CIRCULAR LAW FOR RANDOM BLOCK BAND MATRICES WITH GENUINELY SUBLINEAR BANDWIDTH

VISHESH JAIN

Department of Statistics, Stanford University

INDRAJIT JANA

Indian Institute of Technology, Bhubaneswar

KYLE LUH

Department of Mathematics, University of Colorado Boulder

SEAN O'ROURKE

Department of Mathematics, University of Colorado Boulder

Abstract. We prove the circular law for a class of non-Hermitian random block band matrices with genuinely sublinear bandwidth. Namely, we show there exists $\tau \in (0,1)$ so that if the bandwidth of the matrix X is at least $n^{1-\tau}$ and the nonzero entries are iid random variables with mean zero and slightly more than four finite moments, then the limiting empirical eigenvalue distribution of X, when properly normalized, converges in probability to the uniform distribution on the unit disk in the complex plane. The key technical result is a least singular value bound for shifted random band block matrices with genuinely sublinear bandwidth, which improves on a result of [N. Cook, Ann. Probab., 46, 3442 (2018)] in the band matrix setting.

1. Introduction

Random band matrices play an important role in mathematics and physics. Unlike many classical matrix ensembles, band matrices with small bandwidth are not of mean-field type and involve short-range interactions. As such, band matrices interpolate between classical mean field models with delocalized eigenvectors (when the bandwidth is large) and models with localized eigenvectors and poisson eigenvalue statistics (when the bandwidth is small) [22]. In addition, random band matrices

Date: July 19, 2021.

Corresponding author email: Sean.D.ORourke@colorado.edu.

have been studied in the context of nuclear physics, quantum chaos, theoretical ecology, systems of interacting particles, and neuroscience [3, 4, 5, 6, 29, 49, 50, 83]. Many mathematical results have been established for the eigenvalues and eigenvectors of random band matrices, especially Hermitian models; we refer the reader to [3, 8, 12, 13, 15, 16, 24, 27, 28, 29, 36, 38, 39, 40, 41, 52, 53, 54, 55, 56, 58, 63, 64, 68, 74, 75, 76, 78, 86] and references therein.

In this paper, we focus on non-Hermitian random block band matrices. Before we introduce the model, we define some notation and recall some previous results for non-Hermitian random matrices with independent entries. For an $n \times n$ matrix A, we let $\lambda_1(A), \ldots, \lambda_n(A) \in \mathbb{C}$ denote the eigenvalues of A (counted with algebraic multiplicity). μ_A is the empirical spectral measure of A defined as

$$\mu_A := \frac{1}{n} \sum_{i=1}^{n} \delta_{\lambda_i(A)},$$

where δ_z denotes a point mass at z.

The circular law describes the limiting empirical spectral measure for a class of random matrices with independent and identically distributed (iid) entries.

Definition 1.1 (iid matrix)**.** Let ξ be a complex-valued random variable. An $n \times n$ matrix X is called an *iid random matrix with atom variable (or atom distribution)* ξ if the entries of X are iid copies of ξ.

The circular law asserts that if X is an $n \times n$ iid random matrix with atom variable ξ having mean zero and unit variance, then the empirical spectral measure of $X/\sqrt{n}$ converges almost surely to the uniform probability measure on the unit disk centered at the origin in the complex plane. This was proved by Tao and Vu in [80, 81], and is the culmination of a large number of results by many authors [10, 37, 42, 43, 45, 47, 61, 62]. We refer the reader to the survey [20] for more complete bibliographic and historical details. Local versions of the circular law have also been established [7, 25, 26, 85, 87]. The eigenvalues of other models of non-Hermitian random matrices have been studied in recent years; see, for instance, [1, 2, 14, 17, 18, 19, 31, 33, 34, 35, 44, 46, 57, 65, 66, 67, 69, 70, 72, 84] and references therein.

Another model of non-Hermitian random matrices takes the form $X \odot A$, where the entries of the $n \times n$ matrix X are iid random variables with mean zero and unit variance and A is a deterministic matrix. Here, $A \odot B$ denotes the Hadamard product of the matrices A and B, with elements given by $(A \odot B)_{ij} = A_{ij} B_{ij}$. The matrix A provides the variance profile for the model, and this model includes band matrices when A has a band structure. The empirical eigenvalue distribution of such matrices was studied in [34]. For example, the following result from [34] describes sufficient conditions for the limiting empirical spectral distribution to be given by the circular law.

Theorem 1.2 (Theorem 2.4 from [34])**.** *Let ξ be a complex-valued random variable with mean zero, unit variance, and $\mathbb{E}|\xi|^{4+\epsilon} < \infty$ for some $\epsilon > 0$. Let X be an $n \times n$ iid matrix with atom variable ξ, and let $A = (\sigma_{ij}^{(n)})$ be an $n \times n$ matrix with non-negative entries which satisfy*

$$\sup_{n \geq 1} \max_{1 \leq i,j \leq n} \sigma_{ij}^{(n)} \leq \sigma_{\max} \tag{1}$$

for some $\sigma_{\max} \in (0, \infty)$ *and*

$$\frac{1}{n}\sum_{i=1}^{n}(\sigma_{ij}^{(n)})^2 = \frac{1}{n}\sum_{j=1}^{n}(\sigma_{ij}^{(n)})^2 = 1 \tag{2}$$

for all $1 \le i, j \le n$. *Then the empirical spectral measure of* $\frac{1}{\sqrt{n}}X \odot A$ *converges in probability as* $n \to \infty$ *to the uniform probability measure on the unit disk in the complex plane centered at the origin.*

More generally, the results in [34] also apply to cases when conditions (1) and (2) are relaxed and the limiting empirical spectral measure is not given by the circular law. However, the results in [34], unlike the results in this paper, require the number of non-zero entries to be proportional to n^2 for the limit to be non-trivial.

1.1. **The model and result.** In this paper, we focus on a model where the number of non-zero entries is *polynomially* smaller than n^2. We now introduce the model of random block band matrices we will study.

Definition 1.3 (Periodic block band matrix)**.** Let $b_n \ge 1$ be an integer that divides n, and let ξ be a complex-valued random variable. We consider the $n \times n$ *periodic block-band matrix* $\tilde{X}$ *with atom variable (or atom distribution)* ξ *and bandwidth* b_n defined to be the tri-diagonal periodic block band matrix $\tilde{X}$ given by

$$\tilde{X} := \begin{pmatrix} \tilde{D}_1 & \tilde{U}_2 & & & \tilde{T}_m \\ \tilde{T}_1 & \tilde{D}_2 & \tilde{U}_3 & & \\ & \tilde{T}_2 & \tilde{D}_3 & \ddots & \\ & & \ddots & \ddots & \tilde{U}_m \\ \tilde{U}_1 & & & \tilde{T}_{m-1} & \tilde{D}_m \end{pmatrix} \tag{3}$$

where the entries not displayed are taken to be zero. Here, $\tilde{D}_1, \tilde{U}_1, \tilde{T}_1, \ldots, \tilde{D}_m, \tilde{U}_m, \tilde{T}_m$ are $b_n \times b_n$ independent iid random matrices each having atom variable ξ and $m := n/b_n$. For convenience, we use the convention that the indices wrap around; meaning for example that $\tilde{U}_{-1} = \tilde{U}_m$.

Note that each row and column of $\tilde{X}$ has $3b_n$ many nonzero random variables. Using the notation $[m] := \{1, \ldots, m\}$ for the discrete interval, we define

$$\begin{aligned} c_n &:= 3b_n & D_i &:= \frac{1}{\sqrt{c_n}}\tilde{D}_i, \ \ \forall\, i \in [m] \\ U_i &:= \frac{1}{\sqrt{c_n}}\tilde{U}_i, \ \ \forall\, i \in [m] & T_i &:= \frac{1}{\sqrt{c_n}}\tilde{T}_i, \ \ \forall\, i \in [m] \\ X &:= \frac{1}{\sqrt{c_n}}\tilde{X} \end{aligned} \tag{4}$$

One motivation for the periodic block band matrix introduced above comes from theoretical ecology. Population densities and food webs, for example, can be modeled by a system involving a large random matrix [6, 59]. The eigenvalues of this random matrix play an important role in the analysis of the the stability of the system, and the circular law and elliptic law have previously been exploited for this purpose [6]. It has been observed that many of these systems correspond to

sparse random matrices with block structures (known as "modules" or "compartments") [6, 79]. The periodic block band matrix introduced above is one such model with a very specific network structure.

Our main result below establishes the circular law for the periodic block band model defined above when b_n is genuinely sublinear. To the best of our knowledge, this is the first result to establish the circular law as the limiting spectral distribution for matrices with genuinely sublinear bandwidth.

Theorem 1.4 (Circular law for random block band matrices)**.** *There exists $c > 0$ such that the following holds. Let ξ be a complex-valued random variable with mean zero, unit variance, and $\mathbb{E}|\xi|^{4+\epsilon} < \infty$ for some $\epsilon > 0$. Assume $\tilde{X}$ is an $n \times n$ periodic block-band matrix with atom variable ξ and bandwidth b_n, where $cn \geq b_n \geq n^{32/33} \log n$. Then the empirical spectral measure of $X := \tilde{X}/\sqrt{3b_n}$ converges in probability as $n \to \infty$ to the uniform probability measure on the unit disk in the complex plane centered at the origin.*

We prove Theorem 1.4 by showing that there exists constants $c, \tau > 0$ so that the empirical spectral measure of X converges to the circular law under the assumption that the bandwidth b_n satisfies $cn \geq b_n \geq n^{1-\tau} \log n$. In fact, the proof reveals that τ can be taken to be $\tau := 1/33$, as stated in Theorem 1.4, although this particular value can likely be improved by optimizing some of the exponents in the proof.

A few remarks concerning the assumptions of Theorem 1.4 are in order. First, the restriction on the bandwidth $b_n \geq n^{1-\tau} \log(n)$ with $\tau = 1/33$ is of a technical nature and we believe this condition can be significantly relaxed. For instance, we give an exponential lower bound on the least singular value of $X - zI$ for $z \in \mathbb{C}$ in Theorem 2.1 below. If this bound could be improved to say polynomial in n, then we could improve the value of τ to $1/2$. It is possible that other methods could also improve this restriction even further. Second, the assumption that the entries have finite $4+\epsilon$ moments is due to the sublinear bandwidth growth rate. Our calculation requires higher moment assumptions for slower bandwidth growth, as can be seen from the proof of Theorem 3.1.

A numerical simulation of Theorem 1.4 is presented in Figure 1.

1.2. **Notation and overview.** We use asymptotic notation under the assumption that $n \to \infty$. The notations $X = O(Y)$ and $Y = \Omega(X)$ denote the estimate $|X| \leq CY$ for some constant $C > 0$ and all $n \geq C$. We write $X = o(Y)$ if $|X| \leq c_n Y$ for some c_n that goes to zero as n tends to infinity.

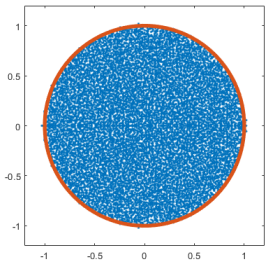
For convenience, we do not always indicate the size of a matrix in our notation. For example, to denote an $n \times n$ matrix A, we simply write A instead of A_n when the size is clear. We use b_n to denote the size of each block matrix and $c_n := 3b_n$ for the number of non-zero entries per row and column. We let $[n] := \{1, 2, 3, \ldots, n\}$ and $e_1, e_2, \ldots, e_n$ be the standard basis elements of $\mathbb{C}^n$. For a matrix A, a_{ij} will be the (i, j)-th entry, a_k will be the kth column, $A^{(k)}$ represents the matrix A with its kth column set to zero and $\mathcal{H}_k$ will be the span of the columns of $A^{(k)}$. Furthermore, A^* is the complex conjugate transpose of the matrix A, and when A is a square matrix, we let

$$A_z := A - zI$$

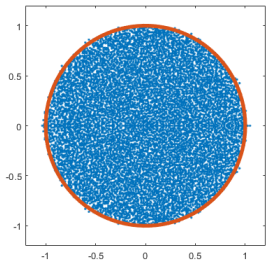
where I denotes the identity matrix and $z \in \mathbb{C}$.

For the spectral information of an $n \times n$ matrix A, we designate

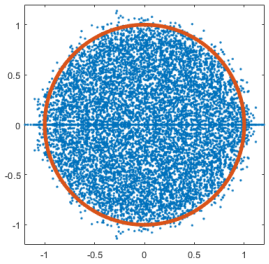
$$\lambda_1(A), \lambda_2(A), \ldots, \lambda_n(A) \in \mathbb{C}$$



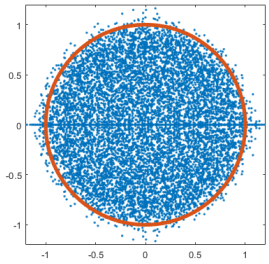
(A) $\tilde{X}$ has Gaussian atom variable with $n = 10,000$ and $b_n = 100$.



(B) $\tilde{X}$ has Rademacher atom variable with $n = 10,000$ and $b_n = 100$.



(C) $\tilde{X}$ has Gaussian atom variable with $n = 10,000$ and $b_n = 10$.



(D) $\tilde{X}$ has Rademacher atom variable with $n = 10,000$ and $b_n = 10$.

FIGURE 1. Numerical simulations for the eigenvalues of $X := \tilde{X}/\sqrt{3b_n}$ when $\tilde{X}$ is an $n \times n$ period block-band matrix with bandwidth b_n for various atom distributions.

to be the eigenvalues of A (counted with algebraic multiplicity) and

$$\mu_A := \frac{1}{n}\sum_{i=1}^{n} \delta_{\lambda_i(A)}$$

to be the empirical measure of the eigenvalues. Here, δ_z represents a point mass at $z \in \mathbb{C}$. Similarly, we denote the singular values of A by

$$s_1(A) \geq s_2(A) \geq \ldots \geq s_n(A) \geq 0$$

and the empirical measure of the squared-singular values as

$$\nu_A := \frac{1}{n}\sum_{i=1}^{n} \delta_{s_i^2(A)}.$$

Additionally, we use $\|A\|$ to mean the standard $\ell_2 \to \ell_2$ operator norm of A.

