CIRCULAR LAW FOR NON-HERMITIAN BLOCK BAND MATRICES WITH SLOWLY GROWING BANDWIDTH

YI HAN

Abstract. We consider the empirical eigenvalue distribution for a class of non-Hermitian random block tridiagonal matrices T with independent entries. The matrix has n blocks on the diagonal and each block has size ℓ_n, so the whole matrix has size $n\ell_n$. We assume that the nonzero entries are i.i.d. with mean 0, variance 1 and having sufficiently high moments. We prove that when the entries have a bounded density, then whenever $\lim_{n\to\infty}\ell_n=\infty$ and $\ell_n=O(\mathrm{Poly}(n))$, the normalized empirical spectral distribution of T converges almost surely to the circular law. The growing bandwidth condition $\lim_{n\to\infty}\ell_n=\infty$ is the optimal condition of circular law with small bandwidth. This confirms the folklore conjecture that the circular law holds whenever the bandwidth increases with the dimension, while all existing results for the circular law are only proven in the delocalized regime $\ell_n\gg n$.

1. Introduction

Let A be an $N\times N$ matrix with eigenvalues $\lambda_1,\cdots,\lambda_N$. We let $\mu_A:=\frac{1}{N}\sum_{i=1}^N\delta_{\lambda_i}$ denote the empirical measure of its eigenvalues, where $\delta.$ is the delta measure. For a symmetric random matrix A with independent entries of mean 0 and variance $\frac{1}{n}$, then μ_A converges to the celebrated Wigner semicircle law [45]. For non-Hermitian random matrices, proving the convergence of μ_A is a much more challenging task because the method of moments no longer applies. For a random matrix A with i.i.d. entries, circular law was proven through a long list of partial results [1], [40], [18] until Tao and Vu [41] obtained the optimal condition. Later, the convergence of limiting spectral distribution is proven for many other random matrix ensembles, most notably for heavy-tailed random matrices [3], sparse directed graphs [29], and sparse i.i.d. matrices [33], [36].

Much less is known about the convergence of μ_A for a non-Hermitian random matrix with a highly structured variance profile, and the technical reason behind it is that least singular value estimates are notoriously hard to obtain when the matrix does not have a flat variance profile. The most important examples in this category are non-Hermitian random band matrices, which can simply be obtained from a Hermitian band matrix by making every entry independently distributed. However, there is a crucial difference between a Hermitian and a non-Hermitian band matrix model: assuming that the entries have a density, then the random potentials on the diagonal immediately provide a least singular value estimate for the Hermitian model (see for instance [38], [32]) which is polynomially small in the dimension. For non-Hermitian band matrices however, if we take the Hermitization, we cannot apply the same proof because the diagonal entries of the Hermitization are all zero! Thus very little is known in the latter setting, even for existence of the limiting density.

The most crucial parameter that governs a band matrix is its bandwidth versus size. For a band matrix A of size N with bandwidth W, the spectral properties of A, including the convergence of μ_A, depend in a crucial way on the relative magnitude of W and N. The study of circular law on non-Hermitian band matrices was initiated in [11] and [27], but these works focus on the range $W \sim N$ or $W \geq N^{1-c}$ for a small $c > 0$. This is because the proof techniques use properties that hold in the mean-field (or Ginibre) case that degenerates when W gets smaller. In the ICM survey [44], the study of limiting ESDs for inhomogeneous random matrices was also listed as a major open problem, about which we currently have little understanding. Recently, the threshold in W for circular law convergence has been pushed much further in [21], where now the sufficient condition is that $W \geq N^{1/2+c}$ for any $c > 0$ whenever the variance profile satisfies a certain regularity condition. The scale $W \sim \sqrt{N}$ is the critical scale separating localized and delocalized regimes of band matrix models, and the existing machinery cannot prove circular law deep into the localized regime. This leaves open the question whether μ_A also converges when W is much smaller, say $W \sim N^\alpha$ for any $\alpha > 0$. For very small α, all the methods in these papers do not apply.

In this work, we push beyond the current boundary for the circular law by showing that for a special class of models, the circular law holds whenever W is slowly growing in N. This confirms the widespread belief that the ESD should converge to the circular law whenever W grows, and provides the first instance of circular law proof in any part of the localized regime with growing W: $\Omega(1) \leq W \leq \sqrt{N}$.

1.1. **Models and main results.** Let A_n be an ensemble of random matrices with empirical eigenvalue density μ_{A_n}, defined on a common probability space . Let μ_c be the uniform probability measure on the unit disk in the center of the complex plane (i.e., the circular law). Then we say μ_{A_n} converges almost surely (resp., in probability) to μ_c if, for any smooth and compactly supported $f : \mathbb{C} \to \mathbb{R}$, the following expression

$$\int_{\mathbb{C}} f(z) d\mu_{A_n}(z) - \int_{\mathbb{C}} f(z) d\mu_c(z)$$

converges to 0 almost surely (resp., in probability).

Theorem 1.1. *Let ζ be a random variable satisfying that $\mathbb{E}[\zeta] = 0, \mathbb{E}[|\zeta|^2] = 1$, and having all moments finite: for all $p \geq 2$, $\mathbb{E}[|\zeta|^p] < c_p < \infty$ for a $c_p > 0$. We further assume that*

- *Either ζ is a real-valued random variable with distributional density on $\mathbb{R}$ bounded by some $L > 0$;*
- *or ζ has independent real and imaginary parts $\Re\zeta, \Im\zeta$, and at least one of $\Re\zeta, \Im\zeta$, has a distributional density on $\mathbb{R}$ bounded by some $L > 0$.*

Take two integers $n, \ell \in \mathbb{N}$ and we denote $\ell = \ell_n$. Consider the following block tridiagonal matrix T

$$T = \begin{bmatrix} A_1 & B_1 & & & \\ C_2 & A_2 & B_2 & & \\ & \ddots & \ddots & \ddots & \\ & & C_{n-1} & A_{n-1} & B_{n-1} \\ & & & C_n & A_n \end{bmatrix} \tag{1.1}$$

where each block A_i, B_i, C_i has size $\ell \times \ell$ and T has size ℓn. Assume that each entry of A_i, B_i, C_i is an i.i.d. copy of $\frac{1}{\sqrt{3\ell}}\zeta$ so that the variance is normalized. Then whenever $\lim_{n\to\infty} \ell_n = \infty$ with $n \geq \ell_n^d$ for a small constant $d > 0$, then μ_T converges almost surely to the circular law.

The matrix T has dimension $N := \ell_n n$ and bandwidth $W := \ell_n$. Theorem 1.1 shows that with a continuous density, the circular law holds for μ_T whenever the bandwidth ℓ_n increases with the matrix dimension and whenever n, the number of blocks, is not too small.

The condition on ℓ_n relative to n in Theorem 1.1 is essentially optimal. First, when ℓ_n is a fixed constant, then μ_T does not converge to the circular law in general. Convergence of ESD in the special case $\ell = 1$ was studied before in [16], and the structure of the limiting density can be remarkably complex [25] despite simplicity of the model. It is fairly reasonable to expect that for any finite bandwidth $\ell = O(1)$, μ_T would have a distribution-dependent limiting density which is not the circular law. Therefore, universality kicks in only if ℓ_n is growing in n, and Theorem 1.1 justifies that this is exactly the case: the minimal condition $\ell_n \to \infty$ is sufficient to guarantee global universality to the circular law (ζ is not fixed to be a Gaussian). Our matrix T shares properties both in the ergodic side with length of transfer governed by n, and in the mean-field side governed by the square blocks of size ℓ. Universality in entry distribution arises from the mean-field side, and we have proven that in this non-Hermitian setting, even very mild mean-field mixing with scale $\ell_n \to \infty$ can restore global universality no matter how large the ergodic scale n is. Finally, we remark that the condition $n \geq \ell_n^d$ for a small $d > 0$ is also natural and almost necessary because our matrix T is not translational invariant for lacking two corner blocks, so we need n to be not too small to make the boundary effect negligible.

Remark 1.2. (The density assumption) We use the continuous density of ζ at two places of the proof: the first is to factor out the determinant as a transfer matrix recursion in Corollary 2.4, which uses invertibility of off-diagonal blocks B_k. When $\ell \gg \log n$, Bernoulli entries are admissible since almost surely all the matrices are nonsingular. But when $\ell \ll \log n$, almost surely many matrix blocks are singular and the determinant factorization does not hold. The second and more crucial usage is in the proof of the least singular value estimate for the block band matrices in Theorem 3.1 and 3.10. Although part of the estimate may still carry over without a bounded density, the estimates we can prove via the current available techniques are at best valid with probability $1 - \Theta(\ell^{-1/2})$ (see [27], Theorem 2.1), so that on the complementary event of probability $\Theta(\ell^{-1/2})$ we have no control over the log determinant at all. The transfer matrix approach requires a joint control over all these small blocks, and thus the estimate cannot work through when n is not small relative to ℓ. This technical difficulty disappears under a bounded density assumption, see Theorem 3.10, where we can control the least singular value with probability 1.

The main focus of this paper is the ergodic regime $n \gg \ell$, and a bounded density is typically assumed to guarantee sufficient regularity in this ergodic, or localized regime. For a Hermitian random band matrix, the Wegner estimate and dynamical localization for $W \ll \sqrt{N}$ are all proven [12] [9] [38] only when assuming a bounded density, and no Wegner estimate is known without a density assumption when $\Omega(1) \leq W \ll \sqrt{N}$. For the non-Hermitian tridiagonal matrix model [16], which corresponds to $W = O(1)$ in our paper, their proof of ESD convergence also uses a similar non-singularity assumption on off-diagonal entries.

Finally, we can assume that ζ has a density with a polynomially (in ℓ) bounded ℓ^∞ norm, say we assume that ζ has a density on $\mathbb{R}$ bounded by $L\ell^C$ for any $C > 0$. For example, this can originate from a vanishing smoothing with $\zeta = \sqrt{1 - \ell^{-2C}}\zeta_1 + \ell^{-C} g$, where ζ_1 is a centered Rademacher variable and g is an independent Gaussian. The result of Theorem 1.1 continues to hold under this assumption and the modification is straightforward.

Remark 1.3. (Model) The proof of Theorem 1.1 crucially uses the block tridiagonal structure of T, and thus does not immediately generalize to other band matrix models with a more general variance profile, such as the periodic case where entry (i,j) has variance $\frac{1}{2W+1}$ if $\min(|i-j|,|N-i+j|)\leq W$ and is identically 0 otherwise. Interestingly, the periodic block band case (see (4.1)), where we have an additional matrix C_1 on the top right corner and a matrix B_n on the lower left corner of T, is not covered by Theorem 1.1 either. For all these more general models, the current best result for the convergence of μ_T to the circular law is $\ell \geq (n\ell)^{1/2+c}$ for any $c>0$ [21], using the parameterization of Theorem 1.1.

From a universality perspective, we expect the circular law to be proven for μ_T for a much more general class of variance profiles whenever the bandwidth is growing, and Theorem 1.1 provides the first rigorous justification of this heuristics for at least one type of variance profile.

Remark 1.4. (On the high moment assumption) The high moment assumptions in Theorem 1.1 are most essentially used in the singular value rigidity result of Proposition 4.3. We have not tried to optimize the moment condition here, but all existing papers on circular law for inhomogeneous matrices [11], [27],[21] require at least a finite $4+\epsilon$ moment to obtain quantitative rigidity estimates. Merely a finite variance condition does not work here. The moment assumption is also used in bounding the transfer matrix in Section 2.4 and estimating the least singular value in Section 3, although a weak moment condition might work in these cases. As rigidity results are indispensable in all these papers, proving a circular law for these inhomogeneous models under only a two moment condition as in [41] might require a complete reworking and is beyond the current scope of this paper.

Remark 1.5. (On almost sure circular law) Theorem 1.1 proves circular law for μ_T in the almost sure sense. If one adopts the normalization $N=n\ell$ and $W=\ell$ to align with prior works, one can check that the almost sure circular law is proven whenever $\Omega(1)\leq W\leq N^{1-d}$ for any small $d>0$. For the square case $W=N$, the almost sure circular law was proven much earlier, see [41]. In contrast, previous circular law papers for band matrices [27], [43] only prove the weak (in probability) circular law. (But see [21], Theorem 1.8 where the proof indeed leads to almost sure circular law because the estimates of Theorem 2.9 therein are much stronger). The reason why we get strong circular law is twofold: first, a bounded density assumption leads to least singular value estimates with good tails (see Theorem 3.1), and more importantly, we adopt a transfer matrix approach with n transfer operators and apply martingale concentration inequality, so the larger n is the stronger probabilistic concentration we gain.

The last 12 months have seen spectacular progress in the study of random band matrices. With the highest relevance to this work, Shcherbina and Shcherbina [39] recently studied the second correlation function of the characteristic polynomial of a non-Hermitian band matrix model and showed distinct behaviors for $W\gg\sqrt{N}$ and $W\ll\sqrt{N}$, which serves as a first step towards the proof of Anderson type transition in non-Hermitian band matrices. For Hermitian band matrix models, Yau and Yin [46] proved for the block band model that whenever $W\geq N^{1/2+c}$ for any $c>0$, the eigenvectors of this band matrix are completely delocalized and quantum unique ergodicity holds. For more general 1-d random band matrices, the same results are proven by Erdős and Riabov in [15]. Higher-dimensional band matrices are also studied in [14] and [13]. In the localized regime $W\leq N^{1/2-c}$, the exponential localization of eigenfunctions was recently proven in Reuben [12], following a long line of previous papers [9], [8], [38],[17]. The invertibility of general non-Hermitian random band matrices with Bernoulli entries was also studied in [19]. Although this paper does not directly touch upon localization/delocalization of eigenvectors, Theorem 1.1 studies the most fundamental spectral statistics, the ESD, in particular in the complete localized regime $W \ll \sqrt{N}$. To our best knowledge, almost no spectral results are known in the localized regime for the non-Hermitian band matrix models, and this work may serve as a first step to the investigation in this localized regime.

1.2. **A brief outline of the proof.** As with all modern proofs of circular law, we begin with the following replacement principle by Tao, Vu and Krishnapur ([41], Theorem 2.1), which offers a sufficient condition for the circular law via convergence of the log determinant:

Theorem 1.6. *(Replacement principle) Suppose that for each $n \in \mathbb{N}$, there are two ensembles of random matrices $A_n, B_n \in M_n(\mathbb{C})$ such that*

(1) *The expression*
$$\|A_n\|_F^2 + \|B_n\|_F^2 \tag{1.2}$$
is bounded in probability (resp. almost surely), where $\|\cdot\|_F$ denotes the matrix Frobenius norm, and

(2) *And that for almost all complex numbers $z \in \mathbb{C}$, the quantity*
$$\frac{1}{n}\log|\det(A_n - zI)| - \frac{1}{n}\log|\det(B_n - zI)|$$
converges in probability (resp. almost surely) to zero. Then $\mu_{A_n} - \mu_{B_n}$ converges in probability (resp. almost surely) to zero.

To verify condition (2) of Theorem 1.6, we need to estimate $\log|\det A_z|$ where $A_z := A_n - zI$. The standard method for doing this (see the survey [4]) is as follows: Let $s_1 \geq \cdots \geq s_n$ denote the singular values of A_z, then $\frac{1}{n}\log|\det A_n| = \frac{1}{n}\sum_{i=1}^n \log s_i$. One can prove that the empirical measure for singular values of A_z converges weakly to a deterministic limit, so the convergence of $\log\det A_z$ follows once we show that the smallest (and a $o(n)$ number of small-ish) singular values among s_i make negligible contribution to the summation. Then we can truncate the summation at a threshold and apply the weak convergence result for a bounded test function. The control of s_n and small-ish s_i is equivalent to obtaining a least (and small-ish) singular value estimate for A_z, which typically has size n^{-C} for some $C > 0$ with high probability when A has i.i.d. entries [41], [35].

For a non-Hermitian random band matrix of size N and bandwidth W, the currently known lower bounds for the least singular values [27], [21], [43] are all exponentially small and has the order $s_N \geq \exp(-\frac{N}{W}N^\epsilon)$. Whether or not these bounds are sharp or can be considerably improved remains an open problem, especially when $W \ll \sqrt{N}$. Thus to guarantee the convergence of log determinant, a popular method is to get a rigidity estimate for the singular values, namely can we guarantee that with high probability only an N^c numbers of singular values are smaller than N^{-1}, say? Rigidity estimates can be proven via a local law, as in [21], which allows us to prove the circular law whenever $W \gg N^{1/2+c}$. However, it is clear that the local law estimate deteriorates when $W \ll \sqrt{N}$ and may be completely useless for small W, say when $W = \log\log N$.

The limitation of this rigidity estimate is that it requires all the singular values be simultaneously controlled, but in reality we may still guarantee convergence of the log determinant without uniform control over all singular values. Indeed, when $W \ll \sqrt{N}$, we are more in an ergodic regime than a mean field regime, and ideas from ergodic theory becomes more relevant. It is enlightening to observe that block tridiagonal matrices, a special class of band matrices, have a clear connection to ergodic theory via the transfer matrix recursion formula (Corollary 2.4) when computing its determinant. The recursion formula was also used in [16] for the tridiagonal case. Thus a natural option is to try to generalize the strategy of [16]. However, an immediate generalization is not feasible. The work [16] uses in an essential way ergodic theory of matrix products in $\mathrm{GL}_2(\mathbb{C})$ (see [5] and [31] for an account), but here we have the matrix product of size $2W$ and a total number of N/W matrices, with both quantities tending to infinity. No general quantitative ergodic theory exists for a general random matrix that controls all Lyapunov exponents (but see [10], [26], [30] for the i.i.d. setting), and when $N/W \gg W$ this matrix product is notoriously difficult to study. As a reference, [23] studied the Lyapunov exponents of products of Gaussian matrices with growing size, but our transfer matrix $M_k^{(B)}$ is a lot more complicated with general entry law, scalar shift and even an inverse matrix in it, so a direct analysis of Lyapunov exponent is clearly out of reach.

A surprisingly effective way to tackle this problem is that we can decompose the transfer matrix product in Corollary 2.4 as the concatenation of many independent systems of transfer matrix products of a smaller size. The starting phase of one system (a unitary ℓ-frame in $\mathbb{C}^{2\ell}$) depends on the output of the previous system, and the difference in every system is that we start the transfer matrix product at a different phase. Taking a filtration and conditioning, we can prove a martingale concentration theorem and show that the log determinant converges with high probability to its mean. By ergodic theory heuristics, the long term behavior should be independent of the initial phase, so we expect each such smaller system should have approximately the same large scale limit. Why does this idea of decomposing into smaller systems become helpful? This is because, by Proposition 2.2, we can naturally associate each smaller system with a tridiagonal block matrix with the same block size $\ell = W$ but having a much smaller number of blocks n. When ℓ is large and n is small, say $\ell \gg n^{12}$, then we are back to the case where circular law can be proven for this smaller system [21], [27],[43] by combining the (potentially suboptimal) least singular value estimate and the rigidity estimate, and what we indeed prove is convergence of the log determinant to the same fixed limit. The place to be cautious is that in Proposition 2.2, we associate these subsystems to a tridiagonal block matrix with an upper and lower boundary layer determined by the initial phase of recursion, and a large portion of technical effort is to obtain analogous, boundary-independent estimates for this matrix with a general boundary. Combining the convergence in each subsystem allows us to prove the convergence of logarithmic determinant for the large tridiagonal block matrix even if $W \ll \sqrt{N}$.

In short, the proof takes an effective interplay between two distinct methods of computing the log determinant: as the summation of log-singular values and as a transfer matrix recursion. The proof combines ideas from ergodic theory and the circular law proof for near-mean-field band matrices, by introducing a general boundary to account for the different initial phase and showing boundary independence in all ensuing estimates.

This paper is organized as follows. Section 2 outlines the main body of proof, conditioning on Theorem 2.16 which states the convergence for a system of smaller scale. The later sections are devoted to the proof of Theorem 2.16, where Section 3 derives the least singular value estimate, Section 4 completes the proof of Theorem 2.16 conditioning on the rigidity estimate in Proposition 4.3, and Section 5 completes the proof of this rigidity estimate.

