_{\mathcal{G}_N} \int_0^\epsilon |\log x| d\nu_N^z(x) \Big) = 0.$$

On the event $\mathcal{G}_N$,

$$\int_0^\epsilon \log x d\nu_N^z(x) = \int_{N^{-\beta}}^\epsilon \log x d\nu_N^z(x),$$

so we only need to show

$$\lim_{\epsilon\to 0} \limsup_{n\to\infty} \mathbb{E}\Big| \int_{N^{-\beta}}^\epsilon \log x d\nu_N^z(x) \Big| = 0. \tag{8.4}$$

We denote $\bar{\nu}_n^z = \mathbb{E}[\nu_N^z]$ and g (resp. g_N) the Stieljes transform of ν^z (resp. $\bar{\nu}_N^z$) given for $\zeta \in \mathbb{C}$ by

$$g^z(\zeta) = \int_0^\infty \frac{1}{\zeta - x} d\nu^z(x), \quad g_N^z(\zeta) = \mathbb{E}[\int_0^\infty \frac{1}{\zeta - x} d\nu_N^z(x)] .$$

The next lemma is the key to prove (8.4).

Lemma 8.1. *Let $z \in \mathbb{C}$ be fixed. There exist $C, N_0 > 0$ and $c_1, c_2 \in (0,1)$ such that for $\eta \in [N^{-c_1}, 1]$ and $N \geq N_0$,*

$$|\Im(g_N^z(i\eta))| \leq C\eta^{-c_2}.$$

We postpone the proof of this lemma to deduce first (8.4). This lemma implies that for $x \in [N^{-c_1}, 1]$,

$$F_N^z(x) := \bar{\nu}_N^z([0,x]) \leq 2\int_{[0,x]} \frac{x^2}{y^2+x^2} d\bar{\nu}_N^z(y) \leq 2x\Im(g_N^z(ix)) \leq 2Cx^{1-c_2}.$$

Hence, we find for $\alpha < c_1$ and $\epsilon < 1$:

$$\begin{aligned}
&\mathbb{E}\Big| \int_{N^{-\beta}}^{\epsilon} \log x d\nu_N^z(x)\Big| \\
&\quad \leq -\int_{N^{-\beta}}^{N^{-\alpha}} \log x d\bar{\nu}_N^z(x) - \int_{N^{-\alpha}}^{\epsilon} \log x d\bar{\nu}_N^z(x) \\
&\quad \leq \beta(\log N)\bar{\nu}_N^z([0,N^{-\alpha}]) + \int_{N^{-\alpha}}^{\epsilon} \frac{1}{x} F_N^z(x)dx - \log(\epsilon)F_N^z(\epsilon) + F_N^z(N^{-\alpha})\log(N^{-\alpha}) \\
&\quad \leq 2C\beta(\log N)N^{-\alpha(1-c_1)} + 2C\frac{\epsilon^{1-c_2}}{1-c_2} - 2(\log\epsilon)\epsilon^{1-c_2},
\end{aligned}$$

which gives (8.4).

Proof of Lemma 8.1. First we observe that Haagerup and Thorbjørnsen [31] proved the convergence of the Stieltjes transform of ν_N^z close to the real axis. Indeed, recall that Ginibre matrices can be decomposed as the sum of two independent GUE matrices: $X_j^N = (Y_j^N + iZ_j^N)/\sqrt{2}$ where $(Y_j, Z_j)_{1\leq j\leq n}$ are independent GUE matrices. Hence $(z-P^N)(z-P^N)^*$ can be seen as a polynomial in independent GUE matrices so that Haagerup and Thorbjørnsen result applies and, see e.g [4, Lemma 5.5.4], implying that there exists c_1 finite such that for $\Im\zeta \in [N^{-c_1}, 1]$ and N large enough,

$$|g^z(\zeta) - g_N^z(\zeta)| \leq \frac{c_2}{N^2(\Im\zeta)^{c_3}}.$$

Up to take a smaller c_1, it is therefore enough to show that

$$\Im g^z(i\epsilon) \leq K\epsilon^q$$

for some $q > -1$ and $K > 0$. Following [46, Corollary 1.2], ν^z has no atoms. Moreover, by [46, Theorem 1.1], g^z is bounded close to the real line except possibly on a discrete set A. Assuming at worst that A contains the origin, the same theorem shows that there exists $q \in \mathbb{Q}$ and a constant $K \neq 0$ such that $g^z(\zeta) \simeq K\zeta^q$ for ζ close to the origin. But clearly, since ν^z has no atoms, $q > -1$. A more quantitative proof could have used that [7] implies that the partition function of ν^z is Hölder with exponent $1/15$. □

