Universality for 1d random band matrices: sigma-model approximation

Mariya Shcherbina [*] Tatyana Shcherbina [†]

This paper is dedicated to Tom Spencer on the occasion of his 70th birthday.

Abstract

The paper continues the development of the rigorous supersymmetric transfer matrix approach to the random band matrices started in [19], [20]. We consider random Hermitian block band matrices consisting of $W \times W$ random Gaussian blocks (parametrized by $j, k \in \Lambda = [1, n]^d \cap \mathbb{Z}^d$) with a fixed entry's variance $J_{jk} = \delta_{j,k} W^{-1} + \beta \Delta_{j,k} W^{-2}$, $\beta > 0$ in each block. Taking the limit $W \to \infty$ with fixed n and β, we derive the sigma-model approximation of the second correlation function similar to Efetov's one. Then, considering the limit $\beta, n \to \infty$, we prove that in the dimension $d = 1$ the behaviour of the sigma-model approximation in the bulk of the spectrum, as $\beta \gg n$, is determined by the classical Wigner – Dyson statistics.

1 Introduction

Random band matrices (RBM) represent quantum systems on a large box in $\mathbb{Z}^d$ with random quantum transition amplitudes effective up to distances of order W, which is called a bandwidth. They are natural intermediate models to study eigenvalue statistics and quantum propagation in disordered systems as they interpolate between Wigner matrices and random Schrödinger operators: Wigner matrix ensembles represent mean-field models without spatial structure, where the quantum transition rates between any two sites are i.i.d. random variables; in contrast, random Schrödinger operator has only a random diagonal potential in addition to the deterministic Laplacian on a box in $\mathbb{Z}^d$.

The density of states ρ of a general class of RBM with $W \gg 1$ is given by the well-known Wigner semicircle law (see [3, 16]):

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

The main feature of RBM is that they can be used to model the celebrated Anderson metal-insulator phase transition in $d \geq 3$. Moreover, the crossover for RBM can be investigated even in $d = 1$ by varying the bandwidth W.

More precisely, the key physical parameter of RBM is the localization length ℓ_ψ, which describes the length scale of the eigenvector $\psi(E)$ corresponding to the energy $E \in (-2, 2)$. The system is called delocalized if for all E in the bulk of spectrum ℓ_ψ is comparable with the system

[*]Institute for Low Temperature Physics, Kharkiv, Ukraine& Karazin Kharkiv National University, Kharkiv, Ukraine, e-mail: shcherbi@ilt.kharkov.ua

[†]Department of Mathematics, Princeton University, Princeton, USA, e-mail: tshcherbyna@princeton.edu. Supported in part by NSF grant DMS-1700009.

size, $\ell_\psi \sim n$, and it is called localized otherwise. Delocalized systems correspond to electric conductors, and localized systems are insulators.

In the case of 1d RBM there is a fundamental conjecture stating that for every eigenfunction $\psi(E)$ in the bulk of the spectrum ℓ_ψ is of order W^2 (see [5, 14]). In $d = 2$, the localization length is expected to be exponentially large in W, in $d \geq 3$ it is expected to be macroscopic, $\ell_\psi \sim n$, i.e. system is delocalized (for more details on these conjectures see [24]).

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 $O(N^{-1})$. According to the Wigner – Dyson universality conjecture, this local behaviour 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). In terms of eigenvalue statistics the conjecture about the localization length of RBM in $d = 1$ means that 1d RBM in the bulk of the spectrum changes the spectral local behaviour of random operator type with Poisson local eigenvalue statistics (for $W \ll \sqrt{N}$) to the local spectral behaviour of the GUE/GOE type (for $W \gg \sqrt{N}$).

The conjecture supported by physical derivation due to Fyodorov and Mirlin (see [14]) based on supersymmetric formalism, and also by the so-called Thouless scaling. However, there are only a few partial results on the mathematical level of rigour. At the present time only some upper and lower bounds for ℓ_ψ for the general class of 1d RBM are proved rigorously. It is known from the paper [18] that $\ell_\psi \leq W^8$. Recently this bound was improved in [17] to W^7. On the other side, for the general Wigner matrices (i.e. $W = n$) the bulk universality has been proved in [12, 25], which gives $\ell_\psi \geq W$. By a development of the Erdős-Yau approach, there were also obtained some other results, where the localization length is controlled in a rather weak sense, i.e. the estimates hold for *"most"* eigenfunctions ψ only: $\ell_\psi \geq W^{7/6}$ in [10] and $\ell_\psi \geq W^{5/4}$ in [11]. GUE/GOE gap distributions for $W \sim n$ was proved recently in [4].

The study of the decay of eigenfunctions is closely related to properties of the Green function $(H - E - i\varepsilon)^{-1}$ with a small ε. For instance, if $(H - E - i\varepsilon)^{-1}_{ii}$ (without expectation) is bounded for all i and some $E \in (-2, 2)$, then the normalized eigenvector $\psi(E)$ of H is delocalized on scale ε^{-1} in a sense that

$$\max_i |\psi_i(E)|^2 \lesssim \varepsilon,$$

and so ψ is supported on at least ε^{-1} sites. In particular, if $(H - E - i\varepsilon)^{-1}_{ii}$ can be controlled down to the scale $\varepsilon \sim 1/N$, then the system is in the complete delocalized regime. Moreover, in view of the bound

$$\mathbb{E}\{|(H - E - i\varepsilon)^{-1}_{jk}|^2\} \sim C\varepsilon^{-1}\, e^{-\|j-k\|/\ell}$$

which is supposed to be valid for localized regime, the problem of localization/delocalization reduces to controlling

$$\mathbb{E}\{|(H - E - i\varepsilon)^{-1}_{jk}|^2\}$$

for $\varepsilon \sim 1/N$. As will be shown below, similar estimates of $\mathbb{E}\{|\mathrm{Tr}\,(H - E - i\varepsilon)^{-1}|^2\}$ for $\varepsilon \sim N^{-1}$ are required to work with the correlation functions of RBM.

Despite many attempts, such control has not been achieved so far. The standard approaches of [12] and [11] do not seem to work for $\varepsilon \leq W^{-1}$, and so cannot give an information about the strong form of delocalization (i.e. for *all* eigenfunctions). Classical moment methods, even with a delicate renormalization approach [23], could not break the barrier $\varepsilon \sim W^{-1}$ either.

Another method, which allows to work with random operators with non-trivial spatial structures, is supersymmetry techniques (SUSY) based on the representation of the determinant as an integral over the Grassmann variables. Combining this representation with the representation of the inverse determinant as an integral over the Gaussian complex field, SUSY allows to obtain an integral representation for the main spectral characteristics (such as density of states, second correlation functions, or the average of an elements of the resolvent) 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). For instance, according to the properties of the Stieljes transform, the second correlation function R_2 defined by the equality

$$\mathbf{E}\Big\{\sum_{j_1\neq j_2}\varphi(\lambda_{j_1},\lambda_{j_2})\Big\}=\int_{\mathbb{R}^2}\varphi(\lambda_1,\lambda_2)R_2(\lambda_1,\lambda_2)d\lambda_1 d\lambda_2, \tag{1.2}$$

where $\{\lambda_j\}$ are eigenvalues of a random matrix, the function $\varphi:\mathbb{R}^2\to\mathbb{C}$ is bounded, continuous and symmetric in its arguments, and the summation is over all pairs of distinct integers $j_1,j_2\in\{1,\dots,N\}$, can be rewritten as follows

$$\begin{aligned}R_2(\lambda_1,\lambda_2)=&(\pi N)^{-2}\lim_{\varepsilon\to 0}\mathbb{E}\{\Im\,\mathrm{Tr}\,(H-\lambda_1-i\varepsilon)^{-1}\Im\,\mathrm{Tr}\,(H-\lambda_2-i\varepsilon)^{-1}\} \\ =&(2i\pi N)^{-2}\lim_{\varepsilon\to 0}\mathbb{E}\Big\{\Big(\mathrm{Tr}\,(H-\lambda_1-i\varepsilon)^{-1}-\mathrm{Tr}\,(H-\lambda_1+i\varepsilon)^{-1}\Big) \\ &\times\Big(\mathrm{Tr}\,(H-\lambda_2-i\varepsilon)^{-1}-\mathrm{Tr}\,(H-\lambda_2+i\varepsilon)^{-1}\Big)\Big\},\end{aligned} \tag{1.3}$$

and since

$$\mathbb{E}\{\mathrm{Tr}\,(H-z_1)^{-1}\mathrm{Tr}\,(H-z_2)^{-1}\}=\frac{d^2}{dz_1'dz_2'}\mathbb{E}\Big\{\frac{\det(H-z_1)\det(H-z_2))}{\det(H-z_1')\det(H-z_2'))}\Big\}\Big|_{z'=z}, \tag{1.4}$$

R_2 can be represented as a sum of derivatives of the expectation of the ratio of four determinants, which we will call the generalized correlation function.

The derivation of SUSY integral representation is basically an algebraic step, and usually can be done by the standard algebraic manipulations. SUSY is widely used in the physics literature, but the rigorous analysis of the obtained integral representation is a real mathematical challenge. Usually it is quite difficult, and it requires a powerful analytic and statistical mechanics techniques, such as a saddle point analysis, transfer operators, cluster expansions, renormalization group methods, etc. However, it can be done rigorously for some special class of RBM. For instance, by using SUSY the detailed information about the averaged density of states of a special case of Gaussian RBM in dimension 3 including local semicircle low at arbitrary short scales and smoothness in energy (in the limit of infinite volume and fixed large band width W) was obtained in [7]. The techniques of that paper were used in [6] to obtain the same result in 2d. A similar result in 1d was obtained by the SUSY transfer matrix approach in [19]. Moreover, by applying the SUSY approach in [21], [20] the crossover in this model (in 1d) was proved for the correlation functions of characteristic polynomials. In addition, the rigorous application of SUSY to the Gaussian RBM which has the special block-band structure was developed in [22], where the universality of the bulk local regime for $W\sim n$ was proved. The block band matrices are the special class of Wegner's orbital models (see [27]), i.e. Hermitian matrices H_N with complex zero-mean random Gaussian entries $H_{jk,\alpha\beta}$, where $j,k\in\Lambda=[1,n]^d\cap\mathbb{Z}^d$ (they parameterize the lattice sites) and $\alpha,\gamma=1,\dots,W$ (they parameterize the orbitals on each site), such that

$$\langle H_{j_1k_1,\alpha_1\gamma_1}H_{j_2k_2,\alpha_2\gamma_2}\rangle=\delta_{j_1k_2}\delta_{j_2k_1}\delta_{\alpha_1\gamma_2}\delta_{\gamma_1\alpha_2}J_{j_1k_1} \tag{1.5}$$

with

$$J = 1/W + \beta\Delta/W, \tag{1.6}$$

where $W \gg 1$ and Δ is the discrete Laplacian on Λ. The probability law of H_N can be written in the form

$$P_N(dH_N) = \exp\Big\{ -\frac{1}{2}\sum_{j,k\in\Lambda}\sum_{\alpha,\gamma=1}^{W}\frac{|H_{jk,\alpha\gamma}|^2}{J_{jk}}\Big\}dH_N. \tag{1.7}$$

Combining the approach of [22] with Green's function comparison strategy the delocalization (in a strong sense) for $W \gg n^{6/7}$ has been proved in [1] for the block band matrices (1.5) with rather general non-Gaussian element's distribution.

As it was mentioned above, the main advantage of SUSY techniques is that the main spectral characteristics of the model (1.5) – (1.6) such as a density of states, R_2, $\mathbb{E}\{|G_{jk}(E+i\varepsilon)|^2\}$, etc. can be expressed via SUSY as the averages of certain observables in nearest-neighbour statistical mechanics models on Λ. This in particular in 1d case 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 [9]) and on a heuristic level it was adapted specifically for RBM in [15] (see also references therein), although its rigorous application to the main spectral characteristics is quite difficult due to the complicated structure and non self-adjointness of the corresponding transfer operator. The rigorous application of this method to the density of states and correlation function of characteristic polynomials was done in [19], [20]. In this paper we make the next step in the developing of this approach and apply the technique 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 (± 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 detailes see, e.g., [24]). The sigma-model approximation for RBM was introduced by Efetov (see [9]), and the spins there are 4×4 matrices with both complex and Grassmann entries (this approximation was studied in [14], [15]). Let us mention also that the average conductance for 1d Efetov's sigma-model for RBM was computed in [8]. The aim of this paper is to derive the sigma-model approximation for the second correlation function for RBM and then analyse it rigorously in the dimension one by the transfer matrix formalism.

The mechanism of the crossover for the sigma-model is essentially the same as for the correlation functions of characteristic polynomials (see [20]). It is based on the fact that the spectral gap between two largest eigenvalues of the transfer operator is β^{-1} (it corresponds to W^{-2} in [20]). This implies that for $n/\beta \gg 1$ the n-th degree of the transfer operator converges to the rank one projection on the eigenvector corresponding to the largest eigenvalue, while for $n/\beta \ll 1$ the n-th degree of the transfer operator behaves like the multiplication operator. But the structure of the transfer operator for the sigma-model is more complicated: now it is a 6×6 matrix kernel whose entries are kernels depending on two unitary 2×2 matrices U, U' and two hyperbolic 2×2 matrices S, S'. Hence the spectral analysis in the case of sigma-model is much more involved (see Section 5). We would like to mentioned that in the case of the second generalized correlation function of the 1d block band matrices ((1.5)-(1.6) with $\beta = \alpha W$), the transfer operator becomes 70×70 matrix, whose spectral analysis provides serious structural problems. Thus the analysis of the sigma-model approximation is an important intermediate step.

Set

$$z_1 = E + i\varepsilon/N + \xi_1/N\rho(E), \quad z_2 = E + i\varepsilon/N + \xi_2/N\rho(E), \tag{1.8}$$
$$z_1' = E + i\varepsilon/N + \xi_1'/N\rho(E), \quad z_2' = E + i\varepsilon/N + \xi_2'/N\rho(E),$$

where $E \in (-2,2)$, $\varepsilon > 0$, $\rho(E)$ is defined in (1.1), and $\xi_1, \xi_2, \xi_1', \xi_2' \in [-C, C] \subset \mathbb{R}$ and define

$$\mathcal{R}^{+-}_{Wn\beta}(E,\varepsilon,\xi) = \mathbf{E}\Big\{\frac{\det(H_N - z_1)\det(H_N - \overline{z}_2)}{\det(H_N - z_1')\det(H_N - \overline{z}_2')}\Big\}, \tag{1.9}$$
$$\mathcal{R}^{++}_{Wn\beta}(E,\varepsilon,\xi) = \mathbf{E}\Big\{\frac{\det(H_N - z_1)\det(H_N - z_2)}{\det(H_N - z_1')\det(H_N - z_2')}\Big\}$$

for $\xi = (\xi_1, \xi_2, \xi_1', \xi_2')$.

To derive the sigma-model approximation for the model (1.5) – (1.6), we take β in (1.6) of order $1/W$, i.e. put

$$J = 1/W + \beta\Delta/W^2, \quad \beta > 0. \tag{1.10}$$

The main result states that in the model (1.10) with fixed β and $|\Lambda|$, and with $W \to \infty$, the correlators $\mathcal{R}^{+-}_{Wn\beta}$ and $\mathcal{R}^{++}_{Wn\beta}$ of (1.9) converge to the values given by the sigma-model approximation. More precisely, we get

Theorem 1.1. *Given $\mathcal{R}^{+-}_{Wn\beta}$ of (1.9) ,(1.5) and (1.10), with any dimension d, any fixed β, $|\Lambda|$, $\varepsilon > 0$, and $\xi = (\xi_1, \bar{\xi}_2, \xi_1', \bar{\xi}_2') \in \mathbb{C}^4$ $(|\Im \xi_j| < \varepsilon \cdot \rho(E)/2)$ we have, as $W \to \infty$:*

$$\mathcal{R}^{+-}_{Wn\beta}(E,\varepsilon,\xi) \to \mathcal{R}^{+-}_{n\beta}(E,\varepsilon,\xi), \quad \frac{\partial^2 \mathcal{R}^{+-}_{Wn\beta}}{\partial\xi_1'\partial\xi_2'}(E,\varepsilon,\xi) \to \frac{\partial^2 \mathcal{R}^{+-}_{n\beta}}{\partial\xi_1'\partial\xi_2'}(E,\varepsilon,\xi), \tag{1.11}$$
$$\textit{where} \quad \mathcal{R}^{+-}_{n\beta}(E,\varepsilon,\xi) = C_{E,\xi}\int \exp\Big\{\frac{\tilde{\beta}}{4}\sum \operatorname{Str} Q_j Q_{j-1} - \frac{c_0}{2|\Lambda|}\sum \operatorname{Str} Q_j \Lambda_{\xi,\varepsilon}\Big\} dQ,$$

$\tilde{\beta} = (2\pi\rho(E))^2\beta$, $U_j \in \mathring{U}(2)$, $S_j \in \mathring{U}(1,1)$ (see notation (1.19) below),

$$C_{E,\xi} = e^{E(\xi_1+\xi_2-\xi_1'-\xi_2')/2\rho(E)},$$

and Q_j are 4×4 supermatrices with commuting diagonal and anticommution off-diagonal 2×2 blocks

$$Q_j = \begin{pmatrix} U_j^* & 0 \\ 0 & S_j^{-1} \end{pmatrix} \begin{pmatrix} (I + 2\hat{\rho}_j\hat{\tau}_j)L & 2\hat{\tau}_j \\ 2\hat{\rho}_j & -(I - 2\hat{\rho}_j\hat{\tau}_j)L \end{pmatrix} \begin{pmatrix} U_j & 0 \\ 0 & S_j \end{pmatrix}, \tag{1.12}$$

$$dQ = \prod dQ_j, \quad dQ_j = (1 - 2n_{j,1}n_{j,2})\, d\rho_{j,1}d\tau_{j,1}\, d\rho_{j,2}d\tau_{j,2}\, dU_j\, dS_j$$

with

$$n_{j,1} = \rho_{j,1}\tau_{j,1}, \quad n_{j,2} = \rho_{j,2}\tau_{j,2},$$
$$\hat{\rho}_j = \operatorname{diag}\{\rho_{j1}, \rho_{j2}\}, \quad \hat{\tau}_j = \operatorname{diag}\{\tau_{j1}, \rho_{j2}\}, \quad L = \operatorname{diag}\{1, -1\}$$

Here $\rho_{j,l}$, $\tau_{j,l}$, $l = 1,2$ are anticommuting Grassmann variables,

$$\operatorname{Str}\begin{pmatrix} A & \sigma \\ \eta & B \end{pmatrix} = \operatorname{Tr} A - \operatorname{Tr} B,$$

and

$$\Lambda_{\xi,\varepsilon} = \operatorname{diag}\{\varepsilon - i\xi_1/\rho(E), -\varepsilon - i\xi_2/\rho(E), \varepsilon - i\xi_1'/\rho(E), -\varepsilon - i\xi_2'/\rho(E)\}.$$

Theorem 1.2. *Given $\mathcal{R}^{++}_{Wn\beta}$ of (1.9) ,(1.5) and (1.10), with any dimension d, any fixed β, $|\Lambda|$, $\varepsilon > 0$, and $\xi = (\xi_1, \xi_2, \xi_1', \xi_2') \in \mathbb{C}^4$ $(|\Im \xi_j| < \varepsilon \cdot \rho(E)/2)$ we have, as $W \to \infty$:*

$$\mathcal{R}^{++}_{Wn\beta}(E, \varepsilon, \xi) \to e^{ia_+(\xi_1'+\xi_2'-\xi_1-\xi_2)/\rho(E)}, \tag{1.13}$$
$$\frac{\partial^2 \mathcal{R}^{++}_{Wn\beta}}{\partial \xi_1' \partial \xi_2'}(E, \varepsilon, \xi) \to -a_+^2/\rho^2(E) \cdot e^{ia_+(\xi_1'+\xi_2'-\xi_1-\xi_2)/\rho(E)}, \qquad a_+ = (iE + \sqrt{4 - E^2})/2.$$

Note that $Q_j^2 = I$ for Q_j of (1.12) and so the integral in the r.h.s of (1.11) is a sigma-model approximation similar to Efetov's one (see [9]).