2. Proof of circular law: main structure of the argument

This section has three components. First, we present the linear algebra result that reduces the determinant computation of block matrices to the transfer matrix problem and set up an inverse problem that, for a given initial value, produces a block tridiagonal matrix. Then we prove martingale concentration inequalities for the transfer matrix, and reduce the convergence result to a smaller scale system in Theorem 2.16.

2.1. **Linear algebra results.** We first introduce a convenient notion I_{mid} for certain block diagonal matrices:

Notation 2.1. *For given integers n,ℓ, we let I_{mid} be the following $(n+2)\ell\times(n+2)\ell$ diagonal matrix: it has entries 0 on the first ℓ rows and the last ℓ rows, and it has entries* 1 *on the middle $n\ell$ rows.*

We shall need the following computation to handle general boundary conditions:

Proposition 2.2. *Let $A_1,\cdots,A_n,B_1,\cdots,B_n,C_1,\cdots,C_n$ be $\ell\times\ell$ matrices such that B_i are invertible for all $1\le i\le n$. Let $\Pi\in\mathbb{C}^{\ell\times 2\ell}$ and $\Xi\in\mathbb{C}^{2\ell\times\ell}$ be fixed matrices such that*

$$\det(\Pi\Pi^*)=\det(\Xi^*\Xi)=1,$$

where throughout the paper we use Ξ^* to denote the complex conjugate transpose of Ξ. Then we can find a block tridiagonal matrix $\mathcal{T}_{n+2}$ of size $(n+2)\ell\times(n+2)\ell$ such that

$$\det\left(\mathcal{T}_{n+2}-zI_{mid}\right)=(-1)^{(n+1)\ell}\Big(\prod_{k=1}^{n}\det B_k\Big)\det\Big(\Pi\Big(\prod_{k=n}^{1}M_k^{(B)}(z)\Big)\Xi\Big)$$

for any $z\in\mathbb{C}$, and where for each k,

$$M_k^{(B)}(z)=\begin{bmatrix}-B_k^{-1}(A_k-zI_\ell) & -B_k^{-1}C_k\\ I & 0\end{bmatrix},$$

and $\mathcal{T}_{n+2}$ has the form

$$\mathcal{T}_{n+2}=\begin{bmatrix}V & U & & & & \\ C_1 & A_1 & B_1 & & & \\ & C_2 & A_2 & B_2 & & \\ & & \ddots & \ddots & \ddots & \\ & & & C_n & A_n & B_n\\ & & & & S_+^{bd} & C_+\end{bmatrix}\in\mathbb{C}^{(n+2)\ell\times(n+2)\ell}, \tag{2.1}$$

where $\left[V,U\right],\left[S_+^{bd},C_+\right]\in\mathbb{C}^{\ell\times 2\ell}$ are determined only by Π and Ξ that satisfy

$$\left[V,U\right]\left[V,U\right]^*=\left[S_+^{bd},C_+\right]\left[S_+^{bd},C_+\right]^*=I_\ell.$$

Notation 2.3. *We use the non-standard notation $\prod_{k=n}^{1}M_k$ to denote the matrix product $M_nM_{n-1}\cdots M_2M_1$, since matrix multiplication is not commutative.*

The symbols Ξ and Π denote the boundary conditions of this tridiagonal matrix iteration. We can specialize to block tridiagonal matrices as follows:

Corollary 2.4. *With the same assumption as Proposition 2.2, define the following matrix*

$$\mathcal{T}_n^{tri} = \begin{bmatrix} A_1 & B_1 & & & \\ C_2 & A_2 & B_2 & & \\ & \ddots & \ddots & \ddots & \\ & & C_{n-1} & A_{n-1} & B_{n-1} \\ & & & C_n & A_n \end{bmatrix} \tag{2.2}$$

then we have for any $z \in \mathbb{C}$*,*

$$\det(\mathcal{T}_n^{tri} - zI_{n\ell}) = (-1)^{n\ell}(\prod_{k=1}^{n} \det B_k) \det\left(\begin{bmatrix} I_\ell & 0 \end{bmatrix} \prod_{k=n}^{1} M_k^{(B)}(z) \begin{bmatrix} I_\ell \\ 0 \end{bmatrix} \right).$$

Proof of Proposition 2.2. **Step 1: forming unitary boundary frames**. We define

$$Q_- := \Xi(\Xi^*\Xi)^{-1/2} \in \mathbb{C}^{2\ell\times\ell}, \quad \widetilde{Q}_+ := (\Pi\Pi^*)^{-1/2}\Pi \in \mathbb{C}^{\ell\times 2\ell},$$

so that we necessarily have

$$Q_-^*Q_- = I_\ell, \quad \widetilde{Q}_+\widetilde{Q}_+^* = I_\ell.$$

Then for any $2\ell \times 2\ell$ matrix K, we use the following elementary identity to simplify:

$$\det(\widetilde{Q}_+KQ_-) = \frac{\det(\Pi K\Xi)}{\sqrt{\det(\Pi\Pi^*)\det(\Xi^*\Xi)}} = \det(\Pi K\Xi), \tag{2.3}$$

using the assumption $\det(\Pi\Pi^*) = \det(\Xi^*\Xi) = 1$.

From Q_- we choose a unitary complement $Q_-^\perp \in \mathbb{C}^{2\ell\times\ell}$ (and thus $(Q_-^\perp)^*Q_-^\perp = I_\ell$) so that we have

$$\begin{bmatrix} Q_-^\perp, Q_- \end{bmatrix} \in \mathcal{U}(2\ell),$$

the group of unitary matrices of size 2ℓ. We then partition $(Q_-^\perp)^*$ into the following form

$$(Q_-^\perp)^* = \begin{bmatrix} U & V \end{bmatrix} \in \mathbb{C}^{\ell\times 2\ell},$$

so that $VV^* + UU^* = I_\ell$ and thus $\|V\|, \|U\| \leq 1$. In this proof, we use column labels $0, 1, \cdots, n+1$ to index the columns of T_{n+2},

We also define the following decomposition

$$\widetilde{Q}_+ = \begin{bmatrix} C_+, S_+^{bd} \end{bmatrix}. \tag{2.4}$$

Step 2: Elementary reductions of $\mathcal{T}_{n+2}$ that preserve the determinant. We introduce three operations that can help us to compute the determinant of $\mathcal{T}_{n+2}$, using $0, 1, \cdots, n+1$ to index the column blocks and the row blocks of the matrix $\mathcal{T}_{n+2}$. The following column pairs $(k, k+1)$ all refer to this labeling of the k-th and the $k+1$-th block.

2.1: Right multiplication removing A_k and B_k. For each $k = 1, \cdots, n$, we act on the column pair $(k, k+1)$ by the following right multiplier

$$G_k := \begin{bmatrix} I_\ell & 0 \\ -B_k^{-1}(A_k - zI_\ell) & B_k^{-1} \end{bmatrix}.$$

Then we extend G_k to the whole matrix via placing it at the $(k, k+1)$-block columns and block rows, while we preserve the identity elsewhere. Let $\mathcal{G}_k$ denote this constructed matrix, then we have

$$\det \mathcal{G}_k = \det G_k = \det(B_k)^{-1}.$$

Note that our choice of G_k always satisfies

$$\begin{bmatrix} A_k - zI_\ell, B_k \end{bmatrix} G_k = \begin{bmatrix} 0, I_\ell \end{bmatrix}.$$

Then the entry in column $k+1$ is exactly I_ℓ. After the product $\mathcal{P}$ is applied, the rows $2, \cdots, n+1$ each contain a unique identity block in distinct columns.

2.2: Right multiplication removing C_k. For each $k = 1, 2, \cdots, n$, we perform the following elementary column update

$$\mathrm{Col}_{k-1} \leftarrow \mathrm{Col}_{k-1} - C_k \,\mathrm{Col}_{k+1},$$

this is obtained by right multiplication of $\mathcal{T}_{n+2}$ by an elementary block matrix $\mathcal{E}_k$ of determinant 1.

2.3: Clearing up the identity via left multiplication. For each $1 \le k \le n$ define the following row operation

$$\mathcal{L}_k : \ \mathrm{Row}_i \leftarrow \mathrm{Row}_i - (\mathrm{Row}_i \ \text{ at column } k+1) \cdot \mathrm{Row}_k,$$

and this is applied for all indices $i \neq k$ having a nonzero entry in the column $k+1$. In our sequential application (see below), each time we apply $\mathcal{L}_k$, the only such row to be modified by $\mathcal{L}_k$ is the bottom row $i = n+1$. Again, we have $\det \mathcal{L}_k = 1$.

Sequential cleaning. For each $k = n, \cdots, 1$ in decreasing order, we apply, in this order, the following operations on the matrix $\mathcal{T}_{n+2}$:

(1) Right pivot on B_k: we multiply from the right the matrix $\mathcal{G}_k$, $\det \mathcal{G}_k = \det(B_k)^{-1}$.
(2) Column clean-up: we multiply from the right the matrix $\mathcal{E}_k$ with $\det \mathcal{E}_k = 1$.
(3) Row clean-up: we multiply from the left the matrix $\mathcal{L}_k$ with $\det \mathcal{L}_k = 1$.

Set $\mathcal{L} = \prod_{k=1}^n \mathcal{L}_k$, $\quad \mathcal{P} = \prod_{k=1}^n (\mathcal{E}_k \mathcal{G}_k)$ and we denote

$$\mathcal{T}^{red} := \mathcal{L}(\mathcal{T}_{n+2}\mathcal{P}).$$

To compute the determinant $\det \mathcal{T}^{red}$, we first check that a number of properties are preserved under this sequential cleaning procedure. In the following we let $\rho_0, \rho_1, \cdots, \rho_{n+1}$ denote the entries of the last block row of $\mathcal{T}_{n+2}$, which changes after every iteration.

We claim that after the processing steps labeled $n, n-1, \cdots, k$ (i.e., after we apply $\mathcal{G}_n, E_n, \mathcal{L}_n, \cdots, \mathcal{G}_k, E_k, \mathcal{L}_k$), the following induction hypothesis hold for this value k:

- $\mathrm{Ind}(k)$ for interior levels: For each $n \ge j \ge k$,the row j is a pure pivot in that

$$\text{row } j = \text{(only) } I_\ell \text{ at column } j+1, \text{ zeros elsewhere.}$$

- Ind_{top}: The top row of $\mathcal{T}_{n+2}$ is unchanged and remains $[V, U]$ supported on the columns $(0, 1)$.
- $\mathrm{Ind}_{bot}(k)$, the bottom row sliding pair: We assume that the row $n+1$ is supported only on columns $(k-1, k)$ with the 2-block of the form

$$[\rho_k, \rho_{k-1}] = \left[C_+, S_+^{bd}\right] \big(\prod_{m=n}^{k} M_m^{(B)}(z)\big).$$

Verification of induction hypotheses. For the initial value $k = n+1$, we have $\mathrm{Ind}(n+1)$ is vacuous, Ind_{top} is trivially true and $\mathrm{Ind}_{bot}(n+1)$ follows from definition. Then we prove that all these properties hold for all k via taking a backward induction on k. In each inductive step $k \to k-1$, we assume the inductive hypothesis $\mathrm{Ind}(k), \mathrm{Ind}_{top}, \mathrm{Ind}_{bot}(k)$.

The application of $\mathcal{G}_{k-1}$ changes only columns $(k-1, k)$, and then the interior $k-1$-th row becomes $[C_{k-1}, 0, I_\ell]$ on columns labeled $(k-2, k-1, k)$. The bottom pair $[\rho_{k-1}, \rho_k]$ now becomes

$$\left[\rho_{k-1} - \rho_k B_{k-1}^{-1}(A_{k-1} - zI_\ell), \rho_k B_{k-1}^{-1}\right]$$

on entries $(k-1,k)$. Then applying $\mathcal{E}_{k-1}$, this would zero out C_{k-1} in row $k-1$ and adds $-\rho_k B_{k-1}^{-1} C_{k-1}$ into column $k-2$ of the bottom row, which sends ρ_k to 0. Finally $\mathcal{L}_{k-1}$ clears the column k from the bottom row. Then we conclude that

$$[\rho_{k-1},\rho_{k-2}] = [\rho_k,\rho_{k-1}]M_{k-1}^{(B)} = [C_+, S_+^{bd}] \prod_{m=n}^{k} M_m^{(B)}(z) \cdot M_{k-1}^{(B)}(z),$$

thus verifying $\mathrm{Ind}(k-1)$ and $\mathrm{Ind}_{bot}(k-1)$. The top row never moves under these maps, so we verify Ind_{top}. Thus verifies all the inductive hypothesis for each $k = 1, \cdots, n+1$.

From this inductive proof, we can see that $\mathcal{T}^{red}$ has the following form

$$\mathcal{T}^{red} = \begin{pmatrix} [V,U], & & & \\ & I_\ell & & \\ & & \ddots & \\ & & & I_\ell \\ [\rho_0,\rho_1] & & & \end{pmatrix},$$

and the last row is

$$[\rho_1,\rho_0] = [C_+, S_+^{bd}] \prod_{k=n}^{1} M_k^{(B)}(z).$$

Then let us denote by

$$\mathcal{B} := \begin{bmatrix} U & V \\ \rho_1 & \rho_0 \end{bmatrix} = \begin{bmatrix} U, \quad V \\ [C_+, S_+^{bd}] \left(\prod_{k=n}^{1} M_k^{(B)}(z)\right) \end{bmatrix}.$$

We have $\det \mathcal{T}^{red} = (-1)^{(n+1)\ell} \det \mathfrak{B}$, so that we conclude

$$\det \mathcal{T}_{n+2} = (-1)^{(n+1)\ell} \left(\prod_{k=1}^{n} \det B_k\right) \det \mathfrak{B}.$$

Step 3: computing $\det \mathfrak{B}$**.** By the following Fact 2.5, we have

$$\det \mathfrak{B} = \det \left[[C_+, S_+^{bd}] \textstyle\prod_{k=n}^{1} M_k^{(B)}(z) Q_- \right] = \det \left[\widetilde{Q}_+ (\textstyle\prod_{k=n}^{1} M_k^{(B)}(z)) Q_- \right],$$

which completes the proof by equation (2.3). □

Fact 2.5. *Let $R \in \mathbb{C}^{\ell\times 2\ell}$ have orthonormal rows $RR^* = I_\ell$. Let $Q_- \in \mathbb{C}^{2\ell\times\ell}$ have unitary columns and $RQ_- = 0$. Then we have, for any $X \in \mathbb{C}^{\ell\times 2\ell}$,*

$$\det \begin{bmatrix} R \\ X \end{bmatrix} = \det(XQ_-),$$

whenever the $2\ell\times 2\ell$ matrix block $\begin{bmatrix} R \\ Q_-^* \end{bmatrix}$ is in $SU(2\ell)$.

Proof. We shall complete R into the following unitary matrix

$$\mathcal{Q} := \begin{bmatrix} R \\ R_\perp \end{bmatrix} \in SU(2\ell), \quad R_\perp^* = Q_-.$$

Since $\det Q = 1$, and by left multiplication we have

$$\begin{bmatrix} R \\ X \end{bmatrix} \mathcal{Q}^* = \begin{bmatrix} RR^* & RR_\perp^* \\ XR^* & XR_\perp^* \end{bmatrix} = \begin{bmatrix} I_\ell & 0 \\ XR^* & XQ_- \end{bmatrix}.$$

Thus we have

$$\det \begin{bmatrix} R \\ X \end{bmatrix} = \det(XQ_-),$$

which completes the proof. □

Now we specialize these computations to the special case in Corollary 2.4.

Proof of Corollary 2.4. For the determinant of $\mathcal{T}_n^{tri}$, one can check that with $\Pi = \Xi^* = [I_\ell, 0]$, we have $V = -I_\ell, U = 0, C_+ = I_\ell, S_+^{bd} = 0$. Taking these special values into definition of $\mathcal{T}_{n+2}$, we have $\det(\mathcal{T}_{n+2} - zI_{mid}) = (-1)^\ell \det(\mathcal{T}_n^{tri} - zI_{n\ell})$, completing the proof. □

2.2. **Least singular value and operator norm bounds.** We collect here several standard results on the least singular value and operator norm of a random matrix.

For entry distribution of bounded density, we use the following result of [42]:

Lemma 2.6. *([42], Corollary 1.2) Let A be an $n \times n$ matrix with i.i.d. real entries of mean 0, variance 1 and having distributional density bounded by $L > 0$. Then we can find $C = C(L) > 0$ depending only on L such that*

$$\mathbb{P}(s_{min}(A) \le \epsilon n^{-1/2}) \le C\epsilon, \quad \epsilon > 0. \tag{2.5}$$

The main advantage of Lemma 2.6 is that it does not have the exponential error term e^{-cn} which appears when no density assumption is made. For Gaussian matrices, these estimates without exponential error were known much earlier [37].

For the complex case, we can dispense with the following weaker estimate:

Lemma 2.7. *Let ζ be a mean 0, variance 1 random variable satisfying the density assumptions in Theorem 1.1. Then we can find $C = C(L) > 0$ such that*

$$\mathbb{P}(s_{min}(A) \le \epsilon n^{-3/2}) \le C\epsilon, \quad \epsilon > 0. \tag{2.6}$$

Lemma 2.7 can be proven in a way similar to Lemma 3.2 and is fairly standard by now, so we omit its proof since it is a straightforward modification of Lemma 3.2.

We shall need a quantitative operator norm estimate, which follows from taking a truncation to [2], Theorem 5.9:

Lemma 2.8. *Let A be an $n \times n$ matrix with i.i.d. complex-valued entries ζ having mean 0, variance 1 and all p-th moment finite. Then we can find $C_\zeta > 0$ and $C_k > 0$ for each $k \in \mathbb{N}_+$, depending only on the moments of ζ such that the following two estimates hold:*

$$\mathbb{E}\|A\|_{op} \le C_\zeta \sqrt{n},, \quad \mathbb{P}(\|A\| \ge C_\zeta t\sqrt{n}) \le C_k (tn)^{-k} \quad \forall k \in \mathbb{N}_+, \forall t \ge 1.$$

Finally, we need a result on convergence of log determinants of random matrices. We use the normalization $(\frac{1}{3n})^{1/2}$ since the entries in T are normalized by $(\frac{1}{3\ell})^{1/2}$.

Proposition 2.9. *Let A be an $n \times n$ matrix with i.i.d. complex-valued entries ζ having mean 0 and variance 1. Then*

$$\frac{1}{n} \log \left|\det \left((3n)^{-1/2} A\right)\right| \to_{n\to\infty} -\frac{1}{2}\log 3 - \frac{1}{2}$$

where the convergence holds both in probability and almost surely. Assuming moreover that ζ satisfies the assumptions in Theorem 1.1, we also have the convergence in expectation

$$\frac{1}{n}\mathbb{E} \log \left|\det \left((3n)^{-1/2} A\right)\right| \to_{n\to\infty} -\frac{1}{2}\log 3 - \frac{1}{2}.$$

The statement for almost sure convergence follows from the circular law proof in [41], but we will actually use the version of convergence in expectation when applying martingale concentration inequalities. We will justify this by proving uniform integrability of the log determinant, so that stronger assumptions (such as a bounded density) are needed. The proof of Proposition 2.9 is deferred to Section 4.4.

2.3. **Identification with a transfer operator iteration.** We will essentially be studying the quantity $\det(\Pi(\prod_{k=n}^1 M_k^{(B)}(z))\Xi)$.

It is convenient to re-express this quantity in the wedge space $\wedge^\ell \mathbb{C}^{2\ell}$, and write it in the form of an ergodic dynamical system acting on $\wedge^\ell \mathbb{C}^{2\ell}$.

