Forum of Mathematics, Sigma (2019), Vol. 7, e3, 72 pages
doi:10.1017/fms.2018.29

REGULARIZATION OF NON-NORMAL MATRICES BY GAUSSIAN NOISE—THE BANDED TOEPLITZ AND TWISTED TOEPLITZ CASES

ANIRBAN BASAK[1,3], ELLIOT PAQUETTE[2] and OFER ZEITOUNI[3,4]

[1] International Centre for Theoretical Sciences, Tata Institute of Fundamental Research, Bangalore 560089, India;
email: anirban.basak@icts.res.in

[2] Department of Mathematics, The Ohio State University, Tower 100, 231 W 18th Ave, Columbus, Ohio 43210, USA;
email: paquette.30@osu.edu

[3] Department of Mathematics, Weizmann Institute of Science, POB 26, Rehovot 76100, Israel

[4] Courant Institute, New York University, 251 Mercer St, New York, NY 10012, USA;
email: ofer.zeitouni@weizmann.ac.il

Received 17 April 2018; accepted 4 December 2018

Abstract

We consider the spectrum of additive, polynomially vanishing random perturbations of deterministic matrices, as follows. Let M_N be a deterministic $N \times N$ matrix, and let G_N be a complex Ginibre matrix. We consider the matrix $\mathcal{M}_N = M_N + N^{-\gamma} G_N$, where $\gamma > 1/2$. With L_N the empirical measure of eigenvalues of $\mathcal{M}_N$, we provide a general deterministic equivalence theorem that ties L_N to the singular values of $z - M_N$, with $z \in \mathbb{C}$. We then compute the limit of L_N when M_N is an upper-triangular Toeplitz matrix of finite symbol: if $M_N = \sum_{i=0}^{\mathfrak{d}} a_i J^i$ where $\mathfrak{d}$ is fixed, $a_i \in \mathbb{C}$ are deterministic scalars and J is the nilpotent matrix $J(i,j) = \mathbf{1}_{j=i+1}$, then L_N converges, as $N \to \infty$, to the law of $\sum_{i=0}^{\mathfrak{d}} a_i U^i$ where U is a uniform random variable on the unit circle in the complex plane. We also consider the case of slowly varying diagonals (twisted Toeplitz matrices), and, when $\mathfrak{d} = 1$, also of independent and identically distributed entries on the diagonals in M_N.

2010 Mathematics Subject Classification: 60B20 (primary); 15B05, 15B52 (secondary)

1. Introduction

Write G_N for an $N \times N$ random matrix whose entries are independent and identically distributed standard *complex* Gaussian variables (a *Ginibre* matrix), and let $\{M_N\}_{N=1}^{\infty}$ be a sequence of deterministic $N \times N$ matrices. Consider a noisy counterpart given by

$$\mathcal{M}_N := M_N + N^{-\gamma} G_N, \tag{1.1}$$

where $\gamma \in (1/2, \infty)$ is fixed, noting that by standard estimates, see [8, Corollary 1.2],

$$\|N^{-\gamma} G_N\| \to_{N\to\infty} 0 \quad \text{a.s.}, \tag{1.2}$$

where $\|\cdot\|$ denotes the operator norm. Let λ_i, $i = 1, \ldots, N$ denote the eigenvalues of $\mathcal{M}_N$, and let

$$L_N := N^{-1} \sum_{i=1}^{N} \delta_{\lambda_i} \tag{1.3}$$

denote the associated empirical measure. In this paper, we study the convergence of L_N for a class of matrices M_N. Discussions of background and related approaches are deferred to Sections 1.3 and 1.4.

1.1. Main results. For a probability measure μ on $\mathbb{C}$ which integrates the log function at infinity, and $z \in \mathbb{C}$, denote the *Logarithmic potential* associated with μ by

$$\mathcal{L}_\mu(z) := \int \log|z - y|\mu(dy). \tag{1.4}$$

The importance of the logarithmic potentials lies in the fact that the pointwise convergence of $\mathcal{L}_{\mu_n}(z)$ to a limit $\mathcal{L}_\mu(z)$ implies the weak convergence $\mu_n \to \mu$.

Our first main result is a deterministic equivalence theorem for $\mathcal{L}_{L_N}(z)$. We formulate here a simplified version under more stringent conditions than necessary, and refer to Theorem 2.1 for the general statement, which also has an explicit description of the functions g_N appearing in the statement of Theorem 1.1. Let $\mathrm{Id} = \mathrm{Id}_N$ stand for the identity matrix of dimension N. Let $J = J_N$ denote the nilpotent matrix with $J_{ij} = \mathbf{1}_{j=i+1}$ for $1 \leqslant i < j \leqslant N$.

THEOREM 1.1. *Fix $\gamma > 1/2$. Fix $z \in \mathbb{C}$ and let $s_N(z)$ denote the number of singular values of $M_N - z\,\mathrm{Id}$ smaller than $N^{-\gamma+1/2+\delta_N}$, where $0 < \delta_N \to_{N\to\infty} 0$. Suppose $s_N(z)\log N/N \to_{N\to\infty} 0$. Then, there exist explicit, deterministic functions $g_N(z)$ so that*

$$|\mathcal{L}_{L_N}(z) - g_N(z)| \to_{N\to\infty} 0 \quad \textit{in probability}.$$

The importance of Theorem 1.1 (and its more elaborate version, Theorem 2.1) lies in the fact that it reduces the question of weak convergence of the random

https://doi.org/10.1017/fms.2018.29 Published online by Cambridge University Press

empirical measure L_N to computations involving the deterministic matrices M_N. Still, these computations are, in general, nontrivial. The other results in this paper are instances in which these computations can be carried through and the limit of L_N can be described explicitly.

Our second main result deals with upper-triangular Toeplitz matrices of finite symbol, that is banded upper-triangular Toeplitz matrices.

THEOREM 1.2. *Let $a_i, i = 0, 1, \ldots, \mathfrak{d}$ be complex (deterministic) numbers. Let $M_N := \sum_{i=0}^{\mathfrak{d}} a_i J^i$, and let $\mathcal{M}_N$ and L_N be as in (1.1) and (1.3). Then L_N converges weakly in probability to the law of $\sum_{i=0}^{\mathfrak{d}} a_i U^i$, where U is uniformly distributed on the unit circle.*

A generalization of Theorem 1.2 to the twisted Toeplitz setup appears in Section 4, see Theorem 4.1 there. As the next theorem shows, in the case of two diagonals in M_N more can be said. For $x_1, x_2 \in \mathbb{R}$ we denote $x_1 \vee x_2 := \max\{x_1, x_2\}$.

THEOREM 1.3. *Let D_N be a diagonal matrix with entries d_i, set $M_N = D_N + J$ and let $\mathcal{M}_N$ be as in (1.1).*

(a) *Let d_i be independent and identically distributed random variables of law ν supported on a subset of a simply connected compact set with Lebesgue area 0. Then L_N converges weakly in probability to a measure μ characterized by $\mathcal{L}_\mu(z) = (\mathbb{E}_\nu \log|z - d_1|) \vee 0$.*

(b) *Let $f : [0, 1] \to \mathbb{C}$ be Hölder continuous and set $d_i = f(i/n)$. Then L_N converges weakly in probability to a probability measure μ satisfying*

$$\mu = \int_0^1 \mathrm{unif}_{f(z),1} dz,$$

where $\mathrm{unif}_{a,b}$ denotes the uniform law on a circle in the complex plane of radius b and center a.

See Section 3 for details and further examples, and note that Theorem 1.3(a) is Corollary 3.6, while Theorem 1.3(b) is Corollary 3.9.

An illustration of Theorems 1.2 and 1.3 is provided in Figure 1.

REMARK 1.4. We chose to consider throughout the paper only the case of perturbation matrices G_N which are complex Ginibre matrices. We believe that the results should carry over in a rather straightforward way to the case of real Ginibre matrices, and with a significant effort to the independent and identically distributed setup, in the same spirit as [27]. To avoid additional technicalities, we did not pursue these extensions here. Some recent results concerning perturbation of Toeplitz matrices by general noise appear in [2].

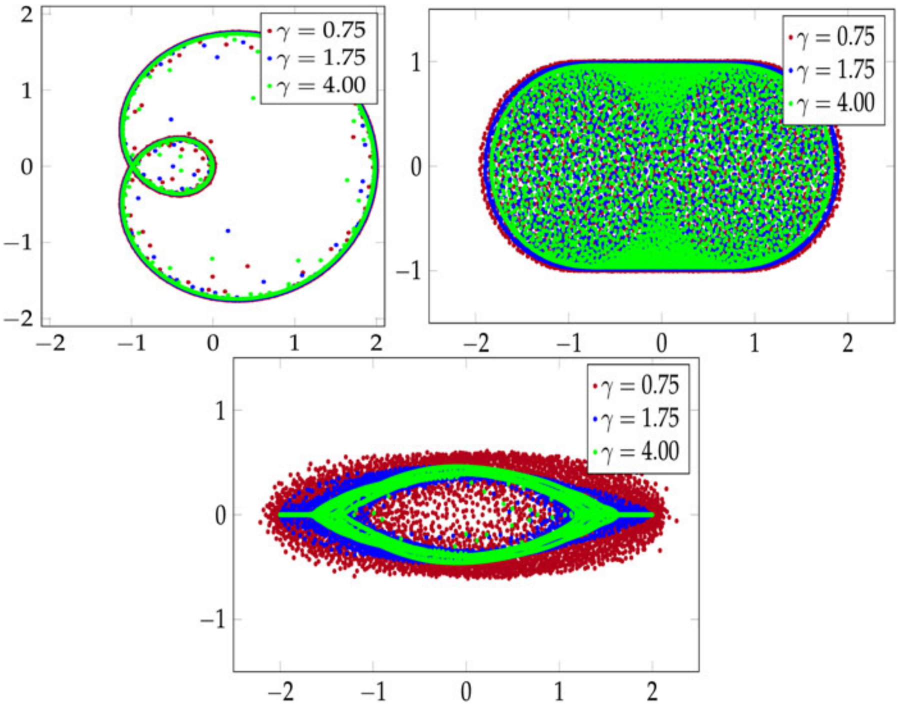

Figure 1. The eigenvalues of $\mathcal{M}_N$, with $N = 4000$ and various γ. On the top left, $M_N = J + J^2$. On the top right, $M_N = D_N + J$ with $D_N(i, i) = -1 + 2i/N$. On the bottom, $M_N = D_N + J$ with D_N independent and identically distributed uniform on $[-2, 2]$.

1.2. A Thouless-type formula. Both Theorems 1.2 and 1.3(a) can also be formulated in terms of Lyapunov exponents. Consider the vector space

$$V := \{x \in \mathbb{R}^N : ((M_N - z\,\mathrm{Id})x)_j = 0, 1 \leqslant j \leqslant N - \mathfrak{d}\}.$$

This space is $\mathfrak{d}$-dimensional. Further, to find $x \in V$, having chosen $x_1, x_2, \ldots, x_{\mathfrak{d}}$, one can solve for the remaining entries of x using the equations $((M_N - z\,\mathrm{Id})x)_j = 0$ for $1 \leqslant j \leqslant N - \mathfrak{d}$ to propagate the solution. Concisely, we can find $\mathfrak{d} \times \mathfrak{d}$ *transfer* matrices $(T_j(z))_1^{N-\mathfrak{d}}$ so that for all $1 \leqslant j \leqslant N - \mathfrak{d}$

$$(x_\ell)_{\ell=j+1}^{j+\mathfrak{d}} = T_j(z) \cdot (x_\ell)_{\ell=j}^{j+\mathfrak{d}-1}.$$

For details, see Definition 5.1; for the exposition here, the explicit form of this matrix will not be necessary.

In the setup of Theorem 1.2, these matrices will all be identical. In the setup of Theorem 1.3 part (a), the matrices will be independent and identically distributed scalars. In either case, the sequence $(T_j(z))_1^{N-\mathfrak{d}}$ is stationary, and we can consider the Lyapunov spectra as the set of values

$$\{\mu_1(z), \mu_2(z), \ldots, \mu_{\mathfrak{d}}(z)\} := \left\{\lim_{n\to\infty} \frac{1}{n} \log \|T_n(z)T_{n-1}(z)\cdots T_1(z)v\| : v \in \mathbb{C}^{\mathfrak{d}}\right\}.$$

In the setup of Theorem 1.3 part (a), there is a single Lyapunov eigenvalue, given by $\mu_1(z) = \mathbb{E}\log|d_1 - z|$. This allows us to write that

$$\mathcal{L}_{L_N}(z) \to \mu_1(z) \vee 0.$$

In the setup of Theorem 1.2, if we set $P(x) = \sum_{i=0}^{\mathfrak{d}} a_i x^i$ we have that

$$\mathcal{L}_{L_N}(z) \to \int_0^{2\pi} \log|P(e^{i\theta}) - z| \frac{d\theta}{2\pi}.$$

On the other hand, factorizing $P(x) - z = a_{\mathfrak{d}} \prod_{i=1}^{\mathfrak{d}} (x - \lambda_i(z))$, we can write

$$\int_0^{2\pi} \log|P(e^{i\theta}) - z| \frac{d\theta}{2\pi} = \sum_{i=1}^{\mathfrak{d}} ((\log|\lambda_i(z)|) \vee 0) + \log|a_{\mathfrak{d}}|.$$

Furthermore, in the Toeplitz case, the eigenvalues of $T_j(z) = T_1(z)$ are just the roots of the symbol $P(x)$, and the Lyapunov spectra are nothing but $\log|\lambda_i(z)|$ for $1 \leqslant i \leqslant \mathfrak{d}$. Hence, we have that in both Theorem 1.3 part (a) and Theorem 1.2,

$$\mathcal{L}_{L_N}(z) \to \sum_{i=1}^{\mathfrak{d}} (\mu_i(z) \vee 0) + \log|a_{\mathfrak{d}}|.$$

A similar result had appeared previously in the study of the Thouless formula for the strip [4, Theorem 2.4]. For the twisted Toeplitz cases (such as in Theorem 1.3 part (b)), the formula must be replaced by an average over local Lyapunov exponents.

1.3. Connection to pseudospectra. The fact that the spectrum of non-normal matrices and operators is not stable with respect to perturbations is well known, see for example [26] for a comprehensive account and [5] for a recent study. To illustrate the issue, we attach in Figure 2 an actual simulation of $U_N M_N U_N^*$ where M_N is as in Figure 1 and U_N is a random Haar-distributed unitary matrix. While the spectrum of M_N is real, the numerical simulations produce errors that make

https://doi.org/10.1017/fms.2018.29 Published online by Cambridge University Press

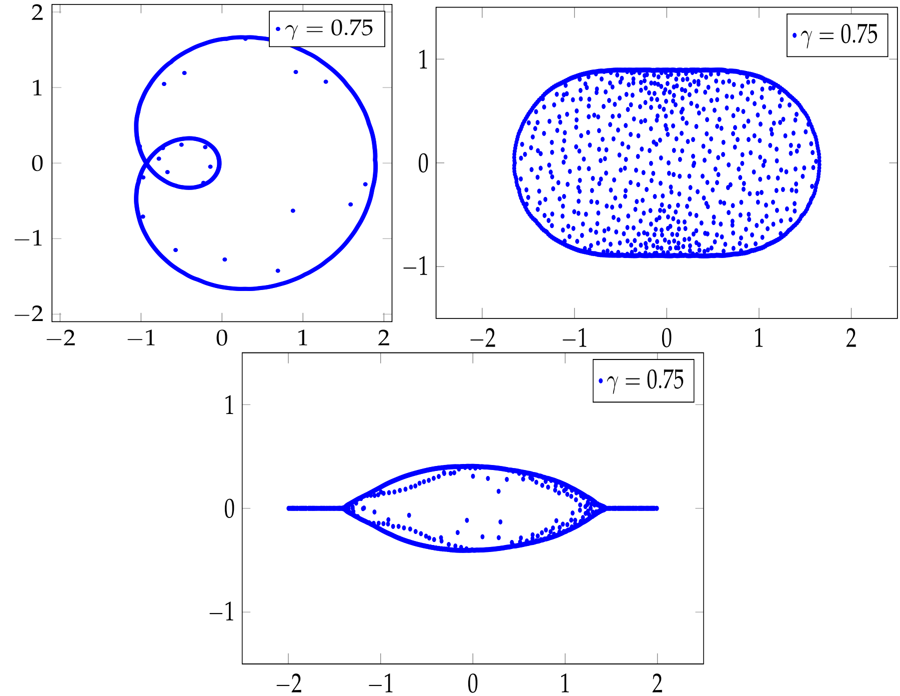

Figure 2. Numerical evaluation of eigenvalues of $U_N M_N U_N^*$ with U_N Haar-distributed unitary and $N = 1000$. On the top left, $M_N = J + J^2$. On the top right, $M_N = D_N + J$ with $D_N(i, i) = -1 + 2i/N$. On the bottom, $M_N = D_N + J$ with D_N independent and identically distributed uniform on $[-2, 2]$. The computation was performed in Scipy with a 64-bit floating point precision. Note the similarity with Figure 1.

the spectrum look similar to the one for the noisy perturbed model $\mathcal{M}_N$, compare with Figure 1. See [22] for early examples of the same phenomenon.

The ϵ–*pseudospectrum*, defined by

$$\Lambda_\epsilon(M_N) := \{z \in \mathbb{C} : \sigma_N(M_N - z\,\mathrm{Id}) \leqslant \epsilon\},$$

with $\sigma_N(\cdot)$ the smallest singular value, is a type of worst-case quantification of the instability of the spectrum. See [22] for the original formulation and [26] for an extensive background and applications in numerical analysis and beyond.

In the literature on pseudospectra, an outsize importance is placed on *exponentially good pseudospectra*, that is, Λ_ϵ where $\epsilon < e^{-\delta N}$ for some δ. Of particular relevance here, simulations of randomly perturbed non-normal

https://doi.org/10.1017/fms.2018.29 Published online by Cambridge University Press

matrices suggest that their spectra concentrate on sets that strongly resemble exponentially good pseudospectral level lines, see for example [15]. In particular, in the upper-triangular Toeplitz case, these curves are precisely the image of the unit circle in the complex plane by the Toeplitz symbol.

Furthermore, all of the models of non-normal matrices that we consider have been, not coincidentally, the subjects of study from the point of view of exponentially good pseudospectra. The work [22] describes many examples of non-normal matrices and gives plots of pseudospectral level lines adjacent to their perturbed eigenvalues. The top two plates in Figures 1 and 2 are Examples 2 and 4 from [22]. Subsequent work [15] proved, using transfer matrices, some estimates for the locations of the ϵ-pseudospectrum and exponentially good pseudospectrum of large Toeplitz matrices, and showed in the upper-triangular case that the latter converges to the spectrum of the limit Toeplitz operator, namely to the image of the unit circle by the symbol; our Theorem 1.2 shows that indeed, for upper-triangular symbols of finite support and under small Gaussian perturbations, the empirical measure converges to a limit with precisely this support. The work [25], motivated in part by the Hatano–Nelson model, considers the pseudospectra of random bidiagonal matrices, identifying four regions of distinct pseudospectral growth. Finally [24] computes the exponentially good pseudospectrum of some classes of twisted Toeplitz matrices, including the top-right example of Figures 1 and 2 (the 'Wilkinson' matrix). See also [23] for related results in the continuous setup.

As we shall see in Section 2, adding small Gaussian noise to $M_N - z\,\mathrm{Id}$ roughly has the effect of boosting any exponentially small singular values to the order of unity. Hence in situations in which there are only a few singular values of $M_N - z\,\mathrm{Id}$ that are exponentially small, the log potential at $M_N - z\,\mathrm{Id} + N^{-\gamma} G_N$ can be approximated by computing the log potential of $M_N - z\,\mathrm{Id}$ and subtracting from it the contribution of exponentially small singular values. Indeed, if the exponential growth rates of these extremal singular values are harmonic as functions of z away from the spectra, then a discontinuity in the Laplacian of the log potential occurs exactly where the exponential growth of extremal singular values changes signs. In particular, an exponentially good pseudospectral level line would be contained in the limiting spectral support of $M_N - z\,\mathrm{Id} + N^{-\gamma} G_N$. See Figure 3 for an illustration.

However, as a consequence of the theorems we show in this paper, we see that pseudospectrum alone is in general not sufficient for understanding the limiting spectral distribution of randomly perturbed non-normal matrices, except for special cases, for example, when only one singular value of $M_N - z\,\mathrm{Id}$ is exponentially small, or in the Toeplitz upper-triangular case.

https://doi.org/10.1017/fms.2018.29 Published online by Cambridge University Press

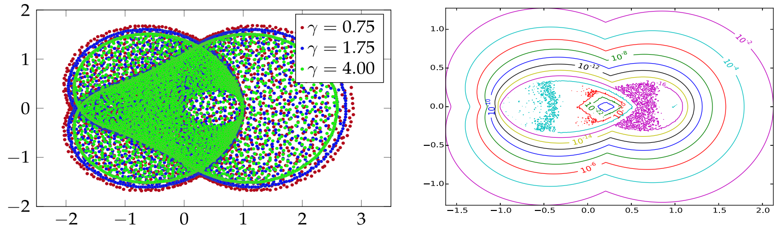

Figure 3. The matrix $M_N = D_N - D_N J + J^2$, with $D_N(i,i) = -1 + 2i/N$. The leftmost image shows eigenvalues of $M_N + N^{-\gamma} G_N$ for $N = 4000$. Theorem 4.1 gives the distributional convergence of the spectra of this matrix on adding Gaussian noise. The rightmost image shows pseudospectral level lines for $N = 100$. The levels displayed are 100^{-k}, for k ranging from 1 to 12. Pseudospectral lines generated by André Gaul's pseudopy package, based off Eigtool by Thomas G. Wright.

1.4. Previous results on typical perturbations and strategy.

Our work is an attempt to address the same issue from a 'typical' perturbation point of view, and in this sense continues the line of research initiated in [10, 27] and [6], which we now describe.

In [10], the authors consider the case where M_N converges in $*$-moments, that is, there exists an operator a in a noncommutative probability space $(\mathcal{A}, \mathrm{tr})$ so that for any noncommutative polynomial P, $N^{-1} \operatorname{tr} P(M_N, M_N^*) \to \operatorname{tr} P(a, a^*)$. Under a regularity assumption on a and the existence of polynomially vanishing perturbations of M_N with empirical measure converging to the spectral measure μ_a associated with a, they show that L_N converges to μ_a in probability. They further show that their assumptions are satisfied when $M_N = J$. The paper [27] shows that this result is stable in the sense that replacing G_N by a matrix with independent and identically distributed entries satisfying mild assumptions does not change the result. The proofs in [10] and [27] control the log potential of $\mathcal{M}_N$ by methods inspired by free probability, and in particular break down if the $*$-moment limit does not coincide with the limit of L_N. Further, in [10] one can find an example of a matrix M_N (with only one nonzero diagonal) where the latter is indeed the case—namely, $M_N(i,j) = J(i,j) \cdot \mathbf{1}_{j \neq 0 \mod \log N}$.

In [6], the authors consider the latter situation and prove a limit theorem, where the limit does depend on γ. The method of proof is very different from [10, 27]—it involves a combinatorial analysis of $\det \mathcal{M}_N(z)$ where $\mathcal{M}_N(z) := \mathcal{M}_N - z\,\mathrm{Id}$. Noting that $\mathcal{L}_{L_N}(z) = 1/N \log |\det \mathcal{M}_N(z)|$, concentration of measure arguments then identifies the terms contributing to the determinant.

https://doi.org/10.1017/fms.2018.29 Published online by Cambridge University Press

Our approach in this paper is related to [6] in that we also compute $|\det(\mathcal{M}_N(z))|$. However, our starting point is to relate the latter to a truncation of $M_N(z) = M_N - z\,\mathrm{Id}$, where the lowest lying singular values of $M_N(z)$ are eliminated (we refer to this as a 'deterministic equivalent model', using terminology borrowed from [11]). The level of truncation depends on γ, which parametrizes the strength of the perturbation. Once this step has been established, we can study the small singular values of $M_N(z)$ using transfer-matrix techniques, in case M_N is Toeplitz with a finite symbol or a slowly varying version of such a matrix. This analysis was not present in [6, 10, 27] and seems to be new in the context of the stability that we study.

We note that other approaches to the study of perturbations of non-normal operators exist. In particular, Sjöstrand and Vogel [17], [18] identify the limit of the empirical value of a random perturbation of a banded Toeplitz matrix with two nonzero diagonals, one above and one below the main diagonal. Their methods, which are quite different from ours, are limited to $\gamma > 5/2$, and yield more quantitative estimates on the empirical measure and its outliers.

The structure of the paper is as follows. In the next section, we introduce and prove the deterministic equivalence. Various standard algebraic facts needed for the proofs are collected in Appendix A. Section 3 presents the analysis of the deterministic equivalent model in the case where only two diagonals are present; the latter restriction simplifies the analysis because transfer matrices reduce to a scalar in that case. Section 4 treats the case of $\mathfrak{d} > 1$, and reduces the twisted Toeplitz case to piecewise constant twisted Toeplitz matrices, described in Theorem 4.4. The proof of the latter appears in Section 5.

1.5. Notation. We use throughout the following standard notation. For two real-valued sequences $\{a_n\}$ and $\{b_n\}$ we write $a_n = o(b_n)$ if $\limsup_{n\to\infty} b_n/a_n = 0$, $a_n = O(b_n)$ if $\limsup_{n\to\infty} |b_n|/|a_n| < \infty$, and $a_n \gg b_n$ if $\limsup_{n\to\infty} |a_n|/|b_n| = \infty$. For any $x \in \mathbb{R}$, we denote $x_- := -\min\{x, 0\}$ and $x_+ := \max\{x, 0\}$. For any $x > 0$, we denote $\log_+(x) := (\log(x))_+$. For $x_1, x_2 \in \mathbb{R}$ we denote $x_1 \vee x_2 := \max\{x_1, x_2\}$ and $x_1 \wedge x_2 = \min\{x_1, x_2\}$. We use e_j to denote the standard unit vector, all of whose entries are 0 except for 1 at the jth entry. For two random variables X_i, $i = 1, 2$, we write $X_1 \stackrel{\mathcal{L}}{=} X_2$ to denote that X_1 and X_2 have the same law. For any matrix M, denote by $\|M\|_2 := [\mathrm{Tr}(MM^*)]^{1/2}$ its Hilbert–Schmidt norm, and for any vector v denote by $\|v\|_2$ its Euclidean norm.

2. The deterministic equivalent

Let $\Sigma_N(M_N(z))$ be the diagonal matrix of singular values of $M_N(z) := M_N - z\,\mathrm{Id}$. Suppose the entries of $\Sigma_N(M_N(z))$ are arranged to be nondecreasing

https://doi.org/10.1017/fms.2018.29 Published online by Cambridge University Press

going down the diagonal. That is, $\Sigma_{ii}(z) = \sigma_{N-i+1}(M_N(z))$ for $i \in [N] := \{1, 2, \ldots, N\}$, where $\Sigma_{ii}(z)$ is the ith diagonal entry of $\Sigma_N(M_N(z))$, and $\sigma_i(M_N(z))$ is the ith largest singular value of $M_N(z)$. By invariance of the Gaussian matrix,

$$\det(z\,\mathrm{Id} - \mathcal{M}_N) \stackrel{\mathcal{L}}{=} \det(\Sigma_N(M_N(z)) + N^{-\gamma} G_N).$$

Suppose that $\Sigma_N(M_N(z))$ is decomposed into two blocks of sizes $N_1 + N_2 = N$, so that

$$\Sigma_N(M_N(z)) := \begin{pmatrix} S_N(M_N(z)) & \\ & B_N(M_N(z)) \end{pmatrix} \quad \text{and} \quad N^{-\gamma} G_N := \begin{pmatrix} X_1 & X_2 \\ X_3 & X_4 \end{pmatrix}. \tag{2.1}$$

For ease of writing, when the matrix M_N is clear from the context, we simply write $\Sigma_N(z)$, $S_N(z)$, and $B_N(z)$ instead of $\Sigma_N(M_N(z))$, $S_N(M_N(z))$, and $B_N(M_N(z))$. Now by the Schur complement formula,

$$\det(\Sigma_N(z) + N^{-\gamma} G_N) \stackrel{\mathcal{L}}{=} \det(B_N(z) + X_4) \cdot \det(\widetilde{S}_N(z)), \tag{2.2}$$

where

$$\widetilde{S}_N(z) := S_N(z) + X_1 - X_2(B_N(z) + X_4)^{-1} X_3. \tag{2.3}$$

(Since the entries of X_4 are independent and identically distributed Gaussian, the matrix $(B_N(z) + X_4)$ is a.s. invertible and hence $\widetilde{S}_N(z)$ is well defined.) The decomposition in (2.2) proves useful when we choose the decomposition so that the entries of $B_N(z)$ are somewhat large with respect to the noise. For such a decomposition as we will see later (see Theorem 2.1) the log determinant of $B_N(z)$ correctly characterizes the log potential of the limiting spectral distribution of $\mathcal{M}_N$. So we need to define $B_N(z)$ appropriately.

Fix a sequence of $\{\varepsilon_N\}$ going down to zero. We define $N^* := N^*(z, \gamma, \varepsilon_N)$ as the largest integer i so that

$$\Sigma_{ii}(z) < \varepsilon_N^{-1} N^{-\gamma} (N - i + 1)^{1/2}. \tag{2.4}$$

If no such $1 \leqslant i \leqslant N$ exists then we let $N^* = 1$. Now set $N_1 = N^*$. This defines $B_N(z)$. With this choice of $B_N(z)$ we have the following result.

THEOREM 2.1. *Fix $z \in \mathbb{C}$. Suppose that $N^* \log N / N \to \alpha < \infty$. Then for any $\{\varepsilon_N\}$ that tends to* 0 *slowly enough that* $\log(\varepsilon_N^{-1})/\log N \to 0$,

$$\frac{1}{N} \log|\det(\Sigma_N(z) + N^{-\gamma} G_N)| - \frac{1}{N} \log|\det(B_N(z))| \to -\alpha\left(\gamma - \frac{1}{2}\right), \tag{2.5}$$

in probability, as $N \to \infty$. If $\alpha = 0$, we may take $\varepsilon_N = N^{-\eta}$ for any $\eta > 0$.

https://doi.org/10.1017/fms.2018.29 Published online by Cambridge University Press

The proof of Theorem 2.1 requires a two-fold argument. First we show that the truncation level chosen above assures that X_4 is negligible with respect to $B_N(z)$. In particular, we obtain the following result.