Proof of (3): Denoting

$$h_N(z) := \int_0^\infty \log|x| d\nu_N^z(x) \quad \text{and} \quad h(z) = \int_0^\infty \log|x| d\nu^z(x)$$

we have shown that for every fixed $z \in \mathbb{C}$, $h_N(z)$ converges in probability to $h(z)$, the latter now being well defined by steps (1) and (2). In particular we have

$$h(z) = \int_{\mathbb{C}} \log|z-\lambda| d\nu_{\mathfrak{p}(c_1,\dots,c_n)}(\lambda)$$

by definition of the Brown measure. We next prove that for any compact set K, and on the events $\mathcal{A}_N = \{\|P^N\| < B\}$, h_N converges as well in L^1 in the sense that

$$\lim_{N\to\infty} \mathbb{E}\Big(\mathbb{1}_{\mathcal{A}_N} \int_{z\in K} |h_N(z) - h(z)| dz\Big) = 0. \tag{8.5}$$

This is enough to conclude the proof of Theorem 1.2 by (8.1) for any twice continuously differentiable function ψ with compact support. The last condition is finally removed since the eigenvalues are almost surely bounded by B according to (8.2). To prove (8.5), it is enough to notice that h_N and h belong to L^2 in the sense that

$$\mathbb{E}[1_{\mathcal{A}_N} \int_K |h_N(z)|^2 dz] + \int_K |h(z)|^2 dz$$

is bounded independently of N, so that the bounded convergence theorem applies. But, Jensen's inequality and Fubini's theorem imply that

$$\mathbb{E}[1_{\mathcal{A}_N} \int_K |h_N(z)|^2 dz] \le \mathbb{E}[1_{\mathcal{A}_N} \int_{\mathbb{C}} \int_K |\log|z-\lambda||^2 dz d\mu_{P^N}(\lambda)] \le \sup_{|\lambda|\le B} \int_K |\log|z-\lambda||^2 dz$$

is finite, and a similar estimate holds for $h(z)$ (one obtains from the boundedness of the circular elements $c_1,\dots,c_n$ that $\nu_{\mathfrak{p}(c_1,\dots,c_n)}$ has compact support). Therefore, in probability h_N converge to h in $L^1(K, \mathrm{Leb})$ and so $\mu_{P^N} = \frac{1}{4\pi}\Delta h_N$ converges to $\nu_{\mathfrak{p}(c_1,\dots,c_m)} = \frac{1}{4\pi}\Delta h$ in the sense of distributions on K. Again, taking K that contains the support of $\nu_{\mathfrak{p}(c_1,\dots,c_m)}$ and μ_{P^N}, the convergence in the sense of distributions implies weak convergence and the result is proved.

References

[1] J. Alt, L. Erdős, and T. Krüger, *Local inhomogeneous circular law*, Ann. App. Probab., 28 (2018), pp. 148–203.

[2] J. Alt and T. Krüger, *Inhomogeneous circular law for correlated matrices*. Preprint, arXiv:2005.13533.

[3] G. W. Anderson, *A local limit law for the empirical spectral distribution of the anticommutator of independent Wigner matrices*, Ann. Inst. Henri Poincaré Probab. Stat., 51 (2015), pp. 809–841.

[4] G. W. Anderson, A. Guionnet, and O. Zeitouni, *An introduction to random matrices*, vol. 118 of Cambridge Studies in Advanced Mathematics, Cambridge University Press, Cambridge, 2010.

[5] J. Banks, A. Kulkarni, S. Mukherjee, and N. Srivastava, *Gaussian regularization of the pseudospectrum and davies' conjecture*. Preprint, arXiv:1906.11819.

[6] J. Banks, J. G. Vargas, A. Kulkarni, and N. Srivastava, *Pseudospectral shattering, the sign function, and diagonalization in nearly matrix multiplication time*. Preprint, arXiv:1912.08805.

[7] M. Banna and T. Mai, *Hölder continuity of cumulative distribution functions for noncommutative polynomials under finite free Fisher information*, J. Funct. Anal., 279 (2020), p. 108710.

[8] A. Basak, N. Cook, and O. Zeitouni, *Circular law for the sum of random permutation matrices*, Electron. J. Probab., 23 (2018), pp. Paper No. 33, 51.

[9] A. Basak and A. Dembo, *Limiting spectral distribution of sums of unitary and orthogonal matrices*, Electron. Commun. Probab., 18 (2013), pp. no. 69, 19.

