SPECTRUM AND PSEUDOSPECTRUM FOR QUADRATIC POLYNOMIALS IN GINIBRE MATRICES
Nicholas A. Cook, Alice Guionnet, Jonathan Husson

To cite this version:

HAL Id: hal-03865667
https://hal.science/hal-03865667
Submitted on 22 Nov 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
spectral and pseudospectrum for quadratic polynomials in ginibre matrices

nicholas cook∗, alice guionnet†, and jonathan husson

abstract. for a fixed quadratic polynomial \( p \) 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 \( P^N = p(X_1^N, \ldots, X_n^N) \) to the brown measure of \( p \) 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 \( p \) 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 \( P^N = p(X_1^N, \ldots, X_n^N) \) vers la mesure de brown de \( p \) évaluée en \( n \) éléments circulaires librement indépendants \( c_1, \ldots, c_n \) 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 algébriko-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

\[
\mu_A := \frac{1}{N} \sum_{j=1}^{N} \delta_{\lambda_j(A)}.
\]

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) \( \mu_{A^N} \) converges in probability to some deterministic probability measure \( \mu \). 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.

---

∗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 $\mu_{\sqrt{N}X^N}$ converges to Girko’s circular law $\mu_{\text{circ}}$, 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 $\xi_{ij}, \xi_{ji}$, 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 $W_1^N, \ldots, W_n^N$ are independent $N \times N$ Wigner matrices and $p$ is a non-commutative self-adjoint polynomial, then the ESD of $p(W_1^N, \ldots, W_n^N)$ converges towards the spectral distribution of $(s_1, \ldots, s_n)$, where $s_1, \ldots, s_n$ 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
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 \cite{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 \cite{47,30,58,10,55} and references therein. This regularizing effect has been exploited for applications in numerical analysis in the recent works \cite{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) $\mu_A$ 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 $\mu_{A^N}$ to the measure $\mu_a$ 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 $\nu_a$ (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 $\varepsilon$-pseudospectrum of a square matrix $A$ is the set
\[
\Lambda_\varepsilon(A) = \{z \in \mathbb{C} : \sigma_{\text{min}}(A - z) \leq \varepsilon\} \tag{1.3}
\]
where $\sigma_{\text{min}}(A) = \sqrt{\lambda_{\text{min}}(AA^*)}$ 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 $\varepsilon$ (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 $\mu_{A^N}$ 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 $\sup_{z \in \mathbb{C}} \mathbb{P}\{\sigma_{\text{min}}(A - z) \leq \varepsilon\}$ 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 $\mathbb{C}(x_1, \ldots, x_n)$ 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 
$(\sqrt{N} \Re (X_{ij}))_{i,j \in [N]}$ and $(\sqrt{N} \Im (X_{ij}))_{i,j \in [N]}$ are independent i.i.d. families of random variables of law $N(0, 1/2)$.

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

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

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

weakly in probability, where $c_1, \ldots, c_n$ are *-free circular elements of a $W^*$-probability space and $\nu_{p(c_1, \ldots, c_n)}$ is the Brown measure of $p(c_1, \ldots, c_n)$ (defined in Section 1.3 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 $p \in \mathbb{C}(x_1, \ldots, x_n)$ be a non-commutative polynomial of degree two. There is an absolute constant $C_0 > 0$ and $C(p), c(p) > 0$ depending only on $p$ such that the following holds. Let $X_1, \ldots, X_n$ be $n$ independent $N \times N$ complex Ginibre matrices and set $P = p(X_1, \ldots, X_n)$. For any $z \in \mathbb{C}$ and any $\varepsilon > 0$,

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

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 $c(p) = 1/(n_0 + 1)$, where $n_0$ is the rank of the quadratic form associated to the homogeneous degree-two part of $p$ (see Lemma 4.1). The constant $C$ depends on $p$ only through the rank parameter $n_0$ and the size of the coefficients, quantified by the norm of the matrix $(s_0, \ldots, s_{n_0})$ 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

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

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

$$
\text{Leb}(\Lambda_{N^{-c}}(P) \cap \Omega) \leq N^{-K} \text{Leb}(\Omega).
$$

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 $w_0 \in \mathbb{C}$, $u = (u_1, \ldots, u_N) \in \mathbb{C}^N$ and i.i.d. copies $\xi_1, \ldots, \xi_N$ of a random variable $\xi \in \mathbb{C}$, the problem is to bound the probability that the scalar random variable

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

(1.4)
lies in a small neighborhood of the origin. When $\xi$ 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 \[20\] \[33\] \[54\], 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 $\xi$ 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 $L$ with dependent entries, for some bounded $d$ determined by $p$, with coefficients given by linear combinations of the Ginibre matrices. The matrix $L$ 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 $W_0, U_1, \ldots, U_N \in \mathbb{C}^{d \times d}$ and i.i.d. copies $L_j$ of a $d \times d$ random matrix $L$, bound the probability that the matrix random walk

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

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

For Theorem \[1.3\] we specifically seek a bound of the form $N^{C} e^c + e^{-N}$ on the probability that $W$ is within distance $\varepsilon$ of $S_d$ (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 $\mathbb{P}(W \in S_d) = 1$. We also have $\mathbb{P}(W \in S_d) = 1$ in the “simple” case that $L = \xi I_d$ is a random scalar multiple of the identity and all of the matrices $W_0, U_1, \ldots, U_N$ lie in a linear subspace contained in $S_d$. 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 $L = \xi I_d$ 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 $S_d$ we must make use of the geometry of the sequence $U_1, \ldots, U_N$.

For the smallest singular value problem, the matrices $L_j$ are taken from an $Nd \times d$ column of $L$, and the sequence of matrices $U_1, \ldots, U_N$ 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 $U_1, \ldots, U_N$ 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 $p$ (indeed, the dependency structure and support of the step variable $L$ is determined by $p$).

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 \((A, \tau)\) is a von Neumann algebra \(A\) equipped with a tracial faithful normal state \(\tau\). According to Voiculescu’s central result [50], the non-commutative distribution of \((X_1^N, \ldots, X_N^N)\) converges in \(\ast\)-moments towards the non-commutative distribution of \(n\) free circular elements \((c_1, \ldots, c_n)\) in \(A\) in the sense that for any polynomial \(p \in \mathbb{C}(x_1, \ldots, x_n, x_1^*, \ldots, x_n^*)\)

\[
\lim_{N \to \infty} \frac{1}{N} \text{Tr} \left( p(X_1^N, \ldots, X_n^N, (X_1^N)^*, \ldots, (X_n^N)^*) \right) = \tau (p(c_1, \ldots, c_n, c_1^*, \ldots, c_n^*)) \quad \text{a.s.}
\]

The right hand side is a linear map on the set of polynomials in \(\mathbb{C}(x_1, \ldots, x_n, x_1^*, \ldots, x_n^*)\) which is uniquely defined by its value on monomials given, for any choices of \(i_1, \ldots, i_k \in [1, n]\) and any \(\varepsilon_1, \ldots, \varepsilon_k \in \{1, \ast\}\), 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 \(\{i_1, \ldots, i_k\}\) so that each block \(b = (b_1, b_2)\) is such that \(i_{b_1} = i_{b_2}\) and \((\varepsilon_{b_1}, \varepsilon_{b_2}) = (1, \ast)\) or \((\ast, 1)\).

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 \(f\)

\[
\lim_{N \to \infty} \frac{1}{N} \text{Tr} \left( f \left( p(X_1^N, \ldots, X_n^N, (X_1^N)^*, \ldots, (X_n^N)^*) \right) \right) = \tau (f (p(c_1, \ldots, c_n, c_1^*, \ldots, c_n^*))) \quad \text{a.s.}
\]

If \(p\) is self-adjoint, this guarantees the convergence of the empirical spectral distribution of \(p(X_1^N, \ldots, X_n^N, (X_1^N)^*, \ldots, (X_n^N)^*)\) towards the distribution of \(p(c_1, \ldots, c_n, c_1^*, \ldots, c_n^*)\) which is specified by its moments as given above. If \(p\) 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 \(\psi\), any complex numbers \(\lambda_i, 1 \leq i \leq N,\)

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

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

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

Noting that \(|z - Q|^2\) is a self-adjoint polynomial in \(X_1^N, \ldots, X_n^N, (X_1^N)^*, \ldots, (X_n^N)^*\), 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 \(\psi\) on \(\mathbb{C}\) by

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

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 \(c_1, \ldots, c_n\) 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 $j$th column, which we will denote by $\text{col}_j(X)$, is close to the span $V_{(j)}$ 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 $e_2, \ldots, e_N$, and the distance is simply the magnitude of the first coordinate of $\text{col}_j(X)$. 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 $p$.

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 $(L^z)^{-1}$, where $L^z$ is the $3N \times 3N$ block matrix

$$L^z = \begin{pmatrix} -z & X_1 & X_2 \\ X_2 & -I & 0 \\ X_1 & 0 & -I \end{pmatrix}.$$ (2.1)

In particular we have the deterministic bound

$$\|(P - z)^{-1}\|_{\text{op}} \leq \|(L^z)^{-1}\|_{\text{op}}$$ (2.2)

and so it suffices to show

$$\mathbb{P}\{\sigma_{\text{min}}(L^z) \leq \varepsilon\} \lesssim N^C \varepsilon^c + e^{-N}.$$ (2.3)
The advantage of this new problem is that $L^z$ can be viewed as an $N \times N$ matrix with entries that are independent $3 \times 3$ random matrices:

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

(To avoid ambiguity in the proof we will actually write $\hat{L}^z_x$ for the element of $M_N(M_3(\mathbb{C}))$ associated to $P$, whereas we write $L^z$ for the associated element of $M_3(M_N(\mathbb{C}))$ 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 $\hat{L}^z_x$ for $3 \times 3$ matrix obtained by projecting the three columns of $\text{col}_j(L^z) =: (L_{i,j}^z)_i^N \in M_3(\mathbb{C})^N \cong M_3(M_N(\mathbb{C}))$ to the orthogonal complement of the span $V_{(j)}^z$ of the remaining $3N - 3$ columns. Specifically, we let $U_j \in M_{3N,3}(\mathbb{C})$ have columns that are orthonormal in $(V_{(j)})^\perp$, and put

$$\hat{L}^z_x = U_j^* \text{col}_j(L^z) = \sum_{i=1}^N U_{i,j}^* L_{i,j}^z \quad (2.5)$$

where $U_{i,j} \in M_3(\mathbb{C})$ are the corresponding $3 \times 3$ blocks of $U_j$. (We remark that, *a posteriori*, with high probability $(V_{(j)})^\perp$ has dimension 3 and the columns of $U_j$ are in fact a basis.) The key is that $\text{col}_j(L^z)$ is independent of $V_{(j)}^z$, and hence $U_j$ can be chosen independently of $\text{col}_j(L^z)$.

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

$$\varepsilon \geq \|L^z v\|_2 = \left\| \sum_{j=1}^N \text{col}_j(L^z) v_j \right\|_2.$$

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

$$\varepsilon \geq \left\| \sum_{j=1}^N U_{j_0}^* \text{col}_j(L^z) v_j \right\|_2 = \|U_{j_0}^* \text{col}_{j_0}(L^z) v_{j_0}\|_2 = \|\hat{L}_{j_0}^z v_{j_0}\|_2 \geq N^{-1/2} \sigma_{\min}(\hat{L}_{j_0}^z).$$

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

$$\mathbb{P}\{\sigma_{\min}(\hat{L}^z_x) \leq \varepsilon\} \lesssim N^C \varepsilon e^e + e^{-N} \quad (2.6)$$

for an adjusted constant $C$.

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

$$\mathbb{P}\{|\det(\hat{L}^z_x)| \leq \varepsilon\} \lesssim N^C \varepsilon e^e + e^{-N}. \quad (2.7)$$

We condition on an arbitrary realization of $U_1$ (recall it is independent of $\text{col}_1(L^z)$). From (2.5) we can view the random matrix $\hat{L}^z_x$ as a random walk in $M_3(\mathbb{C})$ with independent steps $U_{i,j}^* L_{i,j}^z$, 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 $\mathbb{P}\{|u^* \text{col}_1(X)| \leq \varepsilon\}$ 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 $U_1$: consider for
Lemma 6.2) gives an anticoncentration bound of the form

\[ \text{det}(U_{i,1}) \leq N^{-O(1)} \varepsilon^{1/3} \]

where we view the determinant as a degree-3 polynomial \( p \) of all possible structured orthonormal bases as soon as we can find a monomial of \( p \) that satisfies (2.10) we aim to construct an \( \varepsilon \)-net \( N \) for the set of all possible structured orthonormal bases \( U = (U_1, \ldots, U_N) \in M_{3}(\mathbb{C})^N \cong M_{3N,3}(\mathbb{C}) \). Since \( U_1^* \text{col}_j(L^2) = 0_{3 \times 3} \) for every \( 2 \leq j \leq N \), then if \( U_1 \) is approximated within \( \varepsilon \) (in the Hilbert–Schmidt metric) by an element \( \hat{U} \) of the net, we have \( \|\hat{U}^* \text{col}_j(L^2)\|_{\text{HS}} = O(\varepsilon) \) for all \( 2 \leq j \leq N \).
(here we use that $L^z$ has bounded operator norm with high probability). Viewing the $3 \times 3$ random matrix $\tilde{U}^* \text{col}_j(L^z)$ as a random vector in $\mathbb{C}^9$, after a bit of algebra one sees

$$\|\tilde{U}^* \text{col}_j(L^z)\|_{\text{HS}} = \|w_0 + W^*\xi\|_2$$

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

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

is a Gaussian vector and

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

(2.11)

with $\tilde{U}^0, \tilde{U}^1, \tilde{U}^2 \in \mathcal{M}_{N,3}(\mathbb{C})$ denoting the blocks of $\tilde{U}$ under the partitioning of coordinates as in (2.1). Thus, the small ball probability for the matrix random walk $\tilde{U}^* \text{col}_j(L^z) \in \mathcal{M}_3(\mathbb{C})$ is controlled by that of the Gaussian vector $W^*\xi \in \mathbb{C}^9$, which in turn is determined by the effective rank $r$ of $W$ (one can take $r$ to be the number of singular values of $W$ 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 $U$. Supposing we can find a net of size $|\mathcal{N}| = O(1/\varepsilon^2)^dN$, where $0 \leq d \leq 9$ is the effective metric dimension of Struct (which is a subset of the 9-N-dimensional complex vector space $\mathcal{M}_{3N,3}(\mathbb{C})$, so that this estimate holds trivially for $d = 9$) we can then bound the probability that $U_1$ is structured by

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

We thus hope to show that the effective rank $r$ of matrices $W$ 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 $\tilde{U}$, 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 $U$ – 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 $p, q, \ldots$, we indicate this by writing $C = C(p, q, \ldots)$. 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 $|f| \leq Cg$ 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) $g$, we write $f = O_q(g)$ to mean $|f| \leq Cg$ for some $C = C(q) \in (0, +\infty)$, and similarly for $f \lesssim_q g, f \gtrsim_q g$.

