Ann. Henri Poincaré 22 (2021), 4205–4269
© 2021 The Author(s)
1424-0637/21/124205-65
published online August 5, 2021
https://doi.org/10.1007/s00023-021-01085-6

Annales Henri Poincaré

Scattering in Quantum Dots via Noncommutative Rational Functions

László Erdős, Torben Krüger and Yuriy Nemish

Abstract. In the customary random matrix model for transport in quantum dots with M internal degrees of freedom coupled to a chaotic environment via $N \ll M$ channels, the density ρ of transmission eigenvalues is computed from a specific invariant ensemble for which explicit formula for the joint probability density of all eigenvalues is available. We revisit this problem in the large N regime allowing for (i) arbitrary ratio $\phi := N/M \leq 1$; and (ii) general distributions for the matrix elements of the Hamiltonian of the quantum dot. In the limit $\phi \to 0$, we recover the formula for the density ρ that Beenakker (Rev Mod Phys 69:731–808, 1997) has derived for a special matrix ensemble. We also prove that the inverse square root singularity of the density at zero and full transmission in Beenakker's formula persists for any $\phi < 1$ but in the borderline case $\phi = 1$ an anomalous $\lambda^{-2/3}$ singularity arises at zero. To access this level of generality, we develop the theory of global and local laws on the spectral density of a large class of noncommutative rational expressions in large random matrices with i.i.d. entries.

Mathematics Subject Classification. 15B52, 46L54, 60B20, 81V65.

1. Introduction and Results

Since the pioneering discovery of E. Wigner on the universality of eigenvalue statistics of large random matrices [38], random matrix theory has become one of the most successful phenomenological theories to study disordered quantum systems, see [2] for a broad overview. Among many other applications, it has been used for open quantum systems and quantum transport, in particular to

L. Erdős and Yu. Nemish: Partially funded by ERC Advanced Grant RANMAT No. 338804.
T. Krüger: Partially supported by VILLUM FONDEN research grant no. 29369 and the Hausdorff Center for Mathematics.

predict the distribution of transmission eigenvalues of scattering in quantum dots and wires. The theory has been developed over many excellent works starting with the ground-breaking papers by Mello et al. [30] and by Verbaarschot et al. [37]; for a complete overview with extensive references, see reviews by Beenakker [7,8], Fyodorov and Savin [21] and Schomerus [35].

We will focus on quantum dots, i.e., systems without internal structure, coupled to an environment (electron reservoir) via scattering channels. Quantum wires, with a typically quasi one-dimensional internal structure, will be left for further works. In the simplest setup the quantum dot is described by a self-adjoint Hamiltonian (complex Hermitian matrix[1]) $H \in \mathbb{C}^{M\times M}$ acting on an M-dimensional state space $\mathbb{C}^M$. It is coupled to an environment with N_0 effective degrees of freedom via an $M \times N_0$ complex coupling matrix W. Following Wigner's paradigm, both the Hamiltonian H and the coupling matrix W are drawn from random matrix ensembles respecting the basic symmetries of the model. Typically, the entries of W are independent, identically distributed (*i.i.d.*), while H is a Wigner matrix, i.e., it has i.i.d. entries on and above the diagonal. We allow for general distributions in contrast to most existing works in the literature that assume H has Gaussian or Lorentzian distribution.

The Hamiltonian of the total system at Fermi energy $E \in \mathbb{R}$ is given by (see [7, Eq. (80)])

$$\mathcal{H} = \sum_{a=1}^{N_0} |a\rangle E \langle a| + \sum_{\mu,\nu=1}^{M} |\mu\rangle H_{\mu\nu} \langle \nu| + \sum_{\mu=1}^{M}\sum_{a=1}^{N_0} \Big[|\mu\rangle W_{\mu a} \langle a| + |a\rangle W^*_{\mu a} \langle \mu| \Big].$$

One common assumption is that the interaction W is independent of the Fermi energy E.

At any fixed energy $E \in \mathbb{R}$, we define the *scattering matrix* (see [7, Eq. (81)])

$$S(E) := I - 2\pi\mathrm{i}\, W^*(E\cdot I - H + \mathrm{i}\,\pi W W^*)^{-1} W \in \mathbb{C}^{N_0\times N_0}. \tag{1.1}$$

This is the finite-dimensional analogue of the Mahaux–Weidenmüller formula in nuclear physics [28] that can be derived from $\mathcal{H}$ in the $N_0 \to \infty$ limit. The definition (1.1) will be the starting point of our mathematical analysis. Since $H = H^*$, one can easily check that $S(E)$ is unitary.

To distinguish between transmission and reflection, we assume that the scattering channels are split into two groups, left and right channels, with dimensions $N_1 + N_2 = N_0$ and the interaction Hamiltonian is also split accordingly; $W = (W_1, W_2) \in \mathbb{C}^{M\times(N_1+N_2)}$. Therefore, $S(E)$ has a natural 2×2 block structure and we can write it as (see [7, Eq. (23)])

$$S(E) = \begin{pmatrix} R & T' \\ T & R' \end{pmatrix}, \tag{1.2}$$

where $R \in \mathbb{C}^{N_1\times N_1}$, $R' \in \mathbb{C}^{N_2\times N_2}$ are the *reflection matrices* and $T \in \mathbb{C}^{N_2\times N_1}$, $T' \in \mathbb{C}^{N_1\times N_2}$ are the *transmission matrices.*

[1] Our method works for the real symmetric case as well, but for simplicity, we stay in the complex Hermitian symmetry class.

As a consequence of unitarity, one finds that TT^*, $T'(T')^*$, $I - RR^*$ and $I - R'(R')^*$ have the same set of nonzero eigenvalues. For simplicity, we assume that $N_1 = N_2 = N$, i.e., generically these four matrices have no zero eigenvalues, the general $N_1 \neq N_2$ case has been considered in [11]. We denote these *transmission eigenvalues* by $\lambda_1, \lambda_2, \ldots, \lambda_N$. They express the rate of the transmission through each channel. By unitarity of S, clearly $\lambda_i \in [0, 1]$ for all i; $\lambda_i = 0$ means the channel is closed, while $\lambda_i = 1$ corresponds to a fully open channel. The transmission eigenvalues carry important physical properties of the system. For example, $\mathrm{Tr} TT^* = \sum_i \lambda_i$ gives the *zero temperature conductance* (*Landauer formula* [7, Eq. (33)]), while

$$\sum_i \lambda_i(1 - \lambda_i) = \mathrm{Tr} TT^* - \mathrm{Tr}(TT^*)^2 \tag{1.3}$$

is the *shot noise power*, giving the zero temperature fluctuation of the current (*Büttiker's formula*, [12], [7, Eq. (35)]). The dimensionless ratio of the shot noise power and the conductance is called the *Fano factor* (see [8, Eq. (2.15)])

$$F := \frac{\sum_i \lambda_i(1 - \lambda_i)}{\sum_i \lambda_i}. \tag{1.4}$$

The current fluctuation is therefore given by a certain linear statistics of the transmission eigenvalues, and thus, it can be computed from the density ρ of these eigenvalues. Therefore, determining ρ is a main task in the theory of quantum dots.

In many physical situations, it is found that ρ has a *bimodal* structure with a peak at zero and a peak at unit transmission rates. Furthermore, ρ exhibits a power law singularity at the edges of its support $[0, 1]$. One main result of the theory in [7] is that in the $M \gg N \gg 1$ regime at energy $E = 0$ the density of transmission eigenvalues for a quantum dot is given by

$$\rho_{\mathrm{Bee}}(\lambda) = \frac{1}{\pi\sqrt{\lambda(1 - \lambda)}}, \tag{1.5}$$

(the answer is different for quantum wires), i.e., it has an inverse square root singularity at both edges, see [8, Eq. (3.12)]. In this case, the Fano factor is $F = 1/4$ which fits well the experimental data.

The goal of this paper is to revisit and substantially generalize the problem of transmission eigenvalues with very different methods than Beenakker and collaborators used. While those works used invariant matrix ensembles for H and relied on explicit computations for the circular ensemble, we consider very general distribution for the matrix elements of H. In particular, we show that in the regime $\phi := N/M \in (0, 1]$, $M \to \infty$, the empirical density of transmission eigenvalues has a deterministic limit $\rho = \rho_\phi$ and we give a simple algebraic equation to compute it. The solution can be continuously extended as $\phi \to 0$, the equation simplifies for $\phi = 0$, the density becomes explicitly computable and it coincides with (1.5); $\lim_{\phi \to 0} \rho_\phi = \rho_{\phi=0} = \rho_{\mathrm{Bee}}$ for $E = 0$ and our formula holds for any E in the bulk spectrum of H. While no short explicit formula is available for the general case $\phi \in (0, 1]$, we can analyze the singularities of ρ_ϕ for any fixed $\phi \in (0, 1]$ in detail.

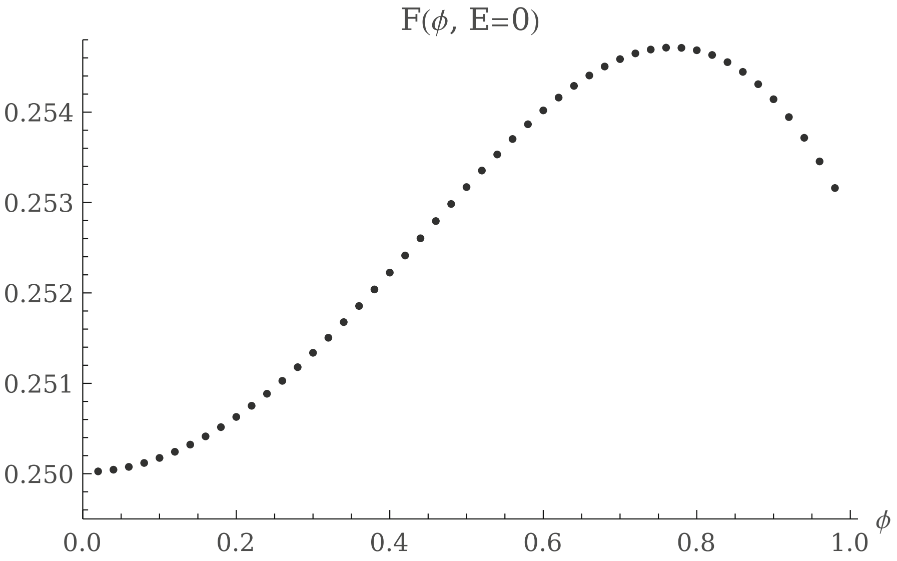

Figure 1. Fano factor for $E = 0$ and $\phi \in (0, 1)$

More precisely, we rigorously prove that for any fixed $\phi \in (0, 1)$ the density ρ_ϕ has an inverse square root singularity both at 0 and at 1,

$$\begin{aligned} &\rho_\phi(\lambda) \sim \lambda^{-1/2}, \qquad &&\text{for } 0 < \lambda \ll 1, \qquad \text{and} \\ &\rho_\phi(\lambda) \sim (1-\lambda)^{-1/2}, \qquad &&\text{for } 0 < 1-\lambda \ll 1, \end{aligned} \tag{1.6}$$

qualitatively in line with ρ_{Bee} from (1.5). However, ρ_ϕ is not symmetric around $\lambda = \frac{1}{2}$ and the Fano factor at $E = 0$ slightly differs from $\frac{1}{4}$ when $\phi \neq 0$. Figure 1 shows the Fano factor for different values of ϕ numerically computed from our theory. We mention that the deviation from 1/4 is within 2% for the entire range of $\phi \in [0, 1)$ which is well below the error bar of the experimental results presented in Fig. 6 of [7] and adapted from [31].

The value $\phi = 1$ is special, since the singularity of ρ at $\lambda \approx 0$ changes to

$$\rho_{\phi=1}(\lambda) \sim \lambda^{-2/3}, \qquad \text{for } 0 < \lambda \ll 1 \tag{1.7}$$

from $\lambda^{-1/2}$ in (1.6), while the inverse square root singularity at 1 persists. This enhancement of singularity signals the emergence of a δ-function component at 0 in ρ_ϕ as ϕ becomes larger than 1, which is a direct consequence of TT^* not having full rank when $N > M$. We remark that this regime is quite unphysical since it corresponds to more scattering channels than the total number of internal states of the quantum dot. In realistic quantum dots, N is smaller than M since not every mode of the dot may participate in scattering. We therefore do not pursue the detailed analysis of ρ_ϕ for $\phi > 1$, although our method can easily be extended to these ϕ's as well.

There are two main differences between our model and that of Beenakker et al. First, the distribution imposed on the random matrix H is different and we consider random W. Second, our current method works in the entire regime $\phi = N/M \in (0,1]$, while Beenakker assumes $M \gg N$, i.e., $\phi \ll 1$. We now explain both differences.

Following Wigner's original vision, any relevant distribution must respect the basic symmetry of the model; in our case, this demands that H be complex Hermitian, while no symmetry constraint is imposed on W. Respecting basic symmetries, one may define ensembles essentially in two ways. Invariant ensembles are defined by imposing that the entire distribution be invariant; it is typically achieved via a global Gibbs factor times the natural flat measure on the space of matrices satisfying the basic symmetry. Wigner ensembles and their generalizations impose distributions directly on the matrix elements and often demand independence (up to the basic symmetry constraint). These two procedures typically yield different ensembles.

While in the simplest case of random Hermitian matrices both types of ensembles have been actively investigated, for ensembles with more complicated structure, like our S that is a rational function of the basic ingredients H and W (1.1), up to recently only the invariant approach was available. Sophisticated explicit formulas have been developed to find the joint distribution of eigenvalues for more and more complicated structured ensembles (see [18]), which could then be combined with orthogonal polynomial methods to obtain local correlation functions. The heavy reliance on explicit formulas, however, imposes a serious limitation on how complicated functions of random matrices, as well as how general distributions on these matrices can be considered. For example, the Gibbs factor is often restricted to Gaussian or closely related ensembles to remain in the realm of explicitly computable orthogonal polynomials.

There have been considerable developments in the other type of ensembles in the recent years. Departing from the invariant world, about 10 years ago the Wigner–Dyson–Mehta universality of local eigenvalue statistics has been proven for Wigner matrices with i.i.d. entries, see [16] and references therein. Later, the i.i.d. condition was relaxed and even matrices with quite general correlation structures among their entries can be handled [14]. One of the key ingredients was to better understand the *Matrix Dyson equation (MDE)*, the basic equation governing the density of states [1]. Together with the linearization trick, this allows us to handle arbitrary polynomials in i.i.d. random matrices [13] and in the current work we extend our method to a large class of rational functions. Note that even if the building block matrices have independent entries, the linearization of their rational expressions will have dependence, but the general MDE can handle it [see (2.13)]. In our work, we deal only with bounded rational functions and relatively straightforward regularizations of unbounded rational functions, the general theory of unbounded rational functions is still in development, see [29] and references therein. We stress that the distribution of the matrix elements of H and W can be practically arbitrary. In particular, our result is not restricted to the Gaussian world.

In comparison, the result of Beenakker et al. [8, Sect. III.B], see also [10, Sect. IV], postulates that H is GUE, while W may even be deterministic and only its singular values are relevant. For example, if all nonzero singular values are equal one and $N_0 = 2N \le M$, i.e., $\phi \le \frac{1}{2}$, then S can be written as

$$S = \frac{I + i\pi \widetilde{H}}{I - i\pi \widetilde{H}}, \qquad \widetilde{H} := Q^T (H - E \cdot I)^{-1} Q,$$

where $Q \in \mathbb{C}^{M \times 2N}$ with $Q_{ij} = \delta_{ij}$. In fact, for the sake of explicit calculations, it is necessary to replace the GUE by a Lorentzian distribution (irrelevant constants ignored),

$$P(H) \sim \det[I + H^2]^{-M}, \tag{1.8}$$

since in this way $\widetilde{H}$ is also Lorentzian, and argue separately that in the sense of correlation functions (1.8) is close to a GUE when M is very large [10, Section III]. Under these conditions S becomes Haar distributed on $U(2N)$ as $M \to \infty$ and N is fixed; this is the step where $M \gg N$ is necessary. Furthermore, one can verify [7, Section II.A.1] that at energy $E = 0$ the transmission eigenvalues of a Haar distributed scattering matrix follow the circular ensemble on $[0, 1]$. Therefore, λ_i's have a well-known joint distribution

$$P\big(\{\lambda_i\}\big) \sim \prod_{i<j} (\lambda_j - \lambda_i)^2 \quad \text{on} \quad [0, 1]^N,$$

and their density can be easily computed, yielding (1.5) in the $N \to \infty$ limit.

While Beenakker's result relies on an impressive identity, it allows little flexibility in the inputs: H needs to be Lorentzian with very large dimension; moreover, $M \gg N$ and $E = 0$ are required. In contrast, our setup allows for a large freedom in choosing the distribution of H, it covers the entire range $M \ge N$ and also applies to any E in the bulk; however, for computational simplicity we model the contacts W also by a random matrix. While the most relevant regime for scattering on quantum dots is still $M \gg N$, as scattering involves surface states only, a very recent work [20] introduces absorbing channels well inside the quantum dot that leads to physical models with $M \sim N$.

The flexibility in our result stems from the fact that our method directly aims at the density of states via an extension of the MDE theory to linearizations of rational functions of random matrices. It seems unnecessarily ambitious, hence requiring too restrictive conditions, to attempt to find the joint distribution of all eigenvalues. Even for Wigner matrices, this is a hopeless task beyond the Gaussian regime. We remark that the present analysis of the density of transmission eigenvalues for the quantum dot is only one convenient application of our approach. This method is powerful enough to answer many related questions concerning the density of states such as the analysis of the scattering poles [19] as well as extensions from quantum dots to quantum wires that we will address in future work.

1.1. Model and Main Theorem

In order to accommodate all parameters that will appear in our analysis, we introduce a generalized version of the scattering matrix (1.1)

$$S(w) := I - 2\,\mathrm{i}\,\gamma\, W^*(w \cdot I - H + \mathrm{i}\,\gamma\, WW^*)^{-1} W \tag{1.9}$$

with $\gamma > 0$ and $w \in \mathbb{C}_+$ for $\phi \in (0, 1/2]$ and $w \in \mathbb{C}_+ \cup \mathbb{R}$ for $\phi \in (1/2, 1]$, where $\mathbb{C}_+ := \{z \in \mathbb{C} : \mathrm{Im} z > 0\}$ denotes the complex upper half plane. The constant $\gamma > 0$ quantifies the coupling between the internal quantum dot Hamiltonian H and the external leads W_1, W_2, i.e., the effective Hamiltonian of the open quantum systems becomes $H - \mathrm{i}\,\gamma WW^*$. The spectral parameter w is used to regularize the potentially unstable inverse in (1.9). Note that for $\phi \in (0, 1/2)$ the matrix WW^* has a nontrivial kernel (and even for $\phi = 1/2$ it has a very small eigenvalue), hence initially a regularization with a small positive imaginary part is necessary that will later be carefully removed. For the technically easier $\phi \in (1/2, 1]$ regime, we may directly choose w to be real since the eigenvalues of WW^* are bounded away from zero with very high probability.

The general formula (1.1) recovers the scattering matrix (1.1) by setting $\gamma = \pi$, $\mathrm{Re} w = E$ and $\mathrm{Im} w = 0$. The regularized scattering matrix still has the 2×2-block structure from (1.2) with $T \in \mathbb{C}^{N\times N}$. We consider the following random matrix model [see (1.1) and (1.2)]

$$\boldsymbol{T}_{w,\phi,\gamma} := TT^* = 4\gamma^2 W_2^* \frac{1}{w - H + \mathrm{i}\gamma WW^*} W_1 W_1^* \frac{1}{\overline{w} - H - \mathrm{i}\gamma WW^*} W_2, \tag{1.10}$$

where the triple of parameters (w, ϕ, γ) belongs to the same set as for (1.9). Furthermore, for $M, N \in \mathbb{N}$, $\phi := N/M$ the matrices $H \in \mathbb{C}^{M\times M}$ and $W_1, W_2 \in \mathbb{C}^{M\times N}$ are three independent random matrix ensembles satisfying the following assumptions

(**H1**qd) H is a Hermitian random matrix having independent (up to symmetry constraints) centered entries of variance $1/M$;

(**H2**qd) W_1 and W_2 are (non-Hermitian) random matrices having independent centered entries of variance $1/M$;

(**H3**qd) entries of H, W_1 and W_2, denoted by $H(i, j)$, $W_1(i, j)$ and $W_2(i, j)$ correspondingly, have finite moments of all orders, i.e., there exist $\varphi_n > 0$, $n \in \mathbb{N}$, such that

$$\begin{aligned} &\max_{1\,\le\, i,\, j\,\le\, M} \mathbb{E}\big[\,|\sqrt{M} H(i,j)|^n\big] \\ &\quad + \max_{\substack{1\,\le\, i\,\le\, M\\ 1\,\le\, j\,\le\, N}} \Big(\mathbb{E}\big[\,|\sqrt{M} W_1(i,j)|^n\big] + \mathbb{E}\big[\,|\sqrt{M} W_2(i,j)|^n\big]\Big) \le \varphi_n. \end{aligned} \tag{1.11}$$

Remark 1.1. (Constant matrices) In (1.10) and later in the paper, for $B \in \mathbb{C}$, $n \in \mathbb{N}$ and $I_n \in \mathbb{C}^{n\times n}$ the identity matrix of size n, we use the shorthand notation $B \cdot I_n = B$. This notation is used only when the dimension of I_n can be unambiguously determined from the context.

Remark 1.2. In the sequel, we consider the coupling constant γ to be a fixed positive number; therefore, we will omit the dependence on γ in our notation.

Denote by $\mu_{\boldsymbol{T}_{w,\phi}}(d\lambda) := \frac{1}{N}\sum_{i=1}^{N}\delta_{\lambda_i}$ the empirical spectral measure of $\boldsymbol{T}_{w,\phi}$, where λ_i are the eigenvalues of the Hermitian matrix $\boldsymbol{T}_{w,\phi}$. To simplify the presentation, we will assume in this paper that the dimensions of the matrices H, W_1 and W_2 grow over a subsequence $(N,M) = (kn, ln)$, $n \in \mathbb{N}$, i.e., $N/M = \phi$ is kept fixed. One could easily extend our argument to include the general situation when one considers two sequences $N = N_n$ and $M = M_n$ tending to infinity such that $\phi_n = N_n/M_n \to \phi$.

We now formulate our two main results. The first one, Theorem 1.3, is the *global law* for the empirical eigenvalue density; it shows that $\mu_{\boldsymbol{T}_{w,\phi}}(d\lambda)$ has a deterministic limit denoted by $\rho_{w,\phi}(d\lambda)$. The second result, Theorem 1.4, contains key properties of $\rho_{w,\phi}(d\lambda)$. It shows that the regularization in the spectral parameter w can be removed and that $\rho_{w,\phi}$ can be continuously extended to $\phi \to 0$; moreover, it explicitly identifies its singularities at zero and one.

Theorem 1.3 (Global law). *Fix a positive real number $\gamma > 0$ and a rational number $\phi \in (0,1] \cap \mathbb{Q}$. Let w be a fixed spectral parameter: for $\phi \in (0,1/2]$ we assume that $w \in \mathbb{C}_+$, while for $\phi \in (1/2,1]$ we assume $w \in \mathbb{C}_+ \cup \mathbb{R}$. Then, there exists a deterministic probability measure $\rho_{w,\phi}(d\lambda)$ with* $\operatorname{supp}\rho_{w,\phi} \subset [0,1]$ *such that $\mu_{\boldsymbol{T}_{w,\phi}}(d\lambda)$ converges weakly to $\rho_{w,\phi}(d\lambda)$ in probability (and almost surely) as $M, N \to \infty$ with $N/M = \phi$.*

Theorem 1.4 (Properties of the transmission eigenvalue density). *Let the numbers γ, w, ϕ and the measure $\rho_{w,\phi}(d\lambda)$ be as in Theorem 1.3.*

(i) *The function*

$$\Big(\big((0,1/2] \cap \mathbb{Q}\big) \times \mathbb{C}_+\Big) \cup \Big(\big((1/2,1] \cap \mathbb{Q}\big) \times \big(\mathbb{C}_+ \cup \mathbb{R}\big)\Big) \ni (\phi, w) \mapsto \rho_{w,\phi}(d\lambda)$$

from Theorem 1.3 with values in probability measures on $[0,1]$ can be extended to a function on $(0,1] \times (\mathbb{C}_+ \cup \mathbb{R})$ that is continuous in the weak topology. In particular, the weak limit

$$\rho_{E,\phi}(d\lambda) = \lim_{\mathbb{C}_+ \ni w \to E} \rho_{w,\phi}(d\lambda) \tag{1.12}$$

exists for all $E \in \mathbb{R}$. For every $E \in \mathbb{R}$ and $\phi \in (0,1]$ the measure $\rho_{E,\phi}(d\lambda)$ is absolutely continuous, i.e., $\rho_{E,\phi}(d\lambda) = \rho_{E,\phi}(\lambda)d\lambda$. The function $\rho_{E,\phi}(\lambda)$ is bounded when λ is away from 0 and 1. Furthermore, the weak limit $\rho_{E,0}(d\lambda) := \lim_{\phi\downarrow 0}\rho_{E,\phi}(d\lambda)$ exists for every $E \in \mathbb{R}$.

(ii) *If $\phi = 1$ and $\gamma > 0$, then $\rho_{E,\phi}(\lambda)$ has the following asymptotics near 0 and 1:*

(a) *for $E \in \mathbb{R}$*

$$\rho_{E,1}(\lambda) = \frac{1}{\pi}\sqrt[3]{\frac{1+E^2}{4\gamma^2}}\,\lambda^{-2/3} + O\left(\lambda^{-1/3}\right) \quad \text{as } \lambda \to 0_+, \tag{1.13}$$

(b) *for* $|E| < \frac{1}{\gamma}(2\sqrt{1+6\gamma^2+\gamma^4}-2\gamma^2-2)^{1/2}$

$$\rho_{E,1}(\lambda) = \frac{1}{\pi}\,\frac{-4\xi_0}{\xi_0^2+\gamma^2E^2+4}\,(1-\lambda)^{-1/2} + O\,(1) \ \ \textit{as } \lambda \to 1_-, \tag{1.14}$$

where $\xi_0 = -\sqrt{2\sqrt{1+6\gamma^2+\gamma^4}-2\gamma^2-2-\gamma^2E^2}$.

(iii) *If* $\phi \in (0,1)$ *and* $\gamma > 0$, *then*

(a) *for* $E \in \mathbb{R}$

$$\rho_{E,\phi}(\lambda) = \frac{1}{\pi}\,\frac{\gamma^2E^2\xi_0^4 + \big(\xi_0^2+2\phi\big)^2\big(4+\gamma^2\xi_0^2\big)}{4\gamma\xi_0\big(\xi_0^2+2\phi\big)^2}\,\lambda^{-1/2} + O\,(1) \ \ \textit{as } \lambda \to 0_+, \tag{1.15}$$

where $\xi_0 > 0$ *is the unique positive solution of*

$$\xi_0^6 + \big(E^2+8\phi-4\big)\xi_0^4 + 4\phi\big(E^2+5\phi-4\big)\xi_0^2 + 16(\phi-1)\phi^2 = 0. \tag{1.16}$$

(b) *for* $|E| < E_* := \frac{1}{\gamma}\Big(2\sqrt{1+2(1+2\phi)\gamma^2+(2\phi-1)^2\gamma^4} - 2\gamma^2(2\phi-1)-2\Big)^{1/2}$ *the density satisfies*

$$\rho_{E,\phi}(\lambda) = \frac{1}{\pi}\,\frac{-4\xi_1}{\xi_1^2+\gamma^2E^2+4}\,(1-\lambda)^{-1/2} + O\,(1) \ \ \textit{as } \lambda \to 1_-, \tag{1.17}$$

where

$$\xi_1 = -\gamma\sqrt{E_*^2-E^2}. \tag{1.18}$$

(iv) *If* $\phi = 0$, $\gamma > 0$ *and* $|E| < 2$ *the density is given globally by the explicit formula*

$$\rho_{E,0}(\lambda) = \frac{1}{\pi}\,\frac{\gamma\big(1+\gamma^2\big)}{\big(1+\gamma^2\big)^2\lambda + \gamma^2\big(4-E^2\big)(1-\lambda)}\,\sqrt{\frac{4-E^2}{\lambda(1-\lambda)}}, \qquad \lambda \in (0,1). \tag{1.19}$$

Proofs of Theorems 1.3 and 1.4. We now explain the structure of the paper that contains the proof of both theorems. The density function $\rho_{E,\phi}(\lambda)$ will be derived from a deterministic equation, the Dyson equation of the linearizaton of (1.10) [see Eqs. (2.13) and (2.27) for $\phi = 1$, as well as (3.7) and (3.18) for $\phi \in (0,1)$]. This equation plays a central role in our analysis. Sections 2 and 3 contain the proofs of the model specific parts of Theorem 1.4 which use some key conclusions of the general theory on noncommutative rational functions developed in Appendix A. The $\phi = 1$ case is treated in Sect. 2 with full details, while in Sect. 3 we explain the modifications for the general rational $\phi \in (0,1)$ case. Theorem 1.3 is proven in Lemma 2.5 for $\phi = 1$ and Lemma 3.4 for general rational $\phi \in (0,1)$ using the general global law from Appendix A. Part (ii) is proven in Sect. 2.5 after having established key properties of the solution for the Dyson equation. Section 3 follows the same structure as Sect. 2 and proves parts (i), (iii) and (iv) of Theorem 1.4 for the $\phi \in [0,1)$ case. □

Remark 1.5. The restriction $|E| < E_*$ in Theorem 1.4 is used to identify the regime in which the density $\rho_{E,\phi}$ has two singularities. For $|E| > E_*$, the singularity at $\lambda = 1$ disappears and the support of the density becomes bounded away from $\lambda = 1$. Physically, this indicates that there are no fully open channels in this regime. This effect is most severe in the case $\phi = 0$, where the system becomes completely reflective as $|E|$ increases above the threshold $E_* = 2$. Our approach is also applicable for $|E| > E_*$, but for the simplicity of the presentation, we restrict our analysis to the physically more relevant situation when two singularities exist simultaneously.

We formulated Theorem 1.3 about the specific matrix (1.10). However, our method works for very general noncommutative (NC) rational expressions in large matrices with i.i.d. entries (with or without Hermitian symmetry) generalizing our previous work [13] on polynomials. For convenience of the readers interested only in the concrete scattering problem, the main part of our paper focuses on this model and we defer the general theory to Appendix A. The details of this appendix are not necessary in order to follow the main arguments in Sects. 2 and 3 provided some basic facts from the appendix are accepted. These facts will be re-stated for the specialization to our concrete operator $\boldsymbol{T}_{w,\phi}$ from (1.10). On the other hand, Appendix A is written in a self-contained form, so readers interested in the general theory can directly go to it, skipping the concrete application.

In Appendix A, we first give the precise definition of the set of admissible rational expressions that requires some technical preparation, see Sect. A.1 for details. Roughly speaking, we can consider any rational expression whose denominators are stably invertible with overwhelming probability. We remark that this property holds for (1.10) since the imaginary part of $w - H + \mathrm{i}\,\gamma W W^*$ has a positive lower bound as long as $\phi > 1/2$ or $w \in \mathbb{C}_+$. With this definition at hand, we develop the theory of global and local laws as well as the identification of the pseudospectrum for such rational expressions in Sects. A.6 and A.5, respectively.

2. Proof of Theorem 1.4 for the Special Case $\phi = 1$

In this section, we study the model (1.10) for fixed $\gamma > 0$, $\phi = 1$ and $w \in \mathbb{C}_+ \cup \mathbb{R}$. This special choice of parameter ϕ ensures that the linearization of $\boldsymbol{T}_{w,\phi}$ has a fairly simple structure, which makes the proof of Theorems 1.3 and 1.4 more transparent and streamlined. Generalization to $\phi \in (0, 1)$ is postponed to Sect. 3. Since the parameter ϕ is fixed to be equal to 1, we will omit the dependence on ϕ in the notation throughout the current section. The information about linearizations of general rational functions is collected in the Appendix A.1. Here, we often refer to specialization of these results to $\boldsymbol{T}_w$.

2.1. Linearization Trick and the Dyson Equation for Linearization

We consider the random matrix model $\boldsymbol{T}_w$ defined in (1.10) as a self-adjoint rational function of random matrices H, W_1 and W_2. In order to study its eigenvalues, we introduce the self-adjoint linearization matrix $\boldsymbol{H}_w \in \mathbb{C}^{8N\times 8N}$

$$\boldsymbol{H}_w := \left(\begin{array}{ccc|ccc|cc} 0 & 0 & 0 & 0 & 0 & 0 & 0 & W_2^* \\ 0 & 0 & \frac{\mathrm{i}}{\gamma} & 0 & 0 & 0 & 0 & W_2^* \\ 0 & -\frac{\mathrm{i}}{\gamma} & 0 & 0 & 0 & 0 & W_2^* & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & \frac{\mathrm{i}}{\gamma} & 0 & W_1^* \\ 0 & 0 & 0 & 0 & -\frac{1}{4\gamma^2} & 0 & W_1^* & 0 \\ 0 & 0 & 0 & -\frac{\mathrm{i}}{\gamma} & 0 & 0 & W_1^* & 0 \\ \hline 0 & 0 & W_2 & 0 & W_1 & W_1 & 0 & w-H \\ W_2 & W_2 & 0 & W_1 & 0 & 0 & \overline{w}-H & 0 \end{array}\right) \tag{2.1}$$

with $w \in \mathbb{C}_+ \cup \mathbb{R}$. Denote by $J_m \in \mathbb{C}^{m\times m}$, $m \in \mathbb{N}$, a matrix whose $(1,1)$ entry is equal to 1 and all other entries are equal to 0. For any $n \in \mathbb{N}$ and $R \in \mathbb{C}^{n\times n}$, we define $\|R\|$ to be the operator norm of R. The following definition is a specialization of the notion of generalized resolvent from Appendix A.3.