For a vector $v \in \mathbb{C}^n$,

$$\|v\| := \left(\sum_{k=1}^{n} |v_k|^2\right)^{1/2} \quad \text{and} \quad \|v\|_\infty := \max_k |v_k|.$$

Finally, we use the following standard notation from analysis and linear algebra. The set of unit vectors in $\mathbb{C}^n$ will be denoted by S^{n-1} i.e. $S^{n-1} := \{v \in \mathbb{C}^n : \|v\| = 1\}$ and the disk of radius r by $\mathbb{D}_r := \{z \in \mathbb{C} : |z| < r\}$. For any set $\mathcal{S} \subset \mathbb{C}^n$ and

$u \in \mathbb{C}^n$,

$$\operatorname{dist}(u, \mathcal{S}) := \inf_{v \in \mathcal{S}} \|u - v\|.$$

$|S|$ denotes the cardinality of the finite set S.

The rest of the paper is devoted to the proof of Theorem 1.4. The proof proceeds via Girko's Hermitization procedure (see [20]) which is now a standard technique in the study of non-Hermitian random matrices. Following [54], we study the empirical eigenvalue distribution of $X_z X_z^*$ for $z \in \mathbb{C}$. In particular, we establish a rate of convergence for the Stieltjes transform $X_z X_z^*$ to the Stieltjes transform of the limiting measure in Section 3. The key technical tool in our proof is a lower bound on the least singular value of X_z presented in Section 2. In Section 4, following the method of Bai [10], these two key ingredients are combined and the proof of Theorem 1.4 is given. The appendix contains a number of auxiliary results.

2. Least singular value

In this section, we present our key least singular value bound, Theorem 2.1. The crucial feature of our result is that the lower bound on the least singular value is only singly exponentially small in m. While this is most likely suboptimal, and indeed, we conjecture that our bound can be substantially improved, it is still significantly better than previous results in the literature. Notably, the work of Cook [32] provides lower bounds on the least singular value for more general structured sparse random matrices; however, specialized to our setting, the lower bound there is *doubly* exponentially small in m (see Equation 3.8 in [32]), which only translates to a circular law for bandwidth (at best) $\Omega(n/\log n)$.

We consider the translated periodic block-band model $X_z = X - zI$, where X is as defined in (4) and $z \in \mathbb{C}$ is fixed. Recall that $m = n/b_n$. Throughout this section, we will assume that $b_n \geq m \geq m_0$, where m_0 is a sufficiently large constant. Recall that for an $n \times n$ matrix A, we let $s_1(A) \geq s_2(A) \geq \cdots \geq s_n(A) \geq 0$ denote its singular values.

Theorem 2.1. *Fix $\epsilon, K' > 0$. Suppose $\tilde{X}$ is an $n \times n$ periodic block band matrix (as defined in (3)) with atom variable ξ satisfying $\mathbb{E}[\xi] = 0$, $\mathbb{E}[|\xi|^2] = 1$, and $\mathbb{E}[|\xi|^{4+\epsilon}] \leq C$, for some absolute constant $C > 0$. Then, for any $z \in \mathbb{C}$ such that $|z| \leq K'$,*

$$\mathbb{P}(s_n(X_z) \leq c_n^{-25m}) \leq \frac{C_\xi}{\sqrt{c_n}},$$

where C_ξ is a constant depending only on ϵ, C and K'.

Let us define the event

$$\mathcal{E}_K = \left\{ \forall i \in [m] : \ \ \|U_i\|, \|(D_i)_z\|, \|T_i\| \leq K, \text{ and } \ s_{b_n}(U_i), s_{b_n}(T_i) \geq b_n^{-5} \right\}.$$

We begin by showing (Lemma 2.4) that $\mathbb{P}(\mathcal{E}_K^c) = O(1/c_n)$. This will allow us to restrict ourselves to the event $\mathcal{E}_K$ for the remainder of this section.

In order to bound the probability of the event $\mathcal{E}_K^c$, we will need the following two results on the smallest and largest singular values of (shifts of) complex random matrices with iid entries.

Proposition 2.2 (Theorem 1.1 from [51])**.** *Let A be an $n \times n$ matrix whose entries are iid copies of a complex random variable ξ satisfying $\mathbb{E}[\xi] = 0$ and $\mathbb{E}[|\xi|^2] = 1$.*

Let F be a fixed $n \times n$ complex matrix whose operator norm is at most $n^{0.51}$. Then, for any $\varepsilon \geq 0$,

$$\mathbb{P}(s_n(F+A) \leq \varepsilon n^{-5/2}) \leq C\varepsilon + C\exp(-\gamma n^{1/50})$$

for two constants $C > 0, \gamma \in (0,1)$ depending only on the distribution of the random variable ξ.

The next proposition can be readily deduced from Theorem 5.9 in [9] along with the standard Chernoff bound.

Proposition 2.3. *Fix $\epsilon > 0$. Let A be an $n \times n$ matrix whose entries are iid copies of a complex random variable ξ satisfying $\mathbb{E}[\xi] = 0$, $E[|\xi|^2] = 1$ and $\mathbb{E}[|\xi|^{4+\epsilon}] \leq M$. Then*

$$\mathbb{P}[\|A\| > K\sqrt{n}] \leq Kn^{-2},$$

where $K > 0$ is a sufficiently large constant depending only on ξ (and hence, also the parameter $\epsilon > 0$).

Applying the above two propositions (along with the triangle inequality for $\|(D_i)_z\|$) and using the union bound, we immediately obtain:

Lemma 2.4. *There exists a constant $K > 0$, depending only on $|z|$ and the random variable ξ (and hence also on the parameter $\epsilon > 0$) such that*

$$\mathbb{P}(\mathcal{E}_K^c) \leq Kb_n^{-1}.$$

For the remainder of this section, we will restrict ourselves to the event $\mathcal{E}_K$. For any $v \in \mathbb{C}^n$, we let

$$v = \begin{pmatrix} v_{[1]} \\ v_{[2]} \\ \vdots \\ v_{[m]} \end{pmatrix}$$

be the division of the coordinates into m vectors $v_{[i]} \in \mathbb{C}^{b_n}$. We will use v_i to denote the i-th coordinate of v. For convenience, we use the convention that the indices wrap around meaning, for example, that $v_{[m+1]} = v_{[1]}$.

For $\alpha, \beta \in (0,1)$, let

$$L_{\alpha,\beta} := \{v \in S^{n-1} : |\{i \in [n] : |v_i| \geq \beta b_n^{-10m} n^{-1/2}\}| \geq \alpha n\},$$

i.e. $L_{\alpha,\beta}$ consists of those unit vectors that have sufficiently many large coordinates. For us, α and β are constants depending on K which will be specified later. Then, as $s_n(X_z) = \inf_{v \in S^{n-1}} \|X_z v\|$, we can decompose the least singular value problem into two terms:

$$\begin{aligned} &\mathbb{P}(\mathcal{E}_K \cap \{s_n(X_z) \leq tb_n^{-10m} n^{-1/2}\}) \qquad (5) \\ \leq &\mathbb{P}(\mathcal{E}_K \cap \{\inf_{v \in L_{\alpha,\beta}} \|X_z v\| \leq tb_n^{-10m} n^{-1/2}\}) + \mathbb{P}(\mathcal{E}_K \cap \{\inf_{v \in L_{\alpha,\beta}^c} \|X_z v\| \leq tb_n^{-10m} n^{-1/2}\}). \end{aligned}$$

2.1. **Reduction to the Distance Problem.** We begin with a lemma due to Rudelson and Vershynin, which converts the first term in (5) into a question about the distance of a random vector to a random subspace.

Lemma 2.5 (Lemma 3.5 from [73]). *Let $x_1 - ze_1, \dots, x_n - ze_n$ be the columns of X_z and let $\mathcal{H}_i$ be the span of all the columns except the i-th. Then,*

$$\mathbb{P}(\mathcal{E}_K \cap \{\inf_{v \in L_{\alpha,\beta}} \|X_z v\| \le t b_n^{-10m} n^{-1/2}\}) \le \frac{1}{\alpha n} \sum_{k=1}^{n} \mathbb{P}(\mathcal{E}_K \cap \{\mathrm{dist}(x_k - ze_k, \mathcal{H}_k) \le \beta^{-1} t\}).$$

Proof. Let

$$p_k := \mathbb{P}(\mathcal{E}_K \cap \{\mathrm{dist}(x_k - ze_k, \mathcal{H}_k) \le \beta^{-1} t\}).$$

By the linearity of expectation, we have

$$\mathbb{E}|\{k \in [n] : \mathcal{E}_K \text{ and } \{\mathrm{dist}(x_k - ze_k, \mathcal{H}_k) \le \beta^{-1} t\}\}| = \sum_{k=1}^{n} p_k.$$

Therefore, if we let

$$\Xi = \mathcal{E}_K \cap \{|\{k \in [n] : \mathrm{dist}(x_k - ze_k, \mathcal{H}_k) \le \beta^{-1} t\}| < \alpha n\},$$

it follows from Markov's inequality that

$$\mathbb{P}(\mathcal{E}_K \cap \Xi^c) \le \frac{\sum_{k=1}^{n} p_k}{\alpha n}.$$

By definition, any vector $v \in L_{\alpha,\beta}$ has at least αn coordinates with absolute value larger than $\beta b_n^{-10m} m^{-1/2} b_n^{-1/2}$. Therefore, on the event Ξ, for any $v \in L_{\alpha,\beta}$, there exists some $k \in [n]$ such that $|v_k| \ge \beta b_n^{-10m} m^{-1/2} b_n^{-1/2}$ and $\mathrm{dist}(x_k - ze_k, \mathcal{H}_k) > \beta^{-1} t$. Hence, on the event Ξ, for all $v \in L_{\alpha,\beta}$,

$$\|X_z v\| \ge |v_k| \mathrm{dist}(x_k - ze_k, \mathcal{H}_k) \ge t b_n^{-10m} m^{-1/2} b_n^{-1/2}.$$

Thus, we see that the probability of the event in the statement of the lemma is at most the probability of $\mathcal{E}_K \cap \Xi^c$, which gives the desired conclusion. □

The distance of $x_k - ze_k$ to $\mathcal{H}_k$ can be bounded from below by $|\langle x_k - ze_k, \hat{n}\rangle|$ where $\hat{n}$ is a unit vector orthogonal to $\mathcal{H}_k$. Our next goal is to obtain some structural information about any vector normal to $\mathcal{H}_k$. For convenience of notation, we will henceforth assume that $k = 1$; the same arguments are readily seen to hold for other values of k as well. Moreover, since the distribution of X_z is invariant under transposition, we may as well assume that $x_1 - ze_1$ is the first *row* of X_z and that $\mathcal{H}_1$ is the subspace spanned by all the rows except for the first.

2.2. **Structure of Normal Vectors and Approximately Null Vectors.** Recall that $\mathcal{H}_1$ is the subspace generated by all the rows of X_z except for the first row. The next proposition establishes that if v is normal to $\mathcal{H}_1$, then there are sufficiently many $v_{[i]}$ with large enough norm. Our approach to lower bounding the coordinates of v is similar to the methods used in [23]; our proof is also similar in spirit to the proof of Proposition 2.9 in [30].

Proposition 2.6. *On the event $\mathcal{E}_K$, for any vector $v \in S^{n-1}$ that is orthogonal to $\mathcal{H}_1$ and for all sufficiently large n (depending on K), either*

$$\|v_{[i]}\| \ge b_n^{-10m} m^{-1/2} \text{ or } \|v_{[i+1]}\| \ge b_n^{-10m} m^{-1/2}.$$

for all $i \in [m-1]$.

Proof. By definition, v must satisfy the following collection of equations:

$$\begin{aligned} T_1 v_{[1]} + (D_2)_z v_{[2]} + U_3 v_{[3]} &= 0 \\ &\vdots \\ T_{i-1} v_{[i-1]} + (D_i)_z v_{[i]} + U_{i+1} v_{[i+1]} &= 0 \\ &\vdots \\ T_{m-2} v_{[m-2]} + (D_{m-1})_z v_{[m-1]} + U_m v_{[m]} &= 0 \\ T_{m-1} v_{[m-1]} + (D_m)_z v_{[m]} + U_1 v_{[1]} &= 0 \end{aligned} \tag{6}$$

Moreover, since $v \in S^{n-1}$, there exists a smallest index $j_0 \in [m]$ such that $\|v_{[j]}\| \geq m^{-1/2}$. If $j_0 \geq 3$, then the following equation (which is a part of (6))

$$T_{j_0-2} v_{[j_0-2]} + (D_{j_0-1})_z v_{[j_0-1]} + U_{j_0} v_{[j_0]} = 0$$

implies that

$$\|T_{j_0-2} v_{[j_0-2]} + (D_{j_0-1})_z v_{[j_0-1]}\| = \|U_{j_0} v_{[j_0]}\|.$$

On the event $\mathcal{E}_K$, we have from the triangle inequality that

$$\|T_{j_0-2} v_{[j_0-2]} + (D_{j_0-1})_z v_{[j_0-1]}\| \leq K(\|v_{[j_0-2]}\| + \|v_{[j_0-1]}\|)$$

and

$$\|U_{j_0} v_{[j_0]}\| \geq b_n^{-5} \|v_{[j_0]}\| \geq b_n^{-5} m^{-1/2}.$$

Therefore, for n sufficiently large compared to K, either

$$\|v_{[j_0-2]}\| \geq b_n^{-10} m^{-1/2} \text{ or } \|v_{[j_0-1]}\| \geq b_n^{-10} m^{-1/2}. \tag{7}$$

Now, let j_{-1} be the smaller of the two indices $j_0 - 1$ and $j_0 - 2$ that satisfies (7). Recall that, for convenience, we are considering indices modulo m. If $j_{-1} \geq 3$ then iterating the argument with j_{-1} and the equation

$$T_{j_{-1}-2} v_{[j_{-1}-2]} + (D_{j_{-1}-1})_z v_{[j_{-1}-1]} + U_{j_{-1}} v_{[j_{-1}]} = 0,$$

we can find $j_{-2} \in \{j_{-1} - 1, j_{-1} - 2\}$ such that

$$\|v_{[j_{-2}]}\| \geq b_n^{-20} m^{-1/2}.$$

Continuing in this manner, we will generate a sequence of indices $j_0, j_{-1}, \ldots, j_{-k}$, $k \leq m$, such that $j_{-k} \in \{1, 2\}$ and such that for all $i \in [k]$

$$|j_{-i} - j_{-i-1}| \leq 2 \text{ and } \|v_{[j_{-i}]}\| \geq b_n^{-10i} m^{-1/2}.$$

We may apply a similar argument to handle indices larger than j_0. Indeed, if $j_0 \leq m - 3$, then we have from (6) that,

$$T_{j_0} v_{[j_0]} + (D_{j_0+1})_z v_{[j_0+1]} + U_{j_0+2} v_{[j_0+2]} = 0.$$

Once again, on the event $\mathcal{E}_K$,

$$\|(D_{j_0+1})_z v_{[j_0+1]} + U_{j_0+2} v_{[j_0+2]}\| \leq K(\|v_{[j_0+1]}\| + \|v_{[j_0+2]}\|)$$

and

$$\|T_{j_0} v_{[j_0]}\| \geq b_n^{-5} m^{-1/2}.$$

As before, this implies that either

$$\|v_{[j_0+1]}\| \geq b_n^{-10} m^{-1/2} \text{ or } \|v_{[j_0+2]}\| \geq b_n^{-10} m^{-1/2}.$$

By iterating this process as above, we obtain a sequence of indices such that $j_0, j_1, \dots, j_{k'}$, $k' \le m$, such that $j_{k'} \in \{m-1, m\}$ and such that for all $i \in [k']$

$$|j_i - j_{i-1}| \le 2 \text{ and } \|v_{[j_i]}\| \ge b_n^{-10i} m^{-1/2}.$$

This completes the proof. □

Note that in the above proof, it is not important that v is precisely normal to $\mathcal{H}_1$. Indeed, exactly the same proof allows us to obtain a similar conclusion for approximately null vectors as well.

