Universality for 1d Random Band Matrices

Mariya Shcherbina\textsuperscript{1,2}, Tatyana Shcherbina\textsuperscript{3,4,15}

\textsuperscript{1} Institute for Low Temperature Physics, Kharkiv, Ukraine.
E-mail: shcherbi@ilt.kharkov.ua
\textsuperscript{2} Karazin Kharkiv National University, Kharkiv, Ukraine
\textsuperscript{3} Department of Mathematics, Princeton University, Princeton, USA
\textsuperscript{4} Department of Mathematics, University of Wisconsin-Madison, Madison, USA
E-mail: tshcherbyna@wisc.edu

Received: 24 March 2020 / Accepted: 2 June 2021
Published online: 21 June 2021 – © The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2021

Abstract: We consider 1d random Hermitian $N \times N$ 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 such matrices in the bulk of the spectrum, as $W \gg \sqrt{N}$, is determined by the Wigner–Dyson statistics. The method of the proof is based on the rigorous application of supersymmetric transfer matrix approach developed in Shcherbina and Shcherbina (J Stat Phys 172:627–664, 2018).

1. Introduction

Random band matrices (RBM) provide a natural and important model to study eigenvalue statistics and quantum transport in disordered systems as they interpolate between classical Wigner matrices, i.e. Hermitian random matrices with all independent identically distributed elements, and random Schrödinger operators, where only a random on-site potential is present in addition to the deterministic Laplacian on a regular box in $d$-dimension lattice. Such matrices have various application in physics: the eigenvalue statistics of RBM is of relevance in quantum chaos, the quantum dynamics associated with RBM can be used to model conductance in thick wires, etc.

The density of states $\rho$ of a general class of RBM with $W \gg 1$ is given by the well-known Wigner semicircle law (see \cite{2,17}):

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

The main feature of RBM is that they can be used to model the celebrated Anderson metal-insulator phase transition in $d \geq 3$ (see the review \cite{28} for the details). Moreover, the crossover for RBM can be investigated even in $d = 1$ by varying the bandwidth $W$. Supported in part by NSF Grant DMS-1700009.
More precisely, the key physical parameter of RBM is the localization length, which describes the length scale of the eigenvector corresponding to the energy $E \in (-2, 2)$. The system is called delocalized if for all $E$ in the bulk of spectrum the localization length is comparable with the system size, and it is called localized otherwise. Physically, delocalized systems correspond to electric conductors, and localized systems are insulators.

The questions of the localization length are closely related to the universality conjecture of the bulk local regime of the random matrix theory. The bulk local regime deals with the behaviour of eigenvalues of $N \times N$ random matrices on the intervals whose length is of the order $O(N^{-1})$. The main objects of the local regime are $k$-point correlation functions $R_k$ ($k = 1, 2, \ldots$), which can be defined by the equalities:

$$\mathbb{E} \left\{ \sum_{j_1 \neq \ldots \neq j_k} \varphi_k (\lambda_{j_1}^{(N)}, \ldots, \lambda_{j_k}^{(N)}) \right\} = \int \varphi_k (\lambda_1^{(N)}, \ldots, \lambda_k^{(N)}) R_k (\lambda_1^{(N)}, \ldots, \lambda_k^{(N)}) d\lambda_1^{(N)} \ldots d\lambda_k^{(N)},$$

where $\varphi_k : \mathbb{R}^k \rightarrow \mathbb{C}$ is bounded, continuous and symmetric in its arguments and the summation is over all $k$-tuples of distinct integers $j_1, \ldots, j_k \in \{1, \ldots, N\}$. According to the Wigner–Dyson universality conjecture, the local behaviour of the eigenvalues does not depend on the matrix probability law (ensemble) and is determined only by the symmetry type of matrices (real symmetric, Hermitian, or quaternion real in the case of real eigenvalues and orthogonal, unitary or symplectic in the case of eigenvalues on the unit circle). For example, the conjecture states that for Hermitian random matrices in the bulk of the spectrum and in the range of parameters for which the eigenvectors are delocalized

$$\frac{1}{(N \rho(E))^k} R_k \left( E + \frac{\xi_1}{\rho(E) N}, \ldots, E + \frac{\xi_k}{\rho(E) N} \right) \longrightarrow \det \left\{ \frac{\sin \pi (\xi_i - \xi_j)}{\pi (\xi_i - \xi_j)} \right\}_{i,j=1}^k \quad N \rightarrow \infty,$$

for any fixed $k$. This means that the limit coincides with that for GUE.

One of the main long standing problems in the field is to prove a fundamental physical conjecture formulated in late 80th (see [7,14]). 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 $W$, and by Poisson statistics for a small $W$ (with exponentially localized eigenvectors). This is an analogue of the celebrated Anderson metal-insulator transition for random Schrödinger operators.

For 1d RBM the transition is conjectured to occur around the critical value $W = \sqrt{N}$. The conjecture is supported by the physical derivation due to Fyodorov and Mirlin (see [14]) based on supersymmetric formalism, and also by the so-called Thouless scaling. However, so far there were only a few partial results on the mathematical level of rigour. Localization of eigenvectors in the bulk of the spectrum was first shown for $W \ll N^{1/8}$ in [20], and then the bound was improved to $N^{1/7}$ in [18]. On the other side, the development of the Erdős–Schlein–Yau approach to Wigner matrices (see [13]) led to results where the weaker form of delocalization was proved for $W \gg N^{6/7}$ in [11], $W \gg N^{4/5}$ in [12], $W \gg N^{7/9}$ in [16]. More recently, this approach together with the new idea of quantum unique ergodicity gave first the GUE/GOE gap distributions for RBM with $W \sim N$ [4], and then it was developed in [5,6,31] to obtain bulk universality and complete delocalization in the range $W \gg N^{3/4}$ (see review [3] for the details). We
mention also that at the edge of the spectrum the transition for 1d band matrices (with critical exponent \( N^{5/6} \)) was understood in [27] by the method of moments.

The main aim of this paper is to prove (1.3) in the range \( W \gg \sqrt{N} \) for the Gaussian Hermitian block RBM, which are RBM with some specific covariance profile. More precisely, we consider Hermitian matrices \( H_N, N = nW \) with elements \( H_{jk,\alpha\beta} \), where \( j, k \in 1, \ldots, n \) (they parametrize the lattice sites) and \( \alpha, \beta = 1, \ldots, W \) (they parametrize the orbitals on each site). The entries \( H_{jk,\alpha\beta} \) are random Gaussian variables with mean zero such that

\[
\langle H_{j_1k_1,\alpha_1\beta_1} H_{j_2k_2,\alpha_2\beta_2} \rangle = \delta_{j_1j_2} \delta_{k_1k_2} \delta_{\alpha_1\alpha_2} \delta_{\beta_1\beta_2} J_{j_1k_1} J_{j_2k_2}.
\]

(1.4)

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

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

Such models were first introduced and studied by Wegner (see [19, 30]).

We restricted ourselves to the case

\[
J = 1/W + \beta \Delta/W, \quad \beta < 1/4,
\]

(1.5)

where \( W \gg 1 \) and \( \Delta \) is the discrete Laplacian on \([1, n] \cap \mathbb{Z}\) with Neumann boundary conditions. This model is one of the possible realizations of the Gaussian random band matrices with the band width \( 2W + 1 \) (notice that the model can be defined similarly in any dimensions \( d > 1 \) taking \( j, k \in [1, n]^d \cap \mathbb{Z}^d \) in (1.4)).

The main result of the paper is the following theorem

**Theorem 1.1.** In the dimension \( d = 1 \) the behaviour of the second order correlation function (1.2) of (1.4)–(1.5), as \( W \geq C n \log^5 n \), in the bulk of the spectrum coincides with that for the GUE. More precisely, if \( \Lambda = [1, n] \cap \mathbb{Z} \) and \( H_N, N = Wn \) are matrices (1.4) with \( J \) of (1.5), then for any \( E \in (-2, 2) \)

\[
(N \rho(E))^{-2} R_2 \left( E + \frac{\xi_1}{\rho(E) N}, E + \frac{\xi_2}{\rho(E) N} \right) \to 1 - \frac{\sin^2(\pi(\xi_1 - \xi_2))}{\pi^2(\xi_1 - \xi_2)^2},
\]

(1.6)

in the limit \( n, W \to \infty, W \geq C n \log^5 n \).

In order to prove Theorem 1.1, we apply a rigorous form of the supersymmetric (SUSY) transfer matrix approach. The approach is based on the fact that the main spectral characteristics of RBM (such as density of states, second correlation functions, or the average of an elements of the resolvent) can be written as the averages of certain observables in some SUSY statistical mechanics models containing both complex and Grassmann variables (so-called dual representation in terms of SUSY). The rigorous analysis of such integral representation is usually very complicated and requires powerful analytic and statistical mechanics tools. In our case the specific form of the covariance (1.5) allows to combine the SUSY techniques with a transfer matrix approach. The supersymmetric transfer matrix formalism in this context was first suggested by Efetov (see [10]) and on a heuristic level it was adapted specifically for RBM in [15] (see also references therein). However the rigorous application of the method to the main spectral characteristics of RBM is quite difficult due to the complicated structure and non self-adjointness of the corresponding transfer operator. During the last years, the techniques were developed
step by step (see [24] for details). First we applied it in [21] to obtain the precise estimate for the density of state. Then the method was elaborated in [22] to study the localized regime of the second correlation function of characteristic polynomials, which together with the result of [25] finished the proof of the transition around \( W \sim N^{1/2} \) on the level of characteristic polynomials. The next crucial step was done in [23], where we applied the techniques to the so-called sigma-model approximation, which is often used by physicists to study complicated statistical mechanics systems. In such approximation spins take values in some symmetric space (\( \pm 1 \) for Ising model, \( S^1 \) for the rotator, \( S^2 \) for the classical Heisenberg model, etc.). It is expected that sigma-models have all the qualitative physics of more complicated models with the same symmetry (for more details see, e.g., [28]). The sigma-model approximation for RBM was introduced by Efetov (see [10]), and the spins there are \( 4 \times 4 \) matrices with both complex and Grassmann entries. As it was shown in [23], the mechanism of the crossover for the sigma-model is essentially the same as for the correlation functions of characteristic polynomials (see [22]), but the structure of the transfer operator for the sigma-model is more complicated: it is a \( 6 \times 6 \) matrix kernel whose entries are kernels depending on two unitary \( 2 \times 2 \) matrices \( U, U' \) and two hyperbolic \( 2 \times 2 \) matrices \( S, S' \). As it will be shown below, in the case of the second correlation function of (1.4)–(1.5) which is the main point of interest in this paper, the transfer operator \( \mathcal{K} \) becomes a \( 70 \times 70 \) matrix whose elements are kernels defined on \( L^2(\mathcal{H}_2^+ L) \), where \( U(2) \) is \( 2 \times 2 \) unitary group, \( \mathcal{H}_2^+ \) is the space of \( 2 \times 2 \) positive Hermitian matrices, and \( L = \text{diag}(1, -1) \), and so the spectral analysis of \( \mathcal{K} \) poses serious structural problems. The key idea of this analysis is to prove that the main part of \( \mathcal{K} \) is still the \( 6 \times 6 \) matrix kernel that appeared in the transfer operator corresponding to the sigma-model approximation.

We would like to mention also that the model (1.4)–(1.5) in any dimension but with a finite number of blocks was analysed in [26] via SUSY techniques combined with a delicate steepest descent method. Combining the approach of [26] with the Green’s function comparison strategy, the delocalization for \( W \gg N^{6/7} \) has been proved in [1] for the block band matrices (1.4) with a rather general (non-Gaussian) distribution of the elements.

Let us add also that SUSY approach together with cluster expansion was used in [9] to obtain the local semicircle low at arbitrary short scales and smoothness in energy (in the limit of infinite volume and fixed large band width \( W \)) for the averaged density of states of a special case of Gaussian RBM in dimension 3. The technique of that paper was used in [8] to obtain the same result in 2d. As was mentioned above, a similar result in 1d was obtained in [21] by the SUSY transfer matrix approach.

Notice that according to the Stieltjes-Perron formula for the inversion of the Stieltjes transform

\[
\text{if } m(z) = \int \tilde{\rho}(\lambda) \frac{d\lambda}{\lambda - z}, \text{ then } \tilde{\rho}(\lambda) = \pi^{-1} \lim_{\varepsilon \to 0} \Im m(\lambda + i\varepsilon).
\]

Hence, to prove Theorem 1.1, it suffices to show that

\[
\lim_{\varepsilon \to 0} \lim_{n \to \infty} F_2(z_1, z_2) = 1 - \frac{\sin^2(\pi(\xi_1 - \xi_2))}{\pi^2(\xi_1 - \xi_2)^2}, \quad (1.7)
\]

where

\[
F_2(z_1, z_2) := (2\pi i N \rho(E))^{-2} E \left\{ \text{Tr} \left( (H - z_1)^{-1} - (H - z_1)^{-1} \right) \cdot \text{Tr} \left( (H - z_2)^{-1} - (H - z_2)^{-1} \right) \right\}. \quad (1.8)
\]
with
\[ z_1 = E + i\varepsilon/N + \xi_1/N\rho(E), \quad z_2 = E + i\varepsilon/N + \xi_2/N\rho(E), \]
\[ z'_1 = E + i\varepsilon/N + \xi'_1/N\rho(E), \quad z'_2 = E + i\varepsilon/N + \xi'_2/N\rho(E), \]
(1.9)

\(\varepsilon > 0,\) and \(\xi_1, \xi_2, \xi'_1, \xi'_2 \in [-C, C] \subseteq \mathbb{R}.\) Since

\[(2\pi i N\rho(E))^2 F_2(z_1, z_2) \]
\[= E\left\{\text{Tr}(H - z_1)^{-1} \cdot \text{Tr}(H - z_2)^{-1}\right\} + E\left\{\overline{\text{Tr}(H - z_1)^{-1} \cdot \text{Tr}(H - z_2)^{-1}}\right\} - E\left\{\text{Tr}(H - z_1)^{-1} \cdot \text{Tr}(H - \overline{z_2})^{-1}\right\} - E\left\{\overline{\text{Tr}(H - z_1)^{-1} \cdot \text{Tr}(H - \overline{z_2})^{-1}}\right\},\]

we get
\[ F_2(z_1, z_2) = (2\pi)^{-2} \frac{\partial^2}{\partial \xi_1^2 \partial \xi_2^2}\left(\mathcal{R}^+_{Wn}(E, \varepsilon, \bar{\xi}) + \overline{\mathcal{R}^+_{Wn}(E, \varepsilon, \bar{\xi})} - \mathcal{R}^+_{Wn}(E, \varepsilon, \bar{\xi}) - \overline{\mathcal{R}^+_{Wn}(E, \varepsilon, \bar{\xi})}\right)|_{\xi'' = \xi}, \]
(1.10)

where \(\xi'' = \xi\) means \(\xi'_1 = \xi_1, \xi'_2 = \xi_2,\) and
\[ \mathcal{R}^+_{Wn}(E, \varepsilon, \bar{\xi}) = E\left\{\frac{\det(H - z_1)\det(H - z_2)}{\det(H - z'_1)\det(H - z'_2)}\right\}, \]
\[ \mathcal{R}^+_{Wn}(E, \varepsilon, \bar{\xi}) = E\left\{\frac{\det(H - z_1)\det(H - z_2)}{\det(H - z'_1)\det(H - z'_2)}\right\}, \]
(1.11)

are the generalized correlation functions evaluated at \(\tilde{\xi} = (\xi_1, \xi_2, \xi'_1, \xi'_2).\)

Introduce also
\[ a_{\pm} = \frac{iE \pm \sqrt{4 - E^2}}{2}, \quad a_{\pm} = : e^{i\varphi_0}. \]
(1.12)

Similarly to [23], the main result of Theorem 1.1 follows from two theorems dealing with the behaviour of generalized correlation functions (1.11):

**Theorem 1.2.** Given \(\mathcal{R}^+_{Wn}\) of (1.11), (1.4) and (1.5), with any fixed \(\beta, \varepsilon > 0,\) and \(\xi = (\xi_1, \xi_2, \xi'_1, \xi'_2) \in \mathbb{C}^4 (|3\xi_j| < \varepsilon \cdot \rho(E)/2)\) we have, as \(n, W \rightarrow \infty, W \geq Cn \log^5 n:\)
\[ \mathcal{R}^+_{Wn}(E, \varepsilon, \bar{\xi}) \rightarrow e^{ia_{\pm}(\xi'_1 + \xi'_2 - \xi_1 - \xi_2)/\rho(E)}, \]
\[ \frac{\partial^2 \mathcal{R}^+_{Wn}(E, \varepsilon, \bar{\xi})}{\partial \xi'_1 \partial \xi'_2}|_{\xi'' = \xi} \rightarrow -a_{\pm}^2/\rho^2(E). \]
(1.13)

**Theorem 1.3.** Given \(\mathcal{R}^+_{Wn}\) of (1.11), (1.4), and (1.5), with any fixed \(\beta, \varepsilon > 0,\) and \(\xi = (\xi_1, \xi_2, \xi'_1, \xi'_2) \in \mathbb{C}^4 (|3\xi_j| < \varepsilon \cdot \rho(E)/2)\) we have, as \(n, W \rightarrow \infty, W \geq Cn \log^5 n:\)
\[ \mathcal{R}^+_{Wn}(E, \varepsilon, \bar{\xi}) \rightarrow e^{iE(\sigma_1 - \sigma_2)} e^{-2\pi \rho(E)\alpha_2} \left(\frac{\alpha_1}{2\alpha_2} + \frac{\alpha_2}{2\alpha_1}\right) \sinh(2\pi \rho(E)\alpha_1) \]
\[ + \cosh(2\pi \rho(E)\alpha_1) - (\sigma_1 - \sigma_2)^2 \frac{\sinh(2\pi \rho(E)\alpha_1)}{2\alpha_1\alpha_2}, \]
(1.14)
where
\[ \sigma_1 = \frac{\xi_1 + \xi_2}{2i\rho(E)}, \quad \sigma_2 = \frac{\xi'_1 + \xi'_2}{2i\rho(E)}, \quad \alpha_1 = \varepsilon + \frac{\xi_1 - \xi_2}{2i\rho(E)}, \quad \alpha_2 = \varepsilon + \frac{\xi'_1 - \xi'_2}{2i\rho(E)}. \]

In addition,
\[ \frac{\partial^2}{\partial \xi'_1 \partial \xi'_2} \mathcal{R}_{pW}^+(E, \varepsilon, \xi) \bigg|_{\xi' = \xi} \to \frac{1}{\rho^2(E)} + \frac{1 - e^{-4i\alpha_1}}{4\alpha_1^2 \rho^2(E)}. \]

Notice that Theorems 1.2–1.3 and (1.8) imply
\[ \lim_{\varepsilon \to 0} \lim_{n, W \to \infty} F_2(z_1, z_2) = \frac{2 + a_+^2 + a_-^2}{4\pi^2 \rho^2(E)} + \frac{1 - e^{-4\pi\rho(E)\alpha_1}}{4\pi^2 (2\alpha_1 \rho(E))^2} + \frac{1 - e^{-4\pi\rho(E)\tilde{\alpha}_1}}{4\pi^2 (2\tilde{\alpha}_1 \rho(E))^2} \]
\[ = \frac{(a_+ - a_-)^2}{4\pi^2 \rho^2(E)} - \frac{4e^{i2\pi(\xi_1 - \xi_2)} - e^{-i2\pi(\xi_1 - \xi_2)}}{4\pi^2 (\xi_1 - \xi_2)^2} \]
\[ = 1 + \frac{(e^{i\pi(\xi_1 - \xi_2)} - e^{-i\pi(\xi_1 - \xi_2)})^2}{4\pi^2 (\xi_1 - \xi_2)^2} = 1 + \frac{\sin^2(\pi(\xi_1 - \xi_2))}{\pi^2 (\xi_1 - \xi_2)^2}, \]
and so Theorem 1.1 indeed follows from Theorems 1.2–1.3. Here we used the equalities \( a_+a_- = -1, \ a_+ - a_- = 2\pi \rho(E) \) that follow from (1.12).

2. Representation of \( \mathcal{R}_{pW}^+ \) and \( \mathcal{R}_{pW}^{++} \) in the Operator Form and the Preliminary Transformation of the Transfer Operator

The goal of the current section is to rewrite the integral representations of \( \mathcal{R}_{pW}^+ \) and \( \mathcal{R}_{pW}^{++} \) (obtained in [26] via SUSY techniques (see Sect. 2)) in the transfer operator notation and to perform the preliminary transformation of the transfer operator. The brief introduction to the SUSY techniques can be found e.g. in [10], Chapter 1.

Set
\[ L_\pm = L_\pm^{(1)} \cup L_\pm^{(2)}, \quad L_\pm^{(1)} = \{te^{\pm i\varphi_0}, \ t \in [0, 1]\}, \]
\[ L_\pm^{(2)} = \{e^{\pm i\varphi_0} + te^{\pm i\psi}, \ t \in [0, \infty]\}, \quad \psi = \begin{cases} \varphi_0, & \varphi_0 \leq \pi/4, \\ \max\{\varphi_0 - \pi/4; \pi/2 - \varphi_0\}, & \varphi_0 > \pi/4, \end{cases} \]
where \( \varphi_0 \) is defined in (1.12).

2.1. Operator expression for \( \mathcal{R}_{pW}^+ \). Denote
\[ L = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \]
and let
\[ \hat{U}(2) = U(2)/U(1) \times U(1), \quad \hat{U}(1, 1) = U(1, 1)/U(1) \times U(1), \]
where \( U(p) \) is the group of \( p \times p \) unitary matrices, and \( U(1, 1) \) is the group of \( 2 \times 2 \) hyperbolic matrices \( S \) such that \( S^*LS = L; \)
Proposition 2.1. [26] For $E \in (-2, 2)$, $\hat{\xi} = \text{diag} \{\xi_1, \xi_2\}$, $\hat{\xi}' = \text{diag} \{\xi'_1, \xi'_2\}$, and $\varepsilon > 0$ we have

\[
\mathcal{R}^+_{Wn}(E, \varepsilon, \hat{\xi}) = W^{4n} \int \prod_{j=1}^n dX_j dY_j (\frac{\pi}{2}) \prod_{j=1}^n d\rho_j d\tau_j \\
\times \exp \left\{ \frac{\beta W}{2} \sum_{j=1}^{n-1} \text{Tr} (X_j - X_{j+1})^2 - \frac{\beta W}{2} \sum_{j=1}^{n-1} \text{Tr} (Y_j - Y_{j+1})^2 \right\} \\
\times \exp \left\{ \frac{W}{2} \sum_{j=1}^n \left( \text{Tr} X_j^2 - \text{Tr} Y_j^2 - i W \text{Tr} X_j (\Lambda_\varepsilon + \frac{\hat{\xi}}{N\rho(E)}) + i W \text{Tr} Y_j (\Lambda_\varepsilon + \frac{\hat{\xi}'}{N\rho(E)}) \right) \right\} \\
\times \exp \left\{ \beta \sum_{j=1}^{n-1} \text{Tr} (\rho_j - \rho_{j+1}) (\tau_j - \tau_{j+1}) - \sum_{j=1}^n \text{Tr} \rho_j \tau_j \right\} \prod_{j=1}^n \frac{\det W Y_j}{\det W (X_j + W^{-1} \rho_j Y_j^{-1} \tau_j)}
\]

(2.3)

where $\Lambda_\varepsilon = E \cdot I_2 + i \varepsilon L/N$, $X_j$ is $2 \times 2$ unitary matrix, $\rho_j$, $\tau_j$ are $2 \times 2$ matrices whose entries are independent Grassmann variables, and $Y_j = T_j^{-1} \hat{B}_j T_j$, $T_j \in \hat{U}(1, 1)$.

\[
\hat{B}_j = \begin{pmatrix} b_{j,1} & 0 \\ 0 & b_{j,2} \end{pmatrix}, \quad b_{j,1} \in \mathcal{L}_+, \quad b_{j,2} \in \mathcal{L}_-.
\]

Here

\[
dY_j = \frac{\pi}{2} (b_{j,1} - b_{j,2})^2 db_{j,1} db_{j,2} d\mu(T_j), \quad d\rho_j d\tau_j = \prod_{l,s=1}^2 d\rho_{j,ls} d\tau_{j,ls},
\]

and $dX_j, d\mu(T_j)$ are Haar measures over $U(2)$ and $\hat{U}(1, 1)$ respectively.

Let $\mathcal{F}_1 : U(2) \to U(2)$, $\mathcal{F}_2 : \mathcal{H}_2^+ L \to \mathcal{H}_2^+ L$ be the operators of multiplication by

\[
\mathcal{F}_1(X) = \exp \left\{ \frac{W}{4} \text{Tr} X^2 - \frac{i W}{2} \text{Tr} X (\Lambda_\varepsilon + \frac{\hat{\xi}^\prime}{N\rho(E)}) - \frac{W}{2} \log \det X \right\}, \\
\mathcal{F}_2(Y) = \exp \left\{ - \frac{W}{4} \text{Tr} Y^2 + \frac{i W}{2} \text{Tr} Y (\Lambda_\varepsilon + \frac{\hat{\xi}^\prime}{N\rho(E)}) + \frac{W}{2} \log \det Y \right\}.
\]

(2.4)

Introduce compact integral operators $K_1$ and $K_2$ in $L_2[U(2)]$ and $L_2[\mathcal{H}_2^+ L]$ with the kernels

\[
K_1(X, X') = W^2 \mathcal{F}_1(X) \exp \left\{ - \frac{\beta W}{2} \text{Tr} (X - X')^2 \right\} \mathcal{F}_1(X'), \\
K_2(Y, Y') = W^2 \mathcal{F}_2(Y) \exp \left\{ \frac{\beta W}{2} \text{Tr} (Y - Y')^2 \right\} \mathcal{F}_2(Y'),
\]

\[
\hat{Q}(\rho, \tau; \rho', \tau') = \exp \left\{ \beta \text{Tr} (\rho - \rho')(\tau - \tau') - \text{Tr} \rho' \tau' \right\} \det^{-W}(I + (X')^{-1} \rho' (Y')^{-1} \tau' / W),
\]

(2.6)
Then Proposition 2.1 yields

$$\mathcal{K} = K_1 \otimes K_2 \cdot \hat{Q}. \quad (2.7)$$

Then Proposition 2.1 yields

$$\mathcal{R}^{+-}_{Wn}(E, \epsilon, \xi) = W^4 \int \mathcal{F}_1(X) \mathcal{F}_2(Y) K^{n-1}(X, Y, \rho, \tau; X', Y', \rho', \tau') \mathcal{F}_1(X') \mathcal{F}_2(Y') \times \det^{-W}(I + (X)^{-1} \rho'(Y)^{-1} \tau' / W) \frac{dXdYd\rho d\tau}{(-\pi^2)} \frac{dX'dY'd\rho'd\tau'}{(-\pi^2)}. \quad (2.8)$$

Notice that $\hat{Q}$ and the operator of multiplication by $\det^{-W}(I + (X)^{-1} \rho'(Y)^{-1} \tau' / W)$ can be considered as operators acting on the space $\Omega_{256} \cong (L_2(U(2)) \otimes L_2[H^2_{1/2}L])^{256}$ of polynomials of Grassmann variables $\rho_{ls}^1, \tau_{ls}^1, l, s = 1, 2$ with coefficients from $L_2(U(2)) \otimes L_2[H^2_{1/2}L]$. Hence, in the natural basis of monomials in Grassmann space, all our operators can be considered as $256 \times 256$ matrices whose entries are operators on $L_2(U(2)) \otimes L_2[H^2_{1/2}L]$.