Denote by $\mathcal{W} := \wedge^\ell \mathbb{C}^{2\ell}$. Then every linear map $g : \mathbb{C}^{2\ell} \to \mathbb{C}^{2\ell}$ lifts to the following linear map

$$\wedge^\ell g : \mathcal{W} \to \mathcal{W}, \quad \wedge^\ell g(v_1 \wedge \cdots \wedge v_\ell) = (gv_1) \wedge (gv_2) \wedge \cdots \wedge (gv_\ell),$$

and similarly for any linear gap $g : V_1 \to V_2$ between two linear spaces V_1, V_2, the map $\wedge^\ell g$ is the lift: $\wedge^\ell g : \wedge^\ell V_1 \mapsto \wedge^\ell V_2$. We identify $\wedge^\ell \mathbb{C}^\ell$ with $\mathbb{C}$.

For the given boundary value Π and Ξ, we let $\hat{f} := \wedge^\ell \Pi$ and let $\hat{g} := \wedge^\ell \Xi$, where we can identify $\hat{g} \in \mathcal{W}$ and $\hat{f} \in \mathcal{W}^*$, where $\mathcal{W}^*$ is the dual space of $\mathcal{W}$. Then we have the identity

$$\det(\Pi(\prod_{k=n}^{1} M_k^{(B)}(z))\Xi) = \langle \hat{f}, \prod_{k=n}^{1} \wedge^\ell M_k^{(B)}(z) \hat{g} \rangle. \tag{2.7}$$

This equality follows from writing the Cauchy-Binet formula in the language of exterior algebra.

The following property will be frequently used:

Fact 2.10. *Let $g : \mathbb{C}^{2\ell} \to \mathbb{C}^{2\ell}$ be a linear map with singular values $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_{2\ell}$. Then the singular values of $\wedge^\ell g$ are given by the following collection*

$$\{\sigma_{i_1}\sigma_{i_2}\cdots\sigma_{i_\ell} : \quad 1 \leq i_1 < i_2 < \cdots < i_\ell \leq 2\ell\}.$$

Then we introduce two frequently used subspaces of $\mathcal{W}$. Let $\mathrm{Gr} := \mathrm{Gr}(\ell, 2\ell)$ be the Grassmannian of ℓ-planes in $\mathbb{C}^{2\ell}$, then $\mathrm{Gr}(\ell, 2\ell)$ is naturally identified as a subspace of $\mathbb{P}(\mathcal{W})$, the projective space of $\mathcal{W}$, via Plücker embedding. We also define the Stiefel manifold

$$\mathrm{St}(\ell, 2\ell) := \{\Pi \in \mathbb{C}^{\ell \times 2\ell} : \Pi\Pi^* = I_\ell\}.$$

Then we have the natural identification, for $U(\ell)$ the unitary group,

$$\mathrm{Gr}(\ell, 2\ell) = \mathrm{St}(\ell, 2\ell)/U(\ell).$$

We also introduce the decomposable cone of $\mathcal{W}$, which forms a basis of $\mathcal{W}$:

$$\mathcal{C}_{dec} := \{v_1 \wedge \cdots \wedge v_\ell \in \mathcal{W} \setminus \{0\}\}.$$

There is a natural identification of its projectivization with the Grassmannian:

$$\mathbb{P}\mathcal{C}_{dec} \simeq \mathrm{Gr}(\ell, 2\ell).$$

We start from $Q_0 := \hat{g} \in \mathcal{C}_{dec} \subseteq \mathcal{W}$ and proceed inductively. Given a current ℓ-frame $Q_{k-1} \in \mathcal{C}_{dec} \subseteq \mathcal{W}$, we take one further matrix multiplication by $\wedge^\ell M_k^{(B)}$ and set

$$Q_k := \wedge^\ell M_k^{(B)} Q_{k-1} \in \mathcal{C}_{dec} \subseteq \mathcal{W}.$$

Let $X_k = [Q_k]$ be the projective chain on the projective space $\mathbb{P}\mathcal{C}_{dec} \simeq \mathrm{Gr}(\ell, 2\ell)$. We define the following function

$$g_\ell(M_k^{(B)}(z), \hat{g}) := \log \frac{\| \wedge^\ell M_k^{(B)}(z)\hat{g}\|}{\|\hat{g}\|}$$

for any $\hat{g} \in \mathcal{C}_{dec}$. Then we may decompose the additive cocycle as

$$\log \left| \det(\Pi(\prod_{k=n}^{1} M_k^{(B)}(z))\Xi) \right| = \sum_{k=1}^{n-1} g_\ell(M_k^{(B)}, X_{k-1}) + \log \left| \langle \hat{f}, \wedge^\ell M_n^{(B)}(z) X_{n-1}\rangle \right|. \tag{2.8}$$

Now by the independence of entries from $M_k^{(B)}(z)$ for distinct k, the above summation, after subtracting its mean, forms a martingale sequence. We first show that the martingale sequence, after dividing by $1/n\ell$ and subtracting its mean, is quantitatively very close to 0 for large n and ℓ. This is done via computing exponential moments and applying concentration inequalities. The harder part is to actually determine the mean.

2.4. **Exponential moments for intermediate increment steps.**

Lemma 2.11. *Consider the function*

$$g_\ell(M_k^{(B)}(z), \hat{g}) := \log \frac{\| \wedge^\ell M_k^{(B)}(z)\hat{g}\|}{\|\hat{g}\|}$$

for any $\hat{g} \in \mathcal{C}_{dec}$. Assume that the random blocks A_k, B_k, C_k in $M_k^{(B)}(z)$ satisfy the same assumption as in Theorem 1.1. Then we can find two constants c_0 and C_0 (depending only on $|z|$ and the entry law of $M_k^{(B)}(z)$) such that, uniformly for any $\hat{g} \in \mathcal{C}_{dec}$, the following estimate holds for all $-\frac{c_0}{\ell} \leq c \leq \frac{c_0}{\ell}$:

$$\mathbb{E} \exp\left(c g_\ell(M_k^{(B)}(z), \hat{g})\right) \leq C_0 \ell^2.$$

We also have the following moment computations:

$$\mathbb{E}[|g_\ell(M_k^{(B)}(z), \hat{g})|^2] \leq C_0(\ell \log \ell)^2, \qquad \mathbb{E}[|g_\ell(M_k^{(B)}(z), \hat{g})|^6] \leq C_0(\ell \log \ell)^6$$

Proof. By Fact 2.10, we have

$$s_{min}(M_k^{(B)}(z))^\ell \leq \| \wedge^\ell M_k^{(B)}(z)\| \leq s_{max}(M_k^{(B)}(z))^\ell,$$

and similarly we have

$$\|(\wedge^\ell M_k^{(B)}(z))^{-1}\| \leq s_{min}(M_k^{(B)}(z))^{-\ell} = (s_{max}(M_k^{(B)}(z))^{-1})^\ell.$$

We can bound

$$\|M_k^{(B)}(z)\| \leq \|B_k^{-1}\|(1 + \|A_k\| + |z| + \|C_k\|).$$

Then by Lemma 2.6, 2.7 and 2.8 and applying Cauchy-Schwarz inequality, we can find constants $c_0 > 0, C_0 > 0$ depending only on z and ζ such that for any $c \in (0, c_0)$, we have

$$\mathbb{E}\|M_k^{(B)}(z)\|^c \leq C_0 \ell^2.$$

Since the entries C_k have bounded density, C_k is invertible almost surely. We then compute

$$(M_k^{(B)}(z))^{-1} = \begin{bmatrix} 0 & I_\ell \\ -C_k^{-1} B_k & -C_k^{-1}(A_k - zI_\ell) \end{bmatrix},$$

so that for the same choice of c_0, C_0 we have for all $c \in (0, c_0)$,

$$\mathbb{E}\|(M_k^{(B)}(z))^{-1}\|^c \leq C_0 \ell^2.$$

Combining these two bounds yields the first estimate. For the estimate on variance, by these moment computations we verify that $\mathbb{E}\left[\left|\log \| \wedge^\ell M_k^{(B)}(z)\|\right|^2\right] \leq C_0\ell^2\log^2\ell$ for a sufficiently large C_0. The computation for the sixth moment is exactly the same. □

Via a more involved argument, we can show that

Lemma 2.12. *We take $\hat{f} = \wedge^\ell[I_\ell, 0]$. For the two constants c_0, C_0 in Lemma 2.11, we have uniformly for any $\hat{g} \in \mathcal{C}_{dec}$ with $\|\hat{g}\| = 1$, for all $-\frac{c_0}{\ell} \leq c \leq \frac{c_0}{\ell}$:*

$$\mathbb{E}\exp\left(c\log|\langle \hat{f}, \wedge^\ell M_n^{(B)}(z)\hat{g}\rangle|\right) \leq C_0\ell^2.$$

We also have the following moment computations:

$$\mathbb{E}[|\log|\langle \hat{f}, \wedge^\ell M_n^{(B)}(z)\hat{g}\rangle||^2] \leq C_0(\ell\log\ell)^2, \quad \mathbb{E}[|\log|\langle \hat{f}, \wedge^\ell M_n^{(B)}(z)\hat{g}\rangle||^6] \leq C_0(\ell\log\ell)^6.$$

Proof. We can find some $\Xi \in \mathbb{C}^{2\ell\times\ell}$, $\det(\Xi^*\Xi) = 1$, such that $\hat{g} = \wedge^\ell\Xi$ since $\hat{g}$ is decomposable with unit norm. Then by (2.7), we only need to compute the moments of $\det(\Pi M_n^{(B)}(z)\Xi)$, with $\Pi = [I_\ell, 0]$. Then by Proposition 2.2 with $n = 1$ and $\Pi = [I_\ell, 0]$, so that Proposition 2.2 yields $[S_+^{bd}, C_+] = [0, I_\ell]$ and $[V, U]$ depending on Ξ, we have

$$\log|\det(\Pi M_n^{(B)}(z)\Xi)| = -\log|\det B_n| + \log\left|\det\begin{bmatrix} V & U & 0 \\ C_n & A_n - zI_\ell & B_n \\ 0 & 0 & I_\ell\end{bmatrix}\right|.$$

We simply use

$$\log|\det B_n| \leq \ell|\log s_{max}(B_n)| + \ell|\log s_{min}(B_n)|, \quad |\det B_n|^c \leq \max(s_{min}(B_n)^c, s_{max}(B_n)^c) \tag{2.9}$$

for the first term, and it suffices to bound $\log|\det E_n|$ where

$$E_n = \begin{bmatrix} V & U \\ C_n & A_n - zI_\ell\end{bmatrix}.$$

By (2.9) with B_n replaced by E_n, we only need a lower tail estimate for $s_{min}(E_n)$ as the upper tail estimate for $s_{max}(E_n)$ follows from Lemma 2.8. This lower tail for $s_{min}(E_n)$ can be derived as a special case of Theorem 3.1 but we outline a simpler proof here. The idea is as follows: since the rows of $[V, U]$ are unitary, it maps by isometry on the subspace W_1 of $\mathbb{C}^{2\ell}$ spanned by these rows. Then we consider the vectors in $\mathbb{C}^{2\ell}$ orthonormal to the rows of $[V, U]$, and restrict the mapping $[C_n, A_n - zI_\ell]$ to this subspace W_2. By Lemma 3.2, the restricted mapping has least singular value bounded by (3.2), so that we have $\mathbb{P}(s_{min}([C_n, A_n - zI_\ell]\mid_{W_2}) \leq \epsilon\ell^{-2}) \leq C_L\epsilon$ for all $\epsilon > 0$ (we switch from $\ell^{-1.5}$ there to ℓ^{-2} here since the entries are normalized by $(3\ell)^{-1/2}$.) We also need an operator norm estimate: by Lemma 2.8 we can assume that $\mathbb{P}(\|E_n\| \geq 4K_z\epsilon^{-1}) \leq \epsilon$ for all $\epsilon \leq 1$, where $K_z > 1$ is a constant depending only on $|z|$ and ζ. Now for any unit vector $w \in \mathbb{C}^{2\ell}$ with orthonormal decomposition $w = w_1 + w_2, w_1 \in W_1, w_2 \in W_2$, if $\|w_1\| \leq \epsilon^2\ell^{-2}(16K_z)^{-1}$, then on the event where s_{min} is not too small and $\|E_n\|$ is bounded as above, we have that

$$\|E_nw\| \geq s_{min}([C_n, A_n - zI_\ell]\mid_{W_2})\|w_2\| - \|[C_n, A_n - zI_\ell]\|\|w_1\| \geq \epsilon\ell^{-2}(16K_z)^{-1}.$$

If $\|w_1\| \geq \epsilon^2\ell^{-2}(16K_z)^{-1}$ we also trivially have $\|E_nw\| \geq \epsilon^2\ell^{-2}(16K_z)^{-1}$. This completes the proof that $\mathbb{P}(s_{min}(E_n) \leq \epsilon^2\ell^{-2}(16K_z)^{-1}) \leq 2C_L\epsilon$ for any $\epsilon > 0$.

Combining all these estimates leads us to the result of Lemma 2.12. Finally, we can slightly modify the constants c_0, C_0 in Lemma 2.11, so they also works here. □

Now we recall some standard facts of moment generating functions. For a random variable X, we denote by

$$K_X(t) = \log \mathbb{E}e^{t(X-\mathbb{E}X)}$$

its moment generating function. Let

$$\alpha := \inf\{t : \mathbb{E}e^{tX} < \infty\}, \quad \beta := \sup\{t : \mathbb{E}e^{tX} < \infty\},$$

then $K_X(t)$ is real analytic on (α, β).

We have $K_X(0) = 0, K'_X(0) = \mathbb{E}[X - \mathbb{E}X] = 0$, $K''_X(0) = \mathrm{Var}(X)$. Assume that $K_X(t)$ is analytic in $(-\alpha, \alpha)$, then Taylor's theorem gives, for any $|t| \leq \alpha/2$, the expansion

$$K_X(t) = \frac{\mathrm{Var}(X)}{2}t^2 + \frac{K_X^{(3)}(\zeta_t)}{6}t^3 \quad \text{for some } \zeta_t \in (-\alpha, \alpha), \tag{2.10}$$

where $K_X^{(3)}(\zeta_t)$ is the third derivative of $K_X^{(3)}$ at ζ_t. To compute the third derivative, since $K_X^{(3)}(t) = \mathbb{E}_t[(X - \mathbb{E}_t X)^3]$ via the tilted law $d\mathbb{P}_t \propto e^{t(X-\mathbb{E}X)}d\mathbb{P}$, we have

$$\begin{aligned} |K_X^{(3)}(t)| &= \frac{\mathbb{E}[(X - \mathbb{E}X - \frac{\mathbb{E}[(X-\mathbb{E}X)e^{t(X-\mathbb{E}X)}]}{\mathbb{E}e^{t(X-\mathbb{E}X)}})^3 e^{t(X-\mathbb{E}X)}]}{\mathbb{E}e^{t(X-\mathbb{E}X)}} \\ &\leq 4\frac{\mathbb{E}[X^3 e^{t(X-\mathbb{E}X)}]}{\mathbb{E}e^{t(X-\mathbb{E}X)}} + 4\left(\frac{\mathbb{E}[(X-\mathbb{E}X)e^{t(X-\mathbb{E}X)}]}{\mathbb{E}e^{t(X-\mathbb{E}X)}}\right)^3. \end{aligned} \tag{2.11}$$

We can immediately convert the moment computations into the following asymptotics of moment generating functions for the random variables:

Fact 2.13. *We take $\hat{f} = \wedge^\ell[I_\ell, 0]$. for any $\epsilon > 0$, we can find a constant $c'_0 > 0$ depending on ζ, $|z|$ and ϵ, and a constant $C'_0 > 0$ depending only on ζ and $|z|$ such that for any $-\frac{c'_0}{\ell} \leq t \leq \frac{c'_0}{\ell}$ and any $\hat{g} \in \mathcal{C}_{dec}$ with $\|\hat{g}\| = 1$, we have*

$$|K_{g_\ell(M_k^{(B)}(z),\hat{g})}(t)| \leq C'_0(\ell \log \ell)^2(t^2 + (\ell^{1+\epsilon}\log \ell)t^3),$$

$$|K_{\log|\langle \hat{f}, \wedge^\ell M_n^{(B)}(z)\hat{g}\rangle|}(t)| \leq C'_0(\ell \log \ell)^2(t^2 + (\ell^{1+\epsilon}\log \ell)t^3).$$

Proof. Take $\epsilon \in (0, \frac{1}{2})$. By Jensen inequality we always have $\mathbb{E}e^{t(X-\mathbb{E}X)} \geq 1$. By expansion (2.10), it suffices to compute $K_X^3(\zeta_t)$ for $X = g_\ell(M_k^{(B)}(z), \zeta)$. Then we apply Hölder's inequality (twice) to the first term on the second line of (2.11) and bound

$$\mathbb{E}[X^3 e^{t(X-\mathbb{E}X)}] \leq (\mathbb{E}e^{\frac{2}{\epsilon}t(X-\mathbb{E}X)})^{\frac{\epsilon}{2}}\mathbb{E}[X^6]^{\frac{1}{2}} \leq (C_0\ell^2)^{\frac{\epsilon}{2}}(C_0\ell\log\ell)^3.$$

To use Lemma 2.11 we need $\frac{2}{\epsilon}t < \frac{c_0}{\ell}$ so that we can take $c'_0 = \frac{c_0\epsilon}{2}$, thus the last inequality above follows from Lemma 2.11. The proof for the second inequality is analogous. □

Then we have the following concentration theorem:

Corollary 2.14. *Under the same assumptions in Lemma 2.11 and Lemma 2.12, denote by*

$$\mathrm{Pro}^z_{n,\ell}(\Pi, \Xi) := \log|\det(\Pi(\prod_{k=n}^{1} M_k^{(B)}(z))\Xi)|, \quad \mathrm{Pro}^z_{n,\ell}(\Xi) := \log\|(\prod_{k=n}^{1}\wedge^\ell M_k^{(B)}(z)\cdot\wedge^\ell\Xi)\|, \tag{2.12}$$

where $\Pi = [I_\ell, 0]$ and Ξ^* is taken from the Stiefel manifold $\mathrm{St}(\ell, 2\ell)$. Then the value of these maps does not depend on the representative of $[\Xi^*] \in \mathrm{Gr}(\ell, 2\ell)$ under the identification

$\mathrm{Gr}(\ell, 2\ell) = \mathrm{St}(\ell, 2\ell)/U(\ell)$. *Moreover, the following large deviation estimates hold: for any* $\epsilon > 0$ *we can find a constant* $c_0' > 0$ *depending only on* ϵ, $|z|$ *and* ζ *such that*

$$\mathbb{P}(\frac{1}{n\ell}|\operatorname{Pro}^z_{n,\ell}(\Pi, \Xi) - \mathbb{E}\operatorname{Pro}^z_{n,\ell}(\Pi, \Xi)| \geq \ell^{-\epsilon}) \leq \exp(-c_0' \frac{n}{\ell^{2\epsilon}(\log \ell)^2}),$$

$$\mathbb{P}(\frac{1}{n\ell}\left|\operatorname{Pro}^z_{n,\ell}(\Xi) - \mathbb{E}\operatorname{Pro}^z_{n,\ell}(\Xi)\right| \geq \ell^{-\epsilon}) \leq \exp(-c_0' \frac{n}{\ell^{2\epsilon}(\log \ell)^2}).$$

Proof. The independence of values on the representation follows from Cauchy–Binet formula and the fact that for any $R \in \mathbb{C}^{\ell\times\ell}$, we have that $\wedge^\ell(\Xi R) = (\det R) \wedge^\ell \Xi$.

