Commun. Math. Phys. 214, 249 – 286 (2000)

Communications in
Mathematical Physics

On the Law of Addition of Random Matrices

L. Pastur[1,2,⋆], **V. Vasilchuk**[2,3]

[1] Centre de Physique Théorique de CNRS, Luminy–case 907, 13288 Marseille, France.
E-mail: pastur@cpt.univ-mrs.fr
[2] U.F.R. de Mathématiques, Université Paris 7, 2, place Jussieu, 75251 Paris Cedex 05, France
[3] Mathematical Division, Institute for Low Temperature Physics, 47, Lenin Ave., 310164, Kharkov, Ukraine.
E-mail: vasilchuk@ilt.kharkov.ua

Received: 27 October 1999/ Accepted: 22 March 2000

Abstract: Normalized eigenvalue counting measure of the sum of two Hermitian (or real symmetric) matrices A_n and B_n rotated independently with respect to each other by the random unitary (or orthogonal) Haar distributed matrix U_n (i.e. $A_n + U_n^* B_n U_n$) is studied in the limit of large matrix order n. Convergence in probability to a limiting nonrandom measure is established. A functional equation for the Stieltjes transform of the limiting measure in terms of limiting eigenvalue measures of A_n and B_n is obtained and studied.

1. Introduction

The paper deals with the eigenvalue distribution of the sum of two $n \times n$ Hermitian or real symmetric random matrices as $n \to \infty$. Namely we express the limiting normalized counting measure of eigenvalues of the sum via the same measures of its two terms, assuming that the latter exist and that terms are randomly rotated one with respect another by an unitary or an orthogonal random matrix uniformly distributed over the group $U(n)$ or $O(n)$ respectively.

One may mention several motivations of the problem. First, it can be regarded in the context of the general problem to describe the eigenvalues of the sum of two matrices in terms of eigenvalues of two terms of the sum. The latter problem dates back at least to the paper of H. Weyl [33], and is related to a number of interesting questions of combinatorics, geometry, algebra, etc. (see e.g. review [8] for recent results and references). The problem is also of considerable interest for mathematical physics because of its evident links with spectral theory and quantum mechanics (perturbation theory in particular).

It is clear that one cannot expect in general a simple and closed expression for eigenvalues of the sum of two given matrices via eigenvalues of terms. Hence, it is natural to look for a "generic" asymptotic answer, studying a randomized version of the

⋆ On leave from the U.F.R. de Mathématiques, Université Paris 7.

problem in which at least one of the two terms is random and both behave rather regularly as $n \to \infty$. Particular results of this type were given in [16, 19] where it was proved that under certain conditions the normalized eigenvalue counting measure of the sum converges in probability to the nonrandom limit that can be found as a unique solution of a certain functional equation, determined by the both term of the sum. Thus, a randomized version of the problem admits a rather constructive and explicit solution in ceratin cases. These results were developed in several directions (see e.g. [9]–[11] and the recent work [21]). Similar problems arose recently in operator algebras studies, known now as the free (non-commutative) probability (see [28, 31, 29] for results and references). In particular, the notion of the R-transform and the free convolution of measures were introduced by Voiculescu and allowed the limiting eigenvalue distributions of the sum to be given in a rather general and simple form. From the point of view of the random matrix theory the problem that we are going to consider is a version of the problem of the deformation (see e.g. [7] for this term) of a given random matrix (that can be a non-random matrix in particular) by another random matrix in the case when "randomness" of the latter includes as an independent part the random choice of the basis in which this matrix is diagonal. We will discuss this topic in more detail in Sect. 2.

In this paper we present a simple method of deriving functional equations for the limiting eigenvalue distribution in a rather general situation. The method is based on certain differential identities for expectations of smooth matrix functions with respect to the normalized Haar measure of $U(n)$ (or $O(n)$) and on elementary matrix identities, the resolvent identity first of all. The basic idea is the same as in [16, 19]: to study not the moments of the counting measure, as it was proposed in the pioneering paper by Wigner [34], but rather its Stieltjes (called also the Cauchy or the Borel) transform, playing the role of an appropriate generating (or characteristic) function of the moments. However, the technical implementation of the idea in this paper is different and simpler than in [16, 19] (see Remark 1 after Theorem 2.1).

The paper is organized as follows. In Sect. 2 we present and discuss our main results (Theorem 2.1). In Sect. 3 we prove Theorems 3.1 and 3.2 giving the solution of the problem under the conditions of the uniform in n boundedness of the fourth moments of the normalized counting measure of the terms. These conditions are more restrictive than those for our principal result, given in Theorem 2.1. Their advantage is that they allow us to use the main ingredients of our approach in more transparent form, free of technicalities. In Sect. 4 we prove Theorem 2.1, whose main condition is the uniform boundedness of the first absolute moment of the normalized counting measure of one of the two terms of the sum. In Sect. 5 we study certain properties of solutions of the functional equation and of the limiting counting measure. In Sect. 6 we discuss topics related to our main result and our technique.

2. Model and Main Result

We consider the ensemble of n-dimensional Hermitian (or real symmetric) random matrices H_n of the form

$$H_n = H_{1,n} + H_{2,n}, \tag{2.1}$$

where

$$H_{1,n} = V_n^* A_n V_n, \;\; H_{2,n} = U_n^* B_n U_n.$$

We assume that A_n and B_n are random Hermitian (or real symmetric) matrices having arbitrary distributions, V_n and U_n are unitary (or orthogonal) random matrices uniformly distributed over the unitary group $U(n)$ (or over the orthogonal group $O(n)$) with respect to the Haar measure, and A_n, B_n, V_n and U_n are mutually independent. For the sake of definiteness we will restrict ourselves to the case of Hermitian matrices and the group $U(n)$ respectively. The results for symmetric matrices and for the group $O(n)$ have the same form, although their proof is more involved technically (see Sect. 6).

We are interested in the asymptotic behavior as $n \to \infty$ of the *normalized eigenvalue counting measure* (NCM) N_n of the ensemble (2.1), defined for any Borel set $\Delta \subset \mathbb{R}$ by the formula

$$N_n(\lambda) = \frac{\#\{\lambda_i \in \Delta\}}{n}, \tag{2.2}$$

where λ_i, $i = 1, \ldots, n$ are the eigenvalues of H_n.

