Electron. J. Probab. **27** (2022), article no. 24, 1–29.
ISSN: 1083-6489 https://doi.org/10.1214/22-EJP747

SUSY transfer matrix approach for the real symmetric 1d random band matrices*

Tatyana Shcherbina†‡

Abstract

This paper adapts the recently developed rigorous application of the supersymmetric transfer matrix approach for the Hermitian 1d band matrices to the case of the orthogonal symmetry. We consider $N \times N$ block band matrices consisting of $W \times W$ random Gaussian blocks (parametrized by $j, k \in \Lambda = [1, n] \cap \mathbb{Z}$, $N = nW$) with a fixed entry's variance $J_{jk} = W^{-1}(\delta_{j,k} + \beta\Delta_{j,k})$ in each block. Considering the limit $W, n \to \infty$, we prove that the behaviour of the second correlation function of characteristic polynomials of such matrices in the bulk of the spectrum exhibit a crossover near the threshold $W \sim \sqrt{N}$.

Keywords: SUSY; random band matrices; real symmetric case; universality; characteristic polynomials.
MSC2020 subject classifications: 60B20.
Submitted to EJP on June 15, 2021, final version accepted on January 29, 2022.

1 Introduction

Starting from the works of Erdős, Yau, Schlein with co-authors (see [16] and reference therein) and Tao and Vu (see, e.g., [35]), significant progress in understanding of universal behaviour of local eigenvalues statistics of many random graph and random matrix models were achieved. However, for the random matrices with spacial structure our understanding is much more limited.

One of the most important such models is the ensemble of random band matrices (RBM), i.e. $N \times N$ matrices having non-zero entries only in a strip of width $2W$ near the main diagonal. Such matrices interpolate between mean-field type Wigner matrices (Hermitian or real symmetric matrices with i.i.d. random entries) and random Schrödinger operators, which have only a random diagonal potential in addition to the deterministic

*Supported in part by NSF grant DMS-1700009.
†Department of Mathematics, University of Wisconsin-Madison, Madison, WI, USA.
E-mail: tshcherbyna@wisc.edu
‡Former affiliation: Department of Mathematics, Princeton University, Princeton, NJ, USA.

Laplacian on a box in $\mathbb{Z}^d$. The density of states ρ of a general class of RBM with $W \gg 1$ is given by the well-known Wigner semicircle law (see [3, 24]):

$$\rho(E) = \frac{1}{2\pi}\sqrt{4 - E^2}, \quad E \in [-2, 2]. \tag{1.1}$$

The main long standing problem in the field is to prove a fundamental physical conjecture formulated in late 80s (see [10], [17]). The conjecture states that the eigenvectors of $N \times N$ RBM are completely delocalized and the local spectral statistics governed by the Wigner-Dyson statistics for large bandwidth W (i.e. the local behaviour is the same as for Wigner matrices), and by Poisson statistics for a small W (with exponentially localized eigenvectors). This is the analogue of the celebrated Anderson metal-insulator transition for random Schrödinger operators (see [34] for more details).

The transition (crossover) for RBM in one spacial dimension is conjectured to occur around the critical value $W = \sqrt{N}$. The conjecture is supported by physical derivation due to Fyodorov and Mirlin (see [17]), and also by the so-called Thouless scaling. On the mathematical level of rigour, localization of eigenvectors in the bulk of the spectrum was first shown for $W \ll N^{1/8}$ [27], and then the bound was improved to $N^{1/7}$ [26]. On the other side, by a development of the Erdős-Schlein-Yau approach to Wigner matrices (see [16]), there were obtained some results where the weaker form of delocalization was proved for $W \gg N^{6/7}$ in [14], $W \gg N^{4/5}$ in [15], $W \gg N^{7/9}$ in [18]. The combination of this approach with the new ideas based on quantum unique ergodicity gave first GUE/GOE gap distributions for RBM with $W \sim N$ [5], and then were developed in [6]–[7], [38] to obtain bulk universality and complete delocalization in the range $W \gg N^{3/4}$ (see review [4] for the details).

There is a completely different approach which allows to work with random operators with a non-trivial spacial structure based on supersymmetry techniques (SUSY). It is widely used in the physics literature (see e.g. reviews [13], [23]) but its rigorous mathematical application is usually quite difficult and it requires to incorporate various analytic and statistical mechanics techniques. However, for the 1d Hermitian RBM of a certain type it was successfully done both for correlation functions of characteristic polynomials and for usual correlation functions. More precisely, combining SUSY with a delicate steepest descent method and transfer matrix techniques, we were able to perform a complete study of the local regime of characteristic polynomials for Hermitian Gaussian 1d RBM (see [29] for the regime $W \gg \sqrt{N}$, [30] for the regime $W \ll \sqrt{N}$, and [32] for the regime $W \sim \sqrt{N}$), and also obtain the first rigorous universality result for the second order correlation function for the whole delocalized region $W \gg \sqrt{N}$ (see [31]).

Let us mention also that SUSY approach was also applied to obtain the detailed information of the density of states for the Hermitian RBM of higher dimensions (see [11], [12]).

There are much less rigorous applications of SUSY techniques for the case of real symmetric matrices, since the SUSY integral representations are more complicated for the case of orthogonal symmetry. However, the technique of [29] was successfully adapted in [33] to the study of characteristic polynomials for real symmetric Gaussian 1d RBM in the delocalized regime $W \gg \sqrt{N}$. In this paper we want to perform the complete study of characteristic polynomials for real symmetric Gaussian 1d RBM adapting the SUSY transfer matrix techniques of [30], [32] to the case of orthogonal symmetry. This is an important step towards the proof of the universality of the usual correlation functions for the case of real symmetric 1d RBM, as well as for the general development of rigorous application of SUSY approach for the real symmetric case.

The model we are going to consider is different from the model of 1d RBM considered in [29]–[30], [32] and in [33], but coincides with the model considered in [31]. Namely,

we consider real symmetric block band matrices, i.e. real symmetric matrices H_N, $N = nW$ with elements $H_{jk,\alpha\gamma}$, where $j, k \in 1, \dots, n$ (they parametrize the lattice sites) and $\alpha, \beta = 1, \dots, W$ (they parametrize the orbitals on each site). The entries $H_{jk,\alpha\gamma}$ are random Gaussian variables with mean zero such that

$$\langle H_{j_1k_1,\alpha_1\beta_1} H_{j_2k_2,\alpha_2\beta_2} \rangle = \delta_{j_1k_2}\delta_{j_2k_1}\delta_{\alpha_1\beta_2}\delta_{\beta_1\alpha_2} J_{j_1k_1}. \tag{1.2}$$

Here $J_{jk} \geq 0$ are matrix elements of the positive-definite symmetric $n \times n$ matrix J, such that

$$\sum_{j=1}^{n} J_{jk} = 1/W.$$

The probability law of H_N can be written in the form

$$P_N(dH_N) = \exp\Big\{ -\frac{1}{4} \sum_{j,k\in\Lambda} \sum_{\alpha,\gamma=1}^{W} \frac{H_{jk,\alpha\gamma}^2}{J_{jk}} \Big\} dH_N, \tag{1.3}$$

where

$$dH_N = \prod_{j<k} \prod_{\alpha\gamma} \frac{dH_{jk,\alpha\gamma}}{\sqrt{2\pi J_{jk}}} \prod_{j} \prod_{\alpha<\gamma} \frac{dH_{jj,\alpha\gamma}}{\sqrt{2\pi J_{jj}}} \prod_{j} \prod_{\alpha} \frac{dH_{jj,\alpha\alpha}}{\sqrt{4\pi J_{jj}}}.$$

Such models were first introduced and studied by Wegner (see [28], [37]) (and sometimes they are also called Wegner's orbital models).

As in [31], we consider the case

$$J = 1/W + \beta\Delta^{(0)}/W, \quad \beta < 1/4, \tag{1.4}$$

where $W \gg 1$ and $\Delta^{(0)}$ is the discrete Laplacian on $[1, n] \cap \mathbb{Z}$ with Neumann boundary conditions. Clearly, this model is one of the possible realizations of the Gaussian random band matrices with the band width $2W + 1$ (note that the model can be defined similarly in any dimensions $d > 1$ taking $j, k \in [1, n]^d \cap \mathbb{Z}^d$ in (1.2)).

The main interest of this paper is to study the behaviour of correlation functions (or the mixed moments) of characteristic polynomials which can be defined as

$$F_{2k}(\Lambda) = \int \prod_{s=1}^{2k} \det(\lambda_s - H_N) P_n(d\,H_N), \tag{1.5}$$

where $P_n(d\,H_N)$ is defined in (1.3), and $\Lambda = \operatorname{diag}\{\lambda_1, \dots, \lambda_{2k}\}$ are real or complex parameters that may depend on N. As in the Hermitian case, correlation functions of characteristic polynomials of real symmetric 1d RBM are expected to exhibit a crossover near the threshold $W \sim \sqrt{N}$: it is expected that they the same local behaviour as for GOE for $W \gg \sqrt{N}$, and the different behaviour for $W \ll \sqrt{N}$.

The asymptotic local behaviour in the bulk of the spectrum of the $2k$-point mixed moment for GOE is well-known. It was proved for $k = 1$ by Brézin and Hikami [8] (based on SUSY approach), and for general k by Borodin and Strahov [9] (with a different techniques) that

$$F_{2k}\left(\Lambda_0 + \hat{\xi}/N\rho(E)\right) = C_{N,k} \frac{\operatorname{Pf}\left\{DS(\pi(\xi_i - \xi_j))\right\}_{i,j=1}^{2k}}{\triangle(\xi_1, \dots, \xi_{2k})} (1 + o(1)),$$

where $C_{N,k}$ is some multiplicative constant depending on N, k,

$$DS(x) = -\frac{3}{x}\frac{d}{dx}\frac{\sin x}{x} = 3\Big(\frac{\sin x}{x^3} - \frac{\cos x}{x^2}\Big), \tag{1.6}$$

$\triangle(\xi_1,\dots,\xi_k)$ is the Vandermonde determinant of $\xi_1,\dots,\xi_k$, and

$$\hat{\xi} = \operatorname{diag}\{\xi_1,\dots,\xi_{2k}\}, \quad \Lambda_0 = E\cdot I.$$

In particular, for $k=1$ we have

$$F_2\left(\Lambda_0+\hat{\xi}/N\rho(E)\right) = C_N\Big(\frac{\sin(\pi(\xi_1-\xi_2))}{\pi^3(\xi_1-\xi_2)^3} - \frac{\cos(\pi(\xi_1-\xi_2))}{\pi^2(\xi_1-\xi_2)^2}\Big)(1+o(1)),$$

The last formula was proved also for real symmetric Wigner and general sample covariance matrices (see [20]).

Set

$$\lambda_1 = E + \frac{\xi}{2N\rho(E)}, \quad \lambda_2 = E - \frac{\xi}{2N\rho(E)}, \tag{1.7}$$

where $E\in(-2,2)$, ρ is defined in (1.1), and ξ is a real parameter varying in any compact set $K\subset\mathbb{R}$, and define

$$D_2 = F_2^{1/2}(E,E). \tag{1.8}$$

The main result of the paper is the following theorem:

Theorem 1.1. *For the real symmetric 1d block random band matrices H_N, $N=nW$ of (1.2)–(1.4) we have*

$$\lim_{n\to\infty}\bar{F}_2\Big(E+\frac{\xi}{2N\rho(E)},E-\frac{\xi}{2N\rho(E)}\Big) = \begin{cases} DS(\pi\xi), & W\gg n\gg 1;\\ (e^{-C^*\Delta-i\xi\pi\hat{\nu}}\cdot 1,1), & n=C_*W\\ 1, & 1\ll W\le n/\log^2 n,\end{cases}$$

where $DS(x)$ is defined in (1.6), $C^*=C_*/(2\pi\rho(E))^2$ with $\rho(E)$ of (1.1), and ε is any sufficiently small positive number. In this formula Δ is a Laplace-Bertrami operator on $\mathring{Sp}(2)=Sp(2)/Sp(1)\times Sp(1)$, $Sp(n)$ is a compact symplectic group of $2n\times 2n$ unitary symplectic matrices, and $(\cdot,\cdot)$ is an inner product on $L_2[\mathring{Sp}(2),d\mu]$, where $d\mu$ is the Haar measure on $\mathring{Sp}(2)$. $\hat{\nu}$ is an operator of multiplication by

$$\nu(Q) = 1-2(|Q_{12}|^2+|Q_{14}|^2) \tag{1.9}$$

on $\mathring{Sp}(2)$. Notice that the since $N=nW$, the transition happens at $W\sim\sqrt{N}$.

1.1 Notation

We denote by C, C_1, etc. various W and N-independent quantities below, which can be different in different formulas. Integrals without limits denote the integration (or the multiple integration) over the whole real axis, or over the Grassmann variables.

Moreover,

- W is a size of the block, and n is the number of blocks in a row, so $N=nW$ is the size of the matrix H of (1.2);
- $\mathbf{E}\{\dots\}$ is an expectation with respect to the measure (1.3);
- $a_\pm = \dfrac{iE\pm\sqrt{4-E^2}}{2} = e^{\pm i\alpha_0}$;
- $\sigma = \begin{pmatrix} 0 & 1\\ -1 & 0\end{pmatrix}$, $\sigma' = \begin{pmatrix} 0 & 1\\ 1 & 0\end{pmatrix}$;
- $D_0 = \begin{pmatrix} a_+ & 0\\ 0 & a_-\end{pmatrix}$, $\Lambda = \begin{pmatrix} \lambda_1 & 0\\ 0 & \lambda_2\end{pmatrix}$, $\hat{\xi} = \begin{pmatrix} \xi & 0\\ 0 & -\xi\end{pmatrix}$, $L = \begin{pmatrix} 1 & 0\\ 0 & -1\end{pmatrix}$;
- $D_{0,4} = \begin{pmatrix} D_0 & 0\\ 0 & D_0\end{pmatrix}$, $\hat{\xi}_4 = \begin{pmatrix} \hat{\xi} & 0\\ 0 & \hat{\xi}\end{pmatrix}$, $L_4 = \begin{pmatrix} L & 0\\ 0 & L\end{pmatrix}$;

- $\Lambda_0 = E \cdot I_2, \quad \Lambda_{0,4} = E \cdot I_4$;
- $U(n)$ is a group of $n \times n$ unitary matrices; unitary symplectic group $Sp(n)$ is a group of $2n \times 2n$ unitary matrices Q which admit the relation

$$Q \begin{pmatrix} 0 & I_n \\ -I_n & 0 \end{pmatrix} Q^t = \begin{pmatrix} 0 & I_n \\ -I_n & 0 \end{pmatrix}.$$

- $\mathring{U}(2) = U(2)/\big(U(1) \times U(1)\big), \quad \mathring{Sp}(2) = Sp(2)/\big(Sp(1) \times Sp(1)\big)$;
- $\mathbb{T} = \{z \in \mathbb{C} : |z| = 1\}, \quad \omega_A = \{z \in \mathbb{C} : |z| = 1 + A/n\}$;
- $d\mu$ is the Haar measure on $\mathring{U}(2)$, $d\mu$ is the Haar measure on $\mathring{Sp}(2)$;
- $c_\pm = 1 + a_\pm^{-2}; \quad t_* = (2\pi\rho(E))^2$
- We denote by $\bar{a}$ the vector (a_1, a_2);

2 Integral representation

The main aim of this section is to derive the following proposition

Proposition 2.1. *The second correlation function (1.5) of the characteristic polynomials for 1d real symmetric Gaussian block band matrices (1.2)–(1.4) can be represented as follows:*

$$F_2\Big(\Lambda_0 + \frac{\hat{\xi}}{2N\rho(E)}\Big) = C_{n,W} \int \exp\Big\{\frac{\beta W}{4} \sum_{j=2}^{n} \mathrm{Tr}\ (F_j - F_{j-1})^2\Big\} \tag{2.1}$$
$$\times \exp\Big\{\frac{W}{4} \sum_{j=1}^{n} \Big(\mathrm{Tr}\, F_j^2 - 2i\mathrm{Tr}\, F_j\big(\Lambda_{0,4} + \frac{\hat{\xi}_4}{2N\rho(E)}\big)\Big)\Big\} \prod_{j=1}^{n} \big(\det F_j\big)^{-W/2} \prod_{j=1}^{n} dF_j,$$

where $\Lambda_{0,4}$ and $\hat{\xi}_4$ are defined in Notation, $N = nW$, C_{nW} is some constant depending on W and n but not on $\hat{\xi}$, and $F_j \in Sp(2)$ are unitary symplectic 4×4 matrices.