3.2. Matrices. For integers $n \geq m \geq 1$ we use the common abbreviations $[m,n] := \{m, \ldots, n\}$ 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 $\mathcal{M}_{n,m}(\mathcal{X})$ for the set of $n \times m$ matrices with entries in a set $\mathcal{X}$, and abbreviate $\mathcal{M}_n(\mathcal{X}) := \mathcal{M}_{n,n}(\mathcal{X})$. The $n \times n$ identity matrix is denoted $\mathbb{I}_n$, with the subscript sometimes omitted if it is clear from the context. For $a \in \mathbb{C}$ we often write $a$ for $a \mathbb{I}_n$ 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 $M^T_S(\mathcal{X})$ for the set of $|S| \times |T|$ matrices with entries in $\mathcal{X}$ and rows and columns indexed by $S$ and $T$, respectively. Thus, $M_{n,m}(\mathcal{X}) = M_{[n]}^{[m]}(\mathcal{X})$. We write $M_S(\mathcal{X}) := M^S_S(\mathcal{X})$ when there can be no confusion (we avoid this when $S$ is itself a product set). For $A \in M^T_S(\mathcal{X})$ and $(s,t) \in S \times T$ we write $A(s,t)$ for the $(s,t)$ entry of $A$. A block matrix $\mathbf{A}$ that is an $n \times n'$ array of blocks $A^{k,\ell} \in M_{N,N'}(\mathcal{X})$ is viewed as an element of $M^{[n'] \times [N']}_{[n] \times [N]}(\mathcal{X})$ and is naturally associated with an element of $M_{n,n'}(M_{N,N'}(\mathcal{X}))$; we abusively write $\mathbf{A}$ for both. Thus,

$$\mathbf{A}(k,\ell) = A^{k,\ell}, \quad A((k,i), (\ell,j)) = A^{k,\ell}(i,j)$$

(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, $\mathbf{A}$ is naturally associated to an element

$$\widetilde{\mathbf{A}} \in M^{[N'] \times [n']}_{[N] \times [n]}(\mathcal{X}) \cong M_{N,N'}(M_{n,n'}(\mathcal{X}))$$

(3.2)

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

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 $p$). In this case we always use boldface for the block matrix $\mathbf{A}$, with its large submatrices $A^{k,\ell}$ indexed with superscripts and its small submatrices indexed with subscripts:

$$\widetilde{\mathbf{A}}(i,j) = A_{i,j} , \quad (i,j) \in [N] \times [N'] .$$

(3.3)

Thus

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

(3.4)

We always use indices $i, j, i_0, t_1$, etc. for the large dimensions and $k, \ell$, etc. for the small dimensions.

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

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

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

$$\text{col}_t(A) = (A^{1,\ell}, \ldots, A^{n,\ell}) \in M_{N,N'}(\mathbb{C})^n , \quad \ell \in [n']$$

$$\text{col}_j(\widetilde{\mathbf{A}}) = (A^{1,j}, \ldots, A^{n,j}) \in M_{n,n'}(\mathbb{C})^N , \quad \ell \in [N']$$

are vectors of matrices, whereas

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

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

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

For $A \in M^T_S(\mathcal{A})$ and $B \in M^U_T(\mathcal{A})$ with $\mathbf{A}$ a $*$-algebra we define the product

$$AB = (\text{row}_s(A) \cdot \text{col}_u(B))_{s \in S, u \in U} = \left( \sum_{t \in T} A(s,t)B(t,u) \right)_{s \in S, u \in U} \in M^U_T(\mathcal{A})$$

and the conjugate transpose

$$\mathbf{A}^* = (A(t,s)^*)_{t \in T, s \in S} \in M^S_T(\mathcal{A})$$

in the usual way. For $A \in M^T_S(\mathcal{A})$ and $B \in M^V_U(\mathcal{A})$ we denote the tensor product $A \otimes B \in M^T_S \otimes M^V_U(\mathcal{A})$ 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 \( \ell_2 \)-norm \( \| \cdot \|_2 \), and \( \mathbb{M}^T_S(\mathbb{C}) \) with the Hilbert–Schmidt norm
\[
\|A\|_{\text{HS}} = \left( \sum_{s \in S, t \in T} |A(s, t)|^2 \right)^{1/2}.
\]

We let \( \mathbb{B}^T \subset \mathbb{C}^T \) and \( \mathbb{B}^T_S \subset \mathbb{M}^T_S(\mathbb{C}) \) denote the closed unit balls under the Euclidean \( \ell_2 \) 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 \( \text{dist}(u, v) := \|u - v\|_2 \) for the Euclidean metric on \( \mathbb{C}^T \). We also equip \( \mathbb{M}^T_S(\mathbb{C}) \) with the \( \ell_2(T) \to \ell_2(S) \) operator norm
\[
\|A\|_{\text{op}} = \sup_{v \in \mathbb{S}^T} \|Av\|_2.
\]

For \( A \in \mathbb{M}^T_S(\mathbb{C}) \) we label its singular values in non-increasing order:
\[
\|A\|_{\text{op}} = \sigma_1(A) \geq \cdots \geq \sigma_{|S|\vee|T|}(A) \geq 0,
\]
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 \( A \in \mathbb{M}^{[n]\times [N]}_{[n]\times [N]}(\mathbb{C}) \) we have \( \sigma_l(A) = \sigma_l(\tilde{A}) \) 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 \( \text{Span}(E_1, \ldots, E_m) := \text{Span}(E_1 \cup \cdots \cup E_m) \) 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 \( A \in \mathbb{M}^T_S(\mathbb{C}) \), by \( \text{Span}(A) \) 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 \( \langle v_j : j \in J \rangle = \text{Span}(\{v_j\}_{j \in J}) \) and \( \langle v_j \rangle = \text{Span}(v_j) \). \( v_j \wedge v_k \) denotes the wedge product of \( v_j \) and \( v_k \).

3.4. Basic facts. Recall that an \( \varepsilon \)-net for a subset \( E \) of a metric space \( (X, d) \) 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 \( \varepsilon \)-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,
\[
\mathbb{P}\{\|X\|_{\text{op}} > C_0\} \lesssim e^{-N}.
\]

**Proof.** This holds more generally for matrices with independent complex uniformly-sub-Gaussian entries; see for instance [4] (one reduces to the real Ginibre case with the triangle inequality). \( \square \)

**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 \( \sigma_{\text{min}}(A) \leq \varepsilon \), 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 \( A_{-i} := \text{Span}\{\text{col}_j(A) : j \neq i\} \) that

\[
\text{dist}(\text{col}_i(A), A_{-i}) \leq \varepsilon \sqrt{N}.
\]

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

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

for a vector \( y \) that depends only on \( M \) and \( A_{-i} \), and hence is independent of \( \text{col}_i(X) \). Conditioning on the columns \( \text{col}_j(X) \) 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. \( \square \)

4. Linearization

Recall that \( \mathbb{C}(x) = \mathbb{C}(x_1, \ldots, x_n) \) is the set of polynomials with complex coefficients in \( n \) noncommuting indeterminates \( x_1, \ldots, x_n \). We express \( p \in \mathbb{C}(x) \) of degree two as

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

Lemma 4.1 (Linearization). Let \( \mathcal{A} \) be a unital \( \ast \)-algebra over \( \mathbb{C} \), let \( x_1, \ldots, x_n \in \mathcal{A} \), and let \( p \in \mathbb{C}(x) \) 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

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

and for \( z \in \mathbb{C} \) set

\[
\mathcal{L}^z = \mathcal{L}^z_p : \mathcal{A}^n \to \mathbb{M}_{[0,n_0]}(\mathcal{A}), \quad \mathcal{L}^z(x) := \mathcal{L}(x) + K^z \otimes 1_\mathcal{A} \tag{4.3}
\]

where

\[
K^z = K^z_p := \begin{pmatrix}
a_0 - z & 0 \\
0 & -I_{n_0}
\end{pmatrix} \in \mathbb{M}_{[0,n_0]}(\mathbb{C}). \tag{4.4}
\]

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

\[
(p - z)^{-1} = ((\mathcal{L}^z(x))^{-1})_{1,1}. \tag{4.5}
\]

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
\[
(M^{-1})_{1,1} = (M_{1,1} - M_{1,2}^{-1} M_{2,2} M_{1,2})^{-1}.
\]

**Proof.** For $s_0 = (s_{0,1}, \ldots, s_{0,n})$ we take $s_{0,\ell} = \overline{t_\ell}$. Write $p_2(x) = \sum_{\ell,m=1}^n a_{\ell,m} x_\ell x_m$ for the homogeneous degree-2 part of $p(x)$. Let $A = U\Sigma V^*$ be the singular value decomposition for $A$, with $U = (u_{\ell,k}), V = (v_{m,k}) \in \mathbb{C}^{n \times n}$, and $\Sigma = \text{diag}(\sigma_1, \ldots, \sigma_n)$ with $\sigma_1 \geq \cdots \geq \sigma_n > 0$. For $1 \leq k \leq n$ and $1 \leq \ell \leq n$, we set $r_{k,\ell} := u_{k,\ell}$ and $s_{k,\ell} = \sigma_k v_{k,\ell}$, so that
\[
p_2(x) = \sum_{k=1}^n \sigma_k \sum_{\ell=1}^n u_{k,\ell} x_\ell \sum_{m=1}^n v_{m,k} x_m
\]
\[
= \sum_{k=1}^n \left( \sum_{\ell=1}^n r_{k,\ell} x_\ell \right) \left( \sum_{m=1}^n s_{k,m} x_m \right)
\]
\[
= \sum_{k=1}^n \langle r_k, x \rangle \langle s_k, x \rangle.
\]
With $L^z(x)$ as in (4.3), by the Schur complement formula we have
\[
((L^z(x))^{-1})_{1,1} = \left( \langle s_0, x \rangle + a_0 - z + \sum_{k=1}^n \langle r_k, x \rangle \langle s_k, x \rangle \right)^{-1} = (p(x) - z)^{-1}
\]
as desired. \hfill \Box

**Corollary 4.2.** Let $X = (X_1, \ldots, X_n)$, and $P$ be as in Theorem 1.3 with $p$ as in (4.1). With $n_0$ the rank of $A = (a_{\ell,m})$, there exist vectors $s_0, s_1, \ldots, s_{n_0} \in \mathbb{C}^n$ depending only on $p$, with $s_1, \ldots, s_{n_0}$ nonzero and mutually orthogonal, such that the following holds. Let
\[
L^z := L + K^z \otimes I_N
\]
with $K^z$ as in (4.4) and
\[
L := \begin{pmatrix}
Y_0 & X_1 & \cdots & X_{n_0} \\
\vdots & \vdots & \ddots & \vdots \\
Y_{n_0} & 0 & \cdots & 0
\end{pmatrix},
\]
where
\[
Y_k := \sum_{\ell=1}^n s_{k,\ell} X_\ell, \quad k \in [0, n_0].
\]
For all $\varepsilon > 0$ and $z \in \mathbb{C}$,
\[
\mathbb{P}\{\sigma_{\min}(P - z) \leq \varepsilon\} \leq \mathbb{P}\{\sigma_{\min}(L^z) \leq \varepsilon\}.
\]

**Proof.** We apply Lemma 4.1 to get a linearized matrix $L^z(X)$ as in (4.3), with
\[
(P - z)^{-1} = ((L^z(X))^{-1})_{1,1}
\]
where the right hand side is the $N \times N$ top-left block of $(L^z(X))^{-1}$. In particular,
\[
\|(P - z)^{-1}\|_{\text{op}} = \|(L^z(X))^{-1}\|_{1,1} \leq \|(L^z(X))^{-1}\|_{\text{op}}.
\]
Now extend $r_1, \ldots, r_{n_0}$ to an orthonormal basis $\mathbb{C}^n$ and let $R$ be the unitary matrix with rows $r_1, \ldots, r_n$. Replacing the $N^2$ independent complex standard Gaussian vectors $X(i,j) = (X_1(i,j), \ldots, X_n(i,j))$ with $R^* X(i,j)$ and $s_0, \ldots, s_{n_0}$ with $s_0 R^*, \ldots, s_{n_0} R^*$ 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. \hfill \Box
It will be convenient later in the proof to assume $n_0 ≥ 2$, so we now dispense with the case that $n_0 = 1$.