The problem was studied recently in [31,26,30] in the context of free (non-commutative) probability. In particular, it follows from results of [26] that if the matrices A_n and B_n are non-random, their norms are uniformly bounded in n, i.e. their NCM $N_{1,n}$ and $N_{2,n}$ have uniformly in n compact supports, and if these measures have weak limits as $n \to \infty$,

$$N_{1,n} \to N_1, \ N_{2,n} \to N_2, \tag{2.3}$$

then the NCM (2.2) of random matrix (2.1) converges weakly with probability 1 to a non-random measure N. Besides, if

$$f(z) = \int_{-\infty}^{\infty} \frac{N(d\lambda)}{\lambda - z}, \quad \mathrm{Im} z > 0, \tag{2.4}$$

is the *Stieltjes transform* of this limiting measure and

$$f_r(z) = \int_{-\infty}^{\infty} \frac{N_r(d\lambda)}{\lambda - z}, \quad r = 1, 2, \tag{2.5}$$

are the Stieltjes transforms of N_r, $r = 1, 2$ of (2.3), then according to [18] $f(z)$ satisfies the functional equation

$$f(z) = f_1(z + R_2(f(z))), \tag{2.6}$$

where $R_2(f)$ is defined by the relation

$$z = -\frac{1}{f_2(z)} - R_2(f_2(z))) \tag{2.7}$$

and is known as the R-transform of the measure N_2 of (2.3) (see Remark 3 after Theorem 2.1 and [31,29] for the definition and properties of this transform taking into account that our definition (2.7) differs from that of [31] by the sign). The proof of this result in [26,18] was based on the asymptotic analysis of the expectations $m_k^{(n)}$ of moments of measure (2.2). Since, according to the spectral theorem and the definition (2.2),

$$m_k^{(n)} = \mathbf{E}\{M_k^{(n)}\}, \ M_k^{(n)} = n^{-1}\mathrm{Tr} H_n^k, \tag{2.8}$$

one can study the averaged moments $m_k^{(n)}$ by computing asymptotically the expectations of the divided by n traces of the powers of (2.1), i.e. of corresponding multiple sums. This direct method dates back to the classic paper by Wigner [34] and requires a considerable amount of combinatorial analysis, existence of all moments measures $N_{1,2}^{(n)}$ and their rather regular behavior as $n \to \infty$ to obtain the convergence of expectations (2.8) for all integer k and to guarantee that limiting moments determine uniquely corresponding measure. By using this method it was proved in [26,18] that the expectation of N_n converges to the limit, determined by (2.6)–(2.7) and in [26] that the variance $\mathrm{Var}\{M_k^{(n)}\} = \mathbf{E}\{(M_k^{(n)})^2\} - \mathbf{E}^2\{M_k^{(n)}\}$ admits the bound

$$\mathrm{Var}\{M_k^{(n)}\} \le \frac{C_k}{n^2}, \tag{2.9}$$

where C_k is independent of n. This bound yields evidently the convergence of all moments with probability 1, thereby the weak convergence with probability 1 of random measures (2.2) to the non-random limit, determined by (2.6), (2.7). The convergence with probability 1 here and below is understood as that in the natural probability space

$$\Omega = \prod_n \Omega_n, \tag{2.10}$$

where Ω_n is the probability space of matrices (2.1), that is the product of respective spaces of A_n and B_n and two copies of the group $U(n)$ for U_n and V_n.

In this paper we obtain the analogous result under weaker assumptions and by using a method that does not involve combinatorics. This is because we work with the Stieltjes transforms of measures (2.2) and (2.3) and derive directly the functional equations for their limits and the bound analogous to (2.9) for the rate of their convergence (rather well known in random matrix theory, see e.g. [23,11]) by using certain simple identities for expectations of matrix functions with respect to the Haar measure (Proposition 3.2 below) and elementary facts on resolvents of Hermitian matrices.

The Stieltjes transform was first used in studies of the eigenvalue distribution of random matrices in paper [16] and proved to be an efficient tool in the field (see e.g. [9–14,19–21,24,25]). We list the properties of the Stieltjes transform that we will need below (see e.g.[1]).

Proposition 2.1. *Let m be a non-negative and normalized to unity measure and*

$$s(z) = \int \frac{m(d\lambda)}{\lambda - z}, \quad \mathrm{Im}\ z \neq 0 \tag{2.11}$$

be the Stieltjes transform of m (here and below integrals without limits denote the integrals over the whole axis). Then:

(i) *$s(z)$ is analytic in $\mathbb{C} \setminus \mathbb{R}$ and*

$$|s(z)| \le |\mathrm{Im}\ z|^{-1}; \tag{2.12}$$

(ii)

$$\mathrm{Im}\ s(z)\mathrm{Im}\ z > 0,\ \mathrm{Im}\ z \neq 0; \tag{2.13}$$

(iii)

$$\lim_{y\to\infty} y|s(iy)| = 1; \tag{2.14}$$

(iv) *for any continuous function* φ *with compact support we have the inversion (Frobenius–Perron) formula*

$$\int \phi(\lambda)N(\mathrm{d}\lambda) = \lim_{\varepsilon\to 0}\frac{1}{\pi}\int \phi(\lambda)\mathrm{Im}\ s(\lambda+i\varepsilon); \tag{2.15}$$

(v) *conversely, any function verifying (2.12)–(2.14) is the Stieltjes transform of a nonnegative and normalized to unity measure and this one-to-one correspondence between measures and their Stieltjes transforms is continuous if one will use the topology of weak convergence for measures and the topology of convergence on compact sets of* $\mathbb{C}\setminus\mathbb{R}$ *for their Stieltjes transforms.*

We formulate now our main result. Since eigenvalues of a Hermitian matrix are unitary invariant we can replace matrices (2.1) by

$$H_n = A_n + U_n^* B_n U_n, \tag{2.16}$$

where A_n, B_n and U_n are the same as in (2.1). However, it is useful to keep in mind that the problem is symmetric in A_n and B_n. We prove

Theorem 2.1. *Let* H_n *be the random* $n\times n$ *matrix of the form (2.1). Assume that the normalized eigenvalue counting measures* $N_{r,n}$, $r=1,2$ *of matrices* A_n *and* B_n *converge weakly in probability as* $n\to\infty$ *to the non-random nonnegative and normalized to 1 measures* N_r, $r=1,2$ *respectively and that*

$$\sup_n \int |\lambda|\mathbf{E}N_{r,n}^*(\mathrm{d}\lambda) \equiv m_1 < \infty, \tag{2.17}$$

where $N_{r,n}^*$ *is one of the measures* $N_{1,n}$ *or* $N_{2,n}$. *Then the normalized eigenvalue counting measure* N_n *of* H_n *converges in probability to a non-random nonnegative and normalized to 1 measure* N *whose Stieltjes transform (2.4) is a unique solution of the system*

$$f(z) = f_1\left(z - \frac{\Delta_2(z)}{f(z)}\right),$$

$$f(z) = f_2\left(z - \frac{\Delta_1(z)}{f(z)}\right), \tag{2.18}$$

$$f(z) = \frac{1-\Delta_1(z)-\Delta_2(z)}{-z}$$

in the class of functions $f(z)$ *satisfying (2.12)–(2.14) and functions* $\Delta_r(z), r=1,2$ *analytic for* $\mathrm{Im}\ z\neq 0$ *and satisfying conditions*

$$\Delta_{1,2}(z)\to 0 \text{ as } \mathrm{Im}\ z\to\infty, \tag{2.19}$$

where $f_r(z), r=1,2$ *are Stieltjes transforms (2.5) of the measures* N_r, $r=1,2$ *and* $\mathbf{E}\{\cdot\}$ *denotes the expectation with respect to the probability measure, generated by* A_n, B_n, U_n *and* V_n.