LEMMA 2.2. *For any given sequence of* $\{\varepsilon_N\}$, *such that* $\varepsilon_N \in (0, 1)$ *for all* N, *we have*

$$\mathbb{E}\det(B_N(z) + X_4) = \det(B_N(z)) \tag{2.6}$$

and

$$\operatorname{Var}\det(B_N(z) + X_4) \leqslant \frac{\varepsilon_N^2}{1-\varepsilon_N^2}(\det(B_N(z)))^2. \tag{2.7}$$

Proof. Since the entries of X_4 are independent with zero mean, (2.6) follows from Lemma A.1. To compute the second moment, we again use Lemma A.1 and the fact that entries of X_4 are independent with zero mean and variance $N^{-2\gamma}$ to obtain

$$\begin{aligned}\mathbb{E}|\det(B_N(z) + X^4)|^2 &= \sum_{k=0}^{N-N^*} \sum_{\substack{S\subset[N]\setminus[N^*]\\ |S|=N-N^*-k}} \mathbb{E}|\det X^4[\check{S}]|^2 \cdot \left(\prod_{i\in S}|\Sigma_{ii}(z)|^2\right)\\ &= \sum_{k=0}^{N-N^*}(k!)N^{-2\gamma k} \sum_{\substack{S\subset[N]\setminus[N^*]\\ |S|=N-N^*-k}} \prod_{i\in S}|\Sigma_{ii}(z)|^2,\end{aligned}$$

where $\check{S} := ([N]\setminus[N^*])\setminus S$. As the diagonal entries of $\Sigma_N(z)$ are arranged in nondecreasing order, recalling the definition of $B_N(z)$ and N^* we find that

$$\prod_{i\in S}|\Sigma_{ii}(z)|^2 \leqslant \frac{|\det B_N(z)|^2}{(\varepsilon_N^{-1}N^{-\gamma})^{2k}(N-N^*)^k},$$

for any $S \subset [N]\setminus[N^*]$ with $|S| = N - N^* - k$. Therefore,

$$\begin{aligned}\mathbb{E}|\det(B_N(z) + X^4)|^2 &\leqslant \sum_{k=0}^{N-N^*}\binom{N-N^*}{k}(k!)N^{-2\gamma k}\frac{|\det(B_N(z))|^2}{(\varepsilon^{-1}N^{-\gamma})^{2k}(N-N^*)^k}\\ &\leqslant \frac{1}{1-\varepsilon_N^2}|\det(B_N(z))|^2.\end{aligned}$$

The last display together with (2.6) yields (2.7). □

REMARK 2.3. Bounding higher (centered) moments of $\det(B_N(z) + X_4)$ and applying the Borel–Cantelli lemma one can strengthen the conclusion of

https://doi.org/10.1017/fms.2018.29 Published online by Cambridge University Press

Theorem 2.1 and show that (2.5) holds almost surely. This in turn shows that the conclusions of the main theorems of the paper, such as Theorems 1.2, 1.3, and 4.4 hold almost surely. We do not pursue this direction here.

Lemma 2.2 shows that if $\varepsilon_N \downarrow 0$ then the log determinant of $B_N(z) + X_4$ is asymptotically the same as that of $B_N(z)$. To establish Theorem 2.1 we also need to show that the log determinant of the Schur complement, $\log\det(\widetilde{S}_N(z))$, is asymptotically negligible, see (2.2). To this end, we obtain the following lower bound.

PROPOSITION 2.4. *Set $\widetilde{N} := N^* \vee \lceil \sqrt{N} \rceil$. If $N^* \leqslant N/2$, then there exist absolute constants $c_2, c_1 > 0$ so that*

$$\mathbb{P}[|\det(\widetilde{S}_N(z))| \leqslant N^{-\gamma N^*} \sqrt{(N^*!)} e^{-c_1 \widetilde{N}}] \leqslant e^{-c_2 \widetilde{N}}. \tag{2.8}$$

Before bringing the proof of Proposition 2.4, we recall the following lemma, which is proved in [6, Lemma 4.4] (in the real case, but the proof carries over to the complex case). An alternative proof can be given based on [19, Theorem 4].

LEMMA 2.5. *Suppose that E is a $Q \times Q$ standard Gaussian matrix. Then for any $Q \times Q$ matrix M independent of E and all $t \geqslant 0$,*

$$\mathbb{P}[|\det(E + M)| \leqslant t] \leqslant \mathbb{P}[|\det E| \leqslant t].$$

We will also need the following lemma, whose proof is an adaptation of the proof of [6, Lemma 2.5].

LEMMA 2.6. *Let E be an $N_0 \times N_0$ matrix of independent complex standard Gaussians. There are absolute constants $c_1' > 0$ and $c_2' > 0$ so that for all $N_0' \geqslant N_0$,*

$$\mathbb{P}[|\det E| \leqslant \sqrt{N_0!} e^{-c_1' N_0'}] \leqslant \frac{1}{c_2'} e^{-c_2' N_0'}.$$

Proof. We begin by recalling from [7] that if E is a complex Ginibre matrix of dimension $N \times N$ matrix, then

$$|\det E|^2 \stackrel{\mathcal{L}}{=} 2^{-N} \prod_{r=1}^{N} \chi_{2r}^2,$$

where χ_r^2 are independent chi-square random variables with r degrees of freedom, that is, they have the distribution of the square of the length of an r-dimensional standard real Gaussian vector. Now fix a large integer K and denote

$$F_K := \prod_{r=1}^{K} \chi_{2r}.$$

Then, $F_K \leqslant 2^{-N_0'}$ with exponentially (in N_0') small probability. Then, proceeding as in the proof of [6, Lemma 2.5], we find that

$$\mathbb{E}L_K^{-2} \leqslant \frac{K!}{N_0!}$$

for a sufficiently large K, where

$$L_K := \prod_{r=K+1}^{N_0} \chi_{2r}.$$

The rest follows from Markov's inequality. □

Proof of Proposition 2.4. By Lemma 2.5, for all $t \geqslant 0$

$$\mathbb{P}[|\det(\widetilde{S}_N(z))| \leqslant t] \leqslant \mathbb{P}[|\det(X_1)| \leqslant t],$$

where X_1 is an $N^* \times N^*$ matrix of independent and identically distributed complex Gaussians of variance $N^{-2\gamma}$. Hence, the desired conclusion follows from Lemma 2.6. □

We turn to finding an upper bound on the determinant of $\widetilde{S}_N(z)$. To this end, we first derive an upper bound on the norm of the inverse of $B_N(z) + X_4$.

LEMMA 2.7. *Fix $\{\varepsilon_N\}$ such that $\varepsilon_N < 1/8$ for all N, and assume that $N^* \leqslant N/2$. Then,*

$$\mathbb{P}(\|(B_N(z) + X_4)^{-1}\| \leqslant 2\varepsilon_N N^\gamma (N - N^*)^{-1/2}) \geqslant 1 - \exp(-cN),$$

for some absolute constant c.

Proof. Gordon's theorem for Gaussian matrices (see [8, Corollary 1.2]) and the triangle inequality give

$$\mathbb{E}\|X_4\| \leqslant 2\sqrt{2}N^{-\gamma}\sqrt{N - N^*}.$$

Since $A \mapsto \|A\|$ is a 1-Lipschitz function, the standard concentration inequality for Gaussian random variables applies and yields

$$\mathbb{P}(\|X_4\| \geqslant 4N^{-\gamma}\sqrt{N - N^*}) \leqslant \exp(-2c(N - N^*)) \leqslant \exp(-cN),$$

https://doi.org/10.1017/fms.2018.29 Published online by Cambridge University Press

for some absolute constant c. On the other hand, by our definition, see (2.4),

$$\sigma_{\min}(B_N(z)) \geqslant \varepsilon_N^{-1} N^{-\gamma} \sqrt{N - N^*},$$

where $\sigma_{\min}(B)$ denotes the minimum singular value of B. Since $\varepsilon_N < 1/8$, Weyl's inequality (see [12, Theorem 3.3.16(c)]) gives that

$$\sigma_{\min}(B_N(z) + X^4) \geqslant (2\varepsilon_N)^{-1} N^{-\gamma} \sqrt{N - N^*},$$

with probability at least $1 - \exp(-cN)$. This completes the proof. □

Building on Lemma 2.7 and using a standard concentration inequality we have the following result.

LEMMA 2.8. *Fix $\{\varepsilon_N\}$ such that $\varepsilon_N < 1/8$ for all N, and assume that $N^* \leqslant N/2$. Then there exist absolute constants c' and C' such that the ℓ_2-norm of each of the rows of $X_2(B_N(z) + X_4)^{-1} X_3$ is at most*

$$C' \varepsilon_N N^{-\gamma+1/2},$$

with probability at least $1 - \exp(-c'N)$.

Proof. By the rotation invariance of X_2 and X_3,

$$X_2(B_N(z) + X_4)^{-1} X_3 \stackrel{\mathcal{L}}{=} X_2 D X_3,$$

where D is a diagonal matrix with entries equal to the singular values of $(B_N(z) + X_4)^{-1}$. Hence, a row of $X_2(B_N(z) + X_4)^{-1} X_3$ is equal in distribution to

$$N^{-\gamma} (D\boldsymbol{x})^t X_3,$$

where $\boldsymbol{x}$ is an $(N - N^*)$-dimensional standard complex Gaussian vector independent of X_3 and for any vector $\boldsymbol{y}$ the notation $\boldsymbol{y}^t$ denotes its transpose. From the rotation invariance of X_3, it follows that

$$(D\boldsymbol{x})^t X_3 \stackrel{\mathcal{L}}{=} \|D\boldsymbol{x}\|_2 e_1^t X_3.$$

The law of $\|D\boldsymbol{x}\|_2$ is stochastically dominated by the law of $\|D\| \|\boldsymbol{x}\|_2$, and so we conclude that the law of the ℓ_2-norm of a row of $X_2(B_N(z) + X_4)^{-1} X_3$ is stochastically dominated by $\frac{1}{4} \cdot N^{-2\gamma} \chi_{2N^*} \cdot \chi_{2(N-N)*} \cdot \|(B_N(z) + X_4)^{-1}\|$, where again $\{\chi_r\}$ are random variables distributed as the length of an r-dimensional standard real Gaussian vector. Applying Lemma 2.7 and standard tail bounds for χ–variables, the result follows. □

https://doi.org/10.1017/fms.2018.29 Published online by Cambridge University Press

Using Lemma 2.8 we now find an upper bound on the determinant of the Schur complement.

LEMMA 2.9. *Fix $\{\varepsilon_N\}$ such that $\varepsilon_N \to 0$ and $\varepsilon_N < 1/8$ for all N, and assume that $N^* \leqslant N/2$. Then there are absolute constants $\bar{c}$ and $\bar{C}$ such that*

$$|\det(\widetilde{S}_N(z))| < \bar{C}^{N^*} N^{-\gamma N^*} \varepsilon_N^{-N^*} (N)^{N^*/2}$$

with probability at least $1 - \exp(-\bar{c}N)$.

Proof. Note that the rows of X_1 have ℓ_2 norm at most $2N^{1/2-\gamma}$ with probability at least $1 - \exp(-\bar{c}_1 N)$ where $\bar{c}_1$ is some absolute constant. By construction of the diagonal matrix $S_N(z)$, its entries (and hence the ℓ_2 norm of its rows) are bounded by $\varepsilon_N^{-1} N^{-\gamma+1/2}$. It follows from these facts and Lemma 2.8 that the ℓ_2 norms of the rows of $\widetilde{S}_N(z)$ are bounded by a constant multiple of $\varepsilon_N^{-1} N^{1/2-\gamma}$, with probability at least $1 - \exp(-\bar{c}N)$. Hadamard's bound on the determinant yields the desired conclusion. □

Equipped with all the ingredients we are now ready to prove Theorem 2.1.

Proof of Theorem 2.1. From Lemma 2.2, upon using Markov's inequality we find

$$\mathbb{P}\left(\left| \frac{\det(B_N(z) + X_4)}{\det(B_N(z))} - 1 \right| \geqslant \frac{1}{2} \right) \leqslant \frac{4\varepsilon_N^2}{1 - \varepsilon_N^2}.$$

Therefore we see that there is a set $\Omega_{N,1}$ such that on $\Omega_{N,1}$ we have

$$\left| \frac{1}{N} \log\det(B_N(z) + X_4) - \frac{1}{N} \log\det B_N(z) \right| \leqslant \frac{2}{N} \left| \frac{\det(B_N(z) + X_4)}{\det(B_N(z))} - 1 \right| \leqslant \frac{1}{N}, \tag{2.9}$$

and

$$\mathbb{P}(\Omega_{N,1}^c) \leqslant \frac{4\varepsilon_N^2}{1 - \varepsilon_N^2}. \tag{2.10}$$

Since $\gamma > 1/2$, from Proposition 2.4 and Lemma 2.9 we also obtain that

$$\begin{aligned} &\left| \frac{1}{N} \log|\det(\widetilde{S}_N(z))| - \frac{N^* \log N}{N} \left(-\gamma + \frac{1}{2} \right) \right| \\ &\leqslant \frac{N^*}{N} \left[\log\bar{C} + \log\left(\frac{1}{\varepsilon_N} \right) + \frac{1}{2} \log\frac{N}{N^*} \right] + (c_1 + 1) \frac{\widetilde{N}}{N} \\ &\leqslant \frac{N^* \log N}{N} \left[\frac{\log\bar{C} + \log(1/\varepsilon_N) + 1/2 \log(N/N^*)}{\log N} + \gamma \right] + (c_1 + 1) \frac{\widetilde{N}}{N} \end{aligned} \tag{2.11}$$

https://doi.org/10.1017/fms.2018.29 Published online by Cambridge University Press

on an event $\Omega_{N,2}$ such that

$$\mathbb{P}(\Omega_{N,2}^c) \leqslant \exp(-\bar{c}N) + \exp(-c_2\widetilde{N}). \tag{2.12}$$

Hence when $N^* \log N/N \to \alpha < \infty$, then taking $|\log(\varepsilon_N)| = o(\log N)$, the desired conclusion holds. With $\varepsilon_N = N^{-\eta}$ and $N^* = o(N/\log N)$ we deduce from (2.11) that

$$\left|\frac{1}{N} \log \det(\widetilde{S}_N(z))\right| = o(1), \tag{2.13}$$

on the event $\Omega_{N,2}$. Now combining (2.9)–(2.10) and (2.12)–(2.13) we see that the convergence in (2.5) holds in probability. □

3. Bidiagonal matrices: rigidity theorems for small singular vectors of $\boldsymbol{D_N + J}$

In this section, we develop estimates for the small singular values of the bidiagonal matrix $M_N = D_N + J$ where $D_N = -\text{diag}(d_1, d_2, \ldots, d_N)$ is a diagonal matrix. These are then used to prove Theorem 1.3. For all $i \leqslant j$, let $\mathcal{D}^{i,j}$ be defined by

$$\mathcal{D}^{i,j} := \prod_{i \leqslant \ell < j} d_\ell.$$

Note that $\mathcal{D}^{i,i} = 1$ by this definition. For all $i \leqslant j$ define a vector by, for each k,

$$\boldsymbol{v}_k^{i,j} := \begin{cases} 0 & k < i, \\ \mathcal{D}^{i,k} & i \leqslant k \leqslant j, \\ 0 & j < k. \end{cases} \tag{3.1}$$

The point of these vectors is that they solve $(M_N \boldsymbol{v}^{i,j})_k = 0$ for k between the boundaries. Precisely:

$$(M_N \boldsymbol{v}^{i,j})_k = \begin{cases} 1 & k = i - 1, \\ -\mathcal{D}^{i,j+1} & k = j, \\ 0 & \text{otherwise}. \end{cases} \tag{3.2}$$

Hence if $\mathcal{D}^{i,j+1}$ is much smaller than some $\mathcal{D}^{i,k} \gg 1$ for $i < k \leqslant j$, then this will be nearly a small singular vector, that is, a singular vector corresponding to a small singular value.

Note that for $i \leqslant k \leqslant j - 1$, we have the identity

$$\boldsymbol{v}_{k+1}^{i,j} = d_k \boldsymbol{v}_k^{i,j}. \tag{3.3}$$

https://doi.org/10.1017/fms.2018.29 Published online by Cambridge University Press

Using the vectors $\boldsymbol{v}^{i,j}$ we now construct approximate small singular vectors. Fix an integer $L := L_N$, the choice of which will be determined later. Consider integers

$$0 = i_1 < i_2 < i_3 < \cdots < i_L < i_{L+1} = N.$$

The vectors $\{\boldsymbol{w}^j = \boldsymbol{v}^{i_j+1,i_{j+1}}\|\boldsymbol{v}^{i_j+1,i_{j+1}}\|_2^{-1},\ j = 1, 2, \ldots, L\}$ have disjoint supports and are therefore orthogonal. We use these vectors as approximations for the small singular vectors and quantify the approximation. To this end, define

$$\begin{aligned} \mathcal{D}_+^{i_j,i_{j+1}} &:= \max_{i_j < s \leqslant r \leqslant i_{j+1}} \sum_{p=s}^{r} \{|\mathcal{D}^{p,r}| + |\mathcal{D}^{s,p}|\}, \quad \text{and} \\ \mathcal{D}_-^{i_j,i_{j+1}} &:= \max_{i_j < s \leqslant r \leqslant i_{j+1}} \sum_{p=s}^{r} \left\{ \frac{1}{|\mathcal{D}^{p,r}|} + \frac{1}{|\mathcal{D}^{s,p}|} \right\}. \end{aligned} \tag{3.4}$$

Provided the entries $\{|d_k| : i_j \leqslant k \leqslant i_{j+1}\}$ are consistently larger than 1 or consistently smaller than 1, at least one of these quantities will be close to 1. Even in the case of independent d_is, in which there may exist a relatively long string of diagonal entries with possibly atypical magnitude, it is unlikely that both of these will be large. We then let

$$\mathfrak{D} := \max_{1 \leqslant j \leqslant L} [\min\{\mathcal{D}_+^{i_j,i_{j+1}}, \mathcal{D}_-^{i_j,i_{j+1}}\}] \geqslant 1. \tag{3.5}$$

When $\mathfrak{D}$ is small, the approximation will be good. Let π_j denote the coordinate projection from $\mathbb{C}^N$ to the coordinates that support $\boldsymbol{w}^j$. Our main result in this section on singular values of M_N is the following.

THEOREM 3.1. *The $(L + 1)$-st smallest singular value satisfies*

$$\sigma_{N-L}(M_N) \geqslant \mathfrak{D}^{-1}.$$

There is an absolute constant $C > 1$ so that the product of the L smallest singular values of M_N satisfies

$$(C\|M_N\|\mathfrak{D}\sqrt{L})^{-L} \prod_{k=1}^{L} \|\pi_k M_N \boldsymbol{w}^k\|_2 \leqslant \prod_{k=0}^{L-1} \sigma_{N-k}(M_N) \leqslant \prod_{k=1}^{L} \|\pi_k M_N \boldsymbol{w}^k\|_2. \tag{3.6}$$

The proof of Theorem 3.1 is deferred to Section 3.2.

https://doi.org/10.1017/fms.2018.29 Published online by Cambridge University Press

3.1. Applications. Theorem 1.3 regarding the behavior of the eigenvalues of $\mathcal{M}_N$ in the bidiagonal case follows from direct applications of Theorems 3.1 and 2.1. We show now how these applications follow.

We begin with a particularly simple case to consider that is not directly related to Theorem 1.3.

COROLLARY 3.2. *Consider a Jordan block*

$$J_N(z) = \begin{bmatrix} z & 1 & & & \\ & z & 1 & & \\ & & \ddots & \ddots & \\ & & & z & 1 \\ & & & & z \end{bmatrix}.$$

Setting $\mathfrak{F}^{-1} = \min\{|1-|z||, N^{-1}, |1-|z|^{-1}|\}/2$, *there is a constant* $C(z) > 1$, *depending only on* z, *so that for all* N, *we have*

$$\sigma_{N-1}(J_N(z)) \geqslant \mathfrak{F}^{-1}, \quad \sigma_N(J_N(z)) \geqslant \mathfrak{F}^{-1}(|z| \wedge 1)^N / C(z),$$

and

$$\sigma_N(J_N(z)) \leqslant C(z)(|z| \wedge 1)^N.$$

Proof. We apply Theorem 3.1 with a single block, that is, $L = 1, i_1 = 0$, and $i_2 = N$. We bound $\mathfrak{D} \leqslant \mathfrak{F}$ upon observing that $|\mathcal{D}^{p,r}| = |z|^{r-p}$ for $r \geqslant p$. By the Gershgorin circle theorem we also have that $\|J_N(z)\| = \|J_N J_N^*\|^{1/2} = O(|z| \vee 1)$. Therefore, the proof now finishes upon using (3.1)–(3.2). □

REMARK 3.3. Observe that the same conclusions as in Corollary 3.2 hold if the diagonal of J_N were replaced by arbitrary complex numbers having the same modulus.

REMARK 3.4. By combining Corollary 3.2 with Theorem 2.1, we get a new proof of [**6**, Theorem 1.4].

When D_N is a diagonal matrix of independent and identically distributed random variables the outcome is similar.

COROLLARY 3.5. *Suppose the entries of* D_N *are independent and identically distributed complex random variables with*

$$\mathbb{E}|d_1|^{\pm\beta_0} < \infty \quad \text{for some } \beta_0 > 0 \textit{ and } \mathbb{E}\log|d_1| \neq 0. \tag{3.7}$$

https://doi.org/10.1017/fms.2018.29 Published online by Cambridge University Press

Then, for every $\varepsilon > 0$*, there is a* $\delta > 0$ *and an* $L \leqslant 5N^{1-\delta}$ *so that with probability approaching* 1 *as* $N \to \infty$*,* $\sigma_{N-L}(M_N) \geqslant N^{-\varepsilon}$ *and*

$$\prod_{k=0}^{L-1} \sigma_{N-k}(M_N) = e^{-N(\mathbb{E}\log|d_1|)_- + o(N)}.$$

Proof. The key to the proof is to construct a partition of $[N]$ so that $\mathfrak{D}$ is small and the upper and the lower bounds of (3.6) are evaluated easily. To this end, we first note that

$$\lim_{\beta\to 0} (\mathbb{E}|d_1|^{\pm\beta})^{1/\beta} = e^{\pm\mathbb{E}\log|d_1|}, \tag{3.8}$$

which follows from Taylor's theorem and dominated convergence using (3.7). Now we focus on the case $\mathbb{E}\log|d_1| > 0$. By (3.8) there exists a $\beta > 0$ so that $p_\beta := \mathbb{E}|d_1|^{-\beta} < 1$. We fix this β for the remainder of the proof. Next we recall that if $\{Z_k\}$ is a non-negative martingale with $Z_0 = 1$ then Doob's maximal inequality implies that

$$\mathbb{P}\left[\sup_{1\leqslant k} Z_k \geqslant t\right] \leqslant \frac{1}{t}. \tag{3.9}$$

Let G_δ be the set of $j \in [N]$ so that there exists a $k \in [N]$ for which

$$\prod_{i=j\wedge k}^{j\vee k} (|d_i|^{-\beta} p_\beta^{-1}) \geqslant N^{2\delta}.$$

Then

$$\begin{aligned}
\mathbb{P}(j \in G_\delta) &\leqslant \mathbb{P}\left(\sup_{k\geqslant 0} \prod_{i=j}^{j+k} (|d_i|^{-\beta} p_\beta^{-1}) \geqslant N^{2\delta}\right) \\
&\quad + \mathbb{P}\left(\sup_{2\leqslant k\leqslant j} \prod_{i=k-1}^{j} (|d_i|^{-\beta} p_\beta^{-1}) \geqslant N^{2\delta}\right) \\
&=: P_1 + P_2.
\end{aligned}$$

Setting $Z_k := \prod_{i=j}^{j+k-1} (|d_i|^{-\beta} p_\beta^{-1})$, one sees that $\{Z_k\}$ is a non-negative martingale with $Z_0 := 1$. Applying (3.9) one gets that $P_1 \leqslant N^{-2\delta}$. Similarly, setting $Z_k = \prod_{i=j-k+1}^{j} (|d_i|^{-\beta} p_\beta^{-1})$, one gets $P_2 \leqslant N^{-2\delta}$. Thus, $\mathbb{E}|G_\delta| \leqslant 2N^{1-2\delta}$, and hence by Markov's inequality, with probability at least $1 - N^{-\delta}$, $|G_\delta| \leqslant 2N^{1-\delta}$. Let $S \subset \mathbb{Z}$ be defined by

$$S := \{\lfloor N^\delta k\rfloor; k \in [\lfloor N^{1-\delta}\rfloor]\} \cup \left\{\bigcup_{x\in G_\delta} (x + \{0, 1\})\right\}, \tag{3.10}$$

https://doi.org/10.1017/fms.2018.29 Published online by Cambridge University Press

where for two sets S_1 and S_2 we denote $S_1 + S_2 := \{s_1 + s_2 : s_1 \in S_1, s_2 \in S_2\}$. Enumerate the elements of S as $\{i_2 < i_3 < i_4 < \cdots < i_{|S|+1}\}$. Extend the collection i's by letting $i_1 = 0$, $L = |S|+1$, and $i_{L+1} = N$. Then $L \leqslant 5N^{1-\delta}$ with probability at least $1 - N^{-\delta}$, and the separation $i_{k+1} - i_k \leqslant N^{\delta}$ for all k.

From the definition of the set S, it follows that when $i_j \notin G_\delta$, we have for any $i_j \leqslant r < i_{j+1}$ and any $k \geqslant r$,

$$\prod_{i=r}^{k} |d_i|^{-1} \leqslant N^{2\delta/\beta} p_\beta^{k/\beta} \leqslant N^{2\delta/\beta},$$

where the last inequality follows from the fact that $p_\beta < 1$. Hence, $\mathcal{D}_-^{i_j, i_{j+1}} \leqslant 2N^{\delta(1+2/\beta)}$. Recalling the definition of S once more we see that when $i_j \in G_\delta$, we have $i_{j+1} = i_j + 1$. Thus, in that case $\mathcal{D}_-^{i_j, j_{j+1}} \leqslant 2$. Therefore, we conclude that if $\mathbb{E}\log|d_1| > 0$ then $\mathfrak{D} \leqslant 2N^{\delta(1+2/\beta)}$, with probability at least $1 - N^{-\delta}$. Using (3.8), a similar argument yields that

$$\mathfrak{D}_+^{i_j, i_{j+1}} \leqslant 2N^{\delta(1+2/\beta)} \quad \text{for all } j = 1, 2, \ldots, L, \tag{3.11}$$

with probability at least $1 - N^{-\delta}$, when $\mathbb{E}\log|d_1| < 0$. Therefore, the same bound as above holds for $\mathfrak{D}$ in this case.

Now taking $\delta = \delta(\varepsilon)$ small enough and invoking the first part of Theorem 3.1, one concludes that with the same probability, $\sigma_{N-L}(M_N) \geqslant N^{-\varepsilon}$, as claimed.

To show the second part of the corollary, we claim that for any $\eta > 0$ there is a $\delta' := \delta'(\eta) > 0$, with $\lim_{\eta\to 0} \delta'(\eta) = 0$, such that for all $t > 0$

$$\max\left\{\mathbb{P}\left[\sum_{i=1}^{k} \log|d_i| \leqslant -t + k(\mathbb{E}\log|d_1| - \eta)\right],\right.$$
$$\left.\mathbb{P}\left[\sum_{i=1}^{k} \log|d_i| \geqslant t + k(\mathbb{E}\log|d_1| + \eta)\right]\right\} \leqslant e^{-\delta' t}. \tag{3.12}$$

To see the above, applying Markov's inequality, we note that for any $\beta' \in (0, \beta_0]$,

$$\mathbb{P}\left[\sum_{i=1}^{k} \log|d_i| \geqslant t + k(\mathbb{E}\log|d_1| + \eta)\right] \leqslant \frac{\mathbb{E}\left[\prod_{i=1}^{k} |d_i|^{\beta'}\right]}{\exp(\beta' k \mathbb{E}\log|d_1|)} \cdot e^{-\beta' t - \beta' k\eta}.$$

Now using (3.8) and choosing β' sufficiently small (depending on η) we deduce

$$\mathbb{P}\left[\sum_{i=1}^{k} \log|d_i| \geqslant t + k(\mathbb{E}\log|d_1| + \eta)\right] \leqslant e^{-\delta' t}.$$

https://doi.org/10.1017/fms.2018.29 Published online by Cambridge University Press

Using a similar argument for the second probability in the left hand side of (3.12), we conclude that (3.12) holds.

We continue with the proof of the second part of the corollary. From (3.12) we conclude that for any $\eta > 0$, there is a constant $\bar{C}(\eta) > 0$, with $\lim_{\eta\to 0} \bar{C}(\eta)^{-1} = 0$, so that, with probability $1 - N^{-100}$, for all $1 \leqslant j \leqslant L$,

$$e^{2(i_{j+1}-i_j)(\mathbb{E}\log|d_1|-\eta)-\bar{C}(\eta)\log N} \leqslant \prod_{i=i_j+1}^{i_{j+1}} |d_i|^2 = \|\pi_j M_N \boldsymbol{v}^{i_j+1,i_{j+1}}\|_2^2 \leqslant e^{2(i_{j+1}-i_j)(\mathbb{E}\log|d_1|+\eta)+\bar{C}(\eta)\log N}. \tag{3.13}$$

In the case that $\mathbb{E}\log|d_1| > 0$, using the fact that $i_{j+1} - i_j \leqslant N^\delta$ we see that

$$\prod_{i=i_j+1}^{i_{j+1}} |d_i|^2 \leqslant \|\boldsymbol{v}^{i_j+1,i_{j+1}}\|_2^2 \leqslant N^\delta \max_{i_j<r\leqslant i_{j+1}} \prod_{i>i_j+1}^{r} |d_i|^2.$$

Proceeding similarly to the proof of (3.13) and applying a union bound we conclude that with probability at least $1 - N^{-99}$, for all $1 \leqslant j \leqslant L$

$$e^{2(i_{j+1}-i_j)(\mathbb{E}\log|d_1|-\eta)-\bar{C}(\eta)\log N} \leqslant \|\boldsymbol{v}^{i_j+1,i_{j+1}}\|_2^2 \leqslant e^{2(i_{j+1}-i_j)(\mathbb{E}\log|d_1|+\eta)+(\bar{C}(\eta)+\delta)\log N}. \tag{3.14}$$

Recall that $\|\pi_k M_N \boldsymbol{w}^k\|_2 = \|\pi_k M_n \boldsymbol{v}^{i_k+1,i_{k+1}}\|_2 / \|\boldsymbol{v}^{i_k+1,i_{k+1}}\|_2$. Hence combining (3.13)–(3.14), using the fact $i_{j+1} - i_j \leqslant N^\delta$ again, and taking $\eta \to 0$ sufficiently slowly with N, so that $\bar{C}(\eta)\log N = o(N^\delta)$, we conclude that if $\mathbb{E}\log|d_1| > 0$ then

$$e^{-o(N)} \leqslant \prod_{k=1}^{L} \|\pi_k M_N \boldsymbol{w}^k\|_2 \leqslant e^{o(N)}, \tag{3.15}$$

with probability approaching one as $N \to \infty$.

Turning to the case $\mathbb{E}\log|d_1| < 0$ we begin with the estimate

$$1 \leqslant \|\boldsymbol{v}^{i_j+1,i_{j+1}}\|_2^2 \leqslant N^\delta \mathfrak{D}_+^{i_j,i_{j+1}} \leqslant 2N^{2\delta(1+1/\beta)} \quad \text{for all } j = 1, 2, \ldots, L,$$

with probability at least $1 - N^{-\delta}$, where the last step follows from (3.11). Therefore

$$e^{\sum_{i=1}^N \log|d_i| - o(N)} \leqslant \prod_{k=1}^{L} \|\pi_k M_N \boldsymbol{w}^k\|_2 = \frac{\prod_{i=1}^N |d_i|}{\prod_{j=1}^L \|\boldsymbol{v}^{i_j+1,i_{j+1}}\|_2} \leqslant e^{\sum_{i=1}^N \log|d_i|}, \tag{3.16}$$

with probability approaching one, as $N \to \infty$.

To complete the proof using Markov's inequality and the union bound we note that $\mathbb{P}(\max_i |d_i| \geqslant N^{2/\beta_0}) = O(1/N)$. Therefore, using the Gershgorin circle

https://doi.org/10.1017/fms.2018.29 Published online by Cambridge University Press

theorem we derive that $\|M_N\|^2 = \|M_N^* M_N\| = O(N^{4/\beta_0})$ with probability tending to one. Now applying Theorem 3.1, the derived bound on $\mathfrak{D}$, using (3.15) when $\mathbb{E}\log|d_1| > 0$, and (3.16) and Chebycheff's inequality when $\mathbb{E}\log|d_1| < 0$, the corollary follows. □

The next corollary and Remark 3.7 following it, combine Corollary 3.5 with Theorem 2.1 to obtain Theorem 1.3(b).