Proof. Introduce the following Grassmann fields:

$$\Psi_l = \{\psi_{jl}^t\}_{j=1,..,n}^t, \quad \psi_{jl} = (\psi_{jl1}, \psi_{jl2}, \dots, \psi_{jlW})^t, \quad l = 1, 2.$$

Using (A.7) (see Appendix A) we obtain

$$F_2(\lambda_1, \lambda_2) = \mathbf{E}\Big\{ \int \exp\{-\Psi_1^+(\lambda_1 - H_N)\Psi_1 - \Psi_2^+(\lambda_2 - H_N)\Psi_2\} d\Psi \Big\}$$
$$= \int d\Psi\ \exp\Big\{ -\lambda_1 \Psi_1^+ \Psi_1 - \lambda_2 \Psi_2^+ \Psi_2 \Big\}$$
$$\times \mathbf{E}\Big\{ \exp\Big\{ \sum_{j<k} \sum_{\alpha,\gamma} H_{jk,\alpha\gamma}(\eta_{jk,\alpha\gamma} + \eta_{kj,\gamma\alpha}) + \sum_{j} \sum_{\alpha \le \gamma} H_{jk,\alpha\gamma}(\eta_{jk,\alpha\gamma} + \eta_{kj,\gamma\alpha}) \Big\} \Big\},$$

where

$$d\Psi = \prod_{j=1}^{n} \prod_{\alpha=1}^{W} \prod_{l=1}^{2} d\overline{\psi}_{jl\alpha} d\psi_{jl\alpha},$$
$$\eta_{jk,\alpha\gamma} = \overline{\psi}_{j1\alpha}\psi_{k1\gamma} + \overline{\psi}_{j2\alpha}\psi_{k2\gamma}, \quad \text{if } j \neq k \text{ or } \alpha \neq \gamma;$$
$$\eta_{jj,\alpha\alpha} = (\overline{\psi}_{j1\alpha}\psi_{j1\alpha} + \overline{\psi}_{j2\alpha}\psi_{j2\alpha})/2.$$

Averaging over (1.3), we get

$$F_2(\lambda_1, \lambda_2) = \int d\Psi\ \exp\{-\lambda_1 \Psi_1^+ \Psi_1 - \lambda_2 \Psi_2^+ \Psi_2\}$$
$$\times \exp\Big\{ \tfrac{1}{2} \sum_{j<k,\alpha,\gamma} J_{jk}\ (\eta_{jk,\alpha\gamma} + \eta_{kj,\gamma\alpha})^2 + \tfrac{1}{2} \sum_{j,\alpha<\gamma} J_{jj}\ (\eta_{jj,\alpha\gamma} + \eta_{jj,\gamma\alpha})^2 + \sum_{j,\alpha} J_{jj}\ \eta_{jj,\alpha\alpha}^2 \Big\}.$$

It is easy to see that

$$\begin{aligned}\frac{1}{2}\sum_{\alpha,\gamma}(\eta_{jk,\alpha\gamma}+\eta_{kj,\gamma\alpha})^2 =& -(\psi^+_{j1}\bar\psi_{j2})(\psi^t_{k1}\psi_{k2})-(\psi^+_{k1}\bar\psi_{k2})(\psi^t_{j1}\psi_{j2})-(\psi^+_{j1}\psi_{j1})(\psi^+_{k1}\psi_{k1})\\ &-(\psi^+_{j2}\psi_{j2})(\psi^+_{k2}\psi_{k2})-(\psi^+_{j1}\psi_{j2})(\psi^+_{k2}\psi_{k1})-(\psi^+_{k1}\psi_{k2})(\psi^+_{j2}\psi_{j1})\\ =&-\frac{1}{2}\mathrm{Tr}\,\tilde F_j\tilde F_k,\\ \sum_{l=1,2}\lambda_l\Psi^+_l\Psi_l =&\frac{1}{2}\sum_{j=1}^n\mathrm{Tr}\,\tilde F_j\Lambda_4\end{aligned}$$

where

$$\tilde F_j=\begin{pmatrix}\psi^+_{j1}\psi_{j1} & \psi^+_{j1}\psi_{j2} & 0 & \psi^+_{j1}\bar\psi_{j2}\\ \psi^+_{j2}\psi_{j1} & \psi^+_{j2}\psi_{j2} & \psi^+_{j2}\bar\psi_{j1} & 0\\ 0 & \psi^t_{j1}\psi_{j2} & \psi^+_{j1}\psi_{j1} & \psi^+_{j2}\psi_{j1}\\ \psi^t_{j2}\psi_{j1} & 0 & \psi^+_{j1}\psi_{j2} & \psi^+_{j2}\psi_{j2}\end{pmatrix},\quad \Lambda_4=\begin{pmatrix}\lambda_1&0&0&0\\0&\lambda_2&0&0\\0&0&\lambda_1&0\\0&0&0&\lambda_2\end{pmatrix}.$$

Applying the superbosonization formula (see Proposition A.1, Appendix A), we obtain

$$F_2(\lambda_1,\lambda_2)=C'_{nW}\int\exp\Big\{-\frac{1}{4}\sum_{j,k=1}^n J_{jk}\mathrm{Tr}\,F_jF_k-\frac{1}{2}\sum_{j,k=1}^n\mathrm{Tr}\,F_j\Lambda_4\Big\}\prod_{j=1}^n(\mathrm{det}F_j)^{-W/2}\prod_{j=1}^n dF_j,$$

where $\{F_j\}_{j=1}^n$ are unitary symplectic 4×4 matrices from $Sp(2)$, and C'_{nW} is some constant depending on W and n but not on λ_1, λ_2. Shifting $F_j\to iWF_j$ and plugging in (1.7), we get Proposition 2.1. □

3 Representation in the operator form

To study (2.1), we are going to apply the transfer matrix approach.
Namely, introduce

$$\mathcal{F}(X)=\exp\Big\{W\Big(\frac{1}{8}\,\mathrm{Tr}\,X^2-\frac{iE}{4}\mathrm{Tr}\,X-\frac{1}{4}\,\mathrm{Tr}\,\log X-C_+\Big)\Big\},\tag{3.1}$$

$$\mathcal{F}_\xi(X)=\mathcal{F}(X)\cdot\mathcal{F}_{n,\xi}(X),\quad \mathcal{F}_{n,\xi}(X):=\exp\Big\{-\frac{i}{8n\rho(E)}\,\mathrm{Tr}\,X\hat\xi_4\Big\},$$

where

$$C_+=\frac{a_+^2}{2}-iEa_+-\log a_+$$

is chosen in such a way that $|\mathcal{F}(X)|=1$ in the saddle-points (see (4.2) later).

Let also $K, K_\xi: Sp(2)\to Sp(2)$ be the operators with the kernels

$$K(X,Y)=\frac{W^3}{2\pi^3}\,\mathcal{F}(X)\,\exp\Big\{\frac{\beta W}{4}\mathrm{Tr}\,(X-Y)^2\Big\}\,\mathcal{F}(Y);\tag{3.2}$$

$$K_\xi(X,Y)=\frac{W^3}{2\pi^3}\,\mathcal{F}_\xi(X)\,\exp\Big\{\frac{\beta W}{4}\mathrm{Tr}\,(X-Y)^2\Big\}\,\mathcal{F}_\xi(Y).\tag{3.3}$$

Then Proposition 2.1 can be reformulated as

$$F_2\Big(E+\frac{\xi}{2N\rho(E)},E-\frac{\xi}{2N\rho(E)}\Big)=\tilde C_{n,W}(K_\xi^{n-1}\mathcal{F}_\xi,\bar{\mathcal{F}}_\xi),\tag{3.4}$$

where $(\cdot,\cdot)$ is a standard inner product in $Sp(2)$ with respect to the Haar measure $d\mu$, and $\tilde C_{nW}$ is some constant depending on W and n but not on ξ.

For an arbitrary compact operator M denote by $\lambda_j(M)$ the jth (by modulus) eigenvalue of M, so that $|\lambda_0(M)| \geq |\lambda_1(M)| \geq \dots$.

Since K_ξ is a compact operator, one can rewrite

$$(K_\xi^{n-1}\mathcal{F}_\xi, \bar{\mathcal{F}}_\xi) = \sum_{j=0}^{\infty} \lambda_j^{n-1}(K_\xi) c_j, \quad with \quad c_j = (\mathcal{F}_\xi, \psi_j)(\bar{\mathcal{F}}_\xi, \tilde{\psi}_j),$$

where $\{\psi_j\}$ are eigenvectors corresponding to $\{\lambda_j(K_\xi)\}$, and $\{\tilde{\psi}_j\}$ are the eigenvectors of $K_\xi^\dagger$. Similar equality is true if we replace K_ξ and $\mathcal{F}_\xi$ by K and $\mathcal{F}$. Hence, to study (2.1), it suffices to study the eigenvalues and eigenvectors of K_ξ, K.

4 Sketch of the proof of Theorem 1.1

As was mentioned above, we are interested in the analysis of the spectral properties of K_ξ of (3.3) (see (3.4)). It appears that it is simpler to work with the resolvent analog of (3.4)

$$(K_\xi^{n-1} f, g) = -\frac{1}{2\pi i} \oint_{\mathcal{L}} z^{n-1} (G_\xi(z) f, g) dz, \quad G_\xi(z) = (K_\xi - z)^{-1}, \tag{4.1}$$

where $\mathcal{L}$ is any closed contour which encloses all eigenvalues of K_ξ.

The idea of the proof is very close to [30]–[32]. To outline it, we start with the following definition

Definition 4.1. *We say that the operator $\mathcal{A}_{n,W}$ is equivalent to $\mathcal{B}_{n,W}$ ($\mathcal{A}_{n,W} \sim \mathcal{B}_{n,W}$) on some contour $\mathcal{L}$ if*

$$\int_{\mathcal{L}} z^{n-1} ((\mathcal{A}_{n,W} - z)^{-1} f, \bar{g}) dz = \int_{\mathcal{L}} z^{n-1} ((\mathcal{B}_{n,W} - z)^{-1} f, \bar{g}) dz \, (1 + o(1)), \quad n, W \to \infty,$$

with some particular functions f, g depending of the problem.

The aim is to find some operator equivalent to K_ξ whose spectral analysis is more accessible. Now we are going to discuss how this was done on the ideological level. The specific choice of the contour $\mathcal{L}$ and functions f, g for each step will be discussed in details in Section 6.

It is easy to check that the stationary points of the function $\mathcal{F}$ of (3.1) are

$$\begin{aligned} X_+ = a_+ \cdot I_4, \quad X_- = a_- \cdot I_4; \\ X_\pm(Q) = Q D_{0,4} Q^*, \quad Q \in \mathring{Sp}(2) \end{aligned} \tag{4.2}$$

where $a_\pm$, $D_{0,4}$ are defined in Notation. Recall also that the value of $|\mathcal{F}|$ at points (4.2) is 1.

The first step in the proof of Theorem 1.1 is to apply the saddle-point approximation. Roughly speaking, we show that if we introduce the projection $\Pr_s$ onto the $W^{-1/2}\log W$-neighbourhoods of the saddle points X_+, X_- and the saddle "surface" $X_\pm$, then in the sense of Definition 4.1

$$K_\xi \sim \underset{s}{\Pr} K_\xi \underset{s}{\Pr} =: K_{s,\xi}.$$

Moreover, one can show that only the neighbourhood of the saddle "surface" $X_\pm$ gives the main contribution to the integral. The proof is based on a study of a quadratic approximation of a function $\mathcal{F}$ of (3.1). Let us also emphasize, that for the block band matrices (1.2)–(1.4) this step is much simpler than for the model considered in [30]–[32] due to the large coefficient W in the exponent of $\mathcal{F}$. This analysis will be performed in details in Section 5.

To study the operator $K_{s,\xi}$ near the saddle "surface" $X_\pm$ we use the "polar coordinates". Namely, the matrices from $Sp(2)$ have two eigenvalues $a_j, b_j \in \mathbb{T} = \{z : |z| = 1\}$ of the multiplicity two and can be considered as quaternion 2×2 matrices. In this language F_j are quaternion unitary matrices, and so they can be diagonalized by the quaternion unitary 2×2 matrices from $\mathring{Sp}(2)$ (see, e.g., [22], Chapter 2.4).