The theorem will be proved in Sect. 4. Here we make several remarks related to the theorem (see also Sect. 5).

Remark 1. The historically first example of a random matrix ensemble representable in the form (2.16) was proposed in [16] and has the form

$$H_{m,n} = H_{0,n} + \sum_{i=1}^{m} \tau_i P_{q_i}, \tag{2.20}$$

where $H_{0,n}$ is a non-random $n \times n$ Hermitian matrix such that its normalized eigenvalue counting measure converges weakly to a limiting non-negative and normalized to 1 measure N_0, τ_i, $i = 1, \dots, m$ are i.i.d. random variables and P_{q_i} are orthogonal projections on unit vectors q_i, $i = 1, \dots, m$, that are independent of one another and of $\{\tau_i\}_{i=1}^m$, and uniformly distributed over the unit sphere in $\mathbb{C}^n$ [1]. It is clear that the matrix

$$\sum_{i=1}^{m} \tau_i P_{q_i} \tag{2.21}$$

can be written in the form $U_n^* B_n U_n$ of the second term of (2.1) or (2.16). According to [16] the NCM of the random matrix (2.21) converges in probability as $n \to \infty$, $m \to \infty$, $m/n \to c \geq 0$ to a non-random nonnegative and normalized to 1 measure whose Stieltjes transform $f_{MP}(z)$ satisfies the equation

$$f_{MP}(z) = -\left(z - c \int \frac{\tau \sigma(\mathrm{d}\tau)}{1 + \tau f_{MP}(z)}\right)^{-1}, \tag{2.22}$$

where σ is the probability law of τ_i in (2.20). Assume that σ has the finite first moment

$$\int |\tau| \sigma(\mathrm{d}\tau) < \infty. \tag{2.23}$$

Then taking (2.21) as the second term of (2.1) we get, in view of inequality

$$\mathbf{E}\left\{\int |\lambda| N_{2,n}(\mathrm{d}\lambda)\right\} \leq n^{-1} \sum_{i=1}^{m} \mathbf{E}\{|\tau_i|\} = \frac{m}{n} \mathbf{E}\{|\tau|\} < \infty,$$

the condition (2.17) of Theorem 2.1. Applying then Theorem 2.1 in which $f_2(z)$ is given by (2.22), we obtain from the two last equations of the system (2.18) that

$$\frac{\Delta_1(z)}{f(z)} = c \int \frac{\tau \sigma(\mathrm{d}\tau)}{1 + \tau f_{MP}(z)}.$$

This and the first equation of (2.18) yield the functional equation for the Stieltjes transform of the limiting eigenvalue distribution of ensemble (2.20)

$$f(z) = f_0\left(z - c \int \frac{\tau \sigma(\mathrm{d}\tau)}{1 + \tau f(z)}\right), \tag{2.24}$$

[1] In fact, in [16] a more general class of independent random vectors was considered, but we restrict ourselves here to the unit vectors, in order to have an example of an ensemble of form (2.1).

where $f_0(z)$ is the Stieltjes transform of the limiting NCM N_0 of the non-random matrix $H_{0,n}$. This equation was obtained in [16] by another method, whose main ingredient was careful analysis of changes of the resolvent of matrices (2.20) induced by addition of the $(m+1)$th term, i.e. by a rank-one perturbation. This allowed the authors to prove that the sequence $g_{i,n}(z) = n^{-1}\mathrm{Tr}(H_{i,n} - z)^{-1},\ i = 1, \ldots, m$ converges in probability to the non-random limit $f(z,t),\ z \in \mathbb{C}\backslash\mathbb{R},\ t \in [0,1]$, as $n \to \infty, m \to \infty, i \to \infty,\ m/n \to c,\ i/m \to t$, and that the limiting function $f(z,t)$ satisfies the quasilinear PDE,

$$\frac{\partial f}{\partial t} + c\frac{\tau(t)}{1+\tau(t)f}\frac{\partial f}{\partial z},\ \ f(z,0) = f_0(z), \tag{2.25}$$

where $\tau(t)$ is the inverse of the probability distribution $\sigma(\tau) = \mathbf{P}\{\tau_i \le \tau\}$. It can be shown that the solution of (2.25) at $t = 1$ coincides with (2.20) [16]. Equation (2.25) with $\tau(t) \equiv$ const is a particular case of the so-called complex Burgers equation which appeared in free probability [31], where the random matrices (2.20) provide an analytic model for the stationary processes with free increments, like in the conventional probability the heat equation and sums of i.i.d. random variables comprise an important ingredient of the theory of random processes with independent increments.

Remark 2. Consider the ensemble known as the deformed Gaussian ensemble [19]:

$$H_n = H_{0,n} + M_n, \tag{2.26}$$

where $H_{0,n}$ is a non-random matrix such that its normalized eigenvalue counting measure converges weakly to the limit N_0 and $M_n = \{M_{jk}\}_{j,k=1}^n$ is a random Hermitian matrix whose matrix elements M_{jk} are complex Gaussian random variables satisfying conditions:

$$\overline{M_{jk}} = M_{kj},\ \mathbf{E}\{M_{jk}\} = 0,\ \mathbf{E}\{M_{j_1k_1}\overline{M_{j_2k_2}}\} = \frac{2w^2}{n}\delta_{j_1j_2}\delta_{k_1k_2}. \tag{2.27}$$

In other words, the ensemble is defined by the distribution

$$\mathbf{P}(\mathrm{d}M) = Z_n^{-1} exp\left\{-\frac{n}{4w^2}\mathrm{Tr}M^2\right\}\mathrm{d}M, \tag{2.28}$$

$$\mathrm{d}M = \prod_{j=1}^{n}\mathrm{d}M_{jj}\prod_{1\le j<k\le n}\mathrm{dRe}\,M_{jk}\mathrm{dIm}\,M_{jk},$$

where Z_n is the normalization constant. The distribution defines the Gaussian Unitary Ensemble (GUE) [17]. This is why ensemble (2.26) is called the deformed GUE [7]. It is known [17] that M_n can be written in the form

$$M_n = U_n^*\Lambda_n U_n, \tag{2.29}$$

where U_n are unitary matrices whose probability law is the Haar measure on $U(n)$ and Λ_n is independent of the U_n diagonal random matrix whose normalized eigenvalue counting measure converges with probability 1 to the semicircle law. The Stieltjes transform $f_{sc}(z)$ of the latter satisfies the simple functional equation [19]

$$f_{sc}(z) = -(z + 2w^2 f_{sc}(z)), \tag{2.30}$$

whose solution yields the semicircle law by Wigner

$$N_{sc}(d\lambda) = (4\pi w^2)^{-1}\sqrt{8w^2 - \lambda^2}\chi_{[-2\sqrt{2}w, 2\sqrt{2}w]}(\lambda)d\lambda, \tag{2.31}$$

where $\chi_{[a,b]}(\lambda)$ is the indicator of the interval $[a, b] \subset \mathbb{R}$. It is easy to see that

$$\mathbf{E}\{n^{-1}\mathrm{Tr}M_n^2\} = 2w^2 < \infty.$$

Denoting by $N_{sc,n}$ the NCM of the random matrices defined by (2.28) we can rewrite this inequality in the form

$$\int_{-\infty}^{\infty} \lambda^2 \mathbf{E}\{N_{sc,n}(d\lambda)\} < \infty. \tag{2.32}$$