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

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

$$P = X_1(α_1X_1 + α_2X_2) + \sum_{k=1}^{3} β_kX_k + γ.$$ 

If $β_3 ≠ 0$ then, conditioning on $X_1, X_2$, we have from Lemma 3.3 that

$$P\{σ_{min}(P) ≤ ε\} ≤ P\{σ_{min}(X_3 + M) ≤ ε/|β_3|\} ≲ (ε/|β_3|)^2N^3 = O_p(N^3ε^2),$$

where $M$ is a deterministic shift depending on $X_1, X_2$. Having obtained the claim in this case, we henceforth assume that $β_3 = 0$. Now write

$$P = (α_2X_1 + β_2)X_2 + (α_1X_1^2 + β_1X_1 + γ).$$

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

$$P\{σ_{min}(P) ≤ ε\} ≤ P\{σ_{min}(α_2X_1 + β_2) ≤ δ\} + P\{σ_{min}(X_2 + M') ≤ ε/δ\}$$

where $M' = (α_2X_1 + β_2)^{-1}(α_1X_1^2 + β_1X_1 + γ)$ is a deterministic shift depending on $X_1$ (note that $M'$ is well defined off a null event since $α_2X_1 + β_2$ is almost-surely invertible). Taking $δ = \min\{√ε, |β_2|/2\}$, the claim again follows from Lemma 3.3.

Finally, assuming $β_3 = β_2 = α_2 = 0$, we have

$$P = α_1X_1^2 + β_1X_1 + γ.$$ 

If $α_1 = 0$ we can conclude along the same lines as in previous cases. Otherwise we can factorize $P = α_1(X_1 - a_+)(X_1 - a_-)$ for some $a_± ∈ ℂ$. Then since $σ_{min}((X_1 - a_+)(X_1 - a_-)) ≥ σ_{min}(X_1 - a_+σ_{min}(X_1 - a_-)$ we have

$$P\{σ_{min}(P) ≤ ε\} ≤ P\{σ_{min}(X_1 - a_+) ≤ (ε/|α_1|)1/2\} + P\{σ_{min}(X_1 - a_-) ≤ (ε/|α_1|)1/2\} ≲ εN^3$$

and the lemma is proved. □

5. Reduction to a Bounded-Dimensional Test Projection

In view of Corollary 4.2 we henceforth write

$$L^z(x) = L(x) + K^z ⊗ 1ₐ, \quad \text{with} \quad L(x) := \begin{pmatrix}
⟨s_0, x⟩ & x_1 & \cdots & x_{n_0} \\
⟨s_1, x⟩ & 0 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
⟨s_{n_0}, x⟩ & 0 & \cdots & 0
\end{pmatrix} \quad \text{(5.1)}$$

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

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

$$L((k,i),(l,j)) = L^{k,l}(i,j) = L_i,j(k,l) = \tilde{L}(i,k),(j,l), \quad k,l ∈ [0,n_0], \quad i,j ∈ [N]$$

with

$$L_{i,j} = L(X_1(i,j), \ldots, X_n(i,j)) ∈ ℳ_{[0,n_0]}(ℂ),$$

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

$$L^z_{i,j} = L^z(X_1(i,j), \ldots, X_n(i,j)) = L_{i,j} + K^zδ_{i,j}.$$
Note that as an element of $\mathbb{M}_N(\mathbb{M}_{[0,n_0]}(\mathbb{C}))$, $\tilde{L}^z = (L_{i,j})_{i,j \in [N]}$ has independent entries, and $\tilde{L} = (L_{i,j})_{i,j \in [N]}$ 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 $1 \leq j \leq N$ we write $\mathbf{L}_{(j)} \in \mathbb{M}_{N,N-1}(\mathbb{M}_{[0,n_0]}(\mathbb{C}))$ (resp. $\mathbf{L}_{(j)}$) for the submatrix of $\mathbf{L}^z$ (resp. $\mathbf{L}$) obtained by removing the $j$th column from each $N \times N$ submatrix. We let $V_j^z \subset \mathbb{C}^{[0,n_0] \times [N]}$ denote the span of the columns of $\mathbf{L}_{(j)}$, viewed as an element of $\mathbb{M}_{[0,n_0] \times [N-1]}(\mathbb{C})$, that is

$$V_j^z := \text{Span}\{\text{col}_{(i,j')}(\mathbf{L}^z) : j' \in [N] \setminus \{j\}, \; i \in [0,n_0]\}. \tag{5.2}$$

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

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

$$\tilde{L}_j^z := \tilde{U}_j^* \text{col}_j(\mathbf{L}^z) = \sum_{i=1}^N U_{i,j}^* L_{i,j} \in \mathbb{M}_{[0,n_0]}(\mathbb{C}). \tag{5.3}$$

Then for any $\varepsilon > 0$,

$$\mathbb{P}\{\sigma_{\min}(\mathbf{L}^z) \leq \varepsilon\} \leq \mathbb{P}\{|\det(\tilde{L}_1^z)| \leq \varepsilon(|z| + C_p)\sqrt{n} + \varepsilon^N\}$$

for some $C_p < \infty$ depending only on $p$.

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

$$\varepsilon \geq \|\mathbf{L}^z \mathbf{v}\|_2 = \left\| \sum_{j=1}^N \text{col}_j(\mathbf{L}^z) v_j \right\|_2.$$ 

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

$$\varepsilon \geq \|\tilde{U}_{j_0}^* \text{col}_j(\mathbf{L}^z) v_{j_0}\|_2 = \|\tilde{L}_{j_0}^z v_{j_0}\|_2 \geq \frac{1}{\sqrt{N}} \sigma_{\min}(\tilde{L}_{j_0}^z).$$

Now note that

$$\|\tilde{L}_{j_0}^z\|_{\text{op}} \leq \|\mathbf{L}^z\|_{\text{op}} \leq \|\mathbf{L}\|_{\text{op}} + |z| + O_p(1).$$

From Lemma 3.2 and the triangle inequality,

$$\mathbb{P}\{|\mathbf{L}\|_{\text{op}} > B\} \leq e^{-N}.$$
for some $B = O_p(1)$. Thus, by the union bound we have for some $B' = O_p(1)$,
\[
\mathbb{P}\{\sigma_{\min}(L^z) \leq \varepsilon\} \leq e^{-N} + \sum_{j=1}^{N} \mathbb{P}\{\sigma_{\min}(\tilde{L}^z_j) \leq \varepsilon \sqrt{N}, \|\tilde{L}^z_j\|_{op} \leq B' + |z|\}
\leq e^{-N} + \sum_{j=1}^{N} \mathbb{P}\{|\det(\tilde{L}^z_j)| \leq (B' + |z|)^{n_0} \varepsilon \sqrt{N}\}.
\]
Since the distribution of $\tilde{L}^z = (L^z_{i,j})_{i,j \in [N]}$ is invariant under simultaneous permutation of the $N$ row and $N$ column indices, the claim follows. \hfill \Box

### 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 $\tilde{L}^z_k$ defined in (5.3). The advantage of this perspective is that for a fixed realization of $U_1$ (which we recall is independent of $\text{col}_1(\tilde{L}^z)$), the summands $U^*_{i,1}L^z_{i,1}$ are $N$ independent random matrices. We can thus view $\tilde{L}^z_k$ as a random walk in $\mathbb{M}_{[0,n_0]}(\mathbb{C})$.

In this section we consider an arbitrary fixed (deterministic) matrix
\[
U \in \mathbb{M}_{[0,n_0]}^{[0,n_0] \times [N]}(\mathbb{C}) \cong \mathbb{M}_N(\mathbb{C})^{0,n_0}
\]
with orthonormal columns in $\mathbb{C}^{[0,n_0] \times [N]}$, identified with the sequence of its block submatrices as $U = (U^0, \ldots, U^{n_0}) \in \mathbb{M}_N(\mathbb{C})^{0,n_0}$ and with the sequence of its square submatrices denoted $U = (U_1, \ldots, U_N) \in \mathbb{M}_{[0,n_0]}(\mathbb{C})^N$. Thus, denoting the rows of $U_i$ by $(v^k_i)$, $k \in [0, n_0]$, we have that $(\nu^k_i)^*$ is the $i$th row of $U^k$.

Let $(g^k_i)_{i \in [N], k \in [n]}$, $(h^k_i)_{i \in [N], k \in [n]} \in \mathbb{C}^{[N] \times [n]}$ be independent arrays of $Nn$ i.i.d. standard real Gaussians and denote $g_i = (g^1_i, \ldots, g^n_i)$, $g^k = (g^k_i)_{i \in [N]}$, and similarly for $h_i, h^k$. We set $\xi^k_i = (2N)^{-1/2}(g^k_i + ih^k_i)$, $\xi_i = (\xi^1_i, \ldots, \xi^n_i)$, $\xi^k = (2N)^{-1/2}(g^k + ih^k)$. Then $\tilde{L}^z_k$ from (5.3) is identically distributed to the $(n_0 + 1) \times (n_0 + 1)$ random matrix
\[
W = M + \sum_{i=1}^{N} U^*_i L_i
\]
where
\[
M = M(U, z, a_0) := U^*_1 K^2 \in \mathbb{M}_{[0,n_0]}(\mathbb{C})
\]
is deterministic and the square matrices
\[
L_i := \mathcal{L}(\xi_i^1, \ldots, \xi_i^n) = \begin{pmatrix}
\langle s_0, \xi_i \rangle & \xi_i^1 & \cdots & \xi_i^{n_0} \\
\langle s_1, \xi_i \rangle & 0 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
\langle s_n, \xi_i \rangle & 0 & \cdots & 0
\end{pmatrix} \in \mathbb{M}_{[0,n_0]}(\mathbb{C})
\]
are i.i.d., with $a_0, s_0, \ldots, s_n$ 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 $g^k$, of the form
\[
\mathbb{P}\{|\det(W)| \leq \varepsilon\} \leq N^C \varepsilon^c.
\]
In this section we identify sufficient structural conditions on the matrix $U$ in order to have (6.4). In Section 7 we will show that such conditions hold with high probability for the matrix $U_1$.\hfill \Box
For $1 \leq \ell \leq n$ let
\[ Q^\ell = Q^\ell(U) = \sum_{k=0}^{n_0} s_{k,\ell} U^k \in M^{[0,n_0]}_{[N]}(\mathbb{C}) \]  
(6.5)
and for $i \in [N]$ denote the $i$th row of $Q^\ell$ by $(w^\ell_i)^*$; thus,
\[ w^\ell_i = \sum_{k=0}^{n_0} s_{k,\ell} v^k_i. \]  
(6.6)

For $\ell \in [n]$ and $i_0, i_1, \ldots, i_{n_0} \in [N]$ denote
\[ \Delta^\ell_{i_0,i_1,\ldots,i_{n_0}} = \Delta^\ell_{i_0,i_1,\ldots,i_{n_0}}(U) := \det(w^\ell_{i_0}, v^0_{i_1}, \ldots, v^0_{i_{n_0}}). \]  
(6.7)

For a parameter $\delta > 0$ we define sets of structured matrix-columns:
\[ \text{Struct}_1(\delta) := \bigcap_{i_0,i_1,\ldots,i_{n_0} \in [N]} \bigcap_{\ell \in [n_0]} \left\{ U \in M^{[0,n_0]}_{[0,n_0] \times [N]}(\mathbb{C}) : |\Delta^\ell_{i_0,i_1,\ldots,i_{n_0}}| < \delta \right\}, \]  
(6.8)
\[ \text{Struct}_2(\delta) := \bigcap_{i_1,\ldots,i_{n_0} \in [N]} \bigcap_{\ell \in [n_0]} \left\{ U \in M^{[0,n_0]}_{[0,n_0] \times [N]}(\mathbb{C}) : |\Delta^\ell_{i_1,i_2,\ldots,i_{n_0}}| < \delta \right\}, \]  
(6.9)
\[ \text{Struct}(\delta) := \text{Struct}_1(\delta) \cap \text{Struct}_2(\delta). \]  
(6.10)

**Lemma 6.1** (Anti-concentration for the determinant of a matrix random walk). Let $U = (U_1, \ldots, U_N) \in M^{[0,n_0]}_{[0,n_0] \times [N]}(\mathbb{C})$ and suppose $U \notin \text{Struct}(\delta)$ for some $\delta > 0$. Then for any $\varepsilon > 0$,
\[ \sup_{M \in M^{[0,n_0]}_{[0,n_0]}} \mathbb{P}\left( \left| \det\left( M + \sum_{i=1}^N U^*_i L_i \right) \right| \leq \varepsilon \right) \lesssim_{n_0} N^{1/2}(\varepsilon/\delta)^{1/(n_0+1)}. \]  
(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 $g_1, \ldots, g_n$ 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 $\varepsilon > 0$,
\[ \sup_{t \in \mathbb{R}} \mathbb{P}\left( |p(g_1, \ldots, g_n) - t| \leq \varepsilon \right) \lesssim_d (\varepsilon/|a|)^{1/d}. \]