COROLLARY 3.6. *Suppose that D_N is a diagonal matrix of independent and identically distributed complex random variables of law ν, and let $M_N = D_N + J$. Suppose that $\mathbb{E}|d_1|^{\pm\beta_0} < \infty$ for some $\beta_0 > 0$ and that $\mathbb{E}\log|d_1 - z| \neq 0$, for Lebesgue a.e. $z \in \mathbb{C}$. Then L_N converges weakly in probability to the probability measure with log potential $(\mathbb{E}\log|d_1 - z|)_+$.*

REMARK 3.7. Let $\mathscr{S} := \mathrm{Supp}\,(d_1)$. The condition that $\mathcal{L}_\nu(z) = \mathbb{E}\log|d_1 - z| \neq 0$, for Lebesgue a.e. z is satisfied, for example, when $\mathscr{S}$ is contained in a compact simply connected set of two-dimensional Lebesgue measure 0. Indeed, note that $\mathcal{L}_\nu$ is harmonic on $\mathscr{S}^c$. $\mathcal{L}_\nu$ cannot vanish identically in $\mathscr{S}^c$ because $\mathscr{S}^c$ is connected and $\limsup_{z\to\infty}\mathcal{L}_\nu(z) = \infty$. By [16, Theorem 3.1.18], it follows that the zero set of $\mathcal{L}_\nu$ in $\mathscr{S}^c$ has zero Lebesgue measure.

Proof of Corollary 3.6. We note first that by Weyl's inequalities for singular values, if $d_i^\downarrow$, $i = 1, \ldots, N$ denote the monotone decreasing reordering of the variables $|d_1|, \ldots, |d_N|$ then $|\sigma_i(\mathcal{M}_N) - d_i^\downarrow| \leqslant 2$ for all i, with probability approaching 1 as $N \to \infty$. (In the last statement, we used that $\|J\| \leqslant 1$ and that by (1.2), $\|N^{-\gamma}G_N\| \to_{N\to\infty} 0$.) By Weyl's majorant theorem [3, Theorem II.3.6] it follows that, with probability tending to 1 as $N \to \infty$,

$$\sum_{i=1}^N \log_+|\lambda_i(\mathcal{M}_N)| \leqslant \sum_{i=1}^N \log_+\sigma_i(\mathcal{M}_N) \leqslant \sum_{i=1}^N \log_+(|d_i| + 2).$$

Since $\mathbb{E}|d_1|^{\beta_0} < \infty$ it follows that $\mathbb{E}\log_+(|d_1| + 2) < \infty$. Therefore, denoting by $B_{\mathbb{C}}(0, R)$ the ball of radius R in the complex plane centered at zero and using the law of large numbers, we find that

$$\begin{aligned} L_N(B_{\mathbb{C}}(0, R)^c) = \frac{1}{N}\sum_{i=1}^N \mathbf{1}_{\{|\lambda_i(\mathcal{M}_N)|>R\}} &\leqslant (\log_+ R)^{-1}\cdot\frac{1}{N}\sum_{i=1}^N \log_+|\lambda_i(\mathcal{M}_N)| \\ &\leqslant 2\cdot\frac{\mathbb{E}\log_+(|d_1| + 2)}{\log_+ R}, \end{aligned}$$

with probability approaching one. Since $R < \infty$ is arbitrary, this in turn implies that the sequence of random probability measures $\{L_N\}_{N\in\mathbb{N}}$ is tight. Therefore, by [20, Theorem 2.8.3], the corollary follows once one proves that $\mathcal{L}_{\mathcal{M}_N}(z) \to \mathbb{E}\log|d_1 - z|_+$ in probability for Lebesgue almost every $z \in \mathbb{C}$.

To prove the convergence of $\mathcal{L}_{\mathcal{M}_N}$, we check that $D_N - z\,\mathrm{Id} + J$ satisfies the hypotheses of Corollary 3.5 for Lebesgue a.e. every z. By assumption $\mathbb{E}|d_1|^{\beta_0} < \infty$ and hence $\mathbb{E}|d_1 - z|^{\beta_0} < \infty$ for all $z \in \mathbb{C}$. Observe that for any $M > 0$

$$\int_{\mathbb{C}}\int_{|z-y|<M} \frac{1}{|z-y|}|dz|^2\nu(dy) = 2\pi M$$

with $|dz|^2$ the 2-dimensional Lebesgue volume element. In particularly, as this is locally integrable, by Jensen's inequality, we have that

$$\mathbb{E}|z - d_1|^{-\beta_0} \leqslant (\mathbb{E}|z - d_1|^{-1})^{\beta_0} < \infty$$

is finite for Lebesgue a.e. z, where without loss of generality we have assumed $\beta_0 \leqslant 1$. Hence $M_N - z\,\mathrm{Id}$ satisfies the hypotheses of Corollary 3.5 for Lebesgue a.e. $z \in \mathbb{C}$.

Let $z \in \mathbb{C}$ be a point at which these hypotheses are satisfied. Choosing $\varepsilon, \eta > 0$ sufficiently small so that $\varepsilon < \gamma - \frac{1}{2} - \eta$, we get from Corollary 3.5 that there is a $\delta > 0$ and an $L \leqslant 5N^{1-\delta}$ so that $\sigma_{N-L}(M_N - z\,\mathrm{Id}) \geqslant N^{-\varepsilon}$ and

$$\prod_{k=0}^{L-1} \sigma_{N-k}(M_N - z\,\mathrm{Id}) = e^{-N(\mathbb{E}\log|d_1 - z|)_- + o(N)}, \tag{3.17}$$

with probability approaching to one, as $N \to \infty$. Applying Theorem 2.1, with $\varepsilon_N = N^{-\eta}$, we get that

$$\frac{1}{N}\log|\det(\mathcal{M}_N - z\,\mathrm{Id})| - \frac{1}{N}\log\frac{|\det(M_N - z\,\mathrm{Id})|}{\prod_{k=0}^{N^*-1}\sigma_{N-k}(M_N - z\,\mathrm{Id})} \to 0 \tag{3.18}$$

in probability, where we recall that N^* was defined in (2.4) as the largest i so that

$$\sigma_{N-i} < N^{\eta-\gamma}(N-i)^{1/2} < N^{-\varepsilon}.$$

So $N^* \leqslant L$, and therefore

$$(N^{\eta-\gamma}(N - N^*)^{1/2})^L \leqslant \frac{\prod_{k=0}^{L-1}\sigma_{N-k}(M_N - z\,\mathrm{Id})}{\prod_{k=0}^{N^*-1}\sigma_{N-k}(M_N - z\,\mathrm{Id})} \leqslant \|M_N - z\,\mathrm{Id}\|^L. \tag{3.19}$$

As we have already seen during the proof of Corollary 3.5 that $\|M_N\| = O(N^{2/\beta_0})$ with probability approaching one, by the triangle inequality the same bound

https://doi.org/10.1017/fms.2018.29 Published online by Cambridge University Press

continues to hold for $\|M_N - z\,\mathrm{Id}\,\|$. Therefore, using the fact that $L = O(N^{1-\delta})$ we deduce that both the upper and lower bounds in (3.19) are subexponential in N. On the other hand, since M_N is an upper-triangular matrix, by Chebycheff's inequality it also follows that $|\det(M_N - z\,\mathrm{Id})| = e^{N\mathbb{E}\log|d_1-z|+o(N)}$, with probability tending to one. Hence, combining (3.17)–(3.18) we deduce

$$\frac{1}{N}\log|\det(\mathcal{M}_N - z\,\mathrm{Id})| - \frac{1}{N}\log\frac{e^{N\mathbb{E}\log|d_1-z|+o(N)}}{e^{-N(\mathbb{E}\log|d_1-z|)_-+o(N)}} \to 0$$

in probability. As this holds for Lebesgue almost every $z \in \mathbb{C}$, the claimed statement follows. □

The next corollary deals with $M_N = D_N + J$ where the entries in D_N vary slowly.

COROLLARY 3.8. *Suppose that $f : [0, 1] \to \mathbb{C}$ is α-Hölder continuous, and that $d_i = f(i/N)$ for $1 \leqslant i \leqslant N$. For every $\varepsilon > 0$, there is a $\delta > 0$ and an $L = O(N^{1-\delta})$ so that $\sigma_{N-L}(M_N) \geqslant N^{-\varepsilon}$ and*

$$\prod_{k=0}^{L-1}\sigma_{N-k}(M_N) = e^{\sum_{i=1}^{N} -(\log|f(i/N)|)_- + o(N)}.$$

Proof. The key to the proof is again to construct a suitable partition of $[N]$. Fix any $\delta \in (0, 1)$. Inductively define $a_j \in \mathbb{N}$ by setting $a_1 := 1$ and letting

$$a_{j+1} := \begin{cases} \inf\{a_j < k \leqslant N : \log|f(k/N)| > N^{-\delta}\} & \text{if } \log|f(a_j/N)| < 0, \\ \inf\{a_j < k \leqslant N : \log|f(k/N)| < -N^{-\delta}\} & \text{if } \log|f(a_j/N)| > 0. \end{cases}$$

From the definition of $\{a_j\}$ and the Hölder continuity of f we see that there is a constant C_0 so that

$$\frac{1}{2}N^{-\delta} \leqslant |f(a_j/N) - f(a_{j+1}/N)| \leqslant C_0\left(\frac{a_{j+1} - a_j}{N}\right)^{\alpha}$$

and note that $|\{a_j\}| = O(N^{\delta/\alpha})$. Let

$$S := \bigcup_{k=1}^{\lfloor N^{1-\delta}\rfloor} \lfloor N^{\delta}k\rfloor \cup \{a_j : j \in \mathbb{N}\}.$$

Enumerate the elements of S as $\{i_2 < i_3 < i_4 < \cdots < i_{|S|+1}\}$. Extend this by letting $i_1 = 0$, $L = |S| + 1$, and $i_{L+1} = N$. Then, upon reducing δ if necessary, we obtain $L = O(N^{1-\delta})$, and the separation $i_{k+1} - i_k < N^{\delta}$ for all k.

https://doi.org/10.1017/fms.2018.29 Published online by Cambridge University Press

Next, recalling the definition of $\mathfrak{D}^{s,r}$, by the construction of $\{a_j\}$, and the triangle inequality we also have that either

$$|\mathcal{D}^{s,r}| \leqslant 2e^{(r-s)N^{-\delta}} \quad \text{for all } i_k < s \leqslant r \leqslant i_{k+1} \text{ or } |\mathcal{D}^{s,r}| \geqslant \tfrac{1}{2}e^{-(r-s)N^{-\delta}},$$

for all $i_k < s \leqslant r \leqslant i_{k+1}$. Hence $\mathfrak{D} \leqslant 4eN^{\delta}$, which upon an application of Theorem 3.1 yields that given any $\varepsilon > 0$, there exists $\delta > 0$, sufficiently small such that $\sigma_{N-L}(M_N) \geqslant N^{-\varepsilon}$. As $\|M_N\| = \|M_N M_N^*\|^{1/2} = O(\sup_{x\in[0,1]} |f(x)|)$, by the Gershgorin circle theorem, applying Theorem 3.1 once again, it remains to show that

$$\prod_{k=1}^{L} \|\pi_k M_N \boldsymbol{w}^k\|_2 = e^{\sum_{i=1}^{N} -(\log|f(i/N)|)_- + o(N)} \tag{3.20}$$

where we recall that $\boldsymbol{w}^k = \boldsymbol{v}^{i_k+1,i_{k+1}}\|\boldsymbol{v}^{i_k+1,i_{k+1}}\|_2^{-1}$ for $1 \leqslant k \leqslant L$. To this end, we note that for each $1 \leqslant k \leqslant L$ we have that either

$$\log|f(s/N)| \leqslant N^{-\delta} \quad \text{for all } i_k < s < i_{k+1} \text{ or } \log|f(s/N)| \geqslant -N^{-\delta}$$
$$\text{for all } i_k < s < i_{k+1}.$$

Using the Hölder continuity of f we further have that $|f(i_{k+1}/N)| \leqslant 2e^{N^{-\delta}}$ in the first case and $|f(i_{k+1}/N)| \geqslant \frac{1}{2}e^{-N^{-\delta}}$ in the second case. Therefore, in the first case, we have that

$$1 \leqslant \|\boldsymbol{v}^{i_k+1,i_{k+1}}\|_2^2 \leqslant 2e^2N^{\delta},$$

and

$$\begin{aligned} |\mathcal{D}^{i_k+1,i_{k+1}+1}| &= \prod_{j=i_k+1}^{i_{k+1}} |f(j/N)| \\ &= \exp\left(\sum_{j=i_k+1}^{i_{k+1}} (\log|f(j/N)|)_+ - \sum_{j=i_k+1}^{i_{k+1}} (\log|f(j/N)|)_-\right) \\ &= \exp\left(O(1) - \sum_{j=i_k+1}^{i_{k+1}} (\log|f(j/N)|)_-\right). \end{aligned}$$

Hence,

$$\|\pi_k M_N \boldsymbol{w}^k\|_2 = \exp\left(O(\log N) - \sum_{j=i_k+1}^{i_{k+1}} (\log|f(j/N)|)_-\right). \tag{3.21}$$

Arguing similarly, in the second case, we have that

$$|\mathcal{D}^{i_k+1,i_{k+1}}|^2 \leqslant \|\boldsymbol{v}^{i_k+1,i_{k+1}}\|_2^2 \leqslant 2e^2N^{\delta}|\mathcal{D}^{i_k+1,i_{k+1}}|^2,$$

and therefore

$$\|\pi_k M_N \boldsymbol{w}^k\|_2 = \exp(O(\log N)) = \exp\left(O(\log N) - \sum_{j=i_k+1}^{i_{k+1}} (\log|f(j/N)|)_- \right). \tag{3.22}$$

Combining (3.21)–(3.22) we arrive at (3.20). This completes the proof. □

Building on Corollary 3.8, the next corollary is Theorem 1.3(b).

COROLLARY 3.9. *Suppose $f : [0,1] \to \mathbb{C}$ is Hölder continuous, and that $d_i = f(i/N)$ for $1 \leqslant i \leqslant N$. Set $M_N = D_N + J$. Then L_N converges weakly in probability to a probability measure with log potential*

$$\int_0^1 \big(\log|f(t) - z|\big)_+ \, dt.$$

The proof follows a very similar track as the derivation of Corollary 3.6 from Corollary 3.5. Moreover, as we will see later in Section 5, the proof of Theorem 4.4 also follows a very similar line of arguments. Hence, the proof of Corollary 3.9 is omitted.

3.2. Proof of Theorem 3.1: estimates for the small singular values of $D_N + J$.

The proof is divided into three separate claims: Corollary 3.13, which is a bound on the $(L+1)$-st smallest singular value, Proposition 3.11, which is an upper bound on the product of small singular values, and Proposition 3.14, which is a lower bound on the product of small singular values.

Recall the notation in (3.3) and (3.5), and that π_j is the coordinate projection from $\mathbb{C}^N$ to the coordinates that support $\boldsymbol{w}^j$. Let ρ_j be the same coordinate projection that in addition kills the i_{j+1} coordinate, that is, the last entry of the support of $\boldsymbol{w}^j$.

LEMMA 3.10. *For all $1 \leqslant j \leqslant L$, and any vector $x \in \mathbb{C}^N$,*

$$\frac{\inf_{a\in\mathbb{C}} \|\pi_j(x - a\boldsymbol{w}^j)\|_2}{\|\rho_j(M_N x)\|_2} \leqslant \min\{\mathcal{D}_+^{i_j,i_{j+1}}, \mathcal{D}_-^{i_j,i_{j+1}}\},$$

where $\mathcal{D}_+^{i_j,i_{j+1}}$ and $\mathcal{D}_-^{i_j,i_{j+1}}$ were defined in (3.4).

Proof of Lemma 3.10. By definition of M_N, we have for any $1 \leqslant p < N$ and any vector $x \in \mathbb{C}^N$,

$$x_{p+1} = d_p x_p - (M_N x)_p.$$

By iterating this identity and using (3.2), we have for $i_{j+1} \geqslant p > k \geqslant i_j + 1$

$$x_p - a\boldsymbol{w}_p^j = (x_k - a\boldsymbol{w}_k^j)\mathcal{D}^{k,p} + \sum_{r=k}^{p-1} (M_N x)_r \mathcal{D}^{r+1,p}.$$

Reversing the roles of k and p and rearranging the formula, this also shows that for $i_{j+1} \geqslant k > p \geqslant i_j + 1$

$$x_p - a\boldsymbol{w}_p^j = \frac{(x_k - a\boldsymbol{w}_k^j)}{\mathcal{D}^{p,k}} + \sum_{r=p}^{k-1} \frac{(M_N x)_r}{\mathcal{D}^{p,r+1}}.$$

Picking a so that $(x_{i_j+1} - a\boldsymbol{w}_{i_j+1}^j) = 0$ or $(x_{i_{j+1}} - a\boldsymbol{w}_{i_{j+1}}^j) = 0$ we have that

$$|x_p - a\boldsymbol{w}_p^j| \leqslant \begin{cases} \sum_{r=i_j+1}^{p-1} |(M_N x)_r| \cdot |\mathcal{D}^{r+1,p}|, & p > i_j + 1, \\ \sum_{r=p}^{i_{j+1}-1} |(M_N x)_r| \cdot |\mathcal{D}^{p,r+1}|^{-1}, & p < i_{j+1} - 1. \end{cases} \tag{3.23}$$

Hence, using the first inequality of (3.23), upon applying Jensen's inequality,

$$|x_p - a\boldsymbol{w}_p^j|^2 \leqslant \left\{ \sum_{r=i_j+1}^{p-1} |(M_N x)_r|^2 \cdot |\mathcal{D}^{r+1,p}| \right\} \cdot \left\{ \sum_{r=i_j+1}^{p-1} |\mathcal{D}^{r+1,p}| \right\}.$$

Summing this bound from $p = i_j + 1$ to $p = i_{j+1}$, rearranging the terms, and using the definition of $\mathcal{D}_+^{i_j,i_{j+1}}$,

$$\|\pi_j(x - a\boldsymbol{w}^j)\|_2^2 = \sum_{p=i_j+1}^{i_{j+1}} |x_p - a\boldsymbol{w}_p^j|^2 \leqslant \left\{ \sum_{r=i_j+1}^{i_{j+1}-1} |(M_N x)_r|^2 \right\} \cdot \{\mathcal{D}_+^{i_j,i_{j+1}}\}^2$$
$$= \|\rho_j(M_N x)\|_2^2 \cdot \{\mathcal{D}_+^{i_j,i_{j+1}}\}^2.$$

Next using the second inequality of (3.23) and proceeding similarly as above we complete the proof. □

We now proceed to using these estimates in order to control the product of the small singular values of M_N. We begin with obtaining an upper bound on the product of small singular values. To this end, we use Lemma A.2, which is a multivariate generalization of Courant–Fischer–Weyl min–max principle for singular values.

PROPOSITION 3.11.

$$\prod_{k=0}^{L-1} \sigma_{N-k}(M_N) \leqslant \prod_{k=1}^{L} \|\pi_k M_N \boldsymbol{w}^k\|_2. \tag{3.24}$$

https://doi.org/10.1017/fms.2018.29 Published online by Cambridge University Press

Proof. Denote $\boldsymbol{W} := [\boldsymbol{w}^1 \ \boldsymbol{w}^2 \ \cdots \ \boldsymbol{w}^L]$. Since the columns of $\boldsymbol{W}$ are orthonormal, from Lemma A.2 it follows that

$$\prod_{k=0}^{L-1} \sigma_{N-k}(M_N) \leqslant \det(\boldsymbol{W}^* M_N^* M_N \boldsymbol{W}). \tag{3.25}$$

To evaluate the RHS of (3.25) we define the $L \times L$ matrix $\mathcal{M}$ by

$$\mathcal{M}_{j,k} := (e_{i_{j+1}})^t M_N \boldsymbol{w}^k,$$

where $\{e_\ell\}_{\ell=1}^N$ are canonical basis vectors in $\mathbb{C}^N$. Recalling (3.2) we note that $\boldsymbol{W}^* M_N^* M_N \boldsymbol{W} = \mathcal{M}^* \mathcal{M}$. On the other hand, the matrix $\mathcal{M}$ being upper triangular we also have

$$\prod_{j=1}^{L} \|\pi_j M_N \boldsymbol{w}^j\|_2^2 = \prod_{j=1}^{L} |\mathcal{M}_{j,j}|^2 = |\det \mathcal{M}|^2 = \det(\boldsymbol{W}^* M_N^* M_N \boldsymbol{W}). \tag{3.26}$$

The proof concludes by invoking (3.25). □

We next show that vectors v which have a sizeable component orthogonal to $\mathcal{S} := \operatorname{span}\{\boldsymbol{w}^j\}$ will necessarily have $\|M_N v\|_2$ large. To this end, let ψ be the orthogonal projection map from $\mathbb{C}^N$ to $\mathcal{S}$. Let ρ be the projection

$$\rho := \sum_{j=1}^{\ell} \rho_j. \tag{3.27}$$

The projections ρ and ψ interact in that

$$\rho M_N \psi = 0, \tag{3.28}$$

which follows immediately from the definition of $\boldsymbol{w}^j$. We can then combine this observation with the earlier Lemma 3.10 to obtain the next lemma.

LEMMA 3.12. *For any vector $v \in \mathbb{C}^N$,*

$$\|v\|_2^2 - \|\psi v\|_2^2 = \|(1-\psi)v\|_2^2 \leqslant \|\rho M_N v\|_2^2 \mathfrak{D}^2.$$

Proof. On the one hand, using the orthogonality of π_j,

$$\begin{aligned} \|v\|_2^2 - \|\psi v\|_2^2 = \inf_{\{a_j\} \subset \mathbb{C}} \|(v - \sum_j a_j \boldsymbol{w}^j)\|_2^2 = \inf_{\{a_j\} \subset \mathbb{C}} \sum_j \|\pi_j (v - a_j \boldsymbol{w}^j)\|_2^2 \\ = \sum_j \inf_{a_j \in \mathbb{C}} \|\pi_j (v - a_j \boldsymbol{w}^j)\|_2^2. \end{aligned}$$

Hence using Lemma 3.10 and (3.5), we get

$$\|v\|_2^2 - \|\psi v\|_2^2 \leqslant \sum_j \|\rho_j M_N v\|_2^2 \mathfrak{D}^2.$$

The stated conclusion of the lemma follows by the orthogonality of ρ_j. □

The last lemma immediately implies a lower bound for the $(L+1)$-st smallest singular value.

COROLLARY 3.13.

$$\sigma_{N-L}(M_N) \geqslant \mathfrak{D}^{-1}$$

Proof. We recall the standard variational characterization of this singular value in the maximin form:

$$\sigma_{N-L}(M_N) = \sup_{V_L} \inf_{x \perp V_L} \|M_N x\|_2,$$

where V_L is an L-dimensional space and x is of unit ℓ_2-norm. Setting $V_L = \mathcal{S}$, the stated corollary immediately follows from Lemma 3.12. □

Now it remains to find a lower bound on the product of the small singular values; this is slightly more involved.

PROPOSITION 3.14. *With notation as above,*

$$\prod_{k=0}^{L-1} \sigma_{N-k}(M_N) \geqslant (8(\|M_N\| \vee 1)\mathfrak{D}\sqrt{L})^{-L} \prod_{k=1}^{L} \|\pi_k M_N \boldsymbol{w}^k\|_2.$$

Proof. Using Lemma A.2 we see that it is enough to find a uniform lower bound on $\prod_{k=1}^{L} \|M_N v_k\|_2$ over all collections of orthonormal vectors $\{v_k\}_{k=1}^{L}$. We bound each $\|M_N v_k\|_2$ below in one of two ways. If $1 - \|\psi v_k\|_2^2 \geqslant 1/2L$ then v_k has a large enough component in the $\mathcal{S}^\perp$ direction that we apply Lemma 3.12 to conclude

$$\|M_N v_k\|_2 \geqslant \|\rho M_N v_k\|_2 \geqslant \frac{1}{\sqrt{2L}\mathfrak{D}} \geqslant \frac{1}{8\mathfrak{D}\sqrt{L}} \frac{\|\pi_k M_N \boldsymbol{w}^k\|_2}{\|M_N\|}, \tag{3.29}$$

where ρ is as in (3.27). Without loss of generality, we may permute the ordering of the vectors so that the first $v_1, v_2, \ldots, v_p$ are those that satisfy $1 - \|\psi v_k\|_2^2 < 1/2L$. For these vectors, we have that

$$\|M_N v_k\|_2^2 = \|\rho M_N (1-\psi) v_k\|_2^2 + \|(1-\rho) M_N v_k\|_2^2, \tag{3.30}$$

https://doi.org/10.1017/fms.2018.29 Published online by Cambridge University Press

where we have used that $\rho M_N \psi = 0$. We now consider two cases. First suppose

$$\|(1-\rho)M_N\psi v_k\|_2 \leqslant 4\|(1-\rho)M_N(1-\psi)v_k\|_2.$$

Since

$$\|(1-\rho)M_N(1-\psi)v_k\|_2 \leqslant \|M_N\| \cdot \|(1-\psi)v_k\|_2,$$

we obtain

$$\|(1-\rho)M_N\psi v_k\|_2 \leqslant 4\|M_N\| \cdot \|(1-\psi)v_k\|_2.$$

Hence by (3.30) and Lemma 3.12 we deduce

$$\begin{aligned}\|M_N v_k\|_2 &\geqslant \|\rho M_N(1-\psi)v_k\|_2 = \|\rho M_N v_k\|_2 \geqslant \mathfrak{D}^{-1}\|(1-\psi)v_k\|_2 \\ &\geqslant \frac{1}{4\|M_N\|\mathfrak{D}}\|(1-\rho)M_N\psi v_k\|_2. \end{aligned} \tag{3.31}$$

On the other hand, if

$$\|(1-\rho)M_N\psi v_k\|_2 > 4\|(1-\rho)M_N(1-\psi)v_k\|_2,$$

then by (3.30) and the triangle inequality we see that

$$\begin{aligned}\|M_N v_k\|_2 &\geqslant \|(1-\rho)M_N v_k\|_2 \\ &\geqslant \|(1-\rho)M_N\psi v_k\|_2 - \|(1-\rho)M_N(1-\psi)v_k\|_2 \\ &\geqslant \tfrac{3}{4}\|(1-\rho)M_N\psi v_k\|_2. \end{aligned} \tag{3.32}$$

Hence, combining (3.31)–(3.32) and noting that $\mathfrak{D} \geqslant 1$ we conclude that in either case,

$$\|M_N v_k\|_2 \geqslant \frac{1}{4(\|M_N\| \vee 1)\mathfrak{D}}\|M_N\psi v_k\|_2,$$

where we have again applied the fact that $\rho M_N \psi = 0$. Thus

$$\prod_{k=1}^{p}\|M_N v_k\|_2 \geqslant \left(\frac{1}{4(\|M_N\| \vee 1)\mathfrak{D}}\right)^p \prod_{k=1}^{p}\|M_N\psi v_k\|_2. \tag{3.33}$$

The rest of the proof boils down to finding a lower bound on the RHS of (3.33). To this end, let Y_1 be the matrix whose columns are $\{\psi v_k\}_{m=1}^{p}$. Since the columns of $\boldsymbol{W} := [\boldsymbol{w}^1\ \boldsymbol{w}^2 \cdots \boldsymbol{w}^L]$ span the subspace $\mathcal{S}$, there must exist an $L \times p$ matrix A_1 such that $Y_1 = \boldsymbol{W}A_1$. We extend the matrix A_1 to an $L \times L$ matrix A so that the last $L-p$ columns of A are orthonormal and are also orthogonal to the first p columns of A_1. Such an extension is always possible by first extending arbitrarily to a basis and then running Gram–Schmidt on the final $L-p$ columns. Set $Y := \boldsymbol{W}A$ and denote the columns of Y by $\boldsymbol{y}_m$, for $m \in [L]$.

Turning to bound the RHS of (3.33), by Hadamard's inequality we now find that

$$\prod_{k=1}^{p} \|M_N \psi v_k\|_2^2 \geqslant \frac{\det(Y^* M_N^* M_N Y)}{\prod_{k=p+1}^{L} \|M_N \boldsymbol{y}_k\|_2^2}. \tag{3.34}$$

We separately bound the numerator and the denominator of (3.34).

Note that $\boldsymbol{y}_k = \sum_{m=1}^{L} a_{m,k} \boldsymbol{w}^m$ where $a_{m,k}$ is the (m,k)th entry of A. Since $\{\boldsymbol{w}^m\}_{m=1}^{L}$ are orthonormal, for $k = p+1, p+2, \ldots, L$, we have

$$\|M_N \boldsymbol{y}_k\|_2 \leqslant \|M_N\| \cdot \|\boldsymbol{y}_k\|_2 = \|M_N\| \cdot \sqrt{\sum_{m=1}^{L} |a_{m,k}|^2 \|\boldsymbol{w}^m\|_2^2} = \|M_N\|, \tag{3.35}$$

where in the last step we also use the fact that the last $L-p$ columns of A have unit ℓ_2-norm. The inequality (3.35) takes care of the denominator of (3.34). Thus it remains to find a lower bound of the numerator of (3.34). To obtain such a bound, we observe that

$$\det(Y^* M_N^* M_N Y) = \det(\boldsymbol{W}^* M_N^* M_N \boldsymbol{W}) \det(AA^*) = \left[\prod_{j=1}^{L} \|\pi_j M_N \boldsymbol{w}^j\|_2^2\right] \cdot \det(AA^*). \tag{3.36}$$

where the last step follows from (3.26). It now remains to bound $\det(AA^*)$.

Let $\mathfrak{a}_{m,m'}$ be the (m,m')th entry of A^*A. Using the orthonormality of $\{\boldsymbol{w}^k\}_{k=1}^{L}$ we have that for any $1 \leqslant m, m' \leqslant p$,

$$\mathfrak{a}_{m,m'} = \sum_{r=1}^{L} \bar{a}_{r,m} a_{r,m'} = \left[\sum_{r=1}^{L} a_{r,m} \boldsymbol{w}^r\right]^* \left[\sum_{r=1}^{L} a_{r,m'} \boldsymbol{w}^r\right] = (\psi v_m)^* (\psi v_{m'}).$$

Since $v_m \perp v_{m'}$, for $m \neq m'$, we see that

$$(\psi v_m)^* (\psi v_{m'}) = -((\mathrm{Id} - \psi) v_m)^* ((\mathrm{Id} - \psi) v_{m'}).$$

By our construction we have that $\|\psi v_m\|_2^2 > 1 - (1/2L)$, for $m = 1, 2, \ldots, p$. Thus

$$|\mathfrak{a}_{m,m'}| = |(\psi v_m)^* (\psi v_{m'})| \leqslant \|(1-\psi) v_m\|_2 \|(1-\psi) v_{m'}\|_2 \leqslant \frac{1}{2L},$$

for all $1 \leqslant m \neq m' \leqslant p$. By a similar reasoning we also obtain that

$$|\mathfrak{a}_{m,m}| \geqslant 1 - \frac{1}{2L},$$

https://doi.org/10.1017/fms.2018.29 Published online by Cambridge University Press

for $m = 1, 2, \ldots, p$. Since the last $L - p$ columns of A are orthonormal and orthogonal to the first p columns we further obtain

$$\mathfrak{a}_{m,m} = 1, \quad m = p+1, p+2, \ldots, L; \quad \mathfrak{a}_{m,m'} = 0, \quad p+1 \leqslant m \neq m' \leqslant L,$$

and

$$\mathfrak{a}_{m,m'} = 0, \quad m \in [p], \quad m' = p+1, p+2, \ldots, L.$$

So in the first p rows of the matrix A^*A we find that the diagonal entries are at least $1 - 1/2L$ and the sum of off-diagonal entries in a row is at most $(p-1)/(2L)$. The last $L-p$ rows of A^*A are simply $\{e_{m'}\}_{m'=p+1}^{L}$. Hence by the Gershgorin circle theorem all eigenvalues of A^*A are at least $\frac{1}{2}$ which implies

$$\det(AA^*) = \det(A^*A) \geqslant 2^{-L}. \tag{3.37}$$

Therefore, from (3.33)–(3.36), we derive

$$\prod_{k=1}^{p} \|M_N v_k\|_2 \geqslant \left(\frac{1}{8(\|M_N\| \vee 1)}\right)^L \cdot \mathfrak{D}^{-p} \cdot \prod_{k=1}^{L} \|\pi_k M_N \boldsymbol{w}^k\|_2^2.$$

Combining this bound with (3.29) finishes the proof of the proposition. □

4. Limiting spectrum of noisy version of banded twisted Toeplitz matrices

We consider in this section upper-triangular twisted Toeplitz matrices of finite symbols, namely upper-triangular matrices with a finite number of slowly varying diagonals; a particular case is the case of upper-triangular Toeplitz matrices of finite symbol.