Proposition 2.7. *Restricted to $\mathcal{E}_K$, for any vector $v \in S^{n-1}$ such that $\|X_z v\| \le b_n^{-10m} m^{-1/2}$ and for all sufficiently large n (depending on K), either*

$$\|v_{[i]}\| \ge b_n^{-10m} m^{-1/2} \text{ or } \|v_{[i+1]}\| \ge b_n^{-10m} m^{-1/2}$$

for all $i \in [m-1]$,

Our next goal is to show that for α, β sufficiently small depending on K (indeed, the proof shows that we can take $\alpha < \gamma'/(K^2 \log K)$ and $\beta < \gamma'/K$, where $\gamma' > 0$ is a constant depending only on the distribution of the random variable ξ), we have

$$\mathbb{P}(\mathcal{E}_K \cap \{\inf_{v \in L^c_{\alpha,\beta}} \|X_z v\| \le \gamma b_n^{-10m} m^{-1/2}\}) \le m \exp(-\gamma b_n), \tag{8}$$

where $\gamma \in (0,1)$ is a constant depending only on the distribution of the random variable ξ.

For this, we begin with a standard decomposition of the unit sphere, due to Rudelson and Vershynin [73].

Definition 2.8. For $k \in \mathbb{N}$ and $a, \kappa \in (0,1)$, let $\mathrm{Sparse}_k(a)$ denote the sparse vectors $\{v \in S^{k-1} : |\mathrm{supp}(v)| \le ak\}$. We define *compressible* vectors by

$$\mathrm{Comp}_k(a,\kappa) := \{v \in S^{k-1} : \exists u \in \mathrm{Sparse}_k(a) \text{ such that } \|v - u\| \le \kappa\}.$$

and *incompressible* vectors by

$$\mathrm{Incomp}_k(a,\kappa) := S^{k-1} \setminus \mathrm{Comp}_k(a,\kappa).$$

Lemma 2.9. *Let M_i denote the $b_n \times c_n$ block matrix given by $(T_{i-1} \quad (D_i)_z \quad U_{i+1})$. There exists a constant $\gamma \in (0,1)$, depending only on the distribution of the random variable ξ, such that*

$$\mathbb{P}\left(\mathcal{E}_K \cap \{\inf_{w \in \mathrm{Comp}_{c_n}(a,\kappa)} \|M_i w\| \le \gamma\}\right) \le \exp(-\gamma b_n),$$

where $a = \gamma/\log K$ and $\kappa = \gamma/K$.

Proof. This is (by now) a standard argument; we include the short proof for the reader's convenience. We begin with the set $\mathrm{Sparse}_{c_n}(a)$. For any vector $v \in S^{c_n - 1}$, there exist positive constants γ, γ', depending only on the distribution of the entries of M_i such that

$$\mathbb{P}(\|M_i v\| \le \gamma) \le e^{-\gamma' b_n}$$

(cf. Lemma 2.4 in [51]). Recall that an ε-net of a set U is a subset $\mathcal{N} \subset U$ such that for any $w \in U$, there exists a $w' \in \mathcal{N}$ satisfying $\|w - w'\| \le \varepsilon$. By a simple volumetric argument, one can construct an ε-net $\mathcal{N}$ of $\mathrm{Sparse}_{c_n}(a)$ with

$$|\mathcal{N}| \le \binom{c_n}{ac_n} \left(\frac{3}{\varepsilon}\right)^{ac_n} \le \exp(ac_n \log(e/a) + ac_n \log(3/\varepsilon)).$$

We set $\varepsilon = \frac{\gamma}{20K}$. Then, by a union bound,

$$\begin{aligned}\mathbb{P}(\inf_{v\in\mathcal{N}} \|M_i v\| \le \gamma) &\le \sum_{v\in\mathcal{N}} \mathbb{P}(\|M_i v\| \le \gamma) \\ &\le \exp(ac_n \log(e/a) + ac_n \log(3/\varepsilon) - \gamma' b_n) \\ &\le \exp(-\tilde{\gamma} b_n),\end{aligned}$$

where the last inequality holds for $a < \gamma''/\log K$ (for an absolute constant $\gamma'' > 0$). Let $v \in \mathrm{Sparse}_{c_n}(a)$. Then, by definition, there exists some $v' \in \mathcal{N}$ such that $\|v - v'\| \le \varepsilon$. Therefore, on the event $\inf_{v\in\mathcal{N}} \|M_i v\| > \gamma$, we have for any $v \in \mathrm{Sparse}_{c_n}(a)$ that

$$\|M_i v\| \ge \|M_i v\| - \|v - v'\|\|M_i\| \ge \gamma - \frac{\gamma}{20K} 10K = \frac{\gamma}{2}.$$

We can then conclude that

$$\mathbb{P}\left(\inf_{v\in\mathrm{Sparse}_{c_n}(a)} \|M_i v\| \le \frac{\gamma}{2}\right) \le \exp(-\tilde{\gamma} b_n).$$

To extend this to compressible vectors, we simply choose $\kappa = \frac{\gamma}{40K}$. For any $y \in \mathrm{Comp}_{c_n}(a,\kappa)$, there exists $v \in \mathrm{Sparse}_{c_n}(a)$ such that $\|y - v\| \le \kappa$. Thus, if $\|M_i v\| \ge \gamma/2$ then

$$\|M_i y\| \ge \|M_i v\| - \|M_i\|\|v - y\| \ge \frac{\gamma}{2} - 10K\frac{\gamma}{40K} \ge \frac{\gamma}{4}. \qquad \square$$

We will also need the following lemma from [73].

Lemma 2.10 (Lemma 3.4 from [73]). *If $v \in Incomp_k(a,\kappa)$, then there exist constants γ_1 and γ_2 depending only on a and κ such that there are at least $\gamma_1 k$ coordinates with $\gamma_3 k^{-1/2} \ge |v_i| \ge \gamma_2 k^{-1/2}$. In fact, we can take $\gamma_1 = \kappa^2 a/2$, $\gamma_2 = \kappa/\sqrt{2}$, and $\gamma_3 = \kappa^{-1/2}$.*

Now, we are ready to prove (8). Consider a vector $v \in S^{n-1}$ such that $\|X_z v\| \le t b_n^{-10m} m^{-1/2}$, where $0 \le t \le 1$. Then, on the event $\mathcal{E}_K$, it follows from Proposition 2.7 that for any $i \in [m]$,

$$\|(v_{[i-1]}, v_{[i]}, v_{[i+1]})\| \ge b_n^{-10m} m^{-1/2}.$$

Moreover, since for every $i \in [m]$,

$$\left\| (U_{i-1}, (D_i)_z, T_{i+1}) \frac{(v_{[i-1]}, v_{[i]}, v_{[i+1]})^T}{\|(v_{[i-1]}, v_{[i]}, v_{[i+1]})^T\|} \right\| \le \frac{t b_n^{-10m} m^{-1/2}}{\|(v_{[i-1]}, v_{[i]}, v_{[i+1]})^T\|},$$

it follows that

$$\left\| (U_{i-1}, (D_i)_z, T_{i+1}) \frac{(v_{[i-1]}, v_{[i]}, v_{[i+1]})^T}{\|(v_{[i-1]}, v_{[i]}, v_{[i+1]})^T\|} \right\| \le t.$$

Let $\mathcal{E}$ denote the event $\mathcal{E}_K \cap \left(\cap_{i\in[m]} \{\inf_{w\in\mathrm{Comp}_{c_n}(a,\kappa)} \|M_i w\| > \gamma\}\right)$, where a, κ, γ are as in Lemma 2.9. On the event $\mathcal{E}$, if $t \le \gamma$, then

$$\frac{(v_{[i-1]}, v_{[i]}, v_{[i+1]})^T}{\|(v_{[i-1]}, v_{[i]}, v_{[i+1]})^T\|} \in \mathrm{Incomp}_{c_n}(a,\kappa).$$

Therefore, we can conclude from Lemma 2.10 that on the event $\mathcal{E}$, any vector $v \in S^{n-1}$ such that $\|X_z v\| \le \gamma b_n^{-10m} m^{-1/2}$ will have at least αn coordinates larger

than $\beta b_n^{-10m} m^{-1/2} b_n^{-1/2}$, where $\alpha = \gamma'/(K^2 \log K)$, $\beta = \gamma'/K$, and $\gamma' > 0$ is a constant depending only on γ.

Hence, with this choice of α, β, γ', the probability of the event in (8) is bounded by

$$\mathbb{P}(\mathcal{E}_K \cap \mathcal{E}^c) \le \sum_{i=1}^{m} \mathbb{P}\left(\mathcal{E}_K \cap \{\inf_{w \in \mathrm{Comp}_{c_n}(a,\kappa)} \|M_i w\| \le \gamma\}\right) \le m \exp(-\gamma b_n),$$

where the last inequality follows by Lemma 2.9. This proves (8).

The next lemma is a direct consequence of Lemma 2.10 and Lemmas 2.5 and 2.7 from [32].

Lemma 2.11. *Let $\xi_1, \dots, \xi_k$ be independent copies of a complex random variable ξ satisfying $\mathbb{E}[|\xi|^2] = 1$. Then, for any $v \in Incomp_k(a,\kappa)$ and for all $\varepsilon \ge 0$,*

$$\sup_{r \in \mathbb{R}} \mathbb{P}\left(\left|\sum_{i=1}^{k} v_i \xi_i - r\right| \le \varepsilon\right) \le C\kappa^2 a \left(\varepsilon + \frac{1}{\sqrt{\kappa k}}\right),$$

where C is a constant depending only on ξ.

2.3. **Proof of Theorem 2.1.**

Proof of Theorem 2.1. By (5) and (8), it suffices to bound

$$\mathbb{P}(\mathcal{E}_K \cap \{\inf_{v \in L_{\alpha,\beta}} \|X_z v\| \le t b_n^{-10m} m^{-1/2}\}),$$

for $t = b_n^{-11m}$. By Lemma 2.5,

$$\mathbb{P}(\mathcal{E}_K \cap \{\inf_{v \in L_{\alpha,\beta}} \|X_z v\| \le t b^{-10m} m^{-1/2}\}) \le \frac{1}{\alpha} \max_{k \in [n]} \mathbb{P}(\mathcal{E}_K \cap \{\mathrm{dist}(x_k - z e_k, \mathcal{H}_k) \le \beta^{-1} t\}).$$

We will obtain a uniform (in k) bound on $\mathbb{P}(\mathcal{E}_K \cap \{\mathrm{dist}(x_k - z e_k, \mathcal{H}_k) \le \beta^{-1} t\})$. For convenience of notation, we show this bound for $k = 1$. Also, recall from before that we may assume that $x_1 - z e_1$ is the first row of the matrix, and that $\mathcal{H}_1$ is the span of all the rows except for the first row.

Let $\mathcal{E}$ denote the event that

$$\inf_{w \in \mathrm{Comp}_{c_n}(a,\kappa)} \|M_1 w\| \ge \gamma.$$

Then, by Lemma 2.9, $\mathbb{P}(\mathcal{E}^c \cap \mathcal{E}_K) \le \exp(-\gamma b_n)$. Let $\hat{n}$ denote a unit normal vector to $\mathcal{H}_1$, let $v := (\hat{n}_{[1]}, \hat{n}_{[2]}, \hat{n}_{[m]})$, and let $\widehat{v} := v/\|v\|$. If $\widehat{v} \in \mathrm{Comp}_{c_n}(a,\kappa)$, then on the event $\mathcal{E} \cap \mathcal{E}_K$, we have

$$\begin{aligned} |\langle x_1 - z e_1, \hat{n} \rangle| &= |\langle x_1 - z e_1, v \rangle| \\ &= \|M_1 v\| \\ &= \|M_1 \widehat{v}\| \|v\| \\ &\ge \gamma \|v\| \\ &\ge \gamma b_n^{-10m} m^{-1/2}. \end{aligned}$$

On the other hand, if $\widehat{v} \in \mathrm{Incomp}_{c_n}(a,\kappa)$, then it follows from Lemma 2.11 that

$$\begin{aligned} \mathbb{P}(|\langle x_1 - z e_1, \hat{n} \rangle| \le \delta) &= \mathbb{P}(|\langle x_1 - z e_1, v \rangle| \le \delta) \\ &= \mathbb{P}(|\langle x_1 - z e_1, \widehat{v} \rangle| \le \delta/\|v\|) \end{aligned}$$

$$\leq C\kappa^2 a\left(\frac{\delta}{\|v\|}+\frac{1}{\sqrt{\kappa b_n}}\right)$$
$$\leq C\kappa^2 a\left(\delta b^{10m}\sqrt{m}+\frac{1}{\sqrt{\kappa b_n}}\right).$$

Taking $\delta=\beta^{-1}b_n^{-11m}$, and combining with the compressible case, we may conclude that

$$\mathbb{P}(\mathcal{E}_K\cap\{\operatorname{dist}(x_1-ze_1,\mathcal{H}_1)\leq\beta^{-1}b_n^{-11m}\})\leq C_K\frac{1}{\sqrt{b_n}}.$$

The same argument can be used to conclude that

$$\max_{k\in[n]}\mathbb{P}(\mathcal{E}_K\cap\{\operatorname{dist}(x_k-ze_k,\mathcal{H}_k)\leq\beta^{-1}b_n^{-11m}\})\leq C_K\frac{1}{\sqrt{b_n}},$$

which completes the proof. □

3. Convergence of ν_{X_z}

In this section, we establish a rate of convergence for the Stieltjes transform of the empirical eigenvalue distribution of $X_zX_z^*$.

