



**HAL**  
open science

# Localization of eigenvectors of non-Hermitian banded noisy Toeplitz matrices

Anirban Basak, Martin Vogel, Ofer Zeitouni

► **To cite this version:**

Anirban Basak, Martin Vogel, Ofer Zeitouni. Localization of eigenvectors of non-Hermitian banded noisy Toeplitz matrices. *Probability and Mathematical Physics*, 2023, 4 (3), pp.477-607. 10.2140/pmp.2023.4.477 . hal-03365834

**HAL Id: hal-03365834**

**<https://hal.science/hal-03365834>**

Submitted on 5 Nov 2021

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

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# LOCALIZATION OF EIGENVECTORS OF NON-HERMITIAN BANDED NOISY TOEPLITZ MATRICES

ANIRBAN BASAK, MARTIN VOGEL, AND OFER ZEITOUNI

ABSTRACT. We prove localization with high probability on sets of size of order  $N/\log N$  for the eigenvectors of non-Hermitian finitely banded  $N \times N$  Toeplitz matrices  $P_N$  subject to small random perturbations, in a very general setting. As perturbation we consider  $N \times N$  random matrices with independent entries of zero mean, finite moments, and which satisfy an appropriate anti-concentration bound. We show via a Grushin problem that an eigenvector for a given eigenvalue  $z$  is well approximated by a random linear combination of the singular vectors of  $P_N - z$  corresponding to its small singular values. We prove precise probabilistic bounds on the local distribution of the eigenvalues of the perturbed matrix and provide a detailed analysis of the singular vectors to conclude the localization result.

## 1. INTRODUCTION AND STATEMENT OF RESULTS

1.1. **The setting.** The spectrum of non-Hermitian operators is inherently sensitive to tiny perturbations due to the fact that their resolvent may be large even far away from the spectrum. This is in stark contrast to the Hermitian case, where due to the spectral theorem, the norm of the resolvent is effectively controlled by the distance of the spectral parameter to the spectrum. This spectral instability of non-Hermitian operators, although traditionally an adversary for numerical analysis [70, 72], has recently shown itself at the origin of beautiful new results in a variety of contexts. For instance in the theory of non-linear partial differential equations for instance, spectral instability may help to explain the blow up in finite time of solutions to certain non-linear diffusion equations which, when solely studying the spectrum of the linearized operator, were expected to have stable solutions [58, 54, 30].

In mathematical physics non-Hermitian operators appear in a large variety of subjects, such as open quantum systems [50, 51, 52]. In quantum mechanics, the study of scattering systems [26, 37, 60, 66, 48, 25] naturally leads to the concept of *quantum resonances* which can be described by the eigenvalues of a non-Hermitian operator obtained from a complex deformation [66] of a Hermitian quantum Hamiltonian. In physical models an “ideal” operator can be perturbed by many different sources, some of which are uncontrolled by experimentalists. To account for these error terms disorder is introduced and, in view of the phenomenon of spectral instability, it is therefore relevant to investigate the influence of random perturbations on the spectral data of non-Hermitian operators. The recent works [62, 41, 24] investigate for instance the distribution of the quantum resonances of random Schrödinger operators such as the celebrated Anderson model [6].

In this paper we consider large deterministic non-Hermitian  $N \times N$  Toeplitz matrices  $P_N$  with small additive random perturbations. It was shown in a series of recent results that the spectra of such matrices, apart from finitely many fluctuating outliers [7, 8, 9, 65, 64], mimic the absolutely continuous spectra of the associated infinite dimensional Laurent operator on  $\mathbb{Z}$ . This is particularly striking since a perturbation of size  $O(N^{-\infty})$  is sufficient to produce this effect, whereas the spectrum of the unperturbed matrix is far from the spectrum of the Laurent operator.

The aim of this paper is to discuss the eigenvectors associated with the eigenvalues of such perturbed Toeplitz matrices. Are the eigenvectors *localized* or *delocalized*? The precise meanings of these notions vary over different subjects, however, they all serve to capture how much an  $\ell^2$  normalized eigenvectors *concentrates on* or *spreads out over* certain parts of its support.

---

*Date:* November 5, 2021.

*2010 Mathematics Subject Classification.* 47A10, 47B80, 47H40, 47A55, 60B20.

*Key words and phrases.* Spectral theory; non-self-adjoint operators; random perturbations.

In random matrix theory there are several ways of testing for localization or delocalization of  $\ell^2$  normalized eigenvectors  $\psi$ . One way is by comparing their  $\ell^p$  norms for  $2 < p \leq +\infty$  with  $N^{1/p-1/2}$ . Complete delocalization is said to occur when  $\|\psi\|_p = O(N^{1/p-1/2})$  (up to some logarithmic factors) since  $N^{1/p-1/2}$  is the  $\ell^p$  norm of the fully delocalized vector  $(N^{-1/2}, \dots, N^{-1/2})$ . Conversely localized eigenvectors have a large  $\ell^p$  norm, as for instance the fully localized vector  $(0, 0, 1, 0, \dots, 0)$  has  $\ell^p$  norm equal to one. These notions were used for instance to prove delocalization via optimal  $\ell^p$  bounds of the eigenvectors of Wigner matrices [29, 28], for non-Hermitian random matrices [56], and for the adjacency matrix of Erdős-Rényi graphs [27]. Recently, localization and delocalization of eigenvectors for the adjacency matrix of *critical* Erdős-Rényi graphs were established in [2].

There is a complementary notion of delocalization, known as *no-gaps delocalization*, which asserts that for any subset  $I \subset [N]$ , with  $|I|$  reasonably large, one has  $\|\psi\|_{\ell^2(I)} \gtrsim |I|$  (again allowing for logarithmic factors). Recently, such delocalizations have been established for Wigner matrices and matrices with independent and identically distributed (i.i.d.) entries (cf. [56, 43, 45]).

In the field of *quantum chaos* [59], in the setting of Hermitian pseudo-differential operators, localization and delocalization of normalized eigenvectors are studied via their associated *semi-classical defect measures*. Translated to the matrix setting [5], we note that  $\sum_{x=1}^N |\psi(x)|^2 \delta_x$ , where  $\delta_x$  denotes the Dirac measure at  $x$ , defines a probability measure. One says that *quantum ergodicity* occurs when  $\sum_{x=1}^N a(x)|\psi(x)|^2$  is close to  $\frac{1}{N} \sum_{x=1}^N a(x)$  for most eigenvectors  $\psi$ , and *uniquely quantum ergodicity* occurs when this holds for all eigenvectors. In contrast, *scarring* occurs when we have concentration of the form  $\sum_{x \in \Lambda} |\psi(x)|^2 \geq 1 - \varepsilon$ ,  $\varepsilon > 0$ , of the eigenvector on some small set  $\Lambda$ . On the other hand, if  $\|\psi\|_p \asymp N^{f(p)}$  for some  $f(p) \neq 1/2 - 1/p$ , then the eigenstate  $\psi$  is termed to be *non-ergodic* and *multi-fractal* [44]. These notions were recently applied to the study of the eigenfunctions of the discrete Laplacian on large regular graphs [3, 4], and to the proof of delocalization of eigenvectors of generalized Wigner matrices [16]. See also [10] for results on deformed Wigner matrices.

*In this paper we prove that the eigenvectors of non-selfadjoint Toeplitz matrices subject to small random perturbations localize on a set of cardinality  $N/\log N$  in the sense that they scar on a set of size  $N/\log N$  with probability close to one.*

To the best of our knowledge, this is the first instance where localization results are proved in the setting of noisy perturbations of non-Hermitian matrices.

It will be seen below that for eigenpairs  $(\lambda, \psi)$ , the length of the localizing set for  $\psi$  and the rate of decay of the slowest decaying *pure state* associated to  $\lambda$  have the same order of magnitude (see Remark 9.4). On the other hand, pure states can be related to the *Lyapunov spectra* of the associated *transfer matrices* (cf. [7, Section 1.2]). Therefore, the reader may note that our result on the localization has the same flavor as the one predicted in the case of the *random Schrödinger operator* on a strip, where it is conjectured that the rate of decay of eigenfunctions is neither slower nor faster than the one prescribed by the slowest *Lyapunov exponent* (see [33] and the references therein).

**1.2. The results.** Let  $N_{\pm} \in \mathbb{Z}$  be such that  $-N_- \leq N_+$  and either  $N_+ \neq 0$  or  $N_- \neq 0$ . Let  $a_i \in \mathbb{C}$ ,  $i \in \mathbb{Z}$ , be such that  $a_{N_+} \neq 0$ ,  $a_{-N_-} \neq 0$ , and  $a_i = 0$  for  $i \notin [-N_-, N_+]$ . Introduce the symbol  $p(\zeta) = \sum_{-N_-}^{N_+} a_j \zeta^{-j}$  and the associated  $N \times N$  *Toeplitz* matrix  $P_N$  with entries  $P_N(i, j) = a_{i-j}$ , that is

$$P_N = \begin{pmatrix} a_0 & a_{-1} & \dots & a_{-N_-} & \dots \\ a_1 & a_0 & a_{-1} & \dots & \dots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ a_{N_+} & \dots & \dots & \dots & \dots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \dots & \dots & a_{N_+} & \dots & a_0 \end{pmatrix}. \quad (1.1)$$

(We refer to Section 4 for an introduction to the terminology, especially with respect to symbols.)

We consider in this paper noisy perturbations of  $P_N$  of the form

$$P_{N,\gamma}^Q = P_N + N^{-\gamma}Q_N, \quad \gamma > 1, \quad (1.2)$$

with  $Q_N$  an  $N \times N$  (random) matrix satisfying Assumptions 1.1 and 1.2 below. The first assumption is on the existence of finite moments.

**Assumption 1.1.** Let  $\{Q_{i,j}\}_{i,j=1}^N$  be the entries of the  $N \times N$  noise matrix  $Q = Q_N$ .

(i) The entries of  $Q$  are jointly independent and have zero mean.

(ii) For any  $h \in \mathbb{N}$  there exists an absolute constant  $\mathfrak{C}_h < \infty$  such that

$$\max_{i,j=1}^N \mathbb{E}[|Q_{i,j}|^{2h}] \leq \mathfrak{C}_h.$$

For notation convenience, we take the sequence  $\mathfrak{C}_h$  increasing in  $h$ . To introduce the second assumption, recall Lévy's concentration function, defined for any complex-valued random variable  $\mathbb{X}$  and  $\varepsilon > 0$  by

$$\mathcal{L}(\mathbb{X}, \varepsilon) \stackrel{\text{def}}{=} \sup_{w \in \mathbb{C}} \mathbf{P}(|\mathbb{X} - w| \leq \varepsilon). \quad (1.3)$$

**Assumption 1.2.** Assume that there exist absolute constants  $\eta \in (0, 1]$  and  $C_{1.2} < \infty$ , such that

$$\mathcal{L}(Q_{i,j}, \varepsilon) \leq C_{1.2} \varepsilon^{1+\eta}, \quad (1.4)$$

for all sufficiently small  $\varepsilon > 0$ , uniformly for all  $N$  and  $i, j \in \{1, 2, \dots, N\}$ .

(The standard example of a noise matrix satisfying Assumptions 1.1 and 1.2 is the complex Ginibre matrix, i.e. with i.i.d. entries that are standard complex Gaussian variables.)

It was recently shown in [7, 8, 64, 65] that all but  $o(N)$  of the eigenvalues  $\{\lambda_i^N\}$  of  $P_{N,\gamma}^Q$  lie in a small neighborhood of the curve  $p(S^1)$ , where  $S^1 := \{z \in \mathbb{C} : |z| = 1\}$ ; in fact, it was shown in those references that the empirical measure of eigenvalues of  $P_{N,\gamma}^Q$ ,

$$L_N := N^{-1} \sum_{i=1}^N \delta_{\lambda_i^N}, \quad (1.5)$$

converges weakly to the push forward of the uniform measure on  $S^1$  by  $p$ . As part of our study, we will obtain more precise information, and show (see Theorem 1.5 and Sections 5-7) that most of the eigenvalues lie in certain neighborhoods of width of order  $\log N/N$  that are separated from  $p(S^1)$  by distance of the same order.

Our goal in this paper is to study the eigenvectors associated with the latter (random) eigenvalues. Roughly speaking, we will show that those eigenvalues  $\hat{z}$  away from certain isolated bad points of  $p(S^1)$  have corresponding eigenvectors which are close to a random linear combination of the eigenvectors  $e_j$  of  $(P_N - \hat{z}I)^*(P_N - \hat{z}I)$  associated with its smallest eigenvalues. In particular we will show that this random linear combination of vectors localizes at scale  $N/\log N$ . To state our results precisely requires the introduction of some machinery, which we now do.

Sometimes, the symbol  $p$  possesses a natural contraction, defined as follows. Set

$$\mathbf{g}(p) := \gcd\{|j| : j \neq 0 \text{ and } a_j \neq 0\}. \quad (1.6)$$

If  $\mathbf{g}_0 := \mathbf{g}(p) > 1$  then  $p(\zeta) = q_p(\zeta^{\mathbf{g}_0})$  for some Laurent polynomial  $q_p$ . If  $\mathbf{g}_0 = 1$  then  $q_p = p$ . For  $\varepsilon > 0$  and a set  $\mathcal{B} \subset \mathbb{C}$ ,  $\mathcal{B}^\varepsilon$  denotes the  $\varepsilon$ -blow up of  $\mathcal{B}$ , that is the Minkowski sum of the sets  $\mathcal{B}$  and  $D(0, \varepsilon)$ , the open disc of radius  $\varepsilon$  centered at zero.

**Definition 1.3** (Set of bad points). Let  $\mathcal{B}_1$  be the collection of self intersection points of  $q_p(S^1)$ , and let  $\mathcal{B}_2$  be the set of branch points, i.e. points  $z$  where the Laurent polynomial  $p(\cdot) - z$  has double roots. Set  $\mathcal{B}_p := \mathcal{B}_1 \cup \mathcal{B}_2$  and  $\mathcal{G}_{p,\varepsilon} := p(S^1) \setminus \mathcal{B}_p^\varepsilon$ .

In Definition 1.3, a point  $w \in \mathbb{C}$  is a self intersection point of  $q_p(S^1)$  if there exist  $\zeta_1 \neq \zeta_2 \in S^1$  so that  $q_p(\zeta_1) = q_p(\zeta_2) = w$ .

Throughout the paper, we make the following assumption on the symbol  $p$ .

**Assumption 1.4.** The symbol  $p$  satisfies  $a_{-N_-}, a_{N_+} \neq 0$ , and  $\mathcal{B}_1$  is a finite set.

Under Assumption 1.4,  $\mathcal{B}_p$  is a finite set. Indeed,  $\mathcal{B}_2$  is precisely the set of all  $z$ 's such that the discriminant of the polynomial  $\zeta \mapsto \zeta^{N_+} p(\zeta) - z$  vanishes, and [15, Lemma 11.4] yields that  $\mathcal{B}_2$  is a finite set. We note that by [40], unless  $N_- = N_+$  and  $|a_{-N_-}| = |a_{N_+}|$ ,  $\mathcal{B}_1$  has cardinality bounded above by  $(N_+ + N_- - 1)^2$ , so symbols avoiding this situation satisfy Assumption 1.4. For  $z \in \mathbb{C}$  let

$$d(z) = \text{ind}_{p(S^1)}(z)$$

denote the winding number of the curve  $p(S^1)$  around  $z$ . We now describe the collection of eigenvalues of interest to us. For  $0 < \varepsilon, C < \infty$  and  $N$  large enough so that  $2C \log N/N < \varepsilon$ , set

$$\Omega(\varepsilon, C, N) := \{z \in \mathbb{C} : C^{-1} \log N/N < \text{dist}(z, \mathcal{G}_{p,\varepsilon}) < C \log N/N, d(z) \neq 0\}, \quad (1.7)$$

where for a set  $\mathcal{B} \subset \mathbb{C}$  and  $w \in \mathbb{C}$  we denote  $\text{dist}(w, \mathcal{B}) := \inf_{w' \in \mathcal{B}} |w - w'|$ . Let  $\mathcal{N}_{\Omega(\varepsilon, C, N), N, \gamma} := |\{\lambda_i^N \in \Omega(\varepsilon, C, N)\}|$  denote the number of eigenvalues of  $P_{N, \gamma}^Q$  that lie in  $\Omega(\varepsilon, C, N)$ .

The following theorem, a combination of the convergence of  $L_N$  discussed above and Theorems 5.4, 5.7, and 7.1 below, shows that *most* eigenvalues of  $P_{N, \gamma}^Q$  lie in  $\Omega(\varepsilon, C, N)$  for appropriate  $\varepsilon, C$ .

**Theorem 1.5.** *Let Assumptions 1.1, 1.2, and 1.4 hold. Fix  $\mu > 0$  and  $\gamma > 1$ . Then there exist  $0 < \varepsilon_{1.5}, C_{1.5} < \infty$  (depending on  $\gamma, \mu$  and  $p$  only) so that*

$$\mathbf{P}(\mathcal{N}_{\Omega(\varepsilon_{1.5}, C_{1.5}, N), N, \gamma} < (1 - \mu)N) \rightarrow_{N \rightarrow \infty} 0. \quad (1.8)$$

Theorem 1.5 may be of independent interest since it improves upon previous results [7, 8, 64, 65] by providing a much sharper estimate on the position of the eigenvalues of  $P_{N, \gamma}^Q$ . We refer the reader to Sections 5–7 for its proof. In what follows, we fix  $\mu > 0$ ,  $\gamma > 1$  and consider the  $\varepsilon_{1.5}$  and  $C_{1.5}$  determined by Theorem 1.5. We then consider eigenvalues  $\lambda_i^N \in \Omega(\varepsilon_{1.5}, C_{1.5}, N)$ . By Theorem 1.5, *most* eigenvalues are of this type. Notice also that for any  $z \in \Omega(\varepsilon_{1.5}, C_{1.5}, N)$ , we have that  $d = d(z) \neq 0$ .

The main result of this paper is the following description of the (right) eigenvectors of  $P_{N, \gamma}^Q$ .

**Theorem 1.6.** *Fix  $\varepsilon_{1.5}, C_{1.5}$  and the notation as above.*

1. *The following occurs with probability approaching one as  $N \rightarrow \infty$ . For each  $\hat{z} \in \Omega(\varepsilon_{1.5}, C_{1.5}, N)$  which is an eigenvalue of  $P_{N, \gamma}^Q$ , let  $v = v(\hat{z})$  denote the corresponding (right) eigenvector, normalized so that  $\|v\|_2 = 1$ . Then there exists a vector  $w$  with  $\|w\|_2 = 1$  such that*

$$\|v - w\|_2 = o(1), \quad (1.9)$$

and a constant  $c > 0$ , depending on  $\gamma$ , so that for any  $\ell \in [N]$ ,

$$\begin{aligned} \|w\|_{\ell^2([\ell, N])} &\leq e^{-c\ell \log N/N}/c, & \text{if } d > 0, \\ \|w\|_{\ell^2([1, N-\ell])} &\leq e^{-c\ell \log N/N}/c, & \text{if } d < 0. \end{aligned} \quad (1.10)$$

The vector  $w$  can be taken as a (random) linear combination of the  $|d|$  eigenvectors of  $(P_N - \hat{z}I)^*(P_N - \hat{z}I)$  corresponding to the  $|d|$  smallest eigenvalues.

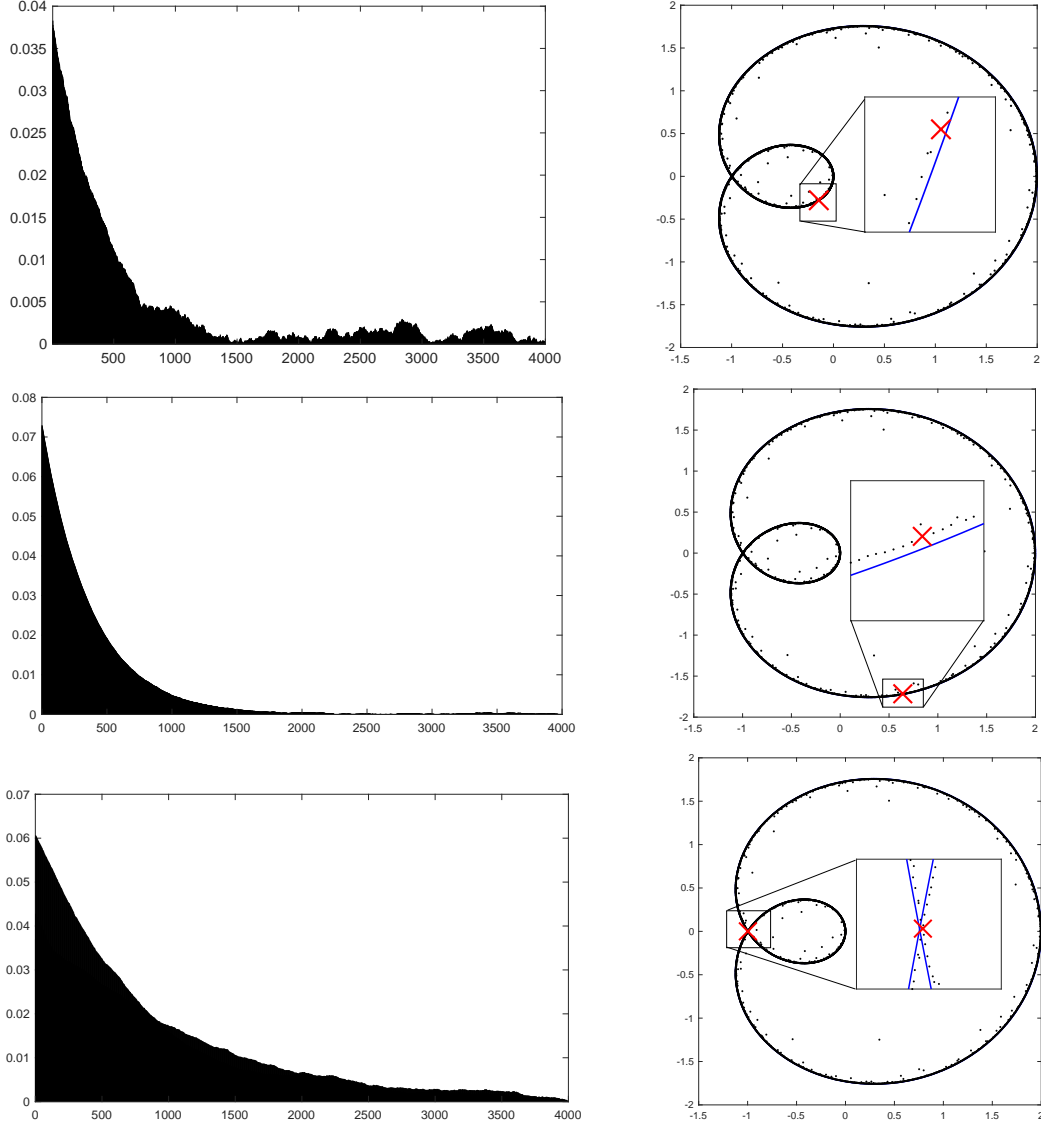
2. *Fix  $z_0 = z_0(N) \in \Omega(\varepsilon_{1.5}, C_{1.5}, N)$  deterministic,  $C_0, \tilde{C}_0$  large, and  $\eta > 0$  small. Then, there exist constants  $c_1 = c_1(\eta, C_0, \tilde{C}_0)$  and  $c_0 = c_0(\gamma) \in (0, 1)$ , with  $c_0 \rightarrow 1$  as  $\gamma \rightarrow 1$  and  $c_0 \rightarrow 0$  as  $\gamma \rightarrow \infty$ , so that, with probability at least  $1 - \eta$ , for every  $\hat{z} = \lambda_i^N \in D(z_0, C_0 \log N/N)$ , any  $0 < \ell \leq \ell' \leq \tilde{C}_0 N / \log N$  satisfying  $\ell' - \ell > N^{c_0}$  and all large  $N$ ,*

$$\begin{aligned} \|w\|_{\ell^2([\ell, \ell'])}^2 &\geq c_1(\ell' - \ell) \log N/N, & \text{if } d > 0, \\ \|w\|_{\ell^2([N-\ell', N-\ell])}^2 &\geq c_1(\ell' - \ell) \log N/N, & \text{if } d < 0. \end{aligned} \quad (1.11)$$

Further, for any  $0 < c' \leq \tilde{C}_0$ ,

$$\begin{aligned} \|v\|_{\ell^2([1, c'N/\log N])}^2 &\geq c'c_1/2, & \text{if } d > 0, \\ \|v\|_{\ell^2([N-c'N/\log N, N])}^2 &\geq c'c_1/2, & \text{if } d < 0. \end{aligned} \quad (1.12)$$

Theorem 1.6 shows a *localization* phenomenon, numerically illustrated in the examples of Figure 1: for all eigenvalues in the good regions, the corresponding eigenvectors localize at scale  $N/\log N$ , and for most eigenvalues, this is the scale at which the eigenvector is ‘‘spread out’’. (Contrast the situation with the regime  $\gamma < 1$ , where delocalization is observed in simulations, see Figure 2;



**Figure 1.** Eigenvectors (left panel) and eigenvalues (right panel) for  $N = 4000$ ,  $\gamma = 1.2$  and symbol  $\zeta + \zeta^2$ . Plotted are the moduli of the entries of the eigenvector that corresponds to the eigenvalue marked with a red  $\times$ . The top two rows correspond to situations covered by Theorem 1.6; note the localization, which occurs at scale  $N/\log N$ . The bottom row is not covered by Theorem 1.6, because the chosen eigenvalue is at vanishing distance from  $\mathcal{B}_1$ .

we discuss predictions for that regime in Section 3.1, after we introduce relevant notions and in particular the relevant *Grushin problem*.)

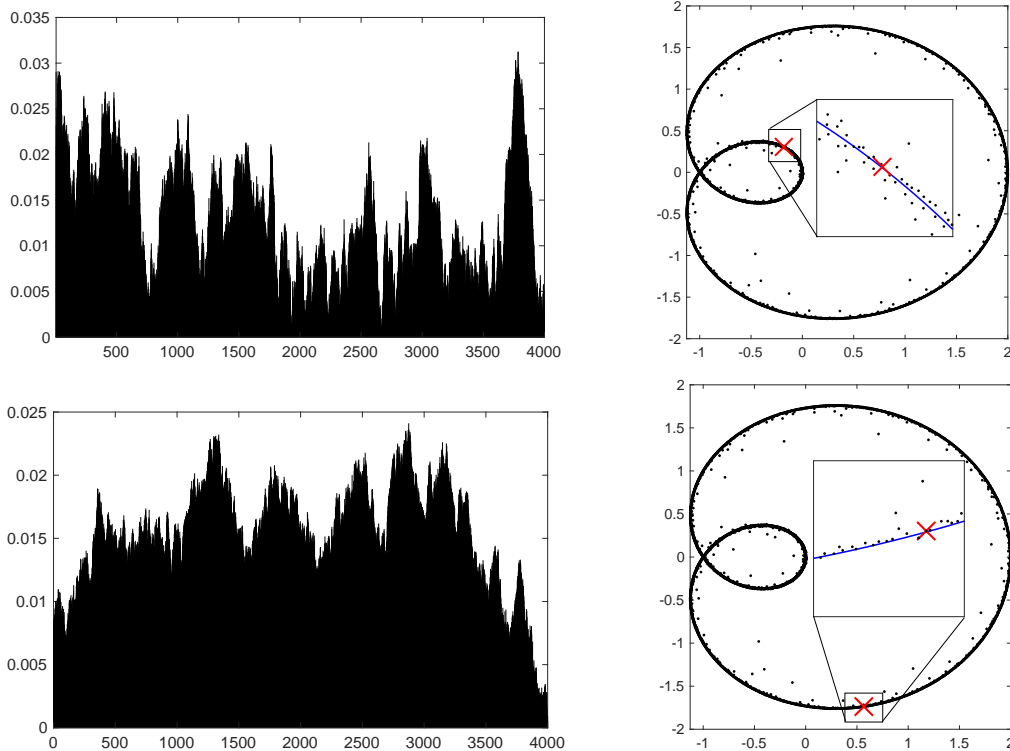
Building on Theorem 1.6, equipped with a local estimate on the number of eigenvalues  $\lambda_i^N$  in regions of diameters  $O(\log N/N)$  (see Theorem 7.4) and applying a Fubini type argument, one can show that except for an arbitrarily small fraction of the eigenvalues, the corresponding eigenvectors  $v(\lambda_i^N)$  localize at scale  $N/\log N$ . In particular, we prove the following result.

**Corollary 1.7.** *Let Assumptions 1.1, 1.2, and 1.4 hold. Then, for any  $\mu > 0$  there exists  $\mu_1, \mu_2 > 0$  so that with  $|\text{supp}_{\mu_1}(v)| := \min\{|I| : \|v\|_{\ell^2(I)} > 1 - \mu_1\}$ ,*

$$\limsup_{N \rightarrow \infty} \frac{1}{N} \mathbf{E} \#\{i : \text{supp}_{\mu_1}(v(\lambda_i^N)) < \mu_2 N / \log N\} \leq \mu.$$

Theorem 1.6 states that the eigenvectors of  $P_{N,\gamma}^Q$  corresponding to most eigenvalues  $\hat{z}$  can be approximated by a random linear combination of the eigenvectors  $e_j$  of  $(P_N - \hat{z}I)^*(P_N - \hat{z}I)$





**Figure 2.** Eigenvectors (left panel) and eigenvalues (right panel) for  $N = 4000$ ,  $\gamma = 0.8$  and symbol  $\zeta + \zeta^2$ . These cases are not covered by Theorem 1.6. Note the stark difference with the corresponding situations in Figure 1, in both the location of eigenvalues relative to  $p(S^1)$  and in the localization properties of eigenvectors.

associated with its  $|d|$  smallest eigenvalues. These are precisely the right singular vectors of  $P_N - \hat{z}I$  associated with its  $|d|$  smallest singular values. However, these singular vectors  $e_j$  are (in general) difficult objects to study and do not admit an easy description. Therefore, we will approximate these singular vectors with certain *quasimodes* of the operator  $P_N$ . The term quasimode for  $P_N$  and a (quasi-)eigenvalue  $z$  refers to a approximate  $\ell^2$ -normalized eigenvectors  $\psi \in \mathbb{C}^N$  of  $P_N - z$  in the sense that

$$\|(P_N - z)\psi\| \rightarrow 0, \quad N \rightarrow \infty.$$

In the literature quasimodes are also referred to as *pseudo-eigenvectors* or *pseudomodes*. In Section 9, we describe a  $|d|$ -dimensional space of quasimodes associated with  $P_N - zI$ , and in Section 10.2, we show that the eigenvectors  $e_j$  of  $(P_N - zI)^*(P_N - zI)$  associated with its  $|d|$  smallest eigenvalues are close to these quasimodes. We refer to Propositions 9.6 and 10.3 for the precise construction, which we do not repeat here. We emphasize however that the *construction* of the  $e_j$ -s depends on  $p$ ,  $N$  and  $z$  only and not on  $Q$  (even if eventually the value of  $z$  to which it will be applied will depend on  $Q$ ).

The upper bound (1.10) is due to the decaying nature of these  $|d|$  linearly independent quasimodes either to the left (when  $d < 0$ ) or to the right (when  $d > 0$ ). These quasimodes decay exponentially quickly,  $|u(n)| \asymp e^{-rn}$  or  $|u(n)| \asymp e^{-r(N-n)}$ , however, at different rates  $r > 0$ . Out of these  $|d|$  quasimodes the first  $(|d| - g_0)$  (recall (1.6)) quasimodes decay at a constant rate  $r > 0$ , resulting in them being completely localized to a point, i.e. either on the left or right hand side of the interval  $[1, N]$ . In contrast, the rest decay at a rate  $r \asymp \log N/N$ , which implies that they localize at a scale  $N/\log N$ . In contrast, the lower bound (1.11) follows upon showing that  $w$  has a non-negligible projection (in  $\ell^2$ ) onto  $\mathcal{S}$ , the subspace spanned by the last  $g_0$  quasimodes that decay precisely at rate  $\log N/N$ .

Note that Theorem 1.6 and Corollary 1.7 establish absences of quantum ergodicity and no-gaps delocalization, and show that the semiclassical defect measure in this setting is the Dirac measure at zero or one, depending on whether  $d$  is positive or negative.

One may wonder whether the assumption  $\hat{z} \in D(z_0, C_0 \log N/N)$  in the second part of Theorem 1.6 is optimal. It will be clear from the proof that to derive (1.11) one needs to control the supremum of the random field  $z \mapsto \tilde{f}(z)^t Q \tilde{e}(z)$  for some  $\tilde{e}(z), \tilde{f}(z) \in \mathbb{C}^N$  such that the Lipschitz norms of the functions  $z \mapsto \tilde{e}(z)$  and  $z \mapsto \tilde{f}(z)$  are  $O(N/\log N)$  and  $O(1)$ , respectively. It is then standard to check that the supremum of the field  $\{\tilde{e}(z)^t Q \tilde{f}(z)\}$  can only be bounded by an  $O(1)$  quantity in discs of radius  $O(\log N/N)$ . The boundedness of this random field is crucial in deriving that  $w$  has non-negligible projection onto  $\mathcal{S}$ . Repeating the same reasoning one can also observe that for all  $\hat{z}$  simultaneously in the good region, with probability approaching one, the  $\ell^2$  norm of the projection of  $w$  onto  $\mathcal{S}$  is at least of the order  $\log N/N$ , and we believe that this is the correct picture.

**1.3. Connection to pseudospectra and pseudo-eigenvectors.** Roughly speaking, the pseudospectrum of an operator represent the spectrum of the operator when subjected to the worst-case perturbation. However, here we study the spectrum under a *typical perturbation*. Nevertheless, in many scenarios the spectrum of random perturbations of a non-self-adjoint operator closely resembles the *pseudospectral level lines* of the unperturbed operator (cf. [7, Section 1.3] and the references therein). There have been attempts to understand pseudospectral properties of Toeplitz matrices. It was proved that the  $\varepsilon$ -pseudospectrum of an unperturbed Toeplitz matrix converges to an  $\varepsilon$ -neighborhood of the spectrum of the limiting Toeplitz operator [55]. On the other hand, in [14] it is shown that for  $P_N$  any *asymptotically good* pseudo-eigenvector must be *asymptotically strongly localized*, meaning that almost all its mass is carried either by the subset  $[1, j_N]$  or the subset  $[N - j_N, N]$  for any sequence  $\{j_N\}$  such that  $j_N \rightarrow \infty$  as  $N \rightarrow \infty$ . Note the contrast with the localization behavior of the eigenvectors of the noisy version of  $P_N$ : they localize with  $j_N \asymp N/\log N$ .

For a more general model, the twisted Toeplitz matrices, a special case of the Berezin-Toeplitz quantization of the two dimensional torus, it has been shown in [71] that if certain *(anti-)twist condition* is satisfied by the *symbol* of the operator, then the pseudo-eigenvectors are localized in the form of *localized wave packets*. This has been generalized in [13] to the Berezin-Toeplitz quantizations of compact symplectic Kähler manifolds. In the field of microlocal analysis of pseudo-differential operators, the analogue of this (anti-)twist condition is the non-vanishing of the Poisson bracket of the real and imaginary parts of the *principal symbol* of the operator. This condition, also known as Hörmander's commutator condition [39], is a corner stone in the now classical theory of local (non-)solvability of partial differential equations. In the works [19, 20, 21, 76, 23, 53] it was linked via a WKB construction to the construction of wave packet pseudomodes of Schrödinger operators with complex potentials, and more generally to nonselfadjoint pseudo-differential operators.

**1.4. Extensions.** A natural extension of our results would be to the region  $\gamma \in (1/2, 1]$ , see Section 3.1 for a discussion. We mention a couple of additional potentially interesting extensions, of broad interest.

**Sup-norm delocalization versus no-gaps delocalization.** As already mentioned in Section 1.1, there are two complementary notions of delocalizations in the random matrix literature. It is shown in [11, 57] that both these notions of delocalization hold for Wigner matrices under various assumptions on its entries. There is no reason to believe that these two notions of delocalizations should hold simultaneously in any given setting.

Indeed, from the proof of Theorem 1.6 it follows that when  $P_N$  is a Jordan block, i.e. its symbol is  $p(\zeta) = \zeta$ , and  $\gamma > 3/2$  we have that  $\|v\|_\infty \asymp \sqrt{\log N/N}$  for an eigenvector  $v$  corresponding to a bulk eigenvalue. See Remark 11.2 for further details. *Thus, in this simple setting, the eigenvectors corresponding to most of the eigenvalues are completely delocalized according to the sup-norm criterion. However, they do not satisfy no-gaps delocalization.* It is worth investigating whether one indeed has that  $\|v\|_\infty \asymp \sqrt{\log N/N}$  for all  $\gamma > 1$  and any finitely banded  $P_N$ .

**Localization for the outlier eigenvalues and multi-fractal structure.** Based on simulations and some heuristic arguments, we predict that the eigenvector  $\psi$  corresponding to an eigenvalue



residing at a distance of order  $N^{\alpha-1}$ ,  $\alpha \in (0, 1]$ , from the spectral curve localizes at scale  $N^{1-\alpha}$ . This shows in particular that for such a  $\psi$  one has  $\|\psi\|_p \asymp N^{(\alpha-1)\cdot(\frac{1}{2}-\frac{1}{p})}$ , establishing that such eigenvectors are multi-fractal. The same reasoning shows that the eigenvectors corresponding to outlier eigenvalues, i.e. those are at a distance of order one from  $p(S^1)$ , would be completely localized. That is, most of their mass is carried by finitely many entries. It seems plausible that the methods of our current work could be adapted to prove these results.

**1.5. Structure of the paper.** In Section 2, we introduce a Grushin problem for a matrix  $P$ , which allows us to represent the null space of  $P$ , if non-empty, in terms of a certain resolvent expansion, see Lemma 2.1. Section 3 then provides a sketch of the proof of Theorem 1.6. Section 4 discusses results on Toeplitz and related operators that are used later in the paper. Sections 5-7 are devoted to estimates on the location of “most” eigenvalues of  $P_{N,\gamma}^Q$ , which together yield Theorem 1.5. The following Section 8 derives estimates on the resolvent of  $P_N - zI$ , for appropriate  $z$  in the “good” region. Section 9 is devoted to the study of quasimodes of Toeplitz matrices, which are the building blocks for the eigenvectors of  $P_{N,\gamma}^Q$  and eventually give the vector  $w$  in the statement of Theorem 1.6. Section 10 gives estimates on the singular values and vectors of  $P_N - z$ , and bounds on various norms of matrices appearing in the Grushin problem. Finally, Section 11 collects all preparatory material and provides the proof of Theorem 1.6.

**1.6. Notation.** We use the following set of notation throughout this paper. The notation  $a \ll b$  means that  $Ca \leq b$  for some sufficiently large constant  $C > 0$ . Writing  $a \asymp b$  means that there exists a constant  $C > 1$  such that  $C^{-1}a \leq b \leq Ca$ . The notation  $a \gtrsim b$  is used to denote that  $a \geq C^{-1}b$ , while we write  $a = o(b)$  to denote  $a \leq C^{-1}b$  for all  $C < \infty$ . Constants  $C, \tilde{C} > 0$  denoted explicitly in the proofs may change value from line to line, while constants such as  $C_{1.2}$  will be kept fixed throughout. The notation  $f = O(N)$  means that there exists a constant  $C > 0$  (independent of  $N$ ) such that  $|f| \leq CN$ . When we want to emphasize that the constant  $C > 0$  depends on some parameter  $k$ , then we write  $C_k$ , or with the above notation  $O_k(N)$ .

We use the standard notation of ceiling and floor: for  $x \in \mathbb{R}$  we write  $\lceil x \rceil := \max\{n \in \mathbb{Z}; n \geq x\}$  and  $\lfloor x \rfloor := \max\{n \in \mathbb{Z}; n \leq x\}$ , respectively. For  $m, n \in \mathbb{Z}$  the notation  $[m, n]$  is used to denote the discrete interval  $\{m, m+1, \dots, n\}$  and we use the shorthand  $[n]$  to denote the discrete interval  $[1, n]$ .

We identify  $\ell^2(K) \simeq \ell_K^2 := \{u \in \ell^2(\mathbb{Z}); \text{supp } u \subset K\}$  and we will frequently identify  $\ell_{[0, N-1]}^2 \simeq \mathbb{C}^N$  so the norm  $\|\cdot\|$  and Hermitian scalar product  $\langle \cdot | \cdot \rangle$  on  $\mathbb{C}^N$  will be the ones of  $\ell_{[0, N-1]}^2$ . When the Euclidean norm and inner products are considered over some specific discrete interval  $I$  we write  $\|\cdot\|_{\ell^2(I)}$  and  $\langle \cdot | \cdot \rangle_{\ell^2(I)}$ . The notation  $\|\cdot\|_\infty$  is used to denote the supremum norm for a vector. For a matrix  $A$  the notations  $\|A\|$  and  $\|A\|_{\text{HS}}$  denote its operator norm, and Hilbert-Schmidt norms, respectively. For a vector  $v \in \mathbb{C}^N$  we write  $v^* \in (\mathbb{C}^N)^*$  for its natural dual with respect to the Hermitian scalar product, i.e.  $v^*u = \langle u | v \rangle$  for any  $v \in \mathbb{C}^N$ .

For a  $N \times N$  matrix  $A$  we denote by  $t_1(A) \leq \dots \leq t_N(A)$  the eigenvalues of  $\sqrt{A^*A}$ . The singular values  $s_1(A) \geq \dots \geq s_N(A)$  of  $A$  are then given by

$$s_{N-n+1}(A) = t_n(A), \quad n = 1, \dots, N. \quad (1.13)$$

For brevity, we sometimes write  $s_{\max}(A)$  and  $s_{\min}(A)$  for the maximum and the minimum singular values of  $A$ . We let  $I_d$  denote the identity matrix, viewed as element in  $\mathbb{C}^{d \times d}$ .

The following standard conventions are followed:

$$a \vee b \stackrel{\text{def}}{=} \max\{a, b\}, \quad a, b \in \mathbb{R}, \quad \prod_1^0 a_j = 1, \quad \text{and} \quad \sum_1^0 a_j = 0.$$

We also use the Dirac notation  $\delta_{n,m} = \mathbf{I}(m = n)$ . The notation  $\Theta(\cdot)$  denotes the Heaviside function, that is  $\Theta(x) = 0$  when  $x \leq 0$ , and  $\Theta(x) = 1$  for  $x > 0$ . For  $z \in \mathbb{C}$  and  $r > 0$ , we write  $D(z, r) \subset \mathbb{C}$  for the (open) disc of radius  $r$  centered at  $z$ . For  $\varepsilon > 0$  and a set  $\mathcal{B} \subset \mathbb{C}$ ,  $\mathcal{B}^\varepsilon$  denotes the  $\varepsilon$ -blow up of the set  $\mathcal{B}$ , that is the Minkowski sum of the sets  $\mathcal{B}$  and  $D(0, \varepsilon)$ . The notation  $\mathcal{B} \dot{\cup} \mathcal{B}'$  is used to denote a disjoint union. For a set  $D \subset \mathbb{C}$  we use both  $\text{cl}(D)$  and  $\overline{D}$  to denote its closure. The complement of an event  $\mathcal{A}$  is denoted  $\mathcal{A}^c$ .

**Acknowledgements.** We thank Elliot Paquette and Nick Trefethen for useful discussions at the beginning of this project, and thank Nicholas Cook for comments on an early draft of this paper. The research of AB was partially supported by DAE Project no. RTI4001 via ICTS, the Infosys Foundation via the Infosys-Chandrashekharan Virtual Centre for Random Geometry, an Infosys–ICTS Excellence Grant, a Start-up Research Grant (SRG/2019/001376) and a MATRICS grant (MTR/2019/001105) from Science and Engineering Research Board. The research of MV was partially supported by a CNRS Momentum 2017 grant and by the Agence Nationale de la Recherche under the grant ANR-20-CE40-0017. The research of OZ was partially supported by an Israel Science Foundation grant # 421/20 and by the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme (grant agreement No. 692452). Part of this work was carried out when AB and MV visited the Mathematics department of the Weizmann Institute of Science whose hospitality is gratefully acknowledged.

## 2. THE GRUSHIN PROBLEMS

We begin by setting up, in some generality, a well-posed Grushin problem, based on [73, 36], see also [65, 64]. It, and its behavior under perturbations, will play a crucial role in our analysis.

Roughly speaking, the Grushin problem amounts to replacing an operator of interest by an enlarged bijective system. In the context of linear partial differential equations, the study of such enlarged system of operators can be traced back to Grushin [35], where it was used to study hypoelliptic operators. In a different setting, such an enlarged system was used by Sjöstrand [61], whose notation we use. It has also been quite useful in bifurcation theory, numerical analysis, and for treatments of spectral problems arising in electromagnetism and quantum mechanics. See the review paper [67].

**2.1. Grushin problem for the unperturbed operator.** Let  $P$  be a complex  $N \times N$ -matrix. (In our application, we will often take  $P = P_N - zI$  where  $P_N$  is the (deterministic) Toeplitz matrix with symbol  $p$  and  $z$  is a *random* parameter close to the spectral curve  $p(S^1)$ . Then, all objects implicitly depend on  $z$ , and we suppress this dependence in notation when not needed.) Let

$$0 \leq t_1^2 \leq \dots \leq t_N^2 \quad (2.1)$$

denote the eigenvalues of  $P^*P$  with associated orthonormal basis of eigenvectors  $e_1, \dots, e_N \in \mathbb{C}^N$ . The spectra of  $P^*P$  and  $PP^*$  are equal and we can find an orthonormal basis  $f_1, \dots, f_N \in \mathbb{C}^N$  of eigenvectors of  $PP^*$  associated with the eigenvalues (2.1) such that

$$P^*f_i = t_i e_i, \quad P e_i = t_i f_i, \quad i = 1, \dots, N. \quad (2.2)$$

Let  $0 < \alpha \ll 1$  and let  $M > 0$  be the number of singular values  $t_i \in [0, \alpha]$ , i.e.

$$0 \leq t_1 \leq \dots \leq t_M \leq \alpha < t_{M+1} \leq \dots \leq t_N. \quad (2.3)$$

Let  $\delta_i$ ,  $1 \leq i \leq M$ , denote an orthonormal basis of  $\mathbb{C}^M$ . Put

$$R_+ := \sum_{i=1}^M \delta_i \circ e_i^*, \quad R_- := \sum_{i=1}^M f_i \circ \delta_i^*, \quad (2.4)$$

Then the Grushin problem

$$\mathcal{P} := \begin{pmatrix} P & R_- \\ R_+ & 0 \end{pmatrix} : \mathbb{C}^N \times \mathbb{C}^M \longrightarrow \mathbb{C}^N \times \mathbb{C}^M \quad (2.5)$$

is bijective. To see this we take  $(v, v_+) \in \mathbb{C}^N \times \mathbb{C}^M$  and proceed to solve

$$\mathcal{P} \begin{pmatrix} u \\ u_- \end{pmatrix} = \begin{pmatrix} v \\ v_+ \end{pmatrix}. \quad (2.6)$$

We write  $u = \sum_1^N u(j)e_j$  and  $v = \sum_1^N v(j)f_j$ . Similarly, we express  $u_-, v_+$  in the basis  $\delta_1, \dots, \delta_M$ . The relation (2.2) then shows that (2.6) is equivalent to

$$\begin{cases} \sum_1^N t_i u(i) f_i + \sum_1^M u_-(j) f_j = \sum_1^N v(j) f_j, \\ u(j) = v_+(j), \quad j = 1, \dots, M, \end{cases}$$

which can be written as

$$\begin{cases} t_i u(i) = v(i), & i = M+1, \dots, N, \\ \begin{pmatrix} t_i & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} u(i) \\ u_-(i) \end{pmatrix} = \begin{pmatrix} v(i) \\ v_+(i) \end{pmatrix}, & i = 1, \dots, M. \end{cases} \quad (2.7)$$

Since

$$\begin{pmatrix} t_i & 1 \\ 1 & 0 \end{pmatrix}^{-1} = \begin{pmatrix} 0 & 1 \\ 1 & -t_i \end{pmatrix},$$

we see that

$$\mathcal{P}^{-1} =: \mathcal{E} = \begin{pmatrix} E & E_+ \\ E_- & E_{-+} \end{pmatrix} \quad (2.8)$$

with

$$E = \sum_{M+1}^N \frac{1}{t_i} e_i \circ f_i, \quad E_+ = \sum_1^M e_i \circ \delta_i^*, \quad E_- = \sum_1^M \delta_i \circ f_i^*, \quad \text{and} \quad E_{-+} = - \sum_1^M t_j \delta_j \circ \delta_j^*. \quad (2.9)$$

From (2.3) and (2.9) it follows that we have the following norm estimates

$$\|E(z)\| \leq \frac{1}{\alpha}, \quad \|E_{\pm}\| = 1, \quad \|E_{-+}\| \leq \alpha. \quad (2.10)$$

Next, we recall a general fact on well-posed Grushin problems.

**Lemma 2.1.** *Let  $\mathcal{H}$  be an  $N$ -dimensional complex Hilbert space, and let  $N \geq M > 0$ . Suppose that*

$$\mathcal{P} = \begin{pmatrix} P & R_- \\ R_+ & 0 \end{pmatrix} : \mathcal{H} \times \mathbb{C}^M \longrightarrow \mathcal{H} \times \mathbb{C}^M$$

is a bijective matrix of linear operators, with inverse

$$\mathcal{E} = \begin{pmatrix} E & E_+ \\ E_- & E_{-+} \end{pmatrix}.$$

Then,  $E_+ : \mathcal{N}(E_{-+}) \rightarrow \mathcal{N}(P)$  is bijective with inverse  $R_+ \upharpoonright_{\mathcal{N}(P)}$ , and  $E_-^* : \mathcal{N}(E_{-+}^*) \rightarrow \mathcal{N}(P^*)$  is bijective with inverse  $R_-^* \upharpoonright_{\mathcal{N}(P^*)}$ .

*Proof.* From  $\mathcal{P}\mathcal{E} = 1$ , we get that  $PE_+ + R_-E_{-+} = 0$  and so

$$E_+ : \mathcal{N}(E_{-+}) \rightarrow \mathcal{N}(P). \quad (2.11)$$

Similarly, we get from  $\mathcal{E}\mathcal{P} = 1$  the equation  $E_-P + E_{-+}R_+ = 0$ , and hence

$$R_+ : \mathcal{N}(P) \rightarrow \mathcal{N}(E_{-+}). \quad (2.12)$$

The identity  $EP + E_+R_+ = 1$  yields that  $E_+R_+ = 1$  on  $\mathcal{N}(P)$ , which, together with  $R_+E_+ = 1$ , shows that (2.11), (2.12), are bijective and inverser to each other. The proof of the second claim is similar, one can follow the same arguments applied to  $\mathcal{P}^*\mathcal{E}^* = \mathcal{E}^*\mathcal{P}^* = 1$ .  $\square$

**2.2. Grushin problem for the perturbed operator.** Now we turn to the perturbed operator

$$P^\delta := P + \delta Q, \quad 0 \leq \delta \ll 1. \quad (2.13)$$

where  $Q$  is a complex  $N \times N$ -matrix (eventually, random). Let  $R_{\pm}$  be as in (2.4), and put

$$\mathcal{P}^\delta := \begin{pmatrix} P^\delta & R_- \\ R_+ & 0 \end{pmatrix} : \mathbb{C}^N \times \mathbb{C}^M \longrightarrow \mathbb{C}^N \times \mathbb{C}^M \quad (2.14)$$

with  $\mathcal{P} = \mathcal{P}^0$ . Applying  $\mathcal{E}$  (see (2.8)) from the right to (2.14) yields

$$\mathcal{P}^\delta \mathcal{E} = I_{N+M} + \begin{pmatrix} \delta QE & \delta QE_+ \\ 0 & 0 \end{pmatrix}. \quad (2.15)$$

Suppose that  $(I + \delta QE)$  is invertible. It is then straightforward to check that  $\mathcal{P}^\delta$  is invertible, with inverse

$$(\mathcal{P}^\delta)^{-1} =: \mathcal{E}^\delta = \begin{pmatrix} E^\delta & E_+^\delta \\ E_-^\delta & E_{-+}^\delta \end{pmatrix}, \quad (2.16)$$

where

$$E^\delta = E(I + \delta QE)^{-1}, \quad E_-^\delta = E_-(I + \delta QE)^{-1}, \quad (2.17)$$

$$E_{-+}^\delta = E_{-+} - E_-(I + \delta QE)^{-1}\delta QE_+, \quad (2.18)$$

and

$$E_+^\delta = E_+ - E(I + \delta QE)^{-1}\delta QE_+. \quad (2.19)$$

We note that if one takes  $P = P_N - zI_N$  with  $z$  an eigenvalue of  $P_N + \delta Q$ , then Lemma 2.1 applied to  $P^\delta$  gives a convenient description of the null-space of  $P$ , which is precisely the eigenspace of  $P_N + \delta Q$  corresponding to the eigenvalue  $z$ . This observation will be a crucial part of our analysis, see (3.1) below.

**Remark 2.2.** *Under the additional assumption that*

$$2\delta\|Q\|\alpha^{-1} \leq 1, \quad (2.20)$$

*which will occur in our setup of  $Q$  as in Assumption 1.1 if  $\delta = N^{-\gamma}$  with  $\gamma > 3/2$  (and with  $\alpha \asymp N^{-1}$ ), we obtain by a Neumann series argument that*

$$\mathcal{E}^\delta = \begin{pmatrix} E_-^\delta & E_+^\delta \\ E_-^\delta & E_{-+}^\delta \end{pmatrix} = \mathcal{E} + \sum_{n=1}^{\infty} (-\delta)^n \begin{pmatrix} E(QE)^n & (EQ)^n E_+ \\ E_-(QE)^n & E_-(QE)^{n-1} QE_+ \end{pmatrix}, \quad (2.21)$$

where by (2.20), (2.10),

$$\begin{aligned} \|E^\delta\| &= \|E(1 + \delta QE)^{-1}\| \leq 2\|E\| \leq 2\alpha^{-1}, \\ \|E_+^\delta\| &= \|(1 + \delta QE)^{-1}E_+\| \leq 2\|E_+\| \leq 2, \\ \|E_-^\delta\| &= \|E_-(1 + \delta QE)^{-1}\| \leq 2\|E_-\| \leq 2, \\ \|E_{-+}^\delta - E_{-+}\| &= \|E_-(1 + \delta QE)^{-1}\delta QE_+\| \leq 2\|\delta Q\| \leq \alpha. \end{aligned} \quad (2.22)$$

In particular, in that case,

$$E_+^\delta = E_+ - \delta QE(1 + O(\delta\|Q\|\alpha^{-1}))E_+ \quad (2.23)$$

and

$$E_{-+}^\delta = E_{-+} - \delta E_-(1 + O(\delta\|Q\|\alpha^{-1}))QE_+. \quad (2.24)$$

### 3. STRUCTURE OF THE PROOF OF THEOREM 1.6

A key ingredient for the proof Theorem 1.6 is Theorem 1.5. The proof of the latter result splits into two parts: In the first part we show that all eigenvalues must be separated from  $p(S^1)$  by a distance of the order  $\log N/N$ . At a very high level it involves an expansion of the determinant of  $P_{N,\gamma}^Q - zI_N$ , with  $z \in \mathbb{C}$ , identifying the dominant term in that expansion, and showing that the dominant cannot be equal to zero (with probability approaching one) when  $z$  is in the vicinity of the spectral curve. We refer the reader to Section 5-6 for further details on these steps. The second part of Theorem 1.5 requires us to show that most of the eigenvalues must be within a distance  $O(\log N/N)$  from spectral curve, again with probability approaching one. This is achieved by an application of Jensen's formula together with upper and lower bounds on the log-potential of  $L_N$  (see (1.5)). See Section 7 for details.

In the remainder of this section we describe the structure of the proof of Theorem 1.6, taking for granted Theorem 1.5 and various technical estimates. The proof of Theorem 1.6 begins with the Grushin problem for  $P_z^\delta = P_{N,\gamma}^Q - zI$ , see (2.16), for  $\delta = N^{-\gamma}$ ,  $z$  which is roughly an eigenvalue, and  $M = |\text{ind}_{p(S^1)}(z)|$  (this will lead to  $t_{M+1} \gtrsim \log N/N$  and  $\alpha$  bounded below by a constant multiple of  $\log N/N$ , see Proposition 10.1). To keep track of the dependence on  $z$ , throughout this section we write  $E(z)$ ,  $E_+(z)$ , etc. To relate the null-space of  $P_z^\delta$  with the null space of  $E_{-+}^\delta(z)$  we will use Lemma 2.1 in an indirect manner: As in its proof note that from (2.14) and (2.16),

$$E^\delta(z)P_z^\delta + E_+^\delta(z)R_+(z) = I \quad \text{and} \quad E_-^\delta(z)P_z^\delta + E_{-+}^\delta(z)R_+(z) = 0. \quad (3.1)$$

If  $z = \hat{z}$  were an eigenvalue of  $P_{N,\gamma}^Q$  with corresponding normalized eigenvector  $v$  then, with notation as in Section 2 and recalling the definition of  $E_+^\delta(z)$ , we would obtain from (3.1) that since  $\{e_i(\hat{z})\}$  forms an orthonormal basis of  $\mathbb{C}^N$ ,

$$\begin{aligned} \sum_{i=M+1}^N (e_i(\hat{z})^* v) \cdot e_i(\hat{z}) &= (I - E_+(\hat{z})R_+(\hat{z}))v \\ &= (I - E_+^\delta(\hat{z})R_+(\hat{z}))v - E(\hat{z})(I + \delta QE(\hat{z}))^{-1} \delta QE_+(\hat{z})R_+(\hat{z})v \\ &= E(\hat{z})(I + \delta QE(\hat{z}))^{-1} \delta QE_+(\hat{z})R_+(\hat{z})v. \end{aligned} \quad (3.2)$$

Consider first the case where  $\gamma$  is large ( $\gamma > 3/2$  will do). Since  $\|Q\| = O(N^{1/2+\epsilon})$ , for any  $\epsilon > 0$ , with high probability, we obtain that  $N^{-\gamma}\|Q\|\alpha^{-1} = o(1)$  and therefore (2.20) holds. Using then (2.23)-(2.24), the projection of  $v$  on  $\text{span}(e_i(z), i \geq M+1)$  is negligible, which yields the first part of Theorem 1.6.

To see the second part, still in the case of large  $\gamma$  (here we will need  $\gamma > 2$ ) and  $z = \hat{z}$ , we obtain from (3.1) that

$$\begin{aligned} 0 &= -E_{-+}^\delta(\hat{z})R_+(\hat{z})v - E_{-+}(\hat{z})R_+(\hat{z})v + \delta E_-(\hat{z})QE_+(\hat{z})R_+(\hat{z})v \\ &\quad - \delta^2 E_-(\hat{z})(I + \delta QE(\hat{z}))^{-1}QE(\hat{z})QE_+(\hat{z})R_+(\hat{z})v, \end{aligned} \quad (3.3)$$

where we also have used the resolvent expansion. By the same reasoning as above, the third term in (3.3) turns out to be of order  $o(\delta)$ , hence negligible compared to the first two terms. Therefore, recalling the definitions of  $E_\pm(z)$ ,  $E_{-+}(z)$ , and  $R_+(z)$  we obtain that, with  $a_j = (e_j(\hat{z})^* v)$ ,

$$\begin{aligned} -E_{-+}(\hat{z})R_+(\hat{z})v + \delta E_-(\hat{z})QE_+(\hat{z})R_+(\hat{z})v \\ = \sum_{i=1}^M a_i t_i \delta_i + \delta \sum_{i=1}^M \left[ \sum_{j=1}^M a_j (f_i(\hat{z})^* Q e_j(\hat{z})) \right] \delta_i = o(\delta). \end{aligned} \quad (3.4)$$

Now, again by Proposition 10.1, there exists  $M > M_0 \geq 0$  so that  $t_j$  decay exponentially in  $N$  for  $j \in [M_0]$ . Thus, we obtain from (3.4) that for  $\gamma$  large,

$$\left\| \sum_{i=1}^{M_0} \left[ \sum_{j=1}^M a_j \cdot (f_i(\hat{z})^* Q e_j(\hat{z})) \right] \delta_i \right\| = o(1).$$

Assume now that  $a_j = o(1)$  for  $j = M_0 + 1, \dots, M$ . Using a basic chaining argument we would then conclude that

$$\left\| \sum_{i=1}^{M_0} \left[ \sum_{j=1}^{M_0} a_j \cdot (f_i(\hat{z})^* Q e_j(\hat{z})) \right] \delta_i \right\| = \|a^\top A\| = o(1),$$

where  $A$  is the  $M_0 \times M_0$  matrix with entries  $A_{i,j} = f_i(\hat{z})^* Q e_j(\hat{z})$ . If  $\hat{z}$  were deterministic, we would have that the smallest singular value of  $A$  is  $o(1)$  and this would lead to a contradiction. Since  $\hat{z}$  is actually random, we will proceed by using the fact that the functions  $f_i, e_i$  are localized, which makes the minimal singular value of  $A$  continuous in  $z$ .

When  $\gamma \in (1, 2]$ , we cannot use in (3.2) and (3.3) an a-priori bound of the form (2.20). Instead, we use a lower bound on the minimum singular value of  $I + \delta QE$ , see (10.10), and the resolvent expansion to replace  $(I + \delta QE)^{-1}$  by  $\sum_{i=0}^L (-\delta QE)^i + (-\delta QE)^{L+1} (I - \delta QE)^{-1}$  for an appropriate  $L$ . Proposition 10.11 and the a-priori bounds on the minimal singular value of  $I + \delta QE$  suffice to control the sum, and again we use a net in order to work with deterministic  $z$ 's.

To carry out this program necessitates a fair amount of auxillary results. We need a-priori estimates for the singular values and associated quasimodes of  $P_N - zI$  for  $z$  close to the location of eigenvalues of  $P_{N,\gamma}^Q$ . So we begin in Section 4 with a fairly detailed analysis of the singular values and vectors of  $P_N - zI$  (for general  $z$ ), and relate the singular values to the winding number of  $p$  around  $z$  (we will stay away of the bad set  $\mathcal{B}_p$  of Definition 1.3, so that the winding number is locally constant and the distance of  $z$  from  $p(S^1)$  is controlled). In Sections 5-7 we provide precise estimates for the range of  $z$ s that we need to consider, that is, for the location of the eigenvalues of  $P_{N,\gamma}^Q$ . Those precise estimates are then used in Section 9 in constructing the quasimodes of

$(P_N - zI)(P_N - zI)^*$ , in terms of the roots of the symbol  $p - z$ . These quasimodes are then used in Section 10, where we show that the eigenvectors are appropriate linear combinations of the localized quasimodes.

**3.1. The case  $\gamma < 1$  - discussion and speculations.** We end this section with some brief remarks concerning  $\gamma < 1$ . In that regime, the single entries of  $\delta Q = N^{-\gamma}Q$  are larger than  $N^{-1}$ , and in particular are asymptotically larger than the distance of the eigenvalue from the spectral curve. In particular, when writing the Grushin problem (2.14), one is forced to take  $M$  growing with  $N$  (in fact, essentially  $M \asymp N^{2(1-\gamma)}$ ; This is forced by the requirement that  $\|E_-(\delta Q)E_+\| < M/N$ ). The resulting eigenvector of  $P^\delta$  are expected to be a combination of the  $M$  bottom quasimodes, with random coefficients. Since the quasimodes oscillate at scale  $N/M = N^{2\gamma-1}$ , the combination is expected to converge to a  $\gamma$ -dependent Gaussian process with correlation length of that scale. The simulations in Figure 2 are in line with this picture, although proving it require ideas going beyond the methods of this paper.

#### 4. ANALYSIS OF TOEPLITZ MATRICES

In this section we begin with presenting some fundamental results about the calculus of Toeplitz matrices. In particular we will focus on symbols given by Laurent polynomials and we will discuss the quantization procedure which maps such a symbol to an operator acting on functions on  $\mathbb{Z}$ ,  $\ell^2(\mathbb{Z})$ ,  $\ell^2([0, \infty[)$ , and  $\ell^2(\mathbb{Z}/N\mathbb{Z})$ .

**4.1. Toeplitz matrices.** We begin by recalling some well known facts about Toeplitz matrices, see for instance [15] and the references therein.

For a  $u \in \ell^2(\mathbb{Z})$  we define the Fourier transform by

$$\mathcal{F}u(\xi) = \sum_{n \in \mathbb{Z}} u_n e^{-in\xi}, \quad \xi \in \mathbb{R}/2\pi\mathbb{Z}, \quad (4.1)$$

so that  $\mathcal{F} : \ell^2(\mathbb{Z}) \rightarrow L^2(\mathbb{R}/2\pi\mathbb{Z}, \frac{d\xi}{2\pi})$  is unitary, and

$$\mathcal{F}^{-1}(f)(n) = f_n = \frac{1}{2\pi} \int_{\mathbb{R}/2\pi\mathbb{Z}} f(\xi) e^{in\xi} d\xi. \quad (4.2)$$

We consider the symbol class of continuous functions on the unit circle  $S^1$  with absolutely convergent Fourier series called the Wiener algebra

$$W := \{a \in C(S^1); \sum_{n \in \mathbb{Z}} |a_n| < +\infty\}. \quad (4.3)$$

Using the Fourier transform we can represent a symbol  $p \in W$  by

$$p(e^{i\xi}) = \sum_{n \in \mathbb{Z}} p_n e^{-in\xi}, \quad \xi \in \mathbb{R}/2\pi\mathbb{Z}. \quad (4.4)$$

We can *quantize* the symbol  $p \in W$  by

$$\text{Op}(p) \stackrel{\text{def}}{=} \mathcal{F}^{-1} p \mathcal{F}. \quad (4.5)$$

Using Parseval's equality it is easy to see that  $p(\tau) : \ell^2(\mathbb{Z}) \rightarrow \ell^2(\mathbb{Z})$  is a bounded operator; it is also straight forward to check that the  $\ell^2$  adjoint of  $\text{Op}(p)$  is given by

$$\text{Op}(p)^* = \text{Op}(\bar{p}), \quad \bar{p}(\zeta) \stackrel{\text{def}}{=} p(1/\zeta). \quad (4.6)$$

Let  $\tau$  denote the right shift operator  $(\tau\psi)(n) = \psi(n-1)$  on  $\ell^2(\mathbb{Z})$  or more generally on functions  $\psi : \mathbb{Z} \rightarrow \mathbb{C}$ . Notice that  $\tau e^{in\theta} = e^{-i\theta} e^{in\theta}$ , so the symbol of  $\tau$  is given by  $e^{-i\xi} = 1/\zeta$ ,  $\zeta \in S^1$ . So we can express (4.5) as well as

$$\text{Op}(p) = \sum_{n \in \mathbb{Z}} p_n \tau^n. \quad (4.7)$$

which can be seen to act more generally on function  $\psi : \mathbb{Z} \rightarrow \mathbb{C}$ .



Let  $K \subset \mathbb{Z}$  be a finite subset or an infinite interval. We identify  $\ell^2(K) \simeq \ell_K^2 := \{u \in \ell^2(\mathbb{Z}); \text{supp } u \subset K\}$ , and we define

$$P_K \stackrel{\text{def}}{=} P_K(p) \stackrel{\text{def}}{=} \mathbf{1}_K \text{Op}(p) \mathbf{1}_K : \ell_K^2 \rightarrow \ell_K^2. \quad (4.8)$$

When  $K$  is finite we call  $P_K$  a *Toeplitz matrix* and when  $K$  is infinite, for example  $K = ]-\infty, 0]$  or  $K = [0, \infty[$  we call  $P_K$  an *infinite Toeplitz matrix*. For simplicity we will sometimes write  $P_N = P_N(p)$  when  $K = [0, N-1]$  and  $P(p)$  when  $K = [0, \infty[$ .

Notice that the matrix elements of  $P_N$  are given by

$$P_N(\nu, \mu) = p_{\nu-\mu}, \quad \nu, \mu \in [0, N-1]. \quad (4.9)$$

A counter part to the Toeplitz matrices are the *Hankel matrices*. Let  $R$  be the reflection operator defined by  $(R\psi)(n) = \psi(-n)$  on functions  $\psi : \mathbb{Z} \rightarrow \mathbb{C}$ . Let

$$\chi_n(\zeta) = \zeta^n, \quad \zeta \in S^1, n \in \mathbb{Z}. \quad (4.10)$$

For  $a \in W$  we define the Hankel matrix of  $a$  by

$$H(a) \stackrel{\text{def}}{=} \sum_1^\infty a_n H(\chi_n), \quad H(\chi_n) = \mathbf{1}_{[0, \infty[} \tau^{n-1} R \mathbf{1}_{[0, \infty[}. \quad (4.11)$$

Represented as infinite matrices we see that Toeplitz matrices carry the same entry on the diagonals, whereas Hankel matrices carry the same entry on the anti-diagonals, for instance

$$T(a) = \begin{pmatrix} a_0 & a_{-1} & a_{-2} & \dots \\ a_1 & a_0 & a_{-1} & \dots \\ a_2 & a_1 & a_0 & \dots \\ \dots & \dots & \dots & \dots \end{pmatrix}, \quad H(a) = \begin{pmatrix} a_1 & a_2 & a_3 & \dots \\ a_2 & a_3 & \dots & \dots \\ a_3 & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots \end{pmatrix}. \quad (4.12)$$

Note that  $H(\chi_n) = 0$  for  $n \leq 0$ , and that  $H(\chi_n)$  has rank  $n$  for  $n \geq 1$ .

The quantization procedures  $p \mapsto \text{Op}(p)$  and  $p \mapsto P_K(p)$  are clearly linear. Composition of such operators is however more complicated. Given  $a, b \in W$  we have that

$$\text{Op}(ab) = \text{Op}(a)\text{Op}(b). \quad (4.13)$$

However, the composition of two Toeplitz matrices is in general not a Toeplitz matrix but only a Toeplitz matrix modulo two products of Hankel matrices, see for instance [15, Proposition 3.10]. Given  $a, b \in W$  we have, with  $\Pi_N = \mathbf{1}_{[0, N-1]}$  and  $\tilde{\Pi}_N = \Pi_N H(\chi_N) \Pi_N$ , that

$$T_N(ab) = T_N(a)T_N(b) + \Pi_N H(a)H(\tilde{b})\Pi_N + \tilde{\Pi}_N H(\tilde{a})H(b)\tilde{\Pi}_N. \quad (4.14)$$

In this paper we mainly consider the operator

$$\text{Op}(p) = \sum_{-N_-}^{N_+} a_j \tau^j, \quad a_{-N_-}, a_{-N_-+1}, \dots, a_{N_+} \in \mathbb{C}, a_{\pm N_{\pm}} \neq 0 \quad (4.15)$$

acting on  $\ell^2(\mathbb{Z})$  or, more generally, on functions  $\psi : \mathbb{Z} \rightarrow \mathbb{C}$ , whose coefficients satisfy Assumption 1.4, and with symbol  $p$  given by

$$\mathbb{C} \setminus \{0\} \ni \zeta \mapsto p(\zeta) = \sum_{-N_-}^{N_+} a_j \zeta^{-j}. \quad (4.16)$$

**4.2. Circulant matrices.** In this section we discuss circulant matrices, which are close relatives of Toeplitz matrices that play an important role in our analysis.