The next theorem describes the behaviour of $\mathcal{R}^{+-}_{n\beta}(E, \varepsilon, \xi)$ of the sigma-model (1.11) in the regime $n \to \infty$, $\beta > Cn \log^2 n$:

Theorem 1.3. *If $n, \beta \to \infty$ in such a way that $\beta > Cn \log^2 n$, then for any fixed $\varepsilon > 0$ and $\xi = (\xi_1, \xi_2, \xi_1', \xi_2') \in \mathbb{C}^4$ $(|\Im \xi_j| < \varepsilon \cdot \rho(E)/2)$ we have*

$$\mathcal{R}^{+-}_{n\beta} \to C_{E,\xi} \cdot e^{-c_0(\alpha_1+\alpha_2)}\Big(\delta_1\delta_2(e^{2c_0\alpha_1} - 1)/\alpha_1\alpha_2 - (\delta_1 + \delta_2)e^{2c_0\alpha_1}/\alpha_2 + e^{2c_0\alpha_1}\alpha_1/\alpha_2\Big), \tag{1.14}$$
$$\text{where} \quad \alpha_1 = \varepsilon - i(\xi_1 - \xi_2)/2\rho(E), \quad \alpha_2 = \varepsilon - i(\xi_1' - \xi_2')/2\rho(E), \tag{1.15}$$
$$\delta_1 = i(\xi_1' - \xi_1)/2\rho(E), \quad \delta_2 = i(\xi_2 - \xi_2')/2\rho(E).$$

Now Theorems 1.1 – 1.3 and (1.3) – (1.4) imply the main result of the paper:

Theorem 1.4. *In the dimension $d = 1$ the behavior of the sigma-model approximation of the second order correlation function (1.2) of (1.5), as $\beta \gg n$, in the bulk of the spectrum coincides with those for the GUE. More precisely, if $\Lambda = [1, n] \cap \mathbb{Z}$ and H_N, $N = Wn$ are matrices (1.5) with J of (1.10), then for any $|E| < \sqrt{2}$*

$$(N\rho(E))^{-2} R_2\left(E + \frac{\xi_1}{\rho(E)\,N}, E + \frac{\xi_2}{\rho(E)\,N}\right) \longrightarrow 1 - \frac{\sin^2(\pi(\xi_1 - \xi_2))}{\pi^2(\xi_1 - \xi_2)^2}, \tag{1.16}$$

in the limit first $W \to \infty$, and then $\beta, n \to \infty$, $\beta \geq Cn \log^2 n$.

Remark 1.1. *Notice that to prove universality of bulk local regime from the delocalization side of random block band matrices (1.5) – (1.7) without a sigma-model approximation one have to take J of (1.6), fix β, and prove (1.16) in the limit $W, n \to \infty$, $W \gg n$, which is different from the asymptotic regime considered in the current paper (first $W \to \infty$ with fixed β, then $\beta \gg n$, $\beta, n \to \infty$).*

The paper is organized as follows. In Section 2 we obtain a convenient SUSY integral representation for $\mathcal{R}^{+-}_{Wn\beta}$ and $\mathcal{R}^{++}_{Wn\beta}$ of (1.9). In Section 3 we prove Theorems 1.1 and 1.2, in Section 4 we derive Theorem 1.4 from Theorems 1.2 and 1.3, in Section 5 we prove Theorem 1.3 modulo some auxiliary result proven in Section 6.

Acknowledgement. We are grateful to Yan Fyodorov for his suggestion of this particular model for the derivation of sigma-model approximation for RBM. TS would like to thank Tom Spencer for his explanation of the nature of sigma-model approximation and for many fruitful discussions without that this paper would never have been written.

1.1 Notation

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

Moreover,

- $N = W|\Lambda|$;
- indices i, j, k vary in Λ and correspond to the number of the site (or the number of the block), index l is always 1 or 2 (this is the field index), and Greek indices α, γ vary from 1 to W and correspond to the position of the element in the block;
- variables ϕ and Φ with different indices are complex variables or vectors correspondingly; if x_j means some variable (vector or matrix) which corresponds to the site $j \in \Lambda$, then x means vector $\{x_j\}_{j\in\Lambda}$, $dx = \prod dx_j$, and dx_j means the product of the differentials which correspond to functionally independent coefficients of x_j;
- variables ψ, Ψ, ρ, and τ with different indices are Grassmann variables or vectors or matreces correspondingly; if ρ_j corresponds to the site $j \in \Lambda$, then ρ means vector $\{\rho_j\}_{j\in\Lambda}$, $d\rho = \prod d\rho_j$, and $d\rho_j$ means the product of the differentials which correspond the components (for vectors) or entries (for matrices) taken into the lexicographic order;
- $$a_\pm = \frac{iE \pm \sqrt{4 - E^2}}{2}, \quad c_\pm = 1 + a_\pm^{-2}, \quad c_0 = \sqrt{4 - E^2} = 2\pi\rho(E); \tag{1.17}$$
  $$L = \mathrm{diag}\,\{1, -1\}, \quad L_\pm = \mathrm{diag}\,\{a_+, a_-\}; \tag{1.18}$$
- $$\mathring{U}(2) = U(2)/U(1) \times U(1), \quad \mathring{U}(1,1) = U(1,1)/U(1) \times U(1), \tag{1.19}$$

  where $U(p)$ is a group of $p \times p$ unitary matrices, and $U(1,1)$ is a group of 2×2 hyperbolic matrices S such that $S^* LS = L$;
- $$\mathcal{L}_\pm(E) = \Big\{ r\Big(iE/2 \pm \sqrt{4 - E^2}/2\Big) | r \in [0, +\infty)\Big\}; \tag{1.20}$$
- $$\tilde{\beta} = c_0^2\,\beta; \tag{1.21}$$
- $$Z_1 = E \cdot I + i\varepsilon \cdot L/N + \hat{\xi}/N\rho(E), \quad Z_2 = E \cdot I + i\varepsilon \cdot L/N + \hat{\xi}'/N\rho(E), \tag{1.22}$$
  $$Z_1^+ = E \cdot I + i\varepsilon \cdot I/N + \hat{\xi}/N\rho(E), \quad Z_2^+ = E \cdot I + i\varepsilon \cdot I/N + \hat{\xi}'/N\rho(E), \tag{1.23}$$
  $$\hat{\xi} = \mathrm{diag}\{\xi_1, \xi_2\}, \quad \hat{\xi}' = \mathrm{diag}\{\xi_1', \xi_2'\}. \tag{1.24}$$

2 Integral representations

In this section we perform the standard algebraic manipulations to obtain an integral representation for the determinant ratio $\mathcal{R}^{+-}_{Wn\beta}(E, \varepsilon, \xi)$ of (1.9).

Proposition 2.1. *For any dimension d, the determinant ratio $\mathcal{R}^{+-}_{Wn\beta}(E,\varepsilon,\xi)$ of (1.9) can be written as follows:*

$$\mathcal{R}^{+-}_{Wn\beta}(E,\varepsilon,\xi) = \frac{\det^2 J\cdot(-1)^{|\Lambda|W}}{(2\pi^3)^{|\Lambda|}\big((W-1)!(W-2)!\big)^{|\Lambda|}}\int dXdY\cdot\exp\{i\sum_{j\in\Lambda}\mathrm{Tr}\,Y_jLZ_2\} \tag{2.1}$$
$$\times\exp\Big\{-\frac{1}{2}\sum_{j,k\in\Lambda}J_{jk}\mathrm{Tr}\,(Y_jL)(Y_kL)-\frac{1}{2}\sum_{j,k\in\Lambda}(J^{-1})_{jk}\mathrm{Tr}\,X_jX_k\Big\}$$
$$\times\det\big\{J^{-1}_{jk}\mathbf{1}_4-\delta_{jk}(iZ_1+X_j)^{-1}\otimes(Y_jL)\big\}_{j,k\in\Lambda}\prod_{j\in\Lambda}\frac{\det^W(iZ_1+X_j)\det^W Y_j}{\det^2 Y_j},$$

where $\{X_j\}_{j\in\Lambda}$ are Hermitian 2×2 matrices with standard dX_j , $\{Y_j\}_{j\in\Lambda}$ are 2×2 positive Hermitian matrices with dY_j of Proposition 7.1, and $Z_{1,2}$ are defined in (1.22).

A similar formula is valid for $\mathcal{R}^{++}_{Wn\beta}(E,\varepsilon,\xi)$ with Y_j instead of Y_jL and Z_l^+ instead of Z_l, $l=1,2$ (see (1.23)).

Proof. Introduce complex and Grassmann fields:

$$\Phi_l=\{\phi_{jl}\}^t_{j\in\Lambda},\quad \phi_{jl}=(\phi_{jl1},\phi_{jl2},\ldots,\phi_{jlW}),\quad l=1,2,\quad -\quad \text{complex},$$
$$\Psi_l=\{\psi_{jl}\}^t_{j\in\Lambda},\quad \psi_{jl}=(\psi_{jl1},\psi_{jl2},\ldots,\psi_{jlW}),\quad l=1,2,\quad -\quad \text{Grassmann}.$$

Using (7.3) – (7.4) (see Appendix) we can write

$$\mathcal{R}^{+-}_{Wn\beta}(E,\varepsilon,\xi)=\pi^{-2W|\Lambda|}\mathbf{E}\Big\{\int\exp\{i\Psi_1^+(z_1'-H_N)\Psi_1-i\Psi_2^+(\overline{z}_2'-H_N)\Psi_2\}$$
$$\times\exp\{i\Phi_1^+(z_1-H_N)\Phi_1-i\Phi_2^+(\overline{z}_2-H_N)\Phi_2\}d\Phi d\Psi\Big\}$$
$$=\int d\Phi d\Psi\ \exp\Big\{i(z_1'\Psi_1^+\Psi_1+z_1\Phi_1^+\Phi_1)-i(\overline{z}_2'\Psi_2^+\Psi_2+\overline{z}_2\Phi_2^+\Phi_2)\Big\}$$
$$\times\mathbf{E}\Big\{\exp\Big\{-\sum_{j\le k}\sum_{\alpha,\gamma}\Big(i\Re H_{jk,\alpha\gamma}\chi^+_{jk,\alpha\gamma}-\Im H_{jk,\alpha\gamma}\chi^-_{jk,\alpha\gamma}\Big)\Big\}\Big\},$$

where z_l, z_l' are defined in (1.8),

$$\chi^{\pm}_{jk,\alpha\gamma}=\eta_{jk,\alpha\gamma}\pm\eta_{kj,\gamma\alpha},$$
$$\eta_{jk,\alpha\gamma}=\overline{\psi}_{j1\alpha}\psi_{k1\gamma}-\overline{\psi}_{j2\alpha}\psi_{k2\gamma}+\overline{\phi}_{j1\alpha}\phi_{k1\gamma}-\overline{\phi}_{j2\alpha}\phi_{k2\gamma},$$
$$\eta_{jj,\alpha\alpha}=(\overline{\psi}_{j1\alpha}\psi_{j1\alpha}-\overline{\psi}_{j2\alpha}\psi_{j2\alpha}+\overline{\phi}_{j1\alpha}\phi_{j1\alpha}-\overline{\phi}_{j2\alpha}\phi_{j2\alpha})/2.$$

Averaging over (1.7), we get

$$\mathcal{R}^{+-}_{Wn\beta}(E,\varepsilon,\xi)=\pi^{-2W|\Lambda|}\int d\Phi d\Psi\ \exp\Big\{i(z_1'\Psi_1^+\Psi_1+z_1\Phi_1^+\Phi_1)-i(\overline{z}_2'\Psi_2^+\Psi_2+\overline{z}_2\Phi_2^+\Phi_2)\Big\}$$
$$\times\exp\Big\{-\sum_{j<k}\sum_{\alpha,\gamma}J_{jk}\ \eta_{jk,\alpha\gamma}\eta_{kj,\gamma\alpha}-\frac{1}{2}\sum_{j,\alpha}J_{jj}\ \eta^2_{jj,\alpha\alpha}\Big\}.$$

Thus,

$$\mathcal{R}^{+-}_{Wn\beta}(E,\varepsilon,\xi)=\pi^{-2W|\Lambda|}\int d\Phi d\Psi\ \exp\Big\{i\sum_{j\in\Lambda}\mathrm{Tr}\,\tilde{X}_jLZ_1+i\sum_{j\in\Lambda}\mathrm{Tr}\,\tilde{Y}_jLZ_2\Big\}$$
$$\times\exp\Big\{\frac{1}{2}\sum_{j,k\in\Lambda}J_{jk}\mathrm{Tr}\,(\tilde{X}_jL)(\tilde{X}_kL)-\frac{1}{2}\sum_{j,k\in\Lambda}J_{jk}\mathrm{Tr}\,(\tilde{Y}_jL)(\tilde{Y}_kL)\Big\} \tag{2.2}$$
$$\times\exp\Big\{-\sum_{j,k\in\Lambda}J_{jk}\big(\overline{\psi}_{j1}\psi_{k1}(\overline{\phi}_{k1}\phi_{j1}-\overline{\phi}_{k2}\phi_{j2})+\overline{\psi}_{j2}\psi_{k2}(\overline{\phi}_{k2}\phi_{j2}-\overline{\phi}_{k1}\phi_{j1})\big)\Big\},$$

where L, $Z_{1,2}$ are defined in (1.18), (1.22), and

$$\tilde{X}_j = \left(\begin{array}{cc} \psi^+_{j1}\psi_{j1} & \psi^+_{j1}\psi_{j2} \\ \psi^+_{j2}\psi_{j1} & \psi^+_{j2}\psi_{j2} \end{array} \right), \quad \tilde{Y}_j = \left(\begin{array}{cc} \phi^+_{j1}\phi_{j1} & \phi^+_{j1}\phi_{j2} \\ \phi^+_{j2}\phi_{j1} & \phi^+_{j2}\phi_{j2} \end{array} \right).$$

Using the standard Hubbard-Stratonovich transformation, we obtain

$$\begin{aligned}(2\pi^2)^{|\Lambda|}\det{}^2 J \cdot \exp\Big\{\frac{1}{2}\sum_{j,k\in\Lambda} J_{jk}\mathrm{Tr}\,(\tilde{X}_jL)(\tilde{X}_kL)\Big\} \\ = \int \exp\Big\{ -\frac{1}{2}\sum_{j,k\in\Lambda}(J^{-1})_{jk}\mathrm{Tr}\,X_jX_k + \sum_{j\in\Lambda}\mathrm{Tr}\,X_j\big(\tilde{X}_jL\big)\Big\}dX, \end{aligned} \quad (2.3)$$

where X_j are 2×2 Hermitian matrices with the standard measure dX_j.

Substituting (2.3) to (2.2) and integrating over $d\Psi$ (see (7.4)), we get

$$\begin{aligned}\mathcal{R}^{+-}_{Wn\beta}(E,\varepsilon,\xi) = \frac{\det^{-2}J}{\big(2\pi^{2(1+W)}\big)^{|\Lambda|}}\int \exp\Big\{i\sum_{j\in\Lambda}\mathrm{Tr}\,\tilde{Y}_jLZ_2 - \frac{1}{2}\sum_{j,k\in\Lambda}J_{jk}\mathrm{Tr}\,(\tilde{Y}_jL)(\tilde{Y}_kL)\Big\} \\ \times \exp\Big\{ -\frac{1}{2}\sum_{j,k\in\Lambda}(J^{-1})_{jk}\mathrm{Tr}\,X_jX_k\Big\}\cdot \det M \cdot\, d\Phi\; dX\end{aligned} \quad (2.4)$$

with $M = M^{(1)} - M^{(2)}$, where $M^{(1)}$ and $M^{(2)}$ are $2W|\Lambda|\times 2W|\Lambda|$ matrices with entries

$$\begin{aligned} M^{(1)}_{j\alpha l,k\gamma l'} &= \delta_{jk}\delta_{\alpha\gamma}(iZ_1+X_j)_{ll'}L_{ll}, \quad j,k\in\Lambda,\ \alpha,\gamma=1,\dots,W,\ l,l'=1,2, \\ M^{(2)}_{j\alpha l,k\gamma l'} &= J_{jk}\delta_{ll'}L_{ll}\sum_{\nu=1}^{2}\varphi_{j\alpha\nu}\overline{\varphi}_{k\gamma\nu}L_{\nu\nu}. \end{aligned} \quad (2.5)$$

We can rewrite

$$\det M = \det M^{(1)}\cdot \det\Big(1 - \big(M^{(1)}\big)^{-1}M^{(2)}\Big) =: \det M^{(1)}\cdot\det\Big(1-\mathcal{M}\Big) \quad (2.6)$$

with

$$\mathcal{M}_{j\alpha l,k\gamma l'} = J_{jk}(iZ_1+X_j)^{-1}_{ll'}\sum_{\nu=1}^{2}\varphi_{j\alpha\nu}\overline{\varphi}_{k\gamma\nu}L_{\nu\nu}. \quad (2.7)$$

Note that $\mathcal{M} = AB$, where

$$\begin{aligned} A_{j\alpha l,k\sigma l'} &= J_{jk}(iZ_1+X_j)^{-1}_{ll'}\,\varphi_{j\alpha\sigma}, \quad j,k\in\Lambda,\ \alpha,\gamma=1,\dots,W,\ l,l',\sigma=1,2, \\ B_{j\sigma l,k\alpha l'} &= \delta_{jk}\delta_{ll'}L_{\sigma\sigma}\,\overline{\varphi}_{k\alpha\sigma}. \end{aligned} \quad (2.8)$$

Therefore, using that $\det(1-AB)=\det(1-BA)$, (2.7), and (2.8), we get

$$\det\Big(1-\mathcal{M}\Big) = \det\Big(1-BA\Big) =: \det\Big(1-\tilde{\mathcal{M}}\Big), \quad (2.9)$$

where

$$\begin{aligned}\tilde{\mathcal{M}}_{j\sigma l,k\sigma' l'} = \sum_{p,\alpha,\nu} B_{j\sigma l,p\alpha\nu}A_{p\alpha\nu,k\sigma' l'} &= J_{jk}(iZ_1+X_j)^{-1}_{ll'}\sum_{\alpha=1}^{W}\overline{\varphi}_{j\alpha\sigma}\varphi_{j\alpha\sigma'}L_{\sigma\sigma} \\ &= J_{jk}(iZ_1+X_j)^{-1}_{ll'}(\tilde{Y}_jL)_{\sigma\sigma'}. \end{aligned} \quad (2.10)$$

This yields

$$\det\Big(1-\tilde{\mathcal{M}}\Big) = \det\big\{\delta_{j,k}\mathbf{1}_4 - J_{j,k}(iZ_1+X_j)^{-1}\otimes(\tilde{Y}_jL)\big\}_{j,k\in\Lambda}$$
$$= \det{}^4 J\cdot\det\big\{J^{-1}_{jk}\mathbf{1}_4 - \delta_{jk}(iZ_1+X_j)^{-1}\otimes(\tilde{Y}_jL)\big\}_{j,k\in\Lambda}. \tag{2.11}$$

Besides,

$$\det M^{(1)} = (-1)^{|\Lambda|W}\prod_{j\in\Lambda}\det{}^W(iZ_1+X_j). \tag{2.12}$$

Now substituting (2.5) – (2.7) and (2.9) – (2.12) to (2.4) and applying the bosonization formula (see Proposition 7.1), we obtain (2.1).

The formula for $\mathcal{R}^{++}_{Wn\beta}(E,\varepsilon,\xi)$ can be obtained by the same way. □

3 Derivation of the sigma-model approximation

3.1 Proof of Theorem 1.1

Let β and $|\Lambda|$ be fixed, and $W\to\infty$.