Thus, introducing the resolvent $G(z) = (\mathcal{K} - z)^{-1}$, one can write

$$\mathcal{R}^{+-}_{Wn}(E, \epsilon, \xi) = W^4(K^{n-1} f, g) = -\frac{W^4}{2\pi i} \oint_\omega z^{n-1}(G(z)f, g)dz,$$

$$f(X, Y, \rho, \tau) = \mathcal{F}_1(X) \mathcal{F}_2(Y)e^{(0)}, \quad g(X, Y, \rho, \tau) = \mathcal{F}_1(X) \mathcal{F}_2(Y) \cdot \left((\det^{-W}(I + (X)^{-1} \rho(Y)^{-1} \tau / W))^t e^{(c)} \right) \quad (2.9)$$

for any contour $\omega$ which contains all points of spectrum $\mathcal{K}$. Vectors $e^{(0)}$ and $e^{(c)}$ here are the vectors in the space of Grassmann variables $\Omega_{256}$ corresponding to 1 and to $\prod_{l,s=1}^{2} \rho_{ls}^1 \tau_{ls}^1$ respectively, and $(\det^{-W}(I + (X)^{-1} \rho(Y)^{-1} \tau / W))^t$ is the transposed operator to the operator of multiplication by $\det^{-W}(I + (X)^{-1} \rho(Y)^{-1} \tau / W)$ in the space $\Omega_{256}$. The appearance of $e^{(c)}$ in the inner product in the r.h.s. reflects the fact that, by definition, the integral over the Grassmann variables of any polynomial from $\Omega_{256}$ gives the coefficient of $\prod_{l,s=1}^{2} \rho_{ls}^1 \tau_{ls}^1$.

2.2. Operator expression for $\mathcal{R}^{++}_{Wn}$

**Proposition 2.2.** [26] For $E \in (-2, 2)$, $\xi = \text{diag} \{\xi_1, \xi_2\}$, $\xi' = \text{diag} \{\xi'_1, \xi'_2\}$, and $\epsilon > 0$ we have

$$\mathcal{R}^{++}_{Wn}(E, \epsilon, \xi) = W^{4n} \int \exp \left\{ \Theta_{n, W} \right\} \prod_{j=1}^{n} \frac{\det^W Y_j dX_j dY_j^+ d\rho_j d\tau_j}{(-\pi^2) \det^W (X_j + W^{-1} \rho_j(Y_j)^{-1} \tau_j)}, \quad (2.10)$$
where

\[
\Theta_{n,W} = \frac{\beta W}{2} \sum_{j=1}^{n-1} \text{Tr} (X_j - X_{j+1})^2 - \frac{\beta W}{2} \sum_{j=1}^{n-1} \text{Tr} (Y_j^* - Y_{j+1}^*)^2 + \frac{W}{2} \sum_{j=1}^{n} \left( \text{Tr} X_j^2 - \text{Tr} (Y_j^*)^2 \right)
\]

\[+ \sum_{j=1}^{n} \left( -iW \text{Tr} X_j ((iE + \varepsilon/N)I_2 + \frac{\xi}{N\rho(E)}) + iW \text{Tr} Y_j^* ((iE + \varepsilon/N)I_2 + \frac{\xi'}{N\rho(E)}) \right)\]

\[+ \beta \sum_{j=1}^{n-1} \text{Tr} (\rho_j - \rho_{j+1})(\tau_j - \tau_{j+1}) - \frac{n}{2} \text{Tr} \rho_j \tau_j,
\]

with \( Y_j^+ \in \mathcal{H}_2^+ \),

\[dY_j^+ = 1_{Y_j^+ > 0} \cdot d\Omega_{12,j} \cdot d\Omega_{1,j} \cdot dY_{11,j} \cdot dY_{22,j},\]

and \( X_j, \rho_j, \tau_j \) are the same as in Proposition 2.1.

Similarly to (2.4) we define \( \mathcal{F}_1^+ : U(2) \to U(2), \mathcal{F}_2^+ : \mathcal{H}_2^+ \to \mathcal{H}_2^+ \) as the operators of multiplication by

\[
\mathcal{F}_1^+(X) = \exp \left\{ \frac{W}{4} \text{Tr} X^2 - \frac{iW}{2} \text{Tr} X ((iE + \varepsilon/N)I_2 + \frac{\tilde{\xi}}{N\rho(E)}) - \frac{W}{2} \log \det X \right\},
\]

\[
\mathcal{F}_2^+(Y) = \exp \left\{ -\frac{W}{4} \text{Tr} Y^2 + \frac{iW}{2} \text{Tr} Y ((iE + \varepsilon/N)I_2 + \frac{\tilde{\xi}'}{N\rho(E)}) + \frac{W}{2} \log \det Y \right\},
\]

and introduce compact integral operators \( K_1^+ \) and \( K_2^+ \) in \( L_2[U(2)] \) and \( L_2[\mathcal{H}_2^+] \) with the kernels

\[
K_1^+(X, X') = W^2 \mathcal{F}_1^+(X) \exp \left\{ -\frac{\beta W}{2} \text{Tr} (X - X')^2 \right\} \mathcal{F}_1^+(X'),
\]

\[
K_2^+(Y, Y') = W^2 \mathcal{F}_2^+(Y) \exp \left\{ \frac{\beta W}{2} \text{Tr} (Y - Y')^2 \right\} \mathcal{F}_2^+(Y').
\]

Set

\[
\mathcal{K}^+ = K_1^+ \otimes K_2^+ \cdot Q.
\]

Similarly to Sect. 2.1, introducing the resolvent \( \mathcal{G}^+(z) = (\mathcal{K}^+ - z)^{-1} \), one can write

\[
\mathcal{R}_{Wn}^{\mathcal{K}^+}(E, \varepsilon, \tilde{\xi}) = W^4((\mathcal{K}^+)^n f, g) = \frac{W^4}{2\pi i} \int_\omega z^{n-1} (\mathcal{G}^+(z)f, g)dz,
\]

\[
f^+(X, Y, \rho, \tau) = \mathcal{F}_1^+(X)\mathcal{F}_2^+(Y)e^{(0)},
\]

\[
g^+(X, Y, \rho, \tau) = \mathcal{F}_1^+(X)\mathcal{F}_2^+(Y) \cdot (\det W(I + (X)^{-1} \rho (Y)^{-1} \tau / W))^{(c)},
\]

where \( \omega \) is some contour enclosing the spectrum of \( \mathcal{K}^+ \).
2.3. Preliminary transformation of $K$. Consider (2.3). Diagonalizing $X_i, Y_i$ by matrices $U_i \in \mathcal{U}(2), S_i \in \mathcal{U}(1, 1)$ such that

$$X_i = U_i^* D_{ai} U_i, \quad D_{ai} = \text{diag}(a_{i1}, a_{i2}), \quad Y_i = S_i^{-1} D_{bi} S_i, \quad D_{bi} = \text{diag}(b_{i1}, b_{i2}),$$

(2.15)
we change the measure $dX_i dY_i/(-\pi^2)$ from (2.3) to

$$(2\pi)^{-2} d\tilde{a}_i \, d\tilde{b}_i \, dU_i \, dS_i \, (a_{i1} - a_{i2})^2 (b_{i1} - b_{i2})^2,$$

where $d\tilde{a}_i = da_{i1} da_{i2}, a_{i1}, a_{i2}$ belong to the unit circle $\mathbb{T} = \{z : |z| = 1\}, d\tilde{b}_i = db_{i1} db_{i2}$ with $b_{i1} \in L_+, b_{i2} \in L_-$, and $dU_i, dS_i$ are the Haar measures on $\mathcal{U}(2), \mathcal{U}(1, 1)$ correspondingly.

Consider also the change of Grassmann variables

$$(U_i \rho_i S_i^{-1})_{a\beta} \to \rho_i, a\beta, \quad (S_i \tau_i U_i^*)_{a\beta} \to \tau_i, a\beta,$$

(2.16)
and let $\tilde{C}'(U_i, S_i)$ be the matrix corresponding to the above change in the space $\Omega_{256}$ of all polynomials of Grassmann variables $\rho_i, \tau_i, l, s = 1, 2$ with coefficients on $L_2[\mathcal{U}(2)] \otimes L_2[\mathcal{U}(1, 1)]$. Notice that the Berezinian of (2.16) is 1.

One can see that

$$K_{n-1} = C^{(n-1)/n}_{\xi, \epsilon} \prod_{i=2}^{n-1} dU_i dS_i d\tilde{a}_i d\tilde{b}_i d\rho_j d\tau_j \prod_{i=1}^{n-1} F(\tilde{a}_i, \tilde{b}_i, U_i, S_i) F(\tilde{a}_{i+1}, \tilde{b}_{i+1} U_{i+1}, S_{i+1}) \times A_{ab}(\tilde{a}_{i+1}, \tilde{a}_i; \tilde{b}_{i+1}, \tilde{b}_i) K_{U S}(U_i^* U_i, S_i^{-1} S_{i+1}) (\tilde{C}'(U_i, S_i))^{-1} \beta^{-2} \tilde{Q}(\rho, \tau; \rho', \tau') \tilde{C}'(U_i, S_i) C_{\xi, \epsilon} \epsilon^{E(\sigma_1 - \sigma_2) + 2\pi \rho(E)(\alpha_1 - \alpha_2)},$$

(2.17)
where $\alpha_i, \sigma_i$ are defined in (1.15), $\tilde{Q}$ is defined by (2.6) with $X'$ and $Y'$ replaced by diagonal matrices $\text{diag}(a_1', a_2')$ and $\text{diag}(b_1', b_2')$, $\tilde{a}_i = \{a_{i1}, a_{i2}\}, \tilde{b}_i = \{b_{i1}, b_{i2}\}$, and

$$K_U(U^* U') := (t_a W) e^{-t_a W(U^* U')_{12}^2}, \quad t_a = \beta(a_1 - a_2)(a_1' - a_2');$$

$$K_S(S^{-1} S') := (t_b W) e^{-t_b W(S^{-1} S')_{12}^2}, \quad t_b = \beta(b_1 - b_2)(b_1' - b_2');$$

$$K_{U S}(U^* U', S^{-1} S') = K_U(U^* U') K_S(S^{-1} S').$$

(2.18)
The constant $C_{\xi, \epsilon}$ is a kind of “normalization constant” here, and it appears because of our definition of $F_U(\tilde{a}, U)$ and $F_S(\tilde{b}, U)$ (see (2.21)).

We also defined

$$A_{a}(a, a') = \left(\frac{W}{2\pi}\right)^{1/2} e^{-W \Lambda(a, a')}; \quad A_{b}(b, b') = \left(\frac{W}{2\pi}\right)^{1/2} e^{W \Lambda(b, b')};$$

(2.19)
$$\Lambda(x, y) = \frac{\beta}{2} (x - y)^2 - \frac{1}{2} \phi_0(x) - \frac{1}{2} \phi_0(y) + \Re \phi_0(a_+);$$

$$\phi_0(x) = x^2/2 - ix E - \log x;$$

$$A_{ab}(\tilde{a}, \tilde{b}; \tilde{a}', \tilde{b}') = A_{a}(a_1, a_1') A_{a}(a_2, a_2') A_{b}(b_1, b_1') A_{b}(b_2, b_2');$$

(2.20)
and put
\[ F_U(\tilde{a}, U) = \exp \left\{ \frac{\sigma_1(a_1 + a_2)}{n} + \frac{\alpha_1(a_1 - a_2)}{2n} \varphi_U - \frac{\eta_1}{n} \right\}, \quad \varphi_U = \text{Tr} ULU^* L, \]
\[ F_S(\tilde{b}, S) = \exp \left\{ -\frac{\sigma_2(b_1 + b_2)}{n} - \frac{\alpha_2(b_1 - b_2)}{2n} \varphi_S + \frac{\eta_2}{n} \right\}, \quad \varphi_S = \text{Tr} SLS^{-1} L, \]
\[ \eta_{1,2} = \sigma_{1,2}(a_+ + a_-) + \alpha_{1,2}(a_+ - a_-), \quad \tilde{F}(\tilde{a}, \tilde{b}, U, S) = F_U^{1/2}(\tilde{a}, U) F_S^{1/2}(\tilde{b}, S), \]
(2.21)
where \( \alpha_{1,2} \) and \( \sigma_{1,2} \) are defined in (1.15).

Notice that \( \tilde{Q} \) of (2.6) and the operator multiplication by \( \det^{-W}(I + (X)^{-1} \rho(Y)^{-1} \tau / W) \) preserves the difference between the numbers of \( \rho \) and \( \tau \). Thus for all Grassmann operators below we can consider the restriction of these operators to the subspace \( \mathcal{Q}_{70} \subset \mathcal{Q}_{256} \) corresponding to the vectors with equal numbers of \( \rho \) and \( \tau \) (it is easy to see that there are 70 such monomials). To simplify notation, all such restrictions will be denoted by the same symbols.

Since \( K_{US}, A_{a,b} \) do not contain Grassmann variables, we can move \( (\tilde{C}'(U_i, S_i))^{-1} \) in each multiplier of (2.17) to the left. Moreover, since \( \tilde{C}'(U, S) \) corresponds to the change of variables (2.16), we have
\[ \tilde{C}'(U_{i+1}, S_{i+1})(\tilde{C}'(U_i, S_i))^{-1} = \tilde{C}'(U_{i+1}U_i^*, S_{i+1}S_i^{-1}). \]
(2.22)
Therefore (2.9) yields the following

**Proposition 2.3.** \( \mathcal{R}_W^{+\rightarrow}(E, \varepsilon, \tilde{\xi}) \) can be rewritten in the form
\[ \mathcal{R}_W^{+\rightarrow}(E, \varepsilon, \tilde{\xi}) = W^4 C_{\tilde{\xi}, \varepsilon}^{+}\left(\tilde{K}^{n-1} f, g\right) = -\frac{C_{\tilde{\xi}, \varepsilon}}{2\pi i} \int_\omega z^{n-1} (G(z) f, g) dz, \]
(2.23)
\[ G(z) = (\tilde{K} - z)^{-1}, \quad f = v(\tilde{a}, \tilde{b}, U, S)(a_1 - a_2)(b_1 - b_2)e^{(0)}, \]
\[ g = v(\tilde{a}, \tilde{b}, U, S)(a_1 - a_2)(b_1 - b_2)(D(\tilde{a}, \tilde{b}))^{(1)} e^{(c)}, \]
\[ v(\tilde{a}, \tilde{b}, U, S) := (2\pi)^{-1} e^{W(\varphi_1(\tilde{a}) + \varphi_2(\tilde{b})) / 2} e^{-W(\varphi_1(\tilde{b}) / 2 + \varphi_2(\tilde{b})) / 2} \tilde{F}(\tilde{a}, \tilde{b}, U, S), \]
\[ \tilde{K}(\tilde{a}, \tilde{b}, U, S; \tilde{a}^{'}, \tilde{b}^{'}, \tilde{\rho}, \tilde{\tau}, \tilde{U}^{(1)}, \tilde{U}^{(2)}) = A_{ab}(\tilde{a}, \tilde{a}^{(1)}; \tilde{b}, \tilde{b}^{(1)}) \tilde{F}(\tilde{a}, \tilde{b}, U, S) Q(\rho, \tau; \tilde{\rho}, \tilde{\tau}) \tilde{C}(\tilde{U}, \tilde{S}) \tilde{C}(\tilde{a}^{'}, \tilde{b}^{'}, \tilde{U}^{(1)}, \tilde{U}^{(2)}), \]
(2.24)
where the operator \( (\tilde{D}(\tilde{a}, \tilde{b}))^{(1)} \) being the transposed operator to \( \tilde{D}(\tilde{a}, \tilde{b}) \) which corresponds to the multiplication in the Grassmann space by \( (\det(1 + W^{-1} D^{-1}_a \rho D^{-1}_b \tau))^{-W} \) and we set
\[ \tilde{C}(\tilde{U}, \tilde{S}) = K_{US}(\tilde{U}, \tilde{S}) \tilde{C}'(\tilde{U}, \tilde{S}), \quad \tilde{U} = U(U')^*, \quad \tilde{S} = S(S')^{-1}. \]
(2.25)
Here \( \tilde{Q} \) is a main Grassmannian part of \( \tilde{K} \) defined in (2.6), and \( A_{ab}, \tilde{F}, K_{US}, \) and \( \tilde{C}' \) are defined in (2.18)–(2.21). Vectors \( e^{(0)} \) and \( e^{(c)} \) are the vectors in the space of Grassmann variables \( \mathcal{Q}_{70} \) corresponding to \( 1 \) and to \( \prod_{i,s=1}^2 \rho_{is} \tau_{is} \).

To study the entries of \( \tilde{C} \), it is convenient to introduce “difference” operators.

**Definition 2.1.** Given a function \( v \) defined on the space of \( 2 \times 2 \) matrices, we denote by \( \left(v(U)\right)_U \) the integral operator with the kernel \( v(U(U'^*) K_{US}(U(U'^*) \) and by \( \left(v(S)\right)_S \) the integral operator with the kernel \( v(S(S')^{-1} K_S(S(S')^{-1}) \), where \( K_{US} \) and \( K_S \) were defined in (2.18).
Recall that $G(z)$ acts in the Grassmann space $Q_{70}$ and it can be considered as a $70 \times 70$ block matrix whose entries are operators on $L_2(U(2)) \otimes L_2[H_1^* L]$. Our strategy of the proof of Theorem 1.3 is to replace the resolvent $G(z)$ by some $6 \times 6$ block matrix whose entries are operators in $L_2(U(2)) \otimes L_2(\hat{U}(1, 1))$. To this aim, we will use multiple times the following simple proposition

**Proposition 2.4.** Let the matrix $H(z)$ have the block form

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

(i) If $\|H_{12}\|\|H_{22}^{-1}\|\|H_{21}\|\|H_{11}^{-1}\| < 1$, then there exists $H^{-1}(z)$ and

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

$$G_{11} = M_1^{-1}, \quad M_1 = H_{11} - H_{12}H_{22}^{-1}H_{21}, \quad (2.26)$$

(ii) If $H_{22}^{-1}(z)$ is an analytic function for $|z| > 1 - \delta$, $\|H_{22}^{-1}\| \leq C$, $H_{11}^{-1}(z)$ is an analytic function for $|z| \geq 1 + 1/n$, $\|H_{11}\| \leq Cn$, and $\|H_{12}\|\|H_{21}\| \ll n^{-1}$, then $H^{-1}(z)$ is an analytic function for $|z| \geq 1 + 1/n$ and

$$\oint_{\omega_0} z^{n-1}(G(z)f, g)dz = \oint_{\omega_0} z^{n-1}(G_{11}f^{(1)}(z), g^{(1)}(z))dz + O(e^{-nc}), \quad (2.27)$$

$$f^{(1)}(z) = f_0 - H_{12}H_{22}^{-1}f_1, \quad g^{(1)}(z) = g_0 - H_{21}^T(H_{22}^{-1})^{-1}g_1, \quad (2.28)$$

where $f = (f_0, f_1)$, $g = (g_0, g_1)$ where $f_0$ and $g_0$ are the projections 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 $H_{22}$ respectively.

**Proof.** Formula (2.26) is the well-known Schur block matrix inversion formula. The condition $\|H_{12}\|\|H_{22}^{-1}\|\|H_{21}\|\|H_{11}^{-1}\| < 1$ guarantees that $M_1$ is invertible, hence we get (i).

Similarly, conditions of (ii) guarantee that $M_1$ is invertible for $|z| \geq 1 + 1/n$, and $G_{11}(z)$ is an analytic function here. Applying the formula (2.26) we have

$$\oint_{\omega_0} z^{n-1}(G(z)f, g)dz = \oint_{\omega_0} z^{n-1}(G_{11}f^{(1)}(z), g^{(1)}(z))dz + \oint_{\omega_0} z^{n-1}(H_{22}^{-1}f_1, g_1)dz.$$ 

For the second integral change the integration contour from $\omega_0$ to $|z| = 1 - \delta$. Then the inequality

$$|z|^{n-1} \leq (1 - \delta)^{n-1} \leq C e^{-nc}$$

yields (2.27). \( \square \)

It is easy to check that points $a_\pm$ (see (1.12)) are the stationary points of the function $\phi_0$ of (2.19). We start the analysis of $K$ of Proposition 2.3 from the restriction of the
integration with respect to \( \tilde{a}_i, \tilde{b}_i \) by the neighbourhood of \( a_\pm \). Saddle-point analysis in this case is easy. Set

\[
\Omega_+ = \{ x : |x - a_+| \leq \log W / W^{1/2} \}, \quad \Omega_- = \{ x : |x - a_-| \leq \log W / W^{1/2} \},
\]

\[
\tilde{\Omega}_\pm = \{ a_1, a'_1, b_1, b'_1 \in \Omega_+, a_2, a'_2, b_2, b'_2 \in \Omega_- \},
\]

\[
\Omega_+ = \{ a_1, a'_1, a_2, a'_2, b_1, b'_1 \in \Omega_+, b_2, b'_2 \in \Omega_- \},
\]

\[
\tilde{\Omega}_- = \{ b_1, b'_1 \in \Omega_+, a_1, a'_1, a_2, a'_2, b_2, b'_2 \in \Omega_- \},
\]

(2.29)

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

**Lemma 2.1.** Let \( \mathcal{L}_\pm \) be as defined in (2.1). Then

\[
\left( \int_{\mathcal{L}_+ \setminus \Omega_+} + \int_{\mathcal{L}_- \setminus \Omega_-} \right) |A_b(b, b')||db'| \leq C e^{-c \log^2 W},
\]

(2.30)

\[
\int_{\mathcal{T} \setminus (\Omega_+ \cup \Omega_-)} |A_a(a, a')||da'| \leq C e^{-c \log^2 W}.
\]

(2.31)

The proof of the lemma is given in Sect. 9.2.

Lemma 2.1 yields that

\[
\int dU'dS'd\tilde{a}'d\tilde{b}'(1 - \mathbf{1}_{\Omega_\pm} - \mathbf{1}_{\tilde{\Omega}_+} - \mathbf{1}_{\tilde{\Omega}_-}) \| \tilde{\mathcal{K}} \| \leq e^{-c \log^2 W}.
\]

(2.32)

Set

\[
H_{11} = (\mathbf{1}_{\Omega_\pm} \tilde{\mathcal{K}} \mathbf{1}_{\Omega_\pm}) \oplus (\mathbf{1}_{\tilde{\Omega}_+} \tilde{\mathcal{K}} \mathbf{1}_{\Omega_+}) \oplus (\mathbf{1}_{\tilde{\Omega}_-} \tilde{\mathcal{K}} \mathbf{1}_{\Omega_-}) = \mathcal{K}_\pm \oplus \mathcal{K}_+ \oplus \mathcal{K}_-.
\]

(2.33)

Then (2.32) yields

\[
\| H_{22} \| + \| H_{12} \| + \| H_{21} \| \leq C e^{-c \log^2 W}.
\]

Therefore for any \( |z| > \frac{1}{2} \)

\[
\| H_{12}(H_{22} - z)^{-1} H_{21} \| \leq C e^{-c \log^2 W}.
\]

Moreover, it will be proven below that (see Lemma 4.1 and Remark 4.1)

\[
\| (H_{11} - z)^{-1} \| \leq C n,
\]

and so for \( G_{11} \) of (2.26) we have

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

Thus we obtain by Proposition 2.4

\[
\mathcal{R}^{+,-}_{Wn}(E, \varepsilon, \bar{\xi}) = -\frac{W^4 C_{\xi,\varepsilon}}{2\pi i} \oint_{\omega} \int z^{-1}(H_{11} - z)^{-1} f, g) dz + O(e^{-c \log^2 W / 2}) + O(e^{-nc_1}).
\]

(2.34)
In view of the block structure of $H_{11}$, its resolvent also has a block structure, hence

$$
\mathcal{R}_{Wn}^{±}(E, \epsilon, \xi) = - \frac{W^4 C_{\xi, \epsilon}^±}{2 \pi i} \int_{\mathcal{S}_0} z^{n-1}(\mathcal{G}_±(z)f_±, g_±)dz - \frac{W^4 C_{\xi, \epsilon}^±}{2 \pi i} \int_{\mathcal{S}_0} z^{n-1}(\mathcal{G}_+(z)f_+, g_+)dz
$$

$$
- \frac{W^4 C_{\xi, \epsilon}^±}{2 \pi i} \int_{\mathcal{S}_0} z^{n-1}(\mathcal{G}_-(z)f_-, g_-)dz = I_± + I_+ + I_-
$$

(2.35)

the spectrum of $\mathcal{K}_+, \mathcal{K}_-$, and $\mathcal{K}_±$.

$$
\mathcal{G}_± = (\mathcal{K}_± - z)^{-1}, \quad \mathcal{G}_+(z) = (\mathcal{K}_+ - z)^{-1}, \quad \mathcal{G}_-(z) = (\mathcal{K}_- - z)^{-1},
$$

(2.36)

and $f_±, f_+, f_-, g_+, g_-$ are projections of $f$ and $g$ onto the subspaces corresponding to $\mathcal{K}_±, \mathcal{K}_+, \mathcal{K}_-$.

2.4. Analysis of $\hat{Q}$ in $\tilde{\Omega}_±$. It is not so difficult to prove that the second and the third terms $I_+$ and $I_-$ in the representation (2.35) do not contribute in the limit (see Sect. 7). Hence our main goal is to analyse the term $I_±$ of this representation.