For  $N \geq 1$  and  $p \in W$  we consider  $\text{Op}(p)$  (4.7) acting on  $\ell^2(\mathbb{Z}/N\mathbb{Z})$  which we identify with the space of  $N$ -periodic functions on  $\mathbb{Z}$ . To distinguish this case from the other operators considered in this paper, we write

$$P_{\mathbb{Z}/N\mathbb{Z}} \stackrel{\text{def}}{=} \text{Op}(p) : \ell^2(\mathbb{Z}/N\mathbb{Z}) \rightarrow \ell^2(\mathbb{Z}/N\mathbb{Z}). \quad (4.17)$$

For  $\nu \in \mathbb{Z}/N\mathbb{Z}$  we get that

$$(P_{\mathbb{Z}/N\mathbb{Z}}u)(\nu) = \sum_{n \in \mathbb{Z}} u(\nu - n) = \sum_{\mu \in \mathbb{Z}/N\mathbb{Z}} \sum_{k \in \mathbb{Z}} p_{\nu - \mu + kN} u(\mu),$$

so the matrix elements of  $P_{\mathbb{Z}/N\mathbb{Z}}$  are given by

$$P_{\mathbb{Z}/N\mathbb{Z}}(\nu, \mu) = \sum_{k \in \mathbb{Z}} p_{\nu - \mu + kN}, \quad \mu, \nu \in \mathbb{Z}/N\mathbb{Z}. \quad (4.18)$$

Identifying  $[0, N - 1] \simeq \mathbb{Z}/N\mathbb{Z}$ , we see by (4.9) that for  $p$  as in (4.16),

$$P_{\mathbb{Z}/N\mathbb{Z}}(\nu, \mu) - P_N(\nu, \mu) = p_{\nu - \mu + N} + p_{\nu - \mu - N} \stackrel{\text{def}}{=} B(\nu, \mu), \quad \mu, \nu \in [0, N - 1], \quad (4.19)$$

where  $B$  is of rank  $N_+ + N_-$ . Using the discrete Fourier transform, with

$$\mathcal{F}_N = N^{-1/2}(\exp(-2\pi i n m / N))_{0 \leq n, m \leq N-1}$$

we may represent  $P_{\mathbb{Z}/N\mathbb{Z}}$  by

$$P_{\mathbb{Z}/N\mathbb{Z}} = \mathcal{F}_N^* \text{diag}(p(e^{2\pi i l / N}); 0 \leq l \leq N - 1) \mathcal{F}_N, \quad (4.20)$$

which immediately shows that the spectrum of  $P_{\mathbb{Z}/N\mathbb{Z}}$  is given by  $\sigma(P_{\mathbb{Z}/N\mathbb{Z}}) = \{p(e^{2\pi i l / N}); 0 \leq l \leq N - 1\}$ .

**4.3. Roots of a Laurent polynomial.** In this section we discuss the roots of the Laurent polynomial

$$p(\zeta) = \sum_{-n_-}^{n_+} a_j \zeta^{-j}, \quad -n_- \leq n_+, \quad a_{-n_-}, a_{-n_-+1}, \dots, a_{n_+} \in \mathbb{C}, \quad a_{\pm n_{\pm}} \neq 0. \quad (4.21)$$

It will be convenient to extend the symbol  $p$  to a holomorphic map  $\tilde{p} : \widehat{\mathbb{C}} \rightarrow \widehat{\mathbb{C}}$  on the extended complex plane  $\widehat{\mathbb{C}} \stackrel{\text{def}}{=} \mathbb{C} \dot{\cup} \{\infty\}$ . Here  $\widehat{\mathbb{C}}$  is a complex manifold, equipped with the topology given by  $U \in \widehat{\mathbb{C}}$  open if either  $U \subset \mathbb{C}$  open or  $\widehat{\mathbb{C}} \setminus U \subset \mathbb{C}$  is compact, and with the equivalence class of holomorphic atlases represented by  $\{\text{id} : \mathbb{C} \rightarrow \mathbb{C}, \phi : \widehat{\mathbb{C}} \setminus \{0\} \rightarrow \mathbb{C}\}$  where  $\phi(\infty) = 0$  and  $\phi(z) = 1/z$  for  $z \notin \{0, \infty\}$ . In what follows we will drop the tilde notation and denote by  $p$  also the extension.

We exclude the case of constant Laurent polynomials, i.e. we assume that

$$n_{\pm} > 0 \text{ when } n_{\mp} = 0. \quad (4.22)$$

We say that  $\zeta \in \mathbb{C} \dot{\cup} \{\infty\}$  is a root of (4.21) when

$$p(\zeta) = \sum_{-n_-}^{n_+} a_j \zeta^{-j} = 0, \quad (4.23)$$

and, by using the change of coordinates  $\zeta = 1/\omega$ , we have that  $\zeta = \infty$  is a root if  $\omega = 0$  is a root of  $p(1/\omega)$ . We keep in mind that in the sequel we will be interested in symbols whose coefficients  $a_j$  may depend on a spectral parameter  $z \in \mathbb{C}$ . For instance the coefficient of order zero of the symbol  $p(\zeta) - z$  is  $a_0 - z$ , and all the other coefficients remain independent of  $z$ .

**Lemma 4.1.** *The Laurent polynomial (4.21), assuming (4.22), has  $n = n_+ \vee 0 + n_- \vee 0$  roots in  $\mathbb{C} \dot{\cup} \{\infty\}$ . Assume that  $0 \notin p(S^1)$ , and let*

$$\zeta_1^+, \dots, \zeta_{m_+}^+ \text{ denote the roots in } D(0, 1), \quad (4.24)$$

and

$$\zeta_1^-, \dots, \zeta_{m_-}^- \text{ denote the roots in } (\mathbb{C} \dot{\cup} \{\infty\}) \setminus \overline{D(0, 1)}, \quad (4.25)$$

where  $m_+ + m_- = n$  (working with the convention that when  $m_+ = 0$  resp.  $m_- = 0$ , we have no roots (4.24) resp. (4.25)). We can distinguish the following three cases:

**Case 1** If  $n_{\pm} \geq 0$ , then we have that

$$0 < |\zeta_1^+| \leq \dots \leq |\zeta_{m_+}^+| < 1 < |\zeta_1^-| \leq \dots \leq |\zeta_{m_-}^-| < +\infty \quad (4.26)$$

**Case 2** When  $\mathfrak{N}_+ < 0$ , then we have that  $\infty$  is not a root but 0 is a root of (4.23) of multiplicity  $m_0 \in [1 \vee |\mathfrak{N}_+|, |\mathfrak{N}_-|]$ . In this case, we have that

$$0 = |\zeta_1^+| = \cdots = |\zeta_{m_0}^+| < |\zeta_{m_0+1}^+| \leq \cdots \leq |\zeta_{m_+}^+| < 1 < |\zeta_1^-| \leq \cdots \leq |\zeta_{m_-}^-| < +\infty. \quad (4.27)$$

**Case 3** When  $\mathfrak{N}_- < 0$ , then 0 is no root, but  $\infty$  is a root of multiplicity  $m_\infty \in [1 \vee |\mathfrak{N}_-|, |\mathfrak{N}_+|]$ , and we may order the roots as

$$0 < |\zeta_1^+| \leq \cdots \leq |\zeta_{m_+}^+| < 1 < |\zeta_1^-| \leq \cdots \leq |\zeta_{m_- - m_\infty}^-| < +\infty = |\zeta_{m_- - m_\infty + 1}^-| = \cdots = |\zeta_{m_-}^-|. \quad (4.28)$$

*Proof.* 1. If  $\mathfrak{N}_\pm \geq 0$ , we see that neither 0 nor  $\infty$  can be a root of (4.23), which therefore has the same roots as

$$\zeta^{\mathfrak{N}_+} p(\zeta) = \sum_0^{\mathfrak{N}_+ + \mathfrak{N}_-} a_{\mathfrak{N}_+ - j} \zeta^j = 0, \quad (4.29)$$

a polynomial of degree  $\mathfrak{N}_+ + \mathfrak{N}_-$ . We order its  $\mathfrak{N}_+ + \mathfrak{N}_-$  roots, counted with their multiplicities, as in (4.24), (4.25), with  $m_+ + m_- = \mathfrak{N}_+ + \mathfrak{N}_-$ , and we conclude (4.26).

2. When  $\mathfrak{N}_+ < 0$  then (4.23) is given by

$$p(\zeta) = \sum_{|\mathfrak{N}_+|}^{\mathfrak{N}_-} a_{-j} \zeta^j = 0, \quad (4.30)$$

a polynomial of degree  $\mathfrak{N}_-$  and so we have  $\mathfrak{N}_-$  roots in  $\mathbb{C}$  and  $\infty$  cannot be a root. It follows from (4.30) that 0 is a root with multiplicity

$$m_0 = \min\{j \in \{|\mathfrak{N}_+|, \dots, \mathfrak{N}_-\}; a_j - z\delta_{j,0} \neq 0\}, \quad (4.31)$$

where we recall the Dirac notation from Section 1.6. Notice that  $\max\{1, |\mathfrak{N}_+|\} \leq m_0 \leq |\mathfrak{N}_-|$ . In this case, we denote the roots as in (4.24), (4.25) and (4.27).

3. If  $\mathfrak{N}_- < 0$  then (4.23) is given by

$$p(\zeta) = \sum_{|\mathfrak{N}_-|}^{\mathfrak{N}_+} a_j \zeta^{-j} = 0. \quad (4.32)$$

So 0 is not a root, however,  $\infty$  maybe be one. Performing the change of variables  $\omega = 1/\zeta$  we see that  $p(1/\omega)$  satisfies the assumptions of Step 2 with the roles  $\mathfrak{N}_-$  and  $\mathfrak{N}_+$  exchanged and roots  $\omega_j^\pm = 1/\zeta_j^\pm$ . Hence, we conclude (4.28) when Case 3 holds.  $\square$

**4.4. Kernels of  $P_{[0, \infty[ - z}$  and  $P_{]-\infty, 0]} - z$ .** In this section we discuss how to construct the elements of the kernels of  $P_{[0, \infty[ - z}$  and  $P_{]-\infty, 0]} - z$ , see (4.8), where  $p$  is as in (4.16), (4.15),  $z \in \mathbb{C} \setminus p(S^1)$ .

Recall that the index of a Fredholm operator  $A$  on a Hilbert space is given by

$$\text{Ind}(A) = \dim \mathcal{N}(A) - \dim \mathcal{N}(A^*).$$

Here  $\mathcal{N}(A)$  denotes the kernel of  $A$ , and note that the kernel of  $A^*$  is isomorphic to the cokernel of  $A$ . We know from [15, Theorem 1.9], [65] that  $P_K - z$ , for  $K = ]-\infty, 0]$  and  $[0, \infty[$ , are Fredholm operators, and that the winding number  $\text{ind}_p(S^1)(z)$  of the curve  $p(S^1)$  around  $z$  is related to the Fredholm index of  $P_K - z$  as follows:

$$\text{Ind}(P_{[0, \infty[ - z}) = \text{ind}_p(S^1)(z), \quad \text{Ind}(P_{]-\infty, 0]} - z) = -\text{ind}_p(S^1)(z). \quad (4.33)$$

Using Lemma 4.1 one can express the winding number of  $p(S^1)$  around  $z$  by

$$\text{ind}_p(S^1)(z) = \frac{1}{2\pi i} \int_{S^1} \frac{d}{d\eta} \log(p(\eta) - z) d\eta = m_+ - (N_+ \vee 0) = (N_- \vee 0) - m_-. \quad (4.34)$$

**Remark 4.2.** *If the symbol  $p - z$  satisfies Case 2 of Lemma 4.1, then  $\text{ind}_{p(S^1)}(z) = m_+ > 0$ , and if it satisfies Case 3, then  $\text{ind}_{p(S^1)}(z) = -m_- < 0$ .*

*When comparing with the literature, for instance [15], then the right hand sides of the equations in (4.34) and (4.33) have the opposite sign. This discrepancy comes from our choice to write the Fourier series (4.1) with a minus instead of a plus sign in the exponential.*

Our aim is now to give explicit expressions for the kernel vectors of  $P_{[0,\infty[} - z$  and  $P_{]-\infty,0]} - z$  depending on the roots of the symbol  $p - z$ . We will only treat the case when the symbol satisfies the assumptions of Case 1 of Lemma 4.1 since the other two cases are non-generic. Furthermore, we will only work in the case where all roots of  $p - z$  are simple. We will see that under these assumptions the kernel vectors will be given by exponential solutions as in (4.37) below, corresponding to the roots in  $\mathbb{C} \setminus \{0\}$ .

4.4.1. *Exponential solutions.* We begin with a slightly more general discussion on exponential solutions. Let  $\zeta \in \mathbb{C} \setminus \{0\}$  be a root of (4.23) with multiplicity  $\text{mult}(\zeta)$ , then the exponential functions

$$\mathbb{Z} \ni \nu \mapsto f_{\zeta,k}(\nu) \stackrel{\text{def}}{=} \nu^k (\zeta^\nu), \quad 0 \leq k \leq \text{mult}(\zeta) - 1 \quad (4.35)$$

are solutions to

$$\text{Op}(p)f_{\zeta,k} = 0, \quad \text{on } \mathbb{Z}, \quad (4.36)$$

and these functions are linearly independent, by the following proposition.

**Proposition 4.3** ([65, Proposition 3.1]). *Let  $\zeta_1, \dots, \zeta_{\mathfrak{N}} \in \mathbb{C} \setminus \{0\}$  be distinct numbers and let  $1 \leq m_j < \infty$ ,  $1 \leq j \leq \mathfrak{N}$ . The functions  $f_{\zeta_j,k} : \mathbb{Z} \rightarrow \mathbb{C}$ ,  $1 \leq j \leq \mathfrak{N}$ ,  $0 \leq k \leq m_j - 1$  are linearly independent. More precisely, if  $K \subset \mathbb{Z}$  is an interval with  $\#K = m_1 + m_2 + \dots + m_{\mathfrak{N}}$ , then  $f_{\zeta_j,k}|_K$  form a basis in  $\ell^2(K)$ .*

This result together with Lemma 4.1 immediately yield the following result.

**Proposition 4.4** ([65, Proposition 3.7]). *Let  $p$  be as in (4.21), (4.22) with  $\mathfrak{N}_{\pm} \geq 0$ . Suppose that  $0 \notin p(S^1)$ , that all roots of  $p(\zeta)$  are simple and ordered as in (4.26).*

*Then, the space of exponential solutions to  $\text{Op}(p)u = 0$  is of dimension  $m_+ + m_-$ , and the general solution is of the form*

$$u(\nu) = \sum_{j=1}^{m_+} a_j^+ (\zeta_j^+)^{\nu} + \sum_{j=1}^{m_-} a_j^- (\zeta_j^-)^{\nu}, \quad a_j^{\pm} \in \mathbb{C}. \quad (4.37)$$

*The subspace of solutions decaying as  $\nu \rightarrow +\infty$  is given by*

$$a_j^- = 0, \quad \text{for } 0 \leq j \leq m_- \quad (4.38)$$

*and the subspace of solutions decaying as  $\nu \rightarrow -\infty$  is given by*

$$a_j^+ = 0, \quad \text{for } 1 \leq j \leq m_+ \quad (4.39)$$

*Proof.* The statement on the dimension and (4.37) are an immediate consequence of Lemma 4.1 and Proposition 4.3. The last statement is a consequence of the fact that  $|\zeta_j^+| < 1$  and  $|\zeta_j^-| > 1$ .  $\square$

4.4.2. *Eigenvectors in Case 1.* We recall [65, Proposition 3.6] in a slightly modified form to fit our somewhat different notation.

**Proposition 4.5.** *Let  $p$  be as in (4.21), (4.22) with  $\mathfrak{N}_{\pm} \geq 0$ . Let  $K \subset \mathbb{Z}$  be an interval of length  $\leq \mathfrak{N}_+ + \mathfrak{N}_-$ . Any function  $u : K \rightarrow \mathbb{C}$  can be extended to a solution  $\tilde{u} : \mathbb{Z} \rightarrow \mathbb{C}$  to  $\text{Op}(p)\tilde{u} = 0$ . The space of such extensions is affine of dimension  $\mathfrak{N}_+ + \mathfrak{N}_- - \#K$ . In particular the extension is unique when  $\mathfrak{N}_+ + \mathfrak{N}_- = \#K$ .*

Let  $p$  be as in (4.15), (4.16) and let  $z \in \mathbb{C} \setminus p(S^1)$ . We assume that the symbol  $p(\zeta) - z$  satisfies Case 1 of Lemma 4.1 (with  $p(\zeta) - z$  replacing  $p(\zeta)$  in (4.21)). This translates to

- $N_{\pm} > 0$ ,
  - or  $z \neq a_0$  when  $N_- = 0$ , or  $z \neq 0$  when  $N_- < 0$ ,
  - or  $z \neq a_0$  when  $N_+ = 0$ , or  $z \neq 0$  when  $N_+ < 0$ .
- (4.40)

By Lemma 4.1, we have  $m_+ + m_- = N_+ \vee 0 + N_- \vee 0$  roots of  $p(\zeta) - z$ .

We now turn to the operators  $P_{[0, \infty[} - z$  and  $P_{]-\infty, 0]} - z$ . The two cases are similar, so we focus on the first one and identify  $[0, \infty[ \simeq \mathbb{N}$  whenever convenient. The following discussion is a modified version of the one in [65], presented here for the reader's convenience.

Let  $u \in \ell^2(\mathbb{N})$  be so that  $(P_{[0, \infty[} - z)u = 0$  on  $\mathbb{N}$ . When  $N_+ > 0$  we put

$$u(-N_+) = \cdots = u(-1) = 0, \quad (4.41)$$

and when  $N_+ \leq 0$ , we do not put (4.41). Then we see that, for  $\nu = 0, 1, \dots$ ,

$$\left( \sum_{-N_+}^{N_+} a_j \tau^j - z \right) u(\nu) = 0. \quad (4.42)$$

Continuing, we then know how to extend  $u \upharpoonright_{[-(N_+ \vee 0), \infty[}$  to a function  $\tilde{u} : \mathbb{Z} \rightarrow \mathbb{C}$  satisfying  $(\text{Op}(p) - z)\tilde{u} = 0$  on  $\mathbb{Z}$ , by solving (4.42) with  $u$  replaced by  $\tilde{u}$  and for  $\nu = -1, -2, \dots$ . More precisely, the equation for  $\nu = -1$  defines uniquely  $\tilde{u}(-N_+ \vee 0 - 1)$  and the next one gives  $\tilde{u}(-N_+ \vee 0 - 2)$ . Continuing in this way we get a solution  $\tilde{u}$  of  $(\text{Op}(p) - z)\tilde{u} = 0$  on  $\mathbb{Z}$ . Since we are in Case 1, Proposition 4.4 implies that  $\tilde{u}$  is of the form (4.37).

Since  $u \in \ell^2(\mathbb{N})$ , it follows by (4.38) that

$$\tilde{u}(\nu) = \sum_{j=1}^{m_+} a_j^+ (\zeta_j^+)^{\nu}. \quad (4.43)$$

When  $N_+ > 0$  we have by construction that  $\tilde{u}(\nu) = u(\nu) = 0$  for  $\nu \in [-N_+, -1]$ , which implies that the coefficients  $a_j^+$  in (4.43) are determined by

$$0 = \mathfrak{A} \begin{pmatrix} a_1^+ \\ \vdots \\ a_{m_+}^+ \end{pmatrix}, \quad \mathfrak{A} = (\mathfrak{A}_1^+, \dots, \mathfrak{A}_{m_+}^+) \in \mathbb{C}^{N_+ \times m_+}, \quad (4.44)$$

$$\mathfrak{A}_j^+ = ((\zeta_j^+)^{\nu})_{-N_+ \leq \nu \leq -1}, \quad \text{for } j = 1, \dots, m_+.$$

Notice that  $\mathfrak{A}$  is a rectangular matrix of size  $N_+ \times m_+$ . Recall from Proposition 4.3 that the exponential functions (4.35) restricted to an interval  $K \Subset \mathbb{Z}$  of length  $|K| = m_+$  are linearly independent. Thus, the linear system in the first line of (4.44) has  $m_+ - N_+$  linearly independent solutions if  $m_+ - N_+ \geq 0$ , and none when  $m_+ - N_+ < 0$ . This implies that  $\dim \mathcal{N}(P_{[0, +\infty[} - z) = (m_+ - N_+) \vee 0$ .

Similarly, when  $N_+ \leq 0$ , we have no constraints on the coefficients in (4.43), which yields  $m_+ \vee 0$  linearly independent solutions, so  $\dim \mathcal{N}(P_{[0, +\infty[} - z) = m_+ \vee 0$ . Thus

$$\dim \mathcal{N}(P_{[0, +\infty[} - z) = (m_+ - N_+ \vee 0) \vee 0. \quad (4.45)$$

Similarly one also obtains the corresponding statements for the kernel of  $(P_{]-\infty, 0]} - z)$  with the  $N_+, m_+$  replaced by  $N_-, m_-$ . Summing up what we have proven so far, we get in view of (4.33), (4.34), see [65] for similar statements:

**Proposition 4.6.** *Let  $p$  be as in (4.15), (4.16). Let  $z \in \mathbb{C} \setminus p(S^1)$ . Assume that the assumptions of Case 1 in Lemma 4.1 hold for  $p(\zeta) - z$ , see also (4.40). Then*

$$\dim \mathcal{N}(P_{[0, +\infty[} - z) = \text{ind}_{p(S^1)}(z) \vee 0$$

and  $u \in \mathcal{N}(P_{[0, +\infty[} - z)$  if and only if  $u \in \ell^2([0, \infty[)$  is such that  $\tilde{u}$  defined above is of the form

$$\tilde{u}(\nu) = \sum_{j=1}^{m_+} a_j^+ (\zeta_j^+)^{\nu}, \quad a_j^+ \in \mathbb{C}, \quad (4.46)$$

with  $a_j^+$  additionally satisfying (4.44) when  $N_+ > 0$ . Furthermore,

$$\dim \mathcal{N}(P_{]-\infty, 0]} - z) = (-\text{ind}_{p(S^1)}(z)) \vee 0.$$

and  $u \in \mathcal{N}(P_{[-\infty, 0]} - z)$  if and only if  $u \in \ell^2([-\infty, 0])$  is of the form

$$\tilde{u}(\nu) = \sum_{j=1}^{m_-} a_j^- (\zeta_j^-)^\nu, \quad a_j^- \in \mathbb{C}, \quad (4.47)$$

with  $a_j^-$  additionally satisfying the analog of (4.44) (with  $m_+, N_+, \zeta_j^+$  replaced by  $m_-, N_-, \zeta_j^-$  and the corresponding vectors  $\mathfrak{A}_j^-$  indexed over  $1 \leq \nu \leq N_-$ ), when  $N_- > 0$ .

## 5. SEPARATION OF THE EIGENVALUES FROM THE SPECTRAL CURVE

As discussed in the introduction, in our setup all but  $o(N)$  of the eigenvalues of  $P_{N, \gamma}^Q$  lie in a small neighborhood of  $p(S^1)$ . In this section we obtain complementary, precise estimates on the location of (all) the eigenvalues for the finitely banded case: we will show that if the perturbation is small in the sense that  $\gamma > 1$ , then the eigenvalues avoid certain  $N$ -dependent regions, referred to as *forbidden tubes* (see Definition 5.1 below), around the spectral curve. These tubes are determined in terms of the roots of the Laurent polynomial  $p_z(\cdot) = 0$ , where

$$p_z(\zeta) \stackrel{\text{def}}{=} p(\zeta) - z = \sum_{-N_-}^{N_+} a_j \zeta^{-j} - z = 0, \quad (5.1)$$

with  $\zeta \in \mathbb{C} \cup \{\infty\}$  and  $z \in \mathbb{C}$ . As before, we will exclude the trivial case  $N_\pm = 0$ . Furthermore, if  $N_+ < 0$  then we modify  $p$  by setting  $a_j = 0$  for  $j = N_+ + 1, \dots, 0$ . This allows us, without loss of generality, to assume that  $N_+ \geq 0$ .

To state the main result of this section we need to introduce some notation. With the goal of being consistent with the notation of [9], we write  $\{\eta_j(z)\}_{j=1}^{\tilde{m}}$  for the negative of the roots of  $p_z(\cdot) = 0$  arranged in a non-increasing order of their moduli. For  $z \notin p(S^1)$  we let  $m_+ = m_+(z)$  denote the number of roots of  $p_z(\cdot)$  that are in  $D(0, 1)$ , while the number of roots in  $(\mathbb{C} \cup \{\infty\}) \setminus \overline{D(0, 1)}$  is denoted  $m_- = m_-(z)$ . Therefore,

$$\tilde{m} \stackrel{\text{def}}{=} m_+ + m_- = N_+ + \max\{N_-, 0\},$$

and

$$|\eta_1(z)| \geq |\eta_2(z)| \geq \dots \geq |\eta_{m_-}(z)| > 1 > |\eta_{m_-+1}(z)| \geq \dots \geq |\eta_{\tilde{m}}(z)|.$$

Compared with the previous notation, see e.g. Lemma 4.1, we have that  $\eta_i = -\zeta_{m_- - i + 1}^-$  for  $i = 1, \dots, m_-$  and  $\eta_{m_- + i} = -\zeta_{m_+ - i + 1}^+$  for  $i = 1, \dots, m_+$ . Recall the notation  $\mathbf{g}(\cdot)$  from (1.6). The parameter  $\mathbf{g}(\cdot)$  will determine the width of forbidden tubes, see Remark 5.6.

The definition of the forbidden tubes also involves the winding number of  $p(S^1)$ . For  $d \in \mathbb{Z}$  we set

$$\mathcal{S}_d \stackrel{\text{def}}{=} \{z \in \mathbb{C} \setminus p(S^1) : m_+(z) - N_+ = d\}. \quad (5.2)$$

Note that for  $z \in \mathcal{S}_d$ , the winding number  $\text{ind}_{p(S^1)}(z)$  equals  $d$ . Also observe that on  $\mathcal{S}_d$  the functions  $z \mapsto m_\pm(z)$  are constants and those common values are determined by  $d$  and  $N_\pm$ .

Recall that  $\text{cl}(A)$  denotes the closure of a set  $A \subset \mathbb{C}$ . We can now define the forbidden tubes.

**Definition 5.1** (Forbidden tubes). *Let  $p(\cdot)$  be a Laurent polynomial. For  $z \in \mathbb{C}$ , let  $p_z(\cdot)$  be as in (5.1), and write  $\mathbf{g}(p) = \mathbf{g}_0 \in \mathbb{Z}$ . Fix  $\varepsilon_0 > 0$ ,  $\gamma' > 1$ , and  $\varepsilon'_0 > 0$  such that  $\varepsilon'_0 < \varepsilon_0$ .*

- For  $d \in [0, \tilde{m}] \cap \mathbb{Z}$  we set  $\mathcal{T}_{\varepsilon'_0, \varepsilon_0}^d \stackrel{\text{def}}{=} \mathcal{T}_{\varepsilon'_0, \varepsilon_0}^{d, (1)} \cup \mathcal{T}_{\varepsilon'_0, \varepsilon_0}^{d, (2)}$ , where for  $d > 0$

$$\mathcal{T}_{\varepsilon'_0, \varepsilon_0}^{d, (1)} \stackrel{\text{def}}{=} \text{cl}\left(\left\{z \in \mathcal{S}_d : \max\{|\eta_{m_- + \mathbf{g}_0 + 1}(z)|, |\eta_{m_-}(z)|^{-1}\} \leq 1 - \varepsilon_0, \right. \right. \\ \left. \left. 1 - \varepsilon'_0 \leq |\eta_{m_- + \mathbf{g}_0}(z)| \leq |\eta_{m_- + 1}(z)| < 1\right\}\right),$$



$$\mathcal{T}_{\varepsilon'_0, \varepsilon_0}^{0, (1)} = \emptyset, \text{ and}$$

$$\mathcal{T}_{\varepsilon'_0, \varepsilon_0}^{d, (2)} \stackrel{\text{def}}{=} \left\{ z \in \mathcal{S}_d : \max \left\{ |\eta_{m_-+1}(z)|, |\eta_{m_- - g_0}(z)|^{-1} \right\} \leq 1 - \varepsilon_0, \right. \\ \left. 1 - \varepsilon'_0 \leq |\eta_{m_- - g_0+1}(z)|^{-1} \leq |\eta_{m_-}(z)|^{-1} < 1 \right\}.$$

- For  $d \in [-\tilde{m}, 0) \cap \mathbb{Z}$  we set  $\mathcal{T}_{\varepsilon'_0, \varepsilon_0}^d \stackrel{\text{def}}{=} \mathcal{T}_{\varepsilon'_0, \varepsilon_0}^{d, (1)} \cup \mathcal{T}_{\varepsilon'_0, \varepsilon_0}^{d, (2)}$ , where

$$\mathcal{T}_{\varepsilon'_0, \varepsilon_0}^{d, (1)} \stackrel{\text{def}}{=} \text{cl} \left( \left\{ z \in \mathcal{S}_d : \max \left\{ |\eta_{m_-+1}(z)|, |\eta_{m_- - g_0}(z)|^{-1} \right\} \leq 1 - \varepsilon_0, \right. \right. \\ \left. \left. 1 - \varepsilon'_0 \leq |\eta_{m_- - g_0+1}(z)|^{-1} \leq |\eta_{m_-}(z)|^{-1} < 1 \right\} \right),$$

$$\mathcal{T}_{\varepsilon'_0, \varepsilon_0}^{d, (2)} \stackrel{\text{def}}{=} \left\{ z \in \mathcal{S}_d : \max \left\{ |\eta_{m_-+g_0+1}(z)|, |\eta_{m_-}(z)|^{-1} \right\} \leq 1 - \varepsilon_0, \right. \\ \left. 1 - \varepsilon'_0 \leq |\eta_{m_-+g_0}(z)| \leq |\eta_{m_-+1}(z)| < 1 \right\}.$$

For all  $d$ , set  $\widehat{\mathcal{T}}_{\gamma', \varepsilon_0}^{d, (s)} \stackrel{\text{def}}{=} \mathcal{T}_{(\gamma'-1) \log N/N, \varepsilon_0}^{d, (s)}$ ,  $s \in \{1, 2\}$ , and write  $\widehat{\mathcal{T}}_{\gamma', \varepsilon_0}^d \stackrel{\text{def}}{=} \widehat{\mathcal{T}}_{\gamma', \varepsilon_0}^{d, (1)} \cup \widehat{\mathcal{T}}_{\gamma', \varepsilon_0}^{d, (2)}$ .

Note that  $\widehat{\mathcal{T}}_{\gamma', \varepsilon_0}^d$  is a (union of) tubes of vanishing width, whereas the width of  $\mathcal{T}_{\varepsilon'_0, \varepsilon_0}^d$  is small but fixed. When the choices of the parameters  $\gamma'$ ,  $\varepsilon_0$ ,  $\varepsilon'_0$  are clear from the context we will suppress the dependence of the tubes on these parameters, and write  $\mathcal{T}^d$ ,  $\widehat{\mathcal{T}}^d$ ,  $\mathcal{T}^{d, (s)}$ , and  $\widehat{\mathcal{T}}^{d, (s)}$ , for  $s \in \{1, 2\}$ .

**Remark 5.2.** The map  $z \mapsto \{\eta_j(z)\}_{j=1}^{\tilde{m}}$  is continuous in the symmetric product topology [74, Appendix 5, Theorem 4A]. Therefore, the map  $z \mapsto (|\eta_1(z)|, |\eta_2(z)|, \dots, |\eta_{\tilde{m}}(z)|)$  is continuous as well. This implies in particular that for  $d > 0$  and any  $z \in \mathcal{T}_{\varepsilon_0, \varepsilon_0}^{d, (1)}$ , one has that

$$\max \left\{ |\eta_{m_-+g_0+1}(z)|, |\eta_{m_-}(z)|^{-1} \right\} \leq 1 - \varepsilon_0, \quad \text{and} \quad 1 - \varepsilon'_0 \leq |\eta_{m_-+g_0}(z)| \leq |\eta_{m_-+1}(z)| \leq 1,$$

where  $m_{\pm}$  are determined by  $d$  and  $N_{\pm}$ . When  $d < 0$  a similar assertion holds for  $z \in \mathcal{T}_{\varepsilon_0, \varepsilon'_0}^{d, (1)}$ .

**Remark 5.3.** While proving the results in this section, and Sections 6 and 7 we will need to consider  $\varepsilon_0$  and  $\varepsilon'_0$  such that  $\varepsilon'_0/\varepsilon_0$  is a small fraction depending only on the degree of the Laurent polynomial  $p(\cdot)$ . The specific choices will be spelled out during the proofs.

We illustrate Definition 5.1 in some examples. Consider first the simplest setup when the Toeplitz matrix  $P_N$  is the Jordan block  $J_N$  given by  $(J_N)_{i,j} = \mathbf{1}_{j=i+1}$  for  $i, j \in [N]$ . In this case,  $\mathcal{T}^{1, (1)} = \overline{D(0, 1)} \setminus D(0, 1 - \varepsilon'_0)$ , whereas  $\mathcal{T}^{0, (2)}$  is  $\overline{D(0, 1 + \varepsilon'_0)} \setminus D(0, 1)$ . The rest of the tubes are empty. When the spectral curve  $p(S^1)$  is a Limaçon, i.e.  $P_N = J_N + J_N^2$  has symbol  $p(\zeta) = \zeta + \zeta^2$ , the union of the tubes is a region around the entire spectral curve, except for a small neighborhood, determined by  $\varepsilon_0$ , around the point on  $p(S^1)$  where the curve intersects itself. See Figure 3.

Below is the first main result of this section. Hereafter, for any  $z \in \mathbb{C}$  we use the shorthand  $P_z^\delta := P_{N, \gamma}^Q - zI_N$ .

**Theorem 5.4.** Fix  $\gamma > 1$ . Set  $\delta = N^{-\gamma}$ . Let  $p(\cdot)$  be a Laurent polynomial and  $P_N$  be the  $N \times N$  Toeplitz matrix with symbol  $p(\cdot)$ , and the entries of the noise matrix  $Q$  satisfies Assumption 1.1. Then, for any  $\tilde{\varepsilon}_0 > 0$  there exists some constant  $\widehat{c}_{\gamma, \tilde{\varepsilon}_0} > 0$ , depending on  $\tilde{\varepsilon}_0$  and  $\gamma$ , such that

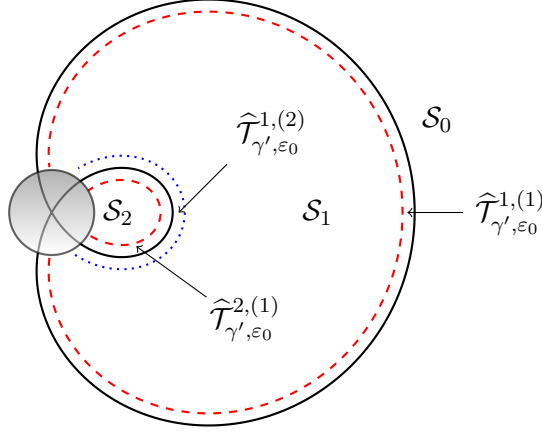
$$\lim_{N \rightarrow \infty} \mathbf{P} \left( \exists z \in \mathcal{S}_0^{\widehat{c}_{\gamma, \tilde{\varepsilon}_0} \log N/N} \setminus (\mathcal{B}_1^{\tilde{\varepsilon}_0} \cup \mathcal{B}_2^{\tilde{\varepsilon}_0}) : \det(P_z^\delta) = 0 \right) = 0.$$

**Remark 5.5.** The proof of Theorem 5.4 yields that one can take

$$\widehat{c}_{\gamma, \tilde{\varepsilon}_0} = \frac{\gamma' - 1}{\sup_{(p(S^1))^{\tilde{\varepsilon}_0} \setminus \mathcal{B}_2^{\tilde{\varepsilon}_0}} \max_{j \in [\tilde{m}]} \left| \frac{d}{dz} \eta_j(z) \right|}, \quad (5.3)$$

for any  $\gamma'$  such that  $1 < \gamma' < \gamma$ . By the Implicit Function Theorem we have that

$$\sup_{\mathbb{D}} \max_{j \in [\tilde{m}]} \left| \frac{d}{dz} \eta_j(z) \right| = O(1), \quad (5.4)$$



**Figure 3.** The different tubes for the Limaçon, where  $p(\zeta) = \zeta + \zeta^2$ ,  $N_+ = 0$ , and  $d = 0, 1, 2$  according to whether  $z$  is outside the curve, inside the larger loop, or inside the inner loop. The shaded disc is  $\mathcal{B}_1^{\tilde{\varepsilon}_0}$ , for some  $\tilde{\varepsilon}_0 > 0$  (recall Definition 1.3). Theorem 5.4 precludes the appearance of eigenvalues of small random additive perturbations of  $P_N = J_N + J_N^2$  in a region away of the shaded area that includes  $\mathcal{S}_0 \cup \hat{\mathcal{T}}_{\gamma', \varepsilon_0}^{1, (1)}$ , while Theorem 5.7 handles  $\hat{\mathcal{T}}_{\gamma', \varepsilon_0}^{1, (2)} \cup \hat{\mathcal{T}}_{\gamma', \varepsilon_0}^{2, (1)}$ .

for any bounded domain  $\mathbb{D}$  such that  $\text{dist}(\mathbb{D}, \mathcal{B}_2) > 0$ . Therefore, if we impose the additional assumption that  $\text{dist}(p(S^1), \mathcal{B}_2) > 0$ , then for  $\tilde{\varepsilon}_0 > 0$  sufficiently small the constant  $\hat{c}_{\gamma, \tilde{\varepsilon}_0}$  in Theorem 5.4 can be chosen to be free of  $\tilde{\varepsilon}_0$ . In particular, in the simplest case when  $P_N = J_N$ , the elementary Jordan block, Theorem 5.4 holds with any  $\hat{c}_{\gamma, \tilde{\varepsilon}_0} = \hat{c}_\gamma < (\gamma - 1)$ . We note that in that example with  $Q_N$  a complex Ginibre matrix, Davies and Hager [22] showed that for  $\gamma \geq 7$ , with high probability all eigenvalues are contained in  $D(0, 1 - (\gamma - 3)\log N/N)$ . Thus, Theorem 5.4 provides a sharper result for general finitely banded Toeplitz matrices, under a fairly generic assumption on  $Q_N$ .

**Remark 5.6.** Let  $\mathbf{g}_0 \in \mathbb{N}$  and consider the symbol  $p(\zeta) = \zeta^{\mathbf{g}_0}$ . Then the empirical measures of the eigenvalues of random perturbations of the corresponding Toeplitz matrices converge to the uniform measure on  $S^1$ , see [7, 8, 65]. However, using (5.3), one obtains from Theorem 5.4 that for  $\gamma > 1$ , with high probability the eigenvalues must be inside the disk of radius  $1 - \mathbf{g}_0 \cdot (\gamma' - 1)\log N/N$ , for any  $\gamma' < \gamma$ . The radius of the disk can be shown to tight upto  $o(\log N/N)$ . Thus, although the parameter  $\mathbf{g}(p)$  plays no role in determining the limit of the empirical measure of the eigenvalues of randomly perturbed Toeplitz matrices, it indeed plays a role in determining the separation of the eigenvalues from the limiting spectral curve.

Theorem 5.4 shows that all eigenvalues of  $P^\delta$  must be at a distance at least  $\hat{c}_{\gamma, \tilde{\varepsilon}_0} \log N/N$  from the portion of  $p(S^1)$  that intersects  $\text{cl}(\mathcal{S}_0)$ . To prove the localization of the eigenvectors we need to extend this picture to the regions with non-zero winding number. This is done in the result below.

**Theorem 5.7.** Fix  $\varepsilon_0 > 0$ ,  $\gamma' + 1 > \gamma > \gamma' > 1$  and  $d \in [-\tilde{m}, \tilde{m}] \cap \mathbb{Z}$ . Assume that the noise matrix  $Q$  satisfies Assumptions 1.1 and 1.2. Then, for all  $\tilde{\varepsilon}_0 > 0$ , we have

$$\lim_{N \rightarrow \infty} \mathbf{P} \left( \exists z \in \hat{\mathcal{T}}_{\gamma', \varepsilon_0}^d \setminus \mathcal{B}_2^{\tilde{\varepsilon}_0} : \det(P_z^\delta) = 0 \right) = 0.$$

While proving Theorems 5.4 and 5.7, for convenience, we will avoid neighborhoods of bad points as defined in Definition 1.3. With additional efforts, it seems possible to extend our results also to neighborhoods of some bad points, in particular those in  $\mathcal{B}_1$ . We do not pursue that direction here.

**Remark 5.8.** Theorems 5.4 and 5.7 hold for  $\gamma > 1$ . We believe, and simulations suggest (see Figure 2), that this threshold for  $\gamma$  is sharp. That is, for  $\gamma < 1$  one may find eigenvalues in  $\text{cl}(\mathcal{S}_0)$  with non-negligible probability. Furthermore, the width of the tube, namely a constant multiple of  $\log N/N$ , is also sharp. See also Theorem 7.1.

**Remark 5.9.** The papers [63, 64, 65] provide complementary estimates on the location of the eigenvalues around the curve  $p(S^1)$ , showing that for the finitely banded case,  $\gamma$  sufficiently large and any  $\delta_0 > 0$ , all but  $O(N^{2\delta_0})$  eigenvalues reside inside a tube around the spectral curve of width  $O(N^{-\delta_0})$ , with high probability.

**Remark 5.10.** The proofs of Theorems 5.4 and 5.7 yield quantitative bounds on probabilities of the events considered. From the proof it follows that, for any  $K_* \in \mathbb{N}$ , the probability that there is no eigenvalue in  $\mathcal{S}_0^{\tilde{c}_{\gamma, \tilde{\varepsilon}_0} \log N/N} \setminus (\mathcal{B}_1^{\tilde{\varepsilon}_0} \cup \mathcal{B}_2^{\tilde{\varepsilon}_0})$  can be bounded by  $O(N^{-K_*})$ . If additional assumptions on the moments of entries of  $Q$  are imposed then this bound can be further improved. For example, the assumption that the entries of  $Q$  have a uniform stretched exponential tail decay would imply a stretched exponential tail decay of the probability as well. On the other hand, the proof of Theorem 5.7 yields that the probability that there are no eigenvalues of  $P_z^\delta$  in  $\widehat{\mathcal{T}}_{\gamma', \varepsilon_0}^d \setminus \mathcal{B}_2^{\tilde{\varepsilon}_0}$  can be bounded by  $O(N^{-(\gamma-\gamma')\eta/8})$ , for  $\eta > 0$  as in Assumption 1.2.

**Remark 5.11.** The determinant of a matrix and its transpose are the same. Furthermore, if  $Q$  satisfies Assumptions 1.1 and 1.2 then so does  $Q^\top$ . Therefore, whenever  $P_N$  is a lower triangular Toeplitz matrix, to prove Theorems 5.4 and 5.7 we can work with  $P_N^\top$ . Thus, while proving these two theorems, without loss of generality, we may and will assume that  $N_- \geq 1$ .

**5.1. Decomposition of the determinant.** In this section we decompose the determinant of  $P_z^\delta$  into the sum of homogeneous polynomials  $\{\det_k(z)\}_{k=0}^N$  in the entries of the noise matrix  $Q$ , see (5.6) and (5.5) below. We then suitably preprocess  $\det_k(z)$  so that various bounds on those terms can be derived, leading to a proof of Theorems 5.4 and 5.7. At a high level, this decomposition and some of the preprocessing steps are as in [8, 9], where these were used to study the empirical measure of the eigenvalues and the process of outlier eigenvalues, respectively. However, in [8, 9] it was sufficient to derive bounds when  $z$  avoids an  $N$ -independent region around the spectral curve  $p(S^1)$ . Here, we need to consider an  $N$ -dependent region around it, and this requires better bounds and a more sophisticated analysis than in [8, 9], especially in the proof of Theorem 5.7.

For  $k \in [N]$  we set

$$\det_k(z) = \det_{k,N}(z) := \sum_{\substack{X, Y \subset [N] \\ |X|=|Y|=k}} (-1)^{\text{sgn}(\sigma_X) \text{sgn}(\sigma_Y)} \det(P_z[X^c; Y^c]) N^{-\gamma k} \det(Q[X; Y]), \quad (5.5)$$

where for a matrix  $A$  the notation  $A[X; Y]$  denotes the sub-matrix of  $A$  induced by the rows in  $X$  and columns in  $Y$ ,  $X^c := [N] \setminus X$ ,  $Y^c := [N] \setminus Y$ , and for  $Z \in \{X, Y\}$ ,  $\sigma_Z$  is the permutation on  $[N]$  which places all the elements of  $Z$  before all the elements of  $Z^c$ , but preserves the order of the elements within the two sets. Additionally denote  $\det_0(z) := \det(P_z)$ . From [8, Lemma A.1] it follows that

$$\det(P_z + N^{-\gamma}Q) = \sum_{k=0}^N \det_k(z). \quad (5.6)$$

We next represent  $\{\det_k(z)\}_{k=1}^N$  as linear combinations of products of determinants of certain bidiagonal and upper triangular Toeplitz matrices with coefficients that are determinants of sub-matrices of  $Q$ . If  $P_z$  is an upper triangular matrix, as  $\{-\eta_j(z)\}_{j=1}^{\tilde{m}}$  are the roots of the equation  $p_z(\cdot) = 0$ , it is immediate to check that

$$P_z = a_{-N_-} \cdot \prod_{j=1}^{\tilde{m}} (J_N + \eta_j(z)I_N), \quad (5.7)$$

where we recall that  $J_N$  is the elementary Jordan block given by  $(J_N)_{i,j} = \delta_{i+1,j}$  for  $i, j \in [N]$ , using the Dirac notation from Section 1.6. Then the desired representation is a consequence of the Cauchy-Binet theorem. For a general Toeplitz matrix, (5.7) does not hold. However, as in [8],  $P_z = P_N(z)$  can be viewed as a certain sub-matrix of an upper triangular finitely banded Toeplitz matrix with a slightly larger dimension, and one can use a product form for the latter coupled with the Cauchy-Binet theorem. To use this idea efficiently we borrow the following definition from [8]. As already alluded to above, without loss of generality, we will assume that  $N_+ \geq 0$  and  $N_- \geq 1$ .

**Definition 5.12** (Toeplitz with a shifted symbol). *Let  $P_N(p)$  be a Toeplitz matrix with the finite symbol*

$$p(\tau) = \sum_{j=-N_-}^{N_+} a_j \tau^j, \quad N_+ \geq 0, N_- \geq 1.$$

For  $n > N_+ + N_-$ ,  $z \in \mathbb{C}$  and  $\bar{N}_+, \bar{N}_- \in \mathbb{N}_0$  such that  $\bar{N}_+ + \bar{N}_- = N_+ + N_- \stackrel{\text{def}}{=} \tilde{N}$ , let  $P_n(p, z; \bar{N}_-)$  denote the  $n \times n$  Toeplitz matrix with symbol

$$\tilde{p}(\tau) \stackrel{\text{def}}{=} \sum_{j=-N_-}^{N_+} a'_j \tau^{j+(N_- - \bar{N}_-)},$$

where  $a'_j := a_j - z\delta_{j,0}$ ,  $j = -N_-, -N_- + 1, \dots, N_+$ .

From Definition 5.12 it follows that

$$P_{\hat{N}_+}(p, z; \tilde{N}) = \begin{bmatrix} a_{N_+} & \cdots & a_0 - z & \cdots & a_{-N_-} & 0 & \cdots & 0 \\ 0 & a_{N_+} & & a_0 - z & & \ddots & & \vdots \\ \vdots & \ddots & \ddots & & \ddots & & \ddots & \vdots \\ \vdots & & \ddots & \ddots & \ddots & & \ddots & \vdots \\ \vdots & & & \ddots & \ddots & & \ddots & a_{-N_-} \\ \vdots & & & & \ddots & \ddots & \ddots & \vdots \\ \vdots & & & & \ddots & \ddots & a_0 - z & \vdots \\ \vdots & & & & & \ddots & \ddots & \vdots \\ 0 & \cdots & \cdots & \cdots & \cdots & \cdots & 0 & a_{N_+} \end{bmatrix}, \quad \hat{N}_+ := N + N_+. \quad (5.8)$$

Therefore,

$$P_N(p_z) = P_N(p, z; N_-) = P_{\hat{N}_+}(p, z; \tilde{N})[[N]; [\hat{N}_+] \setminus [N_+]].$$

Recalling (5.5) and writing  $S + \ell$  for the Minkowski sum of the sets  $S$  and  $\{\ell\}$ , we obtain that

$$\det_k(z) = \sum_{\substack{X, Y \subset [N] \\ |X|=|Y|=k}} (-1)^{\text{sgn}(\sigma_X) \text{sgn}(\sigma_Y)} \mathfrak{D}(X, Y) \cdot N^{-\gamma k} \cdot \det(Q[X; Y]), \quad (5.9)$$

where

$$\mathfrak{D}(X, Y) = \mathfrak{D}(X, Y, z) \stackrel{\text{def}}{=} \det(P_{\hat{N}_+}(p, z; \tilde{N})[X^c; Y^c + N_+]). \quad (5.10)$$

To derive upper bounds on  $\{\det_k(z)\}_{k=0}^N$  we will use Assumption 1.1 to compute high moments of the former (and also use Markov's inequality). By (5.9), this requires the computation of high moments of  $\mathfrak{D}(X, Y)$  for any  $X, Y \subset [N]$ . To this end, we first simplify the expression for  $\mathfrak{D}(X, Y)$ . Because  $P_{\hat{N}_+}(p, z; \tilde{N})$  is an upper triangular Toeplitz matrix and  $\{\eta_j(z)\}_{j=1}^{\tilde{m}}$  are the negative of the roots of the equation  $p_z(\cdot) = 0$ , we obtain as in (5.7) that

$$P_{\hat{N}_+}(p, z; \tilde{N}) = \sum_{\ell=0}^{\tilde{N}} (a_{N_+ - \ell} - z\delta_{\ell, N_+}) J_{\hat{N}_+}^\ell = a_{-N_-} \prod_{j=1}^{\tilde{m}} (J_{\hat{N}_+} + \eta_j(z) I_{\hat{N}_+}). \quad (5.11)$$

We split the product in (5.11) into three blocks: in the first block we consider the product over the roots that are greater than one in moduli and separated from one, in the second block we will have product over the roots that are close to one in moduli, and in the third we consider the product over the rest. To write this decomposition efficiently we set

$$\hat{m}_1 = \hat{m}_1(z) \stackrel{\text{def}}{=} \begin{cases} m_- & \text{for } z \in \mathcal{T}^{d,(1)}, \\ m_- - \mathfrak{g}_0 & \text{for } z \in \mathcal{T}^{d,(2)}, \end{cases} \quad (5.12)$$

and introduce, recalling (1.6) and that  $\mathfrak{g}(p) = \mathfrak{g}_0$ ,

$$\mathcal{P}_z = \mathcal{P}_{\hat{N}_+}(z) \stackrel{\text{def}}{=} \prod_{j=\hat{m}_1+1}^{\hat{m}_1 + \mathfrak{g}_0} (J_{\hat{N}_+} + \eta_j(z) I_{\hat{N}_+}). \quad (5.13)$$

(Note that if  $\mathfrak{g}_0 = 1$  then  $\mathcal{P}_z$  is a bidiagonal matrix. The case  $\mathfrak{g}_0 > 1$ , in which  $\mathcal{P}_z$  is an upper triangular finitely banded Toeplitz matrix with bandwidth  $\mathfrak{g}_0$ , requires a special combinatorial analysis contained in Section 5.3 below.) We now obtain from (5.11) that

$$P_{\widehat{N}_+}(p, z; \widetilde{N}) = a_{-N_-} \prod_{j=1}^{\widehat{m}_1} (J_{\widehat{N}_+} + \eta_j(z) I_{\widehat{N}_+}) \cdot \mathcal{P}_z \cdot \prod_{j=\widehat{m}_1+\mathfrak{g}_0+1}^{\widetilde{m}} (J_{\widehat{N}_+} + \eta_j(z) I_{\widehat{N}_+}), \quad (5.14)$$

where, as is our convention, an empty product equals one. Notice that in the special case  $\mathfrak{g}_0 = 1$ , the decompositions (5.11) and (5.14) are identical. Before going further let us explain the advantage of the decomposition (5.14) over (5.11).

When  $z$  avoids a  $N$ -independent region around the spectral curve all the roots of  $p_z(\cdot) = 0$  are also bounded away from the unit circle. For such  $z$ 's using the decomposition (5.11) appropriate bounds on quantities analogous to  $\mathfrak{D}(X, Y, z)$  were derived in [8, 9]. Under the setup of Theorem 5.7, when  $\gamma > \mathfrak{g}_0$ , suitable adaptations of the arguments in [8, 9] yield good bounds on  $\mathfrak{D}(X, Y, z)$ , and these adaptations fail when  $\gamma \leq \mathfrak{g}_0$ . At a high level this is due to the fact that the arguments in [8, 9], while applying the Cauchy-Binet theorem to obtain a bound on  $\mathfrak{D}(X, Y, z)$ , bound each of the roots  $\{-\eta_j(z)\}_{j=1}^{\widehat{m}}$  by their moduli. Since, for any  $z \in \mathcal{T}^d$ , there are  $\mathfrak{g}_0$  many roots close to one in moduli, this requires  $\gamma > \mathfrak{g}_0$ . To push the analysis to the case  $\mathfrak{g}_0 \geq \gamma > 1$ , we will use that the roots of the symbol of the Toeplitz matrix  $\mathcal{P}_z$  are those of  $p_z(\cdot) = 0$  that are close to the unit circle. Instead of bounding each of these roots by their moduli, obtaining a bound on minors of  $\mathcal{P}_z$  by combinatorial means yields cancellations. Therefore, (5.14) is more useful than (5.11) in the proof of Theorem 5.7, when  $1 < \gamma \leq \mathfrak{g}_0$ .

We now turn to deriving a tractable representation of  $\mathfrak{D}(X, Y, z)$ . Set

$$\widehat{m}_2 = \widehat{m}_2(z) := \widetilde{m} - \widehat{m}_1(z) - \mathfrak{g}_0. \quad (5.15)$$

Fix  $k \in [N]$ , and sets  $\{X_i\}_{i=1}^{\widehat{m}_1+1}$  and  $\{Y_{i'}\}_{i'=1}^{\widehat{m}_2+1}$  such that  $|X_i| = k + N_+$ , for  $i \in [\widehat{m}_1 + 1]$ , and  $|Y_{i'}| = k + N_+$ , for  $i' \in [\widehat{m}_2 + 1]$ . Write

$$X_i := \{x_{i,1} < x_{i,2} < \cdots < x_{i,k+N_+}\}, \quad \mathcal{X}_k := (X_1, X_2, \dots, X_{\widehat{m}_1+1}), \quad (5.16)$$

$$Y_{i'} := \{y_{i',1} < y_{i',2} < \cdots < y_{i',k+N_+}\}, \quad \text{and} \quad \mathcal{Y}_k := (Y_1, Y_2, \dots, Y_{\widehat{m}_2+1}). \quad (5.17)$$

Set

$$\mathfrak{D}_1(\mathcal{X}_k) = \mathfrak{D}_1(\mathcal{X}_k, z) := \prod_{i=1}^{\widehat{m}_1} \det \left( (J_{\widehat{N}_+} + \eta_i(z) I_{\widehat{N}_+}) [\check{X}_i; \check{X}_{i+1}] \right) \quad (5.18)$$

and

$$\mathfrak{D}_2(\mathcal{Y}_k) = \mathfrak{D}_2(\mathcal{Y}_k, z) := \prod_{i'=1}^{\widehat{m}_2} \det \left( (J_{\widehat{N}_+} + \eta_{i'+\widehat{m}_1+\mathfrak{g}_0}(z) I_{\widehat{N}_+}) [\check{Y}_{i'}; \check{Y}_{i'+1}] \right), \quad (5.19)$$

where  $\check{Z} := [\widehat{N}_+] \setminus Z$  for any  $Z \subset [\widehat{N}_+]$ . We emphasize the notational difference between  $\check{Z}$  and  $Z^c$ . The former will be used to write the complement of  $Z$  when viewed as a subset of  $[\widehat{N}_+]$ , whereas for the latter  $Z$  will be viewed as a subset of  $[N]$ .

Now using the Cauchy-Binet theorem and the representation (5.14) we find that

$$\mathfrak{D}(X, Y) = \sum_{\mathcal{X}_k, \mathcal{Y}_k} a_{-N_-}^{N-k} \cdot \mathfrak{D}_1(\mathcal{X}_k) \cdot \det(\mathcal{P}_z[\check{X}_{\widehat{m}_1+1}; \check{Y}_1]) \cdot \mathfrak{D}_2(\mathcal{Y}_k), \quad (5.20)$$

where the sum above is taken over  $(\mathcal{X}_k, \mathcal{Y}_k)$  such that

$$X_1 = X \cup [\widehat{N}_+] \setminus [N] \quad \text{and} \quad Y_{\widehat{m}_2+1} = (Y + N_+) \cup [N_+]. \quad (5.21)$$

Next using [8, Lemma A.3] we find that

$$\det((J_{\widehat{N}_+} + \eta_i(z) I_{\widehat{N}_+}) [\check{X}_i; \check{X}_{i+1}]) = \eta_i(z)^{x_{i+1,1}-1} \cdot \left( \prod_{\ell=2}^{k+N_+} \eta_i(z)^{x_{i+1,\ell}-x_{i,\ell-1}} \right) \cdot \eta_i(z)^{\widehat{N}_+-x_{i,k+N_+}} \cdot \mathbf{1} \{x_{i+1,\ell} \leq x_{i,\ell} < x_{i+1,\ell+1}, \ell \in [k+N_+]\}, \quad (5.22)$$

where we have set  $x_{i+1,k+N_++1} = \widehat{N}_+ + 1$  for convenience. Thus, in light of (5.22), we may restrict the sum in (5.20) over  $\mathcal{X}_k$  belonging to

$$\widehat{L}_k^{(1)} := \{\mathcal{X}_k : 1 \leq x_{i+1,1} \leq x_{i,1} < x_{i+1,2} \leq x_{i,2} < \cdots < x_{i+1,k+N_+} \leq x_{i,k+N_+} \leq \widehat{N}_+\}. \quad (5.23)$$

It will also be convenient to further partition the set of all  $\mathcal{X}_k \in \widehat{L}_k^{(1)}$  so that  $\mathfrak{D}_1(\mathcal{X}_k)$  is constant in each block of that partition. To this end, for any  $\ell := (\ell_1, \ell_2, \dots, \ell_{\widehat{m}_1})$  with  $0 \leq \ell_i \leq \widehat{N}_+$  for  $i \in [\widehat{m}_1]$ , and  $k \in [N]$ , we let

$$L_{\ell,k}^{(1)} := \{\mathcal{X}_k \in \widehat{L}_k^{(1)} : x_{i+1,1} + \sum_{j=2}^{k+N_+} (x_{i+1,j} - x_{i,j-1}) + (\widehat{N}_+ - x_{i,k+N_+}) = \ell_i + k + N_+, \text{ for all } i = 1, 2, \dots, \widehat{m}_1\}.$$

If  $\mathcal{X}_k \in L_{\ell,k}^{(1)}$  for some  $\ell$ , by (5.22), we have that

$$\prod_{i=1}^{\widehat{m}_1} \det \left( (J_{\widehat{N}_+} + \eta_i(z) I_{\widehat{N}_+}) [\check{X}_i; \check{X}_{i+1}] \right) = \prod_{i=1}^{\widehat{m}_1} \eta_i(z)^{\ell_i}. \quad (5.24)$$

Recall that for  $z \in \mathcal{T}^d$  the roots  $\{\eta_i(z)\}_{i=1}^{\widehat{m}_1}$  are greater than one in moduli and bounded away from one. Thus, for large values of  $\{\ell_i\}_{i=1}^{\widehat{m}_1}$  the RHS of (5.24) will be exponentially (in  $N$ ) large. It would be useful to factor out this exponential factor. So, using the observation that

$$x_{i+1,1} + \sum_{j=2}^{k+N_+} (x_{i+1,j} - x_{i,j-1}) + \sum_{j=1}^{k+N_+} (x_{i,j} - x_{i+1,j}) + (\widehat{N}_+ - x_{i,k+N_+}) = \widehat{N}_+,$$

we have the following equivalent representation of  $L_{\ell,k}^{(1)}$ :

$$L_{\ell,k}^{(1)} = \left\{ \mathcal{X}_k \in \widehat{L}_k : \sum_{j=1}^{k+N_+} (x_{i,j} - x_{i+1,j} + 1) = \hat{\ell}_i; i \in [\widehat{m}_1] \right\}, \text{ where } \hat{\ell}_i := \widehat{N}_+ - \ell_i. \quad (5.25)$$

Since  $x_{i+1,j} \leq x_{i,j}$  for any  $i \in [\widehat{m}_1]$ , and  $j \in [k + N_+]$ , we further note from (5.25) that

$$\hat{\ell}_i \geq k + N_+, \quad \text{for } i = 1, 2, \dots, \widehat{m}_1. \quad (5.26)$$

This lower bound will be used later in the proof.

Equipped with the above set of notation we now find that, for any  $\mathcal{X}_k \in L_{\ell,k}^{(1)}$ ,

$$\mathfrak{D}_1(\mathcal{X}_k) = \prod_{i=1}^{\widehat{m}_1} \eta_i(z)^{\widehat{N}_+} \cdot \prod_{i=1}^{\widehat{m}_1} \eta_i(z)^{-\ell_i}. \quad (5.27)$$

Thus we indeed have that  $\mathfrak{D}_1(\mathcal{X}_k)$  is constant for  $\mathcal{X}_k \in L_{\ell,k}^{(1)}$ . We now adopt a similar decomposition of the set of indices  $\mathcal{Y}_k$  so that again inside each of the blocks  $\mathfrak{D}_2(\mathcal{Y}_k)$  will remain the same. This necessitates the following notation: By a similar reasoning to the above the sum over  $\mathcal{Y}_k$  in (5.20) can be restricted to the following set

$$\widehat{L}_k^{(2)} := \{\mathcal{Y}_k : 1 \leq y_{i'+1,1} \leq y_{i',1} < y_{i'+1,2} \leq y_{i',2} < \cdots < y_{i'+1,k+N_+} \leq y_{i',k+N_+} \leq \widehat{N}_+\}. \quad (5.28)$$

Next, for  $\mathbf{q} := (q_1, q_2, \dots, q_{\widehat{m}_2})$  with  $0 \leq q_{i'} \leq \widehat{N}_+$  for  $i' \in [\widehat{m}_2]$  we write

$$L_{\mathbf{q},k}^{(2)} := \{\mathcal{Y}_k \in \widehat{L}_k^{(2)} : y_{i'+1,1} + \sum_{j=2}^{k+N_+} (y_{i'+1,j} - y_{i',j-1}) + (\widehat{N}_+ - y_{i',k+N_+}) = q_{i'} + k + N_+, \text{ for } i' \in [\widehat{m}_2]\}. \quad (5.29)$$



Using [8, Lemma A.3] we observe that for  $\mathcal{Y}_k \in L_{\mathbf{q},k}^{(2)}$ ,

$$\mathfrak{D}_2(\mathcal{Y}_k) = \prod_{i'=1}^{\widehat{m}_2} \eta_{i'+\widehat{m}_1+\mathbf{g}_0}(z)^{q_{i'}}. \quad (5.30)$$

Now upon recalling the restriction on  $X_1$  and  $Y_{\widehat{m}_2+1}$  from (5.21), we deduce from (5.20), (5.27), and (5.30) that, for any  $X, Y \subset [N]$  such that  $|X| = |Y| = k$ , we have the representation

$$\begin{aligned} \mathfrak{D}(X, Y, z) &= a_{-N_-}^{N-k} \cdot \left[ \sum_{\mathcal{X}_k \in L_k^{(1)}} \sum_{\mathcal{Y}_k \in L_k^{(2)}} \mathfrak{D}_1(\mathcal{X}_k) \cdot \det(\mathcal{P}_z[\check{X}_{\widehat{m}_1+1}; \check{Y}_1]) \cdot \mathfrak{D}_2(\mathcal{Y}_k) \right] \\ &= a_{-N_-}^{N-k} \cdot \prod_{i=1}^{\widehat{m}_1} \eta_i(z)^{\widehat{N}_+} \cdot \left[ \sum_{\ell, \mathbf{q}} \sum_{\mathcal{X}_k \in L_{\ell,k}^{(1)}} \sum_{\mathcal{Y}_k \in L_{\mathbf{q},k}^{(2)}} \prod_{i=1}^{\widehat{m}_1} \eta_i(z)^{-\widehat{\ell}_i} \right. \\ &\quad \left. \cdot \det(\mathcal{P}_z[\check{X}_{\widehat{m}_1+1}; \check{Y}_1]) \cdot \prod_{i'=1}^{\widehat{m}_2} \eta_{i'+\widehat{m}_1+\mathbf{g}_0}(z)^{q_{i'}} \cdot \mathbf{1}_{X_1=X \cup [\widehat{N}_+] \setminus [N]} \cdot \mathbf{1}_{Y_{\widehat{m}_2+1}=(Y+N_+) \cup [N_+]} \right]. \quad (5.31) \end{aligned}$$

**5.2. Bound on  $\mathfrak{D}(X, Y)$ .** In this section we state bounds on  $\mathfrak{D}(X, Y, z)$  which will be used in Section 6.1 to compute high moments of the non-dominant terms in the expansion  $\det(P_z^\delta)$ . While stating (and deriving) such bounds it will be convenient to scale  $\mathfrak{D}(X, Y, z)$  (and  $\det_k(z)$ ) by the order of magnitude of the dominant term in the expansion of  $\det(P_z^\delta)$ . Therefore, for  $z \in \mathcal{T}^d$  we define

$$\mathfrak{K}(z) = \mathfrak{K}(z, d, \varepsilon'_0, \varepsilon_0) := a_{-N_-}^N \cdot N^{-\gamma \widehat{d}} \cdot \prod_{j=1}^{\widehat{m}_1+\mathbf{g}_0} \eta_j(z)^N, \quad (5.32)$$

where  $\widehat{m}_1$  is as in (5.12) and

$$\widehat{d} = \widehat{d}(z) := \begin{cases} d - \mathbf{g}_0 & \text{if } z \in \mathcal{T}^{d,(1)}, \\ d & \text{if } z \in \mathcal{T}^{d,(2)}. \end{cases} \quad (5.33)$$

Recalling (5.12) and (5.15), this implies that

$$\widehat{d}(\cdot) = \widehat{m}_2(\cdot) - N_+ \quad \text{on } \mathcal{T}^d. \quad (5.34)$$

For  $z \in \mathcal{T}^d$ , further set

$$\widehat{\det}_k(z) := \det_k(z) / \mathfrak{K}(z), \quad k = 0, 1, 2, \dots, N, \quad (5.35)$$

and for any  $X, Y \subset [N]$  such that  $|X| = |Y| = k$ ,

$$\widehat{\mathfrak{D}} = \widehat{\mathfrak{D}}(X, Y, z) := \frac{\mathfrak{D}(X, Y, z)}{N^{\gamma(\widehat{m}_2 - N_+)} \cdot \mathfrak{K}(z)} = \frac{\mathfrak{D}(X, Y, z)}{a_{-N_-}^N \cdot \prod_{j=1}^{\widehat{m}_1+\mathbf{g}_0} \eta_j(z)^N}, \quad (5.36)$$

where the last equality is due to (5.32) and (5.34).

The following is the first main result of this subsection. Its proof spans over this and the next subsection.

**Lemma 5.13.** *Let  $p(\cdot)$  be a Laurent polynomial with  $N_+ \geq 0, N_- \geq 1$  and  $\mathbf{g}(p) = \mathbf{g}_0 \in \mathbb{N}$ . Fix  $\varepsilon'_0, \varepsilon_0 > 0$  such that  $\varepsilon'_0/\varepsilon_0$  is sufficiently small. Fix  $X, Y \subset [N]$  such that  $|X| = |Y| = k$  so that  $X = \{x_1 < x_2 < \dots < x_k\}$  and  $Y = \{y_1 < y_2 < \dots < y_k\}$ . Assume  $\widehat{d} \geq 0$ . Then we have the following bounds on  $\widehat{\mathfrak{D}}(X, Y, z)$ , for all  $z \in \mathcal{T}_{\varepsilon'_0, \varepsilon_0}^d$ .*

(i) *There exists a constant  $C_{5.13} < \infty$  depending only on  $\varepsilon_0$  and  $p$  so that for any  $k \in \mathbb{N}$  satisfying  $k + N_+ \geq \widehat{m}_2$ ,*

$$\begin{aligned} |\widehat{\mathfrak{D}}(X, Y, z)| &\leq C_{5.13}^{\widehat{m}_1(k+N_+)} \cdot |\eta_{\widehat{m}_1+1}(z)|^{-NG(k,z)} \\ &\quad \cdot \left\{ \sum_{q \geq 0} \binom{q(\widehat{m}_2+1) + \widehat{m}_2(k+N_++1)}{\widehat{m}_2(k+N_++1)} \left(1 - \frac{\varepsilon_0}{2}\right)^q \cdot \mathfrak{J}(X, Y, q) \right\}, \quad (5.37) \end{aligned}$$

where

$$G(k, z) := \begin{cases} k + N_+ - \widehat{m}_2 & \text{if } \widehat{m}_2 \leq k + N_+ < \widehat{m}_2 + \mathbf{g}_0 \text{ and } z \in \mathcal{T}^{d,(1)}, \\ \mathbf{g}_0 & \text{if } k + N_+ \geq \widehat{m}_2 + \mathbf{g}_0 \text{ and } z \in \mathcal{T}^{d,(1)}, \\ 0 & \text{if } z \in \mathcal{T}^{d,(2)}, \end{cases} \quad (5.38)$$

and

$$\mathfrak{J}(X, Y, q) := \prod_{j=1}^{\widehat{m}_2 - N_+} \mathbf{1}_{\{y_j \leq q\}} \cdot \prod_{j=k+N_+-\widehat{m}_2+1}^k \mathbf{1}_{\{N-x_j \leq q\}}. \quad (5.39)$$

(ii) Let  $k \in \mathbb{N}$  be such that  $k + N_+ < \widehat{m}_2$ . Then, for all large  $N$ ,

$$|\widehat{\mathfrak{D}}(X, Y, z)| \leq \left(1 - \frac{\varepsilon_0}{4}\right)^N.$$

**Remark 5.14.** The roots of  $p_z(\cdot) = 0$  can be partitioned into blocks of cardinality  $\mathbf{g}(p) = \mathbf{g}_0$  such that all  $\mathbf{g}_0$  roots in any of those blocks have the same modulus. One also has that  $N_+$  is a multiple of  $\mathbf{g}_0$ . Therefore,  $d = m_+ - N_+ \geq 1$  implies that  $d \geq \mathbf{g}_0$ . On the other hand, for  $d = 0$  we have that  $\mathcal{T}^{d,(1)} = \emptyset$ . Hence, whenever needed, we will use the bound on  $\widehat{\mathfrak{D}}(X, Y, z)$  derived in Lemma 5.13 for all  $z \in \mathcal{T}^{d,(1)}$  and all  $d \geq 0$ .

Some explanations of the bounds stated in Lemma 5.13 are in order. We will see in Section 6 that, for  $k_0 \leq N$  such that  $k_0 + N_+ = \widehat{m}_2(z)$ ,  $\det_{k_0}(z)$  will be the dominant term in the expansion (5.6), and we will show in Corollary 6.13 that it has the same order of magnitude as that of  $\mathfrak{R}(z)$ . The other  $\det_k(z)$  will be negligible compared to  $\det_{k_0}(z)$ . Since the entries of the noise matrix are independent and of zero mean, one gets from (5.9) that

$$\mathbf{E}[|\det_k(z)|^2] = N^{-2\gamma k} \sum_{\substack{X, Y \subset [N] \\ |X|=|Y|=k}} |\mathfrak{D}(X, Y, z)|^2 \cdot \mathbf{E}(|\det(Q[X; Y])|^2). \quad (5.40)$$

Therefore, the bound in Lemma 5.13(ii) indeed shows that, for  $k$  such that  $k + N_+ < \widehat{m}_2(z)$  one has that  $\det_k(z)$  is *exponentially small* compared to  $\det_{k_0}(z)$ .

The implication of the bound in Lemma 5.13(i) is a bit more delicate. Notice from (5.37)-(5.39) that the number of choices of the largest  $(\widehat{m}_2 - N_+)$  elements of  $X$  is bounded by  $O(q^{\widehat{m}_2 - N_+})$ . Same holds for the smallest  $(\widehat{m}_2 - N_+)$  elements of  $Y$ . Since there is a factor in (5.37) which is exponential in  $q$  those elements of  $X$  and  $Y$  are *essentially fixed* for the purpose of computation of the second moment of  $\det_k(z)$ . This observation, as well as the factor  $|\eta_{\widehat{m}_2+1}(z)|^{-N\mathbf{G}(k,z)}$  in (5.37), are crucial in determining the correct order of magnitude for  $\det_k(z)$  for  $k > \widehat{m}_2(z) - N_+$ . Now summing over the allowable ranges of the rest of the elements of  $X$  and  $Y$ , i.e. those which are *free*, using (5.40) one obtains that, for  $k > \widehat{m}_2(z) - N_+$  and  $z \in \widehat{\mathcal{T}}_{\gamma', \varepsilon_0}^d$  with  $1 < \gamma' < \gamma$ ,  $\det_k(z)$  is *polynomially small* compared to  $\det_{k_0}(z)$ . See the proof of Lemma 6.2 for details.