Theorem 3.1. *Let $\tilde{X}$ be an $n\times n$ periodic block band matrix as defined in Definition 1.3 with atom variable ξ. Take $A>1$, and let $z\in\mathbb{C}$ be a fixed complex number. Assume $m_{n,z}(\zeta)=\frac{1}{n}\sum_{i=1}^n[\lambda_i(X_zX_z^*)-\zeta]^{-1}$ is the Stieltjes transform for the empirical spectral measure of $X_zX_z^*$. Suppose that ξ is centered with variance one and $\omega_{4p}:=\mathbb{E}[|\xi|^{4p}]<\infty$ for some integer $p\geq 1$. Then there exists a non random probability measure ν_z on $[0,\infty)$ such that for any $\zeta\in\{\zeta\in\mathbb{C}:-A<\Re(\zeta)<A,0<\Im(\zeta)<1\}$*

$$\mathbb{E}[|m_{n,z}(\zeta)-m_z(\zeta)|^{2p}]\leq\frac{C(p)A^p\omega_{4p}}{|\Im(\zeta)|^{8p}}\left[\left(\frac{n}{c_n^2}\right)^p+\frac{1}{c_n^{p/2}}\right],$$

where $m_z(\zeta)=\int_{\mathbb{R}}\frac{d\nu_z(x)}{x-\zeta}$ and $C(p)>0$ is a constant that depends only on p. Moreover, $m_z(\zeta)$ is the unique solution to the equation

$$m_z(\zeta)=\left[\frac{|z|^2}{1+m_z(\zeta)}-(1+m_z(\zeta))\zeta\right]^{-1},\tag{9}$$

satisfying $\Im(\zeta m_z(\zeta^2))>0$ and $\Im(m_z(\zeta))>0$ when $\Im(\zeta)>0$.

Remark 3.2. We state and prove the above theorem under more general conditions than those of Theorem 1.4. In particular, we allow random variables with no moments above four. Although, the quantitative estimate improves with the number of existing moments. Furthermore, we do not make use of the lower bound on c_n in Theorem 1.4.

We follow the proof strategy from [54]. This previous work demonstrated the convergence of the Stieltjes transform for band matrices rather than block band matrices so we necessarily make some adaptations. More significantly, we deduce an explicit rate of convergence, which does not appear in [54].

Our main object of study will be

$$P_{z,\zeta}:=(X_zX_z^*)_\zeta=(X-zI)(X-zI)^*-\zeta I.$$

Define $X_z^{(k)}$ to be the matrix X_z with the k-th column set to zero. We define

$$P_{z,\zeta}^{(k)}:=(X_z^{(k)}X_z^{(k)*})-\zeta I$$

$$
\begin{aligned}
&= [(X - zI) - (x_k - ze_k)e_k^T][(X - zI) - (x_k - ze_k)e_k^T]^* - \zeta I \\
&= (X - zI)(X - zI)^* - \zeta I - (x_k - ze_k)(x_k - ze_k)^* \\
&= P_{z,\zeta} - (x_k - ze_k)(x_k - ze_k)^*
\end{aligned}
$$

We also denote

$$m_{n,z}^{(k)}(\zeta) := \frac{1}{n} \operatorname{tr}(P_{z,\zeta}^{(k)})^{-1}.$$

Additionally, we use the shorthand

$$\alpha_k := 1 + (x_k - ze_k)^* [P_{z,\zeta}^{(k)}]^{-1} (x_k - ze_k) \tag{10}$$

as this term appears repeatedly in our initial calculations.

For $s_z(\zeta) = m_{n,z}(\zeta)$ or $m_z(\zeta)$, let us define

$$f(s_z) := \left[\frac{|z|^2}{1 + s_z(\zeta)} - (1 + s_z(\zeta))\zeta \right]^{-1}.$$

The motivation for this definition is that $m_z(\zeta)$ is known to be a fixed point of this function when the spectrum obeys the circular law; see Section 11.4 in [9]. The proof of Theorem 3.1 can be divided into several key computations. Since we expect $m_{n,z}(\zeta)$ to also converge to the fixed point of f, we first relate $m_{n,z}(\zeta) - m_z(\zeta)$ to $f(m_{n,z}(\zeta)) - m_{n,z}(\zeta)$.

Lemma 3.3. *Under the assumptions of Theorem 3.1,*

$$m_{n,z}(\zeta) - m_z(\zeta) = [1 - r_{n,z}(\zeta)]^{-1} [m_{n,z}(\zeta) - f(m_{n,z}(\zeta))] \tag{11}$$

where

$$r_{n,z}(\zeta) = f(m_{n,z}(\zeta)) f(m_z(\zeta)) \left[\frac{|z|^2}{(1 + m_{n,z}(\zeta))(1 + m_z(\zeta))} + \zeta \right]. \tag{12}$$

Proof. We have that

$$m_{n,z}(\zeta) - m_z(\zeta) = m_{n,z}(\zeta) - f(m_{n,z}(\zeta)) + f(m_{n,z}(\zeta)) - f(m_z(\zeta)), \tag{13}$$

where we have used the fact that $f(m_z(\zeta)) = m_z(\zeta)$, which is known to characterize the circular law; see Section 11.4 and (11.4.1) in [9]. On the other hand,

$$
\begin{aligned}
&f(m_{n,z}(\zeta)) - f(m_z(\zeta)) \\
&\quad = f(m_{n,z}(\zeta)) f(m_z(\zeta)) \left(\frac{1}{f(m_z(\zeta))} - \frac{1}{f(m_{n,z}(\zeta))} \right) \\
&\quad = f(m_{n,z}(\zeta)) f(m_z(\zeta)) \left[\frac{|z|^2 (m_z(\zeta) - m_{n,z}(\zeta))}{(1 + m_{n,z}(\zeta))(1 + f(m_z(\zeta)))} + \zeta (m_{n,z}(\zeta) - m_z(\zeta)) \right] \\
&\quad = [m_{n,z}(\zeta) - m_z(\zeta)] f(m_{n,z}(\zeta)) f(m_z(\zeta)) \left[\frac{|z|^2}{(1 + m_{n,z}(\zeta))(1 + m_z(\zeta))} + \zeta \right] \\
&\quad =: r_{n,z}(\zeta)[m_{n,z}(\zeta) - m_z(\zeta)].
\end{aligned}
$$

Therefore, by (13),

$$m_{n,z}(\zeta) - m_z(\zeta) = [1 - r_{n,z}(\zeta)]^{-1} [m_{n,z}(\zeta) - f(m_{n,z}(\zeta))]$$

with $r_{n,z}(\zeta)$ given in (12). □

The strategy of our proof is to control the moments of $m_{n,z}(\zeta) - f(m_{n,z}(\zeta))$ and then provide a deterministic bound for $[1 - r_{n,z}(\zeta)]^{-1}$.

We begin with the moments of $f(m_{n,z}(\zeta)) - m_{n,z}(\zeta)$.

Lemma 3.4. *Under the assumptions of Theorem 3.1,*

$$\mathbb{E}[|f(m_{n,z}(\zeta)) - m_{n,z}(\zeta)|^{2p}] \leq \frac{C(p)\omega_{4p}}{|\Im(\zeta)|^{6p}} \left[\left(\frac{n}{c_n^2} \right)^p + \frac{1}{c_n^{p/2}} \right], \tag{14}$$

Proof. We begin by finding a convenient expression to allow us to compute the moments. By the resolvent identity, [1]

$$\begin{aligned} &f(m_{n,z}(\zeta))I - P_{z,\zeta}^{-1} \\ &\quad = f(m_{n,z}(\zeta)) \left[P_{z,\zeta} - f(m_{n,z}(\zeta))^{-1} I \right] P_{z,\zeta}^{-1} \\ &\quad = f(m_{n,z}(\zeta)) \left[(X - zI)(X - zI)^* - \frac{|z|^2}{1 + m_{n,z}(\zeta)} I + \zeta m_{n,z}(\zeta) I \right] P_{z,\zeta}^{-1}. \end{aligned} \tag{15}$$

To simplify this expression, we make the following observation. Since $P_{z,\zeta} = X_z X_z^* - \zeta I$, by Lemma A.1,

$$\begin{aligned} I + \zeta P_{z,\zeta}^{-1} &= X_z X_z^* P_{z,\zeta}^{-1} \\ &= \sum_{k=1}^n (x_k - ze_k)(x_k - ze_k)^* P_{z,\zeta}^{-1} \\ &= \sum_{k=1}^n (x_k - ze_k)(x_k - ze_k)^* [P_{z,\zeta}^{(k)}]^{-1} \alpha_k^{-1}, \end{aligned} \tag{16}$$

where α_k is defined in (10). Taking the normalized trace of (16) yields

$$\begin{aligned} 1 + \zeta m_{n,z}(\zeta) &= \frac{1}{n} \sum_{k=1}^n \frac{1}{\alpha_k} \operatorname{tr}((x_k - ze_k)(x_k - ze_k)^* [P_{z,\zeta}^{(k)}]^{-1}) \\ &= \frac{1}{n} \sum_{k=1}^n \frac{1}{\alpha_k} (x_k - ze_k)^* [P_{z,\zeta}^{(k)}]^{-1} (x_k - ze_k) \\ &= \frac{1}{n} \sum_{k=1}^n \frac{\alpha_k - 1}{\alpha_k} \\ &= 1 - \frac{1}{n} \sum_{k=1}^n \frac{1}{\alpha_k}. \end{aligned}$$

From this, we can conclude that

$$\zeta m_{n,z}(\zeta) = -\frac{1}{n} \sum_{k=1}^n \frac{1}{\alpha_k}. \tag{17}$$

Plugging (17) into (15) gives

$$\begin{aligned} &f(m_{n,z}(\zeta))I - P_{z,\zeta}^{-1} \\ &\quad = f(m_{n,z}(\zeta)) \left[(X - zI)(X - zI)^* - \frac{|z|^2}{1 + m_{n,z}(\zeta)} I - \frac{1}{n} \sum_{k=1}^n \frac{1}{\alpha_k} I \right] P_{z,\zeta}^{-1}. \end{aligned}$$

[1] For two invertible matrices A and B of the same dimension, the resolvent identity is the observation that

$$A^{-1} - B^{-1} = A^{-1}(B - A)B^{-1}.$$

Taking the normalized trace of this equation we find that

$$f(m_{n,z}(\zeta)) - m_{n,z}(\zeta) = \frac{1}{n} f(m_{n,z}(\zeta)) \sum_{k=1}^{n} \Big[(x_k - ze_k)^* P_{z,\zeta}^{-1} (x_k - ze_k) - \frac{|z|^2}{1 + m_{n,z}(\zeta)} e_k^T P_{z,\zeta}^{-1} e_k - \frac{1}{\alpha_k} m_{n,z}(\zeta) \Big]. \tag{18}$$

We will take the $2p$-th moment of this expression.

Let us introduce the following notation to organize the terms on the right hand side of (18). Let

$$\beta_k := x_k^* [P_{z,\zeta}^{(k)}]^{-1} e_k, \qquad \gamma_k := e_k^T [P_{z,\zeta}^{(k)}]^{-1} x_k,$$
$$\delta_k := e_k^T [P_{z,\zeta}^{(k)}]^{-1} e_k, \qquad \tau_k := x_k^* [P_{z,\zeta}^{(k)}]^{-1} x_k.$$

Recall the definition of α_k given in (10). Since

$$\begin{aligned} \alpha_k &= 1 + (x_k - ze_k)^* [P_{z,\zeta}^{(k)}]^{-1} (x_k - ze_k) \\ &= 1 + \tau_k - z\beta_k - \bar{z}\gamma_k + |z|^2 \delta_k, \end{aligned}$$

again by Lemma A.1, we can write

$$\begin{aligned} (x_k - ze_k)^* P_{z,\zeta}^{-1} (x_k - ze_k) &= \alpha_k^{-1} (x_k - ze_k)^* [P_{z,\zeta}^{(k)}]^{-1} (x_k - ze_k) \\ &= \alpha_k^{-1} [\tau_k - z\beta_k - \bar{z}\gamma_k + |z|^2 \delta_k]. \end{aligned}$$

Expanding similarly,

$$\begin{aligned} e_k^T P_{z,\zeta}^{-1} e_k &= e_k^T [P_{z,\zeta}^{(k)}]^{-1} e_k - \alpha_k^{-1} e_k^T [P_{z,\zeta}^{(k)}]^{-1} (x_k - ze_k)(x_k - ze_k)^* [P_{z,\zeta}^{(k)}]^{-1} e_k \\ &= \delta_k - \alpha_k^{-1} (\gamma_k - z\delta_k)(\beta_k - \bar{z}\delta_k) \\ &= \alpha_k^{-1} [(1 + \tau_k - z\beta_k - \bar{z}\gamma_k + |z|^2 \delta_k)\delta_k - (\gamma_k - z\delta_k)(\beta_k - \bar{z}\delta_k)] \\ &= \alpha_k^{-1} [(1 + \tau_k)\delta_k - \gamma_k \beta_k]. \end{aligned}$$

Therefore, (18) can be more succinctly written as

$$\begin{aligned} f(m_{n,z}(\zeta)) - m_{n,z}(\zeta) &= \frac{1}{n} f(m_{n,z}(\zeta)) \sum_{k=1}^{n} \frac{1}{\alpha_k} \Big[(\tau_k - z\beta_k - \bar{z}\gamma_k + |z|^2 \delta_k) \\ &\qquad - \frac{|z|^2}{1 + m_{n,z}(\zeta)} \{(1 + \tau_k)\delta_k - \gamma_k \beta_k\} - m_{n,z}(\zeta) \Big] \\ &= \frac{1}{n} f(m_{n,z}(\zeta)) \sum_{k=1}^{n} \frac{1}{\alpha_k} \Big[(\tau_k - m_{n,z}(\zeta)) \left\{ 1 - \frac{|z|^2 \delta_k}{1 + m_{n,z}(\zeta)} \right\} \\ &\qquad - z\beta_k - \bar{z}\gamma_k + \frac{|z|^2}{1 + m_{n,z}(\zeta)} \beta_k \gamma_k \Big]. \end{aligned} \tag{19}$$

For any $z_1, \ldots, z_n \in \mathbb{C}$ and $\ell \in \mathbb{N}$, by Jensen's inequality,

$$\left| \frac{1}{n} \sum_{i=1}^{n} z_i \right|^{\ell} \le \frac{1}{n} \sum |z_i|^{\ell}. \tag{20}$$