Definition 2.1 (*Generalized resolvent*). We call the matrix-valued function $z \mapsto (\boldsymbol{H}_w - zJ_8 \otimes I_N)^{-1}$ defined for $z \in \mathbb{C}_+$ the *generalized resolvent* of $\boldsymbol{H}_w$.

The results below are formulated using the notion of *asymptotically overwhelming probability.*

Definition 2.2. We say that a sequence of events $\{\Omega_N\}_{N\in\mathbb{N}}$ holds asymptotically with overwhelming probability (*a.w.o.p.* for short) if for any $D > 0$ there exists $C_D > 0$ such that

$$\mathbb{P}[\,\Omega_N] \geq 1 - \frac{C_D}{N^D}. \tag{2.2}$$

Consider the set

$$\Theta_N := \Big\{ \|H\| \leq 3,\ \|W_1\| \leq 3,\ \|W_2\| \leq 3,\ \|(WW^*)^{-1}\| \leq 12 \Big\}\,. \tag{2.3}$$

Note that WW^* is a sample covariance matrix with concentration ratio $1/2$, hence its spectrum is asymptotically supported on an arbitrarily small neighborhood of $[(1-\frac{1}{\sqrt{2}})^2, (1+\frac{1}{\sqrt{2}})^2]$ with very high probability. The limiting support follows from the classical Bai–Yin theorem [6], the corresponding large deviation result under somewhat different conditions follows, e.g., from Corollary V.2.1 of [17]. In fact, the boundedness of $\|(WW^*)^{-1}\|$ and $\|W_i\|$ also follows from [33], at least for subgaussian distributions. Alternatively, under our conditions (1.11) this result also follows from Lemma B.1 in the Appendix by choosing $n = M$, $l = 1$ and $m = 2$. Similarly, from the properties of classical Wigner and *iid* ensembles (see, e.g., [5, Section 5]), we obtain that the event $\{\,\|H\| \leq 3,\ \|W_1\| \leq 3,\ \|W_2\| \leq 3\,\}$ also holds *a.w.o.p.*, and we conclude that for any $D > 0$ there exists $C_D >$ such that

$$\mathbb{P}[\Theta_N] \geq 1 - \frac{C_D}{N^D}. \tag{2.4}$$

In the rest of this section, we consider the random matrix models (1.9) and (1.10) *restricted* to the set Θ_N. Since on this set the smallest eigenvalue of WW^* cannot be smaller that $1/12$, we have that $\|(w-H+\mathrm{i}\,\gamma WW^*)^{-1}\| \leq \frac{1}{12\gamma}$ for all $w \in \mathbb{C}_+ \cup \mathbb{R}$, and thus, the model (1.10) is well defined on Θ_N. In the next lemma, we establish an a priori bound for the generalized resolvent $(\boldsymbol{H}_w - zJ_8 \otimes I_N)^{-1}$.

Lemma 2.3 (Basic properties of the generalized resolvent).

(i) *For any $\gamma > 0$ there exists $C_\gamma > 0$ such that* a.w.o.p.

$$\left\|(\boldsymbol{H}_w - zJ_8 \otimes I_N)^{-1}\right\| \le C_\gamma\left(1 + \frac{1}{\operatorname{Im} z}\right) \tag{2.5}$$

uniformly for all $z \in \mathbb{C}_+$ and $w \in \mathbb{C}_+ \cup \mathbb{R}$.

(ii) *For all $w \in \mathbb{C}_+ \cup \mathbb{R}$, $z \in \mathbb{C}_+$ and $1 \le i, j \le N$*

$$\left[\left(\boldsymbol{H}_w - zJ_8 \otimes I_N\right)^{-1}\right]_{ij} = \left[\left(\boldsymbol{T}_w - zI_N\right)^{-1}\right]_{ij}, \quad 1 \le i, j \le N. \tag{2.6}$$

Proof. Let $\boldsymbol{H}_w^{(\mathrm{init})}$ be the linearization matrix obtained via the linearization algorithm described in Appendix A.2

$$\boldsymbol{H}_w^{(\mathrm{init})} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 2\gamma W_2^* & 0 & 0 \\ 0 & 0 & 0 & 0 & W_1 & -(w-H) & -\sqrt{\gamma}W_1 & -\sqrt{\gamma}W_2 \\ 0 & 0 & 0 & 0 & 0 & -\sqrt{\gamma}W_1^* & -\mathrm{i} & 0 \\ 0 & 0 & 0 & 0 & 0 & -\sqrt{\gamma}W_2^* & 0 & -\mathrm{i} \\ 0 & W_1^* & 0 & 0 & -1 & 0 & 0 & 0 \\ 2\gamma W_2 & -(\overline{w}-H) & -\sqrt{\gamma}W_1 & -\sqrt{\gamma}W_2 & 0 & 0 & 0 & 0 \\ 0 & -\sqrt{\gamma}W_1^* & \mathrm{i} & 0 & 0 & 0 & 0 & 0 \\ 0 & -\sqrt{\gamma}W_2^* & 0 & \mathrm{i} & 0 & 0 & 0 & 0 \end{pmatrix}. \tag{2.7}$$

Denote by $\{E_{ij}, 1 \le i, j \le 8\}$ the standard basis of $\mathbb{C}^{8\times 8}$

$$E_{ij} = \left(\delta_{ki}\delta_{jl}\right)_{k,l=1}^{8},$$

where $\delta_{\alpha\beta}$ is the Kronecker delta. Then, one can easily check that $\boldsymbol{H}_w$ in (2.1) can be obtained from $\boldsymbol{H}_w^{(\mathrm{init})}$ by applying the following transformation

$$\boldsymbol{H}_w = \widetilde{T}\widetilde{P}_{78}\widetilde{P}_{23}\widetilde{P}_{34}\widetilde{P}_{67}\widetilde{P}_{28}\boldsymbol{H}_w^{(\mathrm{init})}\widetilde{P}_{28}\widetilde{P}_{67}\widetilde{P}_{34}\widetilde{P}_{23}\widetilde{P}_{78}\widetilde{T} \tag{2.8}$$

with $\widetilde{T} = \operatorname{diag}\left(1, -2\sqrt{\gamma}, \frac{1}{2\gamma^{3/2}}, -2\sqrt{\gamma}, -\frac{1}{2\gamma}, \frac{1}{2\gamma^{3/2}}, -2\gamma, \frac{1}{2\gamma}\right) \otimes I_N$, $\widetilde{P}_{ij} = \left(E_{ij} + E_{ji} + \sum_{l \notin \{i,j\}} E_{ll}\right) \otimes I_N$. Note that all transposition matrices $\widetilde{P}_{ij}$ in (2.8) leave the upper-left $N \times N$ block intact. Thus, (2.6) follows from the Definition A.5 of linearization and the Schur complement formula [see, e.g., (A.20)] by taking $\mathcal{A} = \mathbb{C}^{N\times N}$.

In order to prove the bound (2.5), consider the set Θ_N defined in (2.3). One can see that $\boldsymbol{H}_w^{(\mathrm{init})}$ satisfies the bound (2.5) by specializing Lemma A.10 for $\mathcal{A} = \mathbb{C}^{N\times N}$, $x_1 = H$, $y_1 = W_1$, $y_2 = W_2$, $C = 12/\gamma$ and the rational expression q being (1.10) on the set Θ_N, as well as using the standard relation between the operator and max norms, similarly as in, e.g., (A.47). On the other hand, $\boldsymbol{H}_w^{(\mathrm{init})}$ and $\boldsymbol{H}_w$ are related by (2.8). For any $R \in \mathbb{C}^{8\times 8}$, applying the transformation $R \mapsto \widetilde{P}_{ij}R\widetilde{P}_{ij}$ does not change the norm of R, while applying the map $R \mapsto \widetilde{T}R\widetilde{T}$ might change the norm by an irrelevant nonzero constant factor only. We thus conclude that $\boldsymbol{H}_w$ also satisfies the bound (2.5) with a constant C_γ being possibly different than the one for $\boldsymbol{H}_w^{(\mathrm{init})}$. Since the set Θ_N satisfies (2.4), the bound (2.5) holds *a.w.o.p.* □

Define

$$\kappa_1 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & \frac{\mathrm{i}}{\gamma} \\ 0 & -\frac{\mathrm{i}}{\gamma} & 0 \end{pmatrix}, \quad \kappa_2 = \begin{pmatrix} 0 & 0 & \frac{\mathrm{i}}{\gamma} \\ 0 & -\frac{1}{4\gamma^2} & 0 \\ -\frac{\mathrm{i}}{\gamma} & 0 & 0 \end{pmatrix} \tag{2.9}$$

and

$$\kappa_3 = \begin{pmatrix} 0 & w \\ \overline{w} & 0 \end{pmatrix}, \quad \kappa_4 = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 0 \end{pmatrix}, \quad \kappa_5 = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}. \tag{2.10}$$

With this notation $\boldsymbol{H}_w$ can be rewritten as

$$\boldsymbol{H}_w = K_0(w)\otimes I_N + K_1\otimes H + L_1\otimes W_1 + L_1^*\otimes W_1^* + L_2\otimes W_2 + L_2^*\otimes W_2^*, \tag{2.11}$$

where $K_0 = K_0(w), K_1, L_1, L_2 \in \mathbb{C}^{8\times 8}$ are given by their block structures as

$$K_0 = \begin{pmatrix} \kappa_1 & 0 & 0 \\ 0 & \kappa_2 & 0 \\ 0 & 0 & \kappa_3 \end{pmatrix}, \quad K_1 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -\sigma_1 \end{pmatrix}, \quad L_1 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & \kappa_4 & 0 \end{pmatrix},$$

$$L_2 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ \kappa_5 & 0 & 0 \end{pmatrix}, \tag{2.12}$$

and $\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ is the usual Pauli matrix. Consider the *Dyson equation for linearization (DEL)*

$$-\frac{1}{M} = zJ_8 - K_0(w) + \Gamma[M] \tag{2.13}$$

with a linear map $\Gamma : \mathbb{C}^{8\times 8} \to \mathbb{C}^{8\times 8}$ given by

$$\Gamma[R] := K_1RK_1 + L_1RL_1^* + L_1^*RL_1 + L_2RL_2^* + L_2^*RL_2, \qquad R \in \mathbb{C}^{8\times 8}. \tag{2.14}$$

Lemma 2.4. (Existence and basic properties of the solution to the DEL (2.13)) *For any $\gamma > 0$, $w \in \mathbb{C}_+ \cup \mathbb{R}$ and $z \in \mathbb{C}_+$ define $M_{z,w} \in \mathbb{C}^{8\times 8}$ as*

$$\begin{aligned} M_{z,w} := (\mathrm{id}_8 \otimes \tau_{\mathcal{S}})\Big(\big(K_0(w) - zJ_8\big) \otimes \mathbb{1}_{\mathcal{S}} + K_1 \otimes s \\ + L_1 \otimes c_1 + L_1^* \otimes c_1^* + L_2 \otimes c_2 + L_2^* \otimes c_2^*\Big)^{-1}, \end{aligned} \tag{2.15}$$

where s, c_1, c_2 are freely independent semicircular and circular elements and $\tau_{\mathcal{S}}$ is a tracial state on a C^*-probability space $(\mathcal{S}, \tau_{\mathcal{S}})$ with the unit element $\mathbb{1}_{\mathcal{S}}$. Then,

(i) *For any $\gamma > 0$ there exists $C_\gamma > 0$ such that $M_{z,w}$ satisfies the a priori bound*

$$\|M_{z,w}\| \le C_\gamma\Big(1 + \frac{1}{\mathrm{Im} z}\Big) \tag{2.16}$$

uniformly for all $w \in \mathbb{C}_+ \cup \mathbb{R}$ and $z \in \mathbb{C}_+$.

(ii) *For any $\gamma > 0$, $w \in \mathbb{C}_+ \cup \mathbb{R}$ and $z \in \mathbb{C}_+$, matrix $M_{z,w}$ satisfies the DEL* (2.13) *and has positive semidefinite imaginary part,* $\mathrm{Im} M_{z,w} \geq 0$. *Moreover, for all $\gamma > 0$ and $w \in \mathbb{C}_+ \cup \mathbb{R}$, the matrix-valued function $z \mapsto M_{z,w}$ is analytic on $\mathbb{C}_+$.*

(iii) *For any $\gamma > 0$, $w \in \mathbb{C}_+ \cup \mathbb{R}$ and $z \in \mathbb{C}_+$ function $z \mapsto M_{z,w}$ admits the representation*

$$M_{z,w} = M_w^\infty + \int_{\mathbb{R}} \frac{V_w(d\lambda)}{\lambda - z}, \tag{2.17}$$

where $M_w^\infty \in \mathbb{C}^{8\times 8}$, and $V_w(d\lambda)$ is a positive semidefinite matrix-valued measure on $\mathbb{R}$ with compact support. In particular, $\lim_{z\to\infty} M_{z,w} = M_w^\infty$.

Proof. Fix $\gamma > 0$ and denote the noncommutative rational (in fact, polynomial) expression

$$q_{1,w}(x, y_1, y_2, y_1^*, y_2^*) := w - x + \mathrm{i}\,\gamma(y_1 y_1^* + y_2 y_2^*). \tag{2.18}$$

Using the fact that $c_1 c_1^* + c_2 c_2^*$ has free Poisson distribution of rate 2, which in particular implies that $\|(c_1 c_1^* + c_2 c_2^*)^{-1}\| \leq (1 - \frac{1}{\sqrt{2}})^{-1}$, we conclude that the triple (s, c_1, c_2) belongs to the effective domain $\mathcal{D}_{\boldsymbol{q}_0, \{q_{1,w}\}; C}$ with $C = \gamma^{-1}(1 - \frac{1}{\sqrt{2}})^{-1}$ *for all* $w \in \mathbb{C}_+ \cup \mathbb{R}$ (see Sect. A.1 for the corresponding definitions).

Following the proofs of Lemmas A.7 and A.10 and specializing them to our concrete case, we obtain that for any fixed $\gamma > 0$ there exists $C_\gamma > 0$ such that

$$\begin{aligned} \Big\| \big((K_0(w) - zJ_8) \otimes \mathbb{1}_{\mathcal{S}} + K_1 \otimes s + L_1 \otimes c_1 + L_1^* \otimes c_1^* \\ + L_2 \otimes c_2 + L_2^* \otimes c_2^* \big)^{-1} \Big\|_{\mathbb{C}^{8\times 8} \otimes \mathcal{S}} \leq C_\gamma \Big(1 + \frac{1}{\mathrm{Im} z} \Big) \end{aligned} \tag{2.19}$$

uniformly for all $w \in \mathbb{C}_+ \cup \mathbb{R}$, which yields the a priori bound (2.16). Part (ii) of Lemma 2.4 now follows directly from parts (ii)–(iv) of Lemma A.12. Finally, (2.17) follows from the representation (A.27) in Lemma A.12. □

With *DEL* (2.13), we associate the corresponding *stability operator* $\mathscr{L}_{z,w} : \mathbb{C}^{8\times 8} \to \mathbb{C}^{8\times 8}$ given by

$$\mathscr{L}_{z,w}\big[R\big] := R - M_{z,w}\,\Gamma\big[R\big]\,M_{z,w}, \quad R \in \mathbb{C}^{8\times 8}. \tag{2.20}$$

The following lemma is directly obtained from Proposition A.19 and establishes Theorem 1.3 and the weak limits in the part (i) of Theorem 1.4 for $\phi = 1$.

Lemma 2.5 (Global law for $\mu_{\boldsymbol{T}_w}$). *For $w \in \mathbb{C}_+ \cup \mathbb{R}$ the empirical spectral measure $\mu_{\boldsymbol{T}_w}(d\lambda)$ converges weakly in probability (and almost surely) to $\rho_w(d\lambda)$, where $\rho_w(d\lambda) := \langle e_1, V_w(d\lambda)\, e_1 \rangle$ is the $(1,1)$ component of the matrix-valued measure $V_w(d\lambda)$ from* (2.17). *Moreover,* $\mathrm{supp}\, \rho_w(d\lambda) \subset [0,1]$ *for all $w \in \mathbb{C}_+ \cup \mathbb{R}$ and $\rho_w(d\lambda)$ converges weakly to $\rho_E(d\lambda)$ as $w \in \mathbb{C}_+$ tends to $E \in \mathbb{R}$.*

Proof. We apply Proposition A.19 to the rational expression in random matrices (1.10). By (A.78), for any $w \in \mathbb{C}_+$ and fixed $\theta > 0$ the generalized resolvent of the linearization $(\boldsymbol{H}_w - zJ_8 \otimes I_N)^{-1}$, corresponding to $\boldsymbol{G}_z$ in (A.78), converges uniformly on $\{\, z : \mathrm{Im} z \geq \theta^{-1},\ |z| \leq \theta \,\}$ to $M_{z,w} \otimes I_N$, corresponding to $M_z^{(\mathrm{sc})} \otimes I_N$ in (A.78). In particular, the identity (2.6), similarly to (A.79) (see also Definition A.17), implies the pointwise convergence of the Stieltjes transform of the empirical spectral measure of $\boldsymbol{T}_w$: for any $\theta, \varepsilon, D > 0$ there exists a constant $C_{\theta,\varepsilon,D} > 0$ such that

$$\mathbb{P}\Big[\Big|\frac{1}{N}\mathrm{Tr}\,\Big(\boldsymbol{T}_w - zI_N\Big)^{-1} - M_{z,w}(1,1)\Big| \geq \frac{N^\varepsilon}{N}\Big] \leq \frac{C_{\theta,\varepsilon,D}}{N^D} \tag{2.21}$$

uniformly on $\{\, z \, : \, \mathrm{Im} z \geq \theta^{-1},\ |z| \leq \theta \,\}$ and $w \in \mathbb{C}_+ \cup \mathbb{R}$. The convergence in (2.21) together with (2.17) imply that the weak convergence

$$\lim_{N\to\infty} \mu_{\boldsymbol{T}_w}(d\lambda) = \rho_w(d\lambda) \tag{2.22}$$

holds in probability (see, e.g., [4, Theorem 2.4.4]).

Now we prove the almost sure convergence. Take any $z_0 \in \mathbb{C}_+$, a sequence $\{z_i\} \subset \mathbb{C}_+$ of different complex numbers with $z_i \to z_0$ such that z_0 is the accumulation point of $\{z_i\}$. Define the sequence of events

$$A_N := \Big\{ \max_{1\leq i\leq N} \Big\{\Big|\frac{1}{N}\mathrm{Tr}\,\Big(\boldsymbol{T}_w - z_i I_N\Big)^{-1} - M_{z_i,w}(1,1)\Big|\Big\} \geq \frac{1}{\sqrt{N}}\Big\}. \tag{2.23}$$

Then it follows from (2.21) and the Borel–Cantelli lemma applied to $\{A_N\}$ that with probability 1

$$\lim_{N\to\infty} \frac{1}{N}\mathrm{Tr}\,\Big(\boldsymbol{T}_w - z_i I_N\Big)^{-1} = M_{z_i,w}(1,1) \tag{2.24}$$

for all $i \in \mathbb{N}$. Finally, applying the Vitali–Porter theorem we conclude that the weak convergence (2.22) holds almost surely.

To prove the bound on the support of $\rho_w(d\lambda)$ note that the scattering matrix $S(w)$, related to $\boldsymbol{T}_w = TT^*$ via (1.2), is unitary. This implies that all singular values of T are located in the interval $[0,1]$, thus $\mathrm{supp}\,\mu_{\boldsymbol{T}_w} \subset [0,1]$. But from (2.22), we know that the empirical spectral measure of TT^* converges weakly to $\rho_w(d\lambda)$, which yields $\mathrm{supp}\rho_w \subset [0,1]$.

It follows from (2.19) and the definition of $M_{z,w}$ in (2.15) that for any $w \in \mathbb{C}_+ \cup \mathbb{R}$, $E \in \mathbb{R}$ and $z \in \mathbb{C}_+$

$$\|M_{z,w} - M_{z,E}\| \leq 2\, C_\gamma^2\, \big|w - E\big| \left(1 + \frac{1}{\mathrm{Im} z}\right)^2. \tag{2.25}$$

This implies the pointwise convergence

$$\lim_{w\to E} M_{z,w} = M_{z,E} \tag{2.26}$$

for all $z \in \mathbb{C}_+$. By (2.17) and $\rho_w(d\lambda) = \langle e_1, V_w(d\lambda)\, e_1\rangle$, the $(1,1)$-components $M_{z,w}(1,1)$ and $M_{z,E}(1,1)$ define the Stieltjes transforms of the measures $\rho_w(d\lambda)$ and $\rho_E(d\lambda)$ correspondingly, from which we conclude that (2.26) yields the weak convergence of $\rho_w(d\lambda)$ to $\rho_E(d\lambda)$ as $w \to E$. □

Note that $M_{z,w}$ is a matrix-valued Herglotz function. Therefore, from the properties of the (matrix-valued) Herglotz functions (see, e.g., [22, Theorems

2.2 and 5.4]), the *absolutely continuous* part of $\rho_w(d\lambda)$ is given by the inverse Stieltjes transform of $M_{z,w}(1,1)$ (see Lemma A.12)

$$\rho_w(\lambda) := \lim_{\eta\downarrow 0} \frac{1}{\pi} \mathrm{Im} M_{\lambda+\mathrm{i}\,\eta,w}(1,1). \tag{2.27}$$

We call the function $\rho_w(\lambda)$, defined in (2.27), the *self-consistent density of states* of the solution to the *DEL* (2.13). It will be shown in Sect. 2.5 that $\rho_w(d\lambda)$ is in fact purely absolutely continuous, i.e., $\rho_w(d\lambda) = \rho_w(\lambda)d\lambda$. The statements (a) and (b) of part (ii) of Theorem 1.4 will be derived from the study of $M_{z,w}$ for the spectral parameter z being close to the real line.

Note that our particular choice of linearization (2.1) allows rewriting the original *DEL* (2.13) in a slightly simpler form. More precisely, if

$$R = \begin{pmatrix} R_{11} & R_{12} & R_{13} \\ R_{21} & R_{22} & R_{23} \\ R_{31} & R_{32} & R_{33} \end{pmatrix} \in \mathbb{C}^{8\times 8} \tag{2.28}$$

with $R_{11}, R_{22} \in \mathbb{C}^{3\times 3}$ and $R_{33} \in \mathbb{C}^{2\times 2}$, then (2.12) yields

$$\Gamma[R] = \begin{pmatrix} \kappa_5^t R_{33}\kappa_5 & 0 & 0 \\ 0 & \kappa_4^t R_{33}\kappa_4 & 0 \\ 0 & 0 & \sigma_1 R_{33}\sigma_1 + \kappa_4 R_{22}\kappa_4^t + \kappa_5 R_{11}\kappa_5^t \end{pmatrix}, \tag{2.29}$$

so that the image $\Gamma[R]$ is a block-diagonal matrix. Together with the definition of K_0 in (2.12), this implies that the right-hand side in (2.13) is a block-diagonal matrix with blocks of sizes 3, 3, and 2 correspondingly. We conclude that any solution to the DEL (2.13) has a block-diagonal form, which, in particular, allows us to write

$$M_{z,w} = \begin{pmatrix} M_1 & 0 & 0 \\ 0 & M_2 & 0 \\ 0 & 0 & M_3 \end{pmatrix} \tag{2.30}$$

with $M_1, M_2 \in \mathbb{C}^{3\times 3}$ and $M_3 \in \mathbb{C}^{2\times 2}$, where we omit the dependence of the blocks on z and w. Now DEL (2.13) can be decomposed into a system of three matrix equations of smaller dimensions

$$-\frac{1}{M_1} = zJ_3 - \kappa_1 + \kappa_5^t M_3 \kappa_5, \quad -\frac{1}{M_2} = -\kappa_2 + \kappa_4^t M_3 \kappa_4 \tag{2.31}$$

and

$$\begin{aligned} -\frac{1}{M_3} = -\kappa_3(w) &- \frac{1}{-\frac{1}{2\gamma^2 z}(I_2+\sigma_3) - \frac{1}{\gamma}\sigma_2 + M_3} \\ &- \frac{1}{-2(I_2-\sigma_3) - \frac{1}{\gamma}\sigma_2 + M_3} + \sigma_1 M_3 \sigma_1, \end{aligned} \tag{2.32}$$

where $\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$, $\sigma_2 = \begin{pmatrix} 0 & -\mathrm{i} \\ \mathrm{i} & 0 \end{pmatrix}$ and $\sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$ are the standard Pauli matrices. The proof of Theorem 1.4 is based on the study of matrices M_1, M_2 and M_3.

2.2. Useful Identities

From now on and until the end of Sect. 2, we study the matrix-valued function $M_{z,w}$ with $w = E \in \mathbb{R}$. We start by collecting several important relations between the components of $M_{z,E}$.

Lemma 2.6. *For all $E \in \mathbb{R}$ and $z \in \mathbb{C}_+$*

$$M_{z,-E} = (Q^- \, M_{z,E} \, Q^-)^t, \tag{2.33}$$

where $Q^- = \operatorname{diag}(-1,-1,1,-1,1,1,-1,1) \in \mathbb{C}^{8\times 8}$. In particular, for all $E \in \mathbb{R}$, $z \in \mathbb{C}_+$ and $1 \le k \le 8$

$$M_{z,E}(k,k) = M_{z,-E}(k,k). \tag{2.34}$$

Proof. Let $\boldsymbol{H}_E^{(\mathrm{sc})} \in \mathcal{S}^{8\times 8}$ be given by

$$\boldsymbol{H}_E^{(\mathrm{sc})} := \left(\begin{array}{ccc|ccc|cc} 0 & 0 & 0 & 0 & 0 & 0 & 0 & c_2^* \\ 0 & 0 & \frac{\mathrm{i}}{\gamma}\mathbb{1}_\mathcal{S} & 0 & 0 & 0 & 0 & c_2^* \\ 0 & -\frac{\mathrm{i}}{\gamma}\mathbb{1}_\mathcal{S} & 0 & 0 & 0 & 0 & c_2^* & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & \frac{\mathrm{i}}{\gamma}\mathbb{1}_\mathcal{S} & 0 & c_1^* \\ 0 & 0 & 0 & 0 & -\frac{1}{4\gamma^2}\mathbb{1}_\mathcal{S} & 0 & c_1^* & 0 \\ 0 & 0 & 0 & -\frac{\mathrm{i}}{\gamma}\mathbb{1}_\mathcal{S} & 0 & 0 & c_1^* & 0 \\ \hline 0 & 0 & c_2 & 0 & c_1 & c_1 & 0 & E\mathbb{1}_\mathcal{S} - s \\ c_2 & c_2 & 0 & c_1 & 0 & 0 & E\mathbb{1}_\mathcal{S} - s & 0 \end{array}\right), \tag{2.35}$$

where s is a semicircular element, c_1, c_2 are circular elements, all freely independent in a C^*-probability space $(\mathcal{S}, \tau_\mathcal{S})$, so that

$$M_{z,E} := (\mathrm{id}_8 \otimes \tau_\mathcal{S})\big(\boldsymbol{H}_E^{(\mathrm{sc})} - zJ_8 \otimes \mathbb{1}_\mathcal{S}\big)^{-1}. \tag{2.36}$$

Using the fact that $-s$, $-c_1^*$ and $-c_2^*$ form again a freely independent family of one semicircular and two circular elements, we can easily check that (here $\times$ denotes multiplication in $\mathcal{S}^{8\times 8}$)

$$(\mathrm{id}_8 \otimes \tau_\mathcal{S})\Big(\big((Q^- \otimes \mathbb{1}_\mathcal{S}) \times \big(\boldsymbol{H}_{-E}^{(\mathrm{sc})} - zJ_8 \otimes \mathbb{1}_\mathcal{S}\big) \times (Q^- \otimes \mathbb{1}_\mathcal{S})\big)^t\Big)^{-1} = M_{z,E}, \tag{2.37}$$

from which (2.33) follows after factorizing Q^-. □

Lemma 2.7. *For all $E \in \mathbb{R}$ and $z \in \mathbb{C}_+$*

$$M_{z,E}(8,8) = 4\gamma^2 z \, M_{z,E}(7,7). \tag{2.38}$$

Proof. Using the Schur complement formula, the lower-right 2×2 block of the inverse of $\boldsymbol{H}_E^{(\mathrm{sc})} - zJ_8 \otimes \mathbb{1}_\mathcal{S}$ can be written as

$$\begin{pmatrix} 4\gamma^2 c_1 c_1^* & E\mathbb{1}_\mathcal{S} - s + \mathrm{i}\,\gamma(c_1c_1^* + c_2c_2^*) \\ E\mathbb{1}_\mathcal{S} - s - \mathrm{i}\,\gamma(c_1c_1^* + c_2c_2^*) & \frac{1}{z} c_2 c_2^* \end{pmatrix}^{-1}. \tag{2.39}$$

For convenience, change the rows in the above matrix, so that

$$M_3 \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = (\mathrm{id}_2 \otimes \tau_\mathcal{S}) \left[\begin{pmatrix} E\mathbb{1}_\mathcal{S} - a & \frac{1}{z} c_2 c_2^* \\ 4\gamma^2 c_1 c_1^* & E\mathbb{1}_\mathcal{S} - a^* \end{pmatrix}^{-1}\right], \tag{2.40}$$

where we introduced

$$a := s + \mathrm{i}\,\gamma\big(c_1c_1^* + c_2c_2^*\big). \tag{2.41}$$

Notice, that since $c_1c_1^* + c_2c_2^*$ has a free Poisson distribution of rate 2, $c_1c_1^* + c_2c_2^* \geq (1-\frac{1}{\sqrt{2}})^2\mathbb{1}_{\mathcal{S}}$ and thus both diagonal elements of the matrix on the right-hand side of (2.40) are invertible. Rewrite the matrix in the square brackets in the following way: for the entries of the first row apply the Schur complement formula with respect to the $(1,1)$-component, and for the second row apply the Schur complement formula with respect to the $(2,2)$-component. This leads to the following expressions for $M_{z,E}(7,7)$ and $M_{z,E}(8,8)$

$$M_{z,E}(7,7) = \frac{1}{z}\tau_{\mathcal{S}}\left(- \frac{1}{E\mathbb{1}_{\mathcal{S}} - a} c_2c_2^* \frac{1}{E\mathbb{1}_{\mathcal{S}} - a^* - \frac{4\gamma^2}{z} c_1c_1^* \frac{1}{E\mathbb{1}_{\mathcal{S}} - a} c_2c_2^*}\right), \tag{2.42}$$

$$M_{z,E}(8,8) = 4\gamma^2\tau_{\mathcal{S}}\left(- \frac{1}{E\mathbb{1}_{\mathcal{S}} - a^*} c_1c_1^* \frac{1}{E\mathbb{1}_{\mathcal{S}} - a - \frac{4\gamma^2}{z} c_2c_2^* \frac{1}{E\mathbb{1}_{\mathcal{S}} - a^*} c_1c_1^*}\right). \tag{2.43}$$

Under $\tau_{\mathcal{S}}$, we can swap the labels of c_1 and c_2 and replace a with $-a^*$ without changing the value in (2.43). After completing these operations, we obtain

$$M_{z,E}(8,8) = 4\gamma^2\tau_{\mathcal{S}}\left(- \frac{1}{E\mathbb{1}_{\mathcal{S}} + a} c_2c_2^* \frac{1}{E\mathbb{1}_{\mathcal{S}} + a^* - \frac{4\gamma^2}{z} c_1c_1^* \frac{1}{E\mathbb{1}_{\mathcal{S}} + a} c_2c_2^*}\right). \tag{2.44}$$

Multiplying both fractions under $\tau_{\mathcal{S}}$ in (2.44) by -1, and swapping E to $-E$ by (2.34), a comparison with (2.42) yields (2.38). □

Lemma 2.8. *For all $E \in \mathbb{R}$ and $z \in \mathbb{C}_+$*

$$M_{z,E}(8,7) - M_{z,E}(7,8) = \frac{\mathrm{i}}{\gamma} M_{z,E}(8,8) \tag{2.45}$$

Proof. Denote

$$T_1 := \begin{pmatrix} -\frac{1}{\gamma^2 z} & \frac{\mathrm{i}}{\gamma} \\ -\frac{\mathrm{i}}{\gamma} & 0 \end{pmatrix} + M_3, \quad T_2 := \begin{pmatrix} 0 & \frac{\mathrm{i}}{\gamma} \\ -\frac{\mathrm{i}}{\gamma} & -4 \end{pmatrix} + M_3, \tag{2.46}$$

so that (2.32) can be rewritten as

$$\frac{1}{M_3} - E\sigma_1 - \frac{1}{T_1} - \frac{1}{T_2} + \sigma_1 M_3 \sigma_1 = 0. \tag{2.47}$$

Then from (2.38), we have that

$$\det T_1 - \det T_2 = 4M_{z,E}(7,7) - \frac{1}{\gamma^2 z} M_{z,E}(8,8) = 0 \tag{2.48}$$

Rewrite (2.47) componentwise using (2.48)

$$\Big(\frac{1}{\det M_3} - \frac{2}{\det T_1} + 1\Big) M_{z,E}(8,8) = -\frac{4}{\det T_1}, \tag{2.49}$$

$$\Big(- \frac{1}{\det M_3} + \frac{2}{\det T_1}\Big) M_{z,E}(7,8) + M_{z,E}(8,7) = E - \frac{2\mathrm{i}}{\gamma}\frac{1}{\det T_1}, \tag{2.50}$$

$$\Big(-\frac{1}{\det M_3}+\frac{2}{\det T_1}\Big)M_{z,E}(8,7)+M_{z,E}(7,8)=E+\frac{2\mathrm{i}}{\gamma}\frac{1}{\det T_1}, \tag{2.51}$$

$$\Big(\frac{1}{\det M_3}-\frac{2}{\det T_1}+1\Big)M_{z,E}(7,7)=-\frac{1}{\gamma^2 z\det T_1}. \tag{2.52}$$

Subtracting (2.51) from (2.50) gives

$$\Big(\frac{1}{\det M_3}-\frac{2}{\det T_1}+1\Big)(M_{z,E}(8,7)-M_{z,E}(7,8))=-\frac{\mathrm{i}}{\gamma}\frac{4}{\det T_1}, \tag{2.53}$$

which together with (2.49) implies (2.45). □

2.3. Boundedness of $M_{z,E}$ Away from $z=0$ and $z=1$

Lemma 2.9. (Boundedness of $M_{z,E}$) *For any small $\theta>0$, there exists $C_\theta>0$ such that*

$$\sup\Big\{\|M_{z,E}\|\,:\,|z|\geq\theta,|1-z|\geq\theta,\,\mathrm{Im}z>0,\,|E|\leq\frac{1}{\theta}\Big\}\leq C_\theta. \tag{2.54}$$

Proof. Introduce the following notation for the entries of M_3

$$\begin{pmatrix} m_{11} & m_{12}\\ m_{21} & m_{22}\end{pmatrix}:=\begin{pmatrix} M_{z,E}(7,7) & M_{z,E}(7,8)\\ M_{z,E}(8,7) & M_{z,E}(8,8)\end{pmatrix}, \tag{2.55}$$

so that, in particular, (2.38) and (2.45) can be rewritten as

$$m_{22}=4\gamma^2 z m_{11}, \tag{2.56}$$

$$m_{21}-m_{12}=\frac{\mathrm{i}}{\gamma}m_{22}. \tag{2.57}$$

Our goal is to show that $M_{z,E}$ and in particular $M_{z,E}(1,1)$ is bounded everywhere if z is away from 0 or 1. From (2.31), we have that

$$\begin{aligned} M_1&=-\begin{pmatrix} m_{22}+z & m_{22} & m_{21}\\ m_{22} & m_{22} & m_{21}-\frac{\mathrm{i}}{\gamma}\\ m_{12} & m_{12}+\frac{\mathrm{i}}{\gamma} & m_{11}\end{pmatrix}^{-1},\\ M_2&=-\begin{pmatrix} m_{22} & m_{21} & m_{21}-\frac{\mathrm{i}}{\gamma}\\ m_{12} & m_{11}+\frac{1}{4\gamma^2} & m_{11}\\ m_{12}+\frac{\mathrm{i}}{\gamma} & m_{11} & m_{11}\end{pmatrix}^{-1}, \end{aligned} \tag{2.58}$$

which after some elementary computations yields

$$\det M_1=\frac{-1}{z\det T_1},\quad \det M_2=\frac{-\gamma^2}{\det T_1} \tag{2.59}$$

and

$$M_{z,E}(1,1)=-\frac{1}{z}-\frac{4m_{11}}{z\det T_1}, \tag{2.60}$$

where T_1 was defined in (2.46). The functions $\{m_{ij},1\leq i,j\leq 2\}$, $\{M_i,1\leq i\leq 3\}$ and $\{T_i,1\leq i\leq 2\}$ defined above all depend on the variables z and E, but in order to make the exposition lighter we drop the explicit dependence on these variables from the notation. Using (2.58)–(2.60) it is enough to show that for any fixed (z_∞, E_∞) with $z_\infty \in \overline{\mathbb{C}_+}$, $z_\infty \notin \{0,1\}$ and $E_\infty \in \mathbb{R}$ we have that $\lim_{(z,E)\to(z_\infty,E_\infty)} |m_{ij}| < \infty$ for $i,j \in \{1,2\}$ and $\lim_{(z,E)\to(z_\infty,E_\infty)} |\det T_1| > 0$.