Change the variables to $F_j = Q_j^* A_{j,4} Q_j$, where $A_{j,4} = \operatorname{diag}\{a_{1j}, a_{2j}, a_{1j}, a_{2j}\}$, eigenvalues $a_{1j}, a_{2j} \in \mathbb{T}$, and $Q_j \in \mathring{Sp}(2)$. Then dF_j of (2.1) becomes (see, e.g., [22])

$$\frac{\pi^2}{12}(a_{1j} - a_{2j})^4 \, d\bar{a}_j \, d\mu(Q_j),$$

where

$$d\bar{a}_j = \frac{da_{1j}}{2\pi i} \frac{da_{2j}}{2\pi i},$$

and $d\mu(Q_j)$ is the normalized to unity Haar measure on the symplectic group $\mathring{Sp}(2)$. Thus we get

$$(K_\xi^{n-1}\mathcal{F}_\xi, \bar{\mathcal{F}}_\xi) = \frac{\pi^{2n}}{12^n} \int (a_{11} - a_{21})^2 F_\xi(a_{11}, a_{21}, Q_1)(a_{1n} - a_{2n})^2 F_\xi(a_{1n}, a_{2n}, Q_n)$$
$$\times \prod_{j=1}^{n-1} \Big((a_{1j} - a_{2j})^2 (a_{1,j+1} - a_{2,j+1})^2 K_\xi(F_j, F_{j+1}) \Big) \prod_{j=1}^{n} d\bar{a}_j \, d\mu(Q_j).$$

Introduce

$$t = (a_1 - a_2)(a_1' - a_2'). \tag{4.3}$$

Then we obtain

$$F_2\Big(E + \frac{\xi}{2N\rho(E)}, E - \frac{\xi}{2N\rho(E)}\Big) = \tilde{C}'_{n,W}(K_\xi^{n-1} f, \bar{f}), \tag{4.4}$$

where now $(\cdot,\cdot)$ is a standard inner product in $L_2[\mathbb{T}^2] \times L_2[\mathring{Sp}(2), d\mu(Q)]$, and $\tilde{C}'_{nW}$ is some constant depending on W and n but not on ξ. Here

$$f(a_1, a_2, Q) = (a_1 - a_2)^2 F_\xi(a_1, a_2, Q), \tag{4.5}$$

and $K_\xi = F_{n,\xi} K F_{n,\xi}$ is an integral operator in $L_2[\mathbb{T}^2] \times L_2[\mathring{Sp}(2), d\mu(Q)]$ defined by the kernel

$$K_\xi(X, Y) = F_{n,\xi}(a_1, a_2, Q) K(a_1, a_2, Q; a_1', a_2', Q') F_{n,\xi}(a_1', a_2', Q'), \tag{4.6}$$

where

$$K(a_1, a_2, Q; a_1', a_2', Q') = A_a(\bar{a}, \bar{a}') K_*(t, Q_1, Q_2);$$
$$K_*(t, Q, Q') := \frac{\beta^2 W^2 t^2}{6} \cdot \exp\{-t\beta W S\big(Q(Q')^*\big)\}, \quad S(Q) = |Q_{12}|^2 + |Q_{14}|^2; \tag{4.7}$$
$$F_{n,\xi}(a, b, Q) = \exp\{-i\xi\pi \cdot \nu(a - b, Q)/n\};$$
$$\nu(p, Q) = \frac{p}{4\pi\rho(E)} \operatorname{Tr} Q L_4 Q^* L_4 = \frac{p}{2\pi\rho(E)}(1 - 2S(Q)) \tag{4.8}$$

with t of (4.3). K_* here is a contribution of the symplectic group $\mathring{Sp}(2)$ into operator K, and $\exp\{-i\xi\pi \cdot \nu(x, Q)/n\}$ comes from the $1/n$-order perturbation $\mathcal{F}_{n,\xi}$ of $\mathcal{F}$ appearing in

$\mathcal{F}_\xi$ (see (3.1)). Operator A_a is a contribution of eigenvalues a_1, a_2 and it has the form

$$A_a(\bar{a};\bar{a}') = A(a_1,a_1')A(a_2,a_2'), \tag{4.9}$$
$$A(a,a') = \Big(\frac{W}{2\pi}\Big)^{1/2} e^{-W\Phi(a,a')};$$
$$\Phi(x,y) = \frac{\beta}{2}(x-y)^2 - \frac{1}{2}\varphi_0(x) - \frac{1}{2}\varphi_0(y) + \Re\varphi_0(a_+);$$
$$\varphi_0(x) = x^2/2 - ixE - \log x. \tag{4.10}$$

Observe that the operator $K_*(t,Q,Q')$ with some $t>0$ is self-adjoint and its kernel depends only on $S\big(Q(Q')^*\big)$. Thus by the standard representation theory arguments (see e.g. [19], [36]), its eigenfunctions are the the same as for Laplace-Bertrami operator on $Sp(2)$. More precisely:

Proposition 4.2. *Consider any self-adjoint integral operator M in $L_2[\mathring{Sp}(2), d\mu(Q)]$. If its kernel $M(Q,Q')$ depends only on $Q(Q')^*$, then its eigenvectors coincide with eigenvectors of Laplace-Bertrami operator on $\mathring{Sp}(2)$. Moreover, if the subspace*

$$L_2[S, d\mu(Q)] \subset L_2[\mathring{Sp}(2), d\mu(Q)]$$

of the functions depending on $S(Q)$ (see (4.7)) only is invariant under M, then it can be diagonalized by the eigenfunctions

$$\phi_j(Q) = (-1)^j P_{2j}(\sqrt{S(Q)}), \tag{4.11}$$

where $P_{2j}(x)$ are orthogonal with respect to the weight $(1-x^2)x^3$ on $[0,1]$ polynomials of degree $2j$, $\phi_0(x)=1$ (polynomials P_{2j} can be written as $P_{2j}(x) = c_j F_{hg}(-j, j+3, 2; 1-x^2)$, where F_{hg} is a hypergeometric function, and c_j is a normalization constant, see [19], Ch. 5). In addition, the following holds

$$(2x^2-1)P_{2j}(x) = \frac{j+3}{2j+3}P_{2j+2}(x) + \frac{j}{2j+3}P_{2j-2}(x), \tag{4.12}$$

so the operator $\hat{\nu}$ of multiplication on $\nu(x,Q)$ of (4.8) is three diagonal in basis (4.11), and

$$(\hat{\nu}\cdot\phi_0, \phi_0) = 0. \tag{4.13}$$

If $M(Q_1,Q_2) = K_*(t,Q_1,Q_2)$ of (4.7), then the corresponding eigenvalues $\{\lambda_j(t)\}_{j=0}^\infty$, if $t>d>0$, where d is some absolute positive constant, have the form

$$\lambda_j(t) = 1 - \frac{(j+1)(j+2)}{Wt} + O((j^2/Wt)^2) + O(e^{-tW}). \tag{4.14}$$

The proof of the proposition can be found in Appendix B.

Notice that, according to Proposition 4.2, since $\mathcal{F}(Q)$, $\mathcal{F}_\xi(Q)$ are the functions of $S(Q)$ only, in what follows we can consider restrictions of K_ξ, K, and K_* of (4.7) to $L_2[S, d\mu(Q)]$ (to simplify notations we will denote these restrictions by the same letters).

In addition, it follows from Proposition 4.2 that if we introduce the following basis in $L_2[\mathbb{R}^2]\times L_2[S,d\mu(Q)]$

$$\Psi_{\bar{k},j}(\bar{a},Q) = \Psi_{\bar{k}}(\bar{a})\phi_j(Q),$$
$$\Psi_{\bar{k}}(\bar{a}) = \psi_{k_1}(a_1)\psi_{k_2}(a_2),$$

where $\bar{k}=(k_1,k_2)$, and $\{\psi_k(x)\}_{k=0}^{\infty}$ is a certain basis in $L_2[\mathbb{R}]$, then the matrix of K of (4.7) in this basis has a "block diagonal structure", which means that

$$\begin{aligned}&(K\Psi_{\bar{k}',j},\Psi_{\bar{k},j_1})=0,\quad j\neq j_1 \\ &(K\Psi_{\bar{k}',j},\Psi_{\bar{k},j})=(K_j\Psi_{\bar{k}'},\Psi_{\bar{k}}) \\ &=\int\lambda_j(t)A_a(\bar{a},\bar{a}')\psi_{k_1}(a_1)\psi_{k_2}(a_2)\psi_{k_1'}(a_1')\psi_{k_2'}(a_2')\frac{da_1da_2da_1'da_2'}{(2\pi i)^4}.\end{aligned} \tag{4.15}$$

The next step in the proof of Theorem 1.1 is to show that we can restrict the number of ϕ_j to

$$l=\max\{1,[\log n\cdot\sqrt{W}/\sqrt{n}]\}. \tag{4.16}$$

l is chosen in such a way that $l^2n/W\gg\log n$. More precisely, we are going to show that in the sense of Definition 4.1

$$K_{s,\xi}\sim\mathcal{P}_lK_{s,\xi}\mathcal{P}_l=:K_{s,l,\xi},$$

where $\mathcal{P}_l$ is the projection on the linear span of $\{\Psi_{\bar{k},j}(\bar{a},Q)\}_{j\leq l-1}$.

For the further resolvent analysis we want to integrate out $\bar{a}$ to change t in the definition of K_* and a_1-a_2, $a_1'-a_2'$ in the definition of $F_{n,\xi}$ (see (4.3), (4.6) – (4.7)) by their saddle-point values $t_*=(a_+-a_-)^2=4\pi^2\rho(E)^2$ and $a_+-a_-=2\pi\rho(E)$ correspondingly. We are going to show that only the top eigenvalue of A gives a contribution. More precisely we want to show that in the sense of Definition 4.1

$$\lambda_0(K_{s,l})^{-1}K_{s,l,\xi}\sim\hat{K}_{*\xi,l} \tag{4.17}$$

where

$$\begin{aligned}&\hat{K}_{*\xi,l}=(\lambda_0(K_{*0}))^{-1}P_l\,K_{*\xi}\,P_l, \\ &K_{*\xi}(Q_1,Q_2)=\frac{W^2t_*^2\beta^2}{6}\cdot e^{-\beta t_*WS(Q_1Q_2^*)}\cdot e^{-i\xi\pi(\nu(2\pi\rho(E),Q_1)+\nu(2\pi\rho(E),Q_2))/n}\end{aligned} \tag{4.18}$$

and P_l is the projection on $\{\phi_j(Q)\}_{j\leq l-1}$. Here K_{*0} is $K_{*\xi}$ with $\xi=0$.

Now (4.17), (4.1) and Definition 4.1 give

$$\begin{aligned}F_2\Big(E+\frac{\xi}{2N\rho(E)},E-\frac{\xi}{2N\rho(E)}\Big)&=C_{n,W}\Big(\hat{K}_{*\xi,l}^{n-1}f_\xi,\bar{f}_\xi\Big)(1+o(1)) \\ &=C_{n,W}\lambda_0(K_{s,l})^{n-1}f_0^2(\hat{K}_{*\xi,l}^{n-1}1,1)(1+o(1)),\end{aligned}$$

where $f_0=(f,\Psi_{\bar{0}})$, and we used that f_ξ asymptotically can be replaced by $f\otimes 1$, where f does not depend on ξ and Q_j. Similarly

$$D_2=C_{n,W}\Big(\hat{K}_{*0,l}^{n-1}f_\xi,\bar{f}_\xi\Big)(1+o(1))=C_{n,W}\lambda_0(K_{s,l})^nf_0^2(\hat{K}_{*0,l}^{n-1}1,1)(1+o(1)).$$

According to Proposition 4.2, $\phi_0(Q)=1$ is an eigenvector of $\hat{K}_{*0}$ of (4.18) with $\xi=0$ and the corresponding eigenvalue is 1, thus

$$(\hat{K}_{*0,l}^{n-1}1,1)=1.$$

Hence

$$\bar{F}_2\Big(E+\frac{\xi}{2N\rho(E)},E-\frac{\xi}{2N\rho(E)}\Big)=(\hat{K}_{*\xi,l}^{n-1}1,1)(1+o(1)). \tag{4.19}$$

Recall that according to Proposition 4.2 the eigenvectors of $\hat{K}_{*0,l}$ are (4.11) and the corresponding eigenvalues are (see (4.14))

$$\lambda_j:=\lambda_j(t_*)=1-j(j+3)/t_*W+O((j(j+3)/W)^2),\quad j=0,1\ldots,l-1. \tag{4.20}$$

Moreover, it follows from (4.6)–(4.7) that

$$\hat{K}_{*\xi,l} = \hat{K}_{*0,l} - n^{-1}\pi i\xi\hat{\nu}_l + o(n^{-1}), \quad \hat{\nu}_l = P_l\hat{\nu}P_l, \tag{4.21}$$

where $\hat{\nu}$ is the operator of multiplication by (1.9), and $o(1/n)$ means some operator whose norm is $o(1/n)$. Thus the eigenvalues of $\hat{K}_{*\xi,l}$ are in the n^{-1}-neighbourhood of λ_j.

In the localized regime $W^{-1} \gg n^{-1}$ we have $l = 1$, thus only $\lambda_0(\hat{K}_{*\xi})$ contributes to (4.19). Since (see Proposition 4.2)

$$(\hat{\nu}\,1, 1) = 0,$$

we get

$$\lambda_0(\hat{K}_{*\xi}) = 1 + o(n^{-1}),$$

and so the limit of (4.19) is 1 (see the end of Section 6 for more details).

In the delocalized regime all eigenvalues of $K_{*\xi,l}$ contribute to (4.19), but $K_{*0,l}^{n-1} \to I$ (roughly speaking, this means that the second term in the r.h.s. of (4.20) does not give a contribution). Hence we have

$$\hat{K}_{*\xi,l} \approx 1 - n^{-1}i\xi\pi\hat{\nu}_l \Rightarrow (K_{*\xi}^{n-1}\,1, 1) \to (e^{-i\xi\pi\hat{\nu}}1, 1) = DS(\pi\xi)$$

with $DS(\pi\xi)$ of (1.6) (see (B.2) and the end of Section 6 for more details).

In the critical regime $W^{-1} = C_*n^{-1}$ all eigenvalues of $\hat{K}_{*\xi,l}$ contribute, but now both second term in the r.h.s. of (4.20) and $1/n$-order term in the r.h.s. of (4.21) make an impact.

As it was mentioned above, the Laplace-Bertrami operator Δ on $L_2[S, d\mu]$ has eigenvectors (4.11) with corresponding eigenvalues

$$\lambda_j^* = j(j+3).$$

Thus, $1 - n^{-1}C^*\Delta$ with $C^* = C_*/t_*$ has the same basis of eigenvectors with eigenvalues $1 - j(j+3)/t_*W$.

Recall that we are interested in $j \le l - 1 \sim \log W$ (since P_l is the projection on $\{\phi_j\}_{j\le l-1}$). Hence, according to (4.20)– (4.21), in the regime $W^{-1} = C_*n^{-1}$ we can write

$$\hat{K}_{*\xi,l} = P_l\big(1 - n^{-1}(C^*\Delta + i\xi\pi\nu)\big)P_l + o(n^{-1}),$$

which implies

$$(\hat{K}_{*\xi,l}^{n-1}1, 1) \to (e^{-C^*\Delta - i\xi\pi\hat{\nu}}1, 1), \tag{4.22}$$

and finishes the proof of Theorem 1.1. The detailed proof of (4.22) is given in Section 6 (see Lemma 6.5).

5 Saddle-point analysis

Recall that the stationary points of the function $\mathcal{F}$ of (3.1) are defined in (4.2).

We start the proof from the restriction of the integration with respect to $\bar{a}_i, \bar{a}_i'$ by the neighbourhood of $a_\pm$. Set

$$\begin{aligned}
\Omega_+ =&\{x : |x - a_+| \le \log W/W^{1/2}\}, \quad \Omega_- = \{x : |x - a_-| \le \log W/W^{1/2}\},\\
\widetilde{\Omega}_\pm =&\{a_1, a_1' \in \Omega_+, a_2, a_2' \in \Omega_-\},\\
\widetilde{\Omega}_+ =&\{a_1, a_1', a_2, a_2' \in \Omega_+\},\\
\widetilde{\Omega}_- =&\{a_1, a_1', a_2, a_2' \in \Omega_-\}
\end{aligned} \tag{5.1}$$

and let $\mathbf{1}_{\widetilde{\Omega}_\pm}$, $\mathbf{1}_{\widetilde{\Omega}_+}$ $\mathbf{1}_{\widetilde{\Omega}_-}$ be indicator functions of the above domains.