[10] A. Basak, E. Paquette, and O. Zeitouni, *Spectrum of random perturbations of Toeplitz matrices with finite symbols*, Trans. Amer. Math. Soc., 373 (2020), pp. 4999–5023.

[11] S. T. Belinschi, T. Mai, and R. Speicher, *Analytic subordination theory of operator-valued free additive convolution and the solution of a general random matrix problem*, J. Reine Angew. Math., 732 (2017), pp. 21–53.

[12] S. T. Belinschi, P. Śniady, and R. Speicher, *Eigenvalues of non-hermitian random matrices and brown measure of non-normal operators: Hermitian reduction and linearization method*, Linear Algebra Appl., 537 (2018), pp. 48– 83.

[13] P. Biane and F. Lehner, *Computation of some examples of Brown's spectral measure in free probability*, Colloq. Math., 90 (2001), pp. 181–211.

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

[15] L. G. Brown, *Lidskiĭ's theorem in the type* II *case*, in Geometric methods in operator algebras (Kyoto, 1983), vol. 123 of Pitman Res. Notes Math. Ser., Longman Sci. Tech., Harlow, 1986, pp. 1–35.

[16] M. Campos, M. Jenssen, M. Michelen, and J. Sahasrabudhe, *The singularity probability of a random symmetric matrix is exponentially small*. Preprint, arXiv:2105.11384.

[17] M. Campos, L. Mattos, R. Morris, and N. Morrison, *On the singularity of random symmetric matrices*. Preprint, arXiv:1904.11478, 2019.

[18] A. Carbery and J. Wright, *Distributional and L_q norm inequalities for polynomials over convex bodies in $\mathbb{R}^n$*, Math. Res. Lett., 8 (2001), pp. 233–248.

[19] N. Cook, W. Hachem, J. Najim, and D. Renfrew, *Non-Hermitian random matrices with a variance profile (I): deterministic equivalents and limiting ESDs*, Electron. J. Probab., 23 (2018), pp. Paper No. 110, 61.

[20] K. P. Costello, T. Tao, and V. Vu, *Random symmetric matrices are almost surely nonsingular*, Duke Math. J., 135 (2006), pp. 395–413.

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

[22] L. Erdős, T. Krüger, and Y. Nemish, *Local laws for polynomials of Wigner matrices*, J. Funct. Anal., 278 (2020), pp. 108507, 59.

[23] J. Feinberg and A. Zee, *Non-Gaussian non-Hermitian random matrix theory: phase transition and addition formalism*, Nuclear Phys. B, 501 (1997), pp. 643–669.

[24] A. Ferber and V. Jain, *Singularity of random symmetric matrices—a combinatorial approach to improved bounds*, Forum Math. Sigma, 7 (2019), pp. Paper No. e22, 29.

[25] J. Fox, M. Kwan, and L. Sauermann, *Combinatorial anti-concentration inequalities, with applications*. Preprint, arXiv:1905.12142.

[26] O. Friedland and S. Sodin, *Bounds on the concentration function in terms of the Diophantine approximation*, C. R. Math. Acad. Sci. Paris, 345 (2007), pp. 513–518.

[27] V. L. Girko, *Circular law*, Theory Probab. Appl., 29 (1984), pp. 694–706.

[28] F. Götze and A. Tikhomirov, *The circular law for random matrices*, Ann. Probab., 38 (2010), pp. 1444–1491.

[29] A. Guionnet, M. Krishnapur, and O. Zeitouni, *The single ring theorem.*, Ann. Math. (2), 174 (2011), pp. 1189–1217.

[30] A. Guionnet, P. M. Wood, and O. Zeitouni, *Convergence of the spectral measure of non-normal matrices*, Proc. Amer. Math. Soc., 142 (2014), pp. 667–679.

[31] U. Haagerup and S. Thorbjørnsen, *A new application of random matrices:* $\mathrm{Ext}(C^*_{\mathrm{red}}(F_2))$ *is not a group*, Ann. of Math. (2), 162 (2005), pp. 711–775.

[32] H. Kösters and A. Tikhomirov, *Limiting spectral distributions of sums of products of non-Hermitian random matrices*, Probab. Math. Statist., 38 (2018), pp. 359–384. [On table of contents: Vol. 33 (2013)].

[33] J. E. Littlewood and A. C. Offord, *On the number of real roots of a random algebraic equation. III*, Rec. Math. [Mat. Sbornik] N.S., 12(54) (1943), pp. 277–286.

[34] L. Lovász, *Singular spaces of matrices and their application in combinatorics*, Bol. Soc. Brasil. Mat. (N.S.), 20 (1989), pp. 87–99.