Defining $|\Lambda|\times|\Lambda|$ matrix R as

$$J^{-1} = W\big(1-\frac{\beta}{W}\triangle+\frac{\beta^2}{W^2}\triangle^2-\dots\big) =: W\big(1-\frac{\beta}{W}\triangle+\frac{1}{W^2}R\big),$$

putting $B_j = W^{-1}Y_jL$, and shifting $iZ_1+X_j\to X_j$, we can rewrite (2.1) of Proposition 2.1 as

$$\mathcal{R}^{+-}_{Wn\beta}(E,\varepsilon,\xi) = Q^{(1)}_{W,|\Lambda|}\int dXdB\cdot\det\Big\{\big(\mathbf{1}-\frac{\beta}{W}\triangle+\frac{1}{W^2}R\big)_{jk}\mathbf{1}_4-\delta_{jk}\cdot X_j^{-1}\otimes B_j\Big\}_{j,k\in\Lambda}$$
$$\times\exp\Big\{-\frac{W}{2}\sum_{j\in\Lambda}\Big(\mathrm{Tr}\,(B_j-iZ_2)^2+\mathrm{Tr}\,(X_j-iZ_1)^2\Big)\Big\}\cdot\prod_{j\in\Lambda}\frac{\det^W X_j\det^W B_j}{\det^2 B_j} \tag{3.1}$$
$$\times\exp\Big\{\frac{\beta}{2}\sum_{j\sim k}\Big(\mathrm{Tr}\,(B_j-B_k)^2-\mathrm{Tr}\,(X_j-X_k)^2\Big)+\frac{1}{2W}\sum_{j,k}R_{jk}\,\mathrm{Tr}\,(X_j-iZ_1)(X_k-iZ_1)\Big\},$$

where

$$Q^{(1)}_{W,|\Lambda|} = \frac{\det^2 J\cdot W^{2(W+1)|\Lambda|}\cdot e^{-W|\Lambda|\mathrm{Tr}\,Z_2^2/2}}{(2\pi^3)^{|\Lambda|}\big((W-1)!(W-2)!\big)^{|\Lambda|}}$$
$$= \frac{W^{4|\Lambda|}\cdot e^{2W|\Lambda|-W|\Lambda|\mathrm{Tr}\,Z_2^2/2}}{(2\pi^2)^{2|\Lambda|}}\cdot\Big(1+O\big(W^{-1}\big)\Big).$$

Change the variables to

$$X_j = U_j^*\hat{X}_jU_j,\quad \hat{X}_j=\mathrm{diag}\,\{x_{j,1},x_{j,2}\},\quad U_j\in\mathring{U}(2),\qquad x_{j,1},x_{j,2}\in\mathbb{R},$$
$$B_j = S_j^{-1}\hat{B}_jS_j,\quad \hat{B}_j=\mathrm{diag}\,\{b_{j,1},b_{j,2}\},\quad S_j\in\mathring{U}(1,1),\quad b_{j,1}\in\mathbb{R}^+,\ b_{j,2}\in\mathbb{R}^-.$$

The Jacobian of such a change is

$$2^{|\Lambda|}(\pi/2)^{2|\Lambda|}\prod_{j\in\Lambda}(x_{j,1}-x_{j,2})^2\prod_{j\in\Lambda}(b_{j,1}-b_{j,2})^2.$$

This and (3.1) yield

$$\mathcal{R}^{+-}_{Wn\beta}(E,\varepsilon,\xi)=Q^{(2)}_{W,|\Lambda|}\int dSdU\int dx\int_{\mathbb{R}_+^{|\Lambda|}\times\mathbb{R}_-^{|\Lambda|}}db\cdot\prod_{j\in\Lambda}\frac{(x_{j,1}-x_{j,2})^2(b_{j,1}-b_{j,2})^2}{b_{j,1}^2b_{j,2}^2} \tag{3.2}$$

$$\times\det\mathcal{D}(\hat{X},\hat{B},U,S)\cdot\exp\Big\{-W\sum_{j\in\Lambda}\sum_{l=1}^{2}(f(x_{j,l})+f(b_{j,l}))\Big\}$$

$$\times\exp\Big\{\frac{\beta}{2}\sum_{j\sim k}\Big(\mathrm{Tr}\,(S_j^{-1}\hat{B}_jS_j-S_k^{-1}\hat{B}_kS_k)^2-\mathrm{Tr}\,(U_j^*\hat{X}_jU_j-U_k^*\hat{X}_kU_k)^2\Big)\Big\}$$

$$\times\exp\Big\{\frac{1}{2W}\sum_{j,k}R_{jk}\,\mathrm{Tr}\,(U_j^*\hat{X}_jU_j-iZ_1)(U_k^*\hat{X}_kU_k-iZ_1)\Big\}$$

$$\times\exp\Big\{\frac{i}{|\Lambda|}\sum_{j\in\Lambda}\Big(\mathrm{Tr}\,U_j^*\hat{X}_jU_j\big(i\varepsilon L+\hat{\xi}/\rho(E)\big)+\mathrm{Tr}\,S_j^{-1}\hat{B}_jS_j\big(i\varepsilon L+\hat{\xi}'/\rho(E)\big)\Big)\Big\},$$

where

$$\det\mathcal{D}(\hat{X},\hat{B},U,S)=\det\Big\{\Big(\mathbf{1}-\frac{\beta}{W}\triangle+\frac{1}{W^2}R\Big)_{jk}\mathbf{1}_4-\delta_{jk}\cdot X_j^{-1}\otimes B_j\Big\}_{j,k\in\Lambda} \tag{3.3}$$

$$=\det\Big\{\delta_{jk}\big(\mathbf{1}-\hat{X}_j^{-1}\otimes\hat{B}_j\big)+\frac{1}{W}\Big(-\beta\triangle+\frac{1}{W}R\Big)_{jk}\cdot U_jU_k^*\otimes S_jS_k^{-1}\Big\}_{j,k\in\Lambda},$$

$$\begin{aligned}Q^{(2)}_{W,|\Lambda|}&=2^{|\Lambda|}(\pi/2)^{2|\Lambda|}\cdot e^{W|\Lambda|(\mathrm{Tr}\,Z_1^2+\mathrm{Tr}\,Z_2^2)/2-W|\Lambda|(2+E^2)}\cdot Q^{(1)}_{W,|\Lambda|}\\&=\frac{W^{4|\Lambda|}\cdot e^{E(\xi_1+\xi_2)/\rho(E)}}{2^{3|\Lambda|}\pi^{2|\Lambda|}}\cdot\Big(1+O\big(W^{-1}\big)\Big),\\f(x)&=x^2/2-iEx-\log x-(2+E^2)/4.\end{aligned}$$

The constant in $f(x)$ is chosen in such a way that $\Re f(a_\pm)=0$. Measures dU_j, dS_j in (3.2) are the Haar measures over $\mathring{U}(2)$ and $\mathring{U}(1,1)$ correspondingly.

Also it is easy to see that for $|E|\le\sqrt{2}$ we can deform the contours of integration as

- for $x_{j,1}$, $x_{j,2}$ to $iE/2+\mathbb{R}$;
- for $b_{j,1}$ to $\mathcal{L}_+(E)$ of (1.20);
- for $b_{j,2}$ to $\mathcal{L}_-(E)$ of (1.20).

To prove Theorem 1.1, we are going to integrate (3.2) over the "fast" variables: $\{x_{j,l}\},\{b_{j,l}\}$, $l=1,2$, $j\in\Lambda$. The first step is the following lemma:

Lemma 3.1. *The integral (3.2) over $\{x_{j,l}\},\{b_{j,l}\}$, $l=1,2$, $j\in\Lambda$ can be restricted to the integral over the $W^{-(1-\kappa)/2}$-neighbourhoods (with a small $\kappa>0$) of the points*

I. $x_{j,1}=a_+$, $x_{j,2}=a_-$ or $x_{j,1}=a_-$, $x_{j,2}=a_+$, $b_{j,1}=a_+$, $b_{j,2}=a_-$ for any $j\in\Lambda$;

II. $x_{j,1}=x_{j,2}=a_+$, $b_{j,1}=a_+$, $b_{j,2}=a_-$ for any $j\in\Lambda$;

III. $x_{j,1}=x_{j,2}=a_-$, $b_{j,1}=a_+$, $b_{j,2}=a_-$ for any $j\in\Lambda$.

Moreover, the contributions of the points II and III are $o(1)$, as $W\to\infty$.

Proof. The proof of the first part of the lemma is straightforward and based on the fact that $\Re f(z)$ for $z = x + iE/2$, $x \in \mathbb{R}$ has two global minimums at $z = a_\pm$, and for $z \in \mathcal{L}_\pm(E)$ has one global minimum at $z = a_\pm$.

To prove the second part of the lemma, consider the neighbourhood of the point II (the point III can be treated in a similar way). Change the variables as

$$\begin{aligned} x_{j,1} &= a_+ + \tilde{x}_{j,1}/\sqrt{W}, & x_{j,2} &= a_+ + \tilde{x}_{j,2}/\sqrt{W}, \\ b_{j,1} &= a_+\big(1 + \tilde{b}_{j,1}/\sqrt{W}\big), & b_{j,2} &= a_-\big(1 + \tilde{b}_{j,2}/\sqrt{W}\big). \end{aligned} \tag{3.4}$$

This gives the Jacobian $(-1)^{|\Lambda|}W^{-2|\Lambda|}$ and also the additional $W^{-|\Lambda|}$ since

$$x_{j,1} - x_{j,2} = (\tilde{x}_{j,1} - \tilde{x}_{j,2})/\sqrt{W}.$$

Together with $Q^{(2)}_{W,|\Lambda|}$ this gives $W^{|\Lambda|}$ in front of the integral (3.2). In addition, expanding f into the series, we get

$$\begin{aligned} f(x_{j,l}) &= f(a_+) + \frac{c_+}{2}\frac{\tilde{x}_{j,l}^2}{W} - \frac{1}{2a_+^3}\frac{\tilde{x}_{j,l}^3}{W^{3/2}} + O\Big(\frac{\tilde{x}_{j,l}^4}{W^2}\Big), \quad l = 1, 2 \\ f(b_{j,1}) &= f(a_+) + \frac{a_+^2 c_+}{2}\cdot\frac{\tilde{b}_{j,1}^2}{W} - \frac{1}{2}\cdot\frac{\tilde{b}_{j,1}^3}{W^{3/2}} + O\Big(\frac{\tilde{b}_{j,1}^4}{W^2}\Big), \\ f(b_{j,2}) &= f(a_-) + \frac{a_-^2 c_-}{2}\cdot\frac{\tilde{b}_{j,2}^2}{W} - \frac{1}{2}\cdot\frac{\tilde{b}_{j,2}^3}{W^{3/2}} + O\Big(\frac{\tilde{b}_{j,2}^4}{W^2}\Big), \end{aligned} \tag{3.5}$$

where

$$c_\pm = 1 + a_\pm^{-2}, \quad f(a_+) = -f(a_-) \in i\mathbb{R}. \tag{3.6}$$

We are going to compute the leading order of the integral over $\{\tilde{x}_{j,l}\}, \{\tilde{b}_{j,l}\}$, $l = 1, 2$, $j \in \Lambda$. To this end, we leave the quadratic part of f (see (3.5)) in the exponent, expand everything else into the series of $\tilde{x}_{j,l}/\sqrt{W}, \tilde{b}_{j,l}/\sqrt{W}$ around the saddle-point $\tilde{x}_{j,l} = \tilde{b}_{j,l} = 0$, and compute the Gaussian integral of each term of this expansion. We are going to prove that all this terms are $o(1)$.

Indeed, consider the expansion of the diagonal elements of $\mathcal{D}(\hat{X}, \hat{B}, U, S)$ of (3.3):

$$\begin{aligned} d_{j,l1} &= 1 - x_{j,l}^{-1}b_{j,1} = (\tilde{x}_{j,l}/a_+ - \tilde{b}_{j,1})/\sqrt{W} + (\tilde{x}_{j,l}\tilde{b}_{j,1}/a_+ - \tilde{x}_{j,l}^2/a_+^2)W + O\big(W^{-3(1-\kappa)/2}\big), \\ d_{j,l2} &= 1 - x_{j,l}^{-1}b_{j,2} = c_- - (\tilde{x}_{j,l}/a_+ - \tilde{b}_{j,2})/a_-^2\sqrt{W} + O\big(W^{-1+\kappa}\big), \quad l = 1, 2. \end{aligned} \tag{3.7}$$

If we rewrite the determinant of $\mathcal{D}(\hat{X}, \hat{B}, U, S)$ in a standard way, then each summand has strictly one element from each row and column. Because of (3.7), each element in the rows $(j, 11)$ and $(j, 21)$ has at least $W^{-1/2}$, and so the expansion of $\det \mathcal{D}(\hat{X}, \hat{B}, U, S)$ starts from $W^{-|\Lambda|}$. Moreover, to obtain $W^{-|\Lambda|}$ (i.e. non-zero contribution) we must consider the summands of the determinant expansion that have only diagonal elements $d_{j,ls}$ (since non-diagonal elements of $\mathcal{D}(\hat{X}, \hat{B}, U, S)$ are $O(W^{-1})$ or less), and furthermore only the first terms in the expansions (3.7) and all other function in (3.2). Thus we get

$$C \cdot \Big\langle \prod_{j\in\Lambda} \frac{\tilde{x}_{j,1}/a_+ - \tilde{b}_{j,1}}{\sqrt{W}} \cdot \frac{\tilde{x}_{j,2}/a_+ - \tilde{b}_{j,1}}{\sqrt{W}} \cdot (\tilde{x}_{j,1} - \tilde{x}_{j,2})^2 \Big\rangle_{++} + o(1), \tag{3.8}$$

where

$$\Big\langle\,\cdot\,\Big\rangle_{++}=\int\Big(\,\cdot\,\Big)\exp\Big\{-\frac{1}{2}\sum_{j\in\Lambda}\Big(c_+(\tilde{x}_{j,1}^2+\tilde{x}_{j,2}^2)+a_+^2c_+\tilde{b}_{j,1}^2+a_-^2c_-\tilde{b}_{j,2}^2\Big)\Big\}d\tilde{x}\,d\tilde{b}.$$

But it is easy to see that the Gaussian integral in (3.8) is zero, which completes the proof of the lemma. □

According to Lemma 3.1 the main contribution to (3.2) is given by the neighbourhoods of the saddle points $x_{j,1}=a_+$, $x_{j,2}=a_-$ or $x_{j,1}=a_-$, $x_{j,2}=a_+$. All such points can be obtained from each other by rotations of U_j, so we can consider only $x_{j,1}=a_+$, $x_{j,2}=a_-$ for all $j\in\Lambda$. Similarly to the proof of Lemma 3.1, change variables as

$$\begin{array}{ll} x_{j,1}=a_++\tilde{x}_{j,1}/\sqrt{W}, & x_{j,2}=a_-+\tilde{x}_{j,2}/\sqrt{W},\\ b_{j,1}=a_+\big(1+\tilde{b}_{j,1}/\sqrt{W}\big), & b_{j,2}=a_-\big(1+\tilde{b}_{j,2}/\sqrt{W}\big). \end{array}\tag{3.9}$$

That slightly change the expansions (3.5) and (3.7). We get

$$f(x_{j,2})=f(a_-)+\frac{c_-}{2}\cdot\frac{\tilde{x}_{j,2}^2}{W}-\frac{1}{2a_-^3}\cdot\frac{\tilde{x}_{j,2}^3}{W^{3/2}}+O\Big(\frac{\tilde{x}_{j,2}^4}{W^2}\Big),\tag{3.10}$$

$$\begin{aligned}
d_{j,11}&=1-x_{j,1}^{-1}b_{j,1}=\frac{\tilde{x}_{j,1}/a_+-\tilde{b}_{j,1}}{\sqrt{W}}+\frac{a_+\tilde{x}_{j,1}\tilde{b}_{j,1}-\tilde{x}_{j,1}^2}{a_+^2W}+O\big(W^{-3(1-\kappa)/2}\big),\\
d_{j,22}&=1-x_{j,2}^{-1}b_{j,2}=\frac{\tilde{x}_{j,2}/a_--\tilde{b}_{j,2}}{\sqrt{W}}+\frac{a_-\tilde{x}_{j,2}\tilde{b}_{j,2}-\tilde{x}_{j,2}^2}{a_-^2W}+O\big(W^{-3(1-\kappa)/2}\big),\\
d_{j,12}&=1-x_{j,1}^{-1}b_{j,2}=c_+-\frac{\tilde{x}_{j,1}/a_+-\tilde{b}_{j,2}}{a_+^2\sqrt{W}}-\frac{a_+\tilde{x}_{j,1}\tilde{b}_{j,2}-\tilde{x}_{j,1}^2}{a_+^4W}+O\big(W^{-3(1-\kappa)/2}\big),\\
d_{j,21}&=1-x_{j,2}^{-1}b_{j,1}=c_--\frac{\tilde{x}_{j,2}/a_--\tilde{b}_{j,1}}{a_-^2\sqrt{W}}-\frac{a_-\tilde{x}_{j,2}\tilde{b}_{j,1}-\tilde{x}_{j,2}^2}{a_-^4W}+O\big(W^{-3(1-\kappa)/2}\big).
\end{aligned}\tag{3.11}$$

The change (3.9) gives the Jacobian $W^{-2|\Lambda|}$, which together with $Q^{(2)}_{W,|\Lambda|}$ gives $W^{2|\Lambda|}$ in front of the integral (3.2). Similarly to the proof of Lemma 3.1, we are going to compute the leading order of the integral (3.2) over $\{\tilde{x}_{j,l}\}$, $\{\tilde{b}_{j,l}\}$, $l=1,2$, $j\in\Lambda$, and so we leave the quadratic part of f (see (3.5) and (3.10)) in the exponent, expand everything else into the series of $\tilde{x}_{j,l}/\sqrt{W},\tilde{b}_{j,l}/\sqrt{W}$ around the saddle-point $\tilde{x}_{j,l}=\tilde{b}_{j,l}=0$, and compute the Gaussian integral of each term of this expansion. We are going to prove, that the non-zero contribution is given by the terms having at least $W^{-2|\Lambda|}$.