We start the study of $I_±$ from the analysis of the main Grassmannian part $\hat{Q}$ of $\tilde{K}$ (see Proposition 2.3) in the domain $\tilde{\Omega}_±$ of (2.29). We shall call the product of Grassmann variables “good”, if it is composed only from the multipliers of the form

$$
n_{\mu\nu} = \rho_{\mu\nu} \tau_{\nu\mu}.
$$

(2.37)

We shall call “semi-good” the expressions

$$
\eta_{1,2} p(n_{11}, n_{22}), \quad \eta_1 = \rho_{12} \tau_{12}, \quad \eta_2 = \rho_{21} \tau_{21}
$$

(2.38)

with a polynomial $p$. All the rest Grassmann expressions we call “non-good”. By (2.6) (recall that now $X'$ and $Y'$ are diagonal matrices $D_{a'}$ and $D_{b'}$ of (2.15))

$$
\beta^{-2} \hat{Q}(\rho, \tau; \rho', \tau') = \beta^{-2} \left( \prod_{\mu, \nu=1,2} e^{\beta (\rho_{\mu\nu} - \rho_{\nu\mu})(\tau_{\nu\mu} - \tau_{\mu\nu})} e^{-(1 + a_1^{-1} b_1^{-1}) \rho_{\mu\nu} \tau_{\nu\mu}} \right)(1 + W^{-1} X + W^{-2} Y)
$$

$$
= \hat{\Pi}(1 + W^{-1} X + W^{-2} Y),
$$

(2.39)

where

$$
\hat{\Pi} = \left( \begin{array}{ccc}
P & 0 & 0 \\
0 & \Pi' & 0 \\
0 & 0 & \hat{\Pi} \end{array} \right).
$$

(2.40)

with $\Pi$ corresponding to “good” vectors, $\Pi'$ to “semi-good” ones, and $\hat{\Pi}$ to “non-good”, i.e.

$$
\Pi := \beta^{-2} (Q_{12} \otimes Q_{21}) \otimes Q_{11} \otimes Q_{22}, \quad \Pi' = I_2 \otimes Q_{11} \otimes Q_{22},
$$

$$
Q_{\mu\nu} := Q(c_{\mu\nu}), \quad Q(c) = \left( \begin{array}{cc} \beta - c & 1 \\
-\beta c & \beta \end{array} \right), \quad c_{\mu\nu} := 1 + a_\mu^{-1} b_\nu^{-1}, \quad \mu, \nu = 1, 2.
$$

(2.41)

Matrices $X$, $Y$ correspond to the multiplication by

$$
X = - \frac{1}{2} \text{Tr} \left( D_{a'}^{-1} \rho' D_{b'}^{-1} \tau' \right)^2
$$

$$
Y = \frac{1}{8} \left( \text{Tr} \left( D_{a'}^{-1} \rho' D_{b'}^{-1} \tau' \right)^2 \right)^2 - \frac{1}{3} \text{Tr} \left( D_{a'}^{-1} \rho' D_{b'}^{-1} \tau' \right)^3 + O(W^{-1})
$$

(2.42)
with $D_{\alpha'}$, $D_{\beta'}$ of (2.15).

Let us make the change of variables

$$
a_{1i} = a_+(1 + i\theta_0 \tilde{a}_{1i}/\sqrt{W}), \quad a_{2i} = a_-(1 + i\theta_- \tilde{a}_{2i}/\sqrt{W}),
$$

$$
b_{1i} = a_+(1 + \theta_0 \tilde{b}_{1i}/\sqrt{W}), \quad b_{2i} = a_-(1 + \theta_- \tilde{b}_{2i}/\sqrt{W}),
$$

(2.43)

where $a_\pm$ is defined in (1.12), and $\theta_\pm$ will be chosen later (see (9.1)). Notice that this change of variables replaces the factor $W^4$ in front of the first integral in the r.h.s. of (2.35) to $W^2$.

Then we get

$$
\Pi = C_\ast \Pi_0 (1 + O(W^{-1/2} \log W)), \quad C_\ast = \lambda_0^+ \lambda_0^-
$$

$$
\Pi' = C_\ast \Pi'_0 (1 + O(W^{-1/2} \log W)),
$$

$$
\widetilde{\Pi} = C_\ast \widetilde{\Pi}_0 (1 + O(W^{-1/2} \log W)),
$$

(2.44)

where

$$
\Pi_0 = C_\ast^{-1} \Pi \Big| _{\tilde{a}_1 = \tilde{a}_2 = \tilde{b}_1 = \tilde{b}_2 = 0}, \quad \Pi'_0 = C_\ast^{-1} \Pi' \Big| _{\tilde{a}_1 = \tilde{a}_2 = \tilde{b}_1 = \tilde{b}_2 = 0}, \quad \widetilde{\Pi}_0 = C_\ast^{-1} \widetilde{\Pi} \Big| _{\tilde{a}_1 = \tilde{a}_2 = \tilde{b}_1 = \tilde{b}_2 = 0}.
$$

In addition, notice that for $X_0 = X|_{\tilde{a}_1 = \tilde{a}_2 = \tilde{b}_1 = \tilde{b}_2 = 0}$ we have

$$
X_0 = -(a_+^{-2} n'_{11} - a_-^{-2} n'_{22}) (n'_{21} - n'_{12}) + (\rho'_{11} \tau'_{12} \rho'_{22} \tau'_{21} + \rho'_{12} \tau'_{22} \rho'_{21} \tau'_{11}).
$$

(2.45)

Evidently

$$
\Pi_0 = (Q_0 \otimes Q_0 \otimes Q_+ \otimes Q_-), \quad Q_0 := \beta^{-1} Q(0), \quad Q_\pm := Q(c_\pm)/\lambda_0^\pm.
$$

(2.46)

where

$$
c_\pm = 1 + a_\pm^{-2},
$$

(2.47)

and $\lambda_0^+ := \lambda_0(Q(c_+))$ and $\lambda_0^- := \lambda_0(Q(c_-))$ are the biggest eigenvalues of $Q(c_\pm)$. It is easy to see that eigenvalues of $Q(c_\pm)$ are solutions of the equations

$$
\lambda^2 - (2\beta - c_\pm)\lambda + \beta^2 = 0
$$

$$
\Rightarrow \lambda_0^\pm = \beta - c_\pm/2 - \sqrt{c_\pm^2/4 - \beta c_\pm}, \quad |\lambda_0^\pm| > \beta.
$$

(2.48)

Let us study the structure of $\Pi_0$. One can see that the vectors $\{e_0^\pm, e_1^\pm\}$

$$
e_0^+ = e^{e_0^- n_{11}}, \quad e_1^+ = e^{e_1^- n_{11}},
$$

$$
e_0^- = e^{e_0^- n_{22}}, \quad e_1^- = e^{e_1^- n_{22}},
$$

$$
c_0^\pm = c_\pm/2 + \sqrt{c_\pm^2/4 - \beta c_\pm}, \quad c_1^\pm = c_\pm/2 - \sqrt{c_\pm^2/4 - \beta c_\pm}
$$

(2.49)

are eigenvectors for $Q_\pm$, such that $\lambda_0^\pm = 1$, $\lambda_1^\pm < 1$. Evidently they make a basis in which

$$
Q_\pm = \text{diag}\{1, \lambda_1^\pm\}.
$$
Thus
\[
\Pi_0 = Q_0 \otimes Q_0 \otimes \left( P_0 \oplus (\lambda_1 P_{10}) \oplus (\lambda_1 - P_{01}) \oplus (\lambda_1 + \lambda_1 - P_{11}) \right),
\]  
(2.50)
where \( P_{\alpha \beta} \) means the projection corresponding to the vector \( e^+_{\alpha} \otimes e^-_{\beta} \) \((\alpha, \beta = 0, 1)\) and we set \( P_0 = P_{00} \). More precisely, if we consider any polynomial \( p(n_{11}, n_{22}) \), then
\[
P_0 p = (p, s^+ \otimes s^-) e^+_{0} \otimes e^-_{0} \quad \text{with} \quad s^\pm = \left( c^\pm_0 \right) \left( c^\pm_0 - c^\pm_1 \right)^{-1},
\]  
(2.51)
Other projectors of (2.50) can be defined similarly.

The important information about the eigenvalues of \( \tilde{\Pi}_0 \) and \( \Pi'_0 \) is given by

**Lemma 2.2.** Given \( \tilde{\Pi}_0 \) and \( \Pi'_0 \) of (2.44), we have
\[
\lambda_0(\tilde{\Pi}_0) < 1 - \delta \quad (\delta > 0), \quad \lambda_0(\Pi'_0) = 1,
\]  
(2.52)
and there are only two eigenvectors of \( \Pi'_0 \) which correspond to 1: \( \eta_1 e^+_{0} e^-_{0} \) and \( \eta_2 e^+_{0} e^-_{0} \) with \( n_1 \) and \( n_2 \) of (3.38).

**Proof.** Any “non-good” product can be represented as an “absolutely non-good” part \( \tilde{n} \) and a “good” part \( p(n_{11}, n_{22}, n_{12}, n_{21}) \). Here “absolutely non-good” are the products which do not contain any \( n_{\alpha \beta} \). It is easy to see that the exponential part of \( \tilde{Q} \) (see (2.39)) transforms \( \tilde{n} \to \beta^m \tilde{n} \), where \( m \geq 2 \) is the degree \( \tilde{n} \). Since a “good” part cannot be \( e^+_{0} e^-_{0} \) (we exclude these two vectors by the condition of the lemma), the corresponding eigenvalue is less than \( \lambda_0^+ \lambda_0^- \). Thus, after multiplication by \( \beta^{-2} C_*^{-1} \) it becomes less than 1. \( \square \)

### 3. Sketch of the Proof of Theorem 1.3

Notice that below we assume that \( \alpha_1, \alpha_2 \) of (1.15) are real and \( \alpha_1 > \varepsilon / 2, \alpha_2 > \varepsilon / 2 \), since it suffices to prove Theorem 1.3 only for \( \xi \) such that
\[
\Re \xi_1 = \Re \xi_2, \quad \Re \xi'_1 = \Re \xi'_2, \quad \xi_1, \xi_2, \xi'_1, \xi'_2 \in \Omega_{ce},
\]
\[
\Omega_{ce} = \{ \xi : \Re \xi > -c\varepsilon \}, \quad \text{with some} \quad 0 < c < 1.
\]  
(3.1)
Indeed, assume that \( \{ R^+_{Wn}(E, \varepsilon, \xi) \} \) are uniformly bounded in \( n, W \) for \( \xi_1, \xi_2, \xi'_1, \xi'_2 \in \Omega_{ce} \). Consider \( \{ R^+_{Wn}(E, \varepsilon, \xi) \} \) as functions on \( \xi_1 \) with fixed \( \xi_2, \xi'_1, \xi'_2 \) such that \( \Re \xi'_1 = \Re \xi'_2 \). From the uniqueness theorem of complex analysis we know that if the sequence of analytic in some domain \( \Omega \) and uniformly bounded functions \( f_n \) is convergent \( (f_n(z) \to f(z)) \) for \( z \in \Omega' \subset \Omega \) and \( \Omega' \) has at least one accumulation point in \( \Omega \), then \( f_n(z) \to f(z) \) uniformly in each compact set of \( \Omega \). Since these functions \( \{ R^+_{Wn}(E, \varepsilon, \xi) \} \) are analytic in \( \Omega_{ce} \), the above argument yields that (1.14) on the segment \( \Re \xi_1 = \Re \xi_2 \) implies (1.14) for any \( \xi_1, \xi_2 \in \Omega_{ce} \), hence for any \( \xi_1, \xi_2 \in \Omega_{ce} \). Then, fixing any \( \xi_1, \xi_2, \xi'_1, \xi'_2 \), we can consider \( \{ R^+_{nB}(E, \varepsilon, \xi) \} \) as a sequence of analytic functions on \( \xi'_1 \). Since, by the above argument, (1.14) is valid on the segment \( \Re \xi'_1 = \Re \xi'_2 \), the same argument yields that (1.14) is valid for any \( \xi'_1, \xi'_2 \).
To check that $\mathcal{R}_{W_n}^{\pm}(E, \varepsilon, \xi)$ is uniformly bounded in $n$, $W$ for $\xi_1, \xi_2, \xi'_1, \xi'_2 \in \Omega_{\varepsilon \xi}$, we apply the Cauchy-Schwarz inequality to $\mathcal{R}_{W_n}^{\pm}(E, \varepsilon, \xi)$ in the form (1.11). Then we get

$$|\mathcal{R}_{W_n}^{\pm}(E, \varepsilon, \xi)|^2 \leq |\mathcal{R}_{W_n}^{\pm}(E, \varepsilon, \xi_1)| |\mathcal{R}_{W_n}^{\pm}(E, \varepsilon, \xi_2)|$$

where $\xi_1 = (\xi_1, \xi_1', \xi'_1, \xi'_1')$, $\xi_2 = (\xi_2, \xi_2', \xi'_2, \xi'_2')$. Since $\xi_1, \xi_2$ satisfy (3.1), the uniform boundedness of the r.h.s. follows from the uniform convergence (in $\xi$ satisfying (3.1)) of (1.14).

It will be shown below that if $\bar{a}, \bar{b} \in \Omega_{\varepsilon \xi}$, then the matrix $\hat{C}$ of (2.25) becomes diagonal to the leading order (see Lemma 6.3). Moreover, it will be shown also (see Proposition 9.1) that the main part of $A_{ab}$ is the product of operators whose kernel can be obtained from $A_a$ and $A_b$ by leaving only quadratic terms in $\Lambda(x, y)$ (see (2.19)). Hence, according to the analysis of leading order of $\hat{\varphi}$ in $\Omega_{\varepsilon \xi}$ performed in Sect. 2.4, to the leading order

$$\mathcal{K}_{\pm} \sim A_+^* \otimes A_+^* \otimes A_-^* \otimes A_-^* \otimes (\Pi_0 \otimes \Pi_0' \otimes \tilde{\Pi}_0),$$

(3.2)

where $A_+^*$ is defined in (9.2), and $\Pi_0$, $\Pi_0'$, $\tilde{\Pi}_0$ are defined in (2.44). By Proposition 9.1 $\lambda_0(\hat{A}_+^*) = \lambda_0(\hat{A}_-^*) = 1$, and all other eigenvalues are less than 1. Hence, it is naturally to expect that, as $n \to \infty$, only the projection onto the corresponding eigenvector $A_{\pm}^*$ will give non-zero contribution. Similarly, by our preliminary analysis of $\Pi_0$ in Sect. 2.4, the largest eigenvalue of its main part is 1 and the corresponding root subspace is

$$\mathcal{L} = \mathcal{L}^{(0)} \otimes (e_0^+ \otimes e_-^0), \quad \mathcal{L}^{(0)} = \text{Lin}\{1, n_{12}, n_{21}, n_{21}n_{12}, \eta_1, \eta_2\},$$

(3.3)

(see (2.38), (2.49) and (2.51)), and the part of $\Pi_0$, corresponding to this eigenvalue, is $((\mathcal{Q}_0 \otimes \mathcal{Q}_0) \otimes I_2) \otimes P_0$ (it is a matrix of rank 6). However, we cannot consider only the main order of the operator $\mathcal{K}_{\pm}$ because of the factor $W^4$ in front of the integral in the expression for $I_{\pm}$ in (2.35). The integration over “space variables” (see (2.43)) “kills” $W^2$, but we still have the factor $W^2$ there. It means that in the expansion of the coefficients of the operator $\mathcal{K}_{\pm}$ we need to keep the terms up to the order $W^{-2}$ which is a rather involved problem. This is one of the main technical difficulties in the proof of Theorem 1.3.

The main idea is to prove that we can replace $\mathcal{G}_{\pm}(z)$ of (2.36) by the resolvent of some $6 \times 6$ matrix (corresponding to the space $\mathcal{L}$ of (3.3)) with coefficients which do not depend on “space variables” and depend only on $U$, $U'$ and $S$, $S'$. From the analysis of the sigma-model approximation (see [23]) we know that the main part of this matrix should be the $4 \times 4$ block which, after some transformation, takes the form

$$M = K_F \begin{pmatrix} 1 & F_1 + \tilde{K}_1 & F_2 + \tilde{K}_2 & \tilde{K}_5 \\ 0 & 1 & 0 & F_2 + \tilde{K}_3 \\ 0 & 0 & 1 & F_1 + \tilde{K}_4 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$  

(3.4)

Here and below

$$K_F = F_0 K_0 F_0, \quad K_0 = K_{0U} K_{0S}, \quad F_0 = F_{0U} F_{0S} = e^{-\pi \rho(E)(a_1 |U_{12}|^2 + a_2 |S_{12}|^2)/n},$$

(3.5)

where $K_{0U}$, $K_{0S}$ are the “difference operators” defined by (2.18) for $t_a = t_b = t_s := \beta(a_+ - a_-)^2$, and $F_{0U}$, $F_{0S}$ are “multiplication operators” defined in (2.21) with $\bar{a} = \bar{b} = \bar{c} = \bar{d} = \bar{e} = \bar{f}$. 

Universality for 1d Random Band Matrices 683
\( \tilde{b} = (a_+, a_-) \). It is important for us that for \( \tilde{\xi}, \tilde{\xi}' \in \Omega_{\epsilon, \kappa} \) (see (3.1)) we have \( \alpha_1, \alpha_2 > 0 \) (see (1.15)), and hence the multiplication operator \( F_0 \) satisfies the bound

\[
0 \leq F_0 \leq 1. \tag{3.6}
\]

Norms of operators \( \tilde{K}_\alpha \) \( (\alpha = 1, \ldots 5) \) are not small, but \( \tilde{K}_\alpha \) in a certain sense can be bounded by \( 1 - K_0 \). Then, repeating the argument of [23], one can obtain that for smooth enough \( f \rightarrow K_0 f \rightarrow f \) as \( W \rightarrow \infty \), hence in (3.4) we can replace \( K_F \) by \( F_0^2 \) and \( \tilde{K}_\alpha \) by \( 0 \). It was shown in [23], that the replacement of \( \mathcal{G}_\pm(z) \) with \( (M - z)^{-1} \) allows to prove Theorem 1.3.

Unfortunately, this idea cannot be realized exactly in the form described above, since \( \mathcal{G}_\pm(z) \) can be replaced by \( (M - z)^{-1} \) only in a small neighbourhood of the point \( z = 1 \). For other \( |z| \geq 1 + 1/n \) we prove that \( \mathcal{G}_\pm(z) \) can be replaced by \( (\mathcal{M}(z) - z)^{-1} \), where the operator matrix \( \mathcal{M}(z) \) has a more complicated form (see Proposition 4.1) than \( M \) of (3.4). One can see that \( \mathcal{M} \) (the upper left \( 4 \times 4 \) block of \( \mathcal{M}(z) \)) contains additional terms \( u(K_F/z) \) and \( u_0(K_F/z) \) with some analytic functions \( u(\zeta) \) and \( u_0(\zeta) \) which become zero at \( \zeta = 1 \). One can hope that these terms are small as \( z \sim 1 \), since we know that \( K_F \sim 1 \) if it is applied to a sufficiently smooth function. The remainder operators \( \tilde{K}_\alpha \) are also more complicated than in (3.4), but they are still small if applied to sufficiently smooth functions. Section 6 is devoted to the derivation of Theorem 1.3 from Proposition 4 modulo two lemmas whose proofs are given in Sect. 5.

Hence, the main technical tool for the proof of Theorem 1.3 is Proposition 4.1. The proof of Proposition 4.1 is given in Sect. 6. It is based on the application of Proposition 2.4 to the operator \( K \). First we consider \( K_\pm \) as a block matrix in appropriate basis. Denote

\[
\Psi_k(\tilde{a}, \tilde{b}) = \psi_+^k(\tilde{a}_1)\psi_-^k(\tilde{a}_2)\psi_+^k(\tilde{b}_1)\psi_-^k(\tilde{b}_2), \quad \tilde{k} = (k_1, k_2, k_3, k_4). \tag{3.7}
\]

where \( \psi_+^k(\tilde{a}), \psi_-^k(\tilde{b}) \) are vectors diagonalizing the quadratic approximation of \( A_{ab} \) (see (9.3)). Set

\[
\langle K_\pm \rangle_{kk'} := \int \mathcal{K}_\pm \Psi_k(\tilde{a}', \tilde{b}')\Psi_k(\tilde{a}, \tilde{b})d\tilde{a}'d\tilde{b}'d\tilde{a}d\tilde{b}. \tag{3.8}
\]

Then the block operator \( K_\pm \) becomes a semi-infinite block matrix whose blocks \( \langle K_\pm \rangle_{kk'} \) are \( 70 \times 70 \) matrices. The entries of each block are integral operators in \( L_2(U_2) \otimes L_2(U(1, 1)) \). Now we consider the decomposition of the Grassmann space \( \Omega_{70} \) into 3 subspaces

\[
\Omega_{70} = \mathcal{L} + (\mathcal{L}_1 + \mathcal{L}_2 + \mathcal{L}_3) + \mathcal{L}' \tag{3.9}
\]

where \( \mathcal{L} \) is defined in (3.3), subspaces \( \mathcal{L}_1, \mathcal{L}_2, \mathcal{L}_3 \) are defined similarly to (3.3) but with \( e_0^1e_0^- \) replaced by \( e_0^1e_1^- \), \( e_0^1e_1^- \), and \( e_0^1e_1^- \) respectively, and \( \mathcal{L}' \) is a space of "non-good" vectors. Consider each \( 70 \times 70 \) matrix \( \langle K \rangle_{kk'} \) as a block matrix corresponding to the decomposition (3.9) with blocks \( \langle K \rangle_{kk'}^{(\mu \nu)} (\mu, \nu = 1, 2, 3) \), and use Proposition 2.4 with \( H^{(11)} = \langle K \rangle_{00}^{(11)} \). The matrix \( \mathcal{M}' \) of (4.2) is obtained as

\[
\mathcal{M}' = T \left( H^{(11)} - H^{(12)}(H^{(22)} - z)^{-1}H^{(21)} \right) T, \tag{3.10}
\]

where the transformation matrix \( T \) has the form

\[
T = \left( \begin{array}{ccc} T_0 & 0 & 0 \\ 0 & T_0 & I_2 \\ I_2 & 0 & 0 \end{array} \right), \quad T_0 = \left( \begin{array}{cc} 0 & W^{1/2} \\ W^{-1/2} & 0 \end{array} \right), \quad T_0^2 = I_2, \quad T^2 = I_6. \tag{3.10}
\]
4. Proposition 4.1 and Derivation of Theorem 1.3 from It

Let \( P_\Sigma \) be the projection onto the subspace (3.3) such that transforms all “non-good” vectors into 0 and for \( x \otimes p(n_{11}, n_{22}) \) with \( x \in \mathcal{L}^{(0)} \)

\[
P_\Sigma(x \otimes p) = x \otimes P_0 p, \tag{4.1}
\]

with \( P_0 \) of (2.51).

We shall use also the following definition.

**Definition 4.1.** Let \( p(U) \) (or \( p(S) \)) be products of matrix entries of \( U \) (or \( S \)). We say that the operator \((p)_U \) (or \((p)_S \)) (see (2.1) for the definition of \((p)_U \)) is of the type \( m \), if the number of non-diagonal entries of \( U \) (or \( S \)) in \( p \) is \( m \). We say that the operator \((\prod_{i=1}^p P_{0r_i} F_i F_{0} B_i)\) is of the joint type \( m \), if the operators \( r_i \) (of the form \((p)_U \)) and \( r'_i \) (of the form \((p)_S \)) are of the type \( m_i \), \( m'_i \), respectively, \( m_1 + m_1' + \cdots + m_p + m'_p = m \), and \( B_i = \varphi_i(F_0 K_i F_0) \), where \( K_i \) are of the type 0, \( F_0 \) is defined in (3.5), and \( \varphi_i \) are analytic functions in \( \mathbb{B}_{1+\delta} = \{ z : |z| \leq 1 + \delta \} \). We denote \( O_{\ast}(v_1^m) \) the linear combinations of the operators of the joint type at least \( m \).

Notice also that by (2.16) and (2.25) we obtain that the non-diagonal entries of \( \hat{C} \) are linear combinations of operators of non-zero type.

**Proposition 4.1.** For the resolvent \( G_\pm \) of (2.36) there exist matrices \( \mathcal{M}(z) \), \( \mathcal{M}(z) \) such that for \( T \) of (3.10) and \( P_\Sigma \) of (4.1), we have

\[
P_\Sigma G_\pm(z) P_\Sigma = T(\mathcal{M}(z) - z)^{-1}T, \quad \mathcal{M}(z) = \mathcal{M}(z) + O(W^{-1}), \quad \mathcal{M}(z) := \left( \begin{array}{cc} \mathcal{M} & \mathcal{M}' \\ \mathcal{M}'' & \mathcal{M}''' \end{array} \right) \tag{4.2}
\]

where

\[
\mathcal{M} = \begin{pmatrix} K_F & u_1(z) + K_F F_1 + \tilde{K}_1 & -u_2(z) + K_F F_2 + \tilde{K}_2 & u_0(z) + \tilde{K}_5 \\ 0 & K_F & 0 & -u_2(z) + K_F F_2 + \tilde{K}_3 \\ 0 & 0 & K_F & u_0(z) + \tilde{K}_5 \\ 0 & 0 & 0 & K_F \end{pmatrix}, 
\tag{4.3}
\]

\[
\mathcal{M}' = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad \mathcal{M}'' = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad \mathcal{M}''' = \begin{pmatrix} F_0 K_1 F_0 & 0 \\ 0 & F_0 K_1 F_0 \end{pmatrix}. 
\tag{4.4}
\]

Here \( K_F \) and \( F_0 \) are defined in (3.5) and \( K_1, K_2 \) are any operators of the type 0. Operators \( F_1 \) and \( F_2 \) have the form

\[
F_1 = n^{-1} \left( y_1(K_F/z)(2\sigma_1 + \alpha_1 \varphi_U) + y_2(K_F/z)(2\sigma_2 - \alpha_2 \varphi_S) \right), \\
F_2 = n^{-1} \left( y_2(K_F/z)(2\sigma_1 - \alpha_1 \varphi_U) + y_1(K_F/z)(2\sigma_2 + \alpha_2 \varphi_S) \right), 
\tag{4.5}
\]

with \( \varphi_U, \varphi_S, \alpha_1, \alpha_2, \sigma_1, \sigma_2 \) defined in (2.21). Functions \( u \) and \( u_0 \) have zero at \( \xi = 1 \)

\[
u(\xi) = (\xi - 1) u_1(\xi), \quad u_0(\xi) = (\xi - 1) u_2(\xi), \tag{4.6}
\]
and \( u_1(\xi), u_2(\xi), v_1(\xi), v_2(\xi) \) are analytic in \( B_{1+\delta} = \{ \xi : |\xi| < 1 + \delta, \delta > 0 \} \). Moreover,

\[
\tilde{K}_i = W O_*(v^{(1)}) O_* (v^{(1)}) \quad (i \neq 5), \quad \tilde{K}_5 = W O_*(v^{(1)}) O_* (v^{(1)}) + W^2 O_*(v^{(\mu)}) O_*(v^{(\nu)}),
\]

where \( \mu, \nu > 0 \). In addition,

\[
I_{\pm} = -\frac{C_{\varepsilon, \delta}}{2 \pi i} \int z^{n-1} ((M' - z)^{-1} \hat{f} - \hat{g}) dz + o(1), \quad \hat{f} = \hat{f}_0 + \hat{f}, \quad \hat{g} = \hat{g}_0 + \hat{g},
\]

\[
\hat{f}_0 = (f_{01}, f_{02}, -f_{02}, d_{1*}, 0, 0) F_0, \quad \hat{g}_0 = (d_{2*}, g_{02}, -g_{02}, g_{04}, 0, 0) F_0,
\]

\[
\hat{f} = W O_*(v^{(1)}) O_* (v^{(1)}) \varphi_1 (K_F) F_0 f' + O(W^{-1}) F_0 f'', \quad \hat{g} = W O_*(v^{(1)}) O_* (v^{(1)}) \varphi_2 (K_F) F_0 g' + O(W^{-1}) F_0 g'' \]

where \( f_{0i}, g_{0i} \) are analytic in \( B_{1+\delta} \). Functions of \( z^{-1} K_F, f_{0i} F_0 \) is a result of the application of \( f_{0i} (K_F / z) \) to \( F_0 \). \( O(W^{-1}) F_0 \) is a result of the application of the operator \( O(W^{-1}) \) norm to \( F_0 \), and \( O_*(v^{(2)}) \varphi_{1,2} (K_F) F_0 \) is a result of the application of the operator \( O_*(v^{(2)}) \varphi_{1,2} (K_F) \) to \( F_0 \), and \( f', g', f'', g'' \in \mathbb{C}^6 \).