Our main result is the following theorem.

THEOREM 4.1. *Fix $\mathfrak{d} \in \mathbb{N}$, $\alpha_0 > 1/2$ and $\alpha_\ell \in (0, 1]$ for $\ell \in [\mathfrak{d}]$. For each $\ell \in [\mathfrak{d}] \cup \{0\}$, let $f_\ell : [0, 1] \mapsto \mathbb{C}$ be an α_ℓ-Hölder-continuous function, and let $D_N^{(\ell)}$ be the diagonal matrix with entries $\{f_\ell(i/N)\}_{i\in[N]}$. Set $M_N := \sum_{\ell=0}^{\mathfrak{d}} D_N^{(\ell)} J^\ell$ and set $\mathcal{M}_N$ as in (1.1). Then L_N converges weakly in probability to $\mu_{\mathfrak{d},f}$, the law of $\sum_{\ell=0}^{\mathfrak{d}} f_\ell(X)U^\ell$, where $X \sim \mathrm{Unif}(0, 1)$, U is uniformly distributed on the unit circle in $\mathbb{C}$, and X and U are independent of each other.*

REMARK 4.2. Theorem 1.2 follows from Theorem 4.1 by taking $f_\ell(\cdot) = a_\ell$.

Recall the notation $\mathcal{L}_\mu$ for the log potential of a measure μ, see (1.4). Similar to Theorem 1.3, we prove Theorem 4.1 by showing that for Lebesgue a.e.

https://doi.org/10.1017/fms.2018.29 Published online by Cambridge University Press

$z \in \mathbb{C}$, $\mathcal{L}_{L_N}(z) \to \mathcal{L}_{\mu_{\mathfrak{d},f}}(z)$ in probability. Toward this end we begin by identifying $\mathcal{L}_{\mu_{\mathfrak{d},f}}(\cdot)$. For $z \in \mathbb{C}$ and $x \in [0, 1]$, introduce the symbol

$$P_{z,x}(\lambda) := P_{z,x,\mathfrak{d},f}(\lambda) := f_{\mathfrak{d}}(x)\lambda^{\mathfrak{d}} + f_{\mathfrak{d}-1}(x)\lambda^{\mathfrak{d}-1} + \cdots + f_1(x)\lambda + f_0(x) - z. \quad (4.1)$$

Let $\hat{\mathfrak{d}} := \hat{\mathfrak{d}}(x)$ denote the degree of $P_{z,x}(\cdot)$. If $\hat{\mathfrak{d}} > 0$ then $P_{z,x}(\cdot)$ has $\hat{\mathfrak{d}}$ roots $\lambda_1(z, x), \lambda_2(z, x), \ldots, \lambda_{\hat{\mathfrak{d}}}(z, x)$ (multiplicities allowed). Partition $[0, 1]$ as follows: for $\ell \in \{1, \ldots, \mathfrak{d}\}$ set

$$\mathcal{A}_\ell := \{x \in [0, 1] : f_\ell(x) \neq 0\}, \mathcal{B}_\ell := \mathcal{A}_\ell \Big\backslash \left(\bigcup_{j=\ell+1}^{\mathfrak{d}} \mathcal{A}_j \right), \quad (4.2)$$

and in particular $\mathcal{B}_{\mathfrak{d}} := \mathcal{A}_{\mathfrak{d}}$. Set $\mathcal{B}_0 := [0, 1] \setminus (\bigcup_{\ell=1}^{\mathfrak{d}} \mathcal{B}_\ell)$.

LEMMA 4.3. *For Lebesgue a.e. $z \in \mathbb{C}$ we have*

$$\mathcal{L}_{\mu_{\mathfrak{d},f}}(z) = \sum_{\ell=1}^{\mathfrak{d}} \left[\int_0^1 \left\{ \sum_{j=1}^{\ell} \log_+ |\lambda_j(z, x)| + \log |f_\ell(x)| \right\} \cdot \mathbf{1}_{\mathcal{B}_\ell}(x)\, dx \right] + \int_0^1 \log |f_0(x) - z| \mathbf{1}_{\mathcal{B}_0}(x)\, dx. \quad (4.3)$$

Setting $Y := \sum_{\ell=0}^{\mathfrak{d}} f_\ell(X) U^\ell$, we see that the law of Y is compactly supported in $\mathbb{C}$. Therefore, for any $z \in \mathbb{C}$ and $\varepsilon > 0$,

$$\mathbb{E}_{\mu_{\mathfrak{d},f}} |\log |Y - z| \mathbf{1}_{\{|Y-z| \geqslant \varepsilon\}}| < \infty.$$

On the other hand from Lemma 4.5 we will see that for Lebesgue a.e. $z \in \mathbb{C}$,

$$\lim_{\varepsilon \downarrow 0} \mathbb{E}_{\mu_{\mathfrak{d},f}} |\log |Y - z| \mathbf{1}_{\{|Y-z| \leqslant \varepsilon\}}| = 0.$$

Therefore, by Fubini's theorem one can use iterated integrals to evaluate $\mathcal{L}_{\mu_{\mathfrak{d},f}}(z)$, for Lebesgue almost every $z \in \mathbb{C}$. Note that this does not imply the integrability of individual terms under the integral sign in the RHS of (4.3).

Proof of Lemma 4.3. Following the discussion above we proceed to evaluate $\mathcal{L}_{\mu_{\mathfrak{d},f}}(z)$ using iterated integrals. To this end, we have

https://doi.org/10.1017/fms.2018.29 Published online by Cambridge University Press

$$\mathcal{L}_{\mu_{\mathfrak{d},f}}(z) = \mathbb{E}\left(\log\left|z - \sum_{\ell=0}^{\mathfrak{d}} f_\ell(X)U^\ell\right|\right) = \frac{1}{2\pi}\int_{\mathbb{S}^1}\int_0^1 \log|P_{z,x}(\lambda)|\,dx\,d\lambda$$
$$= \sum_{\ell=1}^{\mathfrak{d}}\int_0^1\left[\frac{1}{2\pi}\int_{\mathbb{S}^1}\left\{\sum_{j=1}^{\ell}\log|\lambda - \lambda_j(z,x)|\right\}d\lambda + \log|f_\ell(x)|\right]\mathbf{1}_{\mathcal{B}_\ell}(x)\,dx$$
$$+ \int_0^1 \log|f_0(x) - z|\mathbf{1}_{\mathcal{B}_0}(x)\,dx.$$

Since

$$\frac{1}{2\pi}\int_{\mathbb{S}^1}\log|z' - \lambda|\,d\lambda = \begin{cases}\log|z'| & \text{if } |z'| \geqslant 1\\ 0 & \text{otherwise,}\end{cases}$$

the claim follows. □

4.1. Reduction to piecewise constant $\{f_\ell\}_{\ell=0}^{\mathfrak{d}}$. To prove Theorem 4.1 we adopt a strategy similar to the proof of Theorem 1.3. Namely, we find approximate singular vectors corresponding to small singular values of $M_N - z\,\mathrm{Id}$ for *almost* all $z \in \mathbb{C}$. To this end, note that

$$((M_N - z\,\mathrm{Id})w)_i = (f_0(i/N) - z)w_i + \sum_{\ell=1}^{\mathfrak{d}} f_\ell(i/N)w_{\ell+i}, \tag{4.4}$$

for $i \in [N - \mathfrak{d}]$. Therefore, given any arbitrary values of $\{w_\ell\}_{\ell=j}^{j+\mathfrak{d}-1}$ one can construct a N-dimensional vector w such that $((M_N - z\,\mathrm{Id})w)_i = 0$ for $i \in [N - \mathfrak{d}] \setminus [j + \mathfrak{d} - 1]$. Such choices of w will be candidates for approximate singular vectors. To study these vectors we note from (4.4) that $(w_{j+\mathfrak{d}}, \ldots, w_{j+1})^{\mathsf{T}} = T_j(w_{j+\mathfrak{d}-1}, \ldots, w_j)^{\mathsf{T}}$ for some transfer matrix T_j. Iterating, we have that

$$(w_N, \ldots, w_{N-\mathfrak{d}+1})^{\mathsf{T}} = \left(\prod_k T_k\right)\cdot(w_{j+\mathfrak{d}-1}, \ldots, w_j)^{\mathsf{T}}.$$

Unlike Theorem 1.3, where the transfer matrices are actually scalars, here the transfer matrices $\{T_k\}$ are in general noncommuting if $\{f_\ell\}_{\ell=0}^{\mathfrak{d}}$ are varying. This complicates the study of the approximate singular value vector w.

To overcome this difficulty we employ the following two-fold argument. We introduce a regularized model where the $\{f_\ell\}_{\ell=0}^{\mathfrak{d}}$ are piecewise constant and hence $\{D_N^{(\ell)}\}_{\ell=0}^{\mathfrak{d}}$ have constant diagonal blocks. Then, the transfer matrices $\{T_k\}$ are constant, and hence commute, within each block. This will be sufficient to derive the necessary properties of the small approximate singular vectors, which in turn allows us to deduce that if the sizes of the blocks are chosen carefully then the

https://doi.org/10.1017/fms.2018.29 Published online by Cambridge University Press

empirical spectral distribution (ESD) of the regularized model admits the limit as described in Theorem 4.1. To complete the proof of Theorem 4.1 we then show that the limits of the ESDs of the regularized model and the original model must be the same.

We now introduce the regularized model. Let $\{f_\ell\}_{\ell=0}^{\mathfrak{d}}$ be as in Theorem 4.1. Fix some $\delta_1, \delta_2 \in (0, 1)$. For $\ell \in [\mathfrak{d}] \cup \{0\}$, let $\hat{D}_N^{(\ell)}$ be a diagonal matrix with

$$(\hat{D}_N^{(\ell)})_{i,i} := f_\ell\left(\frac{\lfloor iN^{\delta_1-1}\rfloor}{N^{\delta_1}}\right) \cdot \mathbf{1}\left(\left|f_\ell\left(\frac{\lfloor iN^{\delta_1-1}\rfloor}{N^{\delta_1}}\right)\right| \geqslant N^{-\delta_2}\right), \quad i \in [N],$$

and define the regularized version of M_N as

$$\hat{M}_N := \sum_{\ell=0}^{\mathfrak{d}} \hat{D}_N^{(\ell)} J^\ell, \quad \hat{\mathcal{M}}_N = \hat{M}_N + N^{-\gamma} G_N. \tag{4.5}$$

Note that in $\hat{M}_N$ we have an additional truncation $\mathbf{1}(|f_\ell(\cdot)| \geqslant N^{-\delta_2})$. This means that if in a certain block $\{f_\ell\}_{\ell=\mathfrak{d}_\star+1}^{\mathfrak{d}}$ are smaller than $N^{-\delta_2}$ then in that block $\hat{M}_N$ can be treated as a matrix with $\mathfrak{d}_\star$ nonzero off-diagonal entries. This, in particular, implies that if $\mathfrak{d}_\star = 0$ then in that block $\hat{M}_N$ becomes a diagonal matrix. Furthermore the truncation at $N^{-\delta_2}$ allows to derive bounds on the operator norm of the transfer matrices, which will be later used during the proofs.

Now we can state our main result for $\hat{\mathcal{M}}_N$. Its proof, which is the main technical part of the proof of Theorem 4.1, is deferred to Section 5.

THEOREM 4.4. *Fix $\mathfrak{d} \in \mathbb{N}$, $\alpha_0 > 1/2$, $\alpha_\ell \in (0, 1]$ for $\ell \in [\mathfrak{d}]$ and $\delta_1, \delta_2 \in (0, 1/2)$ such that* $\max\{\delta_1, \delta_2\} \leqslant (\gamma - 1/2)/(20\mathfrak{d}^2)$. *For each $\ell \in [\mathfrak{d}] \cup \{0\}$, let $f_\ell : [0, 1] \mapsto \mathbb{C}$ be an α_ℓ-Hölder-continuous function. Let $\hat{\mathcal{M}}_N$ be as in* (4.5). *Then $L_{\hat{\mathcal{M}}_N}$ converges weakly in probability to $\mu_{\mathfrak{d},f}$.*

4.2. Proof of Theorem 4.1 assuming Theorem 4.4.

The proof is motivated by the proof of [10, Theorems 4, 5] and the replacement principle, which was introduced in [21, Theorem 2.1]. (We will use a version of the replacement principle from [1, Lemma 10.1].)

We begin with some preparatory material. To apply the replacement principle we will need the following 'regularity' property of the limit, closely related to [10, Definition 1].

LEMMA 4.5. *For Lebesgue almost every $z \in \mathbb{C}$,*

$$\lim_{\varepsilon \to 0} \mathbb{E}_{\mu_{\mathfrak{d},f}}[\log(|X - z|)\mathbf{1}_{\{|X-z|\leqslant\varepsilon\}}] = 0.$$

https://doi.org/10.1017/fms.2018.29 Published online by Cambridge University Press

Proof. Applying Tonelli's theorem for any probability measure μ on $\mathbb{R}$ and $0 \leqslant a_1 < a_2 < 1$,

$$\log(a_2)\mu((a_1,a_2)) - \int_{a_1}^{a_2} \log(x)\, d\mu(x) = \int_{a_1}^{a_2} \left[\int_x^{a_2} \frac{1}{t}\, dt\right] d\mu(x)$$
$$= \int_{a_1}^{a_2} \frac{\mu((a_1,t))}{t}\, dt.$$

As $0 \leqslant a_1 < a_2 < 1$, rearranging the above we obtain

$$\int_{a_1}^{a_2} |\log(t)| d\mu(t) \leqslant |\log(a_2)|\mu((0,a_2)) + \int_{a_1}^{a_2} \frac{\mu((0,t))}{t}\, dt.$$

Therefore, recalling the definitions of $\mu_{\mathfrak{d},f}$ from the statement of Theorem 4.1 and $P_{z,x}(\cdot)$ from (4.1), we have

$$\begin{aligned}
&|\mathbb{E}_{\mu_{\mathfrak{d},f}}[\log(|X-z|)\mathbf{1}_{\{|X-z|\leqslant\varepsilon\}}]| \\
&\quad\leqslant -\frac{1}{2\pi}\log\varepsilon \int_0^1\int_0^{2\pi} \mathbf{1}_{\{|P_{z,x}(e^{\mathrm{i}\theta})|\leqslant\varepsilon\}}\, d\theta\, dx \\
&\qquad + \frac{1}{2\pi}\int_0^{\varepsilon}\int_0^1\int_0^{2\pi} \frac{\mathbf{1}_{\{|P_{z,x}(e^{\mathrm{i}\theta})|\leqslant t\}}}{t}\, d\theta\, dx\, dt \\
&\quad =: -\frac{1}{2\pi}\log\varepsilon \int_0^1\int_0^{2\pi} \mathbf{1}_{\{|P_{z,x}(e^{\mathrm{i}\theta})|\leqslant\varepsilon\}}\, d\theta\, dx + \mathrm{Term}_{\mathfrak{d}}(\varepsilon).
\end{aligned} \tag{4.6}$$

We will show that

$$\lim_{\varepsilon\to 0} \mathrm{Term}_{\mathfrak{d}}(\varepsilon) = 0. \tag{4.7}$$

This will take care of the second term in the RHS of (4.6), and a similar argument (which we omit) applies to the first term, completing the proof of the lemma.

Turning to prove (4.7), fix $1 > \alpha_{\mathfrak{d}} > \alpha_{\mathfrak{d}-1} > \cdots > \alpha_1 > 0$. Recalling that $\{\lambda_i(z,x)\}_{i=1}^d$ are the roots of the equation $P_{z,x}(\lambda) = 0$ on the set $\mathcal{B}_{\mathfrak{d}}$, see (4.2), we write there $|P_{z,x}(e^{\mathrm{i}\theta})| = |f_{\mathfrak{d}}(x)| \cdot \prod_{j=1}^{\mathfrak{d}} |\lambda_j(z,x) - e^{\mathrm{i}\theta}|$. Splitting the integral into the two parts $|f_{\mathfrak{d}}(x)| \geqslant t^{\alpha_{\mathfrak{d}}}$ and $|f_{\mathfrak{d}}(x)| \leqslant t^{\alpha_{\mathfrak{d}}}$, we obtain

$$\mathrm{Term}_{\mathfrak{d}}(\varepsilon) \leqslant \frac{1}{2\pi}\int_0^{\varepsilon}\int_0^1\int_0^{2\pi} \frac{\sum_{j=1}^{\mathfrak{d}} \mathbf{1}_{\{|\lambda_j(z,x)-e^{\mathrm{i}\theta}|\leqslant t^{((1-\alpha_{\mathfrak{d}})/\mathfrak{d})}\}}}{t}\, d\theta\, dx\, dt + \mathrm{Term}_{\mathfrak{d}-1}(\varepsilon),$$

where

$$\mathrm{Term}_{\mathfrak{d}-1}(\varepsilon) := \frac{1}{2\pi}\int_0^{\varepsilon}\int_0^1\int_0^{2\pi} \frac{\mathbf{1}_{\{|P_{z,x}(e^{\mathrm{i}\theta})|\leqslant t\}} \cdot \mathbf{1}_{\{|f_{\mathfrak{d}}(x)|\leqslant t^{\alpha_{\mathfrak{d}}}\}}}{t}\, d\theta\, dx\, dt$$

Since for any $\lambda \in \mathbb{C}$, and $s > 0$ sufficiently small, $\int_0^{2\pi} \mathbf{1}_{\{|\lambda - e^{i\theta}| \leqslant s\}} \, d\theta \leqslant 4s$, we conclude that

$$\limsup_{\varepsilon \to 0} \mathrm{Term}_{\mathfrak{d}}(\varepsilon) \leqslant \limsup_{\varepsilon \to 0} \mathrm{Term}_{\mathfrak{d}-1}(\varepsilon). \tag{4.8}$$

Denote $P_{z,x}^{(\mathfrak{d})}(\lambda) := P_{z,x}(\lambda) - f_{\mathfrak{d}}(x)\lambda^{\mathfrak{d}}$ and, when $f_{\mathfrak{d}-1}(x) \neq 0$, let $\{\lambda_i^{(\mathfrak{d})}(z,x)\}_{i=1}^{d-1}$ denote its roots. Using the triangle inequality, for any $t \in (0,1)$, we further have that

$$\begin{aligned}
\{|P_{z,x}(e^{i\theta})| \leqslant t\} \cap \{|f_{\mathfrak{d}}(x)| \leqslant t^{\alpha_{\mathfrak{d}}}\} &\subseteq \{|P_{z,x}^{(\mathfrak{d})}(e^{i\theta})| \leqslant 2t^{\alpha_{\mathfrak{d}}}\} \\
&= (\{|P_{z,x}^{(\mathfrak{d})}(e^{i\theta})| \leqslant 2t^{\alpha_{\mathfrak{d}}}\} \cap \{|f_{\mathfrak{d}-1}(x)| \geqslant t^{\alpha_{\mathfrak{d}-1}}\}) \\
&\times \cup (\{|P_{z,x}^{(\mathfrak{d})}(e^{i\theta})| \leqslant 2t^{\alpha_{\mathfrak{d}}}\} \cap \{|f_{\mathfrak{d}-1}(x)| \leqslant t^{\alpha_{\mathfrak{d}-1}}\}).
\end{aligned}$$

Therefore arguing as in the lines leading to (4.8) we obtain

$$\begin{aligned}
&\limsup_{\varepsilon \to 0} \mathrm{Term}_{\mathfrak{d}-1}(\varepsilon) \\
&\quad \leqslant \limsup_{\varepsilon \to 0} \frac{1}{2\pi} \int_0^{\varepsilon} \int_0^1 \int_0^{2\pi} \frac{\sum_{j=1}^{\mathfrak{d}-1} \mathbf{1}_{\{|\lambda_j^{(\mathfrak{d})}(z,x) - e^{i\theta}| \leqslant 2^{(1/\mathfrak{d}-1)} t^{(\alpha_{\mathfrak{d}} - \alpha_{\mathfrak{d}-1})/(\mathfrak{d}-1)}\}}}{t} \, d\theta \, dx \, dt \\
&\qquad + \limsup_{\varepsilon \to 0} \mathrm{Term}_{\mathfrak{d}-2}(\varepsilon) \leqslant \limsup_{\varepsilon \to 0} \mathrm{Term}_{\mathfrak{d}-2}(\varepsilon),
\end{aligned}$$

where

$$\mathrm{Term}_{\mathfrak{d}-2}(\varepsilon) := \frac{1}{2\pi} \int_0^{\varepsilon} \int_0^1 \int_0^{2\pi} \frac{\mathbf{1}_{\{|P_{z,x}^{(\mathfrak{d})}(e^{i\theta})| \leqslant 2t^{\alpha_{\mathfrak{d}}}\}} \cdot \mathbf{1}_{\{|f_{\mathfrak{d}-1}(x)| \leqslant t^{\alpha_{\mathfrak{d}-1}}\}}}{t} \, d\theta \, dx \, dt.$$

Iterating the above argument and using induction we deduce that

$$\limsup_{\varepsilon \to 0} \mathrm{Term}_{\mathfrak{d}}(\varepsilon) \leqslant \limsup_{\varepsilon \to 0} \int_0^{\varepsilon} \int_0^1 \frac{\mathbf{1}_{\{|f_0(x) - z| \leqslant (\mathfrak{d}+1)t^{\alpha_1}\}}}{t} \, dx \, dt. \tag{4.9}$$

Since f_0 is an α_0-Hölder-continuous function with $\alpha_0 > 1/2$, we have that $f_0([0,1]) := \{f_0(x) : x \in [0,1]\}$ has zero Lebesgue measure (in $\mathbb{C}$), by a volumetric argument. Moreover the set $f_0([0,1])$ being a closed set, for every $z \notin f_0([0,1])$ we have $\mathrm{dist}(z, f_0([0,1])) := \inf_{x \in [0,1]} |z - f_0(x)| > 0$. Therefore, given a $z \notin f_0([0,1])$ there exists $\varepsilon > 0$ sufficiently small such that $\{x \in [0,1] : |f_0(x) - z| \leqslant (\mathfrak{d}+1)\varepsilon^{\alpha_1}\} = \emptyset$. Thus, from (4.9) we deduce (4.7). This completes the proof of the lemma. □

To prove Theorem 4.1 we need another ingredient.

https://doi.org/10.1017/fms.2018.29 Published online by Cambridge University Press

LEMMA 4.6. *Fix $z \in \mathbb{C}$. Let $\tilde{\nu}^z_{\mathcal{M}_N}$ be the symmetrized version of empirical measure of the singular values of $\mathcal{M}_N - z\,\mathrm{Id}$. Define $\tilde{\nu}^z_{\hat{\mathcal{M}}_N}$ similarly. Then both $\tilde{\nu}^z_{\mathcal{M}_N}$ and $\tilde{\nu}^z_{\hat{\mathcal{M}}_N}$ converge weakly in probability, as $N \to \infty$, to the symmetrized version $\tilde{\nu}^z$ of the law of $|\sum_{\ell=0}^{\eth} f_\ell(X)U^\ell - z|$, where X and U are as in Theorem 4.1.*

Proof of Lemma 4.6. For any probability measure μ supported on $[0, \infty)$, we let $\tilde{\mu}$ denote its symmetrized version, given by $\tilde{\mu}((-y, -x)) = \tilde{\mu}((x, y)) = \frac{1}{2}\mu((x, y))$ for any $0 < x < y < \infty$. We let $\tilde{\nu}^z_{M_N}$ and $\tilde{\nu}^z_{\hat{M}_N}$ be the symmetrized versions of the empirical measures of the singular values of $M_N - z\,\mathrm{Id}$ and of $\hat{M}_N - z\,\mathrm{Id}$, respectively. Note first that $\|N^{-\gamma}G_N\| \to_{N\to\infty} 0$ in probability (and in fact, a.s.) and that there exists a constant $C = C(\{f_\ell\}, z)$ so that $\|M_N - z\,\mathrm{Id}\| \leqslant C$, $\|\hat{M}_N - z\,\mathrm{Id}\| \leqslant C$. Therefore, it follows from Weyl's inequalities that for any metric on the space of probability measures on $\mathbb{R}$ compatible with the weak topology,

$$d(\tilde{\nu}^z_{\mathcal{M}_N}, \tilde{\nu}^z_{M_N}) \to_{N\to\infty} 0, \quad d(\tilde{\nu}^z_{\hat{\mathcal{M}}_N}, \tilde{\nu}^z_{\hat{M}_N}) \to_{N\to\infty} 0,$$

in probability. On the other hand, by definition, $\|M_N - \hat{M}_N\| \leqslant N^{-\min\{\alpha\delta_1, \delta_2\}}$, where $\alpha := \min_{\ell=0}^{\eth}\{\alpha_\ell\}$ and therefore, by Weyl's inequalities,

$$d(\tilde{\nu}^z_{M_N}, \tilde{\nu}^z_{\hat{M}_N}) \to_{N\to\infty} 0.$$

Combining the last two displays, we deduce that it is enough to show that $\tilde{\nu}^z_{M_N}$ converges weakly to $\tilde{\nu}^z_f$, the symmetrized version of the law of $|\sum_{\ell=0}^{\eth} f_\ell(X)U^\ell - z|$. To this end, we will employ the method of moments. We will show that for every $k \in \mathbb{N}$,

$$\lim_{N\to\infty} \frac{1}{N} \mathrm{tr}[(M_N - z\,\mathrm{Id})(M_N - z\,\mathrm{Id})^*]^k = \mathbb{E}\left[\left|\sum_{\ell=0}^{\eth} f_\ell(X)U^\ell - z\right|^{2k}\right]. \tag{4.10}$$

This will complete the proof. Since we can absorb z in $f_0(\cdot)$ it is enough to prove (4.10) only for $z = 0$.

We begin by evaluating the RHS of (4.10) for $z = 0$. As $U^{-1} = U^*$, we have

$$\left|\sum_{\ell=0}^{\eth} f_\ell(X)U^\ell\right|^{2k} = \sum_{\ell=-k\eth}^{k\eth} F_\ell^{(k)}(X)U^\ell, \tag{4.11}$$

for some functions $\{F_\ell^{(k)}(\cdot)\}$. Since $\mathbb{E}U^\ell = \mathbb{E}U^{-\ell} = 0$ for any $0 \neq \ell \in \mathbb{N}$, we get

$$\mathbb{E}\left[\left|\sum_{\ell=0}^{\eth} f_\ell(X)U^\ell\right|^{2k}\right] = \mathbb{E}[F_0^{(k)}(X)] = \int_0^1 F_0^{(k)}(x)\,dx. \tag{4.12}$$

https://doi.org/10.1017/fms.2018.29 Published online by Cambridge University Press

Expanding the sum in the LHS of (4.11) and collecting the coefficient of U^0 it follows that

$$F_0^{(k)}(x) = \sum_{\{\ell_i\}_{i=1}^{2k}:\sum_{i=1}^k(\ell_{2i-1}-\ell_{2i})=0} \prod_{i=1}^k f_{\ell_{2i-1}}(x)\cdot \prod_{i=1}^k \overline{f_{\ell_{2i}}(x)}. \tag{4.13}$$

Turning to identifying the LHS of (4.10) we see that

$$(M_N M_N^*)^k = \sum_{i=1}^k \sum_{\ell_i=0}^{\mathfrak{d}} D_N^{(\ell_1)} J^{\ell_1}(J^*)^{\ell_2}\overline{D_N^{(\ell_2)}} D_N^{(\ell_3)}\cdots D_N^{(\ell_{2k-1})} J^{\ell_{2k-1}}(J^*)^{\ell_{2k}}\overline{D_N^{(\ell_{2k})}}. \tag{4.14}$$

As $\{D_N^{(\ell)}\}$ are diagonal matrices, using the facts that

$$(J^\ell)_{i,j} = \mathbf{1}_{\{1\leqslant j=i+\ell\leqslant N\}}, \quad [(J^*)^\ell]_{i,j} = \mathbf{1}_{\{1\leqslant j=i-\ell\leqslant N\}}, \quad i,j\in[N],$$

we have that

$$\begin{aligned}\Delta(\boldsymbol{\ell})_{i,i} &= (D_N^{(\ell_1)})_{i,i}\cdot\overline{(D_N^{(\ell_2)})_{i+s_1,i+s_1}}\cdot(D_N^{(\ell_3)})_{i+s_1,i+s_1}\cdots(D_N^{(\ell_{2k-1})})_{i+s_{k-1},i+s_{k-1}}\\ &\quad\cdot\overline{(D_N^{(\ell_{2k})})_{i+s_k,i}}\cdot\prod_{j=1}^k \mathbf{1}_{\{i+s_{j-1}+\ell_j\in[N]\}}\cdot\mathbf{1}_{\{i+s_j\in[N]\}},\end{aligned} \tag{4.15}$$

where

$$\Delta(\boldsymbol{\ell})_{i,i} := \left(D_N^{(\ell_1)} J^{\ell_1}(J^*)^{\ell_2}\overline{D_N^{(\ell_2)}} D_N^{(\ell_3)}\cdots D_N^{(\ell_{2k-1})} J^{\ell_{2k-1}}(J^*)^{\ell_{2k}}\overline{D_N^{(\ell_{2k})}}\right)_{i,i},$$

and for $j\in[k]$ $s_j := \sum_{i=1}^j(\ell_{2i-1}-\ell_{2i})$ and $s_0 := 0$. Using the fact that $\{D_N^{(\ell)}\}$ are diagonal matrices again we deduce from (4.15) that if $s_k\neq 0$ then $\Delta(\boldsymbol{\ell})_{i,i}=0$ for any $i\in[N]$.

Thus to establish (4.10) we only need to consider the sum over $\{\ell_i\}_{i=1}^{2k}$ such that $s_k=0$. Fixing such a sequence of $\{\ell_i\}_{i=1}^{2k}$ we observe that for any $i\in[N-2k\mathfrak{d}]\setminus[2k\mathfrak{d}]$

$$\begin{aligned}\Delta(\boldsymbol{\ell})_{i,i} &= (D_N^{(\ell_1)})_{i,i}\cdot(D_N^{(\ell_2)})_{i+s_1,i+s_1}\\ &\quad\cdot\overline{(D_N^{(\ell_3)})_{i+s_1,i+s_1}}\cdots(D_N^{(\ell_{2k-1})})_{i+s_{k-1},i+s_{k-1}}\cdot(D_N^{(\ell_{2k})})_{i,i},\end{aligned} \tag{4.16}$$

and $\Delta(\boldsymbol{\ell})_{i,i}$ is bounded for the remaining indices in $[N]$. Since $(D_N^{(\ell)})_{i,i}=f_\ell(i/N)$, for any $\{\ell_i\}_{i=1}^{2k}$ we find that

$$\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^N\Delta(\boldsymbol{\ell})_{i,i} = \int_0^1\prod_{i=1}^k f_{\ell_{2i-1}}(x)\cdot\prod_{i=1}^k\overline{f_{\ell_{2i}}(x)}\,dx.$$

Substituting in (4.14) and using (4.12)–(4.13), we arrive at (4.10) (with $z=0$). This completes the proof. □

https://doi.org/10.1017/fms.2018.29 Published online by Cambridge University Press

We next introduce further notation. The *Stieltjes transform* of a probability measure μ on $\mathbb{R}$ is defined as

$$G_\mu(\xi) := \int \frac{1}{\xi - y}\,\mu(dy), \quad \xi \in \mathbb{C}\setminus\mathbb{R}.$$

We use the following standard bounds on the Stieltjes transform, see [10, (6)–(8)], in order to integrate the logarithmic function: for any $\tau, \varrho > 0$, and $a, b \in \mathbb{R}$ such that $b - a > \varrho$ we have

$$\mu([a, b]) \leqslant \int_{a-\varrho}^{b+\varrho} |\Im G_\mu(x + i\tau)|\,dx + \frac{\tau}{\varrho}, \tag{4.17}$$

and

$$\mu([a, b]) \geqslant \int_{a+\varrho}^{b-\varrho} |\Im G_\mu(x + i\tau)|\,dx - \frac{\tau}{\varrho}. \tag{4.18}$$

We need also the symmetrized form of the Stieltjes transform, as follows. For an $N \times N$ matrix C_N, define

$$\widetilde{C}_N := \begin{bmatrix} 0 & C_N \\ C_N^* & 0 \end{bmatrix}$$

and the Stieltjes transform

$$G_{C_N}(\xi) := \frac{1}{2N}\operatorname{tr}(\xi - \widetilde{C}_N)^{-1}, \quad \xi \in \mathbb{C}\setminus\mathbb{R}.$$

$G_{C_N}(\cdot)$ is the Stieltjes transform of the symmetrized version of the empirical measure of the singular values of C_N. Using the resolvent identity, we have that for two $N \times N$ matrices C_N and D_N,

$$|G_{C_N}(\xi) - G_{D_N}(\xi)| \leqslant \frac{\|C_N - D_N\|}{(\Im(\xi))^2}. \tag{4.19}$$

We are finally ready to prove Theorem 4.1.