Lemma 3.2. *Formula (3.2) can be rewritten as*

$$\begin{aligned}
\mathcal{R}^{+-}_{Wn\beta}(E,\varepsilon,\xi)&=(c_0/2\pi)^{2|\Lambda|}C_{E,\varepsilon}\int dz\,d\tilde{\rho}\,d\tilde{\tau}\,dU\,dS\\
&\times\exp\Big\{-\frac{1}{2}(Mz,z)+W^{1/2}(z,h^0)+W^{-1/2}(z,h+\zeta/|\Lambda)|)\Big\}\\
&\times\exp\Big\{\beta\sum\mathrm{Tr}\,\Big(U_j^*\tilde{\rho}_jS_j-U_{j-1}^*\tilde{\rho}_{j-1}S_{j-1}\Big)\Big(S_j^{-1}\tilde{\tau}_jU_j-S_{j-1}^{-1}\tilde{\tau}_{j-1}U_{j-1}\Big)\Big\}\\
&\times\exp\Big\{\sum\Big(c_+n_{j,12}+c_-n_{j,21}-n_{j,1}/c_0a_++n_{j,2}/c_0a_-\Big)-\beta c_0^2\sum(v_j^2+t_j^2)\Big\}\\
&\times\exp\Big\{\frac{ic_0}{2|\Lambda|}\sum_{j\in\Lambda}\Big(\mathrm{Tr}\,U_j^*LU_j\big(i\varepsilon L+\hat{\xi}/\rho(E)\big)+\mathrm{Tr}\,S_j^{-1}LS_j\big(i\varepsilon L+\hat{\xi}'/\rho(E)\big)\Big)\Big\}+o(1),
\end{aligned}\tag{3.12}$$

where

$$\tilde\rho_j=\begin{pmatrix}\rho_{j,11} & \rho_{j,12}/\sqrt{W}\\ \rho_{j,21}/\sqrt{W} & \rho_{j,22}\end{pmatrix},\quad \tilde\tau_j=\begin{pmatrix}\tau_{j,11} & \tau_{j,12}/\sqrt{W}\\ \tau_{j,21}/\sqrt{W} & \tau_{j,22}\end{pmatrix} \tag{3.13}$$
$$n_{j,12}=\rho_{j,12}\tau_{j,12},\quad n_{j,21}=\rho_{j,21}\tau_{j,21},$$
$$n_{j,1}=\rho_{j,11}\tau_{j,11},\quad n_{j,2}=\rho_{j,22}\tau_{j,22},$$
$$z=(z_{j,11},z_{j,22},z_{j,12},z_{j,21})=(\tilde x_{j,1},\tilde x_{j,2},\tilde b_{j,1},\tilde b_{j,1}),$$

and

$$M=M_0+W^{-1}\tilde M \tag{3.14}$$
$$(M_0z,z)=\sum_{j\in\Lambda}\Big(c_+\tilde x_{j,1}^2+c_-\tilde x_{j,2}^2+a_+^2c_+\tilde b_{j,1}^2+a_-^2c_-\tilde b_{j,2}^2\Big) \tag{3.15}$$
$$\begin{aligned}(\tilde Mz,z)&=-2\beta\sum\Big(\tilde x_{j,1}\tilde x_{j-1,1}+\tilde x_{j,2}\tilde x_{j-1,2}-a_+^2\tilde b_{j,1}\tilde b_{j-1,1}-a_-^2\tilde b_{j,2}\tilde b_{j-1,2}\Big)\\ &+2\beta\sum\Big(v_j^2\,(\tilde x_{j,1}-\tilde x_{j,2})(\tilde x_{j-1,1}-\tilde x_{j-1,2})+t_j^2\,(a_+\tilde b_{j,1}-a_-\tilde b_{j,2})(a_+\tilde b_{j-1,1}-a_-\tilde b_{j-1,2})\Big)\\ &-\sum\Big(\frac{4}{c_0^2}(\tilde x_{j,1}\tilde x_{j,2}-\tilde b_{j,1}\tilde b_{j,2})-2(a_+^{-3}n_{j,12}\tilde x_{j,1}\tilde b_{j,2}+a_-^{-3}n_{j,21}\tilde x_{j,2}\tilde b_{j,1})\Big).\end{aligned} \tag{3.16}$$

Here $\zeta=\{\zeta_j\}_{j\in\Lambda}$, $\zeta_j=(\zeta_{j,11},\zeta_{j,22},a_+\zeta_{j,12},a_-\zeta_{j,21})$ *with*

$$\zeta_{j,11}=-\varepsilon+i\xi_1/\rho(E)+2\alpha_1u_j^2,\quad \zeta_{j,22}=\varepsilon+i\xi_2/\rho(E)-2\alpha_1u_j^2,$$
$$\zeta_{j,12}=-\varepsilon+i\xi_1'/\rho(E)-2\alpha_2s_j^2,\quad \zeta_{j,21}=\varepsilon+i\xi_2'/\rho(E)+2\alpha_2s_j^2,$$

where $\alpha_{1,2}$ *are defined in (1.15). We also denoted*

$$\begin{array}{ll} h=\{h_{j,ls}\}_{j\in\Lambda,l,s=1,2}, & h^0=\{h^0_{j,ls}\}_{j\in\Lambda,l,s=1,2},\\ h_{j,11}=2/c_0-\beta c_0v_j^2-\beta c_0v_{j+1}^2+a_-n_{j,12}/a_+^2, & h^0_{j,11}=n_{j,1}/a_+,\\ h_{j,22}=-2/c_0+\beta c_0v_j^2+\beta c_0v_{j+1}^2+a_+n_{j,21}/a_-^2, & h^0_{j,22}=n_{j,2}/a_-,\\ h_{j,12}=2a_+/c_0-2-\beta c_0a_+t_j^2-\beta c_0a_+t_{j+1}^2-n_{j,21}a_+/a_-, & h^0_{j12}=-n_{j,1},\\ h_{j,21}=-2a_-/c_0-2+\beta c_0a_-t_j^2+\beta c_0a_-t_{j+1}^2-n_{j,12}a_-/a_+, & h^0_{j,21}=-n_{j,2},\end{array} \tag{3.17}$$

and

$$u_j=|(U_j)_{12}|,\quad v_j=|(U_jU_{j-1}^*)_{12}|,\quad s_j=|(S_j)_{12}|,\quad t_j=|(S_jS_{j-1}^{-1})_{12}|.$$

Proof. Rewriting the determinant in (3.3) in a standard way, we obtain

$$\det\mathcal{D}(\hat X,\hat B,U,S)=\sum_{\bar\sigma}(-1)^{|\sigma|}\prod_{j\in|\Lambda|}P_{j,\bar\sigma_j}(\tilde x_{j,1},\tilde x_{j2},\tilde b_{j,1},\tilde b_{j,1}), \tag{3.18}$$

where $\bar\sigma$ is a permutation of $\{(j,ls)\}$, $l,s=1,2$, $j\in\Lambda$, $\bar\sigma_j$ is its restriction on $\{(j,ls)\}_{l,s=1}^2$, $(-1)^{|\sigma|}$ is a sign of σ and $P_{j,\bar\sigma_j}$ is an expansion in $\tilde x_{j,1},\tilde x_{j2}$, $\tilde b_{j,1},\tilde b_{j,1}$ of the product of four elements from the rows $\{(j,ls)\}_{l,s=1}^2$ taken with respect to $\bar\sigma_j$.

Let us prove that for each $j\in\Lambda$ and any $\bar\sigma$ each term of $P_{j,\bar\sigma_j}(\tilde x_{j,1},\tilde x_{j2},\tilde b_{j,1},\tilde b_{j,1})$ of (3.18) belongs to one of the three following groups:

i. has a coefficient W^{-2} or lower;

ii. has a coefficient $W^{-3/2}$ and at least one of variables $\tilde{x}_{j,1}, \tilde{x}_{j2}$, $\tilde{b}_{j,1}, \tilde{b}_{j,1}$ of the odd degree;

iii. has a coefficient W^{-1} and at least two variables of $\tilde{x}_{j,1}, \tilde{x}_{j2}$, $\tilde{b}_{j,1}, \tilde{b}_{j,1}$ of the odd degree;

Note that each element in the expansion of the coefficients of the rows $(j, 11)$ and $(j, 22)$ has a coefficient $W^{-1/2}$ or lower, and so $P_{j,\bar{\sigma}_j}(\tilde{x}_{j,1}, \tilde{x}_{j2}, \tilde{b}_{j,1}, \tilde{b}_{j,1})$ has a coefficient W^{-1} or lower. In addition, if $P_{j,\bar{\sigma}_j}(\tilde{x}_{j,1}, \tilde{x}_{j,2}, \tilde{b}_{j,1}, \tilde{b}_{j,1})$ contains any terms with R_{jk} (see (3.3)), or at least one off-diagonal elements in $(j, 12)$ and $(j, 21)$, we get a coefficient W^{-2} or lower (and so obtain the group (i)).

We are left to consider terms with $d_{j,12}d_{j,21}$. If $P_{j,\bar{\sigma}_j}(\tilde{x}_{j,1}, \tilde{x}_{j,2}, \tilde{b}_{j,1}, \tilde{b}_{j,1})$ contains two off-diagonal elements in rows $(j, 11)$ and $(j, 11)$, we get group (i). One off-diagonal element and $d_{j,11}$ (or $d_{j,22}$) gives group (ii) or group (i) (since off-diagonal elements do not depend on $\tilde{x}_{j,1}, \tilde{x}_{j,2}$, $\tilde{b}_{j,1}, \tilde{b}_{j,1}$), and it is easy to see from (3.11) that all the terms in expansion of $d_{j,11}d_{j,22}d_{j,12}d_{j,21}$ belongs to groups (i) – (iii).

To get a non-zero contribution, we have to complete the expression $P_{j,\bar{\sigma}_j}(\tilde{x}_{j,1}, \tilde{x}_{j,2}, \tilde{b}_{j,1}, \tilde{b}_{j,1})$ by some other terms of the expansion of the exponent of (3.2) in order to get an even degree of each variable $\tilde{x}_{j,1}, \tilde{x}_{j,2}$, $\tilde{b}_{j,1}, \tilde{b}_{j,1}$. But all such a terms have the coefficient $W^{-1/2}$ or lower, and therefore Lemma 3.2 yields that the coefficient near each j in terms that gives a non-zero contribution must be W^{-2} or lower. Since we have a coefficient $W^{2|\Lambda|}$ in (3.2) after the change (3.9), this means that to get a non-zero contribution each coefficient must be exactly W^{-2}. Note that the terms of $P_{j,\bar{\sigma}_j}(\tilde{x}_{j,1}, \tilde{x}_{j,2}, \tilde{b}_{j,1}, \tilde{b}_{j,1})$ that can be completed to the monomial with all even degrees and with a coefficients W^{-2} does not contain any terms with R_{jk}, and any terms of the expansion $d_{j,ls}$, $l, s = 1, 2$ of order $W^{-3/2}$ or lower. They also cannot be completed to the monomial with all even degrees and with a coefficients W^{-2} by any terms of the exponent of (3.2) that has a coefficient lower then $W^{-1/2}$ for some j. Thus we need to consider the terms up to the third order in the expansions (3.5) and (3.10), the linear terms of the functions in the second and the forth exponents of (3.2), and the linear terms in $-2\log b_{j,l}$, $l = 1, 2$ coming from

$$b_{j,l}^{-2} = e^{-2\log b_{j,l}}, \quad l = 1, 2.$$

Note that the terms containing $\tilde{x}_{j,1}\tilde{b}_{j,1}/W$ in $d_{j,11}$ (see (3.11)) cannot contribute to the limit, since if we complete them to the monomial with even degrees of $\tilde{x}_{j,1}, \tilde{b}_{j,1}$, then it will contain W^{-2} and an additional W^{-1} should come from the line containing $d_{j,22}$. Moreover, the terms containing $\tilde{x}_{j,1}^2$ in $d_{j,11}$ can give a non-zero contribution only if the resulting monomial contains only $\tilde{x}_{j,1}^2$, since otherwise, taking into account the contribution of the line containing $d_{j,22}$, we again obtain at least W^{-3}. Thus we can replace $\tilde{x}_{j,1}^2$ by its average via Gaussian measure $(2\pi/c_+)^{-1/2}e^{-c_+\tilde{x}_{j,1}^2/2}$, i.e. by c_+^{-1}. The same is true for $\tilde{x}_{j,2}\tilde{b}_{j,2}/W$ and for $\tilde{x}_{j,2}^2$ which could be replaced by c_-^{-1}. Similar argument yields that the contribution of the terms with $\tilde{x}_{j,1}^2$ in the line containing $d_{j,12}$ and $\tilde{x}_{j,2}^2$ in the line containing $d_{j,21}$ disappear in the limit $W \to \infty$. Thus the term corresponding to $W^{2|\Lambda|}\det\mathcal{D}$ in (3.2) can be replaced by the term

$$\begin{aligned} \int d\rho\, d\tau \exp\Big\{ &\beta \sum \mathrm{Tr}\, \Big(U_j^*\tilde{\rho}_j S_j - U_{j-1}^*\tilde{\rho}_{j-1}S_{j-1}\Big)\Big(S_j^{-1}\tilde{\tau}_j U_j - S_{j-1}^{-1}\tilde{\tau}_{j-1}U_{j-1}\Big) \\ &+ \sum \Big(c_+ n_{j,12} + c_- n_{j,21} - n_{j,1}/c_0 a_+ + n_{j,2}/c_0 a_-\Big) \\ &+ W^{1/2}\sum \Big((\tilde{x}_{j,1}/a_+ - \tilde{b}_{j,1})n_{j,1} + (\tilde{x}_{j,2}/a_- - \tilde{b}_{j,2})n_{j,2}\Big) \\ &- W^{-1/2}\sum \Big(a_+^{-2}(\tilde{x}_{j,1}/a_+ - \tilde{b}_{j,2})n_{j,12} + a_-^{-2}(\tilde{x}_{j,2}/a_- - \tilde{b}_{j,1})n_{j,21}\Big)\Big\} + O(W^{-1/2}), \end{aligned} \tag{3.19}$$

where $\tilde\rho_j$, $\tilde\tau_j$, $n_{j,12}$, $n_{j,21}$, $n_{j,1}$, $n_{j,2}$ are defined in (3.13). Here we have used Grassmann variables $\{\rho_{j,ls}\}$, $\{\tau_{j,ls}\}$, $j\in\Lambda$, $l,s=1,2$ to rewrite the determinant (3.3) with respect to (7.4), have substituted (3.11) and left only terms that give the contribution (according to arguments above), and then have changed $\rho_{j,11}\to\sqrt{W}\rho_{j,11}$, $\tau_{j,11}\to\sqrt{W}\rho_{j,11}$. Note also

$$c_+a_+^2=c_0a_+,\quad c_-a_-^2=-c_0a_-. \tag{3.20}$$

Now let us prove that the contribution of the third order in the expansions (3.5) and (3.10) is small. Indeed, the terms $P_{j,\bar\sigma_j}(\tilde x_{j,1},\tilde x_{j,2},\tilde b_{j,1},\tilde b_{j,1})$ that can be completed to the monomial with all even degrees and with a coefficients W^{-2} by these cubic terms can be one of two types

1. terms $\big(\tilde x_{j,1}/a_+-\tilde b_{j,1}\big)\cdot x\cdot c_+\cdot c_-$, where c_+, c_- come from the zero terms of $d_{j,12}$, $d_{j,21}$ (see (3.11)) and x is an element of the row $(j,22)$ and so does not depend on $\tilde x_{j,1}$, $\tilde b_{j,1}$ (or similar terms with $\big(\tilde x_{j,2}/a_--\tilde b_{j,2}\big)$);

2. terms of $\big(\tilde x_{j,1}/a_+-\tilde b_{j,1}\big)\big(\tilde x_{j,2}/a_--\tilde b_{j,2}\big)\big(\tilde x_{j,1}/a_+-\tilde b_{j,2}\big)\cdot c_-$ with $\tilde x_{j,1}^2$ or $\tilde b_{j,2}^2$ (or similar terms with c_+ coming from $d_{j,12}$) ;

But it is easy to see that

$$\int(\tilde x_{j,1}^4/(3a_+^4)-\tilde b_{j,1}^4/3)\cdot e^{-\frac{c_+\tilde x_{j,1}^2}{2}-\frac{a_+^2c_+\tilde b_{j,1}^2}{2}}\,d\tilde x_{j,1}\,d\tilde b_{j,1}=\frac{2\pi}{a_+c_+}\Big(\frac{1}{a_+^4c_+^2}-\frac{1}{a_+^4c_+^2}\Big)=0,$$

and so the contribution of (1) is zero. Similarly the contribution (2) is zero.

Therefore, the contribution of the third order in the expansions (3.5) is small, and using (3.19) and also

$$\begin{aligned}&\exp\Big\{\frac{i}{|\Lambda|}\sum_{j\in\Lambda}\Big(\mathrm{Tr}\,U_j^*L_\pm U_j\big(i\varepsilon L+\hat\xi_1/\rho(E)\big)+\mathrm{Tr}\,S_j^{-1}L_\pm S_j\big(i\varepsilon L+\hat\xi_2/\rho(E)\big)\Big)\Big\}\\&=\exp\big\{-E(\xi_1+\xi_2+\xi_1'+\xi_2')/2\rho(E)\big\}\\&\times\exp\Big\{\frac{ic_0}{2|\Lambda|}\sum_{j\in\Lambda}\Big(\mathrm{Tr}\,U_j^*LU_j\Big(i\varepsilon L+\frac{\hat\xi}{\rho(E)}\Big)+\mathrm{Tr}\,S_j^{-1}LS_j\Big(i\varepsilon L+\frac{\hat\xi'}{\rho(E)}\Big)\Big)\Big\}\end{aligned}$$

for $L_\pm$, L defined in (1.18), we get (3.12).

□

Denoting the exponent in the second line of (3.12) by $\mathcal{E}(z)$ and taking the Gaussian integral over dz with z of (3.13), we get

$$\begin{aligned}\int_{\mathbb{R}^{4|\Lambda|}}\mathcal{E}(z)dz&=(2\pi)^{2|\Lambda|}\mathrm{det}^{-1/2}M\\&\exp\Big\{\frac12(M^{-1}(W^{1/2}h^0+W^{-1/2}(h+\zeta/\Lambda)),W^{1/2}h^0+W^{-1/2}(h+\zeta/|\Lambda|))\Big\}.\end{aligned} \tag{3.21}$$

It is easy to see from (3.14) – (3.16) that

$$\det\,M=\det\,M_0(1+O(W^{-1}))=(c_+^2c_-^2a_+^2a_-^2)^{|\Lambda|}(1+O(W^{-1}))=c_0^{4|\Lambda|}(1+O(W^{-1}))$$

with $c_\pm$ of (3.6). Note now that

$$M^{-1}=\big(M_0+\frac1W\tilde M\big)^{-1}=M_0^{-1}-\frac1WM_0^{-1}\tilde MM_0^{-1}+O(W^{-2}).$$

Since M_0 is diagonal and $h^0_{j,ls}$ is proportional to $n_{j,1}$ or $n_{j,2}$ and $n^2_{j,l}=0$, we have

$$(M_0^{-1}h^0,h^0)=0.$$

Hence, the exponent in the r.h.s. of (3.21) takes the form

$$\frac{1}{2}\Big((M_0^{-1}h^0,h+\zeta/\Lambda)+(M_0^{-1}(h+\zeta/\Lambda),h^0)\\ -(M_0^{-1}\tilde{M}M_0^{-1}h^0,h^0)\Big)+o(1)=I_1+I_2-I_3+o(1).$$

Then we can rewrite (recall (3.17) and (3.20))

$$\begin{aligned}I_1+I_2=&\sum\Big(\frac{(h_{j,11}+\zeta_{j,11}/|\Lambda|)n_{j,1}}{a_+c_+}+\frac{(h_{j,22}+\zeta_{j,22}/\Lambda)n_{j,2}}{a_-c_-}\\ &-\frac{(h_{j,12}+a_+\zeta_{j,12}/|\Lambda|)n_{j,1}}{a_+^2c_+}-\frac{(h_{j,21}+a_-\zeta_{j,21}/|\Lambda|)n_{j,2}}{a_-^2c_-}\Big)\\ =&\sum n_{j,1}\Big(\frac{2}{a_+c_0}+\beta\big(t_j^2+t_{j+1}^2-v_j^2-v_{j+1}^2\big)+\frac{a_-n_{j,12}}{a_+^2c_0}+\frac{n_{j,21}}{a_-c_0}+\frac{\zeta_{j,11}-\zeta_{j,12}}{c_0|\Lambda|}\Big)\\ &+\sum n_{j,2}\Big(-\frac{2}{a_-c_0}+\beta\big(t_j^2+t_{j+1}^2-v_j^2-v_{j+1}^2\big)-\frac{a_+n_{j,21}}{a_-^2c_0}-\frac{n_{j,12}}{a_+c_0}-\frac{\zeta_{j,22}-\zeta_{j,21}}{c_0|\Lambda|}\Big)+O(W^{-1}),\end{aligned}\tag{3.22}$$

$$\begin{aligned}I_3=&\frac{4}{c_0^4}\sum n_{j,1}n_{j,2}-\frac{1}{a_+^2c_0^2}\sum n_{j,12}n_{j,1}n_{j,2}-\frac{1}{a_-^2c_0^2}\sum n_{j,21}n_{j,1}n_{j,2}\\ &+\sum\frac{\beta(v_j^2+t_j^2)}{c_0^2}\big(n_{j,1}n_{j+1,1}+n_{j,1}n_{j+1,2}+n_{j,2}n_{j+1,1}+n_{j,2}n_{j+1,2}\big)+O(W^{-1}).\end{aligned}\tag{3.23}$$

Moreover,

$$\begin{aligned}&\exp\Big\{\beta\sum\mathrm{Tr}\,\Big(U_j^*\tilde{\rho}_jS_j-U_{j-1}^*\tilde{\rho}_{j-1}S_{j-1}\Big)\Big(S_j^{-1}\tilde{\tau}_jU_j-S_{j-1}^{-1}\tilde{\tau}_{j-1}U_{j-1}\Big)\Big\}\\ &=\exp\Big\{\frac{\beta}{W}\sum\mathrm{Tr}\,\Big(U_j^*\hat{\rho}_jS_j-U_{j-1}^*\hat{\rho}_{j-1}S_{j-1}\Big)\Big(S_j^{-1}\hat{\tau}_jU_j-S_{j-1}^{-1}\hat{\tau}_{j-1}U_{j-1}\Big)\Big\}+O(W^{-1/2}),\end{aligned}\tag{3.24}$$

where

$$\hat{\rho}_j=\mathrm{diag}\{\rho_{j,11},\rho_{j,22}\},\quad \hat{\tau}_j=\mathrm{diag}\{\tau_{j,11},\tau_{j,22}\}.$$