We postpone the proof of Proposition 4.1 to Sect. 6. Now we prove Theorem 1.3 on the basis of Proposition 4.1. We need also two following lemmas, whose proofs can be found in Sect. 5.

**Lemma 4.1.** Given \( M', M \) of (4.2), there is \( C > 0 \) which does not depend on \( n, W \), such that for any \( |z| \geq 1 + 1/n \) we have

\[
\| (M' - z)^{-1} \| \leq C/|z - 1|, \quad \| G(z) \| \leq C/|z - 1|, \quad (G(z) = (M - z)^{-1}).
\]

**Remark 4.1.** Lemma 4.1 combined with Propositions 4.1 and 2.4 implies that the spectrum of \( K_{\pm} \) belongs to the circle \( |z| \leq 1 + 1/n \).

Denote by \( M_0 \) an upper triangular matrix which is obtained from \( M \) of (4.3) by replacing all \( \tilde{K}_i \) with zeros and \( K_F \) with \( F_0^2 \). Set

\[
M_0 = \begin{pmatrix} M_0 & 0 \\ 0 & M'' \\ \end{pmatrix} \Rightarrow G_0(z) := (M_0 - z)^{-1} = \begin{pmatrix} G_0(z) & 0 \\ 0 & (M'' - z)^{-1} \\ \end{pmatrix},
\]

where \( G_0(z) := (M_0 - z)^{-1} \).

Introduce also

\[
L = \log^2 n, \quad P_L = P_L U \otimes P_L S,
\]

where \( P_L U \) is an orthogonal projection of \( L_2 (\hat{U}_2) \) onto the subspace \( \oplus_{l < L} L_{2(U^1, 1)} \) and \( P_L S \) is a similar orthogonal projection in \( L_2 (\hat{U}(1, 1)) \) (see (9.8)).

**Lemma 4.2.** Let \( M \) and \( M_0 \) be defined by (4.2) and of (4.10), and \( E_0 \) be projection in \( L_2 = L_2 (\hat{U}_2) \otimes L_2 (U(1, 1)) \) onto the subspace of function on \( |U|_2^2 \) and \( |S|_2^2 \). Then

\[
\|(M_0 - M) P_L E_0 \| \leq L^2 / W, \quad \|(M_0 - M)^{(i)} P_L E_0 \| \leq L^2 / W
\]
and

\[ \| (1 - \mathcal{P}_L) \hat{f}_0 \| \leq e^{-c \log^{4/3} n}, \quad \| (1 - \mathcal{P}_L) \mathcal{G}_0 \hat{f}_0 \| \leq e^{-c \log^{4/3} n}, \quad \| \mathcal{G}_0 \hat{f}_0 \| \leq n/|z - 1|, \]

\[ \| \hat{f}_0 \| \leq C n, \quad \| \hat{\mathcal{G}} \| \leq C n L^4 / W^2 + O(e^{-c \log^{4/3} n}). \]  \hspace{1cm} (4.13)  \hspace{1cm} (4.14)

where \( \hat{f}_0, \hat{f} \) are defined in (4.8). Similar bounds hold for \( \hat{g}_0, \hat{g} \) in place of \( f_0, f \).

**Proof of Theorem 1.3.** By Lemma 4.1 we conclude that we can choose \( \omega_0 \) of (2.28) as an integration contour in (2.35). In addition, Lemma 4.1 and (4.14) imply for \( z \in \omega_0 \)

\[ |((\mathcal{M}' - z)^{-1} \hat{f}, \hat{g}) - ((\mathcal{M}' - z)^{-1} \hat{f}_0, \hat{g}_0)| \]

\[ \leq C |z - 1|^{-1} \left( \| \hat{f} \| \| \hat{g}_0 \| + \| \hat{f}_0 \| + \| \hat{f} \| \| \hat{g} \| \right) + O(e^{-c \log^{4/3} n}) \]

\[ \leq L^2 n / W |z - 1| + O(e^{-c \log^{4/3} n}). \]  \hspace{1cm} (4.15)

Then, using that \( |z|^n \leq C \) for \( z \in \omega_0 \),

\[ \int_{\omega_0} \frac{|dz|}{|z - 1|} \leq C \log n, \]  \hspace{1cm} (4.16)

and \( L^2 n \log n / W \to 0 \) (since \( W \geq C n \log^{5} n \) by assumptions of Theorem 1.3), we get from (4.8) and (4.15)

\[ I_{\pm} = -\frac{C \xi_{\pm}}{2 \pi i} \int_{\omega_0} z^{n-1} ((\mathcal{M}' - z)^{-1} \hat{f}_0, \hat{g}_0) dz + o(1). \]  \hspace{1cm} (4.17)

To analyse the r.h.s. we use the resolvent identity

\[ A_1^{-1} - A_2^{-1} = -A_1^{-1} (A_1 - A_2) A_2^{-1} = -A_2^{-1} (A_1 - A_2) A_1^{-1} \]  \hspace{1cm} (4.18)

with \( A_1 = \mathcal{M}' - z, A_2 = \mathcal{M} - z. \) Since \( \mathcal{M}' = \mathcal{M} + O(W^{-1}), \) applying twice (4.18) we obtain

\[ \left| (\mathcal{M}' - z)^{-1} \mathcal{G}_0 \hat{f}_0, \hat{g}_0 \right| \leq \left| \mathcal{G}_0 (O(W^{-1}) + \mathcal{M} - \mathcal{M}_0) \mathcal{G}_0 \hat{f}_0, \hat{g}_0 \right| \]

\[ + \| (\mathcal{M}' - z)^{-1} \cdot \| (O(W^{-1}) + \mathcal{M} - \mathcal{M}_0) \mathcal{G}_0 (z) \hat{f}_0 \| \cdot \| (O(W^{-1}) + \mathcal{M} - \mathcal{M}_0) (z) \mathcal{G}_0 \| \]

\[ + O(e^{-c \log^{4/3} n}) = T_1 + T_2 + O(e^{-c \log^{4/3} n}). \]

Relations (4.12), (4.13), and \( \mathcal{G}_0 (z) \hat{f}_0 \in \mathcal{E}_0 L_2, \) yield for \( z \in \omega_0 \)

\[ \| (O(W^{-1}) + \mathcal{M} - \mathcal{M}_0) \mathcal{G}_0 (z) \hat{f}_0 \| = \| (O(W^{-1}) + \mathcal{M} - \mathcal{M}_0) \mathcal{P}_L \mathcal{E}_0 \mathcal{G}_0 (z) \hat{f}_0 \| + O(e^{-c \log^{4/3} n}) \]

\[ \leq \frac{CL^2}{W} \| \mathcal{G}_0 \hat{f}_0 \| + O(e^{-c \log^{4/3} n}). \]

Hence, using Lemma 4.2 for \( \| (\mathcal{M}' - z)^{-1} \| \) and the last bound of (4.13), we get

\[ T_1 \leq \frac{CL^2}{W} \| \mathcal{G}_0 \hat{f}_0 \| \| \mathcal{G}_0 \hat{g}_0 \| + O(e^{-c \log^{4/3} n}) \leq \frac{CnL^2}{W \cdot |z - 1|} + O(e^{-c \log^{4/3} n}), \]

\[ T_2 \leq C |z - 1|^{-1} \frac{L^4}{W^2} \| \mathcal{G}_0 \hat{f}_0 \| \| \mathcal{G}_0 \hat{g}_0 \| + O(e^{-c \log^{4/3} n}) \leq \frac{Cn^2L^4}{W^2 \cdot |z - 1|} + O(e^{-c \log^{4/3} n}). \]
These bounds and (4.16) give

\[
\left| \int_{\omega_0} z^{n-1} ((\mathbf{M}' - z)^{-1} - \mathbf{G}_0) f_0, g_0) dz \right|
\leq C \left( \frac{L^2 n}{W} + \frac{L^4 n^3}{W^2} \right) \int_{\omega_0} \left| \frac{dz}{z} \right| + O(e^{-c \log^{4/3} n}) \leq C \frac{n L^2 \log n}{W} + O(e^{-c \log^{4/3} n}) = o(1). \tag{4.19}
\]

Thus, in view of (4.17), (4.19), and (4.10), we get

\[
I_{\pm} = -\frac{C_{\xi, \varepsilon}}{2\pi i} \int_{\omega_0} z^{n-1} (\mathbf{G}_0 f_0, g_0) dz + o(1) = -\frac{C_{\xi, \varepsilon}}{2\pi i} \int_{\omega_0} z^{n-1} (\mathbf{G}_0 f_0, g_0) dz + o(1)
\]

Using the representation (2.35) for \( \mathcal{R}_{nW}^{++} \), the inverse matrix formula for \( \mathbf{G}_0(z) \) (see (5.5) below), the form of \( \widehat{f}_0, \widehat{g}_0 \) (see (4.8)), and taking into account the bound (7.3) for \( I_+ \) and similar bound for \( I_- \), we obtain

\[
C_{\xi, \varepsilon}^{-1} \mathcal{R}_{nW}^{++}(E, \varepsilon, \xi) = -\frac{1}{2\pi i} \int_{\omega_0} \left( (G_F(z)(f_{01, d_{2a}} + 2 f_{02, 2g_{02}} + d_{1* g_{04}})F_0^2
\right.
\]

\[
- 2G_\mathbf{F}'(z)uf_0(z)(f_{02, 2g_{02} - d_{1* g_{04}}})F_0^2 - G_\mathbf{F}'(z)(F_1 - F_2)(f_{02, 2g_{02} - d_{1* g_{04}}})F_0^2
\]

\[
- G_\mathbf{F}'(z)(f_{02, 2g_{02} - d_{1* g_{04}}})^2 + 2G_\mathbf{F}'(z)(F_1 + u(F_0^2/z))(F_2 - u(F_0^2/z))d_{1* g_{04}}^2 F_0^2 + o(1). \tag{4.20}
\]

where \( G_F(z) = (F_0^2 - z)^{-1} \).

Taking the integral with respect to \( z \) and using (4.6), we get

\[
\mathcal{R}_{nW}^{++}(E, \varepsilon, \xi) = C_{\xi, \varepsilon} \int \left( k_1 \tilde{F}_1 \tilde{F}_2 + k_2(\tilde{F}_1 - \tilde{F}_2) + k_3 \right) F_0^{2n} dU dS + o(1),
\]

where

\[
\tilde{F}_1 = y_1(1)(2\sigma_1 + \alpha_1 \varphi_U) + y_2(1)(2\sigma_2 - \alpha_2 \varphi_S),
\]

\[
\tilde{F}_2 = y_2(1)(2\sigma_1 - \alpha_1 \varphi_U) + y_1(1)(2\sigma_2 + \alpha_2 \varphi_S),
\]

\[
F_0^{2n} = e^{i E(\sigma_1 - \sigma_2)} e^{\pi \rho(E) \sigma_1 \varphi_U - \pi \rho(E) \sigma_2 \varphi_S},
\]

and \( k_1, k_2, k_3 \) are some constants which we find using the fact that for \( \alpha_1 = \alpha_2 = \alpha \) and \( \sigma_1 = \sigma_2 = \sigma \) the above expression is 1 (see the definition of \( \mathcal{R}_{nW}^{++}(E, \varepsilon, \xi) \) in (1.11)). It implies immediately that the coefficient of \( \sigma^2 \) is 0, hence

\[
y_1(1) = -y_2(1) = y \Rightarrow \tilde{F}_1 \tilde{F}_2 = y^2(\alpha_1 \varphi_U + \alpha_2 \varphi_S)^2 - 4y^2(\sigma_1 - \sigma_2)^2.
\]

Performing the integration with respect to \( dU, dS \) we obtain

\[
\int dU dS e^{\pi \rho(E) \alpha_1 \varphi_U - \pi \rho(E) \alpha_2 \varphi_S} (\alpha_1 \varphi_U + \alpha_2 \varphi_S)^2
\]

\[
eq e^{-2 \pi \rho(E) \alpha_2} \frac{2}{(\pi \rho(E))^2} \left( \frac{\alpha_1}{2 \alpha_2} + \frac{\alpha_2}{2 \alpha_1} \right) \sinh(2 \pi \rho(E) \alpha_1) + \cosh(2 \pi \rho(E) \alpha_1) + \frac{\sinh(2 \pi \rho(E) \alpha_1)}{4(\pi \rho(E))^2 \alpha_1 \alpha_2}
\]

\[
\int dU dS e^{\pi \rho(E) \alpha_1 \varphi_U - \pi \rho(E) \alpha_2 \varphi_S} = e^{-2 \pi \rho(E) \alpha_2} \sinh(2 \pi \rho(E) \alpha_1) \frac{8(\pi \rho(E))^2 \alpha_1 \alpha_2}{2 \alpha_1 \alpha_2}.
\]
Now putting $\alpha_1 = \alpha_2 = \alpha$ and $\sigma_1 = \sigma_2 = \sigma$, we can conclude that $k_1 y^2 (\pi \rho(E))^{-2} = 1$ and $k_3 = -2$. Hence
\[
\mathcal{R}^{+-}_n(W, \varepsilon, \xi) = e^{iE(\sigma_1 - \sigma_2)} e^{-2\pi \rho(E)\alpha_2} \left( \left( \frac{\alpha_1}{2\alpha_2} + \frac{\alpha_2}{2\alpha_1} \right) \sinh(2\pi \rho(E)\alpha_1) + \cosh(2\pi \rho(E)\alpha_1) \right) - \left( k_2 y (\sigma_1 - \sigma_2) + (\sigma_1 - \sigma_2)^2 \right) \frac{\sinh(2\pi \rho(E)\alpha_1)}{2\alpha_1 \alpha_2}.
\]
Taking the derivative with respect to $\xi'$ and then putting $\bar{\xi} = \bar{\xi}' = 0$, we get
\[
(N \rho(E))^{-1} \mathbb{E}\{ \text{Tr} (H_N - E - i\varepsilon/N)^{-1} \} = \left. \frac{\partial}{\partial \xi'_1} \mathcal{R}^{+-}_n(W, \varepsilon, \xi) \right|_{\xi = \bar{\xi}' = 0} = \frac{ia_+}{\rho(E)} + k_2 y \frac{1 - e^{-4\pi \rho(E)\varepsilon}}{8\varepsilon^2 \rho(E)i} + o(1).
\]
But it follows from Theorem 1.2 that
\[
\mathbb{E}\{ N^{-1} \text{Tr} (H_N - E - i\varepsilon/N)^{-1} \} = \frac{ia_+}{\rho(E)} + o(1).
\]
Therefore we conclude that $k_2 = 0$ and (1.14) holds. As a corollary, we obtain (1.16). \hfill \Box

5. Proof of Lemmas 4.1 and 4.2

Proof of Lemmas 4.1. To obtain the bound for $\|G\|$, we start from the analysis of
\[
\mathcal{M}_1 = \mathcal{M} - \mathcal{M}'(\mathcal{M}'' - z)^{-1} \mathcal{M}'', \quad \widetilde{G}_1 := (\mathcal{M}_1 - z)^{-1}
\]
and prove that
\[
\|\widetilde{G}_1\| \leq C/|z - 1|, \quad |z| > 1 + 1/n, \quad (5.1)
\]
or, equivalently,
\[
\| (\widetilde{G}_1)_{ij} \| \leq C/|z - 1|.
\]
Since by (4.7) $\widetilde{K}_1 = W F_0 r_1 B r_2 F_0$, where $r_1$ and $r_2$ are of the type 1, at least, we have by (6.6) with $K_\alpha = K_0$
\[
| (\widetilde{K}_1 f, g) |^2 \leq C W^2 (F_0 r_2)^* r_2 F_0 f, f) (F_0 r_1 B B^* r_1^* F_0 g, g) \leq C (F_0 (1 + C/W - K_0) F_0 f, f) (F_0 (1 + C/W - K_0) F_0 g, g). \quad (5.2)
\]
Let us check that the operators which come from $\mathcal{M}'(\mathcal{M}'' - z)^{-1} \mathcal{M}''$ satisfy (5.2). Since $\widetilde{K}_1 = W F_0 r_1 B r_2 F_0$, we have
\[
| ( (\mathcal{M}'(\mathcal{M}'' - z)^{-1} \mathcal{M}'')_{ij} f, g ) |^2 \leq W^2 (F_0 r_1 B r_2 F_0 G_{\sigma^*} F_0 r_2^* B^* r_1^* F_0 g, g) \cdot W^2 (F_0 (r_2)^* (r_1)^* F_0 G_{\sigma^*} F_0 (r_1)' B' (r_2)' F_0 f, f), \quad (5.3)
\]
$G_{\sigma^*} := (1 + 1/n - F_0 K_\sigma F_0)^{-1}$, $\sigma = 1, 2$.\hfill \Box
Assume that \( r_2 \) contains \( U_1 \) or \( U_2 \). Then by (6.6)
\[
\| Br_2 F_0 G_{\sigma^*} F_0 r_2^* B^* \| \leq \| B \|^2 \| G_{\sigma^*}^{1/2} F_0 r_2^* F_0 G_{\sigma^*}^{1/2} \|
\]
\[
\leq C W^{-1} \| G_{\sigma^*}^{1/2} F_0 (1 + C / W - K_{\sigma}) F_0 G_{\sigma^*}^{1/2} \|
\]
\[
\leq C W^{-1} \| G_{\sigma^*}^{1/2} (1 + C / W - F_0 K_{\sigma} F_0) G_{\sigma^*}^{1/2} \| \leq C' W^{-1}. \tag{5.4}
\]

Here we used that \( W \gg n \).

Hence
\[
W^2 (F_0 r_1 Br_2 F_0 G_{\sigma^*} F_0 r_2^* B^* r_1^* F_0 g, g) \leq C W (F_0 r_1 F_0 g, g) \leq C'' (F_0 (1 + C / W - K_0) F_0 g, g)
\]

This bound combined with a similar bound for \((F_0 (r_1')^* (B')^* (r_1')^* F_0 G_{\sigma^*} F_0 (r_1') B' (r_2') F_0 f, f)\) yields (5.2).

It is easy to see that if some operator \( \hat{A} \) has the form similar to (4.3) (i.e., its diagonal entries are the same (equal to \( A \)) and the other non-zero entries are only in the first line and the last column) then the resolvent \( \hat{G} = (\hat{A} - z)^{-1} \) has the same form and
\[
\hat{G}_{12} = -G A_{12} G, \quad \hat{G}_{13} = -G A_{13} G, \quad \hat{G}_{24} = -G A_{24} G, \quad \hat{G}_{34} = -G A_{34} G,
\]
\[
\hat{G}_{14} = -G A_{14} G + G A_{12} G A_{24} G + G A_{13} G A_{34} G, \quad \text{where} \ G := (A - z)^{-1} = \hat{G}_{ii}. \tag{5.5}
\]

Applying (5.5) to \( M_1 \), we conclude that it is sufficient to prove the corresponding bounds for the operators
\[
G, \ G \tilde{K}_a G, \ G \tilde{K}_a G \tilde{K}_\beta G, \ G K_F F_v G, \ G K_F F_v G K_F F_v G, \ G \tilde{K}_a G K_F F_v G, \ G K_F F_{\mu} G \tilde{K}_a G, \ G u_0 \left( \frac{K_F}{z} \right) G, \ G u \left( \frac{K_F}{z} \right) G, \ G u \left( \frac{K_F}{z} \right) G, \ G u \left( \frac{K_F}{z} \right) G \tilde{K}_a G, \ G u \left( \frac{K_F}{z} \right) G F_{\alpha} G. \tag{5.6}
\]

with \( F_v \) (\( v = 1, 2 \)) of (4.5) and
\[
G(z) := (K_F - z)^{-1}. \tag{5.7}
\]

Observe that by (4.6) \( \| G u (K_F / z) \| \leq C \) and \( \| G u_0 (K_F / z) \| \leq C \), hence the bounds for the operators of the second line of (5.6) follow from those of the first line.

By (3.5)–(3.6) \( 0 \leq K_F \leq 1 \) hence the spectrum of \( K_F \) is contained in \([0, 1]\). Thus, it is evident that for \(|z| > 1 + 1/n\)
\[
\|G(z)\| \leq \text{dist}[z, [0, 1]] \leq C / |z - 1|. \tag{5.8}
\]

To estimate other entries, we set
\[
G_* := G(z) \big|_{|z| = 1 + 1/n}, \tag{5.9}
\]

and prove the bounds
\[
\| G_{\sigma^*}^{1/2} \tilde{K}_a G_{\sigma^*}^{1/2} \| \leq C, \quad \| G_{\sigma^*}^{-1/2} G_{\sigma^*}^{1/2} \| \leq C, \quad \| G_{\sigma^*}^{1/2} K_{F} F_v G_{\sigma^*}^{1/2} \| \leq C,
\]
\[
\| G_{\sigma^*}^{1/2} (\varphi / n) K_{F} \tilde{x}(K_F / z) G_{\sigma^*}^{1/2} \| \leq C \quad \| G_{\sigma^*}^{1/2} (\varphi / n) K_{F} \tilde{y}(K_F / z) G_{\sigma^*}^{1/2} \| \leq C, \tag{5.10}
\]

where \( \tilde{x}(\zeta) \) and \( \tilde{y}(\zeta) \) are some analytic in \( B_{1+\delta} \) functions. Notice also that \( G \) and \( G_* \) commute.
It is easy to see that operators from (5.6) can be represented as a sum of $G^{1/2} \prod G^{1/2}$, where $\prod$ is some product of the operators from (5.10).

The definition of the operator norm, (5.2), and the bound $F \leq 1$ yield

$$\|G^{1/2} \tilde{K}_\alpha G^{1/2}\| = \sup_{\|f\|\leq 1} |(\tilde{K}_\alpha G^{1/2} f, G^{1/2} g)|^2$$

$$\leq \sup_{\|f\|\leq 1} |(F_0(1 + C/W - K_0) F_0 G^{1/2} f, G^{1/2} f)| \sup_{\|g\|\leq 1} |(F_0(1 + C/W - K_0) F_0 G^{1/2} g, G^{1/2} g)|$$

$$= \|G^{1/2} F_0(1 + C/W - K_0) F_0 G^{1/2}\|^2 \leq \|G^{1/2} F_0(1 + C/W - K_0) F_0 G^{1/2}\|^2 \leq 1.$$  (5.11)

Moreover, since $G$ and $G_*$ commute, we have

$$\|G^{1/2}(z)G^{-1/2}\| = \|G(z)G^{-1}\| \leq \max_{|z| \geq 1, n_0 \leq \lambda \leq 1} \frac{1 + 1/n - \lambda}{|z - \lambda|} \leq C,$$

which gives the second inequality of (5.10).

To prove the inequalities of the second line of (5.10), observe that since $G_*$ and $\tilde{K}(F/z)$ commute, we have

$$\|G^{1/2}(\varphi_U/n)K_\alpha \tilde{K}(F/z)G^{1/2}\| \leq C\|G^{1/2}(\varphi_U/n)K_\alpha G^{1/2}\|$$

$$\leq C\|G^{1/2}(\varphi_U/n)F_0 G^{1/2}\| + C\|G^{1/2}(\varphi_U/n) F_0(1 - K_0) F_0 G^{1/2}\|$$

$$\leq C\|G^{1/2}(\varphi_U/n)F_0 G^{1/2}\| + C\|G^{1/2}(\varphi_U/n)^2 F_0^2 G^{1/2}\| \leq \|G^{1/2} F_0(1 - K_0) F_0 G^{1/2}\|^2.$$

Moreover, since

$$(\varphi_U/n) F_0^2 \leq C(1 + 1/n - F_0^2), \quad (1 - K_0) \leq 1 - K_0),$$

using the last line of (5.11) and the bound

$$\|G^{1/2}(\varphi_U/n) F_0 G^{1/2}\| \leq C\|G^{1/2} F_0(1 + 1/n - F_0^2) G^{1/2}\| \leq C\|G^{1/2} F_0(1 + 1/n - K_0) F_0 G^{1/2}\| = C,$$

we obtain the third bound of (5.10). The last bound (5.10) can be proved similarly. Thus, we get (5.1).

Now let us prove (4.9) for $\|\tilde{G}_1 M_+ (M^{'''} - z)^{-1}\|$. The structure of $\tilde{G}_1$ (see (5.5), (5.6)), and the bounds (5.10) combined with the structure of $M_+$ and $M^{'''}$ (see (4.7)) imply that it suffices to find the bound for

$$W\|G^{1/2} F_0 r_1 B r_2 F_0 G^{1/2}\| \leq ||B||\|G^{1/2} F_0 r_1 ||r_2 F_0 G^{1/2}\|$$

with $r_1, r_2$ of non-zero type (see (5.9) and (5.3) for the definition of $G_*$ and $G_{\alpha\sigma}$). But in view of (4.7)

$$\|G^{1/2} F_0 r_1\|^2 \leq \|G^{1/2} F_0 r_1 F_0 G^{1/2}\| \leq C W^{-1}\|G^{1/2} F_0(1 + C/W - K_0) F_0 G^{1/2}\| \leq CW^{-1}$$

The same bound for $\|r_2 F_0 G^{1/2}\|^2$ was obtained in (5.4). Hence $\|\tilde{G}_1 M_+ (M^{'''} - z)^{-1}\|$ satisfies (4.9).

The bounds for $\|(M^{'''} - z)^{-1} M'' \tilde{G}_1\|$ and $\|(M^{'''} - z)^{-1} M'' \tilde{G}_1 M_+ (M^{'''} - z)^{-1}\|$ can be obtained similarly. Then, using Proposition 2.4, we obtain that the block matrix $\mathfrak{m}$ satisfies the bound (4.9).

Since by (4.2) $M_+ = \mathfrak{m} + O(W^{-1})$, we have

$$(\mathfrak{m} - z)^{-1} = (1 + (\mathfrak{m} - z)^{-1} O(W^{-1})^{-1} (\mathfrak{m} - z)^{-1} = (\mathfrak{m} - z)^{-1} (1 + O(n/W)).$$

$\square$
Proof of Lemma 4.2. We start from the proof of (4.12) with operators of a bit more general form than (4.7), since these operators appear in (4.8). Consider

\[ \tilde{K} = W \tilde{K}', \quad \tilde{K}' = F_0(q_1)_U(p_1)_SF_0\varphi_1(F_0K_mF_0)F_0(q_2)_U(p_2)_SF_0\varphi_2(K_F)F_0, \]

where \( p_{1,2} \) are of the type 1 and \( q_{1,2} \) are of the type 0, \( \varphi_{1,2}(\zeta) \) are analytic in \( \mathbb{B}_{1+\delta} \) function and \( K_m \) is some operator of the type 0 acting in \( E_m(\mathcal{L}_2) \). Here and below \( E_m \) is an orthogonal projection on the space of the functions

\[ \phi(S_{12}^2, |U_{12}|^2)S_{11}^{m_1-k_1}S_{12}^{k_1}U_{11}^{m_2-k_2}U_{12}^{k_2}, \quad \tilde{m} = (m_1, m_2), \quad k_1, k_2 \in \mathbb{Z}, \]

and, if some of the above exponents are negative, then we replace the respective matrix entry with its conjugate. Remark that the operators \( q_1)_U(p_1)_S, \tilde{K}_m, (q_2)_U(p_2)_S, K_0 \) should feet to each other (otherwise their product is zero) hence one could insert \( E_m \) after \( \varphi_1 \) and \( E_0 \) after \( \varphi_2 \).