Proof of Theorem 4.1. To show that $L_{\mathcal{M}_N}$ and $L_{\hat{\mathcal{M}}_N}$ admit the same limit we need to show that for every continuous bounded function $f : \mathbb{C} \mapsto \mathbb{R}$

$$\int f(z)\,dL_{\mathcal{M}_N}(z) - \int f(z)\,dL_{\hat{\mathcal{M}}_N}(z) \to 0, \tag{4.20}$$

as $N \to \infty$, in probability. By continuity we have that $\max_\ell \sup_{x\in[0,1]} |f_\ell(x)| < \infty$, and then $\|M_N\|, \|\hat{M}_N\| < \infty$. Therefore, using (1.2), we have that $\|\mathcal{M}_N\|$, $\|\hat{\mathcal{M}}_N\| < \infty$ with probability approaching one. Thus, it suffices to show that

https://doi.org/10.1017/fms.2018.29 Published online by Cambridge University Press

(4.20) holds for compactly supported functions f. Furthermore, an application of the Stone–Weierstrass theorem yields that one can restrict attentions to smooth functions, namely it is enough to consider $f \in C_c^2(\mathbb{C})$.

Turning to the proof of (4.19) for such f, we use [1, Lemma 8.1] to note that we need to show that:

(i) The expression
$$\frac{1}{N}\|\mathcal{M}_N\|_2^2 + \frac{1}{N}\|\hat{\mathcal{M}}_N\|_2^2$$
is bounded in probability.

(ii) For Lebesgue almost all $z \in B_{\mathbb{C}}(0, R)$,
$$\frac{1}{N}\log|\det(\mathcal{M}_N - z\,\mathrm{Id})| - \frac{1}{N}\log|\det(\hat{\mathcal{M}}_N - z\,\mathrm{Id})| \to 0,$$
as $N \to \infty$, in probability.

As already noted above, $\|\mathcal{M}_N\|, \|\hat{\mathcal{M}}_N\| < \infty$ with probability approaching one, condition (i) is immediate.

It remains to establish condition (ii). By the definitions of $\tilde{\nu}^z_{\mathcal{M}_N}$ and $\tilde{\nu}^z_{\hat{\mathcal{M}}_N}$, we have

$$\frac{1}{N}\log|\det(\mathcal{M}_N - z\,\mathrm{Id})| - \frac{1}{N}\log|\det(\hat{\mathcal{M}}_N - z\,\mathrm{Id})| \\ = \int \log|x|\,d\tilde{\nu}^z_{\mathcal{M}_N}(x) - \int \log|x|\,d\tilde{\nu}^z_{\hat{\mathcal{M}}_N}(x). \tag{4.21}$$

We point out to the reader that both $\mathcal{M}_N - z\,\mathrm{Id}$ and $\hat{\mathcal{M}}_N - z\,\mathrm{Id}$ being Gaussian perturbations of some deterministic matrices are nonsingular almost surely, and therefore both sides of (4.21) are well defined on a set with probability one. Now from Lemma 4.6 we have that for any $\varepsilon > 0$ and $R_0 \geqslant 1$,

$$\int_{(\varepsilon,R_0)\cup(-R_0,-\varepsilon)} \log|x|\,d\tilde{\nu}^z_{\mathcal{M}_N}(x) - \int_{(\varepsilon,R_0)\cup(-R_0,-\varepsilon)} \log|x|\,d\tilde{\nu}^z_{\hat{\mathcal{M}}_N}(x) \to 0, \\ \text{in probability.} \tag{4.22}$$

We observe that

$$\mathbb{E}\int x^2\,d\tilde{\nu}^z_{\mathcal{M}_N}(x) = \mathbb{E}\left[\frac{1}{N}\operatorname{tr}\{(\mathcal{M}_N - z\,\mathrm{Id})(\mathcal{M}_N - z\,\mathrm{Id})^*\}\right] \\ \leqslant 3\left[|z|^2 + \frac{1}{N}\operatorname{tr}(M_N M_N^*) + N^{-(1+2\gamma)}\operatorname{tr}(G_N G_N^*)\right] \leqslant C,$$

https://doi.org/10.1017/fms.2018.29 Published online by Cambridge University Press

for some $C < \infty$. Thus using the fact that $|\log|x||/x^2$ is decreasing for $|x| \geqslant \sqrt{e}$, we have that, for any $R_0 > \sqrt{e}$,

$$\mathbb{E}\left[\int_{(-R_0,R_0)^c} \log|x|\, d\tilde{\nu}^z_{\mathcal{M}_N}(x)\right] \leqslant \frac{\log R_0}{R_0^2}\mathbb{E}\left[\int_{(-R_0,R_0)^c} x^2 d\tilde{\nu}^z_{\mathcal{M}_N}(x)\right] \leqslant \frac{\log R_0}{R_0^2}\cdot C.$$

By the same argument

$$\mathbb{E}\left[\int_{(-R_0,R_0)^c} \log|x|\, d\tilde{\nu}^z_{\hat{\mathcal{M}}_N}(x)\right] \leqslant \frac{\log R_0}{R_0^2}\cdot C. \tag{4.23}$$

So, applying Markov's inequality we see that for any $\eta > 0$, there exists $R_0(\eta)$ such that

$$\begin{aligned}\lim_{\eta\to 0}\limsup_{N\to\infty}\mathbb{P}\Bigg(\bigg|\int_{(-R_0(\eta),R_0(\eta))^c} &\log|x|\, d\tilde{\nu}^z_{\mathcal{M}_N}(x)\\ &- \int_{(-R_0(\eta),R_0(\eta))^c} \log|x|\, d\tilde{\nu}^z_{\hat{\mathcal{M}}_N}(x)\bigg| > \eta\Bigg)\\ = 0.\end{aligned} \tag{4.24}$$

Combining (4.21)–(4.24) we see that to establish condition (ii) it remains to show that for any $\eta > 0$, there exists $\varepsilon(\eta)$ such that

$$\lim_{\eta\to 0}\limsup_{N\to\infty}\mathbb{P}\left(\left|\int_{-\varepsilon(\eta)}^{\varepsilon(\eta)} \log|x|\, d\tilde{\nu}^z_{\mathcal{M}_N}(x) - \int_{-\varepsilon(\eta)}^{\varepsilon(\eta)} \log|x| d\tilde{\nu}^z_{\hat{\mathcal{M}}_N}(x)\right| > 2\eta\right) = 0. \tag{4.25}$$

To prove the above it is enough to show that

$$\lim_{\eta\to 0}\limsup_{N\to\infty}\mathbb{P}\left(\left|\int_{-\varepsilon(\eta)}^{\varepsilon(\eta)} \log|x|\, d\tilde{\nu}^z_{\mathcal{M}_N}(x)\right| > \eta\right) = 0 \tag{4.26}$$

and

$$\lim_{\eta\to 0}\limsup_{N\to\infty}\mathbb{P}\left(\left|\int_{-\varepsilon(\eta)}^{\varepsilon(\eta)} \log|x|\, d\tilde{\nu}^z_{\hat{\mathcal{M}}_N}(x)\right| > \eta\right) = 0, \tag{4.27}$$

where $\varepsilon(\eta)$ is as in (4.25). It follows from Theorem 4.4 that for Lebesgue almost every $z \in \mathbb{C}$, $\mathcal{L}_{L_{\hat{\mathcal{M}}_N}}(z) \to_{N\to\infty} \mathcal{L}_{\mu_{d,f}}(z)$ in probability, which is equivalent to the statement that

$$\int \log|x|\, d\tilde{\nu}^z_{\hat{\mathcal{M}}_N}(x) \to \int \log|x|\, d\tilde{\nu}^z_f(x) \quad \text{in probability}. \tag{4.28}$$

Applying Lemma 4.6 we have that

$$\int_{(-\varepsilon,\varepsilon)^c\cap(-R_0,R_0)} \log|x|\, d\tilde{\nu}^z_{\hat{\mathcal{M}}_N}(x) \to \int_{(-\varepsilon,\varepsilon)^c\cap(-R_0,R_0)} \log|x|\, d\tilde{\nu}^z_f(x) \quad \text{in probability,} \tag{4.29}$$

for any $\varepsilon > 0$ and $R_0 < \infty$. As $\tilde{\nu}^z_f$ is compactly supported using (4.23) we obtain that

$$\begin{aligned}\lim_{\eta\to 0}\limsup_{N\to\infty}\mathbb{P}\bigg(\bigg|\int_{(-R_0(\eta),R_0(\eta))^c} &\log|x|\, d\tilde{\nu}^z_{\hat{\mathcal{M}}_N}(x)\\ &- \int_{(-R_0(\eta),R_0(\eta))^c} \log|x|\tilde{\nu}^z_f(x)\bigg| > \eta\bigg)\\ = \lim_{\eta\to 0}\limsup_{N\to\infty}\mathbb{P}\bigg(\bigg|\int_{(-R_0(\eta),R_0(\eta))^c} &\log|x|\, d\tilde{\nu}^z_{\hat{\mathcal{M}}_N}(x)\bigg| > \eta\bigg) = 0,\end{aligned}$$

where $R_0(\eta)$ is as in (4.24). Therefore, from (4.28)–(4.29) we now deduce that, for every $\varepsilon > 0$,

$$\int_{-\varepsilon}^{\varepsilon} \log|x|\, d\tilde{\nu}^z_{\hat{\mathcal{M}}_N}(x) \to \int_{-\varepsilon}^{\varepsilon} \log|x|\, d\tilde{\nu}^z_f(x) \quad \text{in probability.}$$

This together with Lemma 4.5 yields (4.27). It remains to establish (4.26).

To this end, from [9, Proposition 16] we have that, for any $t > 0$,

$$\mathbb{P}(\sigma_N(\mathcal{M}_N - z\,\mathrm{Id}) < t) \leqslant C_0 N^{1+2\gamma} t^2,$$

for some constant C_0. Therefore, for any $\eta > 0$,

$$\begin{aligned}\limsup_{N\to\infty}\mathbb{P}\bigg(\bigg|\int_{-N^{-(1+\gamma)}}^{N^{-(1+\gamma)}} \log|x|\, d\tilde{\nu}^z_{\mathcal{M}_N}(x)\bigg| \geqslant \eta/2\bigg)&\\ \leqslant \limsup_{N\to\infty}\mathbb{P}(\sigma_N(\mathcal{M}_N - z\,\mathrm{Id}) < N^{-(1+\gamma)}) = 0.&\end{aligned} \tag{4.30}$$

Since $\|\mathcal{M}_N - \hat{\mathcal{M}}_N\| \leqslant N^{-\delta'}$, where $\delta' := \min\{\alpha\delta_1, \delta_2\}$, from (4.19) and setting $\tau = N^{-\delta'/4}$ we obtain

$$|G_{\mathcal{M}_N - z\,\mathrm{Id}}(x + \mathrm{i}\tau) - G_{\hat{\mathcal{M}}_N - z\,\mathrm{Id}}(x + \mathrm{i}\tau)| \leqslant N^{-\delta'/2}.$$

https://doi.org/10.1017/fms.2018.29 Published online by Cambridge University Press

Setting $\varrho = N^{-\delta'/8}$, $\kappa = \delta'/16$, and using (4.17)–(4.18) in the second inequality, we have that

$$\begin{aligned}
-\int_{N^{-(1+\gamma)}}^{N^{-\kappa}} \log(x)\, d\tilde{\nu}^z_{\mathcal{M}_N}(x) &\leqslant (1+\gamma)\log N \cdot \nu^z_{\mathcal{M}_N}([N^{-(1+\gamma)}, N^{-\kappa}]) \\
&\leqslant (1+\gamma)\log N (2N^{-\delta'/8} + \tilde{\nu}^z_{\hat{\mathcal{M}}_N}([N^{-(1+\gamma)} - 2\varrho, N^{-\kappa} + 2\varrho])) \\
&\leqslant (1+\gamma)\log N (2N^{-\delta'/8} + 2\tilde{\nu}^z_{\hat{\mathcal{M}}_N}([0, 2N^{-\kappa}])) \\
&\leqslant 2(1+\gamma)\log N \cdot N^{-\delta'/8} - \frac{4(1+\gamma)}{\kappa}\int_0^{2N^{-\kappa}} \log(x)\, d\tilde{\nu}^z_{\hat{\mathcal{M}}_N}(x), \qquad (4.31)
\end{aligned}$$

for all large N, where in the third inequality we used the symmetry of $\tilde{\nu}^z_{\hat{\mathcal{M}}_N}$ and $\varrho = o(N^{-\kappa})$.

It remains to bound the integral of $\log(\cdot)$ over $(N^{-\kappa}, \varepsilon)$. Toward this, using integration by parts we note that, for $0 \leqslant a_1 < a_2 < 1$ and any probability measure μ on $\mathbb{R}$,

$$-\int_{a_1}^{a_2} \log(x)\, d\mu(x) = -\log(a_2)\mu([a_1, a_2]) + \int_{a_1}^{a_2} \frac{\mu([a_1, t])}{t}\, dt. \qquad (4.32)$$

Arguing as in (4.31) we obtain

$$\begin{aligned}
\int_{N^{-\kappa}}^{\varepsilon} \frac{\tilde{\nu}^z_{\mathcal{M}_N}([N^{-\kappa}, t])}{t}\, dt &\leqslant 2N^{-\delta'/8}\int_{N^{-\kappa}}^{\varepsilon} \frac{1}{t}\, dt + \int_{N^{-\kappa}}^{\varepsilon} \frac{\tilde{\nu}^z_{\hat{\mathcal{M}}_N}([N^{-\kappa}/2, t + N^{-\kappa}])}{t}\, dt \\
&\leqslant 2\kappa N^{-\delta'/8} \cdot \log N + 2\int_{N^{-\kappa}/2}^{2\varepsilon} \frac{\tilde{\nu}^z_{\hat{\mathcal{M}}_N}([N^{-\kappa}/2, t])}{t}\, dt,
\end{aligned} \qquad (4.33)$$

where in the last step we have used the fact that $t + N^{-\kappa} \leqslant 2t$ for any $t \geqslant N^{-\kappa}$, and a change of variable. A similar reasoning yields that

$$-\log(\varepsilon)\tilde{\nu}^z_{\mathcal{M}_N}([N^{-\kappa}, \varepsilon]) \leqslant -\log(\varepsilon)(2N^{-\delta'/8} + \tilde{\nu}^z_{\mathcal{M}_N}([N^{-\kappa}/2, 2\varepsilon])). \qquad (4.34)$$

Thus combining (4.31), and (4.33)–(4.34), and using (4.32) we deduce that for $\varepsilon > 0$ sufficiently small and all large N,

$$-\int_{N^{-(1+\gamma)}}^{\varepsilon} \log(x)\, d\tilde{\nu}^z_{\mathcal{M}_N}(x) \leqslant C_0'\left[\log N \cdot N^{-\delta'/8} - \int_0^{2\varepsilon} \log(x)\, d\tilde{\nu}^z_{\hat{\mathcal{M}}_N}(dx)\right],$$

where C_0' is some large constant. Now, combining this with (4.27)–(4.30), the claim in (4.26) follows. This completes the proof of the theorem. □

https://doi.org/10.1017/fms.2018.29 Published online by Cambridge University Press

5. The piecewise constant case—proof of Theorem 4.4

Similar to the proof of Theorem 1.3, the main step in the proof of Theorem 4.4 is the proof of convergence of the log potentials $\mathcal{L}_{L_{\hat{\mathcal{M}}_N}}(z)$, which will use Theorem 2.1. To verify the assumptions of Theorem 2.1 we need to find an analogue of Theorem 3.1. To this end, we need to identify approximate singular vectors of $\hat{M}_N(z) := \hat{M}_N - z\,\mathrm{Id}$ corresponding to small singular values, establish that $\|\hat{M}_N(z)w\|_2$ cannot be small for any vector w orthogonal to these approximate singular vectors, and obtain matching upper and lower bounds, up to subexponential factors, on the product of the small singular values. Overall, we follow a scheme similar to the one in Section 3. However, as we will see below, some significant changes are necessary when treating the case $\mathfrak{d} > 1$, even in the constant diagonal setup.

We begin by fixing additional notation. Set

$$\mathfrak{b}_k := \{i \in [N] : \lfloor iN^{\delta_1 - 1}\rfloor = k\}, \quad k = 0, 1, \ldots, L_0 := \lfloor N^{\delta_1}\rfloor. \tag{5.1}$$

Note that if $i \in \mathfrak{b}_k$, for some k, then for any $\ell \in [\mathfrak{d}] \cup \{0\}$, $(\hat{D}_N^{(\ell)})_{i,i} = f_\ell(kN^{-\delta_1})\mathbf{1}_{\{|f_\ell(kN^{-\delta_1})| \geqslant N^{-\delta_2}\}} =: t_\ell(k)$, that is, the diagonals of $\hat{M}_N$ are constant within each $\mathfrak{b}_k$. Therefore, for any $w \in \mathbb{C}^N$ and $i \in \mathfrak{b}_k$,

$$(\hat{M}_N(z)w)_i = (t_0(k) - z)w_i + \sum_{\ell=1}^{\mathfrak{d}} t_\ell(k) w_{\ell+i}. \tag{5.2}$$

Using (5.2) we will construct vectors w for which $\hat{M}_N(z)w \approx 0$. It will be easier to reformulate (5.2) as a system of linear equations, for which we need to define the following notion of transfer matrix.

DEFINITION 5.1. Fix $k \in \mathbb{N} \cup \{0\}$ such that $\mathfrak{b}_k \neq \emptyset$. Denote

$$\hat{\mathfrak{d}}_N := \hat{\mathfrak{d}}_N(k) := \max\{\ell : |t_\ell(k)| \neq 0\}.$$

For $\hat{\mathfrak{d}}_N(k) > 0$ denote

$$\hat{\boldsymbol{t}}(k) := \left(\frac{t_{\hat{\mathfrak{d}}_N - 1}(k)}{t_{\hat{\mathfrak{d}}_N}(k)}, \frac{t_{\hat{\mathfrak{d}}_N - 2}(k)}{t_{\hat{\mathfrak{d}}_N}(k)}, \ldots, \frac{t_1(k)}{t_{\hat{\mathfrak{d}}_N}(k)}\right)$$

and define the following $\hat{\mathfrak{d}}_N \times \hat{\mathfrak{d}}_N$ matrix:

$$T_k(z) := \begin{bmatrix} -\hat{\boldsymbol{t}}(k) & \dfrac{z - t_0(k)}{t_{\hat{\mathfrak{d}}_N}(k)} \\ I_{\hat{\mathfrak{d}}_N - 1} & \mathbf{0}_{\hat{\mathfrak{d}}_N - 1} \end{bmatrix},$$

where $\mathbf{0}_n$ is the n-dimensional vector of all zeros.

Recall the symbol $P_{z,x}$, see (4.1). The next result shows that the eigenvalues of the transfer matrix $T_k(z)$ are the roots of the equations $P_{z,x}(\lambda) = 0$ for some appropriate choices of x.

LEMMA 5.1. *Fix $k \in \mathbb{N} \cup \{0\}$ such that $\mathfrak{b}_k \neq \emptyset$. Let $\hat{\mathfrak{d}}_N$ and $T_k(z)$ be as in Definition 5.1. Assume $\hat{\mathfrak{d}}_N > 0$. Fix any $\ell \in [\hat{\mathfrak{d}}_N]$. Denote*

$$\boldsymbol{v}_\ell(k)^{\mathsf{T}} := \left(\hat{\lambda}_\ell^{\hat{\mathfrak{d}}_N-1}(z,k)\ \hat{\lambda}_\ell^{\hat{\mathfrak{d}}_N-2}(z,k)\ \cdots\ 1\right),$$

where $\{\hat{\lambda}_\ell(z,k)\}_{\ell=1}^{\hat{\mathfrak{d}}_N}$ are the roots of the equation

$$\hat{P}_{z,k}(\lambda) := t_{\mathfrak{d}}(k)\lambda^{\mathfrak{d}} + t_{\mathfrak{d}-1}(k)\lambda^{\mathfrak{d}-1} + \cdots + t_1(k)\lambda + t_0(k) - z = 0.$$

Then for any $\ell \in [\hat{\mathfrak{d}}_N]$, $\boldsymbol{v}_\ell(k)$ is an eigenvector of $T_k(z)$ corresponding to $\hat{\lambda}_\ell(z,k)$.

Proof. Since $\{\hat{\lambda}_\ell(z,k)\}_{\ell=1}^{\hat{\mathfrak{d}}_N}$ are the roots of the equation $\hat{P}_{z,k}(\lambda) = 0$, from the definition of $T_k(z)$ we note that

$$T_k(z)\boldsymbol{v}_\ell(k) = \begin{pmatrix} \hat{\lambda}_\ell^{\hat{\mathfrak{d}}_N}(z,k) - \dfrac{1}{t_{\hat{\mathfrak{d}}_N}(k)}\hat{P}_{z,k}(\hat{\lambda}_\ell(z,k)) \\ \hat{\lambda}_\ell^{\hat{\mathfrak{d}}_N-1}(z,k) \\ \vdots \\ \hat{\lambda}_\ell(z,k) \end{pmatrix} = \hat{\lambda}_\ell(z,k)\cdot \boldsymbol{v}_\ell(k).$$

This completes the proof. □

Lemma 5.1 shows that the eigenvalues of $T_k(z)$ are the roots of the polynomial equation $\hat{P}_{z,k}(\lambda) = 0$. If $\min_\ell\{|f_\ell(x)|\} \geqslant N^{-\delta_2}$, then it is easy to see that $P_{z,x}(\lambda) = \hat{P}_{z,k}(\lambda)$ where $x = kN^{-\delta_1}$. Therefore, in such cases, the roots of $P_{z,x}(\lambda)$ and $\hat{P}_{z,k}(\lambda)$ coincide. This property will be later used in the proof of Theorem 4.4.

Note that Lemma 5.1 also provides the eigenvectors of $T_k(z)$. Using these eigenvectors we now construct approximate singular vectors of $\hat{M}_N(z)$, corresponding to small singular values.

Construction of approximate singular vectors. Recall, see (5.1), that $\{\mathfrak{b}_k\}_{k=0}^{L_0}$ is a partition of $[N]$ with $L_0 = \lfloor N^{\delta_1}\rfloor$, so that within each block $\mathfrak{b}_k$, the entries of the diagonals of $\hat{M}_N$ remain constant. However, for certain values of N, the last block $\mathfrak{b}_{L_0}$ may have a small length. To overcome this, we slightly modify $\hat{M}_N$. Namely, we replace the (off)-diagonal entries of $\hat{M}_N$ in the last two blocks by

https://doi.org/10.1017/fms.2018.29 Published online by Cambridge University Press

their average. By a slight abuse of notation, we continue to write $\{\mathfrak{b}_k\}$ to indicate the blocks in which the (off)-diagonal entries of the modified $\hat{M}_N$ are constant. Note that we now have that $N^{1-\delta_1}/2 \leqslant |\mathfrak{b}_k| \leqslant 2N^{1-\delta_1}$, for all k. Since this extra modification results in a change of operator norm of $O(N^{-\alpha\delta_1} + N^{-\delta_2})$, where we recall $\alpha := \min_{\ell=0}^{\mathfrak{d}}\{\alpha_\ell\}$, it is enough to prove Theorem 4.4 for the modified $\hat{M}_N(z)$.

Fix $\delta_3 \in (0, 1/3)$. Next we choose a refinement of the above partition $\{\{\mathfrak{b}_k^{(j')}\}_{j'=1}^{L_k'}\}_{k=0}^{L_0} \subset \{\mathfrak{b}_k\}_{k=0}^{L_0}$ such that $N^{\delta_3}/2 \leqslant |\mathfrak{b}_k^{(j')}| \leqslant 2N^{\delta_3}$ for all $k \in [L_0] \cup \{0\}$ and $j' \in [L_k']$. Since $\delta_1 < 1/2$ and $\delta_3 < 1/3$, such a property can be ensured. Fix

$$L := \sum_{k=0}^{L_0} L_k' = O(N^{1-\delta_3}). \tag{5.3}$$

Further let

$$0 = i_1 < i_2 < i_3 < \cdots < i_{L+1} = N,$$

be the end points of the partition $\{\{\mathfrak{b}_k^{(j)}\}_{j=1}^{L_k'}\}_{k=0}^{L_0}$. That is, for any $k \in [L_0] \cup \{0\}$ and $j' \in [L_k']$ we have $\mathfrak{b}_k^{(j')} = [i_{j+1}] \setminus [i_j]$ for some $j \in [L] \cup \{0\}$, which in turn implies that $N^{\delta_3}/2 \leqslant i_{j+1} - i_j \leqslant 2N^{\delta_3}$ for all $j \in [L] \cup \{0\}$.

Next fix $k \in [L_0] \cup \{0\}$ and $j' \in [L_k']$ and assume that $\mathfrak{b}_k^{(j')} = [i_{j+1}] \setminus [i_j]$ for some $j \in [L] \cup \{0\}$. For $\ell \in [\hat{\mathfrak{d}}_N(k)]$ define the N-dimensional vectors $\boldsymbol{w}_\ell^j(k)$ as follows

$$(\boldsymbol{w}_\ell^j(k))_m := \begin{cases} \hat{\lambda}_\ell(z,k)^{m-i_j-1} & \text{for } m = i_j+1, i_j+2, \ldots, i_{j+1} \\ 0 & \text{otherwise} \end{cases}$$

where $\hat{\lambda}_1(z,k), \hat{\lambda}_2(z,k), \ldots, \hat{\lambda}_{\hat{\mathfrak{d}}_N}(z,k)$ are the roots of the equation $\hat{P}_{z,k}(\lambda) = 0$.

REMARK 5.2. Note that when $\hat{\mathfrak{d}}_N(k) = 0$ then a block in $\hat{M}_N(z)$ becomes diagonal. Since the singular values of a diagonal matrix are the absolute values of its diagonal entries, we do not need to bother with constructing approximate singular vectors. Therefore, when computing bounds on the small singular values we will assume that $\hat{\mathfrak{d}}_N(k) > 0$, and define the candidates for small singular vectors $\{\boldsymbol{w}_\ell^j(k)\}$ only in that case.

The next lemma gives a simple but useful property of the vectors $\{\boldsymbol{w}_\ell^j(k)\}$. Before stating the result let us introduce one more notation. Fix any

https://doi.org/10.1017/fms.2018.29 Published online by Cambridge University Press

N-dimensional vector u. For $k \in [L_0] \cup \{0\}$, denote

$$u[\![k, m]\!] := \begin{pmatrix} u_{m+\hat{\mathfrak{d}}_N(k)-1} \\ u_{m+\hat{\mathfrak{d}}_N(k)-2} \\ \vdots \\ u_m \end{pmatrix}, \quad m = 1, 2, \ldots, N - \hat{\mathfrak{d}}_N(k) + 1.$$

LEMMA 5.3. *Fix $k \in [L_0] \cup \{0\}$ and $j' \in [L'_k]$ such that $\mathfrak{b}_k^{(j')} = [i_{j+1}] \setminus [i_j]$ for some $j \in [L] \cup \{0\}$. Assume that $\hat{\mathfrak{d}}_N(k) > 0$.*

(i) *Let u be an N-dimensional vector such that*

$$u[\![k, m+1]\!] = T_k(z)u[\![k, m]\!], \quad m \in [i_{j+1} - \hat{\mathfrak{d}}_N(k)] \setminus [i_j]. \tag{5.4}$$

Then

$$(\hat{M}_N(z)u)_m = 0, \quad m \in [i_{j+1} - \hat{\mathfrak{d}}_N(k)] \setminus [i_j].$$

(ii) *For any $\ell \in [\hat{\mathfrak{d}}_N(k)]$ we have*

$$(\hat{M}_N(z)\boldsymbol{w}_\ell^j(k))_m = 0, \quad m \in [i_{j+1} - \hat{\mathfrak{d}}_N(k)] \setminus [i_j].$$

Proof. The condition (5.4) implies that

$$t_{\hat{\mathfrak{d}}_N(k)}(k)u_{m+\hat{\mathfrak{d}}_N(k)} = (z - t_0(k))u_m - \sum_{m'=1}^{\hat{\mathfrak{d}}_N(k)-1} t_{m'}(k)u_{m+m'},$$

for $m \in [i_{j+1} - \hat{\mathfrak{d}}_N(k)] \setminus [i_j]$. The conclusion of part (i) is now immediate from (5.2) and the definition of $\hat{\mathfrak{d}}_N(k)$. To prove (ii) we note that $\boldsymbol{w}_\ell^j(k)[\![k, m]\!] = \hat{\lambda}_\ell(z, x)^{m-i_j-1}\boldsymbol{v}_\ell(k)$, for $m \in [i_{j+1} - \hat{\mathfrak{d}}_N(k)] \setminus [i_j]$, where $\boldsymbol{v}_\ell$ is as in Lemma 5.1. Thus from Lemma 5.1 it follows that

$$\begin{aligned} T_k(z)\boldsymbol{w}_\ell^j(k)[\![k, m]\!] &= \hat{\lambda}_\ell(z, k)^{m-i_j-1}T_k(z)\boldsymbol{v}_\ell(k) = \hat{\lambda}_\ell(z, k)^{m-i_j}\boldsymbol{v}_\ell(k) \\ &= \boldsymbol{w}_\ell^j(k)[\![k, m+1]\!]. \end{aligned}$$

Now the proof completes by applying part (i). □

Identification of the set of bad z's. Recall that to prove Theorem 4.4 we only need to find the limit of the log potential for Lebesgue almost every $z \in \mathbb{C}$. As we will see below, our methods to control the small singular values of $\hat{M}_N(z)$ break

https://doi.org/10.1017/fms.2018.29 Published online by Cambridge University Press

down when $P_{z,x}(\lambda) = 0$ has roots near the unit circle or when the Vandermonde matrix

$$V(k) := \left[v_1(k)\ v_2(k)\ \cdots\ v_{\hat{\mathfrak{d}}_N}(k)\right] \tag{5.5}$$

loses invertibility. In the next lemma we show that the collection of all such bad z's has small Lebesgue measure.