As we plan to invoke this inequality, it suffices for our purposes to bound the moment of each summand in (19). Using Corollary A.4,

$$\begin{aligned}
\mathbb{E}[|\beta_k|^{2p}] &= \mathbb{E}\left[|x_k^*[P_{z,\zeta}^{(k)}]^{-1}e_k e_k^T[P_{z,\zeta}^{(k*)}]^{-1}x_k|^p\right] \\
&\le \frac{2^{p-1}}{c_n^{p/2}}\mathbb{E}\left|(\sqrt{c_n}x_k^*[P_{z,\zeta}^{(k)}]^{-1}e_k e_k^T[P_{z,\zeta}^{(k*)}]^{-1}(\sqrt{c_n}x_k) - \operatorname{tr}\left([P_{z,\zeta}^{(k)}]^{-1}e_k e_k^T[P_{z,\zeta}^{(k*)}]^{-1}\right)\right|^p \\
&\quad + \frac{2^{p-1}}{c_n^{p/2}}\left|\operatorname{tr}\left([P_{z,\zeta}^{(k)}]^{-1}e_k e_k^T[P_{z,\zeta}^{(k*)}]^{-1}\right)\right|^p \\
&\le C(p)\frac{\omega_{2p}+1}{c_n^{p/2}|\Im(\zeta)|^{2p}} \\
&\le \frac{C(p)\omega_{4p}}{c_n^{p/2}|\Im(\zeta)|^{2p}} \qquad (21)
\end{aligned}$$

where $C(p)$ is a constant that only depends on p and may vary from line to line. An identical computation yields

$$\mathbb{E}[|\gamma_k|^{2p}] \le \frac{C(p)\omega_{4p}}{c_n^{p/2}|\Im(\zeta)|^{2p}}. \qquad (22)$$

By Lemma A.2, we have

$$\left|m_{n,z}(\zeta) - \frac{1}{n}\operatorname{tr}[P_{z,\zeta}^{(k)}]^{-1}\right| = \frac{1}{n}\left|\operatorname{tr}(P_{z,\zeta}^{-1} - [P_{z,\zeta}^{(k)}]^{-1})\right| \le \frac{1}{n|\Im(\zeta)|}.$$

Therefore

$$\begin{aligned}
&\mathbb{E}\left[|\tau_k - m_{n,z}(\zeta)|^{2p}\right] \\
&\le 2^{2p}\mathbb{E}\left|\tau_k - \frac{1}{n}\operatorname{tr}[P_{z,\zeta}^{(k)}]^{-1}\right|^{2p} + \frac{2^{2p}}{n^{2p}|\Im(\zeta)|^{2p}} \\
&\le 2^{4p}\mathbb{E}\left|\tau_k - \frac{1}{c_n}\sum_{i\in I_k}[P_{z,\zeta}^{(k)}]_{ii}^{-1}\right|^{2p} + 2^{4p}\mathbb{E}\left|\frac{1}{c_n}\sum_{i\in I_k}[P_{z,\zeta}^{(k)}]_{ii}^{-1} - \frac{1}{n}\operatorname{tr}[P_{z,\zeta}^{(k)}]^{-1}\right|^{2p} + \frac{2^{2p}}{n^{2p}|\Im(\zeta)|^{2p}}. \qquad (23)
\end{aligned}$$

We recall that $\tau_k = x_k^*[P_{z,\zeta}^{(k)}]^{-1}x_k$, where x_k is a band vector already scaled by $1/\sqrt{c_n}$. So, from Corollary A.4, we can conclude that

$$\mathbb{E}\left|\tau_k - \frac{1}{c_n}\sum_{i\in I_k}[P_{z,\zeta}^{(k)}]_{ii}^{-1}\right|^{2p} \le \frac{C(p)\omega_{4p}}{c_n^{p/2}|\Im(\zeta)|^{2p}}$$

where I_k denotes the indices in the support of x_k.

To estimate the second term of (23), we use Lemma A.5 to write

$$\mathbb{E}\left|\frac{1}{c_n}\sum_{i\in I_k}[P_{z,\zeta}^{(k)}]_{ii}^{-1} - \frac{1}{n}\operatorname{tr}[P_{z,\zeta}^{(k)}]^{-1}\right|^{2p} \le \mathbb{E}\left|\frac{1}{c_n}\sum_{i\in I_k}[P_{z,\zeta}]_{ii}^{-1} - \frac{1}{n}\operatorname{tr}[P_{z,\zeta}]^{-1}\right|^{2p} + \frac{2^{2p}}{c_n^{2p}\Im(\zeta)^{2p}}. \qquad (24)$$

The first expectation on the right hand side can be further decomposed as

$$\mathbb{E}\left|\frac{1}{c_n}\sum_{i\in I_k}[P_{z,\zeta}]_{ii}^{-1} - \frac{1}{n}\operatorname{tr}[P_{z,\zeta}]^{-1}\right|^{2p}$$

$$\leq \frac{2^{2p}}{c_n^{2p}}\mathbb{E}\left|\sum_{i\in I_k}[P_{z,\zeta}]_{ii}^{-1}-\sum_{i\in I_k}\mathbb{E}[P_{z,\zeta}]_{ii}^{-1}\right|^{2p}+\frac{2^{2p}}{n^{2p}}\mathbb{E}\left|\mathrm{tr}\{P_{z,\zeta}\}^{-1}]-\mathbb{E}\,\mathrm{tr}[P_{z,\zeta}]^{-1}\right|^{2p}. \tag{25}$$

In the above estimate, we have used the fact that we have a periodic block band matrix with iid entries, therefore $\mathbb{E}[\{P_{z,\zeta}\}_{ii}^{-1}]=\mathbb{E}[\{P_{z,\zeta}\}_{11}^{-1}]$ for all $1\leq i\leq n$, which is the conclusion of Lemma A.8. Now, we estimate the first term of (25) via a simple martingale decomposition.

Let $\mathcal{F}_k=\sigma\left(\{x_i:1\leq i\leq k\}\right)$ be the sigma algebra generated by the first k columns of X. Let us define

$$h(X)=\sum_{i\in I_k}[P_{z,\zeta}^{(k)}]_{ii}^{-1}. \tag{26}$$

Then we have the telescoping sum

$$h(X)-\mathbb{E}[h(X)]=\sum_{k=1}^{n}\left[\mathbb{E}[h(X)|\mathcal{F}_k]-\mathbb{E}[h(X)|\mathcal{F}_{k-1}]\right],$$

where $\mathcal{F}_0$ is the trivial sigma algebra. Using Lemma A.5, we have

$$|\mathbb{E}[h(X)|\mathcal{F}_k]-\mathbb{E}[h(X)|\mathcal{F}_{k-1}]|\leq 2/|\Im(\zeta)|.$$

Now by Corollary A.7,

$$\mathbb{E}[|h(X)-\mathbb{E}[h(X)]|^{2p}]\leq\frac{C(p)n^p}{|\Im(\zeta)|^{4p}},$$

where $C(p)$ is a constant that depends only on p.

As above, using Lemma A.5 and Result A.6, we estimate the second term of (25) by

$$\mathbb{E}\left|\mathbb{E}[\mathrm{tr}\{P_{z,\zeta}^{(k)}\}^{-1}]-\mathrm{tr}[P_{z,\zeta}^{(k)}]^{-1}\right|^{2p}\leq\frac{C(p)n^p}{|\Im(\zeta)|^{4p}}.$$

Using the above estimates in (23), we obtain

$$\mathbb{E}\left[|\tau_k-m_{n,z}(\zeta)|^{2p}\right]\leq\frac{C(p)}{|\Im(\zeta)|^{4p}}\left(\frac{n}{c_n^2}\right)^p, \tag{27}$$

To complete the estimates of (19), we need to lower bound $(f(m_{n,z}(\zeta)))^{-1}$ and α_k (recall that α_k is defined in (10)).

Since $\Im(\zeta)>0$, it follows that

$$\delta:=\int_0^\infty\frac{1}{|\lambda-\zeta|^2}\,d\mu_{X_zX_z^*}(\lambda)>0.$$

As a result, for any $\zeta\in\mathbb{C}$ with $\Im(\zeta)>0$,

$$\Im(\zeta m_{n,z}(\zeta))=\int_0^\infty\frac{\Im(\zeta)\lambda}{|\lambda-\zeta|^2}\,d\mu_{X_zX_z^*}(\lambda)\geq 0$$
$$\Im(m_{n,z}(\zeta))=\int_0^\infty\frac{\Im(\zeta)}{|\lambda-\zeta|^2}\,d\mu_{X_zX_z^*}(\lambda)\geq\Im(\zeta)\delta>0.$$

Using the above estimates, we have

$$\begin{aligned}|\Im(f(m_{n,z}(\zeta))^{-1})| &= \left|\Im\left[\frac{|z|^2}{1+m_{n,z}(\zeta)} - (1+m_{n,z}(\zeta))\zeta\right]\right| \\ &= \left|\left[\frac{|z|^2\Im(\overline{m_{n,z}(\zeta)})}{|1+m_{n,z}(\zeta)|^2} - \Im(\zeta) - \Im(\zeta m_{n,z}(\zeta))\right]\right| \\ &\geq |\Im(\zeta)|.\end{aligned}$$

Therefore

$$|f(m_{n,z}(\zeta))^{-1}| \geq |\Im(f(m_{n,z}(\zeta))^{-1})| \geq |\Im(\zeta)|. \tag{28}$$

Following the similar computation as (A2), we can also conclude that

$$|\alpha_k| \geq \delta|\Im(\zeta)|. \tag{29}$$

Finally, plugging (29), (28), (21), (22), (27) into (19) gives the desired bound (14). □

Next, we provide a deterministic upper bound on $|1 - r_{n,z}(\zeta)|$.

Lemma 3.5. *Under the assumptions of Theorem 3.1,*

$$|1 - r_{n,z}(\zeta)| \geq \frac{|\Im(\zeta)|}{4\sqrt{A}}. \tag{30}$$

Proof. Let us denote

$$\begin{aligned} A_{n,z}(\zeta) &:= 1 + m_{n,z}(\zeta) & A_z(\zeta) &:= 1 + m_z(\zeta) \\ B_{n,z}(\zeta) &:= |z|^2 - \zeta A_{n,z}(\zeta)^2 & B_z(\zeta) &:= |z|^2 - \zeta A_z(\zeta)^2 \\ \epsilon_{n,z}(\zeta) &:= m_{n,z}(\zeta) - f(m_{n,z}(\zeta)). \end{aligned}$$

Let $m_z(\zeta)$ be the solution of the equation $m_z(\zeta) = A_z(\zeta)/B_z(\zeta)$ satisfying $\Im(\sqrt{\zeta}m_z(\zeta)) > 0$ when $\Im(\sqrt{\zeta}) > 0$, where we have used the negative real axis for the branch cut of the square root function. The existence of such a solution is well-known in the circular law literature (see Section 11.4 in [9]).

Observe that as per the above notations, we may write

$$\begin{aligned} f(m_{n,z}(\zeta)) &= \frac{A_{n,z}(\zeta)}{B_{n,z}(\zeta)}, \\ m_{n,z}(\zeta) &= \frac{A_{n,z}(\zeta)}{B_{n,z}(\zeta)} + \epsilon_{n,z}(\zeta). \end{aligned}$$

Using the fact that $|ab| \leq \frac{1}{2}(|a|^2 + |b|^2)$ for $a, b \in \mathbb{C}$, and employing a similar calculation as in [47], we write

$$\begin{aligned} |1 - r_{n,z}(\zeta)| &= \left|1 - \frac{|z|^2 + \zeta A_z(\zeta)A_{n,z}(\zeta)}{B_z(\zeta)B_{n,z}(\zeta)}\right| \\ &\geq 1 - \left|\frac{|z|^2 + \zeta A_z(\zeta)A_{n,z}(\zeta)}{B_z(\zeta)B_{n,z}(\zeta)}\right| \\ &\geq 1 - \frac{|z|^2 + |\zeta A_z(\zeta)A_{n,z}(\zeta)|}{|B_z(\zeta)B_{n,z}(\zeta)|} \\ &\geq \frac{1}{2}\left(1 - \frac{|z|^2 + |\sqrt{\zeta}A_z(\zeta)|^2}{|B_z(\zeta)|^2}\right) + \frac{1}{2}\left(1 - \frac{|z|^2 + |\sqrt{\zeta}A_{n,z}(\zeta)|^2}{|B_{n,z}(\zeta)|^2}\right). \end{aligned} \tag{31}$$

Now, we estimate lower bounds for each expression of (31). We proceed as follows:

$$\begin{aligned}
\Im(\sqrt{\zeta}A_{n,z}(\zeta)) &= \Im(\sqrt{\zeta}m_{n,z}(\zeta)) + \Im(\sqrt{\zeta}) \\
&= \Im\left[\frac{\sqrt{\zeta}A_n, z(\zeta)\bar{B}_{n,z}(\zeta)}{|B_{n,z}(\zeta)|^2}\right] + \Im(\sqrt{\zeta}\epsilon_{n,z}(\zeta)) + \Im(\sqrt{\zeta}) \\
&= \Im\left[\frac{\sqrt{\zeta}A_{n,z}(\zeta)(|z|^2 - \overline{\zeta A_{n,z}(\zeta)^2}}{|B_{n,z}(\zeta)|^2}\right] + \Im(\sqrt{\zeta}\epsilon_{n,z}(\zeta)) + \Im(\sqrt{\zeta}) \\
&= \frac{\Im\big[\sqrt{\zeta}A_{n,z}(\zeta)|z|^2 - |\sqrt{\zeta}A_{n,z}(\zeta)|^2\overline{\sqrt{\zeta}A_{n,z}(\zeta)}\big]}{|B_{n,z}(\zeta)|^2} + \Im(\sqrt{\zeta}\epsilon_{n,z}(\zeta)) + \Im(\sqrt{\zeta}) \\
&= \Im(\sqrt{\zeta}A_{n,z}(\zeta))\left[\frac{|z|^2 + |\sqrt{\zeta}A_{n,z}(\zeta)|^2}{|B_{n,z}(\zeta)|^2}\right] + \Im(\sqrt{\zeta}\epsilon_{n,z}(\zeta)) + \Im(\sqrt{\zeta}).
\end{aligned}$$