The proof of Lemma 5.13 requires two auxiliary lemmas. Before stating them we introduce the notation

$$\mathbf{x}_i = \mathbf{x}_i(\mathcal{X}_k) := \sum_{j=1}^{k+N_+} x_{i,j}; \text{ for } i \in [\widehat{m}_1 + 1], \text{ and } \mathbf{y}_{i'} = \mathbf{y}_{i'}(\mathcal{Y}_k) := \sum_{j=1}^{k+N_+} y_{i',j}; \text{ for } i' \in [\widehat{m}_2 + 1]. \quad (5.41)$$

Below is the first auxiliary lemma. In its statement we use the notation of Lemma 5.13.

**Lemma 5.15.** Fix  $\ell \in [\widehat{N}_+]^{\widehat{m}_1}$  and  $\mathbf{q} \in [\widehat{N}_+]^{\widehat{m}_2}$ . For  $\mathcal{X}_k \in L_{\ell,k}^{(1)}$  and  $\mathcal{Y}_k \in L_{\ell,k}^{(2)}$ , set

$$\mathfrak{F}_1 = \mathfrak{F}_1(\mathcal{X}_k, \ell, z) := \prod_{i=1}^{\widehat{m}_1} \eta_i(z)^{-\ell_i}, \quad (5.42)$$

$$\mathfrak{F}_2 = \mathfrak{F}_2(\mathcal{Y}_k, \mathbf{q}, z) := \prod_{i=1}^{\widehat{m}_2} \eta_{\widehat{m}_1 + \mathbf{g}_0 + i}(z)^{q_i} \quad \text{and} \quad \widetilde{\mathfrak{F}}_2 = \widetilde{\mathfrak{F}}_2(\mathcal{Y}_k, \mathbf{q}, z) := \prod_{i=1}^{\widehat{m}_2} \left( \frac{\eta_{\widehat{m}_1 + \mathbf{g}_0 + i}(z)}{\eta_{\widehat{m}_1 + \mathbf{g}_0}(z)} \right)^{q_i}. \quad (5.43)$$

Assume  $\varepsilon'_0 \leq \varepsilon_0/3$ . Then, for any  $z \in \mathcal{T}_{\varepsilon'_0, \varepsilon_0}^d$  we have

$$\sum_{\ell} \sum_{\mathcal{X}_k \in L_{\ell, k}^{(1)}} |\mathfrak{F}_1| \cdot \mathbf{1}_{\{X_1 = X \cup ([\widehat{N}_+] \setminus [N])\}} \leq \left(\frac{1}{\varepsilon_0}\right)^{\widehat{m}_1(k+N_+)}, \quad (5.44)$$

$$\sum_{\mathbf{q}} \sum_{\mathcal{Y}_k \in L_{\mathbf{q}, k}^{(2)}} |\mathfrak{F}_2| \cdot \mathbf{1}_{\{Y_{\widehat{m}_2+1} = (Y+N_+) \cup [N_+]\}} \leq \left(\frac{2(\widehat{m}_2+1)}{\varepsilon_0}\right)^{\widehat{m}_2(k+N_++1)+1}, \quad (5.45)$$

and

$$\sum_{\mathbf{q}: \bar{q}=\mathbf{q}} \sum_{\mathcal{Y}_k \in L_{\mathbf{q}, k}^{(2)}} |\widetilde{\mathfrak{F}}_2| \cdot \mathbf{1}_{\{Y_{\widehat{m}_2+1} = (Y+N_+) \cup [N_+]\}} \leq \binom{q(\widehat{m}_2+1) + \widehat{m}_2(k+N_++1)}{\widehat{m}_2(k+N_++1)} \left(1 - \frac{\varepsilon_0}{2}\right)^q, \quad (5.46)$$

where

$$\bar{q} = \bar{q}(\mathbf{q}) := \sum_{i'=1}^{\widehat{m}_2} q_{i'}. \quad (5.47)$$

*Proof.* We first prove (5.44). We will iteratively sum over the collection of indices  $\mathcal{X}_k$  such that  $X_1 = X \cup ([\widehat{N}_+] \setminus [N])$ , starting from  $X_{\widehat{m}_1+1}$ . For  $\mathcal{X}_k \in L_{\ell, k}^{(1)}$  we note that  $\mathfrak{F}_1$  depends on  $\mathcal{X}_k$  only through the values of  $\ell$ . Thus, to compute the sum over  $\mathcal{X}_k$  such that  $\mathcal{X}_k \in L_{\ell, k}^{(1)}$  it is enough to find a bound on the number of possible choices for all those indices.

In the first step we keep the collection of indices  $(X_1, X_2, \dots, X_{\widehat{m}_1})$  frozen. Then, upon recalling (5.25) we observe that the number of choices of  $\{x_{\widehat{m}_1+1, j}\}_{j=1}^{k+N_+}$  is bounded above by  $\binom{\widehat{\ell}_{\widehat{m}_1-1}}{k+N_+-1}$ . We iterate this argument for any  $i \in [\widehat{m}_1]$ . The number of choices of the indices  $\{x_{i+1, j}\}_{j=1}^{k+N_+}$ , upon keeping the collection of indices  $(X_1, X_2, \dots, X_i)$  frozen, is at most  $\binom{\widehat{\ell}_i-1}{k+N_+-1}$ . This yields that

$$\sum_{\mathcal{X}_k \in L_{\ell, k}^{(1)}} |\mathfrak{F}_1| \cdot \mathbf{1}_{X_1 = X \cup ([\widehat{N}_+] \setminus [N])} \leq \prod_{i=1}^{\widehat{m}_1} \binom{\widehat{\ell}_i-1}{k+N_+-1} \cdot (1-\varepsilon_0)^{\widehat{\ell}_i}, \quad (5.48)$$

where we recalled (5.12) and used that for  $z \in \mathcal{T}^d$ ,

$$\max \left\{ \max_{j=\widehat{m}_1+g_0+1}^{\widehat{m}} |\eta_j(z)|, \max_{j=1}^{\widehat{m}_1} |\eta_j(z)|^{-1} \right\} = \max \left\{ |\eta_{\widehat{m}_1+g_0+1}(z)|, |\eta_{\widehat{m}_1}(z)|^{-1} \right\} \leq 1 - \varepsilon_0. \quad (5.49)$$

Finally, to compute the sum over  $\widehat{\ell}_1, \widehat{\ell}_2, \dots, \widehat{\ell}_{\widehat{m}_1}$  we use the following combinatorial identity: For any  $\lambda \in (0, 1)$  and  $m \in \mathbb{N}$

$$\sum_{s \geq m} \binom{s-1}{m-1} \cdot \lambda^{s-m} = (1-\lambda)^{-m}. \quad (5.50)$$

Indeed, applying the above identity with  $\lambda = (1-\varepsilon_0)$ ,  $m = k+N_+$  and  $s = \widehat{\ell}_i$  for  $i = 1, 2, \dots, \widehat{m}_1$  we deduce (5.44). Note also that we also used the lower bound  $\widehat{\ell}_i \geq k+N_+$  (see (5.26)).

We now turn to the proof of (5.45). We first prove the following intermediate step.

$$\sum_{\mathbf{q}: \bar{q}=\mathbf{q}} \sum_{\mathcal{Y}_k \in L_{\mathbf{q}, k}^{(2)}} |\mathfrak{F}_2| \cdot \mathbf{1}_{\{Y_{\widehat{m}_2+1} = (Y+N_+) \cup [N_+]\}} \leq \binom{q(\widehat{m}_2+1) + \widehat{m}_2(k+N_++1)}{\widehat{m}_2(k+N_++1)} (1-\varepsilon_0)^q. \quad (5.51)$$

To obtain (5.51), we sum over the indices  $\{Y_{i'}\}_{i'=1}^{\widehat{m}_2}$  iteratively, starting with  $Y_1$ . It is straightforward to see that, upon keeping the other indices frozen, the number of choices for the indices  $\{y_{1, j}\}_{j=1}^{k+N_+}$  such that  $\mathcal{Y}_k \in L_{\mathbf{q}, k}^{(2)}$  is bounded above by  $\binom{q_1+k+N_+}{k+N_+}$ . We use the same reasoning to

successively bound the number of allowable choices of  $Y_2, \dots, Y_{\widehat{m}_2}$ . It yields that the number of choices of  $\mathcal{Y}_k$  such that  $\mathcal{Y}_k \in L_{\mathbf{q},k}^{(2)}$  and  $Y_{\widehat{m}_2+1} = (Y + N_+) \cup [N_+]$  is bounded above by

$$\prod_{i'=1}^{\widehat{m}_2} \binom{q_{i'} + k + N_+}{k + N_+} \leq \binom{\bar{\mathbf{q}} + k + N_+}{k + N_+}^{\widehat{m}_2} \leq \binom{\widehat{m}_2(\bar{\mathbf{q}} + k + N_+)}{\widehat{m}_2(k + N_+)}$$

Thus, by (5.49), we have that the LHS of (5.51) is bounded above by

$$\sum_{\mathbf{q}: \bar{\mathbf{q}}=q} \sum_{\mathcal{Y}_k \in L_{\mathbf{q},k}^{(2)}} |\mathfrak{F}_2| \cdot \mathbf{1}_{\{Y_{\widehat{m}_2+1}=(Y+N_+)\cup[N_+]\}} \leq \binom{\widehat{m}_2(q + k + N_+)}{\widehat{m}_2(k + N_+)} \cdot |\{\mathbf{q} : \bar{\mathbf{q}} = q\}| \cdot (1 - \varepsilon_0)^q. \quad (5.52)$$

Noting that, for any  $q \geq 0$ ,

$$|\{\mathbf{q} : \bar{\mathbf{q}} = q\}| = \binom{q + \widehat{m}_2 - 1}{\widehat{m}_2 - 1} \leq \binom{q + \widehat{m}_2}{\widehat{m}_2}, \quad (5.53)$$

that  $\binom{a}{b} \binom{c}{d} \leq \sum_{k'=0}^{b+d} \binom{a}{k'} \binom{c}{b+d-k'} = \binom{a+c}{b+d}$ , and substituting in (5.52) we obtain (5.51). To derive (5.45) from (5.51) we observe that

$$(1 - \varepsilon) \leq \left(1 - \frac{\varepsilon}{2m_\star}\right)^{m_\star}, \quad \text{for } \varepsilon \in (0, 1/2) \text{ and } m_\star \in \mathbb{N}, \quad (5.54)$$

and use (5.50) again.

Finally, to prove (5.46) applying (5.49) we find that whenever  $\varepsilon'_0 \leq \varepsilon_0/3$ ,

$$\sup_{z \in \mathcal{T}^d} \left\{ \max_{i=\widehat{m}_1+\mathbf{g}_0+1}^{\widetilde{m}} \left| \frac{\eta_i(z)}{\eta_{\widehat{m}_1+\mathbf{g}_0}(z)} \right| \right\} \leq 1 - \frac{\varepsilon_0}{2},$$

Repeating the proof of (5.51), the bound in (5.46) follows.  $\square$

The next lemma is the second auxiliary result to be used in the proof of Lemma 5.13, and uses its notation.

**Lemma 5.16.** *Fix  $X_{\widehat{m}_1+1}, Y_1 \subset [N + N_+]$  such that  $|X_{\widehat{m}_1+1}| = |Y_1| = k + N_+$ , where  $k \leq N$ . If*

$$x_{\widehat{m}_1+1,j} \geq y_{1,j}; j \in [k + N_+], \quad \text{and} \quad x_{\widehat{m}_1+1,j} < y_{1,j+\mathbf{g}_0}; j \in [k + N_+ - \mathbf{g}_0], \quad (5.55)$$

then, for all  $z \in \mathcal{T}^d$ ,

$$|\det(\mathcal{P}_z[\check{X}_{\widehat{m}_1+1}; \check{Y}_1])| \leq \mathbf{g}_0^{(2+\mathbf{g}_0)(k+N_+)} \cdot |\eta_{\widehat{m}_1+1}(z)|^{\mathbf{g}_0(N-k)+\mathbf{y}_1-\mathbf{x}_{\widehat{m}_1+1}}. \quad (5.56)$$

On the other hand, if (5.55) does not hold then the LHS of (5.56) vanishes.

The proof of Lemma 5.16 is postponed to Section 5.3. In that section a Toeplitz minor would be represented as a certain *skew Schur polynomial* and using that representation Lemma 5.16 will be proved. We now prove Lemma 5.13.

*Proof of Lemma 5.13(i) for  $z \in \mathcal{T}^{d,(1)}$ .* Fix  $k \in \mathbb{N}$  such that  $k + N_+ \geq \widehat{m}_2$ . Fix  $\ell, \mathbf{q}, \mathcal{X}_k$ , and  $\mathcal{Y}_k$  such that  $\mathcal{X}_k \in L_{\ell,k}^{(1)}$ ,  $\mathcal{Y}_k \in L_{\mathbf{q},k}^{(2)}$ , and (5.21) holds. We also assume that (5.55) holds, for otherwise by the second half of Lemma 5.16 there is no contribution to the sum in (5.31).

We first show that for such pairs  $(\mathcal{X}_k, \ell)$  and  $(\mathcal{Y}_k, \mathbf{q})$  one must have

$$\mathfrak{I}(X, Y, \bar{\mathbf{q}}) = 1, \quad (5.57)$$

where we recall (5.39) and (5.47) for the definitions of  $\mathfrak{I}$  and  $\bar{\mathbf{q}}$ . To see (5.57), as  $\mathcal{Y}_k \in L_{\mathbf{q},k}^{(2)}$ , we note that for any  $j \in [\widehat{m}_2]$

$$y_{\widehat{m}_2+1,j} = \sum_{i=0}^{j-2} (y_{\widehat{m}_2+1-i,j-i} - y_{\widehat{m}_2-i,j-i-1}) + y_{\widehat{m}_2-j+2,1} \leq \sum_{i=0}^{j-1} q_{\widehat{m}_2-i} = \sum_{\lambda=\widehat{m}_2-j+1}^{\widehat{m}_2} q_\lambda \leq \bar{\mathbf{q}}. \quad (5.58)$$

Similarly, as  $\mathcal{X}_k \in L_{\ell,k}^{(1)}$  and  $\mathcal{Y}_k \in L_{\mathbf{q},k}^{(2)}$ , using that (5.55) holds, we also observe that

$$\begin{aligned} (N + N_+ - x_{1,j}) &\leq (\widehat{N}_+ - x_{\widehat{m}_1+1,j}) \stackrel{(5.55)}{\leq} (\widehat{N}_+ - y_{1,j}) \\ &= (\widehat{N}_+ - y_{k+N_++1-j,k+N_+}) + \sum_{i=0}^{k+N_+-j-1} (y_{i+2,j+i+1} - y_{i+1,j+i}) \leq \sum_{\lambda=1}^{k+N_++1-j} q_\lambda \leq \bar{\mathbf{q}}, \end{aligned} \quad (5.59)$$

where in the last step we also used that

$$j \geq k + N_+ - \widehat{m}_2 + 1 \geq 1 \quad \implies \quad k + N_+ + 1 - j \leq \widehat{m}_2.$$

The inequalities (5.58)-(5.59) together with (5.21) now establish (5.57). This implies that while summing over the indices in (5.31), without loss of generality we can assume that (5.57) holds.

Before evaluating that sum we do some more simplifications to the representation in (5.31). As  $\mathcal{X}_k \in L_{\ell,k}^{(1)}$ , recalling the definition of  $\{\mathbf{x}_i\}_{i=1}^{\widehat{m}_1+1}$  from (5.41) we see that

$$\mathbf{x}_{\widehat{m}_1+1} = \sum_{i=1}^{\widehat{m}_1} (\mathbf{x}_{i+1} - \mathbf{x}_i) + \mathbf{x}_1 = \widehat{m}_1(k + N_+) - \sum_{i=1}^{\widehat{m}_1} \widehat{\ell}_i + \mathbf{x}_1.$$

Thus, for any  $\mathcal{X}_k \in L_{\ell,k}^{(1)}$ ,

$$\begin{aligned} \widehat{\mathcal{D}}_1 &= \widehat{\mathcal{D}}_1(\mathcal{X}_k, \ell, z) := \prod_{i=1}^{\widehat{m}_1} \eta_i(z)^{-\widehat{\ell}_i} \cdot \eta_{\widehat{m}_1+1}(z)^{-\mathbf{x}_{\widehat{m}_1+1}} \\ &= \prod_{i=1}^{\widehat{m}_1} \left( \frac{\eta_i(z)}{\eta_{\widehat{m}_1+1}(z)} \right)^{-\widehat{\ell}_i} \cdot \eta_{\widehat{m}_1+1}(z)^{-\widehat{m}_1(k+N_+)} \cdot \eta_{\widehat{m}_1+1}(z)^{-\mathbf{x}_1}. \end{aligned} \quad (5.60)$$

On the other hand, for any  $\mathcal{Y}_k \in L_{\mathbf{q},k}^{(2)}$ ,

$$\mathbf{y}_1 = \mathbf{y}_2 - q_1 + (N - k) = \mathbf{y}_3 + 2(N - k) - q_1 - q_2 = \cdots = \mathbf{y}_{\widehat{m}_2+1} + \widehat{m}_2 \cdot (N - k) - \sum_{i=1}^{\widehat{m}_2} q_i.$$

This yields that

$$\begin{aligned} \widehat{\mathcal{D}}_2 &= \widehat{\mathcal{D}}_2(\mathcal{Y}_k, \ell, z) := \prod_{i=1}^{\widehat{m}_2} \eta_{\widehat{m}_1+\mathbf{g}_0+i}(z)^{q_i} \cdot \eta_{\widehat{m}_1+1}(z)^{\mathbf{y}_1} \\ &= \prod_{i=1}^{\widehat{m}_2} \left( \frac{\eta_{\widehat{m}_1+\mathbf{g}_0+i}(z)}{\eta_{\widehat{m}_1+1}(z)} \right)^{q_i} \cdot \eta_{\widehat{m}_1+1}(z)^{\widehat{m}_2(N-k) + \mathbf{y}_{\widehat{m}_2+1}}. \end{aligned} \quad (5.61)$$

Fix  $k$  such that  $k + N_+ \geq \widehat{m}_2 + \mathbf{g}_0$ . For  $\mathcal{X}_k \in L_{\ell,k}^{(1)}$  and  $\mathcal{Y}_k \in \widehat{L}_k^{(2)}$ , upon using that  $\widehat{m}_2 \geq 0$  and (5.55), we find that

$$\begin{aligned} \mathbf{y}_{\widehat{m}_2+1} &\stackrel{(5.41)}{\geq} \sum_{j=\widehat{m}_2+\mathbf{g}_0+1}^{k+N_+} y_{\widehat{m}_2+1,j} \stackrel{(5.28)}{\geq} \sum_{j=\mathbf{g}_0+1}^{k+N_+-\widehat{m}_2} y_{1,j} \stackrel{(5.55)}{\geq} \sum_{j=1}^{k+N_+-\widehat{m}_2-\mathbf{g}_0} x_{\widehat{m}_1+1,j} \\ &\stackrel{(5.26)}{\geq} \sum_{j=1}^{k+N_+-\widehat{m}_2-\mathbf{g}_0} x_{1,j} - \sum_{i=1}^{\widehat{m}_1} \widehat{\ell}_i + (k + N_+) \widehat{m}_1 \\ &\stackrel{(5.41)}{\geq} \mathbf{x}_1 - \sum_{i=1}^{\widehat{m}_1} \widehat{\ell}_i + (k + N_+) \widehat{m}_1 - \widehat{N}_+(\widehat{m}_2 + \mathbf{g}_0). \end{aligned} \quad (5.62)$$

Note that for  $k \in \mathbb{N}$  such that  $\widehat{m}_2 \leq k + N_+ \leq \widehat{m}_2 + \mathbf{g}_0$ , the sum over  $j$  in the left side of (5.62) is empty. Therefore, by (5.26) and as  $X_1 \subset [\widehat{N}_+]$  we obtain that for such values of  $k$ ,

$$\mathbf{y}_{\widehat{m}_2+1} \geq \mathbf{x}_1 - \sum_{i=1}^{\widehat{m}_1} \widehat{\ell}_i + (k + N_+) \widehat{m}_1 - \widehat{N}_+(k + N_+). \quad (5.63)$$

Now as  $z \in \mathcal{T}^{d,(1)}$  we have  $|\eta_{\widehat{m}_1+1}(z)| \leq 1$ . Thus, recalling (5.42)-(5.43) and (5.38), and using (5.62)-(5.63) we deduce that

$$|\widehat{\mathfrak{D}}_1| \cdot |\widehat{\mathfrak{D}}_2| \leq |\mathfrak{F}_1| \cdot |\mathfrak{F}_2| \cdot |\eta_{\widehat{m}_1+1}(z)|^{-NG(k,z) - \widetilde{m}(k+N_+)}. \quad (5.64)$$

On the other hand, from (5.60)-(5.61), (5.31), and (5.36) it follows that

$$\begin{aligned} \left| \widehat{\mathfrak{D}}(X, Y, z) \right| &\leq |a_{-N_-}|^{-k} \prod_{i=1}^{\widehat{m}_1} |\eta_i(z)|^{N_+} \prod_{i=\widehat{m}_1+1}^{\widehat{m}_1+\mathbf{g}_0} |\eta_i(z)|^{-N} \cdot \left[ \sum_{\ell, \mathbf{q}} \sum_{\mathcal{X}_k \in L_{\ell, k}^{(1)}} \sum_{\mathcal{Y}_k \in L_{\mathbf{q}, k}^{(2)}} |\widehat{\mathfrak{D}}_1| \right. \\ &\quad \cdot \left. \frac{|\det(\mathcal{P}_z[\check{X}_{\widehat{m}_1+1}; \check{Y}_1])|}{|\eta_{\widehat{m}_1+1}(z)^{\mathbf{y}_1 - \mathbf{x}_{\widehat{m}_1+1}}|} \cdot |\widehat{\mathfrak{D}}_2| \cdot \mathbf{1}_{X_1 = X \cup [\widehat{N}_+] \setminus [N]} \cdot \mathbf{1}_{Y_{\widehat{m}_2+1} = (Y+N_+) \cup [N_+]} \cdot \mathfrak{J}(X, Y, \bar{\mathbf{q}}) \right], \end{aligned} \quad (5.65)$$

where we also used that  $\mathbf{g}(p) = \mathbf{g}_0$  (recall (1.6) for the definition of  $\mathbf{g}(\cdot)$ ) implies that

$$|\eta_{\widehat{m}_1+1}(z)| = |\eta_{\widehat{m}_1+2}(z)| = \cdots = |\eta_{\widehat{m}_1+\mathbf{g}_0}(z)|. \quad (5.66)$$

It is clear that if  $z \in \mathcal{T}^d$  for some  $d$  then  $z$  lies in compact domain in  $\mathbb{C}$ . As the map  $z \mapsto \max_{j=1}^{\widetilde{m}} |\eta_j(z)|$  is continuous, we therefore have that

$$\max_d \sup_{z \in \mathcal{T}^d} \max_{j=1}^{\widetilde{m}} |\eta_j(z)| < \infty.$$

Therefore, continuing from above, applying Lemma 5.16, from (5.57) and (5.64) we deduce that, there exists some constant  $\widetilde{C}_{5.13} < \infty$  such that

$$\begin{aligned} \left| \widehat{\mathfrak{D}}(X, Y, z) \right| &\leq \widetilde{C}_{5.13}^{\widehat{m}_1(k+N_+)} \cdot |\eta_{\widehat{m}_1+1}(z)|^{-NG(k,z) - \widetilde{m}(k+N_+)} \\ &\quad \left[ \sum_{\ell, \mathbf{q}} \sum_{\mathcal{X}_k \in L_{\ell, k}^{(1)}} \sum_{\mathcal{Y}_k \in L_{\mathbf{q}, k}^{(2)}} |\mathfrak{F}_1| \cdot |\mathfrak{F}_2| \cdot \mathbf{1}_{X_1 = X \cup [\widehat{N}_+] \setminus [N]} \cdot \mathbf{1}_{Y_{\widehat{m}_2+1} = (Y+N_+) \cup [N_+]} \cdot \mathfrak{J}(X, Y, \bar{\mathbf{q}}) \right]. \end{aligned}$$

Finally we use Lemma 5.15 and

$$1 - \varepsilon'_0 \leq \min_{j=\widehat{m}_1+1}^{\widehat{m}_1+\mathbf{g}_0} |\eta_j(z)| = \max_{j=\widehat{m}_1+1}^{\widehat{m}_1+\mathbf{g}_0} |\eta_j(z)| \leq 1, \quad \text{for } z \in \mathcal{T}^{d,(1)}, \quad (5.67)$$

to derive the desired bound. This completes the proof of Lemma 5.13(i) for  $z \in \mathcal{T}^{d,(1)}$ .  $\square$

*Proof of Lemma 5.13(ii) for  $z \in \mathcal{T}^{d,(1)}$ .* As  $k + N_+ < \widehat{m}_2$ , for any  $\mathcal{Y}_k \in L_{\mathbf{q}, k}^{(2)}$  we have that

$$\begin{aligned} \bar{\mathbf{q}} &\stackrel{(5.47)}{\geq} \sum_{i=1}^{k+N_++1} q_i \\ &\stackrel{(5.29)}{\geq} y_{2,1} + \sum_{j=2}^{k+N_+} (y_{j+1,j} - y_{j,j-1}) + (\widehat{N}_+ - y_{k+N_++1, k+N_+}) - (k + N_+)(k + N_+ + 1) \\ &\geq \widehat{N}_+ - \widehat{m}_2^2, \end{aligned} \quad (5.68)$$

where in the second inequality we used that

$$y_{2,1} \leq q_1 + k + N_+, y_{3,2} - y_{2,1} \leq q_2 + k + N_+, \dots, \widehat{N}_+ - y_{k+N_++1, k+N_+} \leq q_{k+N_++1} + k + N_+,$$

and in the last we telescoped the sum and used again that  $k + N_+ < \widehat{m}_2$ . This implies that the set  $L_{\mathbf{q}, k}^{(2)}$  is empty unless (5.68) holds.



On the other hand, recalling (5.42)-(5.43), and using (5.31) and (5.36) we note that

$$\left| \widehat{\mathfrak{D}}(X, Y, z) \right| \leq |a_{-N_-}|^{-k} \cdot \prod_{i=1}^{\widehat{m}_1} |\eta_i(z)|^{N_+} \cdot \prod_{i=\widehat{m}_1+1}^{\widehat{m}_1+\mathfrak{g}_0} |\eta_i(z)|^{-N} \cdot \left[ \sum_{\ell, \mathfrak{q}} \sum_{\mathcal{X}_k \in L_{\ell, k}^{(1)}} \sum_{\mathcal{Y}_k \in L_{\mathfrak{q}, k}^{(2)}} |\mathfrak{F}_1| \cdot |\det(\mathcal{P}_z[\check{X}_{\widehat{m}_1+1}; \check{Y}_1])| \cdot |\mathfrak{F}_2| \cdot \mathbf{1}_{X_1=X \cup [\widehat{N}_+] \setminus [N]} \cdot \mathbf{1}_{Y_{\widehat{m}_2+1}=(Y+N_+) \cup [N_+]} \right].$$

From Lemma 5.16 we have that

$$\left| \frac{\det(\mathcal{P}_z[\check{X}_{\widehat{m}_1+1}; \check{Y}_1])}{\prod_{i=\widehat{m}_1+1}^{\widehat{m}_1+\mathfrak{g}_0} \eta_i(z)^N} \right| \leq \mathfrak{g}_0^{(\mathfrak{g}_0+2)(k+N_+)} |\eta_{\widehat{m}_1+1}(z)|^{\mathbf{y}_1 - \mathbf{x}_{\widehat{m}_1+1} - k\mathfrak{g}_0} \leq (1 + 2\varepsilon'_0)^{\widehat{m}_2 N}, \quad (5.69)$$

for all large  $N$ , where in the first step we have used (5.66) and in the second nequality we used (5.67), that  $\mathbf{y}_1 \geq 0$ ,  $\mathbf{x}_{\widehat{m}_1+1} \leq \widehat{m}_2 \widehat{N}_+$ , and  $k + N_+ \leq \widehat{m}_2$ .

Therefore, applying Lemma 5.15, and (5.68)-(5.69), we deduce that

$$\left| \widehat{\mathfrak{D}}(X, Y, z) \right| \stackrel{(5.69), (5.44)}{\leq} \left( \frac{1}{\varepsilon_0} \right)^{\widehat{m}_1(k+N_+)} \cdot (1 + 2\varepsilon'_0)^{\widehat{m}_2 N} \cdot \left[ \sum_{\mathfrak{q}} \sum_{\mathcal{Y}_k \in L_{\mathfrak{q}, k}^{(2)}} |\mathfrak{F}_2| \cdot \mathbf{1}_{Y_{\widehat{m}_2+1}=(Y+N_+) \cup [N_+]} \right] \stackrel{(5.46)}{\leq} \left( \frac{1}{\varepsilon_0} \right)^{\widehat{m}_1(k+N_+)} \cdot (1 + 2\varepsilon'_0)^{\widehat{m}_2 N} \cdot \left[ \sum_{\bar{\mathfrak{q}}} \left( \frac{\bar{\mathfrak{q}}(\widehat{m}_2 + 1) + \widehat{m}_2(k + N_+ + 1)}{\widehat{m}_2(k + N_+ + 1)} \right) \left( 1 - \frac{\varepsilon_0}{2} \right)^{\bar{\mathfrak{q}}} \cdot \mathbf{1}_{\{\bar{\mathfrak{q}} \geq \widehat{N}_+ - \widehat{m}_2^2\}} \right]. \quad (5.70)$$

Since for any fixed  $n_1 \in \mathbb{N}$  and  $n_2 \in \mathbb{N}$  large enough, depending only on  $\varepsilon_0$  and  $n_1$ , one has

$$\sum_{\bar{\mathfrak{q}} \geq n_2} \bar{\mathfrak{q}}^{n_1} \left( 1 - \frac{\varepsilon_0}{2} \right)^{\bar{\mathfrak{q}}} \leq \left( 1 - \frac{3\varepsilon_0}{8} \right)^{n_2},$$

and  $\varepsilon'_0/\varepsilon_0$  is sufficiently small, upon using (5.53), the claimed bound on  $\widehat{\mathfrak{D}}(X, Y, z)$  now follows from (5.70). This completes the proof of part (ii) for  $z \in \mathcal{T}^{d, (1)}$ .  $\square$

Next we prove Lemma 5.13 for  $z \in \mathcal{T}^{d, (2)}$ . It requires some minor modifications compared to the case  $z \in \mathcal{T}^{d, (1)}$ .

*Proof of Lemma 5.13 for  $z \in \mathcal{T}^{d, (2)}$ .* First let us prove part (i). We begin by noting that (5.57) continues to hold even for  $z \in \mathcal{T}^{d, (2)}$  and  $k \leq N$  such that  $k + N_+ \geq \widehat{m}_2$ . Recall (5.42)-(5.43). We use (5.31) and (5.36) to note that

$$\left| \widehat{\mathfrak{D}}(X, Y, z) \right| \leq |a_{-N_-}|^{-k} \prod_{i=1}^{\widehat{m}_1} |\eta_i(z)|^{N_+} \prod_{i=\widehat{m}_1+1}^{\widehat{m}_1+\mathfrak{g}_0} |\eta_i(z)|^{-N} \cdot \left[ \sum_{\ell, \mathfrak{q}} \sum_{\mathcal{X}_k \in L_{\ell, k}^{(1)}} \sum_{\mathcal{Y}_k \in L_{\mathfrak{q}, k}^{(2)}} |\mathfrak{F}_1| \cdot |\det(\mathcal{P}_z[\check{X}_{\widehat{m}_1+1}; \check{Y}_1])| \cdot |\mathfrak{F}_2| \cdot \mathbf{1}_{X_1=X \cup [\widehat{N}_+] \setminus [N]} \cdot \mathbf{1}_{Y_{\widehat{m}_2+1}=(Y+N_+) \cup [N_+]} \cdot \mathfrak{I}(X, Y, \mathfrak{q}) \right]. \quad (5.71)$$

As  $z \in \mathcal{T}^{d, (2)}$  implies that  $|\eta_{\widehat{m}_1+1}(z)| \geq 1$ , applying Lemma 5.16, and using that  $\mathbf{y}_1 \leq \mathbf{x}_{\widehat{m}_1+1}$ , we obtain that

$$\left| \frac{\det(\mathcal{P}_z[\check{X}_{\widehat{m}_1+1}; \check{Y}_1])}{\eta_{\widehat{m}_1+1}(z)^{N\mathfrak{g}_0}} \right| \leq \mathfrak{g}_0^{\mathfrak{g}_0+2(k+N_+)}. \quad (5.72)$$

Thus, from (5.71), we now derive that

$$\left| \widehat{\mathfrak{D}}(X, Y, z) \right| \leq |a_{-N_-}|^{-k} \cdot \mathfrak{g}_0^{\mathfrak{g}_0 + 2(k + N_+)} \cdot \prod_{i=1}^{\widehat{m}_1} |\eta_i(z)|^{N_+} \cdot \left[ \sum_{\ell, \mathfrak{q}} \sum_{\mathcal{X}_k \in L_{\ell, k}^{(1)}} \sum_{\mathcal{Y}_k \in L_{\mathfrak{q}, k}^{(2)}} |\mathfrak{F}_1| \cdot |\mathfrak{F}_2| \cdot \mathbf{1}_{X_1 = X \cup [\widehat{N}_+] \setminus [N]} \cdot \mathbf{1}_{Y_{\widehat{m}_2+1} = (Y + N_+) \cup [N_+]} \cdot \mathfrak{J}(X, Y, \mathfrak{q}) \right].$$

The claimed bound on  $\widehat{\mathfrak{D}}(X, Y, z)$  now follows upon using Lemma 5.15, in the same way as for  $z \in \mathcal{T}^{d, (1)}$ . This completes the proof of part (i) of the lemma for  $z \in \mathcal{T}^{d, (2)}$ .

We now turn to the proof of the second part. To this end, we note that all the steps of the proof of Lemma 5.13(ii) continues to hold under the current setup, except for the bound (5.69). In fact, under our current setup, the bound (5.69) can be improved to (5.72). Thus, repeating rest of the arguments in the proof of Lemma 5.13(ii) used for  $z \in \mathcal{T}^{d, (1)}$ , the proof of the second part of this lemma completes for  $z \in \mathcal{T}^{d, (2)}$ .  $\square$

**5.3. Toeplitz minors as skew Schur polynomials.** The goal of this section is to prove Lemma 5.16. This will follow from the following bound on minors of finitely banded Toeplitz matrices.

**Proposition 5.17.** *Let  $\mathfrak{g}_*, \widehat{N} \in \mathbb{N}$ . Fix a compact domain  $\mathbb{D} \subset \mathbb{C}$  and let  $z \in \mathbb{D}$ . Let  $\widehat{T}_{\widehat{N}}(z) = \widehat{T}_{\widehat{N}}(\widehat{p}_z)$  be the  $\widehat{N} \times \widehat{N}$  Toeplitz matrix with symbol*

$$\widehat{p}_z(\tau) := \sum_{j=0}^{\mathfrak{g}_*} \widehat{a}_{-j}(z) \tau^{-j}, \quad a_0(z), a_{-1}(z), \dots, a_{-\mathfrak{g}_*}(z) \in \mathbb{C}; \quad a_0(z) \neq 0, \quad a_{-\mathfrak{g}_*}(z) = 1 \quad \forall z \in \mathbb{D}. \quad (5.73)$$

Denote  $\{\widehat{\eta}_j(z)\}_{j=1}^{\mathfrak{g}_*}$  to be the roots of the polynomial  $\widehat{p}_z(\cdot) = 0$ . Assume that, for some constant  $c_1 \in (0, 1)$ ,

$$\inf_{z \in \mathbb{D}} \min_{k \neq j} \left| 1 - \frac{\widehat{\eta}_j(z)}{\widehat{\eta}_k(z)} \right| \geq c_1. \quad (5.74)$$

Then for any

$$\mathfrak{X} := \{\mathfrak{x}_1 < \mathfrak{x}_2 < \dots < \mathfrak{x}_n\} \subset [\widehat{N}] \quad \text{and} \quad \mathfrak{Y} := \{\mathfrak{y}_1 < \mathfrak{y}_2 < \dots < \mathfrak{y}_n\} \subset [\widehat{N}]$$

we have that

$$\left| \det(\widehat{T}_{\widehat{N}}(z)[\mathfrak{X}, \mathfrak{Y}]) \right| \leq c_1^{-\mathfrak{g}_* n} \mathfrak{g}_*^{2n} \cdot \prod_{j=1}^{\mathfrak{g}_*} |\widehat{\eta}_j(z)|^{\widehat{N} - n} \cdot \left( \min_{j=1}^{\mathfrak{g}_*} |\widehat{\eta}_j(z)| \right)^{\bar{\mathfrak{y}} - \bar{\mathfrak{x}}} \cdot \prod_{i=1}^n \mathbf{1}_{\{\mathfrak{x}_i \geq \mathfrak{y}_i\}} \cdot \prod_{i=1}^{n - \mathfrak{g}_*} \mathbf{1}_{\{\mathfrak{x}_i < \mathfrak{y}_{i + \mathfrak{g}_*}\}}, \quad (5.75)$$

where  $\check{\mathfrak{X}} := [\widehat{N}] \setminus \mathfrak{X}$ ,  $\check{\mathfrak{Y}} := [\widehat{N}] \setminus \mathfrak{Y}$ ,

$$\bar{\mathfrak{x}} := \sum_{i=1}^n \mathfrak{x}_i, \quad \text{and} \quad \bar{\mathfrak{y}} := \sum_{i=1}^n \mathfrak{y}_i.$$

**Remark 5.18.** *The bound on the order of magnitude of Toeplitz minors derived via Proposition 5.17 can be seen to be optimal when the roots of  $\widehat{p}_z(\cdot) = 0$  have the same moduli. However, it is suboptimal when the roots have different moduli. For example, consider the case  $\mathfrak{g}_* = 2$  and choose  $z \in \mathbb{C}$  such that  $|\widehat{\eta}_1(z)|/|\widehat{\eta}_2(z)| \geq (1 + \varepsilon')$  for some  $\varepsilon' > 0$ . To see the sub optimality of the bound now let  $\mathfrak{X} = \{\widehat{N} - 1, \widehat{N}\}$  and  $\mathfrak{Y} = \{1, 2\}$ .*

*Since we will apply Proposition 5.17 for a Toeplitz matrix for which the roots of its symbols have the same moduli, we have not tried to derive a version of Proposition 5.17 that captures the optimal order of magnitude even when roots have different moduli.*

As already mentioned in Section 5.2, to prove Proposition 5.17 (and hence Lemma 5.16) we will use the fact a Toeplitz minor can be expressed as certain *skew Schur polynomial*, see Remark 5.25 below for historical background. To state the relevant result and carry out the proof of Proposition 5.17 we need to borrow some notations from the theory of symmetric functions. The references [46, 68] are excellent resources for this purpose.

We start with the definition of *complete homogeneous symmetric polynomials*.

**Definition 5.19** (Complete homogeneous symmetric polynomial). *Fix  $w, r \in \mathbb{N}$ . The  $r$ -th complete homogeneous polynomial in  $w$  variables, to be denoted by  $h_r(\cdot)$ , is the sum of all monomials of total degree  $r$ . That is, for  $(t_1, t_2, \dots, t_w) \in \mathbb{R}^n$ ,*

$$h_r(t_1, t_2, \dots, t_w) := \sum_{1 \leq s_1 \leq s_2 \leq \dots \leq s_r \leq w} \prod_{u=1}^r t_{s_u}.$$

We further set  $h_0 := 1$  and  $h_r := 0$  for  $r < 0$ .

We next proceed to define the notion of skew Schur polynomials.

**Definition 5.20** (Skew partitions and skew Schur polynomials). *Fix  $w, m \in \mathbb{N}$ . A vector  $\lambda = (\lambda_1, \lambda_2, \dots, \lambda_m)$  with  $\lambda_i \in \mathbb{Z}$ ,  $i \in [m]$ , and  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_m \geq 0$  is called an integer partition (or in short a partition), of length  $\ell(\lambda) = m$  and weight  $|\lambda| = \sum_{i=1}^m \lambda_i$ .*

*For two partitions  $\mu$  and  $\lambda$  we write  $\mu \subseteq \lambda$  if  $\ell(\mu) \leq \ell(\lambda)$  and  $\mu_s \leq \lambda_s$  for all  $s \in [\ell(\mu)]$ . In that case, the pair  $(\lambda, \mu)$  is called a skew partition and is often denoted by  $\lambda/\mu$ .*

*Given a skew partition  $\lambda/\mu$  we define the skew Schur polynomial in  $w$  variables as follows:*

$$s_{\lambda/\mu}(\cdot) := \det \left[ (h_{\lambda_u - \mu_v - u + v}(\cdot))_{u,v=1}^{\ell(\lambda)} \right], \quad (5.76)$$

where  $\{h(\cdot)\}$  are the complete homogeneous symmetric polynomials in  $w$  variables as in Definition 5.19, and the partition  $\mu$  is extended to have length  $\ell(\lambda)$  by appending zeros. When  $\mu$  is an empty partition the polynomial  $s_{\lambda/\emptyset}$  is called a Schur polynomial and is written as  $s_\lambda$ . Thus, for any given partition  $\lambda$ , we have

$$s_\lambda = \det \left[ (h_{\lambda_u - u + v})_{u,v=1}^{\ell(\lambda)} \right]. \quad (5.77)$$

**Remark 5.21.** *Skew Schur polynomials admit a combinatorial description in terms of skew semi-standard Young tableaux. The equivalence of these two definitions is due to the first Jacobi-Trudi identity (see [68, Theorem 7.16.1]). Since we do not require the notion of Young tableaux elsewhere in this paper we have chosen (5.76) as the definition of the skew Schur polynomial  $s_{\lambda/\mu}$ . We refer the reader to [68, Chapter 7.10] for a detailed overview of these matters.*

The following identity, known already to Jacobi (see [68, Exercise 7.4] or [18, Theorem 3.2]) provides an efficient representation of complete homogeneous symmetric polynomials.

**Lemma 5.22.** *Let  $r, w \in \mathbb{N}$  and  $(t_1, t_2, \dots, t_w) \in \mathbb{R}^n$ . If  $\{t_s\}_{s=1}^w$  are pairwise distinct then*

$$h_r(t_1, t_2, \dots, t_w) = \sum_{j=1}^w \frac{t_j^{r+w-1}}{\prod_{k \in [w] \setminus \{j\}} (t_j - t_k)}.$$

The next lemma, which is a key to the proof of Proposition 5.17, states that a Toeplitz minor can be expressed in terms of skew Schur polynomials.

**Lemma 5.23** ([47, Theorem 2.1]). *Consider the setup of Proposition 5.17. Then*

$$\det(\widehat{T}_{\widehat{N}}(z)[\check{\mathfrak{X}}; \check{\mathfrak{Y}}]) = (-1)^{|\lambda_0| + |\mu_0|} \cdot \widehat{a}_0(z)^{\widehat{N}-n} \cdot s_{\lambda_0/\mu_0}(\widehat{\eta}_1(z)^{-1}, \widehat{\eta}_2(z)^{-1}, \dots, \widehat{\eta}_{\mathfrak{g}^*}(z)^{-1}),$$

where

$$\lambda_0 := (\widehat{N} - n + 1 - \eta_1, \widehat{N} - n + 2 - \eta_2, \dots, \widehat{N} - \eta_n) \text{ and } \mu_0 := (\widehat{N} - n + 1 - \mathfrak{r}_1, \widehat{N} - n + 2 - \mathfrak{r}_2, \dots, \widehat{N} - \mathfrak{r}_n).$$

**Remark 5.24.** *The reader may note that the expression for the Toeplitz minor stated in Lemma 5.23 is somewhat different than the one in [47, Theorem 2.1]. This is due to the fact that [47] expresses the minor in terms of the roots of the polynomial  $\widehat{p}_z(\tau) := \tau^{\mathfrak{g}^*} \cdot \widehat{p}_z(\tau)$ , where  $\widehat{p}_z(\cdot)$  is as in (5.73), while  $\{\widehat{\eta}_j(z)\}_{j=1}^{\mathfrak{g}^*}$  are the roots of  $\widehat{p}_z(\cdot) = 0$  which are indeed the reciprocals of those of  $\widehat{p}_z(\cdot) = 0$ .*

**Remark 5.25.** *It was first noted by Bump and Diaconis [17] that any Toeplitz minor can be expressed in terms of skew Schur polynomials. They expressed such minors as an integral of certain functions involving skew Schur polynomials over the unitary group (see [17, Theorem 3]). Later Alexandersson in [1] expressed minors of triangular Toeplitz matrices as certain skew Schur polynomials (see [1, Proposition 3]). Maximenko and Moctezuma-Salazar [47] gave a different proof*

of such a representation for non-triangular banded Toeplitz matrices. Here we use the formulation of [47] which is the most convenient for our setup.

Before proving Proposition 5.17 we make the following elementary observation.

**Lemma 5.26.** *Consider the setup of Proposition 5.17. If  $\mathfrak{X}, \mathfrak{Y} \subset [\widehat{N}]$  are such that either there exists an  $i \in [n]$  for which  $\mathfrak{x}_i < \mathfrak{y}_i$  or there exists an  $i \in [n - \mathfrak{g}_*]$  such that  $\mathfrak{x}_i \geq \mathfrak{y}_{i+\mathfrak{g}_*}$ , then*

$$\det(\widehat{T}_{\widehat{N}}(z)[\check{\mathfrak{X}}; \check{\mathfrak{Y}}]) = 0.$$

*Proof.* We note that  $\widehat{T}_{\widehat{N}}(z)$  is an upper triangular Toeplitz matrix with the symbol  $\widehat{p}_z(\cdot)$  (see (5.73)). As  $\{\widehat{\eta}_j\}_{j=1}^{\mathfrak{g}_*}$  are the roots of  $\widehat{p}(1/\cdot) = 0$ , it follows that

$$\widehat{T}_{\widehat{N}}(z) = \prod_{j=1}^{\mathfrak{g}_*} (J_{\widehat{N}} - \widehat{\eta}_j(z)I_{\widehat{N}}).$$

Therefore, by the Cauchy-Binet theorem we have that

$$\det(\widehat{T}_{\widehat{N}}(z)[\check{\mathfrak{X}}; \check{\mathfrak{Y}}]) = \sum_{\ell=2}^{\mathfrak{g}_*} \sum_{\substack{Z_\ell \subset [\widehat{N}] \\ |Z_\ell|=n}} \prod_{j=1}^{\mathfrak{g}_*} \det((J_{\widehat{N}} - \widehat{\eta}_j(z)I_{\widehat{N}})[\check{Z}_j; \check{Z}_{j+1}]), \quad (5.78)$$

where we set  $Z_1 = \mathfrak{X}$  and  $Z_{\mathfrak{g}_*+1} = \mathfrak{Y}$ . For  $\ell \in [\mathfrak{g}_* + 1]$ , we index the elements of  $Z_\ell$  as follows:

$$Z_\ell := \{z_{\ell,1} < z_{\ell,2} < \cdots < z_{\ell,n}\}.$$

By [8, Lemma A.3] we note that

$$\det((J_{\widehat{N}} - \widehat{\eta}_\ell(z)I_{\widehat{N}})[\check{Z}_\ell; \check{Z}_{\ell+1}]) \neq 0$$

if and only if

$$z_{\ell+1,1} \leq z_{\ell,1} < z_{\ell+1,2} \leq z_{\ell,2} < \cdots < z_{\ell+1,n} \leq z_{\ell,n} \leq n. \quad (5.79)$$

Therefore, in (5.78) we can restrict the sum over  $\{Z_2, Z_3, \dots, Z_{\mathfrak{g}_*}\}$  such that (5.79) holds for all  $\ell \in [\mathfrak{g}_*]$ . On the other hand, if (5.79) holds for all  $\ell \in [\mathfrak{g}_*]$ , as  $Z_1 = \mathfrak{X}$  and  $Z_{\mathfrak{g}_*+1} = \mathfrak{Y}$ , we derive that for all  $i \in [n]$

$$\mathfrak{y}_i = z_{\mathfrak{g}_*+1,i} \leq z_{\mathfrak{g}_*,i} \leq \cdots \leq z_{1,i} = \mathfrak{x}_i,$$

and for any  $i \in [n - \mathfrak{g}_*]$

$$\mathfrak{x}_i = z_{1,i} < z_{2,i+1} < z_{3,i+2} < \cdots < z_{\mathfrak{g}_*+1,i+\mathfrak{g}_*} = \mathfrak{y}_{i+\mathfrak{g}_*}.$$

Therefore, if either of the above two conditions are violated for some  $i \in [n]$ , then there must be an  $\ell \in [\mathfrak{g}_*]$  such that (5.79) is violated. In that case, the sum  $\{Z_2, Z_3, \dots, Z_{\mathfrak{g}_*}\}$  in (5.78) is an empty sum and hence  $\det(\widehat{T}_{\widehat{N}}(z)[\check{\mathfrak{X}}; \check{\mathfrak{Y}}]) = 0$ .  $\square$

We are now ready to prove Proposition 5.17.

*Proof of Proposition 5.17.* From (5.76) we see that

$$\mathfrak{s}_{\lambda_0/\mu_0}(\cdot) = \det \left[ (h_{\mathfrak{r}_v - \mathfrak{y}_u}(\cdot))_{u,v=1}^n \right], \quad (5.80)$$

where  $\lambda_0$  and  $\mu_0$  are as in Lemma 5.23. Set  $\widehat{\xi} := (\widehat{\xi}_1, \widehat{\xi}_2, \dots, \widehat{\xi}_{\mathfrak{g}_*})$ , where  $\widehat{\xi}_i := \widehat{\eta}_i^{-1}(z)$  for  $i \in [n]$ . Using (5.80) in the first equality and Lemma 5.22 in the second, we have that

$$\begin{aligned} \mathfrak{s}_{\lambda_0/\mu_0}(\widehat{\xi}) &= \sum_{\pi} \operatorname{sgn}(\pi) \prod_{v=1}^n h_{\mathfrak{r}_v - \mathfrak{y}_{\pi(v)}}(\widehat{\xi}) \\ &= \sum_{j_1, j_2, \dots, j_n=1}^{\mathfrak{g}_*} \prod_{u=1}^n \frac{\widehat{\xi}_{j_u}^{\mathfrak{g}_* - 1}}{\prod_{k_u \in [\mathfrak{g}_*] \setminus \{j_u\}} (\widehat{\xi}_{j_u} - \widehat{\xi}_{k_u})} \cdot \left[ \sum_{\pi} \operatorname{sgn}(\pi) \cdot \prod_{v=1}^n \widehat{\xi}_{j_v}^{\mathfrak{r}_v - \mathfrak{y}_{\pi(v)}} \cdot \prod_{v=1}^n \mathbf{1}_{\{\mathfrak{r}_v \geq \mathfrak{y}_{\pi(v)}\}} \right], \quad (5.81) \end{aligned}$$

where the sum in the first step and the innermost sum in the second step are over all permutations  $\pi$  on  $[n]$ , and in the last step we also used the fact that  $h_r(\cdot) = 0$  for  $r < 0$ .

We claim that there are at most  $\mathbf{g}_\star^n$  many permutations  $\pi$  such that the summand in the innermost sum in (5.81) is nonzero. That is, there are at most  $\mathbf{g}_\star^n$  many permutations  $\pi$  such that

$$\mathfrak{r}_v \geq \eta_{\pi(v)}, \quad \text{for all } v \in [n]. \quad (5.82)$$

Equipped with this claim, upon crudely bounding each term appearing in the sums in (5.81) by its maximum, it now follows that

$$\left| \mathfrak{s}_{\lambda_0/\mu_0}(\widehat{\xi}) \right| \leq \mathbf{g}_\star^{2n} \left( \max_{j=1}^{\mathbf{g}_\star} \left\{ \prod_{k \in [\mathbf{g}_\star] \setminus \{j\}} \left| \frac{1}{1 - (\widehat{\xi}_k/\widehat{\xi}_j)} \right| \right\} \right)^n \cdot \left( \max_{j=1}^{\mathbf{g}_\star} |\widehat{\xi}_j| \right)^{\bar{\mathfrak{r}} - \bar{\eta}} \leq \mathbf{g}_\star^{2n} \cdot c_1^{-\mathbf{g}_\star n} \cdot \left( \max_{j=1}^{\mathbf{g}_\star} |\widehat{\xi}_j| \right)^{\bar{\mathfrak{r}} - \bar{\eta}}, \quad (5.83)$$

where the last inequality is due to (5.74). Now, using Lemmas 5.23 and 5.26, and noting that

$$\widehat{a}_0(z) = (-1)^{\mathbf{g}_\star} \prod_{j=1}^{\mathbf{g}_\star} \widehat{\eta}_j(z), \quad (5.84)$$

we obtain (5.75) from (5.83).

Thus, to finish the proof it remains to prove the claim regarding the number of permutations  $\pi$  satisfying (5.82). To do that, by Lemma 5.26, without loss of generality, we may assume that  $\mathfrak{r}_v < \eta_{v+\mathbf{g}_\star}$  for all  $v \in [n - \mathbf{g}_\star]$ .

Now we bound the number of choices of  $\pi$  satisfying (5.82) as follows:

- As  $\mathfrak{r}_1 < \eta_{1+\mathbf{g}_\star}$ ,  $\eta_1 < \eta_2 < \dots < \eta_n$ , and  $\mathfrak{r}_1 \geq \eta_{\pi(1)}$  we find that  $\pi(1) \in [\mathbf{g}_\star]$ . That is, the number of choices for  $\pi(1)$  is  $\mathbf{g}_\star$ .
- Having chosen  $\pi(1)$  we now choose  $\pi(2)$ . Using the same argument as above we note that  $\pi(2) \in [\mathbf{g}_\star + 1]$ . As  $\pi(1) \in [\mathbf{g}_\star] \subset [\mathbf{g}_\star + 1]$  the number of choices for  $\pi(2)$  is again at most  $\mathbf{g}_\star$ .
- Continuing from the above, we find that for any given  $v \in [n - \mathbf{g}_\star]$  we must have  $\pi(v) \in [v + \mathbf{g}_\star - 1]$ . On the other hand  $\pi(1), \pi(2), \dots, \pi(v-1) \in [v + \mathbf{g}_\star - 2]$ . Thus the number of choices for  $\pi(v)$  is at most  $\mathbf{g}_\star$ .
- Finally  $\pi(n - \mathbf{g}_\star + 1), \pi(n - \mathbf{g}_\star + 2), \dots, \pi(n)$  can be chosen in at most  $\mathbf{g}_\star!$  ( $\leq \mathbf{g}_\star^{\mathbf{g}_\star}$ ) ways.

This proves the claim and hence we have the desired bound.  $\square$

We end this section with the proof of Lemma 5.16. This is a direct consequence of Proposition 5.17.

*Proof of Lemma 5.16.* We recall from (5.13) that  $\mathcal{P}_z$  is an upper triangular Toeplitz matrix of dimension  $\widehat{N}_+$  such that the roots of its symbols are  $\{-\eta_j(z)\}_{j=\widehat{m}_1+1}^{\widehat{m}_1+\mathbf{g}_0}$ . Note that  $\mathbf{g}(p) = \mathbf{g}_0$  (see (1.6) for a definition of  $\mathbf{g}(\cdot)$ ) implies that there exists  $\widetilde{\eta}(z) \in \mathbb{C}$  such that

$$\eta_{\widehat{m}_1+j}(z) = \widetilde{\eta}(z) \cdot e^{2\pi\sqrt{-1}j/\mathbf{g}_0}, \quad j \in [\mathbf{g}_0]. \quad (5.85)$$

Upon recalling the definition of the tube  $\mathcal{T}^d$  we see that

$$1 - \frac{\varepsilon_0}{3} \leq \inf_{z \in \mathcal{T}^d} |\eta_{\widehat{m}_1+\mathbf{g}_0}(z)| \leq \sup_{z \in \mathcal{T}^d} |\eta_{\widehat{m}_1+1}(z)| \leq 1 + \frac{\varepsilon_0}{3}.$$

This further implies that for any  $z \in \mathcal{T}^d$  we have that  $\widetilde{\eta}(z) \neq 0$ . Therefore, we indeed have that

$$\min_{i \neq j \in [\mathbf{g}_0]} \left| \frac{\eta_{\widehat{m}_1+i}(z)}{\eta_{\widehat{m}_1+j}(z)} - 1 \right| \geq \mathbf{g}_0^{-1}. \quad (5.86)$$

Recall also, see e.g. (5.66), that

$$|\eta_{\widehat{m}_1+1}(z)| = |\eta_{\widehat{m}_1+2}(z)| = \dots = |\eta_{\widehat{m}_1+\mathbf{g}_0}(z)|.$$

As  $|X_{\widehat{m}_1+1}| = |Y_1| = k + N_+$  the bound (5.56), as well as the fact that that the LHS of (5.56) equals zero when (5.55) is violated, is now immediate from Proposition 5.17. This completes the proof.  $\square$

## 6. PROOFS OF THE SEPARATION THEOREMS 5.4 AND 5.7

In this section, we complete the proofs of Theorems 5.4 and 5.7, based on Lemma 5.13. The proof is split into four parts: first, upon employing Lemma 5.13, we derive correct order of magnitudes (of moments) of  $\det_k(z)$ , per fixed  $z$  and  $k$  corresponding to non-dominant terms, in the regions of interest. Next, in order to control  $\sup_z \det_k(z)$ , we derive bounds on the moments on the supremum of the derivative of  $\det_k(z)$ . Combining these steps and a covering argument, we get the desired bounds for  $\sup_z \det_k(z)$  for non-dominant terms in the expansion (5.6). The third part of the proof derives a uniform lower bound on the dominant term, using Assumption 1.2. In the last step, we combine the above ingredients with geometric information on the forbidden regions to deduce Theorems 5.4 and 5.7.

**6.1. Step 1: moment bounds on the non-dominant terms (per fixed  $z$ ).** In this section we derive bounds on the moments of  $\det_k(z)$ . Recall (5.6), (5.35), (5.12), and (5.15), and note that  $\widehat{m}_1$  and  $\widehat{m}_2$  (which depend on  $z$ ) are constant on the tubes  $\mathcal{T}^{d,(1)}$  and  $\mathcal{T}^{d,(2)}$  with fixed  $d$ . We first consider the case of *small*  $k$ .

**Lemma 6.1.** *Let Assumption 1.1 hold. Fix  $\varepsilon'_0, \varepsilon_0 > 0$  such that  $\varepsilon'_0/\varepsilon_0$  is sufficiently small. Fix  $d \geq 0$  and  $s \in \{1, 2\}$ . Then, for all large  $N$ ,*

$$\max_{k < \widehat{m}_2 - N_+} \sup_{z \in \mathcal{T}_{\varepsilon'_0, \varepsilon_0}^{d,(s)}} \mathbf{E} \left[ \left| \widehat{\det}_k(z) \right|^2 \right] \leq \left( 1 - \frac{\varepsilon_0}{8} \right)^N. \quad (6.1)$$

*Proof.* Fix  $s \in \{1, 2\}$ . Since the entries of  $Q$  are independent with zero mean and bounded variance, we get that for any  $\mathbb{X}_*, \mathbb{Y}_*, \mathbb{X}', \mathbb{Y}' \subset [N]$

$$\mathbf{E} \left[ \det(Q[\mathbb{X}_*; \mathbb{Y}_*]) \cdot \overline{\det(Q[\mathbb{X}'; \mathbb{Y}'])} \right] \leq \mathfrak{C}_1 \begin{cases} k! & \text{if } \mathbb{X}_* = \mathbb{X}', \mathbb{Y}_* = \mathbb{Y}', \text{ and } |\mathbb{X}_*| = |\mathbb{Y}_*| = k, \\ 0 & \text{otherwise.} \end{cases} \quad (6.2)$$

Therefore, upon recalling (5.9), (5.35)-(5.36), and (5.40) we find that

$$\mathbf{E} \left[ \left| \widehat{\det}_k(z) \right|^2 \right] \leq \mathfrak{C}_1 N^{2\gamma(\widehat{m}_2 - N_+ - k)} \cdot k! \sum_{\substack{X, Y \subset [N] \\ |X| = |Y| = k}} |\widehat{\mathcal{D}}(X, Y, z)|^2. \quad (6.3)$$

Since there are at most  $\binom{N}{k}^2$  choices for the sets  $X, Y \subset [N]$  such that  $|X| = |Y| = k$ , the claimed upper bound now follows from Lemma 5.13(ii). This completes the proof.  $\square$

Next we derive bound on the second moment of  $\det_k(z)$  for  $k \geq \widehat{m}_2 - N_+$ .

**Lemma 6.2.** *In the setup of Lemma 6.1, we have the following bounds, for all large  $N$ .*

(i) *Fix  $\gamma'$  with  $1 < \gamma' < \gamma$ . Then for any  $k \in \mathbb{N}$  such that  $\widehat{m}_2 - N_+ \leq k \leq N$  we have*

$$\sup_{z \in \widehat{\mathcal{T}}_{\gamma', \varepsilon_0}^{d,(s)}} \mathbf{E} \left[ \left| \widehat{\det}_k(z) \right|^2 \right] = O \left( N^{-(\gamma - \gamma') \cdot (k + N_+ - \widehat{m}_2)} \right), \quad s = 1, 2.$$

(ii) *For  $k_0 = \widehat{m}_2 - N_+$  we have*

$$\sup_{z \in \mathcal{T}_{\varepsilon'_0, \varepsilon_0}^{d,(s)}} \mathbf{E} \left[ \left| \widehat{\det}_{k_0}(z) \right|^2 \right] = O(1), \quad s = 1, 2.$$

(iii) *For any  $k \in \mathbb{N}$  such that  $\widehat{m}_2 - N_+ \leq k \leq N$  we have*

$$\sup_{z \in \mathcal{T}_{\varepsilon'_0, \varepsilon_0}^{d,(2)}} \mathbf{E} \left[ \left| \widehat{\det}_k(z) \right|^2 \right] = O \left( N^{-(\gamma - 1) \cdot (k + N_+ - \widehat{m}_2)} \right).$$

**Remark 6.3.** *Lemma 6.2(ii) shows that the supremum of the second moment of  $\widehat{\det}_k(z)$ , for  $k = k_0$ , is well controlled in the tubes  $\mathcal{T}^d$  that are of small but fixed width, whereas Lemma 6.2(i) shows that the same can be said for any  $k > k_0$  only in the tubes  $\widehat{\mathcal{T}}^d$  that are of vanishing width.*

In part (iii) we see that one has such a control for  $z \in \mathcal{T}^{d,(2)}$  but, as will be seen during the course of the proof, those bounds fail for  $z \in \mathcal{T}^{d,(1)} \setminus \widehat{\mathcal{T}}^{d,(1)}$ .

*Proof of Lemma 6.2.* We start with the proof of part (i), beginning with  $s = 1$ . Denote

$$\mathfrak{d}_q := \binom{q(\widehat{m}_2 + 1) + \widehat{m}_2(k + N_+ + 1)}{\widehat{m}_2(k + N_+ + 1)} \left(1 - \frac{\varepsilon_0}{2}\right)^q. \quad (6.4)$$

As

$$|\eta_{\widehat{m}_1+1}(z)| \geq 1 - (\gamma' - 1) \frac{\log N}{N}, \quad \text{for } z \in \widehat{\mathcal{T}}_{\gamma', \varepsilon_0}^{d,(1)}, \quad (6.5)$$

by Lemma 5.13(i) we have that

$$\sum_{\substack{X, Y \subset [N] \\ |X|=|Y|=k}} |\widehat{\mathfrak{D}}(X, Y, z)|^2 \leq 2\widehat{C}_{5.13}^{2\widehat{m}(k+N_+)} \cdot N^{2(\gamma'-1)G(k,z)} \left[ \sum_{\substack{X, Y \subset [N] \\ |X|=|Y|=k}} \sum_{q, q' \geq 0} \mathfrak{d}_q \mathfrak{d}_{q'} \cdot \mathfrak{J}(X, Y, q) \right]. \quad (6.6)$$

We claim that, for any  $q \geq 0$ ,

$$|\{X, Y \subset [N] : |X| = |Y| = k \text{ and } \mathfrak{J}(X, Y, q) = 1\}| \leq (2\widehat{m}_2)! \binom{q + 2\widehat{m}_2}{2\widehat{m}_2} \cdot \binom{N}{k + N_+ - \widehat{m}_2}^2. \quad (6.7)$$

Indeed, upon recalling (5.39), we note that for a given  $q \geq 0$ ,

$$\mathfrak{J}(X, Y, q) = 1 \quad \implies \quad \max \left\{ \max_{j=k+N_+-\widehat{m}_2+1}^k (N - x_j), \max_{j=1}^{\widehat{m}_2-N_+} y_j \right\} \leq q.$$

Thus the number of choices of  $\{x_j\}_{j=k+N_+-\widehat{m}_2+1}^k$  and  $\{y_j\}_{j=1}^{\widehat{m}_2-N_+}$  is at most

$$q^{2(\widehat{m}_2-N_+)} \leq q^{2\widehat{m}_2} \leq (2\widehat{m}_2)! \binom{q + 2\widehat{m}_2}{2\widehat{m}_2}.$$

Now choose the remaining elements of  $X$  and  $Y$  to obtain (6.7).

Next applying (5.50) and (5.54) we have

$$\sum_{q' \geq 0} \mathfrak{d}_{q'} \leq \left( \frac{4(\widehat{m}_2 + 1)}{\varepsilon_0} \right)^{\widehat{m}_2(k+N_++1)+1}.$$

On the other hand, (5.50), (5.54), and (6.7) yield that

$$\left[ \sum_{\substack{X, Y \subset [N] \\ |X|=|Y|=k}} \sum_{q \geq 0} \mathfrak{d}_q \cdot \mathfrak{J}(X, Y, q) \right] \leq (2\widehat{m}_2)! \cdot \left( \frac{4(\widehat{m}_2 + 2)}{\varepsilon_0} \right)^{\widehat{m}_2(k+N_++3)+1} \cdot N^{2(k+N_+-\widehat{m}_2)} k^{2\widehat{m}_2} (k!)^{-2}.$$

Plugging in the last two bounds in (6.6) and using that  $G(k, z) \leq k + N_+ - \widehat{m}_2$ , we deduce that

$$\sum_{\substack{X, Y \subset [N] \\ |X|=|Y|=k}} |\widehat{\mathfrak{D}}(X, Y, z)|^2 \leq \overline{C}_1^{\widehat{m}(k+N_+)} \cdot N^{2\gamma'(k+N_+-\widehat{m}_2)} \cdot k^{2\widehat{m}_2} \cdot (k!)^{-2},$$

for some constant  $\overline{C}_1 < \infty$ . Upon using (6.3) part(i) of the lemma follows for  $s = 1$ .

Turning to  $s = 2$  we observe that, as  $G(k, \cdot) \equiv 0$  on  $\mathcal{T}_{\varepsilon'_0, \varepsilon_0}^{d,(2)}$  and  $\mathcal{T}_{\varepsilon'_0, \varepsilon_0}^{d,(2)} \supset \widehat{\mathcal{T}}_{\gamma', \varepsilon_0}^{d,(2)}$  for all large  $N$ , the upper bound (6.6) continues to hold in with  $N^{2(\gamma'-1)G(k,z)}$  replaced by one. Thus, repeating the rest of the arguments we derive part (i) for  $s = 2$ , as well as part (iii). The proof of part (ii) is exactly the same, where we again note that  $G(k_0, \cdot) \equiv 0$  on  $\mathcal{T}_{\varepsilon'_0, \varepsilon_0}^{d,(1)}$  and use that  $k_0 + N_+ - \widehat{m}_2 = 0$ . This completes the proof of the lemma.  $\square$



While applying Theorem 5.7 to derive the localization of the eigenvectors we will choose  $1 < \gamma' < \gamma$  such that  $(\gamma - \gamma')$  is (fixed but) arbitrarily close to zero. For such choice of parameters, if  $k + N_+ - \widehat{m}_2$  is *small* then Lemma 6.2 does not yield a sufficiently strong probability bound to carry out the covering argument. To overcome this caveat we control *high moments* of  $\det_k(z)$  for such choices of  $k$ , which when applied together with Markov's inequality produce desired probability bounds.

**Lemma 6.4.** *Consider the same setup as in Lemma 6.2. There exists a constant  $\widehat{C}_{6.4} < \infty$  (depending on  $p(\cdot)$  and  $\varepsilon_0$  only) so that, for all  $h, k, K_0 \in \mathbb{N}$  such that  $\widehat{m}_2 - N_+ \leq k \leq K_0 - N_+$ , we have*

$$\sup_{z \in \widehat{\mathcal{T}}_{\gamma', \varepsilon_0}^{d, (1)}} \mathbf{E} \left[ \left| \widehat{\det}_k(z) \right|^{2h} \right] \leq C_{6.4} N^{-2h(\gamma - \gamma') \cdot (k + N_+ - \widehat{m}_2)} \quad (6.8)$$

and

$$\sup_{z \in \mathcal{T}_{\varepsilon_0, \varepsilon_0}^{d, (2)}} \mathbf{E} \left[ \left| \widehat{\det}_k(z) \right|^{2h} \right] \leq C_{6.4} N^{-2h(\gamma - 1) \cdot (k + N_+ - \widehat{m}_2)}, \quad (6.9)$$

where

$$C_{6.4} = C_{6.4}(h, K_0, \widehat{m}, \widehat{m}_2) := \widehat{C}_{6.4}^{\widehat{m}K_0h} \cdot (K_0h)^{8K_0h} \cdot \mathfrak{C}_{2K_0h}.$$

*Proof.* We start with the proof of (6.8). Note that it suffices to show that for any  $z \in \widehat{\mathcal{T}}_{\gamma', \varepsilon_0}^{d, (1)}$ ,

$$\mathfrak{I}_k(z) := N^{2\gamma h(k + N_+ - \widehat{m}_2)} \mathbf{E} \left[ \left| \widehat{\det}_k(z) \right|^{2h} \right] \leq \widehat{C}_{6.4}^{\widehat{m}K_0h} \cdot (K_0h)^{8K_0h} \cdot \mathfrak{C}_{2K_0h} \cdot N^{2h\gamma'(k + N_+ - \widehat{m}_2)}. \quad (6.10)$$

Turning to prove (6.10), using (5.9), (5.32), and (5.35)-(5.36) we find that

$$\begin{aligned} \mathfrak{I}_k(z) = \mathbf{E} \left[ \sum_{\mathbf{X}} \sum_{\mathbf{Y}} \prod_{i=1}^{2h} (-1)^{\text{sgn}(\sigma_{X^{(i)}}) \text{sgn}(\sigma_{Y^{(i)}})} \cdot \prod_{i=1}^h \widehat{\mathfrak{D}}(X^{(i)}, Y^{(i)}, z) \cdot \prod_{i=1}^h \det(Q[X^{(i)}; Y^{(i)}]) \right. \\ \left. \cdot \prod_{i=h+1}^{2h} \overline{\widehat{\mathfrak{D}}(X^{(i)}, Y^{(i)}, z)} \cdot \prod_{i=h+1}^{2h} \overline{\det(Q[X^{(i)}; Y^{(i)}])} \right], \quad (6.11) \end{aligned}$$

where the sum is taken over

$$\mathbf{X} := \{X^{(1)}, X^{(2)}, \dots, X^{(2h)}\} \quad \text{and} \quad \mathbf{Y} := \{Y^{(1)}, Y^{(2)}, \dots, Y^{(2h)}\},$$

and for  $i \in [2h]$ ,

$$X^{(i)} := \{x_1^{(i)} < x_2^{(i)} < \dots < x_k^{(i)}\} \subset [N] \quad \text{and} \quad Y^{(i)} := \{y_1^{(i)} < y_2^{(i)} < \dots < y_k^{(i)}\} \subset [N].$$

Associate to  $\mathbf{X}$  the partition  $\mathfrak{P}_{\mathbf{X}}$  determined by the equivalence relation  $x_{j'}^{(i')} \sim x_{j''}^{(i'')}$  iff  $x_{j'}^{(i')} = x_{j''}^{(i'')}$ , for some  $i' \neq i'' \in [2h]$  and  $j', j'' \in [k]$ . Similarly, associate to  $\mathbf{Y}$  the partition  $\mathfrak{P}_{\mathbf{Y}}$ .