We will prove the concentration in two steps. First, we form the conditional expectation of $\operatorname{Pro}^z_{n,\ell}(\Pi, \Xi)$ as follows: with $[\wedge^\ell \Xi] = X_0 \in \mathrm{Gr}(\ell, 2\ell)$,

$$\mathbb{E}_c \operatorname{Pro}^z_{n,\ell}(\Pi, \Xi) = \sum_{k=1}^{n-1} \mathbb{E} g_\ell(M_k^{(B)}(z), X_{k-1}) + \mathbb{E}\log|\langle \hat{f}, \wedge^\ell M_n^{(B)}(z) X_{n-1}|.$$

This conditional expectation is again a random variable and we have $\mathbb{E}[\mathbb{E}_c \operatorname{Pro}^z_{n,\ell}(\Pi, \Xi)] = \mathbb{E}\operatorname{Pro}^z_{n,\ell}(\Pi, \Xi)$. Then the difference $\operatorname{Pro}^z_{n,\ell}(\Pi, \Xi) - \mathbb{E}_c \operatorname{Pro}^z_{n,\ell}(\Pi, \Xi)$ forms a martingale sequence as we have subtracted the mean of each increment.

By Markov's inequality, for all $t > 0$ and any $\lambda > 0$ sufficiently small such that all the moment generating functions in the following are finite:

$$\begin{aligned}&\mathbb{P}(\operatorname{Pro}^z_{n,\ell}(\Pi, \Xi) - \mathbb{E}_c \operatorname{Pro}^z_{n,\ell}(\Pi, \Xi) \geq t)\\ &\quad\leq \exp\left(-\lambda t + \sum_{k=1}^{n-1} \sup_{\hat{g}\in\mathcal{C}_{dec}:\|\hat{g}\|=1} K_{g_\ell(M_k^{(B)}(z),\hat{g})}(\lambda) + \sup_{\hat{g}\in\mathcal{C}_{dec}:\|\hat{g}\|=1} K_{\log|\langle\wedge^\ell\Pi, \wedge^\ell M_n^{(B)}(z)\hat{g}\rangle|}(\lambda)\right),\end{aligned}$$

where we use iterative conditioning to decompose the sum in the exponential MGF, and take the sup over $\hat{g} \in \mathcal{C}_{dec} : \|\hat{g}\| = 1$ since the exponential MGF are uniformly bounded. Then we take $\lambda = \frac{c_0'}{\ell^{1+\epsilon}(\log \ell)^2}$ and $t = n\ell^{1-\epsilon}$ and apply Fact 2.13. By applying a similar argument via moment generating functions we can deduce that

$$\begin{aligned}&\mathbb{P}(\mathbb{E}_c \operatorname{Pro}^z_{n,\ell}(\Pi, \Xi) - \mathbb{E}\operatorname{Pro}^z_{n,\ell}(\Pi, \Xi) \geq t)\\ &\quad\leq \exp\left(-\lambda t + \sum_{k=1}^{n-1} K_{\mathbb{E}[g_\ell(M_k^{(B)}(z),X_{k-1})|X_{k-1}]}(\lambda) + K_{\mathbb{E}[\log|\langle\wedge^\ell\Pi,\wedge^\ell M_n^{(B)}(z)X_{n-1}\rangle||X_{n-1}]}(\lambda)\right),\end{aligned}$$

where $\mathbb{E}[g_\ell(M_k^{(B)}(z), X_{k-1}) \mid X_{k-1}]$ is a random variable depending on X_{k-1}, and similarly that $\mathbb{E}[\log|\langle\wedge^\ell\Pi, \wedge^\ell M_n^{(B)}(z)X_{n-1}\rangle| \mid X_{n-1}]$ is a random variable depending on X_{n-1} only. By Jensen's inequality, since x^2, x^6 and $\exp(x)$ are convex, we can easily verify that the moment estimates in Lemma 2.11 and Lemma 2.12, and hence the exponential moment generating function estimate in Fact 2.13 are valid for $K_{\mathbb{E}[g_\ell(M_k^{(B)}(z),X_{k-1})|X_{k-1}]}(\lambda)$ and $K_{\mathbb{E}[\log|\langle\wedge^\ell\Pi,\wedge^\ell M_n^{(B)}(z)X_{n-1}\rangle||X_{n-1}]}(\lambda)$ as well, so the same concentration follows.

The proof for the lower tail is analogous for both estimates and combining them justifies the concentration of $\operatorname{Pro}^z_{n,\ell}(\Pi, \Xi)$. The deviation estimates for $\operatorname{Pro}^z_{n,\ell}(\Xi)$ are exactly the same. □

Note that the upper bound in Lemma 2.11 is pessimistic and the $\log \ell$ factor might be necessary (as we upper bound the first ℓ singular values directly by the single largest one). However, the concentration in Corollary 2.14 is already strong enough for our purpose.

Remark 2.15. We can analogously prove a concentration inequality for

$$\frac{1}{n\ell}\sum_{i=1}^n \log|\det B_i| - \mathbb{E}\left[\frac{1}{n\ell}\sum_{i=1}^n \log|\det B_i|\right],$$

and a version with C_i in place of B_i, which has exactly the same quantitative form as Corollary 2.14. The proof idea is the same, so we omit the statement and proof.

2.5. **Decomposition of the ergodic sum.** Corollary 2.14 and the decomposition in Corollary 2.4 show that the core difficulty to proving the circular law for our band matrix model is to estimate the mean $\mathbb{E}\,\mathrm{Pro}_{n,\ell}^z(\Pi,\Xi)$. As the structure of $\mathrm{Pro}_{n,\ell}^z(\Pi,\Xi)$ is highly reminiscent of an ergodic transfer operator acting on $\mathcal{C}_{dec}\subset\mathcal{W}$, we expect that the value of $\frac{1}{n\ell}\mathbb{E}\,\mathrm{Pro}_{n,\ell}^z(\Pi,\Xi)$ will be asymptotically independent of the precise speed for n and ℓ tending to infinity, and asymptotically independent of the initial frame Ξ. The same asymptotics should also hold for $\mathrm{Pro}_{n,\ell}^z(\Xi)$. We now rigorously justify this heuristic and eventually use it to prove the circular law theorem for slowly growing ℓ. The following theorem is the most essential technical step in verifying these heuristics:

Theorem 2.16. *Take* $\Pi=[I_\ell,0]$ *and consider any frame* $\Xi\in\mathbb{C}^{2\ell\times\ell}$ *satisfying* $\det(\Xi^*\Xi)=1$. *Consider the block tridiagonal matrix* $\mathcal{T}_{n+2}$ *defined in equation* (2.1), *where all assumptions of Theorem 1.1 are verified. We assume that* $\ell^{1/12}\geq n\geq \ell^d$ *for some fixed* $d\in(0,\frac{1}{12})$. *Then for any* $z\in\mathbb{C}$, *the following limit holds in probability as* $n,\ell\to\infty$:

$$\frac{1}{n\ell}\log|\det(\mathcal{T}_{n+2}-zI_{mid})|\to\mathcal{U}(z)=\begin{cases}\frac{|z|^2-1}{2}, & |z|\leq 1,\\ \log|z|, & |z|>1.\end{cases}$$

We also have the convergence in expectation as $n,\ell\to\infty$:

$$\frac{1}{n\ell}\mathbb{E}\log|\det(\mathcal{T}_{n+2}-zI_{mid})|\to\mathcal{U}(z).$$

Moreover, the convergence rate in probability and in expectation can be quantified to be uniform over the choice of frame Ξ, *and the rate depends only on* ζ, $|z|$ *and* $d>0$.

The function $\mathcal{U}(z)$ is the well-known Ginibre potential and Theorem 2.16 is essentially stating the circular law for $\mathcal{T}_{n+2}$. What is important here is the relative magnitude of ℓ and n: it allows n to be polynomial growth in ℓ but still require the block size ℓ to be much larger than the number of blocks n, so that we are close to the delocalization regime and should have mean-field type behavior for the matrix (although this is not yet rigorously proven). We have stated the theorem in the special case $\Pi=[I_\ell,0]$ to simplify its proof as this is the only case we will actually use, but it is not hard to verify that the proof works for general Π via the same argument.

For a large n and a given ℓ, we use the following fact to decompose $[n]$ into smaller components so that Theorem 2.16 can be applied separately to each component.

Fact 2.17. *Let* ℓ *be a sufficiently large integer, and* $n\geq 10\ell^d$ *for some* $d>0$. *Then we can find an integer* n_0 *with* $2\ell^d\leq n_0\leq 4\ell^d$ *such that* $n-n_0\lfloor\frac{n}{n_0}\rfloor\geq\frac{1}{2}n_0$, *where* $\lfloor x\rfloor$ *is the integer part of a real number* x.

Corollary 2.18. *Assume that* $\ell^{1/12}\geq n\geq\ell^d$ *for some* $d>0$, *and take* $\Pi=(I_\ell,0)$. *Then for any* $\Xi\in\mathbb{C}^{2\ell\times\ell}$ *with* $\det(\Xi^*\Xi)=1$, *we have for any fixed* $z\in\mathbb{C}$,

$$\mathbb{E}\frac{1}{n\ell}\mathrm{Pro}_{n,\ell}^z(\Pi,\Xi)\to\mathcal{U}(z)+\frac{1}{2}\log 3+\frac{1}{2},\quad \mathbb{E}\frac{1}{n\ell}\mathrm{Pro}_{n,\ell}^z(\Xi)\to\mathcal{U}(z)+\frac{1}{2}\log 3+\frac{1}{2}$$

and the convergence rate is uniform over the choice of the frame Ξ.

Proof. For the convergence regarding $\mathrm{Pro}^z_{n,\ell}(\Pi,\Xi)$, we simply take this Π and Ξ and apply Theorem 2.16. Combining the determinant expression in Corollary 2.4, Proposition 2.9 with the concentration inequality (Remark 2.15) applied to $\sum_{k=1}^n \log|\det B_k|$, and the concentration inequality of $\mathrm{Pro}^z_{n,\ell}(\Pi,\Xi)$ in Corollary 2.14, we deduce that

$$\frac{1}{n\ell}\mathbb{E}\,\mathrm{Pro}^z_{n,\ell}(\Pi,\Xi)\to\mathcal{U}(z)+\frac{1}{2}\log 3+\frac{1}{2},$$

and the convergence is uniform over the choice of Ξ in the given joint limit of n,ℓ.

For the convergence relating to $\mathrm{Pro}^z_{n,\ell}(\Xi)$, we use that for $\hat{f}=\wedge^\ell[I_\ell,0]$:

$$\langle\hat{f},\prod_{k=n}^{1}\wedge^\ell M_k^{(B)}(z)\cdot\wedge^\ell\Xi\rangle=\langle\hat{f},\wedge^\ell M_n^{(B)}(z)X_{n-1}\rangle\|\prod_{k=n-1}^{1}\wedge^\ell M_k^{(B)}(z)\cdot\wedge^\ell\Xi\|. \tag{2.13}$$

Take the logarithm on both sides and then take the expectation. Take the conditional expectation of $\log|\langle f,\wedge^\ell M_n^{(B)}(z)X_{n-1}\rangle|$ over X_{n-1}, using Lemma 2.12 to get that the expectation of this term is bounded in absolute value by $O(\ell\log\ell)$ uniformly over $X_{n-1}\in\mathcal{C}_{dec}:\|X_{n-1}\|=1$. Then the second multiplicative term on the right hand side of (2.13) is precisely $\mathrm{Pro}^z_{n-1,\ell}(\Xi)$. Combined with the previous estimate, we conclude that

$$\frac{1}{n\ell}\mathbb{E}\,\mathrm{Pro}^z_{n-1,\ell}(\Xi)\to\mathcal{U}(z)+\frac{1}{2}\log 3+\frac{1}{2}-\frac{O(\ell\log\ell)}{n\ell},$$

and the last term vanishes since $n\geq\ell^d$. Finally, it suffices to replace $n-1$ by n. $\square$

With these technical results, we can quickly conclude the proof of Theorem 1.1.

Proof of Theorem 1.1. As we assume the atom variable ζ of the matrix T has finite second moment, it is standard to verify that $\|T\|_F^2$ is bounded in probability almost surely, justifying criterion (1) of Theorem 1.6.

By replacement principle (Theorem 1.6), the determinantal expansion in Corollary 2.4 and the almost sure convergence of $\frac{1}{n\ell}\sum_{k=1}^n\log|\det B_k|$ term (which follows from Proposition 2.9 and Remark 2.15), we only need to show the a.s. convergence of

$$\frac{1}{n\ell}\log\|\langle\wedge^\ell\Pi,\prod_{k=n}^{1}M_k^{(B)}(z)\cdot\wedge^\ell\Xi\rangle\| \tag{2.14}$$

to the deterministic limit $\mathcal{U}(z)+\frac{1}{2}\log 3+\frac{1}{2}$ for $\Pi=\begin{bmatrix}I_\ell,0\end{bmatrix}$ and $\Xi=\begin{bmatrix}0\\I_\ell\end{bmatrix}$.

Since $n\geq\ell_n^d$, once we take $\epsilon=d/3$, we are in the concentration regime of Corollary 2.14 so that (2.14) concentrates around its mean. By Borel-Cantelli, the convergence is almost sure since $\sum_{n\geq 1}\exp(-n^{d/3}/(\log\ell_n)^2)<\infty$. Thus it suffices to prove, as $n,\ell=\ell_n$ tend to infinity with $n\geq\ell_n^d$, the following convergence of deterministic quantities:

$$\frac{1}{n\ell}\mathbb{E}\log|\langle\wedge^\ell\Pi,(\prod_{k=n}^{1}\wedge^\ell M_k^{(B)}(z))\cdot\wedge^\ell\Xi\rangle|\to\mathcal{U}(z)+\frac{1}{2}\log 3+\frac{1}{2}. \tag{2.15}$$

By slightly varying the value of d, we can assume that $n\geq 10\ell^d$. By Fact 2.17, we can choose some n_0 with $2\ell^d\leq n_0\leq 4\ell^d$, noting that $d<\frac{1}{12}$. Denote by $\underline{n}=\lfloor\frac{n}{n_0}\rfloor$ and decompose the integer interval $[1,n]$ as the union $\{[(k-1)n_0+1,kn_0]\}_{1\leq k\leq\underline{n}}\cup[n-n_0\underline{n}+1,n]$.

The last interval has length at least $\frac{1}{2}n_0$ by Fact 2.17. Recall that we denote by $Q_k := \prod_{m=k}^{1} \wedge^\ell M_m^{(B)}(z) \cdot \wedge^\ell \Xi$ and $X_k = [Q_k]$ its projection on the Grassmannian $\mathrm{Gr}(\ell, 2\ell)$.

We work on the filtration generated by $\{A_k, B_k, C_k\}_{1\le k\le n}$, note that X_k depends on $\{A_i, B_i, C_i\}_{1\le i\le k}$, Then we can iteratively take the conditional expectation in (2.15) with respect to variables $X_{n-1}, X_{n-n_0\underline{n}}, X_{n-2n_0\underline{n}}, \cdots, X_{n_0}$ and rewrite the sum of (2.15) as

$$\frac{1}{n\ell}\mathbb{E}\left[\sum_{i=1}^{\underline{n}} \mathbb{E}\,\mathrm{Pro}^z_{n_0,\ell}(X_{(i-1)n_0}) + \mathbb{E}\,\mathrm{Pro}^z_{n-\underline{n}n_0,\ell}(\Pi, X_{n-\underline{n}n_0})\right]$$

(by Corollary 2.14, we can replace $X_k \in \mathrm{Gr}(\ell, 2\ell)$ by any of its representative in $\mathrm{St}(\ell, 2\ell)$),

By corollary 2.18, every term in the first summation in the bracket, divided by $n_0\ell$, converges to the limit $\mathcal{U}(z) + \frac{1}{2}\log 3 + \frac{1}{2}$ (the last term converges to the same limit divided by $(n - n_0\underline{n})\ell$). Moreover, the convergence is uniform for the first summation of i from 1 to $\underline{n}$. This completes the proof. □

3. Least singular value for the auxiliary matrix

In this section, we derive a least singular value estimate in Theorem 3.1 as the first step to the proof of Theorem 2.16.

We first prove the following high probability least singular value bound for $\mathcal{T}_{n+2} - zI_{mid}$, whose matrix structure we recall here:

$$\mathcal{T}_{n+2} - zI_{mid} = \begin{bmatrix} V & U & & & & \\ C_1 & A_1 - zI_\ell & B_1 & & & \\ & C_2 & A_2 - zI_\ell & B_2 & & \\ & & \ddots & \ddots & \ddots & \\ & & & C_n & A_n - zI_\ell & B_n \\ & & & & S_+^{bd} & C_+ \end{bmatrix} \in \mathbb{C}^{(n+2)\ell\times(n+2)\ell}, \quad (3.1)$$

Theorem 3.1. *Take the specialization $\Pi = [I_\ell, 0]$ so that $[S_+^{bd}, C_+] = [0, I_\ell]$. Let ζ satisfy the same conditions as in Theorem 1.1. Then we can find $C > 0$ depending only on $|z|$ and the law of ζ such that, whenever ℓ is sufficiently large,*

$$\mathbb{P}(s_{min}(\mathcal{T}_{n+2} - zI_{mid}) \le \ell^{-10n}(\ell^2 n)^{-10} t) \le Ct + \ell^{-5}, \quad t > 0.$$

The lower bound in Theorem 3.1 is exponentially small in the dimension, and we will only use this estimate when ℓ is much larger than n, say when $\ell \ge n^{12}$.

The proof of Theorem 3.1 is similar to [27], Section 2, and uses the invertibility via distance approach pioneered in [35]. However, the first block row and last block row of $\mathcal{T}_{n+2}$ are purely deterministic and $[V, U]$ is arbitrary: this adds to significant technical difficulty and are new in the literature.

To handle these generic boundary conditions, we first prove the following estimate:

Lemma 3.2. *Fix $L > 0$. Let ζ be a random variable which is either real with distributional density on $\mathbb{R}$ bounded by L, or has independent real and imaginary parts $\Re\zeta, \Im\zeta$ such that at least one of them has distributional density bounded by L. Let F be a $\ell \times 2\ell$ matrix with i.i.d. entries of law ζ, and $S : \mathbb{C}^{2\ell\times\ell}$ be a fixed (complex-valued) matrix with unitary columns. Let $Z \in \mathbb{C}^{\ell\times\ell}$ be an arbitrary fixed matrix. Then we can find $C_L > 0$ depending only on L such that for any $\epsilon > 0$,*

$$\mathbb{P}(s_{min}(FS - Z) \le \epsilon\ell^{-1.5}) \le C_L\epsilon. \quad (3.2)$$

The proof of Lemma 3.2 is given at the end of the section.

For the upper boundary $\begin{bmatrix} V & U \end{bmatrix}$, we denote by $P_{V,U}^{\perp}$ its unitary complement in $\mathbb{C}^{2\ell}$:

$$P_{V,U}^{\perp} = \{v \in \mathbb{C}^{2\ell} : \begin{bmatrix} V & U \end{bmatrix} \cdot v = 0\}. \tag{3.3}$$

We denote by $\begin{bmatrix} C_1, A_1 - zI_\ell \end{bmatrix} |_{P_{V,U}^{\perp}}$ the restriction of this $\ell \times 2\ell$ matrix to the subspace $P_{V,U}^{\perp}$.