Consequently, we have

$$1 - \frac{|z|^2 + |\sqrt{\zeta}A_{n,z}(\zeta)|^2}{|B_{n,z}(\zeta)|^2} = \frac{\Im(\sqrt{\zeta}\epsilon_{n,z}(\zeta)) + \Im(\sqrt{\zeta})}{\Im(\sqrt{\zeta}A_{n,z}(\zeta))} = \frac{\Im(\sqrt{\zeta}\epsilon_{n,z}(\zeta)) + \Im(\sqrt{\zeta})}{\Im(\sqrt{\zeta}) + \Im(\sqrt{\zeta}m_{n,z}(\zeta))}. \tag{32}$$

Similarly,

$$1 - \frac{|z|^2 + |\sqrt{\zeta}A_z(\zeta)|^2}{|B_z(\zeta)|^2} = \frac{\Im(\sqrt{\zeta})}{\Im(\sqrt{\zeta}A_z(\zeta))} = \frac{\Im(\sqrt{\zeta})}{\Im(\sqrt{\zeta}) + \Im(\sqrt{\zeta}m_z(\zeta))}. \tag{33}$$

Recall that we have chosen the solution $m_z(\zeta)$ such that $\Im(\sqrt{\zeta}m_z(\zeta))$ and $\Im(\sqrt{\zeta})$ have the same sign. Therefore,

$$0 \le \frac{\Im(\sqrt{\zeta})}{\Im(\sqrt{\zeta}) + \Im(\sqrt{\zeta}m_z(\zeta))} = 1 - \frac{|z|^2 + |\sqrt{\zeta}A_z(\zeta)|^2}{|B_z(\zeta)|^2} = 1 - \frac{|z|^2}{|B_z(\zeta)|^2} - |\sqrt{\zeta}m_z(\zeta)|^2.$$

As a result,

$$|\sqrt{\zeta}m_z(\zeta)| \le 1.$$

Using the the above estimate in (33) and the fact that $\Im(\sqrt{\zeta}m_z(\zeta))$ and $\Im(\sqrt{\zeta})$ have the same sign, we obtain

$$\begin{aligned}
1 - \frac{|z|^2 + |\sqrt{\zeta}A_z(\zeta)|^2}{|B_z(\zeta)|^2} &= \frac{\Im(\sqrt{\zeta})}{\Im(\sqrt{\zeta}) + \Im(\sqrt{\zeta}m_z(\zeta))} \\
&= \frac{|\Im(\sqrt{\zeta})|}{|\Im(\sqrt{\zeta})| + |\Im(\sqrt{\zeta}m_z(\zeta))|} \\
&\ge \frac{|\Im(\sqrt{\zeta})|}{|\Im(\sqrt{\zeta})| + 1} \\
&= \frac{1}{1 + |\Im(\sqrt{\zeta})|^{-1}} \\
&\ge \frac{1}{1 + 3\sqrt{A}|\Im(\zeta)|^{-1}} \\
&\ge \frac{|\Im(\zeta)|}{4\sqrt{A}},
\end{aligned} \tag{34}$$

where the second to last inequality follows from the fact that $|\Im(\sqrt{\zeta})| > |\Im(\zeta)|/3\sqrt{A}$ which is implied by the assumption $\zeta \in \{\zeta \in \mathbb{C} : -A < \Re(\zeta) < A, 0 < \Im(\zeta) < 1\}$ and $A > 1$.

Similarly,

$$1 - \frac{|z|^2 + |\sqrt{\zeta} A_{n,z}(\zeta)|^2}{|B_{n,z}(\zeta)|^2} \geq \frac{|\Im(\zeta)|}{4\sqrt{A}}. \tag{35}$$

Using the estimates (35), (34) in (31), we have

$$|1 - r_{n,z}(\zeta)| \geq \frac{|\Im(\zeta)|}{4\sqrt{A}}.$$

□

Theorem 3.1 follows easily from the above calculations.

Proof of Theorem 3.1. By Lemma 3.3,

$$\mathbb{E}[|m_{n,z}(\zeta) - m_z(\zeta)|^{2p}] = \mathbb{E}|[1 - r_{n,z}(\zeta)]^{-1}[m_{n,z}(\zeta) - f(m_{n,z}(\zeta))]|^{2p} \tag{36}$$

Therefore, by Lemmas 3.4 and 3.5,

$$\mathbb{E}[|m_{n,z}(\zeta) - m_z(\zeta)|^{2p}] \leq \frac{C(p) A^p \omega_{4p}}{|\Im(\zeta)|^{8p}} \left[\left(\frac{n}{c_n^2} \right)^p + \frac{1}{c_n^{p/2}} \right].$$

Since m_z is the Stieltjes transform of ν_z, it is a well-known property (see, for example, Section 11.4 and (11.4.1) in [9]) that m_z is the unique solution of (9) satisfying $\Im(\zeta m_z(\zeta^2)) > 0$ and $\Im(m_z(\zeta)) > 0$ when $\Im(\zeta) > 0$. □

4. Proof of Theorem 1.4

4.1. **Spectral norm bound.** Before proving Theorem 1.4, we note the following spectral norm bound on X.

Proposition 4.1 (Spectral norm bound)**.** *There exists a constant $K > 0$ such that $\|X\| \leq K$ with probability $1 - o(1)$.*

Proof. For any vector $v \in \mathbb{C}^n$, it follows from the block structure of X that

$$\|Xv\| \leq C\|v\| \left(\max_{1 \leq i \leq m} \|T_i\| + \max_{1 \leq i \leq m} \|U_i\| + \max_{1 \leq i \leq m} \|D_i\| \right),$$

where $C > 0$ is an absolute constant. The claim then follows from Lemma 2.4. □

4.2. **Proof of Theorem 1.4.** In order to complete the proof of Theorem 1.4, we will use the following replacement principle from [81]. Let $\|A\|_2$ denote the Hilbert–Schmidt norm of the matrix A defined by the formula

$$\|A\|_2 := \sqrt{\operatorname{tr}(AA^*)} = \sqrt{\operatorname{tr}(A^*A)}.$$

Theorem 4.2 (Replacement principle; Theorem 2.1 from [81])**.** *Suppose for each n that G and X are $n \times n$ ensembles of random matrices. Assume that:*

(i) the expression

$$\frac{1}{n}\|G\|_2^2 + \frac{1}{n}\|X\|_2^2$$

is bounded in probability (resp. almost surely);

(ii) for almost all complex numbers z,

$$\frac{1}{n}\log|\det(G_z)| - \frac{1}{n}\log|\det(X_z)|$$

converges in probability (resp. almost surely) to zero and, in particular, for fixed z, these determinants are nonzero with probability $1-o(1)$ (resp. almost surely nonzero for all but finitely many n).

Then

$$\mu_G - \mu_X$$

converges in probability (resp. almost surely) to zero.

We will apply the replacement principle to the normalized band matrix X, while the other matrix is taken to be $G := \frac{1}{\sqrt{n}}\tilde{G}$, where the entries of the $n \times n$ matrix $\tilde{G}$ are iid standard Gaussian random variables, i.e., $\tilde{G}$ is a Ginibre matrix. As the limiting behavior of μ_G is known to be almost surely the circular law [81], it will suffice, in order to complete the proof of Theorem 1.4, to check the two conditions of Theorem 4.2.

Condition (i) from Theorem 4.2 follows by the law of large numbers. Thus, it suffices to verify the second condition. To do so, we introduce the following notation inspired by Chapter 11 of [9]. For $z \in \mathbb{C}$, we define the following empirical distributions constructed from the squared singular values of X_z and G_z:

$$\nu_{X_z}(\cdot) := \frac{1}{n}\sum_{i=1}^{n} \delta_{s_i^2(X_z)}(\cdot)$$

and

$$\nu_{G_z}(\cdot) := \frac{1}{n}\sum_{i=1}^{n} \delta_{s_i^2(G_z)}(\cdot).$$

It follows that

$$\frac{1}{n}\log|\det(X_z)| - \frac{1}{n}\log|\det(G_z)| = \frac{1}{2}\int_0^\infty \log x \; \nu_{X_z}(dx) - \frac{1}{2}\int_0^\infty \log x \; \nu_{G_z}(dx).$$

By Theorem 2.1 as well as Proposition 4.1, there exists a constant $K > 0$ (depending on z) such that

$$\int_0^\infty \log x \, \nu_{X_z}(dx) - \int_0^\infty \log x \, \nu_{G_z}(dx) = \int_{c_n^{-25m}}^K \log x \, \nu_{X_z}(dx) - \int_{c_n^{-25m}}^K \log x \, \nu_{G_z}(dx) \tag{37}$$

with probability $1-o(1)$. Here, the largest and smallest singular values of G_z can be controlled by the results in [80, 82]. We will apply the following lemma.

Lemma 4.3. *For any probability measure μ and ν on $\mathbb{R}$ and any $0 < a < b$,*

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

where

$$\|\mu - \nu\|_{[a,b]} := \sup_{x\in[a,b]} |\mu([a,x]) - \nu([a,x])|.$$

Proof. We rewrite

$$\int_a^b \log(x) d\mu(x) = \log(b)\mu([a,b]) - \int_a^b \int_x^b \frac{1}{t} dt d\mu(x).$$

Applying Fubini's theorem, we deduce that

$$\int_a^b \int_x^b \frac{1}{t} dt d\mu(x) = \int_a^b \frac{\mu([a,t])}{t} dt.$$

Similarly, the same equalities apply to ν. Thus, we obtain that

$$\begin{aligned}
&\left| \int_a^b \log(x) d\mu(x) - \int_a^b \log(x) d\nu(x) \right| \\
&\leq |\log(b)||\mu([a,b]) - \nu([a,b])| + \left| \int_a^b \frac{\mu([a,t]) - \nu([a,t])}{t} dt \right| \\
&\leq |\log b| \|\mu - \nu\|_{[a,b]} + \|\mu - \nu\|_{[a,b]} \int_a^b \frac{1}{t} dt,
\end{aligned}$$

from which the conclusion follows. □

Returning to (37) and applying the above lemma, we find that

$$\left| \frac{1}{n} \log |\det(X_z)| - \frac{1}{n} \log |\det(G_z)| \right| \leq C \frac{n}{b_n} \log(n) \|\nu_{X_z}(\cdot) - \nu_{G_z}(\cdot)\|_{[0,\infty)} \tag{38}$$

for a constant $C > 0$, where

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

for any probability measures μ and ν on $\mathbb{R}$. Let $\nu_z(\cdot)$ be the probability measure on $[0,\infty)$ from Theorem 3.1 (or equivalently, the probability measure defined in Section 11.4 of [9]). By the triangle inequality, it suffices to show that

$$\|\nu_{X_z}(\cdot) - \nu_z(\cdot)\|_{[0,\infty)} = O\left(\left(\frac{n \log n}{b_n^2} \right)^{1/31} \right) \tag{39}$$

and

$$\|\nu_{G_z}(\cdot) - \nu_z(\cdot)\|_{[0,\infty)} = O\left(\left(\frac{n \log n}{b_n^2} \right)^{1/31} \right) \tag{40}$$

with probability $1 - o(1)$. The convergence in (40) follows from Lemma 11.16 from [9]; in fact, the results in [9] provide a much better error bound which holds almost surely. Thus, it remains to establish (39), which is a consequence of the following lemma.

Lemma 4.4. *Let $\tilde{X}$ and X be as in Theorem 1.4 with $b_n \geq n^{32/33} \log n$. Then, for any fixed $z \in \mathbb{C}$,*

$$\|\nu_{X_z}(\cdot) - \nu_z(\cdot)\|_{[0,\infty)} = O\left(\left(\frac{n \log n}{b_n^2} \right)^{1/31} \right)$$

with probability $1 - o(1)$.

Proof. Fix $z \in \mathbb{C}$. For notational simplicity define

$$q_n := \frac{n \log n}{b_n^2}.$$

Let $m_{n,z}$ be the Stieltjes transform of $\nu_{X_z}(\cdot)$ and m_z be the Stieltjes transform of $\nu_z(\cdot)$. We consider both Stieltjes transforms only on the upper-half plane $\mathbb{C}^+$. On the upper-half plane, both Stieltjes transforms are Lipschitz:

$$|m_{n,z}(\zeta) - m_{n,z}(\xi)| \leq \frac{|\zeta - \xi|}{\Im\zeta \Im\xi}, \qquad |m_z(\zeta) - m_z(\xi)| \leq \frac{|\zeta - \xi|}{\Im\zeta \Im\xi}. \tag{41}$$

Fix $A > 0$ sufficiently large to be chosen later. Define the line segment in the complex plane:

$$L := \left\{ \zeta = \theta + i q_n^{2/31} \in \mathbb{C}^+ : -A \leq \theta \leq A \right\}. \tag{42}$$

Applying Theorem 3.1 and Markov's inequality, for any $\zeta \in L$, we have

$$\mathbb{P}\left(|m_{n,z}(\zeta) - m_z(\zeta)| \geq q_n^{5/31} \right) \leq \frac{C}{q_n^{26/31}} \frac{n}{b_n^2}$$

for a constant $C > 0$ which depends only on the moments of the atom variable ξ and A. Let $\mathcal{N}$ be a $q_n^{5/31}$-net of L. By a simple covering argument, $\mathcal{N}$ can be chosen so that $|\mathcal{N}| = O(q_n^{-5/31})$. Thus, by the union bound,

$$\mathbb{P}\left(\sup_{\zeta \in \mathcal{N}} |m_{n,z}(\zeta) - m_z(\zeta)| \geq q_n^{5/31} \right) \leq \frac{C}{q_n} \frac{n}{b_n^2} = \frac{C}{\log n} = o(1).$$

Using the Lipschitz continuity (41), this bound can be extended to all of L, and we obtain

$$\sup_{\zeta \in L} |m_{n,z}(\zeta) - m_z(\zeta)| = O(q_n^{1/31}). \tag{43}$$

with probability $1 - o(1)$.