As the entries of  $Q$  are independent and are of zero mean it is straightforward to observe that for any  $(\mathbf{X}, \mathbf{Y})$  such that either  $\mathfrak{P}_{\mathbf{X}}$  or  $\mathfrak{P}_{\mathbf{Y}}$  has an equivalence class of size one, the expectation of the summand on the RHS of (6.11) equals zero. Hence, to compute a bound on the RHS of (6.11) we only need to consider partitions  $\mathfrak{P}_{\mathbf{X}}$  and  $\mathfrak{P}_{\mathbf{Y}}$  such that each of their equivalence classes has size at least two. For brevity we term them *pair partitions*. Fix one such pair of partitions  $(\mathfrak{P}, \widetilde{\mathfrak{P}})$  and a pair  $(\mathbf{X}, \mathbf{Y})$  such that  $\mathfrak{P}_{\mathbf{X}} = \mathfrak{P}$  and  $\mathfrak{P}_{\mathbf{Y}} = \widetilde{\mathfrak{P}}$ .

For any  $i \in [2h]$  the determinant  $\det(Q[X^{(i)}; Y^{(i)}])$  is a linear combination of  $k!$  terms each of which are products of  $k$  independent entries of the sub matrix  $Q[X^{(i)}; Y^{(i)}]$ . Therefore denoting

$$\mathfrak{Q} := \mathbf{E} \left[ \left| \prod_{i=1}^h \det(Q[X^{(i)}; Y^{(i)}]) \cdot \overline{\prod_{i=h+1}^{2h} \det(Q[X^{(i)}; Y^{(i)}])} \right| \right],$$

by the triangle inequality, we note that  $\mathfrak{Q}$  is a sum of at most  $(k!)^{2h}$  terms of the form

$$\mathbf{E} \left[ \left| \prod_{w=1}^b Q_{u_w, v_w}^{c_w} \cdot \overline{Q_{u_w, v_w}^{c_w}} \right| \right], \quad (6.12)$$

for some  $b \in \mathbb{N}$  and collections of positive integers  $\{c_w\}_{w=1}^b$  and  $\{c'_w\}_{w=1}^b$  such that

$$\sum_{w=1}^b c_w = \sum_{w=1}^b c'_w = kh,$$

where  $Q_{u_w, v_w}$  denotes the  $(u_w, v_w)$ -th entry of  $Q$  and the collection  $\{(u_w, v_w)\}_{w=1}^b$  is pairwise disjoint. Thus  $\{Q_{u_w, v_w}\}_{w=1}^b$  are jointly independent. Upon using Hölder's inequality and Assumption 1.1(ii) it follows that (6.12) is bounded above by

$$\prod_{w=1}^b \left( \mathbf{E} \left[ |Q_{u_w, v_w}|^{2kh} \right] \right)^{\frac{c_w + c'_w}{2kh}} \leq \mathfrak{C}_{2kh} \leq \mathfrak{C}_{2K_0h},$$

where in the last step we use that  $k \leq K_0$  and that  $\mathfrak{C}_w$  is increasing in  $w$ . This, in turn implies that  $\Omega \leq K_0^{2K_0h} \cdot \mathfrak{C}_{2K_0h}$ . Plugging this bound in (6.11), and applying Lemma 5.13(i), (6.5), and that  $G(k, z) \leq k + N_+ - \widehat{m}_2$ , we now deduce that

$$\mathfrak{T}_k(z) \leq 2\widehat{\mathcal{C}}_{5.13}^{2h\widehat{m}(k+N_+)} \cdot N^{2h(\gamma'-1)\cdot(k+N_+-\widehat{m}_2)} \cdot K_0^{2K_0h} \cdot \mathfrak{C}_{2K_0h} \cdot \left\{ \sum_{\mathfrak{P}, \widetilde{\mathfrak{P}}} \left[ \sum_{\mathbf{q}\#} \sum_{\mathbf{Y}: \mathfrak{P}_{\mathbf{Y}} = \widetilde{\mathfrak{P}}} \sum_{\mathbf{X}: \mathfrak{P}_{\mathbf{X}} = \mathfrak{P}} \prod_{i=1}^{2h} \mathfrak{d}_{q^{(i)}} \cdot \mathfrak{J}(X^{(i)}, Y^{(i)}, q^{(i)}) \right] \right\}, \quad (6.13)$$

where  $\mathbf{q}\# := (q^{(1)}, q^{(2)}, \dots, q^{(2h)})$ , and the outer sum is over all pair partitions  $\mathfrak{P}$  and  $\widetilde{\mathfrak{P}}$ . To evaluate the sum in (6.13) we will compute the sum in a specific order. To execute this step we need a few definitions.

We split the elements of  $\mathbf{X}$  and  $\mathbf{Y}$  into generating and non-generating elements as follows: For any  $i \in [2h]$  we term the collection of elements  $\{x_j^{(i)}\}_{j=k+N_+-\widehat{m}_2+1}^k$  and  $\{y_j^{(i)}\}_{j=1}^{\widehat{m}_2-N_+}$  *non-generating elements*. The rest of the elements of  $\mathbf{X}$  and  $\mathbf{Y}$  are termed *generating elements*. Let  $(\mathfrak{p}_1, \mathfrak{p}_2, \dots, \mathfrak{p}_b)$ , for some  $b > 0$ , be the equivalence classes of  $\mathfrak{P}_{\mathbf{X}}$ , the partition associated with  $\mathbf{X}$ . For  $w \in [b]$ , we say that  $\mathfrak{p}_w$  is non-generating if it contains *at least one non-generating* element of  $\mathbf{X}$ . Otherwise  $\mathfrak{p}_w$  will be said to be generating. The same convention is adopted for  $\mathbf{Y}$ .

We are now ready to derive the desired upper bound on (6.13). We proceed as follows:

- Fix a couple of pair partitions  $\mathfrak{P}$  and  $\widetilde{\mathfrak{P}}$ . Note that, as  $k \leq K_0$ , the number of choices of  $(\mathfrak{P}, \widetilde{\mathfrak{P}})$  is bounded above by  $(K_0h)^{4K_0h}$ .
- We next evaluate the innermost sum in (6.13), fixing  $\mathbf{q}\#$ . Since  $\mathfrak{P}_{\mathbf{X}} = \mathfrak{P}$  one only needs to sum over the possible (common) values of the equivalence classes of  $\mathfrak{P}$ .
- Let  $(\mathfrak{p}_1, \mathfrak{p}_2, \dots, \mathfrak{p}_b)$  be the equivalence classes of  $\mathfrak{P}$  and without loss of generality assume that  $(\mathfrak{p}_1, \mathfrak{p}_2, \dots, \mathfrak{p}_{b_0})$  are generating for some  $b_0 \leq b$ . Notice that the number of elements of  $\mathbf{X}$  is  $2kh$ . If  $\mathfrak{P}_{\mathbf{X}} = \mathfrak{P}$  for some  $\mathbf{X}$  and  $\mathfrak{P}$  is a pair partition then  $b_0 \leq (k + N_+ - \widehat{m}_2)h$ .
- As  $\mathbf{X} \subset [N]$  the total number of possible values of  $\{\mathfrak{p}_w\}_{w=1}^{b_0}$  is bounded above by  $N^{b_0} \leq N^{(k+N_+-\widehat{m}_2)h}$ .
- On the other hand, upon recalling (5.39) we find that

$$\prod_{i=1}^{2h} \mathfrak{J}(X^{(i)}, Y^{(i)}, q^{(i)}) = 1 \quad \implies \quad \mathfrak{p}_w \geq N - \bar{q}\#, \quad w \in [b] \setminus [b_0 + 1],$$

where  $\bar{q}\# := \sum_{i=1}^{2h} q^{(i)}$ . Thus, the number of choices for the of non-generating equivalences classes is bounded above by  $\bar{q}\#^b \leq \bar{q}\#^{K_0h}$ .

- We repeat the same idea as above to compute the second sum in (6.13).

Putting all the above pieces together we now obtain from (6.13) that

$$\mathfrak{T}_k(z) \leq 2\widehat{\mathcal{C}}_{5.13}^{2h\widehat{m}K_0} \cdot (2K_0h)^{8K_0h} \cdot \mathfrak{C}_{2K_0h} \cdot N^{2(k+N_+-\widehat{m}_2)\gamma'h} \left[ \sum_{\mathbf{q}\#} \bar{q}\#^{K_0h} \prod_{i=1}^{2h} \mathfrak{d}_{q^{(i)}} \right]. \quad (6.14)$$

Applying (5.50) and (5.54) one can note that the sum over  $\mathbf{q}_\#$  in (6.14) is bounded by  $\tilde{C}^{\tilde{m}K_0h}$ , for some constant  $\tilde{C} < \infty$ , depending only on  $\hat{m}_2$  and  $\varepsilon_0$ . Together, this yields (6.10) and completes the proof of (6.8).

To prove (6.9) we simply note that, again by Lemma 5.13, (6.13) continues to hold for any  $z \in \mathcal{T}_{\varepsilon'_0, \varepsilon_0}^{d, (2)}$  with  $N^{2h(\gamma'-1) \cdot (k+N_+ - \hat{m}_2)}$  replaced by one. Therefore, repeating the rest of the arguments we obtain (6.9) and complete the proof of the lemma.  $\square$

**6.2. Step 2: uniform upper bounds on non-dominant terms.** In this section we derive uniform upper bounds on the non-dominant terms in the expansion (5.6). The following is the main result of this section.

**Theorem 6.5.** *Let Assumption 1.1 hold. Fix parameters  $\gamma', \varepsilon_0, \varepsilon'_0, \tilde{\varepsilon}_0 > 0$  such that  $(\gamma - 1) \vee 1 < \gamma' < \gamma$  and  $\varepsilon'_0/\varepsilon_0$  is sufficiently small. Let  $d \geq 0$ . Then, for all large  $N$ ,*

$$\max_{s \in \{1, 2\}} \mathbf{P} \left[ \sup_{z \in \hat{\mathcal{T}}_{\gamma', \varepsilon_0}^{d, (s)}} \left| \sum_{k \neq \hat{m}_2 - N_+} \widehat{\det}_k(z) \right| \geq N^{-(\gamma - \gamma')/4} \right] \leq 1/N. \quad (6.15)$$

Furthermore, for all large  $N$ ,

$$\mathbf{P} \left[ \sup_{z \in \mathcal{T}_{\varepsilon'_0, \varepsilon_0}^{d, (2)}} \left| \sum_{k \neq \hat{m}_2 - N_+} \widehat{\det}_k(z) \right| \geq N^{-(\gamma - \gamma')/4} \right] \leq 1/N. \quad (6.16)$$

The difference between (6.15) and (6.16) is that the former yields bounds for tubes with diminishing width (in  $N$ ), while the latter provides bounds for certain tubes of fixed width. The bound (6.16) will be used in the proof of Theorem 5.4.

To prove Theorem 6.5 we will use the uniform bounds on the moments of  $\widehat{\det}_k(z)$  that were derived in Section 6.1. We will also need a bound on the second moment of the supremum of the derivatives of  $\widehat{\det}_k(z)$ .

**Lemma 6.6.** *Consider the same setup as in Theorem 6.5. Then there exist constants  $0 < c_{6.6}, C_{6.6} < \infty$ , depending only on  $\gamma'$  and  $\tilde{\varepsilon}_0$ , such that for any  $d \geq 0$  and  $k \in [N]$  we have that*

$$\max_{s \in \{1, 2\}} \mathbf{E} \left[ \sup_{z \in (\hat{\mathcal{T}}_{\gamma', \varepsilon_0}^{d, (s)} \setminus \mathcal{B}_2^{\tilde{\varepsilon}_0})^{c_{6.6} \log N/N}} \left| \frac{d}{dz} \widehat{\det}_k(z) \right|^2 \right] \leq C_{6.6} \left( \frac{N}{\log N} \right)^3 \quad (6.17)$$

and

$$\mathbf{E} \left[ \sup_{z \in (\mathcal{T}_{\varepsilon'_0, \varepsilon_0}^{d, (2)} \setminus \mathcal{B}_2^{\tilde{\varepsilon}_0})^{c_{6.6} \log N/N}} \left| \frac{d}{dz} \widehat{\det}_k(z) \right|^2 \right] \leq C_{6.6} \left( \frac{N}{\log N} \right)^4. \quad (6.18)$$

The proof of Theorem 6.5 will also require estimates on the *non-random* term in the expansion of  $\det(P_z^\delta)$ , as follows. Recall the notation  $\hat{d}$ , see (5.33), and that  $P_z^\delta = P_{N, \gamma}^Q - zI_N$ .

**Lemma 6.7.** *Consider the same setup as in Lemma 6.1. We have the following bounds.*

(a) *If  $\hat{d} > 0$  then, for all large  $N$ ,*

$$\sup_{z \in \mathcal{T}_{\varepsilon'_0, \varepsilon_0}^{d, (s)} \setminus \mathcal{B}_2^{\tilde{\varepsilon}_0}} |\widehat{\det}_0(z)| \leq \left(1 - \frac{\varepsilon_0}{4}\right)^N, \quad s \in \{1, 2\}. \quad (6.19)$$

(b) *Fix  $d \geq 0$  and  $\tilde{\varepsilon} > 0$ . Let  $\hat{d} \geq 0$ . Then there exists a constant  $c_\star \in (0, 1)$  so that, with  $X_\star := [N] \setminus [N - \hat{m}_2 + N_+]$  and  $Y_\star := [\hat{m}_2 - N_+]$ ,*

$$c_\star \leq \inf_{z \in \mathcal{T}_{\varepsilon'_0, \varepsilon_0}^{d, (s)} \setminus \mathcal{B}_2^{\tilde{\varepsilon}_0}} |\widehat{\mathfrak{D}}(X_\star, Y_\star, z)| \leq \sup_{z \in \mathcal{T}_{\varepsilon'_0, \varepsilon_0}^{d, (s)} \setminus \mathcal{B}_2^{\tilde{\varepsilon}_0}} |\widehat{\mathfrak{D}}(X_\star, Y_\star, z)| \leq c_\star^{-1}, \quad s \in \{1, 2\}. \quad (6.20)$$

**Remark 6.8.** Note that if  $\widehat{d} = 0$  then  $X_\star = Y_\star = \emptyset$ . Therefore, in that case  $\widehat{\det}_0(z) = \widehat{\mathfrak{D}}(X_\star, Y_\star, z)$  (see (5.9)-(5.10) and (5.34)-(5.36)). Thus, for  $\widehat{d} = 0$  Lemma 6.7(ii) provides a lower bound on the (dominant term)  $\widehat{\det}_0(z)$ .

Equipped with Lemmas 6.6 and 6.7 we now prove Theorem 6.5.

*Proof of Theorem 6.5 (assuming Lemmas 6.6 and 6.7).* Fix  $s \in \{1, 2\}$  and  $K_0 = K_0(\gamma, \gamma', \widehat{m}) = \lceil 30/(\gamma - \gamma') \rceil + \widehat{m}$ . We first show that for any  $k$  such that  $k + N_+ \geq K_0$ ,

$$\max_{s \in \{1, 2\}} \mathbf{P} \left[ \sup_{z \in \widehat{\mathcal{T}}_{\gamma', \varepsilon_0}^{d, (s)} \setminus \mathcal{B}_2^{\varepsilon_0}} \left| \widehat{\det}_k(z) \right| \geq N^{-2} \right] \leq 2/N^3. \quad (6.21)$$

To see (6.21), denote

$$\mathcal{F}(z) := \left\{ \left| \widehat{\det}_k(z) \right| \leq N^{-\frac{(k+N_+-\widehat{m}_2)(\gamma-\gamma')}{4}} \right\}$$

and

$$\mathcal{F}_0 := \left\{ \sup_{z \in \left( \widehat{\mathcal{T}}_{\gamma', \varepsilon_0}^{d, (s)} \setminus \mathcal{B}_2^{\varepsilon_0} \right)^{c_{6.6} \log N/N}} \left| \frac{d}{dz} \widehat{\det}_k(z) \right| \geq N^3 \right\}.$$

Applying Lemma 6.6 (in particular (6.17)) and Markov's inequality we obtain that

$$\mathbf{P}(\mathcal{F}_0) \leq N^{-3}. \quad (6.22)$$

From our choice of  $K_0$  and Lemma 6.2(i) it follows that

$$\sup_{z \in \widehat{\mathcal{T}}_{\gamma', \varepsilon_0}^{d, (s)}} \mathbf{P}(\mathcal{F}(z)^c) \leq N^{-15}. \quad (6.23)$$

Let  $\mathcal{N}$  be a net of  $\widehat{\mathcal{T}}_{\gamma', \varepsilon_0}^{d, (s)}$  of mesh size  $N^{-6}$ , which has cardinality at most  $O(N^{12})$ . Then, (6.22)-(6.23) yield that

$$\mathbf{P}(\cup_{z \in \mathcal{N}} \mathcal{F}(z)^c \cup \mathcal{F}_0) \leq 2/N^3. \quad (6.24)$$

On the other hand, we find that for all large  $N$  and  $z' \in \mathcal{N}$  the ball  $D(z', N^{-6})$  is contained in the  $(c_{6.6} \log N/N)$ -blow up of  $\widehat{\mathcal{T}}_{\gamma', \varepsilon_0}^{d, (s)} \setminus \mathcal{B}_2^{\varepsilon_0}$ . Thus, using the triangle inequality and the first order Taylor series expansion, we see that on the event  $\cap_{z' \in \mathcal{N}} \mathcal{F}(z') \cap \mathcal{F}_0^c$ ,

$$\begin{aligned} \left| \widehat{\det}_k(z) \right| &\leq \sup_{z' \in \mathcal{N}} \left| \widehat{\det}_k(z') \right| + \text{dist}(z, \mathcal{N}) \cdot \sup_{z' \in \left( \widehat{\mathcal{T}}_{\gamma', \varepsilon_0}^{d, (s)} \setminus \mathcal{B}_2^{\varepsilon_0} \right)^{c_{6.6} \log N/N}} \left| \frac{d}{dz} \widehat{\det}_k(z') \right| \\ &\leq N^{-\frac{1}{4}(k+N_+-\widehat{m}_2)(\gamma-\gamma')} + N^{-3} \leq N^{-2}, \end{aligned} \quad (6.25)$$

for all large  $N$ , where in the last step follows from the facts that  $k + N_+ \geq K_0$  and our choice of  $K_0$ . Therefore, from (6.24) we now have (6.21).

Next we claim that

$$\mathbf{P} \left[ \sup_{z \in \widehat{\mathcal{T}}_{\gamma', \varepsilon_0}^{d, (s)} \setminus \mathcal{B}_2^{\varepsilon_0}} \left| \widehat{\det}_k(z) \right| \geq N^{-2} \right] \leq 2/N^3, \quad (6.26)$$

for any  $k \in \mathbb{N}$  such that  $k + N_+ < \widehat{m}_2$ . Indeed, by Lemma 6.1, we see that (6.23) continues to hold in this case. Thus, repeating the same proof as for (6.21), we obtain (6.26).

Now we aim to show that

$$\mathbf{P} \left[ \sup_{z \in \widehat{\mathcal{T}}_{\gamma', \varepsilon_0}^{d, (s)} \setminus \mathcal{B}_2^{\varepsilon_0}} \left| \widehat{\det}_k(z) \right| \geq 2N^{-\frac{1}{2}(\gamma-\gamma')} \right] \leq 2/N^3, \quad (6.27)$$

for any  $k$  such that  $\widehat{m}_2 + 1 \leq k + N_+ \leq K_0$ . To this end, we set  $h_0 = h_0(\gamma, \gamma', \widehat{m}, \mathbf{g}_0) = \lceil 15/(\gamma - \gamma') \rceil$ . We apply Lemma 6.4 with this  $h_0$  and Markov's inequality to find that (6.23) continues to hold for any  $\widehat{m}_2 - N_+ + 1 \leq k \leq K_0 - N_+$ . Thus, repeating the same argument as above, yet again,

we further observe that (6.25) also holds, except for the last step. Since  $\gamma' > \gamma - 1$  the last step there can be replaced by  $2N^{-\frac{1}{2}(\gamma-\gamma')}$ . Therefore we have (6.27).

Combining the probability estimates (6.21), (6.26), and (6.27), and using a union bound we deduce that

$$\mathbf{P} \left[ \sup_{z \in \widehat{\mathcal{T}}_{\gamma', \varepsilon_0}^{d, (s)} \setminus \mathcal{B}_2^{\tilde{\varepsilon}_0}} \left| \sum_{k \notin \{\widehat{m}_2 - N_+, 0\}} \widehat{\det}_k(z) \right| \geq \frac{1}{2} N^{-(\gamma-\gamma')/4} \right] \leq 3/N^2,$$

for all large  $N$ . If  $\widehat{d} = 0$  then, by (6.44),  $\widehat{m}_2 - N_+ = 0$ , and hence we have the desired probability bound. If  $\widehat{d} > 0$  then upon using Lemma 6.7(a) the proof of (6.15) completes. The proof of (6.16), being similar, is omitted.  $\square$

We now proceed to the proof of Lemma 6.6, which uses the bounds derived in Lemmas 6.1 and 6.2, and Cauchy's integral formula for smooth functions. Cauchy's integral formula allows us to control the second moment of the supremum of a random analytic function (and its derivative) on a nice domain by controlling the supremum of the second moment of the same analytic function on a slightly enlarged domain. To carry out this scheme, we first show that the blow up of any tube  $\mathcal{T}^{d, (s)}$ ,  $s = 1, 2$ , away from the bad set  $\mathcal{B}_2^{\tilde{\varepsilon}_0}$ , is again contained in a union of the tubes with slightly modified parameters.

**Lemma 6.9.** *Fix  $0 < \varepsilon'_0 < \varepsilon_0$ , and  $\tilde{\varepsilon}_0 > 0$ . We have the following geometric properties of the tubes.*

(i) *There exists a constant  $C_{6.9} < \infty$  such that*

$$\mathcal{T}_{\varepsilon'_0, \varepsilon_0}^d \subset (p(S^1))^{C_{6.9} \varepsilon'_0}. \quad (6.28)$$

*In particular,  $\mathcal{T}_{\varepsilon'_0, \varepsilon_0}^d$  is a bounded set.*

(ii) *Fix  $\varepsilon''_0 > 0$  such that  $\varepsilon''_0 \leq \varepsilon'_0$ . There exists some constant  $\varepsilon_{6.9}$ , depending only on  $\tilde{\varepsilon}_0$ , such that, for any  $\beta \in (0, 1/2)$ , and  $\varepsilon \leq \varepsilon''_0 \varepsilon_{6.9}$ , we have*

$$\left( \mathcal{T}_{\varepsilon'_0, \varepsilon_0}^{d, (1)} \setminus \mathcal{B}_2^{\tilde{\varepsilon}_0} \right)^{\beta \varepsilon} \subset \left( \mathcal{T}_{(1+\beta)\varepsilon'_0, (1-\beta)\varepsilon_0}^{d, (1)} \cup \mathcal{T}_{\beta\varepsilon''_0, (1-\beta)\varepsilon_0}^{d-\mathbf{g}_0, (2)} \right) \setminus \mathcal{B}_2^{\tilde{\varepsilon}_0/2}, \quad (6.29)$$

*for  $d \geq 1$ , while for  $d \geq 0$  we have*

$$\left( \mathcal{T}_{\varepsilon'_0, \varepsilon_0}^{d, (2)} \setminus \mathcal{B}_2^{\tilde{\varepsilon}_0} \right)^{\beta \varepsilon} \subset \left( \mathcal{T}_{(1+\beta)\varepsilon'_0, (1-\beta)\varepsilon_0}^{d, (2)} \cup \mathcal{T}_{\beta\varepsilon''_0, (1-\beta)\varepsilon_0}^{d+\mathbf{g}_0, (1)} \right) \setminus \mathcal{B}_2^{\tilde{\varepsilon}_0/2}. \quad (6.30)$$

*Proof.* We begin with the proof of part (i). We observe that  $z \in \mathcal{T}_{\varepsilon'_0, \varepsilon_0}^d$  implies that there exists a root  $\eta(z)$  of  $p_z(\cdot) = 0$  such that  $|\eta(z)| - 1 \leq \varepsilon'_0$ . Set  $z_0 := p(\eta(z)/|\eta(z)|) \in p(S^1)$ . By the triangle inequality and uniform boundedness of  $p'(\cdot)$  in a compact neighborhood of  $S^1$ , one has that  $|z - z_0| \leq C_{6.9} \varepsilon'_0$  for some  $C_{6.9} < \infty$  depending on  $p$ , yielding (6.28).

Turning to the proof of part (ii) we pick any  $z_*$  belonging to the set on the LHS of (6.29). By definition, there exists a  $z_0 \in \mathcal{T}_{\varepsilon'_0, \varepsilon_0}^{d, (1)} \setminus \mathcal{B}_2^{\tilde{\varepsilon}_0}$  such that  $|z_* - z_0| \leq \beta \varepsilon$ . Thus, for  $\varepsilon \leq \tilde{\varepsilon}_0/2$ , we have that  $z_* \notin \mathcal{B}_2^{\tilde{\varepsilon}_0/2}$ . On the other hand, noting that for  $j \in [\tilde{m}]$ , the maps  $z \mapsto \eta_j(z)$  are analytic outside  $\mathcal{B}_2^{\tilde{\varepsilon}_0/2}$  and using that  $\mathcal{T}^d$  is a bounded set, we deduce that

$$\max_{j \in [\tilde{m}]} |\eta_j(z_0) - \eta_j(z_*)| \leq O_{\tilde{\varepsilon}_0}(1) \cdot |z_* - z_0| \leq \beta \varepsilon''_0/4, \quad (6.31)$$

where in the last inequality we chose  $\varepsilon \leq \varepsilon_{6.9} \varepsilon''_0$  with  $\varepsilon_{6.9}$  a sufficiently small constant, depending only on  $\tilde{\varepsilon}_0$ .

Since  $z_0 \in \mathcal{T}_{\varepsilon'_0, \varepsilon_0}^{d, (1)}$ , by (6.31) we deduce that

$$\max \{ |\eta_{m_+ + \mathbf{g}_0 + 1}(z_*)|, |\eta_{m_-}(z_*)|^{-1} \} \leq 1 - \varepsilon_0(1 - \beta/2), \quad (6.32)$$

and

$$1 - \varepsilon'_0(1 + \beta/2) \leq |\eta_{m_+ + 1}(z_*)| = |\eta_{m_+ + \mathbf{g}_0}(z_*)| \leq 1 + \varepsilon''_0 \beta/2, \quad (6.33)$$

where we also use that the roots can be partitioned into blocks, each of size  $\mathbf{g}_0$ , such that the roots in each of those block have the same moduli.

Now there are three possibilities: If  $|\eta_{m-+1}(z_*)| < 1$ , in which case  $z_* \in \mathcal{S}_d$ , and therefore by (6.32)-(6.33) we have that  $z_* \in \mathcal{T}_{(1+\beta)\varepsilon'_0, (1-\beta)\varepsilon_0}^{d, (1)}$ . If  $|\eta_{m-+g_0}(z_*)| > 1$  then  $z_* \in \mathcal{S}^{d-g_0}$ , and consequently  $z_* \in \mathcal{T}_{\beta\varepsilon''_0, (1-\beta)\varepsilon_0}^{d-g_0, (2)}$ . Finally, if  $|\eta_{m-+1}(z_*)| = 1$ , by the maximum modulus principle, there exists a sequence  $\{z_n\}_{n \in \mathbb{N}}$  such that  $|z - z_n| \leq 1/n$  and  $|\eta_{m-+1}(z_n)| < 1$  for all  $n \in \mathbb{N}$ . It is clear that  $z_n \in \mathcal{S}_d$ , and by the continuity of the roots the bound in (6.32) and the lower bound in (6.33) holds for  $z_n$ , for all large  $n$ , with  $\beta/2$  replaced by  $\beta$ . This, in particular, shows that  $z_n \in \mathcal{T}_{(1+\beta)\varepsilon'_0, (1-\beta)\varepsilon_0}^{d, (1)}$  for all large  $n$ . Therefore,  $\mathcal{T}_{(1+\beta)\varepsilon'_0, (1-\beta)\varepsilon_0}^{d, (1)}$  being a closed set the limit  $z_*$  must also be in  $\mathcal{T}_{(1+\beta)\varepsilon'_0, (1-\beta)\varepsilon_0}^{d, (1)}$ . This completes the proof of (6.29). The proof of (6.30) being similar, details are omitted.  $\square$

The following lemma is proved by a standard volumetric estimate, that we omit.

**Lemma 6.10.** *For any  $\varepsilon, \beta \in (0, 1]$  there exists a net of  $p(S^1)^\varepsilon$  of mesh size  $\beta\varepsilon$  and of cardinality at most  $O(\beta^{-2}\varepsilon^{-1})$ .*

We are now ready to provide the proof of Lemma 6.6.

*Proof of Lemma 6.6.* Fix  $d \geq 0$  and  $s \in \{1, 2\}$ . This fixes  $\widehat{m}_1$  and  $\widehat{d}$  (see (5.12) and (5.33)). Fix  $\varepsilon_0, \varepsilon'_0, \varepsilon''_0 > 0$  such that  $\varepsilon'_0 < \varepsilon_0$  is sufficiently small and  $\varepsilon''_0 \leq \varepsilon'_0$ . The precise choice of  $\varepsilon''_0$  will be specified later. Set  $\varepsilon = \varepsilon_{6.9}\varepsilon''_0$  and let  $\beta \in (0, 1/16)$ . Pick any  $z \in \mathbb{T}^\beta := (\mathcal{T}_{\varepsilon'_0, \varepsilon_0}^{d, (s)} \setminus \mathcal{B}_2^{\varepsilon_0})^{\beta\varepsilon}$ . From Lemmas 6.9(i)-(ii) and 6.10 we deduce that there exist a finite collection  $\{z_1, z_2, \dots, z_b\} \subset \mathbb{T}^\beta$  such that

$$\cup_{i=1}^b D(z_i, \beta\varepsilon) \supset \mathbb{T}^\beta \quad (6.34)$$

with

$$b = O(\varepsilon'_0 \cdot \varepsilon''_0^{-2}). \quad (6.35)$$

Hence, it suffices to derive bounds on the derivative of  $\widehat{\det}_k(\cdot)$  on each  $\mathbb{D}_i^\beta := D(z_i, \beta\varepsilon)$  for  $i \in [b]$ .

Turning to do that we let  $\mathbb{D} := D(0, R) \setminus \mathcal{B}_2^{\varepsilon_0}$ , where  $R := 2 \max_{\eta \in S^1} |p(1/\eta)|$ , and consider a smooth cutoff function  $\chi : \mathbb{D} \mapsto \mathbb{C}$  such that

$$\chi \equiv \begin{cases} 1 & \text{on } \mathbb{D}_i^{2\beta}, \\ 0 & \text{on } \mathbb{D} \setminus \mathbb{D}_i^{3\beta}. \end{cases}$$

This implies that the derivative of  $\chi(\cdot)$  is non-zero only on  $\mathbb{D}_i^{3\beta} \setminus \mathbb{D}_i^{2\beta}$ . Since  $\text{dist}(\mathbb{D}_i^{2\beta}, \partial\mathbb{D}_i^{3\beta}) = \beta\varepsilon$  one can choose  $\chi$  such that the absolute value of its anti-holomorphic derivative  $\partial_{\bar{w}}\chi(w)$  is at most  $O(\beta^{-1}\varepsilon^{-1})$  in  $\mathbb{D}$ . Further let  $\Xi : \mathbb{D} \mapsto \mathbb{C}$  be some (possibly random) holomorphic function. Then, applying Cauchy's integral formula for the smooth function  $z \mapsto \Xi(z) \cdot \chi(z)$  on the domain  $\mathbb{D}$  (e.g. see [38, Theorem 1.2.1]) we obtain that for any  $z \in \mathbb{D}_i^\beta$ ,

$$\Xi(z) = -\frac{1}{\pi} \int_{\mathbb{D}_i^{3\beta} \setminus \mathbb{D}_i^{2\beta}} \frac{\Xi(w) \cdot \partial_{\bar{w}}\chi(w)}{w - z} dL(w),$$

where  $L(\cdot)$  is the two-dimensional Lebesgue measure. By the bounded convergence theorem we also have that

$$\frac{d}{dz}\Xi(z) = -\frac{1}{\pi} \int_{\mathbb{D}_i^{3\beta} \setminus \mathbb{D}_i^{2\beta}} \frac{\Xi(w) \cdot \partial_{\bar{w}}\chi(w)}{(w - z)^2} dL(w).$$

Now, by the Cauchy-Schwarz inequality and Fubini's Theorem we further deduce that

$$\begin{aligned} \mathbf{E} \left[ \sup_{z \in \mathbb{D}_i^\beta} \left| \frac{d}{dz}\Xi(z) \right|^2 \right] &\leq \frac{L(\mathbb{D}_i^{3\beta})}{\pi^2 \cdot \text{dist}(\mathbb{D}_i^\beta, \partial\mathbb{D}_i^{2\beta})^4} \cdot \int_{\mathbb{D}_i^{3\beta} \setminus \mathbb{D}_i^{2\beta}} \mathbf{E} [|\Xi(w)|^2] \cdot |\partial_{\bar{w}}\chi(w)|^2 dL(w) \\ &= O(\beta^{-2}\varepsilon^{-2}) \cdot \sup_{w \in \mathbb{D}_i^{3\beta}} \mathbf{E} [|\Xi(w)|^2]. \end{aligned} \quad (6.36)$$

To complete the proof of the lemma we now proceed to apply (6.36) with appropriate choices of  $\Xi(\cdot)$  and  $\varepsilon''_0$ .



We note that the roots of  $p_z(\cdot) = 0$  are analytic in  $z$  for  $z \in \mathbb{D}$ . Therefore, there exists a reordering of the indices of the roots  $\{\eta_j(z)\}_{j \in [\widehat{m}]}$ , to be denoted by  $\{\widehat{\eta}_j(z)\}_{j \in [\widehat{m}]}$ , such that the maps  $z \mapsto \widehat{\eta}_j(z)$  are holomorphic on  $\mathbb{D}$ . We set

$$\Xi(z) = \Xi_k(z) := \frac{\det_k(z)}{a_{-N_-}^N N^{-\gamma \widehat{d}} \prod_{j=1}^{\widehat{m}_1 + \mathbf{g}_0} \widehat{\eta}_j(z)^N}. \quad (6.37)$$

We claim that the holomorphic functions  $\{\widehat{\eta}_j(z)\}_{j \in [\widehat{m}]}$  can be chosen in such a way so that

$$\prod_{j=1}^{\widehat{m}_1 + \mathbf{g}_0} \widehat{\eta}_j(z) = \prod_{j=1}^{\widehat{m}_1 + \mathbf{g}_0} \eta_j(z) \quad \text{for } z \in \mathbb{D}_i^{3\beta}. \quad (6.38)$$

Indeed, pick any  $\widehat{z} \in \mathbb{D}_i^{3\beta}$ . By the Implicit function theorem it is immediate that  $\{\widehat{\eta}_j(z)\}_{j \in [\widehat{m}]}$  can be chosen such that (6.38) holds for  $z = \widehat{z}$ . Also, note that by Lemma 6.9(ii) we have that

$$\mathbb{D}_i^{3\beta} \subset \mathbb{T}^{4\beta} \subset \left( \mathcal{T}_{(1+4\beta)\varepsilon'_0, (1-4\beta)\varepsilon_0}^{d, (s)} \cup \mathcal{T}_{\varepsilon''_0, (1-4\beta)\varepsilon_0}^{d+(-1)^s \mathbf{g}_0, (3-s)} \right) \setminus \mathcal{B}_2^{\varepsilon_0/2}. \quad (6.39)$$

The inclusion (6.39) and  $\beta < 1/16$  imply that there are no roots of  $p_z(\cdot) = 0$  in  $\mathbb{D}_i^{3\beta}$  with moduli between  $1 + 2\varepsilon''_0$  and  $1 + \varepsilon_0/2$  (choose  $\varepsilon_0$  and  $\varepsilon'_0$  such that  $\varepsilon'_0/\varepsilon_0 < 1/8$ ). Hence, (6.38) must continue to hold for all  $z$  in the connected domain  $\mathbb{D}_i^{3\beta}$ , for otherwise the image of the continuous map  $|\widehat{\eta}_j(z)|$  from  $\mathbb{D}_i^{3\beta}$  to  $\mathbb{R}_+$  would be disconnected for some  $j$ . It further follows from (6.39) that  $\widehat{d}(\cdot) \equiv \widehat{d}$  and  $\widehat{m}_1(\cdot) \equiv \widehat{m}_1$  on  $\mathbb{D}_i^{3\beta}$ . This means that  $\widehat{\det}_k(\cdot) \equiv \Xi_k(\cdot)$  on  $\mathbb{D}_i^{3\beta}$  (recall (5.32)-(5.35)). Since this holds for every  $i \in [b]$  we use (6.34) and (6.39) to deduce from (6.36) that

$$\mathbf{E} \left[ \sup_{z \in \mathbb{T}^\beta} \left| \frac{d}{dz} \widehat{\det}_k(z) \right|^2 \right] = O(\varepsilon'_0 / (\varepsilon''_0)^4) \cdot \sup_{w \in \mathbb{T}^{4\beta}} \mathbf{E} \left[ |\widehat{\det}_k(w)|^2 \right]. \quad (6.40)$$

To complete the proofs of (6.17)-(6.18) it remains to find an upper bound on the RHS of (6.40).

We first consider the proof of (6.18). Set  $\varepsilon''_0 = (\gamma' - 1) \log N/N$ . With this choice of  $\varepsilon''_0$ , as  $\beta \leq 1/16$ , we obtain from (6.39) that  $\mathbb{T}^{4\beta} \subset \mathcal{T}_{2\varepsilon'_0, \varepsilon_0/2}^{d, (2)} \cup \widehat{\mathcal{T}}_{\gamma', \varepsilon_0/2}^{d + \mathbf{g}_0, (1)}$ . Apply Lemmas 6.1, 6.2(i) and 6.2(iii) with  $\varepsilon'_0$  and  $\varepsilon_0$  replaced by  $2\varepsilon'_0$  and  $\varepsilon_0/2$  to derive that

$$\sup_{w \in \mathbb{T}^{4\beta}} \mathbf{E} \left[ |\widehat{\det}_k(z)|^2 \right] = O(1). \quad (6.41)$$

Plugging this bound in (6.40) we obtain (6.18).

The proof of (6.17) is similar. Set  $\varepsilon'_0 = \varepsilon''_0 = (\gamma' - 1) \log N/N$  and choose  $\beta$  such that  $(1 + 3\beta)\varepsilon'_0 = (\gamma'' - 1) \log N/N$  for some  $\gamma'' < \gamma$ . Apply Lemmas 6.1, 6.2(i) and 6.2(iii) with  $\gamma'', \varepsilon''_0$ , and  $\varepsilon_0/2$  instead of  $\gamma', \varepsilon'_0$ , and  $\varepsilon_0$ , respectively to deduce that (6.41) continues to hold for  $\mathbb{T}^\beta = (\mathcal{T}_{\varepsilon'_0, \varepsilon_0}^{d, (1)} \setminus \mathcal{B}_2^{\varepsilon_0})^{\beta\varepsilon}$ . The desired bound now follows from (6.40).  $\square$

We end this section with the proof of Lemma 6.7.

*Proof of Lemma 6.7.* We first prove part (a). The proof is a direct consequence of Widom's formula for the determinant of a finitely banded Toeplitz matrix. Indeed, by [15, Theorem 2.8], for  $s \in \{1, 2\}$ , we have, recalling that  $\widetilde{N} = N_+ + N_-$ , that

$$\widehat{\det}_0(z) = N^{\gamma(\widehat{m}_2 - N_+)} \cdot \prod_{j=1}^{\widehat{m}_1 + \mathbf{g}_0} \eta_j(z)^{-N} \cdot (-1)^{NN_-} \cdot \left[ \sum_{\mathcal{I} \in \binom{[\widetilde{N}]}{N_-}} \mathcal{C}_{\mathcal{I}}(z) \cdot \prod_{i \in \mathcal{I}} \eta_i(z)^N \right],$$

where for any  $\mathcal{I} \subset [\widetilde{N}]$

$$\mathcal{C}_{\mathcal{I}}(z) := \prod_{\substack{j_1 \in \mathcal{I} \\ j_2 \in [\widetilde{N}] \setminus \mathcal{I}}} \frac{\eta_{j_1}(z)}{\eta_{j_1}(z) - \eta_{j_2}(z)}.$$



Observe that that  $\inf_{z \in \mathcal{T}^d} |\eta_{\widehat{m}_1 + \mathbf{g}_0}(z)| \geq 1/2$ . Thus, now by the continuity of the maps  $z \mapsto \eta_j(z)$ , for  $j \in [\widetilde{m}]$ , as  $\mathcal{T}^d$  is a bounded set and we work off  $\mathcal{B}_2^{\widetilde{\varepsilon}_0}$ , it follows that

$$0 < \inf_{z \in \mathcal{T}^d \setminus \mathcal{B}_2^{\widetilde{\varepsilon}_0}} |\mathcal{C}_{[\widehat{m}_1 + \mathbf{g}_0]}(z)| \leq \sup_{z \in \mathcal{T}^d \setminus \mathcal{B}_2^{\widetilde{\varepsilon}_0}} |\mathcal{C}_{[\widehat{m}_1 + \mathbf{g}_0]}(z)| \leq \max_{\mathcal{I} \subset [N_+ + N_-]} \sup_{z \in \mathcal{T}^d \setminus \mathcal{B}_2^{\widetilde{\varepsilon}_0}} |\mathcal{C}_{\mathcal{I}}(z)| < \infty. \quad (6.42)$$

On the other hand, recalling the definition of the tube  $\mathcal{T}^d$ , it is evident that

$$\sup_{z \in \mathcal{T}^d} \left\{ \max_{j_1 \in [\widehat{m}_1 + \mathbf{g}_0], j_2 \notin [\widehat{m}_1 + \mathbf{g}_0]} \left( \left| \frac{\eta_{j_2}(z)}{\eta_{j_1}(z)} \right| \vee |\eta_{j_2}(z)| \right) \right\} \leq \left( 1 - \frac{\varepsilon_0}{3} \right). \quad (6.43)$$

Notice that, for  $s \in \{1, 2\}$

$$\widehat{d} > 0 \iff \widehat{m}_2 - N_+ > 0 \iff \widehat{m}_1 + \mathbf{g}_0 < N_-. \quad (6.44)$$

Therefore, as  $\widehat{d} > 0$ , we have that  $\mathcal{I} \setminus [\widehat{m}_1 + \mathbf{g}_0] \neq \emptyset$  for any set  $\mathcal{I} \subset [\widetilde{N}]$  of cardinality  $N_-$ . Thus, (6.43) now implies that

$$\sup_{z \in \mathcal{T}^d} \left| \prod_{j=1}^{\widehat{m}_1 + \mathbf{g}_0} \eta_j(z)^{-N} \cdot \prod_{i \in \mathcal{I}} \eta_i(z)^N \right| \leq \left( 1 - \frac{\varepsilon_0}{3} \right)^N,$$

for all such subsets  $\mathcal{I}$ . Using this together with (6.42) we now have part (a).

Turning to the proof of part (b), we recall (5.8) to observe that  $P_{\widehat{N}_+}(p, z; \widetilde{N})[X_\star^c; Y_\star^c + N_+]$  is a  $(\widehat{N}_+ - \widehat{m}_2)$ -dimensional Toeplitz matrix with symbol

$$p_\star(\tau) := \tau^{\widehat{m}_2 - N_+} \left[ \sum_{j=-N_-}^{N_+} a_j \tau^j - z \right].$$

Notice that the roots of  $p_\star(1/\cdot)$  coincide with those of  $p_z(\cdot)$ . Therefore, again by [15, Theorem 2.5], using that  $\widetilde{N} = \widehat{m}_1 + \widehat{m}_2 + \mathbf{g}_0$ , from (5.10) and (5.36) we deduce that

$$\widehat{\mathfrak{D}}(X_\star, Y_\star, z) = (-1)^{\widehat{N}(\widehat{m}_1 + \mathbf{g}_0)} a_{-N_-}^{-(\widehat{m}_2 - N_+)} \cdot \prod_{j=1}^{\widehat{m}_1 + \mathbf{g}_0} \eta_j(z)^{-(\widehat{m}_2 - N_+)} \sum_{\mathcal{I} \in \binom{[\widetilde{N}]}{[\widehat{m}_1 + \mathbf{g}_0]}} \mathcal{C}_{\mathcal{I}}(z) \cdot \frac{\prod_{i \in \mathcal{I}} \eta_i(z)^{\widehat{N}}}{\prod_{j=1}^{\widehat{m}_1 + \mathbf{g}_0} \eta_j(z)^{\widehat{N}}}, \quad (6.45)$$

where we use the shorthand  $\widehat{N} := \widehat{N}_+ - \widehat{m}_2 = N + N_+ - \widehat{m}_2$ .

Upon using (6.42) one finds that the summand in the RHS of (6.45) for  $\mathcal{I} = [\widehat{m}_1 + \mathbf{g}_0]$  is uniformly bounded below by  $3c_\star$  for some constant  $c_\star > 0$ . On the other hand, for  $\mathcal{I} \neq [\widehat{m}_1 + \mathbf{g}_0]$ , using (6.42)-(6.43) we deduce that it is exponentially small (in  $\widehat{N}$  and hence in  $N$ ) compared to  $c_\star$ . Therefore, the sum in (6.45) is uniformly bounded below by  $2c_\star$ . Using now that  $\sup_{z \in \mathcal{T}^d} \max_{j \in [\widetilde{m}]} |\eta_j(z)| < \infty$ , and shrinking  $c_\star$  if needed, complete the proof of the lower bound in (6.20) completes. The proof of the upper bound is immediate from (6.45), the above discussion, and the fact that  $\inf_{z \in \mathcal{T}^d} |\eta_{\widehat{m}_1 + \mathbf{g}_0}(z)| \geq 1/2$ . We omit further details.  $\square$

**6.3. Step 3: uniform lower bound on the dominant term.** In this section we prove a uniform lower bound on the dominant term in the expansion of  $\det(P_z^\delta)$ . The following is the main result of this section.

**Theorem 6.11.** *Fix  $\varepsilon_0, \widetilde{\varepsilon}_0 > 0$ . Assume that  $\widehat{d} > 0$  and that the entries of  $Q$  satisfy Assumption 1.2. Then, there exists  $C_{6.11} = C_{6.11}(p, \eta, \varepsilon_0, \widetilde{\varepsilon}_0) < \infty$  so that for all  $\varepsilon'_0$  sufficiently small,*

$$\mathbf{P} \left( \inf_{z \in \mathcal{T}_{\varepsilon'_0, \varepsilon_0}^{d, (s)} \setminus \mathcal{B}_2^{\widetilde{\varepsilon}_0}} |\widehat{\det}_{\widehat{m}_2 - N_+}^{(s)}(z)| \leq \frac{1}{2} (\varepsilon'_0)^{3\eta/4} \right) \leq C_{6.11} (\varepsilon'_0)^{\eta/3}, \quad s \in \{1, 2\}.$$

The proof of Theorem 6.11 uses the following anti-concentration bound for certain polynomials in independent random variables such that the degree of each of those random variables in those polynomials is at most one. This is a generalization of [9, Proposition 4.1].

**Lemma 6.12.** Fix  $k \in \mathbb{N}$  and let  $\{U_i\}_{i=1}^k$  be a sequence of independent complex-valued random variables, whose Lévy concentration functions satisfy the bound (1.4) with  $\eta \in (0, 1]$ . Let  $\{Z_{\mathcal{I}}; \mathcal{I} \in \binom{[k]}{n}\}$  be another collection of random variables which is jointly independent of the collection  $\{U_i\}_{i=1}^k$ . Define

$$\mathcal{U}_k := \sum_{\mathcal{I} \subset [k]} Z_{\mathcal{I}} \prod_{i \in \mathcal{I}} U_i.$$

Then, for any constant  $c_\star > 0$  and  $\varepsilon \in (0, c_\star e^{-1}]$  we have that

$$\mathbf{P}(|\mathcal{U}_k| \leq \varepsilon) \leq \bar{C}_{6.12} \cdot \left(\frac{\varepsilon}{c_\star}\right)^{(1+\eta)} \cdot \left(\log\left(\frac{c_\star}{\varepsilon}\right)\right)^{k-1} + \mathbf{P}(|Z_{[k]}| \leq c_\star),$$

where  $\bar{C}_{6.12} < \infty$  is some large constant depending only on  $\eta$  and  $C_{1.2}$ .

*Proof of Lemma 6.12.* We will show that

$$\begin{aligned} & \mathbf{P}(|\mathcal{U}_k| \leq \varepsilon | Z_{[k]}) \cdot \mathbf{1}_{\{|Z_{[k]}| > c_\star\}} \\ & \leq \mathbf{P}\left(|\hat{\mathcal{U}}_k| \leq c_\star^{-1} \varepsilon | Z_{[k]}\right) \cdot \mathbf{1}_{\{|Z_{[k]}| > c_\star\}} \leq \bar{C}_{6.12} \cdot \left(\frac{\varepsilon}{c_\star}\right)^{(1+\eta)} \left(\log\left(\frac{c_\star}{\varepsilon}\right)\right)^{k-1}, \end{aligned} \quad (6.46)$$

where

$$\hat{\mathcal{U}}_k := \sum_{\mathcal{I} \subset [k]} \hat{Z}_{\mathcal{I}} \prod_{i \in \mathcal{I}} U_i \quad \text{and} \quad \hat{Z}_{\mathcal{I}} := \frac{Z_{\mathcal{I}}}{Z_{[k]}}, \text{ for } \mathcal{I} \subset [k].$$

(Notice that  $\{\hat{Z}_{\mathcal{I}}; \mathcal{I} \subset [k]\}$  are well defined on the event  $\{|Z_{[k]}| > c_\star\}$ , and hence so is  $\hat{\mathcal{U}}_k$ .) Taking the expectation over  $Z_{[k]}$  in (6.46) together with a union bound immediately yields Lemma 6.12. The first inequality in (6.46) is straightforward from the definition of  $\hat{\mathcal{U}}_k$ . So, we only need to prove the second inequality in (6.46).

Turning to that task, for  $j \in [k]$ , set  $\mathcal{J}_j := \{j, j+1, \dots, k\}$ . Define

$$\hat{\mathcal{U}}_j := \sum_{\mathcal{I} \supset \mathcal{J}_{j+1}} \hat{Z}_{\mathcal{I}} \prod_{i \in \mathcal{I} \setminus \mathcal{J}_{j+1}} U_i \quad \text{and} \quad \tilde{\mathcal{U}}_j := \sum_{\substack{\mathcal{I} \supset \mathcal{J}_{j+2} \\ (j+1) \notin \mathcal{I}}} \hat{Z}_{\mathcal{I}} \prod_{i \in \mathcal{I} \setminus \mathcal{J}_{j+2}} U_i, \quad \text{for } j \in [k-1] \cup \{0\},$$

where  $\mathcal{J}_{k+1} = \emptyset$ . We will prove inductively that, for any  $j \in [k]$ , and all  $t \in (0, e^{-1}]$ ,

$$\mathbf{P}\left(|\hat{\mathcal{U}}_j| \leq t | Z_{[k]}\right) \leq (6(C_{1.4} \vee 1)e^{1+\eta})^j t^{1+\eta} \left(\log\left(\frac{1}{t}\right)\right)^{j-2}. \quad (6.47)$$

Plugging  $t = c_\star^{-1} \varepsilon$  and  $j = k$  in (6.47) yields (6.46).

Proceeding to the proof of (6.47), we start with  $j = 1$ . As  $\hat{Z}_{[k]} = 1$  we have that  $\hat{\mathcal{U}}_1 = U_1 + \tilde{\mathcal{U}}_0$ . Thus, as  $U_1$  is independent of  $\{Z_{\mathcal{I}}, \mathcal{I} \subset [k]\}$ , by (1.4) we have that

$$\mathbf{P}\left(|\hat{\mathcal{U}}_1| \leq t | Z_{[k]}\right) \leq \mathcal{L}(U_1, t) \leq C_{1.2} t^{1+\eta}.$$

To prove (6.47) for an arbitrary  $j \in [k]$ , upon using induction we note that

$$\hat{\mathcal{U}}_j = U_j \cdot \hat{\mathcal{U}}_{j-1} + \tilde{\mathcal{U}}_{j-1}, \quad \text{for } j \in [k]. \quad (6.48)$$

Observe that by our assumption the random variables  $\{U_{j-1}, \tilde{\mathcal{U}}_{j-1}\}$  are independent of  $U_j$  for any  $j \in [k]$ . This, in particular, allows us to use the anti-concentration property of  $U_j$  to derive the same for  $\hat{\mathcal{U}}_j$ .

To complete the proof of (6.47) by induction, assume it holds for  $j = j_*$ , and set  $C_j = (6(C_{1.4} \vee 1)e^{1+\eta})^j$ . The induction hypothesis yields that

$$\begin{aligned} & \mathbf{P} \left( \left| \widehat{\mathcal{U}}_{j_*+1} \right| \leq t \mid Z_{[k]} \right) \\ & \leq \mathbf{P} \left( \left| \widehat{\mathcal{U}}_{j_*} \right| \leq t \mid Z_{[k]} \right) + \mathbf{E} \left[ \mathbf{P} \left( \left| U_{j_*} + \frac{\widetilde{\mathcal{U}}_{j_*}}{\widehat{\mathcal{U}}_{j_*}} \right| \leq \frac{t}{\left| \widehat{\mathcal{U}}_{j_*} \right|} \mid \widehat{\mathcal{U}}_{j_*}, \widetilde{\mathcal{U}}_{j_*}, Z_{[k]} \right) \cdot \mathbf{1} \left( \left| \widehat{\mathcal{U}}_{j_*} \right| \geq t \right) \mid Z_{[k]} \right] \\ & \stackrel{(1.4)}{\leq} C_{j_*} t^{1+\eta} \left( \log \left( \frac{1}{t} \right) \right)^{j_*-2} + C_{1.2} t^{1+\eta} \cdot \mathbf{E} \left[ \left| \widehat{\mathcal{U}}_{j_*} \right|^{-(1+\eta)} \mathbf{1} \left( \left| \widehat{\mathcal{U}}_{j_*} \right| \geq t \right) \mid Z_{[k]} \right], \end{aligned} \quad (6.49)$$

where we have used that the triplet  $\{\widehat{\mathcal{U}}_{j_*}, \widetilde{\mathcal{U}}_{j_*}, Z_{[k]}\}$  is independent of  $U_{j_*}$ . Using integration by parts, for any probability measure  $\mu$  supported on  $[0, \infty)$  we have that

$$\int_t^{e^{-1}} x^{-(1+\eta)} d\mu(x) = e^{1+\eta} \mu([t, 1]) + (1+\eta) \cdot \int_t^{e^{-1}} \frac{\mu([t, s])}{s^{2+\eta}} ds.$$

Therefore, using the induction hypothesis and the fact that  $\eta \in (0, 1]$ , we have that

$$\begin{aligned} & \mathbf{E} \left[ \left| \widehat{\mathcal{U}}_{j_*} \right|^{-(1+\eta)} \mathbf{1} \left( \left| \widehat{\mathcal{U}}_{j_*} \right| \geq t \right) \mid Z_{[k]} \right] \leq e^{1+\eta} + \mathbf{E} \left[ \left| \widehat{\mathcal{U}}_{j_*} \right|^{-(1+\eta)} \mathbf{1} \left( \left| \widehat{\mathcal{U}}_{j_*} \right| \in [t, e^{-1}] \right) \mid Z_{[k]} \right] \\ & \leq 2e^{1+\eta} + 2 \int_t^{e^{-1}} \frac{\mathbf{P} \left( \left| \widehat{\mathcal{U}}_{j_*} \right| \leq s \mid Z_{[k]} \right)}{s^{2+\eta}} ds \leq 2e^{1+\eta} + 2C_{j_*} \int_t^{e^{-1}} s^{-1} \left( \log \left( \frac{1}{s} \right) \right)^{j_*-2} ds \\ & \leq 2e^{1+\eta} + \frac{2C_{j_*}}{(j_*-1) \vee 1} \left( \log \left( \frac{1}{t} \right) \right)^{j_*-1}. \end{aligned}$$

Combining the above with (6.49) and using that  $\log(1/t) \geq 1$  for  $t \leq e^{-1}$ , we establish (6.47) for  $j = j_* + 1$ . This concludes the induction argument and hence the proof of the lemma is complete.  $\square$

Lemmas 6.7(b) and 6.12 yield easily an anti-concentration bound for  $\det_{\widehat{m}_2 - N_+}(\cdot)$  per fixed  $z$ , as follows.

**Corollary 6.13.** *Consider the same setup as in Theorem 6.11. Then for  $s \in \{1, 2\}$  and  $\varepsilon \in (0, c_* e^{-1}]$  we have*

$$\sup_{z \in \mathcal{T}_{\varepsilon'_0, \varepsilon_0}^{d, (s)} \setminus \mathcal{B}_2^{\varepsilon_0}} \mathbf{P} \left( \left| \widehat{\det}_{\widehat{m}_2 - N_+}(z) \right| \leq \varepsilon \right) \leq \bar{C}_{6.12} \cdot \left( \frac{\varepsilon}{c_*} \right)^{(1+\eta)} \cdot \left( \log \left( \frac{c_*}{\varepsilon} \right) \right)^{\widehat{m}_2 - N_+ - 1},$$

where  $c_*$  is as in Lemma 6.7(b).

*Proof.* Recall (5.9), and (5.32)-(5.36). Fix  $s \in \{1, 2\}$ . We note that

$$\begin{aligned} \widehat{\det}_{\widehat{m}_2 - N_+}(z) &= (-1)^{\text{sgn}(\sigma_{X_*}) \text{sgn}(\sigma_{Y_*})} \widehat{\mathfrak{D}}(X_*, Y_*, z) \cdot \det(Q[X_*; Y_*]) \\ &\quad + \sum (-1)^{\text{sgn}(\sigma_X) \text{sgn}(\sigma_Y)} \widehat{\mathfrak{D}}(X, Y, z) \cdot \det(Q[X; Y]), \end{aligned} \quad (6.50)$$

where  $X_*$  and  $Y_*$  are as in Lemma 6.7(b), and the sum in the RHS of (6.50) is over subsets  $X, Y \subset [N]$  such that  $|X| = |Y| = \widehat{m}_2 - N_+$  and  $(X, Y) \neq (X_*, Y_*)$ . It is easy to see that  $\widehat{\det}_{\widehat{m}_2 - N_+}(z)$  is a homogeneous polynomial in the entries of  $Q$ , which are jointly independent, such that the degree of each of its entry in that polynomial is at most one. Hence, Lemma 6.12 is applicable. By Lemma 6.7(b) we have uniform lower bound on  $\widehat{\mathfrak{D}}(X_*, Y_*, z)$ . Therefore, upon applying Lemma 6.12 with  $k = \widehat{m}_2 - N_+$  and the roles of  $\{U_i\}_{i=1}^k$  being played by the diagonal entries of the sub matrix  $Q[X_*, Y_*]$ , we derive the desired anti-concentration bound.  $\square$

Corollary 6.13 is not sufficient to yield Theorem 6.11. As discussed in Section 6.2, we also need a bound on the supremum of the derivative of  $\det_{\widehat{m}_2 - N_+}(z)$ . This is derived below.

**Lemma 6.14.** *Let Assumption 1.1 hold. Fix  $\tilde{\varepsilon}_0$ . There exists some constant  $c_{6.14} > 0$ , depending only on  $p, \tilde{\varepsilon}_0$ , such that for all  $\varepsilon_0, \varepsilon'_0 > 0$  such that  $\varepsilon'_0/\varepsilon_0$  is small enough,*

$$\limsup_{N \rightarrow \infty} \mathbf{E} \left[ \sup_{z \in \left( \mathcal{T}_{\varepsilon'_0, \varepsilon_0}^{d, (s)} \setminus \mathcal{B}_2^{\tilde{\varepsilon}_0} \right)^{c_{6.14} \varepsilon'_0}} \left| \frac{d}{dz} \widehat{\det}_{\widehat{m}_2 - N_+}(z) \right|^2 \right] = O(1), \quad s \in \{1, 2\}. \quad (6.51)$$

*Proof.* We borrow ingredients from the proof of Lemma 6.6. Let  $s = 2$ . The proof for  $s = 1$  being the same will be omitted.

Our starting point is (6.40). Fix  $k = \widehat{m}_2 - N_+$ . As in the proof of Lemma 6.6 we set  $\varepsilon = \varepsilon_{6.9} \varepsilon''_0$  and fix  $\beta \in (0, 1/8)$ . Set  $\varepsilon''_0 = \varepsilon'_0$ . By (6.39)  $\mathbb{T}^{4\beta}$  is contained in the union of  $\mathcal{T}_{2\varepsilon'_0, \varepsilon_0/2}^{d, (2)}$  and  $\mathcal{T}_{2\varepsilon'_0, \varepsilon_0/2}^{d+g_0, (1)}$ . Apply Lemma 6.2(ii) for  $\mathcal{T}^{d, (2)}$  and  $\mathcal{T}^{d+g_0, (1)}$  with  $\varepsilon'_0$  and  $\varepsilon_0$  being replaced by  $2\varepsilon'_0$ , and  $\varepsilon_0/2$ , respectively to deduce that (6.41) holds under the current setup. Therefore, the desired bound follows from (6.40). This finishes the proof.  $\square$

**Remark 6.15.** *From (6.40)-(6.41) we have that the RHS of (6.51) is  $O((\varepsilon'_0)^{-3})$ . On the other hand Lemma 6.2(ii), and hence (6.41), hold as soon as  $\varepsilon'_0/\varepsilon_0 \leq c$  where  $c > 0$  is some constant depending only on the degree of the Laurent polynomial. This allows us to use Theorem 6.14 for  $\varepsilon'_0 = c\varepsilon_0$  and extend the result for  $\varepsilon'_0 < c\varepsilon_0$ , by noting that the sets  $(\mathcal{T}_{\varepsilon'_0, \varepsilon_0}^{d, (s)} \setminus \mathcal{B}_2^{\tilde{\varepsilon}_0})^{c_{6.14} \varepsilon'_0}$  are increasing in  $\varepsilon'_0$ . It in particular enables us to claim that the RHS of (6.51) is  $O(\varepsilon_0^{-3})$ . This observation will be used in the proof of Theorem 6.11 to claim that the constant  $C_{6.11}$  does not depend on  $\varepsilon'_0$ .*

We are now ready to prove Theorem 6.11.

*Proof of Theorem 6.11.* Fix  $s \in \{1, 2\}$ . Similar to the proof of Theorem 6.5 we will also use a covering argument.

Write

$$\tilde{\mathcal{F}}(z) := \left\{ \left| \widehat{\det}_{\widehat{m}_2 - N_+}(z) \right| \leq \varepsilon'_0 \varepsilon_\star^{-1} \right\}$$

and

$$\tilde{\mathcal{F}}_0 := \left\{ \sup_{z \in \left( \mathcal{T}_{\varepsilon'_0, \varepsilon_0}^{d, (s)} \setminus \mathcal{B}_2^{\tilde{\varepsilon}_0} \right)^{c_{6.14} \varepsilon'_0}} \left| \frac{d}{dz} \widehat{\det}_{\widehat{m}_2 - N_+}(z) \right| \geq \varepsilon_\star^{-1} \right\},$$

where  $\varepsilon_\star = (\varepsilon'_0)^{\eta/4}$ . By Lemmas 6.9(i) and 6.10 there exists a net  $\mathcal{N}_\star$  of  $\mathcal{T}_{\varepsilon'_0, \varepsilon_0}^{d, (s)}$  with mesh size  $c_{6.14} \varepsilon'_0/2$  such that  $|\mathcal{N}_\star| = O(\varepsilon_0^{-1})$ . Therefore, applying Corollary 6.13 and Lemma 6.14 (see also Remark 6.15) we have that

$$\mathbf{P} \left( \bigcup_{z \in \mathcal{N}_\star} \tilde{\mathcal{F}}(z) \cup \tilde{\mathcal{F}}_0 \right) = O \left( (\varepsilon'_0)^{\eta/3} \right), \quad (6.52)$$

for any  $\varepsilon'_0 > 0$  sufficiently small. On the other hand, for any  $z \in \mathcal{N}_\star$  we have that  $D(z, c_{6.14} \varepsilon'_0/2) \subset (\mathcal{T}_{\varepsilon'_0, \varepsilon_0}^{d, (s)} \setminus \mathcal{B}_2^{\tilde{\varepsilon}_0})^{c_{6.14} \varepsilon'_0/2}$ . Therefore, by the first order Taylor series expansion and triangle inequality it follows that on the event  $\bigcap_{z \in \mathcal{N}_\star} \tilde{\mathcal{F}}(z)^c \cup \tilde{\mathcal{F}}_0^c$ , for any  $z' \in \mathcal{T}_{\varepsilon'_0, \varepsilon_0}^{d, (s)} \setminus \mathcal{B}_2^{\tilde{\varepsilon}_0}$

$$\begin{aligned} \left| \widehat{\det}_{\widehat{m}_2 - N_+}(z') \right| &\geq \inf_{z \in \mathcal{N}_\star} \left| \widehat{\det}_{\widehat{m}_2 - N_+}(z) \right| - \text{dist}(z', \mathcal{N}_\star) \cdot \sup_{z \in \left( \mathcal{T}_{\varepsilon'_0, \varepsilon_0}^{d, (s)} \setminus \mathcal{B}_2^{\tilde{\varepsilon}_0} \right)^{c_{6.14} \varepsilon'_0}} \left| \frac{d}{dz} \widehat{\det}_{\widehat{m}_2 - N_+}(z) \right| \\ &\geq \frac{1}{2} \varepsilon'_0 \cdot \varepsilon_\star^{-1} = \frac{1}{2} (\varepsilon'_0)^{3\eta/4}. \end{aligned}$$

This together with (6.52) completes the proof of the theorem.  $\square$

**6.4. Step 4: combining bounds from Sections 6.1-6.3.** In this short section we combine the results obtained in the previous three sections and complete the proofs of Theorems 5.4 and 5.7. We start with the proof of Theorem 5.4. The estimates derived Sections 6.1-6.3 are for the tubes which are defined through the roots  $p_z(\cdot) = 0$ , while Theorem 5.4 is to be proved for regions that are defined via the spectral parameter  $z$ . Therefore, we will need to prove some additional geometric features of the tubes which will allow us to relate those to regions defined through  $z$ . These will also be used later in Section 7. Stating these results requires the following notation.

For any  $\epsilon > 0$  and set  $\mathcal{B} \subset \mathbb{C}$  we write  $\mathcal{B}^{-\epsilon} := ((\mathcal{B}^\epsilon)^c)^c$ . For any collection of  $(\mathfrak{g}_0 + 1)$  distinct indices  $\mathbf{i} := \{i_1, i_2, \dots, i_{\mathfrak{g}_0+1}\} \subset [\tilde{m}]$  and  $c > 0$  we set

$$\mathcal{W}_{\mathbf{i}}(c) := \left\{ z \in \mathbb{C} : \|\eta_{i_j}(z)\| - 1 + \|\eta_{i'_j}(z)\| - 1 \leq c, \text{ for all } j \neq j' \in [\mathfrak{g}_0 + 1] \right\}.$$

and

$$\mathcal{W}(c) := \{z : \exists \eta = \eta(z) \text{ such that } p_z(-\eta) = 0 \text{ and } \|\eta\| - 1 \leq c\}.$$

**Lemma 6.16.** *Fix  $\tilde{\epsilon}_0 > 0$ . We have the following geometric properties:*

(i) *There exists a constant  $C_{6.16} < \infty$  such that for any  $\epsilon > 0$ ,*

$$(p(S^1))^\epsilon \setminus \mathcal{B}_2^{\tilde{\epsilon}_0} \subset \mathcal{W}(C_{6.16}\epsilon) \setminus \mathcal{B}_2^{\tilde{\epsilon}_0}.$$

(ii) *There exists a constant  $c_{6.16} > 0$  such that for any  $c \leq c_{6.16}$ ,*

$$\left( \bigcup_{\mathbf{i} \subset [\tilde{m}]: |\mathbf{i}| = \mathfrak{g}_0 + 1} \mathcal{W}_{\mathbf{i}}(c\tilde{\epsilon}_0) \right) \cap (\mathcal{B}_1^{\tilde{\epsilon}_0} \cup \mathcal{B}_2^{\tilde{\epsilon}_0})^c = \emptyset. \quad (6.53)$$

(iii) *Fix  $0 < \epsilon_0 < \epsilon'_0$  such that  $\epsilon'_0 + \epsilon_0 \leq c_{6.16}\tilde{\epsilon}_0$ . Then, for any  $\epsilon > 0$  such that  $\epsilon \leq C_{6.16}^{-1}\epsilon'_0$ ,*

$$(p(S^1))^\epsilon \setminus (\mathcal{B}_1^{\tilde{\epsilon}_0} \cup \mathcal{B}_2^{\tilde{\epsilon}_0}) \subset \left( \bigcup_d \mathcal{T}_{\epsilon'_0, \epsilon_0}^d \right) \setminus (\mathcal{B}_1^{\tilde{\epsilon}_0} \cup \mathcal{B}_2^{\tilde{\epsilon}_0}).$$

(iv) *Let  $\epsilon_0, \epsilon'_0$ , and  $\epsilon$  be as in part (iii). Then,*

$$(\mathcal{S}_0^\epsilon \setminus \mathcal{S}_0) \setminus (\mathcal{B}_1^{\tilde{\epsilon}_0} \cup \mathcal{B}_2^{\tilde{\epsilon}_0}) \subset (\mathcal{T}_{\epsilon'_0, \epsilon_0}^{\mathfrak{g}_0, (1)} \cup \mathcal{T}_{\epsilon'_0, \epsilon_0}^{-\mathfrak{g}_0, (1)}) \setminus (\mathcal{B}_1^{\tilde{\epsilon}_0} \cup \mathcal{B}_2^{\tilde{\epsilon}_0}). \quad (6.54)$$

(v) *Let  $\epsilon_0, \epsilon'_0$ , and  $\epsilon$  be as in part (iii). Then,*

$$(\mathcal{S}_0 \setminus \mathcal{S}_0^{-\epsilon}) \setminus (\mathcal{B}_1^{\tilde{\epsilon}_0} \cup \mathcal{B}_2^{\tilde{\epsilon}_0}) \subset \mathcal{T}_{\epsilon'_0, \epsilon_0}^{0, (2)} \setminus \mathcal{B}_2^{\tilde{\epsilon}_0}.$$

*Proof.* The proof of part (i) is immediate. Fix  $z' \in (p(S^1))^\epsilon$ . Then there exists  $\hat{z} \in p(S^1)$  such that  $|z' - \hat{z}| \leq \epsilon$ . This, in particular implies that there exists  $i \in [\tilde{m}]$  such that  $\eta_i(\hat{z}) \in S^1$ . Therefore, by the triangle inequality,

$$\|\eta_i(z')\| - 1 \leq \|\eta_i(z') - \eta_i(\hat{z})\| \leq \max_{i \in [\tilde{m}]} \sup_{z \in (p(S^1))^{2\epsilon} \setminus \mathcal{B}_2^{\tilde{\epsilon}_0}} \left| \frac{d}{dz} \eta_i(z) \right| \cdot |z' - \hat{z}|. \quad (6.55)$$

Since  $p(S^1)$  is a bounded set, and the roots are analytic outside a neighborhood of  $\mathcal{B}_2$  we have that the supremum of the derivative of the roots is bounded  $(p(S^1))^{2\epsilon} \setminus \mathcal{B}_2^{\tilde{\epsilon}_0}$ . This yields part (i).