Combining (3.22) – (3.24) we can integrate the main term of (3.21) with respect to $\rho_{j,12}$, $\tau_{j,12}$, $\rho_{j,21}$, $\tau_{j,21}$ according to (7.4). This integration gives

$$\begin{aligned}&\prod_{j\in\Lambda}\Big(c_++\frac{a_-n_{j,1}}{a_+^2c_0}-\frac{n_{j,2}}{a_+c_0}+\frac{n_{j,1}n_{j,2}}{a_+^2c_0^2}\Big)\Big(c_-+\frac{n_{j,1}}{a_-c_0}-\frac{a_+n_{j,2}}{a_-^2c_0}+\frac{n_{j,1}n_{j,2}}{a_-^2c_0^2}\Big)\\ &=c_0^2+\frac{c_0n_{j,2}}{a_-}-\frac{c_0n_{j,1}}{a_+}+(1+2/c_0^2)n_{j,1}n_{j,2}=c_0^2\cdot\exp\Big\{-\frac{n_{j,1}}{a_+c_0}+\frac{n_{j,2}}{a_-c_0}\Big\}\cdot\Big(1+\frac{2}{c_0^4}n_{j,1}n_{j,2}\Big),\end{aligned}$$

which together with (3.22) – (3.24) yields

$$
\begin{aligned}
\mathcal{R}^{+-}_{Wn\beta}(E,\varepsilon,\xi) =& c_0^{4|\Lambda|} C_{E,\varepsilon} \int d\hat{\rho}\, d\hat{\tau}\, dU\, dS \prod_{j\in\Lambda} \Big(1 - \frac{2}{c_0^4} n_{j,1} n_{j,2}\Big) \exp\Big\{ -\beta c_0^2 \sum (v_j^2 + t_j^2)\Big\} \\
&\times \exp\Big\{ \beta \sum \mathrm{Tr}\, \Big(U_j^* \hat{\rho}_j S_j - U_{j-1}^* \hat{\rho}_{j-1} S_{j-1}\Big)\Big(S_j^{-1} \hat{\tau}_j U_j - S_{j-1}^{-1} \hat{\tau}_{j-1} U_{j-1}\Big)\Big\} \\
&\times \exp\Big\{ \sum n_{j,1} \Big(\beta\big(t_j^2 + t_{j+1}^2 - v_j^2 - v_{j+1}^2\big) + \frac{\zeta_{j,11} - \zeta_{j,12}}{c_0|\Lambda|} |)\Big)\Big\} \\
&\times \exp\Big\{ \sum n_{j,2} \Big(\beta\big(t_j^2 + t_{j+1}^2 - v_j^2 - v_{j+1}^2\big) - \frac{\zeta_{j,22} - \zeta_{j,21}}{c_0|\Lambda|})\Big)\Big\} + o(1),
\end{aligned}
$$

where we have used

$$
(1 + 2n_{j,1}n_{j,2}/c_0^4) \cdot e^{-4n_{j,1}n_{j,2}/c_0^4} = 1 - 2n_{j,1}n_{j,2}/c_0^4.
$$

Now changing

$$
\rho_{j,11} \to c_0 \rho_{j,1}, \quad \tau_{j,11} \to c_0 \tau_{j,1}, \quad \rho_{j,22} \to c_0 \rho_{j,2}, \quad \tau_{j,22} \to c_0 \rho_{j,2}
$$

with an appropriate change in $n_{j,1}$, $n_{j,2}$, $\hat{\rho}_j$, $\hat{\tau}_j$, and recalling (1.21), we get

$$
\begin{aligned}
\mathcal{R}^{+-}_{Wn\beta}(E,\varepsilon,\xi) =& C_{E,\varepsilon} \int d\hat{\rho}\, d\hat{\tau}\, dU\, dS \prod_{j\in\Lambda} \Big(1 - 2n_{j,1} n_{j,2}\Big) \exp\Big\{ -\tilde{\beta} \sum (v_j^2 + t_j^2)\Big\} \\
&\times \exp\Big\{ \tilde{\beta} \sum \mathrm{Tr}\, \Big(U_j^* \hat{\rho}_j S_j - U_{j-1}^* \hat{\rho}_{j-1} S_{j-1}\Big)\Big(S_j^{-1} \hat{\tau}_j U_j - S_{j-1}^{-1} \hat{\tau}_{j-1} U_{j-1}\Big)\Big\} \\
&\times \exp\Big\{ \sum n_{j,1} \Big(\tilde{\beta}\big(t_j^2 + t_{j+1}^2 - v_j^2 - v_{j+1}^2\big) + c_0(\zeta_{j,11} - \zeta_{j,12})/|\Lambda|)\Big)\Big\} \\
&\times \exp\Big\{ \sum n_{j,2} \Big(\tilde{\beta}\big(t_j^2 + t_{j+1}^2 - v_j^2 - v_{j+1}^2\big) - c_0(\zeta_{j,22} - \zeta_{j,21})/|\Lambda|)\Big)\Big\} \\
&\times \exp\Big\{ \frac{ic_0}{2|\Lambda|} \sum_{j\in\Lambda} \Big(\mathrm{Tr}\, U_j^* L U_j \big(i\varepsilon L + \hat{\xi}/\rho(E)\big) + \mathrm{Tr}\, S_j^{-1} L S_j \big(i\varepsilon L + \hat{\xi}'/\rho(E)\big)\Big)\Big\},
\end{aligned}
$$

which can be rewritten as (1.11). The second relation of (1.11) follows from the uniform in ξ convergence of $\mathcal{R}^{+-}_{Wn\beta}(E,\varepsilon,\xi)$, as $W \to \infty$.

□

3.2 Proof of Theorem 1.2.

Theorem 1.2 can be proved in a similar way. First of all we can write an analogue of (3.2):

$$\mathcal{R}^{++}_{Wn\beta}(E,\varepsilon,\xi)=Q^{(2)}_{W,|\Lambda|}\int dV\,dU\int dx\int_{\mathbb{R}_+^{2|\Lambda|}}db\prod_{j\in\Lambda}\frac{(x_{j,1}-x_{j,2})^2(b_{j,1}-b_{j,2})^2}{b_{j,1}^2b_{j,2}^2} \tag{3.25}$$
$$\times\exp\Big\{-W\sum_{j\in\Lambda}\sum_{\sigma=1}^{2}\big(f(x_{j,\sigma})+f(b_{j,\sigma})\big)\Big\}\cdot\det\mathcal{D}(\hat{X},\hat{B},U,V)$$
$$\times\exp\Big\{\frac{\beta}{2}\sum_{j\sim k}\Big(\mathrm{Tr}\,(V_j^*\hat{B}_jV_j-V_k^*\hat{B}_kV_k)^2-\mathrm{Tr}\,(U_j^*\hat{X}_jU_j-U_k^*\hat{X}_kU_k)^2\Big)\Big\}$$
$$\times\exp\Big\{\frac{1}{2W}\sum_{j,k}R_{jk}\,\mathrm{Tr}\,(U_j^*\hat{X}_jU_j-iZ_1)(U_k^*\hat{X}_kU_k-iZ_1)\Big\}$$
$$\times\exp\Big\{\frac{i}{|\Lambda|}\sum_{j\in\Lambda}\Big(\mathrm{Tr}\,U_j^*\hat{X}_jU_j\big(i\varepsilon\cdot I+\hat{\xi}_1/\rho(E)\big)+\mathrm{Tr}\,V_j^*\hat{B}_jV_j\big(i\varepsilon\cdot I+\hat{\xi}_2/\rho(E)\big)\Big)\Big\}.$$

Note that (3.25) has unitary V_j instead of hyperbolic S_j and $i\varepsilon\cdot I$ instead of $i\varepsilon\cdot L$. Then we deform the contours of integration as

- for $x_{j,1}$, $x_{j,2}$ to $iE/2+\mathbb{R}$;
- for $b_{j,1},b_{j,2}$ to $\mathcal{L}_+(E)$ of (1.20)

and prove the following lemma in the same way as Lemma 3.1):

Lemma 3.3. *The integral (3.25) over $\{x_{j,l}\},\{b_{j,l}\}$, $l=1,2$, $j\in\Lambda$ can be restricted to the integral over the neighbourhood of the points*

I. $x_{j,1}=a_+$, $x_{j,2}=a_-$ *or* $x_{j,1}=a_-$, $x_{j,2}=a_+$, $b_{j,1}=b_{j,2}=a_+$ *for any* $j\in\Lambda$;

II. $x_{j,1}=x_{j,2}=a_+$, $b_{j,1}=b_{j,2}=a_+$ *for any* $j\in\Lambda$;

III. $x_{j,1}=x_{j,2}=a_-$, $b_{j,1}=b_{j,2}=a_+$ *for any* $j\in\Lambda$.

Moreover, the contributions of the points I and II are $o(1)$, *as* $W\to\infty$.

Indeed, the contribution of the point II is small, since after an appropriate change of variables similar to (3.4) (which gives $W^{-2|\Lambda|}$) the expression

$$(x_{j,1}-x_{j,2})^2(b_{j,1}-b_{j,2})^2$$

gives $W^{-2|\Lambda|}$, and the expansion of $\det\mathcal{D}(\hat{X},\hat{B},U,V)$ starts from $W^{-2|\Lambda|}$ (see (3.11)).

For the points I the expression for $\det\mathcal{D}(\hat{X},\hat{B},U,V)$ starts from $W^{-|\Lambda|}$, and another $W^{-|\Lambda|}$ comes from $(b_{j,1}-b_{j,2})^2$. Therefore similarly to (3.8) we get that the main contribution around these saddle-points is given by

$$C\cdot\Big\langle\prod_{j\in\Lambda}(\tilde{x}_{j,1}/a_+-\tilde{b}_{j,1})\cdot(\tilde{x}_{j,1}/a_+-\tilde{b}_{j,2})\cdot(\tilde{b}_{j,1}-\tilde{b}_{j,2})^2\Big\rangle+o(1), \tag{3.26}$$

where

$$\langle \cdot \rangle = \int \big(\cdot \big) \exp\Big\{ -\frac{1}{2} \sum_{j\in\Lambda} \big(c_+ \tilde{x}_{j,1}^2 + c_- \tilde{x}_{j,2}^2 + a_+^2 c_+ (\tilde{b}_{j,1}^2 + \tilde{b}_{j,2}^2)\big)\Big\} d\tilde{x} d\tilde{b}.$$

But it is easy to see that the Gaussian integral in (3.26) is zero.

Thus we are left to compute the contribution of the point III. Doing again an appropriate change of variables similar to (3.4) , we see that the expression

$$(x_{j,1} - x_{j,2})^2 (b_{j,1} - b_{j,2})^2$$

already gives $W^{-2|\Lambda|}$, and hence to obtain a non-zero contribution we have to compute

$$\int \prod_{j\in\Lambda} (\tilde{x}_{j,1} - \tilde{x}_{j,2})^2 (\tilde{b}_{j,1} - \tilde{b}_{j,2})^2 \exp\Big\{ -\frac{1}{2} \sum_{j\in\Lambda} \big(c_- (\tilde{x}_{j,1}^2 + -\tilde{x}_{j,2}^2) + a_+^2 c_+ (\tilde{b}_{j,1}^2 + \tilde{b}_{j,2}^2)\big)\Big\} d\tilde{x} d\tilde{b}$$
$$= \Big((2\pi)^2 \cdot 4 (c_+ c_- a_+)^{-2} \Big)^{|\Lambda|}$$

and take only zero terms in the expansions of all other functions in (3.25). That gives the first relation of (1.13). The second relation of (1.13) follows from the uniform in ξ convergence of $\mathcal{R}^{++}_{Wn\beta}(E, \varepsilon, \xi)$ as $W \to \infty$. $\square$

4 Proof of Theorem 1.4

According (1.3), (1.9), (1.11), and (1.13), to prove Theorem 1.4, it is sufficient to show that

$$(2\pi)^{-2} \lim_{\varepsilon\to 0} \lim_{\beta,n\to\infty} \frac{\partial^2}{\partial \xi_1' \partial \xi_2'} \Big(\mathcal{R}^{+-}_{n\beta}(E, \varepsilon, \xi) + \overline{\mathcal{R}^{+-}_{n\beta}}(E, \varepsilon, \xi) \tag{4.1}$$
$$- \mathcal{R}^{++}_{n\beta}(E, \varepsilon, \xi) - \overline{\mathcal{R}^{++}_{n\beta}}(E, \varepsilon, \xi) \Big)\Big|_{\xi'=\xi} = 1 - \frac{\sin^2(\pi(\xi_1 - \xi_2))}{\pi^2 (\xi_1 - \xi_2)^2}.$$

Using (1.13) we get

$$\lim_{\varepsilon\to 0} \lim_{\beta,n\to\infty} \frac{\partial^2}{\partial \xi_1' \partial \xi_2'} \Big(\mathcal{R}^{++}_{n\beta}(E, \varepsilon, \xi) + \overline{\mathcal{R}^{++}_{n\beta}}(E, \varepsilon, \xi) \Big)\Big|_{\xi'=\xi} = -\frac{a_+^2 + a_-^2}{\rho^2(E)}. \tag{4.2}$$

In addition, $\mathcal{R}^{+-}_{n\beta}(E, \varepsilon, \xi)$ are analytic functions in any of $\xi_1, \xi_2, \xi_1', \xi_2'$ for $\Im\xi_1', \Im\xi_1' > -\varepsilon$, and they are uniformly bounded in n, β for $\xi_1, \xi_2, \xi_1', \xi_2'$ varying in any compacts satisfying this condition. Hence, we can replace the order of the derivative and the limiting transition and by (1.14) obtain

$$\lim_{\beta,n\to\infty} \frac{\partial^2}{\partial \xi_1' \partial \xi_2'} \mathcal{R}^{+-}_{n\beta}(E, \varepsilon, \xi)\Big|_{\xi'=\xi}$$
$$= \frac{\partial^2}{\partial \xi_1' \partial \xi_2'} C_{E,\varepsilon} e^{-c_0(\alpha_1+\alpha_2)} \Big(\delta_1\delta_2 (e^{2c_0\alpha_1} - 1)/\alpha_1\alpha_2 - (\delta_1 + \delta_2) e^{2c_0\alpha_1}/\alpha_2 + e^{2c_0\alpha_1}\alpha_1/\alpha_2 \Big)\Big|_{\xi'=\xi}.$$

Computing the derivative, we get

$$\lim_{\beta,n\to\infty} \frac{\partial^2}{\partial \xi_1' \partial \xi_2'} \mathcal{R}^{+-}_{n\beta}(E, \varepsilon, \xi)\Big|_{\xi'=\xi} = \frac{1}{\rho^2(E)} - \frac{1 - e^{2\pi i \theta_\varepsilon}}{\theta_\varepsilon^2},$$

where

$$\theta_\varepsilon = 2i\alpha_1\rho(E) = 2i\varepsilon\rho(E) + \xi_1 - \xi_2.$$

This yields

$$\lim_{\beta,n\to\infty} \frac{\partial^2}{\partial\xi_1'\partial\xi_2'} \Big(\mathcal{R}_{n\beta}^{+-}(E,\varepsilon,\xi) + \overline{\mathcal{R}_{n\beta}^{+-}}(E,\varepsilon,\xi)\Big)\Big|_{\xi'=\xi} = \frac{2}{\rho^2(E)} + \frac{(e^{i\pi\theta_\varepsilon} - e^{-i\pi\theta_\varepsilon})^2}{\theta_\varepsilon^2},$$

and hence

$$\lim_{\varepsilon\to 0}\lim_{\beta,n\to\infty} \frac{\partial^2}{\partial\xi_1'\partial\xi_2'} \Big(\mathcal{R}_{n\beta}^{+-}(E,\varepsilon,\xi) + \overline{\mathcal{R}_{n\beta}^{+-}}(E,\varepsilon,\xi)\Big)\Big|_{\xi'=\xi} = \frac{2}{\rho^2(E)} - \frac{4\sin^2(\pi(\xi_1-\xi_2))}{(\xi_1-\xi_2)^2},$$

which combined with (4.2), and

$$a_+^2 + a_-^2 + 2 = (a_+ - a_-)^2 = 4\pi^2\rho(E)^2,$$

gives (4.1), thus Theorem 1.4.

5 Proof of Theorem 1.3

Let us note that to prove Theorem 1.3, it suffices to prove it only for ξ such that

$$\begin{aligned} &\Re\xi_1 = \Re\xi_2, \quad \Re\xi_1' = \Re\xi_2', \quad \xi_1,\xi_2,\xi_1',\xi_2' \in \Omega_{c\varepsilon} \\ &\Omega_{c\varepsilon} = \{\xi : \Im\xi > -c\varepsilon\}, \quad (0<c<1). \end{aligned} \tag{5.1}$$

Indeed, assume that $\{\mathcal{R}_{n\beta}^{+-}(E,\varepsilon,\xi)\}$ are uniformly bounded in n,β for $\xi_1,\xi_2,\xi_1',\xi_2' \in \Omega_{c\varepsilon}$. Consider $\{\mathcal{R}_{n\beta}^{+-}(E,\varepsilon,\xi)\}$ as functions on ξ_1 with fixed ξ_2,ξ_1',ξ_2' such that $\Re\xi_1' = \Re\xi_2'$. Since these functions are analytic in $\Omega_{c\varepsilon}$, the standard complex analysis argument yields that (1.14) on the segment $\Re\xi_1 = \Re\xi_2$ implies (1.14) for any $\xi_1 \in \Omega_{c\varepsilon}$, hence for any $\xi_1,\xi_2 \in \Omega_{c\varepsilon}$. Then, fixing any ξ_1,ξ_2,ξ_2', we can consider $\{\mathcal{R}_{n\beta}^{+-}(E,\varepsilon,\xi)\}$ as a sequence of analytic functions on ξ_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 ξ_1',ξ_2'. Therefore, it is enough to prove Theorem 1.3 for real $\alpha_1 > \varepsilon/2$, $\alpha_2 > \varepsilon/2$, which means that we take $c = \rho(E)$ (see the definition (1.15)).

To check that $\{\mathcal{R}_{n\beta}^{+-}(E,\varepsilon,\xi)\}$ are uniformly bounded in n,β for $\xi_1,\xi_2,\xi_1',\xi_2' \in \Omega_{c\varepsilon}$, we apply the Schwartz inequality to $\mathcal{R}_{Wn\beta}^{+-}(E,\varepsilon,\xi)$ in the form (1.9). Then we get

$$\begin{aligned} &|\mathcal{R}_{Wn\beta}^{+-}(E,\varepsilon,\xi)|^2 \le |\mathcal{R}_{Wn\beta}^{+-}(E,\varepsilon,\xi_1)|\,|\mathcal{R}_{Wn\beta}^{+-}(E,\varepsilon,\xi_2)| \\ \Rightarrow &\mathcal{R}_{n\beta}^{+-}(E,\varepsilon,\xi)|^2 \le |\mathcal{R}_{n\beta}^{+-}(E,\varepsilon,\xi_1)|\,|\mathcal{R}_{n\beta}^{+-}(E,\varepsilon,\xi_2)| \end{aligned}$$

where $\xi_1 = (\xi_1,\xi_1,\xi_1',\xi_1')$, $\xi_2 = (\xi_2,\xi_2,\xi_2',\xi_2')$. Since ξ_1,ξ_2 satisfy (5.1), the uniform boundedness of the r.h.s. follows from the uniform convergence (in ξ, satisfying (5.1)) of (1.14) (see Section 5.2).