Lemma 5.1. *Given $A(a,a')$ of (4.9), we have*

$$\int_{\mathbb{T}\setminus(\Omega_+\cup\Omega_-)} |A(a,a')||da'| \le Ce^{-c\log^2 W}. \tag{5.2}$$

Proof. Recall that

$$a_\pm = e^{\pm i\alpha_0},$$

and write for the parametrization $a = e^{i\varphi}$, $a' = e^{i\varphi'}$

$$\begin{aligned}
-\Re\Phi(e^{i\varphi}, e^{i\varphi'}) =& -\beta(\cos\varphi - \cos\varphi')^2/2 + \beta(\sin\varphi - \sin\varphi')^2/2 - \frac{\sin^2\varphi + \sin^2\varphi'}{2} \\
&+ \frac{E(\sin\varphi + \sin\varphi')}{2} + \sin^2\alpha_0 - E\sin\alpha_0 \\
&= -\beta(\cos\varphi - \cos\varphi')^2/2 + \beta(\sin\varphi - \sin\varphi')^2/2 \\
&- (\sin\varphi - \sin\alpha_0)^2/2 - (\sin\varphi' - \sin\alpha_0)^2/2 \\
&\le \beta(\sin\varphi - \sin\varphi')^2/2 - (\sin\varphi - \sin\alpha_0)^2/2 - (\sin\varphi' - \sin\alpha_0)^2/2 \\
&\le -(1-2\beta)(\sin\varphi - \sin\alpha_0)^2/2 - (1-2\beta)(\sin\varphi' - \sin\alpha_0)^2/2.
\end{aligned}$$

Here we have used $\sin\alpha_0 = E/2$. We have also for $a' \in \mathbb{T}\setminus(\Omega_+\cup\Omega_-)$

$$|\sin\varphi' - \sin\alpha_0| \ge C\log W/\sqrt{W}.$$

Since $\beta < 1/4$, this implies (5.2) □

Lemma 5.1 yields that

$$\int dQ'd\bar{a}'(1 - \mathbf{1}_{\widetilde{\Omega}_\pm} - \mathbf{1}_{\widetilde{\Omega}_+} - \mathbf{1}_{\widetilde{\Omega}_-})\|K\| \le e^{-c\log^2 W} \tag{5.3}$$

Let us prove the following simple proposition

Proposition 5.2. *Let the matrix $H(z)$ have the block form*

$$H(z) = \begin{pmatrix} H_{11}(z) & H_{12}(z) \\ H_{21}(z) & H_{22}(z) \end{pmatrix}.$$

Then

$$\begin{aligned}
G(z) := H^{-1}(z) &= \begin{pmatrix} G_{11} & -G_{11}H_{12}H_{22}^{-1} \\ -H_{22}^{-1}H_{21}G_{11} & H_{22}^{-1} + H_{22}^{-1}H_{21}G_{11}H_{12}H_{22}^{-1} \end{pmatrix} \\
G_{11} &= (H_{11} - H_{12}H_{22}^{-1}H_{21})^{-1},
\end{aligned} \tag{5.4}$$

If H_{22}^{-1} is an analytic function for $|z| > 1-\delta$, and $\|H_{22}^{-1}\| \le C$, then

$$\begin{aligned}
&\oint_{\omega_A} z^{n-1}(G(z)f,g)dz = \oint_{\omega_A} z^{n-1}(G_{11}f^{(1)}(z), g^{(1)}(z))dz + O(e^{-nc}) \\
&f^{(1)}(z) = f_0 - H_{12}H_{22}^{-1}f_1, \quad g^{(1)}(z) = g_0 - H_{21}^T(H_{22}^T)^{-1}g_1
\end{aligned} \tag{5.5}$$

where $\omega_A = \{z : |z| = 1 + A/n\}$, $f = (f_0, f_1)$, $g = (g_0, g_1)$ where f_0 and g_0 are the projection of f and g on the subspace corresponding to H_{11}, while f_1 and g_1 are the projection of f and g on the subspace corresponding to H_{22}.

Proof. Formula (5.4) is the well-known block matrix inversion formula. Now apply the formula (5.4) and write

$$\oint_{\omega_A} z^{n-1}(G(z)f,g)dz = \oint_{\omega_A} z^{n-1}(G_{11}f^{(1)}(z), g^{(1)}(z))dz + \oint_{\omega_A} z^{n-1}(H_{22}^{-1}f_1, g_1)dz.$$

For the second integral change the integration contour from ω_A to $|z| = 1 - \delta$. Then the inequality

$$|z|^{n-1} \le (1-\delta)^{n-1} \le Ce^{-nc}$$

yields (5.5). □

Notice that since $\|K\| \le 1$ and $|\mathcal{F}_{n,\xi}| \le 1 + C/n$, we can find such A that all eigenvalues of K_ξ lie inside $\omega_A = \{z : |z| = 1 + A/n\}$.

Set

$$H_{11}(z) = H_{11} - z = (\mathbf{1}_{\widetilde{\Omega}_\pm} K_\xi \mathbf{1}_{\widetilde{\Omega}_\pm}) \oplus (\mathbf{1}_{\widetilde{\Omega}_+} K_\xi \mathbf{1}_{\widetilde{\Omega}_+}) \oplus (\mathbf{1}_{\widetilde{\Omega}_-} K_\xi \mathbf{1}_{\widetilde{\Omega}_-}) - z = K_{\xi,\pm} \oplus K_{\xi,+} \oplus K_{\xi,-} - z.$$

Then (5.3) yields

$$\|H_{22}\| + \|H_{12}\| + \|H_{21}\| \le Ce^{-c\log^2 W}.$$

Therefore, for any $|z| > \frac{1}{2}$

$$\|H_{12}(H_{22} - z)^{-1} H_{21}\| \le Ce^{-c\log^2 W}.$$

Moreover, it will be proven below that

$$\|(H_{11} - z)^{-1}\| \le Cn, \quad z \in \omega_A,$$

and so for G_{11} of (5.4) we have

$$\|G_{11} - (H_{11} - z)^{-1}\| \le e^{-c\log^2 W/2}.$$

Here we have used $W \ge n^\varepsilon$. Thus we obtain by Proposition 5.2

$$\oint_{\omega_A} z^{n-1}(G_\xi(z)f, g)dz = \oint_{\omega_A} z^{n-1}((H_{11} - z)^{-1}f, g)dz + O(e^{-c\log^2 W/2}) + O(e^{-nc_1}), \quad (5.6)$$

where $G_\xi(z)$ is a resolvent of K_ξ (see (4.1)). In view of the block structure of H_{11}, its resolvent also has a block structure, hence

$$\begin{aligned}\oint_{\omega_A} z^{n-1}(G_\xi(z)f, g)dz = \oint_{\omega_A} z^{n-1}(G_{\xi,\pm}(z)f_\pm, g_\pm)dz + \oint_{\omega_A} z^{n-1}(G_{\xi,+}(z)f_+, g_+)dz \\ + \oint_{\omega_A} z^{n-1}(G_{\xi,-}(z)f_-, g_-)dz = I_{\xi,\pm} + I_{\xi,+} + I_{\xi,-}, \quad (5.7)\end{aligned}$$

where

$$G_{\xi,\pm} = (K_{\xi,\pm} - z)^{-1}, \quad G_{\xi,+}(z) = (K_{\xi,+} - z)^{-1}, \quad G_{\xi,-}(z) = (K_{\xi,-} - z)^{-1}$$

and $f_\pm, f_+, f_-, g_\pm, g_+, g_-$ are projections of f and g onto the subspaces corresponding to $K_{\xi,\pm}$, $K_{\xi,+}$, $K_{\xi,-}$. One can perform similar analysis for K instead of K_ξ and define $I_\pm$, I_+, and I_-.

In the next sections we are going to study each integral $I_{\xi,\pm}$, $I_{\xi,+}$, and $I_{\xi,-}$ separately. It will be shown below (see Section 7) that that $I_{\xi,+}$ and $I_{\xi,-}$ are exponentially small comparable to $I_{\xi,\pm}$, so the main task is to study $I_{\xi,\pm}$.

6 Analysis of $I_{\xi,\pm}$

As was mentioned in Section 4, to analyze $K_\pm$ and $K_{\xi,\pm}$ we are going to use the polar decomposition (4.6)–(4.10).

We start with the analysis of operator A_a of (4.9) in the domain $\widetilde{\Omega}_\pm$ of (5.1).

To this end, we are going to consider quadratic approximation of $A(a,a')$ defined in (4.9). Make a change of variables

$$a_{1i}=a_+(1+i\theta_+\widetilde{a}_{1i}/\sqrt{W}),\quad a_{2i}=a_-(1+i\theta_-\widetilde{a}_{2i}/\sqrt{W}), \tag{6.1}$$

where $\theta_\pm$ are some complex constants with $|\theta_\pm|=1$ which will be determined later (see (6.5)). Notice that the Jacobian of (6.1) is a constant depending on n,W but not on ξ, thus it does not contribute to $\tilde{C}'_{n,W}$ (see (4.4)). Define

$$\begin{aligned}A^+(\tilde{a},\tilde{a}')&=1_{\Omega_+}A\Big(a_+(1+i\theta_+\tilde{a}/\sqrt{W}),a_+(1+i\theta_+\tilde{a}'/\sqrt{W})\Big)1_{\Omega_+},\\ A^-(\tilde{a},\tilde{a}')&=1_{\Omega_-}A\Big(a_-(1+i\theta_-\tilde{a}/\sqrt{W}),a_-(1+i\theta_-\tilde{a}'/\sqrt{W})\Big)1_{\Omega_-}.\end{aligned} \tag{6.2}$$

Then

$$\begin{aligned}&K_{\xi,\pm}(a_1,a_2,Q;a_1',a_2',Q')\\ &\qquad=A^+(\tilde{a}_1,\tilde{a}_1')A^-(\tilde{a}_2,\tilde{a}_2')K_*(t,Q,Q')e^{-\frac{i\xi\pi}{n}\left(\nu(a_1-a_2,Q)+\nu(a_1'-a_2',Q')\right)}.\end{aligned} \tag{6.3}$$

Since $\varphi_0''(a_+)=c_+$ (see (4.10) and (1.1)), it is easy to see that the kernel A^+ of (6.2) takes the form

$$\begin{aligned}A^+(\tilde{a},\tilde{a}')&=A^+_*(\tilde{a},\tilde{a}')(1+W^{-1/2}\hat{p}_+(\tilde{a}))(1+W^{-1/2}\hat{p}_+(\tilde{a}'))+O(e^{-c\log^2 W}),\\ A^+_*(\tilde{a},\tilde{a}')&=\frac{a_+\theta_+}{\sqrt{2\pi}}\exp\Big\{(a_+\theta_+)^2\big[\beta(\tilde{a}-\tilde{a}')^2/2-c_+\tilde{a}^2/4-c_+(\tilde{a}')^2/4\big]\Big\}\\ \hat{p}_+(\tilde{a})&=ic_{3+}\tilde{a}^3-c_{4+}\tilde{a}^4W^{-1/2}-ic_{5+}\tilde{a}^5W^{-1}+\dots\end{aligned} \tag{6.4}$$

where the coefficients $c_{3+},c_{4+},\dots$ are expressed in terms of the derivatives of φ_0 at a_+. Similarly A^- of (6.2) can be approximated via A^-_* defined similarly to A^+_* in (6.4).

It is easy to check that for $\beta<1/4$ the real parts of the eigenvalues $\alpha_{1,+}$, $\alpha_{2,+}$ of the quadratic form

$$\begin{pmatrix} a_+^2(\frac{c_+}{2}-\beta) & a_+^2\beta\\ a_+^2\beta & a_+^2(\frac{c_+}{2}-\beta)\end{pmatrix}$$

in the exponent of A^+_* of (6.4) are positive. Same is true for A^-_*. Denote

$$\theta_\pm=(|\kappa_\pm|/\kappa_\pm)^{1/2},\quad \kappa_\pm=(\alpha_{1,\pm}\alpha_{2,\pm})^{1/2}=a_\pm^2\big(c_\pm^2/4-\beta c_\pm\big)^{1/2}, \tag{6.5}$$

with $c_\pm$ of (1.1). Notice that $\theta_\pm$ is defined in such a way that

$$\Re(\theta_\pm^2\alpha_{1,\pm})>0,\quad \Re(\theta_\pm^2\alpha_{2,\pm})>0.$$

Now introduce the orthonormal bases

$$\psi_k^\pm(\tilde{a})=|\kappa_\pm|^{1/4}H_k(|\kappa_\pm|^{1/2}\tilde{a})e^{-|\kappa_\pm|\tilde{a}^2/2}, \tag{6.6}$$

where $\{H_k(x)\}$ are Hermite polynomials which are orthonormal with the weight e^{-x^2}:

$$H_k(x)=(2^{k-1/2}k!\sqrt{2\pi})^{-1/2}e^{x^2}(\frac{d}{dx})^ke^{-x^2}.$$

Below we will need the following lemma

Lemma 6.1. *(i) Let κ_+, κ_- be defined as in (6.5). Then the matrices of the operators A_*^+ and A_*^- are diagonal in the basis $\{\psi_k^+\}$ and $\{\psi_k^-\}$, and corresponding eigenvalues have the form*

$$\lambda_k^\pm = \lambda_k(A_*^\pm) = \lambda_0^\pm \cdot q_\pm^k, \quad k = 0, 1, 2 \dots \tag{6.7}$$

with

$$\lambda_0^\pm = (\kappa_\pm/a_\pm^2 + c_\pm/2 - \beta)^{-1/2}, \tag{6.8}$$
$$q_\pm = \frac{\beta}{\kappa_\pm/a_\pm^2 + c_\pm/2 - \beta}, \quad |q_\pm| < 1.$$

Notice that $|q_\pm| < 1$ implies

$$|\lambda_0^\pm| \le \beta^{-1/2}. \tag{6.9}$$

The matrices of operators A^+ and A^- of (6.2) have the form

$$\begin{aligned}(A^\pm)_{k,k} &= \lambda_0^\pm \cdot q_\pm^k + O(1/W), \\ (A^\pm)_{k,k'} &= O(W^{-1/2})(\delta_{|k-k'|,1} + \delta_{|k-k'|,3}) \\ &\quad + O(W^{-1})\delta_{|k-k'|,2} + O(W^{-(|k-k'|-3)/2}), \quad k \ne k'.\end{aligned} \tag{6.10}$$

(ii) The eigenvalues of operator

$$A_\pm = 1_{\bar\Omega_\pm}\big(\lambda_0(t)A_a\big)1_{\bar\Omega_\pm} \tag{6.11}$$

are $\lambda_0^+\lambda_0^- q_+^k q_-^l + O(1/W)$, $k, l = 0, 1, ..$ and they are solutions of the equation

$$(A_\pm)_{0,0} - z - (A_\pm)^{(12)}((A_\pm)^{(22)} - z)^{-1}(A_\pm)^{(21)} = 0, \tag{6.12}$$

where

$$A_\pm = \begin{pmatrix} A_{00} & A^{(12)} \\ A^{(21)} & A^{(22)} \end{pmatrix}$$

according to the decomposition $\{\psi_{k_1}^+\psi_{k_2}^-\} = \{\psi_0^+\psi_0^-\} \oplus \{\psi_{k_1}^+\psi_{k_2}^-\}_{\bar k \ne 0}$ with $\bar k = (k_1, k_2)$. Here $\lambda_0(t)$ is the top eigenvalue of $K_*(t, Q, Q')$ (see (4.14)).