Thus, if we use (2.29) as the second term in (2.16), it will satisfy condition (2.1). Taking $f_{sc}(z)$ as $f_2(z)$ in (2.18) we find from the two last equations of the system that $\Delta_2(z)/f(z) = -2w^2 f(z)$ and then the first equation of (2.18) takes the form

$$f(z) = f_0(z + 2w^2 f(z)), \tag{2.33}$$

where $f_0(z)$ is the Stieltjes transform of the limiting counting measure of matrices $H_{0,n}$. This functional equation determining the limiting eigenvalue distribution of the deformed GUE was found by another method in [19] (see also [12]) for random matrices (2.26) in which M_n has independent (modulo the Hermitian symmetry conditions) entries, for (2.28) in particular.

Remark 3. Consider now a probability measure $m(\mathrm{d}\lambda)$ and assume that its second moment m_2 is finite. In this case we can write the Stieltjes transform $s(z)$ of m in the form

$$s(z) = -(z + \Sigma(z))^{-1},$$

where $\Sigma(z)$ is the Stieltjes transform of a non-negative measure whose total mass is m_2 (to prove this fact one can use, for example, the general integral representation [1] for functions satisfying (2.13)). Since $s'(z) = z^{-2}(1 + o(1))$, $z \to \infty$, then, according to the local inversion theorem, there exists a unique functional inverse $z(s)$ of $s(z)$ defined and analytic in a neighborhood of zero and assuming its values in a neighborhood of infinity. Denote

$$\Sigma(z(s)) = R_m(s) \tag{2.34}$$

and following Voiculescu [31] call $R_m(s)$ the R-transform of the probability measure m. By using the R-transforms $R_{1,2}$ of measures $N_{1,2}$ we can rewrite the first two equations of system (2.18) in the form

$$\frac{\Delta_{1,2}}{f(z)} = \frac{1}{f(z)} + z + R_{2,1}(f(z)) = -R(f(z)) + R_{2,1}(f(z)), \tag{2.35}$$

where R denotes the R-transform of the limiting normalized counting measure N of the ensemble (2.1) (the measure whose Stieltjes transform is f). These relations and the third equation of system (2.18) lead to the remarkably simple expression of R via R_1 and R_2,

$$R(f) = R_1(f) + R_2(f), \tag{2.36}$$

that "linearizes" the rather complex system (2.18). The relation was obtained by Voiculescu in the context of C^*-algebra studies (see [31, 29] for results and references). Thus, one can regard the system (2.18) as a version of the binary operation on measures defined by (2.36) and known as the non-commutative convolution. A simple precursor of relation (2.36) containing the functional inverses of f and $f_{1,2}$ for real z lying outside of the support of N_0 in (2.24) was used in [16] (see also [25]) to locate the support of N in terms of the support of N_0 in the case of ensemble (2.20). The simplest form of the relation (2.36) for the case when both measures are semicircle measures (2.31), i.e. when $R_{1,2} = 2w_{1,2}^2 f$, was indicated in [19]. Formal derivation of relation (2.36) for the case when the both matrices H_1 and H_2 are distributed according to the laws

$$P_{1,2}^{(n)}(\mathrm{d}H) = Z_{1,2}^{(n)} exp\{-nV_{1,2}(H)\}, \mathrm{d}H, \tag{2.37}$$

where $V_{1,2} : \mathbb{R} \to \mathbb{R}_+$ are polynomials of an even degree was given in [36]. The derivation is based on the perturbation theory with respect to the non-quadratic part of $V_{1,2}$ and the R-transform is related to the sum of irreducible diagrams of the formal perturbation series. Existence of the limiting eigenvalue counting measure for the random matrix ensemble (2.37) was rigorously proved in [6] for a rather broad class of functions V (not necessarily polynomials). It was also proved that the normalized counting measure (2.2) converges in probability to the limiting measure. The form (2.29) of matrices of ensemble (2.37) can be deduced from known results on the ensemble (2.37) (see e.g.[5]) in the same way as for the GUE (2.28), where $V(\lambda) = \lambda^2/4w^2$ (see [17]). Condition (2.17) follows from results of [6, 21]. Thus we can apply Theorem 2.1 to obtain rigorously relation (2.36) in the case when matrices H_r, $r = 1, 2$ in (2.1) are distributed according to (2.37).

Remark 4. The problem of addition of random Hermitian (real symmetric) matrices has natural multiplicative analogues in the case of positive definite Hermitian (real symmetric) or unitary (orthogonal) matrices. Namely, assuming that A_n and B_n are positive definite matrices and U_n is the unitary (orthogonal) Haar distributed random matrix we can consider the positive definite random matrix

$$H_n = A_n^{1/2} U_n^* B_n U_n A_n^{1/2}. \tag{2.38}$$

Likewise, if S_n and T_n are unitary (orthogonal) matrices and U_n is as above we can consider the random unitary matrices

$$V_n = S_n U_n^* T_n U_n. \tag{2.39}$$

In latter case the normalized eigenvalue counting measure is defined as n^{-1} times the number of eigenvalues belonging to a Borel set of the unit circle.

In both cases (2.38) and (2.39) one can study the limiting properties of the NCM's of respective random matrices provided that the "input" matrices A_n, B_n, S_n and T_n have limiting eigenvalue distributions. The first examples of ensembles of the above forms as multiplicative analogues of the ensemble (2.20) were proposed in [16], where the respective functional equations analogous to (2.24) were derived. A general class of the random matrix ensembles of these forms was studied in free probability [28, 31, 2], where the notions of the S-transform and the free multiplicative convolution of measures were proposed and used to give a general form of the limiting eigenvalue distributions of products (2.38) and (2.39). It will be shown in the subsequent paper [27] that a version of the method of this paper leads to results analogous to those given in Theorem 2.1 above.