We now prove some additional relations that can be obtained from (2.56), (2.57) and (2.49)–(2.52). Plugging (2.56) and (2.57) into (2.50) (recall that we are using notation (2.55)) gives

$$\Big(-\frac{1}{\det M_3}+\frac{2}{\det T_1}+1\Big)m_{12}+4\gamma \mathrm{i}\, z m_{11}+\frac{2\mathrm{i}}{\gamma \det T_1}-E=0, \tag{2.61}$$

which, after applying (2.52) to the terms in the parenthesis, can be rewritten as

$$\Big(\frac{1}{\gamma^2 z m_{11}\det T_1}+2\Big)m_{12}=E-\frac{2\mathrm{i}}{\gamma\det T_1}-4\gamma\mathrm{i}\, z m_{11}. \tag{2.62}$$

From the definitions of T_1 and M_3 we have

$$\det T_1=\det M_3+4(z-1)m_{11}-\frac{1}{\gamma^2}, \tag{2.63}$$

while (2.52) gives

$$\frac{1}{\det T_1}\Big(2-\frac{1}{\gamma^2 z m_{11}}\Big)=\frac{1}{\det M_3}+1. \tag{2.64}$$

Combining (2.63) and (2.64), we get the following quadratic equation for $\det M_3$

$$\begin{aligned}\big(\det M_3\big)^2&+\det M_3\bigg(4(z-1)m_{11}+\frac{1}{\gamma^2 z m_{11}}-\Big(1+\frac{1}{\gamma^2}\Big)\bigg)\\&+\bigg(4(z-1)m_{11}-\frac{1}{\gamma^2}\bigg)=0.\end{aligned} \tag{2.65}$$

Note that (2.62)–(2.65) hold for all $E\in\mathbb{R}$ and all $z\in\mathbb{C}_+$.

Using the above relations, we proceed with a proof by contradiction. Assume that there exists a sequence $(z_n,E_n)_{n=1}^\infty \subset \mathbb{C}_+\times[-\theta^{-1},\theta^{-1}]$, such that $|m_{11}^{(n)}|\to\infty$ as $n\to\infty$ (here and below we denote the evaluations at (z_n,E_n) by adding the superscript (n)). Solving (2.65) for $\det M_3$ allows us to express $\det M_3$ in terms of m_{11}

$$\begin{aligned}\det M_3=\frac{1}{2}\bigg\{&-4(z-1)m_{11}+\Big(1+\frac{1}{\gamma^2}\Big)\\&\pm 4(z-1)m_{11}\bigg[1-\frac{1}{2}\bigg(\frac{1}{2(z-1)}\Big(1+\frac{1}{\gamma^2}\Big)+\frac{1}{z-1}\bigg)\frac{1}{m_{11}}\\&+O\bigg(\frac{1}{|m_{11}|^2}\bigg)\bigg]\bigg\}.\end{aligned} \tag{2.66}$$

By passing to a subsequence, we may assume that the choice of the ± sign in (2.66) is constant for all n.

If we take the − sign in (2.66), then

$$\det M_3^{(n)}=-4(z_n-1)m_{11}^{(n)}+O(1) \tag{2.67}$$

and by (2.64)

$$\det T_1^{(n)}\to 2,\quad n\to\infty. \tag{2.68}$$

From (2.62), (2.56), (2.57) and (2.68),

$$m_{12}^{(n)} = -2\gamma \mathrm{i}\, z_n m_{11}^{(n)} + O\left(1\right), \quad m_{21}^{(n)} = 2\gamma \mathrm{i}\, z_n m_{11}^{(n)} + O\left(1\right), \tag{2.69}$$

and therefore

$$\det M_3^{(n)} = 4\gamma^2 z_n \left(m_{11}^{(n)}\right)^2 - 4\gamma^2 z_n^2 \left(m_{11}^{(n)}\right)^2 = 4\gamma^2 z_n(1-z_n)\left(m_{11}^{(n)}\right)^2 + O\left(m_{11}^{(n)}\right), \tag{2.70}$$

which contradicts to (2.67) since $|z_n(1-z_n)|$ is separated away from 0.

If we take + sign in (2.66), then

$$\det M_3^{(n)} = -1 + O\left(\frac{1}{|m_{11}^{(n)}|}\right), \tag{2.71}$$

and from (2.63)

$$\det T_1^{(n)} = 4(z_n - 1)m_{11}^{(n)} + O\left(1\right). \tag{2.72}$$

But then again, from (2.62), (2.56), (2.57) and (2.72),

$$m_{12}^{(n)} = -2\gamma \mathrm{i}\, z_n m_{11}^{(n)} + O\left(1\right), \quad m_{21}^{(n)} = 2\gamma \mathrm{i}\, z_n m_{11}^{(n)} + O\left(1\right), \tag{2.73}$$

so that

$$\det M_3^{(n)} = 4\gamma^2 z_n(1-z_n)(m_{11}^{(n)})^2 + O\left(\left|m_{11}^{(n)}\right|\right), \tag{2.74}$$

which contradicts to (2.71). Therefore, we have proven that m_{11} is bounded everywhere away from the points $z \in \{0,1\}$. It is clear from (2.56) that the boundedness of m_{11} guarantees the boundedness of m_{22}. On the other hand, assuming that m_{12} and m_{21} [see (2.57)] are unbounded implies that $|\det M_3^{(n)}| \to \infty$ and $|\det T_1^{(n)}| \to \infty$ on some sequence $(z_n, E_n)_{n=1}^{\infty}$, which contradicts to the boundedness of m_{11} in (2.61). We conclude that all entries of M_3 are bounded everywhere away from the points $z \in \{0,1\}$.

Assume now that there exists a sequence $(z_n, E_n)_{n=1}^{\infty} \subset \mathbb{C}_+ \times [-\theta^{-1}, \theta^{-1}]$ such that $\det T_1^{(n)} \to 0$ as $n \to \infty$. Then by (2.62)

$$\left(\frac{1}{\gamma^2 z_n m_{11}^{(n)}} + O\left(\det T_1^{(n)}\right)\right) m_{12}^{(n)} = -\frac{2\mathrm{i}}{\gamma} + O\left(\det T_1{}^{(n)}\right), \tag{2.75}$$

which together with the fact that $m_{11}^{(n)}$ is bounded implies that

$$m_{12}^{(n)} = -2\mathrm{i}\,\gamma z_n m_{11}^{(n)} + O\left(\det T_1^{(n)}\right). \tag{2.76}$$

Then by (2.56) and (2.57) we find

$$\det M_3^{(n)} = 4\gamma^2 z_n(1-z_n)\big(m_{11}^{(n)}\big)^2 + O\left(\det T_1^{(n)}\right), \tag{2.77}$$

and conclude from (2.77) and (2.63) that $\det M_3^{(n)}$ does not vanish as $n \to \infty$. But then (2.64) implies

$$m_{11}^{(n)} = \frac{1}{2\gamma^2 z_n} + O\left(\det T_1^{(n)}\right), \tag{2.78}$$

and thus by (2.63) and (2.78) also

$$\det M_3^{(n)} = \frac{1}{\gamma^2} - 2(z_n - 1)\frac{1}{\gamma^2 z_n} + O\left(\det T_1^{(n)}\right), \tag{2.79}$$

as well as

$$\det M_3^{(n)} = 4\gamma^2 z_n(1-z_n)\frac{1}{4\gamma^4 z_n^2} + O\left(\det T_1^{(n)}\right) = (1-z_n)\frac{1}{\gamma^2 z_n} + O\left(\det T_1^{(n)}\right), \tag{2.80}$$

by (2.77) and (2.78), which contradicts to (2.79) since z_n is away from 0. We conclude that $|\det T_1|$ is bounded away from 0, which together with (2.60) establishes (a non-effective version of) (2.54).

In order to keep the presentation simple, above we showed that $\|M_{z,E}\|$ is bounded for $z \in \overline{\mathbb{C}_+}\backslash\{0,1\}$ and $E \in \mathbb{R}$ without providing an explicit effective bound C_θ as formulated in (2.54). Note that the constants hidden in the $O(\cdot)$ terms (for example, in (2.67)–(2.70), (2.71)–(2.74) or (2.75)–(2.80)) depend only on $|z_n|$, $|1 - z_n|$ and E_n. Therefore, using the assumptions $|m_{11}^{(n)}| \geq D_\theta$ and $|\det T_1^{(n)}| \leq 1/D_\theta$ for n large enough and carefully chosen $D_\theta > 0$ instead of $|m_{11}^{(n)}| \to \infty$ and $|\det T_1^{(n)}| \to 0$ as $n \to \infty$, together with the a priori bound (2.16), eventually leads to the uniform bound (2.54). □

2.4. Singularities of $M_{z,E}(1, 1)$

Lemma 2.10. (Behavior of $M_{z,E}(1,1)$ in the vicinities of $z = 0$ and $z = 1$)

(a) *For all* $E \in \mathbb{R}$

$$M_{z,E}(1,1) = \sqrt[3]{\frac{1+E^2}{4\gamma^2}}\, z^{-2/3} + O\left(\frac{1}{|z|^{1/3}}\right) \quad \textit{as } z \to 0; \tag{2.81}$$

(b) *for all* $|E| < \frac{1}{\gamma}\sqrt{2\sqrt{1+6\gamma^2+\gamma^4} - 2 - 2\gamma^2}$

$$M_{z,E}(1,1) = -\frac{16\gamma^2\xi_0}{\gamma^2E^2 + 4 + 16\gamma^4\xi_0^2}\,(z-1)^{-1/2} + O(1) \quad \textit{as } z \to 1, \tag{2.82}$$

where the constant $\xi_0 < 0$ *is as in part (ii) of Theorem 1.4 (see also (2.95)).*

The choices of the branches for $(\cdot)^{1/3}$ *and* $(\cdot)^{1/2}$ *are specified in the course of the proof below.*

Proof. We multiply (2.32) from the left by M_3 and from the right by $(Z_1 + M_3)(Z_1 - Z_2)^{-1}(Z_2 + M_3)$ with the short hand notation

$$Z_1 := -\frac{1}{2\gamma^2 z}(I_2 + \sigma_3) - \frac{1}{\gamma}\sigma_2, \qquad Z_2 := -2(I_2 - \sigma_3) - \frac{1}{\gamma}\sigma_2. \tag{2.83}$$

Subsequently, the equation for M_3 takes the form $\Delta = 0$ with

$$\Delta := \Big(Z_1 - M_3 + M_3\sigma_1(M_3\sigma_1 - E)(Z_1 + M_3)\Big)\frac{1}{Z_1 - Z_2}(Z_2 + M_3) - M_3. \tag{2.84}$$

Using that $(Z_1 - Z_2)^{-1} = -\frac{\gamma^2 z}{2}(I_2 + \sigma_3) + \frac{1}{8}(I_2 - \sigma_3)$ and performing the matrix products we see that the entries of $\Delta \in \mathbb{C}^{2\times 2}$ are polynomials in the entries of M_3, E and z.

Step 1: Expansion around $z = 0$. We will now first construct a solution to (2.32) in a vicinity of $z = 0$ by asymptotic expansion. Later, in *Step 3*, we will show that the constructed solution coincides with M_3 defined in (2.15) and (2.30). For this purpose, we write $t = z^{1/3}$ with an analytic cubic root on $\mathbb{C}\backslash(\mathrm{i}\,(-\infty, 0]))$ such that $(-1)^{1/3} = -1$. Then, we make the ansatz

$$\widetilde{M_3} = \begin{pmatrix} \xi_{11}t^{-1} & 4\gamma^2\xi_{12}t \\ 4\gamma^2\xi_{21}t & 4\gamma^2\xi_{22}t^2 \end{pmatrix}, \tag{2.85}$$

where we will determine the unknown functions $\Xi_t := (\xi_{ij}(t))_{i,j=1}^2$. Plugging (2.85) into (2.84) reveals that

$$\Delta = \begin{pmatrix} 4\gamma^2(q_{11} + tp_{11}) & 4\gamma^2 t(q_{12} + tp_{12}) \\ -4\gamma^2 t(q_{21} + tp_{21}) & -16\gamma^4 t^3(q_{22} + tp_{22}) \end{pmatrix}, \tag{2.86}$$

where the entries of $P := (p_{ij})_{i,j=1}^2$ are polynomials in t, E, ξ_{ij} and where $Q := (q_{ij})_{i,j=1}^2$ is given by

$$Q = \begin{pmatrix} \frac{1}{16\gamma^4} + \xi_{11}^2\xi_{22} - E\xi_{11}\xi_{12} & \xi_{12} + E\xi_{11}\xi_{22} \\ \xi_{21} + E\xi_{11}\xi_{22} & -\frac{1}{16\gamma^4} - \xi_{11}\xi_{22}^2 + E\xi_{21}\xi_{22} \end{pmatrix}. \tag{2.87}$$

Thus, the equation $\Delta = 0$ is equivalent to $Q + tP = 0$. For $t = 0$, the three possible solutions are

$$\Xi_0 = \begin{pmatrix} \zeta & -E\zeta^2 \\ -E\zeta^2 & \zeta \end{pmatrix}, \quad \text{where} \quad (1 + E^2)\zeta^3 = -\frac{1}{16\gamma^4}. \tag{2.88}$$

Since $\det \nabla_\Xi Q|_{t=0} = -3\zeta^4(1 + E^2)^2$ the equation $Q + tP = 0$ is linearly stable at $t = 0$ and can thus be solved for Ξ_t in a neighborhood of Ξ_0 when t is sufficiently small. Since the equation is polynomial, the solution Ξ_t admits a power series expansion in t. Now we define the analytic function $\widetilde{M_3}(z)$ through (2.85) on a neighborhood of $z = 0$ in $\mathbb{C}\backslash(\mathrm{i}\,(-\infty, 0]))$ with Ξ the solution to $Q + tP = 0$ and the choice $\zeta < 0$ in (2.88).

We will now check that for any phase $e^{\mathrm{i}\,\psi} \neq -1$ in the complex upper half-plane the imaginary part of $\widetilde{M_3}(\theta e^{\mathrm{i}\,\psi})$ is positive definite in the sense that for any fixed $\varepsilon > 0$ we have

$$\inf_{\psi\in[-\pi+\varepsilon,0]} \inf_{\|u\|_2=1} \mathrm{Im}\langle u, \widetilde{M_3}(\theta e^{\mathrm{i}\,\psi})u\rangle > 0, \tag{2.89}$$

for sufficiently small $\theta > 0$. Indeed, for any vector $u = (u_1, u_2) \in \mathbb{C}^2$ we find

$$\begin{aligned} \mathrm{Im}\langle u, \widetilde{M_3}u\rangle &\geq \beta\mathrm{Im}(-t^{-1}|u_1|^2 - t^2|u_2|^2) - C|u_1||u_2||t| \\ &\geq \beta_\varepsilon\left(\frac{1}{|t|}|u_1|^2 + |t|^2|u_2|^2\right) - \frac{C}{R}\frac{1}{|t|}|u_1|^2 - CR|u_2||t|^3 \end{aligned} \tag{2.90}$$

for some constants $\beta, C > 0$, small t, an ε-dependent constant β_ε and any $R > 0$. For $R = 2C/\beta_\varepsilon$ and sufficiently small t this is still positive.

Step 2: Expansion around $z = 1$. For the expansion around $z = 1$, we proceed similarly to the discussion at $z = 0$. We set $t = \sqrt{z-1}$, where the square root has a branch cut at $\mathrm{i}\,(-\infty, 0]$ and $\sqrt{1} = 1$. Here, we make an ansatz with a reduced number of unknown functions by exploiting the identities (2.38) and (2.45), namely

$$\widetilde{M}_3 = \begin{pmatrix} \frac{\xi}{4\gamma^2 t} & \frac{E}{2} - \frac{\mathrm{i}\,\xi}{2\gamma}(t^{-1}+t) + \frac{\nu t}{4\gamma} \\ \frac{E}{2} + \frac{\mathrm{i}\,\xi}{2\gamma}(t^{-1}+t) + \frac{\nu t}{4\gamma} & \xi(t^{-1}+t) \end{pmatrix}. \tag{2.91}$$

We will determine the unknown functions ξ and ν. Plugging (2.91) into (2.84), multiplying out everything and simplifying afterwards reveals

$$\Delta = \begin{pmatrix} -\frac{1}{64\gamma^4}(q_1 + tp_1) & -\frac{1}{32\gamma^3}(q_2 + tp_2) \\ \frac{1}{32\gamma^3}(q_2 + tp_2) & \frac{1+t^2}{16\gamma^2}(q_1 + tp_1 - 2\mathrm{i}\,(q_2 + tp_2)) \end{pmatrix}, \tag{2.92}$$

where p_1, p_2 are polynomials in t, E, ξ, ν and where

$$\begin{aligned} q_1 =& \,\xi^4 + (2E^2\gamma^2 + 4 + 4\gamma^2)\xi^2 + \gamma^2(\gamma^2 E^4 + 4E^2(1+\gamma^2) - 16) \\ &- 16\mathrm{i}\,\gamma^3 E + \mathrm{i}\,E^2\gamma^2\xi\nu + 4\mathrm{i}\,\xi\nu + \mathrm{i}\,\xi^3\nu, \\ q_2 =& \,\xi^3\nu + (4 + \gamma^2 E^2)\xi\nu - 16\gamma^3 E\,. \end{aligned} \tag{2.93}$$

The solution to the system $(q_1, q_2) = 0$ at $t = 0$ has the form

$$\begin{aligned} \nu_0 = \frac{16\gamma^3 E}{\xi(4 + \gamma^2E^2 + \xi^2)}, \quad & r := \gamma^2 E^4 + \frac{1}{\gamma^2}\xi^4 \\ & + \frac{4}{\gamma^2}(1+\gamma^2)\xi^2 + 4E^2\left(1 + \gamma^2 + \frac{1}{2}\xi^2\right) - 16 = 0 \end{aligned} \tag{2.94}$$

In $r = 0$ from (2.94) we choose the unique negative solution (as long as the expression inside the square root is positive)

$$\xi_0 = -\sqrt{2\sqrt{1 + 6\gamma^2 + \gamma^4} - 2 - 2\gamma^2 - \gamma^2 E^2}. \tag{2.95}$$

We compute the Jacobian

$$\det \nabla_{\xi,\nu}(q_1, q_2) = 4\xi^2\left(8 + 2\gamma^2(12 + E^2) + \gamma^4\left(\frac{2}{\gamma^4}\xi^2 - 2E^2\right) - 2\gamma^2\xi^2 + \gamma^2 r\right), \tag{2.96}$$

and evaluate at $t = 0$ with $\xi = \xi_0$ to find

$$\det \nabla_{\xi_0,\nu_0}(q_1, q_2) = 16\xi_0^2\left(1{+}\gamma^4{+}\sqrt{1{+}6\gamma^2{+}\gamma^4} - \gamma^2\left(-6{+}\sqrt{1{+}6\gamma^2{+}\gamma^4}\right)\right) < 0. \tag{2.97}$$

In particular, ξ and ν admit a power series expansion in t.

Step 3: Coincidence of $\widetilde{M}_3$ *and* M_3, *asymptotic behavior of* $M_{z,E}(1,1)$. Here, we show that the asymptotic expansion $\widetilde{M}_3$ around $z = 0, 1$ coincides with M_3 defined in (2.15) and (2.30), the solution to *DEL* (2.32) constructed from free probability. For this purpose, for any small $\delta > 0$, we construct a modification $M^\delta_{z,E}$ of the explicit solution $M_{z,E}$ (given by (2.36)) as follows

$$M^\delta_{z,E} := (\mathrm{id}_8 \otimes \tau_{\mathcal{S}})\big(\boldsymbol{H}^{(\mathrm{sc}),\delta}_E - zJ_8 \otimes \mathbb{1}_{\mathcal{S}}\big)^{-1}, \tag{2.98}$$

with

$$\boldsymbol{H}_E^{(\mathrm{sc}),\delta} := \begin{pmatrix} \kappa_1 \otimes \mathbb{1}_{\mathcal{S}} & 0 & \kappa_5^t \otimes c_2^* \\ 0 & \kappa_2 \otimes \mathbb{1}_{\mathcal{S}} & \kappa_4^t \otimes c_1^* \\ \kappa_5 \otimes c_2 & \kappa_4 \otimes c_1 & \sigma_1 \otimes (E \cdot \mathbb{1}_{\mathcal{S}} - (1-\delta)\cdot s) - \delta \sigma_1 \widetilde{M}_3 \sigma_1 \otimes \mathbb{1}_{\mathcal{S}} \end{pmatrix}, \tag{2.99}$$

where c_i are circular and s semicircular elements. Since the original generalized resolvent has the bound $\|M_{z,E}\| \leq C(1 + \frac{1}{\mathrm{Im} z})$, we also get

$$\|M_{z,E}^\delta - M_{z,E}\| \leq C_{E,z}\delta. \tag{2.100}$$

Now, similarly as in (2.31)–(2.32), we derive the *DEL* corresponding to the generalized resolvent of $\boldsymbol{H}_E^{(\mathrm{sc}),\delta}$ and find

$$-\frac{1}{M_1^\delta} = zJ_3 - \kappa_1 + (1-\delta)\kappa_5^t M_3^\delta \kappa_5, \quad -\frac{1}{M_2^\delta} = -\kappa_2 + (1-\delta)\kappa_4^t M_3^\delta \kappa_4, \tag{2.101}$$

$$-\frac{1}{M_3^\delta} = -E\sigma_1 - \frac{1}{Z_1 + M_3^\delta} - \frac{1}{Z_2 + M_3^\delta} + (1-\delta)\sigma_1 M_3^\delta \sigma_1 + \delta \sigma_1 \widetilde{M}_3 \sigma_1, \tag{2.102}$$

where we exploited the block structure of $M_{z,E}^\delta$ and denoted

$$M_{z,E}^\delta = \begin{pmatrix} M_1^\delta & 0 & 0 \\ 0 & M_2^\delta & 0 \\ 0 & 0 & M_3^\delta \end{pmatrix}. \tag{2.103}$$

Notice that $\mathrm{Im}\big(\delta\sigma_1 \widetilde{M}_3 \sigma_1\big) > 0$ and $R \mapsto -E\sigma_1 - (Z_1 + R)^{-1} - (Z_2 + R)^{-1} + (1-\delta)\sigma_1 R \sigma_1$ is a positivity preserving analytic mapping. The existence and uniqueness of the solution to (2.102) now follow from [25, Theorem 2.1]. Inverting both sides of (2.102), we obtain a fixed point equation $M_3^\delta = \Phi_\delta(M_3^\delta)$ for M_3^δ, where the map Φ_δ can be read off from the inverse of the right-hand side of (2.102). Clearly, Φ_δ with any $\delta > 0$ is a contraction with respect to the Carathéodory metric mapping the set of 2×2 matrices with (strictly) positive definite imaginary parts strictly into itself (for more details, see [25]). In particular, there is a unique solution with positive semidefinite imaginary part and thus $M_3^\delta = \widetilde{M}_3$ for $\delta > 0$. Together with (2.100) and taking the limit $\delta \downarrow 0$, we conclude $M_3 = \widetilde{M}_3$. In particular, M_3 has a power law expansion around the singularities at $z = 0, 1$.

Finally, plugging the expansions (2.85) and (2.91) into (2.60) yields the asymptotics (2.81)–(2.82). □

2.5. Proof of (i) and (ii) of Theorem 1.4 for $\phi = 1$

We will need the following classical result from the theory of Herglotz functions (see, e.g., [22, Theorem 2.3])

Lemma 2.11. *Let $m : \mathbb{C}_+ \to \mathbb{C}_+$ be a Herglotz function with representation*

$$m(z) = m^\infty + \int_{\mathbb{R}} \frac{1}{\lambda - z} v(d\lambda) \tag{2.104}$$

for $m^\infty \in \mathbb{R}$ and a Borel measure $v(d\lambda)$ on $\mathbb{R}$. If for some $p > 1$ and interval $I \subset \mathbb{R}$

$$\sup_{0<\eta<1} \int_I |\mathrm{Im} m(\lambda + \mathrm{i}\,\eta)|^p d\lambda < \infty, \tag{2.105}$$

then the measure $v(d\lambda)$ is absolutely continuous on I.

Proof of (i) and (ii) of Theorem 1.4 for $\phi = 1$. The weak limits in the part (i) of Theorem 1.4 have been established in Lemma 2.5.

We know from (2.17) that $M_{z,E}(1,1)$ is a Herglotz function admitting the representation (2.104) with $m^\infty = M_E^\infty(1,1)$ and $v(d\lambda) = \langle e_1, V_E(d\lambda)\, e_1\rangle = \rho_E(d\lambda)$. Therefore, the absolute continuity of $\rho_E(d\lambda) = \rho_E(\lambda)d\lambda$ is a direct consequence of Lemma 2.11 and Lemmas 2.9 and 2.10 establishing together the integrability of $\lambda \mapsto \mathrm{Im} M_{\lambda+\mathrm{i}\,\eta,E}(1,1)$ in the form (2.105) on the whole real axis for any $p < 3/2$.

Finally, the asymptotic behavior of $\rho_E(\lambda)$ (1.13)–(1.14) near its singularities at 0 and 1 follows from Lemma 2.10 and the inverse Stieltjes transform formula (2.27). This, together with the global law established in Lemma 2.5, finishes the proof of Theorem 1.4 for $\phi = 1$. □

3. Proof of Theorem 1.4 for General Rational $\phi \in (0, 1)$

In this section, we explain how the techniques described in Sect. 2 can be used to prove Theorems 1.3 and 1.4 in the case when $\phi = k/l \in (0, 1)$ for fixed $k, l \in \mathbb{N}$, i.e., when $M = ln$ and $N = kn$ for some integer n tending to infinity. In addition to the steps used in Sect. 2, we need to tensorize the setup to accommodate for rectangular matrices. For example, the $M \times N = ln \times kn$ matrices W_1 and W_2 will be viewed as $l \times k$ rectangular block matrices with blocks of dimension $n \times n$. As we will see later, this allows us to treat the linearization matrix as a Kronecker random matrix in independent Wigner and i.i.d. matrices, which in turn makes various probabilistic estimates of the error terms in the corresponding *DEL* readily accessible from [3]. The restriction $\phi = k/l$ makes the presentation conceptually easier.

Note that different values of $\phi = k/l \in (0, 1)$ require slightly different approaches. For $\phi \in (1/2, 1)$, the matrix $\boldsymbol{T}_{E,\phi} \in \mathbb{C}^{kn\times kn}$ given by (1.10) is well defined and bounded with very high probability. Indeed, in this case the product WW^* is a sample covariance matrix with concentration ratio $\frac{1}{2\phi} \in (0, 1)$, therefore similarly as in Sect. 2.1, one can define a sequence of events $\Theta_{\phi,N}$ holding *a.w.o.p.* [see (2.4) and (3.2)] such that for any small $\delta > 0$ and big enough $n \geq n_0(\delta)$, the spectrum of WW^* is contained inside the interval $[(1-\frac{1}{\sqrt{2\phi}})^2(1-\delta), (1+\frac{1}{\sqrt{2\phi}})^2(1+\delta)]$ when restricted to $\Theta_{\phi,N}$ (see, e.g., [5, Section 5]). Thus, as n tends to infinity, the matrices in the denominator of (1.10) have positive imaginary parts and therefore bounded inverses with very high probability.

On the other hand, for $\phi \in (0, 1/2]$ we need to proceed via regularization by replacing $\mathrm{i}\,\gamma WW^*$ with $\mathrm{i}\,\gamma WW^* + \mathrm{i}\,\epsilon\phi I_M$ for some small $\epsilon > 0$ to ensure the invertibility of the denominators in (1.10). This requires a more careful analysis of the ϵ-dependence of various bounds and identities before we can take $\epsilon \to 0$.

Note also that for $\phi > 1$, $\mathrm{Rank}(\boldsymbol{T}_{E,\phi}) \leq M$ and the spectral measure of the matrix $\boldsymbol{T}_{E,\phi}$ has an atom of mass $1 - 1/2\phi$ at zero, while, as we will show, for $\phi < 1$ the spectral measure $\mu_{\boldsymbol{T}_{E,\phi}}$ does not have the pure point component. The regime $\phi = 1$ studied in Sect. 2 is borderline: the limiting spectral measure of $\boldsymbol{T}_{E,\phi}$ does not have atom at 0, but its behavior near the origin is more singular than in the case $\phi < 1$.

3.1. Linearization Trick and the Dyson Equation for Linearization

In order to apply the linearization trick for $\phi = k/l \in (0,1)$, we split H, W_1 and W_2 into blocks of size $n \times n$, so that

$$H = (\widehat{H}_{ij})_{\substack{i=1\ldots l \\ j=1\ldots l}}, \quad W_1 = (\widehat{W}_{1,ij})_{\substack{i=1\ldots l \\ j=1\ldots k}}, \quad W_2 = (\widehat{W}_{2,ij})_{\substack{i=1\ldots l \\ j=1\ldots k}}, \tag{3.1}$$

with $\widehat{H}_{ij}, \widehat{W}_{1,ij}, \widehat{W}_{2,ij} \in \mathbb{C}^{n\times n}$. Note that for any $i \in \{1,\ldots,l\}$, $\sqrt{l}\cdot\widehat{H}_{ii}$ is a (normalized) Wigner matrix of size n, and for any $i \neq j$, $\sqrt{l}\cdot\widehat{H}_{ij}$ is a (normalized) *i.i.d.* matrix of size n. Similarly, all the matrices $\sqrt{l}\cdot\widehat{W}_{1,ij}$ and $\sqrt{l}\cdot\widehat{W}_{2,ij}$ are also *i.i.d.* matrices of size n. All these matrices are independent apart from the natural constraint $\widehat{H}_{ij} = \widehat{H}^*_{ji}$.

Define the events

$$\Theta_{\phi,N} := \begin{cases} \{\|H\| \leq 3,\ \|W_1\| \leq 3,\ \|W_2\| \leq 3,\ \|(WW^*)^{-1}\| \leq \frac{1}{(1-\frac{1}{\sqrt{2\phi}})^2(1-\delta)}\}, & \phi \in (1/2,1), \\ \{\|H\| \leq 3,\ \|W_1\| \leq 3,\ \|W_2\| \leq 3\}, & \phi \in (0,1/2], \end{cases} \tag{3.2}$$

for some $\delta > 0$. Similarly as in Sect. 2.1, one can show that the events $\Theta_{\phi,N}$ hold *a.w.o.p.* and the random matrix models (1.9) and (1.10) for $\phi \in (1/2,1)$ and $w \in \mathbb{C}_+ \cup \mathbb{R}$ restricted to $\Theta_{\phi,N}$ are well defined (see Remark A.21 and Lemma B.1). At the same time, in order to deal with $\phi \in (0,1/2]$, we consider the regularized models (1.9) and (1.10) with $w \in \mathbb{C}_+$, i.e., $\mathrm{Im} w > 0$ strictly positive, which guarantees the invertibility of $w - H + \mathrm{i}\gamma WW^*$ without any additional restrictions on H and W.