The top eigenvalue of $K_\pm$ has the form

$$\lambda_0(K_\pm) = \lambda_0(A_\pm) = \lambda_0^+\lambda_0^- + O(1/W).$$

Proof. To simplify formulas, we consider the kernel (see (6.4)–(6.5))

$$M(x, y) = a_+(2\pi)^{-1/2}e^{-(\mathcal{A}x,x)/2},\ \bar x = (x, y), \quad \mathcal{A} = \begin{pmatrix} \mu & \nu \\ \nu & \mu \end{pmatrix},\ \lambda_\pm = \mu \pm \nu,\ \Re\lambda_\pm > 0.$$

Then, taking $\kappa = \sqrt{\mu^2 - \nu^2} = \sqrt{\lambda_+\lambda_-}$, we obtain

$$\begin{aligned}a_+(2\pi)^{-1/2} \int e^{-(\mathcal{A}x,x)/2 + \kappa y^2/2}\big(\frac{d}{dy}\big)^k e^{-\kappa y^2}\, dy \\ = q^k \cdot a_+(\mu + \kappa)^{-1/2} e^{\kappa x^2/2}\big(\frac{d}{dx}\big)^k e^{-\kappa x^2}, \quad q = \frac{\nu}{\mu + \kappa},\end{aligned}$$

so $e^{\kappa y^2/2}(\frac{d}{dy})^k e^{-\kappa y^2}$, $k = 0, 1, \dots$ are the eigenvectors of M. Since M is compact, we have $|q| < 1$. Notice also that

$$a_+(\mu + \kappa_\pm)^{-1/2} = \lambda_0^\pm.$$

Now if we change the variables

$$x_1 = \theta x,\, y_1 = \theta y, \quad \theta = e^{-i(\arg\lambda_+ + \arg\lambda_-)/4} = e^{-i\arg\kappa/2},$$

then for the new matrix $\widetilde{A} = \theta^2 A$ has eigenvalues $\theta^2\lambda_+, \theta^2\lambda_-$, whose real parts are still positive, $\widetilde{\kappa} = |\kappa|$, and $\widetilde{q} = q$. This finishes the proof of (6.7)–(6.8).

Formula (6.10) follows directly from (6.4) and the fact that the Gaussian integral of x^{2k+1} is zero, and it immediately gives the statement about eigenvalues of $A_\pm$ (it is easy to see that $\lambda_0(t)$ does not change anything since it has only $\tilde{a}/W^{3/2}$ and $\tilde{a}'/W^{3/2}$).

Equation (6.12) can be obtained from the standard Schur inversion formula. The rest of part (ii) follows directly from (i) and Proposition 4.2. □

Now we are going to normalize $K_\pm$, $K_{\xi,\pm}$ by $\lambda_0(K_\pm)$:

$$\hat{K}_\pm = \lambda_0(K_\pm)^{-1}K_\pm, \quad \hat{K}_{\xi,\pm} = \lambda_0(K_\pm)^{-1}K_{\xi,\pm} \tag{6.13}$$

with $K_{\pm,\xi}$ of (6.3). Notice that

$$\hat{K}_\pm = \hat{A}_\pm \cdot \hat{K}_* \tag{6.14}$$

where

$$\hat{A}_\pm = (\lambda_0(A_\pm))^{-1}A_\pm, \quad \hat{K}_*(t,Q,Q') = (\lambda_0(t))^{-1}K_*(t,Q,Q'),$$

so both top eigenvalues of $\hat{A}_\pm$, $\hat{K}_*$ are 1, and

$$\hat{\lambda}_j(\hat{K}_*) = 1 - \frac{j(j+3)}{tW} + O((j^2/tW)^2), \quad j = 1,2,\dots. \tag{6.15}$$

Therefore, it is easy to see that all eigenvalues of $\hat{K}_\pm$, $\hat{K}_{\xi,\pm}$ lie inside $\omega_A = \{z : |z| = 1 + A/n\}$. Thus, we get

$$\begin{aligned} I_{\pm,\xi} &= -2\pi i(K_{\xi,\pm}^{n-1}f, g) = -2\pi i \cdot \lambda_0(K_\pm)^{n-1}(\hat{K}_{\xi,\pm}^{n-1}f, g) \\ &= \lambda_0(K_\pm)^{n-1}\int_{\omega_A} z^{n-1}(\hat{G}_\xi(z)f, g)dz, \end{aligned}$$

where

$$\hat{G}_\xi(z) = (\hat{K}_{\xi,\pm} - z)^{-1}.$$

Similarly we can rewrite $I_\pm$.

Consider the matrix of $\hat{K}_{\xi,\pm}$ in the basis

$$\Psi_{\bar{k},j}(\tilde{a}_1,\tilde{a}_2,Q) = \psi^+_{k_1}(\tilde{a}_1)\psi^-_{k_2}(\tilde{a}_2)\phi_j(Q), \quad k_1,k_2,j \geq 0, \tag{6.16}$$

with $\psi^\pm_k$ of (6.6), and ϕ_j of (4.11). Let $\mathcal{H}_1 = \{\Psi_{\bar{k},j}\}_{j\leq l-1}$ and

$$L_2(\mathbb{R}^2) \times L_2(\mathring{S}p(2), d\mu(Q)) = \mathcal{H}_1 \oplus \mathcal{H}_2, \tag{6.17}$$

and write

$$\hat{K}_\pm = \begin{pmatrix} \hat{K}^{(11)} & \hat{K}^{(12)} \\ \hat{K}^{(21)} & \hat{K}^{(22)} \end{pmatrix}, \quad \hat{K}_{\xi,\pm} = \begin{pmatrix} \hat{K}_\xi^{(11)} & \hat{K}_\xi^{(12)} \\ \hat{K}_\xi^{(21)} & \hat{K}_\xi^{(22)} \end{pmatrix} \tag{6.18}$$

according to this decomposition. We will need the following simple lemma

Lemma 6.2. *Given decomposition (6.18), we have*

$$\hat{K}^{(12)} = \hat{K}^{(21)} = 0, \quad \|\hat{K}_\xi^{(12)}\| \leq \frac{C}{n}, \quad \|\hat{K}_\xi^{(21)}\| \leq \frac{C}{n}; \tag{6.19}$$

and for $|z| \geq 1 + A/n$ with big enough A we have

$$\|(\hat{K}_\xi^{(11)} - z)^{-1}\| \leq Cn, \tag{6.20}$$

$$\|(\hat{K}_\xi^{(22)} - z)^{-1}\| \leq CW/l^2, \tag{6.21}$$

and same is valid for $\hat{K}$.

Proof of Lemma 6.2. The bound (6.19) follows follows from the block-diagonal structure of $\hat{K}_\pm$ with respect to the basis (6.16) (see (4.15)), and the fact that $\hat{K}_{\pm,\xi}$ is $1/n$-order perturbation of $\hat{K}_\pm$.

The bound (6.20) follows from

$$\|\hat{K}_\xi^{(11)}\| \le 1 + c/n,$$

since for big enough A

$$\|(\hat{K}_\xi^{(11)} - z)^{-1}\| \le |z|^{-1} \sum_{k=0}^{\infty} \left(\frac{\|K_\xi^{(11)}\|}{|z|} \right)^k \le Cn.$$

Similarly, according to (4.14)–(4.15), we get

$$\|\hat{K}^{(22)}\| \le 1 - \frac{Cl(l+3)}{W}$$

and, since $l^2/W \sim \log^2 n/n$ for $W \ge Cn$ and $l^2/W \gg n^{-1}$ for $W \ll n$,

$$\|\hat{K}_\xi^{(22)}\| \le 1 - \frac{Cl^2}{W},$$

which implies (6.21). □

The next step is to prove that we can consider only the upper-left block $K_\xi^{(11)}$ of K_ξ (see (6.18)). More precisely, we are going to prove

Lemma 6.3. *We have*

$$\int\limits_{\omega_A} z^{n-1} (\hat{G}_\xi(z) f, \bar{f}) dz = \int\limits_{\omega_A} z^{n-1} (\hat{G}_{1,\xi}(z)\, f_1, \bar{f}_1) dz + O\Big(\frac{W \log n}{l^2 n}\Big),$$

where

$$\hat{G}_{1,\xi}(z) = (\hat{K}_\xi^{(11)} - z)^{-1},$$

and we decomposed $f = (f_1, f_2)$ with respect to the decomposition (6.17). Notice that

$$\frac{W \log n}{l^2 n} \le \frac{1}{\log n}.$$

Proof. Using the well-known Schur inversion formula we get

$$(\hat{K}_\xi - z)^{-1} = \begin{pmatrix} \hat{G}_\xi^{(11)} & -\hat{G}_\xi^{(11)} \hat{K}_\xi^{(12)} \hat{G}_{2,\xi} \\ -\hat{G}_{2,\xi} \hat{K}_\xi^{(21)} \hat{G}_\xi^{(11)} & \hat{G}_{2,\xi} + \hat{G}_{2,\xi} \hat{K}_\xi^{(21)} \hat{G}_\xi^{(11)} \hat{K}_\xi^{(12)} \hat{G}_{2,\xi} \end{pmatrix},$$

where

$$\hat{G}_{2,\xi}(z) = (\hat{K}_\xi^{(22)} - z)^{-1},$$
$$\hat{G}_\xi^{(11)} = (\hat{K}_\xi^{(11)} - z - \hat{K}_\xi^{(12)} \hat{G}_{2,\xi} \hat{K}_\xi^{(21)})^{-1} = (1 - \hat{G}_{1,\xi} \hat{K}_\xi^{(12)} \hat{G}_{2,\xi} \hat{K}_\xi^{(21)})^{-1} \hat{G}_{1,\xi}.$$

Thus

$$\begin{aligned} \int_{\omega_A} z^{n-1} ((K_\xi - z)^{-1} f, \bar{f}) dz = \int_{\omega_A} z^{n-1} \Big((G_\xi^{(11)} f_1, \bar{f}_1) - (G_\xi^{(11)} K_\xi^{(12)} G_{2,\xi} f_2, \bar{f}_1) \Big) dz \\ - \int_{\omega_A} z^{n-1} (G_{2,\xi} K_\xi^{(21)} G_\xi^{(11)} f_1, \bar{f}_2) dz \\ + \int_{\omega_A} z^{n-1} ((G_{2,\xi} + G_{2,\xi} K_\xi^{(21)} G_\xi^{(11)} K_\xi^{(12)} G_{2,\xi}) f_2, \bar{f}_2) dz. \end{aligned} \tag{6.22}$$

Denoting

$$R = (1 - \hat{G}_{1,\xi}\hat{K}_\xi^{(12)}\hat{G}_{2,\xi}\hat{K}_\xi^{(21)})^{-1},$$

we get

$$\hat{G}_\xi^{(11)} = R\hat{G}_{1,\xi}.$$

According to (6.19)–(6.21), we obtain

$$\|\hat{G}_{1,\xi}\hat{K}_\xi^{(12)}\hat{G}_{2,\xi}\hat{K}_\xi^{(21)}\| \le Cn \cdot \frac{1}{n^2} \cdot \frac{W}{l^2} = \frac{CW}{l^2 n}.$$

Therefore,

$$\|1 - R\| \le \frac{CW}{l^2 n},$$

which together with (6.20) implies

$$\begin{aligned}\Big|\int_{\omega_A} z^{n-1}\Big((\hat{G}_\xi^{(11)} - \hat{G}_{1,\xi})f_1, \bar{f}_1\Big)\Big| &= \Big|\int_{\omega_A} z^{n-1}\Big((1-R)G_{1,\xi})f_1, \bar{f}_1\Big)\Big| \\ &\le C\|1-R\| \cdot \|f_1\|^2 \cdot \int_{\omega_A} \frac{|dz|}{|z-1|} \le \frac{CW\log n}{l^2 n}.\end{aligned}$$

It is easy to see also that

$$\|f_2\| \le C/n,$$

and because of the consideration above

$$\begin{aligned}&\|\hat{G}_{2,\xi} + \hat{G}_{2,\xi}\hat{K}_\xi^{(21)}\hat{G}_\xi^{(11)}\hat{K}_\xi^{(12)}\hat{G}_{2,\xi}\| \le CW/l^2, \\ &\|\hat{G}_{2,\xi}\hat{K}_\xi^{(21)}\hat{G}_\xi^{(11)}\| \le CW/l^2, \quad \|\hat{G}_\xi^{(11)}\hat{K}_\xi^{(12)}\hat{G}_{2,\xi}\| \le CW/l^2,\end{aligned}$$

so other terms in (6.22) are also small. □

The next step is to show that we can consider only the projection of $\hat{K}_\xi^{(11)}$, $\hat{K}^{(11)}$ on the linear span of $\{\Psi_{0,j}\}_{j\le l}$ (see (6.16)). We prove

Lemma 6.4. *Let P_l be the projection on $\{\phi_j\}_{j=0}^{l-1}$ of (4.11), $\Delta_l = P_l\Delta P_l$, and $\hat{\nu}_l = P_l\hat{\nu}P_l$ with $\hat{\nu}$ defined in (1.9). Then*

$$\begin{aligned}\int_{\omega_A} z^{n-1}(\hat{G}_{1,\xi}(z)\, f_1, \bar{f}_1)dz &= O\Big(\frac{(l-1)^2 n}{W^{3/2}}\Big) + O\Big(\frac{1}{W^{1/2}}\Big) \\ &+ \int_{\omega_A} \zeta^{n-1}\Big(\Big(P_l - \frac{1}{t_* W}\Delta_l - \frac{i\pi\xi}{n}\hat{\nu}_l - \zeta + O\Big(\frac{(l-1)^4}{W^2}\Big)\Big)^{-1} f_0, \bar{f}_0\Big)d\zeta,\end{aligned}$$

where $O(x)$ is an operator whose norm is bounded by Cx which does not depend on ζ, and

$$f_0 = (f, \Psi_{\bar{0}}).$$

Recall $l = 1$ for $W \ll n$ and $(l-1)^2 n/W^{3/2} \sim \log^2 n/W^{1/2}$ for $W \ge Cn$, and $t_* = (2\pi\rho(E))^2$. A similar formula is true for $\hat{G}_1$ (i.e. for $\hat{K}^{(11)}$ instead of $\hat{K}_\xi^{(11)}$).

Proof. Write $\hat{K}_\xi^{(11)} - z$, $\hat{K}^{(11)} - z$ in the block form

$$\hat{K}^{(11)} - z = \begin{pmatrix} M_1 & M_{12} \\ M_{21} & M_2 \end{pmatrix}, \quad \hat{K}_\xi^{(11)} - z = \begin{pmatrix} M_{1,\xi} & M_{12,\xi} \\ M_{21,\xi} & M_{2,\xi} \end{pmatrix}$$

according to decomposition

$$\mathcal{H}_1 = \mathcal{M}_1 \oplus \mathcal{M}_2,$$

where $\mathcal{M}_1$ is a linear span of $\Psi_0^{(l)} = \{\Psi_{0,j}\}_{j\le l}$ (see (6.16)).