3. Convergence with Probability 1 for Non-Random A_n and B_n

As the first step of the proof of Theorem 2.1 we prove the following

Theorem 3.1. *Let H_n be the random $n \times n$ matrix of the form (2.1) in which A_n and B_n are non-random Hermitian matrices, U_n and V_n are random independent unitary matrices distributed each according to the normalized to 1 Haar measure on $U(n)$. Assume that the normalized counting measures $N_{r,n}, r = 1, 2$ of matrices A_n and B_n converge weakly as $n \to \infty$ to nonnegative and normalized to 1 measures $N_r, r = 1, 2$ respectively and that*

$$\sup_n \int \lambda^4 N_{r,n}(\mathrm{d}\lambda) = m_4 < \infty, r = 1, 2. \tag{3.1}$$

Then the normalized eigenvalue counting measure (2.2) of H_n converges with probability 1 to a non-random and normalized to 1 measure whose Stieltjes transform (2.4) is a unique solution of the system (2.18) in the class of functions $f(z)$, $\Delta_r(z), r = 1, 2$ analytic for $\mathrm{Im}\ z \neq 0$ *and satisfying conditions (2.12)–(2.14) and (2.19) respectively.*

Remark 1. The theorem generalizes the results of [26] proved under the condition that supports of the NCM $N_{r,n}, r = 1, 2$ of A_n and B_n are uniformly bounded in n.

Remark 2. By mimicking the proof of the Glivenko–Cantelli theorem (see e.g. [15]), one can prove that the random distribution functions $N_n(\lambda) = N_n(] - \infty, \lambda[)$ corresponding to measures (2.2) converge uniformly with probability 1 to the distribution function $N(\lambda) = N(] - \infty, \lambda[)$ corresponding to measure N:

$$\mathbf{P}\{\lim_{n\to\infty} \sup_{\lambda\in\mathbb{R}} |N_n(\lambda) - N(\lambda)| = 0\} = 1.$$

We present now our technical means. First is a collection of elementary facts of linear algebra.

Proposition 3.1. *Let $\mathbf{M}_n$ be the algebra of linear transformations of $\mathbb{C}^n$ in itself ($n \times n$ complex matrices) equipped with the norm, induced by the Euclidean norm of $\mathbb{C}^n$.*

We have :

(i) *if $M \in \mathbf{M}_n$ and $\{M_{jk}\}_{j,k=1}^n$ is the matrix of M in any orthonormalized basis of $\mathbb{C}^n$, then*

$$|M_{jk}| \le ||M||; \tag{3.2}$$

(ii) *if $\mathrm{Tr} M = \sum_{j=1}^n M_{jj}$, then*

$$|\mathrm{Tr} M_1 M_2| \le (\mathrm{Tr} M_1 M_1^*)^{1/2} (\mathrm{Tr} M_2 M_2^*)^{1/2}, \tag{3.3}$$

where M^* is the Hermitian conjugate of M, and if P is a positive definite transformation, then

$$|\mathrm{Tr} M P| \le ||M|| \mathrm{Tr} P; \tag{3.4}$$

(iii) *for any Hermitian transformation* M *its resolvent*

$$G(z) = (M - z)^{-1} \tag{3.5}$$

is defined for all non-real z, $\text{Im } z \neq 0$,

$$||G(z)|| \leq |\text{Im } z|^{-1} \tag{3.6}$$

and if $\{G_{jk}(z)\}_{j,k=1}^{n}$ *is the matrix of* $G(z)$ *in any orthonormalized basis of* $\mathbb{C}^n$ *then*

$$|G_{jk}(z)| \leq |\text{Im } z|^{-1}; \tag{3.7}$$

(iv) *if* M_1 *and* M_2 *are two Hermitian transformations and* $G_r(z), r = 1, 2$ *are their resolvents, then*

$$G_2(z) = G_1(z) - G_1(z)(M_2 - M_1)G_2(z) \tag{3.8}$$

(the resolvent identity);

(v) *if* $G(z) = (M - z)^{-1}$ *is regarded as a function of* M, *then the derivative* $G'(z)$ *of* $G(z)$ *with respect to* M *verifies the relation*

$$G'(z) \cdot X = -G(z)XG(z) \tag{3.9}$$

for any Hermitian $X \in \mathbf{M}_n$, *and, in particular,*

$$||G'(z)|| \leq ||G(z)||^2 \leq |\text{Im } z|^{-2}. \tag{3.10}$$

Here is our main technical tool.

Proposition 3.2. *Let* $\Phi : \mathbf{M}_n \to \mathbb{C}$ *be a continuously differentiable function. Then the following relation holds for any* $M \in \mathbf{M}_n$ *and any Hermitian element* $X \in \mathbf{M}_n$*:*

$$\int_{U(n)} \Phi'(U^* M U) \cdot [X, U^* M U] \mathrm{d}U = 0, \tag{3.11}$$

where

$$[M_1, M_2] = M_1 M_2 - M_1 M_2 \tag{3.12}$$

is the commutator of M_1 *and* M_2 *and the symbol*

$$\int_{U(n)} \ldots dU \tag{3.13}$$

denotes integration over $U(n)$ *with respect to the normalized Haar measure* dU.

Proof. To prove (3.11) we use the right shift invariance of the Haar measure: $\mathrm{d}U = \mathrm{d}(UU_0)$, $\forall U_0 \in U(n)$ according to which the integral

$$\int_{U(n)} \Phi\left(e^{-i\varepsilon X} U^* M U e^{i\varepsilon X}\right) \mathrm{d}U$$

is independent of ε for any Hermitian $X \in \mathbf{M}_n$. Thus its derivative with respect to ε at $\varepsilon = 0$ is zero. This derivative is the l.h.s. of (3.11). $\square$

Proposition 3.3. *System (2.18) has a unique solution in the class of functions* $f(z)$, $\Delta_{1,2}(z)$ *analytic for* $\mathrm{Im}\ z \neq 0$ *and satisfying conditions (2.12)–(2.14) and (2.19).*

Proof. Assume that there exist two solutions $(f', \Delta'_{1,2})$ and $(f'', \Delta''_{1,2})$ of the system. Denote $\delta f = f' - f''$, $\delta\Delta_{1,2} = \Delta'_{1,2} - \Delta''_{1,2}$. Then, by using (2.18) and the integral representation (2.5) for $f_{1,2}$, we obtain the linear system for $\delta\phi = z\delta f$, and for $\delta\Delta_{1,2}$,

$$\begin{aligned} \delta\phi(1 - a_1(z)) + b_1(z)\delta\Delta_1 &= 0, \\ \delta\phi(1 - a_2(z)) + b_2(z)\delta\Delta_2 &= 0, \\ \delta\phi - \delta\Delta_1 - \delta\Delta_2 &= 0, \end{aligned} \tag{3.14}$$

where

$$a_1 = \frac{\Delta''_1}{f'f''} I_2,\ b_1 = \frac{z}{f'} I_2,\ I_2 = I_2(z - \Delta'_1/f', z - \Delta''_1/f''), \tag{3.15}$$

$$I_2(z', z'') = \int \frac{N_2(\mathrm{d}\lambda)}{(\lambda - z')(\lambda - z'')}, \tag{3.16}$$

and a_2, b_2 can be obtained from a_1 and b_1 by replacing N_2 and Δ_1 by N_1 and Δ_2 in the above formulas. For any $y_0 > 0$ consider the domain

$$E(y_0) = \{z \in \mathbb{C} : |\mathrm{Im}\ z| \geq y_0,\ |\mathrm{Re}\ z| \leq |\mathrm{Im}\ z|\}. \tag{3.17}$$