To complete the proof, we will use Corollary B.15 from [9] and (43) to bound $\|\nu_{X_z}(\cdot) - \nu_z(\cdot)\|_{[0,\infty)}$. Indeed, take $K > 0$ sufficiently large so that $\nu_{X_z}([0, K]) = 1$ with probability $1 - o(1)$ and $\nu_z([0, K]) = 1$. Such a choice is always possible by Proposition 4.1 and since ν_z has compact support (a fact which can also be deduced from Proposition 4.1). Recall the parameter $A > 0$ used to define the line segment L (see (42)). Taking $A, a > 0$ sufficiently large, setting $\eta := q_n^{2/31}$, and letting $\zeta := \theta + i\eta$, Corollary B.15 from [9] implies that

$$\begin{aligned}
&\|\nu_{X_z}(\cdot) - \nu_z(\cdot)\|_{[0,\infty)} \\
&\leq C \left[\int_{-A}^{A} |m_{n,z}(\zeta) - m_z(\zeta)| d\theta + \frac{1}{\eta} \sup_x \int_{|y| \leq 2\eta a} |\nu_z((-\infty, x + y]) - \nu_z((-\infty, x])| dy \right],
\end{aligned}$$

where $C > 0$ depends only on the choice of A, K, a. The second term is bounded by Lemma 11.9 from [9]:

$$\frac{1}{\eta} \sup_x \int_{|y| \leq 2\eta a} |\nu_z((-\infty, x + y]) - \nu_z((\infty, x])| dy \leq C' \sqrt{\eta}$$

for a constant $C' > 0$ depending only on a. For the first term we apply (43) to obtain

$$\int_{-A}^{A} |m_{n,z}(\zeta) - m_z(\zeta)| d\theta = O\left(q_n^{1/31} \right)$$

with probability $1-o(1)$. Combining the two bounds above, we conclude that, with probability $1-o(1)$,

$$\|\nu_{X_z}(\cdot)-\nu(\cdot,z)\|_{[0,\infty)}=O\left(q_n^{1/31}\right),$$

which completes the proof of the lemma. □

Lemma 4.4 establish (39). Combining (39), (40) with (38) and taking $b_n \geq n^{32/33}\log n$ completes the proof of Theorem 1.4.

Appendix A. Auxiliary tools

Lemma A.1 (Sherman-Morrison formula; see Section 0.7.4 in [48])**.** *Let A and $A+vv^*$ be two invertible matrices, where $v\in\mathbb{C}^n$. Then*

$$v^*(A+vv^*)^{-1}=\frac{v^*A^{-1}}{1+v^*A^{-1}v}.$$

Lemma A.2. *Let $\zeta\in\mathbb{C}\backslash\mathbb{R}_+$, and A be an $n\times n$ non-negative definite matrix. Then for any $v\in\mathbb{C}^n$,*

$$|\operatorname{tr}[(A+vv^*-\zeta I)^{-1}-(A-\zeta I)^{-1}]|\leq\frac{1}{|\Im(\zeta)|}.$$

Proof. The proof is similar to Lemma 2.6 in [77]. Using the resolvent identity and Lemma A.1,

$$\begin{aligned}
&\operatorname{tr}[(A+vv^*-\zeta I)^{-1}-(A-\zeta I)^{-1}]\\
=&-\operatorname{tr}(A+vv^*-\zeta I)^{-1}vv^*(A-\zeta I)^{-1}\\
=&-v^*(A-\zeta I)^{-1}(A+vv^*-\zeta I)^{-1}v\\
=&-\frac{v^*(A-\zeta I)^{-1}(A-\zeta I)^{-1}v}{1+v^*(A-\zeta I)^{-1}v}.
\end{aligned}\tag{A1}$$

Let $A=\sum_{i=1}^n\lambda_i(A)u_iu_i^*$ be the spectral decomposition of A, where $\lambda_i(A)\geq 0$ for all $1\leq i\leq n$. Then

$$\begin{aligned}
|v^*(A-\zeta I)^{-1}(A-\zeta I)^{-1}v|&=\sum_{i=1}^n\frac{|u_i^*v|^2}{|\lambda_i(A)-\zeta|^2},\\
|1+v^*(A-\zeta I)^{-1}v|^2&=\left|1+\sum_{i=1}^n\frac{|u_i^*v|^2}{\lambda_i(A)-\zeta}\right|^2\\
&=\left|1+\sum_{i=1}^n\frac{(\lambda_i(A)-\bar\zeta)|u_i^*v|^2}{|\lambda_i(A)-\zeta|^2}\right|^2\\
&=\left|1+\sum_{i=1}^n\frac{(\lambda_i(A)-\Re(\zeta))|u_i^*v|^2}{|\lambda_i(A)-\zeta|^2}\right|^2+\left|\sum_{i=1}^n\frac{\Im(\zeta)|u_i^*v|^2}{|\lambda_i(A)-\zeta|^2}\right|^2\\
&\geq|\Im(\zeta)|^2\left|\sum_{i=1}^n\frac{|u_i^*v|^2}{|\lambda_i(A)-\zeta|^2}\right|^2.
\end{aligned}\tag{A2}$$

Plugging in the above estimates in (A1), we obtain the result. □

Lemma A.3 (Lemma 2.7 from [11] and Equation (2.5) in [71]). *Let $\xi = (\xi_1, \xi_2, \dots, \xi_n)$ be a random vector such that ξ_i are iid complex valued random variables with $\mathbb{E}[\xi_1] = 0$ and $\mathbb{E}[|\xi_1|^2] = 1$. Then for any deterministic $n \times n$ matrix A,*

$$\mathbb{E}[|\xi^* A\xi - \operatorname{tr} A|^p] \le C_1(p)((\mathbb{E}|\xi_1|^4 \operatorname{tr} A^*A)^{p/2} + \mathbb{E}[|\xi_1|^{2p}] \operatorname{tr}(A^*A)^{p/2}),$$
$$\mathbb{E}[|\xi^* A\xi|^p] \le C_2(p)\mathbb{E}[|\xi_1|^{2p}]((\operatorname{tr} A^*A)^{p/2} + |\operatorname{tr} A|^p),$$

where $C_1(p), C_2(p)$ are constants that depend only on p.

Corollary A.4. *Let $I \subset \{1, 2, \dots, n\}$ be a fixed index set and $\xi_1, \xi_2, \dots, \xi_n$ be a set of iid complex valued random variables with $\mathbb{E}[\xi_1] = 0$ and $\mathbb{E}[|\xi_1|^2] = 1$. Define $v = (v_1, v_2, \dots, v_n)$ where $v_i = \xi_i \mathbf{1}_{\{i \in I\}}$. Then for any fixed $n \times n$ deterministic matrix A we have*

$$\mathbb{E}\left[\left|v^* A v - \sum_{i \in I} a_{ii}\right|^p\right] \le C(p)|I|^{p/2}\mathbb{E}[|\xi_1|^{2p}]\|A\|^p.$$

Proof. Let us define an $n \times n$ matrix $\tilde{A}$ as $(\tilde{A})_{ij} = a_{ij}\mathbf{1}_{\{i \in I\}}\mathbf{1}_{\{j \in I\}}$, where $a_{ij} = (A)_{ij}$. Then, $v^*Av = v^*\tilde{A}v$. In addition, $\operatorname{tr} \tilde{A} = \sum_{i \in I} a_{ii}$. Therefore, using Lemma A.3 and the fact that $\operatorname{tr}(\tilde{A}^*\tilde{A}) \le |I|\|\tilde{A}\|^2 \le |I|\|A\|^2$, the claim of the corollary follows. □

Lemma A.5. *Let P and Q be two $n \times n$ non-negative definite matrices, then for any $\zeta \in \mathbb{C} \backslash \mathbb{R}_+$ and $I \subset \{1, 2, \dots, n\}$,*

$$\left|\sum_{k \in I}(P - \zeta I)^{-1}_{kk} - \sum_{i \in I}(Q - \zeta I)^{-1}_{kk}\right| \le \frac{2}{|\Im(\zeta)|} \operatorname{rank}(P - Q).$$

Proof. The above lemma is similar to Lemma C.3 from [21]. For the readers' convenience, we include the proof here. Using the resolvent identity, we have

$$(P - \zeta I)^{-1} - (Q - \zeta I)^{-1} = (P - \zeta I)^{-1}(Q - P)(Q - \zeta I)^{-1}.$$

Therefore, $r := \operatorname{rank}[(P - \zeta I)^{-1} - (Q - \zeta I)^{-1}] \le \operatorname{rank}(P - Q)$. Let us write the singular value decomposition as

$$(P - \zeta I)^{-1} - (Q - \zeta I)^{-1} = \sum_{i=1}^{r} s_i u_i v_i^*,$$

where $s_1, s_2, \dots, s_r$ are at most r non zero singular values of $(P - \zeta I)^{-1} - (Q - \zeta I)^{-1}$, and $\{u_1, u_2, \dots, u_r\}$, $\{v_1, v_2, \dots, v_r\}$ are two sets of orthonormal vectors. Consequently, we may write

$$(P - \zeta I)^{-1}_{kk} - (Q - \zeta I)^{-1}_{kk} = \sum_{i=1}^{r} s_i (e_k^T u_i)(v_i^* e_k).$$

Using Cauchy-Schwarz inequality,

$$\begin{aligned}\left|\sum_{k \in I}(P - \zeta I)^{-1}_{kk} - \sum_{k \in I}(Q - \zeta I)^{-1}_{kk}\right| &\le \sum_{i=1}^{r} s_i \sum_{k \in I} |e^T u_i||v_i^* e_k| \\ &\le \sum_{i=1}^{r} s_i \sqrt{\sum_{k \in I} |e_k^T u_i|^2} \sqrt{\sum_{k \in I} |v_k^* e_k|^2}\end{aligned}$$

$$\leq \sum_{i=1}^{r} s_i \|u\|\|v\|$$
$$\leq \sum_{i=1}^{r} s_i \leq \frac{2r}{|\Im(\zeta)|} \leq \frac{2}{|\Im(\zeta)|}\operatorname{rank}(P-Q),$$

where the second last inequality follows from the fact that $s_i \leq \|(P-\zeta I)^{-1} - (Q-\zeta I)^{-1}\| \leq 2/|\Im(\zeta)|$ for all $1 \leq i \leq r$. □

Result A.6 (Azuma-Hoeffding inequality; see [60]). *Let $\{\xi_k\}_k$ be a martingale with respect to the filtration $\{\mathcal{F}_k\}_k$ such that for all k, $|\xi_{k+1} - \xi_k| \leq c_k$ almost surely. Then for any $t > 0$*

$$\mathbb{P}(|\xi_n - \mathbb{E}[\xi_n]| > t) \leq 2\exp\left\{-\frac{t^2}{2\sum_{k=1}^{n} c_k^2}\right\}.$$

A simple consequence of the previous concentration inequality is a bound on the moments.

Corollary A.7. *Under the conditions of Result A.6, for $l \in \mathbb{N}$, we have*

$$\mathbb{E}[|\xi_n - \mathbb{E}\xi_n|^l] \leq C(l)\left(\sum_{k=1}^{n} c_k^2\right)^{l/2}$$

where $C(l)$ is a constant only depending on l.

Proof. This result can be deducued from the straightforward calculation using Result A.6,

$$\begin{aligned}\mathbb{E}[|\xi_n - \mathbb{E}\xi_n|^l] &= l\int_0^\infty t^{l-1}\mathbb{P}(|\xi_n - \mathbb{E}\xi_n| > t)\,dt \\ &\leq 2l\int_0^\infty t^{l-1}\exp\left(-\frac{t^2}{2\sum_{k=1}^{n} c_k^2}\right)dt \\ &= l\left(2\sum_{k=1}^{n} c_k^2\right)^{l/2}\int_0^\infty u^{l/2-1}e^{-u}\,du \\ &= l\Gamma(l/2)2^{l/2}\left(2\sum_{k=1}^{n} c_k^2\right)^{l/2},\end{aligned}$$

where Γ is the gamma function. □

Our final lemma is a technical observation which is of use in Section 3.

Lemma A.8. *We let X be the random matrix from Theorem 1.4 (without the restriction on the bandwidth). We recall the notation from Section 3. For fixed $z \in \mathbb{C}$ and ζ in the upper half of the complex plane,*

$$P_{z,\zeta} := (X_z X_z^*)_\zeta = (X - zI)(X - zI)^* - \zeta I.$$

Then for all $1 \leq i \leq n$,

$$\mathbb{E}[\{P_{z,\zeta}\}_{ii}^{-1}] = \mathbb{E}[\{P_{z,\zeta}\}_{11}^{-1}].$$

Proof. We divide $[n]$ into sets $I_1, \dots, I_m$ where $I_i = [(i-1)b_n + 1, ib_n] \cap \mathbb{N}$. Let P_{ij} denote the $n \times n$ permutation matrix that permutes the i-th and j-th column when acting from the left on a matrix. Observe that when $i, j \in I_k$ for some $k \in [m]$, $P_{ij} X_z P_{ij}^{-1}$ has the same distribution as X_z due to the iid assumption and block structure. Therefore, $X_z X_z^*$ has the same distribution as $P_{ij} X_z P_{ij}^T P_{ij} X_z^* P_{ij}^T = P_{ij} X_z X_z^* P_{ij}^T$. Thus,

$$(X_z X_z^* - \zeta I)_{ii}^{-1} \sim (P_{ij}(X_z X_z^* - \zeta I) P_{ij}^T)_{ii}^{-1} = (P_{ij}(X_z X_z^* - \zeta I)^{-1} P_{ij})_{ii} \sim (X_z X_z^* - \zeta I)_{jj}^{-1},$$

where we use $\sim$ to denote equality in distribution. This establishes that the expectation for any two indices in the same index block are identical. It remains to show that the expectations for the various blocks are the same. Here, we define a permutation that exploits the block-band structure. Let P be the permutation that cyclically shifts I_k to I_{k+1} maintaining the order within each block and using the convention that $I_{m+1} = I_1$. By the structure of the matrix and the iid assumption,

$$X_z X_z^* \sim P X_z X_z^* P^{-1}.$$

Thus,

$$\begin{aligned}(X_z X_z^* - \zeta I)_{11}^{-1} &\sim (P(X_z X_z^* - \zeta I) P^{-1})_{11}^{-1} \\ &= (P^{-1}(X_z X_z^* - \zeta I)^{-1} P)_{11} \\ &\sim (X_z X_z^* - \zeta I)_{b+1,b+1}^{-1}.\end{aligned}$$

Continuing inductively establishes the equivalence of all the expectations along the diagonal of $(X_z X_z^* - \zeta I)^{-1}$. □

Data availability statement

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

Acknowledgment

The authors thank the anonymous referees for useful feedback and corrections. K. Luh has been supported in part by NSF grant DMS-1702533. S. O'Rourke has been supported in part by NSF grants ECCS-1913131 and DMS-1810500.