Notice also that the cases when \( p_1 \) and \( p_2 \) both depend on \( U \) (or one of them depends on \( U \), and another one depends on \( S \)), as well as the cases when the joint type of \( \tilde{K}' \) is more than 2, are also possible, but their analysis is similar. For \( W^2 O_s(\psi^{(4)}) \) the proof is similar as well.

Assume that we proved the following relations

\[ F_0(\varphi(F_0K_mF_0) - \varphi(F_0^2))F_0E_m\mathcal{P}_L = O(L^2W^{-1}), \quad \|(K_m - 1)E_m\mathcal{P}_L\| \leq CL^2/W, \tag{5.12} \]

\[ [p_2, \varphi(F_0)] = \varphi'(F_0)D_S + O(W^{-3/2}), \quad D_S = \Phi_1(S_1)(p_2p_*)_S + \Phi_1(S_1)(p_2p_*)_S, \tag{5.13} \]

\[ \|[(p^{(s)})_S, \varphi(F_0)]\| \leq CW^{-(1+s)/2}, \quad \|[(p^{(s)})_S, \Phi_1]\| \leq CW^{-(s+1)/2}, \tag{5.14} \]

\[ \|\varphi_1(F_0K_mF_0), \varphi(F_0)\| \leq C/\sqrt{W}, \quad \|\varphi_1(F_0K_mF_0), \Phi_1\| \leq C/\sqrt{W}, \tag{5.15} \]

\[ \|p_{1,2}\|_S \leq C/\sqrt{W}, \quad \|p_{1,2}\|_S \mathcal{P}_L \leq CL/W, \quad \|D_S\| \leq CW^{-1}, \tag{5.16} \]

where \( \varphi \) is any analytic in \( \mathbb{B}_{1+\delta} \) function, \( p^{(s)} \) is any product of the type \( s \), \( \Phi_1 \) is the operator of multiplication by \( \Phi_1(S) = cn^{-1}S_{11}S_{12}F_0 \) (see (2.21)) and \( p_*(S) = S_{11}S_{12} \).

Then, using the first inequality of (5.12), the first bound of (5.14), and the first bound of (5.16), we get

\[ \tilde{K}'\mathcal{P}_L = B_1(p_1)_S F_0\varphi_1(F_0K_mF_0)F_0(q_2)_U(p_2)_SF_0^2\varphi_2(F_0^2)\mathcal{P}_L + O(L^2W^{-2}), \]

where \( B_1 \) is a bounded operator to the left of \( (p_1)_S \) in \( \tilde{K}' \). Then (5.13) with \( \varphi_2(\zeta) = \zeta^2\varphi_2(\zeta^2) \) combined with the first bound of (5.16), (5.12) for \( (q_2)_U \), and (5.16) yield

\[ \tilde{K}'\mathcal{P}_L = B_1(p_1)_SF_0\varphi_1(F_0K_mF_0)F_0(q_2)_U\left( \varphi_2(F_0)(p_2)_S + \varphi_2(F_0)D_S \right)\mathcal{P}_L + O(L^2W^{-2}) \]

\[ = B_1(p_1)_SF_0\varphi_1(F_0K_mF_0)F_0\left( \varphi_2(F_0)(p_2)_S + \varphi_2(F_0)D_S \right)(q_2)_U\mathcal{P}_L + O(L^2W^{-2}). \tag{5.17} \]

Using consequently the first bound of (5.15) for \( \varphi(\zeta) = \varphi_2(\zeta) \) combined with the first bound of (5.16) and the fact that \( (p_2)_S(q_2)_U \) commutes with \( \mathcal{P}_L \), then the first relation of
(5.12), then the first bound of (5.14) with \( \varphi(\zeta) = \zeta^2 \varphi_1(\zeta^2) \tilde{\varphi_2}(\zeta) \), and finally the second bound of (5.16), we obtain for the term in the r.h.s. of (5.17) which contains \((p_2)_S\):

\[
B_1 F_0(p_1)_S F_0 \varphi_1(F_0K_m F_0) F_0 \tilde{\varphi_2}(F_0)(p_2)_S \varphi(q_2) U P_L
\]

\[
= B_1 F_0(p_1)_S \tilde{\varphi_2}(F_0) F_0 \varphi_1(F_0K_m F_0) F_0 P_L(q_2) U (p_2)_S P_L + O(L^2 W^{-2})
\]

\[
= B_1 F_0(p_1)_S F_0^2 \varphi_2(F_0^2) \tilde{\varphi_2}(F_0) P_L(q_2) U (p_2)_S P_L + O(L^2 W^{-2})
\]

\[
= B_1 F_0^3 \varphi_1(F_0^2) \tilde{\varphi_2}(F_0)(p_1)_S \varphi(q_2) U (p_2)_S P_L + O(L^2 W^{-2}) = O(L^2 W^{-2}).
\]

Similarly for the first term of \( D_S \) (see (5.15)) we get

\[
B_1(p_1)_S F_0 \varphi_1(F_0K_m F_0) F_0 \tilde{\varphi_2}(F_0) \Phi_1(p_2 p_*) (q_2)_U P_L
\]

\[
= B_1 F_0^2 \tilde{\varphi_2}(F_0) F_0 \varphi_1(F_0^2) \Phi_1(p_1)_S (p_1)_S (p_2 p_*) (q_2)_U P_L + O(L^2 W^{-2}) = O(L^2 W^{-2}).
\]

The second term of \( D_S \) can be analysed similarly.

Thus we are left to prove relations (5.12)–(5.16). Remark that (5.16) and the first bound of (5.12) are direct corollaries of Lemma 6.2, where we need only to take into account that \((p_2 p_*)_S\) is also reduced by \(L^{(d)}S\) and it is of the second type, hence its norm is bounded by \(O(W^{-1})\).

To prove the first relation of (5.12), we use first the Cauchy formula and (4.18):\[
\varphi(F_0K_0 F_0) - \varphi(F_0^2) = \frac{1}{2 \pi i} \oint_{|z|=1+\delta} \varphi(z) G_1(z) F_0(K_0 - 1) F_0 G_0(z) dz
\]

\[
= \frac{1}{2 \pi i} \oint_{|z|=1+\delta} \varphi(z) G_1(z) F_0 \left( F_0 G_0(z)(K_0 - 1) + [K_0, F_0 G_0(z)] \right)
\]

\[
G_1 = (F_0K_0F_0 - z)^{-1}, \quad G_0 = (F_0^2 - z)^{-1}, \quad (5.18)
\]

To estimate the above commutator, we are going to expand

\[
F_0 G_0(z) = -z^{-1} \sum_{s=0}^{\infty} z^{-s} F_0^{2s+1}.
\]

Notice that for any \( p > 0 \) \([K_m, F_0^p]\) is an integral operator with the kernel

\[
(F_0^p(S_1) - F_0^p(S_2)) \tilde{\varphi} (S) K_0 S (S) K_0 U,
\]

where \( \tilde{S} = S_1 S_2^{-1} \) and \( K_m \sim (\tilde{p})_S \). Now use the formula

\[
e^{-x} - e^{-y} = -(x - y) e^{-x} - (x - y)^2 \int_0^1 t e^{-tx - (1-t)y} dt \quad (5.19)
\]

for \( x = c(2s + 1)n^{-1}|(S_1)_{12}|^2 \) and \( y = c(2s + 1)n^{-1}|(S_2)_{12}|^2 \) (recall that \( c = \alpha_2(a_+ - a_-) \) by (2.21)). Then

\[
y - x = (2s + 1)cn^{-1} \left(|\tilde{S}_{12}|^2 |(S_1)_{12}|^2 + 1 + (\tilde{S}_{11}^{-1} + 1)|S_{12}(S_1)_{12}(S_1)_{12}(S_1)_{12} + \bar{S}_{11}^{-1} + 1)||S_{12}(S_1)_{12} + \tilde{S}_{12}(S_1)_{12}(S_1)_{12}|^2) \right)
\]

\[
= (2s + 1)cn^{-1} (\bar{p}_s(S_1)p_s(S_1 + p_s(S_1) \bar{p}_s(S_1)) + O((2s + 1)n^{-1}|\tilde{S}_{12}|^2). \quad (5.20)
\]

\( \tilde{\varphi} (\tilde{S}) K_0 (\tilde{U}, \tilde{S}) z^{-s-1} \), summing with respect to \( s \), and using the second bound of (5.16) for \( \tilde{p} \tilde{p}_s \), we obtain the first relation of (5.12). The second relation of (5.12), as well as (5.16), are direct corollaries of Lemma 6.2.
The proof of (5.13) is very similar. We expand $\varphi(F_0)$ into the Taylor series $\sum \varphi_r F_0^r$. The commutator $\left( (p^2)_S, F_0^s \right)$ is an integral operator with the kernel $(F_0^s(S_1) - F_0^s(S_2)) K_0(\widetilde{S}, \widetilde{U}) p_2(\widetilde{S})$. Hence, using (5.19) and (5.20) as previously, we obtain (after multiplication by $K_0(\widetilde{S}) p_2(\widetilde{S})$) that the remainder term will give us $O(W^{-3/2})$ by (5.16). Then, multiplying the relation by $\varphi_s$ and summing with respect to $s$, we get (5.13). The proof of (5.14) is very similar (and even simpler).

For the proof of (5.15) we use again the Cauchy formula and write

$$
\left[ \varphi(\hat{F}_0) - \varphi(0) \right] = \frac{1}{(2\pi i)^2} \int_{|a|=1+\delta} \varphi(z_1)\varphi(z_2) [G(z_1), G(0)] dz_1 dz_2,
$$

where $G_1, G_0$ are defined in (5.18) and $G_{01}$ can be obtained from $G_0$, if we replace $F_0^2$ by $F_0$. The last relation here follows from the first bound of (5.14) with $s = 0$ and bounds for $\|G_1\|$ and $\|G_{01}\|$. The second bond of (5.15) can be obtained similarly.

To prove the point for $(1 - \mathcal{P}_L) \hat{G}_0 \hat{f}$ with $\hat{G}_0$ of (4.10), it suffices to prove similar bounds for $(1 - \mathcal{P}_L S) \hat{G}_0 \hat{f}$ and $(1 - \mathcal{P}_L U) \hat{G}_0 \hat{f}$. From the structure of $\hat{G}_0$ (see (5.5)) we conclude that we need to prove corresponding bounds for

$$
(1 - \mathcal{P}_L S) G_0 f, \quad (1 - \mathcal{P}_L S) G_0 F_f, \quad (1 - \mathcal{P}_L S) G_0 F_f, \quad (1 - \mathcal{P}_L S) G_0 F_f,
$$

with some $c_{1,2}$. By [29] we have for any $\varphi(2|S_1|)$

$$
\| (1 - \mathcal{P}_L S) \varphi \|_S^2 = \int_{\rho > L} \left| a(\rho) \right|^2 \rho \tanh(\pi \rho) d\rho, \quad a(\rho) = \int_0^\infty \varphi(x) \mathcal{P}_L^{1/2 - i\rho} (x - 1) dx
$$

where $\mathcal{P}_L^{1/2}$ is the Legendre function (see the proof of Lemma 6.2 for the definition). Hence,

$$
\| (1 - \mathcal{P}_L S) \varphi \|_S^2 \leq \left( L^2 + \frac{1}{4} \right)^{-2m} \| \Delta_S^m \varphi \|_S^2 = \left( L^2 + \frac{1}{4} \right)^{-2m} \left( \Delta_S^m \varphi, \varphi \right).
$$

\[\square\]

**Proposition 5.1.** For any smooth enough function $\varphi$

$$
|\Delta_S^m \varphi(x)| \leq 2^{2m} (m!)^2 \sum_{k=0}^{2m} (x + 1)^k |\varphi^{(k)}(x)|.
$$

Proposition 5.1 implies

$$
(\Delta_S^m G_0 f, G_0 f) \leq 2^{4m} ((2m)!)^2 \sum_{k=0}^{2m} I_k,
$$

$$
I_k := \int_0^\infty dx (x + 1)^k \left| \frac{\partial^k}{\partial x^k} \frac{e^{-c_1 x/n}}{e^{-c_1 x/n} - z} \right| | e^{-c_1 x/n} - z |.
$$
Expanding \((e^{-c_1 x/n} - z)^{-1}\) into the series with respect to \(e^{-j c_1 x/n}\) and taking the derivative, we get

\[
I_k \leq C n^2 \sum_{j=2}^{\infty} \left| z \right|^{-j} j^k \int_{0}^{\infty} (\tilde{x} + c_1/n)^k e^{-j \tilde{x}} d\tilde{x}
\]

\[
\leq C n^2 \sum_{j=2}^{\infty} \left| z \right|^{-j} j^k \int_{0}^{\infty} \tilde{x}^k e^{-j (\tilde{x} - c_1/n)} d\tilde{x} \leq C n^2 k! \sum_{j=2}^{\infty} j^{-1} |e^{c_1/n} / z| \leq C (2m)! n^2 \log n.
\]

Here we used the change of variable \(\tilde{x} = c_1 x / n\) and take \(|z| = 1 + 1/n\) with sufficiently big \(A\). Thus,

\[
(\Delta_S^{2m} G_0 f, G_0 f) \leq n^2 \log n \cdot n (2m((2m)!)^3 2^{4m} (L^2 + 1/4)^{-2m})
\]

\[
\Rightarrow \|(1 - \mathcal{P}_{LS}) G_0 f \| \leq 2 n^2 (\log n m \exp(4m \log 2 + 4m(\log(2m)^{3/2} - \log L - 1))) \leq e^{-c \log^{4/3} n},
\]

if we take \(m = \tilde{c} L^{2/3} = \tilde{c} \log^{4/3} n\) with sufficiently small \(\tilde{c}\). The bounds for \((1 - \mathcal{P}_{LS}) f\) and other functions from (5.21) and similar bounds with \((1 - \mathcal{P}_{LU}) f\) can be obtained similarly. For the last bound in (4.13) we need to estimate \(\|G_0 f\|, \|G_0^3 F_\alpha f\|, \|G_0^3 F_\alpha F_\beta f\|\).

Using the change of variables \(x = n \tilde{x}\), we get, e.g., for \(\|G_0^3 F_\alpha F_\beta f\|:\)

\[
\|G_0^3 F_\alpha F_\beta f\| \leq n \left( \int_{0}^{1} + \int_{1}^{\infty} \right) \left( \tilde{x} + C/n \right)^4 e^{-c_1 \tilde{x}} \frac{d\tilde{x}}{|z - e^{-c_1 \tilde{x}}|} \leq C n \left( \int_{0}^{1} d\tilde{x} + \frac{e^{-c_1 \tilde{x}}}{|z - e^{-c_1 \tilde{x}}|^2} \right) + C n \leq \frac{nC}{|z - 1|}.
\]

The bounds (4.14) follow from the bound for \(\widetilde{K} \mathcal{P}_L\) obtained above and (4.13), since

\[
\|\widetilde{K} F_0\| \leq \|\widetilde{K} \mathcal{P}_L F_0\| + \|\widetilde{K} (1 - \mathcal{P}_L) F_0\| \leq C \left( L^2 / W \right) \|F_0\| + O(e^{-c \log^{4/3} n}).
\]

\[\square\]

6. Proof of Proposition 4.1

In what follows it will be convenient for us to consider decomposition of the Grassmann space \(\Omega_{70}\) into “good”, “semi-good” and “non-good” subspaces (see (2.39)–(2.40)) and write \(\tilde{Q}\) and \(\tilde{C}\) (see (2.25)) as \(3 \times 3\) block matrices corresponding to this decomposition.

Lemma 6.1. The matrix \(\tilde{C}\) (see (2.25)) in the block representation (2.40) has the form

\[
\begin{align*}
\tilde{C} &= \tilde{C}_d + \tilde{C}, \\
\tilde{C} &= \begin{pmatrix}
O(W^{-1}) & O_*(\mathbf{t}(1))O_*(\mathbf{t}(1)) & O_*(\mathbf{t}(1)) \\
O_*(\mathbf{t}(1))O_*(\mathbf{t}(1)) & O(W^{-1}) & O_*(\mathbf{t}(1)) \\
O_*(\mathbf{t}(1)) & O_*(\mathbf{t}(1)) & O(W^{-1})
\end{pmatrix},
\end{align*}
\]

(6.1)

where \(\tilde{C}_d\) is a diagonal matrix whose entries are \(K_{US}\) in the first block and some operators of the type zero (see Definition (4.1)) for other blocks (their norm is not more than \(1 + C / W\) by Lemma 6.2 below), and \(O_*(\mathbf{t}(1))\) is defined in (4.1).

Moreover, for \(\tilde{C}_0 = \tilde{C}_{\mid_{t_a = t_b = t_*}}\) we have

\[
\begin{align*}
\tilde{C}_0^{(11)}(e_0^+ e_0^-) &= e_0^+ e_0^- - (n_* W)^{-1} (n_{12} - n_{21}) (e_0^- e_0^+ - e_0^- e_0^+) + O(W^{-2}), \\
\frac{\partial^2 \tilde{C}_0^{(11)}(n_{12} e_0^+ e_0^-)}{\partial n_{12} \partial n_{21}} &= (n_* W)^{-1} (e_0^+ e_0^- - c_0^- e_0^+) + O(W^{-2}) = -\frac{\partial^2 \tilde{C}_0^{(11)}(n_{21} e_0^+ e_0^-)}{\partial n_{12} \partial n_{21}} + O(W^{-2}).
\end{align*}
\]

(6.2)
Definition 6.1. We write $M = O_s(C)$ for some $70 \times 70$ matrix $M$, if the entries of $M$ in the block representation as in (2.40) satisfy the same bounds as the entries of the corresponding block of $C$ (see (6.1)).

The proof is based on the following lemma proven in Sect. 9.3.

Lemma 6.2. Operators of the form $(v)_U$ and $(v)_S$ are reduced by the $l$-th subspace of irreducible representation of the shift operator on $U(2)$ and $U(1, 1)$ respectively. If $p(U)$ is some product of matrix entries of $U$ of non-zero type, then we have for the reduced operator $(p)_U^{(l)}$

$$
(p)_U^{(l)} = \lambda_{ij}^{(l)}(p) \xi_{ij}^{(l)U}, \quad \lambda_{ij}^{(l)}(p) = q_p^{1/2}((l)(Wt_a)^{-s(p)}(1 + O(l^2/W)),
$$

where a partial isometric operator $\xi_{ij}^{(l)U}$ (see (9.12) for the definition), exponent $s(p) \geq 1$, and polynomial $q_p$ do not depend on $t$, $W$ and depend only on $p(U)$. If $p$ is of type, then the corresponding $\xi_{ii}^{(l)U}$ is an orthogonal projection and the corresponding $q_p(l)$ is quadratic

$$
\lambda_{ij}^{(l)}(p) = 1 - q_p(l)(Wt_a)^{-1}(1 + O(l^2/W)),
$$

$$
K_U^{(l)} = \xi_{00}^{(l)U} - l(l + 1)(Wt_a)^{-1} + O(l^4W^{-2}), \quad \frac{\partial K_U^{(l)}}{\partial t_a} = W((p)_{U}^{(l)})^*(p)_{U}^{(l)}(1 + O(l^2W^{-1})),
$$

where $p(U) = U_1U_2$. The same is true for $p(S)$. In particular, for any fixed $s = 0, 1, 2, \ldots$

$$
(U_12^{2s})_U^{(l)} = \xi_{00}^{(l)U}(s!(Wt_a)^{-s} + O(l^2/W^{s+1})),
$$

$$
(S_22^{2s})_S^{(l)} = \xi_{00}^{(l)S}(s!(Wt_a)^{-s} + O(l^2/W^{s+1})).
$$

In addition, if $r = B(p)_U$ or $r = B(p)_S$ with $p$ of the type at least 1 and with some bounded operator $B$, then for any $K_\alpha$ corresponding to the product of the type 0

$$
r^*r \leq CW^{-1}(1 + C/W - K_\alpha).
$$

Proof of Lemma 6.1. The result of the application of $\hat{C}$ to the product of the Grassmann variables has the form

$$
\prod_{j=1}^s \rho_{\alpha_j\beta_j} \tau_{\beta_j'} \alpha_j' \rightarrow \prod (U \rho S^{-1})_{\alpha_j\beta_j} (S \tau U^*_s)_{\beta_j' \alpha_j'}
$$

$$
= \prod_{j=1}^s \left( \sum_{\mu_j, v_j = 1}^\infty \sum_{\mu_j', v_j' = 1}^\infty U_{\alpha_j \mu_j} U^*_s_{\mu_j' \alpha_j'} (S^{-1})_{\beta_j \beta_j} S_{\beta_j' \beta_j} \rho_{\mu_j, v_j} \tau_{v_j' \mu_j'} \right).
$$

According to Lemma 6.2, the bound for the operator which is the product of matrix entries of $U$ depends on the number of $U_{12}$ or $U_{21}$ in this product and the same for the product of matrix entries of $S$. In other words, the bound for the entry of $\hat{C}$ corresponding to the transformation $\prod \rho_{\alpha_j\beta_j} \tau_{\beta_j'} \alpha_j'$ to $\prod \rho_{\mu_j, v_j} \tau_{v_j' \mu_j'}$ depends on the number of index changes ($1 \rightarrow 2$ or $2 \rightarrow 1$) which we need to transform

$$
\cup(\{(\alpha_j, \beta_j)\}) \rightarrow \cup(\{\mu_j, v_j\}) \quad \text{and} \quad \cup(\{\beta_j', \alpha_j'\}) \rightarrow \cup(\{v_j', \mu_j\}.
$$
For all non-diagonal entries of \( \hat{C} \) we have at least one transformation of indexes. In view of Lemma 6.2, this means at least \( O_*(\ell^{(1)}) \) operator in the corresponding entry.

If we transform “good product” into another “good” one, then we need at least two transformations and the resulting product of the entries of \( U \) and \( S \) can be written as a function of \( |U_{12}|^2 \) and \( |S_{12}|^2 \). Thus Lemma 6.2 yields \( O(W^{-1}) \) for the non-diagonal entries inside the block \( \hat{C}^{(1)} \), and the diagonal entries from this block have the form \( 1 + O(W^{-1}) \).

To obtain a “semi-good” vector from a “good” one, we need at least two transformations with at least 1 non-diagonal \( U_{ab} \) and at least 1 non-diagonal \( S_{ab} \), so corresponding operators will have the form at least \( O_*(\ell^{(1)})O_*(\ell^{(1)}) \).

Relations (6.2) are simple corollary from the equalities:

\[
\hat{C}_0^{(1)}(e_0^+e_0^-) = \hat{C}_0^{(1)}(e_0^+e_0^-) + O(W^{-2}),
\]

\[
\hat{C}_0^{(1)}(n_{ab}^+e_0^-) = \hat{C}_0^{(1)}(n_{ab}^+e_0^-) + O(W^{-2}),
\]

\[
\hat{C}_0^{(1)}(n_{ab}) = n_{ab} + O(W^{-1}), \quad \hat{C}_0^{(1)}(e_0^+e_0^-) = e_0^+ - (t_2W)^{-1}c_0^+(n_{12} - n_{21}) + O(W^{-2}).
\]

Here, to obtain the last equality, we take into account that

\[
(U_{12}U_{21}^{-1})_U \sim (t_2W)^{-1}, \quad (S_{12}S_{21}^{-1})_S \sim -(t_2W)^{-1}.
\]

Now we consider operator \( K \) as a block matrix of the form (3.8)

**Lemma 6.3.** We have

\[
\langle \hat{F} \hat{Q} A_{ab} \hat{F} \hat{Q} \rangle_{\tilde{k}\tilde{k}'} = \delta_{\tilde{k},\tilde{k}'}(C_dF_0 + C_0F_0 + O(W^{-1})) + (1 - \delta_{\tilde{k},\tilde{k}'})W^{-d(k - \tilde{k})/2}O_*(C),
\]

\[
\langle \hat{F} \hat{Q} A_{ab} \hat{F} \hat{Q} \rangle_{\tilde{k}\tilde{k}'} = \delta_{\tilde{k},\tilde{k}'}(\lambda_{\tilde{k}}F_0\hat{\Pi}_0 + O(W^{-1})) + (1 - \delta_{\tilde{k},\tilde{k}'})O(W^{-d(k - \tilde{k})/2}), \tag{6.7}
\]

where \( A_{ab} \) is defined in (2.20), \( F_0 = \hat{F} \big|_{a_1, a_2 = b_1, b_2 = 0}, C_d = \hat{C}_d \big|_{a_1, a_2 = b_1, b_2 = 0}, C_0 = \hat{C}_0 \big|_{a_1, a_2 = b_1, b_2 = 0}, \hat{\Pi}_0 = \Pi_0 \oplus \Pi'_0 \oplus \hat{\Pi}_0 \) (see (2.44)), and \( \lambda_{\tilde{k}'} = \lambda_{k_1}^{+} \lambda_{k_2}^{-} \lambda_{k_3}^{+} \lambda_{k_4}^{-} \) with \( \lambda_{k}^{\pm} \) defined in (9.5). We set here

\[
d(\tilde{k}) = \begin{cases} 
1, k = \pm \ell_i, \text{ or } k = \pm 3\ell_i, \\
2, |k| - \text{ even}, \\
3, \text{ otherwise},
\end{cases} \tag{6.8}
\]

where \( |k| = |k_1| + |k_2| + |k_3| + |k_4|, \)

\[
\ell_1 := (1, 0, 0, 0), \quad \ell_2 := (0, 1, 0, 0), \quad \ell_3 := (0, 0, 1, 0), \quad \ell_4 := (0, 0, 0, 1). \tag{6.9}
\]

**Proof.** To obtain the assertion for the matrix \( \hat{F} \hat{Q} A_{ab} \), observe that if we expand all integral kernels with respect to \( \tilde{a}, \tilde{b} \), then for entries with \( \tilde{k} - \tilde{k}' = \pm \ell_i \) zero order terms disappear after integration, and the first order terms give \( O(W^{-1/2}) \). The entries with \( \tilde{k} - \tilde{k}' = \pm 3\ell_i \) are also \( O(W^{-1/2}) \), since we have \( \tilde{a}_i^3W^{-1/2} \) and \( \tilde{b}_i^3W^{-1/2} \) in \( A^\pm \) and \( A_b^\pm \) (see (9.2)). The entries with even \( \tilde{k} - \tilde{k}' \) can obtain non-zero contribution only from the terms which are of even order (thus at least quadratic) with respect to \( \tilde{a}, \tilde{b} \), hence they are at least \( O(W^{-1}) \). And for all other entries, the non-zero contribution can be
obtained only from the terms which are of the odd order 3 or more, and which cannot contain only $\tilde{a}_i^3$ or $\tilde{b}_i^3$.