LEMMA 5.4. *Let $\mathcal{B}_N$ denote the collection of $z \in \mathbb{C}$ such that either of the following properties hold:*

(i) *For some $k \in [L_0] \cup \{0\}$, such that $\hat{\mathfrak{d}}_N(k) > 0$,*

$$|t_{\hat{\mathfrak{d}}_N(k)}(k)|^{\hat{\mathfrak{d}}_N(k)-1} |\det(V(k))|^2 \leqslant N^{-2\delta_1 \mathfrak{d}}.$$

(ii) *For some $k \in [L_0] \cup \{0\}$, such that $\hat{\mathfrak{d}}_N(k) > 0$, there exists a root λ of the equation $\hat{P}_{z,k}(\lambda) = 0$ such that $1 - N^{-3\delta_1} \leqslant |\lambda| \leqslant 1 + N^{-3\delta_1}$.*

(iii)

$$\inf_{k=0}^{L_0} |z - t_0(k)| \leqslant N^{-(\alpha_0\delta_1)/(2(2\alpha_0-1))}.$$

Then $\mathrm{Leb}(\mathcal{B}_N) \leqslant N^{-\delta_1}$ *for all large N.*

Proof. We first estimate the area of z satisfying (ii). Fix $k \in [L_0] \cup \{0\}$ such that $\hat{\mathfrak{d}}_N := \hat{\mathfrak{d}}_N(k) > 0$. For any $\varepsilon > 0$, denote $\mathcal{A}_\varepsilon := B_{\mathbb{C}}(0, 1+\varepsilon) \setminus B_{\mathbb{C}}(0, 1-\varepsilon)$, where we recall that $B_{\mathbb{C}}(0, r)$ denote the open disc of radius r in the complex plane centered at zero. Let $Q_k(\lambda) := \hat{P}_{z,k}(\lambda) + z = \sum_{\ell=0}^{\hat{\mathfrak{d}}_N} t_\ell(k)\lambda^\ell$. Then, the set of z for which there exists a $\lambda \in \mathcal{A}_{N^{-3\delta_1}}$ so that $\hat{P}_{z,k}(\lambda) = 0$ are contained in the image $Q_k(\mathcal{A}_{N^{-3\delta_1}})$. Therefore the set of all z's satisfying property (ii) is contained in $\bigcup_k Q_k(\mathcal{A}_{N^{-3\delta_1}})$. The area of such an image $Q_k(\mathcal{A}_{N^{-3\delta_1}})$ can be estimated by $\sup_{|z|<2} |Q_k'(z)| \mathrm{Leb}(\mathcal{A}_{N^{-3\delta_1}})$. Since $L_0 = O(N^{\delta_1})$, it follows that for all N sufficiently large

$$\mathrm{Leb}\left(\bigcup_k Q_k(\mathcal{A}_{N^{-3\delta_1}})\right) \leqslant \frac{1}{3} N^{-\delta_1}.$$

Turning to the set of z satisfying (i), we recall that the *discriminant* of a polynomial $\tilde{P}(\lambda) := \sum_{\ell=1}^m t_\ell \lambda^\ell$ is

$$D(\tilde{P}) := \det[\mathrm{Disc}(\tilde{P})] = t_m^{2m-1} \prod_{1\leqslant \ell < \ell' \leqslant m} (\lambda_\ell - \lambda_{\ell'})^2, \tag{5.6}$$

https://doi.org/10.1017/fms.2018.29 Published online by Cambridge University Press

where

$$\mathrm{Disc}(\tilde{P}) := \begin{bmatrix} t_m & t_{m-1} & \cdots & \cdots & t_0 & & & \\ & \ddots & & & & \ddots & & \\ & & t_m & t_{m-1} & \cdots & \cdots & \cdots & t_0 \\ mt_m & \cdots & 2t_2 & t_1 & & & & \\ & \ddots & & & \ddots & & & \\ & & \ddots & & & \ddots & & \\ & & & mt_m & \cdots & 2t_2 & t_1 & \end{bmatrix}.$$

Since z appears in $\hat{P}_{z,k}(\lambda)$ only as coefficient of λ^0, expanding the determinant we see

$$\det[\mathrm{Disc}(\hat{P}_{z,k})] = (\hat{\mathfrak{d}}_N t_{\hat{\mathfrak{d}}_N}(k))^{\hat{\mathfrak{d}}_N} \cdot z^{\hat{\mathfrak{d}}_N - 1} + P_{\hat{\mathfrak{d}}_N - 2}(x) z^{\hat{\mathfrak{d}}_N - 2} + \cdots + P_0(x),$$

for some continuous functions $\{P_\ell(\cdot)\}_{\ell=0}^{\hat{\mathfrak{d}}_N - 2}$. Let $\{z_\ell(k)\}_{\ell=1}^{\hat{\mathfrak{d}}_N - 1}$ be the roots of the equation $\det[\mathrm{Disc}(\hat{P}_{z,k})] = 0$. Hence,

$$|D(\hat{P}_{z,k})| = (\hat{\mathfrak{d}}_N t_{\hat{\mathfrak{d}}_N}(k))^{\hat{\mathfrak{d}}_N} \prod_{\ell=1}^{\hat{\mathfrak{d}}_N - 1} |z - z_\ell(k)|.$$

Therefore, recalling (5.6) we obtain that

$$\begin{aligned} |t_{\hat{\mathfrak{d}}_N}(k)|^{2\hat{\mathfrak{d}}_N - 1} |\det(\boldsymbol{V}(k))|^2 &= |t_{\hat{\mathfrak{d}}_N}(k)|^{2\hat{\mathfrak{d}}_N - 1} \prod_{1 \leqslant \ell < \ell' \leqslant \hat{\mathfrak{d}}_N(k)} |\hat{\lambda}_\ell(z,k) - \hat{\lambda}_{\ell'}(z,k)|^2 \\ &= |D(\hat{P}_{z,k})| = |\hat{\mathfrak{d}}_N t_{\hat{\mathfrak{d}}_N}(k)|^{\hat{\mathfrak{d}}_N} \prod_{\ell=1}^{\hat{\mathfrak{d}}_N - 1} |z - z_\ell(k)|. \end{aligned}$$

Thus the set of all z's satisfying property (i) is contained in

$$\bigcup_{k=0}^{L_0} \bigcup_{\ell=1}^{\hat{\mathfrak{d}}_N - 1} B_{\mathbb{C}}(z_\ell(k), \hat{\mathfrak{d}}_N^{-1} N^{-2\delta_1}),$$

whose Lebesgue measure is bounded above by $N^{-3\delta_1}$. Taking a union on $O(N^{\delta_1})$ possible k-s we find that the Lebesgue measure of the set of z satisfying property (i) is bounded by $\frac{1}{3} N^{-\delta_1}$. To complete the proof it remains to prove the same for the set of z satisfying property (iii). This follows from a volumetric argument. Indeed, recalling the definition of $t_0(\cdot)$ we find that

$$\inf_{k=0}^{L_0} |z - t_0(k)| \geqslant \min\{\mathrm{dist}(z, f_0([0,1])), |z|\}.$$

Using the fact that f_0 is an α_0-Hölder-continuous function from the triangle inequality we obtain that

$$dist(z, f_0([0,1])) \leqslant N^{-\varepsilon} \Rightarrow z \in \bigcup_{k=0}^{\lceil N^{\varepsilon/\alpha_0} \rceil} B_{\mathbb{C}}(f_0(k/N^{\varepsilon/\alpha_0}), 2N^{-\varepsilon})$$

for $\varepsilon > 0$. As $\alpha_0 > \frac{1}{2}$, setting $\varepsilon = (\alpha_0\delta_1/2(2\alpha_0 - 1))$ it yields that

$$Leb(\{z : dist(z, f_0([0,1])) \leqslant N^{-\varepsilon}\} \cup B_{\mathbb{C}}(0, N^{-\varepsilon})) \leqslant \tfrac{1}{3}N^{-\delta_1}.$$

This completes the proof. □

Next, building on Lemma 5.3(ii) we show that for $z \notin \mathcal{B}_N$ and vectors w not belonging to the span of $\{\boldsymbol{w}_\ell^j(k)\}$, the ℓ_2-norm of $\hat{M}_N(z)w$ cannot be too small. This yields a bound on the number of small singular values of $\hat{M}_N(z)$.

Fix $k \in [L_0] \cup \{0\}$ and $j' \in [L'_k]$. This fixes some $j \in [L] \cup \{0\}$ such that $\mathfrak{b}_k^{(j')} = [i_{j+1}] \setminus [i_j]$. Denote $\mathcal{S}_{k,j} := \text{span}(\boldsymbol{w}_\ell^j(k) : \ell \in [\hat{\mathfrak{d}}_N(k)])$ and let $\psi_{k,j}$ be the orthogonal projection onto $\mathcal{S}_{k,j}$. Further let $\pi_{k,j}$ and $\rho_{k,j}$ be the projections onto the span of $\{e_m\}_{m=i_j+1}^{i_{j+1}}$ and $\{e_m\}_{k=i_j+1}^{i_{j+1}-\hat{\mathfrak{d}}_N(k)}$, respectively, where e_m is the mth canonical basic vector. When needed, we will view $\pi_{k,j}$, $\rho_{k,j}$, and $\psi_{k,j}$ as projection matrices of appropriate dimensions.

LEMMA 5.5. *Fix $R < \infty$ and $z \in B_{\mathbb{C}}(0,R) \setminus \mathcal{B}_N$. Let $k \in [L_0]$ and $j' \in [L'_k]$ such that $\mathfrak{b}_k^{(j')} = [i_{j+1}] \setminus [i_j]$ for some $j \in [L] \cup \{0\}$. Then there exists a positive finite constant $C_1(R, \mathfrak{d}, \boldsymbol{f})$, depending only on $R, \mathfrak{d}$, and $\max_\ell \sup_{x\in[0,1]} |f_\ell(x)|$, such that for any $w \in \mathbb{C}^N$, we have*

$$\|\pi_{k,j}(w - \psi_{k,j}w)\|_2 \leqslant C_1(R, \mathfrak{d}, \boldsymbol{f}) N^{2\mathfrak{d}\delta_1 + \mathfrak{d}^2\delta_2 + 2\delta_3 + (\alpha_0\delta_1)/(2(2\alpha_0-1))} \|\rho_{k,j}\hat{M}_N(z)w\|_2. \tag{5.7}$$

Note that Lemma 5.5 is similar to Lemma 3.10. Analogous to the proof of Lemma 3.10, here also the proof proceeds by identifying a vector $y \in \mathcal{S}_{k,j}$ and showing that $\|\pi_{k,j}(x - y)\|_2$ satisfies the bound (5.7). For $\mathfrak{d} > 1$, the choice of an appropriate vector y is significantly more difficult and requires new ideas.

Proof of Lemma 5.5. We write $\hat{\lambda}_\ell := \hat{\lambda}_\ell(z,k)$ and $\hat{\mathfrak{d}}_N := \hat{\mathfrak{d}}_N(k)$. First let us consider the case $\hat{\mathfrak{d}}_N = 0$. This implies that $\mathcal{S}_{k,j} = \emptyset$. Therefore, $\psi_{k,j} = 0$ and $\rho_{k,j} = \pi_{k,j}$. So, it is enough to show that

$$\|\pi_{k,j}w\|_2 \leqslant N^{(\alpha_0\delta_1)/(2(2\alpha_0-1))} \|\pi_{k,j}\hat{M}_N(z)w\|_2. \tag{5.8}$$

https://doi.org/10.1017/fms.2018.29 Published online by Cambridge University Press

Since $\hat{\mathfrak{d}}_N = 0$, from (5.2) we further have that

$$\pi_{k,j}\hat{M}_N(z)w = (t_0(k) - z)\pi_{k,j}w.$$

Using the fact that $z \notin \mathcal{B}_N$ we have that $\inf_k |t_0(k) - z| \geqslant N^{-(\alpha_0\delta_1/2(2\alpha_0-1))}$. This yields (5.8) and hence (5.7) is established for $\hat{\mathfrak{d}}_N = 0$. It remains to prove the same when $\hat{\mathfrak{d}}_N > 0$.

Without loss of generality, we assume that $\{\hat{\lambda}_\ell\}_{\ell=1}^{\hat{\mathfrak{d}}_N}$ are arranged in decreasing order of moduli, and define $\mathfrak{d}_0$ so that $|\hat{\lambda}_1| \geqslant |\hat{\lambda}_2| \geqslant \cdots \geqslant |\hat{\lambda}_{\mathfrak{d}_0}| \geqslant 1 > |\hat{\lambda}_{\mathfrak{d}_0+1}| \geqslant \cdots \geqslant |\hat{\lambda}_{\hat{\mathfrak{d}}_N}|$, with $\mathfrak{d}_0 = \hat{\mathfrak{d}}_N$ if all $\hat{\lambda}_\ell \geqslant 1$ and $\mathfrak{d}_0 = 0$ if all $\hat{\lambda}_\ell < 1$. Define a $(2\hat{\mathfrak{d}}_N) \times (2\hat{\mathfrak{d}}_N)$ matrix $\mathsf{L} := [\mathsf{L}_1 : \mathsf{L}_2]$ where

$$\mathsf{L} :_1= \begin{bmatrix} \boldsymbol{v}_1(k) & \boldsymbol{v}_2(k) & \cdots & \boldsymbol{v}_{\hat{\mathfrak{d}}_N}(k) \\ \hat{\lambda}_1^{b_j-\hat{\mathfrak{d}}_N}\boldsymbol{v}_1(k) & \hat{\lambda}_2^{b_j-\hat{\mathfrak{d}}_N(k)}\boldsymbol{v}_2(k) & \cdots & \hat{\lambda}_{\hat{\mathfrak{d}}_N}^{b_j-\hat{\mathfrak{d}}_N}\boldsymbol{v}_{\hat{\mathfrak{d}}_N}(k) \end{bmatrix},$$

$$\mathsf{L}_2 := \begin{bmatrix} \boldsymbol{v}_1(k) \cdots \boldsymbol{v}_{\mathfrak{d}_0}(k) & 0 & \cdots & 0 \\ 0 \quad \cdots \quad 0 & \boldsymbol{v}_{\mathfrak{d}_0+1}(k) & \cdots & \boldsymbol{v}_{\hat{\mathfrak{d}}_N}(k) \end{bmatrix},$$

$\{\boldsymbol{v}_\ell(k)\}_{\ell=1}^{\hat{\mathfrak{d}}_N}$ are as in Lemma 5.1, and $b_j := i_{j+1} - i_j$. Since $z \in B_{\mathbb{C}}(0, R) \setminus \mathcal{B}_N$, the eigenvalues $\{\hat{\lambda}_\ell\}_{\ell=1}^{\hat{\mathfrak{d}}_N}$ are all distinct, and hence the vectors $\{\boldsymbol{v}_\ell(k)\}_{\ell=1}^{\hat{\mathfrak{d}}_N}$ are linearly independent. Therefore, $\mathrm{rank}(\mathsf{L}) = 2\hat{\mathfrak{d}}_N$, and the system of linear equations

$$\begin{pmatrix} w[\![k, i_j + 1]\!] \\ w[\![k, i_{j+1} - \hat{\mathfrak{d}}_N + 1]\!] \end{pmatrix} = \mathsf{L} \begin{pmatrix} a_1 \\ \vdots \\ a_{\hat{\mathfrak{d}}_N} \\ a_1' \\ \vdots \\ a_{\hat{\mathfrak{d}}_N}' \end{pmatrix} \tag{5.9}$$

admits a unique solution. Set

$$y := \sum_{\ell=1}^{\hat{\mathfrak{d}}_N} a_\ell \boldsymbol{w}_\ell^j(k) \quad \text{and} \quad \zeta := \zeta(w) := w - y.$$

With this choice of ζ we will show that

$$\|\pi_{k,j}\zeta\|_2 \leqslant C_1(R, \mathfrak{d}, \boldsymbol{f})N^{2\mathfrak{d}\delta_1+\mathfrak{d}^2\delta_2+2\delta_3}\|\rho_{k,j}\hat{M}_N(z)w\|_2, \tag{5.10}$$

for some constant $C_1(R, \mathfrak{d}, \boldsymbol{f})$. This will complete the proof. Indeed, from the definition of the projection operator $\psi_{k,j}$ it follows that

$$\|w - \psi_{k,j}w\|_2 \leqslant \|\zeta\|_2. \tag{5.11}$$

https://doi.org/10.1017/fms.2018.29 Published online by Cambridge University Press

On the other hand, we note that $\mathcal{S}_{k,j} \subset \operatorname{span}(\{e_m\}_{m=i_j+1}^{i_j+1})$. Therefore $\pi_{k,j}\psi_{k,j} = \psi_{k,j}$. Recalling that $y \in \mathcal{S}_{k,j}$, we have

$$\|w - \psi_{k,j}w\|_2^2 = \|(\mathrm{Id} - \pi_{k,j})w\|_2^2 + \|\pi_{k,j}(w - \psi_{k,j}w)\|_2^2$$

and

$$\|\zeta\|_2^2 = \|(\mathrm{Id} - \pi_{k,j})\zeta\|_2^2 + \|\pi_{k,j}\zeta\|_2^2 = \|(\mathrm{Id} - \pi_{k,j})w\|_2^2 + \|\pi_{k,j}\zeta\|_2^2.$$

Thus from (5.11) we obtain

$$\|\pi_{k,j}(w - \psi_{k,j}w)\|_2 \leqslant \|\pi_{k,j}\zeta\|_2,$$

and so it is enough to show that (5.10) holds.

We turn now to the proof of (5.10). Recalling that $y \in \mathcal{S}_{k,j}$, the span of $\{\boldsymbol{w}_\ell^j(k)\}_{\ell\in[\hat{\mathfrak{d}}]}$, an application of Lemma 5.3(ii) implies that $\rho_{k,j}\hat{M}_N(z)y = 0$. So, using (5.2) and recalling the definition of $T_k(z)$ we see that

$$\zeta[\![k, m+1]\!] = T_k(z)\zeta[\![k, m]\!] + \begin{pmatrix} \frac{1}{t_{\hat{\mathfrak{d}}_N}(k)}(\hat{M}_N(z)w)_m \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad m \in [i_{j+1} - \hat{\mathfrak{d}}_N] \setminus [i_j]. \tag{5.12}$$

From the linear independence of $\{\boldsymbol{v}_\ell(k)\}_1^{\hat{\mathfrak{d}}_N}$, there are $\{\tilde{a}_\ell\}_{\ell=1}^{\hat{\mathfrak{d}}_N}$ so that

$$\begin{pmatrix}1 \ 0 \cdots 0\end{pmatrix} = \sum_{\ell=1}^{\hat{\mathfrak{d}}_N} \tilde{a}_\ell \boldsymbol{v}_\ell(k)^{\mathsf{T}}. \tag{5.13}$$

Hence denoting $\beta_m := (t_{\hat{\mathfrak{d}}_N}(k))^{-1}(\hat{M}_N(z)w)_m$ we observe that (5.12) simplifies to

$$\zeta[\![k, m+1]\!] = T_k(z)\zeta[\![k, m]\!] + \beta_m \sum_{\ell=1}^{\hat{\mathfrak{d}}_N} \tilde{a}_\ell \boldsymbol{v}_\ell(k), \quad m \in [i_{j+1} - \hat{\mathfrak{d}}_N] \setminus [i_j + 1]. \tag{5.14}$$

Iterating (5.14) we obtain

$$\begin{aligned} \zeta[\![k, m+1]\!] &= T_k(z)^{m-i_j}\zeta[\![k, i_j+1]\!] + \sum_{m'=i_j+1}^{m} \beta_{m'} T_k(z)^{m-m'}\left[\sum_{\ell=1}^{\hat{\mathfrak{d}}_N} \tilde{a}_\ell \boldsymbol{v}_\ell(k)\right] \\ &= T_k(z)^{m-i_j}\zeta[\![k, i_j+1]\!] + \sum_{\ell=1}^{\hat{\mathfrak{d}}_N} \tilde{a}_\ell \left[\sum_{m'=i_j+1}^{m} \beta_{m'} \hat{\lambda}_\ell^{m-m'}\right] \boldsymbol{v}_\ell(k), \end{aligned} \tag{5.15}$$

https://doi.org/10.1017/fms.2018.29 Published online by Cambridge University Press

for $m \in [i_{j+1} - \hat{\mathfrak{d}}_N] \setminus [i_j + 1]$, where the last step follows from the fact that $\boldsymbol{v}_\ell(k)$ is an eigenvector of $T_k(z)$ corresponding to $\hat{\lambda}_\ell$ (see Lemma 5.1). Recalling that $b_j := i_{j+1} - i_j$ we note that (5.15) in particular implies

$$\zeta[\![k, i_{j+1}-\hat{\mathfrak{d}}_N+1]\!] = T_k(z)^{b_j-\hat{\mathfrak{d}}_N}\zeta[\![k, i_j+1]\!] + \sum_{\ell=1}^{\hat{\mathfrak{d}}_N} \tilde{a}_\ell \left[\sum_{m'=i_j+1}^{i_{j+1}-\hat{\mathfrak{d}}_N} \beta_{m'}\hat{\lambda}_\ell^{i_{j+1}-\hat{\mathfrak{d}}_N-m'}\right]\boldsymbol{v}_\ell(k). \tag{5.16}$$

Now recalling the definitions of $\{\boldsymbol{w}_\ell^j(k)\}_{\ell\in[\hat{\mathfrak{d}}_N]}$ we see that

$$\boldsymbol{w}_\ell^j(k)[\![k, i_j + 1]\!] = \boldsymbol{v}_\ell(k) \quad \text{and} \quad \boldsymbol{w}_\ell^j(k)[\![k, i_{j+1} - \hat{\mathfrak{d}}_N + 1]\!] = \hat{\lambda}_\ell^{b_j-\hat{\mathfrak{d}}_N}\boldsymbol{v}_\ell(k).$$

Since $\zeta = w - \sum_{\ell=1}^{\hat{\mathfrak{d}}_N} a_\ell \boldsymbol{w}_\ell^j(k)$, from (5.9) we obtain

$$\zeta[\![k, i_j + 1]\!] = \sum_{\ell=1}^{\mathfrak{d}_0} a'_\ell \boldsymbol{v}_\ell(k) \quad \text{and} \quad \zeta[\![k, i_{j+1} - \hat{\mathfrak{d}}_N + 1]\!] = \sum_{\ell=\mathfrak{d}_0+1}^{\hat{\mathfrak{d}}_N} a'_\ell \boldsymbol{v}_\ell(k). \tag{5.17}$$

Plugging these in (5.16) we deduce

$$\sum_{\ell=1}^{\mathfrak{d}_0}\left(\tilde{a}_\ell\left[\sum_{m'=i_j+1}^{i_{j+1}-\hat{\mathfrak{d}}_N}\beta_{m'}\hat{\lambda}_\ell^{i_{j+1}-\hat{\mathfrak{d}}_N-m'}\right] + a'_\ell\hat{\lambda}_\ell^{b_j-\hat{\mathfrak{d}}_N}\right)\boldsymbol{v}_\ell(k)$$
$$+ \sum_{\ell=\mathfrak{d}_0+1}^{\hat{\mathfrak{d}}_N}\left(\tilde{a}_\ell\left[\sum_{m'=i_j+1}^{i_{j+1}-\hat{\mathfrak{d}}_N}\beta_{m'}\hat{\lambda}_\ell^{i_{j+1}-\hat{\mathfrak{d}}_N-m'}\right] - a'_\ell\right)\boldsymbol{v}_\ell(k) = 0.$$

Since $\{\boldsymbol{v}_\ell(k)\}_{\ell=1}^{\hat{\mathfrak{d}}_N}$ are linearly independent vectors it further implies that

$$a'_\ell\hat{\lambda}_\ell^{b_j-\hat{\mathfrak{d}}_N} + \tilde{a}_\ell\left[\sum_{m'=i_j+1}^{i_{j+1}-\hat{\mathfrak{d}}_N}\beta_{m'}\hat{\lambda}_\ell^{i_{j+1}-\hat{\mathfrak{d}}_N-m'}\right] = 0, \quad \ell \in [\mathfrak{d}_0];$$
$$a'_\ell = \tilde{a}_\ell\left[\sum_{m'=i_j+1}^{i_{j+1}-\hat{\mathfrak{d}}_N}\beta_{m'}\hat{\lambda}_\ell^{i_{j+1}-\hat{\mathfrak{d}}_N-m'}\right], \quad \ell \in [\hat{\mathfrak{d}}_N] \setminus [\mathfrak{d}_0]. \tag{5.18}$$

https://doi.org/10.1017/fms.2018.29 Published online by Cambridge University Press

Thus from (5.15) and (5.17), using (5.18), we further obtain that for any $m \in [i_{j+1} - \hat{\mathfrak{d}}_N] \setminus [i_j]$,

$$\begin{aligned}
\zeta[\![k, m+1]\!] &= T_k(z)^{m-i_j}\zeta[\![k, i_j+1]\!] + \sum_{\ell=1}^{\hat{\mathfrak{d}}_N} \tilde{a}_\ell \Bigg[\sum_{m'=i_j+1}^{m} \beta_{m'}\hat{\lambda}_\ell^{m-m'}\Bigg] \boldsymbol{v}_\ell(k) \\
&= \sum_{\ell=1}^{\mathfrak{d}_0} \left(a'_\ell \hat{\lambda}_\ell^{m-i_j} + \tilde{a}_\ell \Bigg[\sum_{m'=i_j+1}^{m} \beta_{m'}\hat{\lambda}_\ell^{m-m'}\Bigg]\right) \boldsymbol{v}_\ell(k) \\
&\quad + \sum_{\ell=\mathfrak{d}_0+1}^{\hat{\mathfrak{d}}_N} \tilde{a}_\ell \Bigg[\sum_{m'=i_j+1}^{m} \beta_{m'}\hat{\lambda}_\ell^{m-m'}\Bigg] \boldsymbol{v}_\ell(k) \\
&= -\sum_{\ell=1}^{\mathfrak{d}_0} \tilde{a}_\ell \Bigg[\sum_{m'=m+1}^{i_{j+1}-\hat{\mathfrak{d}}_N} \beta_{m'}\hat{\lambda}_\ell^{m-m'}\Bigg] \boldsymbol{v}_\ell(k) + \sum_{\ell=\mathfrak{d}_0+1}^{\hat{\mathfrak{d}}_N} \tilde{a}_\ell \Bigg[\sum_{m'=i_j+1}^{m} \beta_{m'}\hat{\lambda}_\ell^{m-m'}\Bigg] \boldsymbol{v}_\ell(k).
\end{aligned} \tag{5.19}$$

Since the first coordinate of $\zeta[\![k, m+1]\!]$ is $\zeta_{m+\hat{\mathfrak{d}}_N}$, $|\hat{\lambda}_\ell| \geqslant 1$ for $\ell \in [\mathfrak{d}_0]$, and $|\hat{\lambda}_\ell| \leqslant 1$ for $\ell \in [\hat{\mathfrak{d}}_N] \setminus [\mathfrak{d}_0]$, using the triangle inequality from (5.19) we see that

$$\begin{aligned}
|\zeta_{m+\hat{\mathfrak{d}}_N}| &\leqslant \sum_{\ell=1}^{\mathfrak{d}_0} |\tilde{a}_\ell||\hat{\lambda}_\ell|^{\hat{\mathfrak{d}}_N-1} \sum_{m'=m+1}^{i_{j+1}-\hat{\mathfrak{d}}_N} |\beta_{m'}| + \sum_{\ell=\mathfrak{d}_0+1}^{\hat{\mathfrak{d}}_N} |\tilde{a}_\ell||\hat{\lambda}_\ell|^{\hat{\mathfrak{d}}_N-1} \sum_{m'=i_j+1}^{m} |\beta_{m'}| \\
&\leqslant \|T_k(z)\|^{\hat{\mathfrak{d}}_N-1} \left(\sum_{m'=i_j+1}^{i_{j+1}-\hat{\mathfrak{d}}_N} |\beta_{m'}| \right) \cdot \left(\sum_{\ell=1}^{\hat{\mathfrak{d}}_N} |\tilde{a}_\ell| \right) \quad \text{for } m \in [i_{j+1} - \hat{\mathfrak{d}}_N] \setminus [i_j].
\end{aligned} \tag{5.20}$$

From (5.17) and (5.18) it also follows that

$$\zeta[\![k, i_j+1]\!] = \sum_{\ell=1}^{\mathfrak{d}_0} a'_\ell \boldsymbol{v}_\ell(k) = -\sum_{\ell=1}^{\mathfrak{d}_0} \tilde{a}_\ell \Bigg[\sum_{m'=i_j+1}^{i_{j+1}-\hat{\mathfrak{d}}_N} \beta_{m'}\hat{\lambda}_\ell^{i_j-m'}\Bigg] \boldsymbol{v}_\ell(k).$$

Thus

$$\begin{aligned}
\max_{m=i_j+1}^{i_j+\hat{\mathfrak{d}}} |\zeta(m)| &\leqslant \|\zeta[\![k, i_j+1]\!]\|_2 \leqslant \sum_{\ell=1}^{\mathfrak{d}_0} |\tilde{a}_\ell| \cdot \left(\sum_{m'=i_j+1}^{i_{j+1}-\hat{\mathfrak{d}}_N} |\beta_{m'}| \right) \cdot \|\boldsymbol{v}_\ell(k)\|_2 \\
&\leqslant \sqrt{\mathfrak{d}} \cdot (\|T_k(z)\|^{\mathfrak{d}-1} \vee 1) \cdot \left(\sum_{m'=i_j+1}^{i_{j+1}-\hat{\mathfrak{d}}_N} |\beta_{m'}| \right) \cdot \left(\sum_{\ell=1}^{\hat{\mathfrak{d}}_N} |\tilde{a}_\ell| \right),
\end{aligned} \tag{5.21}$$

where we have used the fact that

$$\|\boldsymbol{v}_\ell(k)\|_2 \leqslant \sqrt{\mathfrak{d}} \cdot (|\hat{\lambda}_\ell|^{\mathfrak{d}-1} \vee 1).$$

Hence, combining (5.20)–(5.21) we obtain

$$\|\pi_{k,j}(\zeta)\|_2 \leqslant b_j \max_{m=i_j+1}^{i_{j+1}} |\zeta_j| \leqslant 2N^{\delta_3}\sqrt{\mathfrak{d}} \cdot (\|T_k(z)\|^{\mathfrak{d}-1} \vee 1)$$
$$\cdot \left(\sum_{m=i_j+1}^{i_{j+1}-\hat{\mathfrak{d}}_N} |\beta_m| \right) \cdot \left(\sum_{\ell=1}^{\hat{\mathfrak{d}}_N} |\tilde{a}_\ell| \right). \tag{5.22}$$

Now to complete the proof we need to find a bound on $\sum_{\ell=1}^{\hat{\mathfrak{d}}_N} |\tilde{a}_\ell|$. To this end, recall that $(\tilde{a}_1, \ldots, \tilde{a}_{\hat{\mathfrak{d}}_N})$ satisfies the system of linear equations (5.13). Using Cramer's rule it is easy to check that $\tilde{a}_\ell = \prod_{\ell' \neq \ell} (\hat{\lambda}_{\ell'})/(\hat{\lambda}_{\ell'} - \hat{\lambda}_\ell)$. Therefore, recalling that $\boldsymbol{V}(k) := [\boldsymbol{v}_1(k)\ \boldsymbol{v}_2(k) \cdots \boldsymbol{v}_{\hat{\mathfrak{d}}_N}(k)]$,

$$|\tilde{a}_\ell| = \prod_{\ell' \neq \ell} \left| \frac{\hat{\lambda}_{\ell'}}{\hat{\lambda}_{\ell'} - \hat{\lambda}_\ell} \right| \leqslant (\|T_k(z)\|^{\mathfrak{d}-1} \vee 1) \cdot \prod_{\ell' \neq \ell} \left| \frac{1}{\hat{\lambda}_{\ell'} - \hat{\lambda}_\ell} \right|$$
$$\leqslant \left((2\|T_k(z)\|)^{((\mathfrak{d}(\mathfrak{d}-1))/2)} \vee 1 \right) \cdot |\det(\boldsymbol{V}(k))|^{-1}$$
$$\leqslant \left((2\|T_k(z)\|)^{\mathfrak{d}(\mathfrak{d}-1)} \vee 1 \right) \cdot |\det(\boldsymbol{V}(k))|^{-2}.$$

Recalling that $\beta_m = (t_{\hat{\mathfrak{d}}_N}(k))^{-1}(\hat{M}_N(z)w)_m$, an application of Cauchy–Schwarz yields

$$\|\pi_{k,j}(\zeta)\|_2 \leqslant 2^{\mathfrak{d}^2}\mathfrak{d}^{3/2}N^{3\delta_3/2} \cdot (\|T_k(z)\|^{\mathfrak{d}^2} \vee 1) \cdot \|\rho_{k,j}\hat{M}_N(z)w\|_2$$
$$\cdot |t_{\hat{\mathfrak{d}}_N}(k)|^{-1}|\det(\boldsymbol{V}(k))|^{-2}. \tag{5.23}$$

Since $z \in \mathcal{B}_N^c$ (cf. Lemma 5.4),

$$|t_{\hat{\mathfrak{d}}_N}(k)| \cdot |\det(\boldsymbol{V}(k))|^2 \geqslant \frac{|t_{\hat{\mathfrak{d}}_N}(k)|^{\hat{\mathfrak{d}}_N(k)-1}|\det(\boldsymbol{V}(k))|^2}{\sup_{x\in[0,1]} |f_{\hat{\mathfrak{d}}_N}(x)|^{\hat{\mathfrak{d}}_N-2}} \geqslant C_0^{-1}N^{-2\delta_1\mathfrak{d}},$$

for some $C_0 < \infty$. By the Gershgorin circle theorem we also have that

$$\|T_k(z)\|^2 = \|T_k(z)^*T_k(z)\| = O((t_{\hat{\mathfrak{d}}_N}(k))^{-2}) = O(N^{2\delta_2}), \tag{5.24}$$

where the last step follows from the fact that the nonzero entries of $\hat{M}_N$ are bounded below by $N^{-\delta_2}$. Therefore, plugging the last two bounds in (5.23) we arrive at (5.10). □

Denote

$$\mathfrak{L} := \sum_{k=0}^{L_0} L'_k \cdot \hat{\mathfrak{d}}_N(k) = O(\mathfrak{d} N^{1-\delta_3}). \tag{5.25}$$

Building on Lemma 5.5 we now prove a lower bound on the $(\mathfrak{L}+1)$-st smallest singular value of $\hat{M}_N(z)$. First we prove an estimate that will also be useful in obtaining a lower bound on the product of the small singular values of $\hat{M}_N(z)$. To state it, we let $\psi := \sum \psi_{k,j}$, that is, ψ is the orthogonal projection operator onto the space spanned by $\cup \mathcal{S}_{k,j}$, and $\rho := \sum \rho_{k,j}$.