Next we turn to prove (ii). Let us first prove (6.53) for Laurent polynomials  $p(\cdot)$  such that  $\mathfrak{g}(p) = \mathfrak{g}_0 = 1$ . To this end, fix an arbitrary constant  $c > 0$ . Assume that (6.53) does not hold for this chosen  $c$ . Then there exist a  $z \notin \mathcal{B}_1^{\tilde{\epsilon}_0} \cup \mathcal{B}_2^{\tilde{\epsilon}_0}$  and a pair  $i \neq j \in [\tilde{m}]$  such that  $\|\eta_i(z)\| - 1 + \|\eta_j(z)\| - 1 \leq c\tilde{\epsilon}_0$ . Therefore, if (6.53) does not hold for any  $c > 0$  then we obtain a pair  $i \neq j \in [\tilde{m}]$ , a sequence  $\{z_n\}$ , and a subsequence of integers  $\{n_k\}$  such that

$$\|\eta_i(z_{n_k})\| - 1 + \|\eta_j(z_{n_k})\| - 1 \leq \tilde{\epsilon}_0/n_k \quad \text{and} \quad z_{n_k} \in \mathbb{C} \setminus (\mathcal{B}_1^{\tilde{\epsilon}_0} \cup \mathcal{B}_2^{\tilde{\epsilon}_0}), \quad \forall k. \quad (6.56)$$

Since for each of these  $z_{n_k}$  there exists a root of  $p_{z_{n_k}}(\cdot) = 0$  that is close to the unit circle, it follows that  $\{z_{n_k}\}$  is a bounded sequence (see also the proof of Lemma 6.9(i)), which thus possesses a converging subsequence, and hence may be assumed to converge to some  $z_\star \notin \mathcal{B}_2^{\tilde{\epsilon}_0}$ . Now, by the continuity of the maps  $z \mapsto \eta_i(z)$  and  $z \mapsto \eta_j(z)$  we derive from (6.56) that  $|\eta_i(z_\star)| = |\eta_j(z_\star)| = 1$ . As  $z_\star \notin \mathcal{B}_2^{\tilde{\epsilon}_0}$  we further deduce that  $\eta_i(z_\star) \neq \eta_j(z_\star)$ . This, in particular, implies that there exist

$\eta \neq \eta' \in S^1$  such that  $p(-\eta) = p(-\eta') = z_*$ , and hence  $z_* \in \mathcal{B}_1$ . This is a contradiction. Therefore, we have (6.53) for  $\mathfrak{g}_0 = 1$ .

To prove (6.53) for a Laurent polynomial  $p(\cdot)$  with  $\mathfrak{g}(p) = \mathfrak{g}_0 > 1$  we note that for such  $p(\cdot)$  there exists a Laurent polynomial  $\widehat{p}(\cdot)$ , with  $\mathfrak{g}(\widehat{p}) = 1$ , such that  $p(\eta) = \widehat{p}(\eta^{\mathfrak{g}_0})$  for  $\eta \in \mathbb{C}$ . This means that the roots of  $p_z(\cdot) = 0$  are obtained by taking the  $\mathfrak{g}_0$ -th roots of those of  $\widehat{p}_z(\cdot) = 0$ . Furthermore, the set of intersection points and branch points of  $p(\cdot)$  and  $\widehat{p}(\cdot)$  are identical. Since (6.53) holds for  $\widehat{p}(\cdot)$ , upon shrinking  $c_{6.16}$  it continues to hold for  $p(\cdot)$ .

The proof of (iii) is also straightforward. Indeed, by parts (i) and (ii) we have

$$(p(S^1))^{C_{6.16}^{-1}\varepsilon'_0} \setminus (\mathcal{B}_1^{\tilde{\varepsilon}_0} \cup \mathcal{B}_2^{\tilde{\varepsilon}_0}) \subset \mathcal{W}(\varepsilon'_0) \setminus (\mathcal{B}_1^{\tilde{\varepsilon}_0} \cup \mathcal{B}_2^{\tilde{\varepsilon}_0}) = \mathcal{W}(\varepsilon'_0) \cap \left( \bigcup_{i \in [\tilde{m}]: |i| = \mathfrak{g}_0 + 1} \mathcal{W}_i(c_{6.16}\tilde{\varepsilon}_0) \right)^c \setminus (\mathcal{B}_1^{\tilde{\varepsilon}_0} \cup \mathcal{B}_2^{\tilde{\varepsilon}_0}) \quad (6.57)$$

Note that by the definition of  $\mathcal{W}(\cdot)$  it follows that any  $z$  belonging to rightmost set in (6.57) must have a root  $-\eta(z)$  of  $p_z(\cdot) = 0$  such that  $|\eta(z)| - 1 \leq \varepsilon'_0$ . As  $\mathfrak{g}(p) = \mathfrak{g}_0$  there are  $\mathfrak{g}_0$  such roots. Recalling the definition of  $\mathcal{W}_i(\cdot)$  we deduce that distance of all other roots from  $S^1$  must be at least  $\varepsilon_0$ . Hence,  $z \in \mathcal{T}_{\varepsilon'_0, \varepsilon_0}^d$  for some  $d$ . This yields part (iii).

Turning to prove part (iv) we begin with the proof of the following intermediate result:

$$(\mathcal{S}_0^\varepsilon \setminus \mathcal{S}_0) \setminus (\mathcal{B}_1^{\tilde{\varepsilon}_0} \cup \mathcal{B}_2^{\tilde{\varepsilon}_0}) \subset (p(S^1))^\varepsilon \setminus \mathcal{S}_0 \setminus (\mathcal{B}_1^{\tilde{\varepsilon}_0} \cup \mathcal{B}_2^{\tilde{\varepsilon}_0}). \quad (6.58)$$

for any  $\varepsilon < \tilde{\varepsilon}_0/2$ . To see (6.58), we note that given any  $z$  belonging to the set on the LHS of (6.58) there exists  $z_* \in \mathcal{S}_0$  such that  $|z - z_*| \leq \varepsilon$ . Define  $z_*(t) := z_*(1-t) + tz$ ,  $t \in [0, 1]$ . We claim that there exists  $t_* \in (0, 1]$  such that  $z_*(t_*) \in p(S^1)$ . This will indeed show that  $z \in (p(S^1))^\varepsilon$ .

To prove the existence of  $t_*$ , as  $z_* \in \mathcal{S}_0$ , we see that,  $|\eta_{N_-}(z_*)| > 1 > |\eta_{N_-+1}(z_*)|$ . As  $z \notin \mathcal{S}_0$ , we have either  $|\eta_{N_-}(z)| \leq 1$  or  $|\eta_{N_-+1}(z)| \geq 1$ . Assume  $|\eta_{N_-}(z)| \leq 1$ . If an equality holds then we set  $t_* = 1$ . So, assume further that  $|\eta_{N_-}(z)| < 1$ . Since the map  $z_0 \mapsto |\eta_{N_-}(z_0)|$  is well defined and continuous for  $z_0 \notin \mathcal{B}_2^{\tilde{\varepsilon}_0/2}$ , the existence of  $t_* \in (0, 1)$  such that  $|\eta_{N_-}(z_*(t_*))| = 1$ , and hence  $z_*(t_*) \in p(S^1)$ , is immediate from the intermediate value theorem. Similarly for the case  $|\eta_{N_-+1}(z)| \geq 1$  one can repeat the above argument to find  $t_* \in (0, 1]$  such that  $|\eta_{N_-+1}(z_*(t_*))| = 1$  implying again  $z_*(t_*) \in p(S^1)$ . This yields (6.58).

Returning to the proof of (6.54) we apply part (iii) to deduce that for any  $\varepsilon > 0$  such that  $\varepsilon \leq C_{6.16}^{-1}\varepsilon'_0$  we have

$$(\mathcal{S}_0^\varepsilon \setminus \mathcal{S}_0) \setminus (\mathcal{B}_1^{\tilde{\varepsilon}_0} \cup \mathcal{B}_2^{\tilde{\varepsilon}_0}) \subset \left( \bigcup_d \mathcal{T}_{\varepsilon'_0, \varepsilon_0}^d \right) \setminus (\mathcal{B}_1^{\tilde{\varepsilon}_0} \cup \mathcal{B}_2^{\tilde{\varepsilon}_0}).$$

Thus, upon noting that as  $\mathfrak{g}(p) = \mathfrak{g}_0$  the winding number should be a multiple of  $\mathfrak{g}_0$ , it suffices to show that

$$(\mathcal{S}_0^\varepsilon \setminus \mathcal{S}_0) \cap \left( \bigcup_{d: |d| \geq 2\mathfrak{g}_0} \mathcal{T}_{\varepsilon'_0, \varepsilon_0}^d \cup \mathcal{T}_{\varepsilon'_0, \varepsilon_0}^{\pm \mathfrak{g}_0, (2)} \right) = \emptyset. \quad (6.59)$$

Assume that there exists  $z \in (\mathcal{S}_0^\varepsilon \setminus \mathcal{S}_0) \cap \mathcal{T}_{\varepsilon'_0, \varepsilon_0}^d$  for some  $d$  such that  $|d| \geq 2\mathfrak{g}_0$ . Again we need to split to the cases  $|\eta_{N_-}(z)| \leq 1$  and  $|\eta_{N_-+1}(z)| \geq 1$ . We first consider the case  $|\eta_{N_-}(z)| \leq 1$ . Recall that  $d = m_+ - N_+ = N_- - m_-$ . By the definition of the tube  $\mathcal{T}_{\varepsilon'_0, \varepsilon_0}^d$  we note that  $d \geq 2\mathfrak{g}_0$  implies that  $|\eta_{N_-}(z)| \leq 1 - \varepsilon_0$ . On the other hand,  $d \leq -2\mathfrak{g}_0$  implies that  $|\eta_{N_-}(z)|^{-1} \leq 1 - \varepsilon_0$ . However, we also recall from above that  $|\eta_{N_-}(z)| \leq 1$  implies that there exists  $z_*(t_*) \in p(S^1)$  such that  $|\eta_{N_-}(z_*(t_*))| = 1$  and  $|z - z_*(t_*)| \leq \varepsilon$ . As  $\varepsilon \leq C_{6.16}^{-1}\varepsilon'_0$ , and  $\varepsilon'_0 \leq \varepsilon_0/2$ , this together with (6.55) indeed yields a contradiction. A similar argument works for the case  $|\eta_{N_-+1}(z)| \geq 1$ . One can repeat the same argument again to further show that  $(\mathcal{S}_0^\varepsilon \setminus \mathcal{S}_0) \cap \mathcal{T}_{\varepsilon'_0, \varepsilon_0}^{\pm \mathfrak{g}_0, (2)} = \emptyset$ . We omit the details.

Finally we proceed to prove part (v). Repeating an argument similar to the proof of (6.58) we obtain that

$$(\mathcal{S}_0 \setminus \mathcal{S}_0^{-\varepsilon}) \setminus (\mathcal{B}_1^{\tilde{\varepsilon}_0} \cup \mathcal{B}_2^{\tilde{\varepsilon}_0}) \subset (p(S^1))^\varepsilon \cap \mathcal{S}_0 \setminus (\mathcal{B}_1^{\tilde{\varepsilon}_0} \cup \mathcal{B}_2^{\tilde{\varepsilon}_0}). \quad (6.60)$$

Now by part (iii) the set on the RHS of (6.60) must be contained in the union of the tubes. However, the only tube that has nonempty intersection with  $\mathcal{S}_0$  is  $\mathcal{T}_{\varepsilon'_0, \varepsilon_0}^{0, (2)}$ . This yields part (v) and hence the proof of the lemma is now complete.  $\square$



We are ready to prove Theorem 5.4.

*Proof of Theorem 5.4.* Fix  $\gamma' > 1$  such that  $\gamma - 1 < \gamma' < \gamma$ . Set  $\varepsilon'_0 = (\gamma' - 1) \log N/N$ ,  $\varepsilon_0 = c_{6.16} \tilde{\varepsilon}_0/2$ , some arbitrary constant  $\varepsilon''_0 > 0$  such that  $\varepsilon''_0/\varepsilon_0$  sufficiently small,  $\varepsilon_1 = C_{6.16}^{-1} \varepsilon'_0$ , and  $\varepsilon_2 = C_{6.16}^{-1} \varepsilon''_0$ . By Lemma 6.16(iv)-(v) we note that

$$(\mathcal{S}_0^{\varepsilon_1} \setminus \mathcal{S}_0) \setminus (\mathcal{B}_1^{\tilde{\varepsilon}_0} \cup \mathcal{B}_2^{\tilde{\varepsilon}_0}) \subset (\mathcal{T}_{\varepsilon'_0, \varepsilon_0}^{\mathbf{g}_0, (1)} \cup \mathcal{T}_{\varepsilon'_0, \varepsilon_0}^{-\mathbf{g}_0, (1)}) \setminus (\mathcal{B}_1^{\tilde{\varepsilon}_0} \cup \mathcal{B}_2^{\tilde{\varepsilon}_0})$$

and

$$(\mathcal{S}_0 \setminus \mathcal{S}_0^{-\varepsilon_2}) \setminus (\mathcal{B}_1^{\tilde{\varepsilon}_0} \cup \mathcal{B}_2^{\tilde{\varepsilon}_0}) \subset \mathcal{T}_{\varepsilon''_0, \varepsilon_0}^{0, (2)} \setminus \mathcal{B}_2^{\tilde{\varepsilon}_0}.$$

Therefore, by [9, Theorem 1.1] and a union bound we now derive that

$$\begin{aligned} & \limsup_{N \rightarrow \infty} \mathbf{P} \left( \exists z \in \mathcal{S}_0^{\varepsilon_1} \setminus (\mathcal{B}_1^{\tilde{\varepsilon}_0} \cup \mathcal{B}_2^{\tilde{\varepsilon}_0}) : \det(P_z^\delta) = 0 \right) \\ & \leq \limsup_{N \rightarrow \infty} \mathbf{P} \left( \exists z \in (\mathcal{S}_0^{\varepsilon_1} \setminus \mathcal{S}_0^{-\varepsilon_2}) \setminus (\mathcal{B}_1^{\tilde{\varepsilon}_0} \cup \mathcal{B}_2^{\tilde{\varepsilon}_0}) : \det(P_z^\delta) = 0 \right) \\ & \leq \limsup_{N \rightarrow \infty} \left[ \mathbf{P} \left( \exists z \in \widehat{\mathcal{T}}_{\gamma', \varepsilon_0}^{\pm \mathbf{g}_0, (1)} \setminus \mathcal{B}_2^{\tilde{\varepsilon}_0} : \det(P_z^\delta) = 0 \right) + \mathbf{P} \left( \exists z \in \mathcal{T}_{\varepsilon''_0, \varepsilon_0}^{0, (2)} \setminus \mathcal{B}_2^{\tilde{\varepsilon}_0} : \det(P_z^\delta) = 0 \right) \right]. \end{aligned} \quad (6.61)$$

It remains to show that each of the terms in the RHS of (6.61) equals zero in the limit. For  $d \in \{0, \mathbf{g}_0\}$ , upon recalling (5.6), this follows from Theorem 6.5, Lemma 6.7(b) (see also Remark 6.8), and the triangle inequality. To see that the same holds for  $d = -\mathbf{g}_0$ , as  $P_N^\Gamma$  is a Toeplitz matrix with symbol  $\tilde{p}(\cdot) = p(1/\cdot)$  and hence  $\text{ind}_{p(S^1)}(\cdot) = -\text{ind}_{\tilde{p}(S^1)}(\cdot)$ , we apply Theorem 6.5 and Lemma 6.7(c) for  $(P_z^\delta)^\Gamma$ . This completes the proof of the theorem.  $\square$

We end the section with the proof of Theorem 5.7.

*Proof of Theorem 5.7.* The case  $d = 0$  is covered by Theorem 5.4. Consider the case  $d > 0$ . Fix  $s \in \{1, 2\}$ . Since  $0 < \gamma - \gamma' < 1$  we observe that, for  $\varepsilon''_0 = N^{-(\gamma - \gamma')/8}$ ,

$$\widehat{\mathcal{T}}_{\gamma', \varepsilon_0}^{d, (s)} \setminus \mathcal{B}_2^{\tilde{\varepsilon}_0} \subset \mathcal{T}_{\varepsilon''_0, \varepsilon_0}^{d, (s)} \setminus \mathcal{B}_2^{\tilde{\varepsilon}_0},$$

for all large  $N$ . Therefore, applying Theorems 6.5 and 6.11 (with  $\varepsilon'_0$  replaced by  $\varepsilon''_0$ ), and upon using the triangle inequality it is immediate that

$$\mathbf{P} \left( \exists z \in \widehat{\mathcal{T}}_{\gamma', \varepsilon_0}^{d, (s)} \setminus \mathcal{B}_2^{\tilde{\varepsilon}_0} : \det(P_z^\delta) = 0 \right) \leq 2N^{-\eta(\gamma - \gamma')/32},$$

for all large  $N$ , yielding the desired result for  $d > 0$ . To prove the same for  $d < 0$ , as  $\det(P_z^\delta) = \det((P_z^\delta)^\Gamma)$  and  $P_N^\Gamma$  is a Toeplitz matrix with symbol  $p(1/\cdot)$ , we work with  $(P_z^\delta)^\Gamma$  and proceed same as above. This finishes the proof.  $\square$

**Remark 6.17.** *Theorem 5.7 and Lemma 6.16(ii) imply, upon recalling the definitions of the tube  $\widehat{\mathcal{T}}_{\gamma', \varepsilon_0}^d$  and  $\mathbf{g}_0$ , that with probability approaching one, for any eigenvalue  $z \notin \mathcal{B}_1^{\tilde{\varepsilon}_0} \cup \mathcal{B}_2^{\tilde{\varepsilon}_0}$  one has that*

$$p^{-1}(z) \cap (S^1 + D(0, C_\gamma N^{-1} \log N)) = \emptyset,$$

for some appropriately chosen  $C_\gamma$  such that  $\lim_{\gamma \rightarrow 1} C_\gamma = 0$  and  $\lim_{\gamma \rightarrow \infty} C_\gamma = \infty$ .

## 7. LOCATION OF THE BULK OF THE EIGENVALUES

In this section we prove that away from the bad sets, the bulk of the spectrum of  $P_{N, \gamma}^Q$  is contained in tubes of width  $O(\log N/N)$  around  $p(S^1)$ , with high probability, see Theorem 7.1. We also prove in Theorem 7.4 an upper bound on the number of eigenvalues of  $P_{N, \gamma}^Q$  residing in the ball  $D(x, \alpha \log N/N)$ , where the distance of  $x$  from the spectral curve is of the order  $\log N/N$  and  $\alpha < \infty$  is some large constant. The latter estimate will be used in the proof of Corollary 1.7.

Before stating the results, recall the bad sets of Definition 1.3, and define for  $\varepsilon, \alpha, \tilde{\varepsilon}_0 > 0$ ,

$$\Omega_\varepsilon := \{z \in \mathbb{C} : \text{dist}(z, p(S^1)) > \varepsilon\}, \quad \mathcal{N}_\alpha := \{z : z \in \Omega_{\alpha \log N/N} \setminus (\mathcal{B}_1^{\tilde{\varepsilon}_0} \cup \mathcal{B}_2^{\tilde{\varepsilon}_0}) \text{ an eigenvalue of } P_{N, \gamma}^Q\}. \quad (7.1)$$

The first main result of this section is the following.



**Theorem 7.1.** Fix  $\gamma > 1$  and  $\tilde{\varepsilon}_0 > 0$ . Let Assumptions 1.1 and 1.2 hold. Then there exists a constant  $C_{7.1} < \infty$ , depending only on  $p(\cdot)$ ,  $\gamma$ , and  $\tilde{\varepsilon}_0$ , so that for any  $\alpha > 0$ , we have

$$\lim_{N \rightarrow \infty} \mathbf{P} \left( |\mathcal{N}_\alpha| \leq \frac{C_{7.1} N}{\alpha} \right) = 1. \quad (7.2)$$

To prove Theorem 7.1 we will use Jensen's formula. For an analytic function  $f$ , let  $n_f(x, r)$  denote the number of roots in a ball of radius  $r$  around  $x$ . If  $f(x) \neq 0$ , then we have

$$\int_0^r \frac{n_f(x, s)}{s} ds = \frac{1}{2\pi} \int_0^{2\pi} \log |f(x + re^{i\theta})| d\theta - \log |f(x)|. \quad (7.3)$$

In particular, we obtain from (7.3) the bound

$$n_f(x, ur) \leq \frac{1}{2(1-u)\pi} \int_0^{2\pi} \log |f(x + re^{i\theta})| d\theta - \frac{1}{1-u} \log |f(x)|, \quad (7.4)$$

valid for  $u \in (0, 1)$ . We will apply (7.4) with  $f(z) = \det(P_{N,\gamma}^Q - zI_N)$ . Introduce the function

$$\phi_\infty(z) := \frac{1}{2\pi} \int_0^{2\pi} \log |z - p(e^{i\theta})| d\theta. \quad (7.5)$$

To prove Theorem 7.1 we will need the following two lemmas. Recall that  $P_z^\delta = P_{N,\gamma}^Q - zI_N$ .

**Lemma 7.2** (Upper bound on determinant). Consider the setup as in Theorem 7.1. Then there exist constants  $0 < c_{7.2}, C_{7.2} < \infty$ , depending only on  $\gamma, \tilde{\varepsilon}_0$  and  $p(\cdot)$ , such that

$$\mathbf{P}(\mathcal{A}_{UB}^c) \leq N^{-3},$$

for all large  $N$ , where

$$\mathcal{A}_{UB} := \bigcap_{z \in \Omega_{1/N} \cap (p(S^1))^{c_{7.2}\tilde{\varepsilon}_0} \setminus (\mathcal{B}_1^{\tilde{\varepsilon}_0} \cup \mathcal{B}_2^{\tilde{\varepsilon}_0})} \left\{ \log |\det(P_z^\delta)| \leq N\phi_\infty(z) + C_{7.2} \log N \right\}.$$

We also need the following complementary lower bounds.

**Lemma 7.3** (Lower bound on determinant). Consider the setup of Theorem 7.1. Then, there exists a constant  $C_{7.3} < \infty$  depending only on  $\gamma, \tilde{\varepsilon}_0$ , and  $p(\cdot)$ , such that

$$\max_{z \in (p(S^1))^{c_{7.2}\tilde{\varepsilon}_0} \setminus (\mathcal{B}_1^{\tilde{\varepsilon}_0} \cup \mathcal{B}_2^{\tilde{\varepsilon}_0})} \mathbf{P}(\mathcal{A}_{z, LB}^c) \leq 1/N^3, \quad (7.6)$$

where

$$\mathcal{A}_{z, LB} := \left\{ \log |\det(P_z^\delta)| \geq N\phi_\infty(z) - C_{7.3} \log N \right\}.$$

Using these two lemmas, whose proofs are postponed, we now prove Theorem 7.1.

*Proof of Theorem 7.1 (assuming Lemmas 7.2 and 7.3).* Let  $R := 2 \max_{z \in S^1} |p(z)|$ . For  $i \in \mathbb{N}$  such that  $i \leq i_\star := \lceil \log(c_{7.2}\tilde{\varepsilon}_0 N / (4\alpha \log N)) / \log 2 \rceil$ , set

$$\Omega_{\alpha, i} := \left\{ z \in D(0, R) \setminus (\mathcal{B}_1^{\tilde{\varepsilon}_0} \cup \mathcal{B}_2^{\tilde{\varepsilon}_0}) : \left( \frac{3\alpha}{4} + 2^{i-3}\alpha \right) \cdot \frac{\log N}{N} \leq \text{dist}(z, p(S^1)) \leq \left( \frac{3\alpha}{4} + 2^{i-2}\alpha \right) \cdot \frac{\log N}{N} \right\},$$

and define  $\mathcal{N}_{\alpha, i} := \mathcal{N}_\alpha \cap \Omega_{\alpha, i}$ . From our choice of  $i_\star$  it is clear that

$$\left( D(0, R) \setminus (\mathcal{B}_1^{\tilde{\varepsilon}_0} \cup \mathcal{B}_2^{\tilde{\varepsilon}_0}) \right) \setminus \left( \bigcup_{i=1}^{i_\star} \Omega_{\alpha, i} \right) \subset \left( D(0, R) \setminus (\mathcal{B}_1^{\tilde{\varepsilon}_0} \cup \mathcal{B}_2^{\tilde{\varepsilon}_0}) \right) \cap \Omega_{c_{7.2}\tilde{\varepsilon}_0/32}. \quad (7.7)$$

By [9, Theorem 1.1] we have that

$$\lim_{N \rightarrow \infty} \mathbf{P} \left( |\mathcal{N}_\alpha \cap D(0, R)^c| > 0 \right) \leq \lim_{N \rightarrow \infty} \mathbf{P} \left( \exists z \notin D(0, R) : z \text{ is an eigenvalue of } P_{N,\gamma}^Q \right) = 0.$$

On the other hand, we have from [65, Corollary 2.2] or [8, Theorem 1.2] that

$$\begin{aligned} & \lim_{N \rightarrow \infty} \mathbf{P} \left( |\mathcal{N}_\alpha \cap \Omega_{c_{7.2}\tilde{\varepsilon}_0/32}| \geq \frac{C'_0 N}{2\alpha} \right) \\ & \leq \lim_{N \rightarrow \infty} \mathbf{P} \left( \left| \left\{ z \in \Omega_{c_{7.2}\tilde{\varepsilon}_0/32} : z \text{ is an eigenvalue of } P_{N,\gamma}^Q \right\} \right| \geq \frac{C'_0 N}{2\alpha} \right) = 0, \end{aligned}$$

for any  $\alpha > 0$ . Hence, in light of (7.7) it suffices to show that

$$\liminf_{N \rightarrow \infty} \mathbf{P} (|\mathcal{N}_{\alpha,i}| = O(2^{-i}\alpha^{-1}N) \text{ for all } i \in [i_\star]) = 1. \quad (7.8)$$

Turning to prove (7.8), we cover  $\Omega_{\alpha,i}$  by a collection of balls  $D_{i,j} := D(z_{i,j}, r_i)$ ,  $j \in N_i$ , with  $z_{i,j} \in \Omega_{\alpha,i}$  and  $r_i = 2^{i-4}\alpha \cdot (\log N/N)$ . As  $z_{i,j} \in \Omega_{\alpha,i}$  we have that  $\text{dist}(z_{i,j}, \mathcal{B}_1 \cup \mathcal{B}_2) \geq \tilde{\varepsilon}_0$ . For  $i \leq i_\star$  we also have that  $r_i \leq r_{i_\star} \leq \tilde{c}_{7.2}\varepsilon_0/32$ . Therefore, there exists an absolute constant  $u \in (0, 1)$ , such that for any  $i \leq i_\star$  and  $j \in [N_i]$ ,

$$\text{dist}(D(z_{i,j}, u^{-1}r_i), \mathcal{B}_1 \cup \mathcal{B}_2) \geq \tilde{3}\varepsilon_0/4.$$

On the other hand, we note that

$$(3\alpha/4) \cdot \log N/N \leq \text{dist}(D(z_{i,j}, u^{-1}r_i), p(S^1)) \leq c_{7.2}\tilde{\varepsilon}_0/4.$$

The last two observations together imply that

$$\bigcup_{i=1}^{i_\star} \bigcup_{j=1}^{N_i} D(z_{i,j}, u^{-1}r_i) \subset \Omega_{1/(2N)} \cap (p(S^1))^{c_{7.2}\tilde{\varepsilon}_0/4} \setminus (\mathcal{B}_1^{\tilde{\varepsilon}_0/2} \cup \mathcal{B}_2^{\tilde{\varepsilon}_0/2}). \quad (7.9)$$

This allows us to apply Lemmas 7.2 and 7.3. Now, using (7.4), we have with

$$n_{i,j} := |\{z \in D_{i,j} : z \text{ is an eigenvalue of } P_{N,\gamma}^Q\}|$$

that

$$n_{i,j} \leq \frac{1}{2(1-u)\pi} \int_0^{2\pi} \log |\det(P^\delta - (z_{i,j} + u^{-1}r_{i,j}e^{i\theta})I_N)| d\theta - \frac{1}{1-u} \log |\det(P_{N,\gamma}^Q - z_{i,j}I_N)|,$$

where  $u \in (0, 1)$  is as above. Let  $\mathcal{A}_{UB}$  be as in Lemma 7.2 with  $\tilde{\varepsilon}_0$  replaced by  $\tilde{\varepsilon}_0/2$ . Then, on the event  $\mathcal{A}_{UB} \cap \mathcal{A}_{z_{i,j},LB}$ , as  $\phi_\infty(z)$  is harmonic off  $p(S^1)$ , by (7.9) we have that

$$n_{i,j} \leq (1-u)^{-1}(C_{7.2} + C_{7.3}) \log N.$$

By Lemma 6.10 it follows that  $N_i$ , the number of balls of radius  $r_i$  needed to cover  $\Omega_{\alpha,i}$  satisfies the bound

$$N_i \leq C2^{-i} \cdot \alpha^{-1} \cdot (N/\log N), \quad (7.10)$$

for some universal constant  $C < \infty$ . Thus, on  $\mathcal{A}_\star := \mathcal{A}_{UB} \cap \bigcap_{i,j} \mathcal{A}_{z_{i,j},LB}$ ,

$$|\mathcal{N}_{\alpha,i}| \leq \sum_{j \in N_i} n_{i,j} \leq (1-u)^{-1}(C_{7.2} + C_{7.3})CN \cdot 2^{-i}\alpha^{-1}.$$

Therefore (7.8) and hence the theorem follows if we can prove  $\mathbf{P}(\mathcal{A}_\star) = 1 - o(1)$ . Note however that

$$\mathbf{P}(\mathcal{A}_\star) \geq 1 - \mathbf{P}(\mathcal{A}_{UB}^c) - \sum_{i=1}^{i_\star} \sum_{j=1}^{N_i} \mathbf{P}(\mathcal{A}_{z_{i,j},LB}^c) \geq 1 - N^{-3} - N^{-3} \left( \sum_{i \leq i_\star} N_i \right) = 1 - o(1),$$

where we used (7.10) and Lemmas 7.2 and 7.3 with  $\tilde{\varepsilon}_0$  replaced by  $\tilde{\varepsilon}_0/2$ .  $\square$

Using the same ideas, we next provide a local upper bound on the number of eigenvalues.

**Theorem 7.4.** *Consider the setup as in Theorem 7.1. Fix  $0 < \underline{\alpha} < \bar{\alpha} < \infty$ . Then, there exists a constant  $C_{7.4} < \infty$  depending only on  $\gamma, \tilde{\varepsilon}_0$ , and  $p(\cdot)$ , such that for any  $z \in (p(S^1))^{\bar{\alpha} \log N/N}$ ,*

$$\mathbf{P} (|\mathcal{N}_{\underline{\alpha}, \bar{\alpha}, z}| \geq C_{7.4} \underline{\alpha}^{-2} \bar{\alpha}^2 \cdot \log N) \leq N^{-2},$$

for all large  $N$ , with

$$\mathcal{N}_{\underline{\alpha}, \bar{\alpha}, z} := \left\{ z' \in D(z, \bar{\alpha} \log N/N) \cap \Omega_{\underline{\alpha} \log N/N} \setminus (\mathcal{B}_1^{\tilde{\varepsilon}_0} \cup \mathcal{B}_2^{\tilde{\varepsilon}_0}) : z' \text{ is an eigenvalue of } P_{N,\gamma}^Q \right\}.$$

*Proof.* We follow the same strategy as in the proof of Theorem 7.1. Fix  $z \in (p(S^1))^{\bar{\alpha} \log N/N}$ . For  $i \in [\bar{i}_*]$ , where  $\bar{i}_* := \lceil \log(8\bar{\alpha}/\underline{\alpha}) / \log 2 \rceil$ , we set

$$\bar{\Omega}_{\underline{\alpha}, \bar{\alpha}, i} := \left\{ z' \in D(z, \bar{\alpha} \log N/N) \cap \Omega_{\underline{\alpha} \log N/N} \setminus (\mathcal{B}_1^{\tilde{\varepsilon}_0} \cup \mathcal{B}_2^{\tilde{\varepsilon}_0}) : \right. \\ \left. \left( \frac{3\underline{\alpha}}{4} + 2^{i-3}\underline{\alpha} \right) \cdot \frac{\log N}{N} \leq \text{dist}(z', p(S^1)) \leq \left( \frac{3\underline{\alpha}}{4} + 2^{i-2}\underline{\alpha} \right) \cdot \frac{\log N}{N} \right\}.$$

Since  $\text{dist}(z, p(S^1)) \leq \bar{\alpha} \log N/N$ , we notice that

$$\bigcup_{i=1}^{\bar{i}_*} \bar{\Omega}_{\underline{\alpha}, \bar{\alpha}, i} \supset D(z, \bar{\alpha} \log N/N) \cap \Omega_{\underline{\alpha} \log N/N} \setminus (\mathcal{B}_1^{\tilde{\varepsilon}_0} \cup \mathcal{B}_2^{\tilde{\varepsilon}_0}).$$

Thus, it suffices to bound the number of eigenvalues in  $\bigcup_{i=1}^{\bar{i}_*} \bar{\Omega}_{\underline{\alpha}, \bar{\alpha}, i}$ .

As in the proof of Theorem 7.1 we cover  $\bar{\Omega}_{\underline{\alpha}, \bar{\alpha}, i}$  by a collection of balls  $D_{i,j} = D(z_{i,j}, \bar{r}_i)$ ,  $j \in \bar{N}_i$ ,  $z_{i,j} \in \bar{\Omega}_{\underline{\alpha}, \bar{\alpha}, i}$  and  $\bar{r}_i = 2^{i-4}\underline{\alpha} \cdot (\log N/N)$ . By our choice of  $\bar{i}_*$  we procure a  $u \in (0, 1)$  such that

$$D(z_{i,j}, u^{-1}\bar{r}_i) \subset (p(S^1))^{3\bar{\alpha} \log N/N} \cap \Omega_{1/(2N)} \setminus (\mathcal{B}_1^{\tilde{\varepsilon}_0/2} \cup \mathcal{B}_2^{\tilde{\varepsilon}_0}), \quad \text{for all } j \in \bar{N}_i \text{ and } i \in [\bar{i}_*].$$

Hence, arguing similarly as in the proof of Theorem 7.1 we deduce that the number of eigenvalues of  $P_{N,\gamma}^Q$  in  $D_{i,j}$  is  $O(\log N)$  for all  $j \in \bar{N}_i$  and  $i \in [\bar{i}_*]$ , on a set with probability at least  $1 - 1/N^2$ .

To complete the proof we use a volumetric argument, yet again, to find that

$$\bar{N}_i = O\left(2^{-2i} \cdot \left(\frac{\bar{\alpha}}{\underline{\alpha}}\right)^2\right).$$

This, together with the choice of  $\bar{i}_*$  indeed yields the desired bound on  $\mathcal{N}_{\underline{\alpha}, \bar{\alpha}, z}$ .  $\square$

We now turn to the proof of Lemma 7.2. Recall  $\phi_\infty(z)$ , see (7.5), and define

$$\widetilde{\det}_k(z) := \frac{\det_k(z)}{a_{-N}^N \prod_{i=1}^{m_-} \eta_i(z)^N} = \frac{\det_k(z)}{\exp(N\phi_\infty(z))},$$

where the last equality follows upon recalling that  $\{-\eta_i(z)\}_{i=1}^{m_-}$  are the roots of  $p_z(\cdot) = 0$  that are greater than or equal to one in modulus, and from the fact that

$$\frac{1}{2\pi} \int_0^{2\pi} \log |\eta - e^{i\theta}| d\theta = \begin{cases} \log |\eta| & \text{if } |\eta| \geq 1, \\ 0 & \text{otherwise.} \end{cases}$$

In the lemma below we derive a bound on the supremum of the second moment  $\widetilde{\det}_k(\cdot)$ .

**Lemma 7.5.** *Let Assumption 1.1 hold. Fix  $\tilde{\varepsilon}_0, \varepsilon_0, \varepsilon'_0 > 0$  such that  $\varepsilon'_0/\varepsilon_0$  is sufficiently small. Fix  $k \in [N] \cup \{0\}$  and  $d \geq 0$ . Then for any  $\varepsilon \leq \varepsilon_{6.9} \varepsilon'_0/2$ , we have*

$$\sup_{z \in \left(\mathcal{T}_{\varepsilon'_0, \varepsilon_0}^{d, (s)} \setminus \mathcal{B}_2^{\tilde{\varepsilon}_0}\right)^\varepsilon} \mathbf{E} \left[ \left| \widetilde{\det}_k(z) \right|^2 \right] = O(1), \quad s \in \{1, 2\}.$$

*Proof.* By Lemma 6.9 it suffices to show that

$$\sup_{z \in \mathcal{T}_{\varepsilon'_0, \varepsilon_0}^{d, (s)}} \mathbf{E} \left[ \left| \widetilde{\det}_k(z) \right|^2 \right] = O(1). \quad (7.11)$$

We begin with the proof of (7.11) for  $k \in [N]$  and  $s = 1$ . Recall from (5.12) that for this choice of  $s$ , we have that  $\hat{m}_1 = m_-$ . Since  $Q$  satisfies Assumption 1.1, we then observe that

$$\mathbf{E} \left[ \left| \widetilde{\det}_k(z) \right|^2 \right] = N^{-2\gamma k} \sum_{\substack{X, Y \subset [N] \\ |X|=|Y|=k}} |\widehat{\mathcal{D}}(X, Y, z)|^2 \cdot |\eta_{\hat{m}_1+1}(z)|^{2g_0 N} \cdot \mathbf{E} \left[ |\det(Q[X; Y])|^2 \right]. \quad (7.12)$$

By Lemma 5.13(i)-(ii), for any  $k \in [N]$ , as  $|\eta_{\hat{m}_1+1}(\cdot)| \leq 1$  on  $\mathcal{T}^{d, (1)}$ , we have that

$$|\widehat{\mathcal{D}}(X, Y, z)|^2 \cdot |\eta_{\hat{m}_1+1}(z)|^{2g_0 N} \leq C^{\tilde{m}(k+N_+)}, \quad (7.13)$$

for some constant  $C < \infty$ . Since the second moment of the determinant of any  $k \times k$  matrix with entries satisfying Assumption 1.1 is  $k!$  and the number of ways one can choose two subsets of  $[N]$  that are cardinality  $k$  is  $\binom{N}{k}^2$  the claimed upper bound, for  $s = 1$ , now follows upon using that  $\gamma > 1$  and plugging the bound (7.13) in (7.12).

To prove (7.11) for  $s = 2$  we observe that  $\widehat{m}_1 + \mathbf{g}_0 = m_-$ , and thus  $\widetilde{\det}_k(\cdot) = N^{-\gamma d} \widehat{\det}_k(\cdot)$  on  $\mathcal{T}^{d,(2)}$ . Hence, the bound for any  $k \in [N]$  is immediate from Lemmas 6.1 and 6.2(iii).

To prove (7.11) for  $k = 0$  we recall (5.12) to notice that  $\widehat{m}_1 + \mathbf{g}_0 = m_-$  on  $\mathcal{T}^{d,(2)}$ , while on  $\mathcal{T}^{d,(1)}$  we have  $\widehat{m}_1 = m_-$ . Therefore, as  $|\widehat{\eta}_{\widehat{m}_1+1}(\cdot)| \leq 1$  on  $\mathcal{T}^{d,(1)}$ , we find that

$$\left| \prod_{j=1}^{\widehat{m}_1+\mathbf{g}_0} \eta_j(\cdot) \right| \leq \left| \prod_{j=1}^{m_-} \eta_j(\cdot) \right|$$

on  $\mathcal{T}^{d,(s)}$  for any  $s \in \{1, 2\}$ . This, in turn, implies that  $|\widetilde{\det}_k(\cdot)| \leq |\widehat{\det}_k(\cdot)|$  on  $\mathcal{T}^{d,(s)}$  (recall (5.32) and (5.35)). Thus, the bound for  $k = 0$  is immediate from Lemma 6.7(i)-(ii). This completes the proof of the lemma.  $\square$

*Proof of Lemma 7.2.* Set  $\varepsilon_0 = c_{6.16} \widetilde{\varepsilon}_0/2$ . Fix  $\varepsilon'_0 > 0$  so that  $\varepsilon'_0/\varepsilon_0$  is small enough for Lemma 7.5 to hold. Now set  $c_{7.2} = C_{6.16}^{-1} c_{6.16} \cdot (\varepsilon'_0/\varepsilon_0)/2$ . Applying Lemma 6.16(iii) and Markov's inequality we find that it suffices to show that for all  $d \in \mathbb{Z}$

$$\max_{s \in \{1, 2\}} \mathbf{E} \left[ \sup_{z \in \mathcal{T}_{\varepsilon'_0, \varepsilon_0}^{d,(s)} \cap \Omega_{1/N} \setminus \mathcal{B}_2^{\widetilde{\varepsilon}_0}} |\widetilde{\det}_k(z)|^2 \right] = O(N^2). \quad (7.14)$$

The case  $d < 0$  can be dealt by consider the transpose of  $P_N$ . Hence, we will prove (7.14) only for  $d \geq 0$ . Toward this end, we use Lemma 7.5 and ideas from the proof of Lemma 6.6. Fix  $d \geq 0$ ,  $s \in \{1, 2\}$ ,  $\varepsilon''_0 = 1/(4N)$ , and  $\beta \in (0, 1/8)$ . Let  $\mathbb{D}, \varepsilon, b$ , and  $\mathbb{D}_i^\beta$ , for  $i \in [b]$  be as in the proof of Lemma 6.6, where now the centers  $z_i$  of the disks  $\mathbb{D}_i^\beta$  are restricted to be inside  $\widetilde{\mathbb{T}} := \mathcal{T}_{\varepsilon'_0, \varepsilon_0}^{d,(s)} \cap \Omega_{1/N}$ . Notice that  $\cup_{i=1}^b \mathbb{D}_i^\beta \supset \widetilde{\mathbb{T}}$ .

Applying Cauchy's integral formula for smooth functions, and proceeding as in the proof of (6.36) we obtain that

$$\mathbf{E} \left[ \sup_{z \in \mathbb{D}_i^\beta} |\Xi(z)|^2 \right] = O(1) \cdot \sup_{w \in \mathbb{D}_i^{3\beta}} \mathbf{E} [|\Xi(w)|^2], \quad (7.15)$$

for any random holomorphic function  $\Xi : \mathbb{D} \mapsto \mathbb{C}$ . To complete the proof it remains to argue that the map  $z \mapsto \widetilde{\det}_k(z)$  is analytic on  $\mathbb{D}_i^{3\beta}$  for each  $i \in [b]$ .

To this end, observe that  $\mathbb{D}_i^{3\beta} \subset \Omega_{1/(2N)}$ . Hence, the map  $z \mapsto m_-(z)$  is constant on  $\mathbb{D}_i^{3\beta}$ . From the definition of the tubes, and Lemma 6.9(ii) it further follows that

$$|\eta_{m_-}(\cdot)| \geq |\eta_{m_-+1}(\cdot)| + \varepsilon'_0/4 \quad \text{on } \mathbb{D}_i^{3\beta} \subset \widetilde{\mathbb{T}}^{3\beta}. \quad (7.16)$$

Therefore, arguing as in the proof of Lemma 6.6 we deduce that the map  $z \mapsto \widetilde{\det}_k(z)$  is indeed analytic on  $\mathbb{D}_i^{3\beta}$ . Thus, using the bound on  $b$  (see (6.35)), Lemma 7.5, and (7.15) the proof of (7.14) is completed.  $\square$

We now turn to the proof of Lemma 7.3. The key ingredient will be the anti-concentration bound derived in Lemma 6.12. We need to argue that  $\det(P_z^\delta)$  admits certain specific representation so that Lemma 6.12 is applicable. To carry out this step we need the following couple of notation.

Fix  $X_0, Y_0 \subset [N]$  such that  $|X_0| = |Y_0| = k_0$  for some  $k_0 \leq N$ . For  $k \geq k_0$  define

$$\det_k^{X_0, Y_0}(z) := N^{-\gamma k} \sum_{\substack{X, Y \subset [N] \\ |X|=|Y|=k \\ X \supset X_0, Y \supset Y_0}} (-1)^{\text{sgn}(\sigma_X) \text{sgn}(\sigma_Y)} \cdot (-1)^{\widehat{\text{sgn}}(\sigma_X) + \widehat{\text{sgn}}(\sigma_Y)} \mathfrak{D}(X, Y) \cdot \det(Q[X \setminus X_0; Y \setminus Y_0]), \quad (7.17)$$

where  $\widehat{\text{sgn}}(X)$  and  $\widehat{\text{sgn}}(Y)$  are the signs of the permutations on  $X$  and  $Y$  that place all elements of  $X_0$  and  $Y_0$  before those of  $X \setminus X_0$  and  $Y \setminus Y_0$ , respectively, but preserves the order of the elements in each of those individual sets. Also define

$$\widetilde{\det}_k^{X_0, Y_0}(z) := \frac{\det_k(z)}{\exp(N\phi_\infty(z))}.$$

**Lemma 7.6.** *Consider the setup as in Lemma 7.3. Fix  $d \geq 0$  and  $s \in \{1, 2\}$ . Let  $X_0, Y_0 \subset [N]$  with  $|X_0| = |Y_0| = d$ . The following moment bounds hold for  $\det_k^{X_0, Y_0}(z)$ :*

(i) *For any  $k$  such that  $d \leq k \leq N$  we have*

$$\sup_{z \in \mathcal{T}_{\varepsilon'_0, \varepsilon_0}^{d, (s)}} \mathbf{E} \left[ \left| \widetilde{\det}_k^{X_0, Y_0}(z) \right|^2 \right] \leq C_{7.6}^{\widetilde{m}(k+N_+)} N^{-2\gamma k + 2(k-k_0)},$$

where  $C_{7.6} < \infty$  is some constant.

(ii) *Fix  $h, K_0 \in \mathbb{N}$ . Then, for any  $k \in \mathbb{N}$  such that  $d \leq k \leq K_0 - N_+$  we have*

$$\sup_{z \in \mathcal{T}_{\varepsilon'_0, \varepsilon_0}^{d, (s)}} \mathbf{E} \left[ \left| \widetilde{\det}_k^{X_0, Y_0}(z) \right|^{2h} \right] \leq C_{7.6}^{\widetilde{m}K_0 h} \cdot (K_0 h)^{8K_0 h} \cdot \mathfrak{C}_{2K_0 h} \cdot N^{-2\gamma h k + 2h(k-k_0)}. \quad (7.18)$$

*Proof.* The proof of part (i) is similar in nature to that of Lemma 7.5. Hence, details are omitted.

To prove part (ii) we employ the same combinatorial argument as in the proof Lemma 6.4. Indeed, similarly to (6.11), one can obtain an analogous expression for the  $(2h)$ -th moment of the absolute value of  $\widetilde{\det}_k^{X_0, Y_0}(z)$ .

Then, using that the entries of  $Q$  are independent and possess zero mean, one observes that only partitions such that each block has size at least two need to be summed. This forces the number of such partitions to be at most  $N^{2(k-d)h}$ . Now, use the bound (7.13) for  $s = 1$ , while for  $s = 2$  use Lemma 5.13(i), and proceed as in the proof of Lemma 6.4. This yields the desired bound. Further details are omitted.  $\square$

We now use Lemmas 6.12 and 7.6 to derive Lemma 7.3.

*Proof of Lemma 7.3.* Similar to the proof of Lemma 7.2 we notice that it suffices to prove the probability bound in (7.6) for  $z \in \mathcal{T}_{\varepsilon'_0, \varepsilon_0}^{d, (s)}$  for  $d \geq 0$ ,  $s \in \{1, 2\}$ , and some appropriately chosen  $\varepsilon'_0$  and  $\varepsilon_0$ .

Fix  $d \geq 0$  and  $s \in \{1, 2\}$ . Set  $\widetilde{X}_\star = [N] \setminus [N-d]$  and  $\widetilde{Y}_\star = [d]$ . We claim that for any  $k \geq d$

$$\widetilde{\det}_k(z) = \widetilde{\det}_k^{\widetilde{X}_\star, \widetilde{Y}_\star}(z) \cdot \det(Q[\widetilde{X}_\star; \widetilde{Y}_\star]) + \widetilde{\det}_k(z), \quad (7.19)$$

where  $\widetilde{\det}_k(z)$  is a homogeneous polynomial of degree  $k$  in the entries of  $Q$  such that the degree of each individual entry is at most one, and the total degree of the entries of  $Q[\widetilde{X}_\star; \widetilde{Y}_\star]$  is strictly less than  $d$ . To see this claim we recall from (5.9) that

$$\det_k(z) = \sum_{\substack{X, Y \subset [N] \\ |X|=|Y|=k}} (-1)^{\text{sgn}(\sigma_X) \text{sgn}(\sigma_Y)} \mathfrak{D}(X, Y) \cdot N^{-\gamma k} \cdot \det(Q[X; Y]). \quad (7.20)$$

Now we split the sum into two parts depending on whether  $X \supset \widetilde{X}_\star$  and  $Y \supset \widetilde{Y}_\star$ . If either  $X \not\supset \widetilde{X}_\star$  or  $Y \not\supset \widetilde{Y}_\star$  then the corresponding term in the RHS of (7.20) is indeed a polynomial in the entries of  $Q$  such that the total degree of the entries of  $Q[\widetilde{X}_\star; \widetilde{Y}_\star]$  is strictly less than  $d$ . Now fix  $X, Y \subset [N]$  such that  $X \supset \widetilde{X}_\star$  and  $Y \supset \widetilde{Y}_\star$ . Expand  $\det(Q[X; Y])$  by writing it as a sum over permutations  $\pi : X \mapsto Y$ . The sum over permutations  $\pi$  such that  $\pi(\widetilde{X}_\star) = \widetilde{Y}_\star$  is indeed a product of  $\det(Q[X \setminus \widetilde{X}_\star; Y \setminus \widetilde{Y}_\star])$  and  $\det(Q[\widetilde{X}_\star; \widetilde{Y}_\star])$  (upto some signs). The sum over the rest of the permutations is again a polynomial such that the total degree of the entries of  $Q[\widetilde{X}_\star; \widetilde{Y}_\star]$  is strictly less than  $d$ . Combining these observations and upon recalling the definition of  $\widetilde{\det}_k^{\widetilde{X}_\star, \widetilde{Y}_\star}(z)$  we arrive at (7.19).

On the other hand,  $\widetilde{\det}_k(z)$  is a polynomial of total degree less than  $d$ , for any  $k < d$ . Since  $\det(P_z^\delta) = \sum_{k=0}^N \det_k(z)$ , this together with (7.19) imply that

$$\frac{\det(P_z^\delta)}{\exp(N\phi_\infty(z))} = \sum_{\mathcal{I} \subset [d]} Z_{\mathcal{I}} \prod_{i \in \mathcal{I}} U_i,$$

where  $\{U_i\}_{i=1}^d$  are the diagonal entries of the sub matrix  $Q[\widetilde{X}_*; \widetilde{Y}_*]$ ,  $\{Z_{\mathcal{I}}, \mathcal{I} \subset [d]\}$  is a collection of random variables that are independent of  $\{U_i\}_{i=1}^d$ , and

$$Z_{[d]} = \sum_{k \geq d} \widetilde{\det}_k^{\widetilde{X}_*, \widetilde{Y}_*}(z).$$

We now apply Lemma 6.12, and obtain that

$$\mathbf{P} \left( \frac{\det(P_z^\delta)}{\exp(N\phi_\infty(z))} \leq N^{-(4+\gamma d)} \right) \leq N^{-4} + \mathbf{P} \left( |Z_{[d]}| \leq \frac{\bar{c}_*}{2} \cdot N^{-\gamma d} \right), \quad (7.21)$$

for all large  $N$ , and any  $\bar{c}_* > 0$ . It remains to bound the probability that  $Z_{[d]}$  exceeds  $(\bar{c}_*/2) \cdot N^{-\gamma d}$  in absolute value.

Turning to do that we set  $K_0 = \lceil \frac{5}{\gamma-1} \rceil + d + N_+$ . Apply Lemma 7.6(i) and Markov's inequality to deduce that

$$\mathbf{P} \left( \left| \widetilde{\det}_k^{\widetilde{X}_*, \widetilde{Y}_*}(z) \right| \geq N^{-(2+\gamma d)} \right) \leq N^{-5},$$

for any  $k \geq K_0 - N_+$ . Apply Lemma 7.6(ii) with  $h = \lceil \frac{6}{\gamma-1} \rceil$  and Markov's inequality to further derive that

$$\mathbf{P} \left( \left| \widetilde{\det}_k^{\widetilde{X}_*, \widetilde{Y}_*}(z) \right| \geq N^{-\frac{(\gamma-1)}{2} - \gamma d} \right) \leq N^{-5},$$

for any  $k \in [N]$  such that  $d < k < K_0 - N_+$ . Hence, by a union bound we obtain that

$$\mathbf{P} \left( N^{\gamma d} \left| \sum_{k > d} \widetilde{\det}_k^{\widetilde{X}_*, \widetilde{Y}_*}(z) \right| \geq N^{-1} + N^{-\frac{(\gamma-1)}{4}} \right) \leq N^{-4}. \quad (7.22)$$

Equipped with (7.21) and (7.22), and upon recalling the definition  $Z_{[d]}$  we notice that to complete the proof of this lemma it is now enough to show that

$$N^{\gamma d} \cdot \inf_{z \in \mathcal{T}_{\varepsilon_0', \varepsilon_0}^{d, (s)}} \left| \widetilde{\det}_d^{\widetilde{X}_*, \widetilde{Y}_*}(z) \right| = \inf_{z \in \mathcal{T}_{\varepsilon_0', \varepsilon_0}^{d, (s)}} \frac{|\mathfrak{D}(\widetilde{X}_*, \widetilde{Y}_*, z)|}{|a_{-N_-}|^N \prod_{j=1}^{m_-} |\eta_j(z)|^N} \geq \bar{c}_* \quad (7.23)$$

for some  $\bar{c}_* > 0$ . To prove (7.23) we argue as in the proof of Lemma 6.7(b). The only difference is that one needs to use (7.16) instead of (6.43). The rest of the argument is the same. This completes the proof of this lemma.  $\square$

We end this section with the proof of Theorem 1.5. It is immediate from Theorems 5.4, 5.7, and 7.1.

*Proof of Theorem 1.5.* We use  $\widehat{\mathcal{N}}_{\widehat{\Omega}}$  to denote the number of eigenvalues of  $P_{N, \gamma}^Q$  in  $\widehat{\Omega}$ . Fix  $\tilde{\varepsilon}_0 > 0$  such that the Lebesgue measure of the set  $p^{-1}(\mathcal{B}_p^{\tilde{\varepsilon}_0} \cap p(S^1))$  is less than  $\mu/(16\pi)$  (recall Definition 1.3). Then, by [8, Theorem 1.2] (see also [65, Theorem 2.1]) it follows that  $\mathbf{P}(\widehat{\mathcal{N}}_{\widehat{\Omega}_1} \geq \mu N/4) \rightarrow 0$  as  $N \rightarrow \infty$ , where  $\widehat{\Omega}_1 := \mathcal{B}_p^{\tilde{\varepsilon}_0}$ . On the other hand, by Theorems 5.4, 5.7, and Lemma 6.16(iii)-(v) it follows that there exists some  $0 < \widehat{c}_\gamma < \infty$  such that  $\mathbf{P}(\widehat{\mathcal{N}}_{\widehat{\Omega}_2} > 0) \rightarrow 0$  as  $N \rightarrow \infty$ , where  $\widehat{\Omega}_2 := (p(S^1))^{\widehat{c}_\gamma \log N/N} \setminus \widehat{\Omega}_1$ . Finally, upon choosing  $\alpha = 4C_{7.1}/\mu$ , and applying Theorem 7.1 with this  $\alpha$  we deduce that  $\mathbf{P}(\widehat{\mathcal{N}}_{\widehat{\Omega}_3} \geq N\mu/4) \rightarrow 0$  as  $N \rightarrow \infty$ , where  $\widehat{\Omega}_3 := \Omega_{\alpha \log N/N}$  (recall (7.1)).

Setting  $\varepsilon_{1.5} = \tilde{\varepsilon}_0$  and  $C_{1.5} = \alpha \sqrt{\widehat{c}_\gamma^{-1}}$  and taking a union bound over the events that  $\widehat{\mathcal{N}}_{\widehat{\Omega}_i} \geq N\mu/4$  for  $i = 1, 2, 3$ , completes the proof.  $\square$



## 8. RESOLVENT ESTIMATES CLOSE TO THE SPECTRAL CURVE

Given a symbol  $p(\zeta) - z$ , with  $z \in \mathbb{C}$ , as in (4.16), we will see that there is an important distinction between the regime where the spectral parameter  $z$  is inside a loop of  $p(S^1)$ , i.e. where the winding number  $\text{ind}_{p(S^1)}(z)$  of the curve  $p(S^1)$  around  $z$  is non-zero, and the regime where  $\text{ind}_{p(S^1)}(z) = 0$ . In the former case, we will provide in Section 9 a construction of quasimodes, which are approximate singular vectors, for finite Toeplitz matrices  $P_N - z$ . In regions where the winding number is zero, we will be interested in obtaining resolvent estimates. The key difficulty in both regimes will be in obtaining sufficiently good estimates when  $z$  is allowed to be at an  $N$ -dependent distance from  $p(S^1)$ .

In this section will prove the following estimate on the resolvent of a parameter dependent Toeplitz matrix which, while being also of independent interest, will be an essential ingredient in proving a spectral gap for the small singular values, see Section 10.1.

**Theorem 8.1.** *Let  $\Omega' \Subset \mathbb{C}$  be a non-empty relatively compact open set. Let  $N \in \mathbb{N}$ , let  $\Omega_N \Subset \Omega'$  be a family of non-empty relatively compact open sets. Let  $q_z$ ,  $z \in \Omega'$ , be a family of Laurent polynomials as in (4.21), with  $\mathfrak{N}_{z,\pm} = \mathfrak{N}_{\pm} \geq 0$  independent of  $z$  in  $\Omega_N$  and satisfying (4.22), and with coefficients  $q_{z,n} \in \mathbb{C}$  for  $-\mathfrak{N}_- \leq n \leq \mathfrak{N}_+$ . Suppose that:*

- *There exists a constant  $0 < C < \infty$  such that for all  $z \in \Omega'$ ;*

$$q_{z,\mathfrak{N}_+} \neq 0, \quad |q_{z,-\mathfrak{N}_-}| \geq 1/C, \quad |q_{z,n}| \leq C, \quad \text{for } -\mathfrak{N}_- \leq n \leq \mathfrak{N}_+. \quad (8.1)$$

- *All roots  $\zeta_z$  of  $q_z(\zeta)$  are simple, for all  $z \in \Omega'$ .*
- *There exists a constant  $0 < C < \infty$  such that for all  $z \in \Omega'$  and for any two distinct roots  $\zeta_z \neq \omega_z$  we have*

$$|\zeta_z - \omega_z| \geq 1/C, \quad (8.2)$$

and for any root  $\zeta_z$

$$0 < |\zeta_z| \leq C. \quad (8.3)$$

- *$0 \notin q_z(S^1)$  and there exists a constant  $C_0 > 0$  such that for  $N > 0$  large enough*

$$\text{dist}(|\zeta_z|, 1) \geq C_0 \frac{\log N}{N}, \quad \text{for all } z \in \Omega_N, \text{ and all roots } \zeta_z. \quad (8.4)$$

- *There exists an  $m_0 \geq 0$  (unrelated to the constants  $m_0$  in (4.27) and (4.28)) and constants  $C_1 > 1, C_2 > 0$  such that for all  $N \in \mathbb{N}$  large enough and all  $z \in \Omega_N$  we have that the roots of  $q_z(\zeta)$  inside  $D(0, 1)$ , of total number  $m_{z,+} \geq m_0$ , satisfy*

$$|\zeta_1^+| \leq \dots \leq |\zeta_{m_{z,+}-m_0}^+| < 1/C_1 < 1 - C_2 \frac{\log N}{N} \leq |\zeta_{m_{z,+}-m_0+1}^+| \leq \dots \leq |\zeta_{m_{z,+}}^+|. \quad (8.5)$$

- *Let  $\hat{m}_0 \in \{0, m_0\}$  and suppose that for all  $N \in \mathbb{N}$  and all  $z \in \Omega_N$  we have  $\text{ind}_{q_z(S^1)}(0) = \hat{m}_0$ .*

Then there exists a constant  $0 < \tilde{C} < +\infty$  such that for  $N > 0$  large enough and all  $z \in \Omega_N$ ,

$$\|P_N(q_z)^{-1}\| \leq \tilde{C} \frac{N^{1+C_2\Theta(\hat{m}_0)}}{\log N}. \quad (8.6)$$

**Remark 8.2.** *It will be clear from the proof that (8.6) holds when  $\text{ind}_{q_z(S^1)}(0) = 0$  even without the structural assumption (8.5). Furthermore, notice that in Theorem 8.1 we only consider the case of non-negative winding numbers. This is all that we need in the sequel, however one could prove the same result in the case of negative winding numbers provided that one assume a similar assumption as (8.5) for the roots in  $\mathbb{C} \setminus D(0, 1)$ .*

*Proof of Theorem 8.1.* Throughout the proof, we will assume that  $N > 0$  is sufficiently large so that the assumptions of the theorem hold. Moreover, all  $O(1)$  terms and constants are understood to be uniform in  $N > 0$  and  $z \in \Omega_N$ , without us mentioning it explicitly at each occurrence. In fact the error terms will only depend on the constants in the hypothesis of the theorem and global constants.



1. Let  $z \in \Omega_N$ . Since  $q_z$  satisfies the assumptions of Case 1 of Lemma 4.1, we may order the roots of  $q_z(\zeta)$  as in (4.26), and using also (8.3) and (8.4),

$$0 < |\zeta_1^+| \leq \dots \leq |\zeta_{m_+,+}^+| < 1 < |\zeta_1^-| \leq \dots \leq |\zeta_{m_-, -}^-| \leq C. \quad (8.7)$$

Since  $\text{ind}_{q_z(S^1)}(0) = \widehat{m}_0$  for all  $z \in \Omega_N$ , it follows from a similar computation as in (4.34) that  $m_{z,\pm} = \mathfrak{N}_\pm \pm \widehat{m}_0$  are independent of  $N$  and  $z$ . Keeping in mind that  $m_{z,\pm}$  are independent of  $z$ , but that the coefficients  $q_{z,n}$  do depend on  $z$ , we suppress from now on the  $z$  subscript of  $q_z$ ,  $q_{z,n}$  and  $m_{z,\pm}$ . Hence, we are interested in the symbol

$$q(\zeta) = \sum_{-(m_- + \widehat{m}_0)}^{m_+ - \widehat{m}_0} q_j \zeta^{-j}, \quad (8.8)$$

where by (8.1)

$$q_{m_+ - \widehat{m}_0} \neq 0, \quad |q_{-(m_- + \widehat{m}_0)}| \geq 1/C \quad (8.9)$$

2. We turn to estimating  $\|P_N(q)^{-1}\|$ . We begin by inverting  $\text{Op}(q) : \ell^2(\mathbb{Z}) \rightarrow \ell^2(\mathbb{Z})$ , see (4.17), which is a convolution operator. We first search for a fundamental solution  $E : \mathbb{Z} \rightarrow \mathbb{C}$  to

$$\text{Op}(q)E = \delta_0 \quad \text{on } \mathbb{Z}, \quad (8.10)$$

where  $\delta_0(\nu) = \delta_{0,\nu}$ , see Section 1.6 for the Dirac notation. Putting

$$E(n) = \mathcal{F}^{-1}\left(\frac{1}{q}\right)(n) = \frac{1}{2\pi} \int_0^{2\pi} \frac{1}{q(e^{i\xi})} e^{in\xi} d\xi. \quad (8.11)$$

it follows from (4.5) that  $E$  solves (8.10). Thus,

$$\begin{aligned} \text{Op}(q)E * v &= v, & v &\in \ell_{\text{comp}}^2(\mathbb{Z}) \\ E * \text{Op}(q)u &= u, & u &\in \ell_{\text{comp}}^2(\mathbb{Z}), \end{aligned} \quad (8.12)$$

where  $*$  denotes the convolution on  $\mathbb{Z}$  and  $\ell_{\text{comp}}^2$  denotes compactly supported functions in  $\ell^2$ .

Notice that we may factor (8.8) as

$$q(\zeta) = q_{-(m_- + \widehat{m}_0)} \prod_{j=1}^{m_+ - \widehat{m}_0} (1 - \zeta_j^+/\zeta) \prod_{j=m_+ - \widehat{m}_0 + 1}^{m_+} (\zeta - \zeta_j^+) \prod_{k=1}^{m_-} (\zeta - \zeta_k^-). \quad (8.13)$$

Performing the change of variables  $e^{i\xi} = \zeta \in S^1$  in (8.11), we get by shrinking  $S^1$  to 0, the residue theorem and (8.13) that for  $n \geq 1$

$$E(n) = \frac{1}{2\pi i} \int_{S^1} \frac{\zeta^{n-1}}{q(\zeta)} d\zeta = \sum_{r=1}^{m_+} \frac{(\zeta_r^+)^{n-1+m_+ - \widehat{m}_0}}{q_{-(m_- + \widehat{m}_0)} \prod_{j \neq r} (\zeta_r^+ - \zeta_j^+) \prod_{k=1}^{m_-} (\zeta_r^+ - \zeta_k^-)}, \quad (8.14)$$

Similarly, we get that

$$E(0) = \sum_{r=1}^{m_+} \frac{(\zeta_r^+)^{-1+m_+ - \widehat{m}_0}}{q_{-(m_- + \widehat{m}_0)} \prod_{j \neq r} (\zeta_r^+ - \zeta_j^+) \prod_{k=1}^{m_-} (\zeta_r^+ - \zeta_k^-)} + \frac{\delta_{m_+, \widehat{m}_0}}{q_{-(m_- + \widehat{m}_0)} \prod_{j=1}^{m_+} (-\zeta_j^+) \prod_{k=1}^{m_-} (-\zeta_k^-)}. \quad (8.15)$$

By deforming  $S^1$  to  $|\zeta| = R$ ,  $R \rightarrow \infty$ , we obtain similarly that for  $n \leq -1$ ,

$$E(n) = - \sum_{r=1}^{m_-} \frac{(\zeta_r^-)^{n-1+m_+ - \widehat{m}_0}}{q_{-(m_- + \widehat{m}_0)} \prod_{j=1}^{m_+} (\zeta_r^- - \zeta_j^+) \prod_{j \neq r} (\zeta_r^- - \zeta_j^-)}. \quad (8.16)$$

Combining (8.14)-(8.16), (8.2), (8.7), (8.1) and (8.9), we have that

$$|E(n)| = O(1) \begin{cases} |\zeta_{m_+}^+|^{n-1}, & n \geq 1, \\ 1, & n = 0, \\ |\zeta_1^-|^{n-1}, & n \leq -1. \end{cases} \quad (8.17)$$

Here, we replace  $(\zeta_1^-)^{-1}$  resp.  $\zeta_{m_+}$  with 0 when  $m_- = 0$  resp.  $m_+ = 0$ . Thus, summing the geometric series and using (8.4) we get that

$$\|E\|_{\ell^1} = O(1) \frac{N}{\log N}, \quad (8.18)$$

and similarly that

$$\|E\|_{\ell^2} = O(1) \sqrt{\frac{N}{\log N}}. \quad (8.19)$$

Since  $E$  is in  $\ell^1$ , see (8.18), the operator of convolution with  $E$ , denoted  $E*$ , is a bounded operator  $\ell^2 \rightarrow \ell^2$ . Therefore, (8.12) can be extended to  $u, v \in \ell^2(\mathbb{Z})$ , and  $E*$  is the inverse of  $\text{Op}(q)$  on  $\ell^2(\mathbb{Z})$ .

3. Next, for  $v \in \ell^2(\mathbb{Z})$  with support in  $[0, N-1]$ , we solve

$$\text{Op}(q)\psi = 0 \quad \text{on } \mathbb{Z} \quad (8.20)$$

with ‘‘boundary’’ conditions

$$\psi \upharpoonright_{[-(m_+ - \widehat{m}_0), -1] \cup [N, N + m_- + \widehat{m}_0 - 1]} = (E * v) \upharpoonright_{[-(m_+ - \widehat{m}_0), -1] \cup [N, N + m_- + \widehat{m}_0 - 1]}. \quad (8.21)$$

We suppose here, and in the sequel, that  $m_{\pm} \mp \widehat{m}_0 > 0$ . The cases when  $m_+ = \widehat{m}_0$  or  $m_- = \widehat{m}_0 = 0$  can be treated similarly, and we shall comment on these cases as we go along. In (8.21) it is understood that when  $m_+ = \widehat{m}_0$ , then we only have the boundary condition on  $[N, N + m_- + \widehat{m}_0 - 1]$ , and similarly when  $m_- = \widehat{m}_0 = 0$ .

Since  $q(\zeta)$  satisfies the assumptions of Case 1 of Lemma 4.1, it follows from Propositions 4.4 and 4.5, and the fact that all roots of  $q(\zeta)$  are simple, that the general solution to (8.20) is of the form

$$\psi(\nu) = \sum_{j=1}^{m_+} a_j^+ (\zeta_j^+)^{\nu} + \sum_{j=1}^{m_-} a_j^- (\zeta_j^-)^{\nu}, \quad a_j^{\pm} \in \mathbb{C}. \quad (8.22)$$

By (8.21), the coefficients are determined by

$$\begin{pmatrix} V_+ \Lambda_+^{\widehat{m}_0 - m_+} & B_+ \Lambda_-^{\widehat{m}_0 - m_+} \\ V_- \Lambda_+^N & B_- \Lambda_-^N \end{pmatrix} a =: Ma = c := (E * v) \upharpoonright_{[\widehat{m}_0 - m_+, -1] \cup [N, N + m_- + \widehat{m}_0 - 1]}. \quad (8.23)$$

Here, the right hand side is seen as a vector in  $\mathbb{C}^{m_+ + m_-}$ ,  $a = (a_+, a_-)^{\top}$ ,  $a_{\pm} = (a_1^{\pm}, \dots, a_{m_{\pm}}^{\pm}) \in \mathbb{C}^{m_{\pm}}$ ,  $\Lambda_+ := \text{diag}(\zeta_1^+, \dots, \zeta_{m_+ - \widehat{m}_0}^+)$ ,  $\Lambda_- := \text{diag}(\zeta_{m_+ - \widehat{m}_0 + 1}^+, \dots, \zeta_{m_+}^+, \zeta_1^-, \dots, \zeta_{m_-}^-)$ ,

$$V_{\pm} := \begin{pmatrix} 1 & \dots & 1 \\ \zeta_1^+ & \dots & \zeta_{m_+ - \widehat{m}_0}^+ \\ \vdots & \vdots & \vdots \\ (\zeta_1^+)^{m_{\pm} \mp \widehat{m}_0 - 1} & \dots & (\zeta_{m_+ - \widehat{m}_0}^+)^{m_{\pm} \mp \widehat{m}_0 - 1} \end{pmatrix}, \quad (8.24)$$

and

$$B_{\pm} := \begin{pmatrix} 1 & \dots & 1 & 1 & \dots & 1 \\ \zeta_{m_+ - \widehat{m}_0 + 1}^+ & \dots & \zeta_{m_+}^+ & \zeta_1^- & \dots & \zeta_{m_-}^- \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ ((\zeta_{m_+ - \widehat{m}_0 + 1}^+)^{m_{\pm} \mp \widehat{m}_0 - 1} & \dots & (\zeta_{m_+}^+)^{m_{\pm} \mp \widehat{m}_0 - 1} & (\zeta_1^-)^{m_{\pm} \mp \widehat{m}_0 - 1} & \dots & (\zeta_{m_-}^-)^{m_{\pm} \mp \widehat{m}_0 - 1} \end{pmatrix}. \quad (8.25)$$

Here, we use the convention that when  $\widehat{m}_0 = 0$  and  $m_+ > 0$ , then the first column of  $B_{\pm}$  is the one containing powers of  $\zeta_1^-$ .

**Remark 8.3.** When  $m_+ = \widehat{m}_0$  then (8.23) is given by  $B_- \Lambda_-^N a = Ma = (E * v) \upharpoonright_{[N, N + m_- - 1]}$ , and when  $m_- = \widehat{m}_0 = 0$  then only the first sum in (8.22) remains and (8.23) is given by  $V_+ \Lambda_+^{-m_+} a_+ = Ma = (E * v) \upharpoonright_{[-m_+, -1]}$ .