The assertion of the lemma for the matrix $\hat{C}$ follows from Lemma 6.2. Indeed, according to the lemma, the eigenvectors (or generalized eigenvectors) of the entries of $\hat{C}$ do not depend on $\tilde{a}$, $\tilde{b}$, $O(1)$ terms of eigenvalues also do not depend on $\tilde{a}$, $\tilde{b}$, and the first terms in the asymptotic expansion depends on $\tilde{a}$, $\tilde{b}$ through the coefficient $(Wt)^{-k}$ only. Hence the expansion of $t_{a}^{-1}$ or $t_{b}^{-1}$ with respect to $\tilde{a}$ and $\tilde{b}$ will add additional $W^{-1/2}$ in each order. □

The lemma implies that in the same basis the operator $\mathcal{K} = \hat{F}^\dagger Q A_{ab} \hat{C} \hat{F}$ has the form

$$\langle \mathcal{K} \rangle_{\tilde{k}\tilde{k}'} = \lambda_{\tilde{k}} (\Pi_0^{(F)} + F_0 \hat{\Pi}_0 C_0 F_0) + O(W^{-1}),$$

$$\langle \mathcal{K} \rangle_{\tilde{k}\tilde{k}'} = \langle \hat{F}^\dagger \Pi A_{ab} \hat{C} \hat{d} \hat{F} \rangle_{\tilde{k}\tilde{k}'} + W^{-d(\tilde{k}-\tilde{k}')/2} O_*(C) + O(W^{-3/2})$$

$$= W^{-d(\tilde{k}-\tilde{k}')/2} (O(1) + O_*(C)) + O(W^{-3/2}), \quad \tilde{k} \neq \tilde{k}',$$

$$\Pi_0^{(F)} := F_0 \hat{\Pi}_0 C_d F_0.$$  \hspace{1cm} (6.10)

Consider now the decomposition (3.9) of the Grassmann space. Let $\langle \mathcal{K} \rangle_{\tilde{k}\tilde{k}}^{(11)}$ correspond to $\mathcal{L}$ and $\langle \mathcal{K} \rangle_{\tilde{k}\tilde{k}'}^{(33)}$ correspond to $\mathcal{L}'$. Then we use Proposition 2.4 with $H^{(11)} = \langle \mathcal{K} \rangle_{00}^{(11)}$. Set $B(z) = (H^{(22)} - z)^{-1}$. Now $B(z)$ is also a block matrix with blocks $B_{\tilde{k}\tilde{k}}$ (they are quadratic $70 \times 70$ matrices for $\tilde{k} \neq 0, \tilde{k}' \neq 0$ or rectangular matrices if $\tilde{k} = 0, \tilde{k}' \neq 0$ or $\tilde{k} = 0, \tilde{k}' = 0$). It follows from (6.10) that

$$\|B_{\tilde{k}\tilde{k}}\| \leq O(1) W^{-d(\tilde{k}-\tilde{k}')/2}. \hspace{1cm} (6.11)$$

Indeed, denoting $G_{\tilde{k}\tilde{k}} = (\langle \mathcal{K} \rangle_{\tilde{k}\tilde{k}} - z)^{-1}$ (remark that $G_{00} = (\langle \mathcal{K} \rangle_{00}^{(1)} - z)^{-1}$ where $\langle \mathcal{K} \rangle_{00}^{(1)}$ is $2 \times 2$ block matrix constructed from $\langle \mathcal{K} \rangle_{00}^{(\mu \nu)}$ with $\mu, \nu = 2, 3$), we have by (6.10)

$$|\lambda_{0}(\langle \mathcal{K} \rangle_{\tilde{k}\tilde{k}})| \leq 1 - \delta, \quad \|\langle \mathcal{K} \rangle_{\tilde{k}\tilde{k}}\| \leq CW^{-1/2}, \quad \tilde{k} \neq \tilde{k}'.$$

$$B_{\tilde{k}\tilde{k}'} = \delta_{\tilde{k}\tilde{k}'} G_{\tilde{k}\tilde{k}} + \sum G_{\tilde{k}\tilde{k}} \langle \mathcal{K} \rangle_{\tilde{k}\tilde{k}} G_{\tilde{k}\tilde{k}'} + \sum G_{\tilde{k}\tilde{k}} \langle \mathcal{K} \rangle_{\tilde{k}\tilde{k}'} G_{\tilde{k}\tilde{k}} G_{\tilde{k}\tilde{k}'} + O(W^{-3/2}). \hspace{1cm} (6.12)$$

$$d(\tilde{k} - \tilde{k}_1) + d(\tilde{k}_1 - \tilde{k}') \geq d(\tilde{k} - \tilde{k}').$$

Hence (6.10) implies (6.11). Similarly, note that

$$d(\tilde{k}) + d(\tilde{k} - \tilde{k}') + d(\tilde{k}') \geq 4,$$

for all $\tilde{k}, \tilde{k}'$ except $\tilde{k} = \tilde{k}' = \ell_i$ or $3\ell_i$, $\tilde{k} = \tilde{k}' = 0$, and $\tilde{k} = 0, \tilde{k}' = \ell_i, 3\ell_i$ or $\tilde{k} = 0, \tilde{k}' = \ell_i, 3\ell_i$. Thus we obtain

$$H^{(12)}(H^{(22)} - z)^{-1} H^{(21)} = P_\mathcal{L} \left( \sum \langle \mathcal{K} \rangle_{0\ell} G_{0\ell} \langle \mathcal{K} \rangle_{\ell0} \right) P_\mathcal{L} + \sum \langle \mathcal{K} \rangle_{0\ell}^{(1\mu)} G_{0\ell}^{(1\mu)} \langle \mathcal{K} \rangle_{\ell0}^{(1\mu)}$$

$$+ \sum \langle \mathcal{K} \rangle_{0\ell}^{(1\mu)} B_{0\ell}^{(1\mu)} \langle \mathcal{K} \rangle_{\ell0}^{(1\mu)} + O(W^{-2}),$$
where summations are with respect to $\tilde{\ell} = \ell^*_i, 3\ell^*_i$ and $\mu, \nu = 2, 3$. But denoting $\Sigma_1$ and $\Sigma_2$ the last two sums above, we have
\[
\Sigma_1 + \Sigma_2 = O(W^{-2}) + W^{-1} O_*(\tau^{(1)}) O_*(\tau^{(1)}) \tag{6.13}
\]
(see Definition 4.1). Indeed, consider $\Sigma_1$. By (6.10) and (6.1) entries of $\langle K \rangle^{(1\mu)}_{00}$ can be $O(W^{-1})$, $O_*(\tau^{(2)})$ for $\mu = 2$, or $O_*(\tau^{(1)})$ for $\mu = 3$. In addition, (6.11) implies
\[
\|B^{(\mu\nu)}_{0k}\| \leq CW^{-1/2},
\]
and according to (6.10) $\langle K \rangle^{(v1)}_{k0}$ is $O(W^{-3/2})$ or at least $W^{-1/2} O_*(\tau^{(1)})$. All together this gives (6.13) for $\Sigma_1$. The estimate for $\Sigma_2$ can be obtained by the same argument.

Now take $T$ of (3.10) and set
\[
\mathcal{M}' = T (H^{(11)} - \tilde{H}) T, \quad \tilde{H} := H^{(12)} (H^{(22)} - z)^{-1} H^{(21)}. \tag{6.14}
\]
By the consideration above
\[
\tilde{H} = P_2 \left( \sum \langle K \rangle_{0k} G_{kk} \langle K \rangle_{k0} \right) P_2 + \sum \langle K \rangle^{(1\mu)}_{00} G^{(\mu\nu)}_{00} \langle K \rangle^{(v1)}_{00} + O(W^{-2}) + W^{-1} O_*(\tau^{(2)}).
\tag{6.15}
\]
Denote by $M$ an upper left $4 \times 4$ block of $(H^{(11)} - \tilde{H})$. Notice that, after above transformation with $T$, only the diagonal entries, the entries $M_{i1}$ and $M_{4i}$ ($i = 2, 3$) multiplied by $W$ and $M_{4i}$ multiplied by $W^2$ may stay $O(1)$ or more. It is easy to see that both $\langle K \rangle^{(1\mu)}_{00}$ and $\langle K \rangle^{(v1)}_{00}$ give at most $O_*(\tau^{(1)})$, and so the contribution of the second term of (6.15) to $WM_{i1}$ and to $WM_{4i}$ ($i = 2, 3$) is at most $WO_*(\tau^{(2)})$ which we include in $\tilde{K}_i$ (see (4.7)). Now we are going to compute the contribution of the second term of (6.15) to $W^2 M_{4i}$. To this end we write (see (6.10))
\[
\langle K \rangle^{(1\mu)}_{00} = F_0 \Pi_0^{(1)} C_0^{(1\mu)} F_0 + O(W^{-1}), \quad \langle K \rangle^{(v1)}_{00} = F_0 \Pi_0^{(v\nu)} C_0^{(v1)} F_0 + O(W^{-1}),
\]
and expand $(\langle K \rangle_{00} - z)^{-1}$ with respect to $C_0$:
\[
(\Pi_0^{(F)} - z)^{-1} \sum_m (-1)^m \left( (F_0 \Pi_0 C_0 F_0 + O(W^{-1}))(\Pi_0^{(F)} - z)^{-1} \right)^m. \tag{6.16}
\]
The entry $M_{4i}$ is the result of the application of the matrix $(H^{(11)} - \tilde{H})$ to the vector $e_0^* e_0^*$ projected onto $n_{12} n_{21} e_0^* e_0^*$. But matrices $\Pi_0, F_0$ and $(\Pi_0^{(F)} - z)^{-1}$ do not increase the total number of $\rho_{12}, \rho_{21}, \tau_{12}, \tau_{21}$ in the Grassmann vector to which they are applied. Hence, if the term in (6.16) does not contain $O(W^{-1})$ we need to take the entries which add into the result $\rho_{12}, \rho_{21}, \tau_{12}, \tau_{21}$ from $C_0$, which gives $O_*(\tau^{(4)})$ (recall that $\|O_*(\tau^{(4)})\| \leq CW^{-2}$ by Lemma 6.2). If the term in (6.16) contains two or more $O(W^{-1})$, then it is $O(W^{-2})$, and if the term contain only one $O(W^{-1})$, then we have at most $W^{-1} O_*(\tau^{(2)})$, since both $\langle K \rangle^{(1\mu)}_{00}$ and $\langle K \rangle^{(v1)}_{00}$ give $O_*(\tau^{(1)})$. Thus contribution of the second term of (6.15) to $W^2 M_{4i}$ can be included in $\tilde{K}_5$ (see (4.7)).

Now rewrite the contribution of the first sum of (6.15) to $M$ as the sum of the terms
\[
P_0 T_k P_0, \quad T_k = P_g \langle K \rangle_{0k} G_{kk} \langle K \rangle_{k0} P_g = \langle (K)_{0k} G_{kk} (K)_{k0}^{(1)} \rangle,
\tag{6.17}
\]
where $P_0$ is defined in (2.50), and $P_g$ is the projection on the "good" vectors in $\Omega_{70}$ (evidently $P_0P_g = P_0$) and the upper index (11) corresponds to the upper left block of the matrix in the decomposition (2.40). Since all matrices $\hat{\Pi}, \hat{F}, F_0, \hat{C}_d, \hat{\Pi}_0, \hat{\Pi}_0^{(F)}$ are block diagonal in decomposition (2.40), the expression for $T_k$ may include non-diagonal blocks of matrices $O_*(C)$ only. But then we should have at least two of such blocks which gives at most $O_*(N^2)$. In addition, according to (6.10), $\langle K \rangle_{\bar{k}\bar{k}}$ and $\langle K \rangle_{\bar{k}0}$ give $O(W^{-1/2})$ each. Thus all such terms are of order $W^{-1}O_*(N^2)$, and we have to take only terms that does not contain any $O_*(C)$ (remark that diagonal block of $O_*(C)$ is $W^{-1}$ so it can give the contribution only to the terms of order $O(W^{-2})$). Hence to check that $(T_0 \otimes T_0)M(T_0 \otimes T_0)$ has the form (4.3), we need to study the structure of two matrices
\[
M_1 = P_0(\hat{F} \hat{\Pi}(1 + W^{-1}\lambda)A_{ab}(\hat{C}_d + C)\hat{F})_{\bar{k}\bar{k}} P_0 + \text{err},
\]
\[
M_2 = P_0 \left( \sum_{\bar{k} = \bar{i}, \bar{3} \bar{i}} (\hat{F} \hat{\Pi} A_{ab} K_{US} \hat{F})_{\bar{k}\bar{k}} (\lambda_{\bar{k}} \Pi_0^{(F)} - \bar{z})^{-1} (\hat{F} \hat{\Pi} A_{ab} K_{US} \hat{F})_{\bar{k}\bar{k}} \right) P_0 + \text{err},
\]
(6.18)
where err means the error terms which we describe above, and $P_0$ is defined in (2.50).

Let us check that up to err terms the matrix $(T_0 \otimes T_0)M_1(T_0 \otimes T_0)$ has the same form as (4.3), but with functions $u$ and $y_1, y_2$ (see (4.5)) replaced by some constants $\bar{u}, \bar{y}_1, \bar{y}_2$. Notice that the block structure of $Q$ (see (2.40)) and the fact that only even degrees of $\bar{a}_1, \bar{a}_2, \bar{a}_1', \bar{a}_2', \bar{b}_1, \bar{b}_2, \bar{b}_1', \bar{b}_2'$ give non-zero contribution yield
\[
M_1 = P_0 P_g \left( \langle \hat{F} \hat{\Pi} K_{US} \hat{F} \rangle_{\bar{k}\bar{k}} + W^{-1}F_0 \Pi_0 \lambda_{\bar{k}}^{(11)} K_0 F_0 + F_0 \Pi_0 C_{\bar{k}}^{(11)} F_0 \right) P_0 P_0 + \text{err},
\]
(6.19)
Indeed, using that $\hat{\Pi}$ in the decomposition (2.40) is block diagonal and $\lambda^{(12)} = 0$ (see (2.42)), we have
\[
(\hat{F} \hat{\Pi}(1 + W^{-1}\lambda)\hat{C} \hat{F})^{(11)} = \hat{F} \left( \hat{\Pi}^{(11)} \hat{C}^{(11)} + W^{-1} \hat{\Pi}^{(11)} \lambda^{(13)} C^{(31)} \right) \hat{F}.
\]
(6.20)
By (2.40) $\lambda^{(13)}$ corresponds to the transformation of "non-good" vectors into "good" ones. But by (2.45) there are only two "non-good vectors" that become "good" after an application of $\lambda$: $\rho_{11} \tau_{12} \rho_{22} \tau_{21}$ and $\rho_{12} \tau_{22} \rho_{21} \tau_{11}$. To obtain these vectors from "good" ones, we need at least two "transformations", hence corresponding entries of $C^{(31)}$ should be at least $O_*(N^2)$, which gives $W^{-1}O_*(N^2)$ and so can be included into an error term.

Let us now consider the contribution of the second and the third matrices in the r.h.s. of (6.19). According to (2.45), $\lambda_{\bar{k}}^{(11)}$ depends on $n_{21} - n_{12}$ and does not contain terms with $n_{12}n_{21}$, and $\Pi_0$ and $F_0$ do not increase the total number of $\rho_{12}, \rho_{21}, \tau_{12}, \tau_{21}$ in the Grassmann vector to which they are applied. Thus the application of $F_0 \Pi_0 \lambda_{\bar{k}}^{(11)} F_0$ to $e_0^+ e_0^-$ gives zero contribution to $M_{41}$, and gives the same constant contribution to $M_{21}$ and $M_{31}$ but with the opposite sign (this contribution thus can be included in $\bar{u}$). Similarly with $M_{42}$ and $M_{43}$. The same statement for $F_0 \Pi_0 C_{\bar{k}}^{(11)} F_0$ follows from (6.2).

To study the first term of the r.h.s. of (6.19), notice that since only even degrees of $\bar{a}_1, \bar{a}_2, \bar{a}_1', \bar{a}_2', \bar{b}_1, \bar{b}_2, \bar{b}_1', \bar{b}_2'$ give non-zero contribution, we need to study the contribution of order $W^{-1}$ only (the next order will be $W^{-2}$ and it may give contribution to $W^2 M_{41}$.
only, thus it will be included in \( u_0 \) (see (4.3)). Observe that to obtain the non-zero contribution of order \( W^{-1} \) we need to consider the term of this order from

\[
\left( Q(c_{12}) \otimes Q(c_{21}) \otimes Q(c_{11}) \otimes Q(c_{22}) \right) \times \mathcal{K}_{US}(t_{a'}, t_{b'}) A_1^+(\widetilde{a}_1, \widetilde{a}_1') A_1^-(\widetilde{b}_1, \widetilde{b}_1') A_2^+(\widetilde{a}_2, \widetilde{a}_2') A_2^-(\widetilde{b}_2, \widetilde{b}_2') \hat{F}(\widetilde{a}, \widetilde{b}) \hat{F}(\widetilde{a}', \widetilde{b}')
\]

(6.21)

(see (9.2), (2.21), (2.41) and (2.43)). Notice first that by (6.4) the terms containing the derivatives of \( \mathcal{K}_{US} \) with respect to \( \widetilde{a}, \widetilde{b} \) will give us \( O_n\left(\epsilon^{(2)}\right) \), so after multiplication by \( W \) they contribute to \( \widetilde{K}_i \). Thus we can change \( \mathcal{K}_{US} \) to \( \mathcal{K}_0 \). Also it is easy to see that we need to take at least one \( \widetilde{a}'_1, \widetilde{a}'_2, \widetilde{b}'_1, \widetilde{b}'_2 \) from \( Q(c_{12}) \otimes Q(c_{21}) \), since otherwise the respective term will contain \( Q_0 \otimes Q_0 \) and will not contribute to the entries which are important for us. The second observation is that if we take, e.g., the term containing \( \widetilde{a}'_1 \) from the first matrix, then we need to complete it by \( \widetilde{a}_1 \) or \( (\widetilde{a}_1')^3 \), or \( \widetilde{a}_1' \) or \( (\widetilde{a}_1)^3 \), otherwise the contribution will be 0. Hence, in this case the the second matrix should stay untouched. The next observation is that \( \mathcal{Q}'(c) \) has the form

\[
\mathcal{Q}'(c) = \begin{pmatrix}
-\beta^{-1} & 0 \\
-1 & 0
\end{pmatrix},
\]

(6.22)

hence \( \mathcal{Q}''(c) = 0 \), and if we take \( \widetilde{a}'_1 \) from the first matrix, the respective term will be of the form

\[
F_0^2 \cdot (\mathcal{Q}' \otimes Q_0 \otimes \widetilde{Q}) + F_0 \left. \frac{\partial F}{\partial a_1} \right|_{\widetilde{a} = \widetilde{b} = (a_1, a_2)} (\mathcal{Q}' \otimes Q_0 \otimes \widetilde{Q}_1)
\]

where \( \widetilde{Q} \) and \( \widetilde{Q}_1 \) are some \( 4 \times 4 \) matrix, whose coefficients depend on

\[
(\widetilde{a}'_1 \cdot \widetilde{a}_1')_{00}, \quad (\widetilde{a}'_1 \cdot (\widetilde{a}_1')^3)_{00}, \quad (\widetilde{a}'_1 \cdot \widetilde{a}_1)_{00}, \quad (\widetilde{a}'_1 \cdot (\widetilde{a}_1)^3)_{00}.
\]

If we consider the contribution of the terms containing \( \widetilde{b}'_1 \) taken from the second matrix in (6.21), then it will have the form

\[
-F_0^2 (Q_0 \otimes Q' \otimes \widetilde{Q}) - F_0 \left. \frac{\partial F}{\partial b_1} \right|_{\widetilde{a} = \widetilde{b} = (a_1, a_2)} (Q_0 \otimes Q' \otimes \widetilde{Q}_1)
\]

with the same \( \widetilde{Q} \) and \( \widetilde{Q}_1 \), since the coefficients of \( \widetilde{a}'_1 \) differs from the respective coefficient of \( \widetilde{b}'_1 \) by the multiplier \( i \), the same for the coefficients of \( \widetilde{a}_1' \) and \( \widetilde{b}_1' \), \( (\widetilde{a}_1')^3 \) and \( (\widetilde{b}_1')^3 \). Repeating the same argument for the terms containing \( \widetilde{a}_2 \) taken from \( Q(c_{21}) \) in (6.21) and \( \widetilde{b}_2 \) taken from \( Q(c_{12}) \), we obtain the form (4.3) for \( TM_1 T \).

Now let us study the matrix \( M_2 \) of (6.18). Define

\[
M_2' = K_F^2 \sum_{k=\ell_1, 3\ell_1} P_g\langle r \rangle_{\ell_0} P_g (\lambda_k P_g \Pi_0^{(F)} P_g - z)^{-1} P_g \langle r \rangle_{\ell_0} P_g + O(W^{-2})
\]

(6.23)

\[
= K_F^2 z^{-1} \sum_{k=\ell_1, 3\ell_1} \sum_m (\lambda_k / z)^m (K_F)^m P_g\langle r \rangle_{\ell_0} \Pi_0^m \langle r \rangle_{\ell_0} P_g + O(W^{-2}),
\]

where \( P_g \) is the projection on the “good” vectors in \( \Omega_{70} \), and \( K_F = F_0 K_0 F_0 \) (see the formulation of Proposition 4.1). Here \( \ell_i \) are defined in (6.9) and \( r \) collects the terms of order \( W^{-1/2} \) from (6.21). It is easy to see that \( M_2 = P_0 M_2' P_0 \). Recall that we are
interested in \((M_2)_{ij}\) and \((M_2)_{i4}\) \((i = 2, 3)\). For \(\tilde{k} = 3\ell_i\) the only terms in \(r\) which can give non-zero contribution are those containing \(\tilde{a}_{j}^{3}\) or \(\tilde{b}_{j}^{3}\). But then in both terms \(P_{g}\langle r \rangle_{\tilde{0}k}\) and \(P_{g}\langle r \rangle_{\tilde{k}0}\) the first two matrices remain untouched and taking into account that

\[
\Pi_{0}^{m} = Q_{0}^{m} \otimes Q_{0}^{m} \otimes (1 + q_{+}^{m}P_{10} + q_{-}^{m}P_{01} + (q_{+}q_{-})^{m}P_{11}), \quad Q_{0}^{m} = \begin{pmatrix} 1 & m\beta^{-1} \\ 0 & 1 \end{pmatrix},
\]

we conclude that the terms with \(\tilde{k} = 3\ell_i\) do not contribute to the entries of \(P_{0}M_{2}P_{0}\) which we are interested in.

For \(\tilde{k} = \ell_1, \ell_3\), repeating the argument used above for \(M_{1}'\) and taking into account that \(\hat{Q}\) depends on \(\tilde{a}_{j}, \tilde{b}_{j}\) but does not depend on \(\tilde{a}_{j}, \tilde{b}_{j}\), we obtain that up to the term which do not contribute to the “important” entries

\[
P_{g}\langle r \rangle_{\tilde{0}\ell_1} \Pi_{0}^{m}\langle r \rangle_{\tilde{0}\tilde{k}}P_{g} = K_{F}\langle 2W^{-1}\rangle\left(v_{+}^{(1)}F_{0}^{-1}\frac{\partial \hat{F}}{\partial t_{1}} + v_{+}^{(2)}\right)\left(Q'Q_{0}^{m-1} \otimes Q_{0}^{m} \otimes Q_{+}^{m} \otimes Q_{m}\right),
\]

\[
P_{g}\langle r \rangle_{\tilde{0}\ell_3} \Pi_{0}^{m}\langle r \rangle_{\tilde{0}\tilde{k}}P_{g} = -K_{F}\langle 2W^{-1}\rangle\left(v_{+}^{(1)}F_{0}^{-1}\frac{\partial \hat{F}}{\partial b_{1}} + v_{+}^{(2)}\right)\left(Q_{0}^{m} \otimes Q'Q_{0}^{m-1} \otimes Q_{+}^{m} \otimes Q_{m}\right),
\]

where the coefficients \(v_{+}^{(1)}\) and \(v_{+}^{(2)}\) are not important for us. Similar relations hold for \(\langle r \rangle_{\tilde{0}\ell_2}\) and \(\langle r \rangle_{\tilde{0}\ell_4}\) with some \(v_{-}^{(1)}\) and \(v_{-}^{(2)}\).

Now after multiplication by \((\lambda_{\tilde{0}}/z)^{m}(K_{F})^{m}\), summation over \(m\), and then transformation (6.14), we get representation (4.2) for \(\mathcal{M}\).

The analysis of

\[
\mathcal{M}' = T_{0} \otimes T_{0} M', \quad \mathcal{M}'' = M''T_{0} \otimes T_{0}, \quad \mathcal{M}''' = M'''',
\]

\[
M' = P_{0}(H^{(11)} - \tilde{H})(1 - P_{0}), \quad M'' = (1 - P_{0})(H^{(11)} - \tilde{H})P_{0},
\]

\[
M''' = (1 - P_{0})(H^{(11)} - \tilde{H})(1 - P_{0})
\]

is simpler than that for \(M\), since we need only to show that the entries of \(M'\) and \(M''\) are \(O_{s}(r^{(1)})O_{s}(r^{(1)})\) or \(O(W^{-2})\). Denote by \(P_{sg}\) the projection onto the subspace of “semi-good” vectors in \(\Omega_{70}\). Similarly to (6.20), using the decomposition (2.40), we obtain

\[
P_{0}H^{(11)}(1 - P_{0}) = P_{g}P_{k}(K)_{00}P_{sg}P_{g} = P_{g}(\kappa^{(12)})_{00}P_{g},
\]

\[
\kappa^{(12)} = (\hat{F}\tilde{\Pi}(1 + W^{-1}\lambda)\hat{F})^{(12)} = \hat{F}\left(\tilde{\Pi}(1 + W^{-1}\lambda^{(11)})\mathcal{C}^{(12)} + W^{-1}\tilde{\Pi}(1 + \lambda^{(13)})\mathcal{C}^{(32)}\right)\hat{F} = O_{s}(r^{(1)})O_{s}(r^{(1)}) + W^{-1}O_{s}(r^{(2)}),
\]

where we have used that the argument given after (6.20) and the fact that by (6.1) \(\mathcal{C}^{(12)} = O_{s}(r^{(1)})O_{s}(r^{(1)})\). For \(P_{0}\tilde{H}(1 - P_{0})\) we have similarly to (6.17)

\[
P_{0}\tilde{H}(1 - P_{0}) = P_{g}P_{g}\tilde{H}P_{sg}P_{g}, \quad P_{g}\tilde{H}P_{sg} = \sum (\langle K \rangle_{\tilde{0}k} G_{\tilde{k}k} \langle K \rangle_{\tilde{k}0})^{(12)} + err.
\]

Repeating the argument given after (6.17), we obtain that the r.h.s. above either contains \(\mathcal{C}^{(12)}\), or contains at least two non-diagonal blocks of \(\mathcal{C}\). In both cases the contribution of the corresponding terms into the r.h.s. above is \(W^{-1}O_{s}(r^{(1)})O_{s}(r^{(1)})\). As for \(M'''\), since \(\tilde{H} = O(W^{-1})\), (4.4) follows from (6.10) and (6.1).
Now let us find $f^{(1)}$ and $g^{(1)}$ from Proposition 2.4. Remark that since $v$ of the (2.23) contains the odd degrees of $\tilde{a}$, $\tilde{b}$ only with the coefficients of order $W^{-k/2}$ ($k \geq 1$), we have (see (6.8)):

\begin{align*}
f_0 &= (v, \Psi_0) P_\Sigma e^{(0)} = (1 + O(W^{-1})) C_0 F_0 P_\Sigma e^{(0)}, \\
f_1 &= (1 + O(W^{-1}))(C_0(1 - P_\Sigma)e^{(0)}, \{f_{\bar{k}k} e^{(0)}\}_{\bar{k}k \neq 0}) F_0, \\
f_{\bar{k}k} := F_0^{-1} (v, \Psi_{\bar{k}}), \quad |f_{\bar{k}k}| \leq C W^{-d(k)/2},
\end{align*}

where the first component of $f_1$ corresponds to $((1 - P_\Sigma) v e^{(0)}, \tilde{\Psi}_0)$. Then, similarly to (6.15),

\begin{equation}
f^{(1)} = f_0 - \sum_{\bar{k}=\ell_1,3\ell_1} P_\Sigma \langle \mathcal{K} \rangle_{\bar{k}\bar{k}} G_{\bar{k}\bar{k}} f_{\bar{k}k} F_0 \\
- \sum_{\mu=2,3} \langle \mathcal{K} \rangle_{\mu00}^{(1)} G_{00}^{(\mu)} f_{1\bar{\mu}} F_0 + O(W^{-2}) F_0 + W^{-1} O_* (r^{(2)}) F_0. \tag{6.25}
\end{equation}

Denoting $\Sigma'_1$ the first sum above, one can see similarly to (6.16)

\begin{align*}
\Sigma'_1 &= P_\Sigma \left( \sum_{\bar{k}=\ell_1,3\ell_1} P_\Sigma (\tilde{\Phi} \tilde{\Pi} \tilde{\Delta} \tilde{\Pi}_0^{(F)} - z)^{-1} f_{\bar{k}k} \right) F_0 + \left( O(W^{-2}) + W^{-1} O_* (r^{(2)}) \right) F_0 \\
&= P_\Sigma \left( \sum_{\bar{k}=\ell_1,3\ell_1} P_\Sigma (r)_{\bar{k}k} (\tilde{\Phi}_0^{(F)} - z)^{-1} P_\Sigma f_{1\bar{k}} \right) F_0 + \left( O(W^{-2}) + W^{-1} O_* (r^{(2)}) \right) F_0. \tag{6.26}
\end{align*}

$P_\Sigma P_\Sigma$, since $f_{1\bar{k}} = P_\Sigma f_{1\bar{k}}$ and $\tilde{\Phi}$ and $\tilde{\Phi}_0^{(F)}$ have a block structure, and use $r$ defined in (6.23).