If $s(z)$ is the Stieltjes transform (2.11) of a probability measure m, then we have for $z \in E(y_0)$,

$$\left| \int \frac{\lambda m(\mathrm{d}\lambda)}{\lambda - z} \right| = \left| \int_{|\lambda| \leq M} + \int_{|\lambda| > M} \right| \leq \frac{M}{y_0} + 2 \int_{|\lambda| > M} m(\mathrm{d}\lambda),$$

i.e.

$$zs(z) = -1 + o(1),\ z \to \infty,\ z \in E(y_0). \tag{3.18}$$

Analogously, by using this asymptotic relation and condition (2.19) we obtain that for $z \to \infty$, $z \in E(y_0)$,

$$z^2 I_{1,2}(z) = 1 + o(1),\ a_{1,2}(z) = o(1),\ b_{1,2}(z) = -1 + o(1).$$

Thus the determinant $b_1 b_2 + b_1 + b_2 - (a_2 b_1 + a_1 b_2)$ of system (3.14) is equal asymptotically to -1. We conclude that if y_0 in (3.17) is big enough, then system (3.14) has only a trivial solution, i.e. system (2.18) is uniquely soluble. □

In what follows we use the notation

$$\int_{U(n)} \ldots \mathrm{d}U = \langle \ldots \rangle. \tag{3.19}$$

Proof of Theorem 3.1. Because of unitary invariance of eigenvalues of Hermitian matrices we can assume without loss of generality that the unitary matrix V in (2.1) is set to unity, i.e. we can work with the random matrix (2.16). We will omit below the subindex

n in all cases when it will not lead to confusion. Write the resolvent identity (3.8) for the pair (H_1, H) of (2.1):

$$G(z) = G_1(z) - G_1(z)H_2G(z), \tag{3.20}$$

where

$$G(z) = (H_1 + H_2 - z)^{-1},\ G_1(z) = (H_1 - z)^{-1}.$$

Consider the matrix $\langle g_n(z)G(z)\rangle$, where

$$g_n(z) = \frac{1}{n}\mathrm{Tr}G(z) = \int \frac{N_n(\mathrm{d}\lambda)}{\lambda - z},\ \mathrm{Im}\ z \neq 0 \tag{3.21}$$

is the Stieltjes transform of random measure (2.2). The resolvent identity (3.20) leads to the relation

$$\langle g_n(z)G(z)\rangle = \langle g_n(z)\rangle G_1(z) - G_1(z)\langle g_n(z)H_2G(z)\rangle. \tag{3.22}$$

By using Proposition 3.2 with the matrix element $\left((H_1 + M - z)^{-1}\right)_{ac}$ as $\Phi(M)$ we have in view of (3.9) and (3.11), (3.12),

$$\langle (G[X, H_2]G)_{ac}\rangle = 0.$$

Choosing the Hermitian matrix X with only (a, b)th and (b, a)th non-zero entries, we obtain

$$\langle G_{aa}(H_2G)_{bc}\rangle = \langle (GH_2)_{aa}G_{bc}\rangle. \tag{3.23}$$

Applying to this relation the operation $n^{-1}\sum_{a=1}^{n}$ and taking into account the definition (3.21) of $g_n(z)$ we rewrite the last relation in the form

$$\langle g_n(z)H_2G(z)\rangle = \langle \delta_{2,n}(z)G(z)\rangle,$$

where

$$\delta_{2,n}(z) = \frac{1}{n}\mathrm{Tr}H_2G(z). \tag{3.24}$$

Thus we can rewrite (3.22) as

$$\langle g_n(z)G(z)\rangle = \langle g_n(z)\rangle G_1(z) - G_1(z)\langle \delta_{2,n}(z)G(z)\rangle. \tag{3.25}$$

Introduce now the centralized quantities

$$g_n^{\circ}(z) = g_n(z) - f_n(z),\ \delta_{2,n}^{\circ}(z) = \delta_{2,n}(z) - \Delta_{2,n}(z), \tag{3.26}$$

where

$$f_n(z) = \langle g_n(z)\rangle,\ \Delta_{2,n}(z) = \langle \delta_{2,n}(z)\rangle. \tag{3.27}$$

With these notations (3.25) becomes

$$f_n(z)\langle G(z)\rangle = f_n(z)G_1(z) - \Delta_{2,n}(z)G_1(z)\langle G(z)\rangle + R_{1,n}(z), \tag{3.28}$$

where

$$R_{1,n}(z) = -\langle g_n^\circ(z)G(z)\rangle - G_1(z)\langle\delta_{2,n}^\circ(z)G(z)\rangle. \tag{3.29}$$

Besides, since

$$\begin{aligned} n^{-1}\mathrm{Tr}H^2 &= n^{-1}\mathrm{Tr}(H_1+H_2)^2 \le 2n^{-1}\mathrm{Tr}H_1^2 + 2n^{-1}\mathrm{Tr}H_2^2 \\ &= 2\int\lambda^2 N_{1,n}(\mathrm{d}\lambda) + 2\int\lambda^2 N_{2,n}(\mathrm{d}\lambda) \le 4m_2 \le 4m_4^{1/2}, \end{aligned} \tag{3.30}$$

we have

$$\mu_2 \equiv \sup_n(n^{-1}\mathrm{Tr}H^2) = \sup_n\int\lambda^2 N_n(\mathrm{d}\lambda) \le 4m_2 \le 4m_4^{1/2} < \infty. \tag{3.31}$$

Thus

$$g_n(z) = \int\frac{N_n(\mathrm{d}\lambda)}{\lambda - z} = -\frac{1}{z} + \widehat{g}_n(z),$$

where

$$\widehat{g}_n(z) = \int\frac{\lambda N_n(\mathrm{d}\lambda)}{(\lambda - z)z}.$$

In view of (3.31)

$$|z\widehat{g}_n(z)| \le |\mathrm{Im}\ z|^{-1}\int|\lambda|N_n(\mathrm{d}\lambda) \le |\mathrm{Im}\ z|^{-1}m_4^{1/4},$$

i.e. the asymptotic relation

$$g_n^{-1}(z) = -z\left(1 + O\left(\frac{1}{|\mathrm{Im}\ z|}\right)\right),\ \mathrm{Im}\ z \to \infty \tag{3.32}$$

holds uniformly in n. We have also the simple bound

$$|g_n(z)| \le |\mathrm{Im} z|^{-1} \tag{3.33}$$

following from (3.4) and (3.7) and, in addition, according to Proposition 3.1 and (3.24), the bounds

$$|\delta_{2,n}(z)| \le m_4^{1/4}|\mathrm{Im}z|^{-1}, \tag{3.34}$$

$$z\delta_{2,n}(z) = n^{-1}\mathrm{Tr}H_2 zG(z) = n^{-1}\mathrm{Tr}H_2(-1 + HG(z)). \tag{3.35}$$

Hence, in view of (3.31),

$$\begin{aligned} |z\delta_{2,n}(z)| &\le (n^{-1}\mathrm{Tr}H_2^2)^{1/2} + (n^{-1}\mathrm{Tr}H_2^2)^{1/2}(n^{-1}\mathrm{Tr}H^2G(z)G^*(z))^{1/2} \\ &\le m_4^{1/4} + 2m_4^{1/2}/y_0, \end{aligned} \tag{3.36}$$

i.e. $z\delta_{2,n}(z)$ is uniformly bounded in n.