The linearization matrix $\boldsymbol{H}_{w,\phi}$ for (1.10) is defined as in (2.1). This is a *Kronecker* random matrix consisting of $(6k+2l)\times(6k+2l)$ blocks of size n. We now introduce a tensorized version of the generalized resolvent that takes into account the additional structure coming from (3.1).

Definition 3.1 (*Generalized resolvent*). Let $(\phi, w) \in ((0,1/2]\times\mathbb{C}_+)\cup((1/2,1)\times(\mathbb{C}_+\cup\mathbb{R}))$ with $\phi = k/l \in \mathbb{Q}$. We call the matrix-valued function $\mathbb{C}_+ \ni z \mapsto (\boldsymbol{H}_{w,\phi} - z\boldsymbol{J}_k \otimes I_n)^{-1}$ the *generalized resolvent* of $\boldsymbol{H}_{w,\phi}$. Here, we denote $\boldsymbol{J}_k := \sum_{i=1}^{k} E_{ii} \in \mathbb{C}^{(6k+2l)\times(6k+2l)}$ with $\{E_{ij}\}$ being the standard basis of $\mathbb{C}^{(6k+2l)\times(6k+2l)}$.

Lemma 3.2 (Basic properties of the generalized resolvent).

(i) *For any* $\gamma > 0$ *and* $\phi \in (1/2,1) \cap \mathbb{Q}$, *there exists* $C_{\gamma,\phi} > 0$ *such that* a.w.o.p.

$$\big\|(\boldsymbol{H}_{w,\phi} - z\boldsymbol{J}_k \otimes I_n)^{-1}\big\| \le C_{\gamma,\phi}\left(1 + \frac{1}{\mathrm{Im} z}\right) \tag{3.3}$$

uniformly for all $z \in \mathbb{C}_+$ *and* $w \in \mathbb{C}_+ \cup \mathbb{R}$.

(ii) *For any* $\gamma > 0$, $\phi \in (0,1/2] \cap \mathbb{Q}$ *and* $w \in \mathbb{C}_+$, *there exists* $C_{\gamma,\phi,w} > 0$ *such that* a.w.o.p.

$$\big\|(\boldsymbol{H}_{w,\phi} - z\boldsymbol{J}_k \otimes I_n)^{-1}\big\| \le C_{\gamma,\phi,w}\left(1 + \frac{1}{\mathrm{Im} z}\right) \tag{3.4}$$

for all $z \in \mathbb{C}_+$.

(iii) *For all* $(\phi, w) \in \big(((0,1/2] \cap \mathbb{Q}) \times \mathbb{C}_+\big) \cup \big(((1/2,1) \cap \mathbb{Q}) \times (\mathbb{C}_+ \cup \mathbb{R})\big)$ *and* $1 \le i,j \le N$

$$\Big[(\boldsymbol{H}_{w,\phi} - z\boldsymbol{J}_k \otimes I_n)^{-1}\Big]_{ij} = \Big[(\boldsymbol{T}_{w,\phi} - zI_N)^{-1}\Big]_{ij}, \quad 1 \le i,j \le N. \tag{3.5}$$

Proof. Denote by $\mathcal{D}_{\boldsymbol{q}_0,\{q_{1,w}\};C}$ the effective domain related to the noncommutative rational function $q_{1,w}(x,y_1,y_2,y_1^*,y_2^*) = w - x + \mathrm{i}\,\gamma(y_1y_1^* + y_2y_2^*)$. More precisely [see (A.4)], the set $\mathcal{D}_{\boldsymbol{q}_0,\{q_{1,w}\};C}$ consists of all triples of elements $(\hat{x},\hat{y}_1,\hat{y}_2)$ such that

$$\left\|\frac{1}{q_{1,w}(\hat{x},\hat{y}_1,\hat{y}_2,\hat{y}_1^*,\hat{y}_2^*)}\right\| \le C. \tag{3.6}$$

As explained at the beginning of this section, for $\phi = k/l \in (1/2,1)$ *a.w.o.p.* the random matrices H, W_1 and W_2 defined in (3.1) belong to the domain $\mathcal{D}_{\boldsymbol{q}_0,\{q_{1,w}\};C}$ with constant [see (3.2) with $\delta = 1/2$] $C = \frac{1}{2\gamma(1-\frac{1}{\sqrt{2\phi}})}$ depending only on γ and ϕ.

For $\phi = k/l \in (0,1/2]$ on the other hand, the evaluation of the function $q_{1,w}$ on random matrices H, W_1 and W_2 is invertible only for $\mathrm{Im} w > 0$, in which case the norm of the inverse of $q_{1,w}$ evaluated on any triple $(\hat{x},\hat{y}_1,\hat{y}_2)$ is bounded by $\frac{1}{\gamma \mathrm{Im} w}$.

With the above choice of the constant C in $\mathcal{D}_{\boldsymbol{q}_0,\{q_{1,w}\};C}$ depending on the value of ϕ, the proof of (i)–(iii) can be obtained from the same argument as in Lemma 2.3 by restricting $\boldsymbol{H}_{w,\phi}$ to the events $\Theta_{\phi,N}$ defined in (3.2) and taking into account the dimensions of H, W_1 and W_2 and the relation $N = kn$ (see Remark A.21). □

From the structure of the linearization, we derive the DEL corresponding to $\boldsymbol{H}_{w,\phi}$

$$-\frac{1}{M} = z\boldsymbol{J}_k - K_0(w) + \Gamma_\phi[M] \tag{3.7}$$

for an unknown matrix-valued function M depending on z, w and ϕ, having the following components:

(i) the expectation matrix is given by

$$K_0(w) := \left(\begin{array}{c|c|c} \kappa_1 \otimes I_k & & \\ \hline & \kappa_2 \otimes I_k & \\ \hline & & \kappa_3(w) \otimes I_l \end{array}\right) \tag{3.8}$$

with matrices $\kappa_1, \kappa_2, \kappa_3(w)$ defined as in (2.9)–(2.10);

(ii) the operator $\Gamma_\phi : \mathbb{C}^{(6k+2l)\times(6k+2l)} \to \mathbb{C}^{(6k+2l)\times(6k+2l)}$ maps an arbitrary matrix

$$R = \begin{pmatrix} R_{11} & R_{12} & R_{13} \\ R_{21} & R_{22} & R_{23} \\ R_{31} & R_{32} & R_{33} \end{pmatrix} \in \mathbb{C}^{(6k+2l)\times(6k+2l)} \tag{3.9}$$

with $R_{11}, R_{22} \in \mathbb{C}^{3k\times 3k}$, $R_{33} \in \mathbb{C}^{2l\times 2l}$ into a block diagonal matrix with the first $3k \times 3k$ block equal to

$$\kappa_5^t\Big(\Big(\mathrm{id}_2 \otimes \frac{1}{l}\mathrm{Tr}_l\Big)R_{33}\Big)\kappa_5 \otimes I_k, \tag{3.10}$$

the second $3k \times 3k$ diagonal block equal to

$$\kappa_4^t\Big(\Big(\mathrm{id}_2 \otimes \frac{1}{l}\mathrm{Tr}_l\Big)R_{33}\Big)\kappa_4 \otimes I_k, \tag{3.11}$$

and the lower-right $2l \times 2l$ block equal to

$$\begin{aligned} &\kappa_5\Big(\Big(\mathrm{id}_3 \otimes \frac{1}{l}\mathrm{Tr}_k\Big)R_{11}\Big)\kappa_5^t \otimes I_l + \kappa_4\Big(\Big(\mathrm{id}_3 \otimes \frac{1}{l}\mathrm{Tr}_k\Big)R_{22}\Big)\kappa_4^t \otimes I_l \\ &\quad +\sigma_1\Big(\Big(\mathrm{id}_2 \otimes \frac{1}{l}\mathrm{Tr}_l\Big)R_{33}\Big)\sigma_1 \otimes I_l. \end{aligned} \tag{3.12}$$

Here, for $n \in \mathbb{N}$, we denote by Tr_n the trace of an $n \times n$ matrix κ_4, κ_5 are defined as in (2.10) and σ_1 is a standard Pauli matrix. The operator Γ_ϕ is the tensorized analogue of Γ from (2.29).

Now we can proceed similarly as for $\phi = 1$ in Sect. 2 just the $(3+3+2)\times(3+3+2)$ structure of the linearized matrices is replaced by larger block matrices structured as $(3k+3k+2l)\times(3k+3k+2l)$.

Lemma 3.3. (Existence and basic properties of the solution to the DEL (3.7)) *For any* $\gamma > 0$, $(\phi, w) \in (((0, 1/2] \cap \mathbb{Q}) \times \mathbb{C}_+) \cup (((1/2, 1) \cap \mathbb{Q}) \times (\mathbb{C}_+ \cup \mathbb{R}))$ *and* $z \in \mathbb{C}_+$ *define* $M_{z,w} \in \mathbb{C}^{(6k+2l)\times(6k+2l)}$ *as*

$$\begin{aligned} M_{z,w} := (\mathrm{id}_{6k+2l} \otimes \tau_{\mathcal{S}})\Big[\Big(&(K_0(w) - z\boldsymbol{J}_k) \otimes \mathbb{1}_{\mathcal{S}} + K_1 \otimes H^{(\mathrm{sc})} + L_1 \otimes W_1^{(\mathrm{sc})} \\ &+ L_1^* \otimes \Big(W_1^{(\mathrm{sc})}\Big)^* + L_2 \otimes W_2^{(\mathrm{sc})} + L_2^* \otimes \Big(W_2^{(\mathrm{sc})}\Big)^*\Big)^{-1}\Big], \end{aligned} \tag{3.13}$$

where $W_1^{(\mathrm{sc})}$ *and* $W_2^{(\mathrm{sc})}$ *are* $l\times k$ *matrices consisting of freely independent circular elements multiplied by* $1/\sqrt{l}$, *and* $H^{(\mathrm{sc})}$ *is an* $l \times l$ *self-adjoint matrix with freely independent semicircular elements multiplied by* $1/\sqrt{l}$ *on the diagonal and freely independent circulars multiplied by* $1/\sqrt{l}$ *above the diagonal.*

Then,

(i) *For any* $\gamma > 0$ *and* $\phi \in (1/2, 1)\cap \mathbb{Q}$, *there exists* $C_{\gamma,\phi} > 0$ *such that* $M_{z,w}$ *satisfies the a priori bound*

$$\|M_{z,w}\| \leq C_{\gamma,\phi}\Big(1 + \frac{1}{\mathrm{Im}z}\Big) \tag{3.14}$$

uniformly for all $w \in \mathbb{C}_+ \cup \mathbb{R}$ *and* $z \in \mathbb{C}_+$.

(ii) *For any* $\gamma > 0$, $\phi \in (0, 1/2]\cap \mathbb{Q}$ *and* $w \in \mathbb{C}_+$, *there exists* $C_{\gamma,\phi,w} > 0$ *such that function* $M_{z,w}$ *satisfies the a priori bound*

$$\|M_{z,w}\| \leq C_{\gamma,\phi,w}\Big(1 + \frac{1}{\mathrm{Im}z}\Big). \tag{3.15}$$

(iii) *For any* $\gamma > 0$, $(\phi, w) \in (((0, 1/2]\cap\mathbb{Q})\times\mathbb{C}_+)\cup(((1/2, 1)\cap\mathbb{Q})\times(\mathbb{C}_+\cup\mathbb{R}))$ *and* $z \in \mathbb{C}_+$, *matrix* $M_{z,w}$ *satisfies the DEL* (3.7) *and has positive semidefinite imaginary part,* $\mathrm{Im}M_{z,w} \geq 0$. *Moreover, for all* $\gamma > 0$, $(\phi, w) \in (((0, 1/2] \cap \mathbb{Q}) \times \mathbb{C}_+) \cup (((1/2, 1) \cap \mathbb{Q}) \times (\mathbb{C}_+ \cup \mathbb{R}))$, *the matrix-valued function* $z \mapsto M_{z,w}$ *is analytic on* $\mathbb{C}_+$.

(iv) *For any* $\gamma > 0$ *and* $(\phi, w) \in (((0, 1/2]\cap\mathbb{Q})\times\mathbb{C}_+)\cup(((1/2, 1)\cap\mathbb{Q})\times(\mathbb{C}_+\cup\mathbb{R}))$ *function* $z \mapsto M_{z,w}$ *admits the representation*

$$M_{z,w} = M^{\infty}_{w,\phi} + \int_{\mathbb{R}} \frac{V_{w,\phi}(d\lambda)}{\lambda - z}, \tag{3.16}$$

where $M^{\infty}_{w,\phi} \in \mathbb{C}^{(6k+2l)\times(6k+2l)}$ *is a self-adjoint matrix, and* $V_{w,\phi}(d\lambda)$ *is a positive-semidefinite matrix-valued measure on* $\mathbb{R}$ *with compact support.*

Proof. The proof follows from parts (i)–(v) of Lemma A.12 (see also Remark A.21). Similarly as in the proof of Lemma 3.2, notice that the noncommutative rational expression $q_{1,w} = w - x + \mathrm{i}\,\gamma(y_1y_1^* + y_2y_2^*)$ evaluated on matrices $x = H^{(\mathrm{sc})}$, $y_1 = W_1^{(\mathrm{sc})}$ and $y_2 = W_2^{(\mathrm{sc})}$ expresses different behavior in variable w for $\phi \in (0, 1/2)$ and $\phi \in [1/2, 1)$.

In the first case, $\phi \in (1/2, 1)$, the invertibility of $q_{1,w}$ evaluated on (s, c_1, c_2) does not depend on w and thus $(s, c_1, c_2) \in \mathcal{D}_{\boldsymbol{q}_0,\{q_{1,w}\};C}$ with $C = C(\gamma, \phi)$. In the case $\phi \in (0, 1/2]$, $q_{1,w}$ is invertible if and only if $\mathrm{Im}w > 0$, and the norm of $(q_{1,w})^{-1}$ depends on $\mathrm{Im}w$. This, in particular, means that $(s, c_1, c_2) \in \mathcal{D}_{\boldsymbol{q}_0,\{q_{1,w}\};C}$ with a w-dependent constant $C(\gamma, \phi, w)$.

This leads to two different a priori estimates: a bound (3.14) uniform in w for $\phi \in (1/2, 1)$ and a w-dependent bound for $\phi \in (0, 1/2]$. The rest of the proof follows directly from Lemma A.12. □

We omit the dependence of $M_{z,w}$ on ϕ for brevity. With these notations, we have the following global law establishing Theorem 1.3 and partially (i) of Theorem 1.4 for $\phi \in (0, 1)$. The proof of the weak limit (1.12) for $\phi \in (0, 1/2)$ is postponed to Sect. 3.6.

Lemma 3.4 (Global law for $\boldsymbol{T}_{w,\phi}$, $\phi \in (0,1)$). *For $(\phi, w) \in ((0,1/2] \times \mathbb{C}_+) \cup ((1/2,1) \times (\mathbb{C}_+ \cup \mathbb{R}))$, $\phi = k/l$, the empirical spectral measure $\mu_{\boldsymbol{T}_{w,\phi}}(d\lambda)$ converges weakly in probability (and almost surely) to $\rho_{w,\phi}(d\lambda)$, where*

$$\rho_{w,\phi}(d\lambda) := \frac{1}{k}\mathrm{Tr}(\boldsymbol{J}_k\, V_{w,\phi}(d\lambda)) \tag{3.17}$$

is the normalized trace of the upper-left $k \times k$ submatrix of the matrix-valued measure $V_{w,\phi}(d\lambda)$ from (3.16). The support of the measure $\rho_{w,\phi}(d\lambda)$ is a subset of the interval $[0,1]$. Moreover, for any $\phi \in (1/2,1)$ and $E \in \mathbb{R}$, the measure $\rho_{w,\phi}(d\lambda)$ converges weakly to $\rho_{E,\phi}(d\lambda)$ as $w \in \mathbb{C}_+$ tends to $E \in \mathbb{R}$.

Proof. The proofs are similar to the case $\phi = 1$ (see Lemma 2.5 and Remark A.21) after taking into account the dimensions of the matrices H, W_1 and W_2 and the additional structure (3.1). □

Definition 3.5 (*Self-consistent density of states*). We call the function

$$\rho_{w,\phi}(\lambda) := \lim_{\eta \downarrow 0} \frac{1}{\pi k}\mathrm{Im}\mathrm{Tr}(\boldsymbol{J}_k M_{\lambda + \mathrm{i}\,\eta, w}) \tag{3.18}$$

that gives the absolutely continuous part of $\rho_{w,\phi}(d\lambda)$, the *self-consistent density of states* of the model (1.10).

Since $\mathrm{supp}(\rho_{w,\phi}) \subset [0,1]$ by unitarity of $S(w)$, part (iii) of Theorem 1.4 can be established by proving the boundedness of the upper-left $k \times k$ minor of $M_{z,w}$ for the spectral parameter z bounded away from 0 and 1 (Sect. 3.3), and analyzing the asymptotic behavior of this upper-left submatrix in the vicinity of the special points $z = 0$ and $z = 1$ (Sect. 3.4). The study of $M_{z,w}$ is simplified by the particular form of $K_0(w)$ and Γ_ϕ, which implies that

$$M_{z,w} = \begin{pmatrix} M_1 \otimes I_k & & \\ & M_2 \otimes I_k & \\ & & M_3 \otimes I_l \end{pmatrix} \tag{3.19}$$

with $M_1, M_2 \in \mathbb{C}^{3\times 3}$ and $M_3 \in \mathbb{C}^{2\times 2}$ satisfying

$$-\frac{1}{M_1} = zJ_3 - \kappa_1 + \kappa_5^t M_3 \kappa_5, \quad -\frac{1}{M_2} = -\kappa_2 + \kappa_4^t M_3 \kappa_4 \tag{3.20}$$

and

$$-\frac{1}{M_3} = -\kappa_3(w) + \phi\kappa_5 M_1 \kappa_5^t + \phi\kappa_4 M_2 \kappa_4^t + \sigma_1 M_3 \sigma_1. \tag{3.21}$$

Similarly as in the case $\phi = 1$, plugging (3.20) into (3.21) leads to the following self-consistent equation for M_3

$$\begin{aligned} -\frac{1}{M_3} = -\kappa_3(w) &- \frac{\phi}{-\frac{1}{2\gamma^2 z}(I_2 + \sigma_3) - \frac{1}{\gamma}\sigma_2 + M_3} \\ &- \frac{\phi}{-2(I_2 - \sigma_3) - \frac{1}{\gamma}\sigma_2 + M_3} + \sigma_1 M_3 \sigma_1, \end{aligned} \tag{3.22}$$

which is the analogue of (2.32) for $\phi \neq 1$.

3.2. Useful Identities

Below we prove that identities similar to (2.34), (2.38) and (2.45) hold for $\phi \in (0,1)$.

Lemma 3.6. *For all* $(\phi, w) \in (((0,1/2] \cap \mathbb{Q}) \times \mathbb{C}_+) \cup (((1/2,1) \cap \mathbb{Q}) \times (\mathbb{C}_+ \cup \mathbb{R}))$, $\gamma > 0$ *and* $z \in \mathbb{C}_+$

(i) $M_{z,w}(i,i) = M_{z,-\overline{w}}(i,i)$ *for all* $1 \le i \le 6k + 2l$;

(ii) $M_{z,w}(6k+l+1, 6k+l+1) = 4\gamma^2 z M_{z,w}(6k+1, 6k+1)$;

(iii) $M_{z,w}(6k+l+1, 6k+1) - M_{z,w}(6k+1, 6k+l+1) = \frac{\mathrm{i}}{\gamma} M_{z,w}(6k+l+1, 6k+l+1)\Big(1 - \frac{\gamma \mathrm{Im} w}{2\phi} \det T_1\Big)$, *where, similarly as in* (2.46), *we denoted*

$$T_1 = \begin{pmatrix} -\frac{1}{\gamma^2 z} & \frac{\mathrm{i}}{\gamma} \\ -\frac{\mathrm{i}}{\gamma} & 0 \end{pmatrix} + M_3 \tag{3.23}$$

Denote the entries of M_3 in (3.19) by m_{ij}, $1 \le i,j \le 2$. Then, the parts (ii) and (iii) of the above lemma can be rewritten as

$$m_{22} = 4\gamma^2 z m_{11}, \tag{3.24}$$

$$m_{21} - m_{12} = \frac{\mathrm{i}}{\gamma} m_{22} \Big(1 - \frac{\gamma \mathrm{Im} w}{2\phi} \det T_1\Big). \tag{3.25}$$

Proof. In order to establish Lemma 3.6, we can follow the proofs of Lemmas 2.6–2.8 and apply them to the matrix

$$(K_0(w) - z\boldsymbol{J}_k) \otimes \mathbb{1}_{\mathcal{S}} + K_1 \otimes H^{(\mathrm{sc})} + L_1 \otimes W_1^{(\mathrm{sc})} + L_1^* \otimes \Big(W_1^{(\mathrm{sc})}\Big)^*$$
$$+ L_2 \otimes W_2^{(\mathrm{sc})} + L_2^* \otimes \Big(W_2^{(\mathrm{sc})}\Big)^*. \tag{3.26}$$

Note that the above matrix (3.26) is obtained from (2.35) by substituting c_1, c_2 and s with matrices $W_1^{(\mathrm{sc})}$, $W_2^{(\mathrm{sc})}$ and $H^{(\mathrm{sc})}$ correspondingly, and taking into account the dimensions of these matrices. For example, if we replace each diagonal entry of the matrix Q^- from the proof of Lemma 2.6 by the tensor product of this entry and a corresponding identity matrix (I_k or I_l), we obtain that the diagonal blocks of $M_{z,w}$ and $M_{z,-\overline{w}}$ coincide.

In order to prove (3.24), similarly as in the proof of Lemma 2.7, use the Schur complement formula with respect to the invertible upper-left $6k \times 6k$ submatrix of (3.26) to write the $2l \times 2l$ lower-right submatrix of its inverse as

$$\begin{pmatrix} 4\gamma^2 W_1^{(\mathrm{sc})} \Big(W_1^{(\mathrm{sc})}\Big)^* & w - H^{(\mathrm{sc})} + \mathrm{i}\,\gamma \Big(W_1^{(\mathrm{sc})} (W_1^{(\mathrm{sc})})^* + W_2^{(\mathrm{sc})} (W_2^{(\mathrm{sc})})^*\Big) \\ \overline{w} - H^{(\mathrm{sc})} - \mathrm{i}\,\gamma \Big(W_1^{(\mathrm{sc})} (W_1^{(\mathrm{sc})})^* + W_2^{(\mathrm{sc})} (W_2^{(\mathrm{sc})})^*\Big) & \frac{1}{z} W_2^{(\mathrm{sc})} (W_2^{(\mathrm{sc})})^* \end{pmatrix}^{-1}. \tag{3.27}$$

Then, we can switch the blocks of (3.27) as in (2.40) and apply the Schur complement formula with respect to

$$w - H^{(\mathrm{sc})} + \mathrm{i}\,\gamma \Big(W_1^{(\mathrm{sc})} (W_1^{(\mathrm{sc})})^* + W_2^{(\mathrm{sc})} (W_2^{(\mathrm{sc})})^*\Big). \tag{3.28}$$

The expression in (3.28) is invertible: for $(\phi, w) \in (((1/2,1) \cap \mathbb{Q}) \times (\mathbb{C}_+ \cup \mathbb{R}))$ the spectrum of $W_1^{(\mathrm{sc})} (W_1^{(\mathrm{sc})})^* + W_2^{(\mathrm{sc})} (W_2^{(\mathrm{sc})})^*$ follows the free Poisson distribution with rate $2\phi \in (1,2)$ and is therefore bounded away from zero, and for $(\phi, w) \in$

$(((0,1/2]\cap\mathbb{Q})\times\mathbb{C}_+)$ the invertibility is guaranteed by Imw being strictly positive. Due to the properties of freely independent circular and semicircular elements, switching the labels of the pair $(W_1^{(\mathrm{sc})},W_2^{(\mathrm{sc})})$ or changing the sign of $H^{(\mathrm{sc})}$ does not change the value of an expression involving these matrices after applying $\mathrm{id}_{6k+2l}\otimes\tau_{\mathcal{S}}$. Therefore, by proceeding as in (2.42)–(2.44) with $M_{z,w}(7,7)$ and $M_{z,w}(8,8)$ replaced by the corresponding $l\times l$ blocks of $M_{z,w}$, and using the diagonal structure of these blocks (3.19) and the part *(i)* of this lemma, we obtain (3.24).

Now it is straightforward to check by plugging (3.24) into (3.22) and following the proof of Lemma 2.8, that (3.25) holds. This proves Lemma 3.6.

□

3.3. Boundedness of $M_{z,w}$ away from $z=0$ and $z=1$ for $\phi\in(0,1)$

The goal of this section is to establish a uniform bound on $\|M_{z,w}\|$ for parameter w close to the real line and parameter z bounded away from 0 and 1. This will be used later in the proof of Theorem 1.4, in particular to show the absolute continuity of the measure $\rho_{E,\phi}(d\lambda)$. In the case $\phi\in(1/2,1)$ we can set $w=E\in\mathbb{R}$ and work directly with $M_{z,E}$. For $\phi\in(0,1/2]$, we prove the uniform bound for $\|M_{z,w}\|$ with $w=E+\mathrm{i}\,\phi\,\epsilon$ and small $\epsilon>0$, which will allow taking the limit $\epsilon\to 0$ in Sect. 3.4.

Lemma 3.7. (Boundedness of $M_{z,w}$)

(i) *Case* $\phi\in(1/2,1]$*: For any* $\gamma>0$ *and small* $\theta>0$*, there exists* $C_{\theta,\gamma}>0$ *such that*

$$\sup\left\{\,\|M_{z,E}\|\;:\;\phi\in(1/2,1)\cap\mathbb{Q},\;|z|\geq\theta,\;|1-z|\geq\theta,\;\mathrm{Im}z>0,\;|E|\leq\frac{1}{\theta}\right\}\leq C_{\theta,\gamma};\tag{3.29}$$

(ii) *Case* $\phi\in(0,1)$*: Let* $w=E+\mathrm{i}\,\phi\epsilon$*,* $\epsilon>0$*. For any* $\gamma>0$*, small* $\theta>0$*,* $\phi_0\in(0,1/2)$ *and* $\epsilon_0>0$ *small enough there exists* $C_{\theta,\gamma,\phi_0,\epsilon_0}>0$ *such that*

$$\|M_{z,w}\|\leq C_{\theta,\gamma,\phi_0,\epsilon_0}\tag{3.30}$$

uniformly on the set

$$\left\{\phi\in[\phi_0,1-\phi_0]\cap\mathbb{Q},|z|\geq\theta,|1-z|\geq\theta,\mathrm{Im}z>0,|E|\leq\frac{1}{\theta},\epsilon\in(0,\epsilon_0]\right\}.\tag{3.31}$$

Proof. Consider first (3.29) for which $\phi\in(1/2,1)$. By setting $\epsilon=0$ in Lemma 3.6, we can proceed by establishing (3.29) in the same manner as in the proof of Lemma 2.9. To guarantee a uniform bound in parameters z, E and ϕ, instead of a sequence $(z_n,E_n)_{n=1}^{\infty}$ as in the proof of Lemma 2.9, we assume the existence of a sequence $((z_n,E_n,\phi_n))_{n=1}^{\infty}$, $z_n\in\mathbb{C}_+$, $|E|\leq 1/\theta$, $\phi_n\in(1/2,1)$, on which $|m_{11}^{(n)}|\to\infty$ or $|\det T_1^{(n)}|\to 0$. Note that for $\phi_n\in(1/2,1)$ the leading terms in the analogues of the mutually contradicting pairs of statements (2.67)/(2.70), (2.71)/(2.74) and (2.79)/(2.80) do not depend on ϕ_n (with the exception of (2.71) where the constant -1 is replaced by ϕ_n-2).

For (3.30), i.e., $\phi\in[\phi_0,1-\phi_0]$, the above argument has to be slightly adjusted to ensure a uniform bound for small $\epsilon>0$. Instead of a sequence

$((z_n, E_n, \phi_n))_{n=1}^{\infty}$ as in the first part of the proof, we now assume the existence of a sequence $((z_n, E_n, \phi_n, \epsilon_n))_{n=1}^{\infty}$, $z_n \in \mathbb{C}_+$, $|E| \leq 1/\theta$, $\phi_n \in [\phi_0, 1-\phi_0]$, $\epsilon_n > 0$, on which $|m_{11}^{(n)}| \to \infty$ or $|\det T_1^{(n)}| \to 0$.

The analogues of (2.67)/(2.70) in this case are

$$\det M_3^{(n)} = -4(z_n - 1)m_{11}^{(n)} + O\left(1\right), \tag{3.32}$$

$$\det M_3^{(n)} = 4\gamma^2 z_n (1 - z_n(1 - \epsilon_n \phi_n \gamma)^2) \left(m_{11}^{(n)}\right)^2 + O\left(m_{11}^{(n)}\right), \tag{3.33}$$

which contradict each other if $|m_{11}^{(n)}| \to \infty$, $\epsilon_n \leq \epsilon_0$ is small enough and z_n is bounded away from 0 and 1.

Instead of the pair (2.71)/(2.74), we take the analogues of (2.71) and (2.72) for $\phi_n \in [\phi_0, 1-\phi_0]$ and $\epsilon_n > 0$

$$\det M_3^{(n)} = \phi_n - 2 + O\left(\frac{1}{|m_{11}^{(n)}|}\right), \tag{3.34}$$

$$\det T_1^{(n)} = 4(z_n - 1)m_{11}^{(n)} + O\left(1\right), \tag{3.35}$$

and observe that since $1 + 1/(\phi_n - 2) \geq \phi_0/(1+\phi_0) > 0$ for $\phi_n \in [\phi_0, 1-\phi_0]$, the above equations contradict to

$$\left(\frac{1}{\det M_3} - \frac{2\phi_n}{\det T_1} + 1\right) m_{11}^{(n)} = -\frac{\phi_n}{\gamma^2 z \det T_1}, \tag{3.36}$$

the analogue of (2.52), in the regime $|m_{11}^{(n)}| \to \infty$.

For the last pair (2.79)/(2.80) note that in the analogue of (2.75)

$$\left(\frac{\phi_n}{\gamma^2 z_n m_{11}^{(n)}} + O\left(\det T_1^{(n)}\right)\right) m_{12}^{(n)} = -\frac{2\phi_n \mathrm{i}}{\gamma} + O\left(\det {T_1}^{(n)}\right) \tag{3.37}$$

with $|\det {T_1}^{(n)}| \to 0$, the parameter ϕ_n disappears after dividing (3.37) by ϕ_n; therefore, we can proceed exactly as in the proof of Lemma 2.9. We conclude, similarly as in Lemma 2.9, that $\|M_{z,w}\|$ is uniformly bounded provided that $\min\{|z|, |1-z|\} \geq \theta$, $|E| \leq 1/\theta$, $\phi_0 \leq \phi \leq 1 - \phi_0$ and $\epsilon \leq \epsilon_0$ for some $\phi_0 > 0$ and $\theta, \epsilon_0 > 0$ small enough. □

3.4. Singularities of $M_{z,E}(1,1)$

For $\phi \in (1/2, 1)$, the solution matrix $M_{z,E}$ with $E \in \mathbb{R}$ is given directly via (3.13). For $\phi \in (0, 1/2]$, this formula cannot be directly applied when $w = E \in \mathbb{R}$ is real; we need an additional regularization argument. Nevertheless, in the next lemma we show the existence of the solution to the Dyson equation (3.20)–(3.22) for $\phi \in (0, 1/2]$ and $w = E \in \mathbb{R}$ and we establish the asymptotic behavior of $M_{z,E}(1,1)$ near $z = 0$ and $z = 1$ for $\phi \in (0,1)$. We start by constructing an expansion of the solution $M_{z,w}$ in the vicinity of $z = 0$ for $\phi \in (0,1)$ and $w = E + \mathrm{i}\,\phi\epsilon$ with $\epsilon > 0$ sufficiently small. We then use this expansion to extend $M_{z,w}$ to $w = E \in \mathbb{R}$ for $\phi \in (0, 1/2]$ by taking $\epsilon \downarrow 0$ and to study the asymptotic behavior of $M_{z,E}$ at special points $z = 0$ and $z = 1$.

Recall that the solution to the Dyson equation has the block structure (3.19), where M_1, M_2 are determined by M_3 through (3.20) and M_3 satisfies (3.22).

Lemma 3.8 (Existence of $M_{z,E}$ for $\phi \in (0,1/2]$ and singularities of $M_{z,E}(1,1)$ for $\phi \in (0,1)$).

(a) *For $\gamma > 0$, let $M_{z,w,\phi}$ denote the function $M_{z,w}$ defined in (3.13) evaluated at a point (z,w,ϕ) with*

$$z \in \mathbb{C}_+ \quad \text{and} \quad (w,\phi) \in \big(\mathbb{C}_+ \times \big((0,1/2] \cap \mathbb{Q}\big)\big) \cup \big((\mathbb{C}_+ \cup \mathbb{R}) \times \big((1/2,1) \cap \mathbb{Q}\big)\big). \tag{3.38}$$

Then $M_{z,w,\phi}$ can be continuously extended to the set

$$z \in \mathbb{C}_+ \quad \text{and} \quad (w,\phi) \in \mathbb{R} \times (0,1) \tag{3.39}$$

in the following sense: for any $(w,\phi) \in \mathbb{R} \times (0,1)$ and any sequence $\{(w_n,\phi_n)\}_{n\geq 1} \subset \big(\mathbb{C}_+ \times \big((0,1/2] \cap \mathbb{Q}\big)\big) \cup \big((\mathbb{C}_+ \cup \mathbb{R}) \times \big((1/2,1) \cap \mathbb{Q}\big)\big)$ with $(w_n,\phi_n) \to (w,\phi)$ as $n \to \infty$, there exists an analytic matrix-valued function $M_{z,w,\phi} : \mathbb{C}_+ \to \mathbb{C}^{(6k+2l)\times(6k+2l)}$, $z \mapsto M_{z,w,\phi}$, such that

$$M_{z,w_n,\phi_n} \to M_{z,w,\phi} \tag{3.40}$$

uniformly on compact z-subsets of $\mathbb{C}_+$ as $n \to \infty$.