By the Cauchy-Schwarz inequality and (8.19), we have that for any  $k \in \mathbb{Z}$

$$|E * v(k)| = \left| \sum_0^{N-1} E(k-m)v(m) \right| \leq \|E\|_{\ell^2(\mathbb{Z})} \|v\|_{\ell^2(\mathbb{Z})} \leq O(1) \sqrt{\frac{N}{\log N}} \|v\|_{\ell^2(\mathbb{Z})}. \quad (8.26)$$

We put  $L := L_+ \oplus L_-$  with

$$\begin{aligned} L_+ &= \text{diag}(G(\zeta_1^+), \dots, G(\zeta_{m_+ - \widehat{m}_0}^+)), \\ L_- &= \text{diag}(G(\zeta_{m_+ - \widehat{m}_0 + 1}^+), \dots, G(\zeta_{m_+}^+), G(\zeta_1^-), \dots, G(\zeta_{m_-}^-)), \quad G(x) = \left( \frac{1 - |x|^{2N}}{1 - |x|^2} \right)^{1/2}, \end{aligned} \quad (8.27)$$

where we work with the same notational convention for  $L_-$  as in (8.25).

Suppose that there exists a constant  $0 < C < \infty$  such that for  $N > 0$  sufficiently large  $M$  is bijective and such that for all  $z \in \Omega_N$

$$\|LM^{-1}\| \leq C \sqrt{\frac{N}{\log N}} N^{C_2 \Theta(\widehat{m}_0)}, \quad (8.28)$$

with  $C_2$  as in (8.5). We will prove this fact in Step 4 below. It then follows from the bijectivity of  $M$  that (8.20) with boundary conditions (8.21) has a unique solutions  $\psi$  of the form (8.22) with coefficients determined by (8.23). Hence, putting

$$\tilde{u} = (E * v - \psi) \upharpoonright_{[0, N-1]},$$

we have that  $P_N(q)\tilde{u} = v$ , so  $P_N(q)$  is surjective, and thus, being a square matrix, bijective. We get that for  $N > 0$  sufficiently large,

$$\begin{aligned} \|\psi\|_{\ell^2([0, N-1])} &\stackrel{(8.22)}{\leq} \sum_{j=1}^{m_+} |a_j^+| G(\zeta_j^+) + \sum_{j=1}^{m_-} |a_j^-| G(\zeta_j^-) \\ &\stackrel{\text{H\"older}}{\leq} \sqrt{m_+ + m_-} \|La\|_{\ell^2([1, m_+ + m_-])} \\ &\stackrel{(8.23)}{\leq} \sqrt{m_+ + m_-} \|LM^{-1}c\|_{\ell^2([1, m_+ + m_-])} \\ &\stackrel{(8.28)}{\leq} O(1) \sqrt{\frac{N}{\log N}} N^{C_2 \Theta(\widehat{m}_0)} \|c\|_{\ell^2([1, m_+ + m_-])} \stackrel{(8.26)}{\leq} O(1) \frac{N^{1+C_2 \Theta(\widehat{m}_0)}}{\log N} \|v\|_{\ell^2}. \end{aligned} \quad (8.29)$$

Using Young's convolution inequality and (8.18), we find that for  $N > 0$  sufficiently large

$$\begin{aligned} \|\tilde{u}\|_{\ell^2([0, N-1])} &\leq \|E * v\|_{\ell^2([0, N-1])} + \|\psi\|_{\ell^2([0, N-1])} \\ &\leq \|E\|_{\ell^1} \|v\|_{\ell^2} + \|\psi\|_{\ell^2([0, N-1])} \stackrel{(8.18), (8.29)}{\leq} O(1) \frac{N^{1+C_2 \Theta(\widehat{m}_0)}}{\log N} \|v\|_{\ell^2}, \end{aligned}$$

uniformly for  $z \in \Omega_N$ . This proves (8.6), provided that (8.28) holds.

4. It remains to prove (8.28). By (8.7), we have that

$$\|V_{\pm}\|, \|B_{\pm}\| = O(1). \quad (8.30)$$

Since  $V_+, B_-$  are Vandermonde matrices, we have that  $\det V_+ = \prod_{1 \leq i < j \leq m_+ - \widehat{m}_0} (\zeta_i^+ - \zeta_j^+)$  and similar expression for  $\det B_-$  given by a finite product of differences of the roots  $\zeta_{m_+ - \widehat{m}_0 + 1}^+, \dots, \zeta_{m_+}^+$  and  $\zeta_1^-, \dots, \zeta_{m_-}^-$  (resp. just  $\zeta_1^-, \dots, \zeta_{m_-}^-$  when  $\widehat{m}_0 = 0$ ). Therefore, (8.2) implies that both matrices are bijective and that their determinants are uniformly bounded from below in modulus.

Together with (8.7), we deduce that

$$\|(V_+)^{-1}\|, \|(B_-)^{-1}\| = O(1). \quad (8.31)$$

To construct the inverse of  $M$ , we proceed via the Schur complement formula and set

$$\Gamma = B_- \Lambda_-^N - V_- \Lambda_+^{N+m_+ - \widehat{m}_0} V_+^{-1} B_+ \Lambda_-^{\widehat{m}_0 - m_+}.$$

Using (8.31), (8.30), (8.4), (8.7) and (8.5), we have that

$$\begin{aligned} & \|V_- \Lambda_+^{N+m_+-\widehat{m}_0} V_+^{-1} B_+ \Lambda_-^{\widehat{m}_0-m_+-N} B_-^{-1}\| \\ &= O(1) \begin{cases} |\zeta_{m_+}|^{N+m_+} |\zeta_1^-|^{-(N+m_+)}, & \widehat{m}_0 = 0, \\ |\zeta_{m_+-m_0}^+|^{N+m_+-m_0} |\zeta_{m_+-m_0+1}^+|^{m_0-(N+m_+)}, & \widehat{m}_0 = m_0, \end{cases} \quad (8.32) \\ &= O(1) \begin{cases} N^{-2C_0}, & \widehat{m}_0 = 0, \\ e^{-N \log C_1} N^{C_2}, & \widehat{m}_0 = m_0 \end{cases} =: \varepsilon_N(\widehat{m}_0). \end{aligned}$$

Hence, for  $N > 0$  large enough (depending only on the constants in the assumptions of the theorem) we have that  $|\varepsilon_N(\widehat{m}_0)| < 1$ , and we assume from now on that this is the case. By a Neumann series argument we see that

$$\Gamma^{-1} = \Lambda_-^{-N} B_-^{-1} (1 + O(1)\varepsilon_N(\widehat{m}_0)). \quad (8.33)$$

Here the  $O(1)$  term means an  $m_- \times m_-$  matrix with operator norm bounded by a constant uniformly in  $N$  and  $z \in \Omega_N$ .

A straightforward computation yields that

$$M^{-1} = \begin{pmatrix} \Lambda_+^{m_+-\widehat{m}_0} V_+^{-1} & -\Lambda_+^{m_+-\widehat{m}_0} V_+^{-1} B_+ \Lambda_-^{\widehat{m}_0-m_+} \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ -\Gamma^{-1} V_- \Lambda_+^{N+m_+-\widehat{m}_0} V_+^{-1} & \Gamma^{-1} \end{pmatrix}. \quad (8.34)$$

Using (8.27) and (8.34), we obtain that

$$\begin{aligned} LM^{-1} &= \quad (8.35) \\ & \begin{pmatrix} L_+ \Lambda_+^{m_+-\widehat{m}_0} V_+^{-1} (1 + B_+ \Lambda_-^{\widehat{m}_0-m_+} \Gamma^{-1} V_- \Lambda_+^{N+m_+-\widehat{m}_0} V_+^{-1}), & -L_+ \Lambda_+^{m_+-\widehat{m}_0} V_+^{-1} B_+ \Lambda_-^{\widehat{m}_0-m_+} \Gamma^{-1} \\ -L_- \Gamma^{-1} V_- \Lambda_+^{N+m_+-\widehat{m}_0} V_+^{-1}, & L_- \Gamma^{-1} \end{pmatrix}. \end{aligned}$$

To prove (8.28) note that it is sufficient to estimate the norm of each block of the matrix (8.35) separately. Since  $|\zeta_{m_+}^+| \leq 1 \leq |\zeta_1^-|$  and (8.5) holds, it follows that

$$\|\Lambda_-^{-n}\|, \|\Lambda_+^n\| \leq O(1), \quad (8.36)$$

for any fixed  $n \in \mathbb{N}$ . Using that  $|\zeta_{m_+}^+| \leq 1$  we find also that

$$\|\Lambda_+^N\| \leq O(1). \quad (8.37)$$

However, in view of (8.5) and since  $1 \leq |\zeta_1^-|$  we have that

$$\|\Lambda_-^{-N}\| \leq O(1) \begin{cases} |\zeta_1^-|^{-N}, & \widehat{m}_0 = 0, \\ |\zeta_{m_+-m_0+1}^+|^{-N}, & \widehat{m}_0 = m_0 \end{cases} \leq O(1) N^{C_2 \Theta(\widehat{m}_0)} \quad (8.38)$$

Continuing, we deduce from (8.31), (8.33) and (8.38) that

$$\|\Gamma^{-1}\| \leq O(1) \|\Lambda_-^{-N} B_-^{-1}\| \leq O(1) N^{C_2 \Theta(\widehat{m}_0)} \quad (8.39)$$

Recall (8.27). Then, by (8.4),

$$\|L_+\| \leq \max_{1 \leq j \leq m_+-\widehat{m}_0} G(\zeta_j^+) = O(1) \sqrt{\frac{N}{\log N}}.$$

Combining this with (8.31) and (8.36), we see that

$$\|L_+ \Lambda_+^{m_+} V_+^{-1}\| = O(1) \sqrt{\frac{N}{\log N}}. \quad (8.40)$$

Hence, using (8.40), (8.31), (8.36), (8.30), (8.39) and (8.37), we get that

$$\|L_+ \Lambda_+^{m_+} V_+^{-1} V_- \Lambda_-^{-m_+} \Gamma^{-1} B_+ \Lambda_+^{N+m_+} V_+^{-1}\| = O(1) \sqrt{\frac{N}{\log N}} N^{C_2 \Theta(\widehat{m}_0)} \quad (8.41)$$

and

$$\|L_+ \Lambda_+^{m_+} V_+^{-1} V_- \Lambda_-^{-m_+} \Gamma^{-1}\| = O(1) \sqrt{\frac{N}{\log N}} N^{C_2 \Theta(\widehat{m}_0)}. \quad (8.42)$$

Using (8.4) and (8.5), we find that

$$\|L_- \Lambda_-^{-N}\| = \max\left\{\max_j |\zeta_j^-|^{-N} G(\zeta_j^-), \max_{m_+ - \widehat{m}_0 + 1 \leq j \leq m_+} |\zeta_j^+|^{-N} G(\zeta_j^+)\right\} = O(1) \sqrt{\frac{N}{\log N}} N^{C_2 \Theta(\widehat{m}_0)}. \quad (8.43)$$

Hence, similarly to (8.41), but using as well (8.43) and (8.33), we get that

$$\|L_- \Gamma^{-1} B_+ \Lambda_+^{N+m_+} V_+^{-1}\| = O(1) \|L_- \Lambda_-^{-N}\| = O(1) \sqrt{\frac{N}{\log N}} N^{C_2 \Theta(\widehat{m}_0)}, \quad (8.44)$$

and

$$\|L_- \Gamma^{-1}\| = O(1) \sqrt{\frac{N}{\log N}} N^{C_2 \Theta(\widehat{m}_0)}. \quad (8.45)$$

Hence, (8.35), (8.40)-(8.42), (8.44), and (8.45) yield (8.28) for  $N > 0$  large enough.

**Remark 8.4.** Recall Remark 8.3. When  $m_+ = \widehat{m}_0$  we have that  $M^{-1} = \Lambda_-^{-N} B_-^{-1}$  and we obtain (8.28) by a similar estimate as in (8.44). When  $m_- = \widehat{m}_0 = 0$  then  $M^{-1} = \Lambda_+^{m_+} V_+^{-1}$  and we obtain (8.28) by (8.40).

This completes the proof of Theorem 8.1.  $\square$

## 9. QUASIMODES FOR BANDED TOEPLITZ MATRICES

In this section we will construct exponentially decaying quasimodes for the operator  $P_N - z$ , for  $z$  close to - however not on - the curve  $p(S^1)$  while avoiding a small neighborhood of the set of bad points, see Definition 1.3. We will see that the sign of the winding number  $d(z)$  will decide whether these quasimodes decay to the left ( $d < 0$ ) or to the right ( $d > 0$ ). The construction will be based on exponential states  $\mathfrak{z}^\pm$ , see Proposition 9.3, determined by the roots of the Laurent polynomial  $p(\zeta) - z$ , with a particular focus on the decay rates of these exponential states. We will separate between those states  $\mathfrak{z}^\pm$  which decay at a fixed rate  $r > 0$  and those which decay at an  $N$ -dependent rate  $r \asymp \log N/N$ , where the former are determined by the roots of  $p(\zeta) - z$  “far” away from  $p(S^1)$  and the latter are determined by the roots “close” to  $p(S^1)$ .

Throughout this section we fix  $z_0 \in p(S^1)$  such that

$$dp \neq 0 \text{ on } p^{-1}(z_0). \quad (9.1)$$

Let  $\Omega \Subset \mathbb{C}$  be an open relatively compact simply connected neighbourhood of  $z_0$  such that

$$|dp| \geq 1/C \text{ on } p^{-1}(z) \text{ for all } z \in \Omega. \quad (9.2)$$

Note that by the implicit function theorem all roots  $\zeta = \zeta(z)$  of  $p(\zeta) - z$  are simple and depend smoothly on  $z \in \Omega$ . Hence, after possibly shrinking  $\Omega$ , there exists  $0 < C < \infty$  such that for any two distinct roots  $\zeta(z) \neq \omega(z)$  we have that

$$|\zeta - \omega| \geq \frac{1}{C}, \quad (9.3)$$

uniformly for  $z \in \Omega$ . Recall the notation of the roots from Lemma 4.1.

Until further notice we will denote the winding number of the curve  $p(S^1)$  around a point  $z \in \mathbb{C}$ , see also (4.34), by

$$d \stackrel{\text{def}}{=} d(z) \stackrel{\text{def}}{=} \text{ind}_{p(S^1)}(z). \quad (9.4)$$

**Assumption 9.1.** Let  $z_0 \in p(S^1)$  be as in (9.1), and let  $\Omega$  be as in (9.2), so that (9.3) holds. Let  $\Omega_N \Subset \Omega \setminus p(S^1)$  be an open relatively compact simply connected non-empty  $N$ -dependent set such that the winding number  $d(z) \neq 0$  is constant for all  $z \in \Omega_N$ , and such that there exist constants  $C_\alpha, C_\beta, C_\gamma > 0$  and  $m_\pm^0 \geq 0$ , independent of  $N$ , such that for  $N > 0$  large enough (depending only on these constants) and for all  $z \in \Omega_N$  the following holds:

- we have that

$$p^{-1}(z) \cap (S^1 + D(0, C_\gamma N^{-1} \log N)) = \emptyset; \quad (9.5)$$

- when  $d > 0$ , then  $0 \leq m_+^0 \leq d \leq m_+$  and

$$0 < |\zeta_1^+| \leq \cdots \leq |\zeta_{m_+ - m_+^0}^+| < \frac{1}{C_\beta} < 1 - C_\alpha \frac{\log N}{N} \leq |\zeta_{m_+ - m_+^0 + 1}^+| \leq \cdots \leq |\zeta_{m_+}^+| < 1; \quad (9.6)$$

- when  $d < 0$ , then  $0 \leq m_-^0 \leq -d \leq m_-$  and

$$1 < |\zeta_1^-| \leq \cdots \leq |\zeta_{m_-^0}^-| < 1 + C_\alpha \frac{\log N}{N} < C_\beta \leq |\zeta_{m_-^0 + 1}^-| \leq \cdots \leq |\zeta_{m_-}^-| < +\infty. \quad (9.7)$$

In (9.6) we work with the notation convention that when  $m_+ = m_+^0$  then the inequalities before  $1/C_\beta$  are void, and when  $m_+^0 = 0$ , then the inequalities on the right hand side of  $1 - C_\alpha \log N/N$  are void. A similar convention is to be understood for (9.7).

**Remark 9.2.** 1. Unless otherwise stated, all future estimates will only depend on  $\Omega$  and the constants in Assumption 9.1, and therefore will be uniform in  $\Omega_N$ . This will be understood implicitly when we state that an estimate is uniform in  $\Omega_N$ .

2. We recall that we will work with  $z \in \Omega(\varepsilon_{1.5}, C_{1.5}, N)$ , see (1.7) and the assumptions of Theorem 1.6. In particular,  $\text{dist}(z, p(S^1)) < C_{1.5} \log N/N$ . Let  $z_0$  be the point on  $p(S^1)$  closest to  $z$ . Then, since  $dp(z_0) \neq 0$  (as we work in  $\mathcal{G}_{p,\varepsilon}$ , as well as the existence of  $C_\alpha$  as in (9.6) or (9.7), e.g. that for  $d > 0$ ,  $\zeta_{m_+}^+ \geq 1 - C_\alpha \log N/N$ ). To see the multiplicity count, i.e. that e.g.  $m_+^0 \leq d$  when  $d > 0$ , note that away from the set  $\mathcal{B}_2$ , if  $\mathbf{g}(p) = 1$  then there is precisely one  $\theta_0$  with  $p(e^{i\theta_0}) = z_0$ , and since  $dp|_{\theta_0} \neq 0$ , there is exactly one root  $\zeta_{m_+}$  near 1. When  $\mathbf{g}(p) > 1$ , there are  $\mathbf{g}(p)$  such roots, while  $|d| \geq \mathbf{g}(p)$  if  $d \neq 0$ , so again the same conclusion holds. On the other hand, by Remark 6.17 we have that for any eigenvalue  $z$  in the good set, with probability approaching one, the assumption (9.5) holds with  $C_\gamma$  as defined in Remark 6.17.

Our aim is to construct quasimodes for  $P_N - z$ ,  $z \in \Omega_N$ , these are approximate  $\ell^2$ -normalized eigenvectors  $\psi \in \mathbb{C}^N$  in the sense that

$$\|(P_N - z)\psi\| \rightarrow 0$$

when  $N \rightarrow \infty$ . Our focus will be on obtaining quantitative estimates for this decay. The eigenvectors of  $P_{[0,\infty[}$  corresponding to the eigenvalue  $z$ , when  $d > 0$ , and shifted versions of the eigenvectors of  $P_{] -\infty, 0]}$  corresponding to the eigenvalue  $z$ , when  $d < 0$ , are excellent candidates for quasimodes (after truncation).

We will construct these quasimodes as follows: Recall from Proposition 4.6 that these eigenvectors are exponential states which are created from powers of the roots of  $p(\zeta) - z$ . Using assumption (9.6) and (9.7) we will take care (whenever possible) to distinguish between eigenvectors which are built up only by roots strictly away from the unit circle, and those eigenvectors which contain also a contribution from roots close to the unit circle. After truncation we orthonormalize the states within the two respective groups. As we shall see below, we will need to slightly modify the behaviour at the boundary of the second group to obtain a sufficiently fast decay. This way we will obtain an almost orthogonal  $|d|$ -dimensional system of quasimodes, see Proposition 9.6 below.

Recall Proposition 4.6. To construct quasimodes out of (4.46) respectively (4.47), we begin with constructing a suitable system of solutions to (4.44) when  $N_+ > 0$  and its analog when  $N_- > 0$  with  $m_+, N_+, \zeta_j^+$  replaced by  $m_-, N_-, \zeta_j^-$  and the corresponding vectors  $\mathfrak{A}_j^-$  indexed over  $1 \leq \nu \leq N_-$ .

We begin with considering the case  $N_+ > 0$ . Notice that this implies that  $0 \notin p^{-1}(\Omega)$ . As discussed after (9.2), the roots  $\zeta_z$  of  $p(\zeta) - z$  depend smoothly on  $z \in \Omega$ , so, after potentially slightly shrinking  $\Omega$ , there exists a constant  $0 < C < \infty$  such that any root  $\zeta_z$  satisfies

$$1/C \leq |\zeta_z|, \text{ for all } z \in \Omega. \quad (9.8)$$

Let  $z \in \Omega_N$ , suppose that  $d > 0$  and recall from (4.34) that  $d = m_+ - N_+$ . In view of (9.6), we find that  $m_+ - m_+^0 \geq N_+ > 0$ . For the sake of the presentation we first suppose that  $d > m_+^0 \geq 1$ . Recall the matrix  $\mathfrak{A}$  from (4.44) and write

$$\mathfrak{B}_+ = (\mathfrak{A}_1^+, \dots, \mathfrak{A}_{N_+}^+) \in \mathbb{C}^{N_+ \times N_+}, \quad \mathfrak{C}_+ = (\mathfrak{A}_1^+, \dots, \mathfrak{A}_{m_+ - m_+^0}^+) \in \mathbb{C}^{N_+ \times (m_+ - m_+^0)}$$

By (9.8) we have that for all  $j = 1, \dots, m_+$

$$\|\mathfrak{A}_j^+\| \leq C^{N_+} \sqrt{N_+}. \quad (9.9)$$

Notice that  $\mathfrak{B}_+$  is a Vandermonde type matrix, so  $|\det \mathfrak{B}_+| = \prod_{k=1}^{N_+} |\zeta_k^+|^{-N_+} \prod_{1 \leq i < j \leq N_+} |\zeta_i^+ - \zeta_j^+|$ . Thus, by (9.8), (9.3), we see that  $\mathfrak{B}_+$  is invertible and, using (9.9), we find

$$\|\mathfrak{B}_+\|, \|\mathfrak{B}_+^{-1}\| = O(1), \quad (9.10)$$

where the constant depends only on  $\Omega$ . Hence, by (9.6), we see that  $\mathfrak{C}_+$  has rank  $N_+$  and has a kernel of dimension  $\dim \mathcal{N}(\mathfrak{C}_+) = m_+ - m_+^0 - N_+ = d - m_+^0 > 0$ . Let  $x_1, \dots, x_{d-m_+^0} \in \mathbb{C}^{m_+ - m_+^0}$  be an orthonormal system spanning  $\mathcal{N}(\mathfrak{C}_+)$ . Clearly,

$$\mathbb{C}^{m_+} \ni a_j^+ := x_j \oplus 0_{m_+^0} \in \mathcal{N}(\mathfrak{A}) \text{ for } j = 1, \dots, d - m_+^0. \quad (9.11)$$

For  $j = d - m_+^0 + 1, \dots, d$ , we put

$$a_j^+ := (-\mathfrak{B}_+^{-1} \mathfrak{A}_{j+N_+}^+, \delta_j) \in \mathbb{C}^{m_+}, \quad (9.12)$$

where  $\delta_k \in \mathbb{C}^{m_+ - N_+}$  is so that  $\delta_k(n) = \delta_{k,n}$  for  $n \in [m_+ - N_+]$ . By construction we have that the  $a_j^+$  are linearly independent and that they span  $\mathcal{N}(\mathfrak{A})$ . Writing  $X_+ = (x_1, \dots, x_{d-m_+^0}) \in \mathbb{C}^{(m_+ - m_+^0) \times (d - m_+^0)}$  and  $\Delta_+ = (\delta_{d-m_+^0+1}, \dots, \delta_d) \in \mathbb{C}^{d \times m_+^0}$ , we put

$$A_1^+ = \begin{pmatrix} X_+ \\ 0 \end{pmatrix}, \quad A_2^+ = \begin{pmatrix} -\mathfrak{B}_+^{-1}(\mathfrak{A}_{m_+ - m_+^0 + 1}^+, \dots, \mathfrak{A}_{m_+}^+) \\ \Delta_+ \end{pmatrix},$$

where on the left hand side  $0 \in \mathbb{C}^{m_+^0 \times (d - m_+^0)}$ . Then,

$$A^+ = (a_1^+, \dots, a_d^+) = (A_1^+, A_2^+). \quad (9.13)$$

When  $m_+^0 = 0$ , then we have no solutions (9.12) and  $A^+ = A_1^+ = X_+$ . When  $d = m_+^0$ , then we have no solutions (9.11) and all vectors  $a_j^+$  are of the form (9.12), so

$$A^+ = A_2^+ = \begin{pmatrix} -\mathfrak{B}_+^{-1}(\mathfrak{A}_{N_+ + 1}^+, \dots, \mathfrak{A}_{m_+}^+) \\ 1_d \end{pmatrix}.$$

Keeping in mind these special cases, we shall keep writing (9.13) for all cases of  $d \geq m_+^0 \geq 0$ , with the convention that  $A_1^+$  is void when  $d = m_+^0$  and that  $A_2^+$  is void when  $m_+^0 = 0$ .

When  $d < 0$  (keeping in mind (4.34)) and  $N_- > 0$ , we work under the assumption (9.7) and we consider the analogue of  $\mathfrak{A}$  from (4.44) with  $m_+, N_+, \zeta_j^+$  replaced by  $m_-, N_-, \zeta_j^-$  and the corresponding vectors  $\mathfrak{A}_j^-$  indexed over  $1 \leq \nu \leq N_-$ . Notice that since  $N_- > 0$ , it follows that  $\infty \notin p^{-1}(\Omega)$ , so, after potentially slightly shrinking  $\Omega$ , we have that any root  $\zeta_z$  of  $p(\zeta) - z$  satisfies  $|\zeta_z| \leq C$  for all  $z \in \Omega$ . Thus, the analogue of (9.9) holds for  $\mathfrak{A}_j^-$ . Putting  $\mathfrak{B}_- = (\mathfrak{A}_{m_- - N_- + 1}^-, \dots, \mathfrak{A}_{m_-}^-)$ , we find by a Van der Monde argument that the analogue of (9.10) holds for  $\mathfrak{B}_-$ . Then, by the exact same steps as above, with the obvious modifications, we can construct  $|d|$  linearly independent solutions  $a_j^-$ , to  $\mathfrak{A}a^- = 0$ , such that

$$A^- = (a_1^-, \dots, a_d^-) = \begin{pmatrix} -\mathfrak{B}_-^{-1}(\mathfrak{A}_1^-, \dots, \mathfrak{A}_{m_-}^-) & X_- \\ \Delta_- & 0 \end{pmatrix} = (A_1^-, A_2^-), \quad (9.14)$$

where  $\delta_k \in \mathbb{C}^{m_- - N_-}$  is so that  $\delta_k(n) = \delta_{k,n}$  for  $n \in [m_- - N_-]$ ,  $\Delta_- = (\delta_1, \dots, \delta_{m_-}) \in \mathbb{C}^{|d| \times m_-}$ , and  $X_- \in \mathbb{C}^{(|d| - m_-^0) \times (|d| - m_-^0)}$  is a matrix with orthonormal columns spanning  $\mathcal{N}((\mathfrak{A}_{m_-^0 + 1}^-, \dots, \mathfrak{A}_{m_-}^-))$ .

In what follows we identify  $\ell^2([0, N - 1]) \simeq \mathbb{C}^N$ , so for the sake of consistency we index vectors and the columns of matrices starting from 0.



**Proposition 9.3.** *Let  $N \geq N_0 > 1$  be integers,  $z_0 \in p(S^1)$  be as in (9.1),  $\Omega \Subset \mathbb{C}$  be a sufficiently small open simply connected relatively compact neighbourhood of  $z_0$ , satisfying (9.2), and  $\Omega_N \Subset \Omega \setminus p(S^1)$  be as in Assumption 9.1. Let  $N > 0$  be sufficiently large, depending only on  $\Omega$  and the constants in Assumption 9.1 and the roots of  $p(\zeta) - z$  be denoted as in Lemma 4.1. Write*

$$\mathfrak{z}_j^+ = (1, \zeta_j^+, \dots, (\zeta_j^+)^{N-1})^t \in \mathbb{C}^N, \quad j = 1, \dots, m_+, \quad (9.15)$$

and

$$\mathfrak{z}_j^- = ((\zeta_j^-)^{1-N}, (\zeta_j^-)^{2-N}, \dots, 1)^t \in \mathbb{C}^N, \quad j = 1, \dots, m_-, \quad (9.16)$$

with the convention that  $1/\infty = 0$ . Moreover, let  $\mathfrak{Z}^\pm = (\mathfrak{z}_1^\pm, \dots, \mathfrak{z}_{m_\pm}^\pm)$ .

- Suppose that  $d > 0$ . When  $N_+ > 0$ , let  $A^+ = (a_1^+, \dots, a_d^+)$  be as in (9.13), and when  $N_+ \leq 0$  put  $A^+ = I_{m_+} \in \mathbb{C}^{m_+ \times m_+}$ . Define  $I_{+,1} := [d - m_+^0]$  and  $I_{+,2} := [d] \setminus [d - m_+^0]$ . Define

$$(\tilde{u}_1^+, \dots, \tilde{u}_d^+) = \mathfrak{Z}^+ A^+ (G_{+,1}^{-1/2} \oplus \mathfrak{L}_+^{-1} G_{+,2}^{-1/2}) \quad (9.17)$$

with  $\mathfrak{L}_+ = \text{diag}(\|\mathfrak{z}_{m_+ - m_+^0 + 1}^+\|, \dots, \|\mathfrak{z}_{m_+}^+\|)$ ,  $G_{+,2} = (\mathfrak{Z}^+ A_2^+ \mathfrak{L}_+^{-1})^* (\mathfrak{Z}^+ A_2^+ \mathfrak{L}_+^{-1})$ , and  $G_{+,1} = (\mathfrak{Z}^+ A_1^+)^* (\mathfrak{Z}^+ A_1^+)$ , using the convention that  $(\tilde{u}_1^+, \dots, \tilde{u}_d^+) = \mathfrak{Z}^+ A^+ G_{+,1}^{-1/2}$  when  $I_{+,2} = \emptyset$ , and that  $(\tilde{u}_1^+, \dots, \tilde{u}_d^+) = \mathfrak{Z}^+ A^+ \mathfrak{L}_+^{-1} G_{+,2}^{-1/2}$  when  $I_{+,1} = \emptyset$ . Then, uniformly for  $z \in \Omega_N$ ,

$$\langle \tilde{u}_n^+ | \tilde{u}_m^+ \rangle = \begin{cases} \delta_{n,m}, & (n, m) \in I_{+,1} \times I_{+,1} \cup I_{+,2} \times I_{+,2} \\ \delta_{n,m} + O((\log N/N)^{1/2}), & (n, m) \in I_{+,1} \times I_{+,2} \cup I_{+,2} \times I_{+,1}, \end{cases} \quad (9.18)$$

and

$$\|\mathbf{1}_{[N_0, N-1]} \tilde{u}_j^+\| \leq O(1) \begin{cases} e^{-N_0 \log C_\beta}, & j \in I_{+,1}, \\ e^{-\frac{N_0}{N} \log N^{C_\gamma}}, & j \in I_{+,2}. \end{cases} \quad (9.19)$$

- Suppose that  $d < 0$ . When  $N_- > 0$ , let  $A^- = (a_1^-, \dots, a_{|d|}^-)$  be as in (9.14), and when  $N_- \leq 0$  put  $A^- = I_{m_-} \in \mathbb{C}^{m_- \times m_-}$ . Define  $I_{-,2} := [m_-^0]$  and  $I_{-,1} := [-d] \setminus [m_-^0]$ . Define

$$(\tilde{u}_1^-, \dots, \tilde{u}_{|d|}^-) = \mathfrak{Z}^- A^- (\mathfrak{L}_-^{-1} G_{-,1}^{-1/2} \oplus G_{-,2}^{-1/2}) \quad (9.20)$$

with  $\mathfrak{L}_- = \text{diag}(\|\mathfrak{z}_{m_+ - m_+^0 + 1}^+\|, \dots, \|\mathfrak{z}_{m_+}^+\|)$ ,  $G_{-,1} = (\mathfrak{Z}^- A_1^- \mathfrak{L}_-^{-1})^* (\mathfrak{Z}^- A_1^- \mathfrak{L}_-^{-1})$  and  $G_{-,2} = (\mathfrak{Z}^- A_2^-)^* (\mathfrak{Z}^- A_2^-)$ , using a similar notation convention as above. Then, uniformly for  $z \in \Omega_N$ ,

$$\langle \tilde{u}_n^- | \tilde{u}_m^- \rangle = \begin{cases} \delta_{n,m}, & (n, m) \in I_{-,1} \times I_{-,1} \cup I_{-,2} \times I_{-,2} \\ \delta_{n,m} + O((\log N/N)^{1/2}), & (n, m) \in I_{-,1} \times I_{-,2} \cup I_{-,2} \times I_{-,1}, \end{cases} \quad (9.21)$$

and

$$\|\mathbf{1}_{[0, N-N_0-1]} \tilde{u}_j^-\| \leq O(1) \begin{cases} e^{-\frac{N_0}{N} \log N^{C_\gamma}}, & j \in I_{-,2}, \\ e^{-N_0 \log C_\beta}, & j \in I_{-,1}. \end{cases} \quad (9.22)$$

**Remark 9.4.** *When  $d > 0$ , the normalized versions of the vectors defined in (9.15) will be termed as the pure states associated with  $z$ . Similarly for  $d < 0$  the normalized version vectors (9.16) will be the pure states corresponding to  $z$ .*

*Thus, for  $z$  satisfying (9.6) we note that the pure states  $\{\mathfrak{z}_j^+ / \|\mathfrak{z}_j^+\|\}_{j=1}^{m_+ - m_+^0}$  are completely localized and its entries decay exponentially in  $N$ , while the remaining pure states decay and localize at a scale  $N / \log N$ . Similar conclusion holds for  $z$ 's satisfying (9.7).*

**Remark 9.5.** *For later use, let us record that for the Jordan block if  $z$  is inside the the unit disc then it has one pure state  $\mathfrak{z}_1^+$ , and  $\tilde{u}_1^+ = \mathfrak{z}_1^+ / \|\mathfrak{z}_1^+\|$ .*

We postpone the proof of Proposition 9.3 to after that of Proposition 9.6. Continuing, we notice that the  $\tilde{u}_j^\pm$  defined in (9.17) resp. (9.20), are the truncated kernel elements of  $P_{[0, +\infty[} - z$  resp.  $P_{]-\infty, N-1]} - z$ , separated into two groups according to their decay behaviour and orthonormalized

within their respective groups. We could use them already as quasimodes for  $P_N - z$ . In fact, it follows from (9.19) with  $N_0 = N - (N_+ \vee 0 + N_- \vee 0)$ , that when  $d > 0$

$$\|(P_N - z)\tilde{u}_j^+\| \leq O(1)\|\mathbf{1}_{[N_0, N-1]}\tilde{u}_j^+\| \leq O(1) \begin{cases} e^{-N \log C_\beta}, & j \in I_{+,1}, \\ N^{-C_\gamma}, & j \in I_{+,2}, \end{cases} \quad (9.23)$$

and similarly from (9.22) when  $d < 0$ . Ultimately we want to use these quasimodes to obtain spectral gaps between the  $(|d| - m_{\text{sign}(d)}^0)$ -th and the  $(|d| - m_{\text{sign}(d)}^0 + 1)$ -th, and between the  $|d|$ -th and  $(|d| + 1)$ -th singular value of  $P_N - z$ . We will see, that these quasimodes will be sufficiently good for our purposes when  $\gamma > 2$ . However, to obtain sufficiently fast decay when  $\gamma > 1$  we need to modify them.

**Proposition 9.6.** *Under the assumptions of Proposition 9.3, we put  $N_0 = (1 - 1/\log N)(N - 1)$ , and we have the following:*

1. If  $d > 0$ , let  $\tilde{u}_j^+$ ,  $j = 1, \dots, d$ , be as in (9.17) and define  $\psi_j^+ \in \mathbb{C}^N$ , by

$$\psi_j^+(\nu) = \begin{cases} \tilde{u}_j^+(\nu), & j \in I_{+,1}, \\ \mathbf{1}_{[0, N_0]}(\nu)\tilde{u}_j^+(\nu) + \mathbf{1}_{[N_0+1, N-1]}(\nu) \frac{\cos(\frac{\pi}{2} \frac{\nu}{N-1})}{\cos(\frac{\pi}{2} \frac{N_0}{N-1})} \tilde{u}_j^+(\nu), & j \in I_{+,2}. \end{cases} \quad (9.24)$$

2. If  $d < 0$ , let  $\tilde{u}_j^-$ ,  $j = 1, \dots, -d$ , be as in (9.20) and define  $\psi_j^- \in \mathbb{C}^N$ , by

$$\psi_j^-(\nu) = \begin{cases} \mathbf{1}_{[0, N-N_0]}(\nu) \frac{\cos(\frac{\pi}{2} \frac{N-1-\nu}{N-1})}{\cos(\frac{\pi}{2} \frac{N_0}{N-1})} \tilde{u}_j^-(\nu) + \mathbf{1}_{[N-N_0+1, N-1]}(\nu) \tilde{u}_j^-(\nu), & j \in I_{-,2}, \\ \tilde{u}_j^-(\nu), & j \in I_{-,1}. \end{cases} \quad (9.25)$$

Then, for  $i, j = 1, \dots, |d|$ , uniformly in  $z \in \Omega_N$ ,

$$\langle \psi_i^{\text{sign}(d)} | \psi_j^{\text{sign}(d)} \rangle = \delta_{i,j} + \begin{cases} 0, & \text{when } i, j \in I_{\text{sign}(d),1} \\ O(\log N)N^{-C_\gamma}, & \text{when } i, j \in I_{\text{sign}(d),2} \\ O(\log N)N^{-C_\gamma} + O((\log N/N)^{1/2}), & \text{else,} \end{cases} \quad (9.26)$$

and

$$\|(P_N - z)\psi_i^{\text{sign}(d)}\| = O(1) \begin{cases} e^{-N \log C_\beta}, & \text{when } i \in I_{\text{sign}(d),1} \\ \frac{\log N}{N} N^{-C_\gamma}, & \text{when } i \in I_{\text{sign}(d),2}. \end{cases} \quad (9.27)$$

We refer to the vectors  $\tilde{\psi}_j^\pm$  as *quasimodes*.

*Proof.* We will consider only the case when  $d > 0$ , the case when  $d < 0$  is similar. Let  $N > 0$  be large enough so that the conclusions of Proposition 9.3 hold.

1. First notice that

$$\cos\left(\frac{\pi}{2} \frac{N_0}{N-1}\right) = \sin \frac{\pi}{2 \log N} = \frac{\pi}{2 \log N} (1 + O((\log N)^{-2})). \quad (9.28)$$

Let  $j = d - m_+^0 + 1, \dots, d$ , then

$$\|\tilde{u}_j^+ - \psi_j^+\|^2 = \sum_{\nu=N_0+1}^{N-1} \left| 1 - \frac{\cos(\frac{\pi}{2} \frac{\nu}{N-1})}{\cos(\frac{\pi}{2} \frac{N_0}{N-1})} \right|^2 |\tilde{u}_j^+(\nu)|^2 \leq O((\log N)^2) \|\mathbf{1}_{[N_0+1, N-1]}\tilde{u}_j^+\|^2.$$

Applying (9.19), we get that

$$\|\tilde{u}_j^+ - \psi_j^+\| \leq O(\log N)N^{-C_\gamma}. \quad (9.29)$$

**Remark 9.7.** *For later use we note here that the analogue of (9.29) in the case when  $d < 0$  holds, i.e.*

$$\|\tilde{u}_j^- - \psi_j^-\| \leq O(\log N)N^{-C_\gamma}, \quad j \in I_{-,2}.$$

This, together with the  $\tilde{u}_j^+$ ,  $j = 1, \dots, d$ , being normalized, see (9.18), yields that

$$\|\psi_j^+\| = 1 + \begin{cases} 0, & \text{when } j \in I_{+,1} \\ O(\log N)N^{-C_\gamma}, & \text{when } j \in I_{+,2}. \end{cases}$$

Using furthermore the fact that the  $\tilde{u}_j^+$  are almost orthogonal, see (9.18), we find that

$$\begin{aligned} \langle \psi_i^+ | \psi_j^+ \rangle &= \langle \psi_i^+ - \tilde{u}_i^+ | \psi_j \rangle + \langle \tilde{u}_i^+ | \psi_j^+ - \tilde{u}_j^+ \rangle + \langle \tilde{u}_i^+ | \tilde{u}_j^+ \rangle \\ &= \delta_{i,j} + \begin{cases} 0, & \text{when } i, j \in I_{+,1}, \\ O(\log N)N^{-C_\gamma}, & \text{when } i, j \in I_{+,2} \\ O(\log N)N^{-C_\gamma} + O((\log N/N)^{1/2}), & \text{else.} \end{cases} \end{aligned}$$

2. It remains to prove (9.27). Recall from (9.17) and Proposition 4.6 that the  $\tilde{u}_j^+ = (\mathfrak{Z}^+ A^+ (G_{+,1}^{-1/2} \oplus \mathfrak{L}_+^{-1} G_{+,2}^{-1/2}))_{-,j}$  are the truncated and normalized (in fact almost orthonormalized) kernel elements of  $P_{[0,+\infty[} - z$ . Here, and in the sequel we denote by  $A_{-,j} \in \mathbb{C}^n$  the  $j$ -th column of some  $n \times m$  matrix  $A$  as a vector in  $\mathbb{C}^n$ . Let  $u_k \in \ell^2([0, +\infty[)$ ,  $k = 1, \dots, d$ , be given by

$$u_k(\nu) = \begin{cases} (\mathfrak{Z}^+ A^+)_{\nu,k}, & 0 \leq \nu \leq N-1, \\ 0, & \nu \geq N. \end{cases} \quad (9.30)$$

Then,

$$((P_{[0,+\infty[} - z)u_k)(\nu) = 0, \quad 0 \leq \nu \leq N-1 - N_- \vee 0.$$

By (4.34) we see that  $N_- > 0$  since we work here with  $d > 0$ . To ease the notation we write  $\tilde{N}_\pm = N_\pm \vee 0$ , and  $\tilde{a}_k = a_k - z\delta_{k,0}$  when  $a_k$  exists, otherwise  $\tilde{a}_k = -z\delta_{k,0}$ , so that

$$\sum_{k=-N_-}^{N_+} a_k \tau^k - z = \sum_{k=-\tilde{N}_-}^{\tilde{N}_+} \tilde{a}_k \tau^k.$$

Thus, for  $0 \leq \nu \leq N-1 - \tilde{N}_-$ ,

$$\begin{aligned} (P_N - z)\tilde{u}_j^+(\nu) &= \sum_{k=-\tilde{N}_-}^{\tilde{N}_+} \tilde{a}_k \mathbf{1}_{[0, N-1]}(\nu - k) \tilde{u}_j^+(\nu - k) \\ &\stackrel{(9.17)}{=} \sum_{k=-\tilde{N}_-}^{\tilde{N}_+} \tilde{a}_k \mathbf{1}_{[0, N-1]}(\nu - k) (\mathfrak{Z}^+ A^+ (G_{+,1}^{-1/2} \oplus \mathfrak{L}_+^{-1} G_{+,2}^{-1/2}))_{(\nu-k),j} \\ &= \sum_{n=1}^d \sum_{k=-\tilde{N}_-}^{\tilde{N}_+} \tilde{a}_k \mathbf{1}_{[0, N-1]}(\nu - k) (\mathfrak{Z}^+ A^+)_{(\nu-k),n} (G_{+,1}^{-1/2} \oplus \mathfrak{L}_+^{-1} G_{+,2}^{-1/2})_{n,j} \\ &\stackrel{(9.30)}{=} \sum_{n=1}^d (G_{+,1}^{-1/2} \oplus \mathfrak{L}_+^{-1} G_{+,2}^{-1/2})_{n,j} ((P_{[0,+\infty[} - z)u_n)(\nu) = 0. \end{aligned} \quad (9.31)$$

Thus, for  $0 \leq \nu \leq N-1 - \tilde{N}_-$  and  $j \in I_{+,1}$

$$(P_N - z)\psi_j^+(\nu) = (P_N - z)\tilde{u}_j^+(\nu) = 0.$$

Recall that  $z \in \Omega \Subset \mathbb{C}$  is relatively compact. Hence, applying (9.19) with  $N_0 = N-1 - \tilde{N}_- - \tilde{N}_+$ , we see that, uniformly in  $z \in \Omega_N$ ,

$$\|(P_N - z)\psi_j^+\| = O(1) \|\mathbf{1}_{[N-1-\tilde{N}_-, \tilde{N}_+, N-1]} \psi_i^+\| = O(1) e^{-N \log C_\beta}, \quad j \in I_{+,1}, \quad (9.32)$$

Until further notice let  $j \in I_{+,2}$ . Equation (9.31) implies that for  $0 \leq \nu \leq N_0 - \tilde{N}_-$

$$(P_N - z)\psi_j^+(\nu) = (P_N - z)\tilde{u}_j^+(\nu) = 0.$$

Next, by (9.28), we have for  $N_0 - (\tilde{N}_+ + \tilde{N}_-) < \nu \leq N_0 + (\tilde{N}_+ + \tilde{N}_-)$  that

$$\left| 1 - \frac{\cos(\frac{\pi}{2} \frac{\nu}{N-1})}{\cos(\frac{\pi}{2} \frac{N_0}{N-1})} \right| \leq O(N^{-1} \log N) |N_0 - \nu| = O(N^{-1} \log N). \quad (9.33)$$

Hence, for  $N_0 - \tilde{N}_- < \nu \leq N_0 + \tilde{N}_+$ , we get by (9.31), (9.33) that

$$\begin{aligned}
(P_N - z)\psi_j^+(\nu) &= \sum_{k=-\tilde{N}_-}^{\tilde{N}_+} \tilde{a}_k \psi_j^+(\nu - k) \\
&= \sum_{k=-\tilde{N}_-}^{\tilde{N}_+} \tilde{a}_k \tilde{u}_j^+(\nu - k) - \sum_{k=-\tilde{N}_-}^{\tilde{N}_+} \tilde{a}_k \left(1 - \frac{\cos(\frac{\pi}{2} \frac{\nu-k}{N-1})}{\cos(\frac{\pi}{2} \frac{N_0}{N-1})}\right) \tilde{u}_j^+(\nu - k) \mathbf{1}_{[N_0+1, N-1]}(\nu - k) \\
&= O(N^{-1} \log N) \|\mathbf{1}_{I(\nu)} \tilde{u}_j^+\|,
\end{aligned} \tag{9.34}$$

where  $I(\nu) := [\nu - \tilde{N}_+, \nu + \tilde{N}_-]$ . For  $N_0 + \tilde{N}_+ < \nu \leq N - 1 - \tilde{N}_-$  we get by (9.31) and a similar estimate as in (9.33), (9.34), that

$$\begin{aligned}
(P_N - z)\psi_j^+(\nu) &= \frac{\cos(\frac{\pi}{2} \frac{\nu}{N-1})}{\cos(\frac{\pi}{2} \frac{N_0}{N-1})} \sum_{k=-\tilde{N}_-}^{\tilde{N}_+} \tilde{a}_k \tilde{u}_j^+(\nu - k) - \sum_{k=-\tilde{N}_-}^{\tilde{N}_+} \tilde{a}_k \frac{\cos(\frac{\pi}{2} \frac{\nu}{N-1}) - \cos(\frac{\pi}{2} \frac{\nu-k}{N-1})}{\cos(\frac{\pi}{2} \frac{N_0}{N-1})} \tilde{u}_j^+(\nu - k) \\
&= O(N^{-1} \log N) \|\mathbf{1}_{I(\nu)} \tilde{u}_j^+\|.
\end{aligned} \tag{9.35}$$

Now let  $N - 1 - \tilde{N}_- < \nu \leq N - 1$ . For  $-\tilde{N}_- \leq k \leq \tilde{N}_+$  we have that  $\cos(\frac{\pi}{2} \frac{\nu-k}{N-1}) \asymp N^{-1}$ , so by (9.28) we get that

$$\begin{aligned}
(P_N - z)\psi_j^+(\nu) &= \sum_{k=-\tilde{N}_-}^{\tilde{N}_+} \tilde{a}_k \frac{\cos(\frac{\pi}{2} \frac{\nu-k}{N-1})}{\cos(\frac{\pi}{2} \frac{N_0}{N-1})} \tilde{u}_j^+(\nu - k) \mathbf{1}_{[N_0+1, N-1]}(\nu - k) \\
&= O(N^{-1} \log N) \|\mathbf{1}_{I(\nu)} \mathbf{1}_{[N_0+1, N-1]} \tilde{u}_j^+\|.
\end{aligned} \tag{9.36}$$

Combining (9.31), (9.34), (9.35) and (9.36) we obtain that

$$\begin{aligned}
\|(P_N - z)\psi_j^+\|^2 &\leq O((N^{-1} \log N)^2) \sum_{\nu=N_0-\tilde{N}_-}^{N-1} \|\mathbf{1}_{I(\nu)} \mathbf{1}_{[N_0-(\tilde{N}_++\tilde{N}_-), N-1]} \tilde{u}_j^+\|^2 \\
&\leq O((N^{-1} \log N)^2) \|\mathbf{1}_{[N_0-(\tilde{N}_++\tilde{N}_-), N-1]} \tilde{u}_j^+\|^2.
\end{aligned} \tag{9.37}$$

Applying now (9.19) with  $N_0 = (1 - 1/\log N)N - (\tilde{N}_+ + \tilde{N}_-)$ , we get that

$$\|(P_N - z)\psi_j^+\| = O(N^{-1} \log N) N^{-C_\gamma} \tag{9.38}$$

which concludes the proof of the proposition.  $\square$

*Proof of Proposition 9.3.* We will only consider the case when  $d > 0$ . The case  $d < 0$  can be proven similarly with the obvious modifications. In what follows, we will assume that  $N > 0$  is sufficiently large, depending only on  $\Omega$  and the constants in Assumption 9.1, without us mentioning this at every occurrence. Similarly, we note that all constants  $C > 0$  and error estimates will be uniform in  $z \in \Omega_N$ , and in fact they depend only on the above mentioned parameters, so we shall only mention this at a few important points.

1. Since the roots of  $p(\zeta) - z$  depend smoothly on  $z \in \Omega$ , see the discussion above (9.3), we have that either  $0 \in p^{-1}(z_0)$  or we may take  $\Omega$  sufficiently small so that  $0 \notin p^{-1}(\Omega)$ . In both cases  $0 \notin p^{-1}(\Omega_N)$ . Hence, we have that the symbol  $p(\zeta) - z$  satisfies for all  $z \in \Omega_N$  the assumptions of Case 1 of Lemma 4.1, see also (4.40). We then know that  $m_+ > 0$  since  $d > 0$ , see (4.34).

2. Recall from (8.24) the  $m_+ \times m_+$  matrices  $V = V_+$  and  $\Lambda = \Lambda_+$  (with  $\hat{m}_0 = 0$ ). Then, for  $r = 1, 2$  and  $\psi_1 \in \mathbb{C}^{d-m_+^0}$ ,  $\psi_2 \in \mathbb{C}^{m_+^0}$ ,

$$\|\mathbb{3}^+ A_r^+ \psi_r\|^2 \geq \sum_{j=0}^{\lfloor (N-2)/m_+ \rfloor - 1} \|V \Lambda^{j m_+} A_r^+ \psi_r\|^2. \tag{9.39}$$

Let  $s_1(V) \geq \dots \geq s_{m_+}(V) \geq 0$  denote the singular values of  $V$ . Since  $|\det V| = \prod_1^{m_+} s_j(V)$ , and  $s_1(V) = \|V\| \leq m_+$ , it follows that

$$s_{m_+}(V) \geq \frac{|\det V|}{m_+^{m_+}} = \frac{\prod_{1 \leq i < j \leq m_+} |\zeta_i^+ - \zeta_j^+|}{m_+^{m_+}} \geq \frac{1}{C}. \quad (9.40)$$

Here, in the equality we used the formula for the determinant of the Van der Monde matrix  $V$ . Moreover, the constant in the last inequality is uniform  $N$  and in  $z \in \Omega_N$  by (9.3). Since  $\Lambda$  is a diagonal matrix, we get from (9.39) that

$$\|\mathfrak{Z}^+ A_r^+ \psi_r\|^2 \geq s_{m_+}^2(V) \sum_{k=1}^{m_+} |(A_r^+ \psi_r)(k)|^2 \sum_{j=0}^{\lfloor (N-2)/m_+ \rfloor - 1} |\zeta_k^+|^{2jm_+}. \quad (9.41)$$

Putting

$$d_k^2 \stackrel{\text{def}}{=} \sum_{j=0}^{\lfloor (N-2)/m_+ \rfloor - 1} |\zeta_k^+|^{2jm_+} = \frac{1 - |\zeta_k^+|^{2\lfloor (N-2)/m_+ \rfloor m_+}}{1 - |\zeta_k^+|^{2m_+}}, \quad (9.42)$$

and  $D = \text{diag}(d_1, \dots, d_{m_+})$ , we get from (9.41) that

$$\|\mathfrak{Z}^+ A_r^+ \psi_r\| \geq s_{m_+}(V) \|DA_r^+ \psi_r\|. \quad (9.43)$$

Observe that the smallest singular value of  $D$  is  $\min_k d_k \geq 1$ . Recall (9.13). Provided that  $d > m_+^0$ , we notice that  $A_1^+$  is an isometry since the columns of  $X_+$  are orthonormal. Thus, it follows from (9.43), (9.40) that

$$\|\mathfrak{Z}^+ A_1^+ \psi_1\| \geq \frac{1}{C} \|\psi_1\|, \quad (9.44)$$

so the smallest singular value of  $G_{+,1}$  is at least  $1/C$ , which,  $G_{+,1}$  being self-adjoint, implies that it is bijective with inverse bounded in norm by  $C$ . In particular, we have that

$$\|A_1^+ G_{+,1}^{-1/2}\| = \sqrt{C}. \quad (9.45)$$

3. Working with  $m_+^0 \geq 1$ , we turn to  $G_{+,2}$  and we will show first that for  $n \geq d - m_+^0 + 1$  the  $u_n^+$  are, up to a small error, given by  $\mathfrak{z}_n / \|\mathfrak{z}_n\|$ . First, we compute

$$F_{n,m} \stackrel{\text{def}}{=} |\langle \mathfrak{z}_n^+ | \mathfrak{z}_m^+ \rangle| = \left| \sum_0^{N-1} (\zeta_n^+ \bar{\zeta}_m^+)^{\nu} \right| = \left( \frac{(1 - |\zeta_n^+|^{2N})(1 - |\zeta_m^+|^{2N}) + |(\zeta_n^+)^N - (\zeta_m^+)^N|^2}{(1 - |\zeta_n^+|^2)(1 - |\zeta_m^+|^2) + |\zeta_n^+ - \zeta_m^+|^2} \right)^{1/2}. \quad (9.46)$$

By (9.6), (9.5) and (9.3) we find that

$$F_{n,m} = \begin{cases} \asymp 1, & \text{when } 1 \leq n, m \leq m_+ - m_+^0, \\ \asymp 1, & \text{when } n \neq m, \\ \asymp \frac{N}{\log N}, & \text{when } m_+ - m_+^0 + 1 \leq n = m \leq m_+. \end{cases} \quad (9.47)$$

Second, notice that by (9.9), (9.10), we have that  $T := -\mathfrak{B}_+^{-1}(\mathfrak{A}_{m_+ - m_+^0 + 1}^+, \dots, \mathfrak{A}_{m_+}^+)$  has norm  $\|T\| = O(1)$ , depending only on  $\Omega$ . Recalling the definition of  $\Delta_+$  from the discussion before (9.13), we see that

$$\mathfrak{Z}^+ A_2^+ \mathfrak{L}_+^{-1} = \tilde{\mathfrak{Z}}^+ T \mathfrak{L}_+^{-1} + \hat{\mathfrak{Z}}^+ \mathfrak{L}_+^{-1}, \quad (9.48)$$

where  $\tilde{\mathfrak{Z}}^+ = (\mathfrak{z}_1^+, \dots, \mathfrak{z}_{N_+}^+)$  and  $\hat{\mathfrak{Z}}^+ = (\mathfrak{z}_{m_+ - m_+^0 + 1}^+, \dots, \mathfrak{z}_{m_+}^+)$ . Using (9.47), we get that

$$\begin{aligned} G_{+,2} &= \mathfrak{L}_+^{-1} (\hat{\mathfrak{Z}}^+)^* \hat{\mathfrak{Z}}^+ \mathfrak{L}_+^{-1} + \mathfrak{L}_+^{-1} \left( (\tilde{\mathfrak{Z}}^+)^* \tilde{\mathfrak{Z}}^+ T + (\tilde{\mathfrak{Z}}^+ T)^* \hat{\mathfrak{Z}}^+ + (\tilde{\mathfrak{Z}}^+ T)^* \tilde{\mathfrak{Z}}^+ T \right) \mathfrak{L}_+^{-1} \\ &= 1 + O(1) \frac{\log N}{N}, \end{aligned} \quad (9.49)$$

where the  $O(1)$  term denotes an  $m_+^0 \times m_+^0$  matrix with norm bounded by  $O(1)$ , depending only on  $\Omega$  and the constants in Assumption 9.1. Hence, for  $N > 0$  sufficiently large (depending only on  $\Omega$  and the constants in Assumption 9.1), we have that  $G_{+,2}$  is invertible and

$$G_{+,2}^{-1/2} = 1 + O(1) \frac{\log N}{N}. \quad (9.50)$$

Hence, using again (9.47) and denoting by  $M_{-,n}$  the  $n$ -th column of a matrix, we get that for  $n \in I_{+,2}$

$$\begin{aligned}\tilde{u}_n^+ &= \left( \mathfrak{Z}^+ A_2^+ \mathfrak{L}_+^{-1} G_{+,2}^{-1/2} \right)_{-,n} \\ &= \left( \tilde{\mathfrak{Z}}^+ T \mathfrak{L}_+^{-1} G_{+,2}^{-1/2} \right)_{-,n} + \left( \hat{\mathfrak{Z}}^+ \mathfrak{L}_+^{-1} G_{+,2}^{-1/2} \right)_{-,n} = \frac{\mathfrak{z}_n^+}{\|\mathfrak{z}_n^+\|} + O(1) \sqrt{\frac{\log N}{N}},\end{aligned}\tag{9.51}$$

where the  $O(1)$  term denotes a vector in  $\mathbb{C}^N$  whose  $\ell^2$  norm is bounded by  $O(1)$ . In particular, (9.44) and (9.50) imply that (9.17) is well-defined.

**Remark 9.8.** For a later use we note that the analogue of (9.51) in the case when  $d < 0$  holds, i.e.

$$\tilde{u}_n^- = \frac{\mathfrak{z}_n^-}{\|\mathfrak{z}_n^-\|} + O(1) \sqrt{\frac{\log N}{N}}, \quad n \in I_{-,2}.$$

4. Since  $(\mathfrak{Z}^+ A_1^+ G_{+,1}^{-1/2})^* (\mathfrak{Z}^+ A_1^+ G_{+,1}^{-1/2}) = 1$ , and  $(\mathfrak{Z}^+ A_2^+ \mathfrak{L}_+ G_{+,2}^{-1/2})^* (\mathfrak{Z}^+ A_2^+ \mathfrak{L}_+ G_{+,2}^{-1/2}) = 1$ , we confirm the first line of (9.18). To prove the second line, let  $(n, m) \in I_{+,1} \times I_{+,2}$ . Equations (9.51) and (9.17) then yield that

$$\begin{aligned}\langle \tilde{u}_n^+ | \tilde{u}_m^+ \rangle &= \left\langle \mathfrak{Z}^+ (A_1^+ G_{+,1}^{-1/2})_{-,n} \middle| \frac{\mathfrak{z}_m^+}{\|\mathfrak{z}_m^+\|} \right\rangle + O(1) \sqrt{\frac{\log N}{N}} \\ &= \sum_{k=1}^{m_+ - m_+^0} \langle \mathfrak{z}_k^+ | \mathfrak{z}_m^+ \rangle \|\mathfrak{z}_m^+\|^{-1} (A_1^+ G_{+,1}^{-1/2})_{k,n} + O(1) \sqrt{\frac{\log N}{N}} \stackrel{(9.45), (9.47)}{\leq} O(1) \sqrt{\frac{\log N}{N}}.\end{aligned}\tag{9.52}$$

The other case of the second line of (9.18) follows from symmetry.

5. At last, we turn to proving (9.19). Let  $j \in I_{+,1}$ . Using Hölder's inequality, we compute

$$\begin{aligned}\|\mathbf{1}_{[N_0, N-1]} \tilde{u}_j^+\|^2 &= \sum_{\nu=N_0}^{N-1} |(\mathfrak{Z}^+ A^+ G_{+,1}^{-1/2})_{\nu,j}|^2 \leq \sum_{\nu=N_0}^{N-1} \left| \sum_{\mu=1}^{m_+ - m_0^+} (\mathfrak{Z}^+)_{\nu,\mu} (A^+ G_{+,1}^{-1/2})_{\mu,j} \right|^2 \\ &\leq \sum_{\nu=N_0}^{N-1} \left( \sum_{\mu=1}^{m_+ - m_0^+} |(\mathfrak{Z}^+)_{\nu,\mu}|^2 \right) \left( \sum_{\mu=1}^{m_+ - m_0^+} |(A^+ G_{+,1}^{-1/2})_{\mu,j}|^2 \right).\end{aligned}\tag{9.53}$$

Since the right most term in the last line is controlled by the Hilbert-Schmidt norm of the corresponding matrices, we get by using (9.45) and summing the geometric series that

$$\begin{aligned}\|\mathbf{1}_{[N_0, N-1]} \tilde{u}_j^+\|^2 &\leq \|A^+ G_{+,1}^{-1/2}\|_{\text{HS}}^2 \sum_{\nu=N_0}^{N-1} \left( \sum_{\mu=1}^{m_+ - m_0^+} |(\mathfrak{Z}^+)_{\nu,\mu}|^2 \right) \\ &\stackrel{(9.45)}{\leq} O(1) \sum_{\nu=N_0}^{N-1} \left( \sum_{\mu=1}^{m_+ - m_0^+} |\zeta_\mu^+|^{2\nu} \right) = O(1) \sum_{\mu=1}^{m_+ - m_0^+} |\zeta_\mu^+|^{2N_0} \frac{1 - |\zeta_\mu^+|^{2(N-N_0)}}{1 - |\zeta_\mu^+|^2}.\end{aligned}\tag{9.54}$$

Since  $|\zeta_{m_+ - m_+^0}| \leq 1/C_\beta < 1$ , we have that for  $\mu = 1, \dots, m_+ - m_+^0$ ,

$$\frac{1 - |\zeta_\mu^+|^{2(N-N_0)}}{1 - |\zeta_\mu^+|^2} \leq \frac{C_\beta}{C_\beta - 1}.\tag{9.55}$$

Similarly we see that

$$|\zeta_\mu^+|^{2N_0} \leq e^{-2N_0 \log C_\beta}.\tag{9.56}$$

Combining this with (9.54) and (9.55) yields that

$$\|\mathbf{1}_{[N_0, N-1]} \tilde{u}_j^+\| \leq O(1) e^{-N_0 \log C_\beta},$$

uniformly in  $z \in \Omega_N$ , and we conclude the first line in (9.19).

It remains to prove the second line of (9.19). Recall the definition of  $A_2^+$  (9.13), and notice that

$$\begin{aligned} A_2^+ \mathfrak{L}_+^{-1} &= \begin{pmatrix} -\mathfrak{B}^{-1}(\mathfrak{A}_{m_+ - m_+^0 + 1}^+, \dots, \mathfrak{A}_{m_+}^+) \mathfrak{L}_+^{-1} \\ \Delta_+ \mathfrak{L}_+^{-1} \end{pmatrix} \\ &= \tilde{\mathfrak{L}}_+^{-1} \begin{pmatrix} -\mathfrak{B}^{-1}(\mathfrak{A}_{m_+ - m_+^0 + 1}^+, \dots, \mathfrak{A}_{m_+}^+) \mathfrak{L}_+^{-1} \\ \Delta_+ \end{pmatrix} \stackrel{\text{def}}{=} \tilde{\mathfrak{L}}_+^{-1} \tilde{A}_2^+ \end{aligned} \quad (9.57)$$

where  $\tilde{\mathfrak{L}}_+ := 1_{N_+} \oplus \text{diag}(\|\mathfrak{z}_{N_+ + 1}^+\|, \dots, \|\mathfrak{z}_{m_+}^+\|)$ . Furthermore, it follows from (9.9), (9.10) and (9.47) that

$$\|\mathfrak{B}^{-1}(\mathfrak{A}_{m_+ - m_+^0 + 1}^+, \dots, \mathfrak{A}_{m_+}^+) \mathfrak{L}_+^{-1}\| = O(1) \sqrt{\frac{\log N}{N}}, \quad (9.58)$$

which implies that

$$\|\tilde{A}_2^+\| = O(1). \quad (9.59)$$

Using the commutation relation (9.57), we compute for  $j \in I_{+,2}$  that

$$\begin{aligned} \|\mathbf{1}_{[N_0, N-1]} \tilde{u}_j^+\|^2 &= \sum_{\nu=N_0}^{N-1} |(\mathfrak{Z}^+ A_2^+ \mathfrak{L}_+^{-1} G_{+,2}^{-1/2})_{\nu,j}|^2 \leq \sum_{\nu=N_0}^{N-1} \left| \sum_{\mu=1}^{m_+} (\mathfrak{Z}^+ \tilde{\mathfrak{L}}_+^{-1})_{\nu,\mu} (\tilde{A}_2^+ G_{+,2}^{-1/2})_{\mu,j} \right|^2 \\ &\leq \sum_{\nu=N_0}^{N-1} \left( \sum_{\mu=1}^{m_+} |(\mathfrak{Z}^+ \tilde{\mathfrak{L}}_+^{-1})_{\nu,\mu}|^2 \right) \left( \sum_{\mu=1}^{m_+} |(\tilde{A}_2^+ G_{+,2}^{-1/2})_{\mu,j}|^2 \right). \end{aligned} \quad (9.60)$$

Thus, by (9.59), (9.50)

$$\begin{aligned} \|\mathbf{1}_{[N_0, N-1]} \tilde{u}_j^+\|^2 &\leq \|\tilde{A}^+ G_+^{-1/2}\|_{\text{HS}}^2 \sum_{\nu=N_0}^{N-1} \left( \sum_{\mu=1}^{m_+} |(\mathfrak{Z}^+ \tilde{\mathfrak{L}}_+^{-1})_{\nu,\mu}|^2 \right) \\ &\leq O(1) \sum_{\nu=N_0}^{N-1} \left( \sum_{\mu=1}^{N_+} |\zeta_\mu^+|^{2\nu} + \sum_{\mu=N_+ + 1}^{m_+} \frac{|\zeta_\mu^+|^{2\nu}}{\|\mathfrak{z}_\mu^+\|^2} \right). \end{aligned} \quad (9.61)$$

Using (9.47), we get by a computation similar to (9.55) and (9.56) that

$$\sum_{\nu=N_0}^{N-1} \left( \sum_{\mu=1}^{N_+} |\zeta_\mu^+|^{2\nu} + \sum_{\mu=N_+ + 1}^{m_+ - m_+^0} \frac{|\zeta_\mu^+|^{2\nu}}{\|\mathfrak{z}_\mu^+\|^2} \right) = O(1) e^{-N_0 \log C_\beta}. \quad (9.62)$$

On the other hand, using (9.46) and the fact that  $|\zeta_{m_+}| < 1$ , we find  $\mu = m_+ - m_+^0 + 1, \dots, m_+$  that

$$\sum_{\nu=N_0}^{N-1} \frac{|\zeta_\mu^+|^{2\nu}}{\|\mathfrak{z}_\mu^+\|^2} = |\zeta_\mu^+|^{2N_0} \frac{1 - |\zeta_\mu^+|^{2(N-N_0)}}{1 - |\zeta_\mu^+|^{2N}} \leq |\zeta_\mu^+|^{2N_0}. \quad (9.63)$$

By (9.5), we see that

$$|\zeta_\mu^+|^{2N_0} \leq \exp \left( 2N_0 \log \left( 1 - C_\gamma \frac{\log N}{N} \right) \right) \leq \exp \left( -2N_0 C_\gamma \frac{\log N}{N} \right) \leq e^{-\frac{2N_0}{N} \log N^{C_\gamma}}, \quad (9.64)$$

which, together with (9.63), (9.62) and (9.61) yields that for  $N > 0$  sufficiently large

$$\|\mathbf{1}_{[N_0, N-1]} \tilde{u}_j^+\|^2 \leq O(1) e^{-\frac{2N_0}{N} \log N^{C_\gamma}}, \quad (9.65)$$

uniformly in  $z \in \Omega_N$ , and we conclude the first second line in (9.19).  $\square$

## 10. SINGULAR VALUES AND VECTORS

We provide in this section various estimates on singular values and vectors of  $P_N - z$ . The section is divided to subsections, dealing respectively with small singular values, singular vectors, large singular values, and the smallest singular value and Hilbert-Schmidt norm of the matrices appearing in the resolvent expansion of the matrices discussed in Section 3.



**10.1. Small singular values.** We will work under the same assumptions as discussed in the beginning of Section 9. We will use the quasimodes constructed in Proposition 9.6 to obtain an upper bound on the  $|d|$ -th smallest singular values of  $P_N - z$  corresponding to the decay speed of the quasimodes. We will also show with the help of Theorem 8.1 that the  $(|d| + 1)$ -th singular value of  $P_N - z$  is significantly larger than the  $|d|$ -th one, which leads to a spectral gap. This is known as *splitting phenomenon*, see for instance [15, Section 9.2]. This phenomenon has typically been investigated for  $z$  being at a fixed distance from  $p(S^1)$  or for fixed symbols. In that case it is known that the smallest  $|d|$  singular values are of order  $O(e^{-N/C})$  whereas the  $(|d| + 1)$ -th singular value is bounded from below by some small  $N$  independent constant. What is new here is that we allow the spectral parameter  $z$  to be at an  $N$ -dependent distance from  $p(S^1)$ , which leads to a significantly smaller spectral gap.

Moreover, we will show that due to the assumptions (9.6), (9.7), there exists a second spectral gap between the first  $(|d| - m_{\text{sign}(d)}^0)$  singular values, which are exponentially small in  $N$ , and the  $(|d| - m_{\text{sign}(d)}^0 + 1)$ -th singular value, which is at most polynomially small in  $N$ .

Both spectral gaps will be crucial in describing the localization of the eigenvectors of  $P_N^\delta$ .