[35] V. A. Marčenko and L. A. Pastur, *Distribution of eigenvalues in certain sets of random matrices*, Math. USSR Sb., 1 (1967), pp. 457–483. English translation of Mat. Sbornik 72 507–536.

[36] R. Meka, O. Nguyen, and V. Vu, *Anti-concentration for polynomials of independent random variables*, Theory Comput., 12 (2016), pp. Paper No. 11, 16.

[37] J. A. Mingo and R. Speicher, *Free probability and random matrices*, vol. 35, The Fields Institute for Research in the Mathematical Sciences, New York, NY, 2017.

[38] H. H. Nguyen, *Inverse Littlewood–Offord problems and the singularity of random symmetric matrices*, Duke Math. J., 161 (2012), pp. 545–586.

[39] H. H. Nguyen and S. O'Rourke, *The elliptic law*, Int. Math. Res. Not. IMRN, (2015), pp. 7620–7689.

[40] ———, *On the concentration of random multilinear forms and the universality of random block matrices*, Probab. Theory Related Fields, 162 (2015), pp. 97–154.

[41] H. H. Nguyen and V. H. Vu, *Small ball probability, inverse theorems, and applications*, in Erdös centennial, vol. 25 of Bolyai Soc. Math. Stud., János Bolyai Math. Soc., Budapest, 2013, pp. 409–463.

[42] S. O'Rourke, D. Renfrew, A. Soshnikov, and V. Vu, *Products of independent elliptic random matrices*, J. Stat. Phys., 160 (2015), pp. 89–119.

[43] S. O'Rourke and A. Soshnikov, *Products of independent non-Hermitian random matrices*, Electron. J. Probab., 16 (2011), pp. no. 81, 2219–2245.

[44] M. Rudelson and R. Vershynin, *The Littlewood-Offord problem and invertibility of random matrices*, Adv. Math., 218 (2008), pp. 600–633.

[45] ———, *Smallest singular value of a random rectangular matrix*, Comm. Pure Appl. Math., 62 (2009), pp. 1707–1739.

[46] D. Shlyakhtenko and P. Skoufranis, *Freely independent random variables with non-atomic distributions*, Trans. Amer. Math. Soc., 367 (2013), pp. 6267–6291.

[47] P. Śniady, *Random regularization of Brown spectral measure*, J. Funct. Anal., 193 (2002), pp. 291–313.

[48] M. Strzelecka, M. Strzelecki, and T. Tkocz, *On the convex infimum convolution inequality with optimal cost function*, Alea, 14 (2017), pp. 903–915.

[49] T. Tao, V. Vu, and M. Krishnapur, *Random matrices: Universality of ESDs and the circular law*, Ann. Probab., 38 (2010), pp. 2023–2065.

[50] T. Tao and V. H. Vu, *Random matrices: the circular law*, Commun. Contemp. Math., 10 (2008), pp. 261–307.

[51] ———, *Inverse Littlewood-Offord theorems and the condition number of random discrete matrices*, Ann. of Math. (2), 169 (2009), pp. 595–632.

[52] A. N. Tikhomirov, *On the asymptotics of the spectrum of the product of two rectangular random matrices*, Sibirsk. Mat. Zh., 52 (2011), pp. 936–954.

[53] L. N. Trefethen and . Embree, Mark, *Spectra and pseudospectra: the behavior of nonnormal matrices and operators*, Princeton University Press, Princeton, N.J, 2005.

[54] R. Vershynin, *Invertibility of symmetric random matrices*, Random Structures Algorithms, 44 (2014), pp. 135–182.

[55] M. Vogel and O. Zeitouni, *Deterministic equivalence for noisy perturbations*. Preprint, arXiv:2001.09024.

[56] D. Voiculescu, *Limit laws for random matrices and free products*, Invent. Math., 104 (1991), pp. 201–220.

[57] E. P. Wigner, *On the distribution of the roots of certain symmetric matrices*, Annals Math., 67 (1958), pp. 325–327.

[58] P. M. Wood, *Universality of the ESD for a fixed matrix plus small random noise: a stability approach*, Ann. Inst. Henri Poincaré Probab. Stat., 52 (2016), pp. 1877–1896.

Duke University, Durham, NC 27708, USA
Email address: nickcook@math.duke.edu

Université de Lyon, ENSL, CNRS, France
Email address: Alice.Guionnet@ens-lyon.fr

Université de Lyon, ENSL, France
Email address: Jonathan.Husson@ens-lyon.fr