LEMMA 5.6. *Fix $R < \infty$ and $z \in B_{\mathbb{C}}(0, R) \setminus \mathcal{B}_N$. Then for any vector $w \in \mathbb{C}^N$ we have*

$$\|w - \psi w\|_2 \leqslant C_1(R, \mathfrak{d}, f) N^{2\mathfrak{d}\delta_1 + \mathfrak{d}^2\delta_2 + 2\delta_3 + (\alpha_0\delta_1)/(2(2\alpha_0 - 1))} \|\rho \hat{M}_N(z) w\|_2,$$

where $C_1(R, \mathfrak{d}, f)$ is as in Lemma 5.5.

Proof. The proof is a simple application of Lemma 5.5. Since $\sum \pi_{k,j} = 1$, $\pi_{k,j}\psi_{k,j} = \psi_{k,j}$, and $\{\pi_{k,j}\}$ are orthogonal it follows that

$$\|w - \psi w\|_2^2 = \left\| \sum (\pi_{k,j} w - \psi_{k,j} w) \right\|_2^2 = \sum \|\pi_{k,j}(w - \psi_{k,j} w)\|_2^2.$$

Now the result follows from Lemma 5.5 upon noting the fact that $\{\rho_{k,j}\}$ are orthogonal. □

From Lemma 5.6 we immediately obtain the following corollary.

COROLLARY 5.7. *Fix $R < \infty$ and $z \in B_{\mathbb{C}}(0, R) \setminus \mathcal{B}_N$. Then*

$$\sigma_{N-\mathfrak{L}}(\hat{M}_N(z)) \geqslant C_1(R, \mathfrak{d}, f)^{-1} N^{-(2\mathfrak{d}\delta_1 + \mathfrak{d}^2\delta_2 + 2\delta_3 + (\alpha_0\delta_1)/(2(2\alpha_0 - 1)))},$$

where $C_1(R, \mathfrak{d}, f)$ is as in Lemma 5.5.

The proof of Corollary 5.7 is similar to that of Corollary 3.13, and hence omitted.

REMARK 5.8. Let $\{\delta_i\}_{i=1}^3$ be such that

$$\max \left\{ \delta_1, \delta_2, \delta_3, \frac{\alpha_0\delta_1}{2(2\alpha_0 - 1)} \right\} \leqslant \frac{1}{40\mathfrak{d}^2} \cdot \left(\gamma - \frac{1}{2} \right) \quad \text{and} \quad \delta_1 < \frac{\delta_3}{4}. \tag{5.26}$$

It follows from Corollary 5.7 that there are only at most $\mathfrak{L} = O(N^{1-\delta_3})$ singular values of $\hat{M}_N(z)$ that are $O(N^{-(1/3)(\gamma - 1/2)})$, which upon choosing

https://doi.org/10.1017/fms.2018.29 Published online by Cambridge University Press

$\varepsilon_N = N^{-(1/3)(\gamma-1/2)}$ in (2.4), implies that $N^* \leqslant \mathfrak{L} = o(N/\log N)$. This verifies that the number of small singular values of $\hat{M}_N(z)$, for $z \in B_{\mathbb{C}}(0, R) \setminus \mathcal{B}_N$ is as desired. In the remainder of this paper we will work with $\{\delta_i\}_{i=1}^3$ satisfying (5.26).

Equipped with Remark 5.8 we note that it remains to find matching upper and lower bounds, up to subexponential factors, on the product of small singular values. In the context of Theorem 3.1 the upper bound on the product of the small singular values was achieved by finding a collection of orthonormal vectors which were approximate singular vectors, and then appealing to Lemma A.2. In the current setup, one notes that the approximate singular vectors, in particular $\{\boldsymbol{w}_\ell^j(k)\}_{\ell=1}^{\hat{\mathfrak{d}}_N}$ for any j and k, are not orthogonal. Therefore we need to work with an orthonormal basis of $\mathcal{S}_{k,j}$. To this end, a key step will be to obtain bounds on the determinant of $\mathfrak{U}_{k,j}^* \mathfrak{U}_{k,j}$ where $\mathfrak{U}_{k,j} := \pi_{k,j} \hat{M}_N(z) U_{k,j}$, and the columns of $U_{k,j}$ are an orthonormal basis of $\mathcal{S}_{k,j}$. We start with bounds on $\det(\mathfrak{W}_{k,j})$ where $\mathfrak{W}_{k,j} := W_{k,j}^* W_{k,j}$, and $W_{k,j}$ is the matrix whose columns are $\{\boldsymbol{w}_\ell^j(k)\}_{\ell=1}^{\hat{\mathfrak{d}}_N(k)}$.

LEMMA 5.9. *Fix $R < \infty$ and $z \in B_{\mathbb{C}}(0, R) \setminus \mathcal{B}_N$. Let $k \in [L_0]$ and $j' \in [L'_k]$ such that $\mathfrak{b}_k^{(j')} = [i_{j+1}] \setminus [i_j]$ for some $j \in [L] \cup \{0\}$. Assume $\hat{\mathfrak{d}}_N(k) > 0$. Then*

$$C_2(R, \mathfrak{d}, \boldsymbol{f})^{-1} N^{-2\mathfrak{d}\delta_1 - \mathfrak{d}^2\delta_2} \prod_{\ell=1}^{\hat{\mathfrak{d}}_N(k)} \left(|\hat{\lambda}_\ell(z, k)| \vee 1\right)^{2b_j} \leqslant \det(\mathfrak{W}_{k,j})$$
$$\leqslant C_2(R, \mathfrak{d}, \boldsymbol{f}) N^{3\delta_1\mathfrak{d}} \prod_{\ell=1}^{\hat{\mathfrak{d}}_N(k)} \left(|\hat{\lambda}_\ell(z, k)| \vee 1\right)^{2b_j},$$

for all large N, uniformly over $k \in [L_0]$ and $j' \in [L'_k]$, where $C_2(R, \mathfrak{d}, \boldsymbol{f})$ is some positive finite constant depending only on $\mathfrak{d}$, R, and $\{f_\ell(\cdot)\}_{\ell=0}^{\mathfrak{d}}$.

Proof. Throughout the proof, for ease of writing, we write $\hat{\mathfrak{d}}_N$ and $\{\hat{\lambda}_\ell\}_{\ell=1}^{\hat{\mathfrak{d}}_N}$ instead of $\hat{\mathfrak{d}}_N(k)$ and $\{\hat{\lambda}_\ell(z, k)\}_{\ell=1}^{\hat{\mathfrak{d}}_N(k)}$.

We first derive the upper bound. Using Hadamard's inequality we observe that

$$\det(\mathfrak{W}_{k,j}) \leqslant \prod_{\ell=1}^{\hat{\mathfrak{d}}_N} (\mathfrak{W}_{k,j})_{\ell,\ell} = \prod_{\ell=1}^{\hat{\mathfrak{d}}_N} \|\boldsymbol{w}_\ell^j(k)\|_2^2. \tag{5.27}$$

We note that $\|\boldsymbol{w}_\ell^j(k)\|_2^2 = \sum_{m=0}^{b_j-1} |\hat{\lambda}_\ell|^{2m}$, where we recall that $b_j = i_{j+1} - i_j \leqslant 2N^{\delta_3}$. Therefore, using the fact that $z \notin \mathcal{B}_N$, which in turn implies that $|\hat{\lambda}_\ell| \neq 1$, we obtain

$$\|\boldsymbol{w}_\ell^j(k)\|_2^2 \leqslant |1 - |\hat{\lambda}_\ell|^2|^{-1} \cdot [2|\hat{\lambda}_\ell|^{2b_j} \mathbf{1}(|\hat{\lambda}_\ell| > 1) + 2\mathbf{1}(|\hat{\lambda}_\ell| < 1)]. \tag{5.28}$$

Since $z \notin \mathcal{B}_N$, the desired upper bound follows from (5.27)–(5.28).

https://doi.org/10.1017/fms.2018.29 Published online by Cambridge University Press

For the lower bound, we apply the Cauchy–Binet formula, which gives

$$\det(\mathfrak{W}_{k,j}) = \sum_{\substack{S\subset[N]\\ |S|=\hat{\mathfrak{d}}_N}} |\det(W_{k,j}[S])|^2, \tag{5.29}$$

where $W_{k,j}[S]$ is the $\hat{\mathfrak{d}}_N \times \hat{\mathfrak{d}}_N$ square submatrix of $W_{k,j}$ with rows indexed by S. Hence for a lower bound, we may pick any S and bound $\det(\mathfrak{W}_{k,j}) \geqslant |\det(W_{k,j}[S])|^2$. Take

$$S = ([\hat{\mathfrak{d}}_N - \mathfrak{d}_0] + i_j) \cup (i_{j+1} - \mathfrak{d}_0 + [\mathfrak{d}_0]).$$

Then we can write

$$\begin{aligned} W_{k,j}[S] &:= \begin{bmatrix} V_1 & V_2 \\ V_3 & V_4 \end{bmatrix} \quad \text{where} \\ (V_1)_{i,\ell} &= \hat{\lambda}_\ell^{i-1}, \quad i \in [\hat{\mathfrak{d}}_N - \mathfrak{d}_0], \ell \in [\mathfrak{d}_0], \\ (V_2)_{i,\ell} &= \hat{\lambda}_{\ell+\mathfrak{d}_0}^{i-1}, \quad i \in [\hat{\mathfrak{d}}_N - \mathfrak{d}_0], \ell \in [\hat{\mathfrak{d}}_N - \mathfrak{d}_0], \\ (V_3)_{i,\ell} &= \hat{\lambda}_\ell^{i+b_j-1-\mathfrak{d}_0}, \quad i \in [\mathfrak{d}_0], \ell \in [\mathfrak{d}_0], \\ (V_4)_{i,\ell} &= \hat{\lambda}_{\ell+\mathfrak{d}_0}^{i+b_j-1-\mathfrak{d}_0}, \quad i \in [\mathfrak{d}_0], \ell \in [\hat{\mathfrak{d}}_N - \mathfrak{d}_0]. \end{aligned}$$

In the cases that either $\mathfrak{d}_0 = 0$ or $\mathfrak{d}_0 = \hat{\mathfrak{d}}_N$, we need only compute the determinant of V_2 or V_3, respectively. Otherwise, by the Schur complement formula,

$$|\det(W_{k,j}[S])| = \left|\det\left(\begin{bmatrix} V_3 & V_4 \\ V_1 & V_2 \end{bmatrix}\right)\right| = |\det(V_3)\det(V_2)\det(\mathrm{Id} - V_2^{-1}V_1V_3^{-1}V_4)|. \tag{5.30}$$

Observe V_2 is a Vandermonde matrix. Writing $V_3 = \tilde{V}_3 \cdot D$, where $D := \operatorname{diag}(\hat{\lambda}_1^{b_j-\mathfrak{d}_0}, \ldots, \hat{\lambda}_{\mathfrak{d}_0}^{b_j-\mathfrak{d}_0})$, we see that $\tilde{V}_3$ is also a Vandermonde matrix.

As $z \notin \mathcal{B}_N$, we can bound the discriminant of $\{\hat{\lambda}_\ell\}$ from below, and as we can bound $|\hat{\lambda}_\ell| \leqslant \|T_k(z)\|$, we have that there is some $C(R, \mathfrak{d}, \boldsymbol{f})$ so that

$$\begin{aligned} |\det(V_2)^2 \det(\tilde{V}_3)^2| &\geqslant (2\|T_k(z)\| \vee 1)^{-\mathfrak{d}(\mathfrak{d}-1)} \prod_{1\leqslant \ell<\ell'\leqslant \hat{\mathfrak{d}}_N} |\hat{\lambda}_\ell(z) - \hat{\lambda}_{\ell'}(z)|^2 \\ &\geqslant C(R, \mathfrak{d}, \boldsymbol{f})^{-1} N^{-2\delta_1\mathfrak{d} - \delta_2\mathfrak{d}^2}. \end{aligned} \tag{5.31}$$

Hence,

$$|\det(V_2)^2 \det(V_3)^2| \geqslant C(R, \mathfrak{d}, \boldsymbol{f})^{-1} N^{-2\delta_1\mathfrak{d} - \delta_2\mathfrak{d}^2} \cdot \prod_{\ell=1}^{\hat{\mathfrak{d}}_N} (|\hat{\lambda}_\ell(z,k)| \vee 1)^{2b_j}. \tag{5.32}$$

https://doi.org/10.1017/fms.2018.29 Published online by Cambridge University Press

Note that the desired lower bound follows from (5.29), (5.30) and (5.32) once we show that

$$\|V_2^{-1}V_1V_3^{-1}V_4\| \leqslant 1/2. \tag{5.33}$$

To this end, first let us recall the following standard inequality:

$$\|\mathsf{M}\| \leqslant \|\mathsf{M}\|_2 = \sqrt{\sum_{i,j} \mathsf{M}_{i,j}^2} \leqslant \max\{n_1, n_2\} \cdot \max_{i,j} |\mathsf{M}_{i,j}|$$

for any matrix M of dimension $n_1 \times n_2$. As $|\lambda_\ell| < 1$ for $\ell \geqslant \mathfrak{d}_0 + 1$ and $|\lambda_\ell| \leqslant \|T_k(z)\|$ for $\ell \leqslant \mathfrak{d}_0$, we note that

$$\|\tilde{V}_3\|, \|V_1\|, \leqslant \mathfrak{d}\|T_k(z)\|^{\mathfrak{d}} \leqslant \mathfrak{d}N^{\delta_2\mathfrak{d}} \quad \text{and} \quad \|V_2\|, \|V_4\| \leqslant \mathfrak{d}.$$

We can then trivially bound

$$\|V_2^{-1}\|^{-1} = \sigma_{\min}(V_2) \geqslant |\det(V_2)|/\|V_2\|^{\mathfrak{d}_0-1}$$

and

$$\|\tilde{V}_3^{-1}\|^{-1} = \sigma_{\min}(\tilde{V}_3) \geqslant |\det(\tilde{V}_3)|/\|\tilde{V}_3\|^{\hat{\mathfrak{d}}_N-\mathfrak{d}_0-1}.$$

Hence, using (5.32)

$$\|V_2^{-1}V_1V_3^{-1}V_4\| \leqslant \frac{\mathfrak{d}^{4\mathfrak{d}}N^{2\mathfrak{d}^2\delta_2}}{|\det(V_2)| \cdot |\det(\tilde{V}_3)|}\|D^{-1}\| \leqslant C(R, \mathfrak{d}, f)\mathfrak{d}^{4\mathfrak{d}}N^{3\mathfrak{d}^2\delta_2+\mathfrak{d}\delta_1} \cdot \|D^{-1}\|. \tag{5.34}$$

Since $z \notin \mathcal{B}_N$, and $4\delta_1 < \delta_3$, the entries of D are bounded below by

$$(1 + N^{-3\delta_1})^{b_j-\hat{\mathfrak{d}}_N} \geqslant \exp\left(\tfrac{1}{2}N^{-3\delta_1}(b_j - \hat{\mathfrak{d}}_N)\right) \geqslant \exp\left(\tfrac{1}{8}N^{-3\delta_1+\delta_3}\right) \geqslant \exp(N^{\delta_1}/8),$$

for all large N. Therefore, $\|D^{-1}\| \leqslant \exp(-N^{\delta_1}/8)$ and hence, in particular, it is smaller than any power of N. Thus, from (5.34) we establish (5.33). This completes the proof of the lemma. □

Building on Lemma 5.9 we now derive bounds on $\det(\mathfrak{U}_{k,j}^*\mathfrak{U}_{k,j})$ where we recall that $\mathfrak{U}_{k,j} := \pi_{k,j}\hat{M}_N(z)U_{k,j}$, and the columns of $U_{k,j}$ are an orthonormal basis of $\mathcal{S}_{k,j}$.

LEMMA 5.10. *Fix $R < \infty$ and $z \in B_{\mathbb{C}}(0, R) \setminus \mathcal{B}_N$. Let $k \in [L_0]$ and $j' \in [L'_k]$ such that $\mathfrak{b}_k^{(j')} = [i_{j+1}]\setminus[i_j]$ for some $j \in [L]\cup\{0\}$. Assume $\hat{\mathfrak{d}}_N(k) > 0$. Then there*

https://doi.org/10.1017/fms.2018.29 Published online by Cambridge University Press

exists a constant $C_3(R, \mathfrak{d}, \boldsymbol{f}) > 1$*, depending only on* $R, \mathfrak{d}$*, and* $\{f_\ell(\cdot)\}_{\ell=0}^{\mathfrak{d}}$*, such that*

$$C_3(R, \mathfrak{d}, \boldsymbol{f})^{-1} N^{-5\mathfrak{d}\delta_1 - 2\mathfrak{d}\delta_2} \prod_{\ell=1}^{\hat{\mathfrak{d}}_N(k)} (|\hat{\lambda}_\ell(z, k)| \wedge 1)^{2b_j} \leqslant \det(\mathfrak{U}_{k,j}^* \mathfrak{U}_{k,j}) \leqslant C_3(R, \mathfrak{d}, \boldsymbol{f}) N^{2\mathfrak{d}^2\delta_2 + 2\mathfrak{d}\delta_1} \prod_{\ell=1}^{\hat{\mathfrak{d}}_N(k)} (|\hat{\lambda}_\ell(z, k)| \wedge 1)^{2b_j}. \tag{5.35}$$

Proof. Since $\{\boldsymbol{w}_\ell^j(k)\}_{\ell=1}^{\hat{\mathfrak{d}}_N}$ span the subspace $\mathcal{S}_{k,j}$, there exists a $\hat{\mathfrak{d}}_N \times \hat{\mathfrak{d}}_N$ matrix Γ such that $U_{k,j} = W_{k,j}\Gamma$. The orthonormality of the columns of $U_{k,j}$ implies $\Gamma^* \mathfrak{W}_{k,j} \Gamma = \mathrm{Id}$. This in particular implies that $\Gamma\Gamma^* = (\mathfrak{W}_{k,j})^{-1}$. Thus

$$\det(\mathfrak{U}_{k,j}^* \mathfrak{U}_{k,j}) = \det(\Gamma^* W_{k,j}^* \hat{M}_N(z)^* \pi_{k,j}^* \pi_{k,j} \hat{M}_N(z) W_{k,j} \Gamma) = \frac{\det(W_{k,j}^* \hat{M}_N(z)^* \pi_{k,j}^* \pi_{k,j} \hat{M}_N(z) W_{k,j})}{\det(\mathfrak{W}_{k,j})}. \tag{5.36}$$

The bound on the denominator of the RHS of (5.36) follows from Lemma 5.9. To evaluate the numerator we recall from Lemma 5.3(ii) that $\rho_{k,j} \hat{M}_N(z) W_{k,j} = \mathbf{0}_{(i_{j+1} - \hat{\mathfrak{d}}_N) \times \hat{\mathfrak{d}}_N}$, where $\mathbf{0}_{n_1 \times n_2}$ is the matrix of zeros of dimension $n_1 \times n_2$. So, we only need to evaluate the next $\hat{\mathfrak{d}}_N$ rows of $\hat{M}_N(z) W_{k,j}$.

To this end, we note that for any $m = i_{j+1} - \hat{\mathfrak{d}}_N + 1, \ldots, i_{j+1}$, and $\ell \in [\hat{\mathfrak{d}}_N]$, we have

$$\begin{aligned}
(\hat{M}_N(z) \boldsymbol{w}_\ell^j(k))_m &= (t_0(k) - z)(\boldsymbol{w}_\ell^j(k))_m + \sum_{m'=m+1}^{i_{j+1}} t_{m'-m}(k) (\boldsymbol{w}_\ell^j(k))_{m'} \\
&= (t_0(k) - z) \hat{\lambda}_\ell^{m - i_j - 1} + \sum_{m'=m+1}^{i_{j+1}} t_{m'-m}(k) \hat{\lambda}_\ell^{m' - i_j - 1} \\
&= \hat{\lambda}_\ell^{m - i_j - 1} \left(\hat{P}_{z,k}(\hat{\lambda}_\ell) - \sum_{m'=i_{j+1}+1}^{m+\hat{\mathfrak{d}}_N} t_{m'-m}(k) \hat{\lambda}_\ell^{m'-m} \right) \\
&= - \sum_{m'=i_{j+1}+1}^{m+\hat{\mathfrak{d}}_N} t_{m'-m}(k) \hat{\lambda}_\ell^{m' - i_j - 1} \\
&= -\hat{\lambda}_\ell^{b_j} \sum_{m'=1}^{m+\hat{\mathfrak{d}}_N - i_{j+1}} t_{m'+i_{j+1}-m}(k) \hat{\lambda}_\ell^{m'-1},
\end{aligned}$$

https://doi.org/10.1017/fms.2018.29 Published online by Cambridge University Press

where the second last step follows from the fact that $\hat{P}_{z,k}(\hat{\lambda}_\ell) = 0$. This implies that

$$\pi_{k,j}\hat{M}_N(z)\boldsymbol{w}_\ell^j(k) = \begin{pmatrix} \mathbf{0}_{(i_{j+1}-\hat{\mathfrak{d}}_N)\times 1} \\ -\hat{\lambda}_\ell^{b_j}\,\Delta\boldsymbol{v}_\ell(k) \\ \mathbf{0}_{(N-i_{j+1})\times 1} \end{pmatrix},$$

where

$$\Delta := \begin{bmatrix} 0 & 0 & \cdots & 0 & t_{\hat{\mathfrak{d}}_N}(k) \\ 0 & 0 & \cdots & t_{\hat{\mathfrak{d}}_N(k)} & t_{\hat{\mathfrak{d}}_N-1}(k) \\ \vdots & \cdots & \cdots & \cdots & \vdots \\ 0 & t_{\hat{\mathfrak{d}}_N}(k) & \cdots & t_3(k) & t_2(k) \\ t_{\hat{\mathfrak{d}}_N}(k) & t_{\hat{\mathfrak{d}}_N-1}(k) & \cdots & t_2(k) & t_1(k) \end{bmatrix}.$$

It further yields that

$$\pi_{k,j}\hat{M}_N(z)W_{k,j} = \begin{bmatrix} \mathbf{0}_{(i_{j+1}-\hat{\mathfrak{d}}_N)\times\hat{\mathfrak{d}}_N} \\ -\Delta\boldsymbol{V}(k)\Lambda^{b_j} \\ \mathbf{0}_{(N-i_{j+1})\times\hat{\mathfrak{d}}_N} \end{bmatrix},$$

where Λ is a diagonal matrix with entries $\{\hat{\lambda}_\ell\}_{\ell=1}^{\hat{\mathfrak{d}}_N}$, and recall $\boldsymbol{V}(k)$ is the $\hat{\mathfrak{d}}_N \times \hat{\mathfrak{d}}_N$ matrix whose columns are $\{\boldsymbol{v}_\ell(k)\}_{\ell=1}^{\hat{\mathfrak{d}}_N}$. Thus

$$\begin{aligned} \det(W_{k,j}^*\hat{M}_N(z)^*\pi_{k,j}^*\pi_{k,j}\hat{M}_N(z)W_{k,j}) &= \det\left((\Lambda^*)^{b_j}\boldsymbol{V}(k)^*\Delta^*\Delta\boldsymbol{V}(k)\Lambda^{b_j}\right) \\ &= \prod_{\ell=1}^{\hat{\mathfrak{d}}_N}|\hat{\lambda}_\ell|^{2b_j}\cdot\det(\boldsymbol{V}(k)\boldsymbol{V}(k)^*)\cdot\det(\Delta^*\Delta). \end{aligned} \tag{5.37}$$

Using (5.24) and that $z \notin \mathcal{B}_N$ (cf. Lemma 5.4), respectively,

$$\begin{aligned} \det(\boldsymbol{V}(k)\boldsymbol{V}(k)^*) = |\det(\boldsymbol{V}(k))|^2 &\leqslant \|T_k(z)\|^{\hat{\mathfrak{d}}_N(\hat{\mathfrak{d}}_N-1)} = O(N^{\mathfrak{d}^2\delta_2}), \\ \times\,|\det(\boldsymbol{V}(k))|^2 &\geqslant N^{-2\delta_1\mathfrak{d}}\cdot\left(\sup_{x\in[0,1]}|f_{\hat{\mathfrak{d}}_N}(x)|^{\hat{\mathfrak{d}}_N-1}\right)^{-1}. \end{aligned}$$

As for Δ, $\det(\Delta) = t_{\hat{\mathfrak{d}}_N}(k)^{\hat{\mathfrak{d}}_N}$, and so

$$N^{-2\mathfrak{d}\delta_2} \leqslant \det(\Delta^*\Delta) \leqslant \sup_{x\in[0,1]}|f_{\hat{\mathfrak{d}}_N}(x)|^{2\hat{\mathfrak{d}}_N}.$$

Now the desired bound on $\det(\mathfrak{U}_{k,j}^*\mathfrak{U}_{k,j})$ follows from (5.36)–(5.37), upon an application of Lemma 5.9. □

Building on Lemma 5.10 we now derive the upper bound on the product of small singular values of $\hat{M}_N(z)$. Before proceeding to the statement of the relevant result let us remind the reader that we chose a partition of $\{\mathfrak{b}_k\}_{k=0}^{L_0}$ of $[N]$ such that for $k \in [L_0] \cup \{0\}$, $\mathfrak{b}_k := \{i \in [N] : \lfloor iN^{\delta_1 - 1} \rfloor = k\}$. We also noted that $N^{1-\delta_1}/2 \leqslant |\mathfrak{b}_k| \leqslant 2N^{1-\delta_1}$ for all $k \in [L_0] \cup \{0\}$. We then considered a refinement $\{\{\mathfrak{b}_k^{(j')}\}_{j'=1}^{L'_k}\}_{k=0}^{L_0}$ of $\{\mathfrak{b}_k\}_{k=0}^{L_0}$ where $N^{\delta_3}/2 \leqslant |\mathfrak{b}_k^{(j')}| \leqslant 2N^{\delta_3}$ for all $k \in [L_0] \cup \{0\}$ and $j' \in [L'_k]$. Finally recall that $0 = i_1 < i_2 < i_3 < \cdots < i_{L+1} = N$, with $L := \sum_{k=0}^{L_0} L'_k$, are the endpoints of the partition $\{\{\mathfrak{b}_k^{(j')}\}_{j'=1}^{L'_k}\}_{k=0}^{L_0}$, and $b_j := i_{j+1} - i_j$. Therefore fixing $k \in [L_0] \cup \{0\}$, and $j' \in [L'_k]$ fixes $j \in [L] \cup \{0\}$ such that $\mathfrak{b}_k^{(j')} = [i_{j+1}] \setminus [i_j]$.

COROLLARY 5.11. *Fix $R < \infty$ and $z \in B_{\mathbb{C}}(0, R) \setminus \mathcal{B}_N$. Recall $\mathfrak{L} := \sum_{k=0}^{L_0} L'_k \cdot \hat{\mathfrak{d}}_N(k)$ and $L = \sum_{k=0}^{L_0} L'_k$. Then*

$$\prod_{m=0}^{\mathfrak{L}-1} \sigma_{N-k}(\hat{M}_N(z)) \leqslant C_3(R, \mathfrak{d}, f)^L N^{(2\mathfrak{d}^2\delta_2 + 2\mathfrak{d}\delta_1)L} \prod_{k=0}^{L_0} \prod_{\ell=1}^{\hat{\mathfrak{d}}_N(k)} (|\hat{\lambda}_\ell(z, k)| \wedge 1)^{|\mathfrak{b}_k|},$$

for all large N, where $C_3(R, \mathfrak{d}, f)$ is as in Lemma 5.10. If, for some k, $\hat{\mathfrak{d}}_N(k) = 0$, then the innermost product becomes empty which, by convention, is set to equal 1.

Proof. Fix $k \in [L_0] \cup \{0\}$, and $j' \in [L'_k]$ such that $\mathfrak{b}_k^{(j')} := [i_{j+1}] \setminus [i_j]$. Let $U_{k,j}$ be the $N \times \hat{\mathfrak{d}}_N(k)$ matrix whose columns form an orthonormal basis of $\mathcal{S}_{k,j}$. Denote

$$U := \left[U_{0,0} \cdots U_{0,L'_0-1}\ U_{1,L'_0} \cdots U_{1,L'_0+L'_1-1} \cdots U_{L_0,L} \right]. \tag{5.38}$$

Note that if $\hat{\mathfrak{d}}_N(k) = 0$ for some k, then $U_{k,j}$ is an empty matrix. Therefore, it is equivalent to ignore such k's while constructing the matrix U. We will show that

$$\begin{aligned}(\det(U^* \hat{M}_N(z)^* \hat{M}_N(z) U))^{1/2} \\ \leqslant \{C_3(R, \mathfrak{d}, f) N^{2\mathfrak{d}^2\delta_2 + 2\mathfrak{d}\delta_1}\}^L \cdot \prod_{k=0}^{L_0} \prod_{\ell=1}^{\hat{\mathfrak{d}}_N(k)} (|\hat{\lambda}_\ell(z, k)| \wedge 1)^{|\mathfrak{b}_k|}.\end{aligned} \tag{5.39}$$

Since, the columns of U are orthonormal, this, together with Lemma A.2, yields the desired upper bound on the product of small singular values.