Set

$$G^0_{1,\xi}(z) = (M_{11,\xi} - M_{12,\xi}M^{-1}_{22,\xi}M_{21,\xi})^{-1} \quad G^0_1(z) = (M_{11} - M_{12}M^{-1}_{22}M_{21})^{-1}$$

Then, using Proposition 5.2, we get

$$\oint_{\omega_A} z^{n-1}(G(z)f_1, \bar f_1)dz = \oint_{\omega_A} z^{n-1}(G^0_1 f^{(1)}(z), g^{(1)}(z))dz + O(e^{-nc}),$$
$$\oint_{\omega_A} z^{n-1}(G_\xi(z)f_1, \bar f_1)dz = \oint_{\omega_A} z^{n-1}(G^0_{1,\xi} f^{(1)}(z), g^{(1)}(z))dz + O(e^{-nc}),$$

where $f^{(1)}$, $g^{(1)}$ are defined as in (5.5).

Recall that

$$\hat K^{(11)} - z = (\hat A P_l \hat K_* P_l \Psi_0^{(l)}, \Psi_0^{(l)}) - z.$$

Set

$$P_l \hat K_* P_l = P_l - \tilde K_l.$$

Then $\tilde K_l$, according to (6.15), is a diagonal matrix with eigenvalues

$$\tilde\lambda_j(t) = j(j+3)/tW + O((j^2/tW)^2), \quad j = 0, \dots, l-1.$$

Since (4.3) and (6.1) imply

$$t = \Big(a_+ - a_- + \frac{i\theta_+ a_+ \tilde a_1 - i\theta_- a_- \tilde a_2)}{\sqrt{W}}\Big)\Big(a_+ - a_- + \frac{i\theta_+ a_+ \tilde a_1' - i\theta_- a_- \tilde a_2')}{\sqrt{W}}\Big),$$

$\tilde K_l$ can be rewritten as

$$\tilde K_l = \Delta_l/t_* W + O_a\Big(\frac{(l-1)^2}{W^{3/2}}\Big) + O\Big(\frac{(l-1)^4}{W^2}\Big) \tag{6.23}$$

with $t_* = (a_+ - a_-)^2 = (2\pi\rho(E))^2$. Here $O_a(X)$ is a diagonal in $\{\phi_j\}_{j=0}^{l-1}$ operator of the type $O(X)$ whose eigenvalues are linear in a, a'.

Now, since, according to Lemma 6.1, $\hat A_{00} = 1 + O(1/W)$, substituting (6.23), we get

$$(\hat A \tilde K_l \Psi_0^{(l)}, \Psi_0^{(l)})_{jj} = (\tilde\lambda_j(t)\hat A \psi_0^+\psi_0^-, \psi_0^+\psi_0^-) = \frac{j(j+3)}{t_* W}\cdot\Big(1 + O\Big(\frac{j^2}{W}\Big)\Big)$$

Therefore,

$$\hat K^{(11)} - z = \hat A_{00} P_l - z - \Delta_l/t_* W + O((l-1)^4/W^2).$$

Similarly

$$M_{12} = \hat A_{12}\otimes P_l + O\Big(\frac{(l-1)^2}{W^{3/2}}\Big),$$
$$M_{21} = \hat A_{21}\otimes P_l + O\Big(\frac{(l-1)^2}{W^{3/2}}\Big),$$
$$M_{22} = \hat A_{22}\otimes P_l - z + O\Big(\frac{(l-1)^2}{W}\Big).$$

Notice also that because of Lemma 6.1

$$\|\hat A_{12}\| \le W^{-1/2}, \quad \|\hat A_{21}\| \le W^{-1/2}, \quad \|(\hat A_{22} - z)^{-1}\| \le C.$$

Hence

$$M_{11} - M_{12}M_{22}^{-1}M_{21} = A_{00}P_l - \Delta_l/t_*W - z - \hat{A}_{12}(\hat{A}_{22} - z)^{-1}\hat{A}_{21}P_l + O_z\Big(\frac{(l-1)^2}{W^2}\Big)$$
$$+ O\Big(\frac{(l-1)^4}{W^2}\Big) = P_l - \Delta_l/t_*W - z - W^{-1}g_1(z)P_l + O\Big(\frac{(l-1)^4}{W^2}\Big),$$

where

$$g_1(z) = \hat{A}_{00} - 1 - \hat{A}_{12}(\hat{A}_{22} - z)^{-1}\hat{A}_{21} + O_z((l-1)^2/W) - O_*((l-1)^2/W)$$

is analytic and bounded in $\{z : |z-1| < \delta\}$ for small enough δ (recall $l^2/W \le \log^2 n/n$ and $\|\hat{A}^{22}\| \le |q_\pm| < 1$ with $q_\pm$ of (6.8) according to Lemma 6.1). Here $O_z(\cdot)$ is an operator of type $O(\cdot)$ which may depend on z, and $O_*(\cdot)$ is an operator $O_z(\cdot)$ with substitution $z = 1$. Lemma 6.1 implies also

$$g_1(1) = 0. \tag{6.24}$$

Now set

$$\zeta(z) = z + W^{-1}g_1(z). \tag{6.25}$$

Since $g_1(z)$ has a bounded derivative in $\{z : |z-1| < \delta\}$, we get

$$\zeta'(z) = 1 + O(1/W),$$

and the Implicit Function Theorem implies that there exists the inverse function $z(\zeta)$ with a derivative of order $1 + O(1/W)$. In addition, by (6.24), $\zeta(1) = 1$, so the image of $\{z : |z-1| < \delta\}$ lies in $\{\zeta : \delta/2 < |\zeta - 1| < 2\delta\}$, and it is easy to show that

$$z(\zeta) = \zeta + W^{-1}\tilde{g}_1(\zeta),$$

where $\tilde{g}_1$ is a bounded analytic in $\{\zeta : |\zeta - 1| < 2\delta\}$, and $\tilde{g}_1(1) = 0$.

Now we consider the contour $\tilde{\omega}_A = \{z : \mathrm{dist}\{z; [1 - C(l-1)(l+2)/W; 1]\} \le A/n\}$ and the contour $\mathcal{L}_2 = \{|z| \le \dfrac{|q_\pm| + 1}{2} < 1\}$ with $q_\pm$ of (6.8). It is easy to see that $\tilde{\omega}_A \cup \mathcal{L}_2$ encircles all the eigenvalues of $\hat{K}^{(11)}$, $\hat{K}_\xi^{(11)}$ (see (4.14) – (4.15) and Lemma 6.1). But for $z \in \mathcal{L}_2$

$$|z|^{n-1} \le e^{-cn},$$

so the contribution of the integral over $\mathcal{L}_2$ is small, and we need to consider integral over $\tilde{\omega}_A$ only. It follows from (6.24)–(6.25) and the consideration above that $\zeta(z), z \in \tilde{\omega}_A$ will be inside $\tilde{\omega}_{2A}$, and so, since $l = 1$ for $W \ll n$ and $l^2/W \sim \log^2 n/n$ for $W \ge Cn$, we get

$$z(\zeta) = \zeta + O((l-1)^2/W^2) + O(1/nW),$$

hence

$$z^{n-1} = \begin{cases} \zeta^{n-1} + O(\log^2 n/W), & W \ge Cn, \\ \zeta^{n-1} + O(1/W), & W \ll n. \end{cases} \tag{6.26}$$

Notice also that for $z \in \tilde{\omega}_A$

$$\|(P_l - \Delta_l/t_*W - \zeta(z))^{-1}\| \le Cn,$$

thus

$$\|G_1^0\| = \|(P_l - \Delta_l/t_*W - \zeta(z) + O((l-1)^4/W^2))^{-1}\| \le \|(P_l - \Delta_l/t_*W - \zeta(z))^{-1}\|$$
$$\times \Big\|\Big(1 + O((l-1)^4/W^2)\cdot(P_l - \Delta_l/t_*W - \zeta(z))^{-1}\Big)^{-1}\Big\| \le Cn.$$

Hence, recalling $l=1$ and $\|\tilde\omega_A\|=C/n$ for $W\ll n$ and $|\tilde\omega_A|\le C(l-1)^2/W$ for $W\ge Cn$, and $\|f^{(1)}(z)-f_0\|\le C/\sqrt{W}$, we obtain

$$\oint_{\tilde\omega_A} z^{n-1}(G_1^0(z)f^{(1)}(z),g^{(1)}(z))dz \\ = \oint_{\tilde\omega_A} z^{n-1}(G_1^0(z)f_0,\bar f_0)dz + O((l-1)^2n/W^{3/2})+O(1/W^{1/2}).$$

According to (6.26), this can be further transformed as

$$\oint_{\tilde\omega_A} z^{n-1}(G_1^0(z)f_0,\bar f_0)dz = O\Big(\frac{(l-1)^4n^2}{W^3}\Big)+O\Big(\frac{1}{W}\Big) \\ + \oint_{\zeta(\tilde\omega_A)} \zeta^{n-1}((P_l-\Delta_l/t_*W-\zeta+O((l-1)^4/W^2))^{-1}f_0,\bar f_0)d\zeta,$$

and the contour now can be changed back to ω_A (notice $l^4n^2/W^3=\log^4 n/W$, $l^4/W^2=\log^2 n/Wn$ for $W\ge Cn$).

In order to perform the same analysis for $\hat K_\xi^{(11)}$ notice that

$$\|M_{12}\|\le C/\sqrt{W},\quad M_{12,\xi}=M_{12}+O\Big(\frac{1}{n\sqrt{W}}\Big);$$
$$\|M_{21}\|\le C/\sqrt{W},\quad M_{21,\xi}=M_{21}+O\Big(\frac{1}{n\sqrt{W}}\Big);$$
$$\|M_{22}^{-1}\|\le C,\qquad M_{22,\xi}=M_{22}+O\Big(\frac{1}{n}\Big),$$

and

$$M_{11,\xi}=(AP_lF_{n,\xi}K_*F_{n,\xi}P_l\Psi_0,\Psi_0)-z=M_{11}-A_{00}\cdot\frac{i\pi\xi}{n}P_l\hat\nu P_l+O\Big(\frac{1}{n\sqrt{W}}\Big).$$

Thus, since $A_{00}=1+O(1/W)$, we have

$$M_{1,\xi}-M_{12,\xi}M_{22,\xi}^{-1}M_{21,\xi}=M_{11}-M_{12}M_{22}^{-1}M_{21}-\frac{i\pi\xi}{n}P_l\hat\nu P_l+O\Big(\frac{1}{n\sqrt{W}}\Big),$$

and hence we can apply same consideration as above. □

Now let us analyze

$$\int_{\omega_A}\zeta^{n-1}\Big(\Big(P_l-\frac{1}{t_*W}\Delta_l-\frac{i\pi\xi}{n}\hat\nu_l-\zeta+O((l-1)^4/W^2)\Big)^{-1}f_0,\bar f_0\Big)d\zeta$$

- **localized regime: $\mathbf{W\ll n}$.** In this regime $l=1$, so we need to study

$$\int_{\omega_A}\zeta^{n-1}\Big(\Big(P_1-\frac{i\pi\xi}{n}\hat\nu_1-\zeta\Big)^{-1}f_0,\bar f_0\Big)d\zeta \\ =-2\pi i\cdot\|f_0\|^2\Big(\Big(P_1-\frac{i\pi\xi}{n}\hat\nu_1\Big)^{n-1}1,1\Big).$$

But since $\phi_0=1$ and $\hat\nu_1=P_1\hat\nu P_1$, $\hat\nu\cdot 1=\phi_1$ (see (1.9) and Proposition 4.2), we obtain

$$(P_1-\frac{i\pi\xi}{n}\hat\nu_1\Big)1=1-\frac{i\pi\xi}{n}P_1\phi_1=1,$$

which implies

$$\Big(\Big(P_1-\frac{i\pi\xi}{n}\hat\nu_1\Big)^{n-1}1,1\Big)=1,$$

thus Theorem 1.1 in the regime $W\ll n$.

- **critical regime: $\mathbf{n = C_* W}$.**

  Again we need to study $\left(K_0^{n-1} f_0, f_0\right)$ with

$$K_0 = P_l - \frac{1}{t_* W}\Delta_l - \frac{i\pi\xi}{n}\hat{\nu}_l + O(l^4/W^2) = P_l - \frac{C^*}{n}\Delta_l - \frac{i\pi\xi}{n}\hat{\nu}_l + O(\log^4 n/n^2), \tag{6.27}$$

  where $C^* = C_*/t_*$. It is enough to prove

  Lemma 6.5. *Given (4.18), if $n = C_* W$, $l = [\log W]$ we have*

$$(K_0^{n-1}1, 1) \to (e^{-C^*\Delta - i\xi\pi\hat{\nu}}1, 1), \quad n, W \to \infty,$$

  with $\hat{\nu}$, Δ as in Theorem 1.1.

  Similar Lemma is proved in [32], but for the sake of completeness we repeat the proof here.

  Proof of Lemma 6.5. Notice that

$$K_0 = P_l - n^{-1}C^*\Delta_l - \frac{i\xi\pi}{n}\hat{\nu}_l + O(\log^4 n/n^2) = P_l e^{-n^{-1}(C^*\Delta_l + i\xi\pi\hat{\nu}_l) + O(\log^4 n/n^2)} P_l.$$

  Thus

$$K_0^{n-1} = P_l e^{-C^*\Delta_l - i\xi\pi\hat{\nu}_l} P_l + O(\log^4 n/n),$$

  ans so

$$(K_0^{n-1}1, 1) = (e^{-C^*\Delta_l - i\xi\pi\hat{\nu}_l}1, 1) + O(\log^4 n/n).$$

  Consider the basis $\{\phi_j\}$ of (4.11). In this basis the Laplace-Bertarami operator Δ is diagonal, and the operator $\hat{\nu}$ is three diagonal (since it corresponds to the multiplication by $2x^2 - 1$, see (1.9) and (4.12)). To simplify notations, let F be an operator of multiplication by $(i\pi\xi\nu)$ and $\tilde{\Delta} = C^*\Delta$. Set

$$D = \tilde{\Delta} + F,$$
$$D^{(l)} = \tilde{\Delta} + F^{(l)},$$

  where $F^{(l)}$ be the matrix F where we put $F_{l-1,l} = F_{l,l-1} = 0$. It is evident that (recall $\phi_0 = 1$)

$$(e^{-D^{(l)}}\phi_0, \phi_0) = \left(e^{-P_l D P_l}\phi_0, \phi_0\right) = (e^{-P_l(C^*\Delta + i\xi\pi\nu)P_l}1, 1).$$

  Thus, we are left to prove that

$$\left(\left(e^{-D} - e^{-D^{(l)}}\right)\phi_0, \phi_0\right) \to 0. \tag{6.28}$$

  Notice that both e^{-D}, $e^{-D^{(l)}}$ are bounded operators, and $|F| \le C$, $|F^{(l)}| \le C$. We will use the well-known Duhamel formula

$$e^{-tA_1} - e^{-tA_2} = \int_0^t e^{-(t-s)A_2}(A_1 - A_2)e^{-sA_1}\,ds. \tag{6.29}$$

For $A_1 = D$, $A_2 = D^{(l)}$ and $t = 1$ it gives

$$\begin{aligned}
&\Big|\big(e^{-D} - e^{-D^{(l)}}\big)\phi_0\Big| = \Big|\int_0^1 e^{-(1-s)D^{(l)}}(F - F^{(l)})e^{-sD}\phi_0\, ds\Big| \\
&= \Big|\int_0^1 e^{-(1-s)D^{(l)}}(F_{l-1}\cdot E_{l-1,l} + F_l \cdot E_{l,l-1})e^{-sD}\phi_0\, ds\Big| \\
&= \Big|\int_0^1 e^{-(1-s)D^{(l)}}\Big(F_l\phi_l\big(e^{-sD}\phi_0, \phi_{l-1}\big) + F_{l-1}\phi_{l-1}\big(e^{-sD}\phi_0, \phi_l\big)\Big)\, ds\Big| \\
&\le C\Big(\big|\big(e^{-sD}\phi_0, \phi_{l-1}\big)\big| + \big|\big(e^{-sD}\phi_0, \phi_l\big)\big|\Big).
\end{aligned}$$

Here $E_{l-1,l}$ is an operator whose matrix in the basis $\{\phi_j\}$ has 1 at $(l-1, l)$ place and zeros everywhere else, and $E_{l,l-1}$ is defined in a similar way. F_{l-1}, F_l are $(l-1, l)$ and $(l, l-1)$ elements of the matrix F in the same basis.