For $\phi \in (0,1)$ and $w = E \in \mathbb{R}$, denote by $M_{z,E} := M_{z,E,\phi}$ the function defined in (3.40) omitting explicitly the dependence on ϕ.

(b) *For all $\phi \in (0,1)$, $\gamma > 0$ and $E \in \mathbb{R}$*

$$M_{z,E}(1,1) = \mathrm{i}\,\frac{4+\gamma^2\nu_0^2+\gamma^2\xi_0^2}{4\gamma\xi_0}\,z^{-1/2} + O(1) \ \text{ as } z \to 0, \tag{3.41}$$

with constants $\xi_0 := \xi_0(\phi,\gamma)$ and $\nu_0 := \nu_0(\phi,\gamma)$ given in (3.49) in the proof below.

(c) *For all $\phi \in (0,1)$, $\gamma > 0$ and $|E| \leq E_0 := E_0(\phi,\gamma)$*

$$M_{z,E}(1,1) = \frac{4\xi_0}{(\xi_0^2+\gamma^2E^2+4)}(z-1)^{-1/2} + O\,(1) \ \text{ as } z \to 1, \tag{3.42}$$

with constants E_0 and $\xi_0 := \xi_0(\phi,\gamma) < 0$ given in (3.52) and (3.54) correspondingly.

The branch of the square root is chosen to be continuous on $\mathbb{C}\backslash(\mathrm{i}(-\infty,0])$ such that $\sqrt{1} = 1$.

Proof. The analysis of (3.22) for $\phi \in (0,1)$ will follow similar steps as the analysis of (2.32) for the $\phi = 1$ case as performed in Sect. 2.4, and we will omit the details of some straightforward albeit tedious calculations. In the first step below, we analyze the solution to (3.22) for rational $\phi \in (0,1)$ and $w = E + \mathrm{i}\,\phi\,\epsilon$ with $E \in \mathbb{R}$ and $\epsilon > 0$ small enough. Using the same procedure that led to (2.84), we rewrite this equation as $\Delta = 0$ with

$$\begin{aligned}\Delta := \Big(&Z_1 + (1-2\phi)M_3 + M_3\sigma_1(M_3\sigma_1 - E + \mathrm{i}\epsilon\phi\sigma_3)(Z_1+M_3)\Big)\\ &\frac{1}{Z_1 - Z_2}(Z_2+M_3) - \phi M_3,\end{aligned} \tag{3.43}$$

and the two 2×2-matrices

$$Z_1 := -\frac{1}{2\gamma^2 z}(I_2 + \sigma_3) - \frac{1}{\gamma}\sigma_2, \qquad Z_2 := -2(I_2 - \sigma_3) - \frac{1}{\gamma}\sigma_2. \tag{3.44}$$

Step 1: Expansion around $z = 0$. We construct an expansion of M_3 as a power series in $t = \sqrt{z}$ in a neighborhood of $z = 0$. We make an ansatz compatible with the symmetries (3.24) and (3.25), namely

$$\widetilde{M}_3(t) = \begin{pmatrix} \frac{\mathrm{i}\xi}{4\gamma t} & \frac{\nu}{2} + \frac{\xi t}{2}(1 + \frac{\epsilon}{\gamma} f) + \epsilon \frac{\mathrm{i}\xi^2}{4} \\ \frac{\nu}{2} - \frac{\xi t}{2}(1 + \frac{\epsilon}{\gamma} f) - \epsilon \frac{\mathrm{i}\xi^2}{4} & \mathrm{i}\gamma\xi t \end{pmatrix}, \tag{3.45}$$

for the two functions $\xi = \xi(t), \nu = \nu(t)$ of t to be determined and where the parameters $E, \gamma, \phi, \epsilon$ are considered fixed. Later, we will show that with the right choice of functions ξ, ν this ansatz coincides with the solution of the Dyson equation, i.e., that $\widetilde{M}_3(t) = M_3(z)$. Here, f solves the equation $q_0 = 0$ for sufficiently small $\epsilon \geq 0$, where

$$q_0 = f + \frac{\gamma^2}{2} \det \left[\begin{pmatrix} -\frac{1}{\gamma^2 t^2} & \frac{\mathrm{i}}{\gamma} \\ -\frac{\mathrm{i}}{\gamma} & 0 \end{pmatrix} + \widetilde{M}_3 \right] + \frac{\mathrm{i}\gamma\xi}{2t} \tag{3.46}$$

is a polynomial in all variables $t, E, \gamma, \xi, \nu, \epsilon$ and f, in which it is quadratic. The choice (3.46) for f ensures that the symmetry condition (3.25) is satisfied.

We plug (3.45) into (3.43) and, after a mechanical but very long calculation that uses $q_0 = 0$ in the $(2, 1)$- and $(2, 2)$-entries of Δ, find

$$\Delta = \begin{pmatrix} -\frac{\mathrm{i}}{16\gamma t} q_{1,\epsilon,t} & \frac{1}{4} q_{2,\epsilon,t} \\ -\frac{1}{4} q_{2,\epsilon,t} & \frac{\mathrm{i}\gamma t}{4} q_{1,\epsilon,t} + (\mathrm{i}\gamma t^2 + \mathrm{i}t\epsilon f t - \frac{\gamma\epsilon\xi t}{2}) q_{2,\epsilon,t} \end{pmatrix}, \tag{3.47}$$

where

$$q_{1,\epsilon,t} = q_1 + \epsilon \widetilde{q}_1 + \frac{t}{4\gamma^3} p_1, \qquad q_{2,\epsilon,t} = q_2 + \frac{t}{\gamma} p_2,$$

and p_1, p_2 are polynomials in $t, E, \gamma, \phi, \xi, \nu, \epsilon$ and $q_1 = q_{1,0,0}, q_2 = q_{2,0,0}, \widetilde{q}_1$ are the following explicitly defined functions of the unknowns (ξ, ν):

$$\begin{aligned} q_1 &= \xi(\xi^2 - \nu^2 + 2E\nu + 4(\phi - 1)), \quad q_2 = \xi^2(\nu - E) + 2\nu\phi, \\ \widetilde{q}_1 &= -\mathrm{i}\,\xi q_2 + \epsilon \left(\frac{\xi^5}{4} + \xi^3 \phi \right). \end{aligned} \tag{3.48}$$

In particular, the equation $\Delta = 0$ is equivalent to $(q_{1,\epsilon,t}, q_{2,\epsilon,t}) = 0$ and thus to $(q_1, q_2) = 0$ in the limit $t \to 0$ and $\epsilon \to 0$. This, in turn, fixes the values for $\nu_0 = \nu|_{t=0}$ and $\xi_0 = \xi|_{t=0}$ through

$$\nu_0 = \frac{E\xi_0^2}{\xi_0^2 + 2\phi}, \quad r := \xi_0^6 + (E^2 + 8\phi - 4)\xi_0^4 + 4\phi(E^2 + 5\phi - 4)\xi_0^2 + 16(\phi - 1)\phi^2 = 0, \tag{3.49}$$

where we choose the positive solution ξ_0 for $r = 0$. The fact that $r = 0$ has a unique positive solution $\xi_0 > 0$ is an explicit elementary calculation. The positivity of ξ_0 will ensure the positive definiteness of $\mathrm{Im}\widetilde{M}_3$ for $z \in \mathbb{C}_+$.

We compute the Jacobian of the function (q_1, q_2) from (3.48) as

$$J(\xi,\nu) := \det \nabla_{\xi,\nu}(q_1,q_2) = 3\xi^4 + ((3\nu^2 - 6\nu E + 4E^2) + 10\phi - 4)\xi^2 \\ +2\phi(4\phi - 4 + 2E\nu - \nu^2). \tag{3.50}$$

Using that at $t = 0$ we have $q_1 = 0$ and $\xi_0 \neq 0$, we can eliminate the quadratic terms in ν and obtain

$$J(\xi_0,\nu_0) = 2\xi_0^2(3\xi_0^2 + 2E^2 + 10\phi - 8).$$

Again an elementary calculation using the defining equation $r = 0$ for ξ_0 shows that $J(\xi_0,\nu_0)$ never vanishes. Thus, the ansatz (3.45) solves the Dyson equation in a small neighborhood of $z = t^2 = 0$. Furthermore, for $t = u(1 + \mathrm{i}u)$ with sufficiently small $u > 0$ it is easy to see that its imaginary part is positive definite. By using a regularization argument analogous to the one from Step 3 of Sect. 2.4 and combining it with the uniqueness of solutions to the Dyson equation with positive definite imaginary part, this implies that $\widetilde{M_3}(t) = M_3(z)$ for all $\sqrt{z} = t = u(1 + \mathrm{i}u)$, $\phi \in (0,1)$ and small enough $\epsilon > 0$. Since both functions are analytic this also implies equality for $z = t^2$ in the complex upper half plane intersected with a neighborhood of $z = 0$.

Step 2: Extending $M_{z,w}$ to $w = E \in \mathbb{R}$ for $\phi \in (0,1)$. Fix $\phi \in (0,1)$, $\gamma > 0$ and $E \in \mathbb{R}$. For $\epsilon_0 > 0$ and a coordinate pair $(i,j) \in \{1,\ldots,6k + 2l\}^2$, consider the family of functions $\{M_{z,E+\mathrm{i}\,\epsilon_n\phi_n}\,(i,j)\}_{n\geq 1}$ analytic in z with $\{\phi_n,\epsilon_n\}_{n\geq 1} \subset ((0,1)\cap\mathbb{Q}) \times (0,\epsilon_0)$ and $(\phi_n,\epsilon_n) \to (\phi,0)$ as $n \to \infty$. It follows from part *(ii)* of Lemma 3.7 that for small enough ϵ_0 and any small $\theta > 0$ the family of functions $\{M_{z,E+\mathrm{i}\,\epsilon_n\phi_n}\,(i,j)\}_{n\geq 1}$ is uniformly bounded on the set $\{z \in \mathbb{C}_+ : |z| \geq \theta, |1-z| \geq \theta\}$, and thus locally bounded on $\mathbb{C}_+$.

It was established in Step 1 above that for any $\gamma > 0$, $E \in \mathbb{R}$, $\phi \in (0,1)$ and $\epsilon > 0$ small enough the solution to Eq. (3.22) with positive semidefinite imaginary part can be explicitly given as $\widetilde{M_3}(\sqrt{z})$ [see (3.45)] in the neighborhood of the origin, i.e., on the set $\{|z| \leq \delta, \mathrm{Im} z > 0\}$ for $\delta > 0$ sufficiently small. Moreover, Step 1 shows that for any $z \in \{|z| \leq \delta, \mathrm{Im} z > 0\}$, there exists a well-defined limit $\lim_{(\phi_n,\epsilon_n)\to(\phi,0)} \widetilde{M_3}(\sqrt{z})$. Together with (3.19) and (3.20), this implies that on the set $\{|z| \leq \delta, \mathrm{Im} z > 0\}$ the limit $M_{z,E} := \lim_{(\phi_n,\epsilon_n)\to(\phi,0)} M_{z,E+\mathrm{i}\,\epsilon_n\phi_n}$ exists as well.

Combining the above information, we see that for any index pair (i,j), $\{M_{z,E+\mathrm{i}\,\epsilon_n\phi_n}\,(i,j)\}_{n\geq 1}$ is a family of analytic functions, locally bounded on $\mathbb{C}_+$ that converges on $\{|z| \leq \delta, \mathrm{Im} z > 0\}$ to $M_{z,E}(i,j)$. Applying the Vitali–Porter theorem (see, e.g., Section 2.4 in [34]), we conclude that for any $z \in \mathbb{C}_+$ the limit $M_{z,E}\,(i,j) := \lim_{(\phi_n,\epsilon_n)\to(\phi,0)} M_{z,E+\mathrm{i}\,\epsilon_n\phi_n}\,(i,j)$ exists, the convergence holds uniformly on the compact subsets of $\mathbb{C}_+$ and, as a result, the function $z \mapsto M_{z,E}\,(i,j)$ is analytic on $\mathbb{C}_+$. Taking $M_{z,E}\,(i,j)$ as the entries of the matrix-valued function $M_{z,E}$ defines the solution to the Dyson equation (3.21) for $\phi \in (0,1)$ and $w = E \in \mathbb{R}$.

To compute the asymptotic behavior of $M_{z,E}(1,1)$, we use (3.45) with $\epsilon = 0$ [see (3.48)], (3.20) and $t = \sqrt{z}$ to find M_1 and its upper left corner element $M_{z,E}(1,1)$ in the neighborhood of $z = 0$

$$M_{z,E}(1,1) = \mathrm{i}\frac{4 + \gamma^2\nu_0^2 + \gamma^2\xi_0^2}{4\gamma\xi_0}\, z^{-1/2} + O(1).$$

Step 3: Expansion around $z = 1$. We apply exactly the same procedure as in Step 2 of Lemma 2.10, just we insert the parameters ϕ and γ into the identities (2.83) and (2.84) and follow them through the analysis. Here, we record the final result of this elementary calculation. Our ansatz is

$$\widetilde{M}_3 = \begin{pmatrix} \frac{\xi}{4\gamma^2 t} & \frac{E}{2} + \frac{\nu t}{4\gamma} - \frac{\mathrm{i}\xi}{2\gamma}(t+t^{-1}) \\ \frac{E}{2} + \frac{\nu t}{4\gamma} + \frac{\mathrm{i}\xi}{2\gamma}(t+t^{-1}) & \xi(t+t^{-1}) \end{pmatrix}. \tag{3.51}$$

The expansion in $t = \sqrt{z-1}$ gives that for all $|E| \le E_0$ with

$$E_0 := \frac{\sqrt{2}}{\gamma}\Big(\gamma^2(1-2\phi) - 1 + \sqrt{1+\gamma^4(1-2\phi)^2 + 2\gamma^2(1+2\phi)}\Big)^{1/2} \tag{3.52}$$

the upper-left component of $M_{z,E}$ is given by

$$M(1,1) = \frac{4\xi_0}{(\xi_0^2+\gamma^2E^2+4)t} + O(1), \tag{3.53}$$

where ξ_0 is defined by

$$\xi_0 = -\gamma\sqrt{E_0^2 - E^2}. \tag{3.54}$$

This finishes the proof of the lemma. □

3.5. Explicit Solution for $\phi \longrightarrow 0$

Lemma 3.9. *Let $\gamma > 0$, $w = E \in \mathbb{R}$ and $\phi = 0$. Then, the Dyson equation (3.21)–(3.22) admits a solution $\mathbb{C}_+ \ni z \mapsto M_{z,E} \in \mathbb{C}^{8\times 8}$ with $\mathrm{Im} M_{z,E} \ge 0$ and the upper-left entry is explicitly given by*

$$M_{z,E}(1,1) = \frac{\gamma^2(4-E^2) - (1+\gamma^2)^2 + \mathrm{i}\gamma(1+\gamma^2)}{\gamma^2(4-E^2) + ((1+\gamma^2)^2 - (4-E^2)\gamma^2)z}\sqrt{\frac{4-E^2}{z(1-z)}}. \tag{3.55}$$

Moreover, this solution can be continuously extended to the set $\phi \in [0,\phi_0]$, $|1-z| \ge \theta$, $\theta \le |z| \le \theta^{-1}$ for $E \in \mathbb{R}$ and sufficiently small $\theta > 0$ and $\phi_0 = \phi_0(\theta) > 0$.

Proof. At $\epsilon = 0$ and setting $\phi = 0$, the Eq. (3.22) simplifies to a quadratic matrix equation for $M_3\sigma_1$, where M_3 satisfies the symmetry constraints (3.24) and (3.25), i.e.,

$$-\frac{1}{M_3} = -E\,\sigma_1 + \sigma_1 M_3 \sigma_1, \qquad M_3 = \begin{pmatrix} \frac{\mathrm{i}\xi}{4\gamma} & \frac{\nu}{2} + \frac{z\xi}{2} \\ \frac{\nu}{2} - \frac{z\xi}{2} & z\gamma\mathrm{i}\xi \end{pmatrix}, \tag{3.56}$$

for two functions ξ and ν that are easily computed to be $\nu = E$ and

$$\xi = \sqrt{\frac{4-E^2}{z(1-z)}}. \tag{3.57}$$

The choice of root is determined by $\mathrm{Re}\xi > 0$. Inserting M_3 into (3.20) leads to

$$M_{z,E}(1,1) = \frac{\gamma^2(4-E^2) - (1+\gamma^2)^2 + \mathrm{i}\gamma(1+\gamma^2)\xi}{\gamma^2(4-E^2) + ((1+\gamma^2)^2 - (4-E^2)\gamma^2)z}. \tag{3.58}$$

We now construct the solution of (3.22) perturbatively for $\phi \in (0,\phi_0)$ with some sufficiently small $\phi_0 = \phi_0(\theta) > 0$ with $|1-z| \ge \theta$ and $\theta \le |z| \le \theta^{-1}$.

Recall that (3.22) is equivalent to $\Delta = 0$ with Δ defined as in (3.43). In particular,

$$\Delta = \widetilde{\Delta}(Z_1 + M_3)\frac{1}{Z_1 - Z_2}(Z_2 + M_3), \tag{3.59}$$

implicitly defining $\widetilde{\Delta}$ that satisfies

$$\widetilde{\Delta}|_{\phi=0} = 1 + M_3(\sigma_1 M_3 \sigma_1 - \sigma_1 E). \tag{3.60}$$

Similarly as we did in the proof of Lemma 3.8 instead of considering the equation $\Delta = 0$ for a solution $M_3 \in \mathbb{C}^{2\times 2}$, we can equivalently consider it as an equation for the two unknown functions ξ and ν from the ansatz (3.56) for M_3. Since clearly the factors $Z_1 + M_3$ and $Z_2 + M_3$ in (3.59) have bounded inverses when M_3 is the explicit solution at $\phi = 0$ from (3.56) with (3.57), we can equivalently consider $\widetilde{\Delta} = 0$ as the equation for ξ and ν. Plugging the ansatz (3.56) into (3.60), we see that $\widetilde{\Delta}_{11} = 0$ and $\widetilde{\Delta}_{12} = 0$ already imply $\widetilde{\Delta} = 0$ and that

$$\begin{aligned}\widetilde{\Delta}_{11}|_{\phi=0} &= \frac{1}{4}(4 + \nu^2 + 2z\nu\xi - z\xi^2 + z^2\xi^2 - 2E(\nu + z\xi)),\\ \widetilde{\Delta}_{12}|_{\phi=0} &= -\frac{\mathrm{i}(E-\nu)\xi}{4\gamma}.\end{aligned}$$

We compute the determinant of the Jacobian of the function $(\xi, \nu) \to (\widetilde{\Delta}_{11}, \widetilde{\Delta}_{12})$ to be

$$\det \nabla_{\xi,\nu}(\widetilde{\Delta}_{11}, \widetilde{\Delta}_{12})|_{\phi=0,\nu=E} = \frac{\mathrm{i}(z-1)z\xi^2}{8\gamma}.$$

Since ξ from (3.57) does not vanish we infer that (3.22) is linearly stable as an equation for ξ, ν for small enough parameters ϕ in a vicinity of the explicit solution M_3 from (3.56) with (3.57) for $\phi = 0$. □

From (3.55), we read off the density of transmission eigenvalues $\rho(\lambda) = \rho_{E,\phi=0}(\lambda) := \frac{1}{\pi}\lim_{\eta\downarrow 0}\mathrm{Im} M_{\lambda+\mathrm{i}\eta,E}(1,1)$. The corresponding Fano factor for $\phi = 0$ is now computable as

$$F(E,\gamma) = 1 - \frac{\int \lambda^2 \rho(\lambda)\mathrm{d}\lambda}{\int \lambda\rho(\lambda)\mathrm{d}\lambda} = \frac{1+\gamma^2}{2(1+\gamma^2+\gamma\sqrt{4-E^2})}. \tag{3.61}$$

For $\gamma = 1$ and $E = 0$, we recover the density (1.5) and the Fano factor $F = \frac{1}{4}$ obtained in [7], see Sect. 4 for more details.

3.6. Proof of Parts (i), (iii) and (iv) of Theorem 1.4

In this section, we collect the results established in Sects. 3.1–3.5 and complete the proof of Theorem 1.4. Recall that Theorem 1.3 was proven in Lemma 2.5 for $\phi = 1$ and Lemma 3.4 for $\phi \in (0,1)$.

Proof of part (i) of Theorem 1.4. The extension of $\rho_{w,\phi}(d\lambda)$ to $w = E \in \mathbb{R}$ for $\phi \in (0,1/2] \cap \mathbb{Q}$ as well as the limit (1.12) and the extension of $\rho_{E,\phi}(d\lambda)$ to irrational $\phi \in (0,1)$ follows from the equivalence between the weak convergence of measures defined by (3.16) and the pointwise convergence of $M_{z,w}$

established in Lemma 3.4 and part (a) of Lemma 3.8 for the corresponding limits.

The weak limit $\lim_{\phi\downarrow 0}\rho_{E,\phi}(d\lambda)=\rho_{E,\phi=0}(\lambda)d\lambda$ follows from the continuity of $\phi\mapsto M_{z,E}$ at $\phi=0$ for all $z\in\mathbb{C}_+$, which was established in Lemma 3.8 in the regimes $|z|\leq\theta$, $|1-z|\leq\theta$ and in Lemma 3.9 in the complementary regimes $|z|\geq\theta$, $\theta\leq|1-z|\leq\theta^{-1}$. Since the Stieltjes transform of $\rho_{E,\phi}$ is given by $M_{z,E}(1,1)$, the exact expression for $\rho_{E,0}$ can be derived as an inverse Stieltjes transform of $M_{z,E}$ from (3.55).

Similarly as for the case $\phi=1$ in Sect. 2.5, Lemma 2.11 and the integrability of $M_{z,E}$ that can be deduced from Lemmas 3.7 and 3.8 yield the absolute continuity of $\rho_{w,\phi}(d\lambda)=\rho_{w,\phi}(\lambda)d\lambda$. □

Proof of part (iii) of Theorem 1.4. Follows from the asymptotic behavior of $M_{z,E}$ near $z=0$ and $z=1$ (3.41)–(3.42) established in Lemma 3.8, block structure of $M_{z,E}$ (3.19) and the definition of the self-consistent density of states (3.18). □

Proof of part (iv) of Theorem 1.4. Follows from the explicit formula for $M_{z,E}$ in the regime $\phi\to 0$ (3.58) established in Sect. 3.5, the block structure of $M_{z,E}$ (3.19) and the definition of the self-consistent density of states (3.18). □

4. Comparison with the Results of Beenakker and Brouwer

Consider the scattering matrix (1.9)

$$S(E):=I-2\gamma\mathrm{i}\,W^*(E\cdot I-H+\mathrm{i}\gamma WW^*)^{-1}W\in\mathbb{C}^{N_0\times N_0}\tag{4.1}$$

with $w=E\in\mathbb{R}$ in the regime $\phi=N/M\to 0$ as $M\to\infty$ with $N_0=2N$. This model was studied in the case of the Gaussian entries by Beenakker and Brouwer in [7,10], and one of the remarkable results of their theory is that in the experimentally relevant setting of the ideal coupling the limiting transmission eigenvalue density is given by the arcsine law (1.5) (see [8, Eq. (3.12)]. The ideal coupling assumption is formulated in terms of the matrix $S(E)$ having zero mean [8, Eq. (3.8)]. Below we show that in the regime $\phi\to 0$ the assumption $\mathbb{E}[S(E)]=0$ is equivalent to $E=0$ and $\gamma=1$. By plugging these values into (1.19) and (3.61), we recover the arcsine distribution and the corresponding Fano factor $F(0,1)=1/4$.

Since the results of the current section do not affect the main outcomes of this paper and are meant to be of expository nature, we will keep the presentation rather informal, focusing only on the crucial steps and omitting the technical details.

For simplicity, we will assume in this section that H is Gaussian. Note that $W\in\mathbb{C}^{M\times N_0}$, so

$$W=U\Gamma V^*\tag{4.2}$$

with unitary matrices $U\in\mathbb{C}^{M\times M}$ and $V\in\mathbb{C}^{N_0\times N_0}$ and

$$\Gamma=\begin{pmatrix}\widetilde{\Gamma}\\ 0\end{pmatrix},\quad \widetilde{\Gamma}=\mathrm{diag}(\gamma_1,\ldots,\gamma_{N_0}),\tag{4.3}$$

where γ_i are the singular values of W and $N_0 \ll M$.

Note that $N_0 = 2N = 2\phi M$, and thus, the eigenvalues of W^*W have Marchenko–Pastur distribution with parameter 2ϕ. For $\phi \to 0$, regime that we are interested in, the eigenvalues of W^*W will be concentrated around point 1, in the neighborhood of size $O(\sqrt{\phi})$, so for simplicity of presentation we will omit the asymptotically small term $O(\sqrt{\phi})$ and assume throughout these computations that all $\gamma_i = 1$, i.e., $\widetilde{\Gamma} = I_{N_0}$.

After applying the singular value decomposition to W and factoring out matrices U and U^* from the inverse, we get

$$S(E) = I - 2\gamma \mathrm{i}\, V\Gamma^t U^*(E\cdot I - H + \mathrm{i}\,\gamma U\Gamma\Gamma^t U^*)^{-1}U\Gamma V^* \tag{4.4}$$

$$= I - 2\gamma \mathrm{i}\, V(\widetilde{\Gamma}^t, 0)\left(E\cdot I - \widetilde{H} + \mathrm{i}\,\gamma \begin{pmatrix} \widetilde{\Gamma}\widetilde{\Gamma}^t & 0 \\ 0 & 0_{(M-N_0)\times(M-N_0)} \end{pmatrix}\right)^{-1} \begin{pmatrix} \widetilde{\Gamma} \\ 0 \end{pmatrix} V^* \tag{4.5}$$

$$= I + 2\gamma \mathrm{i}\, V(\widetilde{\Gamma}^t, 0)\left(\widetilde{H} - E\cdot I - \mathrm{i}\,\gamma \begin{pmatrix} \widetilde{\Gamma}\widetilde{\Gamma}^t & 0 \\ 0 & 0_{(M-N_0)\times(M-N_0)} \end{pmatrix}\right)^{-1} \begin{pmatrix} \widetilde{\Gamma} \\ 0 \end{pmatrix} V^*, \tag{4.6}$$

where $\widetilde{H}$ remains a GUE matrix. Separate the upper-left $N_0 \times N_0$ block of $\widetilde{H}$

$$\widetilde{H} = \begin{pmatrix} \widetilde{H}_1 & \widetilde{H}_2 \\ \widetilde{H}_2^* & \widetilde{H}_3 \end{pmatrix} \tag{4.7}$$

and note that $\frac{1}{2\phi}\widetilde{H}_1$ and $\frac{1}{1-2\phi}\widetilde{H}_3$ are both independent GUE matrices. Now the inverse matrix in (4.6) can be rewritten as

$$\begin{pmatrix} \widetilde{H}_1 - E - \mathrm{i}\,\gamma\widetilde{\Gamma}\widetilde{\Gamma}^t & \widetilde{H}_2 \\ \widetilde{H}_2^* & \widetilde{H}_3 - E \end{pmatrix}^{-1}. \tag{4.8}$$

Using the Schur complement formula we have that the upper-left $N_0 \times N_0$ block of (4.8), the only part that does not vanish after sandwiching (4.8) by $(\widetilde{\Gamma}^t, 0)$ and its transpose, is given by

$$\left(\widetilde{H}_1 - E - \mathrm{i}\,\gamma\widetilde{\Gamma}\widetilde{\Gamma}^t - \widetilde{H}_2\left(\widetilde{H}_3 - E\right)^{-1}\widetilde{H}_2^*\right)^{-1}. \tag{4.9}$$

The semicircular law for the Hermitian (GUE) matrix $\frac{1}{1-2\phi}\widetilde{H}_3$ implies that

$$(\widetilde{H}_3 - E)^{-1} = \frac{1}{1-2\phi}\left(\frac{1}{1-2\phi}\widetilde{H}_3 - \frac{1}{1-2\phi}E\right)^{-1} \approx \frac{1}{1-2\phi} m_{sc}\left(\frac{1}{1-2\phi}E\right) I_{M-N_0}, \tag{4.10}$$

as $M \to \infty$, where $m_{sc}(z)$ denotes the Stieltjes transform of the semicircular distribution and "≈" denotes that the corresponding equality holds asymptotically with a vanishing additive term and with high probability. Note that random matrices $\widetilde{H}_1$, $\widetilde{H}_2$ and $\widetilde{H}_3$ are independent. Therefore, by the concentration for quadratic forms (see, e.g., [15, Theorem C.1]) can be approximated by

$$\widetilde{H}_2\Big(\widetilde{H}_3 - E\Big)^{-1}\widetilde{H}_2^* \approx m_{sc}\Big(\frac{1}{1-2\phi}E\Big)I_{N_0}. \tag{4.11}$$

From (4.9) and (4.11), it remains to check the limiting behavior of

$$\left(\widetilde{H}_1 - E - \mathrm{i}\,\gamma\widetilde{\Gamma}\widetilde{\Gamma}^t - m_{sc}\Big(\frac{1}{1-2\phi}E\Big)\right)^{-1}. \tag{4.12}$$

Recall that since ϕ is small, we assumed that $\widetilde{\Gamma}\widetilde{\Gamma}^t = I_{N_0}$. In this case, using again the semicircular law for the Hermitian matrix $\frac{1}{2\phi}\widetilde{H}_1$, we have

$$\begin{aligned}
&\left(\widetilde{H}_1 - E - \mathrm{i}\,\gamma\widetilde{\Gamma}\widetilde{\Gamma}^t - m_{sc}\Big(\frac{1}{1-2\phi}E\Big)\right)^{-1} \\
&\approx \left(\widetilde{H}_1 - E - \mathrm{i}\,\gamma - m_{sc}\Big(\frac{1}{1-2\phi}E\Big)\right)^{-1} &(4.13)\\
&= \frac{1}{2\phi}\left(\frac{1}{2\phi}\widetilde{H}_1 - \frac{1}{2\phi}\Big(E + \mathrm{i}\,\gamma + m_{sc}\Big(\frac{1}{1-2\phi}E\Big)\Big)\right)^{-1} &(4.14)\\
&\approx \frac{1}{2\phi}m_{sc}\left(\frac{1}{2\phi}\Big(E + \mathrm{i}\,\gamma + m_{sc}\Big(\frac{1}{1-2\phi}E\Big)\Big)\right). &(4.15)
\end{aligned}$$

From the asymptotics $zm_{sc}(z) \to -1, |z| \to \infty$, we get that

$$\left(\widetilde{H}_1 - E - \mathrm{i}\,\gamma\widetilde{\Gamma}\widetilde{\Gamma}^t - m_{sc}\Big(\frac{1}{1-2\phi}E\Big)\right)^{-1} \approx -\frac{1}{E + \mathrm{i}\,\gamma + m_{sc}(E)}, \quad \phi \to 0. \tag{4.16}$$

We conclude that as $N_0, M \to \infty$, $\phi \to 0$ (see (4.6) and (4.16))

$$\mathbb{E}[S(E)] \approx 1 + 2\gamma\mathrm{i}\left(-\frac{1}{E + \mathrm{i}\,\gamma + m_{sc}(E)}\right). \tag{4.17}$$

Now, from $m_{sc}(0) = \mathrm{i}$ we have that

$$\mathbb{E}[S(0)] \approx 1 - \frac{2\gamma\mathrm{i}}{\mathrm{i}\,\gamma + \mathrm{i}}. \tag{4.18}$$

Finally, taking $\gamma = 1$ gives $\mathbb{E}[S(0)] \approx 0$.

Note that from (4.17) we have that in the limit $\phi \to 0$

$$\mathbb{E}[S(0)] = 0 \quad \Leftrightarrow \quad E + m_{sc}(E) = \mathrm{i}\,\gamma, \tag{4.19}$$

and for $E \in \mathbb{R}$, the expression $E + m_{sc}(E)$ is purely imaginary if and only if $E = 0$

$$E + m_{sc}(E) = E + \frac{-E + \sqrt{E^2 - 4}}{2} = \frac{E + \sqrt{E^2 - 4}}{2}. \tag{4.20}$$

Therefore, for $\phi \to 0$, $\mathbb{E}[S(E)] = 0$ if and only if $E = 0$ and $\gamma = 1$.

Acknowledgements

The authors are very grateful to Yan Fyodorov for discussions on the physical background and for providing references, and to the anonymous referee for numerous valuable remarks.

A. Spectral Properties for a General Class of Rational Expressions in Random Matrices

A.1. Noncommutative (NC) Rational Expressions and Their Linearizations

NC rational expressions are formally defined as expressions obtained by applying the four algebraic operations (including taking inverses) to a tuple of NC variables. A systematic overview of the abstract theory of NC rational expression and functions (equivalence classes of rational expressions) can be found in [9]. Note that unlike polynomials, rational expressions do not have a canonical representation, which may lead to a situation when, after evaluating on some algebra, two rational expressions represent the same function. In this paper, we leave aside the question of identification of rational functions, and will work instead directly with rational expressions and their evaluations, specifying each time on which domain the evaluation is taking place. Below we introduce a standard set-up in which we will work and define recursively the classes of rational expressions, denoted by letter $\mathcal{Q}$, together with corresponding domains of evaluation denoted by $\mathcal{D}$.