**Proof.** By rescaling $p$ and $\varepsilon$ by $a$ we may assume $a = 1$. From the Carbery–Wright inequality \[\text{IN}\] we have
\[ \sup_{t \in \mathbb{R}} \mathbb{P}\left( |p(g_1, \ldots, g_n) - t| \leq \varepsilon \sigma \right) \lesssim_d \varepsilon^{1/d} \]
where $\sigma^2 = \text{Var}(p(g_1, \ldots, g_n))$ is the variance of $p$ under the product Gaussian measure. It hence suffices to verify that
\[ \text{Var}(p(g_1, \ldots, g_n)) \gtrsim_d 1. \]

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. \[\square\]

**Proof of Lemma 6.1.** Write
\[ D = \det \hat{L}, \quad \hat{L} := M + \sum_{i=1}^N U^*_i L_i. \]  
(6.12)
Denoting the columns of $M$ by $y^0, \ldots, y^{n_0}$, we have
\[
D = \det \left( y^0 + \sum_{i_0=1}^{N} \sum_{k=0}^{n_0} (s_k, \xi^0_{i_0}) v_{i_0}^k, y^1 + \sum_{i_1=1}^{N} \xi^1_{i_1} v_{i_1}^0, \ldots, y^{n_0} + \sum_{i_{n_0}=1}^{N} \xi_{i_{n_0}}^{n_0} v_{i_{n_0}}^0 \right)
\]
\[
= \det \left( y^0 + \sum_{i_0=1}^{N} \sum_{i_1=1}^{n} \xi_{i_0}^{i_1} w_{i_0}^{i_1}, y^1 + \sum_{i_1=1}^{N} \xi^1_{i_1} v_{i_1}^0, \ldots, y^{n_0} + \sum_{i_{n_0}=1}^{N} \xi_{i_{n_0}}^{n_0} v_{i_{n_0}}^0 \right). \tag{6.13}
\]

Now we condition on the variables $h = (h^k_i)_{i \in [N], k \in [n]}$, the imaginary parts of $\sqrt{2N} \xi$, and view $D$ as a degree-$(n_0 + 1)$ polynomial in the $nN$ i.i.d. standard real Gaussian variables $g = (g^k_i)$, the real parts of $\sqrt{2N} \xi$. Write $\hat{D}$ for the homogeneous degree-$(n_0 + 1)$ part of this polynomial. We have
\[
\hat{D} = (2N)^{-(n_0+1)/2} \sum_{\ell=1}^{n_0} \sum_{i_{0}, \ldots, i_{n_0}=1}^{N} g^\ell_{i_0} g^1_{i_1} \cdots g^{n_0}_{i_{n_0}} \cdot \det(w^\ell_{i_0}, v^0_{i_1}, \ldots, v^0_{i_{n_0}}) \tag{6.14}
\]
\[
= (2N)^{-(n_0+1)/2} \sum_{\ell=1}^{n_0} \sum_{i_{0}, \ldots, i_{n_0}=1}^{N} g^\ell_{i_0} g^1_{i_1} \cdots g^{n_0}_{i_{n_0}} \cdot \Delta^\ell_{i_0, i_1, \ldots, i_{n_0}}. \tag{6.15}
\]

In particular, for $\ell \leq n_0$, the coefficient of the monomial $g^\ell_{i_0} g^1_{i_1} \cdots g^{n_0}_{i_{n_0}}$ is
\[
2 \cdot (2N)^{-(n_0+1)/2} \Delta^\ell_{i_0, i_1, \ldots, i_{n_0}} \tag{6.16}
\]
and for $\ell > n_0$, the coefficient of the monomial $g^\ell_{i_0} g^1_{i_1} \cdots g^{n_0}_{i_{n_0}}$ is
\[
(2N)^{-(n_0+1)/2} \Delta^\ell_{i_0, i_1, \ldots, i_{n_0}}. \tag{6.17}
\]

Since $U \notin \text{Struct}(\delta)$, 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 $\ell \leq n_0$ and $i_1, \ldots, i_{n_0} \in [N]$. 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
\[
\mathbb{P}\{|D| \leq \varepsilon\} \leq \mathbb{P}\{\Re D| \leq \varepsilon\} \lesssim_{n_0} N^{1/2}(\varepsilon/\delta)^{1/(n_0+1)}
\]
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 $V_{(1)} \subset \mathbb{C}^{[0, n_0] \times [N]}$ denotes the span of the $(n_0 + 1)(N-1)$ columns of $L^{(1)}_1$, with the latter viewed as an element of $\mathbb{M}^{[0, n_0] \times [N-1]}(\mathbb{C})$, and that the columns of $U_1$ are a random orthonormal set in $(V_{(1)}^\perp)$. (In view of Theorem 1.3 we have, a posteriori, that with high probability $L^{(1)}_1$ is full rank and the columns of $U_1$ in fact comprise a basis for $(V_{(1)}^\perp)$.) The aim of this section is to establish the following:

**Proposition 7.1** (Structured bases are rare). Assume $n_0 \geq 2$. Then, we have
\[
\mathbb{P}\{U_1 \in \text{Struct}(N^{-\frac{1}{2}n_0-10})\} \lesssim_{p, z} e^{-N} \tag{7.1}
\]
where the sets $\text{Struct}(\delta)$ 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 \( \|P\|_{op} = O_p(1) \). In particular we may assume \( |z| = O_p(1) \) since otherwise we obtain the claim by simply lower bounding \( \sigma_{\min}(P - z) \geq |z| - \|P\|_{op} \). Now by Corollary 4.2
\[
P\{\sigma_{\min}(P - z) \leq \varepsilon \} \leq \mathbb{P}\{\sigma_{\min}(L^*) \leq \varepsilon \}
\]
and from Lemma 5.1 the latter is in turn bounded by
\[
N \mathbb{P}\{ |\det(\tilde{L}_i)\| \leq \varepsilon (|z| + C_p)^{n_0} \sqrt{N} \} + e^{-N}.
\]
Finally, from Proposition 7.1. (6.1) and Lemma 6.1
\[
P\{ |\det(\tilde{L}_i)| \leq \varepsilon (|z| + C_p)^{n_0} \sqrt{N} \} \leq_p e^{-N} + N^{2(1+(n_0+21)/(n_0+1))} (|z| + C_p)^{n_0/(n_0+1)} e^{1/(n_0+1)}
\]
\[
\leq_p e^{-N} + N^{13/3} e^{1/(n_0+1)},
\]
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. \( \square \)

7.1. High-level proof of Proposition 7.1 Our approach is to cut the set \( \text{Struct}(\delta) \) (for sufficiently small \( \delta = N^{-O_p(1)} \)) into several pieces, and to bound the event that \( U_1 \) lies in each piece by taking union bounds over nets. (Recall the definition of a net from Section 3.4.) Once we have approximated \( U_1 \) by a some fixed element \( U \) of a net, our task is then to bound the probability that the columns of \( U \) are nearly contained in \( (V^{(1)})^\perp \), i.e. that \( U^* \col_j(\tilde{L}^*) \approx 0 \) (the zero matrix in \( \mathbb{M}_{[0,n_0]}(\mathbb{C}) \)) for each \( 2 \leq j \leq N \).

We have
\[
\tilde{U}^* \col_j(\tilde{L}^*) = \sum_{i=1}^{N} U_i^* L_{i,j}^* = U_j^* K^* + \sum_{i=1}^{N} U_i^* L_{i,j}
\]
\[
=: M_j(U, z) + \text{Walk}_j(U) \in \mathbb{M}_{[0,n_0]}(\mathbb{C})
\]
where the matrices \( M_j \) are deterministic shifts and \( \text{Walk}_j(U) \) are i.i.d. copies of the centered random walk
\[
\text{Walk}(U) = \sum_{i=1}^{N} U_i^* L_i
\]
with
\[
L_i = \mathcal{L}(\zeta_i^1, \ldots, \zeta_i^n) = \frac{1}{\sqrt{N}} \mathcal{L}(\zeta_i^1, \ldots, \zeta_i^n)
\]
as in (6.3), where \( \zeta_i^k = \frac{1}{\sqrt{2}} (g_i^k + i h_i^k) \) are i.i.d. standard complex Gaussians. Our task is thus reduced to proving an anti-concentration bound for \( \text{Walk}(U) \). Note that whereas in Section 6 we were concerned with anti-concentration for the determinant of \( W = M_1 + \text{Walk}(U) \), here we need to show the matrix \( \text{Walk}(U) \) is anti-concentrated as a random element of the vector space \( \mathbb{M}_{[0,n_0]}(\mathbb{C}) \).

Anti-concentration for \( \text{Walk}(U) \) 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 \( \frac{1}{\sqrt{N}} \sum_{\ell=1}^{n_0} \zeta_i^\ell M_\ell \) of matrices \( M_\ell \), where \( M_\ell \) has zeroth row \( (s_0(\ell), \ldots, s_{n_0}(\ell)) \), \( \ell \)th row \( e_\ell \) (the \( \ell \)th standard basis vector) and all other rows equal to zero. Thus
\[
\text{Walk}(U) = \frac{1}{\sqrt{N}} \sum_{(\ell,i) \in [n] \times [N]} \zeta_i^\ell U_i^* M_\ell
\]
Recalling (6.6), the matrix \( U_i^* M_\ell \) has zeroth column \( w_i^\ell \), \( \ell \)th column \( v_i^\ell \), and all other columns equal to zero. We can hence identify the matrix (7.3) with the Gaussian vector
\[
\frac{1}{\sqrt{N}} \textbf{W}^* \zeta \in \mathbb{C}^{[0,n_0]^2}
\]
where $\zeta = (\zeta^k_p)_{(k,i) \in [n] \times [N]}$, and the walk matrix $W = W(U)$ is defined

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

recalling the matrices $Q^\ell$ 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 $p = x_1 x_2 + x_2 x_1$.) From (7.4) we see that the probability the Gaussian vector $\text{Walk}(U)$ lands in a Hilbert–Schmidt ball of radius $\varepsilon$ scales like $O(\varepsilon^2)^{\text{rank}(W)}$ for small $\varepsilon$ (cf. Lemma 7.7). In particular, from the form of (7.5) we see that a lower bound $\sigma_{\text{min}}(U^0) \gtrsim 1$ 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 $U$ having a well-conditioned zeroth block:

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

Recall that $L_{(1)}$ is obtained by removing the first column from each of the $n_0 + 1$ blocks of $U$ columns of $L$, and that $U_1$ depends on $L$ only through $L_{(1)}$.

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

$$\mathbb{P}\{ U_1 \notin \mathcal{A}(\alpha) \land \|L_{(1)}\|_{\text{op}} \leq B \} \lesssim_p e^{-N} \quad (7.7)$$

for any $\alpha < c_p/(B^2 + |z|)$.

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

$$W_I := \text{Span}(W_{\emptyset}, q_{k_1}, \ldots, q_{k_r}) \quad \text{for} \quad I = \{k_1, \ldots, k_r\} \subset [0,n_0]. \quad (7.8)$$

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

$$\mathbb{P}\{ \|U^* L_{(1)}^z\|_{\text{HS}} \leq \varepsilon \} = O_{p,\alpha}(\varepsilon^2)^{n_0(n_0+1)(N-1)}. \quad (7.9)$$

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,

$$\text{dist}(q_{k_i}, W_{(k_1,\ldots,k_{i-1})}) \geq \beta \quad \forall 1 \leq i \leq r, \quad (7.10)$$

(interpreting the left hand side as $\text{dist}(q_{k_i}, W_{\emptyset})$ when $i = 1$), then for any $\varepsilon > 0$ and $1 \leq j \leq N$,

$$\mathbb{P}\{ \|U^* L_{(1)}^z\|_{\text{HS}} \leq \varepsilon \} = O_{p,\alpha,\beta}(\varepsilon^2)^{n_0(n_0+1)+r(N-1)}. \quad (7.11)$$
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 \( W(U) \) is well conditioned. The following lemma shows we may assume this is the case. For \( \beta > 0 \) we define the set

\[
B(\beta) := \bigcup_{k \in [0,n_0]} \left\{ U \in M_{[0,n_0] \times [N]}(C) : \text{dist}(q_k, W_{[0,n_0] \setminus \{k\}}) < \beta \right\}
\]  

(7.12)

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

\[
\mathbb{P}\{ U_1 \in B(\beta) \cap A(\alpha) \land \|L_z^2\|_{op} \leq B \} \lesssim e^{-N}.
\]

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 \( U_1 \) lies in the set

\[
\mathcal{E} = \mathcal{E}(\delta, \alpha, \beta) := \text{Struct}(\delta) \cap A(\alpha) \cap B(\beta)^c.
\]

(7.13)

We will bound the probability that \( U_1 \) 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 \( \alpha > 0 \), \( \rho_\ast \in (0, 1) \) and \( 0 < \delta \leq \rho_\ast N^{-\frac{n_0-2}{2}} \), \( \mathcal{E} \) has a (Hilbert–Schmidt) \( \rho_\ast \)-net \( \mathcal{N} \subset \mathcal{E} \) of size

\[
|\mathcal{N}| = O_{\rho, \alpha}(1)^N N^{(2n_0 + 3)N} O(1/\rho_\ast)(n_0^2 + \frac{1}{2}n_0 + \frac{1}{2}) N.
\]

(7.14)

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

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 7.2 and the triangle inequality we have that \( \|L\|_{op} = O_p(1) \) with probability \( 1 - O_p(e^{-N}) \), and hence \( \|L(1)\|_{op}, \|L_z^2\|_{op}, \|L(1)^2\|_{op} = O_{p,z}(1) \) with probability \( 1 - O_p(e^{-N}) \). Applying the union bound, for \( \alpha, \beta > 0 \) and some \( B = O_{p,z}(1) \) we have

\[
\mathbb{P}\{ U_1 \in \text{Struct}(\delta) \} \leq \mathbb{P}\{ U_1 \in \mathcal{E}(\delta, \alpha, \beta) \land \|L(1)\|_{op} \leq B \}
\]