**Proposition 10.1.** *Let  $z_0 \in p(S^1)$  be as in (9.1) and let  $\Omega \Subset \mathbb{C}$  be a sufficiently small open simply connected relatively compact neighborhood of a point  $z_0$  satisfying (9.2). Let  $\Omega_N \Subset \Omega \setminus p(S^1)$  be as in Assumption 9.1. For  $z \in \Omega_N$  let  $t_1 \leq \dots \leq t_N$  denote the eigenvalues of  $\sqrt{(P_N - z)^*(P_N - z)}$ . Then there exists a constant  $0 < C < \infty$  such that, for all  $N$  large enough (depending only on  $\Omega$  and the constants in Assumption 9.1) and for all  $z \in \Omega_N$*

$$0 \leq t_1 \leq \dots \leq t_{|d|} \leq C \frac{\log N}{N} N^{-C_\gamma}, \quad (10.1)$$

and

$$t_{|d|+1} \geq \frac{1}{C} \frac{\log N}{N}. \quad (10.2)$$

Additionally, if  $|d| - m_{\text{sign}(d)}^0 > 0$ , then

$$0 \leq t_1 \leq \dots \leq t_{|d| - m_{\text{sign}(d)}^0} \leq C e^{-N \log C_\beta}, \quad (10.3)$$

and if  $m_{\text{sign}(d)}^0 > 0$

$$\frac{1}{C} \frac{\log N}{N} N^{-C_\alpha} \leq t_{|d| - m_{\text{sign}(d)}^0 + 1} \leq \dots \leq t_{|d|} \leq C \frac{\log N}{N} N^{-C_\gamma}. \quad (10.4)$$

**Remark 10.2.** *If in Assumption 9.1 we have that  $|\zeta_{m_+ - m_+^0 + 1}^+| = |\zeta_{m_+}^+|$  (in case  $d > 0$ ) or  $|\zeta_{m_-^0 + 1}^-| = |\zeta_{m_-}^-|$  (in case  $d < 0$ ) then clearly one can take  $C_\alpha = C_\gamma + \epsilon$  for any  $\epsilon > 0$ , improving upon (10.4). By (5.85), this will always be the case for us.*

*Proof.* Recall that for a square  $N \times N$  matrix  $A$ ,  $t_j(A)$  denote the eigenvalues of  $\sqrt{A^*A}$  (ordered non-decreasingly) and  $s_j(A)$  denote the singular values (ordered non-increasingly), see (1.13). By the singular value decomposition of the square matrix  $A$ , we have that  $t_j(A) = t_j(A^*)$  and  $s_j(A) = s_j(A^*)$ . Recall the Ky Fan inequalities, see for instance [32]: for two square matrices  $A, B$  we have

$$\begin{aligned} s_{n+m-1}(A+B) &\leq s_n(A) + s_m(B), \\ s_{n+m-1}(AB) &\leq s_n(A)s_m(B). \end{aligned} \quad (10.5)$$

Further, if  $A$  is an invertible  $N \times N$  matrix, then

$$s_n(A^{-1}) = \frac{1}{s_{N-n+1}(A)}, \quad n = 1, \dots, N.$$

From (4.6) we know that the symbol of the adjoint  $\text{Op}(p)^*$  is given by  $\widetilde{p}(\omega) = \overline{p}(1/\omega)$ . So in view of Lemma 4.1, the roots of  $\overline{p}(1/\omega) - \bar{z}$ , or equivalently the roots of  $p(1/\bar{\omega}) - z$  are given by  $\omega_j^\mp = 1/\bar{\zeta}_j^\pm$ , where  $\zeta_j^\pm$  are the roots of  $p(\zeta) - z$ . Hence, the discussion and results of Section 4 valid for  $p(\zeta) - z$  remain valid for  $\overline{p}(1/\omega) - \bar{z}$  with the roles of  $m_+, N_+$ , and  $m_-, N_-$  exchanged. Furthermore, the roles of the assumption (9.6) and (9.7) are then interchanged. This argument

will allow us to reduce the number cases of symbols  $p - z$  we need to consider, since we may always pass to studying the singular values of the adjoint  $(P_N - z)^*$ . In particular, in the rest of the proof we may and will assume that  $d > 0$  for  $z \in \Omega_N$ .

In what follows, we will work with  $N > 0$  sufficiently large, depending only on  $\Omega$  and the constants in Assumption 9.1, without us mentioning this dependence at every occurrence. Similar, we note that all error estimates will be uniform in  $z \in \Omega_N$ , and in fact they depend only on the above mentioned parameters, so we shall only mention this at a few important points.

1. Let  $z \in \Omega_N$ . We begin with the upper bound on  $t_d$ . Let  $\psi_j^+$ ,  $j = 1, \dots, d$  be as in (9.24), and note that for  $N > 0$  sufficiently large, they are linearly independent. Indeed if they were not, then there would exist an  $0 \neq a = (a_1, \dots, a_d) \in \mathbb{C}^d$  such that  $a_1\psi_1^+ + \dots + a_d\psi_d^+ = 0$ . Let  $j_0$  be such that  $|a_{j_0}| \geq |a_j|$  for  $j \neq j_0$ , with  $|a_{j_0}| > 0$ . Since the  $\psi_j^+$  are almost-orthonormal by (9.26), we find that

$$1 + O(\log N)N^{-C_\gamma} = - \sum_{j \in [d], j \neq j_0} \frac{a_j}{a_{j_0}} \langle \psi_j^+ | \psi_{j_0}^+ \rangle = O(\log N)N^{-C_\gamma} + O((\log N/N)^{1/2}), \quad (10.6)$$

a contradiction.

Let  $K_d \subset \mathbb{C}^N$  denote the linear subspace of  $\mathbb{C}^N$  spanned by the  $\psi_j^+$  and notice that it has dimension  $d$ , for  $N > 0$  sufficiently large. For any  $\psi \in K_d$ , we write  $\psi = \sum_{j=1}^d a_j \psi_j^+$ . If  $\|\psi\| = 1$ , then we get by (9.26) and the Cauchy-Schwarz inequality that for  $N > 0$  sufficiently large,

$$1 = \sum_{j=1}^d |a_j|^2 \|\psi_j^+\|^2 + \sum_{i \neq j} a_i \bar{a}_j \langle \psi_i^+ | \psi_j^+ \rangle = \|a\|^2 (1 + \hat{\epsilon}_N),$$

where  $\hat{\epsilon}_N = O(1)(N^{-C_\gamma} \log N + (\log N/N)^{1/2})$  and the constant is uniform in  $z \in \Omega_N$  and independent of  $\psi$ . Thus, for  $N > 0$  large enough,  $\|a\|^2 = (1 + \hat{\epsilon}_N)$ , uniformly in the choice of  $z \in \Omega_N$  and  $\psi$ . The min-max principle and (9.27) yield that

$$\begin{aligned} t_d^2 &= \min_{L: \dim L = d} \max_{\substack{\psi \in L \\ \|\psi\|=1}} \|(P_N - z)\psi\|^2 \leq \max_{\substack{\psi \in K_d \\ \|\psi\|=1}} \|(P_N - z)\psi\|^2 \\ &\leq \max_{\substack{\psi \in K_d \\ \|\psi\|=1}} \left( \sum_1^d |a_j| \cdot \|(P_N - z)\psi_j^+\| \right)^2 \stackrel{(9.27)}{\leq} O(1) \left( \frac{\log N}{N} N^{-C_\gamma} \right)^2, \end{aligned}$$

and we conclude the upper bound in (10.1) and (10.4).

Suppose that  $d > m_+^0$  and recall (9.6) and (9.7). Letting  $K_{d-m_+^0}$  denote the linear subspace of  $\mathbb{C}^N$  spanned by  $\psi_1^+, \dots, \psi_{d-m_+^0}^+$ , and recalling from (9.26) that these vectors are orthonormal, we deduce similarly as above from the min-max principle and (9.27), that

$$t_{d-m_+^0}^2 \leq \max_{\substack{\psi \in K_{d-m_+^0} \\ \|\psi\|=1}} \|(P_N - z)\psi\|^2 \leq \sum_1^{d-m_+^0} \|(P_N - z)\psi_j^+\|^2 \stackrel{(9.27)}{\leq} O(e^{-2N \log C_\beta}),$$

and we conclude (10.3).

2. We may write  $\Omega$  as the disjoint union

$$\Omega = \Omega' \dot{\cup} (p(S^1) \cap \Omega) \dot{\cup} \Omega''$$

so that  $\Omega_N \subset \Omega'$ . In particular  $\text{ind}_{p(S^1)}(z) = d$  for all  $z \in \Omega'$ .

The rest of the proof will be concerned with the lower bounds on  $t_{d+1}$  and  $t_{d-m_+^0+1}$  for  $z \in \Omega_N$ , and we will have to consider the two cases of  $N_\pm > 0$  or  $N_+ \leq 0$ . Note that by (4.15), these are the only cases we need to consider, since  $d > 0$  while by Remark 4.2 and (4.34), we have that  $d \leq 0$  if  $N_- < 0$ . We note here that in both cases  $\infty \notin p^{-1}(\Omega)$ , see (4.15) and Lemma 4.1.

Recall from the discussion after (9.2) that all roots  $\zeta_z$  of  $p(\zeta) - z$  depend smoothly on  $z \in \Omega$ . So, when either  $N_{\pm} > 0$ , or  $N_+ \leq 0$  and  $0 \notin p^{-1}(z_0)$ , we see that  $\{0, \infty\} \notin p^{-1}(\Omega)$  for  $\Omega$  sufficiently small. On the other hand, when  $N_+ \leq 0$  and  $0 \in p^{-1}(z_0)$ , then by the smooth dependence of the roots on  $z$  that  $0 \notin p^{-1}(\Omega')$ , after possibly shrinking  $\Omega$ . Hence, for  $\Omega$  sufficiently small, there exists a constant  $C > 0$  such that any root  $\zeta_z$  satisfies

$$0 < |\zeta_z| \leq C, \text{ for all } z \in \Omega'. \quad (10.7)$$

Since the symbol  $p(\zeta) - z$  for  $z \in \Omega'$  satisfies the assumptions of Case 1 of Lemma 4.1, we may order its roots as in (4.26). So, in combination with (10.7) we get that

$$0 < |\zeta_1^+| \leq \dots \leq |\zeta_{m_+}^+| < 1 < |\zeta_1^-| \leq \dots \leq |\zeta_{m_-}^-| \leq C, \text{ for all } z \in \Omega'. \quad (10.8)$$

By (4.34) we know that  $m_+ \geq d > 0$ , so we work in (10.8) with the convention that when  $m_- = 0$  then only the estimates on the existing roots in (10.8) apply.

Recall Assumption 9.1, and as in Theorem 8.1, let  $\Theta$  denote the Heaviside function, and let

$$\widehat{m}_0 \in \begin{cases} \{0, m_+^0\}, & m_+^0 > 0 \\ \{0\}, & m_+^0 = 0. \end{cases} \quad (10.9)$$

To shorten the notation we will write  $\Theta := \Theta(\widehat{m}_0)$  whenever convenient. The case  $\widehat{m}_0 = 0$  will be used to prove the lower bound in (10.2), and the case  $\widehat{m}_0 = m_+^0$  for the lower bound in (10.4).

Write  $\widetilde{N}_+ = N_+ \vee 0$ . For  $\zeta \in S^1$  and  $z \in \Omega'$ , we can write the symbol as

$$p(\zeta) - z = \zeta^{-\widetilde{N}_+} \left( \sum_0^{\widetilde{N}_+ + N_-} a_{N_+ - j} \zeta^j - z \zeta^{\widetilde{N}_+} \right) = \zeta^{m_+ - \widetilde{N}_+ - m_+^0 \Theta} q(\zeta) \stackrel{(4.34)}{=} \zeta^{d - m_+^0 \Theta} q(\zeta), \quad (10.10)$$

where

$$q(\zeta) = \sum_{-m_- - m_+^0 \Theta}^{m_+ - m_+^0 \Theta} q_j \zeta^{-j} := a_{-N_-} \prod_1^{m_+ - m_+^0 \Theta} (1 - \zeta_j^+ / \zeta) \prod_{m_+ - m_+^0 \Theta + 1}^{m_+} (\zeta - \zeta_j^+) \prod_1^{m_-} (\zeta - \zeta_j^-) \quad (10.11)$$

Here  $q_{m_+ - m_+^0 \Theta} = a_{-N_-} \prod_1^{m_+} (-\zeta_j^+) \prod_1^{m_-} (-\zeta_j^-)$ , and  $q_{-m_- - m_+^0 \Theta} = a_{-N_-}$ . So, by (10.8) and (4.15), we conclude that there exists some constant  $C > 0$  such that for all  $z \in \Omega'$

$$q_{m_+ - m_+^0 \Theta} \neq 0, \quad |q_{-m_- - m_+^0 \Theta}| \geq 1/C, \quad |q_j| \leq C.$$

Equation (10.11) shows that there are no roots of  $q$  on  $S^1$  and, by performing a computation as in (4.34), that  $\text{ind}_{q(S^1)}(0) = \widehat{m}_0$  for all  $z \in \Omega_N$ . This, together with (10.7), (9.3), (9.5) and (9.6) shows that  $q$ ,  $\Omega_N$  and  $\Omega'$  satisfy the assumptions of Theorem 8.1 with  $\widehat{m}_0$  as in (10.9),  $C_0 = C_\alpha$ ,  $C_1 = C_\beta$ ,  $C_2 = C_\gamma$ , and therefore there exists a constant  $C > 0$  such that for  $N > 0$  sufficiently large and all  $z \in \Omega_N$ ,

$$s_N(P_N(q)) \geq \frac{1}{C} \frac{\log N}{N} N^{-C_\alpha \Theta(\widehat{m}_0)}. \quad (10.12)$$

When  $d = m_+^0 > 0$  then  $s_{N-j+1} = t_j$ , and we immediately conclude the lower bound in (10.4) using that  $\widehat{m}_0 = m_+^0$ . Hence, from now on we may and will assume that  $d > m_+^0 \Theta(\widehat{m}_0)$ , see also Assumption 9.1.

Recall the definition of  $\chi_n$  from (4.10). As noted after (4.12) we have that  $H(\chi_n) = 0$  when  $n \leq 0$  and that  $H(\chi_n)$  has rank  $n$ , when  $n > 0$ . So by (4.14) and (10.10) we get that

$$P_N(p - z) P_N(\chi_{m_+^0 \Theta - d}) = P_N(q) - \Pi_N H(p - z) H(\chi_{d - m_+^0 \Theta}) \Pi_N, \quad (10.13)$$

where the second term on the right hand side has rank  $d - m_+^0 \Theta$ . Since  $s_1(P_N(\chi_{m_+^0 \Theta - d})) = \|P_N(\chi_{m_+^0 \Theta - d})\| = 1$ , we get by (10.13) and the second Ky Fan inequality in (10.5) that for any  $n = 1, \dots, N$ ,

$$s_n(P_N(p - z)) = s_n(P_N(p - z)) s_1(P_N(\chi_{m_+^0 \Theta - d})) \geq s_n(P_N(q) - \Pi_N H(p - z) H(\chi_{d - m_+^0 \Theta}) \Pi_N). \quad (10.14)$$

By interleaving inequalities for singular values, see e.g. [12, Problem III.6.4], for any matrices  $A, K \in \mathbb{C}^{N \times N}$  with  $K$  of rank  $m \in \{0, \dots, N-1\}$ ,

$$s_{N-m}(A) \geq s_N(A + K). \quad (10.15)$$

Applying this estimate with

$$A = P_N(q) - \Pi_N H(p-z) H(\chi_{d-m_+^0 \Theta}) \Pi_N, \quad K = \Pi_N H(p-z) H(\chi_{d-m_+^0 \Theta}) \Pi_N,$$

and  $m = d - m_+^0 \Theta$ , yields together with (10.14) and (10.12) that for  $N > 0$  large enough (so that (10.12) holds)

$$s_{N-d+m_+^0 \Theta}(P_N(p-z)) \geq s_N(P_N(q)) \geq \frac{1}{C} \frac{\log N}{N} N^{-C_\alpha \Theta(\widehat{m}_0)},$$

for all  $z \in \Omega_N$ . Since  $s_{N-j+1} = t_j$ , we get (10.2) by taking  $\widehat{m}_0 = 0$ , and we obtain the lower bound in (10.4), when  $d > m_+^0 > 0$ , by taking  $\widehat{m}_0 = m_+^0$ . This ends the proof of Proposition 10.1.  $\square$

**10.2. Singular vectors.** Throughout this section we will continue to work with  $z \in \Omega_N$  satisfying Assumption 9.1 and we will denote the orthonormal eigenvectors of  $(P_N - z)^*(P_N - z)$  corresponding to the eigenvalues  $0 \leq t_1^2 \leq \dots \leq t_N^2$  by  $e_1, \dots, e_N$ . Using the spectral gap from Proposition 10.1 we will show that the quasimodes from Proposition 9.6 span, up to a small error, the same space as the eigenvectors  $e_1, \dots, e_{|d|}$ . More precisely we will show that the eigenvectors  $e_1, \dots, e_{|d|-m_{\text{sign}(d)}^0}$  are well approximated by a linear combination of the quasimodes  $\psi_1, \dots, \psi_{|d|-m_{\text{sign}(d)}^0}$ , whereas the eigenvectors  $e_{|d|-m_{\text{sign}(d)}^0+1}, \dots, e_{|d|}$  are well approximated by a linear combination of  $\psi_{|d|-m_{\text{sign}(d)}^0+1}, \dots, \psi_{|d|}$ .

**Proposition 10.3.** *Consider the setup as in Proposition 10.1. Let  $\Pi_\kappa = \mathbf{1}_{[0, \kappa^2]}((P_N - z)^*(P_N - z))$ ,  $\kappa \geq 0$ ,  $z \in \Omega_N$ , be the spectral projector onto the eigenspace of  $(P_N - z)^*(P_N - z)$  corresponding to the eigenvalues  $0 \leq t_1^2 \leq \dots \leq t_j^2 \leq \kappa^2$ . Let  $\psi_j^\pm$  and  $I_{\pm,1}, I_{\pm,2}$  be as in Proposition 9.6, let  $N > 0$  be sufficiently large (depending only on the  $\Omega$  and the constants in Assumption 9.1), and put*

$$\tilde{e}_j := \frac{\Pi_{t_{|d|-m_{\text{sign}(d)}^0}^0} \psi_j^{\text{sign}(d)}}{\|\Pi_{t_{|d|-m_{\text{sign}(d)}^0}^0} \psi_j^{\text{sign}(d)}\|}, \quad j \in I_{\text{sign}(d),1},$$

and

$$\tilde{e}_j := \frac{\Pi_{t_{|d|}} \psi_j^{\text{sign}(d)}}{\|\Pi_{t_{|d|}} \psi_j^{\text{sign}(d)}\|}, \quad j \in I_{\text{sign}(d),2}.$$

Then, uniformly in  $z \in \Omega_N$ ,

$$\|\tilde{e}_j - \psi_j^{\text{sign}(d)}\| = O(1) \begin{cases} \frac{N^{(1+C_\alpha)}}{\log N} e^{-N \log C_\beta}, & j \in I_{\text{sign}(d),1}, \\ N^{-C_\gamma} \log N, & j \in I_{\text{sign}(d),2}, \end{cases}$$

and

$$\langle \tilde{e}_i | \tilde{e}_j \rangle = \delta_{i,j} + \begin{cases} \frac{N^{(1+C_\alpha)}}{\log N} e^{-N \log C_\beta}, & \text{when } i, j \in I_{\text{sign}(d),1}, \\ O(\log N) N^{-C_\gamma}, & \text{when } i, j \in I_{\text{sign}(d),2}, \\ O(\log N) N^{-C_\gamma} + O((\log N/N)^{1/2}), & \text{else.} \end{cases} \quad (10.16)$$

*Proof.* To ease the notation we drop the  $\pm$  superscripts and let  $\psi_j$  be either  $\psi_j^+$  or  $\psi_j^-$  depending on  $\text{sign}(d)$ . Furthermore, we write  $B_z := (P_N - z)^*(P_N - z)$  and we let  $N \gg 1$ . In what follows all constants will be uniform in  $N \gg 1$  and  $z \in \Omega_N$  even when we do not state this fact explicitly.

By (9.27), we have that

$$\langle B_z \psi_j | \psi_j \rangle = \|(P_N - z) \psi_j\|^2 = O(1) \begin{cases} e^{-2N \log C_\beta}, & j \in I_{\text{sign}(d),1} \\ \left( \frac{\log N}{N} N^{-C_\gamma} \right)^2, & j \in I_{\text{sign}(d),2}. \end{cases} \quad (10.17)$$

Let  $\kappa = t_{|d|-m_{\text{sign}(d)}^0}$  for  $j \in I_{\text{sign}(d),1}$ , and  $\kappa = t_{|d|}$  for  $j \in I_{\text{sign}(d),2}$ . Let  $\Pi = \Pi_\kappa$  and  $\Pi^\perp = \mathbf{1}_{] \kappa^2, \infty[}(B_z)$ , and notice that both are selfadjoint projectors onto disjoint eigenspaces and that  $\Pi + \Pi^\perp = 1$ . By the spectral theorem and the fact that  $B_z \geq 0$ ,

$$\langle B_z \psi_j | \psi_j \rangle = \langle B_z \Pi \psi_j | \Pi \psi_j \rangle + \langle B_z \Pi^\perp \psi_j | \Pi^\perp \psi_j \rangle \geq \|\Pi^\perp \psi_j\|^2 \begin{cases} t_{|d|-m_{\text{sign}(d)}^0}^2 + 1, & j \in I_{\text{sign}(d),1}, \\ t_{|d|+1}^2, & j \in I_{\text{sign}(d),2}. \end{cases} \quad (10.18)$$

Combining (10.17), (10.18) and (10.2), (10.4) we get that

$$\|\Pi^\perp \psi_j\| \stackrel{(10.17)}{\leq} \begin{cases} O(1) \left( e^{-N \log C_\beta} t_{|d|-m_{\text{sign}(d)}^0}^{-1} \right) \stackrel{(10.4)}{\leq} O(1) \frac{N^{(1+C_\alpha)}}{\log N} e^{-N \log C_\beta}, & j \in I_{\text{sign}(d),1}, \\ O(1) \left( t_{|d|+1}^{-1} \frac{\log N}{N} N^{-C_\gamma} \right) \stackrel{(10.2)}{\leq} O(1) N^{-C_\gamma}, & j \in I_{\text{sign}(d),2}. \end{cases}$$

Hence, by (9.26),

$$|1 - \|\Pi \psi_j\|| \leq \begin{cases} \|\psi_j - \Pi \psi_j\| = \|\Pi^\perp \psi_j\| = O(1) \frac{N^{(1+C_\alpha)}}{\log N} e^{-N \log C_\beta}, & j \in I_{\text{sign}(d),1} \\ \|\psi_j - \Pi \psi_j\| + O(1) N^{-C_\gamma} \log N = O(1) N^{-C_\gamma} \log N, & j \in I_{\text{sign}(d),2} \end{cases} =: \varepsilon_j.$$

Thus, for  $N > 0$  sufficiently large, so that  $\varepsilon_j \ll 1$ ,

$$\|\tilde{e}_j - \psi_j\| = \left\| \frac{\Pi \psi_j}{1 + O(\varepsilon_j)} - \psi_j \right\| \leq O(\varepsilon_j),$$

which, together with (9.26) and

$$\langle \tilde{e}_i | \tilde{e}_j \rangle = \langle \psi_i | \psi_j \rangle + \langle \tilde{e}_i - \psi_i | \psi_j \rangle + \langle \psi_i | \tilde{e}_j - \psi_j \rangle + \langle \tilde{e}_i - \psi_i | \tilde{e}_j - \psi_j \rangle,$$

lets us conclude the proofs of the proposition.  $\square$

Notice that (10.16) implies by a similar argument as before (10.6), that for  $N > 0$  large enough, the vectors  $\tilde{e}_1, \dots, \tilde{e}_{|d|}$  are linearly independent. Hence, we conclude the following.

**Corollary 10.4.** *Under the assumptions of Proposition 10.3, we have that the  $\tilde{e}_1, \dots, \tilde{e}_{|d|-m_{\text{sign}(d)}^0}$  span the range  $\mathcal{R}(\Pi_{t_{|d|-m_{\text{sign}(d)}^0}})$  and that  $\tilde{e}_1, \dots, \tilde{e}_{|d|}$  span  $\mathcal{R}(\Pi_{t_{|d|}})$ .*

**Remark 10.5.** *By Remark 9.5, (9.29), and Proposition 10.3 it follows that for the Jordan block with  $z$  inside the unit disc such that (9.5) holds we have that*

$$\left\| \frac{\mathfrak{z}_1^+}{\|\mathfrak{z}_1^+\|} - e_1 \right\| = O(N^{-C_\gamma} \log N).$$

In what follows we suppose that  $N > 0$  is sufficiently large, so that the conclusion of Proposition 10.3 holds. Furthermore, all error terms will be uniform in  $z \in \Omega_N$  even if we don't state this explicitly.

In view of Proposition 10.3 and Corollary 10.4, we know that there exist  $a_1, \dots, a_{|d|-m_{\text{sign}(d)}^0} \in \ell^2(I_{\text{sign}(d),1})$  and  $b_{|d|-m_{\text{sign}(d)}^0+1}, \dots, b_{|d|} \in \ell^2([|d|])$  such that

$$e_j = \sum_{\nu \in I_{\text{sign}(d),1}} a_j(\nu) \tilde{e}_\nu, \quad j \in I_{\text{sign}(d),1}, \quad \text{and} \quad e_j = \sum_1^{|d|} b_j(\nu) \tilde{e}_\nu, \quad j \in I_{\text{sign}(d),2}.$$

Let  $n, m \in I_{\text{sign}(d),1}$ . Using (10.16) and the fact that the  $e_j$  are orthonormal, we get that

$$\delta_{n,m} = \langle e_n | e_m \rangle = \sum_{\nu, \mu} a_n(\nu) \bar{a}_m(\mu) \langle \tilde{e}_\nu | \tilde{e}_\mu \rangle = \langle a_n | a_m \rangle + O\left(\|a\|^2 N^{(1+C_\alpha)} e^{-N \log C_\beta} / \log N\right),$$

where the constant in the error estimate is independent of  $a$ . We deduce that  $\|a_n\| = 1 + O(N^{(1+C_\alpha)} e^{-N/C} / \log N)$ , and

$$\langle a_n | a_m \rangle = \delta_{n,m} + O\left(N^{(1+C_\alpha)} e^{-N \log C_\beta} / \log N\right). \quad (10.19)$$

Similarly, we deduce from (10.16) that for  $n, m \in I_{\text{sign}(d),2}$

$$\langle b_n | b_m \rangle = \delta_{n,m} + O(N^{-C_\gamma} \log N) + O((\log N/N)^{1/2}), \quad (10.20)$$

where the constants in the error estimate are independent of  $b$ . Decompose

$$b_j = \begin{cases} \tilde{b}_j \oplus \hat{b}_j \in \ell^2(I_{+,1}) \oplus \ell^2(I_{+,2}), & \text{sign}(d) > 0, \\ \tilde{b}_j \oplus \hat{b}_j \in \ell^2(I_{-,2}) \oplus \ell^2(I_{-,1}), & \text{sign}(d) < 0. \end{cases}$$

Exploiting the orthogonality of the  $e_j$  and using (10.19), (10.20), we deduce similarly as above that for  $n \in I_{\text{sign}(d),1}$  and  $m \in I_{\text{sign}(d),2}$ ,

$$\langle a_n | \tilde{b}_m \rangle = O(N^{-C_\gamma} \log N) + O((\log N/N)^{1/2}). \quad (10.21)$$

By a similar argument as before (10.6), we deduce from (10.19) that for  $N > 0$  sufficiently large (depending only on the  $\Omega$  and the constants in Assumption 9.1), the  $a_j$  are linearly independent and span  $\ell^2(I_{\pm,1})$ . Thus, for each  $\tilde{b}_j$  there exists a  $d_j \in \mathbb{C}^{|d|-m_{\text{sign}(d)}^0}$  such that  $\tilde{b}_j = \sum_\nu d_j(\nu) a_\nu$ . Hence, by (10.21), (10.19),

$$\|\tilde{b}_m\|^2 = \sum_{\nu, \mu} d_m(\nu) \overline{d_m(\mu)} \langle a_\nu | a_\mu \rangle = \|d_m\|^2 \left( 1 + O\left( N^{(1+C_\alpha)} e^{-N \log C_\beta} / \log N \right) \right). \quad (10.22)$$

Combining this with (10.21), (10.22), (10.19), we deduce that

$$\|d_k\| = O(N^{-C_\gamma} \log N) + O((\log N/N)^{1/2}) + O\left( N^{(1+C_\alpha)} e^{-N \log C_\beta} / \log N \right),$$

where the error terms are independent of  $a$ ,  $b$ , and  $d$ . Notice that the error term which is exponentially small in  $N$  may be absorbed into the other two terms. Hence,

$$\|\tilde{b}_m\| = O(N^{-C_\gamma} \log N) + O((\log N/N)^{1/2}).$$

Thus, for  $j \in I_{\text{sign}(d),2}$ ,

$$e_j = \sum_{\nu \in I_{\text{sign}(d),2}} \hat{b}_j(\nu) \psi_\nu^{\text{sign}(d)} + O(N^{-C_\gamma} \log N) + O((\log N/N)^{1/2}).$$

In view of (9.29), (9.51) and Remarks 9.8 and 9.7, we may replace  $\psi_\nu^{\text{sign}(d)}$  in the above expression by  $\hat{\mathfrak{z}}_\nu^{\text{sign}(d)} / \|\hat{\mathfrak{z}}_\nu^{\text{sign}(d)}\|$ , see also Proposition 9.3.

Summing up what we have proven so far, we obtain in view of Propositions 10.3 and 9.6 the following. Recall that for  $z \in \Omega_N$ ,  $e_1, \dots, e_N$  are an orthonormal set of eigenvectors of  $(P_N - z)^*(P_N - z)$  corresponding to the eigenvalues  $0 \leq t_1^2 \leq \dots \leq t_N^2$ .

**Proposition 10.6.** *Under the assumptions of Proposition 10.3, we have that for  $j \in I_{\text{sign}(d),1}$ ,*

$$e_j = \sum_{\nu \in I_{\text{sign}(d),1}} a_j(\nu) \psi_\nu^{\text{sign}(d)} + O(1) N^{(1+C_\alpha)} e^{-N \log C_\beta} / \log N,$$

where  $a_j \in \ell^2(I_{\text{sign}(d),1})$  satisfy (10.19), uniformly in  $z \in \Omega_N$ . Further, for  $j \in I_{\text{sign}(d),2}$ ,

$$e_j = \sum_{\nu \in I_{\text{sign}(d),2}} b_j(\nu) \frac{\hat{\mathfrak{z}}_\nu^{\text{sign}(d)}}{\|\hat{\mathfrak{z}}_\nu^{\text{sign}(d)}\|} + O(N^{-C_\gamma} \log N) + O((\log N/N)^{1/2}). \quad (10.23)$$

where  $b_j \in \ell^2(I_{\text{sign}(d),2})$  satisfy

$$\langle b_n | b_m \rangle = \delta_{n,m} + O(\log N N^{-C_\gamma}) + O((\log N/N)^{1/2}), \quad (10.24)$$

uniformly in  $z \in \Omega_N$ .

**Remark 10.7.** *We recall, see points (i), (ii) of Lemma 6.16 and (5.85), that for  $\nu \neq \nu' \in I_{\text{sign}(d),2}$  we have that  $|\zeta_\nu^{\text{sign}(d)} \cdot \overline{\zeta_{\nu'}^{\text{sign}(d)}} - 1|$  is bounded away from 0 by a constant  $c > 0$  independent of  $N$ , while for  $\nu = \nu' \in I_{\text{sign}(d),2}$  it is of order  $\log N/N$  by (9.5)-(9.7).*

Next, we turn to proving localization estimates of the singular vectors.

**Theorem 10.8.** *Consider the setup and notation as in Proposition 10.1. Then, there exist constants  $0 < C, C' < \infty$  such that for  $N > 0$  sufficiently large (depending only on  $\Omega$  and the constants in Assumption 9.1) we have that for any  $z \in \Omega_N$ , for any discrete interval  $I = [\ell_0, \ell_1] \subset [0, N-1]$ , with  $|I| \geq 2(1 + m_+^0)$ , and for  $j \in I_{\text{sign}(d), 2}$*

$$\|e_j\|_{\ell^2(I)}^2 \leq C \min \left\{ |I| \frac{\log N}{N}, \frac{1}{C'} \right\} e^{-2C_\gamma \mathfrak{d} \frac{\log N}{N}} + CN^{-2C_\gamma} (\log N)^2 + C \frac{\log N}{N}, \quad (10.25)$$

and

$$\|e_j\|_{\ell^2(I)}^2 \geq \frac{1}{C} \min \left\{ |I| \frac{\log N}{N}, \frac{1}{C'} \right\} e^{-2C_\alpha \mathfrak{d} \frac{\log N}{N}} - CN^{-2C_\gamma} (\log N)^2 - C \frac{\log N}{N}, \quad (10.26)$$

and for  $j \in I_{\text{sign}(d), 1}$ ,

$$\|e_j\|_{\ell^2(I)}^2 \leq Ce^{-2\mathfrak{d} \log C_\beta} + CN^{2(1+C_\alpha)} e^{-2N \log C_\beta} (\log N)^{-2}. \quad (10.27)$$

Here,

$$\mathfrak{d} := \begin{cases} \ell_0 & \text{when } d > 0, \\ N - \ell_1 & \text{when } d < 0. \end{cases}$$

We remark that we do not use (10.26) in the rest of the paper (instead, we use (10.23)). Its inclusion here is to contrast with (10.25).

*Proof.* We shall only consider the case when the winding number  $d > 0$ , since the case when  $d < 0$  is similar. Throughout we will assume that  $N > 0$  is sufficiently large (depending only on  $\Omega$  and the constants in Assumption 9.1) so that the conclusions of Propositions 10.6, 10.3, 9.6, and 9.3 hold, and we have that all error terms are uniform in  $z \in \Omega_N$  without us mentioning this explicitly.

1. We begin by considering the case where  $d - m_+^0 > 0$  and  $j \in I_{+, 1}$ . By Propositions 10.6 and 9.6, we find that

$$\|e_j\|_{\ell^2(I)}^2 \leq 2 \left\| \sum_{\nu \in I_{+, 1}} a_j(\nu) \psi_\nu^+ \right\|_{\ell^2(I)}^2 + O(1) \frac{N^{2(1+C_\alpha)}}{(\log N)^2} e^{-2N \log C_\beta}. \quad (10.28)$$

Recall from Proposition 9.3 that  $\psi_\nu^+$  is given by the  $\nu$ -th column of  $\mathfrak{Z}^+ A_1^+ G_{+, 1}^{-1/2}$ . Using the Cauchy-Schwarz inequality we get that

$$\begin{aligned} \left\| \sum_{\nu \in I_{+, 1}} a_j(\nu) \psi_\nu^+ \right\|_{\ell^2(I)}^2 &= \sum_{\mu \in I} \left| \sum_{\nu \in I_{+, 1}} \sum_{\eta=1}^{m_+ - m_+^0} \mathfrak{Z}_{\mu, \eta}^+ (A_1^+ G_{+, 1}^{-1/2})_{\eta, \nu} a_j(\nu) \right|^2 \\ &= \sum_{\mu \in I} \left| \sum_{\eta=1}^{m_+ - m_+^0} \mathfrak{Z}_{\mu, \eta}^\pm (A_1^+ G_{+, 1}^{-1/2} a_j)(\eta) \right|^2 \leq \|A_1^+ G_{+, 1}^{-1/2} a_j\|^2 \sum_{\eta=1}^{m_+ - m_+^0} \|\mathfrak{z}_\eta^+\|_{\ell^2(I)}^2. \end{aligned} \quad (10.29)$$

where  $\mathfrak{z}_\eta^+$  is defined as in Proposition 9.3. For  $\eta = 1, \dots, m_+$ , we have that

$$\|\mathfrak{z}_\eta^+\|_{\ell^2(I)}^2 = \sum_{\nu=\ell_0}^{\ell_1} |\zeta_\eta^+|^{2\nu} = |\zeta_\eta^+|^{2\ell_0} \frac{1 - |\zeta_\eta^+|^{2|I|}}{1 - |\zeta_\eta^+|^2}. \quad (10.30)$$

By (9.6) we have that  $|\zeta_\eta^+| \leq 1/C_\beta$  for  $\eta = 1, \dots, m_+ - m_+^0$ . Thus,

$$|\zeta_\eta^+|^{2\ell_0} \leq \|\mathfrak{z}_\eta^+\|_{\ell^2(I)}^2 \leq \frac{C_\beta^2}{C_\beta^2 - 1} |\zeta_\eta^+|^{2\ell_0}. \quad (10.31)$$

Combining (9.45), (10.19), with (10.28) and (10.29), we conclude (10.27).



2. Next, we let  $j \in I_{+,2}$ , so we may assume that  $m_+^0 > 0$ . By Propositions 10.6 and 9.6, we find that

$$\|e_j\|_{\ell^2(I)}^2 \leq 2 \left\| \sum_{\nu \in I_{+,2}} b_j(\nu) \frac{\mathfrak{z}_\nu^+}{\|\mathfrak{z}_\nu^+\|} \right\|_{\ell^2(I)}^2 + \varepsilon_N, \quad (10.32)$$

where, to ease the notation, we put  $\varepsilon_N = O((\log N)^2)N^{-2C_\gamma} + O(\log N/N)$ . For the sake of the presentation we will keep this notation for the error term and we allow for the constant in the estimate to change while remaining uniform in  $z \in \Omega_N$ .

From Proposition 10.6, we see that the coefficients  $b_j$  are uniformly bounded in  $\ell^2$  by  $O(1)$ . Thus by a similar computation as in (10.29), (10.31), we get together with (9.50), (9.59) that

$$\|e_j\|_{\ell^2(I)}^2 \leq O(1) \sum_{\nu \in I_{+,2}} \frac{\|\mathfrak{z}_\nu^+\|_{\ell^2(I)}^2}{\|\mathfrak{z}_\nu^+\|^2} + \varepsilon_N. \quad (10.33)$$

Recall (9.6) and (9.5). Let  $\nu \in I_{+,2}$ , and suppose first that  $|I| \ll N/\log N$ , we find by Taylor expansion that

$$2|I|C_\alpha \frac{\log N}{N} + O\left(|I| \frac{\log N}{N}\right)^2 \leq 1 - |\zeta_\nu^+|^{2|I|} \leq 2|I|C_\gamma \frac{\log N}{N} + O\left(|I| \frac{\log N}{N}\right)^2. \quad (10.34)$$

Thus,

$$\frac{C_\alpha}{C_\gamma} |I| \left(1 + O\left(|I| \frac{\log N}{N}\right)\right) \leq \frac{1 - |\zeta_\nu^+|^{2|I|}}{1 - |\zeta_\nu^+|^2} \leq \frac{C_\gamma}{C_\alpha} |I| \left(1 + O\left(|I| \frac{\log N}{N}\right)\right). \quad (10.35)$$

Plugging the estimates (10.34), (10.35) into (10.30), and using (9.47) as well, we get that there exists a constant  $C > 0$  such that, for  $N > 0$  large enough,

$$\frac{1}{C} |I| \frac{\log N}{N} e^{-2C_\alpha \ell_0 (1 + O(\frac{\log N}{N})) \frac{\log N}{N}} \leq \frac{\|\mathfrak{z}_\nu^+\|_{\ell^2(I)}^2}{\|\mathfrak{z}_\nu^+\|^2} \asymp |I| \frac{\log N}{N} |\zeta_\nu^+|^{2\ell_0} \leq C |I| \frac{\log N}{N} e^{-2C_\gamma \ell_0 \frac{\log N}{N}}. \quad (10.36)$$

When  $N/(C \log N) \leq |I| \leq N$  for some  $C \gg 1$ , then we see by Taylor expansion that

$$\frac{1 - |\zeta_\nu^+|^{2|I|}}{1 - |\zeta_\nu^+|^2} \asymp \frac{N}{\log N}.$$

Hence, by (9.47) and (10.30), we get that there exists a constant  $C > 0$  such that, for  $N > 0$  large enough,

$$\frac{1}{C} e^{-2C_\alpha \ell_0 (1 + O(\frac{\log N}{N})) \frac{\log N}{N}} \leq \frac{\|\mathfrak{z}_\nu^+\|_{\ell^2(I)}^2}{\|\mathfrak{z}_\nu^+\|^2} \asymp |\zeta_\nu^+|^{2\ell_0} \leq C e^{-2C_\gamma \ell_0 \frac{\log N}{N}}. \quad (10.37)$$

By (10.33), (10.36) and (10.37) we conclude (10.25).

3. Next we prove the lower bound (10.26), so we work still with  $j \in I_{+,2}$  and  $m_+^0 > 0$ . Similar to (10.32), we find by Propositions 10.6 and 9.6 that

$$\|e_j\|_{\ell^2(I)}^2 \geq \frac{1}{2} \left\| \sum_{\nu \in I_{+,2}} b_j(\nu) \frac{\mathfrak{z}_\nu^+}{\|\mathfrak{z}_\nu^+\|} \right\|_{\ell^2(I)}^2 - \varepsilon_N.$$

Let

$$V = \begin{pmatrix} 1 & \dots & 1 \\ \zeta_{m_+ - m_+^0 + 1}^+ & \dots & \zeta_{m_+}^+ \\ \vdots & \vdots & \vdots \\ (\zeta_{m_+ - m_+^0 + 1}^+)^{m_+ - m_+^0 - 1} & \dots & (\zeta_{m_+}^+)^{m_+ - m_+^0 - 1} \end{pmatrix}$$

$\Lambda = \text{diag}(\zeta_{m_+ - m_+^0 + 1}^+, \dots, \zeta_{m_+}^+)$  and let  $\mathfrak{L}_+$  be as in Proposition 9.3. By a similar argument as for (9.40), we see that  $s_{m_+ - m_+^0}(V) \geq 1/C$  for some  $C > 0$ , depending only on  $\Omega$ . Let  $|I| \geq 2(1 + m_+^0)$ . Then,

$$\begin{aligned} \left\| \sum_{\nu \in I_{+,2}} b_j(\nu) \frac{\mathfrak{z}_\nu^+}{\|\mathfrak{z}_\nu^+\|} \right\|_{\ell^2(I)}^2 &\geq \sum_{k=0}^{\lfloor (|I|-1)/m_+^0 \rfloor - 1} \|V\Lambda^{\ell_0 + km_+^0} \mathfrak{L}_+^{-1} b\|^2 \geq \frac{1}{C^2} \sum_{k=0}^{\lfloor (|I|-1)/m_+^0 \rfloor - 1} \|\Lambda^{\ell_0 + km_+^0} \mathfrak{L}_+^{-1} b\|^2 \\ &\geq \frac{1}{C^2} \sum_{\nu \in I_{+,2}} |b_j(\nu)|^2 \sum_{k=0}^{\lfloor (|I|-1)/m_+^0 \rfloor - 1} \frac{|\zeta_\nu^+|^{2(\ell_0 + km_+^0)}}{\|\mathfrak{z}_\nu^+\|^2} \end{aligned} \quad (10.38)$$

Let  $\nu \in I_{+,2}$ . A direct computation yields that

$$\sum_{k=0}^{\lfloor (|I|-1)/m_+^0 \rfloor - 1} |\zeta_\nu^+|^{2(\ell_0 + km_+^0)} = |\zeta_\nu^+|^{2\ell_0} \frac{1 - |\zeta_\nu^+|^{2m_+^0(\lfloor (|I|-1)/m_+^0 \rfloor - 1)}}{1 - |\zeta_\nu^+|^{2m_+^0}} \geq |\zeta_\nu^+|^{2\ell_0} \frac{1 - |\zeta_\nu^+|^{2(|I|-1-m_+^0)}}{1 - |\zeta_\nu^+|^{2m_+^0}} \quad (10.39)$$

Recall (9.5) and (9.6). Suppose first that  $2(1 + m_+^0) \leq |I| \ll N/\log N$ , then  $|I|/2 \leq |I| - 1 - m_+^0 \leq |I|$ , and by Taylor expansion we find that

$$1 - |\zeta_\nu^+|^{2(|I|-1-m_+^0)} \geq |I| C_\alpha \frac{\log N}{N} + O\left(\left(|I| \frac{\log N}{N}\right)^2\right).$$

and similarly

$$1 - |\zeta_\nu^+|^{2m_+^0} \leq 2m_+^0 C_\gamma \frac{\log N}{N} + O\left(\left(\frac{\log N}{N}\right)^2\right).$$

Thus,

$$\frac{1 - |\zeta_\nu^+|^{2(|I|-1-m_+^0)}}{1 - |\zeta_\nu^+|^{2m_+^0}} \geq \frac{C_\alpha}{2m_+^0 C_\gamma} |I| \left(1 + O\left(|I| \frac{\log N}{N}\right)\right). \quad (10.40)$$

Combining (10.40), (10.39) and (10.38) with (9.47), yields that there exists a constant  $C > 0$  such that for  $N > 0$  large enough and all  $z \in \Omega_N$ ,

$$\sum_{k=0}^{\lfloor (|I|-1)/m_+^0 \rfloor - 1} \frac{|\zeta_\nu^+|^{2(\ell_0 + km_+^0)}}{\|\mathfrak{z}_\nu^+\|^2} \geq \frac{1}{C} |I| \frac{\log N}{N} |\zeta_\nu^+|^{2\ell_0} \geq \frac{1}{C} |I| \frac{\log N}{N} e^{-2C_\alpha \ell_0 (1 + O(\frac{\log N}{N})) \frac{\log N}{N}}. \quad (10.41)$$

When  $N/(C \log N) \leq |I| \leq N$  for some  $C \gg 1$ , then we see by Taylor expansion that

$$\frac{1 - |\zeta_\nu^+|^{2(|I|-1-m_+^0)}}{1 - |\zeta_\nu^+|^{2m_+^0}} \asymp \frac{1}{C} \frac{N}{\log N}.$$

We then deduce similarly to (10.41) that there exists a constant  $C > 0$  such that for  $N > 0$  large enough and all  $z \in \Omega_N$ ,

$$\sum_{j=0}^{\lfloor (|I|-1)/m_+^0 \rfloor - 1} \frac{|\zeta_\nu^+|^{2(\ell_0 + jm_+^0)}}{\|\mathfrak{z}_\nu^+\|^2} \geq \frac{1}{C} e^{-2C_\alpha \ell_0 (1 + O(\frac{\log N}{N})) \frac{\log N}{N}}. \quad (10.42)$$

From (10.24) we know that for  $N > 0$  sufficiently large  $\|b_j\| \geq 1/2$ . Hence, (10.41), (10.42) together with (10.38) imply (10.26), which concludes the proof of the theorem.  $\square$

**10.3. Large singular values.** The next result gives estimates on the Hilbert-Schmidt and trace norm of  $E$  from (2.9).

**Proposition 10.9.** *Let  $z_0 \in p(S^1)$  be as in (9.1) and let  $\Omega \Subset \mathbb{C}$  be a sufficiently small open relatively compact convex neighbourhood of  $z_0$  satisfying (9.2). Let  $N > 1$  be sufficiently large, and*

let  $\Omega_N \in \Omega \setminus p(S^1)$  satisfy (9.5). Let  $d$  be as in (9.4) and let  $t_j$  be as in Proposition 10.1, then, uniformly in  $N$  and  $z \in \Omega_N$ ,

$$\sum_{j=|d|+1}^N t_j^{-2} = O\left(\frac{N^2}{\log N}\right), \quad (10.43)$$

and

$$\sum_{j=|d|+1}^N t_j^{-1} = O(N \log N). \quad (10.44)$$

*Proof.* 1. Let  $z_0$  and  $\Omega$  as in the assumptions of the proposition and let  $z \in \Omega_N$ . Let  $\tilde{N} = N_+ \vee 0 + N_- \vee 0$  and notice that  $\tilde{N} \geq |d|$ , see Lemma 4.1 and (4.34). By (4.20), (10.15) and (1.13), we see that

$$t_j(P_N - z) \geq t_{j-\tilde{N}}(P_{\mathbb{Z}/N\mathbb{Z}} - z), \quad j = 1 + \tilde{N}, \dots, N. \quad (10.45)$$

From (4.20) we deduce that the set of eigenvalues of  $(P_{\mathbb{Z}/N\mathbb{Z}} - z)^*(P_{\mathbb{Z}/N\mathbb{Z}} - z)$  is given by

$$\{t_j^2(P_{\mathbb{Z}/N\mathbb{Z}} - z); j = 1, \dots, N\} = \{|p(\hat{\zeta}) - z|^2; \hat{\zeta} \in S_N\}, \quad (10.46)$$

where  $S_N \stackrel{\text{def}}{=} \{e^{2\pi i j/N}; j = 0, \dots, N-1\}$ . Note that  $t_1(P_{\mathbb{Z}/N\mathbb{Z}} - z) \geq \text{dist}(z, p(S^1)) > 0$ , since  $z \in \Omega_N$ . Hence, combining (10.45) and (10.2), we get for  $\nu = 1, 2$ ,

$$\sum_{j=|d|+1}^N \frac{1}{t_j^\nu(P_N - z)} \leq \sum_{j=1+\tilde{N}}^N \frac{1}{t_j^\nu(P_N - z)} + \tilde{N} \left(\frac{CN}{\log N}\right)^\nu \leq \sum_{j=1}^N \frac{1}{t_j^\nu(P_{\mathbb{Z}/N\mathbb{Z}} - z)} + \tilde{N} \left(\frac{CN}{\log N}\right)^\nu. \quad (10.47)$$

It remains to deal with the traces  $\text{tr}((P_{\mathbb{Z}/N\mathbb{Z}} - z)^*(P_{\mathbb{Z}/N\mathbb{Z}} - z))^{-\nu/2}$  for  $\nu = 1, 2$ .

2. We know by (4.16) that there are finitely many, say  $N_0 < +\infty$ , points in  $p^{-1}(z_0)$ . We will enumerate them in the following way: for  $j = 1, \dots, N'_0$ , with  $N'_0 \geq 1$ , let  $\zeta_j \in p^{-1}(z_0)$  be so that  $\zeta_j \in S^1$ , and for  $j = N'_0 + 1, \dots, N_0$ , let  $\zeta_j \in p^{-1}(z_0)$  be so that  $\zeta_j \notin S^1$ . If  $N'_0 = N_0$  then the latter set of points is empty.

By (9.2) and the Implicit function theorem, it follows that, after potentially shrinking  $\Omega$  (while keeping it convex), there exist complex open neighbourhoods  $Z_j$  of  $\zeta_j$ , such that  $p : Z_j \rightarrow \Omega$  is a diffeomorphism and such that  $Z_j \cap S^1 = \emptyset$  for  $j = N'_0 + 1, \dots, N_0$ . Furthermore, after potentially further shrinking  $\Omega$  and the sets  $Z_j$ , we can arrange so that  $p(S^1 \cap Z_j) = \Omega \cap p(S^1) = \Gamma$  for all  $j = 1, \dots, N'_0$ . Let  $\Omega' \subset \Omega$  be a smaller open convex neighbourhood of  $z_0$  such that  $\overline{\Omega'} \subset \Omega$ , and put  $Z'_j := (p|_{Z_j \rightarrow \Omega})^{-1}(\Omega') \subset Z_j$ . A convenient choice is

$$\Omega' = \{z \in \mathbb{C}; \text{dist}(z, \Gamma') \leq 1/C_{\Omega'}, \pi(z) \in \Gamma'\},$$

where  $0 < C_{\Omega'} < \infty$  is some sufficiently large constant,  $\Gamma' \subset \Gamma$  is a connected non-empty subsegment, containing  $z_0$ , so that the endpoints of the segments  $\Gamma$  and  $\Gamma'$  are separated by some fixed positive distance, i.e. for some large constant  $C' > 0$  we have that  $\text{dist}(\partial\Gamma, \partial\Gamma') \geq 1/C'$ . Furthermore,  $\pi(z)$  denotes the point in  $p(S^1)$  with  $|\Pi(z) - z| = \text{dist}(z, p(S^1))$ . Notice that this point is unique if  $C_{\Omega'}$  is large enough.

We note that there exists a constant  $C > 0$  (depending only on  $\Omega$  and  $\Omega'$ ) such that for all  $z \in \Omega'$  and all  $\zeta \in S^1 \setminus \bigcup_{j=1}^{N'_0} (Z_j \cap S^1)$

$$|p(\zeta) - z| \geq \text{dist}(\partial\Omega, \partial\Omega') \geq 1/C. \quad (10.48)$$

Furthermore, observe that since two consecutive points in  $S_N$  differ by an angle of  $2\pi/N$ , we have that

$$\#\{S_N \cap Z_j\} = N \int_{(p|_{Z_j \rightarrow \Omega})^{-1}(\Omega')} L_{S^1}(d\theta) + O(1), \quad j = 1, \dots, N'_0, \quad (10.49)$$

where  $L_{S^1}$  denotes the normalized Lebesgue measure on  $S^1$ .

Let  $j \in \{1, \dots, N'_0\}$ ,  $z \in \Omega'_N := \Omega_N \cap \Omega'$ , and  $\zeta_0^j \in Z'_j$  be the unique point in  $Z'_j$  such that  $p(\zeta_0^j) = \Pi(z)$ . For  $N > 1$  sufficiently large (depending only on  $\Omega$  and  $\Omega'$ ) there exists a point  $\widehat{\zeta}_*^j \in Z_j \cap S_N$  which is the closest point in  $Z_j \cap S_N$  to  $\zeta_0^j$ . There may be two such points, in which case we pick one of them. Then,

$$|\zeta_0^j - \widehat{\zeta}_*^j| \leq |1 - e^{2\pi i/N}| \leq O(N^{-1}). \quad (10.50)$$

By Taylor expansion around  $\zeta_0^j$ , we find that

$$|p(\widehat{\zeta}_*^j) - z| \leq |p(\zeta_0^j) - z| + O(N^{-1}),$$

where the constant is independent of  $z$ . In view of (10.46), we have that

$$t_1(P_{\mathbb{Z}/N\mathbb{Z}} - z) = \min_{\widehat{\zeta} \in S_N} |p(\widehat{\zeta}) - z|.$$

Setting  $\tilde{t} := |p(\zeta_0^j) - z|$ , as  $p(\zeta_0^j) = \Pi(z)$  and hence  $|p(\zeta_0^j) - z| = \text{dist}(z, p(S^1))$ , we see that

$$\tilde{t} \leq t_1(P_{\mathbb{Z}/N\mathbb{Z}} - z) \leq \min_{j=1, \dots, N'_0} |p(\widehat{\zeta}_*^j) - z| \leq \max_{j=1, \dots, N'_0} |p(\widehat{\zeta}_*^j) - z| \leq \tilde{t} + O(N^{-1}), \quad (10.51)$$

where the constant is independent of  $z$ . By Taylor expansion and (9.2) we get that for  $\widehat{\zeta}^j \in Z_j$

$$\begin{aligned} |p(\widehat{\zeta}^j) - z| &= |p(\zeta_0^j) + (\partial_\zeta p)(\zeta_0^j)(\widehat{\zeta}^j - \zeta_0^j) + O(|\widehat{\zeta}^j - \zeta_0^j|^2) - z| \\ &\geq \frac{1}{C} |\widehat{\zeta}^j - \zeta_0^j| (1 - C_1 |\widehat{\zeta}^j - \zeta_0^j|) - \tilde{t}. \end{aligned} \quad (10.52)$$

Here the constant  $0 < C < \infty$  only depends on  $\Omega$ , see (9.2), and the constant  $0 < C_1 < \infty$  depends only on  $C$  and the second derivative of  $p$  in the  $Z_j$ 's, which is uniformly bounded. Suppose that

$$|\widehat{\zeta}^j - \zeta_0^j| \leq \frac{1}{2C_1}, \quad |\widehat{\zeta}^j - \zeta_0^j| \geq 8C\tilde{t}, \quad (10.53)$$

where  $C, C_1$  are as in (10.52). Hence, we find by (10.52), (10.53) that

$$|p(\widehat{\zeta}^j) - z| \geq \frac{1}{4C} |\widehat{\zeta}^j - \zeta_0^j|. \quad (10.54)$$

3. We turn to proving the two estimates claimed in the proposition. We let  $\Omega$  and  $\Omega'$  be as in Step 2. Recall (10.51) and note that if  $\tilde{t} \geq 1/C$  for some arbitrarily large but fixed constant  $C$ , then

$$\sum_{j=1}^N \frac{1}{t_j(P_{\mathbb{Z}/N\mathbb{Z}} - z)}, \quad \sum_{j=1}^N \frac{1}{t_j^2(P_{\mathbb{Z}/N\mathbb{Z}} - z)} = O(N),$$

and we conclude (10.43) and (10.44) in view of (10.47).

Continuing, we may assume from now on that  $\tilde{t} \leq 1/C_0$ , for some arbitrarily large but fixed constant  $C_0$ . Since the roots of  $p(\zeta) - z$  depend smoothly on  $z \in \Omega$ , see the discussion after (9.2), it follows from (9.5) that there exists a constant  $C > 0$  such that for all  $N > 0$ ,

$$\text{dist}(\Omega_N, p(S^1)) \geq \frac{1}{C} \frac{\log N}{N},$$

Hence, we may from now on assume that

$$\frac{1}{C} \frac{\log N}{N} \leq \tilde{t} \leq \frac{1}{C_0}. \quad (10.55)$$

for some large  $C > 0$ , independent of  $N > 0$  and  $z \in \Omega_N$ . Let  $\nu \in \{1, 2\}$ . Then, in view of (10.46), (10.48) we get

$$\sum_{j=1}^N \frac{1}{t_j^\nu(P_{\mathbb{Z}/n\mathbb{Z}} - z)} = \sum_{j=1}^{N'_0} \sum_{\widehat{\zeta} \in Z_j \cap S_N} \frac{1}{|p(\widehat{\zeta}) - z|^\nu} + O(N), \quad (10.56)$$

where the constant in the error term depends only on  $\Omega$  and  $\Omega'$ . Using (10.49), we see that by potentially shrinking  $\Omega$ , and the sets  $Z_j$  and  $\Omega'$  accordingly, we may suppose that the first estimate

in (10.53) holds for all  $\widehat{\zeta} \in Z_j \cap S_N$ ,  $j = 1, \dots, N'_0$ . Next, pick a  $j \in \{1, \dots, N'_0\}$ , while keeping in mind that  $N'_0 < \infty$ , and split the corresponding inner sum on the right hand side of (10.56) into two parts: one where the second inequality in (10.53) holds and one where it does not, so

$$\sum_{\widehat{\zeta} \in Z_j \cap S_N} \frac{1}{|p(\widehat{\zeta}) - z|^\nu} = \sum_{\substack{\widehat{\zeta} \in Z_j \cap S_N, \\ |\widehat{\zeta} - \zeta_0^j| \geq 8C\tilde{t}}} \frac{1}{|p(\widehat{\zeta}) - z|^\nu} + \sum_{\substack{\widehat{\zeta} \in Z_j \cap S_N, \\ |\widehat{\zeta} - \zeta_0^j| < 8C\tilde{t}}} \frac{1}{|p(\widehat{\zeta}) - z|^\nu}. \quad (10.57)$$

We begin with treating the second sum on the right hand side. Since the points in  $S_N$  differ by an angle of  $2\pi/N$ , we see that for  $N > 1$

$$\#\{\widehat{\zeta} \in Z_j \cap S_N; |\widehat{\zeta} - \zeta_0^j| < 8C\tilde{t}\} = O(N\tilde{t}),$$

where the constant depends only on the symbol  $p$ ,  $\Omega$  and the constant  $C$  from (10.52). Hence, by (10.46) and (10.51),

$$\sum_{\substack{\widehat{\zeta} \in Z_j \cap S_N, \\ |\widehat{\zeta} - \zeta_0^j| < 8C\tilde{t}}} \frac{1}{|p(\widehat{\zeta}) - z|^\nu} = O(N\tilde{t}^{1-\nu}). \quad (10.58)$$

Next, we deal with the first sum on the right hand side of (10.57). Since for this term (10.53) holds, we get by (10.54) that

$$\sum_{\substack{\widehat{\zeta} \in Z_j \cap S_N, \\ |\widehat{\zeta} - \zeta_0^j| \geq 8C\tilde{t}}} \frac{1}{|p(\widehat{\zeta}) - z|^\nu} \leq \sum_{\substack{\widehat{\zeta} \in Z_j \cap S_N, \\ |\widehat{\zeta} - \zeta_0^j| \geq 8C\tilde{t}}} \frac{(4C)^\nu}{|\widehat{\zeta} - \zeta_0^j|^\nu} \quad (10.59)$$

By (10.50) and (10.55) we find that there exist constants  $0 < C_2, C'_2 < \infty$  such that for  $N > 0$  large enough (depending only on the constants in the referenced inequalities) we have that for all  $\widehat{\zeta} \in Z_j \cap S_N$  with  $|\widehat{\zeta} - \zeta_0^j| \geq 8C\tilde{t}$  and for all  $z \in \Omega_N$ , such that (10.55) holds,

$$|\widehat{\zeta} - \zeta_0^j| \geq \frac{1}{C_2} |\widehat{\zeta} - \widehat{\zeta}_*^j| \geq \frac{1}{C'_2} \tilde{t}. \quad (10.60)$$

By rotational invariance, we may assume that  $\widehat{\zeta}_*^j = 1$ . Then, we have that  $\widehat{\zeta}_k = e^{2\pi ik/N}$ ,  $k = 1, \dots, M_1$  are the points in  $Z_k$  to the left of 1 and  $\widehat{\zeta}_k = e^{2\pi ik/N}$ ,  $k = -1, \dots, -M_2$  are the points in  $Z_k$  to the right of 1. By (10.49) it is clear that after potentially further shrinking  $\Omega$  and taking  $N > 0$  large enough, we have that  $M_1, M_2 \leq N/C_3$ , for  $C_3 \gg 1$ . Hence, we find that there exists a constant  $0 < C_4 < \infty$  such that

$$|\widehat{\zeta}_k - \zeta_*^j| = |e^{\pm 2\pi ik/N} - 1| \geq \frac{1}{C_4 N} |k|, \quad -M_2 \leq k \leq M_1, \quad k \neq 0. \quad (10.61)$$

On the other hand, we see by (10.55) that if  $|k| \leq B^{-1} \log N$ ,  $B \gg 1$ , then

$$|\widehat{\zeta}_k - \widehat{\zeta}_*^j| = |e^{\pm 2\pi ik/N} - 1| \ll \tilde{t}, \quad (10.62)$$

for  $N > 1$ . Combining (10.61), (10.62) with (10.59) and (10.60), we conclude that

$$\sum_{\substack{\widehat{\zeta} \in Z_j \cap S_N, \\ |\widehat{\zeta} - \zeta_0^j| \geq 8C\tilde{t}}} \frac{1}{|p(\widehat{\zeta}) - z|^\nu} \leq \sum_{\substack{k=-M_2 \\ |k| \geq B^{-1} \log N}}^{M_1} \frac{(4CC_2C_4N)^\nu}{|k|^\nu} \leq \begin{cases} O(N \log N), & \nu = 1 \\ O(N^2 (\log N)^{-1}), & \nu = 2. \end{cases}$$

Together with (10.47), (10.55), (10.56), (10.57) and (10.58), we conclude (10.43) and (10.44).  $\square$

**10.4. Smallest singular value and Hilbert-Schmidt norms.** In this short section we provide bounds on minimal singular values and Hilbert-Schmidt norms of matrices appearing in the Grushin problem. Recall the definition of  $E(\cdot)$  from (2.9).

**Lemma 10.10.** *Consider the setup and notation as in Proposition 10.1. Fix  $M \in \mathbb{N}$  and  $\gamma > 1$ . Assume that the entries of  $Q$  are i.i.d. with zero mean, unit variance, and finite fourth moment. If  $\gamma > 3/2$  then*

$$\mathbf{P} \left( \inf_{z \in \Omega_N} s_{\min}(I_N + N^{-\gamma}QE(z)) \leq 1/2 \right) \leq N^{-(\gamma-3/2)/2},$$

for all large  $N$ . Otherwise, if  $\gamma \in (1, 3/2]$  then given any  $\alpha_0 > 0$  there exists  $\beta = \beta(\alpha_0, \gamma) < \infty$ , such that

$$\mathbf{P} \left( s_{\min}(I_N + N^{-\gamma}QE(z)) \leq N^{-\beta} \right) \leq N^{-\alpha_0},$$

uniformly for all  $z \in \Omega_N$  and all large  $N$ .

*Proof.* 1. We first consider the easier case  $\gamma > 3/2$ . By Proposition 10.1, applied with  $M = |d|$ , we have that  $\|E(z)\| = O(N)$  uniformly for  $z \in \Omega_N$ . As the entries of  $Q$  have zero mean and finite fourth moments, applying [42, Theorem 2] and Markov's inequality we obtain that

$$\mathbf{P} \left( N^{-\gamma} \sup_{z \in \Omega_N} \|QE(z)\| \geq 1/2 \right) \leq \mathbf{P} (\|Q\| \geq cN^{\gamma-1}) \leq c^{-1}N^{-(\gamma-1)} \cdot \mathbb{E}\|Q\| \leq N^{-(\gamma-3/2)/2},$$

for all large  $N$ , and some  $c > 0$ . Thus, by the triangle inequality we derive that

$$\mathbf{P} \left( \inf_{z \in \Omega_N} s_{\min}(I_N + N^{-\gamma}QE(z)) \leq 1/2 \right) \leq \mathbf{P} \left( N^{-\gamma} \sup_{z \in \Omega_N} \|QE(z)\| \geq 1/2 \right) \leq N^{-(\gamma-3/2)/2}.$$

This yields the result for  $\gamma > 3/2$ .

2. We turn to the case of  $\gamma \in (1, 3/2]$ . The proof uses a change of basis, a lower bound on the smallest singular value of matrices with i.i.d. entries shifted by a deterministic matrix, and an upper bound on the maximal singular value of a matrix with i.i.d. entries.

To carry out these steps we introduce additional notation. Let  $\mathbf{E}(z)$  and  $\mathbf{F}(z)$  be the two unitary matrices whose columns are  $\{e_1(z), e_2(z), \dots, e_N(z)\}$  and  $\{f_1(z), f_2(z), \dots, f_N(z)\}$ , respectively (Recall the definitions of  $e_i$ 's and  $f_j$ 's from Section 2.1 applied to  $P = P_N$ ). Further define  $\mathbf{E}_1(z)$  and  $\mathbf{E}_2(z)$  to be the  $N \times M$  and  $N \times (N - M)$  matrices consisting of the first  $M$  columns and the remaining  $(N - M)$  columns of  $\mathbf{E}(z)$ , respectively. Similarly define  $\mathbf{F}_1(z)$  and  $\mathbf{F}_2(z)$ . Finally, let  $\mathbf{T}(z)$  be the diagonal matrices with entries  $\{t_{M+1}(z)^{-1}, t_{M+2}(z)^{-1}, \dots, t_N(z)^{-1}\}$ , where  $t_j(z) = t_j$ ,  $j \in [N]$  are as in Proposition 10.1.

Equipped with this notation, upon recalling the definition of  $E(z)$ , we see that

$$I_N + N^{-\gamma}QE(z) = I_N + N^{-\gamma}QE_2(z)\mathbf{T}(z)\mathbf{F}_2(z)^*.$$

Fix  $x \in \mathbb{S}^{N-1}$ . As the columns of  $\mathbf{F}(z)$  form a basis of  $\mathbb{C}^N$ , there exist  $c_1 \in \mathbb{C}^M$  and  $c_2 \in \mathbb{C}^{N-M}$  such that

$$x = \mathbf{F}_1(z)c_1 + \mathbf{F}_2(z)c_2 \quad \text{and} \quad \|c_1\|^2 + \|c_2\|^2 = 1.$$

Therefore, using the orthonormality property of the columns of  $\mathbf{F}(z)$  we derive that

$$\begin{aligned} \|(I_N + N^{-\gamma}QE(z))x\|^2 &= \|\mathbf{F}_1(z)c_1 + \mathbf{F}_2(z)c_2 + N^{-\gamma}QE_2(z)\mathbf{T}(z)c_2\|^2 \\ &= \|\mathbf{F}_1(z)(c_1 + N^{-\gamma}\mathbf{F}_1(z)^*QE_2(z)\mathbf{T}(z)c_2) + \mathbf{F}_2(z)(I_{N-M} + N^{-\gamma}\mathbf{F}_2(z)^*QE_2(z)\mathbf{T}(z))c_2\|^2 \\ &= \|c_1 + N^{-\gamma}\mathbf{F}_1(z)^*QE_2(z)\mathbf{T}(z)c_2\|^2 + \|(I_{N-M} + N^{-\gamma}\mathbf{F}_2(z)^*QE_2(z)\mathbf{T}(z))c_2\|^2 \\ &= \left\| \begin{bmatrix} I_M & N^{-\gamma}\mathbf{F}_1(z)^*QE_2(z)\mathbf{T}(z) \\ 0 & I_{N-M} + N^{-\gamma}\mathbf{F}_2(z)^*QE_2(z)\mathbf{T}(z) \end{bmatrix} \cdot \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} \right\|^2, \end{aligned} \quad (10.63)$$

where in the second step we have used that  $\mathbf{F}_1(z)\mathbf{F}_1(z)^* + \mathbf{F}_2(z)\mathbf{F}_2(z)^* = I_N$  and in the third step we have again used the fact that the columns of  $\mathbf{F}(z) = [\mathbf{F}_1(z) \ \mathbf{F}_2(z)]$  form an orthonormal basis of  $\mathbb{C}^N$ . So (10.63) shows that it is enough to find a lower bound on the smallest singular value of the  $2 \times 2$  block matrix appearing in its RHS.

To this end, we note that any block matrix  $A$  of the form  $A = \begin{bmatrix} I & A_1 \\ 0 & A_2 \end{bmatrix}$  is invertible iff  $A_2$  is, and in that case

$$A^{-1} = \begin{bmatrix} I & -A_1A_2^{-1} \\ 0 & A_2^{-1} \end{bmatrix}.$$

Hence, using the triangle inequality we obtain that

$$\|A^{-1}\| \leq 1 + (\|A_1\| + 1) \cdot \|A_2^{-1}\|.$$

Recalling the fact that  $\|c_1\|^2 + \|c_2\|^2 = 1$ , using the above and (10.63) we therefore deduce that

$$\begin{aligned} (s_{\min}(I_N + N^{-\gamma}QE(z)))^{-1} &= \|(I_N + N^{-\gamma}QE(z))^{-1}\| \\ &\leq 1 + (N^{-\gamma}\|A_1(z)\| + 1) \cdot \|(I_{N-M} + N^{-\gamma}A_2(z))^{-1}\|, \end{aligned} \quad (10.64)$$

where for brevity we write

$$A_1(z) := F_1(z)^*QE_2(z)T(z) \quad \text{and} \quad A_2(z) := F_2(z)^*QE_2(z)T(z).$$

Using [42], Proposition 10.1, and Markov's inequality again we further have that

$$\mathbf{P}\left(N^{-\gamma}\|A_1(z)\| + 1 \geq 2N^{2\alpha_0+1/2}\right) \leq t_{M+1}(z)^{-1} \cdot N^{-(\gamma+2\alpha_0+1/2)} \cdot \mathbb{E}\|Q\| = O(N^{-2\alpha_0}), \quad (10.65)$$

where we have also used the facts that  $t_{M+1}(z)^{-1} = O(N)$  and  $\gamma > 1$ . Hence, by (10.64), we observe that in order to find a lower bound on the smallest singular value of  $I_N + N^{-\gamma}QE(z)$  it remains to find the same for  $I_{N-M} + N^{-\gamma}A_2(z)$ . Upon using Weyl's inequality for singular values it is straightforward to notice that

$$s_{\min}(I_{N-M} + N^{-\gamma}A_2(z)) \geq s_{\min}(I_N + N^{-\gamma}A(z)), \quad (10.66)$$

where

$$A(z) := F(z)^*QE(z)\widehat{T}(z) \quad \text{and} \quad \widehat{T}(z) := \begin{bmatrix} I_M & 0 \\ 0 & T(z) \end{bmatrix}.$$

On the other hand

$$\begin{aligned} s_{\min}(I_N + N^{-\gamma}A(z)) &\geq s_{\min}(\widehat{T}(z)) \cdot s_{\min}(\widehat{T}(z)^{-1} + N^{-\gamma}F(z)^*QE(z)) \\ &= t_N(z)^{-1} \cdot N^{-\gamma} \cdot s_{\min}(N^{\gamma}F(z)\widehat{T}(z)^{-1}E(z)^* + Q) \end{aligned} \quad (10.67)$$

Since  $P_N$  has a bounded norm and  $\Omega_N$  is a bounded set we have that  $\|\widehat{T}(z)^{-1}\| = t_N(z) = O(1)$ . As  $E(z)$  and  $F(z)$  are unitary matrices, an application of [69, Theorem 2.1] together with (10.67) now yield that

$$\mathbf{P}\left(s_{\min}(I_N + N^{-\gamma}A(z)) \leq N^{-\beta_0}\right) \leq N^{-2\alpha_0}, \quad (10.68)$$

for all large  $N$ , where  $\beta_0 < \infty$  is some constant depending only on  $\alpha_0$  and  $\gamma$ . Finally, using (10.64) and (10.66), the probability bounds (10.65) and (10.68), and a union bound we deduce that

$$\mathbf{P}\left(s_{\min}(I_N + N^{-\gamma}QE(z)) \leq \frac{1}{3}N^{-(2\alpha_0+\beta_0+1/2)}\right) = O(N^{-2\alpha_0}),$$

for all large  $N$ . This completes the proof of the lemma.  $\square$

Next we derive bounds on the Hilbert-Schmidt norms of some matrices. This will allow us to control the error terms in the resolvent expansion of various terms appearing in the Grushin problem.