Let $\mathcal{H}$ be a Hilbert space, $\mathcal{A} \subseteq \mathcal{B}(\mathcal{H})$ be a C^*-algebra (of bounded operators on $\mathcal{H}$) with norm $\|\cdot\|_{\mathcal{A}}$ and let $x_1, \ldots, x_{\alpha_*}, y_1, \ldots, y_{\beta_*}$ be the NC variables taking values in $\mathcal{A}$ with $x_\alpha = x_\alpha^*$ for $1 \leq \alpha \leq \alpha_*$. Denote by $\mathcal{A}_{sa} \subset \mathcal{A}$ the set of self-adjoint elements of $\mathcal{A}$ and let $\mathbb{C}\langle \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{y}^* \rangle$ be the set of polynomials in $\boldsymbol{x} := (x_1, \ldots, x_{\alpha_*})$, $\boldsymbol{y} := (y_1, \ldots, y_{\beta_*})$ and $\boldsymbol{y}^* := (y_1^*, \ldots, y_{\beta_*}^*)$. We define the NC rational expressions recursively on their *height* using the following procedure:

(i) Let $\boldsymbol{q}_0 := \mathbb{1}_{\mathcal{A}}$. The set of rational expressions of height 0 is defined to be the set of polynomials in $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{y}^*$ with the domain of definition $\mathcal{A}_{sa}^{\alpha_*} \times \mathcal{A}^{\beta_*}$

$$(\mathcal{Q}_{\boldsymbol{q}_0}, \mathcal{D}_{\boldsymbol{q}_0}) := (\mathbb{C}\langle \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{y}^* \rangle, \mathcal{A}_{sa}^{\alpha_*} \times \mathcal{A}^{\beta_*}). \tag{A.1}$$

(ii) Let $\boldsymbol{q}_1 := (q_{1,1}, \ldots, q_{1,\ell_1}) \in (\mathcal{Q}_{\boldsymbol{q}_0})^{\ell_1}$, $\boldsymbol{q}_2 := (q_{2,1}, \ldots, q_{2,\ell_2}) \in (\mathcal{Q}_{\boldsymbol{q}_0,\boldsymbol{q}_1})^{\ell_2}$, $\ldots$, $\boldsymbol{q}_n := (q_{n,1}, \ldots, q_{n,\ell n}) \in (\mathcal{Q}_{\boldsymbol{q}_0,\ldots,\boldsymbol{q}_{n-1}})^{\ell_n}$, assuming that $(\mathcal{Q}_{\boldsymbol{q}_0}, \mathcal{D}_{\boldsymbol{q}_0})$,

$\ldots$, $(\mathcal{Q}_{\boldsymbol{q}_0,\ldots,\boldsymbol{q}_{n-1}}, \mathcal{D}_{\boldsymbol{q}_0,\ldots,\boldsymbol{q}_{n-1}})$ are defined. Then, we define

$$\mathcal{Q}_{\boldsymbol{q}_0,\ldots,\boldsymbol{q}_{n-1},\boldsymbol{q}_n} := \mathbb{C}\Big\langle \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{y}^*, \frac{1}{\boldsymbol{q}_1}, \frac{1}{\boldsymbol{q}_1^*}, \ldots, \frac{1}{\boldsymbol{q}_n}, \frac{1}{\boldsymbol{q}_n^*}\Big\rangle, \tag{A.2}$$

$$\mathcal{D}_{\boldsymbol{q}_0,\ldots,\boldsymbol{q}_{n-1},\boldsymbol{q}_n} := \Big\{(\boldsymbol{x},\boldsymbol{y}) \in \mathcal{D}_{\boldsymbol{q}_0,\boldsymbol{q}_1,\ldots,\boldsymbol{q}_{n-1}} : \Big\|\frac{1}{q_{n,j}(\boldsymbol{x},\boldsymbol{y},\boldsymbol{y}^*)}\Big\|_{\mathcal{A}} < \infty \text{ for } 1 \le j \le \ell_n\Big\}, \tag{A.3}$$

where $\frac{1}{\boldsymbol{q}_i} := (q_{i,1}^{-1}, \ldots, q_{i,\ell_i}^{-1})$ and $\frac{1}{\boldsymbol{q}_i^*} := ((q_{i,1}^*)^{-1}, \ldots, (q_{i,\ell_i}^*)^{-1})$.
We say that the rational expression $q \in \mathcal{Q}_{\boldsymbol{q}_0,\ldots,\boldsymbol{q}_{n-1},\boldsymbol{q}_n}$ defined on $\mathcal{D}_{\boldsymbol{q}_0,\ldots,\boldsymbol{q}_{n-1},\boldsymbol{q}_n}$ has height n.

For any $C > 0$ and $\boldsymbol{q}_1 \in (\mathcal{Q}_{\boldsymbol{q}_0})^{\ell_1}$, $\boldsymbol{q}_2 \in (\mathcal{Q}_{\boldsymbol{q}_0,\boldsymbol{q}_1})^{\ell_2}$, $\ldots$, $\boldsymbol{q}_n \in (\mathcal{Q}_{\boldsymbol{q}_0,\ldots,\boldsymbol{q}_{n-1}})^{\ell_n}$, define the *effective* domain

$$\begin{aligned}\mathcal{D}_{\boldsymbol{q}_0,\ldots,\boldsymbol{q}_n;C} := \Big\{(\boldsymbol{x},\boldsymbol{y}) \in \mathcal{D}_{\boldsymbol{q}_0,\ldots,\boldsymbol{q}_{n-1};C} : \Big\|\frac{1}{q_{n,j}(\boldsymbol{x},\boldsymbol{y},\boldsymbol{y}^*)}\Big\|_{\mathcal{A}} \le C \text{ for } 1 \le j \le \ell_n\Big\}\\ \subset \mathcal{D}_{\boldsymbol{q}_0,\ldots,\boldsymbol{q}_n}.\end{aligned} \tag{A.4}$$

This domain will allow an effective control of the norm of rational expressions of height n.

Remark A.1. Similar constructions of rational functions/expressions involving their *height* have been exploited actively in the literature (see, e.g., [9, Chapter 4], [32,39]). Note that here we allow the "denominators" without constant terms, so they cannot be automatically expanded into geometric series for small $\boldsymbol{x}, \boldsymbol{y}$. Hence, we need to introduce and follow explicit domains.

Remark A.2. In the sequel, the statement "q is a rational expression of height n" will implicitly mean that there (uniquely) exist $\boldsymbol{q}_1 \in (\mathcal{Q}_{\boldsymbol{q}_0})^{\ell_1}$, $\boldsymbol{q}_2 \in (\mathcal{Q}_{\boldsymbol{q}_0,\boldsymbol{q}_1})^{\ell_2}$, $\ldots$, $\boldsymbol{q}_n \in (\mathcal{Q}_{\boldsymbol{q}_0,\ldots,\boldsymbol{q}_{n-1}})^{\ell_n}$ and $C > 0$ such that $q \in \mathcal{Q}_{\boldsymbol{q}_0,\ldots,\boldsymbol{q}_n}$ and q is evaluated on the effective domain $\mathcal{D}_{\boldsymbol{q}_0,\ldots,\boldsymbol{q}_n;C}$. Note that many of the basic results, in particular about constructing the linearizations, can be formulated in a completely abstract form or without restriction to effective domains.

Remark A.3. When evaluating a rational expression of height n on different C^*-algebras $\mathcal{A}_1, \ldots, \mathcal{A}_k$, we will use the notation $\mathcal{D}_{\boldsymbol{q}_0,\ldots,\boldsymbol{q}_n}(\mathcal{A}_i)$ and $\mathcal{D}_{\boldsymbol{q}_0,\ldots,\boldsymbol{q}_n;C}$ $(\mathcal{A}_i)$ correspondingly.

Definition A.4 (Self-adjoint rational expression). We say that a rational expression $q = q(\boldsymbol{x},\boldsymbol{y},\boldsymbol{y}^*)$ is self-adjoint if $q(\boldsymbol{x},\boldsymbol{y},\boldsymbol{y}^*) = [q(\boldsymbol{x},\boldsymbol{y},\boldsymbol{y}^*)]^*$ for all $(\boldsymbol{x},\boldsymbol{y}) \in \mathcal{D}$.

A.2. Linearizations and Linearization Algorithm

Definition A.5. Let q be a self-adjoint rational expression of height n in NC variables $\boldsymbol{x}$, $\boldsymbol{y}$ and $\boldsymbol{y}^*$. We say that the self-adjoint matrix

$$\boldsymbol{L} = \left(\begin{array}{c|c} \lambda & \boldsymbol{\ell}^* \\ \hline \boldsymbol{\ell} & \widehat{\boldsymbol{L}} \end{array}\right) \in (\mathbb{C}\langle \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{y}^*\rangle)^{m\times m} \tag{A.5}$$

with $\lambda \in \mathbb{C}\langle \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{y}^* \rangle$, $\boldsymbol{\ell} \in (\mathbb{C}\langle \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{y}^* \rangle)^{(m-1)\times 1}$, $\widehat{\boldsymbol{L}} \in (\mathbb{C}\langle \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{y}^* \rangle)^{(m-1)\times (m-1)}$, whose entries are polynomials of degree at most 1, is a *(self-adjoint) linearization* of q if

(i) the submatrix $\widehat{\boldsymbol{L}} \in (\mathbb{C}\langle \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{y}^* \rangle)^{(m-1)\times (m-1)}$ is invertible, and

(ii) $\lambda - \boldsymbol{\ell}^* \widehat{\boldsymbol{L}}^{-1} \boldsymbol{\ell} = q$

for all $(\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{D}_{\boldsymbol{q}_0,\ldots,\boldsymbol{q}_n}$.

The linearization of q can be written as

$$\boldsymbol{L} = K_0 \otimes \mathbb{1}_{\mathcal{A}} - \sum_{\alpha=1}^{\alpha_*} K_\alpha \otimes x_\alpha - \sum_{\beta=1}^{\beta_*} \big(L_\beta \otimes y_\beta + L_\beta^* \otimes y_\beta^* \big), \tag{A.6}$$

where $K_0, K_\alpha, L_\beta \in \mathbb{C}^{m\times m}$.

The idea of studying noncommutative rational functions/expressions via linearizations goes back to Kleene [26]. Since the publication of this work, various approaches and algorithms have been developed for constructing linearizations of general classes of rational functions/expressions (see, e.g., [9] or [24] for a pedagogical presentation of the subject). For reader's convenience, we provide below a simple linearization algorithm based on the method described in [13, Section A.1]. We use the following observation: for matrices $\boldsymbol{A}_i \in (\mathbb{C}\langle \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{y}^* \rangle)^{m_i \times m_i}$ and $\boldsymbol{B}_j \in (\mathbb{C}\langle \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{y}^* \rangle)^{m_j \times m_{j+1}}$, $m_j \in \mathbb{N}$, the lower right $m_1 \times m_k$ submatrix of the inverse of the block matrix

$$\begin{pmatrix} & & & \boldsymbol{B}_1 & -\boldsymbol{A}_1 \\ & & \boldsymbol{B}_2 & -\boldsymbol{A}_2 & \\ & \iddots & \iddots & & \\ \boldsymbol{B}_{k-1} & -\boldsymbol{A}_{k-1} & & & \\ -\boldsymbol{A}_k & & & & \end{pmatrix} \tag{A.7}$$

is equal to

$$-\boldsymbol{A}_1^{-1} \boldsymbol{B}_1 \boldsymbol{A}_2^{-1} \boldsymbol{B}_2 \cdots \boldsymbol{A}_{k-1}^{-1} \boldsymbol{B}_{k-1} \boldsymbol{A}_k^{-1} \in (\mathbb{C}\langle \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{y}^* \rangle)^{m_1 \times m_k}. \tag{A.8}$$

Now, if $q_1, \ldots, q_\ell$ are rational expressions having known linearizations $\boldsymbol{A}_1, \ldots, \boldsymbol{A}_\ell$, then one can easily check that the linearization of the product $w_0 \frac{1}{q_1} w_1 \frac{1}{q_2} \cdots \frac{1}{q_{k-1}} w_{k-1} \frac{1}{q_k} w_k$ with $w_j \in \{\mathbb{1}_{\mathcal{A}}, x_\alpha, y_\beta, y_\beta^* \,|\, 1 \le \alpha \le \alpha_*, 1 \le \beta \le \beta_*\}$ can be given by

$$\left(\begin{array}{c|ccccc} & & & & & \boldsymbol{b}_0 \\ \hline & & & & \boldsymbol{B}_1 & -\boldsymbol{A}_1 \\ & & & \boldsymbol{B}_2 & -\boldsymbol{A}_2 & \\ & & \iddots & \iddots & & \\ & \boldsymbol{B}_{k-1} & -\boldsymbol{A}_{k-1} & & & \\ \boldsymbol{b}_k & -\boldsymbol{A}_k & & & & \end{array}\right) \tag{A.9}$$

with $\boldsymbol{b}_0 = (w_0, 0, \ldots, 0) \in (\mathbb{C}\langle \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{y}^* \rangle)^{1\times m_1}$, $\boldsymbol{b}_k = (w_k, 0, \ldots, 0)^t \in (\mathbb{C}\langle \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{y}^* \rangle)^{m_k \times 1}$ and $\boldsymbol{B}_j$ being matrices of the form

$$\boldsymbol{B}_j = \begin{pmatrix} w_j & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{pmatrix} \in (\mathbb{C}\langle \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{y}^* \rangle)^{m_j \times m_{j+1}}. \tag{A.10}$$

Using (A.9), we can construct a linearization of rational expressions using the induction on the height. In the case of a rational expression of height 0 (polynomial function), linearization can be constructed using, e.g., the algorithm from [13, Section A.1]. If $\boldsymbol{q}_1 \in (\mathcal{Q}_{\boldsymbol{q}_0})^{\ell_1}$ and $q \in \mathcal{Q}_{\boldsymbol{q}_0, \boldsymbol{q}_1}$ is a rational expression of height 1, the linearization can be obtain using the following algorithm:

(B0) write q as a sum of monomials in $\boldsymbol{x}$, $\boldsymbol{y}$, $\boldsymbol{y}^*$, $\frac{1}{\boldsymbol{q}_1}$, $\frac{1}{\boldsymbol{q}_1^*}$ of the form $w_1 \frac{1}{q_1} w_2 \frac{1}{q_2} \cdots w_{k-1} \frac{1}{q_{k-1}} w_k$ with $w_j \in \{\mathbb{1}_{\mathcal{A}}, x_\alpha, y_\beta, y_\beta^* \,|\, 1 \le \alpha \le \alpha_*, 1 \le \beta \le \beta_*\}$ and $q_i \in \{\mathbb{1}_{\mathcal{A}}, q_{1,\gamma}, q_{1,\gamma}^* \,|\, 1 \le \gamma \le \ell_1\}$ using $\frac{1}{\mathbb{1}_{\mathcal{A}}} = \mathbb{1}_{\mathcal{A}}$ if necessary;

(B1) for each polynomial $q_{1,\gamma_1}, q_{1,\gamma_1}^*, 1 \le \gamma_1 \le \ell_1$ construct a linearization (not necessarily self-adjoint) using the algorithm from [13, Section A.1];

(B2) linearization of a monomial in $\boldsymbol{x}$, $\boldsymbol{y}$, $\boldsymbol{y}^*$, $\frac{1}{\boldsymbol{q}_1}$, $\frac{1}{\boldsymbol{q}_1^*}$ of the form $w_1 \frac{1}{q_1} w_2 \frac{1}{q_2} \cdots w_{k-1} \frac{1}{q_{k-1}} w_k$ is given by (A.9) with A_i being either linearizations of polynomials $q_i \in \{q_{1,\gamma_1}, q_{1,\gamma_1}^* \,|\, 1 \le \gamma_1 \le \ell_1\}$, or $\mathbb{1}_{\mathcal{A}}$ (linearization of $\frac{1}{\mathbb{1}_{\mathcal{A}}} = \mathbb{1}_{\mathcal{A}}$), and $w_j \in \{\mathbb{1}_{\mathcal{A}}, x_\alpha, y_\beta, y_\beta^* \,|\, 1 \le \alpha \le \alpha_*, 1 \le \beta \le \beta_*\}$;

(B3) the (possibly not self-adjoint) linearization of a linear combination of monomials (and thus q) is constructed by putting the linearizations of monomials obtained at (B2) into a block-diagonal form using a procedure similar to (R1)–(R2) from [13, Section A.1];

(B4) if after step (B3) the resulting linearization is not self-adjoint, the symmetrized linearization of $q = (q + q^*)/2$ can be obtained by putting the linearization obtained at step (B3) and its conjugate transpose into a block-skew-diagonal form similarly as in (R3) from [13, Section A.1].

Suppose that we know how to construct linearizations for rational expressions of height $\le n-1$ and suppose that $q \in \mathcal{Q}_{\boldsymbol{q}_0, \ldots, \boldsymbol{q}_n}$ is a rational expression of height n. Then, the linearization of q can be constructed using the following algorithm, that is an adaptation of (B0)–(B4):

(S0) write q as a sum of monomials in $\boldsymbol{x}$, $\boldsymbol{y}$, $\boldsymbol{y}^*$, $\frac{1}{\boldsymbol{q}_1}$, $\frac{1}{\boldsymbol{q}_1^*}, \ldots, \frac{1}{\boldsymbol{q}_n}$, $\frac{1}{\boldsymbol{q}_n^*}$ of the form $w_1 \frac{1}{q_1} w_2 \frac{1}{q_2} \cdots w_{k-1} \frac{1}{q_{k-1}} w_k$ with $w_j \in \{\mathbb{1}_{\mathcal{A}}, x_\alpha, y_\beta, y_\beta^* \,|\, 1 \le \alpha \le \alpha_*, 1 \le \beta \le \beta_*\}$ and $q_i \in \{\mathbb{1}_{\mathcal{A}}, q_{t,\gamma_t}, q_{t,\gamma_t}^* \,|\, 1 \le t \le n, 1 \le \gamma_t \le \ell_n\}$ using $\frac{1}{\mathbb{1}_{\mathcal{A}}} = \mathbb{1}_{\mathcal{A}}$ if necessary;

(S1) construct linearizations (not necessarily self-adjoint) of each rational expression q_{t,γ_t} for $1 \le t \le n$, $1 \le \gamma_t \le \ell_t$ of height $\le n-1$ (rational expressions of height $\le n-1$);

(S2) linearization of a monomial in $\boldsymbol{x}$, $\boldsymbol{y}$, $\boldsymbol{y}^*$, $\frac{1}{\boldsymbol{q}_1}$, $\frac{1}{\boldsymbol{q}_1^*}, \dots, \frac{1}{\boldsymbol{q}_n}, \frac{1}{\boldsymbol{q}_n^*}$ of the form $w_1\frac{1}{q_1}w_2\frac{1}{q_2}\cdots w_{k-1}\frac{1}{q_{k-1}}w_k$ is given by (A.9) with A_i being either linearizations of polynomials $q_i \in \{q_{t,\gamma_t}, q^*_{t,\gamma_t} \,|\, 1 \leq t \leq n, 1 \leq \gamma_t \leq \ell_n\}$, or $\mathbb{1}_{\mathcal{A}}$, and $w_j \in \{\mathbb{1}_{\mathcal{A}}, x_\alpha, y_\beta, y^*_\beta \,|\, 1 \leq \alpha \leq \alpha_*, 1 \leq \beta \leq \beta_*\}$.

The steps (S3)–(S4) are identical to (B3)–(B4).

Remark A.6. It is always possible to obtain the specific form of monomials required in steps (B0) and (S0) by adding factors $\mathbb{1}_{\mathcal{A}}$ or $\frac{1}{\mathbb{1}_{\mathcal{A}}}$, as, for example, in

$$\frac{1}{(\mathbb{1}_{\mathcal{A}} - x)^2}x^2 = \mathbb{1}_{\mathcal{A}}\frac{1}{\mathbb{1}_{\mathcal{A}} - x}\mathbb{1}_{\mathcal{A}}\frac{1}{\mathbb{1}_{\mathcal{A}} - x}x\frac{1}{\mathbb{1}_{\mathcal{A}}}x. \tag{A.11}$$

Note that there is some ambiguity at steps (B0) and (S0), representing the fact that we have freedom to choose the order in which the monomials appear in the sum, as well as freedom to put the product of the constant terms $\mathbb{1}_{\mathcal{A}}$ and $\frac{1}{\mathbb{1}_{\mathcal{A}}}$ between the terms w_j and $\frac{1}{q_i}$. All the results of Sect. A hold for any choice of the representation of q as a sum of monomials; therefore, we do not fix any particular order or other rules to guarantee the uniqueness of the representation in (B0) and (S0).

The condition (ii) in the definition of the linearization is satisfied by construction. The condition (i) follows from the following lemma, which gives a bound on $\|(\widehat{\boldsymbol{L}}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{y}^*))^{-1}\|_{\mathbb{C}^{(m-1)\times(m-1)}\otimes\mathcal{A}}$, where for any $n \in \mathbb{N}$ and $\boldsymbol{R} = (\boldsymbol{R}_{ij})_{i,j=1}^n \in \mathbb{C}^{n\times n} \otimes \mathcal{A}$ we denote

$$\|\boldsymbol{R}\|_{\mathbb{C}^{n\times n}\otimes\mathcal{A}} := \max_{1\leq i,j\leq n} \|\boldsymbol{R}_{ij}\|_{\mathcal{A}}. \tag{A.12}$$

Lemma A.7. (Invertibility of $\widehat{\boldsymbol{L}}$) *Let q be a self-adjoint rational expression of height n and let $\boldsymbol{L} = \boldsymbol{L}_q \in (\mathbb{C}\langle \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{y}^*\rangle)^{m\times m}$ be the linearization of q constructed via the above algorithm. Let $\widehat{\boldsymbol{L}}$ be the submatrix of $\boldsymbol{L}$ defined using the decomposition* (A.5)*. Then, there exist $\widehat{C}_{\boldsymbol{L}_q} > 0$ and $\widehat{n}_{\boldsymbol{L}_q} \in \mathbb{N}$ such that for any $(\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{D}_{\boldsymbol{q}_0,\dots,\boldsymbol{q}_n;C}$*

$$\|(\widehat{\boldsymbol{L}}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{y}^*))^{-1}\|_{\mathbb{C}^{(m-1)\times(m-1)}\otimes\mathcal{A}} \leq \widehat{C}_{\boldsymbol{L}_q}(1 + C + \max_{\alpha}\|x_\alpha\|_{\mathcal{A}} + \max_{\beta}\|y_\beta\|_{\mathcal{A}})^{\widehat{n}_{\boldsymbol{L}_q}}. \tag{A.13}$$

Proof. We prove (A.13) by induction on n. For $n = 0$ (the special case of polynomial functions) (A.13) follows from, for example, [13, (3.16)]. Suppose (A.13) holds for all rational expressions of height $k \leq n - 1$. Consider q of height n with linearization obtained via (S0)–(S4). Steps (S3) and (S4) of the linearization algorithm endow $\widehat{\boldsymbol{L}}$ with block-diagonal (S3) or block-skew-diagonal (S4) structure with blocks being the linearizations of monomials obtained at step (S2). Therefore, in order to obtain the bound (A.13) it is enough to consider only the inverses of the blocks of the form

$$\begin{pmatrix} & & & \boldsymbol{B}_1 & -\boldsymbol{A}_1 \\ & & \boldsymbol{B}_2 & -\boldsymbol{A}_2 & \\ & \iddots & \iddots & & \\ \boldsymbol{B}_{k-1} & -\boldsymbol{A}_{k-1} & & & \\ -\boldsymbol{A}_k & & & & \end{pmatrix} \in (\mathbb{C}\langle \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{y}^* \rangle)^{m' \times m'} \tag{A.14}$$

with $\boldsymbol{A}_i$ being the linearizations of the rational expressions $q_i \in \{\frac{1}{q_{t,\gamma_t}}, \frac{1}{q^*_{t,\gamma_t}} \,|\, 0 \leq t \leq n, 1 \leq \gamma_t \leq \ell_t\}$, and $\boldsymbol{B}_j$ being of the form (A.10). One can easily check that the inverse of (A.14) consists of the blocks of the type $\boldsymbol{A}_i^{-1}\boldsymbol{B}_i\boldsymbol{A}_{i+1}^{-1}\boldsymbol{B}_{i+1}\ldots$. The induction step, together with the Schur complement formula for $\boldsymbol{A}_i^{-1}$ and the condition that for $(\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{D}_{\boldsymbol{q}_0,\ldots,\boldsymbol{q}_n;C}$

$$\Big\| \frac{1}{q_i(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{y}^*)} \Big\|_{\mathcal{A}} \leq C \tag{A.15}$$

implies that for each block of type (A.14) there exist $C' > 0$ and $n' \in \mathbb{N}$ such that

$$\left\| \begin{pmatrix} & & & \boldsymbol{B}_1 & -\boldsymbol{A}_1 \\ & & \boldsymbol{B}_2 & -\boldsymbol{A}_2 & \\ & \iddots & \iddots & & \\ \boldsymbol{B}_{k-1} & -\boldsymbol{A}_{k-1} & & & \\ -\boldsymbol{A}_k & & & & \end{pmatrix}^{-1} \right\|_{\mathbb{C}^{m' \times m'} \otimes \mathcal{A}} \leq C'(1 + C + \max_{\alpha} \|x_\alpha\|_{\mathcal{A}} + \max_{\beta} \|y_\beta\|_{\mathcal{A}})^{n'}, \tag{A.16}$$

where $\boldsymbol{A}_i$'s and $\boldsymbol{B}_j$'s in the left-hand side are evaluated at $(\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{D}$. Taking $\widehat{C}_{\boldsymbol{L}_q}$ and $\widehat{n}_{\boldsymbol{L}_q}$, respectively, the maximum over all C''s and the maximum over all n''s in the bounds (A.16) running through all monomials in the representation of q, leads to (A.13) □

Remark A.8. Suppose that $P \in \mathbb{C}^{m \times m}$ is of the form

$$P = \left(\begin{array}{c|ccc} 1 & 0 & \cdots & 0 \\ \hline 0 & & & \\ \vdots & & Q & \\ 0 & & & \end{array}\right) \tag{A.17}$$

with $Q \in \mathbb{C}^{(m-1)\times(m-1)}$ invertible. It is easy to see that if $\boldsymbol{L} \in (\mathbb{C}\langle \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{y}^* \rangle)^{m \times m}$ is a linearization of a rational expression of height q, then so is $(P \otimes \mathbb{1}_{\mathcal{A}})\boldsymbol{L}(P^{-1} \otimes \mathbb{1}_{\mathcal{A}})$. We will use this freedom to bring linearizations to more convenient form.

A.3. A Priori Bound on Generalized Resolvents

Definition A.9 (Generalized resolvent). Let $\boldsymbol{L} \in (\mathbb{C}\langle \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{y}^* \rangle)^{m \times m}$. We call the matrix-valued function $z \mapsto (\boldsymbol{L} - zJ_m \otimes \mathbb{1}_{\mathcal{A}})^{-1}$ defined for $z \in \mathbb{C}_+$ the *generalized resolvent* of $\boldsymbol{L}$.

Lemma A.10. *Let q be a self-adjoint rational expression of height n and let $\boldsymbol{L} \in (\mathbb{C}\langle \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{y}^* \rangle)^{m \times m}$ be the linearization of q constructed using the algorithm from Sect. A.2. Then, there exist $C_{\boldsymbol{L}_q} > 0$ and $n_{\boldsymbol{L}_q} \in \mathbb{N}$ such that for all $C > 1$, $(\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{D}_{\boldsymbol{q}_0, \ldots, \boldsymbol{q}_n; C}$ and $z \in \mathbb{C}_+$*

$$\begin{aligned} &\| (\boldsymbol{L}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{y}^*) - z J_m \otimes \mathbb{1}_{\mathcal{A}})^{-1} \|_{\mathbb{C}^{m \times m} \otimes \mathcal{A}} \\ &\quad \leq C_{\boldsymbol{L}_q} (1 + C + \max_{\alpha} \|x_\alpha\|_{\mathcal{A}} + \max_{\beta} \|y_\beta\|_{\mathcal{A}})^{n_{\boldsymbol{L}_q}} \left(1 + \frac{1}{\operatorname{Im} z} \right). \end{aligned} \tag{A.18}$$

Proof. Rewrite $\boldsymbol{L} - z J_m \otimes \mathbb{1}_{\mathcal{A}}$ using the block decomposition from (A.5)

$$\boldsymbol{L} - z J_m \otimes \mathbb{1}_{\mathcal{A}} = \left(\begin{array}{c|c} \lambda - z \mathbb{1}_{\mathcal{A}} & \boldsymbol{\ell}^* \\ \hline \boldsymbol{\ell} & \widehat{\boldsymbol{L}} \end{array} \right) \tag{A.19}$$

with $\widehat{\boldsymbol{L}} \in (\mathbb{C}\langle \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{y}^* \rangle)^{(m-1) \times (m-1)}$. By (ii) in the definition of the linearization and the Schur complement formula, we have

$$(\boldsymbol{L} - z J_m \otimes \mathbb{1}_{\mathcal{A}})^{-1} = \left(\begin{array}{c|c} (q - z \mathbb{1}_{\mathcal{A}})^{-1} & -(q - z \mathbb{1}_{\mathcal{A}})^{-1} \boldsymbol{\ell}^* \widehat{\boldsymbol{L}}^{-1} \\ \hline -\widehat{\boldsymbol{L}}^{-1} \boldsymbol{\ell} (q - z \mathbb{1}_{\mathcal{A}})^{-1} & \widehat{\boldsymbol{L}}^{-1} + \widehat{\boldsymbol{L}}^{-1} \boldsymbol{\ell} (q - z \mathbb{1}_{\mathcal{A}})^{-1} \boldsymbol{\ell}^* \widehat{\boldsymbol{L}}^{-1} \end{array} \right). \tag{A.20}$$

Now (A.18) follows from Lemma A.7 and the trivial bound for resolvents of self-adjoint elements

$$\left\| \frac{1}{q - z \mathbb{1}_{\mathcal{A}}} \right\|_{\mathcal{A}} \leq \frac{1}{\operatorname{Im} z} \quad \text{uniformly for } z \in \mathbb{C}_+. \tag{A.21}$$

□

A.4. Dyson Equation for Linearizations of NC Rational Expressions

Let q be a self-adjoint rational expression of height n and let $\boldsymbol{L} \in (\mathbb{C}\langle \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{y}^* \rangle)^{m \times m}$ be its linearization constructed using the algorithm from Sect. A.2. Write $\boldsymbol{L}$ as

$$\boldsymbol{L} = K_0 \otimes \mathbb{1}_{\mathcal{A}} - \sum_{\alpha=1}^{\alpha_*} K_\alpha \otimes x_\alpha - \sum_{\beta=1}^{\beta^*} \left(L_\beta \otimes y_\beta + L_\beta^* \otimes y_\beta^* \right) \tag{A.22}$$

with $K_0, K_\alpha, L_\beta \in \mathbb{C}^{m \times m}$ and K_0, K_α self-adjoint. Define the completely positive map $\Gamma : \mathbb{C}^{m \times m} \to \mathbb{C}^{m \times m}$ by

$$\Gamma\big[R\big] = \sum_{\alpha=1}^{\alpha_*} K_\alpha R K_\alpha + \sum_{\beta=1}^{\beta_*} \left(L_\beta R L_\beta^* + L_\beta^* R L_\beta \right), \quad R \in \mathbb{C}^{m \times m}. \tag{A.23}$$

Definition A.11 (Dyson equation for linearizations). We call the equation

$$-\frac{1}{M} = z J_m - K_0 + \Gamma[M] \tag{A.24}$$

the *Dyson equation for the linearization (DEL)* (of a rational expression).

Lemma A.12 (Solution of DEL*: existence and basic properties*). *Let $\mathcal{H}$ be a Hilbert space and let $\mathcal{S} \subset \mathcal{B}(\mathcal{H})$ be a C^*-algebra containing a freely independent family $\{s_1, \ldots, s_{\alpha_*}, c_1, \ldots, c_{\beta_*}\}$ of α_* semicircular and β_* circular elements in a NC probability space $(\mathcal{S}, \tau_{\mathcal{S}})$. Let $q \in \mathcal{Q}_{\boldsymbol{q}_0,\ldots,\boldsymbol{q}_n}$ and assume that $(\boldsymbol{s}, \boldsymbol{c}) \in \mathcal{D}_{\boldsymbol{q}_0,\ldots,\boldsymbol{q}_n;C}$ for $\boldsymbol{s} = (s_1, \ldots, s_{\alpha_*})$, $\boldsymbol{c} := (c_1, \ldots, c_{\beta_*})$ and some $C > 0$. Define*

$$M_z^{(\mathrm{sc})} := (\mathrm{id}_m \otimes \tau_{\mathcal{S}})\Big(\Big[(K_0 - zJ)\otimes \mathbb{1}_{\mathcal{S}} - \sum_{\alpha=1}^{\alpha_*} K_\alpha \otimes s_\alpha - \sum_{\beta=1}^{\beta_*} \big(L_\beta \otimes c_\beta + L_\beta^* \otimes c_\beta^*\big)\Big]^{-1}\Big). \tag{A.25}$$

Then

(i) *there exists $C_{\boldsymbol{L}_q}^{(\mathrm{sc})} > 0$ such that*

$$\|M_z^{(\mathrm{sc})}\|_{\mathbb{C}^{m\times m}} \le C_{\boldsymbol{L}_q}^{(\mathrm{sc})}\Big(1 + \frac{1}{\mathrm{Im} z}\Big); \tag{A.26}$$

(ii) *$M_z^{(\mathrm{sc})}$ satisfies the DEL (A.24);*

(iii) *$M_z^{(\mathrm{sc})}$ depends analytically on z;*

(iv) *$\mathrm{Im} M_z^{(\mathrm{sc})} \ge 0$;*

(v) *$M_z^{(\mathrm{sc})}$ admits the representation*

$$M_z^{(\mathrm{sc})} = M^\infty + \int_{\mathbb{R}} \frac{V(d\lambda)}{\lambda - z}, \tag{A.27}$$

where $M^\infty \in \mathbb{C}^{m\times m}$ is a self-adjoint matrix, and $V(d\lambda)$ is a positive semidefinite matrix-valued measure on $\mathbb{R}$ with compact support;

(vi) *for almost every $\lambda \in R$ the limit $\lim_{\eta\to 0} \pi^{-1} \mathrm{Im} M_{\lambda + \mathrm{i}\,\eta} = V(\lambda) \in \mathbb{C}^{m\times m}$ exists; if the limit is finite on some interval $I \subset \mathbb{R}$ everywhere, then $V(d\lambda)$ is absolutely continuous on I and $V(d\lambda) = V(\lambda)d\lambda$;*

(vii) $\mathrm{supp}(V_{11}) = \mathrm{supp}(\mathrm{Tr} V)$.