We first introduce a family of high probability events on the matrix blocks. For $K > 0$,

$$\begin{aligned} \mathcal{E}_K := \{ &\forall i \in [n] : \|B_i\|, \|C_i\|, \|A_i\| \leq K; s_{min}(B_i) \geq \ell^{-10}, s_{min}(C_i) \geq \ell^{-10}, \\ &s_{min}(\begin{bmatrix} C_1, A_1 - zI_\ell \end{bmatrix} |_{P_{V,U}^{\perp}}) \geq \ell^{-10}\}. \end{aligned} \tag{3.4}$$

Fact 3.3. *For each $z \in \mathbb{C}$ we can find $K > 0$ such that $\mathbb{P}(\mathcal{T}_{n+2} \in \mathcal{E}_K) \geq 1 - K\ell^{-5}$.*

Proof. This follows from combining Lemma 2.7 or Lemma 2.6 with Lemma 2.8 and then apply Lemma 3.2. □

3.1. **Structure of normal vectors.** Let $T_1, \cdots, T_{(n+2)\ell}$ be the rows of $\mathcal{T}_{n+2} - zI_{mid}$ from top to bottom. For each $1 \leq i \leq (n+2)\ell$, let H_i be the linear span of all rows of $\mathcal{T}_{n+2} - zI_{mid}$ in $\mathbb{C}^{(n+2)\ell}$ except the i-th row. We prove the following structural theorem:

Proposition 3.4. *For any unit vector $v \in \mathbb{C}^{(n+2)\ell}$ we denote by $v_{[1]}, \cdots, v_{[n+2]}$ its restriction to the columns of the blocks labeled $1, 2, \cdots, n+2$. For each value $1 \leq k \leq (n+2)\ell$ we denote by $N_k = \lfloor \frac{k-1}{\ell} \rfloor + 1$, so that the k-th row lies in the N_k-th block from top to bottom. Then on the event $\mathcal{E}_K$, for each $\ell + 1 \leq k \leq (n+1)\ell$, let v be a unit vector orthogonal to H_k. Then for sufficiently large ℓ,*

$$\max(\|v_{[N_k-1]}\|_2, \|v_{[N_k]}\|_2, \|v_{[N_k+1]}\|_2) \geq \ell^{-10n}(\ell n)^{-1/2}.$$

Proof. In this proof we write $A_k^z := A_k - zI_\ell$ and we prove a slightly stronger result. Since v is a unit vector, there must exist a $j_0 \in [n+2]$ such that $\|v_{[j_0]}\| \geq (2n)^{-1/2}$. Without loss of generality assume that $j_0 < N_k - 1$ (the opposite case $j_0 > N_k + 1$ is analogous, and the special cases $j_0 = N_k, N_k - 1, N_k + 1$ already verify the claim).

Step 1: To the left of j_0. The normal vector v must satisfy the following equation

$$C_{j_0-2}v_{[j_0-2]} + A_{j_0-2}^z v_{[j_0-1]} + B_{j_0-2}v_{[j_0]} = 0.$$

Then on the event $\mathcal{E}_K$ we get that

$$\|B_{j_0-2}v_{[j_0]}\| \geq \ell^{-10}\|v_{[j_0]}\| \geq \ell^{-10}(2n)^{-1/2},$$

and that

$$\|C_{j_0-2}v_{[j_0-2]} + A_{j_0-2}^z v_{[j_0-1]}\| \leq (K + |z|)(\|v_{[j_0-2]}\| + \|v_{[j_0-1]}\|).$$

(Note that the invertibility of B_{j_0-2} is valid since $j_0 \leq N_k$ and we only remove a row of B_{N_k-1}). Combining both sides, we see whenever ℓ is large enough relative to K, z, then

$$\text{either } \|v_{[j_0-2]}\| \geq \ell^{-10}(2n)^{-1/2} \text{ or } \|v_{[j_0-1]}\| \geq \ell^{-10}(2n)^{-1/2}. \tag{3.5}$$

Then we take j_{-1} to be the smaller one of $j_0 - 1$ or $j_0 - 2$ satisfying (3.5). If $j_{-1} \geq 3$ we iterate the argument with j_{-1} and we can find $j_{-2} \in \{j_{-1} - 1, j_{-1} - 2\}$ with

$$\|v_{[j_{-2}]}\| \geq \ell^{-20}(2n)^{-1/2}.$$

Then we continue in this manner and find a sequence of indices $j_0, j_{-1}, \cdots, j_{-x}, x \leq n$ so that $j_{-x} \in \{1, 2\}$ and that for all $i \in [x]$, we have

$$|j_{-i} - j_{-i-1}| \leq 2, \text{ and } \|v_{[j_{-i}]}\| \geq \ell^{-10i}(2n)^{-1/2}.$$

(If $j_0=1$ or $j_0=2$ then this procedure is unnecessary).

Step 2: To the right of j_0. Since v solves the following equation

$$C_{j_0}v_{[j_0]}+A_{j_0}^z v_{[j_0+1]}+B_{j_0}v_{[j_0+2]}=0, \tag{3.6}$$

so long as $j_0<N_k-1$ so that C_{j_0} is invertible, we conduct exactly the same computation as in Step 1 to deduce that at least one of $\|v_{[j_0+1]}\|$ or $\|v_{[j_0+2]}\|$ has norm at least $\ell^{-10}(2n)^{-1/2}$. Iterating, we can find a sequence of indices $j_1,j_2,\cdots,j_y$, $y\leq n$, so that $j_y\in\{N_k-1,N_k\}$, and for all $i\in[y]$,

$$|j_i-j_{i-1}|\leq 2,\quad \text{and}\|v_{[j_i]}\|\geq \ell^{-10i}(2n)^{-1/2}.$$

This completes the proof. □

In Proposition 3.4, we have excluded values of k in the first and last block row. A special treatment is needed for them. We now introduce an auxiliary matrix:

$$\mathcal{T}_z^{res}:=\begin{bmatrix} C_1 & A_1-zI_\ell & B_1 & & & \\ & C_2 & A_2-zI_\ell & B_2 & & \\ & & \ddots & \ddots & \ddots & \\ & & & C_n & A_n-zI_\ell & \end{bmatrix}\in\mathbb{C}^{n\ell\times(n+1)\ell}.$$

This matrix is obtained from $\mathcal{T}_{n+2}$ by removing the first row and the last row and column. From now on we take the specialization $[S_+^{bd},C_+]=[0,I_\ell]$.

We will evaluate $\mathcal{T}_z^{res}$ on the following subset of $\mathbb{C}^{(n+1)\ell}$:

$$\mathrm{Vec}_{V,U}^\perp:=\{v\in\mathbb{C}^{(n+1)\ell}:(v_{[1]},v_{[2]})\in P_{V,U}^\perp\}.$$

To allow for a more convenient linear algebra treatment, we take an isometry on the first two components of $\mathbb{C}^{(n+1)\ell}$ and reduce $\mathcal{T}_z^{res}$ to the following square matrix:

Fact 3.5. *There exists a matrix C_2^* and a matrix $A_1^*:=[C_1,A_1-zI_\ell]\mid_{P_{V,U}^\perp}$ such that the following defined matrix*

$$\mathcal{T}_z^{res,sq}:=\begin{bmatrix} A_1^* & B_1 & & \\ C_2^* & A_2-zI_\ell & B_2 & \\ & \ddots & \ddots & \ddots \\ & & C_n & A_n-zI_\ell \end{bmatrix}\in\mathbb{C}^{n\ell\times n\ell}$$

satisfies

$$s_{min}(\mathcal{T}_z^{res,sq})=s_{min}(\mathcal{T}_z^{res}\mid_{\mathrm{Vec}_{V,U}^\perp}),$$

where C_2^* is purely a function of C_2 defined by $C_2^*=[0,C_2]\mid_{P_{V,U}^\perp}$.

This fact follows from applying an isometry on the first two components of $\mathbb{C}^{(n+1)\ell}$. We then prove the following geometric property for $\mathcal{T}_z^{res,sq}$:

Proposition 3.6. *Let $T_1^{res},\cdots,T_{n\ell}^{res}$ be the rows of $\mathcal{T}_z^{res,sq}$ from top to bottom. Also denote by H_i^{res} be the linear span of all rows of $\mathcal{T}_z^{res,sq}$ in $\mathbb{C}^{n\ell}$ except the i-th row, for all $1\leq i\leq n\ell$. Then on the event $\mathcal{E}_K$, for each $1\leq k\leq n\ell$, let v be a unit vector orthogonal to H_k^{res}, and that row k lies in the N_k-th block from top to bottom. Then for sufficiently large ℓ,*

$$\max(\|v_{[N_k-1]}\|_2,\|v_{[N_k]}\|_2,\|v_{[N_k+1]}\|_2)\geq \ell^{-10n}(\ell n)^{-1/2}.$$

(When we encounter $v_{[0]}$ or $v_{[n+1]}$, we simply remove them from the maximum bracket).

Proof. If $2 \leq N_k \leq n$, we can always find a $j_0 \in [n]$ such that $\|v_{[j_0]}\| \geq (2n)^{-1/2}$. Then if $j_0 = 1$, we apply the relation

$$A_1^* v_{[1]} + B_1 v_{[2]} = 0$$

and deduce that on $\mathcal{E}_K$, $\|v_{[2]}\| \geq \ell^{-10}(2n)^{-1/2}$. Then apply the relation

$$C_3 v_{[2]} + A_3^z v_{[3]} + B_3 v_{[4]} = 0$$

to deduce that $\max(\|v_{[3]}\|, \|v_{[4]}\|) \geq \ell^{-20}(2n)^{-1/2}$. Applying this iteratively completes the proof. Suppose that $2 \leq j_0 \leq N_k - 2$, then we start from $C_{j_0+1} v_{[j_0]} + A_{j_0+1}^z v_{[j_0+1]} + B_{j_0+1} v_{[j_0+2]} = 0$ and do the iteration. The other case $N_k + 1 \leq j_0 \leq n$ is exactly analogous.

If $N_k = 1$, and that $\|v_{[1]}\| \leq (2n)^{-1/2}$, then we can find some $2 \leq j_0 \leq n$ such that $\|v_{[j_0]}\| \geq (2n)^{-1/2}$. Then we apply the relation

$$C_{j_0-1} v_{j_0-2} + A_{j_0-1}^z v_{[j_0-1]} + B_{j_0-1} v_{[j_0]}$$

to deduce that $\max(\|v_{[j_0-1]}\|, \|v_{[j_0-2]}\|) \geq \ell^{-10}(2n)^{-1/2}$. Then we iterative all the way to the left and deduce that $\max(\|v_{[1]}\|, \|v_{[2]}\|) \geq \ell^{-10n}(2n)^{-1/2}$, completing the proof. □

Before proceeding to the next step, let us conclude that we have gained a good understanding of the geometry of the normal vector to the subspaces H_k^{res} of $\mathcal{T}_z^{res,sq}$ for all k, and the geometry of the normal vector to the subspaces H_k of $\mathcal{T}_{2n+2} - zI_{mid}$ except those in the first and last block. To complete the picture for $\mathcal{T}_{2n+2} - zI_{mid}$, we will first bound the least singular value of $\mathcal{T}_z^{res,sq}$ in the following section.

3.2. **Invertibility via distance.** We use the following simple geometric lemma to reduce singular value to distance estimates. The version we present is actually weaker than the version in [35] and [27], but it relieves us of the task of studying compressible/incompressible vectors separately, and the loss of quantitative bound is negligible for our purpose.

Lemma 3.7. *With the notations above, we have for any $t > 0$,*

$$\mathbb{P}(\mathcal{E}_K \cap \{s_{min}(\mathcal{T}_{n+2} - zI_{mid}) \leq t(2n\ell)^{-1/2}\}) \leq \sum_{k=1}^{\ell(n+2)} \mathbb{P}(\mathcal{E}_K \cap \{\mathrm{dist}(T_k, H_k) \leq t\}),$$

$$\mathbb{P}(\mathcal{E}_K \cap \{s_{min}(\mathcal{T}_z^{res,sq}) \leq t(n\ell)^{-1/2}\}) \leq \sum_{k=1}^{\ell n} \mathbb{P}(\mathcal{E}_K \cap \{\mathrm{dist}(T_k^{res}, H_k^{res}) \leq t\}).$$

Proof. Both estimates follow from the following general principle: let $E \in \mathbb{C}^{n \times n}$ and $v \in \mathbb{C}^n$ with $\|v\| = 1$ such that $\|Ev\| = s_{min}(E)$. There exists a coordinate $i \in [n]$ such that $\|v_i\| \geq n^{-1/2}$, and that we use

$$\|Ev\| \geq \mathrm{dist}(E_i v_i, \mathcal{H}_i) \geq n^{-1/2}\,\mathrm{dist}(E_i, \mathcal{H}_i)$$

where E_i is the i-th column and $\mathcal{H}_i$ the span of all other columns except the i-th. Then we take a union bound for the small ball events of $\mathrm{dist}(E_i, \mathcal{H}_i)$ over all i, and we use $s_{min}(E) = s_{min}(E^*)$ to switch the role of rows and columns. □

Before proceeding to bounding from below $s_{min}(\mathcal{T}_z^{res,sq})$, we would like to insert here the proof of Lemma 3.2 because we will reuse a key computation in the proof.

Proof of Lemma 3.2. Let F_i denote the i-th row of F and Z_i the i-th row of Z, then F_iS-Z_i is the i-th row of $FS-Z$. Let H_i be the subspace of $\mathbb{C}^\ell$ spanned by other rows of $FS-Z$ except the i-th row. Applying the negative second moment identity (see [41],Lemma A.4) to $FS-Z$,

$$\|(FS-Z)^{-1}\|_{HS}^2 \le \sum_{i=1}^{\ell} \operatorname{dist}(F_iS-Z_i,H_i)^{-2}. \tag{3.7}$$

Let $n_i \in \mathbb{C}^\ell$ be a unit normal to H_i, which is independent of F_i. Then

$$\operatorname{dist}(F_iS-Z_i,H_i) \ge |\langle F_iS-Z_i,n_i\rangle| = |\langle F_i,Sn_i\rangle - q| \tag{3.8}$$

where $q=\langle Z_i,n_i\rangle$. We have $\|Sn_i\|_2=1$ since S has unitary columns. When ζ is real-valued, we can choose $\theta\in[0,2\pi]$ such that $\|\Re(e^{-i\theta}Sn_i)\| \ge \frac{1}{2}$. Then

$$|\langle F_i,Sn_i-q\rangle| = |e^{-i\theta}\langle F_i,Sn_i\rangle - e^{-i\theta}q| \ge |\langle F_i,\Re(e^{-i\theta}Sn_i)-\Re(e^{-i\theta}q)\rangle|. \tag{3.9}$$

By the main result of [34], since F_i has i.i.d. coordinates of bounded density L, its projection to any 1-dimensional subspace of $\mathbb{R}^{2\ell}$ has density bounded by $C'L$ for some $C'>0$. Thus

$$\sup_{s\in\mathbb{R}} \mathbb{P}(|\langle F_i,\Re(e^{-i\theta}Sn_i)\rangle - s| \le t) \le C''t \tag{3.10}$$

for a C'' depending only on L. We take $t=\epsilon\ell^{-1}$ and sum up the estimate over all $i=1,\cdots,\ell$ by a union bound. Then we plug the estimate into (3.7) and finally complete the proof.

In the complex case, we condition on one component $\Re\zeta$ or $\Im\zeta$ and apply the argument to the other component with a bounded density to complete the proof. □

Now we state and prove the least singular value lower bound for $\mathcal{T}_z^{res,sq}$:

Proposition 3.8. *With the assumptions above, we have the following estimate for some $C>0$ depending on z and ζ:*

$$\mathbb{P}(s_{min}(\mathcal{T}_z^{res,sq}) \le t\ell^{-10n}(\ell^2 n)^{-1}) \le C(t+\ell^{-5}), \quad t>0. \tag{3.11}$$

Proof. We work on the event $\mathcal{E}_K$ since $\mathbb{P}(\mathcal{E}_K^c) \le \ell^{-5}$, and let n_k denote a unit normal vector to H_k^{res} for each k. Then when $2\le N_k \le n-1$, by Proposition 3.6 we have $\mathrm{Bor}_k := (n_{k[N_k-1]}, n_{k[N_k]}, n_{k[N_k+1]}) \in \mathbb{C}^{3\ell}$ satisfies $\|\mathrm{Bor}_k\|_2 \ge \ell^{-10n}(\ell n)^{-1/2}$, so that by independence,

$$\mathbb{P}(\operatorname{dist}(T_k^{res},H_k^{res}) \le t) \le \sup_{s\in\mathbb{C}} \mathbb{P}(|\langle \mathrm{Bor}_k, {T_k^{res}}_{[N_k-1,N_k,N_k+1]}\rangle - s| \le t) \le Ct\ell^{10n}(\ell^2 n)^{\frac{1}{2}}, \tag{3.12}$$

where ${T_k^{res}}_{[N_k-1,N_k,N_k+1]}$ is the restriction of T_k^{res} to columns in blocks N_k-1, N_k, N_k+1, and we use that this vector has i.i.d. entries of bounded density normalized by $\ell^{-1/2}$. The small ball probability Ct follows exactly as in the proof of Lemma 3.2. When $N_k=n$, we do the same but remove all the vectors with restriction $[N_k+1]=[n+1]$.

Finally, when $N_k=1$, suppose that $\|n_{k[2]}\| \ge \ell^{-10n}(\ell n)^{-1/2}$, we can condition on A_1^* and use the k-th row of B_1 to bound the small ball probability just as in (3.12). If this does not hold then we must have $\|n_{k[1]}\| \ge \ell^{-10n}(\ell n)^{-1/2}$ by Proposition 3.6. Let $A_1^*[k]$ denotes the k-th row of A_1^*, then it suffices to bound $\sup_{s\in\mathbb{C}} \mathbb{P}(|\langle A_1^*[k], n_{k[1]}\rangle - s| \le t)$. By the definition of A_1^*, $A_1^*[k] = \hat{\zeta}\cdot S$ where $\hat{\zeta}$ is a 2ℓ dimensional vector with i.i.d. entry of law $(3\ell)^{-1/2}\zeta$ and $S:\mathbb{C}^{2\ell}\to\mathbb{C}^\ell$ has unitary columns. Then we are precisely in the situation of (3.8), (3.9), (3.10) and thus the same small ball probability follows.

Combining all these estimates with Lemma 3.7, we complete the proof. □

3.3. **Back to the original matrix.** Now we return to $\mathcal{T}_{n+2}$. First, we complete the study of the geometric structure of its normal vectors:

Proposition 3.9. *We take the specialization $[S_+^{bd}, C_+] = [0, I_\ell]$. On the event $\mathcal{E}_K \cap \{s_{min}(\mathcal{T}_z^{res,sq}) \geq \ell^{-10n}(\ell^2 n)^{-4}\}$,*

(1) *For any $1 \leq k \leq \ell$, let v be a unit vector orthogonal to H_k, then $|\langle v, [V,U][k]\rangle| \geq \ell^{-10n}(\ell^2 n)^{-5}$, where $[V,U][k]$ is the k-th row of the matrix $\mathcal{T}_{n+2}$.*

(2) *Meanwhile, for any $(n+1)\ell + 1 \leq k \leq (n+2)\ell$, let v be a unit vector orthogonal to H_k, then the k-th entry of v has absolute value at least $\ell^{-10n}(\ell^2 n)^{-5}$.*

Proof. The proof to the two claims are analogous. For claim (1), we must have $v_{[n+2]} = 0$ for v to be the normal vector, using this lower boundary specialization $[0, I_\ell]$. We can take an orthogonal decomposition $v = v_1 + v_2$, with $v_1 \in P_{V,U}$ and $v_2 \in \mathrm{Vec}_{V,U}^\perp$ since the last block entry vanishes. Using again v is the normal vector of H_k, we see that v_1 must be exactly colinear with the k-th row of $[V, U]$ for this relation to hold. By orthogonality, $\|v_1\| + \|v_2\| = 1$ and $\mathcal{T}_{n+2}$ acts on v_2 just as via $\mathcal{T}_z^{res}$. By Fact 3.5, the latter matrix has the same least singular value as $\mathcal{T}_z^{res,sq}$, which is at least $\ell^{-10n}(\ell^2 n)^{-4}$. Meanwhile, v_1 is supported only on the first two coordinate blocks, so by the operator norm bound we have

$$\|\mathcal{T}_{n+2}^{\backslash\{k\}} \cdot v_1\| \leq 4K\|v_1\|,$$

where we denote by $\mathcal{T}_{n+2}^{\backslash\{k\}}$ as $\mathcal{T}_{n+2}$ with the k-th row removed. This combined with

$$\mathcal{T}_{n+2}^{\backslash\{k\}} \cdot (v_1 + v_2) = 0, \quad \|\mathcal{T}_{n+2}^{\backslash\{k\}} \cdot v_2\| = \|\mathcal{T}_{n+2} \cdot v_2\| \geq \ell^{-10n}(\ell^2 n)^{-4}\|v_2\|$$

yield the desired lower bound for $\|v_1\|$. Finally, recall that v_1 is colinear to $[V, U][k]$.