Now let us bound $\big|\big(e^{-sD}\phi_0, \phi_l\big)\big|$. To this end, apply Duhamel's formula (6.29) $p = [l/2]$ times with $A_1 = D$ and $A_2 = \tilde{\Delta}$. We obtain

$$\begin{aligned}
&\big(e^{-sD}\phi_0, \phi_l\big) = \sum_{j=1}^{p}\int_{s_1+..+s_j\le s}\big(e^{-s_1\Delta}Fe^{-s_2\Delta}F\ldots e^{-s_j\Delta}\phi_0, \phi_l\big)\, ds_1..ds_j \\
&+ \int_{s_1+..+s_p\le s}\big(e^{-s_1 D}Fe^{-s_2\Delta}F\ldots e^{-s_p\Delta}\phi_0, \phi_l\big)\, ds_1..ds_p.
\end{aligned}$$

Since $e^{-s\Delta}$ is diagonal in the basis $\{\phi_j\}$, and F is only three diagonal, the expression $e^{-s_1\Delta}Fe^{-s_2\Delta}F\ldots e^{-s_j\Delta}\phi_0$ is in the linear span of $\{\phi_k\}_{k=0}^{j}$, and thus the sum above is 0. Hence

$$\begin{aligned}
&\Big|\big(e^{-sD}\phi_0, \phi_l\big)\Big| \le \Big|\int_{s_1+..+s_p\le s}\big(e^{-s_1 D}Fe^{-s_2\Delta}F\ldots e^{-s_p\Delta}\phi_0, \phi_l\big)\, ds_1..ds_p\Big| \\
&\le C^l\Big|\int_{s_1+..+s_p\le s} ds_1..ds_p\Big| = \frac{C^l s^l}{l!} \le C_1 e^{-l\log l} \to 0,
\end{aligned}$$

which finishes the proof of (6.28). □

- **delocalized regime: $\mathbf{W \gg n}$.** Since in this regime $l^4/W^2 = C\log^4 n/n^2$, we get

$$\begin{aligned}
&\int_{\omega_A}\zeta^{n-1}\Big(\Big(P_l - \frac{C_*}{W}\Delta_l - \frac{i\pi\xi}{n}\hat{\nu}_l - \zeta + O(l^4/W^2)\Big)^{-1} f_0, \bar{f}_0\Big)d\zeta \\
&= \int_{\omega_A}\zeta^{n-1}\Big(\Big(P_l - \frac{C_*}{W}\Delta_l - \frac{i\pi\xi}{n}\hat{\nu}_l - \zeta\Big)^{-1} f_0, \bar{f}_0\Big)d\zeta + O(\log^4 n/n),
\end{aligned}$$

Hence we need to study

$$|f_0|^2\int_{\omega_A} z^{n-1}\Big(\Big(P_l - \frac{C_*}{W}\Delta_l - \frac{i\pi\xi}{n}\hat{\nu}_l - z\Big)^{-1}1, 1\Big)dz$$

Now let us define

$$m = \sqrt[3]{\frac{W}{n}}, \quad m \to \infty,$$

in order to get

$$\frac{m^2 n}{W} = \frac{1}{m} \to 0.$$

Set

$$G(z) = (P_l - \frac{C_*}{W}\Delta_l - \frac{i\pi\xi}{n}\hat{\nu}_l - z)^{-1}, \quad G^{(m)}(z) = (P_m - \frac{C_*}{W}\Delta_m - \frac{i\pi\xi}{n}\hat{\nu}_m - z)^{-1},$$
$$G^{(m,l)}(z) = (P_l - \frac{C_*}{W}\Delta_l - \frac{i\pi\xi}{n}\hat{\nu}_{l,m} - z)^{-1},$$

where $\hat{\nu}_{l,m}$ has the same matrix as a tridiagonal operator $\hat{\nu}_l$ but with $(m, m+1)$ and $(m+1, m)$ elements equal to 0, and P_m is a projection on $\{\phi_j\}_{j\le m}$, $\Delta_m = P_m\Delta P_m$, $\hat{\nu}_m = P_m\hat{\nu}P_m$. Notice that $\hat{\nu}_{l,m}$ has a block diagonal structure with blocks $m \times m$ and $(l-m)\times(l-m)$, thus

$$(G^{(m,l)}(z)1, 1) = (G^{(m)}(z)1, 1). \tag{6.30}$$

We are going to prove

$$\int\limits_{\omega_A} z^{n-1}\Big(G(z)\cdot 1, 1\Big)dz = \int\limits_{\omega_A} z^{n-1}\Big(G^{(m,l)}(z)\cdot 1, 1\Big)dz + O(1/m). \tag{6.31}$$

Then, if we define

$$G_0^{(m)}(z) = (P_m - \frac{i\pi\xi}{n}\hat{\nu}_m - z)^{-1}, \tag{6.32}$$

we can write using (6.30) and the standard resolvent identity

$$\Big(G^{(m,l)}(z)\cdot 1, 1\Big) = (G^{(m)}(z)1, 1) = (G_0^{(m)}(z)1, 1) + (G^{(m)}(z)(\frac{C_*}{W}\Delta_m)G_0^{(m)}(z)1, 1).$$

But

$$\|\frac{C_*}{W}\Delta_m\| \le \frac{Cm^2}{W} \le \frac{1}{mn},$$

hence, since both resolvent can be bounded by $|z-1|^{-1}$, we get

$$\Big|\int\limits_{\omega_A} z^{n-1}(G^{(m)}(z)(\frac{C_*}{W}\Delta_m)G_0^{(m)}(z)1, 1)dz\Big| \le \frac{C}{mn}\int\limits_{\omega_A}\frac{|dz|}{|z-1|^2} \le \frac{C}{m}, \tag{6.33}$$

where we have used

$$\int\limits_{\omega_A}\frac{|dz|}{|z-1|^2} \le Cn. \tag{6.34}$$

Now (6.30)–(6.33) imply

$$|f_0|^2\int\limits_{\omega_A} z^{n-1}(G(z)1, 1)dz = |f_0|^2\int\limits_{\omega_A} z^{n-1}(G_0^{(m)}(z)1, 1)dz + O(1/m)$$
$$= -2\pi i\cdot|f_0|^2\cdot\Big(\Big(P_m - \frac{i\pi\xi}{n}\hat{\nu}_m\Big)^{n-1}1, 1\Big) + O(1/m).$$

Since $\hat{\nu}$ is bounded, we can easily change $\hat{\nu}_m$ to $\hat{\nu}$ and use

$$\Big(1 - \frac{i\pi\xi}{n}\hat{\nu}\Big)^{n-1} = e^{-i\pi\xi\hat{\nu}} + O(1/n),$$

which implies Theorem 1.1.

Therefore we are left to prove (6.31). First we will need a bound

Lemma 6.6. *For $|z| \ge 1 + A/n$ we have*

$$|G_{ij}(z)| \le \frac{C}{|z-1|}e^{-\delta|i-j|},$$

where C and δ depends only on A. Same is true for $G^{(m,l)}(z)$.

Notice that since $\hat{\nu}$ is bounded 3-diagonal matrix (see (4.12)), and $P_l - \frac{C_*}{W}\Delta_l$ is diagonal, Lemma 6.6 follows from the standard Combes-Thomas arguments (see, e.g., [25], Ch 13, Proposition 13.13.1)).

Using the resolvent identity we can write

$$\int\limits_{\omega_A} z^{n-1}\Big(G(z)\cdot 1,1\Big)dz = \int\limits_{\omega_A} z^{n-1}\Big(G^{(m,l)}(z)\cdot 1,1\Big)dz + \int\limits_{\omega_A} z^{n-1}\Big(G(z)\Big(\frac{\delta\hat{\nu}}{n}\Big)G^{(m,l)}(z)\cdot 1,1\Big)dz, \tag{6.35}$$

where $\delta\hat{\nu} = i\pi\xi(\hat{\nu}_l - \hat{\nu}_{l,m})$, i.e. the matrix with only two non-zero elements $(m,m+1)$ and $(m+1,m)$. Rewrite

$$\Big(G(z)\Big(\frac{\delta\hat{\nu}}{n}\Big)G^{(m,l)}(z)1,1\Big) = \frac{(\delta\hat{\nu})_{m,m+1}}{n}G^{(m,l)}_{m+1,0}(z)G^{*}_{0m}(z) + \frac{(\delta\hat{\nu})_{m+1,m}}{n}G^{(m,l)}_{m,0}(z)G^{*}_{0,m+1}(z). \tag{6.36}$$

But according to Lemma 6.6

$$|G^{(m,l)}_{m+1,0}(z)| \le \frac{C}{|z-1|}e^{-\delta m},$$

and similar bounds hold for other resolvent elements in (6.36). Thus

$$\Big|\int\limits_{\omega_A} z^{n-1}\Big(G(z)\Big(\frac{\delta\hat{\nu}}{n}\Big)G^{(m,l)}(z)1,1\Big)dz\Big| \le \frac{Ce^{-2\delta m}}{n}\int\limits_{\omega_A}\frac{|dz|}{|z-1|^2} \le Ce^{-2\delta m},$$

where we have used (6.34). This and (6.35) yield (6.31).

7 Analysis of I_+ and I_-

Since the integrals I_+ and I_- are similar, we can consider I_+ only. In this case we will consider $\{F_i\}$ of Proposition 2.1 as a $\mathring{Sp}(2)$ matrix which is in $W^{-1/2}$-neigbourhood of a_+I_4. Then F_i can be parametrized as $F_i = a_+(I + i\theta_+X_i/\sqrt{W})$, where X_i is a quaternion Hermitian matrix

$$X_j = \begin{pmatrix} \tilde{a}_{1j} & \tilde{w}_{j1} & 0 & \tilde{w}_{j2} \\ \overline{\tilde{w}}_{j1} & \tilde{a}_{j2} & -\tilde{w}_{j2} & 0 \\ 0 & -\overline{\tilde{w}}_{j2} & \tilde{a}_{j1} & \overline{\tilde{w}}_{j1} \\ \overline{\tilde{w}}_{j2} & 0 & \tilde{w}_{j1} & \tilde{a}_{j2} \end{pmatrix},$$

where $\tilde{w}_{j1} = (x_j + iy_j)/\sqrt{2}$, $\tilde{w}_{j2} = (p_j + iq_j)/\sqrt{2}$. This change transforms the measure dF_i to

$$\frac{(ia_+\theta_+)^6}{4}W^{-3}d\tilde{a}_{1i}d\tilde{a}_{2i}d\tilde{x}_i\,d\tilde{y}_i\,d\tilde{p}_i d\tilde{q}_i.$$

We need to keep the same $\tilde{C}'_{n,W}$ as in (4.4), so in the parametrization above the operator K^+_+ has the form

$$K^+_+(X,X') = \beta^2 A^+_a(\bar{a},x,y,p,q;\bar{a}',x',y',p',q')(1+o(1)),$$

where

$$A^+_a(\bar{a},x,y,p,q;\bar{a}',x',y',p',q') = A^+(\widetilde{a}_1,\widetilde{a}'_1)A^+(\widetilde{a}_2,\widetilde{a}'_2)A^+(x,x')A^+(y,y')A^+(p,p')A^+(q,q').$$

with A^+ of (6.4).

Similarly to Lemma 6.1 one can get that the largest eigenvalue of A_a^+ is $\beta^2(\lambda_0^+)^6 + O(W^{-1})$ (see (6.8)), and the next eigenvalue is smaller then $\beta^2(\lambda_0^+)^6(1-\delta)$. Remember that we have normalization $\lambda_0(K_\pm)^{-1}$, and $\lambda_0(K_\pm) = \lambda_0^+\lambda_0^- + O(1/W)$ (see Lemma 6.1). But, according to (6.9), $\beta|\lambda_0^+|^2 < 1$, thus

$$\|\lambda_0(K_\pm)^{-1}K_+\| < 1-\delta,$$

and so

$$I_+ = O(e^{-cn}).$$

A SUSY techniques

Here we provide the basic formulas and definitions of SUSY approach used in Section 2.

Let us consider two sets of formal variables $\{\psi_j\}_{j=1}^n, \{\overline{\psi}_j\}_{j=1}^n$, which satisfy the anticommutation conditions

$$\psi_j\psi_k + \psi_k\psi_j = \overline{\psi}_j\psi_k + \psi_k\overline{\psi}_j = \overline{\psi}_j\overline{\psi}_k + \overline{\psi}_k\overline{\psi}_j = 0, \quad j,k = 1,\dots,n. \tag{A.1}$$

Note that this definition implies $\psi_j^2 = \overline{\psi}_j^2 = 0$. These two sets of variables $\{\psi_j\}_{j=1}^n$ and $\{\overline{\psi}_j\}_{j=1}^n$ generate the Grassmann algebra $\mathfrak{A}$. Taking into account that $\psi_j^2 = 0$, we have that all elements of $\mathfrak{A}$ are polynomials of $\{\psi_j\}_{j=1}^n$ and $\{\overline{\psi}_j\}_{j=1}^n$ of degree at most one in each variable. We can also define functions of the Grassmann variables. Let χ be an element of $\mathfrak{A}$, i.e.

$$\chi = a + \sum_{j=1}^n (a_j\psi_j + b_j\overline{\psi}_j) + \sum_{j\neq k}(a_{j,k}\psi_j\psi_k + b_{j,k}\psi_j\overline{\psi}_k + c_{j,k}\overline{\psi}_j\overline{\psi}_k) + \dots. \tag{A.2}$$

For any sufficiently smooth function f we define by $f(\chi)$ the element of $\mathfrak{A}$ obtained by substituting $\chi - a$ in the Taylor series of f at the point a. Since χ is a polynomial of $\{\psi_j\}_{j=1}^n$, $\{\overline{\psi}_j\}_{j=1}^n$ of the form (A.2), according to (A.1) there exists such l that $(\chi-a)^l = 0$, and hence the series terminates after a finite number of terms and so $f(\chi) \in \mathfrak{A}$.