Proof. **Proof of (i)**. It follows from Lemma A.10 and the norm bounds for semicircular and circular operators

$$\|s_\alpha\|_{\mathcal{S}} = 2, \quad \|c_\beta\|_{\mathcal{S}} = 2, \quad 1 \le \alpha \le \alpha_*, \quad 1 \le \beta \le \beta_* \tag{A.28}$$

that

$$\Big\|\Big((K_0 - zJ)\otimes \mathbb{1}_{\mathcal{S}} - \sum_{\alpha=1}^{\alpha_*} K_\alpha - \sum_{\beta=1}^{\beta_*}\big(L_\beta \otimes c_\beta + L_\beta^* \otimes c_\beta^*\big)\otimes s_\alpha\Big)^{-1}\Big\|_{\mathbb{C}^{m\times m}\otimes \mathcal{S}} \le C_{\boldsymbol{L}_q}^{(\mathrm{sc})}\Big(1 + \frac{1}{\mathrm{Im} z}\Big) \tag{A.29}$$

for $C_{\boldsymbol{L}_q}^{(\mathrm{sc})} := C_{\boldsymbol{L}_q}(1 + C + 4)^{n_{\boldsymbol{L}\,q}}$.

Proof of (ii). First note that the real and imaginary parts of free circular elements form a freely independent family of semicirculars. Therefore, by defining for each $1 \le \beta \le \beta_*$

$$s_{\alpha_*+\beta} := \sqrt{2}\mathrm{Re} c_\beta, \quad s_{\alpha_*+\beta_*+\beta} := \sqrt{2}\mathrm{Im} c_\beta, \quad K_{\alpha_*+\beta} := \sqrt{2}\mathrm{Re} L_\beta,$$
$$K_{\alpha_*+\beta_*+\beta} := -\sqrt{2}\mathrm{Im} L_\beta, \tag{A.30}$$

$M_z^{(\mathrm{sc})}$ can be rewritten as

$$M_z^{(\mathrm{sc})} = (\mathrm{id}_m \otimes \tau_{\mathcal{S}})\Big(\Big[(K_0 - zJ) \otimes \mathbb{1}_{\mathcal{S}} - \sum_{\alpha=1}^{\alpha_*+2\beta_*} K_\alpha \otimes s_\alpha\Big]^{-1}\Big) \tag{A.31}$$

and it will be enough to show that $M_z^{(\mathrm{sc})}$ satisfies the *DEL* (A.24) with $\Gamma[\,\cdot\,] = \sum_{\alpha=1}^{\alpha+2\beta_*} K_\alpha\, \cdot\, K_\alpha$. This last fact can be established via the argument similar to the proof of the existence of the solution to the *DEL* for the linearizations of *polynomials* in [13, Lemma 2.6 (iv)]. This proof relies on the results of [23, Lemma 5.4] and [27, Proposition 4.1] establishing the existence of the solution to a particular class of *matrix Dyson equations (MDE)*, as well as regularization technique which allows to extend these results from *MDE* to *DEL*. The trivial bound from [13, Lemma 2.5], which justifies the application of the Schur complement formula and relies on the nilpotency structure of the linearizations, in the setting of the current paper can be replaced by the bound (A.29) coming from the specific choice of the domain of evaluation.

Proof of (iii)–(vii). The analyticity of $M_z^{(\mathrm{sc})}$ follows from (A.29) and the positive semidefiniteness of $\mathrm{Im} M_z^{(\mathrm{sc})}$ is a direct consequence of the representation

$$\mathrm{Im} M_z^{(\mathrm{sc})} = \eta\,(\mathrm{id}_m \otimes \tau_{\mathcal{S}})\Big(\big(\boldsymbol{L}^{(\mathrm{sc})} - \bar{z} J_m \otimes \mathbb{1}_{\mathcal{S}}\big)^{-1}(J_m \otimes \mathbb{1}_{\mathcal{S}})\big(\boldsymbol{L}^{(\mathrm{sc})} - z J_m \otimes \mathbb{1}_{\mathcal{S}}\big)^{-1}\Big) \tag{A.32}$$

with $\boldsymbol{L}^{(\mathrm{sc})} := K_0 \otimes \mathbb{1}_{\mathcal{S}} - \sum_{\alpha=1}^{\alpha_*+2\beta_*} K_\alpha \otimes s_\alpha$. Properties (v)–(vii) follow from the general properties of matrix-valued Herglotz functions, Schur formula (A.20) applied to $(\boldsymbol{L}^{(\mathrm{sc})} - zJ_m \otimes \mathbb{1}_{\mathcal{S}})^{-1}$ and the bound (A.29) using the similar argument as in the proof of [13, Lemma 2.7]. □

A.5. Convergence of Spectrum for the Rational Expressions in Random Matrices and the a Priori Bound for the Generalized Resolvent in Random Matrices

The next two sections are devoted to the study of the eigenvalues of a general class of rational expressions evaluated on random Wigner and *iid* matrices.

Assumption A.13 (Wigner and *iid* matrices). Let $X_1, \ldots, X_{\alpha_*} \in \mathbb{C}^{N\times N}$ and $Y_1, \ldots, Y_{\beta_*} \in \mathbb{C}^{N\times N}$ be two independent families of independent random matrices satisfying the following assumptions

(H1) $X_\alpha = (X_\alpha(i,j))_{i,j=1}^N$, $1 \le \alpha \le \alpha_*$, are Hermitian random matrices having independent (up to symmetry constraints) centered entries of variance $1/N$;

(H2) $Y_\beta = (Y_\beta(i,j))_{i,j=1}^N$, $1 \le \beta \le \beta_*$, are (non-Hermitian) random matrices having independent centered entries of variance $1/N$;

(H3) there exist $\varphi_n > 0$, $n \in \mathbb{N}$, such that

$$\max_{1\le i,j\le N}\Big(\max_{1\le\alpha\le\alpha_*} \mathbb{E}\big[|\sqrt{N} X_\alpha(i,j)|^n\big] + \max_{1\le\beta\le\beta_*} \mathbb{E}\big[|\sqrt{N} Y_\beta(i,j)|^n\big]\Big) \le \varphi_n. \tag{A.33}$$

We call X_α *Wigner* matrices and Y_β *iid* matrices.

Denote $\boldsymbol{X} := (X_1, \ldots, X_{\alpha_*})$, $\boldsymbol{Y} := (Y_1, \ldots, Y_{\beta_*})$, $\boldsymbol{Y}^* := (Y_1^*, \ldots, Y_{\beta_*}^*)$ and let q be a (self-adjoint) rational expression in α_* self-adjoint and β_* non self-adjoint noncommutative variables. In order to prove the local law for $q(\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Y}^*)$, we will need to show that the spectrum of $q(\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Y}^*)$ converges to the spectrum of $q(\boldsymbol{s}, \boldsymbol{c}, \boldsymbol{c}^*)$. To this end, for any $\varepsilon > 0$, $m \in \mathbb{N}$ and operator $\boldsymbol{R} \in \mathbb{C}^{mN \times mN}$, denote by $\mathrm{Spec}_\varepsilon(\boldsymbol{R})$ the ε-pseudospectrum of $\boldsymbol{R}$ defined by

$$\mathrm{Spec}_\varepsilon(\boldsymbol{R}) = \mathrm{Spec}(\boldsymbol{R}) \cup \{z \in \mathbb{C} : \|(\boldsymbol{R} - zI_m \otimes I_N)^{-1}\|_{\mathbb{C}^{mN\times mN}} \geq \varepsilon^{-1}\}. \quad \text{(A.34)}$$

It is easy to check that for any $\boldsymbol{R} \in \mathbb{C}^{mN\times mN} \cong \mathbb{C}^{m\times m} \otimes \mathbb{C}^{N\times N}$

$$\|\boldsymbol{R}\|_{\mathbb{C}^{mN\times mN}} \leq m\, \|\boldsymbol{R}\|_{\mathbb{C}^{m\times m} \otimes \mathbb{C}^{N\times N}}, \quad \text{(A.35)}$$

where $\|\cdot\|_{\mathbb{C}^{mN\times mN}}$ and $\|\cdot\|_{\mathbb{C}^{N\times N}}$ denote the operator norms on $\mathbb{C}^{mN\times mN}$ and $\mathbb{C}^{N\times N}$ correspondingly.

For any (not necessarily self-adjoint) rational expression $r(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{y}^*)$ in NC variables $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{y}^*$, denote by $\boldsymbol{L}_r := \boldsymbol{L}_r(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{y}^*) \in (\mathbb{C}\langle \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{y}^* \rangle)^{m_r \times m_r}$ its (not necessarily self-adjoint) linearization, which can be constructed using, for example, the algorithm from Sect. A.2 omitting steps (B4) and (S4) if r is not self-adjoint. Define the corresponding *Hermitized linearization* by

$$\boldsymbol{L}^{r,z}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{y}^*) := \begin{pmatrix} 0 & \boldsymbol{L}_r(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{y}^*) - zJ_{m_r} \otimes \mathbb{1}_{\mathcal{A}} \\ \big(\boldsymbol{L}_r(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{y}^*)\big)^* - \overline{z}J_{m_r} \otimes \mathbb{1}_{\mathcal{A}} & 0 \end{pmatrix} \quad \text{(A.36)}$$

with $\boldsymbol{L}^{r,z}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{y}^*) \in (\mathbb{C}\langle \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{y}^* \rangle)^{2m_r \times 2m_r}$. For the Hermitized linearization (A.36), we define [similarly as in (A.23)] the *self-energy operator* $\Gamma^{r,z} : \mathbb{C}^{2m_r \times 2m_r} \to \mathbb{C}^{2m_r \times 2m_r}$, given by the completely positive map

$$\Gamma^{r,z}[R] = \sum_{\alpha=1}^{\alpha_*} K_\alpha^{r,z} R K_\alpha^{r,z} + \sum_{\beta=1}^{\beta_*} \left(L_\beta^{r,z} R \left(L_\beta^{r,z}\right)^* + \left(L_\beta^{r,z}\right)^* R L_\beta^{r,z} \right), \quad \text{(A.37)}$$

where $K_\alpha^{r,z}$ and $L_\beta^{r,z}$ are the coefficient matrices of $\boldsymbol{L}^{r,z}$ [see, e.g., (A.6)]. Note that if we evaluate $\boldsymbol{L}^{r,z}$ on the tuple of random matrices $(\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Y}^*)$, then $\boldsymbol{L}^{r,z}(\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Y}^*)$ belongs to the class of *Kronecker* random matrices, which were studied in [3]. Therefore, from [3, Lemma 2.2], we have that the corresponding *Matrix Dyson equation*

$$-\frac{1}{M_\omega^{r,z}} = \omega I_{2m_r} - K_0^{r,z} + \Gamma^{r,z}[M_\omega^{r,z}], \quad z \in \mathbb{C}, \quad \omega \in \mathbb{C}_+ \quad \text{(A.38)}$$

has a unique solution with positive semidefinite imaginary part $\mathrm{Im} M_\omega^{r,z} \geq 0$. Moreover, for each $z \in \mathbb{C}$, the solution matrix $M_\omega^{r,z}$ admits the Stieltjes transform representation

$$M_\omega^{r,z} = \int_{\mathbb{R}} \frac{V^{r,z}(d\lambda)}{\lambda - \omega}, \quad \text{(A.39)}$$

where $\{V^{r,z}\}_{z\in\mathbb{C}}$ is a family of measures taking values in the set of positive definite matrices. In the limit $N \to \infty$ the solution $M_\omega^{r,z} \otimes I_N$ well approximates $(\boldsymbol{L}^{r,z}(\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Y}^*) - \omega I)^{-1}$ in the entrywise maximum norm (see [3, Lemma B.1]).

Remark A.14. The statement of [3, Lemma 2.2] is formulated in the more general setting of *Wigner-type* matrices allowing independent but not necessarily identically distributed entries. This model, in general, leads to a system of N matrix equations. In our case, the matrices in $\boldsymbol{X}$ and $\boldsymbol{Y}$ have *i.i.d.* entries (up to symmetry constraints), which reduces the system of N possibly different matrix equations (see, e.g., [3, Eq. (2.6)]) to N identical matrix equations of the form (A.38).

For any rational expression r with linearization $\boldsymbol{L}_r$ define the set $\mathbb{D}_\varepsilon^{\boldsymbol{L}_r} \subset \mathbb{C}$ by

$$\mathbb{D}_\varepsilon^{\boldsymbol{L}_r} := \Big\{ z \,:\, \operatorname{dist}(0, \operatorname{supp} \rho^{r,z}) \le \varepsilon \Big\}, \tag{A.40}$$

where $\rho^{r,z}(d\lambda) := \frac{1}{2m_r} \mathrm{Tr} V^{r,z}(d\lambda)$ and the family of measures $V^{r,z}(d\lambda)$ were defined in (A.39). The set $\mathbb{D}_\varepsilon^{\boldsymbol{L}_r}$ is called the *self-consistent* ε-pseudospectrum of r related to its linearization $\boldsymbol{L}_r$, and $\rho^{r,z}$ is called the *self-consistent density of states* of $\boldsymbol{L}^{r,z}(\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Y}^*)$.

The next lemma contains the main result of this section.

Lemma A.15 (Convergence of the (pseudo)spectrum). *Suppose that $q \in \mathcal{Q}_{\boldsymbol{q}_0,\ldots,\boldsymbol{q}_n}$ is a (not necessarily self-adjoint) rational expression of height n, and $(\boldsymbol{s}, \boldsymbol{c}) \in \mathcal{D}_{\boldsymbol{q}_0,\ldots,\boldsymbol{q}_n;C}(\mathcal{S})$ with some constant $C > 0$. Then, there exists $\widehat{C}^{\mathrm{w}}_{\boldsymbol{L}_q} > 0$ such that*

$$\Big\| \big(\widehat{\boldsymbol{L}}_q(\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Y}^*) \big)^{-1} \big\|_{\mathbb{C}^{(m-1)\times(m-1)} \otimes \mathbb{C}^{N\times N}} \le \widehat{C}^{\mathrm{w}}_{\boldsymbol{L}_q} \quad \textit{a.w.o.p.}, \tag{A.41}$$

where $\widehat{\boldsymbol{L}}_q$ is defined as in (A.5). There exists also a constant $\widetilde{C} > 0$ depending only on the linearization $\boldsymbol{L}_q$ and the constant C such that

$$(\boldsymbol{X}, \boldsymbol{Y}) \in \mathcal{D}_{\boldsymbol{q}_0,\ldots,\boldsymbol{q}_n;\widetilde{C}}\big(\mathbb{C}^{N\times N}\big) \qquad \textit{a.w.o.p.} \tag{A.42}$$

Moreover, for any $\varepsilon \in (0,1)$ the ε-pseudospectrum of $q(\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Y}^*)$ satisfies

$$\operatorname{Spec}_\varepsilon(q(\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Y}^*)) \subset \mathbb{D}_{2\varepsilon}^{\boldsymbol{L}_q} \quad \textit{a.w.o.p.}. \tag{A.43}$$

Proof. We split the proof of this lemma in two parts. First, we show that the condition (A.41) together with (A.42) implies (A.43) for *any*, not necessarily self-adjoint, rational expression and its linearization. After that we prove that (A.41) and (A.42) are satisfied for q if $(\boldsymbol{s}, \boldsymbol{c}) \in \mathcal{D}_{\boldsymbol{q}_0,\ldots,\boldsymbol{q}_n;C}(\mathcal{S})$ using induction on the height n.

Suppose that we have an arbitrary rational expression r and its linearization $\boldsymbol{L}_r$ of size m_r, and suppose that there exists $\widehat{C}^{\mathrm{w}}_{\boldsymbol{L}_r} > 0$ such that *a.w.o.p.*

$$\|(\widehat{\boldsymbol{L}}_r(\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Y}^*))^{-1}\|_{\mathbb{C}^{(m_r-1)\times(m_r-1)} \otimes \mathbb{C}^{N\times N}} \le \widehat{C}^{\mathrm{w}}_{\boldsymbol{L}_r}, \tag{A.44}$$

where $\widehat{\boldsymbol{L}}_r$ is defined similarly as in (A.5). Then from the definition of the linearization (Definition A.5) and the Schur complement formula (A.20), we can choose $C_3, C_4 > 0$ such that

$$\|r(\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Y}^*)\|_{\mathbb{C}^{N\times N}} \le C_3 \tag{A.45}$$

and the sequence of inequalities

$$\|(r(\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Y}^*) - zI_N)^{-1}\|_{\mathbb{C}^{N\times N}}$$

$$\leq \|(\boldsymbol{L}_r(\boldsymbol{X},\boldsymbol{Y},\boldsymbol{Y}^*) - zJ_{m_r} \otimes I_N)^{-1}\|_{\mathbb{C}^{m_rN\times m_rN}} \tag{A.46}$$

$$\leq m_r\|(\boldsymbol{L}_r(\boldsymbol{X},\boldsymbol{Y},\boldsymbol{Y}^*) - zJ_{m_r} \otimes I_N)^{-1}\|_{\mathbb{C}^{m_r\times m_r}\otimes\mathbb{C}^{N\times N}} \tag{A.47}$$

$$\leq C_4\|(r(\boldsymbol{X},\boldsymbol{Y},\boldsymbol{Y}^*) - zI_N)^{-1}\|_{\mathbb{C}^{N\times N}} \tag{A.48}$$

hold *a.w.o.p.* for all $z \in \mathbb{C}$. Here the first and third inequalities follow from the Schur complement formula (A.20) and the norm bounds $\max_{1\leq\alpha\leq\alpha_*}\|X_\alpha\|_{\mathbb{C}^{N\times N}} \leq 3$, $\max_{1\leq\beta\leq\beta_*}\|Y_\beta\|_{\mathbb{C}^{N\times N}} \leq 3$ holding *a.w.o.p.*, the second inequality holds deterministically for all realizations of $\boldsymbol{X}$ and $\boldsymbol{Y}$ [see (A.35)], $C_3 > 0$ depends on $\widehat{C}_r$, m_r and the norms of X_α, Y_β, which we bound by 3, and $C_4 > 1$ additionally depends on C_3.

Note again from (A.20) that if $\widehat{\boldsymbol{L}}_r(\boldsymbol{X},\boldsymbol{Y},\boldsymbol{Y}^*)$ is invertible, then

$$\mathrm{Spec}(r(\boldsymbol{X},\boldsymbol{Y},\boldsymbol{Y}^*)) = \{z\in\mathbb{C} : \boldsymbol{L}_r(\boldsymbol{X},\boldsymbol{Y},\boldsymbol{Y}^*) - zJ_{m_r}\otimes I_N \text{ is not invertible}\}. \tag{A.49}$$

On the other hand, using the definition of $\boldsymbol{L}^{r,z}$ from (A.36), the set on the right-hand side of (A.49) can be described via the spectrum of $\boldsymbol{L}^{r,z}(\boldsymbol{X},\boldsymbol{Y},\boldsymbol{Y}^*)$ as

$$\begin{aligned}&\{z\in\mathbb{C} : \boldsymbol{L}_r(\boldsymbol{X},\boldsymbol{Y},\boldsymbol{Y}^*) - zJ_{m_r}\otimes I_N \text{ is not invertible}\}\\ &\quad = \{z\in\mathbb{C} : 0\in\mathrm{Spec}(\boldsymbol{L}^{r,z}(\boldsymbol{X},\boldsymbol{Y},\boldsymbol{Y}^*))\}\end{aligned} \tag{A.50}$$

with the identity (A.50) holding deterministically for all realizations of $\boldsymbol{X}$ and $\boldsymbol{Y}$.

Under the condition (A.44), the equality (A.49) can be rewritten in terms of the pseudospectrum using (A.46) as

$$\mathrm{Spec}_\varepsilon(r(\boldsymbol{X},\boldsymbol{Y},\boldsymbol{Y}^*)) \subset \Big\{z\in\mathbb{C} : \|(\boldsymbol{L}_r(\boldsymbol{X},\boldsymbol{Y},\boldsymbol{Y}^*) - zJ_{m_r}\otimes I_N)^{-1}\|_{\mathbb{C}^{m_rN\times m_rN}} \geq \frac{1}{\varepsilon}\Big\} \tag{A.51}$$

$$\subset \mathrm{Spec}_{C_4\varepsilon}(r(\boldsymbol{X},\boldsymbol{Y},\boldsymbol{Y}^*)) \tag{A.52}$$

holding *a.w.o.p.* At the same time, from the definition (A.36) we have that the set of the singular values of the Hermitian matrix $\boldsymbol{L}^{r,z}(\boldsymbol{X},\boldsymbol{Y},\boldsymbol{Y}^*)$ coincides with the set of the singular values of $\boldsymbol{L}_r(\boldsymbol{X},\boldsymbol{Y},\boldsymbol{Y}^*) - zJ_{m_r}\otimes I_N$, so that

$$\begin{aligned}&\Big\{z\in\mathbb{C} : \|(\boldsymbol{L}_r(\boldsymbol{X},\boldsymbol{Y},\boldsymbol{Y}^*) - zJ_{m_r}\otimes I_N)^{-1}\|_{\mathbb{C}^{m_rN\times m_rN}} \geq \frac{1}{\varepsilon}\Big\}\\ &\quad = \{z\in\mathbb{C} : \mathrm{dist}(0,\mathrm{Spec}(\boldsymbol{L}^{r,z}(\boldsymbol{X},\boldsymbol{Y},\boldsymbol{Y}^*)) \leq \varepsilon\}.\end{aligned} \tag{A.53}$$

In order to study the spectrum of $\boldsymbol{L}^{r,z}(\boldsymbol{X},\boldsymbol{Y},\boldsymbol{Y}^*)$, we will exploit the fact that $\boldsymbol{L}_r(\boldsymbol{X},\boldsymbol{Y},\boldsymbol{Y}^*) - zJ_{m_r}\otimes I_N$ belongs to the class of *Kronecker* random matrices and thus we can use the results from [3] about the location of spectrum for this class of random matrix ensembles. By applying part (i) of [3, Theorem 4.7] to $\boldsymbol{L}^{r,z}$, we have (similarly as in the proof of [3, Lemma 6.1] for bounded ζ) that for any z satisfying $\mathrm{dist}\,(0,\mathrm{supp}\,\rho^{r,z}) \geq 2\varepsilon$, *a.w.o.p.*

$$\mathrm{Spec}(\boldsymbol{L}^{r,z}(\boldsymbol{X},\boldsymbol{Y},\boldsymbol{Y}^*)) \cap \Big[-\frac{3\varepsilon}{2},\frac{3\varepsilon}{2}\Big] = \emptyset. \tag{A.54}$$

Now we claim that (A.54) holds simultaneously *a.w.o.p.* for all $\{z : |z| \le 2C_3,\ z \notin \mathbb{D}^{\boldsymbol{L}_r}_{2\varepsilon}\}$. To prove this strengthening, we apply the standard grid argument (again analogously as in the proof of [3, Lemma 6.1]) together with the Lipschitz continuity of the eigenvalues of $\boldsymbol{L}^{r,z}(\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Y}^*)$ in z. Together with (A.51) and (A.53) this implies that *a.w.o.p.*

$$\begin{aligned}&\mathrm{Spec}_\varepsilon(r(\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Y}^*))\\&\quad\subset \{z \in \mathbb{C} : \mathrm{dist}(0, \mathrm{Spec}(\boldsymbol{L}^{r,z}(\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Y}^*)) \le \varepsilon\} \subset \Big(\mathbb{D}^{\boldsymbol{L}_r}_{2\varepsilon} \cup \{z : |z| \ge 2C_3\}\Big).\end{aligned} \tag{A.55}$$

We conclude from (A.45) that *a.w.o.p.*

$$\mathrm{Spec}_\varepsilon(r(\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Y}^*)) \subset \mathbb{D}^{\boldsymbol{L}_r}_{2\varepsilon}. \tag{A.56}$$

This finishes the first part of the proof by establishing that for any rational expression the conditions (A.41) and (A.42) imply (A.43).

In order to prove (A.41) and (A.42), we proceed with a proof by induction on the height of q. If q has height 0 (i.e., if q is a polynomial), then (A.42) is trivially true and (A.41) holds by nilpotency [13, Lemma 2.5] and the norm bounds

$$\|X_\alpha\|_{\mathbb{C}^{N\times N}} \le 3, \quad \|Y_\beta\|_{\mathbb{C}^{N\times N}} \le 3 \quad \text{a.w.o.p.} \tag{A.57}$$

Suppose that the statement of the theorem is true for all rational expressions of height $\le n-1$. Together with $(\boldsymbol{s}, \boldsymbol{c}) \in \mathcal{D}_{\boldsymbol{q}_0,\ldots,\boldsymbol{q}_n;C}(\mathcal{S})$ this, in particular, means that *a.w.o.p.*

$$(\boldsymbol{X}, \boldsymbol{Y}) \in \mathcal{D}_{\boldsymbol{q}_0,\ldots,\boldsymbol{q}_{n-1};\widetilde{C}_1}\big(\mathbb{C}^{N\times N}\big) \tag{A.58}$$

for some $\widetilde{C}_1 > 0$, and

$$\|(\widehat{\boldsymbol{L}}_r(\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Y}^*))^{-1}\|_{\mathbb{C}^{(m_r-1)\times(m_r-1)}\otimes\mathbb{C}^{N\times N}} \le \widehat{C}^{\mathrm{w}}_{\boldsymbol{L}_r} \tag{A.59}$$

for all $r \in \mathcal{T}_q := \{q_{i,\gamma_i} : 0 \le i \le n, 1 \le \gamma_i \le \ell_i\}$ with q_{i,γ_i} as in the definition of $\mathcal{Q}_{\boldsymbol{q}_0,\ldots,\boldsymbol{q}_n}$. We now show that there exists $\widetilde{C} > 0$ such that $(\boldsymbol{X}, \boldsymbol{Y}) \in \mathcal{D}_{\boldsymbol{q}_0,\ldots,\boldsymbol{q}_n;\widetilde{C}}\big(\mathbb{C}^{N\times N}\big)$ *a.w.o.p.*, or equivalently, that for all $\gamma_n \in \{1, \ldots, \ell_n\}$ *a.w.o.p.*

$$\Big\|\frac{1}{q_{n,\gamma_n}(\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Y}^*)}\Big\|_{\mathbb{C}^{N\times N}} \le \widetilde{C}. \tag{A.60}$$

Using the result established in the first part of the proof, (A.58) and (A.59) with $r = q_{n,\gamma_n}$ imply that *a.w.o.p.*

$$\mathrm{Spec}_\varepsilon(q_{n,\gamma_n}(\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Y}^*)) \subset \mathbb{D}^{\boldsymbol{L}_{q_{n,\gamma_n}}}_{2\varepsilon}. \tag{A.61}$$

Therefore, it is enough to show that there exists $\widetilde{C} > 0$ such that for all $\gamma_n \in \{1, \ldots, \ell_n\}$ the point $z = 0$ does not belong to $\mathbb{D}^{\boldsymbol{L}_{q_{n,\gamma_n}}}_{2/\widetilde{C}}$. By the definition (A.40), the last condition is equivalent to

$$\mathrm{dist}(0, \mathrm{supp}\,\rho^{q_{n,\gamma_n},0}) \ge \frac{2}{\widetilde{C}}. \tag{A.62}$$

We now show that (A.62) holds for one fixed γ_n. The desired bound (A.60) can then be obtained by taking the maximum over $\widetilde{C}$ for all $\gamma_n \in \{1, \ldots, \ell_n\}$.

Fix $\gamma_n \in \{1, \ldots, \ell_n\}$, and denote by $\boldsymbol{L}_{q_{n,\gamma_n}}(\boldsymbol{s}, \boldsymbol{c}, \boldsymbol{c}^*) \in \mathbb{C}^{m\times m} \otimes \mathcal{S}$ the linearization of q_{n,γ_n} evaluated at $(\boldsymbol{s}, \boldsymbol{c}, \boldsymbol{c}^*)$. By examining carefully the derivation of the a priori bound (A.18), we observe that there exists a constant $t > 0$ depending only on the linearization and the constant C, such that

$$\left\| \Big(\boldsymbol{L}_{q_{n,\gamma_n}}(\boldsymbol{s}, \boldsymbol{c}, \boldsymbol{c}^*) \Big)^{-1} \right\|_{\mathbb{C}^{m\times m} \otimes \mathcal{S}} \leq t. \tag{A.63}$$

Indeed, each entry of $(\boldsymbol{L}_{q_{n,\gamma_n}}(\boldsymbol{s}, \boldsymbol{c}, \boldsymbol{c}^*))^{-1}$ is a polynomial in $\boldsymbol{s}$, $\boldsymbol{c}$, $\boldsymbol{c}^*$ and $(r(\boldsymbol{s}, \boldsymbol{c}, \boldsymbol{c}^*))^{-1}$ with $r \in \mathcal{T}_q$. Therefore, the bounds $\|(r(\boldsymbol{s}, \boldsymbol{c}, \boldsymbol{c}^*))^{-1}\|_{\mathcal{S}} \leq C$ that follow from the assumption $(\boldsymbol{s}, \boldsymbol{c}) \in \mathcal{D}_{\boldsymbol{q}_0,\ldots,\boldsymbol{q}_n;C}$ imply that (A.63) holds for some $t > 0$.

Next notice, that from the definition of $\boldsymbol{L}^{r,z}$ [see (A.36)], we have that

$$\left\| \Big(\boldsymbol{L}^{q_{n,\gamma_n},0}(\boldsymbol{s}, \boldsymbol{c}, \boldsymbol{c}^*) \Big)^{-1} \right\|_{\mathbb{C}^{2m\times 2m} \otimes \mathcal{S}} = \left\| \Big(\boldsymbol{L}_{q_{n,\gamma_n}}(\boldsymbol{s}, \boldsymbol{c}, \boldsymbol{c}^*) \Big)^{-1} \right\|_{\mathbb{C}^{m\times m} \otimes \mathcal{S}}. \tag{A.64}$$

The resolvent identity

$$\begin{aligned} &\Big(\boldsymbol{L}^{q_{n,\gamma_n},z}(\boldsymbol{s}, \boldsymbol{c}, \boldsymbol{c}^*) - \omega I \otimes \mathbb{1}_{\mathcal{S}} \Big)^{-1} \left(I \otimes \mathbb{1}_{\mathcal{S}} - \omega \Big(\boldsymbol{L}^{q_{n,\gamma_n},z}(\boldsymbol{s}, \boldsymbol{c}, \boldsymbol{c}^*) \Big)^{-1} \right) \\ &\quad = \Big(\boldsymbol{L}^{q_{n,\gamma_n},z}(\boldsymbol{s}, \boldsymbol{c}, \boldsymbol{c}^*) \Big)^{-1} \end{aligned} \tag{A.65}$$

together with (A.63) and (A.64) implies that

$$\left\| \Big(\boldsymbol{L}^{q_{n,\gamma_n},0}(\boldsymbol{s}, \boldsymbol{c}, \boldsymbol{c}^*) - \omega I \otimes \mathbb{1}_{\mathcal{S}} \Big)^{-1} \right\|_{\mathbb{C}^{2m\times 2m} \otimes \mathcal{S}} \leq 2t \tag{A.66}$$

for all $\omega \in \mathbb{C}$ satisfying $|\omega| \leq 1/(4mt)$, where we used the relation between the operator and max norms as in (A.47) to estimate the operator norm of $(\boldsymbol{L}^{q_{n,\gamma_n},z}(\boldsymbol{s}, \boldsymbol{c}, \boldsymbol{c}^*))^{-1}$. The bound (A.66) means, in particular, that

$$\operatorname{Spec}\Big(\boldsymbol{L}^{q_{n,\gamma_n},0}(\boldsymbol{s}, \boldsymbol{c}, \boldsymbol{c}^*) \Big) \cap \left[-\frac{1}{4mt}, \frac{1}{4mt} \right] = \emptyset. \tag{A.67}$$

The spectrum of the self-adjoint operator $\boldsymbol{L}^{q_{n,\gamma_n},0}(\boldsymbol{s}, \boldsymbol{c}, \boldsymbol{c}^*)$ is characterized by the solution to the self-consistent Eq. A.38 (see, e.g., [36, Theorem 4.1.12]) via

$$\operatorname{Spec}\Big(\boldsymbol{L}^{q_{n,\gamma_n},0}(\boldsymbol{s}, \boldsymbol{c}, \boldsymbol{c}^*) \Big) = \operatorname{supp} \rho^{q_{n,\gamma_n},0}, \tag{A.68}$$

which establishes (A.62) with $\widetilde{C} = 8mt$.

Finally, (A.42) together with the Schur complement formula, (A.15)–(A.16), and the special form of the linearization blocks (A.9)–(A.10), yields (A.41) for $q \in \mathcal{Q}_{\boldsymbol{q}_0,\ldots,\boldsymbol{q}_n}$. This finishes the proof of the lemma. □

The bound (A.41) together with the Schur complement formula (A.20) applied to $(\boldsymbol{L}_q(\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Y}^*) - zJ_m \otimes I_N)^{-1}$, the trivial bound (A.21) applied to $q(\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Y}^*)$ and the norm bounds $\|X_\alpha\|_{\mathbb{C}^{N\times N}} \leq 3$ and $\|Y_\beta\|_{\mathbb{C}^{N\times N}} \leq 3$ holding [similarly as in (2.3)–(2.4)] $a.w.o.p.$, imply the trivial bound for the generalized resolvent in random matrices.