The proof for case (2) is exactly the same and omitted. □

Finally we complete the proof of the main result on lower bounding $\sigma_{min}(\mathcal{T}_{n+2} - zI_{mid})$:

Proof of Theorem 3.1. This is exactly the same as the proof of Proposition 3.8. By Proposition 3.8, we have that $\mathbb{P}(s_{min}(\mathcal{T}_z^{res,sq}) \leq \ell^{-10n}(\ell^2 n)^{-4}) \leq \ell^{-5}$. Then on the complement of this event and on $\mathcal{E}_K^c$, proceed as follows. For interior rows with $\ell + 1 \leq k \leq (n+1)\ell$, Proposition 3.6 offers the geometric description for the normal vector. One then takes the inner product with the k-th row which has i.i.d. entries with bounded density on blocks $[N_k - 1, N_k, N_k + 1]$. This implies the same small ball probability estimate as in (3.12). For boundaries where $k \leq \ell$ or $k \geq (n+1)\ell + 1$, Proposition 3.9 shows the normal vector v already has an overlap with an inner product of absolute value at least $\ell^{-10n}(\ell^2 n)^{-5}$ with the k-th row of $\mathcal{T}_{n+2} - zI_{mid}$. Combining all these with Lemma 3.7 completes the proof. □

3.4. **An upgraded tail estimate.** Currently, the least singular value bound in Theorem 3.1 has an additive error ℓ^{-5}. When we later deduce uniform integrability of log determinant in Section 4.4, we will need a version of Theorem 3.1 with a much stronger tail estimate. We present here a strengthening of Theorem 3.1 with better tail control:

Theorem 3.10. *In the same setting as Theorem 3.1, the following estimate holds for all $i \in \mathbb{N}_+$: whenever ℓ is sufficiently large,*

$$\mathbb{P}(s_{min}(\mathcal{T}_{n+2} - zI_{mid}) \leq \ell^{-10ni}(\ell^2 n)^{-10} t) \leq Ct + \ell^{-5i}, \quad t > 0. \tag{3.13}$$

Consequently, for each $\alpha \geq 1$ we can find $C_\alpha > 0$ depending only on α such that

$$\mathbb{E}[|\log|s_{min}(\mathcal{T}_{n+2} - zI_{mid})||^\alpha] \leq C_\alpha(n(\log \ell + \log n))^\alpha.$$

Proof. We will prove this theorem by modifying the proof of Theorem 3.1. for each $i \geq 1$ consider the following event $\mathcal{E}_K^i$:

$$\begin{aligned}\mathcal{E}_K^i :=&\{\forall i \in [n] : \|B_i\|, \|C_i\|, \|A_i\| \leq K^i; s_{min}(B_i) \geq \ell^{-10i}, s_{min}(C_i) \geq \ell^{-10i},\\ &s_{min}([C_1, A_1 - zI_\ell] \mid_{P_{V,U}^\perp}) \geq \ell^{-10i}\}.\end{aligned}$$

Then we have $\mathbb{P}(\mathcal{E}_K^i) \geq 1 - K\ell^{-5i}$. We can do the proof of Proposition 3.4 and 3.6 again on the event $\mathcal{E}_K^i$, with the net effect of replacing the $\ell^{-10n}(\ell n)^{-1/2}$ lower bound by $\ell^{-10ni}(\ell n)^{-1/2}$ in both results.

Then going through the proof of Proposition 3.8 we can verify that for each $i \in \mathbb{N}_+$,

$$\mathbb{P}(s_{min}(\mathcal{T}_z^{res,sq}) \leq t\ell^{-10ni}(\ell^2 n)^{-1}) \leq C(t + \ell^{-5i}), \quad t > 0. \tag{3.14}$$

Then working through the proof of Proposition 3.9 on $\mathcal{E}_K^i$ completes the proof of (3.13) for each $i \in \mathbb{N}_+$. The moment estimate is straightforward. $\square$

4. Convergence of the log determinant

With the least singular value estimate in Theorem 3.1, the proof of Theorem 2.16, i.e. the convergence of log determinant, is reduced to the control of small-ish singular values. We will do this via a rigidity estimate as in [27], [21]. However, two new technical difficulties arise here: (1) the tridiagonal block matrix does not have a doubly stochastic variance profile, and (2) we need to treat the arbitrary upper boundary $[V, U]$.

4.1. **Convergence of Stieltjes transform.** We shall derive a convergence rate for the Stieltjes transform to the circular law limit. Since $\mathcal{T}_{n+2}$ is not translational invariant, we consider a translational invariant substitute, the following periodic block matrix model :

$$\mathcal{T}_{n+2}^{per} = \begin{bmatrix} A_0 & B_0 & & & C_0 \\ C_1 & A_1 & B_1 & & \\ & \ddots & \ddots & \ddots & \\ & & C_n & A_n & B_n \\ B_{n+1} & & & C_{n+1} & A_{n+1} \end{bmatrix} \tag{4.1}$$

where $\mathcal{T}_{n+2}^{per}$ shares the same set of random matrices with those appearing in $\mathcal{T}_{n+2}$, but with two additional blocks C_0, B_{n+1} being independent from $\mathcal{T}_{n+2}$ and having the same distribution as the individual blocks of $\mathcal{T}_{n+2}$. Then we clearly see that $\operatorname{rank}(\mathcal{T}_{n+2}^{per} - zI_{(n+2)\ell} - \mathcal{T}_{n+2} + zI_{mid}) \leq 8\ell$, and for any $z \in \mathbb{C}$ we have

$$\operatorname{rank}\left((\mathcal{T}_{n+2}^{per} - zI_{(n+2)\ell})(\mathcal{T}_{n+2}^{per} - zI_{(n+2)\ell})^* - (\mathcal{T}_{n+2} - zI_{mid})(\mathcal{T}_{n+2} - zI_{mid})^*\right) \leq 24\ell. \tag{4.2}$$

The following convergence result is a variant of Proposition 5.3 and a similar result is also proven in [27], Theorem 3.1:

Proposition 4.1. *Assume that $\ell \geq n$, and take any fixed $z \in \mathbb{C}$. Assume that the entry law ζ satisfies the moment assumptions in Theorem 1.1. We denote by*

$$m_{n+2,z}^{per}(\xi) := \frac{1}{(n+2)\ell} \sum_{i=1}^{(n+2)\ell} [\lambda_i\left((\mathcal{T}_{n+2}^{per} - zI_{(n+2)\ell})(\mathcal{T}_{n+2}^{per} - zI_{(n+2)\ell})^*\right) - \xi]^{-1} \tag{4.3}$$

the Stieltjes transform for the empirical measure of $(\mathcal{T}_{n+2}^{per}-zI_{(n+2)\ell})(\mathcal{T}_{n+2}^{per}-zI_{(n+2)\ell})^*$. *Then we can find a deterministic probability measure* ν_z *on* $[0,\infty)$ *such that* $m_z(\xi):=\int_{\mathbb{R}}\frac{d\nu_z(x)}{x-\xi}$ *is the unique solution to the following equation*

$$m_z(\xi)=\left[\frac{|z|^2}{1+m_z(\xi)}-(1+m_z(\xi))\xi\right]^{-1}$$

satisfying $\Im(\xi m_z(\xi^2))>0$ *and* $\Im(m_z(\xi))>0$ *when ever* $\Im(\xi)>0$, *and we have the following convergence for all* $\{\xi\in\mathbb{C}:\Im\xi>0\}$: *for any* $c>0$, *with probability at least* $1-(\ell n)^{-100}$ *the following limit holds:*

$$|m_{n+2,z}^{per}(\xi)-m_z(\xi)|\leq\frac{C}{|\Im\xi|^8}\frac{(\ell n)^c}{\ell^{1/2}},$$

where C *depends only on* ζ *and* c.

The proof of Proposition 4.1 is deferred to the next section. Let $m_{n+2,z}(\xi)$ be the Stieltjes transform of empirical measure of $(\mathcal{T}_{n+2}-zI_{mid})(\mathcal{T}_{n+2}-zI_{mid})^*$, defined similarly as in (4.3) (but subtract I_{mid} instead of the identity matrix). We can derive the convergence of $m_{n+2,z}(\xi)$ via the following low rank update formula:

Lemma 4.2. *With the assumptions above, we have for any* $\xi\in\mathbb{C}:\Im\xi>0$ *and* $z\in\mathbb{C}$,

$$|m_{n+2,z}^{per}(\xi)-m_{n+2,z}(\xi)|\leq\frac{24\pi}{(n+2)|\Im\xi|}.$$

Proof. This follows from the following perturbation formula (see [2], Appendix A and B):

$$|\frac{1}{N}\operatorname{tr}(A-\xi)^{-1}-\frac{1}{N}\operatorname{tr}(B-\xi)^{-1}|\leq\frac{\pi}{|\Im\xi|}\frac{\operatorname{rank}(A-B)}{N}$$

for two square matrices A,B of size N, applied to the rank difference in (4.2). □

4.2. **Singular value rigidity estimate.** When we combine the convergence in Proposition 4.1 with the rank update in Lemma 4.2, a problem arises that the error caused by the rank perturbation, which has the order $1/n\Im\xi$, dominates the error in the convergence in Proposition 4.1 when ℓ is much larger than n. This increased error is not an artifact: since the matrix $\mathcal{T}_{n+2}$ does not have a doubly stochastic variance profile, its Stieltjes transform differs from the circular law density by a correction having the scale $\ell/n\ell=1/n$. In the polynomial regime $\ell=\text{Poly}(n)$, we still get a polynomial convergence rate of $m_{n+2,z}(\zeta)$ to $m_z(\zeta)$. However, since the least singular value estimate we get in Theorem 3.1 is exponentially small, a significantly weakened polynomial convergence rate will not permit us to conclude with the convergence of the log determinant, see the details in Section 4.3.

We will eliminate this problem by proving a rigidity estimate for the small singular values of $\mathcal{T}_{n+2}-zI_{mid}$, which states that with high probability most of its singular values are at least n^{-10}. Similar estimates are proven in [21] when the matrix has a doubly stochastic variance profile and without boundary. The reason why this rigidity holds for $\mathcal{T}_{n+2}$ is that the added upper and lower boundaries actually do not destroy the invertibility of the matrix near 0, and hence do not destroy singular value rigidity near 0. Also, the fact that we change from periodic to Dirichlet boundary condition (we do not have the upper right and lower left blocks) should also have negligible effect in small singular values. The precise rigidity result is the following:

Proposition 4.3. *In the setting of Theorem 2.16, for any $c > 0$, with probability at least $1-(n\ell)^{-100}$, in the interval $[0,(n\ell)^{-50}]$, the random matrix $\mathcal{T}_{n+2}-zI_{mid}$ has at most $\frac{(n\ell)^{1+c}}{\ell^{0.1}}$ numbers of singular values.*

The proof of Proposition 4.3 is rather technical and deferred to Section 5. The value of 50 in the exponent is inessential and can be tuned so long as it is polynomially small. This is the only place throughout the paper where we use the fact that ζ has finite p-th moment for all p: we choose this strong moment condition to considerably simplify the proof.

Proposition 4.3 is mainly used for very large ℓ, say when $\ell = (n\ell)^{0.95}$. Although this estimate does not appear to be strong enough, as it excludes a very large number, say $\frac{(n\ell)^{1+c}}{\ell^{0.1}}$ of singular values, it guarantees that the other singular values are at least polynomially small in ℓ, rather than exponentially small in ℓ as in Theorem 3.1.

4.3. **Passing to the log determinant.** We note here an upper bound for $\|\mathcal{T}_{n+2}\|$:

Lemma 4.4. *We have the following deterministic operator norm upper bound of $\mathcal{T}_{n+2}$:*

$$\|\mathcal{T}_{n+2}\| \leq 2+10(\max_{1\leq i\leq n}\|A_i\|+\max_{1\leq i\leq n}\|B_i\|+\max_{1\leq i\leq n}\|C_i\|).$$

This estimate can be verified by using the tridiagonal block structure of inner rows of $\mathcal{T}_{n+2}$, and that the top and bottom rows have operator norm bounded by 1.

Now we proceed to prove convergence of the log determinant to the Gaussian model. For $z\in\mathbb{C}$, let $G_z := G - zI_{(n+2)\ell}$ where G is a complex Ginibre matrix of size $(n+2)\ell$ and having normalized variance. Denote by $X_z := \mathcal{T}_{n+2}-zI_{mid}$ (we are not subtracting by identity matrix) and denote by $s_1(X_z)\geq s_2(X_z)\geq\cdots$ the singular values of X_z in decreasing order. Define $s_i(G_z)$ analogously. We denote by

$$\nu_{X_z}(\cdot) := \frac{1}{(n+2)\ell}\sum_{i=1}^{(n+2)\ell}\delta_{s_i^2(T_z)}(\cdot),\quad \nu_{G_z}(\cdot) := \frac{1}{(n+2)\ell}\sum_{i=1}^{(n+2)\ell}\delta_{s_i^2(G_z)}(\cdot)$$

the empirical measure of singular values of X_z, G_z. For two probability measures μ,ν supported on $[0,\infty)$, we define their Kolmogorov distance via

$$\|\mu-\nu\|_{[0,\infty)} := \sup_{x\geq 0}|\mu([0,x])-\nu([0,x])|.$$

We can obtain an upper bound for the Kolmogorov distance $\|\nu_{X_z}(\cdot)-\nu_{G_z}(\cdot)\|$ as follows:

Corollary 4.5. *In the setting of Theorem 2.16, where we assume that for some $d\in(0,\frac{1}{2})$ we have $\ell^{1/2}\geq n\geq \ell^d$, we can find a constant $x_d\in(0,1)$ depending only on d such that, with probability at least $1-(n\ell)^{-90}$, the following estimate holds for some constant C depending only on ζ:*

$$\|\nu_{X_z}(\cdot)-\nu_z(\cdot)\|_{[0,\infty)}\leq C\ell^{-x_d}.$$

Exactly the same estimate also holds for $\|\nu_{G_z}(\cdot)-\nu_z(\cdot)\|_{[0,\infty)}$.

The proof of Corollary 4.5 is deferred to the end of this section. We also use a convenient formula for the Kolmogorov distance:

Fact 4.6. *For two probability measures μ,ν on $[0,\infty)$, and $0<a<b$,*

$$\left|\int_a^b \log x d\mu(x) - \int_a^b \log x d\nu(x)\right| \leq 2[|\log b|+|\log a|]\|\mu-\nu\|_{[0,\infty)}.$$

Moreover, for any $\beta > 1$ we have

$$\left|\int_a^b |\log x|^\beta d\mu(x) - \int_a^b |\log x|^\beta d\nu(x)\right| \leq 2[|\log b|^\beta + |\log a|^\beta]\|\mu - \nu\|_{[0,\infty)}.$$

Proof. The first claim can be found in [27], Lemma 4.3. To prove the second claim, for any C^1 function φ on $[a, b]$, we write the integral as

$$\int_a^b \varphi(x)d\mu(x) = \varphi(b)\mu([a,b]) - \int_a^b \int_x^b \varphi'(t)dtd\mu(x) = \varphi(b)\mu([a,b]) - \int_a^b \frac{\mu([a,t]}{t}\varphi'_*(t)dt,$$

where $\varphi'_*(t)$ is the function satisfying $\varphi'(t) = \varphi'_*(t)/t$. Take the same computation for ν and subtract, then upper bound by $\|\mu - \nu\|_{[a,b]}(\varphi(b) + \int_a^b |\varphi'_*(t)|\frac{dt}{t})$. For our choice $\varphi(x) = |\log x|^\beta, \beta > 1$ we can compute that

$$\int_a^b |\varphi'(t)|dt = \beta \int_a^b \frac{|\log t|^{\beta-1}}{t}dt \leq |\log a|^\beta + |\log b|^\beta.$$

Combining both the decomposition at endpoint b as above, and the decomposition at endpoint a and then sum the two bounds yield the final factor $2(|\log a|^\beta + |\log b|^\beta)$. □

Now we can prove Theorem 2.16, the main result of this section:

Proof of Theorem 2.16: convergence in probability. Recall that we assume, for some very small $d > 0$, $\ell^{1/12} \geq n \geq \ell^d$. With probability at least $1 - \ell^{-5}$, we can assume the following four estimates hold simultaneously:

(1) The least singular values of both X_z, G_z are at least ℓ^{-11n}, by Theorem 3.1;
(2) $\|G_z\|, \|X_z\| \leq K < \infty$ for some constant $K > 0$, by Lemma 4.4 and Lemma 2.8;
(3) The convergence rate in Corollary 4.5 holds;
(4) The two matrices X_z, G_z have at most $\frac{(n\ell)^{1+c}}{\ell^{0.1}}$ numbers of singular values in the interval $[0, (n\ell)]^{-50}$ for any $c > 0$, by proposition 4.3.

Then on this event we decompose the log integration as follows:

$$\int_0^\infty \log x\nu_{X_z}(dx) - \int_0^\infty \log x\nu_{G_z}(dx) = \int_{I_1\cup I_2} \log x(\nu_{X_z}(dx) - \nu_{G_z}(dx)), \tag{4.4}$$

where $I_1 = [\ell^{-11n}, (n\ell)^{-50}]$ and $I_2 = [(n\ell)^{-50}, K]$. By Fact 4.6, the integration on I_2 is bounded in absolute value by

$$2[|\log K| + |\log(n\ell)^{-50}|]\|\nu_{X_z}(\cdot) - \nu_{G_z}(\cdot)\|_{[0,\infty)} = o(\ell^{-x_d/2}) \to 0.$$

The integration on I_1 is bounded in absolute value by

$$\frac{1}{n\ell}O(n\log\ell\frac{(n\ell)^{1+c}}{\ell^{0.1}}) = o(\ell^{-0.01}) \to 0,$$

whenever we take $c > 0$ sufficiently small, and we use the assumption $\ell \geq n^{12}$. Combining the two integrations on I_1, I_2 completes the proof of convergence in probability. □

Finally, we complete the proof of Corollary 4.5:

Proof of Corollary 4.5. Recall that we assume, for some $d \in (0, \frac{1}{2})$, that $n \geq \ell^d$ and $\ell \geq n^{12}$. Then we can find some $x_d \in (0, (1-d)/16)$ such that $\frac{x_d+d/2}{x_d+1} \leq d$. For any $D > 0$, consider the strip

$$L := \{\zeta = \theta + i(\frac{n}{\ell})^{x_d} : \quad -D \leq \theta \leq D\}.$$

Then on L we have $n\Im\xi = n(\frac{n}{\ell})^{x_d} \geq \ell^{d/2}$, so that Lemma 4.2 yields, for all $\xi \in L$:

$$|m_{n+2,z}(\xi) - m_{n+2,z}^{per}(\xi)| \leq \frac{1}{\ell^{d/2}}.$$

By Proposition 4.1, we have for each $\xi \in L$, for any $c > 0$ with probability $1 - (\ell n)^{-100}$,

$$|m_{n+2,z}^{per}(\xi) - m_z(\xi)| \leq \frac{C(\ell n)^c}{n^{8x_d}\ell^{1/2-8x_d}} \leq \frac{C(\ell n)^c}{\ell^{d/2}}.$$

We can take $c > 0$ sufficiently small relative to d so that the right hand side is at most $C\ell^{-d/3}$. Since both $m_{n+2,z}^{per}$ and m_z are Lipschitz in ξ with Lipschitz constant bounded by $|\Im\xi|^{-1}$, we can extend the convergence to all of L and deduce that with probability at least $i - (\ell n)^{-90}$, uniformly for all $\xi \in L$ and for a possibly different constant C,

$$|m_{n+2,z}(\xi) - m_z(\xi)| \leq C\ell^{-d/3}. \tag{4.5}$$

Then by [2], Corollary B.15 and Lemma 11.9, just as in the proof of [27], Lemma 4.4, we have the following estimate with $\eta = \Im\xi$:

$$\|\nu_{X_z}(\cdot) - \nu_z(\cdot)\|_{[0,\infty)} \leq C\left[\int_{-D}^{D} |m_{n+2,z}(\xi) - m_z(\xi)|d\xi + \sqrt{\eta}\right]$$

for some universal $C > 0$. Taking the estimate (4.5) inside, we get that with probability at least $1 - (n\ell)^{-90}$,

$$\|\nu_{X_z}(\cdot) - \nu_z(\cdot)\|_{[0,\infty)} \leq C(\ell^{-d/3} + (\frac{n}{\ell})^{x_d}) \leq C(\ell^{-d/3} + \ell^{-x_d/2}),$$

where the last inequality follows from the $\ell \geq n^2$. Then we replace x_d by $\min(d/3, x_d/2)$ and complete the proof. $\square$

4.4. **Uniform integrability of the log determinant.** In this section, we upgrade the convergence of log determinants in Proposition 2.9 and Theorem 2.16 to convergence in expectation, by establishing uniform integrability.