**Proposition 10.11.** *Consider the setup and notation as in Proposition 10.1. Additionally assume that  $\Omega$  is convex. Fix  $\gamma > 1$ ,  $\kappa \in \mathbb{N}$  and  $d \in [-\tilde{m}, \tilde{m}] \cap \mathbb{Z} \setminus \{0\}$ . Let Assumption 1.1 holds. Then for any  $\alpha_0 > 0$ , we have, uniformly in  $z \in \Omega_N$  and all  $N$  large, that*

$$\mathbf{P}\left(N^{-\gamma\kappa}\|(E(z)Q)^\kappa E_+(z)\|_{\text{HS}} \geq N^{-\frac{(\gamma-1)\kappa}{2}}\right) \leq N^{-\alpha_0}, \quad (10.69)$$

$$\mathbf{P}\left(N^{-\gamma\kappa}\|(E(z)Q)^\kappa E(z)\|_{\text{HS}} \geq N^{1-\frac{(\gamma-1)\kappa}{2}}\right) \leq N^{-\alpha_0}, \quad (10.70)$$

$$\mathbf{P}\left(N^{-\gamma\kappa}\|E_-(z)(QE(z))^\kappa\|_{\text{HS}} \geq N^{-\frac{(\gamma-1)\kappa}{2}}\right) \leq N^{-\alpha_0}, \quad (10.71)$$

and

$$\mathbf{P}\left(N^{-\gamma(\kappa+1)}\|E_-(z)(QE(z))^\kappa QE_+(z)\|_{\text{HS}} \geq N^{-\gamma-\frac{(\gamma-1)\kappa}{2}}\right) \leq N^{-\alpha_0}. \quad (10.72)$$

To prove Proposition 10.11 we will rely on bounds on moments of Hilbert-Schmidt norm of matrices appearing in (10.69)-(10.72), as well as bounds on  $\|E_\pm(\cdot)\|_{\text{HS}}$  and on the sum of the inverse of the non-zero singular values of  $E(\cdot)$ .



**Proposition 10.12.** Fix  $\kappa \in \mathbb{N}$ . Let  $\{\mathbf{E}^{(\ell)}\}_{\ell=1}^{\kappa}$  and  $\{\mathbf{F}^{(\ell)}\}_{\ell=1}^{\kappa}$  be two collections of  $N \times N$  deterministic matrices, possibly with complex-valued entries, such that

$$\max \left\{ \max_{\ell=1}^{\kappa} \|\mathbf{E}^{(\ell)}\|_{\text{HS}}, \max_{\ell=1}^{\kappa} \|\mathbf{F}^{(\ell)}\|_{\text{HS}} \right\} \leq 1. \quad (10.73)$$

Define the product

$$\widehat{\mathbf{E}}(\kappa) := \mathbf{E}^{(1)}Q\mathbf{F}^{(1)} \cdot \mathbf{E}^{(2)}Q\mathbf{F}^{(2)} \dots \mathbf{E}^{(\kappa)}Q\mathbf{F}^{(\kappa)}. \quad (10.74)$$

If the entries of  $Q$  satisfies Assumption 1.1 then for any  $h_0 \in \mathbb{N}$

$$\mathbf{E}\|\widehat{\mathbf{E}}(\kappa)\|_{\text{HS}}^{2h_0} \leq C_{10.12} \cdot \mathfrak{C}_{\kappa h_0},$$

where  $C_{10.12} < \infty$  is some constant depending only on  $\kappa$  and  $h_0$ .

The proof of Proposition 10.12 is straightforward. It relies on bounds on moments of quadratic forms of independent random variables, derived in [75].

*Proof of Proposition 10.12.* We begin with an auxiliary computation. Let  $A, B$  be deterministic matrices. Then

$$W := \|AQB\|_{\text{HS}}^2 = \sum_{i,j,k,k',\ell,\ell'} A_{i,k} \overline{A_{i,k'}} Q_{k,\ell} \overline{Q_{k',\ell'}} B_{\ell,j} \overline{B_{\ell',j}} = \sum_{k,\ell,k',\ell'} Q_{k\ell} \overline{Q_{k'\ell'}} C_{k,\ell,k',\ell'},$$

where

$$C_{k,\ell,k',\ell'} := \sum_{i,j} A_{ik} \overline{A_{ik'}} B_{\ell,j} \overline{B_{\ell',j}}.$$

This, in particular, shows that  $W$  is a quadratic form in the entries of  $Q$ . Hence, using [75, Theorem 2]<sup>1</sup> and Assumption 1.1, for any  $s \geq 2$ , we find that

$$\mathbf{E}|W - \mathbf{E}W|^s \leq C(s) \left( \sum_{k,k',\ell,\ell'} |C_{k,\ell,k',\ell'}|^2 \right)^{s/2} \cdot \mathfrak{C}_s \leq C(s) \|A\|_{\text{HS}}^{2s} \|B\|_{\text{HS}}^{2s} \cdot \mathfrak{C}_s,$$

for some absolute constant  $C(s) < \infty$ , depending only on  $s$ , where in the last step we used that

$$|C_{k,\ell,k',\ell'}| \leq \|a_k\| \cdot \|a_{k'}\| \cdot \|b_\ell\| \cdot \|b_{\ell'}\|,$$

and  $a_k$  and  $b_k$  are the  $k$ -th column and row of  $A$  and  $B$ , respectively.

On the other hand we note that

$$|\mathbf{E}W| \leq \mathfrak{C}_1 \sum_{k,\ell} \|a_k\|^2 \|b_\ell\|^2 = \|A\|_{\text{HS}}^2 \|B\|_{\text{HS}}^2 \cdot \mathfrak{C}_1.$$

Thus

$$\mathbf{E}|W|^s \leq 2^{s-1} (\mathbf{E}|W - \mathbf{E}W|^s + |\mathbf{E}W|^s) \leq 2^s C(s) \mathfrak{C}_s \cdot \|A\|_{\text{HS}}^{2s} \|B\|_{\text{HS}}^{2s}, \quad (10.75)$$

where we also used that by Jensen's inequality it follows  $\mathfrak{C}_1^s \leq \mathfrak{C}_s$ . Next by Hölder's inequality we obtain that

$$\mathbf{E}\|\widehat{\mathbf{E}}(\kappa)\|_{\text{HS}}^{2h_0} \leq \mathbf{E} \left( \prod_{j=1}^{\kappa} \|\mathbf{E}^{(j)}Q\mathbf{F}^{(j)}\|_{\text{HS}}^{2h_0} \right) \leq \left( \prod_{j=1}^{\kappa} \mathbf{E}\|\mathbf{E}^{(j)}Q\mathbf{F}^{(j)}\|_{\text{HS}}^{2\kappa h_0} \right)^{1/\kappa} \leq \max_{j=1}^{\kappa} \mathbf{E}\|\mathbf{E}^{(j)}Q\mathbf{F}^{(j)}\|_{\text{HS}}^{2\kappa h_0}.$$

Now the desired bound follows from (10.75), upon setting  $s = \kappa h_0$ ,  $A = \mathbf{E}^{(j)}$ , and  $B = \mathbf{F}^{(j)}$ , for  $j \in [\kappa]$ . This concludes the proof.  $\square$

Equipped with Proposition 10.12, we now prove Proposition 10.11.

<sup>1</sup>Whittle in [75] works with *real* linear and quadratic forms of *real* valued random variables. Upon splitting the sums into real and imaginary parts, analogous bounds can be derived in the complex case. Hence, without loss of generality we continue to use [75] for complex quadratic forms of complex valued random variables.

*Proof of Proposition 10.11.* Recall that by Proposition 10.9, there exists a constant  $C_{10.9}$  so that

$$\sum_{j=|d|+1}^N t_j(z)^{-2} \leq C_{10.9} \frac{N^2}{\log N} \quad \text{and} \quad \sum_{j=|d|+1}^N t_j(z)^{-1} \leq C_{10.9} N \log N. \quad (10.76)$$

For  $j \in [\kappa]$  we let

$$\mathbf{F}^{(j)} := \frac{1}{\sqrt{C_{10.9} N \log N}} \sum_{i=|d|+1}^N \frac{1}{\sqrt{t_i(z)}} e_i(z) \circ e_i^*(z)$$

and

$$\mathbf{E}^{(j+1)} := \frac{1}{\sqrt{C_{10.9} N \log N}} \sum_{i=|d|+1}^N \frac{1}{\sqrt{t_i(z)}} e_i(z) \circ f_i^*(z).$$

Further let

$$\mathbf{E}^{(1)} := |d|^{-1/2} E_-(z) \quad \text{and} \quad \mathbf{F}^{(\kappa+1)} := |d|^{-1/2} E_+(z).$$

By Proposition 10.9 and (2.9)-(2.10) it follows that (10.73) holds with these choices of  $\{\mathbf{E}^{(j)}\}_{j=1}^{\kappa+1}$  and  $\{\mathbf{F}^{(j)}\}_{j=1}^{\kappa+1}$ . On the other hand note that  $E(z) = C_{10.9} N \log N \cdot \mathbf{F}^{(j)} \mathbf{E}^{(j+1)}$  for  $j \in [\kappa]$ . Therefore, applying Proposition 10.12 we now derive that

$$\mathbf{E} \left[ \|\tilde{\mathbf{E}}\|_{\text{HS}}^{2h_0} \right] \leq d^{2h_0} \cdot (C_{10.9} N \log N)^{2h_0 \kappa} \cdot C_{10.12} \cdot \mathfrak{C}_{\kappa h_0}.$$

where  $\tilde{\mathbf{E}} := E_-(z)(QE(z))^\kappa QE_+(z)$ . Hence applying Markov's inequality with  $h_0 = \lceil \frac{2\alpha_0}{\kappa(\gamma-1)} \rceil$  we obtain that

$$\mathbf{P} \left( N^{-\gamma(\kappa+1)} \|E_-(z)(QE(z))^\kappa QE_+(z)\|_{\text{HS}} \geq N^{-\gamma - \frac{(\gamma-1)\kappa}{2}} \right) \leq N^{-(\gamma+1)\kappa h_0} \mathbf{E} \left[ \|\tilde{\mathbf{E}}\|_{\text{HS}}^{2h_0} \right] \leq N^{-\alpha_0},$$

for all large  $N$ . This concludes the proof of (10.72). To obtain (10.71) we let

$$\mathbf{E}_\star := \mathbf{E}^{(1)} Q \mathbf{F}^{(1)} \cdot \mathbf{E}^{(2)} Q \mathbf{F}^{(2)} \dots \mathbf{E}^{(\kappa-1)} Q \mathbf{F}^{(\kappa-1)} \cdot \mathbf{E}^{(\kappa)} Q \mathbf{F}_\star^{(\kappa)},$$

where  $\{\mathbf{E}^{(i)}\}_{i=1}^{\kappa}$  and  $\{\mathbf{F}^{(i)}\}_{i=1}^{\kappa-1}$  are as above, and  $\mathbf{F}^{(\kappa)} := C_{10.9}^{-1/2} N^{-1} (\log N)^{1/2} \cdot E(z)$ . Notice that, by (10.76), we have  $\|\mathbf{F}_\star^{(\kappa)}\|_{\text{HS}} \leq 1$ . Therefore, applying Proposition 10.12 again we obtain that

$$\mathbf{E} \left[ \|\mathbf{E}_\star\|_{\text{HS}}^{2h_0} \right] \leq d^{h_0} \cdot (C_{10.9} N \log N)^{2h_0 \kappa} \cdot C_{10.12} \cdot \mathfrak{C}_{\kappa h_0}.$$

Now (10.71) follows by choosing  $h_0$  as above and applying Markov's inequality. The proofs of (10.69)-(10.70) are similar. For (10.69) we take  $\mathbf{E}^{(1)} = E(z)/(C_{10.9} N \log N)$  and  $\mathbf{F}^{(\kappa+1)} = E_+(z)/\sqrt{|d|}$ , while for (10.70) we take  $\mathbf{F}^{(\kappa+1)} = E(z)/(C_{10.9} N \log N)$ . Further details are omitted.  $\square$

The next corollary will come handy in the proof of Theorem 1.6.

**Corollary 10.13.** *In the setup of Proposition 10.11, we have that for any  $\alpha_0 > 0$ , for all  $N$  large,*

$$\mathbf{P} \left( N^{-\gamma} \|E(z)(I + N^{-\gamma} QE(z))^{-1} QE_+(z)\|_{\text{HS}} \geq N^{-(\gamma-1)/4} \right) \leq N^{-\alpha_0}, \quad (10.77)$$

$$\mathbf{P} \left( N^{-2\gamma} \|E_-(z)(I + N^{-\gamma} QE(z))^{-1} QE(z) QE_+(z)\|_{\text{HS}} \geq N^{-\gamma - (\gamma-1)/4} \right) \leq N^{-\alpha_0}, \quad (10.78)$$

$$\mathbf{P} \left( \|E_-(z)(I + N^{-\gamma} QE(z))^{-1}\|_{\text{HS}} > 2\sqrt{M} \right) \leq N^{-\alpha_0}, \quad (10.79)$$

and

$$\mathbf{P} \left( \|E(z)(I + N^{-\gamma} QE(z))^{-1}\|_{\text{HS}} > 2C_{10.9} N / \sqrt{\log N} \right) \leq N^{-\alpha_0}. \quad (10.80)$$

*Proof.* The proof uses the resolvent expansion and Proposition 10.11. First let us prove (10.77).

Using the resolvent expansion, we write, with  $\hat{Q} = N^{-\gamma} Q$  and  $k$  positive integer and omitting throughout the argument  $z$ ,

$$(I + \hat{Q}E)^{-1} = \sum_{i=0}^{k-1} (-\hat{Q}E)^i + (-\hat{Q}E)^k (I + \hat{Q}E)^{-1}. \quad (10.81)$$

Considering the first term in the right hand side of (10.81), we have by (10.69) that, for  $i \geq 0$ ,

$$\|E(-\hat{Q}E)^i \hat{Q}E_+\|_{\text{HS}} \leq N^{-(\gamma-1)(i+1)}/2, \quad (10.82)$$

with probability  $1 - N^{-2\alpha_0}$ , for all  $N$  large. On the other hand, by (10.70) and the fact that for any two matrices  $A, B$  one has  $\|BA\|_{\text{HS}} \leq \|A\| \cdot \|B\|_{\text{HS}}$ , we obtain that

$$\|E(-\hat{Q}E)^k (I + \hat{Q}E)^{-1} \hat{Q}E_+\|_{\text{HS}} \leq N^{1 - \frac{(\gamma-1)k}{2}} \cdot \|(I + \hat{Q}E)^{-1}\| \cdot \|Q\| \cdot \|E_+\|, \quad (10.83)$$

on a set of event with probability at least  $1 - N^{-2\alpha_0}$ . Now by the second part of Lemma 10.10, [42] (which gives  $\mathbf{E}\|Q\| = O(N^{1/2})$ ), and (2.10), we further obtain that  $\|(I + \hat{Q}E)^{-1}\| \cdot \|Q\| \cdot \|E_+\| \leq N^\beta$  for some finite  $\beta$ , with probability at least  $1 - N^{-2\alpha_0}$ . By choosing  $k$  large enough, we conclude that on these events, the right hand side of (10.83) can be made arbitrarily small. Combining this with (10.81) and (10.82), we obtain (10.77). The proof of (10.78), being similar, is omitted.

To see (10.79), we use the resolvent expansion (10.81) and write

$$E_-(I + N^{-\gamma}QE)^{-1} = E_- + \sum_{i=1}^{k-1} E_-(\hat{Q}E)^i + E_-(\hat{Q}E)^k (I + \hat{Q}E)^{-1}.$$

Recall that  $\|E_-\|_{\text{HS}} = \sqrt{M}$ . As above, the other terms are controlled by (10.71) and Lemma 10.10. This completes the proof of (10.79). To prove (10.80) we observe that by (10.76) one has that  $\|E\|_{\text{HS}} \leq C_{10.9}N/\sqrt{\log N}$  (without loss of generality assume  $C_{10.9} \geq 1$ ). Therefore, arguing similarly as above one derives (10.80). Further details are omitted.  $\square$

**Remark 10.14.** *The reader may inspect the proofs of Proposition 10.11 and Corollary 10.13 to observe that (10.77) continues to hold even if we can replace the term  $N^{-(\gamma-1)/4}$  by  $N^{-(1-\epsilon)\cdot(\gamma-1)}$  for any  $\epsilon > 0$ . This improvement will be necessary later in Section 11 to argue that the eigenvectors of a randomly perturbed Jordan block is completely delocalized in sup-norm.*

**Remark 10.15.** *In Sections 9 and 10.1-10.4 we worked with Assumption 9.1. Since  $p(S^1)$  is compact, it follows that Assumption 9.1 holds uniformly in  $z \in \Omega_{\varepsilon_{1.5}, C_{1.5}, N}$  (recall (1.7)). This observation will be used in Section 11, in the proofs of Theorem 1.6 and Corollary 1.7.*

## 11. PROOFS OF THEOREM 1.6 AND COROLLARY 1.7

We start with the proof of Theorem 1.6.

**11.1. Proof of Theorem 1.6.** We begin by creating a net  $\mathcal{N}_\gamma$  in  $\Omega(\varepsilon_{1.5}, C_{1.5}, N)$ , of spacing  $N^{-\theta}$  and cardinality at most  $O(N^{2\theta-1} \log N)$ , with  $\theta > \gamma \vee 4$ . Note that for any  $\hat{z} \in \Omega(\varepsilon_{1.5}, C_{1.5}, N)$ , there exists a  $z^* = z^*(\hat{z}) \in \mathcal{N}_\gamma$  with  $|z^* - \hat{z}| \leq N^{-\theta}$ .

1. We begin with the proof of the first point in Theorem 1.6, and proceed as in the sketch of Section 3. Let  $\hat{z} \in \Omega(\varepsilon_{1.5}, C_{1.5}, N)$  be an eigenvalue of  $P_{N,\gamma}^Q$ , with associated right eigenvector  $v = v(\hat{z})$ . Recall that  $\delta = N^{-\gamma}$ . For any  $z \in \mathcal{N}_\gamma$ , define  $P_z^\delta := P_{N,\gamma}^Q - z$ , and note that since  $P_z^\delta - P_{\hat{z}}^\delta = (\hat{z} - z)I_N$ , we have that

$$P_z^\delta v = (\hat{z} - z)v. \quad (11.1)$$

Set  $M = |\text{ind}_{p(S^1)}(z)| = |d(z)|$  (this yields  $t_{M+1} \gtrsim \log N/N$  and  $\alpha$  bounded below by a constant multiple of  $\log N/N$ , see Proposition 10.1).

Consider the Grushin problem associated with  $P_z^\delta$ , see Section 2, and recall (3.1). We then obtain, similarly to (3.2) and using (11.1), that

$$\begin{aligned} \sum_{i=M+1}^N (e_i(z)^* v) \cdot e_i(z) &= (I - E_+(z)R_+(z))v \\ &= (I - E_+^\delta(z)R_+(z))v - E(z)(I + \delta QE(z)^{-1} \delta QE_+(z)R_+(z))v \\ &= E^\delta(z)(\hat{z} - z)v + E(z)(I + \delta QE(z)^{-1} \delta QE_+(z)R_+(z))v. \end{aligned} \quad (11.2)$$

We next control the right hand side of (11.2), for  $z = z^*$ . As discussed in Section 3, the control is simpler when  $\gamma > 3/2$ . Indeed, in that case, by (2.10) and Proposition 10.1 we have that  $\|E(z)\| = O(N/\log N)$  uniformly in  $\mathcal{N}_\gamma$  (see also Remark 10.15). Now by the first part of Lemma 10.10,

we have, with probability approaching 1 as  $N \rightarrow \infty$ , that for all  $z \in \mathcal{N}_\gamma$ ,  $\|(I + \delta QE(z))^{-1}\| \leq 2$ . Therefore, by (2.17),  $\|E^\delta(z)\| = O(N/\log N)$  uniformly in  $\mathcal{N}_\gamma$  with the same probability. Hence, for  $z = z^*$ ,

$$\|E^\delta(z^*)(\hat{z} - z^*)v\| \leq O(N/\log N) \cdot N^{-\theta} = o(1). \quad (11.3)$$

It remains to control the second term in (11.2).

By construction,  $\|E_+\|, \|R_+\| \leq 1$ , and  $\|Q\| = O(N^{1/2+\epsilon})$ , for any  $\epsilon > 0$ , with probability approaching 1 as  $N \rightarrow \infty$ . Applying Lemma 10.10 again we obtain that with  $z = z^*$ ,

$$\|E(z^*)(I + \delta QE(z^*))^{-1} \delta QE_+(z^*) R_+(z^*) v\| \leq N^{-\gamma} O(N^{1/2}) \cdot N/\log N = o(1). \quad (11.4)$$

Together with (11.3), we conclude from (11.2) that if  $\gamma > 3/2$  then for  $z^*$  with  $|z^* - \hat{z}| = O(N^{-\theta})$ ,

$$\left\| \sum_{i=M+1}^N (e_i(z^*)^* v) \cdot e_i(z^*) \right\| = o(1). \quad (11.5)$$

We turn next to the case  $\gamma \in (1, 3/2]$ . In that case, we apply Corollary 10.13 and obtain that for all  $z \in \mathcal{N}_\gamma$ , with probability larger than  $1 - N^{-\alpha_0}$ , we have that

$$\|E^\delta(z)\| = O(N/\sqrt{\log N}) \quad \text{and} \quad \|E(z)(I + \delta QE(z))^{-1} \delta QE_+(z) R_+(z) v\| = o(1). \quad (11.6)$$

Together with (11.3), this yields (11.5) also for  $\gamma \in (1, 3/2]$ .

For  $z \in \mathbb{C}$  we define

$$w_z := E_+(z) R_+(z) v. \quad (11.7)$$

We will show below that  $\|v - w_{z^*}\| = o(1)$  and  $\|w_{z^*} - w_{\hat{z}}\| = o(1)$ . This will yield (1.9) with  $w = w_{\hat{z}}$ . The localization estimate (1.10) then will follow from (10.25) and (10.27) in Theorem 10.8. The assertion that  $w$  is a random linear combination of  $\{e_j(\hat{z})\}_{j \in [M]}$  follows from its definition in (11.7).

Turning to proving that  $\|v - w_{z^*}\| = o(1)$ , we note that, by (11.2),  $v - w_{z^*}$  equals the left hand side of (11.2), whose norm is  $o(1)$  by (11.5).

It remains to show that  $\|w_{z^*} - w_{\hat{z}}\| = o(1)$ . This follows from a standard perturbative argument. Indeed, let  $B_z = (P_N - zI)^*(P_N - zI)$ . Using Proposition 10.1, the  $M$ -th singular value of both  $B_{\hat{z}}$  and  $B_{z^*}$  are  $O(N^{-C_\gamma} \log N/N)$  while the  $(M+1)$ -th is at least  $C \log N/N$ , for appropriate  $C, C_\gamma$  independent of  $N$ . Let  $\Pi_z = \Pi_{z,M}$  be the spectral projector of  $B_z$  on the first  $M$  eigenvalues. Then, choosing  $\Gamma = \partial D(0, C \log N/2N)$ , we can write

$$\Pi_z = \frac{1}{2\pi i} \int_{\Gamma} (B_z - \lambda)^{-1} d\lambda,$$

and an easy computation (using the Ky-Fan inequalities and the spectral gap) gives that  $\|\Pi_{\hat{z}} - \Pi_{z^*}\| = O(N^{-\theta} N^2/\log N) = o(N^{-2})$  since  $\theta > 4$ . By (11.7) we notice that  $w_z = \Pi_z v$ . We therefore obtain that  $\|w_{\hat{z}} - w_{z^*}\| = o(N^{-2}) = o(1)$ , as claimed. This completes the proof of the first point of Theorem 1.6, for all  $\gamma > 1$ .

2. We next turn to the proof of the second part of Theorem 1.6. Fix  $z_0 \in \Omega(\varepsilon_{1.5}, C_{1.5}, N)$  and let  $\hat{z}$  be an eigenvalue of  $P_{N,\gamma}^Q$  as in the second part of the theorem. Recall that  $z^* \in \mathcal{N}_\gamma$  is such that is such that  $|\hat{z} - z^*| \leq N^{-\theta}$ .

The starting point of the proof is the observation that by (3.1), (11.1), the definition of  $E_{-+}^\delta(z)$ , and the resolvent expansion, we have that for any  $z \in \mathbb{C}$ ,

$$E_{-+}^\delta(z)(\hat{z} - z)v = -E_{-+}^\delta(z) R_+(z) v \quad (11.8)$$

$$= -E_{-+}(z) R_+(z) v + \delta E_{-+}(z) Q E_+(z) R_+(z) v - \delta^2 E_{-+}(z) (I + \delta QE(z))^{-1} Q E(z) Q E_+(z) R_+(z) v.$$

(Compare with (3.3).) Using the definition of  $E_{-+}(z)$  and of  $R_+(z)$ , we rewrite the first two terms in the right hand side of (11.8) as

$$\begin{aligned} & -E_{-+}(z) R_+(z) v + \delta E_{-+}(z) Q E_+(z) R_+(z) v \\ &= \sum_{i=1}^M t_i \cdot (e_i(z)^* v) \cdot \delta_i + \delta \sum_{i=1}^M \left[ \sum_{j=1}^M (e_j(z)^* v) \cdot (f_i(z)^* Q e_j(z)) \right] \delta_i. \end{aligned} \quad (11.9)$$

Next, recall that  $E_-^\delta = E_-(I + \delta QE)^{-1}$ . Thus, using (10.79), we obtain that

$$\|E_-^\delta(z)(\hat{z} - z)v\| \leq 2\sqrt{M}|\hat{z} - z|, \quad (11.10)$$

with probability  $1 - N^{-2\alpha_0}$ , for any  $\alpha_0 > 0$ , when  $N$  is large enough. By (10.78), we also have with the same probability that

$$\|\delta^2 E_-(z)(I + \delta QE(z))^{-1}QE(z)QE_+(z)R_+(z)v\| \leq N^{-\gamma - (\gamma-1)/4}. \quad (11.11)$$

Upon choosing  $\alpha_0$  sufficiently large, by a union bound, (11.10) and (11.11) hold simultaneously for all  $z \in \mathcal{N}_\gamma$ . Thus, we have that with probability  $1 - N^{-3\alpha_0/2}$ ,

$$\left\| \sum_{i=1}^M t_i \cdot (e_i(z)^*v) \cdot \delta_i + \delta \sum_{i=1}^M \left[ \sum_{j=1}^M (e_j(z)^*v) \cdot (f_i(z)^*Qe_j(z)) \right] \delta_i \right\| \leq 2\sqrt{M}|\hat{z} - z| + N^{-\gamma - (\gamma-1)/4}. \quad (11.12)$$

3. Recall next Proposition 10.1, and that  $|d(z)| = M$  if  $z \in \mathcal{N}_\gamma \cap D(z_0, C_0 \log N/N)$ . In what follows, we assume that  $|z^* - \hat{z}| < N^{-\theta}$ , so that the above conditions hold for  $z = z^*$  and  $\hat{z}$ . Let  $M_0 = |d(z)| - m_{\text{sign}(d(z))}^0$ . Assume, for the time being, that  $M_0 > 0$ . Recall from (10.3) that if  $M_0 > 0$  then for  $j \in [M_0]$ ,  $|t_j(z)| \leq Ce^{-cN}$  for some  $c > 0$ .

Let  $A = A(z)$  denote the  $M_0 \times M$  matrix with entries (in the  $\delta_i$  basis)  $A_{ij} := A_{ij}(z) := f_i(z)^*Qe_j(z)$  and, for  $j \in [M]$ , set  $\alpha_j := e_j(z^*)^*v$ . Write  $\hat{A}$  for the  $M_0 \times M_0$  submatrix of  $A$  consisting of its first  $M_0$  columns. We obtain from (11.12) and (10.3) that

$$\left\| \sum_{i=1}^{M_0} \delta_i \sum_{j=1}^M A_{ij}(z^*)\alpha_j \right\| \leq N^\gamma (e^{-cN} + 2\sqrt{M}|z^* - \hat{z}| + N^{-\gamma - (\gamma-1)/4}) = o(1), \quad (11.13)$$

since  $\theta > \gamma$ . We have the following lemma, whose proof is postponed.

**Lemma 11.1.** *For every  $\eta > 0$ , there exist  $0 < c_\eta, C_\eta < \infty$  depending on  $\eta$  and  $C_0$  only, so that, with probability at least  $1 - \eta$ , for all  $z \in \mathcal{N}_\gamma \cap D(z_0, C_0 \log N/N)$ , we have that  $s_{\min}(\hat{A}(z)) > c_\eta$  and that  $s_{\max}(A(z)) < C_\eta$ .*

Continuing with the proof of Theorem 1.6, we obtain from (11.13) that with  $a = \sqrt{\sum_{j=1}^{M_0} |\alpha_j|^2}$ ,

$$\left\| \sum_{i=1}^{M_0} \delta_j \sum_{j=M_0+1}^M A_{ij}\alpha_j \right\| \geq ac_\eta - o(1). \quad (11.14)$$

Recall now that the norm of the projection of  $v$  on  $\text{span}(e_j, j \in [M])$  is  $1 - o(1)$  by the first part of the theorem. The orthogonality of the  $e_j$ 's then gives that  $\|\alpha\| = 1 + o(1)$ , where  $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_M)$ , and together with Lemma 11.1, (11.14), also that, with  $b = \sqrt{\sum_{j=M_0+1}^M |\alpha_j|^2}$ ,

$$C_\eta b \geq c_\eta a - o(1) = c_\eta \sqrt{1 - b^2} - o(1).$$

It follows that, with  $M_0 > 0$ , for some  $c_0 > 0$  independent of  $N$ ,

$$b^2 = \sum_{j=M_0+1}^M |\alpha_j|^2 > c_0. \quad (11.15)$$

For  $M_0 = 0$ , the lower bound (11.15) also holds, by the first part of Theorem 1.6.

We are now ready to complete the proof of (1.11). To this end, we assume in the sequel that  $d > 0$ , the case  $d < 0$  requiring only notation changes. Set

$$w' := \sum_{j=M_0+1}^M \alpha_j e_j \quad \text{and} \quad w'' := \sum_{j=M_0+1}^M \alpha_j \hat{e}_j, \quad (11.16)$$

with  $\alpha_j = e_j(z^*)^*v$  as before and

$$\hat{e}_j := \sum_{\nu=M_0+1}^M b_j(\nu) \frac{\mathfrak{z}_\nu^+}{\|\mathfrak{z}_\nu^+\|}, \quad (11.17)$$

with  $b_j(\nu)$ 's as in (10.23).

Since  $\|\alpha\| = 1 + o(1)$ , using the triangle inequality, the Cauchy-Schwarz inequality, and (10.23), we observe that

$$\|w'\|_{\ell^2([\ell, \ell'])}^2 \geq \frac{1}{2} \|w''\|_{\ell^2([\ell, \ell'])}^2 - \epsilon_N, \quad (11.18)$$

where  $\epsilon_N = O(\log N/N) + O(N^{-2C_\gamma}(\log N)^2)$ .

We next derive a lower bound on  $\|w'\|_{\ell^2([\ell, \ell'])}^2$ . Set  $\zeta_{\nu\nu'} := \zeta_\nu^+ \overline{\zeta_{\nu'}^+}$ . Recall Proposition 9.3 and note that

$$\langle \mathfrak{z}_\nu^+ | \mathfrak{z}_{\nu'}^+ \rangle_{\ell^2([\ell, \ell'])} = \zeta_{\nu\nu'}^\ell \frac{1 - \zeta_{\nu\nu'}^{\ell' - \ell + 1}}{1 - \zeta_{\nu\nu'}}, \quad \nu, \nu' \in [M] \setminus [M_0]. \quad (11.19)$$

Note that

$$\frac{1}{(1 - |\zeta_\nu^+|^2) \|\mathfrak{z}_\nu^+\|^2} = 1 + o(1) \quad \text{and} \quad \|\mathfrak{z}_\nu^+\|^2 \asymp N/\log N.$$

Thus, by Remark 10.7, from (11.19) we obtain that

$$\frac{\langle \mathfrak{z}_\nu^+ | \mathfrak{z}_{\nu'}^+ \rangle_{\ell^2([\ell, \ell'])}}{\|\mathfrak{z}_\nu^+\| \|\mathfrak{z}_{\nu'}^+\|} = \begin{cases} O(\log N/N) & \text{if } \nu \neq \nu', \\ (1 + o(1))(|\zeta_\nu^+|^{2\ell} - |\zeta_\nu^+|^{2(\ell'+1)}) & \text{if } \nu = \nu'. \end{cases} \quad (11.20)$$

On the other hand, since  $\nu, \nu' > M_0$ , for  $q \in [N]$ ,  $|\zeta_\nu^+|^{2q} = |\zeta_{\nu'}^+|^{2q} = e^{-cq \log N/N(1+o(1))}$  for some constant  $c > 0$  independent of  $\nu$ , and therefore, for  $\ell, \ell' = O(N/\log N)$ ,

$$(|\zeta_\nu^+|^{2\ell} - |\zeta_\nu^+|^{2(\ell'+1)}) \geq c'(\ell' - \ell) \cdot \frac{\log N}{N}, \quad (11.21)$$

for some  $c' > 0$ . Hence, recalling (11.16)-(11.17), from (10.23), (11.20) and (11.21), we deduce that for any  $\ell, \ell' \in [1, N]$ ,

$$\begin{aligned} \|w''\|_{\ell^2([\ell, \ell'])}^2 &= \sum_{j, j'=M_0+1}^M \bar{\alpha}_j \alpha_{j'} \sum_{\nu, \nu'=M_0+1}^M \overline{b_j(\nu)} b_{j'}(\nu') \cdot \frac{\langle \mathfrak{z}_\nu^+ | \mathfrak{z}_{\nu'}^+ \rangle_{\ell^2([\ell, \ell'])}}{\|\mathfrak{z}_\nu^+\| \cdot \|\mathfrak{z}_{\nu'}^+\|} \\ &\geq \sum_{j, j'=M_0+1}^M \bar{\alpha}_j \alpha_{j'} \sum_{\nu=M_0+1}^M \overline{b_j(\nu)} b_{j'}(\nu) \cdot c'(\ell' - \ell) \cdot \frac{\log N}{N} + O(\log N/N) \\ &= c'(\ell' - \ell) \cdot \frac{\log N}{N} \sum_{j, j'=M_0+1}^M \alpha_j \bar{\alpha}_{j'} \langle b_j | b_{j'} \rangle + O(\log N/N) \\ &\stackrel{(10.24)}{=} c'(\ell' - \ell) \cdot \frac{\log N}{N} \left( o(1) + \sum_{j=M_0+1}^M |\alpha_j|^2 \right) + O(\log N/N) \stackrel{(11.15)}{\geq} \frac{c_0 c'(\ell' - \ell)}{2} \cdot \frac{\log N}{N} + O(\log N/N). \end{aligned} \quad (11.22)$$

In the range of  $\ell, \ell'$  of the statement of Theorem 1.6, the first term in the right hand side of (11.22) is of order larger than  $\epsilon_N + O(\log N/N)$ , and therefore, for such  $\ell, \ell'$ , by (11.18)

$$\|w'\|_{\ell^2([\ell, \ell'])} \geq c_1(\ell' - \ell) \log N/N, \quad (11.23)$$

for some  $c_1 > 0$ . It remains to show that (11.23) continues to hold with  $w'$  replaced by  $w = w_z$ .

To see this, note that by (10.27), we have that  $\|w_{z^*} - w'\|_{\ell^2([C \log N, N])} = O(N^{-2})$ , for some  $C < \infty$ . Recall from above that  $\|w_z - w_{z^*}\| = o(N^{-2})$ . Thus, by the triangle inequality we indeed have that (11.23) holds for  $w$  with  $\ell \geq C \log N$  and  $\ell'$  as in the statement of Theorem 1.6. On the other hand, if  $1 \leq \ell \leq C \log N$  we observe that  $\ell' - \ell \leq 2(\ell' - C \log N)$  for any  $\ell' \geq N^{c_0}$ . Thus, upon shrinking  $c_1$  in (11.23), that lower bound continues to hold for  $w$  and such values of  $\ell$ . The proof of (1.12) is now immediate from (1.9). The assertion about the constant  $c_0$  follows from Remark 6.17 and the definition of  $\epsilon_N$ . This finally completes the proof.  $\square$



**11.2. Proof of Lemma 11.1.** Throughout the proof, we assume for notational simplicity that  $d > 0$ , the general case posing no new difficulties. We begin by rewriting  $\hat{A}(z)$ , see (11.13). Fix  $M_1 = m_+ - m_+^0$  and note that  $M_1 \geq M_0$  with equality if  $N_+ = 0$ . Let  $E_0(z)$  denote the  $N \times M_0$  matrix with columns  $e_j(z)$ , and  $F_0(z)$  the  $N \times M_0$  matrix with columns  $f_j(z)$ . It follows from Lemma 9.3 and Proposition 10.3 that the columns of  $E_0(z)$  are almost orthonormal (with inner product  $o(1)$  between different columns) and with  $|e_j(z)(k)| \leq e^{-ck}$ ,  $k = 1, \dots, N$  for some  $c > 0$ . Similarly, the columns of  $F_0(z)$  are almost orthonormal, with  $|f_j(z)(k)| \leq e^{-c(N-k)}$ ,  $k = 1, \dots, N$ . Recall that  $\hat{A}(z) = F_0(z)^* Q E_0(z)$ .

We begin by fixing a particular  $z$  as in the statement of the lemma, and write  $\bar{A} = \hat{A}(z)$ ,  $\bar{E}_0 = E_0(z)$ ,  $\bar{F}_0 = F_0(z)$ . Note that

$$\mathbf{E} \|\bar{A}\|_{\text{HS}}^2 = \sum_{i,j=1}^{M_0} \mathbf{E} |\bar{A}_{ij}|^2 \leq \mathfrak{C}_2 \left( \sum_{i,j=1}^{M_0} \|e_i(z)\|^2 \cdot \|f_j(z)\|^2 \right) \leq 2\mathfrak{C}_2 M_0^2,$$

for all large  $N$ . Therefore, one can find a constant  $C(\eta)$  so that

$$\mathbf{P}(s_{\max}(\bar{A}) > C(\eta)/2) \leq \eta/4. \quad (11.24)$$

We next control  $s_{\min}(\bar{A})$ . We use the inequality

$$s_{\min}(\bar{A}) \geq \frac{|\det(\bar{A})|}{(s_{\max}(\bar{A}))^{M_0-1}}. \quad (11.25)$$

To control the determinant in (11.25), we begin by considering the matrix  $\bar{E}_0$ . Because the columns of  $E_0$  decay exponentially and the columns of  $E_0$  are almost orthogonal, there exists a  $K > 0$  so that the  $K \times M_0$  sub-matrix  $\bar{E}_{0,K}$  consisting of the first  $K$  rows of  $\bar{E}_0$  has columns with inner product smaller than  $1/M_0^2$ , and therefore singular values all larger than  $1/2$ . It follows from [49, Lemma 4.3] (see also [34]) that there exists an  $M_0 \times M_0$  sub-matrix of  $\bar{E}_{0,K}$ , denoted  $\bar{E}_{0,K,M_0}$ , with singular values all larger than  $1/2\sqrt{1 + (K - M_0)M_0}$ . In particular, there exists a (nonrandom) constant  $c_0$  so that  $\det(\bar{E}_{0,K,M_0}) \geq c_0$ . Similarly, with  $\bar{F}_{0,K}$  denoting the sub-matrix of  $F_0$  consisting from its last  $K$  rows, there exists an  $M_0 \times M_0$  sub-matrix of  $\bar{F}_{0,K}$ , denoted  $\bar{F}_{0,K,M_0}$ , with  $\det(\bar{F}_{0,K,M_0}) \geq c_0$ .

Returning to the control of the determinant in (11.25), we note that the latter is a homogeneous polynomial in the entries of  $Q$ , and apply Lemma 6.12, as follows. Let  $I = \{i_1 < i_2 < \dots < i_{M_0}\} \subset [K]$  denote the list of rows of  $\bar{E}_0$  participating in  $\bar{E}_{0,K,M_0}$ . Similarly, let  $J = \{j_1 < j_2 < \dots < j_{M_0}\} \subset [N] \setminus [N - K]$  denote the list of rows of  $\bar{F}_0$  participating in  $\bar{F}_{0,K,M_0}$ . Let  $U_\ell = Q_{i_\ell, j_\ell}$ ,  $\ell = 1, \dots, M_0$ , then by the Cauchy-Binet formula,

$$\det(\bar{A}) = \sum_{\mathcal{I} \subset [M_0]} Z_{\mathcal{I}} \prod_{\ell \in \mathcal{I}} U_\ell \quad (11.26)$$

for some random variables  $Z_{\mathcal{I}}$  (which depend only on  $Q_{a,b}$  with  $(a,b) \notin \cup_{\ell=1}^{M_0} (i_\ell, j_\ell)$ ). Crucially,  $|Z_{[M_0]}| = |\det(\bar{E}_{0,K,M_0}) \cdot \det(\bar{F}_{0,K,M_0})| \geq c_0^{2M_0} =: c_\star$ .

By Lemma 6.12, we obtain that for any  $\varepsilon \in (0, c_\star e^{-1})$ ,

$$\mathbf{P}(\det(\bar{A}) \leq \varepsilon) \leq \bar{C}_{6.12} \cdot \left(\frac{\varepsilon}{c_\star}\right)^{(1+\eta)} \cdot \left(\log\left(\frac{c_\star}{\varepsilon}\right)\right)^{M_0-1}. \quad (11.27)$$

Combining (11.27) with (11.25) and (11.24), we obtain that there exists a  $c_\eta > 0$  so that

$$\mathbf{P}(s_{\min}(\bar{A}) > 2c_\eta) \geq 1 - \eta/4. \quad (11.28)$$

Next, let  $z' \neq z \in \mathcal{N}_\gamma \cap D(z_0, C_0 \log N/N)$ . Since  $|dp(z)|$  is bounded below uniformly in a fixed small ( $N$ -independent) neighborhood of  $z_0$ , we have that the roots of  $p - z$  and  $p - z'$  satisfy, for all  $j \leq N_- + N_+$ , that  $|\zeta_j^+(z) - \zeta_j^+(z')| \leq C_1 \log N/N$ . Using Propositions 9.3, 9.6, and 10.3, this implies in turn that there exists a constant  $C$  so that, for  $j = 1, \dots, M_0$ ,

$$\|e_j(z) - e_j(z')\| \leq C|z - z'| + C e^{-N/C} \leq 2C|z - z'|. \quad (11.29)$$

Indeed, Proposition 9.3 yields the first inequality for the  $\tilde{u}_\ell^+$ ,  $\ell = 1, \dots, M_1$ , Proposition 9.6 allows one to transfer this to the quasimodes  $\psi_j^+$ , and finally Proposition 10.3 transfers it to the  $e_j$ 's.



The right most inequality in (11.29) follows from the fact that  $\mathcal{N}_\gamma$  is a net of mesh size  $N^{-\theta}$ . By the same argument the same bound continues to hold when  $e_j$ 's are replaced by  $f_j$ 's.

Denote  $\mathcal{N}_{\gamma, z_0} := \mathcal{N}_\gamma \cap D(z_0, C_0 \log N/N)$ . Since  $\hat{A}_{i,j}(\cdot)$ ,  $i, j \in [M_0]$ , are linear in the entries of  $Q$ , applying [75, Theorem 2] together with (11.29) we obtain that, for any  $h \in \mathbb{N}$ ,

$$\max_{i,j \in [M_0]} \max_{z' \in \mathcal{N}_{\gamma, z_0}} \mathbf{E} |\hat{A}_{i,j}(z) - \hat{A}_{i,j}(z')|^{2h} \leq C_h \cdot \left( \frac{\log N}{N} \right)^{2h},$$

for some constant  $C_h < \infty$ , independent of  $N$ , where we also used that  $|z - z'| = O(\log N/N)$  for all  $z, z' \in \mathcal{N}_{\gamma, z_0}$ . Therefore, choosing  $h$  sufficiently large so that  $|\mathcal{N}_{\gamma, z_0}| N^{-h} = o(1)$ , applying Markov's inequality, taking a union bound over  $i, j \in [M_0]$  and  $z' \in \mathcal{N}_{\gamma, z_0}$  and recalling Assumption 1.1, we deduce that, on an event of probability approaching one,

$$s_{\max}(\hat{A}(z) - \hat{A}(z')) \leq \|\hat{A}(z) - \hat{A}(z')\|_{\text{HS}} \leq c_\eta, \quad (11.30)$$

uniformly for all  $z, z' \in \mathcal{N}_{\gamma, z_0}$ . By the Ky Fan inequalities,  $s_{\min}(\hat{A}(z')) \geq s_{\min}(\hat{A}(z)) - s_{\max}(\hat{A}(z) - \hat{A}(z'))$ . So we obtain the claim for  $s_{\min}(\hat{A}(z'))$ .

Finally, we turn to the control of  $s_{\max}(A(z))$ , for all  $z \in \mathcal{N}_{\gamma, z_0}$ . Let  $\tilde{E}(z)$  and  $\tilde{E}_0(z)$  be  $N \times (M - M_0)$  matrix whose columns are  $\mathfrak{z}_{j+M_0}^+(z) / \|\mathfrak{z}_{j+M_0}^+(z)\|$  (as in Proposition 9.3) and  $e_{j+M_0}(z)$ , respectively, for  $j = 1, \dots, M - M_0$ . As in Proposition 9.6, let  $\tilde{F}(z)$  be the  $N \times M_0$  matrix with columns  $\psi_i^-(z)$ ,  $i = 1, \dots, M_0$ . Set  $\tilde{A}(z) := \tilde{F}(z)^* Q \tilde{E}(z)$ . Recalling the definition of  $A$  from the discussion above (11.13), we find that in the new notation,  $A(z) = F_0(z)^* Q \tilde{E}_0(z)$ . Further let  $\tilde{B}(z)$  and  $\tilde{C}(z)$  be the matrices whose columns are  $\{b_{j+M_0}\}$  and  $\{a_j\}$ , as defined in Proposition 10.6. By Proposition 10.6 we have that

$$\|\tilde{E}(z)\tilde{B}(z) - \tilde{E}_0(z)\|_{\text{HS}}, \|\tilde{F}(z)\tilde{C}(z) - F_0(z)\|_{\text{HS}} = O(N^{-c}) \quad \text{and} \quad \|\tilde{E}_0(z)\|_{\text{HS}}, \|F_0(z)\|_{\text{HS}} \leq 2M,$$

for all  $z \in \mathcal{N}_{\gamma, z_0}$  and all large  $N$ , where  $c > 0$  is some constant. Therefore, applying Proposition 10.12, the triangle inequality, and proceeding as in the proof of (11.30), we deduce that

$$\max_{z \in \mathcal{N}_{\gamma, z_0}} \|A(z) - \tilde{B}(z)^* \tilde{A}(z) \tilde{C}(z)\| \leq \max_{z \in \mathcal{N}_{\gamma, z_0}} \|A(z) - \tilde{B}(z)^* \tilde{A}(z) \tilde{C}(z)\|_{\text{HS}} = o(1),$$

on an event with probability approaching one, as  $N \rightarrow +\infty$ . Moreover, by (10.19) and (10.24) we have that  $\|\tilde{B}(z)\|$  and  $\|\tilde{C}(z)\|$  are uniformly bounded for all  $z \in \mathcal{N}_{\gamma, z_0}$ . Therefore, it suffices to bound  $\sup_{z \in \mathcal{N}_{\gamma, z_0}} \|\tilde{A}(z)\|$ , and hence in fact it is enough to prove that for fixed  $i \in [M_0], j \in [M] \setminus [M_0]$ ,

$$\lim_{x \rightarrow \infty} \limsup_{N \rightarrow \infty} \mathbf{P} \left( \sup_{z \in \mathcal{N}_{\gamma, z_0}} |\tilde{A}_{ij}(z)| > x \right) = 0. \quad (11.31)$$

Toward this goal, we note that for  $z, z' \in \mathcal{N}_{\gamma, z_0}$ , one has that for some constant  $C_0 < \infty$ , independent of  $N$ ,

$$\left\| \frac{\mathfrak{z}_j^+(z)}{\|\mathfrak{z}_j^+(z)\|} - \frac{\mathfrak{z}_j^+(z')}{\|\mathfrak{z}_j^+(z')\|} \right\| \leq C_0 |z - z'| \cdot \frac{N}{\log N}, \quad j \in [M] \setminus [M_0],$$

and by Proposition 10.6 the bound (11.29) continues to hold with  $e_j(\cdot)$  replaced by  $\psi_j^-(\cdot)$  for  $j \in [M_0]$ . Therefore, upon using Assumption 1.1 and applying [75, Theorem 2], we find that for some constant  $C_h$  independent of  $N$ ,

$$\mathbf{E} |\tilde{A}_{i,j}(z) - \tilde{A}_{i,j}(z')|^h \leq C_h \left( \frac{|z - z'| N}{\log N} \right)^h. \quad (11.32)$$

We now apply a Komogorov-type argument. Namely, for  $k = 0, 1, \dots$ , let  $\mathcal{N}_{\gamma, z_0, k}$  denote a minimal subset of  $\mathcal{N}_{\gamma, z_0}$  so that for any  $z \in \mathcal{N}_{\gamma, z_0}$  there exists  $z' \in \mathcal{N}_{\gamma, z_0, k}$  so that  $|z - z'| \leq 2^{-k} \log N/N$ . Note that  $|\mathcal{N}_{\gamma, z_0, k}| = O(2^{2k})$ . Now, fix a constant  $x$  and introduce the event

$$\mathcal{A}_k = \{\exists z, z' \in \mathcal{N}_{\gamma, z_0, k} : |z - z'| \leq 2^{-k+1} \log N/N, |\tilde{A}_{i,j}(z) - \tilde{A}_{i,j}(z')| > x/j^2\},$$

and set  $\mathcal{B} = \cap_{k=1}^{k_0} \mathcal{A}_k^{\mathbb{C}}$ , where  $2^{k_0-1} \leq C_0 N / \log N \leq 2^{k_0}$ . On the event  $\mathcal{B}$  we have that

$$\max_{z \in \mathcal{N}_{\gamma, z_0}} |\tilde{A}_{i,j}(z)| \leq 10x.$$

On the other hand, by Markov's inequality and (11.32), for some constant  $C'_h$  independent of  $N$ ,

$$\mathbf{P}(\mathcal{B}^{\mathbb{C}}) \leq C'_h x^{-h} \sum_{k=1}^{k_0} 2^{2k} 2^{-hk}. \quad (11.33)$$

Choosing  $h = 4$ , this implies (11.31) and completes the proof of the lemma.  $\square$

**11.3. Proof of Corollary 1.7.** We begin by applying Theorem 1.5 with  $\mu$  replaced by  $\mu/8$ . This yields the existence of constants  $\varepsilon_{1.5}, C_{1.5}$  such that

$$\mathbf{P}(\mathcal{N}_{\Omega(\varepsilon_{1.5}, C_{1.5}, N), N, \gamma} < (1 - \mu/8)N) \rightarrow_{N \rightarrow \infty} 0. \quad (11.34)$$

We now claim the following. Fix  $\bar{C} < \infty$ ,  $z_0 \in \Omega(\varepsilon_{1.5}, C_{1.5}, N)$ , and  $\mu > 0$ . Then there exist  $0 < \mu_1, \mu_2 < \infty$  such that, for all large  $N$ ,

$$\mathbf{P}(\exists i \in [N] \text{ such that } \text{supp}_{\mu_1}(v(\lambda_i^N)) < \mu_2 N / \log N \text{ and } \lambda_i^N \in \mathbb{D}_{z_0}) \leq \mu / (4\bar{C}), \quad (11.35)$$

where  $\mathbb{D}_{z_0} := D(z_0, C_{1.5} \log N / (2N)) \cap \Omega(\varepsilon_{1.5}, C_{1.5}, N)$ . To prove (11.35) we borrow some notation from the proof of Theorem 1.6. Let  $\hat{z} \in \mathbb{D}_{z_0}$  and  $z^* = z^*(\hat{z}) \in \mathcal{N}_{\gamma, z_0} := \mathcal{N}_{\gamma} \cap D_{z_0}$  such that  $|\hat{z} - z^*| \leq N^{-\theta}$ . Recall from the proof of Theorem 1.6 that  $w_z = \sum_{j=1}^M \alpha_j e_j(z)$ , for  $z \in \mathbb{C}$ ,  $w' = w'(z^*) = \sum_{j=M_0+1}^M \alpha_j e_j(z^*)$ , and  $w = w_{\hat{z}}$ . Also recall that  $\|w - w_{z^*}\| = o(1)$ . Applying Lemma 11.1, with  $\eta = \mu / (4\bar{C})$ , and upon choosing  $\mu_1$  sufficiently small, we obtain that, simultaneously for all  $z^*$  in  $\mathcal{N}_{\gamma, z_0}$ , and hence simultaneously for all  $\hat{z} \in \mathbb{D}_{z_0}$ , and any  $I \subset [N]$  such that  $\|v\|_{\ell^2(I)} > 1 - \mu_1$ ,

$$\sum_{j=M_0+1}^M |\alpha_j| \|e_j\|_{\ell^2(I)} \geq \|w'\|_{\ell^2(I)} \stackrel{(1.9)}{\geq} \|v\|_{\ell^2(I)} - \|w_{z^*} - w'\| - o(1) \geq 1 - \mu_1/2 - \sqrt{\sum_{j=1}^{M_0} |\alpha_j|^2} \geq \mu_1, \quad (11.15)$$

with probability at least  $1 - \mu / (4\bar{C})$ , where  $v = v(\hat{z})$  is the right eigenvector corresponding to  $\hat{z}$ . On the other hand by (10.25) it follows that for any such  $I$  we must have that  $|I| \geq \mu_2 N / \log N$ , for some  $\mu_2 > 0$ . This, indeed proves (11.35).

We continue with the proof of the corollary. Introduce the following notation. For  $z_0 \in \mathbb{C}$  set

$$\mathcal{J}_{z_0} := \{\#\{i \in [N] : \lambda_i^N \in \mathbb{D}_{z_0}\} \leq C_{7.4} C_{1.5}^4 \log N\},$$

$$\mathcal{C}_{z_0} := \{\exists i \in [N] \text{ such that } \text{supp}_{\mu_1}(v(\lambda_i^N)) < \mu_2 N / \log N \text{ and } \lambda_i^N \in \mathbb{D}_{z_0}\},$$

and for  $i \in [N]$

$$\mathbb{I}_i := \mathbf{I}(\text{supp}_{\mu_1}(v(\lambda_i^N)) < \mu_2 N / \log N).$$

Let  $\mathcal{R}$  be a net of  $\Omega(\varepsilon_{1.5}, C_{1.5}, N)$  of mesh size  $C_{1.5} \log N / (2N)$ . It is clear that  $|\mathcal{R}| \leq CC_{1.5}^{-1} N / \log N$ , for some  $C < \infty$ . Note also that by Theorem 7.4, for any  $z_0 \in \mathcal{R}$ ,

$$\mathbf{P}(\mathcal{J}_{z_0}^c) \leq N^{-2}. \quad (11.36)$$

We can now complete the proof. Indeed, for all large  $N$ ,

$$\begin{aligned} \sum_{i=1}^N \mathbf{E}[\mathbb{I}_i] &\stackrel{(11.34)}{\leq} \sum_i \mathbf{E}[\mathbb{I}_i \cdot \mathbf{I}(\lambda_i^N \in \Omega(\varepsilon_{1.5}, C_{1.5}, N))] + N\mu/4 \\ &\leq \sum_{z_0 \in \mathcal{R}} \mathbf{E} \left[ \sum_{i: \lambda_i^N \in \mathbb{D}_{z_0}} \mathbb{I}_i \right] + N\mu/4 \leq N \sum_{z_0 \in \mathcal{R}} \mathbf{P}(\mathcal{J}_{z_0}^c) + \sum_{z_0 \in \mathcal{R}} \mathbf{E} \left[ \sum_{i: \lambda_i^N \in \mathbb{D}_{z_0}} \mathbb{I}_i \cdot \mathbf{I}(\mathcal{J}_{z_0}) \right] + N\mu/4 \\ &\stackrel{(11.36)}{\leq} C_{7.4} C_{1.5}^4 \log N \sum_{z_0 \in \mathcal{R}} \mathbf{P}(\mathcal{C}_{z_0}) + N\mu/2 \stackrel{(11.35)}{\leq} N\mu, \end{aligned}$$

where in the penultimate step we used that  $|\mathcal{R}| = o(N)$ , and in last step we chose  $\bar{C} = CC_{7.4} C_{1.5}^3$ , and used the bound on  $|\mathcal{R}|$ . This completes the proof.  $\square$

**Remark 11.2** (Sup-norm delocalization). *We borrow notation from the proof of Theorem 1.6. Let  $P_N$  be the Jordan block. By Remark 5.5, for any fixed  $\gamma' < \gamma$ , all eigenvalues of  $P_{N,\gamma}^Q$  are inside the disc  $D(0, 1 - (\gamma' - 1) \log N/N)$  with probability approaching one. Therefore, for any eigenvalue  $\hat{z}$  and its approximating net point  $z^*$  we can take  $C_\gamma = \gamma' - 1$  (see (9.5)). On the other hand, by Remark 10.14 we observe that the  $o(1)$  term in (11.6) can be replaced by  $O(N^{-(\gamma'-1)})$  and hence the same can be done for (11.5). Furthermore, in this case we have only one pure state, recall Remarks 9.4 and 9.5. Thus  $M = 1$ . Therefore, by Remark 10.5 we next derive that*

$$\left\| \frac{\hat{\mathfrak{z}}_1^+}{\|\hat{\mathfrak{z}}_1^+\|} - v \right\|_\infty \leq \left\| \frac{\hat{\mathfrak{z}}_1^+}{\|\hat{\mathfrak{z}}_1^+\|} - v \right\| \leq \|v - \alpha_1 e_1\| + |1 - |\alpha_1|| + \left\| \frac{\hat{\mathfrak{z}}_1^+}{\|\hat{\mathfrak{z}}_1^+\|} - e_1 \right\| = O(N^{-(\gamma'-1)} \log N).$$

Since,  $\gamma > 3/2$ , upon choosing  $\gamma'$  appropriately, it is now immediate that

$$\sqrt{\frac{\log N}{N}} \asymp \frac{1}{2} \cdot \frac{\|\hat{\mathfrak{z}}_1^+\|_\infty}{\|\hat{\mathfrak{z}}_1^+\|} \leq \|v\|_\infty \leq 2 \frac{\|\hat{\mathfrak{z}}_1^+\|_\infty}{\|\hat{\mathfrak{z}}_1^+\|} \asymp \sqrt{\frac{\log N}{N}}.$$

## REFERENCES

- [1] P. Alexandersson, *Schur polynomials, banded Toeplitz matrices and Widom's formula*, Electronic Journal of Combinatorics, **19**(4), P22 (2012).
- [2] J. Alt, R. Ducatez, and A. Knowles, *Delocalization transition for critical Erdős Rényi graphs*, arXiv preprint arXiv:2005.14180 (2020).
- [3] N. Anantharaman, *Quantum ergodicity on regular graphs*, Communications in Mathematical Physics, **535**, 633–690 (2017).
- [4] N. Anantharaman and E. Le Masson, *Quantum ergodicity on large regular graphs*, Duke Mathematical Journal, **164**, 723–765 (2015).
- [5] N. Anantharaman and M. Sabri, *Quantum ergodicity on graphs: From spectral to spatial delocalization*, Annals of Mathematics, **189**(3), 753–835 (2019).
- [6] P. W. Anderson, *Absence of Diffusion in Certain Random Lattices*, Physical Review, **109**(5), 1492–1505 (1958).
- [7] A. Basak, E. Psquette, and O. Zeitouni, *Regularization of non-normal matrices by Gaussian noise - the banded Toeplitz and twisted Toeplitz cases*, Forum of Mathematics, Sigma 7, paper e3 (2019).
- [8] A. Basak, E. Psquette, and O. Zeitouni, *Spectrum of random perturbations of Toeplitz matrices with finite symbols*, Transactions of the American Mathematical Society **373**, 4999–5023 (2020).
- [9] A. Basak and O. Zeitouni, *Outliers of random perturbation of Toeplitz matrices with finite symbols*, Probability Theory and Related Fields, **178**, 771–826 (2020).
- [10] Benigni, L., *Eigenvectors distribution and quantum unique ergodicity for deformed Wigner matrices*, Ann. Inst. Henri Poincaré Probab. Stat., **56**, (2020).
- [11] L. Benigni and P. Lopatto, *Optimal delocalization for generalized Wigner matrices*, arXiv preprint arXiv:2007.09585 (2020).
- [12] R. Bhatia. *Matrix Analysis*. Springer-Verlag, Berlin (1997).
- [13] D. Borthwick and A. Uribe, *On the Pseudospectra of Berezin–Toeplitz Operators*, Methods and Applications of Analysis, **10**(1), 31–66 (2003).
- [14] A. Böttcher and S. M. Grudsky, *Asymptotically good pseudomodes for Toeplitz matrices and Wiener-Hopf operators*, In Operator Theoretical Methods and Applications to Mathematical Physics, pp. 175–188, Birkhäuser, Basel (2004).
- [15] A. Böttcher and S. M. Grudsky, *Spectral Properties of Banded Toeplitz matrices*, SIAM (2005).
- [16] P. Bourgade and H.-T. Yau, *The eigenvector moment flow and local quantum unique ergodicity*, Communications in Mathematical Physics, **350**, 231–278 (2017).
- [17] D. Bump and P. Diaconis, *Toeplitz minors*, Journal of Combinatorial Theory, Series A, **97**, 252–271 (2002).
- [18] E. F. Cornelius, *Identities for complete homogeneous systems*, Journal of Algebra, Number Theory and Applications, **21**, 109–116 (2011).
- [19] E. B. Davies, *Pseudo-spectra, the harmonic oscillator and complex resonances*, Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences, **455**, 585–599 (1982).
- [20] E. B. Davies, *Semi-classical states for non-self-adjoint Schrödinger operators*, Communications in Mathematical Physics, **200**(1), 35–41 (1999).
- [21] E. B. Davies, *Pseudospectra of differential operators*, Journal of Operator Theory, **43**(2), 243–262 (2000).
- [22] E. B. Davies and M. Hager, *Perturbations of Jordan matrices*, Journal of Approximation Theory, **156**, 82–94 (2009).
- [23] N. Dencker, J. Sjöstrand, and M. Zworski, *Pseudospectra of semiclassical (pseudo-) differential operators*, Communications on Pure and Applied Mathematics **57** (2004), no. 3, 384–415.
- [24] A. Drouot, *Resonances for random highly oscillatory potentials*, J. of Mathematical Physics **59** (2018), no. 101506, 34 pp.

- [25] S. Dyatlov and M. Zworski, *Mathematical Theory of Scattering Resonances*, American Mathematical Society, 2019.
- [26] V. Enss, *Introduction to Asymptotic Observables for Multi-Particle Quantum Scattering*, Lecture Notes in Mathematics, Springer, **1218** (1983), 61–92.
- [27] L. Erdős, A. Knowles, H.-T. Yau, and J. Yin, *Spectral statistics of erdős-rényi graphs I: Local semicircle law*, Annals of Probability **41**(3B), 2279–2375 (2013).
- [28] L. Erdős, B. Schlein, and H.-T. Yau, *Local semicircle law and complete delocalization for Wigner random matrices*, Communications in Mathematical Physics, **287**, 641–655 (2009).
- [29] L. Erdős, B. Schlein, and H.-T. Yau, *Semicircle law on short scales and delocalization of eigenvectors for Wigner random matrices*, Annals of Probability **37**(3), 815–852 (2009).
- [30] J. Galkowski, *Nonlinear Instability in a Semiclassical Problem*, Communications in Mathematical Physics, **316**, 705–722 (2012).
- [31] D. García-García and M. Tierz, *Toeplitz minors and specializations of skew Schur polynomials*, Journal of Combinatorial Theory, Series A, **172**, article 105201 (2020).
- [32] I.C. Gohberg and M.G. Krein, *Introduction to the theory of linear non-selfadjoint operators*, Translations of mathematical monographs, vol. 18, AMS (1969).
- [33] I. Goldsheid and S. Sodin, *Lower bounds on Anderson-localised eigenfunctions on a strip*, arXiv preprint arXiv:2012.03017 (2020).
- [34] S. A. Goreinov, E. E. Tyrtyshnikov, and N. L. Zamarashkin, *A theory of pseudo-skeleton approximations*, Linear Algebra and its Applications, **261**, 1–21 (1997).
- [35] V. V. Grushin. *Les problèmes aux limites dégénérés et les opérateurs pseudodifférentiels*, Actes du Congrès International des Mathématiciens, Tome 2, Gauthier-Villars, 737–743 (1970).
- [36] M. Hager and J. Sjöstrand, *Eigenvalue asymptotics for randomly perturbed non-selfadjoint operators*, Mathematische Annalen **342**, 177–243 (2008).
- [37] B. Helffer and J. Sjöstrand, *Resonances en limite semiclassique*, Bull. Soc. Math. France **114** (1986), no. 24–25.
- [38] L. Hörmander, *An introduction to complex analysis in several variables*, Elsevier (1973).
- [39] L. Hörmander, *The Analysis of Linear Partial Differential Operators IV*, Grundlehren der mathematischen Wissenschaften, vol. 275, Springer-Verlag, 1985.
- [40] S. Kalmykov and L. V. Kovalev, *Self intersections of Laurent polynomials and the density of Jordan curves*, Proceedings of the American Mathematical Society, to appear (2021).
- [41] F. Klopp, *Resonances for large one-dimensional "ergodic" systems*, Anal. PDE **9** (2016), no. 2, 259–352.
- [42] R. Latała, *Some estimates of norms of random matrices*, Proceedings of the American Mathematical Society, **133**(5), 1273–1282 (2004).
- [43] K. Luh and S. O’Rourke, *Eigenvector delocalization for non-Hermitian random matrices and applications*, Random Structures & Algorithms, **57**(1), 169–210 (2020).
- [44] A. De Luca, B. L. Altshuler, V. E. Kravtsov, and A. Scardicchio, *Anderson Localization on the Bethe Lattice: Nonergodicity of Extended States*, Physical Review Letters **113**(4), 046806 (2014).
- [45] A. Lytova and K. Tikhomirov, *On delocalization of eigenvectors of random non-Hermitian matrices*, Probability Theory and Related Fields, **177**, 465–524 (2020).
- [46] I. G. Macdonald, *Symmetric functions and Hall polynomials*, second edition, The Oxford University Press (1995).
- [47] E. A. Maximenko and M. A. Moctezuma-Salazar, *Cofactors and eigenvectors of banded Toeplitz matrices: Trench formulas via skew Schur polynomials*, Operators and Matrices, **11**(4), 1149–1169 (2017).
- [48] R. B. Melrose, *Geometric Scattering Theory*, Cambridge University Press, 1995.
- [49] A. Mikhalev and I. V. Oseledets, *Rectangular maximum-volume sub-matrices and their applications*, Linear Algebra and its Applications, **538**, 187–211 (2018).
- [50] S. Nonnenmacher and M. Zworski, *Distribution of resonances for open quantum maps.*, Comm. Math. Phys (2007), no. 269, 311–365.
- [51] S. Nonnenmacher and M. Zworski, *Quantum decay rates in chaotic scattering*, Acta Math. **203** (2009), no. 2, 149 – 233.
- [52] S. Nonnenmacher, J. Sjöstrand, and M. Zworski, *Fractal weyl law for open quantum chaotic maps*, Annals of Math. **179** (2014), no. 1, 179–251.
- [53] K. Pravda-Starov, *Pseudo-spectrum for a class of semi-classical operators*, Bull. Soc. Math. France **136** (2008), no. 3, 329–372.
- [54] A. Raphael and M. Zworski, *Pseudospectral effects and basins of attraction*, unpublished notes (2005).
- [55] L. Reichel and L. N. Trefethen, *Eigenvalues and pseudo-eigenvalues of Toeplitz matrices*, Linear algebra and its applications, **162**, 153–185 (1992).
- [56] M. Rudelson and R. Vershynin, *Delocalization of eigenvectors of random matrices with independent entries*, Duke Mathematical Journal, **164**(13), 2507–2538 (2015).
- [57] M. Rudelson and R. Vershynin, *No-gaps delocalization for general random matrices*, Geometric and Functional Analysis, **26**, 1716–1776 (2016).
- [58] B. Sandstede and A. Scheel, *Basin boundaries and bifurcations near convective instabilities: a case study*, Journal of Differential Equations, **208**(1), 176–193 (2005).

- [59] P. Sarnak, *Arithmetic quantum chaos*, in *The Schur Lectures (tel aviv)*, Israel Math. Conf. Proc **8**, 183–236 (1992).
- [60] L. M. Sigal and A. Soffer, *The N-particle Scattering Problem: Asymptotic Completeness for Short Range Systems*, *Annals of Math.* **126** (1987), 35–108.
- [61] J. Sjöstrand, *Operators of principle type with interior boundary conditions*, *Acta Math.* **130** (1973), 1–51.
- [62] J. Sjöstrand, *Weyl law for semi-classical resonances with randomly perturbed potentials*, *Mémoires de la SMF* **136** (2014).
- [63] J. Sjöstrand and M. Vogel, *Large bidiagonal matrices and random perturbations*, *Journal of Spectral Theory*, **6**(4), 977–1020 (2016).
- [64] J. Sjöstrand and M. Vogel, *General Toeplitz matrices subject to Gaussian perturbations*, *Annales Henri Poincaré*, **22**, 49–81 (2021).
- [65] J. Sjöstrand and M. Vogel, *Toeplitz band matrices with small random perturbations*, *Indagationes Mathematicae*, **32**(1), 275–322 (2021).
- [66] J. Sjöstrand and M. Zworski, *Complex scaling and the distribution of scattering poles*, *Journal of the American Mathematical Society* **43**(4), 729–769 (1991).
- [67] J. Sjöstrand and M. Zworski, *Elementary linear algebra for advanced spectral problems*, *Annales de l’institut Fourier*, **57**(7), 2095–2141 (2007).
- [68] R. P. Stanley, *Enumerative Combinatorics*, Volume 2. Cambridge University Press (1999).
- [69] T. Tao and V. Vu, *Random Matrices: The Circular Law*, *Communications in Contemporary Mathematics*, **10**(2), 261–307 (2008).
- [70] L.N. Trefethen, *Pseudospectra of linear operators*, *SIAM Rev.* **39**, 383–406 (1997).
- [71] L. N. Trefethen and S. J. Chapman, *Wave packet pseudomodes of twisted Toeplitz matrices*, *Communications on Pure and Applied Mathematics*, **57**(9), 1233–1264 (2004).
- [72] L. N. Trefethen and M. Embree, *Spectra and Pseudospectra: The Behavior of Nonnormal Matrices and Operators*, Princeton University Press (2005).
- [73] M. Vogel, *Almost sure Weyl law for quantized tori*, *Communications in Mathematical Physics*, **378**(2), 1539–1585 (2020).
- [74] H. Whitney, *Complex Analytic Varieties*. Addison-Wesley Pub. Co. Reading, Massachusetts (1972).
- [75] P. Whittle, *Bounds for the Moments of Linear and Quadratic Forms in Independent Variables*. *Teoriya Veroyatnostei i ee Primeneniya*, **5**(3), 331–335 (1960).
- [76] M. Zworski, *A remark on a paper of E.B. Davies*, *Proc. A.M.S.* (2001), no. 129, 2955–2957.

(Anirban Basak) INTERNATIONAL CENTRE FOR THEORETICAL SCIENCES, TATA INSTITUTE OF FUNDAMENTAL RESEARCH, BANGALORE, INDIA.

*Email address:* [anirban.basak@icts.res.in](mailto:anirban.basak@icts.res.in)

(Martin Vogel) INSTITUT DE RECHERCHE MATHÉMATIQUE AVANCÉE - UMR 7501, UNIVERSITÉ DE STRASBOURG ET CNRS, 7 RUE RENÉ-DESCARTES, 67084 STRASBOURG CEDEX, FRANCE.

*Email address:* [vogel@math.unistra.fr](mailto:vogel@math.unistra.fr)

(Ofer Zeitouni) DEPARTMENT OF MATHEMATICS, WEIZMANN INSTITUTE OF SCIENCE, POB 26, REHOVOT 76100, ISRAEL AND COURANT INSTITUTE, NEW YORK UNIVERSITY, 251 MERCER ST, NEW YORK, NY 10012, USA.

*Email address:* [ofer.zeitouni@weizmann.ac.il](mailto:ofer.zeitouni@weizmann.ac.il)