+ \[
\mathbb{P}\{ U_1 \notin A(\alpha) \land \|L(1)\|_{op} \leq B \}
\]

+ \[
\mathbb{P}\{ U_1 \in A(\alpha) \cap B(\beta) \land \|L^2\|_{op} \leq B \} + O_p(e^{-N}).
\]

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

\[
\mathbb{P}\{ U_1 \in \text{Struct}(\delta) \} \leq \mathbb{P}\{ U_1 \in \mathcal{E}(\delta, \alpha, \beta) \land \|L^2\|_{op} \leq B \} + O_p(e^{-N}).
\]

(7.15)

Let \( \rho_\ast > 0 \) to be chosen later and let \( \mathcal{N} \) be as in Lemma 7.5 (assuming \( \delta \) is sufficiently small depending on the choice of \( \rho_\ast \)). By approximating any realization of \( U_1 \in \mathcal{E} \) by some \( U \in \mathcal{N} \) with \( \|U_1 - U\|_{HS} \leq \rho_\ast \) and using that

\[
\|(U_1 - U)^* L(1)^2_s \|_{HS} \leq \|L^2\|_{op} U_1 - U \|_{HS}
\]

we have, since \( U_1^* L(1)_s \) vanishes everywhere by definition,

\[
\mathbb{P}\{ U_1 \in \mathcal{E} \land \|L^2\|_{op} \leq B \} \leq \mathbb{P}\{ \exists U \in \mathcal{E} : U^* L(1)^2_s = 0 \land \|L^2\|_{op} \leq B \}
\]

\[
\leq \mathbb{P}\{ \exists U \in \mathcal{N} : \|U^* L(1)^2_s\|_{HS} \leq B \rho_\ast \}.
\]

(7.16)
Now applying the union bound and (7.11) with \( r = n_0 + 1 \), using that \( \mathcal{N} \subset \mathcal{E} \subset B(\beta)^c \), we have
\[
\mathbb{P}\{ U_1 \in \mathcal{E} \land \| L \|_{op} \leq B \} \leq |N| O_{p,z}(\rho_*^2)^{(n_0+1)^2(N-1)}
\leq O_{p,z}(1)^N N(2n_0+3)N^2 \rho_*^2 \left[(n_0+1)^2-n_0^2-\frac{3}{2}n_0-\frac{3}{2}\right]N-O_p(1)
= O_{p,z}(1)^N N(2n_0+3)N^{(n_0-1)}-O_p(1).
\]
Recalling our assumption that \( n_0 \geq 2 \), one verifies the last expression is at most
\[
O_{p,z}(1)^N (C')^{N-O_p(1)} \lesssim e^{-N}
\]
if we take
\[
\rho_* = c' N^{-(2n_0+3)/(n_0-1)}
\]
for a sufficiently small constant \( c'(p,z) > 0 \). Now one verifies that the condition of Lemma 7.5 is satisfied for all \( N \geq N_0 \) for a sufficiently large constant \( N_0(p,z) > 0 \) 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 \( L_{(1)}^z \) in block form as follows:
\[
L_{(1)}^z = \begin{pmatrix}
\text{row}_1(Y'_0) & \text{row}_1(X'_1) & \text{row}_1(X'_2) & \cdots & \text{row}_1(X'_{n_0}) \\
Y''_0 + (a_0 - z) I_{N-1} & X''_1 & X''_2 & \cdots & X''_{n_0} \\
\text{row}_1(Y'_1) & 0_{1,N-1} & 0 & \cdots & 0 \\
Y''_1 & -I_{N-1} & \text{row}_2(Y'_2) & 0_{1,N-1} & 0 \\
Y''_2 & 0 & \text{row}_2(Y'_2) & -I_{N-1} & 0 \\
\vdots & \vdots & \vdots & \ddots & 0 \\
\text{row}_1(Y'_{n_0}) & 0 & \cdots & 0 & 0_{1,N-1} \\
Y''_{n_0} & -I_{N-1} & \text{row}_1(Y'_{n_0}) & 0 & 0 \\
\end{pmatrix}
\]

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
\[
\hat{Y} = \begin{pmatrix}
\text{row}_1(Y'_1) \\
\vdots \\
\text{row}_1(Y'_{n_0})
\end{pmatrix} \in \mathbb{M}_{n_0,N-1}(\mathbb{C}).
\]

We first show that
\[
\mathbb{P}\{ \sigma_{\min}(\hat{Y}) < c'_p \} \lesssim e^{-N}
\]
if \( c'_p > 0 \) is sufficiently small. On the event that \( \sigma_{\min}(\hat{Y}) < c'_p \), there exists \( x \in \mathbb{S}^{n_0-1} \) such that
\[
c'_p > \| x^* \hat{Y} \|_2 = \left\| \sum_{k=1}^{n_0} x(k) \sum_{\ell \in [n]} s_{k,\ell} \text{row}_1(X'_{\ell}) \right\|_2 = \| y(1) \text{row}_1(X'_1) + \cdots + y(n) \text{row}_1(X'_{n_0}) \|_2
\]
where \( y := \sum_{k=1}^{n_0} x(k) s_{k,\ell_0} \). Since \( s_{1,\ell_0}, \ldots, s_{n_0,\ell_0} \) are mutually orthogonal and nonzero it follows that \( \| y \|_2 \gtrsim 1 \), and in particular \( y \) must have a coordinate \( y(\ell_0) \) of size \( |y(\ell_0)| \gtrsim 1 \). By projecting \( \text{row}_1(X'_{\ell_0}) \) to the orthocomplement of \( \mathcal{W}_{(\ell_0)} = \text{Span}\{\text{row}_1(X'_{\ell})\}_{\ell \neq \ell_0} \) we see that \( \text{dist}(\text{row}_1(X'_{\ell_0}), \mathcal{W}_{(\ell_0)}) \gtrsim c'_p \). Taking a union bound to fix \( \ell_0 \) and conditioning on the rows \( \text{row}_1(X'_{\ell}) \) with \( \ell \neq \ell_0 \), we are left with bounding the probability that the vector \( \text{row}_1(X'_{\ell_0}) \) is within distance \( O_p(c'_p) \) of a fixed subspace of \( \mathbb{C}^{N-1} \) of dimension \( n-1 \). Since \( X'_{\ell} \) has i.i.d. complex Gaussian entries of variance \( 1/N \),
by rotational invariance of the Gaussian measure we may assume \( \tilde{W}(\ell_0) \) is the coordinate subspace spanned by \( e_2, \ldots, e_n \), and we have reduced to bounding the probability of the event that

\[
\sum_{j=n+1}^N |X_{\ell_0}(1,j)|^2 = O_p(c'_p).
\]

Now we can take \( c'_p \) 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 \( L(1) \) such that \( \|L(1)\|_{op} \leq B \) and \( \sigma_{\min}(\tilde{Y}) \geq c'_p \). This fixes the subspace \((V^\perp(1))^\perp\). We claim that

\[
\{ U = (u_0, \ldots, u_{n_0}) : u_0, u_1, \ldots, u_{n_0} \text{ orthonormal in } (V^\perp(1))^\perp \cap A(\alpha) = \emptyset \} \quad (7.20)
\]

for \( \alpha < c_p/(B^2 + |z|) \). From this the lemma clearly follows.

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

\[
\bar{u}^0(1) \text{row}_1(X_k^\perp) + (\bar{u}^0)^* X_k^\perp = (\hat{u}^k)^*.
\]

(7.21)

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

\[
\|\hat{u}^k\|_2 \leq 2B/\alpha, \quad \forall 1 \leq k \leq n_0.
\]

(7.22)

Again from the fact that \( u^* \) is a left null vector for \( L^\perp(1) \), we have

\[
\sum_{k=1}^{n_0} u^k(1) \text{row}_1(Y_k^\perp) = -\bar{u}^0(1) \text{row}_1(Y_0^\perp) - (\bar{u}^0)^* (Y_0^\perp + (a_0 - z) I_{N-1}) - \sum_{k=1}^{n_0} (\hat{u}^k)^* Y_k^\perp
\]

for each \( 1 \leq k \leq n_0 \). Writing \( \hat{u}_1 = (u^1(1), \ldots, u^{n_0}(1)) \), the left hand side has norm at least \( \sigma_{\min}(\tilde{Y}) ||\hat{u}_1||_2 \geq c'_p ||\hat{u}_1||_2 \). Since \( L(1) \), 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 \( O_p(\alpha(B^2 + |z|)) \). Putting these bounds together we have

\[
||\hat{u}_1||_2 \leq (B^2 + |z|) \alpha.
\]

(7.23)

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

\[
1 = \|u\|_2^2 = \|u^0\|_2^2 + \|\hat{u}_1\|_2^2 + \sum_{k=1}^{n_0} \|\hat{u}^k\|_2^2 \leq \alpha^2 (1 + O_p(B^2 + |z|)^2)
\]

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


Lemma 7.7. Let \( M \in \mathbb{M}_{d,d'}(\mathbb{C}) \) and let \( Z = (\zeta_1, \ldots, \zeta_{d'}) \) be a vector of i.i.d. standard complex Gaussians. For \( I \subseteq [d] \) let \( M_I \in \mathbb{M}_{|I|,|I'|}(\mathbb{C}) \) 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

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

(7.23)
Proof. By projecting $MZ$ to the coordinate subspace $C^d$ 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 $\Sigma = \text{diag}(\sigma_1(M), \ldots, \sigma_d(M))$. We have
\[
\sup_{w \in C^d} P(\|MZ - w\|_2 \leq \varepsilon) = \sup_{w \in C^d} P(\|\Sigma V^* Z - U^* w\|_2 \leq \varepsilon) = \sup_{w \in C^d} P(\|\Sigma V^* Z - w\|_2 \leq \varepsilon) = \sup_{w \in C^d} P(\|\Sigma Z - w\|_2 \leq \varepsilon)
\]
where we used the invariance of the Euclidean norm and the distribution of $Z$ under unitary transformations of $C^d$ and $C^{d'}$, respectively. Now we have
\[
P(\|\Sigma Z - w\|_2 \leq \varepsilon) = P\left(\sum_{j=1}^d |\sigma_j(M) \zeta_j - w_j|^2 \leq \varepsilon^2\right) \leq \prod_{j=1}^d P(|\sigma_j(M) \zeta_j - w_j| \leq \varepsilon)
\]
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. \hfill \qed

Lemma 7.8 (Tensorization of anti-concentration (cf. [44])). Suppose that $\xi_1, \ldots, \xi_m$ are independent non-negative random variables such that, for some $\varepsilon_1, \varepsilon_2 > 0$, $P\{\xi_i \leq \varepsilon_1\} \leq \varepsilon_2$ for all $1 \leq i \leq m$. Then
\[
P\left\{\sum_{i=1}^m \xi_i^2 \leq \varepsilon_1^2 m\right\} = O(\varepsilon_2)^m.
\]

We now prove Lemma 7.3 Since $U \in A(\alpha)$ we have $\sigma_{\min}(R) = \sigma_{\min}(U^0) \geq \alpha$, and hence
\[
det(R^* R) \gtrsim_{p, \alpha} 1.
\]
(7.24)

Letting $\varepsilon > 0$ be arbitrary, from (7.4) and Lemma 7.7 with $M = W^*$, $M_I = R^*$ we obtain
\[
\sup_{M \in M_{[0,n_0]}(C)} P\{\|M + \text{Walk}(U)\|_{HS} \leq \varepsilon N^{-1/2}\} = O_{p, \alpha}(\varepsilon^n)^{n_0(n_0+1)}.
\]
In particular, from (7.2) we have that for each $2 \leq j \leq N$,
\[
P\{\|U^* \text{col}_j(\tilde{L}^z)\|_{HS} \leq \varepsilon N^{-1/2}\} = O_{p, \alpha}(\varepsilon^n)^{n_0(n_0+1)}.
\]
Since the columns $\text{col}_j(\tilde{L}^z)$ are independent, from Lemma 7.8 we obtain
\[
P\{\|U^* L^z(1)\|_{HS} \leq \varepsilon\} = P\left\{\sum_{j=1}^N \|U^* \text{col}_j(\tilde{L}^z)\|_{HS}^2 \leq \varepsilon^2\right\} = O_{p, \alpha}(\varepsilon^n)^{n_0(n_0+1)(N-1)}
\]
which is (7.9).