Proof of Proposition 2.9. The part of almost-sure convergence is essentially an intermediate technical step in the proof of circular law in [41], where they proved that, taking G the same size matrix with i.i.d. complex Gaussian entries of unit variance, then the following limit holds both in probability and almost surely:

$$\frac{1}{n}\log|\det(n^{-1/2}A)| - \frac{1}{n}\log|\det(n^{-1/2}G)| \to 0, \quad n \to \infty.$$

It is well-known that, by the Ginibre potential computations,

$$\frac{1}{n}\log|\det(n^{-1/2}G)| \to \int_{|w|\leq 1} \log|w|\frac{dA(w)}{\pi} = -\frac{1}{2}, a.s., \quad n \to \infty,$$

where $dA(w)$ is the two-dimensional Lebesgue measure.

To prove convergence in probability, we check uniform integrability by proving that

$$\mathbb{E}[|\frac{1}{n}\log|\det n^{-1/2}A||^2] < \infty.$$

We let $s_1 \geq s_2 \geq \cdots \geq s_n$ denote the singular values of $n^{-1/2}A$ in decreasing order, then by Cauchy-Schwarz we have

$$|\frac{1}{n}\log|\det n^{-1/2}A||^2 \leq \frac{1}{n}\sum_{i=1}^{n}|\log s_i|^2.$$

Let ν_A denote the empirical measure of singular values of $(n^{-1/2}A)(n^{-1/2}A)^*$. Then by the same argument as in Corollary 4.5, we can find some $c_0 > 0$ such that with probability $1 - n^{-10}$, $\|\nu_A(\cdot) - \nu_0(\cdot)\|_{[0,\infty)} \leq n^{-c_0}$. Let Ω_A denote the event where this convergence holds, and moreover that $\|n^{-1/2}A\| \leq K$ for some large $K > 0$, and $s_{min}(n^{-1/2}A) \geq n^{-9}$. Then by the previous theorems, we have $\mathbb{P}(\Omega_A) \geq 1 - n^{-5}$.

On the event Ω_A, we apply Fact 4.6 with $\beta = 2$ to verify that

$$\left| \frac{1}{n} \sum_{i=1}^{n} |\log s_i|^2 - \int_{n^{-18}}^{K^2} |\log x|^2 \nu_0(dx) \right| \leq 2(2|\log K| + 20 \log n) n^{-c_0} = o(1), \tag{4.6}$$

and it also follows from a standard computation that $\int_0^\infty |\log x|^2 \nu_0(dx) < \infty$.

On the event Ω_A^c, we take the trivial upper bound

$$\frac{1}{n} \log |\det n^{-1/2}A| \leq \max(\log s_{max}(n^{-1/2}A), \log s_{min}(n^{-1/2}A)),$$

and we use Cauchy-Schwarz in the following inequality to get

$$\mathbb{E}[1_{\Omega_A^c} |\frac{1}{n} \log |\det n^{-1/2}A||^2] \leq \mathbb{P}(\Omega_A^c) \mathbb{E}[\max(\log s_{max}(n^{-1/2}A), \log s_{min}(n^{-1/2}A))^4]^{\frac{1}{2}}, \tag{4.7}$$

which is upper bounded by $n^{-9} \log^2 n \to 0$ thanks to Lemma 2.8 and 2.7. This verifies the uniform boundedness of $\mathbb{E}[|\frac{1}{n} \log |\det n^{-1/2}A||^2]$, which completes the proof of uniform integrability and thus the convergence in expectation.

Finally, we note that to obtain a quantitative convergence rate in expectation from a convergence rate in probability, we can use the following simple consequence of Hölder inequality. Suppose that random variables X_n converges to X in probability with a tail $\mathbb{P}(|X_n - X| \geq t) \leq r_n(t)$ for all $t \geq 0$ and that $\mathbb{E}|X_n|^2 + \mathbb{E}|X|^2 \leq M$ for some $M > 0$, then for any $t > 0$, $\mathbb{E}|X_n - X| \leq t + 2\sqrt{M \cdot r_n(t)}$. □

The convergence in expectation of Theorem 2.16 can be proven similarly:

Proof of Theorem 2.16 : convergence in expectation. We let $N = (n+2)\ell$ and $s_1, \cdots, s_N$ the singular value of the matrix $\mathcal{T}_{n+2} - zI_{mid}$. Let Ω_T be the event where all the four criterions (1) to (4) in the (in probability convergence part of the) proof of Theorem 2.16 hold true. Then on Ω_T, we can bound $\sum_{i=1}^{N} \frac{1}{N} \log^2 s_i$ by separating the sum over singular values in I_1 and I_2 defined there, with the summation over I_1 converges to 0 and the summation over I_2 can be estimated similarly to (4.6).

We have $\mathbb{P}(\Omega_T) \geq 1 - N^{-4.5}$, and on Ω_T^c we similarly apply Cauchy-Schwarz to upper bound the quantity as in (4.7). By Theorem 3.10, we can verify that

$$\mathbb{E}[|\log |s_{min}(\mathcal{T}_{n+2} - zI_{mid})||^4] \leq C(N \log N)^4$$

for some $C > 0$ depending only on ζ (and a better tail for s_{max} follows from Lemma 4.4 and 2.8). Since $N^{-4.5}$ dominates $N^2 \log^2 N$, the expectation over Ω_T^c again vanishes. □

5. Singular value rigidity estimate for a general boundary case

In this section, we prove Proposition 4.3, the rigidity bound of small singular values of $\mathcal{T}_{n+2} - zI_{mid}$.

5.1. **Removal of the boundary.** The standard method to prove rigidity estimates is to derive a local law for $\mathcal{T}_{n+2}$ and bound the imaginary part of its Stieltjes transform. However, the general boundary on the first row of $\mathcal{T}_{n+2}$ makes the derivation of local law extremely complicated. Here we take a few steps to remove the boundary and reduce to rigidity estimates of a more canonical model.

Lemma 5.1. *We can find a $2\ell \times \ell$ matrix R with unitary columns, where we denote by $R = \begin{bmatrix} R_1 \\ R_2 \end{bmatrix}$ its two components of size ℓ, such that the following statement holds on the event $\{\|\mathcal{T}_{n+2} - zI_{mid}\| \le K\}$ for some $K > 1$. For any $k \in \mathbb{N}_+$ and $\epsilon \in (0, \frac{1}{36K^2})$, suppose that $\mathcal{T}_{n+2} - zI_{mid}$ has at least k singular values in the interval $[0, \frac{\epsilon}{3K\sqrt{2n\ell}}]$, then the following matrix Y_z^R has at least k singular values in the interval $[0, \epsilon]$,*

$$Y_z^R := \begin{bmatrix} [C_1, A_1 - zI_\ell]\, R & B_1 & & \\ C_2R_2 & A_2 - zI_\ell & B_2 & \\ & & \ddots\ \ddots\ \ddots & \\ & & C_n & A_n - zI_\ell \end{bmatrix}.$$

Proof. Recall that we take the specialization $[S_+^{bd}, C_+] = [0, I_\ell]$, so the lower boundary is much easier to get rid of. Thus we focus on removing the upper boundary, and the process of removing the lower boundary is exactly the same. We take a unitary transform on the first two block columns of $\mathcal{T}_{n+2}$ to transform $[V, U]$ to $[I_\ell, 0]$, and denote by $Y_Z^{R,layer}$ the matrix form of the transformed matrix. This transform does not change the singular values. Let $v_1, \cdots, v_k$ be k mutually orthogonal unit vectors that are associated with a singular value in $[0, \epsilon]$ for this transformed matrix. Then for each vector v_i, we have that the mass on its first component satisfies $\epsilon \ge \|[v_i]_{[1]}\|, \quad \forall 1 \le i \le k$.

Let P_S be the projection onto the first ℓ coordinates of $\mathbb{C}^N$, and denote by $v_i' = (1-P_S)v_i$ for each i. Consider any vector $z = a_1v_1 + \cdots + a_kv_k$ with $\|z\| = 1$, then $\sum_i a_i^2 = 1$. We have

$$\|P_Sz\|_2 \le \sum_i a_i\epsilon \le \sqrt{2n\ell}\epsilon$$

since we always have $k \le (n+2)\ell$. Under the assumption $2n\ell\epsilon^2 \le \frac{1}{4}$, this implies that $\{v_i'\}$ are linearly independent. Let W be the subspace of $\mathbb{C}^{(n+2)\ell}$ spanned by $\{v_i'\}_{i=1}^k$, so that W has dimension k. Note that $Y_z^R \cdot P_Sv = P_SY_z^{R,layer}v$.

We claim that all singular values of Y_Z^R in W are less than $3K\sqrt{2n\ell}\epsilon$. For any $u \in W$ we find $z \in \mathrm{Span}(v_1, \cdots, v_k)$ with $u = (1 - P_S)z$, then $\|z\| \le 1.5\|u\|$ under $2n\ell\epsilon^2 \le \frac{1}{4}$ and

$$\|Y_z^Ru\| \le \|Y_z^{R,layer}u\| + K\|P_Sz\| \le (\epsilon + 2K\sqrt{2n\ell\epsilon^2})|u| \le 3K\sqrt{2n\ell}\epsilon|u|.$$

We then replace ϵ by $3K\sqrt{2n\ell}\epsilon$. By the variational characterization of least singular value, we complete the proof. □

Next, we reduce the rigidity of Y_z^R to the case where $B_1 = C_2R_2 = 0$, by showing that rigidity of Y_z^R can be deduced from rigidity of its lower $[2, n] \times [2, n]$ block-principal minor.

Lemma 5.2. *Let $Y_z^{R,diag}$ denote the block diagonal matrix obtained from Y_z^R by setting $B_1 = 0$ and $C_2R_2 = 0$. Then with probability at least $1-C\ell^{-8}$ with respect to the randomness of B_1 and C_2, the following is true when $\ell \ge n$: Suppose that $Y_z^{R,diag}$ has at most k singular values in the interval $[0, \ell^{-10}]$, then Y_z^R has at most k singular values in the interval $[0, \ell^{-30}]$.*

In other words, this lemma shows that additional randomness in B_1 and C_2 will not destroy the singular value rigidity of $Y_z^{R,diag}$ too much.

Proof. Let v be a unit vector such that $\|Y_z^{R,diag}v\| \geq \ell^{-10}$, then we claim that with probability at least $1-C\ell^{-10}$ we have $\|Y_z^R v\| \geq \ell^{-30}$. Assume for contradiction that $\|Y_z^R v\| \leq \ell^{-30}$, then taking the difference, we have

$$\|(Y_z^R - Y_z^{R,diag})v\| = \sqrt{\|B_1 v_{[2]}\|^2 + \|C_2 R_2 v_{[1]}\|^2} \geq \frac{1}{2}\ell^{-10}.$$

We condition on an event that $\|B_1\| \leq K, \|C_2 R_2\| \leq K$ for a large K, which has probability at least $1-\ell^{-50}$ by Lemma 2.8. Then on this event we must have $\max(\|v_{[2]}\|, \|R_2 v_{[1]}\|) \geq \frac{1}{4K}\ell^{-10}$. Combining the two relations $\|Y_z^{R,diag}v\| \geq \ell^{-10}$ and $\|Y_z^R v\| \leq \ell^{-30}$, this implies that for some $w_1, w_2 \in \mathbb{C}^\ell$ depending only on $Y_z^{R,diag}$ we must have $\|B_1 v_{[2]} - w_1\| \leq \ell^{-30}$ and $\|C_2 R_2 v_{[1]} - w_2\| \leq \ell^{-30}$. Assume without loss of generality that $\|v_{[2]}\| \geq \frac{1}{4K}\ell^{-10}$, then by the least singular value estimate in Lemma 2.7, and conditioning on Y_z^R and the operator norm bound, the latter event has probability at most $C\ell^{-10}$. The matrix has $\ell(n+2) \leq 2\ell^2$ singular values, so taking the union bound over at most $2\ell^2$ events, we conclude that with probability at least $1-C\ell^{-10}$, Y_z^R has singular value at least ℓ^{-30} on the linear subspace spanned by singular vectors of $Y_z^{R,diag}$ associated with singular value at least ℓ^{-10}. □

Now that $Y_z^{R,diag}$ has a block diagonal form and for its top left corner $[C_1, A_1 - zI_\ell]R$ we already have good control of its least singular value by Lemma 3.2, so that we only need to obtain rigidity estimates for its $[2,n]$ principal minor. The latter matrix $[Y_z^{R,diag}]_{[2,n]\times[2,n]}$ has exactly the same form as our block tridiagonal matrix T (1.1).

5.2. **The canonical block tridiagonal case: convergence to MDE.** We proceed using a local law approach to derive rigidity estimates for $[Y_z^{R,diag}]_{[2,n]\times[2,n]}$, which is a block tridiagonal matrix. There has been well-developed techniques to show that the Green function of $[Y_z^{R,diag}]_{[2,n]\times[2,n]}$ converges to the solution to a set of matrix Dyson equations (MDE), but the main challenge in our case is that the system is not translationally invariant due to the boundaries, so the solution to the MDE cannot be obtained in a closed form.

We let $\mathcal{T}_n$ be the block tridiagonal matrix (1.1) of size $n\ell$ and block size ℓ. We use the following set of parameters: $W = \ell$, and $N = n\ell = nW$.

We take the notation

$$Y_z = \mathcal{T}_n - zI_N, \tag{5.1}$$

and we sometimes abbreviate the subscript z by simply writing $Y = Y_z$. We define the entry-wise variance of Y_z as the following matrix $S \in \mathbb{R}^{N\times N}$:

$$S = (s_{ij}), \quad s_{ij} = \operatorname{Var}(\mathcal{T}_n)_{ij},$$

The main object of our study is the Green function of $Y_z^* Y_z$, and its trace:

$$G(w) := G(w,z) = (Y_z^* Y_z - w)^{-1}, \quad m(w) := m(w,z) = \frac{1}{N}\operatorname{Tr} G(w,z), \quad w = E + i\eta. \tag{5.2}$$

We also consider the Green function $Y_z Y_z^*$:

$$\mathcal{G}(w) := \mathcal{G}(w,z) = (Y_z Y_z^* - w)^{-1}. \tag{5.3}$$

A standard route to study $G, \mathcal{G}$ is to take a reduction to the Hermitization

$$\mathcal{H}_z := \begin{bmatrix} 0 & Y_z \\ Y_z^* & 0 \end{bmatrix}$$

and denote by

$$\mathcal{HG}(w) := \mathcal{HG}(w,z) = (\mathcal{H}_z - w)^{-1}.$$

These resolvents satisfy the following linear algebra identities: for any $\lambda \in \mathbb{C} : \Im\lambda > 0$,

$$\mathcal{G}(\lambda^2) = \frac{1}{\lambda}[\mathcal{HG}(\lambda)]_{[1,N]\times[1,N]}, \quad G(\lambda^2) = \frac{1}{\lambda}[\mathcal{HG}(\lambda)]_{[N+1,2N]\times[N+1,2N]}. \tag{5.4}$$

The strategy of this subsection is as follows. We first show, by matrix concentration inequality in Proposition 5.3, that $\mathcal{HG}(w)$ will converge to the unique solution $M^{\mathcal{H}}(w)$ with positive imaginary part to the following equation

$$-M^{\mathcal{H}}(w)^{-1} = wI_{2N} + \begin{bmatrix} 0 & zI_N \\ \bar{z}I_N & 0 \end{bmatrix} + \mathcal{S}^{\mathcal{H}}[M^{\mathcal{H}}(w)], \tag{5.5}$$

where $\mathcal{S}^{\mathcal{H}} : \mathbb{C}^{2N\times 2N} \to \mathbb{C}^{2N\times 2N}$ is the following defined self-energy operator

$$\mathcal{S}^{\mathcal{H}}[X] = \mathbb{E}\left[\begin{bmatrix} 0 & \mathcal{T}_n \\ \mathcal{T}_n^* & 0 \end{bmatrix} X \begin{bmatrix} 0 & \mathcal{T}_n \\ \mathcal{T}_n^* & 0 \end{bmatrix}^*\right]. \tag{5.6}$$

Then one checks that $\mathcal{S}^{\mathcal{H}}$ is block diagonal and acts via two diagonal maps:

$$\mathcal{S}^{\mathcal{H}}\begin{bmatrix} X_{11} & X_{12} \\ X_{21} & X_{22} \end{bmatrix} = \begin{bmatrix} \Phi[X_{22}] & 0 \\ 0 & \widetilde{\Phi}[X_{11}] \end{bmatrix},$$

where each block has size $N \times N$. The two maps $\Phi, \widetilde{\Phi}$ acts diagonally with the expression

$$(\Phi[Z])_{ii} = \sum_c s_{ic} Z_{cc}, \quad (\widetilde{\Phi}[W])_{cc} = \sum_i s_{ic} W_{ii},$$

where the summation are over $[N]$.

Then from (5.4), we deduce that $(\mathcal{G}(w), G(w))$ should converge with high probability to $(M_1(w), M_2(w))$ such that

$$M_1(w^2) = \frac{1}{w}[M^{\mathcal{H}}(w)]_{[1,N]\times[1,N]}, \quad M_2(w^2) = \frac{1}{w}[M^{\mathcal{H}}(w)]_{[N+1,2N]\times[N+1,2N]}. \tag{5.7}$$

The precise matrix concentration result for convergence of $\mathcal{HG}(w)$ towards $M^{\mathcal{H}}(w)$ is stated as follows:

Proposition 5.3. *Let ζ satisfy the moment assumptions in Theorem 3.1, and assume $W \geq \sqrt{N}$ (indeed $W = N^{c'}$ suffices for any c'). Then for any sufficiently small $c > 0$, we can find a constant $C > 0$ depending only on ζ and c such that whenever N is large enough, the following holds with probability at least $1 - N^{-100}$: for any $w \in \mathbb{C}, \Im w > 0$,*

$$\sup_{1\leq i\leq N} |[\mathcal{HG}(w)]_{ii} - [M^{\mathcal{H}}(w)]_{ii}| \leq \frac{CN^c}{\sqrt{W}(\Im w)^5},$$

$$\sup_{1\leq i\leq N} |[G(w^2)]_{ii} - [M_1(w^2)]_{ii}| \leq \frac{CN^c}{\sqrt{W}(\Im w)^6}.$$

Proof of Proposition 5.3. The proof uses exactly the same strategy in [20], [22] and in [21], Section 4.3. We use the matrix concentration inequality in [7], at the resolvent level to compare $\mathcal{HG}(w)$ to the MDE solution $M^{\mathcal{H}}(w)$. As the atom variable ζ has all moments finite, we take the same truncation argument as in [21], Section 4.3 and the quantitative estimates stated there directly carry over to the current setting. The only difference compared to [21] and [22] is that the variance profile of $\mathcal{T}_n$ is not doubly stochastic, but the machinery in [7] does not rely on double stochasticity. Finally, we use (5.7) to project the comparison of

$M^{\mathcal{H}}(w)$ and $\mathcal{H}\mathcal{G}(w)$, to the comparison of $G(w), \mathcal{G}(w)$ with $M_2(w), M_1(w)$ and pick up an additional w^{-1} factor. □

The precise polynomial rate of convergence in Proposition 5.3 is not important, so long as the rate is polynomial in W and $\Im w$.