Corollary A.16. *There exists* $C^{\mathrm{w}}_{\boldsymbol{L}_q} > 0$ *depending only on the linearization* $\boldsymbol{L}_q$*, such that* a.w.o.p.

$$\|(\boldsymbol{L}_q(\boldsymbol{X},\boldsymbol{Y},\boldsymbol{Y}^*) - zJ_m \otimes I_N)^{-1}\|_{\mathbb{C}^{m\times m}\otimes\mathbb{C}^{N\times N}} \leq C^{\mathrm{w}}_{\boldsymbol{L}_q}\Big(1+\frac{1}{\mathrm{Im} z}\Big). \tag{A.69}$$

A.6. Global and Local Laws for Rational Expressions in Random Matrices

Denote, as before, by $\mathcal{S}$ a C^*-algebra containing a freely independent family $\{s_1,\ldots,s_{\alpha_*},c_1,\ldots,c_{\beta_*}\}$ of α_* semicircular and β_* circular elements in a NC probability space $(\mathcal{S},\tau_{\mathcal{S}})$. Let $q \in \mathcal{Q}_{\boldsymbol{q}_0,\ldots,\boldsymbol{q}_n}$ be a rational expression of height n and assume that $(\boldsymbol{s},\boldsymbol{c}) \in \mathcal{D}_{\boldsymbol{q}_0,\ldots,\boldsymbol{q}_n;C}(\mathcal{S})$ for $\boldsymbol{s} = (s_1,\ldots,s_{\alpha_*})$, $\boldsymbol{c} := (c_1,\ldots,c_{\beta_*})$ and some $C>0$. Let

$$\boldsymbol{L} = \boldsymbol{L}(\boldsymbol{x},\boldsymbol{y},\boldsymbol{y}^*) := K_0\otimes\mathbb{1}_{\mathcal{A}} - \sum_{\alpha=1}^{\alpha_*} K_\alpha\otimes x_\alpha - \sum_{\beta=1}^{\beta_*}\big(L_\beta\otimes y_\beta + L_\beta^*\otimes y_\beta^*\big) \tag{A.70}$$

be the linearization of q constructed via the algorithm from Sect. A.2.

In order to formulate the local law, we need to introduce the notion of the *stochastic domination.*

Definition A.17 (*Stochastic domination*). Let $\Phi := (\Phi_N)_{N\in\mathbb{N}}$ and $\Psi := (\Psi_N)_{N\in\mathbb{N}}$ be two sequences of nonnegative random variables. We say that Φ is *stochastically dominated* by Ψ (denoted $\Phi \prec \Psi$), if for any $\varepsilon, D>0$ there exists $C(\varepsilon,D)>0$ such that for all $N\in\mathbb{N}$

$$\mathbb{P}[\Phi_N \geq N^\varepsilon \Psi_N] \leq \frac{C(\varepsilon,D)}{N^D}. \tag{A.71}$$

Let $q(\boldsymbol{X},\boldsymbol{Y},\boldsymbol{Y}^*) \in \mathbb{C}^{N\times N}$ be the evaluation of q on the α_*-tuple of Wigner and β_*-tuple of *iid* random matrices satisfying (**H1**)–(**H3**), and define the linearization matrix

$$\boldsymbol{H} := \boldsymbol{L}(\boldsymbol{X},\boldsymbol{Y},\boldsymbol{Y}^*) = K_0\otimes I_N - \sum_{\alpha=1}^{\alpha_*} K_\alpha\otimes X_\alpha - \sum_{\beta=1}^{\beta_*}\big(L_\beta\otimes Y_\beta + L_\beta^*\otimes Y_\beta^*\big). \tag{A.72}$$

Let $\boldsymbol{G}_z := (\boldsymbol{H} - zJ\otimes I_N)^{-1} \in \mathbb{C}^{mN\times mN}$ be the *generalized* resolvent of $\boldsymbol{H}$. Note that the generalized resolvent $\boldsymbol{G}_z$, when viewed as taking values in $\mathbb{C}^{m\times m}\otimes\mathbb{C}^{N\times N}$, can be written as

$$\boldsymbol{G}_z = \sum_{i,j=1}^N G_{z,ij}\otimes E_{ij}, \tag{A.73}$$

where the collection of matrices $E_{ij} := (\delta_{ki}\delta_{jl})_{1\leq k,l\leq N}$ form a standard basis of $\mathbb{C}^{N\times N}$ and $G_{z,ij}\in\mathbb{C}^{m\times m}$ is an $m\times m$ matrix for each (i,j) pair. In general, we will follow the convention that for any $\boldsymbol{A}\in\mathbb{C}^{m\times m}\otimes\mathbb{C}^{N\times N}$, we denote by $\boldsymbol{A}_{kl}\in\mathbb{C}^{N\times N}$, $k,l = 1,2,\ldots,m$ the (k,l)-th block according to the $\mathbb{C}^{m\times m}$ factor in the tensor product, while $A_{ij}\in\mathbb{C}^{m\times m}$, $i,j = 1,2,\ldots,N$ denotes the (i,j)-th block in the second factor, i.e.,

$$\boldsymbol{A} = \sum_{k,l=1}^m E_{kl}\otimes\boldsymbol{A}_{kl} = \sum_{i,j=1}^N A_{ij}\otimes E_{ij}, \tag{A.74}$$

in particular $A_{ij}(k,l) = \boldsymbol{A}_{kl}(i,j)$.

Let $M_z^{(\mathrm{sc})} : \mathbb{C}_+ \to \mathbb{C}^{m\times m}$ be a matrix valued function given by (A.25). For each $z \in \mathbb{C}_+$ we define the *stability operator* $\mathscr{L}_z : \mathbb{C}^{m\times m} \to \mathbb{C}^{m\times m}$ corresponding to $M_z^{(\mathrm{sc})}$ by

$$\mathscr{L}_z\big[R\big] = R - M_z^{(\mathrm{sc})}\Gamma\big[R\big]M_z^{(\mathrm{sc})}, \quad R \in \mathbb{C}^{m\times m}. \tag{A.75}$$

For $n \in \mathbb{N}$ and an operator $\mathscr{R} : \mathbb{C}^{n\times n} \to \mathbb{C}^{n\times n}$, we denote by $\|\mathscr{R}\|_{\mathbb{C}^{n\times n}\to\mathbb{C}^{n\times n}}$ the operator norm of $\mathscr{R}$ generated by the the operator norm on $\mathbb{C}^{n\times n}$. Then, the following holds.

Theorem A.18 (Local law for rational expressions). *Let $M_z^{(\mathrm{sc})}$ be defined as in* (A.25) *and let $\mathscr{L}_z$ be the stability operator corresponding to $M_z^{(\mathrm{sc})}$. If there exist $C_0 > 0$ and $\mathcal{I} \subset \mathbb{R}$ such that for all z with* $\mathrm{Re}z \in \mathcal{I}$ *and* $0 \le \mathrm{Im}z < \infty$

*(**M1**) $\|M_z^{(\mathrm{sc})}\|_{\mathbb{C}^{m\times m}} \le C_0$;*
*(**M2**) $\|\mathscr{L}_z^{-1}\|_{\mathbb{C}^{m\times m}\to\mathbb{C}^{m\times m}} \le C_0$,*

then the optimal local law holds for $\boldsymbol{G}(z)$ on the set $\mathcal{I}$, i.e., uniformly for $\mathrm{Re}z \in \mathcal{I}$

$$\max_{1\le i,j\le N}\Big\|G_{z,ij} - M_z^{(\mathrm{sc})}\delta_{ij}\Big\|_{\mathbb{C}^{m\times m}} \prec \sqrt{\frac{1}{N\mathrm{Im}z}},$$
$$\Big\|\frac{1}{N}\sum_{i=1}^{N} G_{z,ii} - M_z^{(\mathrm{sc})}\Big\|_{\mathbb{C}^{m\times m}} \prec \frac{1}{N\mathrm{Im}z}. \tag{A.76}$$

In particular, this implies that on the set $\mathcal{I}$ the optimal local law holds for the rational expression in random matrices $q(\boldsymbol{X},\boldsymbol{Y},\boldsymbol{Y}^*)$, i.e., uniformly for $\mathrm{Re}z \in \mathcal{I}$

$$\max_{1\le i,j\le N}\Big|g_{z,ij} - M_z^{(\mathrm{sc})}(1,1)\delta_{ij}\Big| \prec \sqrt{\frac{1}{N\mathrm{Im}z}},$$
$$\Big|\frac{1}{N}\sum_{i=1}^{N} g_{z,ii} - M_z^{(\mathrm{sc})}(1,1)\Big| \prec \frac{1}{N\mathrm{Im}z}, \tag{A.77}$$

where $\boldsymbol{g}_z = (g_{z,ij})_{i,j=1}^N := (q(\boldsymbol{X},\boldsymbol{Y},\boldsymbol{Y}^*) - zI_N)^{-1} \in \mathbb{C}^{N\times N}$.

Note that by the definition of the generalized resolvent and (A.20) and the notational convention (A.74), we have that $g_{z,ij} = G_{z,ij}(1,1)$ for all $1 \le i,j \le N$.

Proof. Our proof of the local law for linearizations of rational expressions (A.76) is analogous to the proof of the corresponding result for linearizations of *polynomials* in Wigner and *iid* matrices [13, Theorem 5.1]. Below we provide a summary of the important steps of that proof and show how these steps are adjusted to the current setting of rational expressions.

1. Restricting analysis to the set where $q(\boldsymbol{X},\boldsymbol{Y},\boldsymbol{Y}^*)$ is well defined. In contrast to the case when q is a polynomial, the evaluation $q(\boldsymbol{X},\boldsymbol{Y},\boldsymbol{Y}^*)$ may not always be well defined and the generalized resolvent $\boldsymbol{G}_z$ may not always be bounded, even when $z \in \mathbb{C}_+$, i.e., the a priori bound analogous to (2.5) in [13, Lemma 2.5] may not hold. But according to Lemma A.15 and Corollary A.16, this bound can be replaced by (A.69) and the existence of $q(\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Y}^*)$ can be guaranteed on an event $\Theta = \Theta_N$ of asymptotically overwhelming probability. The entire analysis is then restricted to this set. In particular, the indicator sets $\chi(\cdot)$ in the proof of [13, Theorem 5.1] should be replaced by $\chi(\cdot) \cap \Theta$.

2. Exploiting the Kronecker structure of $\boldsymbol{H}$ *and regularization of the DEL.* Similarly as in the polynomial case, any linearization of a rational expression evaluated on Wigner and *iid* ensembles belongs to the class of *Kronecker* random matrices. Therefore, in order to obtain the initial estimates on the error term, we use the results of [3, Lemma 4.4]. As in the polynomial setup, these results require introducing a small regularization $\omega = iu$ with $u > 0$ in order to use the stability theory of the MDE (A.38). The bounds in [13, Lemma 5.2] are uniform in this regularization and are a consequence of the a priori estimate [13, Lemma 2.5]. In our current setting, they remain true when this a priori estimate is replaced by (A.18).

3. Effective replacement of the Ward identity. Using again the a priori bound (A.18) instead of [13, Lemma 2.5], one can obtain the Ward identity type estimates (see, e.g., [13, formula (5.13)]) for the error terms involving the generalized resolvent.

4. Finishing the proof. With the above modifications, the proof of Theorem A.18 can be obtained by following the proof of [13, Theorem 5.1] line by line. □

In the same spirit, we can follow line by line the proof of [13, Proposition 2.17], replacing the use of [13, Lemma 2.5] with the trivial bound (A.18), to obtain the following *global law* for (the linearizations of) rational expressions. Note that in the proof of Proposition A.19, we do not have to assume that the conditions (**M1**)–(**M2**) hold as stated in Theorem A.18. For the global law the boundedness of $M_z^{(sc)}$ and $\mathscr{L}_z^{-1}$ is needed only for z away from the real axis, in which case it follows from (A.18) and the representation [13, equation (5.22)] of $\mathscr{L}_z^{-1}$.

Proposition A.19 (Global law). *For any $\theta > 0$, uniformly on $\{ z : \mathrm{Im} z \geq \theta^{-1}, |z| \leq \theta \}$*

$$\max_{1 \leq i,j \leq N} \left\| G_{z,ij} - M_z^{(\mathrm{sc})} \delta_{ij} \right\|_{\mathbb{C}^{m\times m}} \prec_\theta \frac{1}{\sqrt{N}},$$
$$\left\| \frac{1}{N} \sum_{i=1}^{N} G_{z,ii} - M_z^{(\mathrm{sc})} \right\|_{\mathbb{C}^{m\times m}} \prec_\theta \frac{1}{N}, \tag{A.78}$$

where $\prec_\theta$ indicates that the constant in the definition of the stochastic domination (see Definition A.17) may depend on θ. In particular, uniformly on $\{ z : \mathrm{Im} z \geq \theta^{-1}, |z| \leq \theta \}$

$$\left| \frac{1}{N} \mathrm{Tr} \boldsymbol{g}_z - M_z^{(\mathrm{sc})}(1,1) \right| \prec_\theta \frac{1}{N}. \tag{A.79}$$

The global law for the trace of the resolvent of a rational expression, $\mathrm{Tr}\boldsymbol{g}_z$, has already been proven in [39] with a somewhat different method; we comment on this point in Remark A.20.

Proof. Firstly we show that for any rational expression q as defined at the beginning of this section and any $\theta > 0$, the bounds (**M1**)–(**M2**) from Theorem A.18 hold uniformly on $\{\, z \,:\, \mathrm{Im} z \geq \theta^{-1},\, |z| \leq \theta \,\}$, namely that for any $\theta > 0$ there exist $C_\theta > 0$ such that for $\{\, z \,:\, \mathrm{Im} z \geq \theta^{-1},\, |z| \leq \theta \,\}$

$$\| M_z^{(\mathrm{sc})} \|_{\mathbb{C}^{m\times m}} \leq C_\theta, \qquad \| \mathscr{L}_z^{-1} \|_{\mathbb{C}^{m\times m} \to \mathbb{C}^{m \times m}} \leq C_\theta. \tag{A.80}$$

The first estimate in (A.80) follows directly from (A.26) and $(\mathrm{Im} z)^{-1} \leq \theta$. In order to obtain the second estimate, we use the identity

$$\mathscr{L}_z[\, R \,] = (\mathrm{id} \otimes \tau_{\mathcal{S}}) \Big(\Big(\boldsymbol{L}^{\mathrm{sc}} - z J_m \otimes \mathbb{1}_{\mathcal{S}} \Big)^{-1} \Big((M_z^{\mathrm{sc}})^{-1} \, R \, (M_z^{\mathrm{sc}})^{-1} \otimes \mathbb{1}_{\mathcal{S}} \Big) \\ \Big(\boldsymbol{L}^{\mathrm{sc}} - z J_m \otimes \mathbb{1}_{\mathcal{S}} \Big)^{-1} \Big) \tag{A.81}$$

for all $\mathbb{R} \in \mathbb{C}^{m \times m}$, where $\boldsymbol{L}^{\mathrm{sc}} := \boldsymbol{L}(\boldsymbol{s}, \boldsymbol{c}, \boldsymbol{c}^*)$ is the linearization of the rational expression q evaluated on the freely independent semicircular and circular elements (see the proof of [13, Proposition 2.17] for the derivation of this identity). The boundedness of $(M_z^{\mathrm{sc}})^{-1}$ follows from the boundedness of M_z^{sc}, the Dyson equation (A.24) and $|z| < \theta$, while the trivial bound in Lemma A.10 implies the boundedness of $\| (\boldsymbol{L}^{\mathrm{sc}} - z J_m \otimes \mathbb{1}_{\mathcal{S}})^{-1} \|_{\mathbb{C}^{m\times m} \otimes \mathcal{S}}$ holding polynomially in θ. Combining this with (A.81) and $(\mathrm{Im} z)^{-1} \leq \theta$, we obtain the second inequality in (A.80). Note that in the above argument, we rely on the fact that $(\boldsymbol{s}, \boldsymbol{c}) \in \mathcal{D}_{\boldsymbol{q}_0, \ldots, \boldsymbol{q}_n; C}(\mathcal{S})$.

Now we can proceed with the proof (A.78) by following the argument used in Theorem A.18 but restricting the analysis to a subset $\{\, z \,:\, \mathrm{Im} z \geq \theta^{-1},\, |z| \leq \theta \,\}$ bounded away from the real line. The concentration inequalities (A.78) then follow from (A.77) with $(\mathrm{Im} z)^{-1} \leq \theta$. Finally, taking the (1,1)-component in the averaged global law for the linearization in (A.78) yields (A.79). □

From (A.27), the function $M_z^{(\mathrm{sc})}(1,1)$ is a Stieltjes transform of a probability measure $\langle e_1, V(d\lambda) e_1 \rangle$, which together with Proposition A.19 implies that in probability (and almost surely) the empirical spectral measure of $q(\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Y}^*)$ converges weakly to $\langle e_1, V(d\lambda) e_1 \rangle$ as $N \to \infty$.

Remark A.20. In [39] an induction argument on the height of rational functions (similar as in the proof of Lemma A.15) was used to show, that the trace and the norm of a rational expression in GUE (or Wigner) matrices converges almost surely to the trace and norm of the same rational expression in semicircular elements. This result implies in particular the global law for self-adjoint rational expressions in Wigner matrices as in (A.79) (without giving the optimal convergence rate) and the convergence of the (pseudo)spectrum. We use a slightly different approach compared to [39], in particular in the proof of Lemma A.15, by working with the linearization matrix and the generalized resolvent rather than directly with the rational expressions. This allows us to prove not only the global law but also the local laws for the *linearized* models in Sect. A.6, leading to the global and local laws for the rational expressions (A.77) and (A.79) with a precise control of the convergence speed. This also allows us to link the main outcomes of the present paper with the results about the convergence of the (pseudo)spectrum and the global and local laws established in [3,13].

Remark A.21. The results of Sects. A.3–A.6 can be extended to the setting when the NC variables $x_1, \dots, x_{\alpha_*}, y_1, \dots, y_{\beta_*}$ are replaced by the *matrices* of NC variables

$$x_\alpha = \big(x_\alpha(i,j)\big)_{1\le i,j\le l_\alpha} \in \mathcal{A}^{l_\alpha\times l_\alpha}, \quad y_\beta = \big(y_\beta(i,j)\big)_{\substack{1\le i\le l_\beta\\ 1\le j\le k_\beta}} \in \mathcal{A}^{l_\beta\times k_\beta} \tag{A.82}$$

and $q(\boldsymbol{x},\boldsymbol{y},\boldsymbol{y}^*) \in \mathcal{A}^{l_q\times l_q}$ with properly chosen dimensions that make the corresponding matrix operations well defined. The linearization of such model has the same structure (A.6) and we can analyze it by considering the matrix entries $x_\alpha(i,j)$, $1\le i,j\le k_\alpha$, and $y_\beta(i,j)$, $1\le i\le l_\beta, 1\le j\le k_\beta$, as the new NC variables. Note that in this case the diagonal entries of x_α are self-adjoint.

In order to relate the spectrum of the rational expression $q(\boldsymbol{x},\boldsymbol{y},\boldsymbol{y}^*)$ to its linearization, the generalized resolvent should be replaced by

$$z \mapsto (\boldsymbol{L} - z\boldsymbol{J}_{l_q})^{-1}. \tag{A.83}$$

where $\boldsymbol{J}_{l_q} = \sum_{i=1}^{l_q} E_{ii} \in \mathbb{C}^{m\times m}$ with E_{ij} being the standard basis of $\mathbb{C}^{m\times m}$ and m the dimension of the linearization. With the above definition the upper-left $l_q\times l_q$ corner of $(\boldsymbol{L} - z\boldsymbol{J}_{l_q})^{-1}$ is equal to $(q(\boldsymbol{x},\boldsymbol{y},\boldsymbol{y}^*) - I_{l_q}\otimes \mathbb{1}_{\mathcal{A}})^{-1}$, the resolvent of the rational expression $q(\boldsymbol{x},\boldsymbol{y},\boldsymbol{y}^*)$.

We can then proceed with the study of the corresponding random matrix model, with $x_\alpha(i,i)$ independent Wigner matrices and all other elements being independent *i.i.d.* matrices. The linearization of this model is again a *Kronecker* random matrix; therefore, all the probabilistic estimates in the above analysis remain valid. The proof of the main results of Sects. A.3–A.6 follows the same lines as in the $l_q = l_\alpha = l_\beta = k_\beta = 1$ setting; the main changes are notational and are reduced to incorporating the additional structure (A.82) and (A.83).

B. Norm Bounds for Random Sample Covariance Matrices

In this section, we prove the norm bounds for random sample covariance ensembles that are used frequently in the paper. Although various forms of these bounds are well known in the literature (see, e.g., [6,17] or [33]), we provide a short alternative proof to make sure that the assumptions about the random matrix ensembles and the probabilistic estimates match the setting of the present paper.

Lemma B.1. *Let $l, m \in \mathbb{N}$ and let $W \in \mathbb{C}^{ln \times mn}$ be a (non-Hermitian) random matrix with independent centered entries of variance $1/(ln)$. Suppose that the entries of W have finite moments of all orders. Then for any $m < l$ and $\delta > 0$, as $n \to \infty$,* a.w.o.p.

$$\left\|(WW^*)^{-1}\right\|_{\mathbb{C}^{ln\times ln}} \leq \frac{1}{\left(1 - \frac{1}{\sqrt{m/l}}\right)^2 (1-\delta)}. \tag{B.1}$$

Proof. Let $\widehat{W}_{ij} \in \mathbb{C}^{n\times n}$, $1 \leq i \leq l$, $1 \leq j \leq m$, be independent $n \times n$ random matrices having i.i.d. centered entries with variance $1/n$, so that W can be written as a block matrix

$$W = \left(\frac{1}{\sqrt{l}} \widehat{W}_{ij}\right)_{\substack{i=1\ldots l\\ j=1\ldots m}}. \tag{B.2}$$

Denote $Q := WW^*$. For any $z \in \mathbb{C}$, the linearization of $Q - zI_{ln}$ has a simple structure

$$\boldsymbol{L}_{Q,z} = \begin{pmatrix} -zI_{ln} & W \\ W^* & -I_{mn} \end{pmatrix}, \tag{B.3}$$

and using the representation (B.2) can be viewed as a Kronecker matrix in i.i.d. matrices $\widehat{W}_{ij}$. Denote also

$$\boldsymbol{L}_{Q,z}(\boldsymbol{c}, \boldsymbol{c}^*) := \begin{pmatrix} -zI_l \otimes \mathbb{1}_{\mathcal{S}} & W^{(\mathrm{c})} \\ (W^{(\mathrm{c})})^* & -I_m \otimes \mathbb{1}_{\mathcal{S}} \end{pmatrix}, \tag{B.4}$$

where $W^{(\mathrm{c})} := (\frac{1}{\sqrt{l}} c_{ij})_{\substack{i=1\ldots l\\ j=1\ldots m}}$, and $\{c_{ij}\}$ is a family of freely independent circular elements. By using the properties of the Kronecker random matrices from [3, Theorems 2.4 and 6.1], we get that for any $\varepsilon \in (0,1)$ *a.w.o.p.*

$$\mathrm{Spec}\,(WW^*) \subset \left\{z : \mathrm{dist}\big(0, \mathrm{supp}\,\rho^z\big) \leq \varepsilon\right\}, \tag{B.5}$$

where ρ^z is the self-consistent density of states satisfying $\mathrm{supp}\,\rho^z = \mathrm{Spec}$ $(\boldsymbol{L}_0^{Q,z}(\boldsymbol{c}, \boldsymbol{c}^*))$ and

$$\boldsymbol{L}_\omega^{Q,z}(\boldsymbol{c}, \boldsymbol{c}^*) = \begin{pmatrix} -\omega I_{l+m} \otimes \mathbb{1}_{\mathcal{S}} & \boldsymbol{L}_{Q,z}(\boldsymbol{c}, \boldsymbol{c}^*) \\ \big(\boldsymbol{L}_{Q,z}(\boldsymbol{c}, \boldsymbol{c}^*)\big)^* & -\omega I_{l+m} \otimes \mathbb{1}_{\mathcal{S}} \end{pmatrix}. \tag{B.6}$$

Inclusion (B.5) can be obtained by repeating the argument from Lemma A.15 leading to (A.55), and taking into account the $(l+m)\times(l+m)$ decomposition in (B.3) and (B.4).

Fix $\delta \in (0,1)$. Then, we claim that there exists a sufficiently small $\varepsilon > 0$ such that

$$\left\{z : \mathrm{dist}\big(0, \mathrm{supp}\,\rho^z\big) \leq \varepsilon\right\} \cap \left\{z : \mathrm{Re} z < \left(1 - \frac{1}{\sqrt{m/l}}\right)^2 (1-\delta)\right\} = \emptyset. \tag{B.7}$$

Indeed, $W^{(\mathrm{c})}(W^{(\mathrm{c})})^*$ follows the free Poisson distribution with rate $l/m > 1$ (equivalent to Marchenko–Pastur distribution with parameter $m/l < 1$) with

$$\mathrm{Spec}\,(W^{(\mathrm{c})}(W^{(\mathrm{c})})^*) = \left[\left(1 - \sqrt{m/l}\right)^2, \left(1 + \sqrt{m/l}\right)^2\right]. \tag{B.8}$$

There exists $C > 0$ such that for any $z \in \mathbb{C}$ with $\mathrm{Re} z < (1 - \sqrt{m/l})^2(1 - \delta)$

$$\left\| \left(W^{(\mathrm{c})}(W^{(\mathrm{c})})^* - zI_l \otimes \mathbb{1}_{\mathcal{S}} \right)^{-1} \right\|_{\mathbb{C}^{l\times l} \otimes \mathcal{S}} \le C, \tag{B.9}$$

and thus, by the Schur complement formula, there exists $\widetilde{C} > 0$ depending only on C, such that

$$\left\| \left(\boldsymbol{L}_{Q,z}(\boldsymbol{c}, \boldsymbol{c}^*) \right)^{-1} \right\|_{\mathbb{C}^{(l+m)\times(l+m)} \otimes \mathcal{S}} \le \widetilde{C}. \tag{B.10}$$

From (B.6) and (B.10), we get that the resolvent of the Hermitized matrix (B.6) is bounded at $\omega = 0$

$$\left\| \left(\boldsymbol{L}_0^{Q,z}(\boldsymbol{c}, \boldsymbol{c}^*) \right)^{-1} \right\|_{\mathbb{C}^{2(l+m)\times 2(l+m)} \otimes \mathcal{S}} = \left\| \left(\boldsymbol{L}_{Q,z}(\boldsymbol{c}, \boldsymbol{c}^*) \right)^{-1} \right\|_{\mathbb{C}^{(l+m)\times(l+m)} \otimes \mathcal{S}} \le \widetilde{C} \tag{B.11}$$

uniformly for $\mathrm{Re} z < (1 - \sqrt{m/l})^2(1 - \delta)$, from which (B.7) can be obtained using the resolvent identity (similarly as in (A.65)–(A.67)).

Finally, (B.5) and (B.7) imply that for any $\delta \in (0, 1)$ *a.w.o.p.* the smallest eigenvalue of the positive definite matrix WW^* is greater than $(1 - 1/\sqrt{m/l})^2(1 - \delta)$, from which (B.1) follows. □

References

[1] Ajanki, O.H., Erdős, L., Krüger, T.: Stability of the matrix Dyson equation and random matrices with correlations. Probab. Theory Relat. Fields **173**(1–2), 293–373 (2019)

[2] Akemann, G., Baik, J., Di Francesco, P. (eds.): The Oxford Handbook of Random Matrix Theory. Oxford University Press, Oxford (2011)

[3] Alt, J., Erdős, L., Krüger, T., Nemish, Yu.: Location of the spectrum of Kronecker random matrices. Ann. Inst. Henri Poincaré Probab. Stat. **55**(2), 661–696 (2019)

[4] Anderson, G.W., Guionnet, A., Zeitouni, O.: An Introduction to Random Matrices. Cambridge Studies in Advanced Mathematics, vol. 118. Cambridge University Press, Cambridge (2010)

[5] Bai, Z., Silverstein, J.W.: Spectral Analysis of Large Dimensional Random Matrices. Springer Series in Statistics, 2nd edn. Springer, New York (2010)

[6] Bai, Z.D., Yin, Y.Q.: Limit of the smallest eigenvalue of a large-dimensional sample covariance matrix. Ann. Probab. **21**(3), 1275–1294 (1993)

[7] Beenakker, C.W.J.: Random-matrix theory of quantum transport. Rev. Mod. Phys. **69**, 731–808 (1997)

[8] Beenakker, C.W.J.: Condensed matter physics. In: Akemann, G., Baik, J., Di Francesco, P. (eds.) The Oxford Handbook of Random Matrix Theory, pp. 723–743. Oxford Univ. Press, Oxford (2011)

[9] Berstel, J., Reutenauer, C.: Noncommutative Rational Series with Applications. Encyclopedia of Mathematics and Its Applications, vol. 137. Cambridge University Press, Cambridge (2011)

[10] Brouwer, P.W.: Generalized circular ensemble of scattering matrices for a chaotic cavity with nonideal leads. Phys. Rev. B **51**, 16878–16884 (1995)

[11] Brouwer, P.W., Beenakker, C.W.J.: Diagrammatic method of integration over the unitary group, with applications to quantum transport in mesoscopic systems. J. Math. Phys. **37**(10), 4904–4934 (1996)

[12] Büttiker, M.: Scattering theory of thermal and excess noise in open conductors. Phys. Rev. Lett. **65**, 2901–2904 (1990)

[13] Erdős, L., Krüger, T., Nemish, Yu.: Local laws for polynomials of Wigner matrices. J. Funct. Anal. **278**(12), 108507 (2020)

[14] Erdős, L., Krüger, T., Schröder, D.: Random matrices with slow correlation decay. Forum Math. Sigma **7**, e8 (2019)

[15] Erdős, L., Knowles, A., Yau, H.-T., Yin, J.: The local semicircle law for a general class of random matrices. Electron. J. Probab. **18**(59), 58 (2013)

[16] Erdős, L., Yau, H.-T.: A Dynamical Approach to Random Matrix Theory, volume 28 of Courant Lecture Notes in Mathematics. Courant Institute of Mathematical Sciences, New York; American Mathematical Society, Providence, RI (2017)

[17] Feldheim, O.N., Sodin, S.: A universality result for the smallest eigenvalues of certain sample covariance matrices. Geom. Funct. Anal. **20**(1), 88–123 (2010)

[18] Forrester, P.J.: Log-Gases and Random Matrices. London Mathematical Society Monographs Series, vol. 34. Princeton University Press, Princeton (2010)

[19] Fyodorov, Y.V.: Random Matrix Theory of Resonances: An Overview, vol. 9, pp. 666–669. Institute of Electrical and Electronics Engineers Inc., Piscataway (2016)

[20] Fyodorov, Y.V.: Reflection time difference as a probe of s-matrix zeroes in chaotic resonance scattering. Acta Phys. Polon. A **136**(5), 785–789 (2019)

[21] Fyodorov, Y.V., Savin, D.V.: Resonance scattering of waves in chaotic systems. In: Akemann, G., Baik, J., Di Francesco, P. (eds) The Oxford Handbook of Random Matrix Theory, pp. 703–722. Oxford Univ. Press, Oxford (2011)

[22] Gesztesy, F., Tsekanovskii, E.: On matrix-valued Herglotz functions. Math. Nachr. **218**, 61–138 (2000)

[23] Haagerup, U., Thorbjørnsen, S.: A new application of random matrices: $\mathrm{Ext}(C^*_{\mathrm{red}}(F_2))$ is not a group. Ann. Math. (2) **162**(2), 711–775 (2005)

[24] Helton, J.W., Mai, T., Speicher, R.: Applications of realizations (aka linearizations) to free probability. J. Funct. Anal. **274**(1), 1–79 (2018)

[25] Helton, J.W., Rashidi Far, R., Speicher, R.: Operator-valued semicircular elements: solving a quadratic matrix equation with positivity constraints. Int. Math. Res. Not. IMRN **22**, 15 (2007)

[26] Kleene, S.C.: Representation of events in nerve nets and finite automata. In: Shannon, C. E., McCarthy, J. (eds.) Automata Studies, Annals of Mathematics Studies, vol. 34, pp. 3–41. Princeton University Press, Princeton (1956)

[27] Lehner, F.: Computing norms of free operators with matrix coefficients. Am. J. Math. **121**(3), 453–486 (1999)

[28] Mahaux, C., Weidenmüller, H.A.: Comparison between the r-matrix and eigenchannel methods. Phys. Rev. **170**, 847–856 (1968)

[29] Mai, T., Speicher, R., Yin, S.: The free field: realization via unbounded operators and Atiyah property (2019). ArXiv:1905.08187

[30] Mello, P.A., Pereyra, P., Seligman, T.H.: Information theory and statistical nuclear reactions. I. General theory and applications to few-channel problems. Ann. Phys. **161**(2), 254–275 (1985)

[31] Oberholzer, S., Sukhorukov, E.V., Strunk, C., Schönenberger, C., Heinzel, T., Holland, M.: Shot noise by quantum scattering in chaotic cavities. Phys. Rev. Lett. **86**, 2114–2117 (2001)

[32] Reutenauer, C.: Inversion height in free fields. Sel. Math. (N.S.) **2**(1), 93–109 (1996)

[33] Rudelson, M., Vershynin, R.: Smallest singular value of a random rectangular matrix. Commun. Pure Appl. Math. **62**(12), 1707–1739 (2009)

[34] Schiff, J.L.: Normal Families. Universitext. Springer, New York (1993)

[35] Schomerus, H.: Random matrix approaches to open quantum systems. In: Grégory Schehr, Alexander Altland, Yan V. Fyodorov, Neil O'Connell, and Leticia F. Cugliandolo (eds.) Stochastic Processes and Random Matrices, pp. 409–473. Oxford Univ. Press, Oxford (2017)

[36] Speicher, R.: Combinatorial theory of the free product with amalgamation and operator-valued free probability theory. Mem. Am. Math. Soc. **132**(627), x+88 (1998)

[37] Verbaarschot, J., Weidenmüller, H., Zirnbauer, M.: Grassmann integration in stochastic quantum physics: the case of compound-nucleus scattering. Phys. Rep. **129**(6), 367–438 (1985)

[38] Wigner, E.P.: On the distribution of the roots of certain symmetric matrices. Ann. Math. **2**(67), 325–327 (1958)

[39] Yin, S.: Non-commutative rational functions in strongly convergent random variables. Adv. Oper. Theory **3**(1), 178–192 (2018)

László Erdős
IST Austria
Klosterneuburg
Austria
e-mail: lerdos@ist.ac.at

Torben Krüger
University of Copenhagen
Copenhagen
Denmark
e-mail: tk@math.ku.dk

Yuriy Nemish
UC San Diego
La Jolla
USA
e-mail: ynemish@ucsd.edu

Communicated by Vadim Gorin.

Received: October 23, 2020.

Accepted: July 5, 2021.