Repeating the argument which we used for $M$, we get that all entries of $\langle \mathcal{K} \rangle_{\mu00}^{(1)}$ and $G_{00}^{(\mu)}$ have an order $O(W^{-1})$ or $O_* (r^{(1)})$. Then, denoting $\Sigma'_2$ the second sum in (6.25), we have

\[ \Sigma'_2 = \left( O(W^{-2}) + O_* (r^{(1)}) O_* (r^{(1)}) \right) F_0. \]

Since $f_{1\bar{\mu}} \sim p(n_{11}, n_{22})$, where $p$ is some polynomial, applying the argument given after (6.16), we obtain for the forth component $(f^{(1)})_4$ of $f^{(1)}$

\[ (f^{(1)})_4 = \left( O(W^{-2}) + W^{-1} O_* (r^{(2)}) + O_* (r^{(4)}) \right) F_0. \]

Additional terms which come into $(f^{(1)})_1$ from $\Sigma'_1$ and $\Sigma'_2$ are $O(W^{-1}) + O_* (r^{(2)})$, hence

\[ (f^{(1)})_1 = \left( d_1 + O(W^{-1}) + O_* (r^{(2)}) \right) F_0, \quad d_1 = C_0 (P_\Sigma e^{(0)})_1. \]
Moreover, repeating the argument given after (6.23) for the sum in the r.h.s. of (6.26), we obtain

$$W(f^{(1)})_2 = \left( \phi(K_F/z) + O(W^{-1}) + O_*(t^{(2)}) \right) F_0,$$

$$W(f^{(1)})_3 = \left( - \phi(K_F/z) + O(W^{-1}) + O_*(t^{(2)}) \right) F_0$$

with some $\phi(\zeta)$ analytic in $B_{1+\delta}$. Finally, since

$$Te_1 = W^{-1}e_4, \quad Te_2 = e_3, \quad Te_3 = e_2, \quad Te_4 = We_1,$$

we obtain that $\hat{f} = WTf^{(1)}$ satisfies conditions of (4.8). The relations for $\hat{g} = WT^*g^{(1)}$ can be obtained similarly.

To finish the proof of Proposition 4.1, we are left to prove (4.6).

To this end, consider the case when $n \sim \log W$. Since by (6.1) $|\tau| \leq CW^{-1/2}$, we have $\|\hat{K}_i\| \leq CW^{-1/2}$ and using the formula (5.5) for $M_1(z)$ (which is $M(z)$ with $K_i = 0$) we evidently get the trivial bound

$$\|(M_1(z) - z)^{-1}\| \leq C\|G(z)\|^3, \quad G(z) = (K_F - z)^{-1},$$

$$\Rightarrow \|\mathcal{G}(z)\| = \|(M_1(z) - z + O(W^{-1/2}))^{-1}\|$$

$$\leq \|(M_1(z) - z)^{-1}\| \|1 + O(W^{-1/2})(M_1(z) - z)^{-1}\| \leq Cn^3.$$  

Here we used that $\|G(z)\| \leq Cn$ for $z \in \mathcal{W}_0$. Thus, if we define $M_0(z)$ as $M_1(z)$ with $K_F$ replaced by $F_0^*$ in the upper left block and $\mathcal{G}_0(z) = (M_0(z) - z)^{-1}$, then, by (4.18),

$$|(\mathcal{G}(z) - \mathcal{G}_0(z))\hat{f}, \hat{g}| \leq \|\mathcal{G}(z)\| \|(O(W^{-1/2})\mathcal{G}_0(z)\hat{f}\| \|\hat{g}\| + \|(1 - K_F)\mathcal{G}_0(z)\hat{f}\| \|\hat{g}\|)$$

$$\leq Cn^3 \left( (O(W^{-1/2})n + \|(1 - K_F)\mathcal{P}_L\| \|\mathcal{G}_0(z)\hat{f}\| \|\hat{g}\| + \|(1 - K_F)\| \|(1 - \mathcal{P}_L)\mathcal{G}_0(z)\hat{f}\| \|\hat{g}\| \right) \to 0.$$  

Here we used the projection $\mathcal{P}_L$ defined in (4.11), bound $\|(1 - K_F)\mathcal{P}_L\| \leq L^2/W$ which follows from (6.5), the bound (4.13) for the last term here, and $n \sim \log W$. Now, the above argument, bound (7.3) for $I_+$, and a similar bound for $I_-$, yield

$$\mathcal{R}_{nW}^+(E, \varepsilon, \xi) = I_+ + I_+ + I_- = -\frac{1}{2\pi i} \oint_{\mathcal{W}_0} z^{n-1} (\mathcal{G}_0(z)\hat{f}, \hat{g}) dz + o(1),$$

and thus we can rewrite $\mathcal{R}_{nW}^+(E, \varepsilon, \xi)$ according to (4.20).

Observe that by definition of $P_\mathcal{L}$ (see (4.1)) one can see that the constants $d_{1_+}$ and $d_{2_+}$ of (4.8) are not zero. Indeed,

$$d_{1_+} = (e_1 \otimes e_1 \otimes s^+ \otimes s^- e_1 \otimes e_1 \otimes e_1 \otimes e_1) = c_+^- c_1^- (c_0^+ - c_1^-)^{-1} (c_0^- - c_1^-)^{-1} \neq 0,$$

$$d_{2_+} = (e_2 \otimes e_2 \otimes e_+ \otimes e_+ D_0^{(1)}(e_2 \otimes e_2) \otimes e_2 \otimes e_2)$$

$$= \left( D_0(e_2 \otimes e_2) \otimes e_+ \otimes e_+ e_2 \otimes e_2 \otimes e_2 \otimes e_2 \right) = (c_0^+ - c_+)(c_0^- - c_-) \neq 0.$$  

Hence taking the integral with respect to $z$ (see (4.20)) and using that for $u(1) \neq 0$

$$-\frac{1}{2\pi i} \oint_{\mathcal{W}_0} dz z^n G_F^3(z) u^2 (F_0^*) \sim Cn^2 u^2 (F_0^2)^2 \sim O(n)$$

$$\sim Cn^2 u^2 (F_0^2)^2 + O(n)$$
and the other terms of (4.20) are maximum \(O(n)\) (recall that \(F_1\) and \(F_2\) of (4.5) contain the multiplier \(n^{-1}\)), we obtain

\[
\mathcal{R}_{nw}^+(E, \varepsilon, \xi) = Cn^2u^2(1)d_{1^*}d_{2^*} \int F_0^{2(n-2)}dUdS + O(n), \quad C_\ast \neq 0.
\]

On the other hand, we know that \(\mathcal{R}_{nw}^+(E, \varepsilon, 0) = 1\). Thus we conclude that \(u(1) = 0\). But using this in (4.20), we obtain that all the terms in (4.20) are bounded except the one which contain \(u_0\). Repeating the above argument we obtain that \(u_0(1) = 0\). \(\square\)

7. The Analysis of \(I_+\) and \(I_-\)

The analysis of \(I_+\) is much simpler than that for \(I_\pm\). We diagonalize \(Y\) according to (2.15), and use (2.43) for \(b_1, b_2\), but use another parametrization for \(X\). Instead of diagonalizing \(X\) by the unitary transformation, this time it is more convenient to parametrized it as following (cf (2.43))

\[
X_{11} = a_x(1 + i\theta_s\tilde{a}_1/\sqrt{W}), \quad X_{22} = a_x(1 + i\theta_s\tilde{a}_2/\sqrt{W}), \quad X_{12} = \frac{i\theta_+a_x(x + iy)}{\sqrt{2W}} = -X_{12}.
\]

This change transforms the measure \(dX_idY_i/(-\pi^2)\) from (2.3) to

\[
(a_-\theta_-)(a_+\theta_+)^5(2\pi)^{-3}d\tilde{a}_i\ d\tilde{b}_i\ d\tilde{x}_i\ dS_i,
\]

and it reduces the factor \(W^4\) in front of \(I_+\) in (2.35) to \(W\).

In addition, let \(\hat{C}''(S_i)\) be the matrix corresponding to the change of Grassmann variables

\[
(\rho_iS_i^{-1})_{\alpha\beta} \rightarrow \rho_{i,\alpha\beta}, \quad (S_i\tau_i)_{\alpha\beta} \rightarrow \tau_{i,\alpha\beta}, \quad (7.1)
\]

in the space \(\Omega_{70}\). Set also

\[
\hat{F}_+(X, \tilde{b}, S) = F_S^{1/2}(\tilde{b}, S) \cdot \exp \left\{ -\frac{i}{2n}\text{Tr} X(\varepsilon L + \hat{\xi}/\rho(E)) \right\}.
\]

In this parametrization the operator \(\mathcal{K}_+\) has the form

\[
\mathcal{K}_+(\tilde{a}, \tilde{b}, x, y, \rho, \tau, S; \tilde{a}', \tilde{b}', x', y', \rho', \tau', S')
\]

\[
= (-\lambda_0^+)^{-5/2}(\varepsilon - \lambda_0^-)^{-1/2}A_{ab}^+(\tilde{a}, \tilde{a}', x, x', y, y'; \tilde{b}, \tilde{b}')
\]

\[
\times \hat{F}_+(X, \tilde{b}, S)\beta^{-1}\hat{Q}(\rho, \tau; \rho', \tau', S', D_b)\hat{C}''(S(S')^{-1})K_S(S(S')^{-1})\hat{F}_+(X', \tilde{b}', S'),
\]

where

\[
A_{ab}^+(\tilde{a}, \tilde{a}', x, x', y, y'; \tilde{b}, \tilde{b}') = A_{a}^+(\tilde{a}_1, \tilde{a})A_{a}^+(\tilde{a}_2, \tilde{a}')A_{b}^+(\tilde{b}_1, \tilde{b})A_{b}^+(\tilde{b}_2, \tilde{b}')A_{a}(x, x')A_{a}(y, y')
\]

and (cf (9.1) and (9.2))

\[
A_{a}^+(x, x')
\]

\[
= (-\lambda_0^+)^{1/2} \frac{a_x\theta_+}{(2\pi)^{1/2}} \exp \left\{ (a_x\theta_+)^2(\beta(x - x')^2/2 - \frac{c_+}{4}(x^2 + (x')^2)) \right\} (1 + o(W^{-1})). \quad (7.2)
\]
The main order of this operator has the form $A^+_s(\vec{x}, \vec{x}')$, hence the largest eigenvalue of this operator is $1 + O(W^{-1})$. It is easy to see that, similarly to consideration below (see (2.39)–(2.46)), we have

$$\hat{Q}(\rho, \tau; \rho', \tau', X', D_b) = \beta^2 \left((Q_0 \otimes Q_+ \otimes Q_+ \otimes Q_0) \otimes R\right)(1 + o(1)),$$

and so

$$\mathcal{K}_+ = \beta(-\lambda_0^+)^{-1/2}(-\lambda_0^+)^{-1/2}A_s^+ \otimes A_s^+ \otimes A_s^- \otimes A_s^- \otimes A_s^+ \otimes A_s^+ \otimes (Q_0 \otimes Q_+ \otimes Q_+ \otimes Q_0) \otimes R \otimes (K_5^\prime)^n(1 + o(1)).$$

Hence, using that the maximum eigenvalues of $A_s^+, A_s^-, Q_0, Q_s$ are 1, the largest eigenvalue of $K_5^\prime^n$ is $1 + O(W^{-1})$, and $|\lambda_0(R)| < 1$ (similarly to Lemma 2.2), we obtain

$$\lambda_0(\mathcal{K}_+) = \beta(-\lambda_0^+)^{-1/2}(-\lambda_0^+)^{-1/2}(1 + o(1)) = (q_+ q_-)^{1/2}(1 + o(1)) \implies |\lambda_0(\mathcal{K}_+)| < 1 - c.$$

Here we used the definition of $q_\pm$ (see (9.5)).

Thus, choosing the circle of radius $1 - c$ as the integration contour in (2.35), we get that

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

(7.3)

The bound for $I_-$ can be obtained similarly.

8. Proof of Theorem 1.2

The analysis of $\mathcal{K}^+$ of (2.12) is much more simple then that of $\mathcal{K}$, and can be done by similar arguments.

Define

$$\tilde{\Omega}_\pm^+ = \{a_1, a'_1, b_1, b'_1, b_2, b'_2 \in \Omega_+, a_2, a'_2 \in \Omega_-\},$$

$$\tilde{\Omega}_\mp^+ = \{a_1, a'_1, a_2, a'_2, b_1, b'_1, b_2, b'_2 \in \Omega_+,\}$$

$$\tilde{\Omega}_\mp^+ = \{a_1, a'_1, a_2, a'_2 \in \Omega_-, b_1, b'_1, b_2, b'_2 \in \Omega_+\}.$$

Repeating the argument of Sect. 2.3 one can obtain

$$\mathcal{R}_{W_\mathcal{K}}(E, \varepsilon, \bar{\xi}) = -\frac{W_4}{2\pi i} \oint_{\tilde{\omega}_{\mathcal{K}}} z^{n-1} (G_\pm^1(z) f_\pm, g_\pm) dz - \frac{W_4}{2\pi i} \oint_{\tilde{\omega}_{\mathcal{K}}} z^{n-1} (G_\pm^1(z) f_\pm, g_\pm) dz$$

$$- \frac{W_4}{2\pi i} \oint_{\tilde{\omega}_{\mathcal{K}}} z^{n-1} (G_\pm^1(z) f_\pm, g_\pm) dz + o(1) = I_+^1 + I_+^1 + I_+^1 + o(1),$$

(8.1)

where

$$\mathcal{K}_\mp^+ = \mathbf{1}_{\Omega_\pm^+} \mathcal{K}^+ \mathbf{1}_{\Omega_\pm^+}, \quad \mathcal{K}_\mp^+ = \mathbf{1}_{\tilde{\Omega}_\pm^+} \mathcal{K}^+ \mathbf{1}_{\tilde{\Omega}_\pm^+}, \quad \mathcal{K}_\mp^+ = \mathbf{1}_{\tilde{\Omega}_\pm^+} \mathcal{K}^+ \mathbf{1}_{\tilde{\Omega}_\mp^+},$$

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

and $f_\pm^1, f_\pm^1, f_\pm^1$ and $g_\pm^1, g_\pm^1, g_\pm^1$ are projections of $f^+$ and $g^+$ onto the subspaces corresponding to $\mathcal{K}_\pm^+$, $\mathcal{K}_\mp^+$, $\mathcal{K}_\mp^+$. We will prove below that this time the main contribution to (8.1) is given by $I_+^1$. 

M. Shcherbina, T. Shcherbina
8.1. The analysis of $I_+$. Similarly to Sect. 7, in this case it is more convenient to consider $X$ like unitary matrix which is close to the $a_+I_2$, and $Y$ as a Hermitian matrix which is close to $a_+I_2$. Then $X_i$ and $Y_i$ will be parametrized as (cf (2.43))

\[
X_{11} = a_+(1 + i \theta_a \tilde{a}_1 / \sqrt{W}), \quad X_{22} = a_+(1 + i \theta_a \tilde{a}_2 / \sqrt{W}), \quad X_{12} = \frac{i a_+ \theta_a (x + iy)}{\sqrt{2W}} = -X_{12},
\]

\[
Y_{11} = a_+(1 + \theta_a \tilde{b}_1 / \sqrt{W}), \quad Y_{22} = a_+(1 + \theta_a \tilde{b}_1 / \sqrt{W}), \quad Y_{12} = \frac{a_+ \theta_a (p + iq)}{\sqrt{2W}} = Y_{12}.
\] (8.2)

This change transforms the measure $dX_i dY_i / (-\pi^2)$ from (2.10) to

\[
(a_+)^8 (\theta_a)^8 (2\pi)^{-4} d\tilde{a}_i d\tilde{b}_i d\tilde{x}_i d\tilde{y}_i d\tilde{p}_i d\tilde{q}_i,
\]

and it “kills” the factor $W^4$ in front of $I_+^*$ in (8.1). Set also

\[
\hat{F}_+^*(X, Y) = \exp \left\{ \frac{i}{2n} \text{Tr} Y (\epsilon I + \hat{\xi} / \rho(E)) - \frac{i}{2n} \text{Tr} X (\epsilon I + \hat{\xi} / \rho(E)) \right\}.
\]

It is easy to see also that with this parametrization the operator $K_+^*$ has the form

\[
K_+^*(X, Y; X', Y') = (-\lambda_0^*)^{-4} \hat{F}_+^*(X, Y) A_{ab}^+ (\tilde{a}, \tilde{a}', x, x', y, y'; \tilde{b}, \tilde{b}', p, p', q, q')
\]

\[
\times \hat{F}_+^*(X', Y') \hat{Q}(\rho, \tau; \rho', \tau'; X', Y')(1 + o(1)),
\]

where

\[
A_{ab}^+(\tilde{a}, \tilde{a}', x, x', y, y'; \tilde{b}, \tilde{b}') = A_a^+ (\tilde{a}_1, \tilde{a}_1') A_a^+ (\tilde{a}_2, \tilde{a}_2') A_b^+ (\tilde{b}_1, \tilde{b}_1') A_b^+ (\tilde{b}_2, \tilde{b}_2') A_{ab}^+ (x, x') A_{ab}^+ (y, y') A_{ab}^+ (p, p') A_{ab}^+ (q, q')
\]

and $A_+$ is defined in (7.2).

The main order of $A_+^*$ has the form $A_+^*$ (see (9.2)), hence the largest eigenvalue of $A_{ab}^+$ is $1 + O(W^{-1/2})$, and the next eigenvalue is smaller then $1 - \delta$.

Moreover, if we consider $\hat{Q}(\rho, \tau; \rho', \tau'; X', Y')$ in this parametrization, we get

\[
(-\lambda_0^*)^{-4} \hat{Q}(\rho, \tau; \rho', \tau') = (-\lambda_0^*)^{-4} \left( \prod_{\mu, \nu = 1, 2} e^{\beta (\rho_{\mu\nu} - \rho_{\mu\nu}')} (\tau_{\mu\nu} - \tau_{\mu\nu}') e^{-c_{\mu\nu}'} (1 + O(W^{-1/2})) \right)
\]

\[
= (\Pi_{0,+} \oplus R_{0,+}) (1 + O(W^{-1/2})),
\] (8.3)

where

\[
\Pi_{0,+} := Q_+ \otimes Q_+ \otimes Q_+ \otimes Q_+
\] (8.4)

with $Q_+$ of (2.46) corresponding to “good” vectors, and $R_+$ corresponding to all other vectors. Similarly to Lemma 2.2 one can prove

**Lemma 8.1.** Given $R_{0,+}$ of (8.8), we have

\[
\lambda_0(R_{0,+}) < 1 - \delta.
\]
Proof. Indeed, as was mentioned in the proof of Lemma 2.2, any “non-good” product can be represented as an “absolutely non-good” part $\tilde{\eta}$ and a “good” part $p(n_{11}, n_{22}, n_{12}, n_{21})$. Here “absolutely non-good” are the products which do not contain any $n_{\alpha, \beta}$. It is easy to see that the exponential part of $\tilde{Q}$ (see (8.8)) transforms $\tilde{\eta} \to \beta^m \tilde{\eta}$, where $m \geq 2$ is the degree $\tilde{\eta}$. But $\beta/\lambda_0^+ \leq 1 - \delta$, thus the eigenvalue is smaller than $1 - \delta$. □

Now Lemma 8.1 and the consideration above yield that the first eigenvalue $\lambda_0(K^+_\omega)$ of $K^+_\omega$ is $1 + O(W^{-1/2})$, and the next eigenvalue is smaller than $1 - \delta$. Since $K^+_\omega$ is a compact operator, according to the spectral theorem, we can rewrite its resolvent $G^+_\omega(z)$ as

$$G^+_\omega(z) = \frac{P_{\mu}}{\lambda_0(K^+_\omega) - z} + R(z), \quad \|R(z)\| \leq C_\delta, \quad (8.5)$$

where $R(z)$ is analytic operator-functions in $\{z : 1 - \delta/2 < |z| < 1 + \delta\}$, and $P_{\mu}$ is a rank one operator of the form $P_{\mu} = \mu \otimes \mu^*$ with vectors $\mu, \mu^*$ such that

$$K^+_\omega \mu = \lambda_0(K^+_\omega) \mu, \quad (K^+_\omega)^* \mu^* = \frac{\beta}{\lambda_0(K^+_\omega)} \mu^*. \quad (8.6)$$

Thus, if we change the contour $\omega_0$ to $L_1 \cup L_0$

$$L_1 = \{z : |z| \leq 1 - \delta/2\}, \quad L_0 = \{z : |z - \lambda_0(K^+_\omega)| \leq \delta/4\}, \quad (8.7)$$

we get

$$I_\pm = -\frac{1}{2\pi i} \int_{L_0} + \int_{L_1} z^{n-1}(G^+_\omega(z)f_+, g_+)dz = -\frac{1}{2\pi i} \int_{L_0} z^{n-1}(G^+_\omega(z)f_+, g_+)dz + O(e^{-cn})$$

$$= -\frac{1}{2\pi i} \int_{L_0} z^{n-1}\left(\frac{P_{\mu}}{\lambda_0(K^+_\omega) - z}\right)f_+, g_+\right)dz = (\lambda_0(K^+_\omega))^{n-1}(P_{\mu}f_+, g_+) + O(e^{-cn}).$$

Since we will prove below that $I^+_\omega$ and $I^-_\omega$ are of order $O(e^{-cn})$, we have from (8.1)

$$\mathcal{R}_{W_\omega}(E, \varepsilon, \xi) = e^{i(\xi_1 + \xi_2 - \xi_1' - \xi_2')/\rho(E)} \cdot (\lambda_0(K^+_\omega))^{n-1}(P_{\mu}f_+, g_+) + O(e^{-cn}).$$

Notice that according to the consideration above

$$\lambda_0(K^+_\omega) = 1 + C_1/\sqrt{W} + O(W^{-1}), \quad (P_{\mu}f_+, g_+) = c_1 + O(W^{-1/2}).$$

In addition, due to the definition of $\mathcal{R}_{W_\omega}(E, \varepsilon, \xi)$ (see (1.11)), we have

$$\mathcal{R}_{W_\omega}(E, \varepsilon, \xi)|_{\xi' = \xi} = 1.$$
8.2. The analysis of $I^+$. In this case we will consider $X$ like unitary matrix which is close to the $a_- I_2$, and $Y$ as a Hermitian matrix which is close to $a_+ I_2$. Then $X_i$ and $Y_i$ will be parametrized as (cf (8.2))

$$X_{11} = a_-(1 + i \theta_\alpha a_1 / \sqrt{W})$$

$$Y_{11} = a_+(1 + \theta_\alpha b_1 / \sqrt{W})$$

This change transforms the measure $dX_i dY_i / (-\pi^2)$ from (2.10) to

$$(2\pi)^{-4} d\tilde{a}_i d\tilde{b}_i d\tilde{x}_i d\tilde{y}_i d\tilde{p}_i d\tilde{q}_i,$$

and it “kills” the factor $W^4$ in front of $I^+$ in (8.1).