We now derive the self-consistency equations solved by M_1 and M_2:

Lemma 5.4. *Since $M^{\mathcal{H}}(w)$ solves (5.5), then $M_1(w), M_2(w)$ should individually solve the following two (coupled) matrix Dyson equations:*

$$-M_1(w)^{-1} = wI_N + w\Phi[M_2(w)] - |z|^2(I_N + \widetilde{\Phi}[M_1(w)])^{-1}, \tag{5.8}$$

and likewise

$$-M_2(w)^{-1} = wI_N + w\widetilde{\Phi}[M_1(w)] - |z|^2(I_N + \Phi[M_2(w)])^{-1}. \tag{5.9}$$

The proof is presented at the end of this section.

We now check that the solution M_1, M_2 are diagonal matrices with constant diagonal value per block. For each $1 \leq k \leq n$ we denote by m_k the Stieltjes transform of the k-th diagonal block of M_1, i.e.

$$m_k = \frac{1}{\ell} \operatorname{Tr}[M_1(w)]_{[(k-1)\ell+1,k\ell]\times[(k-1)\ell+1,k\ell]}.$$

We define $\widetilde{m}_1, \cdots, \widetilde{m}_n$ similarly as the normalized Stieltjes transform of each diagonal block of $M_2(w)$.

Fact 5.5. *The solution $M_1(w)$ is diagonal and has the form*

$$M_1(w) = \operatorname{diag}(m_1 I_\ell, m_2 I_\ell, \cdots, m_n I_\ell).$$

The other solution $M_2(w)$ also has a similar form

$$M_2(w) = \operatorname{diag}(\widetilde{m}_1 I_\ell, \widetilde{m}_2 I_\ell, \cdots, \widetilde{m}_n I_\ell).$$

Moreover, $m_i(w) = \tilde{m}_i(w)$ for all $1 \leq i \leq n$, so we are reduced to only one set of equation.

Proof. Since each random block in $\mathcal{T}_n$ is i.i.d., one can check that the map $\Phi[M_2(w)]$ depends only on the trace of each block and is constant on each block. Thus $M_1(w)$ is also constant on each diagonal block and is a block identity matrix with diagonal entry m_k on the k-th block. One can check that $M_2(w)$ also has the same form. Then this implies that the four blocks on the right hand side of (5.5), labeled by $[1,N]\times[1,N], [1,N]\times[N+1,2N], [N+1,2N]\times[1,N], [N+1,2N]\times[N+1,2N]$ are all diagonal matrices, so that $\mathcal{H}\mathcal{G}(w)$, being the inverse of this matrix, also satisfies that $\mathcal{H}\mathcal{G}_{12,}, \mathcal{H}\mathcal{G}_{21}$ are diagonal matrices, where we denote by $\mathcal{H}\mathcal{G} = \begin{bmatrix} \mathcal{H}\mathcal{G}_{11} & \mathcal{H}\mathcal{G}_{12,} \\ \mathcal{H}\mathcal{G}_{21} & \mathcal{H}\mathcal{G}_{22} \end{bmatrix}$. By the homogeneous variance structure of $\mathcal{T}_n$, we can verify that $\mathcal{H}\mathcal{G}' := \begin{bmatrix} \mathcal{H}\mathcal{G}_{22} & \mathcal{H}\mathcal{G}_{12,} \\ \mathcal{H}\mathcal{G}_{21} & \mathcal{H}\mathcal{G}_{11} \end{bmatrix}$ is also a solution to the MDE (5.5) with positive imaginary part. By [24], Theorem 2.1, the solution to MDE with $\Im w > 0$ and positive imaginary part is unique. Thus we have $\mathcal{H}\mathcal{G} = \mathcal{H}\mathcal{G}'$, which implies, via (5.4), that $m_i = \tilde{m}_i$ for each $1 \leq i \leq n$. □

We can now extract the equations solved by m_i as follows:

$$\frac{1}{m_i} = -w(1 + m_i^{avg}) + |z|^2(1 + m_i^{avg})^{-1}, \quad 1 \leq i \leq n, \tag{5.10}$$

where we denote by

$$m_i^{avg} = \frac{m_{i-1} + m_i + m_{i+1}}{3}, \tag{5.11}$$

and we set the zero boundary condition $m_0 = m_{n+1} = 0$.

When the model is fully translational invariant with no boundary, we can check that each $m_k, k \in [n+2]$ should take the same value m_c, where $m_c = m_c(w, z)$ is unique solution with positive imaginary part to the following equation

$$m_c^{-1} = -w(1 + m_c) + |z|^2(1 + m_c)^{-1}. \tag{5.12}$$

This can be verified via taking the candidate solution into (5.10) and ignoring all boundary non-isotropic effects. This computation that $m_k = \widetilde{m}_k = m_c$ has been used several times in prior circular law papers such as [6] and [21], and we can obtain the proof of Proposition 4.1 directly from this:

Proof of Proposition 4.1. It suffices to apply the proof of Proposition 5.3 to the periodic matrix $\mathcal{T}_{n+2}^{per}$, and use the fact that the solution to the MDE associated to $\mathcal{T}_{n+2}^{per}$ is precisely m_c on the diagonal. Then the convergence of $m_{n+2,z}^{per}(\zeta)$ to $m_z(\zeta)$ follows from the projection of MDE in (5.7) and standard properties of Stieltjes transform of the circular law density, see [27], section 4 and [28]. □

However, for our matrix Y_z, the boundary is not periodic so we have no closed form solutions. Nevertheless, we can still prove the following uniform upper bound:

Lemma 5.6. *For all $w = i\eta, \eta > 0$, the solution to* (5.10) *satisfies*

$$|\Im m_i| \le \max(2|z|, \sqrt{6})\eta^{-1/2}, \quad \forall 1 \le i \le n.$$

Proof. To upper bound $\Im m_i$, it is enough to show that the real part of $1/m_i$ is not too small, since

$$|m_i| \le \frac{1}{\Re(1/m_i)}, \quad \Im m_i \le |m_i|.$$

We denote by $s_i := m_i^{avg}$, then the dissipative term $-i\eta(1+s_i)$ contributes $\eta\Im(1+s_i)$ to $\Re(1/m_i)$. The coupling term $|z|^2(1+s_i)^{-1}$ may possibly reduce the real part, but by an amount at most $|z|^2/\Im(1+s_i)$. Thus taking the absolute value of the equation, we have

$$\Re\frac{1}{m_i} \ge \eta\Im(1+s_i) - \frac{|z|^2}{\Im(1+s_i)}.$$

Now we denote by $u_i = \Im m_i$ and let $u_* = \max_i u_i$ (take $m_0 = m_{n+1} = 0$). Then let i_* be the index achieving the maximum, we have

$$\Im(1 + s_{i_*}) = \frac{u_{i_*-1} + u_{i_*} + u_{i_*+1}}{3} \ge \frac{u_*}{3}. \tag{5.13}$$

Now, if $\eta\Im(1+s_{i_*}) \le 4\frac{|z|^2}{\Im(1+s_{i_*})}$, this directly yields the desired upper bound

$$\Im(1 + s_{i_*}) \le 2|z|\eta^{-1/2}.$$

If otherwise $\eta\Im(1+s_{i_*}) \ge 4\frac{|z|^2}{\Im(1+s_{i_*})}$, then

$$\Re\frac{1}{m_i} \ge \frac{1}{2}\eta\Im(1+s_i).$$

This implies that, using (5.13) for the last inequality in the next equation,

$$u_* \le |m_{i_*}| \le \frac{2}{\eta \Im(1+s_{i_*})} \le \frac{6}{\eta u_*},$$

so that $u_* \le \sqrt{6}\eta^{-1/2}$. Combining both cases completes the proof. ☐

Combining this uniform upper bound of Stieltjes transform, and matrix concentration inequality in Proposition 5.3, we can deduce singular value rigidity for Y_z:

Corollary 5.7. *With the same assumption as above, for any sufficiently small constant $c > 0$, we have that with probability at least $1 - N^{-100}$, the number of singular values of Y_z in $[0, W^{-1/10}N^{c/5}]$ is at most $N^{1+c/5}W^{-1/10}$.*

Proof. With probability at least $1 - N^{-100}$ where Proposition 5.3 holds, we bound the number of eigenvalues by the Stieltjes transform via the Poisson kernel formula and get

$$|\text{Number of singular values of } Y_z \text{ in } [0,\eta]| \le N\eta^2 \operatorname{Tr} \mathcal{G}(i\eta^2) \le N\eta^2(\eta^{-1} + \frac{N^c}{W^{1/2}\eta^6}).$$

We take $\eta = W^{-1/10}N^{c/5}$ and this leads to the desired estimate. ☐

Now we return to the matrix $\mathcal{T}_{n+2} - zI_{mid}$ and complete the proof of Proposition 4.3.

Proof of Proposition 4.3. We work on an event $\|\mathcal{T}_{n+2} - zI_{mid}\| \le K$ for some K that holds with probability at least $1 - N^{-100}$: the existence of such K follows from the operator norm bound in Lemma 4.4 and 2.8. Then we apply Lemma 5.1 to reduce to singular value rigidity of Y_Z^R, and apply Lemma 5.2 to reduce to singular value rigidity of the block diagonal matrix $Y_z^{R,diag}$ with probability at least $1 - N^{-100}$. The rigidity of the upper block follows from Lemma 3.2 and rigidity for the lower large block follows from Corollary 5.7. ☐

We list here the (standard) proof of Lemma 5.4.

Proof of Lemma 5.4. We denote by

$$M^{\mathcal{H}} = \begin{bmatrix} A & B \\ C & D \end{bmatrix}, \quad R := \begin{bmatrix} wI_N + \Phi[D] & zI_N \\ \bar{z}I_N & wI_N + \widetilde{\Phi}[A] \end{bmatrix},$$

so that we can rewrite $M^{\mathcal{H}}(w) = -R^{-1}$.

Applying Schur complement formula to $R = \begin{bmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{bmatrix}$ with R_{22} invertible, we have

$$(R^{-1})_{11} = (R_{11} - R_{12}R_{22}^{-1}R_{21})^{-1}.$$

Applied with the following choices

$$R_{11} = wI_N + \Phi[D], R_{12} = zI_N, R_{21} = \bar{z}I_N, R_{22} = wI_N + \widetilde{\Phi}[A],$$

we get that

$$-A^{-1} = wI_N + \Phi[D] - |z|^2(wI_N + \widetilde{\Phi}[A])^{-1}.$$

A similar equation can be derived for D^{-1}. Denote $A = A[w]$ to signify the dependence on w, then we use $M_1(w^2) = \frac{1}{w}A[w]$ in (5.7) to rescale and then verify the two identities satisfied by $M_1(w), M_2(w)$. ☐

Acknowledgment

The author thanks Charles Bordenave, Giorgio Cipolloni and Tatyana Shcherbina for many insightful discussions on this topic.

Funding

The author receives a fellowship from IAS provided by the S.S. Chern Foundation for Mathematical Research Fund and the Fund for Mathematics.

Part of this work was completed while the author was financially supported by a Simons Foundation Grant (601948, DJ) at Massachusetts Institute of Technology.

References

[1] Z. D. Bai. "Circular law". In: *Ann. Probab.* 25.1 (1997), pp. 494–529. ISSN: 0091-1798,2168-894X.

[2] Zhidong Bai and Jack W Silverstein. *Spectral analysis of large dimensional random matrices*. Vol. 20. Springer, 2010.

[3] Charles Bordenave, Pietro Caputo, and Djalil Chafai. "Spectrum of non-Hermitian heavy tailed random matrices". In: *Communications in mathematical physics* 307.2 (2011), pp. 513–560.

[4] Charles Bordenave and Djalil Chafaï. "Around the circular law". In: *Probab. Surv.* 9 (2012), pp. 1–89. ISSN: 1549-5787.

[5] Philippe Bougerol et al. *Products of random matrices with applications to Schrödinger operators*. Vol. 8. Springer Science & Business Media, 2012.

[6] Paul Bourgade, Horng-Tzer Yau, and Jun Yin. "Local circular law for random matrices". In: *Probab. Theory Related Fields* 159.3-4 (2014), pp. 545–595. ISSN: 0178-8051,1432-2064.

[7] Tatiana Brailovskaya and Ramon van Handel. "Universality and sharp matrix concentration inequalities". In: *Geometric and Functional Analysis* 34.6 (2024), pp. 1734–1838.

[8] Nixia Chen and Charles K Smart. "Random band matrix localization by scalar fluctuations". In: *arXiv preprint arXiv:2206.06439* (2022).

[9] Giorgio Cipolloni et al. "Dynamical Localization for Random Band Matrices Up to $W \ll N^{1/4}$". In: *Communications in Mathematical Physics* 405.3 (2024), p. 82.

[10] Joel E Cohen and Charles M Newman. "The stability of large random matrices and their products". In: *The Annals of Probability* (1984), pp. 283–310.

[11] Nicholas Cook. "Lower bounds for the smallest singular value of structured random matrices". In: *The Annals of probability* 46.6 (2018), pp. 3442–3500.

[12] Reuben Drogin. "Localization of One-Dimensional Random Band Matrices". In: *arXiv preprint arXiv:2508.05802* (2025).

[13] Sofiia Dubova et al. "Delocalization of Non-Mean-Field Random Matrices in Dimensions $d \geq 3$". In: *arXiv preprint arXiv:2507.20274* (2025).

[14] Sofiia Dubova et al. "Delocalization of two-dimensional random band matrices". In: *arXiv preprint arXiv:2503.07606* (2025).

[15] László Erdős and Volodymyr Riabov. "The Zigzag Strategy for Random Band Matrices". In: *arXiv preprint arXiv:2506.06441* (2025).

[16] Ilya Ya Goldsheid and Boris A Khoruzhenko. "The Thouless formula for random non-Hermitian Jacobi matrices". In: *Israel Journal of Mathematics* 148.1 (2005), pp. 331–346.

[17] Michael Goldstein. "Fluctuations and localization length for random band GOE matrix". In: *arXiv preprint arXiv:2210.04346* (2022).

[18] Friedrich Götze and Alexander Tikhomirov. "The circular law for random matrices". In: *Ann. Probab.* 38.4 (2010), pp. 1444–1491. ISSN: 0091-1798,2168-894X.

[19] Yi Han. "Invertibility for non-Hermitian random band matrices with sublinear bandwidth and discrete entries". In: *arXiv preprint arXiv:2507.05417* (2025).

[20] Yi Han. "Outliers and bounded rank perturbation for non-Hermitian random band matrices". In: *arXiv preprint arXiv:2408.00567* (2024).

[21] Yi Han. "The circular law for non-Hermitian random band matrices up to bandwidth $W^{1/2+c}$". In: *arXiv preprint arXiv:2508.18143* (2025).

[22] Yi Han. "The circular law for random band matrices: improved bandwidth for general models". In: *arXiv preprint arXiv:2410.16457* (2024).

[23] Boris Hanin and Grigoris Paouris. "Non-asymptotic results for singular values of Gaussian matrix products". In: *Geom. Funct. Anal.* 31.2 (2021), pp. 268–324. ISSN: 1016-443X,1420-8970.

[24] J William Helton, Reza Rashidi Far, and Roland Speicher. "Operator-valued semicircular elements: solving a quadratic matrix equation with positivity constraints". In: *International Mathematics Research Notices* 2007.9 (2007), rnm086–rnm086.

[25] Daniel E Holz, Henri Orland, and A Zee. "On the remarkable spectrum of a non-Hermitian random matrix model". In: *Journal of Physics A: Mathematical and General* 36.12 (2003), p. 3385.

[26] Marco Isopi and Charles M Newman. "The triangle law for Lyapunov exponents of large random matrices". In: *Communications in mathematical physics* 143.3 (1992), pp. 591–598.

[27] Vishesh Jain et al. "Circular law for random block band matrices with genuinely sublinear bandwidth". In: *Journal of Mathematical Physics* 62.8 (2021).

[28] Indrajit Jana and Alexander Soshnikov. "Distribution of singular values of random band matrices; Marchenko–Pastur law and more". In: *Journal of Statistical Physics* 168.5 (2017), pp. 964–985.

[29] Alexander E. Litvak et al. "Circular law for sparse random regular digraphs". In: *J. Eur. Math. Soc. (JEMS)* 23.2 (2021), pp. 467–501. ISSN: 1435-9855,1435-9863.

[30] Hoi H. Nguyen. "Asymptotic Lyapunov exponents for large random matrices". In: *Ann. Appl. Probab.* 27.6 (2017), pp. 3672–3705. ISSN: 1050-5164,2168-8737.

[31] Valery Iustinovich Oseledec. "A multiplicative ergodic theorem, Lyapunov characteristic numbers for dynamical systems". In: *Transactions of the Moscow Mathematical Society* 19 (1968), pp. 197–231.

[32] Ron Peled et al. "On the Wegner orbital model". In: *International Mathematics Research Notices* 2019.4 (2019), pp. 1030–1058.

[33] Mark Rudelson and Konstantin Tikhomirov. "The sparse circular law under minimal assumptions". In: *Geom. Funct. Anal.* 29.2 (2019), pp. 561–637. ISSN: 1016-443X,1420-8970.

[34] Mark Rudelson and Roman Vershynin. "Small ball probabilities for linear images of high-dimensional distributions". In: *International Mathematics Research Notices* 2015.19 (2015), pp. 9594–9617.

[35] Mark Rudelson and Roman Vershynin. “The Littlewood–Offord problem and invertibility of random matrices”. In: *Advances in Mathematics* 218.2 (2008), pp. 600–633.
[36] Ashwin Sah, Julian Sahasrabudhe, and Mehtaab Sawhney. “The limiting spectral law for sparse iid matrices”. In: *arXiv preprint arXiv:2310.17635* (2023).
[37] Arvind Sankar, Daniel A Spielman, and Shang-Hua Teng. “Smoothed analysis of the condition numbers and growth factors of matrices”. In: *SIAM Journal on Matrix Analysis and Applications* 28.2 (2006), pp. 446–476.
[38] Jeffrey Schenker. “Eigenvector localization for random band matrices with power law band width”. In: *Communications in Mathematical Physics* 290.3 (2009), pp. 1065–1097.
[39] Mariya Shcherbina and Tatyana Shcherbina. “Characteristic polynomials of non-Hermitian random band matrices”. In: *arXiv preprint arXiv:2510.04255* (2025).
[40] Terence Tao and Van Vu. “Random matrices: the circular law”. In: *Communications in Contemporary Mathematics* 10.02 (2008), pp. 261–307.
[41] Terence Tao, Van Vu, and Manjunath Krishnapur. “Random matrices: universality of ESDs and the circular law”. In: *Annals of Probability* 38.5 (2010), pp. 2023–2065.
[42] Konstantin Tikhomirov. “Invertibility via distance for noncentered random matrices with continuous distributions”. In: *Random Structures & Algorithms* 57.2 (2020), pp. 526–562.
[43] Konstantin Tikhomirov. “On pseudospectrum of inhomogeneous non-Hermitian random matrices”. In: *arXiv preprint arXiv:2307.08211* (2023).
[44] Konstantin Tikhomirov. “Quantitative invertibility of non-Hermitian random matrices”. In: *ICM—International Congress of Mathematicians. Vol. 4. Sections 5–8.* EMS Press, Berlin, 2023, pp. 3292–3313.
[45] Eugene P Wigner. “On the distribution of the roots of certain symmetric matrices”. In: *Annals of Mathematics* 67.2 (1958), pp. 325–327.
[46] Horng-Tzer Yau and Jun Yin. “Delocalization of One-Dimensional Random Band Matrices”. In: *arXiv preprint arXiv:2501.01718* (2025).

Institute for Advanced Study, 1 Einstein Drive, Princeton, NJ
Email address: `hanyi@ias.edu`