References

[1] R. Adamczak and D. Chafaï. Circular law for random matrices with unconditional log-concave distribution. *Commun. Contemp. Math.*, 17(4):1550020, 22, 2015.
[2] R. Adamczak, D. Chafaï, and P. Wolff. Circular law for random matrices with exchangeable entries. *Random Structures Algorithms*, 48(3):454–479, 2016.
[3] J. Aljadeff, D. Renfrew, and M. Stern. Eigenvalues of block structured asymmetric random matrices. *J. Math. Phys.*, 56(10):103502, 14, 2015.
[4] J. Aljadeff, M. Stern, and T. Sharpee. Transition to chaos in random networks with cell-type-specific connectivity. *Phys. Rev. Lett.*, 114:088101, Feb 2015.
[5] S. Allesina, J. Grilli, G. Barabás, S. Tang, J. Aljadeff, and A. Maritan. Predicting the stability of large structured food webs. *Nature Communications*, 6, 2015.
[6] S. Allesina and S. Tang. The stability–complexity relationship at age 40: a random matrix perspective. *Population Ecology*, 57(1):63–75, 2015.
[7] J. Alt, L. Erdős, and T. Krüger. Local inhomogeneous circular law. *Ann. Appl. Probab.*, 28(1):148–203, 2018.
[8] G. W. Anderson and O. Zeitouni. A CLT for a band matrix model. *Probab. Theory Related Fields*, 134(2):283–338, 2006.

[9] Z. Bai and J. W. Silverstein. *Spectral analysis of large dimensional random matrices*. Springer Series in Statistics. Springer, New York, second edition, 2010.
[10] Z. D. Bai. Circular law. *Ann. Probab.*, 25(1):494–529, 1997.
[11] Z. D. Bai and J. W. Silverstein. No eigenvalues outside the support of the limiting spectral distribution of large-dimensional sample covariance matrices. *Ann. Probab.*, 26(1):316–345, 1998.
[12] A. S. Bandeira and R. van Handel. Sharp nonasymptotic bounds on the norm of random matrices with independent entries. *Ann. Probab.*, 44(4):2479–2506, 2016.
[13] A. Basak and A. Bose. Limiting spectral distributions of some band matrices. *Period. Math. Hungar.*, 63(1):113–150, 2011.
[14] A. Basak, N. Cook, and O. Zeitouni. Circular law for the sum of random permutation matrices. *Electron. J. Probab.*, 23:Paper No. 33, 51, 2018.
[15] S. Belinschi, A. Dembo, and A. Guionnet. Spectral measure of heavy tailed band and covariance random matrices. *Comm. Math. Phys.*, 289(3):1023–1055, 2009.
[16] L. V. Bogachev, S. A. Molchanov, and L. A. Pastur. On the density of states of random band matrices. *Mat. Zametki*, 50(6):31–42, 157, 1991.
[17] C. Bordenave, P. Caputo, and D. Chafaï. Spectrum of large random reversible Markov chains: heavy-tailed weights on the complete graph. *Ann. Probab.*, 39(4):1544–1590, 2011.
[18] C. Bordenave, P. Caputo, and D. Chafaï. Spectrum of non-Hermitian heavy tailed random matrices. *Comm. Math. Phys.*, 307(2):513–560, 2011.
[19] C. Bordenave, P. Caputo, and D. Chafaï. Circular law theorem for random Markov matrices. *Probab. Theory Related Fields*, 152(3-4):751–779, 2012.
[20] C. Bordenave and D. Chafaï. Around the circular law. *Probab. Surv.*, 9:1–89, 2012.
[21] C. Bordenave and A. Guionnet. Localization and delocalization of eigenvectors for heavy-tailed random matrices. *Probab. Theory Related Fields*, 157(3-4):885–953, 2013.
[22] P. Bourgade. Random band matrices. In *Proceedings of the International Congress of Mathematicians—Rio de Janeiro 2018. Vol. IV. Invited lectures*, pages 2759–2784. World Sci. Publ., Hackensack, NJ, 2018.
[23] P. Bourgade, L. Erdős, H.-T. Yau, and J. Yin. Universality for a class of random band matrices. *Adv. Theor. Math. Phys.*, 21(3):739–800, 2017.
[24] P. Bourgade, F. Yang, H.-T. Yau, and J. Yin. Random band matrices in the delocalized phase, II: generalized resolvent estimates. *J. Stat. Phys.*, 174(6):1189–1221, 2019.
[25] P. Bourgade, H.-T. Yau, and J. Yin. Local circular law for random matrices. *Probab. Theory Related Fields*, 159(3-4):545–595, 2014.
[26] P. Bourgade, H.-T. Yau, and J. Yin. The local circular law II: the edge case. *Probab. Theory Related Fields*, 159(3-4):619–660, 2014.
[27] G. Casati and V. Girko. Wigner's semicircle law for band random matrices. *Random Oper. Stochastic Equations*, 1(1):15–21, 1993.
[28] G. Casati, F. Izrailev, and L. Molinari. Scaling properties of the eigenvalue spacing distribution for band random matrices. *Journal of Physics A: Mathematical and General*, 24(20):4755–4762, oct 1991.
[29] G. Casati, L. Molinari, and F. Izrailev. Scaling properties of band random matrices. *Phys. Rev. Lett.*, 64(16):1851–1854, 1990.
[30] R. Chaudhuri, V. Jain, and N. S. Pillai. Universality and least singular values of random matrix products: a simplified approach. To appear in Bernoulli. Available at arXiv:2007.03595, 2020.
[31] N. Cook. The circular law for random regular digraphs with random edge weights. *Random Matrices Theory Appl.*, 6(3):1750012, 23, 2017.
[32] N. Cook. Lower bounds for the smallest singular value of structured random matrices. *Ann. Probab.*, 46(6):3442–3500, 2018.
[33] N. Cook. The circular law for random regular digraphs. *Ann. Inst. Henri Poincaré Probab. Stat.*, 55(4):2111–2167, 2019.
[34] N. Cook, W. Hachem, J. Najim, and D. Renfrew. Non-Hermitian random matrices with a variance profile (I): deterministic equivalents and limiting ESDs. *Electron. J. Probab.*, 23:Paper No. 110, 61, 2018.
[35] N. A. Cook, W. Hachem, J. Najim, and D. Renfrew. Non-hermitian random matrices with a variance profile (ii): properties and examples, 2020.

[36] G. Dubach and Y. Peled. On words of non-Hermitian random matrices. *Ann. Probab.*, 49(4):–, 2021.
[37] A. Edelman. The probability that a random real Gaussian matrix has k real eigenvalues, related distributions, and the circular law. *J. Multivariate Anal.*, 60(2):203–232, 1997.
[38] L. Erdős and A. Knowles. Quantum diffusion and eigenfunction delocalization in a random band matrix model. *Comm. Math. Phys.*, 303(2):509–554, 2011.
[39] L. Erdős, A. Knowles, and H.-T. Yau. Averaging fluctuations in resolvents of random band matrices. *Ann. Henri Poincaré*, 14(8):1837–1926, 2013.
[40] L. Erdős, A. Knowles, H.-T. Yau, and J. Yin. Delocalization and diffusion profile for random band matrices. *Comm. Math. Phys.*, 323(1):367–416, 2013.
[41] Y. V. Fyodorov and A. D. Mirlin. Scaling properties of localization in random band matrices: a σ-model approach. *Phys. Rev. Lett.*, 67(18):2405–2409, 1991.
[42] J. Ginibre. Statistical ensembles of complex, quaternion, and real matrices. *J. Mathematical Phys.*, 6:440–449, 1965.
[43] V. L. Girko. The circular law. *Teor. Veroyatnost. i Primenen.*, 29(4):669–679, 1984.
[44] V. L. Girko. The elliptic law. *Teor. Veroyatnost. i Primenen.*, 30(4):640–651, 1985.
[45] V. L. Girko. The circular law: ten years later. *Random Oper. Stochastic Equations*, 2(3):235–276, 1994.
[46] F. Götze, A. Naumov, and A. Tikhomirov. On a generalization of the elliptic law for random matrices. *Acta Phys. Polon. B*, 46(9):1737–1745, 2015.
[47] F. Götze and A. Tikhomirov. The circular law for random matrices. *Ann. Probab.*, 38(4):1444–1491, 2010.
[48] R. A. Horn and C. R. Johnson. *Matrix analysis*. Cambridge University Press, Cambridge, second edition, 2013.
[49] Y. Imry. Coherent propagation of two interacting particles in a random potential. *Europhysics Letters (EPL)*, 30(7):405–408, jun 1995.
[50] P. Jacquod and D. L. Shepelyansky. Hidden breit-wigner distribution and other properties of random matrices with preferential basis. *Phys. Rev. Lett.*, 75:3501–3504, Nov 1995.
[51] V. Jain. The strong circular law: A combinatorial view. *Random Matrices: Theory and Applications*, 0(0):2150031, 0.
[52] I. Jana. Clt for non-hermitian random band matrices with variance profiles. Available at arXiv:1904.11098, 2019.
[53] I. Jana, K. Saha, and A. Soshnikov. Fluctuations of linear eigenvalue statistics of random band matrices. *Theory Probab. Appl.*, 60(3):407–443, 2016.
[54] I. Jana and A. Soshnikov. Distribution of singular values of random band matrices; Marchenko-Pastur law and more. *J. Stat. Phys.*, 168(5):964–985, 2017.
[55] A. Khorunzhy. On spectral norm of large band random matrices. Available at arXiv:math-ph/0404017, 2004.
[56] L. Li and A. Soshnikov. Central limit theorem for linear statistics of eigenvalues of band random matrices. *Random Matrices Theory Appl.*, 2(4):1350009, 50, 2013.
[57] A. E. Litvak, A. Lytova, K. Tikhomirov, N. Tomczak-Jaegermann, and P. Youssef. Circular law for sparse random regular digraphs. *J. Eur. Math. Soc. (JEMS)*, 23(2):467–501, 2021.
[58] D.-Z. Liu and Z.-D. Wang. Limit distribution of eigenvalues for random Hankel and Toeplitz band matrices. *J. Theoret. Probab.*, 24(4):988–1001, 2011.
[59] R. M. May. Will a large complex system be stable? *Nature*, 238:413–414, 1972.
[60] C. McDiarmid. On the method of bounded differences. In *Surveys in combinatorics, 1989 (Norwich, 1989)*, volume 141 of *London Math. Soc. Lecture Note Ser.*, pages 148–188. Cambridge Univ. Press, Cambridge, 1989.
[61] M. L. Mehta. *Random matrices and the statistical theory of energy levels*. Academic Press, New York-London, 1967.
[62] M. L. Mehta. *Random matrices*, volume 142 of *Pure and Applied Mathematics (Amsterdam)*. Elsevier/Academic Press, Amsterdam, third edition, 2004.
[63] A. D. Mirlin, Y. V. Fyodorov, F.-M. Dittes, J. Quezada, and T. H. Seligman. Transition from localized to extended eigenstates in the ensemble of power-law random banded matrices. *Phys. Rev. E*, 54:3221–3230, Oct 1996.
[64] S. A. Molchanov, L. A. Pastur, and A. M. Khorunzhiĭ. Distribution of the eigenvalues of random band matrices in the limit of their infinite order. *Teoret. Mat. Fiz.*, 90(2):163–178, 1992.

[65] A. Naumov. Elliptic law for real random matrices. Available at arXiv:1201.1639, 2012.
[66] H. H. Nguyen. Random doubly stochastic matrices: the circular law. *Ann. Probab.*, 42(3):1161–1196, 2014.
[67] H. H. Nguyen and S. O'Rourke. The elliptic law. *Int. Math. Res. Not. IMRN*, (17):7620–7689, 2015.
[68] S. Olver and A. Swan. Evidence of the Poisson/Gaudin-Mehta phase transition for band matrices on global scales. *Random Matrices Theory Appl.*, 7(2):1850002, 21, 2018.
[69] S. O'Rourke, D. Renfrew, A. Soshnikov, and V. Vu. Products of independent elliptic random matrices. *J. Stat. Phys.*, 160(1):89–119, 2015.
[70] S. O'Rourke and A. Soshnikov. Products of independent non-Hermitian random matrices. *Electron. J. Probab.*, 16:no. 81, 2219–2245, 2011.
[71] B. Rider and J. W. Silverstein. Gaussian fluctuations for non-Hermitian random matrix ensembles. *Ann. Probab.*, 34(6):2118–2143, 2006.
[72] M. Rudelson and K. Tikhomirov. The sparse circular law under minimal assumptions. *Geom. Funct. Anal.*, 29(2):561–637, 2019.
[73] M. Rudelson and R. Vershynin. The Littlewood-Offord problem and invertibility of random matrices. *Adv. Math.*, 218(2):600–633, 2008.
[74] J. Schenker. Eigenvector localization for random band matrices with power law band width. *Comm. Math. Phys.*, 290(3):1065–1097, 2009.
[75] M. Shcherbina. On fluctuations of eigenvalues of random band matrices. *J. Stat. Phys.*, 161(1):73–90, 2015.
[76] D. Shlyakhtenko. Random Gaussian band matrices and freeness with amalgamation. *Internat. Math. Res. Notices*, (20):1013–1025, 1996.
[77] J. W. Silverstein and Z. D. Bai. On the empirical distribution of eigenvalues of a class of large-dimensional random matrices. *J. Multivariate Anal.*, 54(2):175–192, 1995.
[78] S. Sodin. The spectral edge of some random band matrices. *Ann. of Math. (2)*, 172(3):2223–2251, 2010.
[79] D. B. Stouffer and J. Bascompte. Compartmentalization increases food-web persistence. *Proceedings of the National Academy of Sciences*, 108(9):3648–3652, 2011.
[80] T. Tao and V. Vu. Random matrices: the circular law. *Commun. Contemp. Math.*, 10(2):261–307, 2008.
[81] T. Tao and V. Vu. Random matrices: universality of ESDs and the circular law. *Ann. Probab.*, 38(5):2023–2065, 2010. With an appendix by Manjunath Krishnapur.
[82] R. Vershynin. Introduction to the non-asymptotic analysis of random matrices. In *Compressed sensing*, pages 210–268. Cambridge Univ. Press, Cambridge, 2012.
[83] E. P. Wigner. Characteristic vectors of bordered matrices with infinite dimensions. *Ann. of Math. (2)*, 62:548–564, 1955.
[84] P. M. Wood. Universality and the circular law for sparse random matrices. *Ann. Appl. Probab.*, 22(3):1266–1300, 2012.
[85] H. Xi, F. Yang, and J. Yin. Local circular law for the product of a deterministic matrix with a random matrix. *Electron. J. Probab.*, 22:Paper No. 60, 77, 2017.
[86] F. Yang and J. Yin. Random band matrices in the delocalized phase, III: averaging fluctuations. *Probab. Theory Related Fields*, 179(1-2):451–540, 2021.
[87] J. Yin. The local circular law III: general case. *Probab. Theory Related Fields*, 160(3-4):679–732, 2014.