Turning to prove (5.39) we note the following: For any $k \in [L_0] \cup \{0\}$ and $j' \in [L'_k]$ such that $\mathfrak{b}_k^{(j')} := [i_{j+1}] \setminus [i_j]$ and $\hat{\mathfrak{d}}_N(k) > 0$, the columns of $\hat{M}_N(z) U_{k,j}$ belong to the subspace

$$\mathcal{T}_{k,j} := \operatorname{span}(\{e_m\}_{m=i_{j+1}-\hat{\mathfrak{d}}_N(k)+1}^{i_{j+1}}, \{e_m\}_{m=i_j-\hat{\mathfrak{d}}_N(k^-)+1}^{i_j}),$$

where (a) $k^- := k$ if $j' > 1$, and (b) $k^- := k-1$ if $j' = 1$, and we set $\hat{\mathfrak{d}}_N(-1) := 0$.

This, in particular, implies that

$$\det(U^* \hat{M}_N(z)^* \hat{M}_N(z) U) = \det(\mathfrak{U}_{\mathcal{T}}^* \mathfrak{U}_{\mathcal{T}}), \tag{5.40}$$

where $\mathfrak{U}_{\mathcal{T}}$ is the matrix obtained from $\pi_{\mathcal{T}} \hat{M}_N(z) U = \hat{M}_N(z) U$ by deleting its zero rows, and $\pi_{\mathcal{T}}$ is the orthogonal projection onto $\mathrm{span}(\bigcup_{\hat{\mathfrak{d}}_N(k)>0} \mathcal{T}_{k,j})$. For any $v \in \mathbb{C}^N$, let us denote $\hat{\pi}_{k,j} v$ to be the b_j-dimensional vector obtained from $\pi_{k,j} v$ by deleting its zero rows. Equipped with this notation, we also note that $\mathfrak{U}_{\mathcal{T}}$ is a $\mathfrak{L} \times \mathfrak{L}$ block upper-triangular matrix with $\{\hat{\pi}_{k,j} \hat{M}_N(z) U_{k,j}\}$ as its diagonal blocks. This yields that

$$\det(\mathfrak{U}_{\mathcal{T}}) = \prod \det(\hat{\pi}_{k,j} \hat{M}_N(z) U_{k,j}). \tag{5.41}$$

Since

$$\det(U_{k,j}^* \hat{M}_N(z)^* \hat{\pi}_{k,j}^* \hat{\pi}_{k,j} \hat{M}_N(z) U_{k,j}) = \det(U_{k,j}^* \hat{M}_N(z)^* \pi_{k,j}^* \pi_{k,j} \hat{M}_N(z) U_{k,j}),$$

combining (5.40)–(5.41), and applying Lemma 5.10 we arrive at (5.39). This completes the proof of the lemma. □

It remains to find a matching lower bound on the product of the small singular values. Recall the notation L_0, L, $\mathfrak{L}$, see (5.1), (5.3), (5.25).

LEMMA 5.12. *Fix $R < \infty$ and $z \in B_{\mathbb{C}}(0, R) \setminus \mathcal{B}_N$. Then there exists a constant $C_4(R, \mathfrak{d}, \boldsymbol{f})$, depending only on R, $\mathfrak{d}$, and $\{f_\ell\}_{\ell=0}^{\mathfrak{d}}$, such that*

$$\prod_{m=0}^{\mathfrak{L}-1} \sigma_{N-m}(\hat{M}_N(z)) \geqslant \left(C_4(R, \mathfrak{d}, \boldsymbol{f}) N^{7\mathfrak{d}^2\delta_1 + 3\mathfrak{d}^3\delta_2 + 4\mathfrak{d}\delta_3 + (\alpha_0 \mathfrak{d}\delta_1)/(2(2\alpha_0 - 1))}\right)^{-L} \cdot \mathfrak{L}^{-\mathfrak{L}/2} \prod_{k=0}^{L_0} \prod_{\ell=1}^{\hat{\mathfrak{d}}_N(k)} (|\hat{\lambda}_\ell(z, k)| \wedge 1)^{|\mathfrak{b}_k|},$$

for all large N.

The proof of Lemma 5.12 is similar to that of Proposition 3.14. Hence, we provide only a brief outline below.

Proof of Lemma 5.12. Using Lemma A.2 again we see that it is enough to find a uniform lower bound on

$$\prod_{m=1}^{\mathfrak{L}} \| \hat{M}_N(z) w_m \|_2$$

https://doi.org/10.1017/fms.2018.29 Published online by Cambridge University Press

over all collections of orthonormal vectors $\{w_m\}_{m=1}^{\mathfrak{L}}$. Analogous to the proof of Proposition 3.14 we bound each $\|\hat{M}_N(z)w_m\|_2$ in one of two ways. If $1 - \|\psi w_m\|_2^2 \geqslant 1/2\mathfrak{L}$ then applying Lemma 5.6 we deduce

$$\begin{aligned}\|\hat{M}_N(z)w_m\|_2 &\geqslant \|\rho\hat{M}_N(z)w_m\|_2\\ &\geqslant \frac{1}{\sqrt{2\mathfrak{L}}}C_1(R,\mathfrak{d},f)^{-1}N^{-(2\mathfrak{d}\delta_1+\mathfrak{d}^2\delta_2+2\delta_3+(\alpha_0\delta_1)/(2(2\alpha_0-1)))},\end{aligned} \tag{5.42}$$

where we recall that ψ is the orthogonal projection onto the subspace $\mathcal{S} := \mathrm{span}(\cup\mathcal{S}_{k,j})$.

Without loss of generality, assume w_m, $m \in [p]$ satisfies $1 - \|\psi w_m\|_2^2 < 1/2\mathfrak{L}$. Proceeding similarly as in the steps leading to (3.33) we find

$$\begin{aligned}&\prod_{m=1}^{p}\|\hat{M}_N(z)w_m\|_2\\ &\quad\geqslant \left(\frac{C_1(R,\mathfrak{d},f)^{-1}N^{-(2\mathfrak{d}\delta_1+\mathfrak{d}^2\delta_2+2\delta_3+(\alpha_0\delta_1)/(2(2\alpha_0-1)))}}{4(\|\hat{M}_N(z)\|\vee 1)}\right)^p\prod_{m=1}^{p}\|\hat{M}_N(z)\psi w_m\|_2.\end{aligned} \tag{5.43}$$

Let Y_1 be the matrix whose columns are $\{\psi w_m\}_{m=1}^{p}$. Since the columns of U span the subspace $\mathcal{S}$, there must exist an $\mathfrak{L}\times p$ matrix A_1 such that $Y_1 = UA_1$. We extend the matrix A_1 to an $\mathfrak{L}\times\mathfrak{L}$ matrix A so that the last $\mathfrak{L}-p$ columns of A are orthonormal and are also orthogonal to the first p columns of A_1. Set $Y := UA$ and let us denote the columns of Y to be $\boldsymbol{y}_m$, for $m \in [\mathfrak{L}]$.

Turning to bound the RHS of (5.43), by Hadamard's inequality we now find that

$$\prod_{m=1}^{p}\|\hat{M}_N(z)\psi w_m\|_2^2 \geqslant \frac{\det(Y^*\hat{M}_N(z)^*\hat{M}_N(z)Y)}{\prod_{m=p+1}^{\mathfrak{L}}\|\hat{M}_N(z)\boldsymbol{y}_m\|_2^2}. \tag{5.44}$$

We separately bound the numerator and the denominator of (5.44). An argument similar to the proof of (3.36) yields

$$\|\hat{M}_N(z)\boldsymbol{y}_m\|_2 \leqslant \|\hat{M}_N(z)\|. \tag{5.45}$$

It remains to find a lower bound of the numerator of (5.44). To obtain such a bound, we observe that

$$\det(Y^*\hat{M}_N(z)^*\hat{M}_N(z)Y) = \det(U^*\hat{M}_N(z)^*\hat{M}_N(z)U)\det(AA^*). \tag{5.46}$$

https://doi.org/10.1017/fms.2018.29 Published online by Cambridge University Press

Proceeding similarly as in the proof of (5.39), and applying the lower bound derived in Lemma 5.10 we deduce

$$\det(U^*\hat{M}_N(z)^*\hat{M}_N(z)U) \geqslant \{C_3(R, \mathfrak{d}, f)N^{5\mathfrak{d}\delta_1+2\mathfrak{d}\delta_2}\}^{-L}\cdot\prod_{k=0}^{L_0}\prod_{\ell=1}^{\hat{\mathfrak{d}}_N(k)}(|\hat{\lambda}_\ell(z,k)|\wedge 1)^{2|\mathfrak{b}_k|}. \tag{5.47}$$

Arguments analogous to (3.37) further show that $\det(AA^*) \geqslant 2^{-\mathfrak{L}} \geqslant 2^{-\mathfrak{d}L}$. Plugging this bound in (5.46), and using (5.47) we obtain

$$\det(Y^*\hat{M}_N(z)^*\hat{M}_N(z)Y) \geqslant \{C_3(R, \mathfrak{d}, f)2^{\mathfrak{d}}N^{5\mathfrak{d}\delta_1+2\mathfrak{d}\delta_2}\}^{-L} \cdot\prod_{k=0}^{L_0}\prod_{\ell=1}^{\hat{\mathfrak{d}}_N(k)}(|\hat{\lambda}_\ell(z,k)|\wedge 1)^{2|\mathfrak{b}_k|}.$$

Therefore, from (5.43)–(5.45), and using the fact that $p \leqslant \mathfrak{L} \leqslant \mathfrak{d}L$, we derive

$$\prod_{m=1}^{p}\|\hat{M}_N(z)w_m\|_2 \geqslant \left(\frac{C_1(R, \mathfrak{d}, f)C_3(R, \mathfrak{d}, f)N^{5\mathfrak{d}\delta_1+2\mathfrak{d}^2\delta_2+2\delta_3+(\alpha_0\delta_1)/(2(2\alpha_0-1))}}{8(\|\hat{M}_N(z)\|\vee 1)}\right)^{-\mathfrak{d}L} \cdot\prod_{k=0}^{L_0}\prod_{\ell=1}^{\hat{\mathfrak{d}}_N(k)}(|\hat{\lambda}_\ell(z,k)|\wedge 1)^{|\mathfrak{b}_k|}.$$

Since by the Gershgorin circle theorem, $\|\hat{M}_N(z)\| \leqslant |z| + \sum_{\ell=0}^{\mathfrak{d}}\sup_{x\in[0,1]}|f_\ell(x)|$, using (5.42) we complete the proof of the lower bound. □

We are now ready to finish the proof of Theorem 4.4.

Proof of Theorem 4.4. The tightness of the sequence of random probability measures $\{L_{\hat{\mathcal{M}}_N}\}_{N\in\mathbb{N}}$ in $\mathcal{P}(\mathbb{R})$, the set of all probability measures on $\mathbb{R}$ is immediate from the domination by singular values, see the proof of Corollary 3.6. Therefore, by Prokhorov's theorem $\{L_{\hat{\mathcal{M}}_N}\}_{N\in\mathbb{N}}$ admits subsequential limits. We need to show that all subsequential limits coincide and are given by the deterministic probability measure $\mu_{\mathfrak{d},f}$.

Suppose on the contrary that there exists a subsequence $\{N_m\}$ such that the above does not hold, that is, the limit along the subsequence is not $\mu_{\mathfrak{d},f}$. We fix a further arbitrary subsequence $\{N_{m_n}\}\subset\{N_m\}$ with $N_{m_n}\geqslant 2^n$ for all $n\in\mathbb{N}$ and prove for that subsequence that

$$L_{\hat{\mathcal{M}}_{N_{m_n}}} \Rightarrow \mu_{\mathfrak{d},f} \quad \text{as } n\to\infty \quad \text{in probability.} \tag{5.48}$$

This will prove the theorem. Turning to the proof of (5.48), we first apply [20, Theorem 2.8.3] and deduce that it is enough to show that for Lebesgue a.e. $z \in \mathbb{C}$,

$$\mathcal{L}_{\hat{\mathcal{M}}_{N_{m_n}}}(z) \to \mathcal{L}_{\mu_{\mathfrak{d},f}}(z) \quad \text{as } n \to \infty \quad \text{in probability,} \tag{5.49}$$

where $\mathcal{L}_{\hat{\mathcal{M}}_N}(\cdot)$ is the log potential of the ESD of $\hat{\mathcal{M}}_N$. Since $\mu_{\mathfrak{d},f}$ is compactly supported, one can check the proof of [20, Theorem 2.8.3] to deduce that in fact it suffices to establish (5.49) for Lebesgue a.e. $z \in B_{\mathbb{C}}(0, R)$ for some large R.

We will show that given any $\varepsilon > 0$, there exists a set $\hat{\mathcal{B}}_\varepsilon \subset \mathbb{C}$, depending on the subsequence $\{N_{m_n}\}$, with $\text{Leb}(\hat{\mathcal{B}}_\varepsilon) \leqslant \varepsilon$, such that for all $z \in B_{\mathbb{C}}(0, R) \setminus \hat{\mathcal{B}}_\varepsilon$, the convergence in (5.49) holds. Since $\varepsilon > 0$ is arbitrary, this will complete the proof of (5.49).

Toward this end, define $\hat{\mathcal{B}}_\varepsilon := \bigcup_{n \geqslant n_0(\varepsilon)} \mathcal{B}_{N_{m_n}} \cup f_0([0, 1])$ for some $n_0(\varepsilon) \geqslant 1$, where $\mathcal{B}_N$ is as in Lemma 5.4. Since $f_0(\cdot)$ is an α_0-Hölder-continuous function with $\alpha_0 \geqslant 1/2$, a simple volumetric estimate shows that $\text{Leb}(f_0([0, 1])) = 0$. Hence, using Lemma 5.4 and the union bound we see that, given any $\varepsilon > 0$, there exists $n_0 := n_0(\varepsilon)$ such that $\text{Leb}(\hat{\mathcal{B}}_\varepsilon) \leqslant \varepsilon$. With this choice of the set $\hat{\mathcal{B}}_\varepsilon$, we now prove (5.49).

To this end, our goal is to apply Theorem 2.1. We need to show that all the assumptions of Theorem 2.1 are satisfied. From Remark 5.8 we have that for any $z \in B_{\mathbb{C}}(0, R) \setminus \hat{\mathcal{B}}_\varepsilon$, $N^* = o(N/\log N)$ along the subsequence $\{N_{m_n}\}$. Therefore applying Theorem 2.1 we conclude that for $z \in B_{\mathbb{C}}(0, R) \setminus \hat{\mathcal{B}}_\varepsilon$,

$$\left| \frac{1}{N_{m_n}} \log |\det(\hat{\mathcal{M}}_{N_{m_n}}(z))| - \frac{1}{N_{m_n}} \log |\det B(\hat{M}_{N_{m_n}}(z))| \right| \to 0 \quad \text{as } n \to \infty, \tag{5.50}$$

in probability. Thus it remains to find

$$\lim_{n \to \infty} \frac{1}{N_{m_n}} \log \det |B(\hat{M}_{N_{m_n}}(z))|.$$

To this end, we note that

$$|\det B(\hat{M}_N(z))| = \frac{|\det \hat{M}_N(z)|}{\prod_{p=0}^{N^*} \sigma_{N-p}(\hat{M}_N(z))} = \frac{\prod_{k=0}^{L_0} |t_0(k) - z|^{|\mathfrak{b}_k|}}{\prod_{p=0}^{N^*} \sigma_{N-p}(\hat{M}_N(z))}. \tag{5.51}$$

Since $\|\hat{M}_N(z)\| = O(1)$ for any $z \in B_{\mathbb{C}}(0, R)$, and $N^* \leqslant \mathfrak{L} = O(N^{1-\delta_3})$ by (5.25), using the definition of N^* we have that

$$\lim_{n \to \infty} \frac{1}{N_{m_n}} \log \left(\prod_{p=N^*+1}^{\mathfrak{L}-1} \sigma_{N_{m_n}-p}(\hat{M}_{N_{m_n}}(z)) \right) = 0.$$

https://doi.org/10.1017/fms.2018.29 Published online by Cambridge University Press

Hence, it is enough to find $\lim_{n\to\infty} \Upsilon_n(z)$ and show that it equals $\mathcal{L}_{\mu_{\mathfrak{d},f}}(z)$, where

$$\Upsilon_n(z) := \frac{1}{N_{m_n}} \log\left(\sum_{k=0}^{L_0} |\mathfrak{b}_k| \log|t_0(k) - z| - \sum_{p=0}^{\mathfrak{L}-1} \log \sigma_{N_{m_n}-p}(\hat{M}_{N_{m_n}}(z))\right).$$

Since $\mathfrak{L} = O(N^{1-\delta_3})$, we have, using Lemma 5.12, that for any $z \in B_{\mathbb{C}}(0, R) \setminus \hat{\mathcal{B}}_\varepsilon$,

$$\begin{aligned}
&\limsup_{n\to\infty} \Upsilon_n(z) \\
&\leqslant \limsup_{n\to\infty} \frac{1}{N_{m_n}} \log\left(\sum_{k=0}^{L_0} |\mathfrak{b}_k| \left\{\log|t_0(k) - z| - \sum_{\ell=1}^{\hat{\mathfrak{d}}_{N_{m_n}}(k)} \log(|\hat{\lambda}_\ell(z, k)| \wedge 1)\right\}\right) \\
&= \limsup_{n\to\infty} \frac{1}{N_{m_n}} \log\left(\sum_{k=0}^{L_0} |\mathfrak{b}_k| \left\{\sum_{\ell=1}^{\hat{\mathfrak{d}}_{N_{m_n}}(k)} \log(|\hat{\lambda}_\ell(z, k)| \vee 1) + \log|t_{\hat{\mathfrak{d}}_{N_{m_n}}(k)}(k)|\right\}\right),
\end{aligned} \tag{5.52}$$

where the last step follows from the fact that $\{\hat{\lambda}_\ell(z, k)\}_{\ell=1}^{\hat{\mathfrak{d}}_N(k)}$ are the eigenvalues of the matrix $T_k(z)$, and hence

$$\prod_{\ell=1}^{\hat{\mathfrak{d}}_{N_{m_n}}(k)} |\hat{\lambda}_\ell(z, k)| = |\det(T_k(z))| = \frac{|z - t_0(k)|}{|t_{\hat{\mathfrak{d}}_{N_{m_n}}(k)}(k)|}. \tag{5.53}$$

We claim that for any $z \in B_{\mathbb{C}}(0, R) \setminus \hat{\mathcal{B}}_\varepsilon$,

$$\sup_n \sup_{k\in[L_0]\cup\{0\}} \left| \sum_{\ell=1}^{\hat{\mathfrak{d}}_{N_{m_n}}(k)} \log(|\hat{\lambda}_\ell(z, k)| \vee 1) + \log|t_{\hat{\mathfrak{d}}_{N_{m_n}}(k)}(k)| \right| \leqslant C(f, \mathfrak{d}, R) < \infty, \tag{5.54}$$

for some constant $C(f, \mathfrak{d}, R)$ depending only on $\{f_\ell\}_{\ell=0}^{\mathfrak{d}}$, $\mathfrak{d}$, and R. Indeed, noting that the closed set $f_0([0, 1]) \in \mathcal{B}_\varepsilon$, we have $\eta_0 := \text{dist}(z, f_0([0, 1])) > 0$. Upon using the triangle inequality we therefore conclude that there exists $\eta > 0$, such that for every $k \in [L_0] \cup \{0\}$, every root of the polynomial equation $\hat{P}_{z,k}(\lambda) = 0$ is greater than η in absolute value. Thus,

$$\sup_{k\in[L_0]\cup\{0\}} \left| \sum_{\ell=1}^{\hat{\mathfrak{d}}_{N_{m_n}}(k)} \log(|\hat{\lambda}_\ell(z, k)| \wedge 1) \right| \leqslant \mathfrak{d}|\log(\eta)|,$$

for all n. Since $\eta_0 \leqslant |z - t_0(k)| \leqslant R + \sup_{x\in[0,1]} |f_0(x)|$, for all $k \in [L_0] \cup \{0\}$, using (5.53) again the claim in (5.54) follows.

Next fix $x \in [0, 1]$. It is easy to check that for any $x \in (0, 1)$, $\hat{P}_{z,k}(\lambda) \to P_{z,x}(\lambda)$ where $k = \lfloor x N^{\delta_1} \rfloor$. Since the roots of a polynomial are continuous function of its coefficients we have that

$$\sum_{\ell=1}^{\hat{\mathfrak{d}}_{N_{m_n}}(k)} \log(|\hat{\lambda}_\ell(z, k)| \vee 1) + \log |t_{\hat{\mathfrak{d}}_{N_{m_n}}(k)}(k)| \to \sum_{\ell=1}^{\hat{\mathfrak{d}}(x)} \log(|\lambda_\ell(z, x)| \vee 1) + \log |f_{\hat{\mathfrak{d}}(x)}(x)|,$$

as $n \to \infty$, where $k = \lfloor x N^{\delta_1} \rfloor$. Therefore, using (5.54) and the bounded convergence theorem, from (5.52) we deduce that

$$\limsup_{n\to\infty} \Upsilon_n(z) \leqslant \int_0^1 \left\{ \sum_{\ell=1}^{\hat{\mathfrak{d}}(x)} \log(|\lambda_\ell(z, x)| \vee 1) + \log |f_{\hat{\mathfrak{d}}(x)}(x)| \right\} dx = \mathcal{L}_{\mu_{\mathfrak{d},f}}(z).$$

Now applying Corollary 5.11 and using a similar reasoning as above, it also follows that for any $z \in B_{\mathbb{C}}(0, R) \setminus \hat{\mathcal{B}}_\varepsilon$, we have $\liminf_{n\to\infty} \Upsilon_n(z) \geqslant \mathcal{L}_{\mu_{\mathfrak{d},f}}(z)$. This together with (5.50) shows that, for any $\varepsilon > 0$, the convergence in (5.49) holds for all z outside a set of Lebesgue measure at most ε. Hence, the proof of the theorem is now complete. □

Acknowledgements

We thank Nick Trefethen for many useful discussions and examples that motivated this study, and in particular for urging us to consider the twisted Toeplitz case. We also thank him for his comments on a previous version of this article. We are grateful to the anonymous referee for thoughtful comments which led to the improvement of the presentation and clarity of this paper. Research of A.B. and O.Z. are partially supported by grant 147/15 from the Israel Science Foundation and by a US-Israel BSF grant. A.B. also acknowledges the support from ICTS–Infosys Excellence Grant.

Appendix A. Some algebraic facts

In this section we collect a couple of standard matrix results which have been used in the proofs appearing in Sections 2, 3, and 5.

The first result shows that the determinant of the sum of the two matrices can be expressed as a linear combination of products of the determinants of

https://doi.org/10.1017/fms.2018.29 Published online by Cambridge University Press

appropriate submatrices. The proof trivially follows from the definition of the determinant. For a proof we refer the reader to [14]. We adopt the convention that the determinant of the matrix of size zero is one. The following result essentially follows from the definition of the determinant of a matrix.

LEMMA A.1. *For an $N \times N$ matrix A, and $X, Y \subseteq [N]$ we write $A[X, Y]$ for the submatrix of A which consists of the rows in X and the columns in Y. Then for any two $N \times N$ matrices A and B we have*

$$\det(A + B) = \sum_{\substack{X,Y \subset [N] \\ |X|=|Y|}} (-1)^{\operatorname{sgn}(\sigma_X)\operatorname{sgn}(\sigma_Y)} \det(A[\check{X}, \check{Y}]) \det(B[X, Y]), \tag{A.1}$$

where $\check{X} := [N] \setminus X$, $\check{Y} := [N] \setminus Y$ and σ_Z for $Z \in \{X, Y\}$ is the permutation on $[N]$ which places all the elements of Z before all the elements of $\check{Z}$, but preserves the order of elements within the two sets.

The next lemma deals with the characterization of products of singular values.

LEMMA A.2. *Let A be an $N \times N$ matrix. Then for any $k \leqslant N - 1$, we have*

$$\prod_{k'=0}^{k} \sigma_{N-k'}(A) = \inf_{\xi_0,\xi_1,\ldots,\xi_k} (\det(\Xi_k^* A^* A \Xi_k))^{1/2} = \inf_{\xi_0,\xi_1,\ldots,\xi_k} \prod_{k'=0}^{k} \|A\xi_{k'}\|_2, \tag{A.2}$$

where the infimums are taken over set of orthonormal vectors $\{\xi_0, \xi_1, \ldots, \xi_k\}$ and Ξ_k is the matrix whose columns are $\{\xi_{k'}\}_{k'=0}^{k}$.

The equality (A.2) can be thought of as a generalization of Courant–Fischer–Weyl min–max principle. The equality of the leftmost and the rightmost terms in (A.2) follows from [13, Page 200, Ex. 12]. For completeness, we provide a proof.

Proof. Using Hadamard's determinantal inequality we first observe that

$$\inf_{\xi_0,\xi_1,\ldots,\xi_k} (\det(\Xi_k^* A^* A \Xi_k))^{1/2} \leqslant \inf_{\xi_0,\xi_1,\ldots,\xi_k} \prod_{k'=0}^{k} \|A\xi_{k'}\|_2.$$

Next setting $\xi_0, \xi_1, \ldots, \xi_k$ to be the right singular vectors of A corresponding to $\sigma_N(A), \sigma_{N-1}(A), \ldots, \sigma_{N-k}(A)$, respectively, we see that the product of the ℓ_2-norms of $A\xi_{k'}$, for $k' = 0, 1, \ldots, k$, equals the product of $(k + 1)$-st smallest singular values of A. Hence, we deduce

$$\prod_{k'=0}^{k} \sigma_{N-k'}(A) \geqslant \inf_{\xi_0,\xi_1,\ldots,\xi_k} \prod_{k'=0}^{k} \|A\xi_{k'}\|_2.$$

https://doi.org/10.1017/fms.2018.29 Published online by Cambridge University Press

Therefore to prove (A.2) it is enough to show that

$$\prod_{k'=0}^{k} \sigma_{N-k'}(A) \leqslant \inf_{\xi_0,\xi_1,\ldots,\xi_k} (\det(\Xi_k^* A^* A \Xi_k))^{1/2}. \tag{A.3}$$

Let $A = U\Sigma W$ be the singular value decomposition of A. Thus

$$\det(\Xi_k^* A^* A \Xi_k) = \det(\Xi_k^* W^* \Sigma^2 W \Xi_k). \tag{A.4}$$

Instead of taking the infimum over all Ξ_k, whose columns are orthonormal, we may change variables and take the infimum over all $W_k = W\Xi_k$, which is again a collection of $(k+1)$ orthonormal columns. Applying Cauchy–Binet formula,

$$\det(W_k^* \Sigma^2 W_k) = \sum_{\substack{S\subset[N]\\|S|=k+1}} |\det((\Sigma W_k)[S])|^2,$$

where $(\Sigma W_k)[S]$ is the $(k+1)\times(k+1)$ matrix with rows in S. Since Σ is a diagonal matrix we observe that

$$\det((\Sigma W_\ell)[S]) = \det((W_\ell)[S]) \cdot \prod_{i\in S} \sigma_i(A) \geqslant \det((W_\ell)[S]) \prod_{k'=0}^{k} \sigma_{N-k'}(A).$$

Hence

$$\det(W_k^* \Sigma^2 W_k) \geqslant \prod_{k'=0}^{k} \sigma_{N-k'}^2(A) \sum_{\substack{S\subset[N]\\|S|=k+1}} |\det((W_k)[S])|^2 = \prod_{k'=0}^{k} \sigma_{N-k'}^2(A) \det(W_k^* W_k),$$

where we have again applied the Cauchy–Binet formula. Since the columns of W_k are orthonormal, combining the above with (A.4), the inequality (A.3) follows. This completes the proof. □

References

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

[2] A. Basak, E. Paquette and O. Zeitouni, 'Spectrum of random perturbations of Toeplitz matrices with finite symbols', Preprint, 2018, arXiv:1812.06207.

[3] R. Bhatia, *Matrix Analysis* (Springer, Berlin, 1997).

[4] W. Craig and B. Simon, 'Log Hölder continuity of the integrated density of states for stochastic Jacobi matrices', *Comm. Math. Phys.* **90**(2) (1983), 207–218.

[5] R. B. Davies and M. Hager, 'Perturbations of Jordan matrices', *J. Approx. Theory* **156** (2009), 82–94.
[6] O. N. Feldheim, E. Paquette and O. Zeitouni, 'Regularization of non-normal matrices by Gaussian noise', *Int. Math. Res. Not. IMRN* **18** (2015), 8724–8751.
[7] N. R. Goodman, 'Distribution of the determinant of a complex wishart distributed matrix', *Ann. of Math. Stat.* **34**(1) (1963), 178–180.
[8] Y. Gordon, 'On Milman's inequality and random subspaces which escape through a mesh in $\mathbb{R}^n$', *Geom. Aspects Funct. Anal.* (2002), 84–106. (Springer, 1988).
[9] A. Guionnet, M. Krishnapur and O. Zeitouni, 'The single ring theorem', *Ann. of Math.* **174**(2) (2011), 1189–1217.
[10] A. Guionnet, P. M. Wood and O. Zeitouni, 'Convergence of the spectral measure of non-normal matrices', *Proc. Amer. Math. Soc.* **142** (2014), 667–679.
[11] W. Hachem, P. Loubaton and J. Najim, 'Deterministic equivalents for certain random functionals of large random matrices', *Ann. Appl. Probab.* **3** (2007), 875–930.
[12] R. A. Horn and C. R. Johnson, *Topics in Matrix Analysis* (Cambridge University Press, 1991).
[13] R. A. Horn and C. R. Johnson, *Matrix Analysis*, (Cambridge University Press, 2012).
[14] M. Marcus, 'Determinant of sums', *College Math. J.* **21** (1990), 130–135.
[15] L. Reichel and L. N. Trefethen, 'Eigenvalues and pseudo–eigenvalues of Toeplitz matrices', *Linear Algebra Appl.* **162** (1992), 153–185.
[16] B. Simon, *A Comprehensive Course in Analysis, Part 3. Harmonic Analysis* (American Mathematical Society, Providence, RI, 2015).
[17] J. Sjöstrand and M. Vogel, 'Large bi-diagonal matrices and random perturbations', *J. Spectr. Theor.* **6** (2016), 977–1020.
[18] J. Sjöstrand and M. Vogel, 'Interior eigenvalue density of large bi-diagonal matrices subject to random perturbations', in *Microlocal Analysis and Singular Perturbation Theory*, RIMS Kôkyûroku Bessatsu, B61 (Res. Inst. Math. Sci. (RIMS), Kyoto, 2017), 201–227.
[19] P. Śniady, 'Random regularization of Brown spectral measure', *J. Funct. Anal.* **193** (2002), 291–313.
[20] T. Tao, *Topics in Random Matrix Theory*, Graduate Studies in Mathematics, 132 (American Mathematical Society, Providence, RI, 2012).
[21] T. Tao, V. Vu and K. Manjunath, 'Random matrices: universality of ESDs and the circular law', *Ann. Probab.* **38**(5) (2010), 2023–2065.
[22] L. N. Trefethen, 'Pseudospectra of Matrices', in *Numerical Analysis 1991,* Vol. 260 (eds. D. F. Griffiths and G. A. Watson) (Longman Sci. Tech., Harlow, UK, 1992), 234–266.
[23] L. N. Trefethen, 'Wave packet pseudomodes of variable coefficient differential operators', *Proc. R. Soc. Lond. Ser. A* **461** (2005), 3099–3122.
[24] L. N. Trefethen and S. J. Chapman, 'Wave packet pseudomodes of twisted Toeplitz matrices', *Comm. Pure Appl. Math.* **57**(9) (2004), 1233–1264.
[25] L. N. Trefethen, M. Contedini and M. Embree, 'Spectra, pseudospectra, and localization for random bidiagonal matrices', *Commun. Pure Appl. Math.* **54**(5) (2001), 595–623.
[26] L. N. Trefethen and M. Embree, *Spectra and Pseudospectra: the Behavior of Nonnormal Matrices and Operators* (Princeton University Press, Princeton, 2005).
[27] P. M. Wood, 'Universality of the ESD for a fixed matrix plus small random noise: A stability approach', *Ann. Inst. H. Poincaré Probab. Stat.* **52**(4) (2016), 1877–1896.