For example, we have

$$\begin{aligned}
&\exp\{a\,\overline{\psi}_1\psi_1\} = 1 + a\,\overline{\psi}_1\psi_1 + (a\,\overline{\psi}_1\psi_1)^2/2 + \ldots = 1 + a\,\overline{\psi}_1\psi_1,\\
&\exp\{a_{11}\overline{\psi}_1\psi_1 + a_{12}\overline{\psi}_1\psi_2 + a_{21}\overline{\psi}_2\psi_1 + a_{22}\overline{\psi}_2\psi_2\} = 1 + a_{11}\overline{\psi}_1\psi_1\\
&+ a_{12}\overline{\psi}_1\psi_2 + a_{21}\overline{\psi}_2\psi_1 + a_{22}\overline{\psi}_2\psi_2 + (a_{11}\overline{\psi}_1\psi_1 + a_{12}\overline{\psi}_1\psi_2\\
&+ a_{21}\overline{\psi}_2\psi_1 + a_{22}\overline{\psi}_2\psi_2)^2/2 + \ldots = 1 + a_{11}\overline{\psi}_1\psi_1 + a_{12}\overline{\psi}_1\psi_2 + a_{21}\overline{\psi}_2\psi_1\\
&+ a_{22}\overline{\psi}_2\psi_2 + (a_{11}a_{22} - a_{12}a_{21})\overline{\psi}_1\psi_1\overline{\psi}_2\psi_2.
\end{aligned} \tag{A.3}$$

Following Berezin [2], we define the operation of integration with respect to the anticommuting variables in a formal way:

$$\int d\,\psi_j = \int d\,\overline{\psi}_j = 0, \quad \int \psi_j d\,\psi_j = \int \overline{\psi}_j d\,\overline{\psi}_j = 1, \tag{A.4}$$

and then extend the definition to the general element of $\mathfrak{A}$ by linearity. A multiple integral is defined to be a repeated integral. Assume also that the "differentials" $d\,\psi_j$ and $d\,\overline{\psi}_k$ anticommute with each other and with the variables ψ_j and $\overline{\psi}_k$. Thus, according to the definition, if

$$f(\psi_1,\dots,\psi_k) = p_0 + \sum_{j_1=1}^k p_{j_1}\psi_{j_1} + \sum_{j_1<j_2} p_{j_1,j_2}\psi_{j_1}\psi_{j_2} + \ldots + p_{1,2,\dots,k}\psi_1\dots\psi_k,$$

then

$$\int f(\psi_1,\dots,\psi_k)d\,\psi_k\dots d\,\psi_1 = p_{1,2,\dots,k}. \tag{A.5}$$

Let A be an ordinary Hermitian matrix with positive real part. The following Gaussian integral is well-known

$$\int \exp\Big\{-\sum_{j,k=1}^{n} A_{jk}z_j\overline{z}_k\Big\}\prod_{j=1}^{n}\frac{d\,\Re z_j d\,\Im z_j}{\pi} = \frac{1}{\det A}. \tag{A.6}$$

One of the important formulas of the Grassmann variables theory is the analog of this formula for the Grassmann algebra (see [2]):

$$\int \exp\Big\{-\sum_{j,k=1}^{n} A_{jk}\overline{\psi}_j\psi_k\Big\}\prod_{j=1}^{n} d\,\overline{\psi}_j d\,\psi_j = \det A, \tag{A.7}$$

where A now is any $n\times n$ matrix.

For $n=1$ and 2 this formula follows immediately from (A.3) and (A.5).

We will also need the following proposition

Proposition A.1 (see [21] and references therein)**.** *Let $\psi_j=(\psi_{j1},\dots,\psi_{jm})^t$, $j=1,\dots,p$ be the Grassman vectors, and let F be some function that depends only on combinations*

$$\psi^+\psi := \Big\{\sum_{\alpha=1}^{m}\bar\psi_{j\alpha}\psi_{k\alpha}\Big\}_{j,k=1}^{p},\quad \psi\psi^t := \Big\{\sum_{\alpha=1}^{m}\psi_{j\alpha}\psi_{k\alpha}\Big\}_{j,k=1}^{p},\quad \psi^+\bar\psi := \Big\{\sum_{\alpha=1}^{m}\bar\psi_{j\alpha}\bar\psi_{k\alpha}\Big\}_{j,k=1}^{p}$$

and set

$$d\Psi = \prod_{j=1}^{p}\prod_{\alpha=1}^{m} d\bar\psi_{j\alpha}d\psi_{j\alpha}.$$

Assume also that $m\ge p$. Then

$$\int F\begin{pmatrix}\psi^+\psi & \psi^+\bar\psi\\ \psi^t\psi & \psi^+\psi\end{pmatrix} d\Phi d\Psi = C_{p,m}\int F(Q)\cdot\det{}^{-m/2}Q\,d\mu(Q),$$

where $C_{p,m}$ is some constant depending on p and m, $Q\in Sp(p)$, and $d\mu(Q)$ is a Haar measure over $Sp(p)$.

B Proof of Proposition 4.2

The first part of Proposition 4.2 follows from the standard representation theory arguments and can be found e.g. in [19], Ch.5. The recurrence relation (4.12) follows from the recurrence relation for hypergeometric functions, see e.g. [1].

Notice also that operator $\hat\nu$ correspond to the multiplication on $c(2x^2-1)$ with $x=\sqrt{S(Q)}$ (see (4.8)). Thus, (4.12) gives that $\hat\nu\phi_0$ is proportional to ϕ_1, which implies (4.13).

To get the asymptotic expression (4.14) for the eigenvalues of K_* we need the Itzykson-Zuber formula of the integration over $\mathring{Sp}(2)$ (for the proof see, e.g., [33])

Proposition B.1. *If $p\neq 0$, then*

$$\int_{\mathring{Sp}(2)}\exp\{-p\,S\big(Q(Q')^*\big)\}d\,\mu(Q') = \frac{6}{p^2}\Big(1-2/p+e^{-p}\big(1+2/p\big)\Big). \tag{B.1}$$

Moreover,

$$\int_{\mathring{Sp}(2)}\exp\{i\pi\xi - 2i\pi\xi S\big(Q\big)\}d\,\mu(Q) = DS(\pi\xi). \tag{B.2}$$

Given (4.12), it is easy to check that $P_{2n}(0) = (-1)^n$, and the coefficient at x^2 of P_{2n} is $(-1)^{n-1}n(n+3)/2$. Therefore

$$\begin{aligned}\lambda_j(t) &= \frac{p^2}{6}\int_{\mathring{S}p(2)} \exp\{-p\,S\big(Q\big)\}(-1)^j P_{2j}(Q)\,d\mu(Q)\\ &= \frac{p^2}{6}\int_{\mathring{S}p(2)} \exp\{-p\,S\big(Q\big)\}\Big(1-\frac{j(j+3)}{2}S(Q)+\dots\Big)\,d\mu(Q)\\ &= 1-\frac{2}{p}+\frac{p^2}{6}\Big(-\frac{12}{p^3}\Big(1-\frac{2}{p}\Big)+\frac{6}{p^2}\cdot\frac{2}{p^2}\Big)\cdot\frac{j(j+3)}{2}+O((j^2/Wt)^2)\\ &= 1-\frac{(j+1)(j+2)}{Wt}+O((j^2/Wt)^2)\end{aligned}$$

with $p = Wt$. Here we used $j(j+3)+2 = (j+1)(j+2)$, (B.1), and

$$\int_{\mathring{S}p(2)} \exp\{-p\,S\big(Q\big)\}(S(Q))^k\,d\mu(Q) = (-1)^k\Big(\frac{d}{dp}\Big)^k\int_{\mathring{S}p(2)} \exp\{-p\,S\big(Q\big)\}\,d\mu(Q).$$

References

[1] Abramowitz, M., Stegun, I. A.: Handbook of Mathematical Functions, *Dover*, NY, 1965. MR0208797

[2] Berezin, F.A.: Introduction to the algebra and analysis of anticommuting variables. *Moscow State University Publ.*, Moscow, 1983 (Russian) MR0732126

[3] Bogachev, L. V., Molchanov, S. A., and Pastur, L. A.: On the level density of random band matrices. *Mat. Zametki* **50:6**, (1991), 31–42 MR1150631

[4] Bourgade, P.: Random Band Matrices, *Proceedings of the ICM*, Rio de Janeiro, (2018), 2745–2770 MR3966510

[5] Bourgade, P., Erdős, L., Yau, H.-T., Yin, J.: Universality for a class of random band matrices, *Advances in Theor. and Math. Physics* **21:3**, (2017), 739–800 MR3695802

[6] Bourgade, P., Yau, H.-T., Yin, J.: Random band matrices in the delocalized phase, I: Quantum unique ergodicity and universality, *Commun. Pure and Applied Math.* **73:7**, (2020), 1526–1596 MR4156609

[7] Bourgade, P., Yang, F., Yau, H.-T., Yin, J.: Random band matrices in the delocalized phase, II: Generalized resolvent estimates, *J.Stat. Phys.* **174**, (2019), 1189–1221 MR3934695

[8] Brézin, E., Hikami, S.: Characteristic polynomials of real symmetric random matrices. *Commun. Math. Phys.* **223**, (2001), 363–382 MR1864437

[9] Borodin, A., Strahov, E.: Averages of characteristic polynomials in random matrix theory. *Comm. Pure Appl. Math.* **59**, (2006), 161–253 MR2190222

[10] Casati, G., Molinari, L., Israilev, F.: Scaling properties of band random matrices, *Phys. Rev. Lett.* **64**, (1990), 1851–1854 MR1046365

[11] Disertori, M., Lager, M.: Density of states for random band matrices in two dimensions, *Ann. Henri Poincare* **18:7**, (2017), 2367–2413 MR3665217

[12] Disertori, M., Pinson, H., Spencer, T.: Density of states for random band matrices. *Comm. Math. Phys.* **232**, (2002), 83–124 MR1942858

[13] Efetov, K.: Supersymmetry in disorder and chaos. *Cambridge university press*, New York, 1997 MR1628498

[14] Erdős, L., Knowles, A.: Quantum diffusion and eigenfunction delocalization in a random band matrix model. *Commun. Math. Phys.* **303**, (2011), 509–554 MR2782623

[15] Erdős, L., Knowles, A., Yau, H.-T., Yin, J.: Delocalization and diffusion profile for random band matrices, *Commun. Math. Phys.* **323**, (2013), 367–416 MR3085669

[16] Erdős, L., Yau, H.-T., Yin, J.: Bulk universality for generalized Wigner matrices, *Probab. Theory Relat. Fields* **154**, (2012), 341–407 MR2981427

[17] Fyodorov, Y.V., Mirlin, A.D.: Scaling properties of localization in random band matrices: a σ-model approach, *Phys. Rev. Lett.* **67**, (1991), 2405–2409 MR1130103

[18] He, Y., Marcozzi, M., Diffusion Profile for Random Band Matrices: a Short Proof, *J. Stat. Phys.* **177**, (2019), 666–716 MR4027579

[19] Helgason, S.: Groups and Geometric Analysis: Integral Geometry, Invariant Differential Operators, and Spherical Functions, *Mathematical Surveys and Monographs* **83**, 1984, 667 pp. MR1790156

[20] Kösters, H.: On the second-order correlation function of the characteristic polynomial of a real symmetric Wigner matrix, *Elect. Comm. in Probab.* **13**, (2008), 435–447 MR2430711

[21] Littelmann, P., Sommers, H. − J., Zirnbauer, M. R.: Superbosonization of invariant random matrix ensembles. *Commun. Math. Phys.* **283**, (2008), 343–395 MR2430637

[22] Mehta, M.L.: Random Matrices, *Academic Press*, New York, 1991 MR1083764

[23] Mirlin, A. D.: Statistics of energy levels. New Directions in Quantum Chaos, (Proceedings of the International School of Physics "Enrico Fermi", Course CXLIII), ed. by G.Casati, I.Guarneri, U.Smilansky, *IOS Press*, Amsterdam, (2000), 223-298 MR1843504

[24] Molchanov, S. A., Pastur, L. A., Khorunzhii, A. M.: Distribution of the eigenvalues of random band matrices in the limit of their infinite order, *Theor. Math. Phys.* **90**, (1992), 108–118 MR1182292

[25] Pastur L., Shcherbina, M.: Eigenvalue distribution of large random matrices, *American Mathematical Society*, 2011, 634 pp. MR2808038

[26] Peled, R., Schenker, J., Shamis, M., Sodin, A.: On the Wegner orbital model, *International Mathematics Research Notices* **4**, (2019), 1030–1058 MR3915294

[27] Schenker, J.: Eigenvector localization for random band matrices with power law band width, *Commun. Math. Phys.* **290**, (2009), 1065–1097 MR2525652

[28] Schäfer, L., Wegner, F.: Disordered system with n orbitals per site: Lagrange formulation, hyperbolic symmetry, and Goldstone modes. *Z. Phys. B* **38**, (1980), 113-126 MR0575503

[29] Shcherbina, T.: On the second mixed moment of the characteristic polynomials of the 1D band matrices. *Commun. Math. Phys.* **328**, (2014), 45–82 MR3196980

[30] Shcherbina, M., Shcherbina, T.: Characteristic polynomials for 1d random band matrices from the localization side, *Commun. Math. Phys.* **351**, (2017), 1009–1044 MR3623245

[31] Shcherbina, M., Shcherbina, T.: Universality for 1d random band matrices, *Commun. Math. Phys.* **385**, (2021), 667–716 MR4278281

[32] Shcherbina, T.: Characteristic polynomials of random band matrices near the threshold, *J.Stat.Phys.* **179:4**, (2020), 920–944 MR4102449

[33] Shcherbina, T.: Universality of the second mixed moment of the characteristic polynomials of the 1D band matrices: real symmetric case, *J. Math. Phys.* **56**, (2015), pp.29 MR3369897

[34] Spencer, T.: SUSY statistical mechanics and random band matrices. Quantum many body system, Cetraro, Italy 2010, *Lecture notes in mathematics 2051 (CIME Foundation subseries)*, 2012 MR2953867

[35] Tao, T., Vu., V.: Random matrices: universality of local eigenvalue statistics. *Acta Math.* **206**, (2011), 127–204 MR2784665

[36] Vilenkin, N. Ja.: Special Functions and the Theory of Group Representations. *Translations of Mathematical Monographs, AMS*, 1968, 613 pp. MR0229863

[37] Wegner, F.J.: Disordered system with n orbitals per site: $n \to \infty$ limit, *Phys. Rev. B* **19**, (1979), 783–792

[38] Yang, F., Yin, J.: Random band matrices in the delocalized phase, III: Averaging fluctuations, *Probab. Theory Relat. Fields* **179**, (2021), 451–540 MR4221663

Electronic Journal of Probability
Electronic Communications in Probability

Advantages of publishing in EJP-ECP

- Very high standards
- Free for authors, free for readers
- Quick publication (no backlog)
- Secure publication (LOCKSS[1])
- Easy interface (EJMS[2])

Economical model of EJP-ECP

- Non profit, sponsored by IMS[3], BS[4] , ProjectEuclid[5]
- Purely electronic

Help keep the journal free and vigorous

- Donate to the IMS open access fund[6] (click here to donate!)
- Submit your best articles to EJP-ECP
- Choose EJP-ECP over for-profit journals

[1]LOCKSS: Lots of Copies Keep Stuff Safe `http://www.lockss.org/`
[2]EJMS: Electronic Journal Management System `http://www.vtex.lt/en/ejms.html`
[3]IMS: Institute of Mathematical Statistics `http://www.imstat.org/`
[4]BS: Bernoulli Society `http://www.bernoulli-society.org/`
[5]Project Euclid: `https://projecteuclid.org/`
[6]IMS Open Access Fund: `http://www.imstat.org/publications/open.htm`