As a result of the above bounds we have for $|\mathrm{Im}\ z| \ge y_0$ uniformly in n,

$$||\Delta_{2,n}(z)f_n^{-1}(z)G_1(z)|| = O\left(\frac{1}{y_0}\right),\quad y_0 \to \infty,$$

i.e. the matrix $1 - \Delta_{2,n}(z) f_n^{-1}(z) G_1(z)$ is invertible uniformly in n and there is y_0 independent of n and such that for $|\mathrm{Im}\ z| \geq y_0$,

$$||(1 + \Delta_{2,n}(z) f_n^{-1}(z) G_1(z))^{-1}|| \leq 2. \tag{3.37}$$

Thus (3.28) is equivalent to

$$\langle G(z) \rangle = (1 + \Delta_{2,n}(z) f_n^{-1}(z) G_1(z))^{-1} G_1(z) \\ (1 + \Delta_{2,n}(z) f_n^{-1}(z) G_1(z))^{-1} f_n^{-1}(z) R_{1,n}(z)$$

or to

$$\langle G(z) \rangle = G_1 \left(z - \Delta_{2,n}(z) f_n^{-1}(z) \right) + (1 + \Delta_{2,n}(z) f_n^{-1}(z) G_1(z))^{-1} f_n^{-1}(z) R_{1,n}(z).$$

Applying to this relation the operation $n^{-1}\mathrm{Tr}$ we obtain

$$f_n(z) = f_{1,n}(z - \Delta_{2,n}(z) f_n^{-1}(z)) + r_{1,n}(z), \tag{3.38}$$

where

$$f_{1,n}(z) = n^{-1}\mathrm{Tr} G_1(z) = \int \frac{N_{1,n}(\mathrm{d}\lambda)}{\lambda - z} \tag{3.39}$$

is the Stieltjes transform of the normalized counting measure of $H_{1,n}$ in (2.1) and

$$r_{1,n}(z) = n^{-1}\mathrm{Tr}(1 + \Delta_{2,n}(z) f_n^{-1}(z) G_1(z))^{-1} f_n^{-1}(z) R_{1,n}(z), \tag{3.40}$$

where $R_{1,n}(z)$ is defined in (3.29). We show in the next Theorem 3.2 that there exists a sufficiently big $y_0 > 0$ and $C(y_0) > 0$, both independent of n and such that if $z \in E(y_0)$, where $E(y_0)$ is defined in (3.17), then the variances

$$v_1(z) = \langle |g_n^\circ(z)|^2 \rangle, \ v_2(z) = \langle |\delta_{2,n}^\circ(z)|^2 \rangle \tag{3.41}$$

admit the bounds

$$v_1(z) \leq \frac{C(y_0)}{n^2}, \ v_2(z) \leq \frac{C(y_0)}{n^2}. \tag{3.42}$$

These bounds, Proposition 3.1, (3.37), and the Schwartz inequality for the expectation $\langle \dots \rangle$ imply that uniformly in n and in $z \in E(y_0)$,

$$|r_{1,n}(z)| \leq \frac{2C^{1/2}(y_0)}{n} (1 + y_0^{-1}) \langle | f_n^{-2}(z) n^{-1}\mathrm{Tr} G(z) G^*(z)|^2 \rangle^{1/2}.$$

In view of (3.27), (3.32) and the identity $zG(z) = -1 + HG(z)$ we have

$$f_n^{-1}(z) G(z) = -z(1 + O(y_0^{-1})) G(z) = (1 + O(y_0^{-1}))(1 - HG(z)),$$

and since, by (3.3), (3.4) and (3.30),

$$|\langle n^{-1}\mathrm{Tr} HG(z) \rangle| \leq y_0^{-1} \langle n^{-1}\mathrm{Tr} H^2 \rangle \\ \leq 2m_4^{1/4} y_0^{-1}, \ |\langle n^{-1}\mathrm{Tr} H^2 G(z) G^*(z) \rangle| \leq 4m_4^{1/2} y_0^{-2},$$

we obtain that for $z \in E(y_0)$,

$$|r_{1,n}(z)| \leq \frac{C_1(y_0)}{n}, \tag{3.43}$$

where $C_1(y_0)$ is independent of n and is bounded in y_0.

Furthermore, the bounds (3.33) and (3.34) imply that sequences $\{f_n(z)\}$ and $\{\Delta_{2,n}(z)\}$ are analytic and uniformly in n bounded for $|\text{Im } z| \geq y_0 > 0$. Thus the sequences are compact with respect to uniform convergence on compacts of the domain

$$D(y_0) = \{z \in \mathbb{C} : |\text{Im } z| \geq y_0 > 0\}. \tag{3.44}$$

In addition, according to the hypothesis of the theorem, the normalized counting measures $N_{1,n}$ of matrices $H_{1,n}$ converge weakly to a limiting probability measure N_1. Hence, their Stieltjes transforms (3.39) converge uniformly on compacts of (3.44) to the Stieltjes transform f_1 of N_1. Hence, if $y_0 > 0$ is large enough, there exist two analytic in (3.44) functions f and Δ_2 verifying the relation

$$f(z) = f_1\left(z - \frac{\Delta_2(z)}{f(z)}\right), \ |\text{Im } z| \geq y_0.$$

This is the first equation of system (2.18). The second equation of the system follows from the argument above in which the roles H_1 and H_2 are interchanged, in particular the quantity $\langle n^{-1}\text{Tr} H_1 G(z)\rangle$ is denoted $\Delta_{1,n}(z)$. As for the third equation, it is just the limiting form of the identity

$$\langle n^{-1}\text{Tr}(H_{1,n} + H_{2,n} - z)G(z)\rangle = 1. \tag{3.45}$$

Thus, we have derived system (2.18). Its unique solubility in domain (3.17) where y_0 is large enough is proved in Proposition 3.3. Besides, all three functions f_n, $\Delta_{r,n}$, $r = 1, 2$ defined in (3.27) are a priori analytic for $|\text{Im } z| > 0$. Hence, their limits f, Δ_r, $r = 1, 2$ are also analytic for non-real z. In view of the weak compactness of probability measures and the continuity of the one-to-one correspondence between nonnegative measures and their Stieltjes transforms (see Prop. 2.1(v)) there exists a unique nonnegative measure N such that f admit the representation (2.4). The measure N is a probability measure in view of (3.32) and.(2.14).