5.1 Representation of $\mathcal{R}_{n\beta}^{+-}$ in the operator form

Now we are going to represent $\mathcal{R}_{n\beta}^{+-}$ in 1d case in the operator form. Put $n = |\Lambda|$, and set

$$\mathcal{M}(Q,Q') = \mathcal{F}(Q)H(Q,Q')\mathcal{F}(Q'),$$

where

$$H(Q,Q') = \exp\Big\{\frac{\tilde\beta}{4}\mathrm{Str}\,QQ'\Big\}(1-n_1n_2)(1-n_1'n_2') \tag{5.2}$$
$$\mathcal{F}(Q) = \exp\Big\{-\frac{c_0}{4n}\mathrm{Str}\,Q\Lambda_{\xi,\varepsilon}\Big\} = F(U,S)\cdot\exp\Big\{n_1\cdot F_1(U,S)+n_2\cdot F_2(U,S)\Big\}$$

with Q, Q' of the form (1.12) and

$$\begin{aligned} F(U,S) &= \exp\big\{-\frac{c_0}{n}\big(\alpha_1(1-|U_{12}|^2)+\alpha_2\cdot|S_{12}|^2\big)\big\}, \\ F_1(U,S) &= -c_0\big(\delta_1-\alpha_1\cdot|U_{12}|^2-\alpha_2\cdot|S_{12}|^2\big)/n, \\ F_2(U,S) &= -c_0\big(\delta_2-\alpha_1\cdot|U_{12}|^2-\alpha_2\cdot|S_{12}|^2\big)/n, \\ n_l &= \rho_l\tau_l, \quad n_l'=\rho_l'\tau_l', \quad l=1,2, \end{aligned} \tag{5.3}$$

and $\alpha_{1,2},\delta_{1,2}$ defined in (1.15). Hence, by (1.11)

$$\mathcal{R}^{+-}_{n\beta}(E,\varepsilon,\xi) = C_{E,\varepsilon}e^{c_0(\alpha_1-\alpha_2)}\int(1-n_1n_2)\mathcal{F}(Q)\mathcal{M}^{n-1}(Q,Q')\mathcal{F}(Q')(1-n_1'n_2')dQdQ' \tag{5.4}$$

with

$$dQ = d\rho_1 d\tau_1 d\rho_2 d\tau_2 dUdS.$$

Note that $\mathcal{M}$, H, $\mathcal{F}$ can be considered as operators acting on the space of polynomials of Grassmann variables ρ_l', τ_l', $l=1,2$ with coefficients from $L_2(U)\otimes L_2(S)$, where L_2 are taken with respect to the Haar measures on $\mathring{U}(2)$, $\mathring{U}(1,1)$. It is easy to see these that operators transform any even Grassmann polynomial into an even polynomial and an odd one into an odd one. In addition, they preserve the modulo of the difference between the number of ρ_l and the number of τ_l. Since we are going to apply these operators only to even polynomials which contain equal numbers of ρ_l and τ_l, we need to study a restriction of $\mathcal{M}$, H, $\mathcal{F}$ to the space $\mathcal{P}_6\cong(L_2(U(2))\otimes L_2(U(1,1)))^6$ of polynomials

$$\widehat{q} = q_0+q_1n_1'+q_2n_2'+q_3n_1'n_2'+q_4\rho_1'\tau_2'+q_5\rho_2'\tau_1'. \tag{5.5}$$

Thus $\mathcal{M}$ is represented by a 6×6 matrix $\mathcal{P}_6\to\mathcal{P}_6$ (which we also denote $\mathcal{M}$) of the form $\mathcal{F}H\mathcal{F}\big|_{\mathcal{P}_6}$, the entries of the matrix H are the integral operators on $L_2(U)\otimes L_2(S)$ with the kernels of the form $v(U(U')^*,S(S')^{-1})$ (the integrals are taken with respect to $dU'dS'$), and the entries of the matrix $\mathcal{F}$ are operators of multiplication in $L_2(U)\otimes L_2(S)$. Then (5.4) takes the form

$$\begin{aligned} \mathcal{R}^{+-}_{n\beta}(E,\varepsilon,\xi) &= C_{E,\varepsilon}e^{c_0(\alpha_1-\alpha_2)}\int(\mathcal{M}^{n-1}\tilde f(U',S'),\tilde g(U,S))_6\,dUdSdU'dS', \\ \tilde f(U,S) &:= \mathcal{F}\cdot(1-n_1n_2), \quad \tilde g(U,S):=\mathcal{F}\cdot(1-n_1n_2), \end{aligned} \tag{5.6}$$

where by $(\cdot,\cdot)_6$ we mean the "scalar" product in $\mathcal{P}_6$ which gives the coefficient in front of n_1n_2 in the product of two polynomials of the form (5.5).

5.2 Proof of Theorem 1.3 for $\alpha_1,\alpha_2>\varepsilon/2$

As it was mentioned in the beginning of Section 5, it suffices to prove Theorem 1.3 for real $\alpha_1,\alpha_2>\varepsilon/2$.

The proof of (1.14) is based on the following representation of $\mathcal{R}^{+-}_{n\beta}(E,\varepsilon,\xi)$.

Proposition 5.1. *For any ξ such that $\alpha_1, \alpha_2 > \varepsilon/2$ (see (1.15)) we have*

$$\mathcal{R}^{+-}_{n\beta}(E,\varepsilon,\xi) = \frac{C_{E,\varepsilon}e^{c_0(\alpha_1-\alpha_2)}}{2\pi i}\oint_{\omega_A} z^{n-1}(\widehat{G}(z)\widehat{f},\widehat{g})dz, \quad \omega_A = \{z : |z| = 1 + A/n\}, \tag{5.7}$$

$$\widehat{G}(z) = (\widehat{M} - z)^{-1}, \quad \widehat{M} = \widehat{F}\widehat{K}\widehat{F}, \quad \widehat{K} = \widehat{K}_0 + O(\beta^{-1}), \tag{5.8}$$

where operators $\widehat{K}_0$, $\widehat{F}$ and the vectors $\widehat{f}$, $\widehat{g}$ have the form

$$\widehat{K}_0 = \begin{pmatrix} K_{US} & \widetilde{K}_1 & \widetilde{K}_2 & \widetilde{K}_3 \\ 0 & K_{US} & 0 & \widetilde{K}_2 \\ 0 & 0 & K_{US} & \widetilde{K}_1 \\ 0 & 0 & 0 & K_{US} \end{pmatrix}, \quad \widehat{F} = F\begin{pmatrix} 1 & F_1 & F_2 & F_1F_2 \\ 0 & 1 & 0 & F_2 \\ 0 & 0 & 1 & F_1 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{5.9}$$

$$\hat{f} = \widehat{F}(e_4 - e_1), \quad \hat{g} = \widehat{F}^{(t)}(e_1 - e_4)$$

with F and $F_{1,2}$ being the operator on $L_2(U)\otimes L_2(S)$ of multiplication by the functions F and $F_{1,2}$ defined in (5.3), $K_{US} = K_U \otimes K_S$ and K_U and K_S being the integral operators in $L_2(U)$ and $L_2(S)$ with a "difference" kernels

$$K_U(U,U') = K_U(U(U')^*) = \tilde{\beta}e^{-\tilde{\beta}|(U(U')^*)_{12}|^2},$$
$$K_S(S,S') = K_S(S(S')^{-1}) = \tilde{\beta}e^{-\tilde{\beta}|(S(S')^{-1})_{12}|^2}.$$

Here $\widetilde{K}_p$, $p = 1,2,3$ are normal operators on $L_2(U)\otimes L_2(S)$, they commute with K_{US} and with the Laplace operators $\widetilde{\Delta}_U, \widetilde{\Delta}_S$ on the corresponding groups and satisfy the bounds

$$|\widetilde{K}_p| \le C(1 - K_{US}) \le -C(\widetilde{\Delta}_U + \widetilde{\Delta}_S)/\beta, \tag{5.10}$$

where the Laplace operators $\widetilde{\Delta}_U, \widetilde{\Delta}_S$ for the functions depending only on $|S_{12}|^2$ and $|U_{12}|^2$ have the form

$$\widetilde{\Delta}_S(\varphi) = -\frac{d}{dx}x(x+1)\frac{d\varphi}{dx} \quad (x = |S_{12}|^2), \qquad \widetilde{\Delta}_U(\varphi) = -\frac{d}{dx}x(1-x)\frac{d\varphi}{dx} \quad (x = |U_{12}|^2).$$

We postpone the proof of the proposition to Section 6 and now derive (1.14) from it. To this end, set

$$\widehat{M}_0 = \widehat{F}^2, \quad \widehat{G}_0 = (\widehat{M}_0 - z)^{-1},$$

and consider

$$\Delta G := \widehat{G} - \widehat{G}_0 = -\widehat{G}_0(\widehat{M} - \widehat{M}_0)\widehat{G}_0 - \widehat{G}_0(\widehat{M} - \widehat{M}_0)\widehat{G}(\widehat{M} - \widehat{M}_0)\widehat{G}_0.$$

We apply the following lemma, which we will prove later:

Lemma 5.1. *For any $z \in \omega_A$ (see (5.7)) we have the bounds*

$$\begin{aligned} &\|(\widehat{M} - \widehat{M}_0)\widehat{G}_0\widehat{f}\|^2 \le C(n/\tilde{\beta})^2, \quad \|(\widehat{M} - \widehat{M}_0)\widehat{G}_0\widehat{g}\|^2 \le C(n/\tilde{\beta})^2 \\ &|(\widehat{G}_0(\widehat{M} - \widehat{M}_0)\widehat{G}_0\widehat{f},\widehat{g})| \le n\tilde{\beta}^{-1}/|z-1|, \quad \|\widehat{G}\| \le C\log^2 n/|z-1|. \end{aligned} \tag{5.11}$$

The lemma implies that

$$\Big|\frac{1}{2\pi i}\oint_{\omega_A} z^{n-1}(\Delta G\widehat{f},\widehat{g})dz\Big| \le C\oint_{\omega_A} |(\widehat{G}_0(\widehat{M}-\widehat{M}_0)\widehat{G}_0\widehat{f},\widehat{g})|\,|dz|$$
$$+\,C\oint_{\omega_A} \|\widehat{G}(z)\|\cdot\|(\widehat{M}-\widehat{M}_0)\widehat{G}_0(z)\widehat{f}\|\cdot\|(\widehat{M}-\widehat{M}_0)\widehat{G}_0(\bar{z})\widehat{g}\|\,|dz|$$
$$\le C(n/\tilde{\beta})\oint_{\omega_A}\frac{|dz|}{|z-1|}\le Cn\log n/\tilde{\beta}\to 0,$$

where we used $n\log^2 n \ll \tilde{\beta}$ and

$$\oint_{\omega_A}\frac{|dz|}{|z-1|}\le C\log n.$$

Thus we have proved that (recall (5.9))

$$\mathcal{R}^{+-}_{n\tilde{\beta}}(E,\varepsilon,\xi)=\frac{C_{E,\varepsilon}e^{c_0(\alpha_1-\alpha_2)}}{2\pi i}\oint_{\omega_A} z^{n-1}(\widehat{G}_0(z)\widehat{f},\widehat{g})dz+o(1)=C_{E,\varepsilon}e^{c_0(\alpha_1-\alpha_2)}(\widehat{F}^{2n-2}\widehat{f},\widehat{g})+o(1)$$
$$=C_{E,\varepsilon}e^{c_0(\alpha_1-\alpha_2)}\int(4n^2F_1F_2-2)F^{2n}dUdS+o(1).$$

Performing the integration with respect to dU, dS we obtain (1.14). $\square$

Proof of Lemma 5.1. To prove the first inequality of (5.11), observe that since $\widehat{F}$ is bounded we have

$$\|(\widehat{M}-\widehat{M}_0)\widehat{G}_0\widehat{f}\|^2=\|\widehat{F}(\widehat{K}-1)\widehat{F}\widehat{G}_0\widehat{f}\|^2\le C\|(\widehat{K}_0-1)\widehat{F}\widehat{G}_0\widehat{f}\|^2.$$

Moreover, since $\widetilde{K}_\alpha$ and $1-K_{US}$ commute with $\widetilde{\Delta}_U,\widetilde{\Delta}_S$, (5.10) implies

$$(\widehat{K}_0-1)^*(\widehat{K}_0-1)\le C\tilde{\beta}^{-2}(\widetilde{\Delta}_U+\widetilde{\Delta}_S)^2$$
$$\Rightarrow\|(\widehat{M}-\widehat{M}_0)\widehat{G}_0\widehat{f}\|^2\le C\tilde{\beta}^{-2}\|(\widetilde{\Delta}_U+\widetilde{\Delta}_S)\widehat{F}\widehat{G}_0\widehat{f}\|^2\le C'\tilde{\beta}^{-2}(\|\widetilde{\Delta}_U\widehat{G}_0\widehat{F}\widehat{f}\|^2+\|\widetilde{\Delta}_S\widehat{G}_0\widehat{F}\widehat{f}\|^2)$$
$$\le C'\tilde{\beta}^{-2}\max_{\mu,\nu\le 4}(\|\widetilde{\Delta}_U(\widehat{G}_0)_{\mu\nu}(\widehat{F}\widehat{f})_\nu\|^2+(\|\widetilde{\Delta}_S(\widehat{G}_0)_{\mu\nu}(\widehat{F}\widehat{f})_\nu\|^2).$$

It is easy to see that $\widehat{G}_0$ has the same form as the matrices in (5.9) with zeros below the main diagonal and

$$(\widehat{G}_0)_{ii}=G_0:=(F^2-z)^{-1},\quad (\widehat{G}_0)_{23}=0,\quad (\widehat{G}_0)_{12}=(\widehat{G}_0)_{34}=-2F_1G_0^2F^2$$
$$(\widehat{G}_0)_{13}=(\widehat{G}_0)_{24}=-2F_1G_0^2F^2,\quad (\widehat{G}_0)_{14}=8F_1F_2G_0^3F^4-4F_1F_2G_0^2F^2$$

(recall that here all operators commute with each other). In addition $(\widehat{F}\widehat{f})_\nu$, $\nu=1,..,4$ are the linear combinations of the functions $(F_1)^{\gamma_1}(F_2)^{\gamma_2}F^\sigma$ with $\gamma_{1,2}=0,1,2$, $\sigma=1,2$. Let us estimate the term which appears after the application of $\widetilde{\Delta}_SF^4F_1F_2G_0^3$ to the function F^2 (the other terms can be estimated similarly). Rewrite

$$\tilde{\beta}^{-2}\|\widetilde{\Delta}_SF_1F_2(F^2-z)^{-3}F^6\|^2$$
$$=C\tilde{\beta}^{-2}\int_0^\infty dx\Big|\frac{d}{dx}(x^2+x)\frac{d}{dx}\frac{(x+c_1)(x+c_2)}{n^2}\frac{e^{-3\alpha x/n}}{(e^{-\alpha x/n}-z)^3}\Big|^2,$$

where c_1 and c_2 correspond to the terms of (5.3), which do not depend on $x = |S_{12}|^2$, end $\alpha = 2c_0\alpha_2 > 0$. Changing $\tilde{x} = x/n$ we get

$$\begin{aligned}
&\tilde{\beta}^{-2} n \int_0^\infty d\tilde{x} \Big| \frac{d}{d\tilde{x}} \tilde{x}(\tilde{x}+1/n) \frac{d}{d\tilde{x}} (\tilde{x}+c_1/n)(\tilde{x}+c_2/n) \frac{e^{-3\alpha\tilde{x}}}{(e^{-\alpha\tilde{x}}-z)^3} \Big|^2 \\
\le& C\tilde{\beta}^{-2} n \int_0^\infty d\tilde{x} \Big| \frac{(\tilde{x}+c/n)^2}{|e^{-\alpha\tilde{x}}-z|^3} + \frac{(\tilde{x}+c/n)^3}{|e^{-\alpha\tilde{x}}-z|^4} + \frac{(\tilde{x}+c/n)^4}{|e^{-\alpha\tilde{x}}-z|^5} \Big|^2 e^{-6\alpha\tilde{x}} \\
\le& C\tilde{\beta}^{-2} n \int_0^\infty \frac{e^{-6\alpha\tilde{x}}}{|e^{-\alpha\tilde{x}}-z|^2} d\tilde{x} \le C(n/\tilde{\beta})^2, \quad |z| \ge 1 + A/n.
\end{aligned}$$

Here $c = \max\{|c_1|, |c_2|, 1\}$.

The second and the third inequality in (5.11) can be obtained similarly.

To obtain the bound for $\|\widehat{G}\|$, we introduce

$$\widehat{M}_1 := \widehat{F}\widehat{K}_0\widehat{F}, \quad \widehat{G}_1 := (\widehat{M}_1 - z)^{-1}$$

and prove that

$$\|\widehat{G}_1\| \le C\log^2 n/|z-1|,$$

or, equivalently,

$$\|\widehat{G}_{1,ij}\| \le C\log^2 n/|z-1|. \tag{5.12}$$

Observe that $\widehat{M}_1$ have the same form as the matrices in (5.9) with $K_{US} \to FK_{US}F$, $\widetilde{K}_i \to L_i$, where

$$\begin{aligned}
L_1 =& FK_{US}FF_1 + F_1FK_{US}F + F\widetilde{K}_1F, \qquad (5.13)\\
L_2 =& FK_{US}FF_2 + F_2FK_{US}F + F\widetilde{K}_2F \\
L_3 =& F\widetilde{K}_3F + F_1F_2FK_{US}F + FK_{US}FF_1F_2 + F_1FK_{US}FF_2 + F_2FK_{US}FF_1 \\
&+ F\widetilde{K}_1FF_2 + F_1F\widetilde{K}_2F + F\widetilde{K}_2FF_1 + F_2F\widetilde{K}_1F.
\end{aligned}$$

Then the matrix $\widehat{G}_1 := (\widehat{F}\widehat{K}_0\widehat{F} - z)^{-1}$ has zeros at the same places as in (5.9) and

$$\begin{aligned}
&\widehat{G}_{1,ii} = G := (FK_{US}F - z)^{-1}, \quad \widehat{G}_{1,1i} = \widehat{G}_{1,(4-i)4} = -GL_{i-1}G, \quad i = 2,3, \\
&\widehat{G}_{1,14} = GL_1GL_2G + GL_2GL_1G - GL_3G,
\end{aligned}$$

Since the spectrum of $FK_{US}F$ belongs to $[0,1]$, it is evident that

$$\|G_{1,ii}\| = \|G\| \le C/|z-1|. \tag{5.14}$$

To estimate the non-diagonal entries, we set

$$G_* := G(z)\Big|_{z=1+A/n}$$

and prove the bounds

$$\begin{aligned}
&\|G_*^{1/2}F\widetilde{K}_\alpha FG_*^{1/2}\| \le \|G_*^{1/2}F|\widetilde{K}_\alpha|FG_*^{1/2}\| \le C, \quad \|G_*^{-1/2}G^{1/2}\| \le C, \quad \alpha = 1,2,3, \qquad (5.15)\\
&\|G_*^{1/2}F_\alpha FK_{US}FG_*^{1/2}\| \le C\log n, \quad \alpha = 1,2 \qquad (5.16)\\
&\|G_*^{1/2}F_1F_2FK_{US}FG_*^{1/2}\| \le C\log^2 n, \quad \|G_*^{1/2}F_1FK_{US}FF_2G_*^{1/2}\| \le C\log^2 n \\
&\|G_*^{1/2}F_\alpha^2FG_*^{1/2}\| \le C\log^2 n, \quad \alpha = 1,2.
\end{aligned}$$

It is easy to see from (5.13) that $GL_{1,2}G$, GL_1GL_2G and GL_2GL_1G can be represented as a linear combination of the terms $G^{1/2}\Pi G^{1/2}$, where Π is some product of the operators whose bounds are given in (5.15) and the first line of (5.16) or operators similar to them (e.g., $G_*^{1/2}FK_{US}FF_\alpha G_*^{1/2}$ instead of $G_*^{1/2}F_\alpha FK_{US}FG_*^{1/2}$, etc.). For instance,

$$GFK_{US}FF_1G = G^{1/2}\cdot(G^{1/2}G_*^{-1/2})\cdot(G_*^{1/2}FK_{US}FF_1G_*^{1/2})\cdot(G_*^{-1/2}G^{1/2})\cdot G^{1/2}.$$

Therefore (5.14) and the first line of (5.16) yield

$$\|GL_{1,2}G\| \le C\log n\cdot\|G\| \le C\log n/|z-1|, \quad \|GL_1GL_2G\|+\|GL_1GL_2G\| \le C\log^2 n/|z-1|.$$

To estimate GL_3G, we use the bounds from the last two lines of (5.16), combined with the inequality (recall that G_* and F are self-adjoint, and F commutes with F_2)

$$\begin{aligned}\|G_*^{1/2}F\widetilde{K}_1FF_2G_*^{1/2}\| \le&\|G_*^{1/2}F\widetilde{K}_1F\widetilde{K}_1^*FG_*^{1/2}\|^{1/2}\cdot\|G_*^{1/2}F_2F\bar{F}_2G_*^{1/2}\|^{1/2} \\ \le&\|G_*^{1/2}F|\widetilde{K}_1|FG_*^{1/2}\|^{1/2}\cdot\|G_*^{1/2}|F_2|^2FG_*^{1/2}\|^{1/2}.\end{aligned}\tag{5.17}$$

The terms in the r.h.s. above can be estimated with the first inequality of (5.15) and the last inequality of (5.16). In the last inequality of (5.17) we used that since $F\le 1$ and $\widetilde{K}_1\widetilde{K}_1^*\le c\cdot|\widetilde{K}_1|$,

$$G_*^{1/2}F\widetilde{K}_1F\widetilde{K}_1^*FG_*^{1/2} \le G_*^{1/2}F\widetilde{K}_1\widetilde{K}_1^*FG_*^{1/2} \le c\cdot G_*^{1/2}F|\widetilde{K}_1|FG_*^{1/2}.$$

The expression $\|G_*^{1/2}F\widetilde{K}_2FF_1G_*^{1/2}\|$ can be estimated similarly.