For (7.11), under the assumptions $\sigma_{\min}(U^0) \geq \alpha$ and (7.10), with $I = \{k_1, \ldots, k_r\}$ we have
\[
det(W^*_I W_I) \gtrsim \det(R^* R) \cdot \beta^{2r} \gtrsim_{p, \alpha, \beta} 1,
\]
(recalling the notation around (7.8) and using the fact that
\[
det(W^*_I W_I) = |\text{dist}(q_{k_{i+1}}, W_I)|^2 \det(W^*_I W_I)
\]
for $I_i = \{k_1, \ldots, k_i\}$) and the proof concludes by following the same lines as we did for (7.9). \hfill \qed
7.4. Proof of Lemma \[\text{7.4} \]: 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

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

(recall our notation \(\mathbb{B}^T \subset M^T_0(\mathbb{C})\) for the closed unit Hilbert–Schmidt ball). Note that \(U_1 \in \mathbb{B}_*\) almost surely since its \(n_0 + 1\) columns are unit vectors. For \(\beta_1 > 0\) let

\[
\mathcal{B}_0(\beta_1) := \{ U \in \mathbb{B}_*: \text{dist}(q_k, W_0) < \beta_1 \ \forall k \in [0, n_0] \}
\]

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,

\[
\mathcal{B}(k_1, \ldots, k_r)(\beta_0, \beta_1) := \{ U \in \mathbb{B}_*: \text{dist}(q_k, W_{(k_1, \ldots, k_{i-1})}) \geq \beta_0 \ \forall 1 \leq i \leq r \}
\]

\[
\mathcal{R}_r(\beta_0, \beta_1) := \mathcal{B}(k_1, \ldots, k_r)(\beta_0, \beta_1) \setminus \bigcup_{k_{r+1} \notin \{k_1, \ldots, k_r\}} \mathcal{B}(k_1, \ldots, k_{r+1})(\beta_1).
\]

(interpreting \(W_{(k_1, \ldots, k_{i-1})}\) as \(W_0\) for \(i = 1\)).

**Claim 7.9.** For any \(\alpha \in (0, \frac{1}{2}]\) there exists \(c_0(p, \alpha) > 0\) such that for any \(\beta_1 \in (0, c_0]\), there is a set \(\mathcal{N}_0(\beta_1) \subset \mathbb{B}_* \cap \mathcal{A}(\alpha)\) with

\[
|\mathcal{N}_0(\beta_1)| \leq O_p((1/\beta_1^2)^{(n_0+1)N})
\]

and such that for any \(U \in \mathcal{B}_0(\beta_1) \cap \mathcal{A}(\alpha)\) there exists \(\tilde{U} \in \mathcal{N}_0(\beta_1)\) with

\[
\|U - \tilde{U}\|_{\text{HS}} \lesssim_{p, \alpha, \beta_1} 1.
\]

**Claim 7.10.** For any \(\alpha, \beta_0 \in (0, \frac{1}{2}]\) there exists \(c_1(p, \alpha, \beta_0) > 0\) 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

\[
|\mathcal{N}(k_1, \ldots, k_r)(\beta_0, \beta_1)| \leq O_p((1/\beta_1^2)^{(r+1)n_0(N-1)})
\]

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

\[
\|U - \tilde{U}\|_{\text{HS}} \lesssim_{p, \alpha, \beta_0, \beta_1} 1.
\]

We now conclude the proof of Lemma \[\text{7.4} \] on the above claims. By the same lines we used in the proof of Proposition \[\text{7.1} \] (cf. (7.16)–(7.17)) we have from Claim \[\text{7.9} \] and \[\text{7.10} \] that for \(\beta_0 \leq c_0(p, \alpha)\),

\[
\mathbb{P}\{U_1 \in \mathcal{B}_0(\beta_0) \cap \mathcal{A}(\alpha) \land \|L^2\|_{\text{op}} \leq B\} \leq \mathbb{P}\{U_1 \in \mathcal{B}_0(\beta_0) \cap \mathcal{A}(\alpha)\}
\]

\[
\leq |\mathcal{N}_0(\beta_0)|O_p\left(B^2\beta^2n_0(n_0+1)(N-1)\right)
\]

\[
\leq O_p,A,B(1)^N\mathcal{O}(\beta_0)^n_0(1-(n_0+1))N-O_p(1).
\]

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 \(\beta_0 = \beta_0(p, \alpha, B) > 0\) such that

\[
\mathbb{P}\{U_1 \in \mathcal{B}_0(\beta_0) \cap \mathcal{A}(\alpha) \land \|L^2\|_{\text{op}} \leq B\} \lesssim_p e^{-N}.
\]

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

\[
\mathbb{P}\{U_1 \in \mathcal{B}(k_1, \ldots, k_r)(\gamma_0, \gamma_1) \cap \mathcal{A}(\alpha) \land \|L^2\|_{\text{op}} \leq B\}
\]

\[
\leq |\mathcal{N}(k_1, \ldots, k_r)(\gamma_0, \gamma_1)|O_p,A,\gamma_0(2\gamma_0^2n_0(n_0+1)+r)(N-1)
\]

\[
\leq O_p,A,B,\gamma_0(1)^N\mathcal{O}(\gamma_0^2n_0(n_0+1)+r)(n_0+1)N-O_p(1).
\]

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 \(\gamma_1(p, \alpha, B, \gamma_0) \in (0, \gamma_0)\) such that

\[
\mathbb{P}\{U_1 \in \mathcal{B}(k_1, \ldots, k_r)(\gamma_0, \gamma_1) \cap \mathcal{A}(\alpha) \land \|L^2\|_{\text{op}} \leq B\} \lesssim_p e^{-N}.
\]
Now we recursively obtain a sequence $\beta_0 \geq \beta_1 \geq \ldots \geq \beta_{n_0} =: \beta_* > 0$ with $\beta_0 = \beta_0(p, \alpha, B)$ as in (7.32), and $\beta_k = \gamma_1(p, \alpha, B, \beta_{k-1})$ for $1 \leq k \leq n_0$. We cover

$$B(\beta_*) \cap B_* \subseteq \bigcup_{k_1, \ldots, k_{n_0} \in \{0, n_0\}} B((\beta_0, \beta_1) \cup B((\beta_1, \beta_2) \cup \cdots \cup B((\beta_{k_1}, \ldots, k_{n_0}))(\beta_{k-1}, \beta_{n_0}).$$

Applying the union bound followed by (7.32) and (7.33) with $1 \leq k \leq n_0$, we obtain

$$\mathbb{P}\{U_1 \in B(\beta) \cap A(\alpha) \wedge \|A\|_{\text{op}} \leq B \} \leq e^{-N}$$

for any $0 < \beta \leq \beta_*$. \hfill \Box

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 $\sigma_{\text{min}}(A) \geq \alpha > 0$, $\|b - b'\| \leq \varepsilon$, and $\|a_i - a'_i\|_2 \leq \varepsilon$ for each $i \in [k]$. If $\varepsilon \leq \frac{1}{2}k^{-1/2}\alpha$, then

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

**Proof.** From the triangle inequality it suffices to show

$$|\text{dist}(b, \text{Span}(a_1, \ldots, a_k)) - \text{dist}(b, \text{Span}(a'_1, \ldots, a'_k))| \leq k\varepsilon\|b\|_2/\alpha. \quad (7.34)$$

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

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

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 $j$th one. Let $\tilde{b}, \hat{a}_j, \tilde{a}_j'$ denote the projections of $b, a_j, a'_j$ to $V^\perp$. Then the left hand side above is

$$|\text{dist}(\tilde{b}, \langle \hat{a}_j \rangle) - \text{dist}(\tilde{b}, \langle \tilde{a}_j' \rangle)| = \frac{\|\tilde{b} \wedge \hat{a}_j\|_2 - \|\tilde{b} \wedge \tilde{a}_j'\|_2}{\|\hat{a}_j\|_2} \leq \frac{1}{\|\hat{a}_j\|_2} \|\tilde{b} \wedge \hat{a}_j\|_2 - \|\tilde{b} \wedge \tilde{a}_j'\|_2 \leq \frac{1}{\|\hat{a}_j\|_2} \|b\|_2 \varepsilon + \frac{\varepsilon}{\|\hat{a}_j\|_2} \|\tilde{b}\|_2 \leq \frac{2\varepsilon\|b\|_2}{\|\hat{a}_j\|_2}.$$

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

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

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

$$\text{dist}(a_j, V) = \left\|a_j - \sum_{i \neq j} c_i \text{col}_i(A^{(j)}) \right\|_2 \geq \sigma_{\text{min}}(A^{(j)})(1 + \sum_{i \neq j} |c_i|^2)^{1/2} \geq \sigma_{\text{min}}(A^{(j)}).$$
Now since \( \|A^{(j)} - A\|_{\text{op}} \leq \|A^{(j)} - A\|_{HS} \leq \varepsilon \sqrt{k} \) for all \( 1 \leq j \leq k \),
\[
\sigma_{\min}(A^{(j)}) \geq \sigma_{\min}(A) - \|A^{(j)} - A\|_{\text{op}} \geq \sigma_{\min}(A) - \varepsilon \sqrt{k}
\]
and \( (7.36) \) follows from the previous two displays and our assumption on \( \varepsilon \). \( \square \)

**Proof of Claim 7.10.** The key property of \( W = W(U) \) is that the columns of \( Q \) are determined by those of \( U \). Indeed, let \( S \in \mathbb{M}^{[0,n_0]}(\mathbb{C}) \) have columns \( s_0, \ldots, s_{n_0} \) (cf. (6.3)) and let \( \tilde{S} \in \mathbb{M}^{[0,n_0]}(\mathbb{C}) \) be the result of adding the standard basis vector \( e_1 = (1, 0, \ldots, 0) \in \mathbb{C}^n \) as zeroth row. We have
\[
\begin{pmatrix} U^0 \\ Q \end{pmatrix} = (\tilde{S} \otimes I_N)U, \quad \text{in particular} \quad Q = (S \otimes I_N)U. \tag{7.37}
\]
Since \( s_1, \ldots, s_{n_0} \) are linearly independent, it follows that \( \tilde{S} \) has full column rank, so
\[
U = (\tilde{S}^+ \otimes I_N)^{-1}U, \quad \text{where} \quad \tilde{S}^+ = (\tilde{S}^* \tilde{S})^{-1} \tilde{S}^* \text{is the Moore–Penrose pseudoinverse of } \tilde{S}.
\]

For \( U^0 = (u_0^0, \ldots, u_{n_0}^0) \in \mathbb{M}^{[0,n]}(\mathbb{C}) \) and \( u_1', \ldots, u_r' \in \mathbb{C}^{[n_0] \times [N]} \) with \( r \leq n_0 \), we write \( W^{(r)}(U^0, u_1', \ldots, u_r') \) for the span of the columns of \( R(U^0) \) together with \( (S \otimes I_N)u_1, \ldots, (S \otimes I_N)u_r \), where
\[
u_k = \begin{pmatrix} u_k^0 \\ u_k' \end{pmatrix}.
\]
In particular, for \( I = \{k_1, \ldots, k_r\} \subset [0, n_0] \) we have
\[
W_I(U) = W^{(r)}(U^0, u_{k_1}', \ldots, u_{k_r}'). \tag{7.39}
\]
We introduce this notation to make it clear that this subspace is fixed by the partial data \( U^0, u_{k_1}', \ldots, u_{k_r}' \), which will be crucial for the net construction.

For any \( \beta \in (0, 1/2) \) we let \( \Sigma_0(\beta, \alpha) \) be a \( \beta \)-net for the set
\[
\{ U \in 2\sqrt{n_0 + 1} \cdot B^{[0,n_0]}_{[N]} : \sigma_{\min}(U) \geq \alpha \}
\]
and let \( \Sigma_1(\beta) \) be a \( \beta \)-net for \( B^{[n_0] \times [N]} \) (the unit ball in \( \mathbb{C}^{[n_0] \times [N]} \)). (Both nets are taken with respect to the Euclidean metric.) For each \( \tilde{U}^0 \in \Sigma_0(\beta, \alpha) \), \( \tilde{u}_1', \ldots, \tilde{u}_r' \in \Sigma_1(\beta) \) and any \( R \geq 1 \), we let \( \Sigma^{(r)}_{\tilde{U}^0, \tilde{u}_1', \ldots, \tilde{u}_r'}(\beta, R) \) be a (Euclidean) \( \beta \)-net for the ball of radius \( R \) in \( W^{(r)}(\tilde{U}^0, \tilde{u}_1', \ldots, \tilde{u}_r') \). For \( r = 0 \) we let \( \Sigma^{(0)}_{\tilde{U}^0}(\beta, R) \) be a \( \beta \)-net for the ball of radius \( R \) in \( W^{(0)}(\tilde{U}^0) \). By Lemma 3.1 we may choose these nets so that
\[
|\Sigma_0(\beta, \alpha)| \leq O_p(1/\beta^2)(n_0 + 1)^N, \tag{7.40}
\]
\[
|\Sigma_1(\beta)| \leq O(1/\beta^2)n_0N, \tag{7.41}
\]
\[
|\Sigma^{(r)}_{\tilde{U}^0, \tilde{u}_1', \ldots, \tilde{u}_r'}(\beta, R)|, \quad |\Sigma^{(0)}_{\tilde{U}^0}(\beta, R)| \leq O(R^2/\beta^2)O_p(1). \tag{7.42}
\]

Now let \( 1 \leq r \leq n_0 \) and fix distinct \( k_1, \ldots, k_r \in [0, n_0] \). In the sequel we write \( I = \{k_1, \ldots, k_r\} \).

Let \( \mathcal{N} \) be the set of \( \tilde{U} \in \mathbb{B}_* \) such that
\[
\tilde{U}^0 \in \Sigma_0(\beta_1, \alpha), \quad \tilde{u}_k' \in \Sigma_1(\beta_1) \quad \forall k \in I_r, \tag{7.43}
\]
and
\[
\tilde{u}_k' \in \left\{ (S' \otimes I_N) \begin{pmatrix} \tilde{q}_k^0 \\ \tilde{q}_k' \end{pmatrix} : \tilde{q}_k \in \Sigma_{\tilde{U}^0, (\tilde{u}_k')_{k \in I_r}}(\beta_1, R) \right\} \quad \forall k \notin I_r, \tag{7.44}
\]
with \( R = O_p(1) \) to be taken sufficiently large, and where \( S' \) is the matrix \( \tilde{S}^+ \) with its first row removed (see \( (7.38) \)).
Fix an arbitrary $U \in \tilde{B}_{(k_1, \ldots, k_r)}(\beta_0, \beta_1) \cap A(\alpha)$. We claim there exists $\tilde{U} \in N$ such that (7.31) holds and $\tilde{U} \in B_{(k_1, \ldots, k_r)}(\beta_0/2)$. From this the claim will follow by taking $N_{(k_1, \ldots, k_r)}(\beta_0, \beta_1) = N \cap B_{(k_1, \ldots, k_r)}(\beta_0/2)$, noting that $\tilde{U} \in A(\alpha)$ by (7.43), and that the bound (7.30) follows from (7.40)–(7.42).