In this parametrization the operator $K_+^-$ has the form

$$K_+^-(X, Y; X', Y') = (-\lambda_0^-)^{-2}(-\lambda_0^-)^{-2}A^-_{ab}(\tilde{a}, \tilde{a}'; x, x', y, y'; \tilde{b}, \tilde{b}'; p, p', q, q') \hat{Q}(\rho, \tau; \rho', \tau'; X', Y')(1 + o(1)),$$

where

$$A^-_{ab}(\tilde{a}, \tilde{a}'; x, x', y, y'; \tilde{b}, \tilde{b}') = A^-_a(\tilde{a}_1, \tilde{a}_2)A^-_a(\tilde{b}_1, \tilde{b}_2)A^-_b(\tilde{a}_2, \tilde{a}_2)A^-_b(\tilde{b}_1, \tilde{b}_2)A^-_a(x, x')A^-_b(y, y')A^+_p(p, p')A^+_p(q, q'),$$

where $A^+_p(x, x')$ is defined in (7.2), and

$$A^-_a(x, x') = (-\lambda_0^-)^{-1/2} \frac{a_+ - \theta_\alpha}{(2\pi)^{1/2}} \exp \left\{ (a_- - \theta_\alpha)^2 (\beta (x - x')^2 / 2 - \frac{c}{4} (x^2 + (x')^2)) \right\} (1 + o(1)).$$

Similarly to Sect. 9.1 one can get that the largest eigenvalue of $A^-_{ab}$ is $1 + O(W^{-1/2})$, and the next eigenvalue is smaller then $1 - \delta$. Considering $\hat{Q}(\rho, \tau; \rho', \tau'; X', Y')$ in this parametrization, we get

$$\hat{Q}(\rho, \tau; \rho', \tau') = \left( \prod_{\mu, \nu=1,2} e^{\beta (\rho_{\mu\nu} - \rho_{\nu\mu}) (\tau_{\nu\mu} - \tau_{\mu\nu})} \right) (1 + O(W^{-1/2})). \tag{8.8}$$

Here we used that $1 + (a_+ a_-)^{-1} = 0$. In the same way as in Lemma 8.1 it is easy to see that the largest eigenvalue of the matrix in (8.8) is $\beta^4 (1 + O(W^{-1/2}))$, and thus

$$\|(-\lambda_0^+)^{-2}(-\lambda_0^-)^{-2} \hat{Q} \| \leq \frac{\beta^4}{|\lambda_0^+|^4} (1 + O(W^{-1/2})) < 1 - \delta.$$

This implies

$$\|K_+^-\| < 1 - \delta,$$

and so

$$I_+^+ = O(e^{-cn}).$$
8.3. The analysis of $I^\pm_1$. In this case we diagonalize $X$ and use parametrization (2.43) for $X$ and parametrization (8.2) for $Y$. Then repeating almost literally the consideration in Sect. 7 (just changing $K_S$ to $K_U$) we get

$$I^\pm_1 = O(e^{-cn}).$$

9. Auxiliary Results

9.1. The analysis of $A_{ab}$. Take $c_\pm$ of (2.47) and consider the operator $A_b$ (2.19) near the point $a_+$ with $b, b'$ defined as $b_{1i}$ in (2.43), where

$$\theta_\pm = (|\kappa_\pm|/\kappa_\pm)^{1/2}, \quad \kappa_\pm = (c_\pm^2/4 - \beta c_\pm)^{1/2}. \quad (9.1)$$

It is easy to see that since $\phi''(a_+) = c_+$, after this change of variables and normalization by $(-\lambda_0^+)^{1/2}$ of (2.48), the kernel $A_b$ takes the form

$$(-\lambda_0^+)^{1/2} A_b \rightarrow A^+_b = A^+_a(1 + W^{-1/2} p_+({\tilde b}))(1 + W^{-1/2} p_+({\tilde b}')) + O(e^{-c\log^2 W}),$$

$$A^+_a = \frac{(-\lambda_0^+)^{1/2} a_+\theta_+}{\sqrt{2\pi}} \exp \left\{ (a_+\theta_+)^2 / 2 - \beta_+ b^2 / 2 - c_+ b^2 / 4 - c_+ (\tilde b')^2 / 4 \right\},$$

$$p_+({\tilde b}) = c_{3+} {\tilde b}^3 + c_{4+} {\tilde b}^4 W^{-1/2} + c_{5+} b^{-5} W^{-1} + \ldots \quad (9.2)$$

where the coefficients $c_{3+}, c_{4+}, \ldots$ are expressed in terms of the derivatives of $\phi_0$ at $a_+$. Introduce the two orthonormal bases

$$\psi^\pm_k(\tilde b) = |\kappa_\pm|^{1/4} H_k(|\kappa_\pm|^{1/2} b) e^{-|\kappa_\pm|/2}, \quad (9.3)$$

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

Notice that if we make the change of variables for $a, a'$ as $a_{1i}$ in (2.43) with $\theta_+ \theta_-^*$ of (9.1), then

$$(-\lambda_0^+)^{1/2} A_a \rightarrow A^+_a = A^+_a(1 + W^{-1/2} \hat p_+({\tilde a}))(1 + W^{-1/2} \hat p_+({\tilde a}')) + O(e^{-c\log^2 W}),$$

$$\hat p_+({\tilde a}) = i c_{3+} \tilde a^3 - c_{4+} \tilde a^4 W^{-1/2} - i c_{5+} \tilde a^5 W^{-1} + \ldots \quad (9.4)$$

with the same $A^+_a$ and $c_{3+}, c_{4+}$ as in (9.2).

Proposition 9.1. Let $\kappa_+, \kappa_-$ be defined as in (9.1). Then the matrices of the operators $A^+_a$ and $A^-_a$ are diagonal in the basis $\{\psi^+_k\}$ and $\{\psi^-_k\}$ and the corresponding eigenvalues have the form (cf (2.48))

$$\lambda_k^\pm = \lambda_k(A^\pm_a) = q_k^k, \quad k = 0, 1, 2 \ldots, \quad q_k^\pm := \frac{\beta}{\kappa_\pm + c_\pm / 2 - \beta}, \quad |q_k^\pm| < 1. \quad (9.5)$$

The matrices of operators $A^+_a$ and $A^-_a$ have the same (up to the error $W^{-1}$) diagonals as $A^+_a$ and $A^-_a$ respectively, and

$$(A^\pm)_{k,k'} = O(W^{-1/2})(\delta_{|k-k'|,1} + \delta_{|k-k'|,3}) + O(W^{-1})\delta_{|k-k'|,2} + O(W^{-(|k-k'|-3)/2}).$$
Proof. To simplify formulas, we consider the kernel
\[ M(x, y) = (2\pi)^{-1/2} e^{-(Ax, x)/2}, \ \tilde{\kappa} = (x, y), \ \mathcal{A} = \begin{pmatrix} \mu & v \\ v & \mu \end{pmatrix}, \ \lambda_{\pm} = \mu \pm v, \ \Re \lambda_{\pm} > 0. \]

Then, taking \( \kappa = \sqrt{\mu^2 - v^2} = \sqrt{\lambda_{+}\lambda_{-}} \), we obtain that
\[
(2\pi)^{-1/2} \int e^{-(Ax, x)/2} e^{-\kappa y^2/2} dy = q^k (\mu + \kappa)^{-1/2} e^{x^2/2} dy, \quad q = \frac{v}{\mu + \kappa}.
\]

Since the operator with the kernel \( e^{-\langle Ax, x \rangle/2} \) is compact, we have \( |q| < 1 \). Notice also that
\[
(-\lambda_0^+)^{1/2} (\mu + \kappa)^{-1/2} = 1.
\]

If we change the variables as
\[
x_1 = \theta x, \ y_1 = \theta y, \ \theta = e^{-i(\arg \lambda_{+} + \arg \lambda_{-})/4} = e^{-i\arg \kappa/2},
\]
then for the new matrix \( \tilde{A} = \theta^2 A \) has eigenvalues \( \theta^2 \lambda_{+}, \theta^2 \lambda_{-} \), whose real parts are still positive, \( \tilde{\kappa} = |\kappa| \), and \( \tilde{q} = q \).

9.2. Proof of Lemma 2.1. To simplify formulas, set
\[
\Lambda_1(t, s) = \Re \Lambda(e^{\varphi_0} t, e^{i\varphi_0} s) = \frac{\cos 2\varphi_0}{2} (\beta(t - s)^2 - \frac{t^2 + s^2}{2}) - \frac{E \sin \varphi_0}{2} (t + s) \\
+ \frac{\log t}{2} + \frac{\log s}{2} + \Re \phi_0 (a_+) \\
+ \frac{\log (t + s)}{4} + \frac{\log ((t + s)^2 + \sin^2 \phi)}{4},
\]

\[
\Lambda_2(t, s) = \Re \Lambda(e^{i\varphi_0} t e^{i\psi}, e^{i\varphi_0} s e^{i\psi}) = \frac{\cos 2\psi}{2} (\beta(t - s)^2 - \frac{t^2 + s^2}{2}) - \frac{E \sin \psi}{2} (t + s) \\
- \frac{\cos (\psi + \varphi_0)}{2} (t + s) + \frac{\log ((t + s) + \sin \phi + \sin^2 \phi)}{4} + \frac{\log ((s + \cos \phi + \cos^2 \phi)}{4},
\]

with \( \Lambda(x, y) \) defined in (2.19) and \( \phi = \varphi_0 - \psi \).

Then (2.30) for \( |\varphi_0| > \pi/4 \) follows from the inequalities
\[
\Lambda_1(t, s) \leq -c((t - 1)^2 + (s - 1)^2), \ \Lambda_2(t, s) \leq -c(t^2 + s^2) \\
\Re \Lambda(e^{i\varphi_0} t e^{i\psi}, e^{i\varphi_0} s e^{i\psi}) \leq \Lambda_1(t, 1) + \Lambda_2(0, s)
\]
and for \( |\varphi_0| < \pi/4 \) (2.30) follows from the first inequality in (9.6).

The first inequality in (9.6) follows from the relations
\[
\Lambda_1(1, 1) = 0, \ \frac{\partial \Lambda_1(1, 1)}{\partial t} = \frac{\partial \Lambda_1(1, 1)}{\partial s} = 0, \ D \Lambda_1 = \begin{pmatrix} \frac{\partial^2 \Lambda_1(t, s)}{\partial t^2} & \frac{\partial^2 \Lambda_1(t, s)}{\partial t \partial s} \\ \frac{\partial^2 \Lambda_1(t, s)}{\partial s^2} & \frac{\partial^2 \Lambda_2(t, s)}{\partial s^2} \end{pmatrix} < -cI.
\]
The relation for $\Lambda_1$ and its first derivatives follow from the fact that $x = y = e^{i\varphi_0}$ is the stationary point of $\Lambda(x, y)$. To prove the last bound, due to the symmetry $\Lambda_1$, it suffices to check that

$$\max \left\{ \frac{\partial^2 \Lambda_1(t, s)}{\partial t^2}, \frac{\partial^2 \Lambda_1(t, s)}{\partial s^2} \right\} + \frac{\partial^2 \Lambda_1(t, s)}{\partial t \partial s} < -c$$

$$\iff \frac{\cos 2\varphi}{2}(2\beta \pm 2\beta - 1) - \inf_{t < 1} \frac{1}{2t^2} \leq -c$$

(9.7)

The last inequality is valid since the absolute value of the first term above is less than $\frac{1}{2}$, while the second term is less than $-\frac{1}{2}$. Notice, that for $|\varphi_0| < \pi/4$ the last inequality is valid also for all $t, s > 0$, which implies (2.30) in this case.

For $|\varphi_0| > \pi/4$, to prove the second bound of (9.6), it suffices to use that $s = t = 0$ is the stationary point of $\Lambda_2(t, s)$ and the analogue of the first line of (9.7) is valid. Hence we need to check that

$$\frac{\cos 2\psi}{2}(2\beta \pm 2\beta - 1) - \inf_{t > 0} \frac{(t + \cos \phi)^2 - \sin^2 \phi}{2((t + \cos \phi)^2 + \sin^2 \phi)} \leq -c$$

Since the function under inf (we call it $d(t)$) for $t \geq 0$ has only one stationary point (maximum), $d(t)$ can take its minimum either at $t = 0$ or at $t \to \infty$. Since $0 < \phi = \varphi_0 - \psi < \pi/4$, we have $d(0) = \cos(2\phi) \geq 0$, and $d(t) \to 0$ as $t \to \infty, d(t) \to 0$, the above inequality takes the form

$$\cos 2\psi(2\beta \pm 2\beta - 1)/2 < -c$$

Since $\cos 2\psi > 0$ for $\pi/2 > \varphi_0 > \pi/4$ and $4\beta - 1 < 0$, the last inequality is valid. The last bound in (9.6) follows from the relations

$$\Re(e^{i\varphi_0} + se^{i\psi} - te^{i\varphi_0})^2 = \Re e^{2i\varphi_0}(1 - t)^2 + \Re e^{2i\psi} s^2 + 2(1 - t)s \cos(\varphi_0 + \psi)$$

$$\leq \Re e^{2i\varphi_0}(1 - t)^2 + \Re e^{2i\psi} s^2,$$

since $\psi + \varphi_0 \geq \pi/2 \Rightarrow \cos(\varphi_0 + \psi) \leq 0$.

For $\varphi_0 = \pi/4$ we have

$$\Re \Lambda(4e^{i\pi/4}, se^{i\pi/4}) = (-t - s + \log t + \log s + 2)/2$$

which obviously yields (2.30).

To prove (2.31), recall first that now the kernel of the operator has $-\Lambda(e^{i\varphi}, e^{i\varphi'})$ in the exponent, and notice that

$$-\Re \Lambda(e^{i\varphi}, e^{i\varphi'}) = -\beta(\cos \varphi - \cos \varphi')^2/2 + \beta(\sin \varphi - \sin \varphi')^2/2 - (\sin \varphi - \sin \varphi_0)^2/2$$

$$- (\sin \varphi' - \sin \varphi_0)^2/2$$

$$\leq \beta(\sin \varphi - \sin \varphi')^2/2 - (\sin \varphi - \sin \varphi_0)^2/2 - (\sin \varphi' - \sin \varphi_0)^2/2$$

$$\leq -(1 - 2\beta)(\sin \varphi - \sin \varphi_0)^2/2 - (1 - 2\beta)(\sin \varphi - \sin \varphi_0)^2/2.$$
9.3. Proof of Lemma 6.2. It is known that

\[ L_2(U) = \bigoplus_{l=0}^{\infty} L^{(l)} U, \quad L^{(l)} U = \text{Lin } \{ t_{mk}^{(l)} U \}_{m,k=-l}, \]  

(9.8)

where \( \{ t_{mk}^{(l)} (U) \}_{m,k=-l} \) are the coefficients of the irreducible representation of the shift operator \( T_U \tilde{U} = U \tilde{U} \). It follows from the properties of the unitary representation that

\[ t_{mk}^{(l)} (U^{-1}) = t_{km}^{(l)} (U), \quad t_{mk}^{(l)} (U_1 U_2) = \sum t_{mj}^{(l)} (U_1) t_{jk}^{(l)} (U_2). \]

According to [29], Chapter III,

\[ t_{mk}^{(l)} (U) = e^{-i (m \phi + k \psi)/2} P^{(l)}_{mk} (\cos \theta), \quad U = \left( \begin{array}{cc} \cos \theta & i e^{i (\phi + \psi)/2} \\ i e^{-i (\phi - \psi)/2} & \cos \theta \end{array} \right), \]  

(9.9)

where

\[ P_{mk}^{(l)} (\cos \theta) = \frac{c_{mk}}{2 \pi} \int_0^{2\pi} d\phi (\cos \theta + i \sin \theta e^{i \phi})^l (\cos \theta - i \sin \theta e^{-i \phi})^k, \]

\[ c_{mk} = \frac{(l - m)! (l + m)!}{(l - k)! (l + k)!} \]  

(9.10)

It is known also that \( \{ t_{mk}^{(l)} (U) \}_{m,k=-l} \) make an orthonormal basis in \( L^{(l)} U \).

For any function \( v(U) \) consider the matrix \( v^{(l)} U = \{ v_{mk}^{(l)} \} \) defined as

\[ v_{mk}^{(l)} U := \int v(U) t_{mk}^{(l)} U dU. \]  

(9.11)

It is easy to see that if we consider an integral operator \( \tilde{v} \) with the kernel \( v(U_1 U_2^{-1}) \), then

\[ (\tilde{v} t_{mk}) (U) = \sum v_{mj}^{(l)} t_{jk}^{(l)} (U). \]

Hence, we obtain that \( L^{(l)} U \) reduces \( \tilde{v} \) and the reduced operator \( \tilde{v}^{(l)} U \) is uniquely defined by the matrix \( v^{(l)} U \). Moreover, if \( v \) is some product of the matrix entries, then due to the integration with respect to \( \phi, \psi \) in (9.11) there is only one \( k \) and only one \( m \) such that \( v_{mk}^{(l)} \neq 0 \), hence, if we denote \( \mathcal{E}^{(l)}_{ij} \) an isometric operator such that

\[ \mathcal{E}_{ij} (t_{mk}^{(l)} U) = \delta_{j,m} t_{ik}^{(l)} U, \]  

(9.12)

then \( \tilde{v}^{(l)} U = v_{km}^{(l)} \mathcal{E}_{km} \). Let us find the matrices, corresponding to \( (|U_{12}|^{2p})_U \) in \( L^{(l)} U \).

Using (9.10) it is easy to see that

\[ (|U_{12}|^{2p})_U^{(l)} U = \lambda_{00}^{(l)} U \mathcal{E}_{00} \]  

(9.13)

\[ \lambda_{00}^{(l)} (s) = tW \int_{0}^{\pi} e^{-tW \sin^2(\theta/2)} \sin^2(\theta/2) P_{00}^{(l)} (\cos \theta) \sin \theta d\theta \]

Expanding the function under the integral (9.10) for \( \sin(\theta/2) \sim 0 \), it is easy to obtain

\[ P_{00}^{(l)} (1 - x) = 1 - x (l + m) (l + m + 1)/2 + O(l^4 x^2). \]  

(9.14)
Using this asymptotic in the above integral representation of $\lambda_{00}^{(U)}$, we get the first relation in (6.5) Similarly for the product zero type $U$ ($K_{sq}U := (|U_{11}|^{2q}U_{11}^{2q})_U$, and here and below, if $s < 0$, then we replace $U_{11}$ with $U_{22}$), we get

$$
\left( K_{sq}U \right)^{(l)} = \lambda_{-q,-q}^{(l)U} e_{-q,-q},
$$

$$\lambda_{-q,-q} = 1 + O(W^{-1}) - (l + q)(l + q + 1)/(tW) + O(l^2/W^2).$$

(9.15)

In particular, for $s = q = 0$ we obtain the second relation in (6.4) for $K_{00}$. The norm of any operator $(p)_U$ of the type $s$ is bounded because of the inequality

$$
\| (p)_U \| \le Wt \int |U_{11}|^m |U_{12}|^q e^{-Wt[U_{12}]}^2 dU \le \delta_{s,0} + C/W^{s/2}.
$$

To analyse the products of the first type, consider $p_{qs}^{(l)} = |U_{11}|^{2q}U_{12}^{2q+1}$. Then (9.9) and (9.11) yield

$$
\left( p_{qs}^{(l)} \right)_U = \lambda_{l}^{(l)U} (p_{qs}^{(l)}) e_{-1-q,-q},
$$

$$
\left( p_{qs}^{(l)} \right)_U \left( p_{qs}^{(l)} \right)_U = |\lambda_{l}^{(l)U} (p_{qs}^{(l)})|^2 e_{-q,-q},
$$

(9.16)

where

$$
\lambda_{l}^{(l)U} (p_{qs}^{(l)}) = tW \int_0^\pi e^{-W \sin^2(\theta/2)} \sin(\theta/2) \cos^2(s+q+1)(\theta/2) \phi_{-1-q,-q}(\cos \theta) \sin \theta d\theta
$$

$$
= (tW)^{-1} \sqrt{(l + q + 1)(l - q)} \left( 1 + O(l^2/tW) \right).
$$

(9.17)

Comparing (9.15) with $q = 0$ with the last two formulas for $q = 0$, we obtain (6.4).

To prove (6.6), notice that

$$
r^* r \le C \left( (p)_U \right)_U.
$$

By (9.16) $\left( (p)_U \right)_U \left( p \right)_U = O(l^2/W^2) e_{l,q}$ with some $q \in \mathbb{Z}$, and by (9.15) $\left( K_{\alpha} \right)_U = O(l^2/W) e_{l,q}$ if $q = 0$. If $q = q'$, then

$$
\left( p \right)_U \left( p \right)_U \le CW^{-1} (1 + C/W - K_{\alpha}).
$$

(9.18)

And if $q \neq q'$, then $\left( (p)_U \right)_U \left( p \right)_U \left( (p)_U \right)_U$ is not zero only on the image of $e_{l,q}$ and $K_{\alpha}$ is not hence the r.h.s. here is $O(W^{-1})$ and (9.18) is still true. The case of $\left( p(S) \right)_U$ is similar. The analysis of the difference operators in $L_2(S)$ is very similar. The difference is that for the hyperbolic group the irreducible representations are labeled by the continuous parameter $l' = -1/2 + i\rho$, $\rho \in \mathbb{R}$,

$$
t_{mk}^{(l)} = e^{i(m\phi + k\psi)} \mathcal{P}_{mk}^{(l)}(\theta), \quad m, k \in \mathbb{Z},
$$

and $\mathcal{P}_{mk}^{(l)}(\theta)$ has the form (9.10) with $\cos(\theta/2)$ replaced by $\cosh(\theta/2)$, $i \sin(\theta/2)$ replaced by $\sinh(\theta/2)$ and $c_{mk}$ replaced by 1 (see [29], Chapter VI). Then the same argument yields the second line of (6.5) and the last line for $K_{05}$.
9.4. Proof of Proposition 5.1. Consider first $D = \frac{d}{dx} x^2 \frac{d}{dx}$. It is easy to see that for any sufficiently smooth $\varphi(x)$

$$(D)^m \varphi(x) = \sum_{k=1}^{2m} c_{mk} x^k \varphi^{(k)}(x)$$

(9.19)

with some integer coefficients $c_{mk}$. Take $\varphi(x) = e^x$. Then we get

$$(D)^m e^x = Q_{2m}(x)e^x, \quad Q_{2m}(x) = \sum_{k=1}^{2m} c_{mk} x^k.$$ 

(9.20)

It is easy to see also that

$$Q_{2(l+1)}(x) = (x^2 + 2x)Q_{2l}(x) + (2x^2 + 2x)Q'_2(x) + x^2 Q''_2(x).$$

(9.21)

Define $\Sigma_m = \sum_{k=1}^{2m} |c_{mk}|$. Then considering both sides of (9.21) we obtain

$$\Sigma_{l+1} \leq (3 + 4 \cdot (2l) + (2l)(2l - 1)) \Sigma_l \leq 4(l + 1)^2 \Sigma_l.$$ 

Thus, since $\Sigma_1 = 3 < 4$,

$$\Sigma_m \leq 2^{2m} (m!)^2,$$

and so the same bound holds for each individual $c_{km}$. Hence

$$|D^m \varphi(x)| = \left| \sum_{k=1}^{2m} c_{mk} x^k \varphi^{(k)}(x) \right| \leq 2^{2m} (m!)^2 \sum_{k=1}^{2m} x^k |\varphi^{(k)}(x)|.$$ 

(9.22)

Recall now that

$$\Delta_S = -\frac{d}{dx} x(x + 1) \frac{d}{dx}, \quad x = |S_{12}|^2 \geq 0.$$ 

Notice that

$$x(x + 1) \leq (x + 1)^2,$$

$$\frac{d}{dx} x(x + 1) = 2x + 1 < 2x + 2 = \frac{d}{dx} (x + 1)^2,$$

$$\frac{d^2}{dx^2} x(x + 1) = 2 = \frac{d^2}{dx^2} (x + 1)^2.$$

Therefore, Proposition 5.1 follows from (9.22) where we substitute $x + 1$ instead of $x$ into the r.h.s.

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.
References

1. Bao, Z., Erdös, L.: Delocalization for a class of random block band matrices. Probab. Theory Relat. Fields 167, 673–776 (2017)
2. Bogachev, L.V., Molchanov, S.A., Pastur, L.A.: On the level density of random band matrices. Mat. Zametki 50(6), 31–42 (1991)
3. Bourgade, P.: Random band matrices. Proc. Int. Cong. Math. 3, 2745–2770 (2018)
4. Bourgade, P., Erdös, L., Yau, H.-T., Yin, J.: Universality for a class of random band matrices. Adv. Theor. Math. Phys. 21(3), 739–800 (2017)
5. Bourgade, P., Yau, H. T., Yin, J. Random Band Matrices in the Delocalized Phase I: Quantum Unique Ergodicity and Universality. Commun. Pure. Appl. Math. 73(7), 1526–1596 (2020)
6. Bourgade, P., Yang, F., Yau, H.-T., Yin, J.: Random band matrices in the delocalized phase. II: generalized resolvent estimates. J. Stat. Phys. 174, 1189–1221 (2019)
7. Casati, G., Molinari, L., Israilev, F.: Scaling properties of band random matrices. Phys. Rev. Lett. 64, 1851–1854 (1990)
8. Disertori, M., Lager, M.: Density of states for random band matrices in two dimensions. Ann. Henri Poincare 18(7), 2367–2413 (2017)
9. Disertori, M., Pinson, H., Spencer, T.: Density of states for random band matrices. Commun. Math. Phys. 232, 83–124 (2002)
10. Efetov, K.: Supersymmetry in Disorder and Chaos. Cambridge University Press, New York (1997)
11. Erdös, L., Knowles, A.: Quantum diffusion and eigenfunction delocalization in a random band matrix model. Commun. Math. Phys. 303, 509–554 (2011)
12. Erdös, L., Knowles, A., Yau, H.-T., Yin, J.: Delocalization and diffusion profile for random band matrices. Commun. Math. Phys. 323, 367–416 (2013)
13. Erdös, L., Yau, H.-T., Yin, J.: Bulk universality for generalized Wigner matrices. Probab. Theory Relat. Fields 154, 341–407 (2012)
14. Fyodorov, Y.V., Mirlin, A.D.: Scaling properties of localization in random band matrices: a σ-model approach. Phys. Rev. Lett. 67, 2405–2409 (1991)
15. Fyodorov, Y.V., Mirlin, A.D.: Statistical properties of eigenfunctions of random quasi 1d one-particle Hamiltonians. Int. J. Mod. Phys. B 8, 3795–3842 (1994)
16. He, Y., Marcozzi, M. Diffusion profile for random band matrices: A short proof. J. Stat. Phys. 177, 666–716 (2019)
17. 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, 108–118 (1992)
18. Peled, R., Schenker, J., Shamis, M., Sodin, A.: On the Wegner orbital model. Int. Math. Res. Not. 4, 1030–1058 (2019)
19. Schäfer, L., Wegner, F.: Disordered system with n orbitals per site: Lagrange formulation, hyperbolic symmetry, and Goldstone modes. Z. Phys. B 38, 113–126 (1980)
20. Schenker, J.: Eigenvector localization for random band matrices with power law band width. Commun. Math. Phys. 290, 1065–1097 (2009)
21. Shcherbina, M., Shcherbina, T.: Transfer matrix approach to 1d random band matrices: density of states. J. Stat. Phys. 164, 1233–1260 (2016)
22. Shcherbina, M., Shcherbina, T.: Characteristic polynomials for 1d random band matrices from the localization side. Commun. Math. Phys. 351, 1009–1044 (2017)
23. Shcherbina, M., Shcherbina, T.: Universality for 1d random band matrices: sigma-model approximation. J. Stat. Phys. 172, 627–664 (2018)
24. Shcherbina, M., Shcherbina, T.: Transfer matrix approach to 1d random band matrices. Proc. Int. Cong. Math. 2, 2673–2694 (2018)
25. Shcherbina, T.: On the second mixed moment of the characteristic polynomials of the 1D band matrices. Commun. Math. Phys. 328, 45–82 (2014)
26. Shcherbina, T.: Universality of the local regime for the block band matrices with a finite number of blocks. J. Stat. Phys. 155(3), 466–499 (2014)
27. Sodin, S.: The spectral edge of some random band matrices. Ann. Math. 173(3), 2223–2251 (2010)
28. 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)
29. Vilenkin, N. Ja.: Special Functions and the Theory of Group Representations. Translations of Mathematical Monographs, AMS 1968; 613 p
30. Wegner, F.J.: Disordered system with n orbitals per site: n → ∞ limit. Phys. Rev. B 19, 783–792 (1979)
31. Yang, F., Yau, J.: Random band matrices in the delocalized phase, III: averaging fluctuations. Probab. Theory Relat. Fields (2020). https://doi.org/10.1007/s00440-020-01013-5

Communicated by L. Erdos