Now we are left to show (5.15) – (5.16). To prove the first inequality of (5.15), we recall first that for any normal A and B

$$|B^*AB\| \le \|B^*|A|B\| \tag{5.18}$$

Indeed, for any normal A we have

$$|(Ax,y)|^2 \le (|A|x,x)(|A|y,y).$$

and so putting Bx and By instead of x and y we get (5.18).

Now (5.18), the first inequality in (5.10), and the bound $F\le 1$ yield

$$\begin{aligned}&F|\widetilde{K}_\alpha|F \le F(1-K_{US})F \le 1-FK_{US}F \\ &\qquad\Rightarrow \|G_*^{1/2}F|\widetilde{K}_\alpha|FG_*^{1/2}\| \le \|G_*^{1/2}(1-FK_{US}F)G_*^{1/2}\| \le C,\end{aligned}$$

since the spectrum of $FK_{US}F$ belongs to $[0,1]$ and

$$\max_{0\le\lambda\le 1}\frac{1-\lambda}{1+A/n-\lambda} \le 1. \tag{5.19}$$

Moreover, since G and G_* commute we have

$$\|G^{1/2}(z)G_*^{-1/2}\|^2 = \|G(z)G_*^{-1}\| \le \max_{|z|=1+A/n,0\le\lambda\le 1}\frac{1+A/n-\lambda}{|z-\lambda|} \le C,$$

which gives the second inequality of (5.15).

To prove the first inequality (5.16), take n-independent $B>0$ and introduce the projection

$$\Pi_n = \mathbf{1}_{|S_{12}|^2\le Bn\log n}.$$

From the definition (5.3) it is evident that for sufficiently big B we can write

$$\begin{aligned}
&\|(1-\Pi_n)FF_\alpha\| \le C \max_{x>B'\log n} xe^{-x} \le C/n^3, \\
&0 \le \frac{c_0\alpha_2|S_{12}|^2}{n}(1-F^2)^{-1}\Pi_n = \max_{0\le x\le B'\log n} x(1-e^{-2x})^{-1} \le C(B)\log n \\
\Rightarrow &|F_\alpha|\Pi_n \le C/n + C(B)\log n(1-F^2) \le C/n + C(B)\log n(1-FK_{US}F)
\end{aligned} \tag{5.20}$$

with $B' = c_0\alpha_2 B$. Using the first inequality above, the bound $\|G_*\| \le Cn$, and the fact that $FK_{US}F$ commute with G_*, we get

$$\begin{aligned}
G_*^{1/2}F_\alpha FK_{US}FG_*^{1/2} &= G_*^{1/2}F_\alpha((1-\Pi_n)+\Pi_n)FK_{US}FG_*^{1/2} \\
&= O(n^{-2}) + G_*^{1/2}F_\alpha\Pi_n G_*^{1/2}FK_{US}F.
\end{aligned}$$

In addition the third line of (5.20) and (5.19) yield

$$G_*^{1/2}|F_\alpha|\Pi_n G_*^{1/2} \le C + C'\log n\, G_*^{1/2}(1-FK_{US}F)G_*^{1/2} \le C\log n.$$

The proofs of the other inequalities of (5.16) are similar to the proof of the first one.

Thus we obtain (5.12). Since by (5.8) $\hat{K} = \hat{K}_0 + O(\tilde{\beta}^{-1})$, we have

$$\widehat{G} = \widehat{G}_1(1+O(\tilde{\beta}^{-1})\widehat{G}_1)^{-1} = \widehat{G}_1\big(1+O(\log^2 n(n/\tilde{\beta}))\big).$$

Combined with (5.12) the relation finishes the proof of Lemma 5.1. $\square$

6 Proof of Proposition 5.1.

We start with a detailed study of the operator H of (5.2). Set

$$U = U_1U_2^*, \quad S = S_1S_2^{-1}$$

and use two simple formulas, valid for any diagonal 2×2 matrices A and B,

$$\begin{aligned}
\operatorname{Tr} AUBU^* &= \operatorname{Tr} AB - |U_{12}|^2(A_{11}-A_{22})(B_{11}-B_{22}), \\
\operatorname{Tr} ASBS^{-1} &= \operatorname{Tr} AB + |S_{12}|^2(A_{11}-A_{22})(B_{11}-B_{22}).
\end{aligned}$$

Using (1.12) and changing

$$\begin{aligned}
&\hat{\rho} \to \tilde{\beta}^{-1/2}\hat{\rho}, \quad \hat{\rho}' \to \tilde{\beta}^{-1/2}\hat{\rho}' \\
&\hat{\tau} \to \tilde{\beta}^{-1/2}\hat{\tau}, \quad \hat{\tau}' \to \tilde{\beta}^{-1/2}\hat{\tau}',
\end{aligned} \tag{6.1}$$

in (5.2) (note that this gives the Jacobian $\tilde{\beta}^2$), we get

$$\begin{aligned}
H =&\tilde{\beta}^2 \cdot e^{-\tilde{\beta}\cdot w}(1-n_1n_2/\tilde{\beta}^2)(1-n_1'n_2'/\tilde{\beta}^2) \\
&\cdot \exp\Big\{(n_1+n_2+n_1'+n_2')d - \operatorname{Tr}\hat{\rho}U\hat{\tau}'S^{-1} - \operatorname{Tr}\hat{\rho}'U^*\hat{\tau}S - (n_1+n_2)(n_1'+n_2')w/\tilde{\beta}\Big\} \\
d =&1-|U_{12}|^2+|S_{12}|^2, \quad w = |U_{12}|^2+|S_{12}|^2.
\end{aligned}$$

Writing

$$\begin{aligned}&\mathrm{Tr}\,\hat\rho U\hat\tau' S^{-1}=(\rho,A\tau'),\quad \mathrm{Tr}\,\hat\rho' U^*\hat\tau S=(\rho',B\tau)\\ &A_{ij}=U_{ij}S^{-1}_{ji},\quad B_{ij}=U^*_{ij}S_{ji}\end{aligned}$$

and using that

$$(\rho,A\tau')^2=-2\det A\,\rho_1\rho_2\tau_1'\tau_2',\quad (\rho',B\tau)^2=-2\det B\,\rho_1'\rho_2'\tau_1\tau_2,$$

we obtain

$$\begin{aligned}H\big|_{\mathcal{P}_6}=&\tilde\beta^2\cdot\exp\Big\{-\tilde\beta\cdot w+(n_1+n_2+n_1'+n_2')d-(n_1+n_2)(n_1'+n_2')w/\tilde\beta\Big\}\\ &\times\big(1+(\rho,A\tau')(\rho',B\tau)+\det A\cdot\det B\,n_1n_2n_1'n_2'\big)\cdot(1-n_1n_2/\tilde\beta^2)(1-n_1'n_2'/\tilde\beta^2).\end{aligned}\tag{6.2}$$

Introduce the basis $e_1=1$, $e_2=n_1$, $e_3=n_2$, $e_4=n_1n_2$, $e_5=\rho_1\tau_2$, $e_6=\rho_2\tau_1$ of $\mathcal{P}_6$. Denote the space spanned on the first 4 vectors as $\mathcal{P}_4$ and represent H in this basis by the block 6×6 matrix with $H^{(11)}$ corresponding to the projection on $\mathcal{P}_4$. Then using (6.2) we obtain

$$H=\begin{pmatrix}H^{(11)}&H^{(12)}\\ H^{(21)}&H^{(22)}\end{pmatrix},\quad H^{(22)}=K_{US}\begin{pmatrix}A_{11}B_{22}&A_{12}B_{12}\\ A_{21}B_{21}&A_{22}B_{11}\end{pmatrix}=\begin{pmatrix}h_{11}&h_{12}\\ h_{21}&h_{22}\end{pmatrix},\tag{6.3}$$

$$H^{(21)}=\begin{pmatrix}2x_d&x&x&0\\ -2\overline{x}_d&-\overline{x}&-\overline{x}&0\end{pmatrix},\quad H^{(12)}=\begin{pmatrix}0&0\\ y&-\overline{y}\\ y&-\overline{y}\\ 2y_d&-2\overline{y_d}\end{pmatrix}.$$

Here and below h_{ij},x,y,x_d,y_d are "difference" operators whose kernels are defined with the functions

$$\begin{aligned}h_{ij}=&h_{ijU}h_{ijS},\quad h_{ijU}=U_{ij}^2K_U,\quad h_{ijS}=\bar S_{ij}^2K_S\\ x=&x_Ux_S,\quad x_U=U_{11}U_{12}K_U,\quad x_S=\bar S_{11}\bar S_{12}K_S,\quad x_d=x\cdot d,\\ y=&y_Uy_S,\quad y_U=U_{11}\bar U_{12}K_U\quad y_S=\bar S_{11}S_{12}K_S,\quad y_d=y\cdot d,\end{aligned}\tag{6.4}$$

and $\bar x,\bar y,\bar x_d,\bar y_d$ mean the complex conjugate kernels. Now let us study the structure of $H^{(11)}$. Using (6.2) and the relations

$$\det A=\det B=d,\quad (A\rho,\tau')(B\rho',\tau)\big|_{\mathcal{P}_4}=-d(n_1n_1'+n_2n_2')+|U_{12}|^2|S_{12}|^2(n_1+n_2)(n_1'+n_2')$$

we continue to transform H as

$$\begin{aligned}H\big|_{\mathcal{P}_4}=&K_{US}\cdot e^{d(n_1+n_2+n_1'+n_2')}\big(1-w(n_1+n_2)(n_1'+n_2')/\tilde\beta+2w^2n_1n_2n_1'n_2'/\tilde\beta^2\big)\\ &\times\big(1-(n_1n_2+n_1'n_2')/\tilde\beta^2+n_1n_2n_1'n_2'/\tilde\beta^4\big)\\ &\times\big(1-d(n_1n_1'+n_2n_2')+|U_{12}|^2|S_{12}|^2(n_1+n_2)(n_1'+n_2')+d^2n_1n_2n_1'n_2'\big)\\ =&K_{US}\cdot e^{d(n_1+n_2+n_1'+n_2')}\Big(1-w(n_1+n_2)(n_1'+n_2')/\tilde\beta-(n_1n_2+n_1'n_2')/\tilde\beta^2\\ &-d(n_1n_1'+n_2n_2')+|U_{12}|^2|S_{12}|^2(n_1+n_2)(n_1'+n_2')\\ &+(d^2+2w^2/\tilde\beta^2+2dw/\tilde\beta+1/\tilde\beta^4-4w|U_{12}|^2|S_{12}|^2/\tilde\beta)n_1n_2n_1'n_2'\Big).\end{aligned}$$

Represent $H^{(11)} = K_{US} \cdot K$ and observe that to find the coefficients of K we can represent H as a polynomials with respect to n_1, n_2, n_1', n_2' and the coefficients of this polynomials gives the coefficients of K. In particular,

$$K_{11} \sim n_1' n_2', \quad K_{21} \sim n_1 n_1' n_2', \quad K_{31} \sim n_2 n_1' n_2', \quad K_{41} \sim n_1 n_2 n_1' n_2',$$
$$K_{42} \sim n_1 n_2 n_2', \quad K_{43} \sim n_1 n_2 n_1', \quad K_{24} \sim n_1, \quad K_{34} \sim n_2, \quad K_{44} \sim n_1 n_2,$$

Evidently these and the other coefficient of K can be found as the respective derivatives, taken at the point $(n_1, n_2, n_1', n_2') = (0, 0, 0, 0)$.

Now we return to the proof of Proposition 5.1. In order to transform (5.6) to (5.7) with an appropriate $\widehat{M}$ and $\widehat{K}$ satisfying (5.8) – (5.9) we are going to consider the matrix K after the transformation

$$K_T = TKT, \quad T = \begin{pmatrix} 0 & 0 & 0 & \tilde{\beta} \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ \tilde{\beta}^{-1} & 0 & 0 & 0 \end{pmatrix}$$

It is easy to see that

$$K_{T12} = \tilde{\beta} K_{43},\ K_{T13} = \tilde{\beta} K_{42},\ K_{T24} = \tilde{\beta} K_{31},\ K_{T34} = \tilde{\beta} K_{21},\ K_{T14} = \tilde{\beta}^2 K_{41}.$$

All the rest coefficients K change the places or are multiplied by 1, $\tilde{\beta}^{-1}$ or even $\tilde{\beta}^{-2}$. Thus, to obtain representation (5.7)- (5.9), we need to control the elements of $\widehat{K}$ written above. The following lemma allows to understand the order of the operators, which will appear in the coefficients of K.

Lemma 6.1.

$$K_{US}|U_{12}|^2 = \tilde{\beta}^{-1} + O((1 - K_{US})\tilde{\beta}^{-1}), \quad K_{US}|S_{12}|^2 = \tilde{\beta}^{-1} + O((1 - K_{US})\tilde{\beta}^{-1}), \tag{6.5}$$
$$K_{US}|U_{12}|^4 = 2\tilde{\beta}^{-2} + O((1 - K_{US})\tilde{\beta}^{-2}), \quad K_{US}|S_{12}|^4 = 2\tilde{\beta}^{-2} + O((1 - K_{US})\tilde{\beta}^{-2}),$$
$$K_{US}|U_{12}|^2|S_{12}|^2 = \tilde{\beta}^{-2} + O((1 - K_{US})\tilde{\beta}^{-2}).$$

We recall that all operators here are self adjoint and commute with each other, hence the relations mean the ones for the corresponding eigenvalues.

The proof or the lemma will be given at the end of the proof of Lemma 6.2 (see the argument above (6.17)).

Coming back to the coefficients of K, compute first

$$\begin{aligned} K_{US} \cdot K_{41} &= K_{US} \cdot \frac{\partial^4 K}{\partial n_1 \partial n_2 \partial n_1' \partial n_2'}\Big|_{(0,0,0,0)} = K_{US}\big(d^4 - 2d^3 + d^2 + 4d^2|U_{12}|^2|S_{12}|^2 \\ &\quad - 4d^2 w\tilde{\beta}^{-1} - 2d^2\tilde{\beta}^{-2} + 2dw\tilde{\beta}^{-1} + 2w^2\tilde{\beta}^{-2} - 4w|U_{12}|^2|S_{12}|^2/\tilde{\beta} + \tilde{\beta}^{-4}\big) \\ &= K_{US}\Big(d^2w^2 - 2d^2\tilde{\beta}^{-2} - 2dw\tilde{\beta}^{-1}\Big) + O(\tilde{\beta}^{-3}) \\ &= \tilde{\beta}^{-2}\widetilde{K}', \quad \widetilde{K}' = O(1 - K_{US}). \end{aligned}$$

Here we have used the relation (which follows from the definition of d and w)

$$d^4 - 2d^3 + d^2 + 4d^2|U_{12}|^2|S_{12}|^2 = d^2w^2, \quad 4d^2w - 4dw = 4dw(|S_{12}|^2 - |U_{12}|^2),$$

and the lemma above.

Similarly

$$\begin{aligned}&K_{US}\cdot K_{21}=K_{US}\cdot K_{31}=K_{US}\cdot K_{42}=K_{US}\cdot K_{43}\\&=K_{US}(d^3-d^2-2dw/\tilde\beta-d/\tilde\beta^2+2d|U_{12}|^2|S_{12}|^2)=\tilde\beta^{-1}\widetilde K\end{aligned}$$

with

$$\widetilde K=O(1-K_{US}).$$

In addition,

$$\begin{aligned}&K_{ii}=1+O(\tilde\beta^{-1}),\quad i=1,\dots 4,\\&K_{ij}=O(\tilde\beta^{-1}),\quad (i,j)=(2,3)\ or\ (3,2).\end{aligned}$$

Observe now that the operator $\tilde{\mathcal F}=\mathcal F\big|_{\mathcal P_6}$ in (5.6) after the change (6.1) in our basis have the block diagonal form, where a 4×4 upper left block has the form $T\widehat FT$, where $\widehat F$ is given by (5.9), and a 2×2 bottom left block is I. In addition, $\tilde f$ and $\tilde g$ are spanned on e_1,e_2,e_3,e_4 and after the change (6.1) their restriction on $\mathcal P_4$ have the form $\tilde f=\beta^{-1}T\hat f$, $\tilde g=\beta^{-1}T\hat g$. Thus we are interested in the upper left block $G^{(11)}$ of the resolvent $G=(\tilde{\mathcal F}H\tilde{\mathcal F}-z)^{-1}$, and so (5.6) yields

$$\mathcal R^{+-}_{n\beta}(E,\varepsilon,\xi)=\frac{C_{E,\varepsilon}e^{c_0(\alpha_1-\alpha_2)}}{2\pi i}\oint_{\omega_A}z^{n-1}(T\widehat G^{(11)}(z)T\widehat f,\widehat g)dz.$$

But by the Schur compliment formula

$$TG^{(11)}(z)T=\Big(\widehat FT\big(H^{(11)}-H^{(12)}(H^{(22)}-z)^{-1}H^{(21)}\big)T\widehat F-z\Big)^{-1},$$

and so we are left to prove that

$$\widehat M=\widehat FT(H^{(11)}-H^{(12)}(H^{(22)}-z)^{-1}H^{(21)})T\widehat F \tag{6.6}$$

satisfies (5.8) – (5.9).

According to the consideration above, $TH^{(11)}T$ has the form (5.8) – (5.9). The estimate on $H^{(12)}(H^{(22)}-z)^{-1}H^{(21)}$ is given in the following lemma

Lemma 6.2. *Set $G^{(2)}(z):=(H^{(22)}-z)^{-1}$. Then for any $z:|z|=1+A/n$ the operator $H^{(12)}G^{(2)}(z)H^{(21)}$ has the form*

$$H^{(12)}G^{(2)}H^{(21)}=\begin{pmatrix}0&0&0&0\\2R_{1d}&R&R&0\\2R_{1d}&R&R&0\\4R_{dd}&2R_{d1}&2R_{d1}&0\end{pmatrix},$$

where

$$R=yG^{(2)}_{11}x+\overline yG^{(2)}_{22}\overline x-yG^{(2)}_{12}\overline x-\overline yG^{(2)}_{21}x, \tag{6.7}$$

R_{1d} can be obtained from R, if we replace x with x_d, to obtain R_{d1}, one should replace y with y_d, to obtain R_{dd}, one should replace x,y with x_d y_d, and the operators x,y,x_d,y_d are the same as in (6.3).

The operators R,R_{1d},R_{d1},R_{dd} are normal and satisfy the bound

$$|R|+|R_{1d}|+|R_{d1}|+|R_{dd}|\le C\tilde\beta^{-2}(1-K_{US})+O(\tilde\beta^{-3}),\quad 1-K_{US}\le C(\widetilde\Delta_U+\widetilde\Delta_S)/\tilde\beta \tag{6.8}$$

The lemma gives that (6.6) indeed satisfies (5.8) – (5.9), and (5.10), which finishes the proof of Proposition 5.1.