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

$$\|U^0 - \tilde{U}^0\|_2 \leq \beta_1 \quad \text{and} \quad \|u_k' - \tilde{u}_k'\|_2 \leq \beta_1 \quad \forall k \in I_r. \quad (7.45)$$

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

$$\|u_k' - \tilde{u}_k'\|_2 \lesssim_{p, \alpha, \beta_0} \beta_1 \quad (7.46)$$

and so that the resulting matrix $\tilde{U}$ lies in $B_{(k_1, \ldots, k_r)}(\beta_0/2)$.

For arbitrary $k \in [0, n_0]$, since $q_k = (S \otimes I_N)u_k$ and $\|u_k\|_2 \leq 2\sqrt{n_0} + 1$, it follows that

$$\|q_k\|_2 \leq 2\sqrt{n_0 + 1}\|S\|_{op} \lesssim_p 1. \quad (7.48)$$

For $k \in I_r$, with

$$\tilde{q}_k := (S \otimes I_N)\tilde{u}_k, \quad \tilde{u}_k = \left(\begin{array}{c} \tilde{u}_k^0 \\ \tilde{u}_k' \end{array}\right),$$

we similarly have

$$\|q_k - \tilde{q}_k\|_2 \lesssim_p \beta_1 \quad \forall k \in I_r. \quad (7.47)$$

Since $U \in \tilde{B}_{(k_1, \ldots, k_r)}(\beta_0, \beta_1)$,

$$\text{dist}(q_k, W_{I_{i-1}}) = \text{dist}(q_k, W^{(i-1)}(U^0, u_{k_1}', \ldots, u_{k_i-1}')) \geq \beta_0 \quad \forall 1 \leq i \leq r \quad (7.48)$$

and for any $k \notin I_r$,

$$\text{dist}(q_k, W_{I_r}) = \text{dist}(q_k, W^{(r)}(U^0, u_{k_1}', \ldots, u_{k_r}')) \leq \beta_1 \quad (7.49)$$

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

$$\sigma_{\text{min}}(W_{I_{i-1}}) \gtrsim_{\alpha, \beta_0} 1 \quad \forall 1 \leq i \leq r \quad (7.48)$$

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

$$\text{dist}(\tilde{q}_k, W^{(i-1)}(\tilde{U}^0, \tilde{u}_k', \ldots, \tilde{u}_{k_i-1}')) \geq \beta_0 - O_{p, \alpha, \beta_0}(\beta_1) \quad \forall 1 \leq i \leq r. \quad (7.50)$$

We similarly have

$$\text{dist}(q_k, W^{(r)}(\tilde{U}^0, \tilde{u}_k', \ldots, \tilde{u}_{k_r}')) \lesssim_{p, \alpha, \beta_0} \beta_1 \quad \forall k \notin I_r. \quad (7.51)$$

By the triangle inequality, taking $R = O_p(1)$ sufficiently large, there exist $\tilde{q}_k \in \Sigma_{U^0, \tilde{u}_{k_1}', \ldots, \tilde{u}_{k_r}'}(\beta_1, R)$ such that

$$\|q_k - \tilde{q}_k\|_2 \lesssim_{p, \alpha, \beta_0} \beta_1 \quad \forall k \notin I_r. \quad (7.52)$$

Finally, setting $\tilde{u}_k' = (S' \otimes I_N)\tilde{q}_k$, we have for each $k \notin I_r$ that

$$\|u_k' - \tilde{u}_k'\|_2 = \| (S' \otimes I_N) \left( q_k(U) - \tilde{q}_k \right) \|_2 \lesssim_p \left( \left\| q_k(U) - \tilde{q}_k \|_2 \right\|_2 \lesssim_{p, \alpha, \beta_0} \beta_1 \right) \quad (7.53)$$

giving (7.46) as desired. That $\tilde{U} \in B_{(k_1, \ldots, k_r)}(\beta_0/2)$ follows from (7.50) and taking the constant $c_1(p, \alpha, \beta_0)$ sufficiently small.

The proof of Claim 7.9 follows similar lines. The net $N_{\emptyset}(\beta_1)$ is taken as in (7.43)–(7.44) with $I_r = \emptyset$ (so no columns $u_k'$ are fixed at this initial stage). The estimates (7.49) and (7.51) are obtained by the same lines, with the subspaces $W_{\emptyset} = W^{(0)}(U^0)$ and $W^{(0)}(\tilde{U}^0)$, while we skip (7.48) and (7.50).
Lemma 7.12. Let $A \in M_m(\mathbb{C})$ with $|\det A| \geq \kappa$ and columns $x_1, \ldots, x_m \in \mathbb{B}^m$. Suppose that for some $v \in \mathbb{C}^m$ we have
$$|\det((v, A_J))| < \delta \quad \forall J \in \binom{[m]}{m-1}$$
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
$$\delta > |\det((v, A_J))| = |b_j| |\det A| \geq \kappa |b_j| \quad \forall j \in [m],$$
so $\|b\|_2 < \sqrt{m\delta}/\kappa$. From our assumptions we have $\|A\|_{\text{op}} \leq \|A\|_{\text{HS}} \leq \sqrt{m}$, and so
$$\|v\|_2 = \|Ab\|_2 \leq \sqrt{m} \|b\|_2 < m\delta/\kappa$$
as desired. $\square$

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 $U \in A(\alpha^2)$, there exists $I^* = \{i_0, \ldots, i_{n_0}\} \subset [N]$ such that
$$\|v_{i_0}^0 \wedge \cdots \wedge v_{i_{n_0}}^0\|_2 \geq (\alpha/\sqrt{N})^{k+1}$$
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 $\sigma_{\text{min}}(U^0) \leq \|U^0\|_{\text{op}} \leq \|U^0\|_{\text{HS}} < \alpha$. Now for $1 \leq k \leq n_0 - 1$, having picked $i_0, \ldots, i_k$, there exists $i_{k+1} \in [N]$ such that $\text{dist}(v_{i_{k+1}}^0, \langle v_{i_0}^0, \ldots, v_{i_k}^0 \rangle) \geq \alpha/\sqrt{N}$, as otherwise, for any unit vector $u \in \langle v_{i_0}^0, \ldots, v_{i_k}^0 \rangle^\perp$ we would have
$$\|U^0 u\|_2^2 = \sum_{j \in [N]\setminus\{i_0, \ldots, i_k\}} |\langle v_j^0, u \rangle|^2 < \alpha^2,$$
which contradicts $\sigma_{\text{min}}(U^0) \geq \alpha$. (7.53) now follows from the base-times-height formula for the norm of the wedge product. $\square$

Fix an arbitrary $k_0 \in [n_0]$. Our proof of Lemma 7.5 is divided into two cases depending on $p$:
(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 $J \subset [N]$, we let $\Sigma_J(R, \rho) \subset M_{[N]}^{[0,n_0]}(\mathbb{C})$ be a $\rho$-net for the Hilbert–Schmidt ball of radius $R$ on the subspace of matrices supported on the rows indexed by $J$. (We take $\Sigma_0(R, \rho)$ 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
$$|\Sigma_0(R, \rho)| = O(R^2 N/\rho^2)^{n_0+1}, \quad |\Sigma_J(R, \rho)| = O(R^2/\rho^2)^{(n_0+1)|J|}. \quad \text{(7.54)}$$

Proof of Lemma 7.2 under Case A. Let $\mathcal{N}'$ be the set of matrices $U = (U^0, \ldots, U^{n_0}) \in M_{[0,n_0]}^{[0,n_0] \times [N]}(\mathbb{C})$ such that $U^k \in \Sigma_{[N]}(\sqrt{n_0 + 1}, \rho)$ for each $k \in [0, n_0] \setminus \{k_0\}$, and $U^{k_0}$ has rows given by
$$v_{j}^{k_0} = -\frac{1}{s_{k_0,\ell_0}} \sum_{k \in [0, n_0] \setminus \{k_0\}} s_{k,\ell_0} v_{j}^{k}, \quad j \in [N], \quad \text{(7.55)}$$
remains to show

Now for each \( k \neq k_0 \), (Recall that \( \hat{U}^{k_0} \) is determined by (7.60) once \( U^k \) for \( k \neq k_0 \) is fixed.) It only remains to show

\[
\|U^{k_0} - \hat{U}^{k_0}\|_{\text{HS}} \lesssim \rho_\ast.
\] (7.58)

Let \( I^c = \{i_0, \ldots, i_{n_0}\} \subset [N] \) satisfying (7.53) for \( U \). Since \( U \in \text{Struct}(\delta) \subseteq \text{Struct}_1(\delta) \) we have

\[
|\Delta_{j,I}| := |\det(w_j, v^0_j)| < \delta \quad \forall j \in [N], \forall I \in \binom{[N]}{n_0}
\]

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

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

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

\[
\|v^{k_0}_j - \hat{v}^{k_0}_j\|_2 = \frac{1}{|s_{k_0,\ell_0}|^2} \|w_j - \sum_{k \in [0,n_0] \setminus \{k_0\}} s_{k,\ell_0} (v^{k}_j - \hat{v}^{k}_j)\|_2 \\
\lesssim_{\rho,\alpha} \delta^2 N^{n_0+1} + \sum_{k \in [0,n_0] \setminus \{k_0\}} \|v^{k}_j - \hat{v}^{k}_j\|_2^2.
\]

Summing over \( j \) and applying the bounds \( \|U^k - \hat{U}^{k}\|_{\text{HS}} < \rho_\ast \) for \( k \neq k_0 \) and our assumed bound on \( \delta \), we obtain (7.58) as desired.

\( \square \)

Proof of Lemma 7.3 under Case B

Let \( \Lambda_0 \) be a \( \rho_\ast/\sqrt{N} \)-mesh for the interval \([-\bar{R}, \bar{R}]\), for some \( R = O_p(1) \) to be taken sufficiently large. We take \( N'' \) be the set of matrices \( \hat{U} = (\hat{U}^0, \ldots, \hat{U}^{n_0}) \in \mathcal{M}^{[0,n_0]}_{[0,n_0] \times [N]}(\mathbb{C}) \) such that

\[
\hat{U}^k \in \Sigma_{[N]}(\sqrt{n_0 + 1}, \rho_\ast) \quad \forall k \in [n_0] \setminus \{k_0\},
\] (7.59)

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

\[
\hat{U}^0_{j_0} \in \Sigma_{j_0}(\rho_1^{1/2}, \rho_\ast) \quad \text{and} \quad \hat{U}^{k_0}_{j_0} \in \Sigma_{j_0}(\sqrt{n_0 + 1}, \rho_\ast)
\] (7.60)

while for each \( j \in J \),

\[
\hat{v}^{0}_j \in \Sigma_0(1, \rho_1), \quad \rho_1 := \rho_\ast^{3/2}/N
\] (7.61)

and

\[
\hat{v}^{k_0}_j = \frac{1}{s_{k_0,\ell_0}} \left( \hat{w}_j - \sum_{k \in [0,n_0] \setminus \{k_0\}} s_{k,\ell_0} \hat{v}^{k}_j \right)
\] (7.62)

for some \( \hat{w}_j \in \Lambda_0 \cdot (\hat{v}^{0}_j) \). Here we write \( J^c := [N] \setminus J \), and \( \hat{U}^k \) for the matrix obtained from \( \hat{U}^k \) by zeroing out the rows with indices in \( I^c \).
Constructing $N \subset {\mathcal{E}}$ from $N''$ analogously to how $N$ was obtained from $N'$ in the proof for Case [A] we have by construction that

$$|N| \leq |N''|$$

$$\leq |\Sigma_{|N|}(\sqrt{n_0+1}, \rho_*)|^{n_0-1} \sum_{j \in [N]} |\Sigma_{j}(\rho_*^{1/2}, \rho_*)||\Sigma_{j}(\sqrt{n_0+1}, \rho_*)||\Sigma_0(1, \rho_1)||J|\Lambda_0||J|$$

$$= O_p(1/\rho_*^2)^{(n_0-1)(n_0+1)}N \sum_{j \in [N]} O_p(1/\rho_*)^{(n_0+1)(N-J)}O_p(1/\rho_*^2)^{(n_0+1)(N-J)}$$

$$= O_p(1)^N O(1/\rho_*^2)^{(n_0+1/2)(n_0+1)}N \sum_{j \in [N]} O(N)^{2n_0+3}/|J|O(1/\rho_*^2)^{|J|}$$

$$= O_p(N)^{(2n_0+3)}O(1/\rho_*^2)^{(n_0^2+3/2)n_0+3/2}N$$

where in the last line we simply bounded $|J| \leq N$ 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 $U \in {\mathcal{E}}$ there exists $\hat{U} \in N''$ such that (7.57) holds. Fixing now an arbitrary $U \in {\mathcal{E}}$, for each $k \in [n_0] \setminus \{k_0\}$ we let $\hat{U}^k$ be as in (7.59) with $\|U^k - \hat{U}^k\|_{HS} < \rho_*$. It only remains to pick $\hat{U}^0, \hat{U}^{k_0}$ as in (7.60)–(7.62) for some choice of $J = J(U \in [N]$ with

$$\|U^k - \hat{U}^k\|_{HS} \lesssim \rho_*, \quad k \in \{0, k_0\}. \quad (7.63)$$

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

$$J = J(U^0) = \{j \in [N] : \|v^0_j\|_2 \geq \rho_*^{1/2}/\sqrt{N}\}.$$ 

We have $\|U^0_{j_0}\|_{HS} < \rho_*^{1/2}$, and so there exists $\hat{U}^0_{j_0}$ as in (7.60) with $\|U^0_{j_0} - \hat{U}^0_{j_0}\|_{HS} \leq \rho_*$. Completing $\hat{U}^0$ with rows $\hat{v}^0_{j_0}$ as in (7.61) such that $\|v^0_j - \hat{v}^0_j\|_2 \leq \rho_1$ for each $j \in J$, we obtain (7.63) for $k = 0$.

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

$$\sum_{j \in J} \|v^{k_0}_j - \hat{v}^{k_0}_j\|_2^2 \lesssim_{\rho_1, \rho_*} |J|.$$ 

(7.64)

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

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

(7.65)

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

$$|v^0_j \wedge x_{\wedge I^*}| \geq \kappa \rho_*^{1/2}/m\sqrt{N}.$$ 

(7.66)

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

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

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

$$\delta > |w_j \wedge v^0_j \wedge x_{\wedge I}| = \|v^0_j\|_2 \|\bar{w}_j \wedge \bar{x}_{\wedge I}\|_2 \geq \rho_*^{1/2}/\sqrt{N} \|\bar{w}_j \wedge \bar{x}_{\wedge I}\|_2 \quad \forall I \in \binom{\hat{I}}{m-2}.$$ 

(7.67)
On the other hand, since $\|v_j^0\|_2 \leq 1$, we have from (7.66) that
\[ \|x_j^{\lambda}\|_2 \geq |v_j^0 \wedge x_j^{\lambda}| \geq \frac{\kappa \rho_*^{1/2}}{m \sqrt{N}}. \]
Together with (7.67) and Lemma 7.12 (identifying the subspace $\langle v_j^0 \rangle$ with $\mathbb{C}^{m-1}$) this implies
\[ \text{dist}(w_j, \langle v_j^0 \rangle) = \| \tilde{w}_j \|_2 < \frac{m(m-1)\delta N}{\kappa \rho_*} \leq \delta N^{(n_0+3)/2}/\rho_* \lesssim \rho_*/\sqrt{N} \]
using our assumption on $\rho_*$ and recalling $m = n_0 + 1 = O_0(1)$. Now from the estimates
\[ \|w_j\|_2 = O_p(1), \quad \|v_j^0\|_2 \geq \frac{\rho_*^{1/2}}{\sqrt{N}}, \quad \|v_j^0 - \tilde{v}_j^0\|_2 \leq \rho_1 = \rho_*^{3/2}/N, \]
the identity
\[ \text{dist}(w_j, \langle v_j^0 \rangle) = \frac{\|w_j \wedge v_j^0\|_2}{\|v_j^0\|_2} \]
and multilinearity of the wedge product, we have
\[ \text{dist}(w_j, \langle \tilde{v}_j^0 \rangle) = \text{dist}(w_j, \langle v_j^0 \rangle) + O_p(\rho_*/\sqrt{N}) \lesssim \rho_*/\sqrt{N}. \]
Since $\|w_j\|_2 = O_p(1)$, by taking $R$ sufficiently large we have that for every $j \in J$ there exists $\tilde{w}_j \in \Lambda_0 \cdot \tilde{v}_j^0$ such that
\[ \|w_j - \tilde{w}_j\|_2 \lesssim \rho_*/\sqrt{N}. \tag{7.68} \]
Taking $\tilde{v}_j^{k_0}$ 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
\[ \|v_j^0 - \tilde{v}_j^{k_0}\|_2 \lesssim \rho_1 \|w_j - \tilde{w}_j\|_2 + \sum_{k \in \{n_0\}\setminus\{k_0\}} \|v_j^k - \tilde{v}_j^k\|_2. \]
Substituting (7.68), summing over $j \in J$, and applying the bounds $\|U^k - \tilde{U}^k\|_{\text{HS}} \lesssim \rho_*$ for $k \neq k_0$ yields (7.64) and completes the proof. \qed

**Remark 7.14.** Let us explain why we cannot take any linearization for $p$. Consider a linearization of the form
\[ L^* = \begin{pmatrix} -z + T & R_1 & \cdots & R_{n_0} \\ Y_1 & -1 & \cdots & 0 \\ \vdots & \ddots & \ddots \\ Y_{n_0} & \cdots & -1 \end{pmatrix} \]
(compare (4.6)) where the $R_k$, $Y_k$ and $T$ are linear forms in the matrices $X_1, \ldots, X_n$ with respective coefficient vectors $r_k$, $s_t$, $t \in \mathbb{C}^n$. If the family $\{r_1, \ldots, r_{n_0}\}$ 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 $\tilde{x} = (0, x_1, \ldots, x_{n_0})$, the matrices $L_i$ as in (6.3) (with $\langle r_k, \xi_i \rangle$ in place of $\xi_i^k$) all satisfy $L_i \tilde{x} = 0$. When we look for the coefficients of the terms of order $n_0 + 1$ of $\det(M + \sum_{i=1}^{n_0} U_i L_i)$, we find that these are the same as in $\det(\sum_{i=1}^{n_0} U_i L_i)$, and hence they are zero. On the other hand, if $\{s_1, \ldots, s_{n_0}\}$ is not of full rank, then we 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 $\sum x_k \text{row}_1(Y_k) = 0$. Then considering the vector $u$ with first block $u^0 = 0$ and $u^k(1) = x_k$ for $k = 1, \ldots, n_0$, $u^k(i) = 0$ for all $k$ and all $i \in [2, N]$, we have that $u$ is a non-null element of $(V_{(1)}^z)^k$, so that $\sigma_{\min}(U^0) = 0$. And so we cannot have our lower bound on the effective rank $R(U^0)$ (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 $p$ 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 $p = \sum_{i=1}^{n_0} R_i S_i T_i$ where the families $\{R_i\}_{1 \leq i \leq n_0}$ and $\{T_i\}_{1 \leq i \leq n_0}$ 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

$$
\nu_N^z := \mu(z I_N - P^N)(z I_N - P^N)^*,
$$

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

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

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 $\nu_N^z$ converges weakly almost surely to some identifiable probability measure $\nu^z$ 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 $\mu_{P^N}$ converges in distribution to a limit that we identify as $\nu_{p(c_1, \ldots, c_n)}$.

Proof of (1): Because $(z I_N - P^N)(z I_N - P^N)^*$ 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 $\nu^z$ is the distribution of $|z - p(c_1, \ldots, c_n)|^2$ in the sense of $*$-moments, and from the boundedness of the circular elements it follows that $\nu^z$ 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

$$
\limsup_{N \to \infty} \|P^N\| \leq B \quad \text{a.s.}
$$

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

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

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

$$
\limsup_{N \to \infty} \mathbb{P}\left\{ \left| \int_0^\epsilon \log(x) d\nu_N^z(x) \right| \geq \delta \right\} \leq \delta'.
$$

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

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

On the event $\mathcal{G}_N$,\n
$$
\int_0^\epsilon \log(x) d\nu_N^z(x) = \int_0^\epsilon \log(x) d\nu_N^z(x),
$$

so we only need to show

$$
\lim_{\epsilon \to 0} \limsup_{n \to \infty} \mathbb{E}\left| \int_0^\epsilon \log(x) d\nu_N^z(x) \right| = 0.
$$

(8.4)
We denote \( \tilde{\nu}_n^z = \mathbb{E}[\nu_N^x] \) and \( g \) (resp. \( g_N \)) the Stieltjes transform of \( \nu^z \) (resp. \( \nu_N^z \)) given for \( \zeta \in \mathbb{C} \) by

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

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 \( x \in [N^{-c_1}, 1] \),

\[
F_N^z(x) := \tilde{\nu}_N^z([0, x]) \leq 2 \int_{[0,x]} \frac{x^2}{y^2 + x^2} \, dv_N^z(y) \leq 2x^3 g_N^z(ix) \leq 2C x^{1-c_2}.
\]

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

\[
\mathbb{E}\left[ \int_{N^{-\beta}}^\epsilon \log x \, dv_N^z(x) \right] \\
\leq - \int_{N^{-\beta}}^{N^{-\alpha}} \log x \, dv_N^z(x) - \int_{N^{-\alpha}}^\epsilon \log x \, dv_N^z(x) \\
\leq \beta(\log N)\tilde{\nu}_N^z([0, N^{-\alpha}]) + \int_{N^{-\alpha}}^\epsilon \frac{1}{x} F_N^z(x) \, dx - \log(\epsilon) F_N^z(\epsilon) + F_N^z(N^{-\alpha})(\log(N^{-\alpha})) \\
\leq 2C\beta(\log N)N^{-\alpha(1-c_1)} + 2C \frac{\epsilon^{1-c_2}}{1 - c_2} - 2(\log \epsilon)\epsilon^{1-c_2},
\]

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 \( \nu_N^z \) close to the real axis. Indeed, recall that Ginibre matrices can be decomposed as the sum of two independent GUE matrices: \( X^z = (Y_j^+ + iZ_j^N)/\sqrt{2} \) 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 [40, Lemma 5.5.4], implying that there exists \( c_1 \) finite such that for \( 3 \zeta \in [N^{-c_1}, 1] \) and \( N \) large enough,

\[
|g^z(\zeta) - g_N^z(\zeta)| \leq \frac{c_2}{N^2(3\zeta)^{c_3}}.
\]

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

\[
\Im g^z(ie^\theta) \leq K e^\theta
\]

for some \( q > -1 \) and \( K > 0 \). Following [40, Corollary 1.2], \( \nu^z \) has no atoms. Moreover, by [40, 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 \( \zeta \) close to the origin. But clearly, since \( \nu^z \) has no atoms, \( q > -1 \).

A more quantitative proof could have used that [7] implies that the partition function of \( \nu^z \) is Hölder with exponent 1/15.

**Proof of [7]:** Denoting

\[
h_N(z) := \int_0^\infty \log |x| \, d\nu_N^z(x) \quad \text{and} \quad h(z) = \int_0^\infty \log |x| \, dv^z(x)
\]

for \( \epsilon > 0 \),

\[
\mathbb{E}\left[ \int_{N^{-\beta}}^\epsilon \log x \, dv_N^z(x) \right] \\
\leq - \int_{N^{-\beta}}^{N^{-\alpha}} \log x \, dv_N^z(x) - \int_{N^{-\alpha}}^\epsilon \log x \, dv_N^z(x) \\
\leq \beta(\log N)\tilde{\nu}_N^z([0, N^{-\alpha}]) + \int_{N^{-\alpha}}^\epsilon \frac{1}{x} F_N^z(x) \, dx - \log(\epsilon) F_N^z(\epsilon) + F_N^z(N^{-\alpha})(\log(N^{-\alpha})) \\
\leq 2C\beta(\log N)N^{-\alpha(1-c_1)} + 2C \frac{\epsilon^{1-c_2}}{1 - c_2} - 2(\log \epsilon)\epsilon^{1-c_2},
\]

which gives (8.4).
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
\[
h(z) = \int_{\mathbb{C}} \log |z - \lambda| d\nu_{\psi(e_1, \ldots, e_n)}(\lambda)
\]
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
\[
\lim_{N \to \infty} \mathbb{E} \left( \mathbbm{1}_{\mathcal{A}_N} \int_{z \in K} |h_N(z) - h(z)| dz \right) = 0. \tag{8.5}
\]
This is enough to conclude the proof of Theorem 1.2 by (8.1) for any twice continuously differentiable function \( \psi \) 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
\[
\mathbb{E} \left[ \mathbbm{1}_{\mathcal{A}_N} \int_K |h_N(z)|^2 dz + \int_K |h(z)|^2 dz \right]
\]
is bounded independently of \( N \), so that the bounded convergence theorem applies. But, Jensen’s inequality and Fubini’s theorem imply that
\[
\mathbb{E} \left[ \mathbbm{1}_{\mathcal{A}_N} \int_K |h_N(z)|^2 dz \right] \leq \mathbb{E} \left[ \mathbbm{1}_{\mathcal{A}_N} \int_C \int_K |\log |z - \lambda||^2 dz d\mu_{P_N}(\lambda) \right] \leq \sup_{|\lambda| \leq B} \int_K |\log |z - \lambda||^2 dz
\]
is finite, and a similar estimate holds for \( h(z) \) (one obtains from the boundedness of the circular elements \( e_1, \ldots, e_n \) that \( \nu_{\psi(e_1, \ldots, e_n)} \) has compact support). Therefore, in probability \( h_N \) converge to \( h \) in \( L^1 \)(\( K \), Leb) and so \( \mu_{P_N} = \frac{1}{4\pi} \Delta h_N \) converges to \( \nu_{\psi(e_1, \ldots, e_n)} = \frac{1}{4\pi} \Delta h \) in the sense of distributions on \( K \). Again, taking \( K \) that contains the support of \( \nu_{\psi(e_1, \ldots, e_n)} \) and \( \mu_{P_N} \), the convergence in the sense of distributions implies weak convergence and the result is proved.

**References**


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