□

Proof of Lemma 6.2. Let us prove (6.8) for R of (6.7). For R_{1d}, R_{d1}, R_{dd} the proof is the same. To simplify notations set

$$H^{(22)} = h = \hat{h} + \tilde{h},$$

where $\hat{h}$ is the diagonal part of $H^{(22)}$, and $\tilde{h}$ is its off diagonal part, and denote

$$G_0^{(2)} := (\hat{h} - z)^{-1}.$$

It is easy to see that

$$\|h_{12}\| \le \int |U_{12}|^2 |S_{12}|^2 K_U K_S dU dS \le \tilde{\beta}^{-2}, \quad \|x\| \le \tilde{\beta}^{-1}, \quad \|y\| \le \tilde{\beta}^{-1}, \tag{6.9}$$

(recall that by (6.3) $h_{ij} = U_{ij}^2 \bar{S}_{ij}^2 K_U K_S$). Hence, writing

$$G^{(2)} = G_0^{(2)} - G_0^{(2)} \tilde{h} G_0^{(2)} + r, \quad r := G_0^{(2)} \tilde{h} G_0^{(2)} \tilde{h} G^{(2)},$$

and using the bounds above combined with (6.3), we get

$$\|r\| \le C n^3 \tilde{\beta}^{-4} \quad \Rightarrow \quad \|H^{(12)} r H^{(21)}\| \le n^3 \tilde{\beta}^{-6} < \tilde{\beta}^{-3}.$$

Consider $\widehat{R}$ which has the same form as (6.7) but with $G^{(2)}$ replaced by $G_0^{(2)}$. Then the second two terms become zeros and

$$\widehat{R} = y(h_{11} - z)^{-1} x + \bar{y}(h_{22} - z)^{-1} \bar{x} = \widehat{R}_1 + \widehat{R}_2.$$

Let us study the operator

$$\widehat{R}_1 = \sum_{p=0}^{\infty} \frac{y(h_{11})^p x}{z^p} = \sum_{p=0}^{\infty} \frac{(y_U (h_{11U})^p x_U) \otimes (y_S (h_{11S})^p x_S)}{z^p},$$

where y_U, h_{11U}, x_U (see (6.4)) are integral operators on $L_2(U)$ with the "difference" kernels of the form $v(U_1 U_2^{-1})$, and y_S, h_{11S}, x_S are the "difference" integral operators on $L_2(S)$. Here $L_2(U)$ and $L_2(S)$ denote the subspaces of even functions $\varphi(U) = \varphi(-U)$ (or $\varphi(S) = \varphi(-S)$). Since our operators preserve the evenness, it suffices to study only these subspaces. It is known that

$$L_2(U) = \oplus_{l=0}^{\infty} L^{(l)U}, \quad L^{(l)U} = \mathrm{Lin}\,\{t_{mk}^{(l)U}\}_{m,k=-l}^{l}$$

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

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

According to [26], Chapter III,

$$t_{mk}^{(l)U}(U) = e^{-i(m\phi + k\psi)/2} P_{mk}^{(l)}(\theta),$$

where

$$P^{(l)}_{mk}(\cos\theta)=\frac{c_{mk}}{2\pi}\int_0^{2\pi}d\varphi(\cos(\theta/2)+i\sin(\theta/2)e^{i\varphi})^{l+k}(\cos(\theta/2)+i\sin(\theta/2)e^{-i\varphi})^{l-k}e^{i(m-k)\varphi},$$

$$c_{mk}=\Big(\frac{(l-m)!(l+m)!}{(l-k)!(l+k)!}\Big)^{1/2},\quad U=\left(\begin{array}{cc}\cos(\theta/2)e^{i(\phi+\psi)/2} & i\sin(\theta/2)e^{i(\phi-\psi)/2}\\ i\sin(\theta/2)e^{-i(\phi-\psi)/2} & \cos(\theta/2)e^{-i\phi+\psi)/2}\end{array}\right). \tag{6.10}$$

In addition (see [26], Chapter III),

$$P^{(l)}_{mm}(1-x)=1-x(l+m)(l+m+1)/2+O(x^2). \tag{6.11}$$

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

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

$$v^{(l)U}_{mk}:=\int v(U)\overline{t^{(l)U}_{mk}(U)}dU.$$

It is easy to see that if we consider an integral operator $\widehat{v}$ with the kernel $v(U_1U_2^{-1})$, then for any $\varphi(U)\in L^{(l)U}$

$$\begin{aligned}(\widehat{v}\varphi)(U)&=\int v(UU_1^{-1})\sum_{mk}\varphi_{mk}t^{(l)U}_{mk}(U_1)dU_1=\int v(\widetilde{U})\sum_{mk}\varphi_{mk}t^{(l)U}_{mk}(\widetilde{U}^{-1}U)d\widetilde{U}\\ &=\int v(\widetilde{U})\sum_{mkj}\varphi_{mk}t^{(l)U}_{mj}(\widetilde{U}^{-1})t^{(l)U}_{jk}(U)d\widetilde{U}=\sum v^{(l)U}_{jm}\varphi_{mk}t^{(l)U}_{jk}(U).\end{aligned}$$

Hence, denoting Π_l the orthogonal projection on $L^{(l)U}$, one can see that $L^{(l)U}$ reduces $\widehat{v}$ and $\widehat{v}^{(l)U}=\Pi_l\widehat{v}\Pi_l$ is uniquely defined by the matrix $v^{(l)U}$. Moreover, for any functions v_1 and v_2 it is evident that $\widehat{v}_1\widehat{v}_2$ is also a "difference" operator, hence it commutes with Π_l, and if the matrices $v^{(l)U}_1$ and $v^{(l)U}_2$ correspond to $\widehat{v}_1$ and $\widehat{v}_2$, then

$$(\widehat{v}_1\widehat{v}_2)^{(l)U}=v^{(l)U}_1v^{(l)U}_2.$$

Let us find the matrices, corresponding to h_{11U},y_U,x_U (see (6.4)) in $L^{(l)U}$. Using (6.10) it is easy to see that

$$\begin{aligned}(x^{(l)U}_U)_{km}&=\int U_{11}U_{12}K_U(|U_{12}|^2)\overline{t^{(l)U}_{mk}(U)}dU=\tilde{\beta}\int_0^{\pi}\sin\theta d\theta\int_0^{2\pi}d\phi d\psi\sin(\theta/2)\cos(\theta/2)\\ &\times e^{-\tilde{\beta}\sin^2(\theta/2))}e^{i\phi}e^{im\phi+in\psi}\overline{P^{(l)}_{-1,0}(\cos\theta)}=\delta_{m,-1}\delta_{n,0}\lambda^{(l)U}_{-1,0},\end{aligned}$$

where we set

$$\lambda^{(l)U}_{-1,0}=\frac{\tilde{\beta}}{2}\int_0^{\pi}e^{-\tilde{\beta}\sin^2(\theta/2)}P^{(l)}_{-1,0}(\cos\theta)\sin^2\theta d\theta. \tag{6.12}$$

Hence, denoting E_{ij} the matrix which has only ijth entry equal to 1, and all other entries equal to 0, we get

$$x^{(l)U}_U=E_{-1,0}\lambda^{(l)U}_{-1,0}.$$

Introduce also the eigenvalues $\lambda^{(l)U}$ of K_U. Repeating the argument above, we have

$$K_U^{(l)U} = E_{0,0}\lambda^{(l)U},$$

where, using (6.11), we obtain

$$\begin{aligned}\lambda^{(l)U} =&\tilde\beta \int_0^\pi e^{-\tilde\beta \sin^2(\theta/2)} P_{0,0}^{(l)}(\cos\theta)\sin\theta d\theta = \tilde\beta^{-1}\int_0^1 e^{-\tilde\beta x} P_{00}^{(l)}(1-2x)dx \\ =&1 - l(l+1)/\tilde\beta + O(l^4/\tilde\beta^2).\end{aligned} \tag{6.13}$$

To find an asymptotic behaviour of $\lambda_{-1,0}^{(l)U}$, observe that formulas (6.10) and (6.11) yield

$$\begin{aligned}&2i(1+1/l)^{1/2}\sin(\theta/2)\cos(\theta/2)P_{-1,0}^{(l)}(\cos\theta)\\ =&\int_0^{2\pi}\frac{d\varphi}{2\pi}(\cos(\theta/2)+i\sin(\theta/2)e^{i\varphi})^l\cos(\theta/2)+i\sin(\theta/2)e^{-i\varphi})^l 2i\cos\varphi\sin(\theta/2)\cos(\theta/2)\\ =&\int_0^{2\pi}\frac{d\varphi}{2\pi}(\cos(\theta/2)+i\sin(\theta/2)e^{i\varphi})^l(\cos(\theta/2)+i\sin(\theta/2)e^{-i\varphi})^l\\ &\times\Big((\cos(\theta/2)+i\sin(\theta/2)e^{i\varphi})(\cos(\theta/2)+i\sin(\theta/2)e^{-i\varphi})-\cos^2(\theta/2)+\sin^2(\theta/2)\Big)\\ =&P_{0,0}^{(l+1)}(\cos\theta)-P_{0,0}^{(l)}(\cos\theta)(1-2\sin^2(\theta/2)).\end{aligned}$$

Hence

$$\begin{aligned}(1+1/l)^{1/2}\lambda_{-1,0}^{(l)U} =&(\lambda^{(l+1)U}-\lambda^{(l)U})/2+O(\tilde\beta^{-1}) = -(l+1)/\tilde\beta + O(l^2\tilde\beta^{-2})+O(\tilde\beta^{-1})\\ \Rightarrow&|\lambda_{-1,0}^{(l)U}|^2 \le C_0(1-\lambda^{(l)U})/\tilde\beta.\end{aligned}$$

Similarly

$$y_U^{(l)U} = E_{0,-1}\overline{\lambda_{-1,0}^{(l)U}},\quad h_{11U}^{(l)U} = E_{-1,-1}\lambda_{-1,-1}^{(l)U},$$

where we set

$$\lambda_{-1,-1}^{(l)U} = \tilde\beta\int_0^\pi e^{-\tilde\beta\sin^2(\theta/2)}\cos^2(\theta/2)P_{-1,-1}^l(\cos\theta)\sin\theta d\theta = \lambda^{(l)U}+O(\tilde\beta^{-1}) \tag{6.14}$$

and in the last relation we used (6.11). Thus, for any p

$$(y_U(h_{11U})^p x_U)^{(l)U} = |\lambda_{-1,0}^{(l)U}|^2(\lambda_{-1,-1}^{(l)U})^p E_{00}. \tag{6.15}$$

The analysis of $(y_S(h_{11S})^p x_S)$ is very similar, the difference is that for the hyperbolic group the irreducible representations are labelled by the continuous parameter $l' = -\frac{1}{2}+i\rho$, $\rho\in\mathbb{R}$,

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

and $\mathcal{B}_{mk}^{(l)}(\theta)$ has the form (6.10) 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 [26], Chapter VI) . Then the same argument yields that

$$\begin{aligned}&(y_S(h_{11V})^p x_S)^{(l')} = |\lambda_{-1,0}^{(l)S}|^2(\lambda_{-1,-1}^{(l')S})^p E_{00}\\ &|\lambda_{-1,0}^{(l')S}|^2 \le C_0(1-\lambda^{(l')S})/\tilde\beta,\quad \lambda_{-1,-1}^{(l')S} = \lambda^{(l')S}+O(\tilde\beta^{-1}),\end{aligned} \tag{6.16}$$

where $\lambda^{(l)S}_{-1,0}$, $\lambda^{(l')S}$, and $\lambda^{(l')S}_{-1,-1}$, are defined similarly to (6.12), (6.13), and (6.14). Here the bound for $|O(\tilde\beta^{-1})| < C_0\tilde\beta^{-1}$ is uniform in l.

This relation combined with (6.15) yields that $\widehat{R}_1 : L^{(l)U}\otimes L^{(l')S} \to L^{(l)U}\otimes L^{(l')S}$ and the only non zero eigenvalue of $\widehat{R}_1$ in this subspace has the form

$$\lambda^{(ll')} = |\lambda^{(l)U}_{-1,0}|^2|\lambda^{(l')S}_{-1,0}|^2(z-\lambda^{(l)U}_{-1,-1}\lambda^{(l')S}_{-1,-1})^{-1}$$

The bounds (6.15) and (6.16) yield for $|z|>1+2C_0\tilde\beta^{-1}$

$$|\lambda^{(ll')}| \le C\tilde\beta^{-2}\frac{(1-\lambda^{(l)U})\cdot(1-\lambda^{(l')S})}{|z|-\lambda^{(l)U}\lambda^{(l')S}+O(\tilde\beta^{-1})} \le C\tilde\beta^{-2}|1-\lambda^{(l)U}\lambda^{(l')S}|.$$

Here we have used that for any $0<a,b<1$

$$ab \le a^2+b^2-ab < a^2+b^2-a^2b^2,$$

hence, taking $a^2 = 1-\lambda^{(l)U}$, $b^2 = 1-\lambda^{(l')S}$, we obtain the last inequality for $|\lambda^{(ll')}|$.

Note that (6.13) and a similar relation for $\lambda^{(l')S}$ combined with the facts that

$$\widetilde\Delta_U L^{(l)U} = l(l+1)L^{(l)U}, \quad \widetilde\Delta_S L^{(l')S} = -l'(l'+1)L^{(l')S}$$

prove the second inequality in (6.8).

Assertions of Lemma 6.1 can be obtained from the fact that the operators in the l.h.s. of (6.5) are tensor products of the "difference" operators on $L_2(U)$ and $L_2(S)$. Hence they are reduced by $L^{(l)U}\otimes L^{(l')S}$, and since the kernels depend on $|U_{12}|^2$ and $|S_{12}|^2$, the corresponding matrices have the form $\mu^{(l)}\nu^{(l')}E_{00}\otimes E_{00}$, where $\mu^{(l)},\nu^{(l')}$ -are corresponding eigenvalues. For example, for the first operator in (6.5) $\nu^{(l')} = \lambda^{(l')S}$ and

$$\begin{aligned}\mu^{(l)} =& \tilde\beta\int_0^\pi \sin^2(\theta/2)e^{-\tilde\beta\sin^2(\theta/2)}P^{(l)}_{0,0}(\cos\theta)\sin\theta d\theta = \tilde\beta^{-1}\int_0^1 xe^{-\tilde\beta x}P^{(l)}_{00}(1-2x)dx \\ =& 1/\tilde\beta - 2l(l+1)/\tilde\beta^2 + O(l^4/\tilde\beta^3).\end{aligned} \tag{6.17}$$

The first relation of (6.5) follows from the above one combined with the analogue of (6.13) for $\lambda^{(l')S}$. The other relations of (6.5) can be obtained similarly.

To complete the proof of the lemma we are left to consider the part of R which can be obtained if we replace $G^{(2)}$ with $G^{(2)}_0\tilde hG^{(2)}_0$. For this replacement the first two terms of (6.7) are zero. Set

$$R_3 = y(h_{11}-z)^{-1}h_{12}(h_{22}-z)^{-1}\bar x.$$

Repeating the above argument we obtain that $R_3 : L^{(l)U}\otimes L^{(l')S} \to L^{(l)U}\otimes L^{(l')S}$ and the only non zero eigenvalue $\widetilde\lambda^{(ll')}$ of R_3 in this subspace has the form

$$\widetilde\lambda^{(ll')} = \frac{|\lambda^{(l)U}_{-1,0}|^2|\lambda^{(l')S}_{-1,0}|^2\lambda^{(l)U}_{-1,1}\lambda^{(l')S}_{-1,1}}{(z-\lambda^{(l)U}_{-1,-1}\lambda^{(l')S}_{-1,-1})^2},$$

where $\lambda^{(l)U}_{-1,1}$ and $\lambda^{(l')S}_{-1,1}$ by (6.9) satisfy the trivial bound

$$|\lambda^{(l)U}_{-1,1}\lambda^{(l)S}_{-1,1}| \le \|h_{12}\| \le \tilde\beta^{-2}.$$

The bound, (6.15) and (6.16) yield

$$|\widetilde{\lambda}^{(ll')}| \le C\tilde{\beta}^{-4}\frac{|1-\lambda^{(l)U}|\cdot|1-\lambda^{(l)V}|}{||z|-\lambda^{(l)U}\lambda^{(l')V}+O(\tilde{\beta}^{-1})|^2} \le Cn/\tilde{\beta}^4 < \tilde{\beta}^{-3}.$$

The same bound is valid for

$$R_4 = \bar{y}(h_{22}-z)^{-1}h_{21}(h_{11}-z)^{-1}x.$$

These bounds complete the proof of the lemma for R. For R_{1d}, R_{d1}, R_{dd} the proof is the same. □

7 Appendix

7.1 Grassmann integration

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

$$\psi_j\psi_k + \psi_k\psi_j = \overline{\psi}_j\psi_k + \psi_k\overline{\psi}_j = \overline{\psi}_j\overline{\psi}_k + \overline{\psi}_k\overline{\psi}_j = 0, \quad j,k = 1,\dots,n. \tag{7.1}$$

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

$$\chi = a + \sum_{j=1}^n (a_j\psi_j + b_j\overline{\psi}_j) + \sum_{j\ne k}(a_{j,k}\psi_j\psi_k + b_{j,k}\psi_j\overline{\psi}_k + c_{j,k}\overline{\psi}_j\overline{\psi}_k) + \dots. \tag{7.2}$$

For any sufficiently smooth function f we define by $f(\chi)$ the element of $\mathfrak{A}$ obtained by substituting $\chi - a$ in the Taylor series of f at the point a. Since χ is a polynomial of $\{\psi_j\}_{j=1}^n$, $\{\overline{\psi}_j\}_{j=1}^n$ of the form (7.2), according to (7.1) there exists such l that $(\chi - a)^l = 0$, and hence the series terminates after a finite number of terms and so $f(\chi) \in \mathfrak{A}$.

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

$$\int d\,\psi_j = \int d\,\overline{\psi}_j = 0, \quad \int \psi_j d\,\psi_j = \int \overline{\psi}_j d\,\overline{\psi}_j = 1,$$

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

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

then

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

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

$$\int \exp\Big\{-\sum_{j,k=1}^{n} A_{jk} z_j \overline{z}_k\Big\} \prod_{j=1}^{n} \frac{d\,\Re z_j d\,\Im z_j}{\pi} = \frac{1}{\det A}. \tag{7.3}$$

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

$$\int \exp\Big\{-\sum_{j,k=1}^{n} A_{jk} \overline{\psi}_j \psi_k\Big\} \prod_{j=1}^{n} d\,\overline{\psi}_j d\,\psi_j = \det A, \tag{7.4}$$

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

We will also need the following bosonization formula

Proposition 7.1. *(***see [13]** *)*
Let F be some function that depends only on combinations

$$\bar{\phi}\phi := \Big\{\sum_{\alpha=1}^{W} \bar{\phi}_{l\alpha}\phi_{s\alpha}\Big\}_{l,s=1}^{2},$$

and set

$$d\Phi = \prod_{l=1}^{2}\prod_{\alpha=1}^{W} d\Re\phi_{l\alpha} d\Im\phi_{l\alpha}.$$

Assume also that $W \geq 2$. Then

$$\int F\left(\bar{\phi}\phi\right) d\Phi = \frac{\pi^{2W-1}}{(W-1)!(W-2)!}\int F(B)\cdot \det{}^{W-2} B\, dB,$$

where B is a 2×2 positive Hermitian matrix, and

$$dB = \mathbf{1}_{B>0} dB_{11} dB_{22} d\Re B_{12} d\Im B_{12}.$$

References

[1] Bao, J., Erdős, L.: Delocalization for a class of random block band matrices. Probab. Theory Rel. (2016).

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

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

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

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

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

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

[8] Disertori, M., Spencer, T., Zirnbauer, M. R. Supersymmetric 1d sigma model, unpublished

[9] Efetov, K.: Supersymmetry in disorder and chaos. Cambridge university press, New York (1997)

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

[11] Erdös, L., Knowles, A., Yau, H.-T., Yin, J.: Delocalization and diffusion profile for random band matrices. arXiv:1205.5669v1

[12] Erdös, L., Yau, H.-T., Yin, J.: Bulk universality for generalized Wigner matrices, Preprint arXiv:1001.3453.

[13] Fyodorov Y. V.: Negative moments of characteristic polynomials of random matrices: Ingham – Siegel integral as an alternative to Hubbard – Stratonovich transformation, Nuclear Phys. B **621**, 643674 (2002)

[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] 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)

[17] Peled, R., Schenker, J. , Shamis, M., Sodin, A.: On the Wegner orbital model, arXiv:1608.02922

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

[19] Shcherbina, M., Shcherbina, T.: Transfer matrix approach to 1d random band matrices: density of states, J.Stat.Phys., vol. 164, p. 1233 – 1260 (2016)

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

[21] Shcherbina, T. : On the second mixed moment of the characteristic polynomials of the 1D band matrices. Commun. Math. Phys., vol. 328, p. 45 – 82 (2014), arXiv:1209.3385

[22] Shcherbina, T.: Universality of the local regime for the block band matrices with a finite number of blocks. J.Stat.Phys., vol. 155(3), p. 466 – 499 (2014), arXiv:1309.2120

[23] Sodin, S.: An estimate for the average spectral measure of random band matrices. J. Stat. Phys., vol. 144, p. 46 – 59 (2011)

[24] 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)

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

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

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