We conclude that the whole sequence $\{f_n\}$ of expectations (3.27) of the Stieltjes transforms g_n (3.21) of measures (2.2) converges uniformly on compacts of $D(y_0)$, where $D(y_0)$ is defined in (3.44), to the limiting function f verifying (2.18). This result, Theorem 3.2 and the Borel–Cantelli lemma imply that the sequence $\{g_n(z)\}$ converges with probability 1 to $f(z)$ for any fixed $z \in D(y_0)$. Since the convergence of a sequence of analytic functions on any countable set having an accumulation point in their common domain of definition implies the uniform convergence of the sequence on any compact of the domain, we obtain the convergence g_n to f with probability 1 on any compact of $D(y_0)$. Due to the continuity of the one-to-one correspondence between probability measures and their Stieltjes transforms (see Prop.2.1(v)) the normalized eigenvalue counting measure (2.2) of the eigenvalues of random matrix (2.1) converge weakly with probability 1 to the nonrandom measure N whose Stieltjes transform (2.4) satisfies (2.18). □

Theorem 3.2. *Let H_n be the random matrix of the form (2.1) satisfying the condition of Theorem 3.1. Denote*

$$g_n(z) = n^{-1}\mathrm{Tr}(H_n - z)^{-1},\ \delta_{r,n}(z) = n^{-1}\mathrm{Tr}H_{r,n}(H_n - z)^{-1}, r = 1, 2. \tag{3.46}$$

Then there exist y_0 and $C(y_0)$, both positive and independent of n and such that the variances of random variables (3.46) admit the bounds for $|\mathrm{Im}\ z| \geq y_0$,

$$\langle|g_n(z) - \langle g_n(z)\rangle|^2\rangle \leq \frac{C(y_0)}{n^2}, \tag{3.47}$$

$$\langle|\delta_{r,n}(z) - \langle\delta_{r,n}(z)\rangle|^2\rangle \leq \frac{C(y_0)}{n^2},\quad r = 1, 2, \tag{3.48}$$

if $z \in E(y_0)$, where $E(y_0)$ is defined in (3.17).

Proof. Because of the symmetry of the problem with respect to H_1 and H_2 in (2.1) it suffices to prove (3.48) for, say, $\delta_{2,n}(z)$. Besides, we will use below the notations $g(z)$ and $\delta(z)$ for $g_n(z)$ and $\delta_{2,n}(z)$ and the notations 1 and 2 for two values z_1 and z_2 of the complex spectral parameter z. We assume that $|\mathrm{Im}\ z_{1,2}| \geq y_0 > 0$.

We will use the same approach as in the proof of Theorem 3.1, i.e. we will derive and study certain relations obtained by using Proposition 3.2 and the resolvent identity.

Consider the matrix

$$V_1 = \langle g^\circ(1)G(2)\rangle, \tag{3.49}$$

where $g^\circ(1) = g(1) - \langle g(1)\rangle$. It is clear that $n^{-1}\mathrm{Tr}V_1$ for $z_1 = z$ and $z_2 = \overline{z}$ is the variance (3.47) that we denoted by $v_1(z)$ in (3.41):

$$\langle|g^\circ(z)|^2\rangle = n^{-1}\mathrm{Tr}V_1|_{z_1=z,z_2=\overline{z}} = v_1(z). \tag{3.50}$$

In view of the resolvent identity (3.20) for the pair (H_1, H) we have

$$V_1 = -G_1(2)W, \tag{3.51}$$

$$W = \langle g^\circ(1)H_2G(2)\rangle. \tag{3.52}$$

Applying Proposition 3.2 to the function

$$\Phi(M) = G^\circ_{aa}(1)(MG(2))_{cd},$$

where $G(z) = (H_1 + M - z)^{-1}$, and

$$\begin{aligned} G^\circ(z) &= G(z) - \langle G(z)\rangle \\ &= (H_1 + M - z)^{-1} - \int_{U(n)} (H_1 + U^*BU - z)^{-1}dU, \end{aligned}$$

we obtain the relation

$$\begin{aligned} &- \langle(G(1)[X, H_2]G(1))_{aa}(H_2G(2))_{cd}\rangle + \langle G^\circ_{aa}(1)([X, H_2]G(2))_{cd}\rangle \\ &\qquad - \langle G^\circ_{aa}(1)(H_2G(2)[X, H_2]G(2))_{cd}\rangle = 0, \end{aligned}$$

where the operation $[\ldots,\ldots]$ is defined in (3.12). Choosing as X the Hermitian matrix having only the $(c,j)^{\text{th}}$ and $(j,c)^{\text{th}}$ non-zero entries, we obtain from the above relation the following one:

$$\begin{gathered}-\langle G_{ac}(1)(H_2G(1))_{ja}(H_2G(2))_{cd}\rangle+\langle (G(1)H_2)_{ac}G_{ja}(1)(H_2G(2))_{cd}\rangle\\ +\langle G^{\circ}_{aa}(1)\delta_{cc}(H_2G(2))_{jd}\rangle-\langle G^{\circ}_{aa}(1)(H_2)_{cc}G_{jd}(2)\rangle\\ -\langle G^{\circ}_{aa}(1)(H_2G(2))_{cc}(H_2G(2))_{jd}\rangle+\langle G^{\circ}_{aa}(1)(H_2G(2)H_2)_{cc}G_{jd}(2)\rangle=0.\end{gathered}$$

Applying to this relation the operation $n^{-1}\sum\limits_{ac}$ and taking into account that

$$g^{\circ}=n^{-1}\sum_{a}G^{\circ}_{aa},$$

we have

$$\begin{aligned}n^{-2}\langle [G^2(1),H_2]H_2G(2)\rangle+\langle g^{\circ}(1)H_2G(2)\rangle&\\ +\ \langle g^{\circ}(1)k(2)G(2)\rangle-\langle g^{\circ}(1)\delta(2)H_2G(2)\rangle=0,&\qquad(3.53)\end{aligned}$$

where

$$k(z)=n^{-1}\mathrm{Tr}K(z),\ K(z)=BG_U(z)B-B,\quad G_U(z)=UG(z)U^{*}.\qquad(3.54)$$

Introducing the centralized quantity (cf. (3.26))

$$k^{\circ}=k-\langle k\rangle,\qquad(3.55)$$

and using our notations (3.24) and (3.27), we can rewrite (3.53) as

$$(1-\Delta(2))W=-\langle k(2)\rangle V_1+R,\qquad(3.56)$$

where

$$R=\langle g^{\circ}(1)\delta^{\circ}(2)H_2G(2)\rangle-\langle g^{\circ}(1)k^{\circ}(2)G(2)\rangle-T_1,\qquad(3.57)$$

and

$$T_1=n^{-2}\langle [G^2(1),H_2]H_2G(2)\rangle.$$

In view of the uniform in n bound (3.36)), the function $1-\Delta(z)$ is uniformly in n bounded away from zero. Thus we have from (3.51), (3.52) and (3.56),

$$V_1=\Big(1-\langle k(2)\rangle(1-\Delta(2))^{-1}G_1(2)\Big)^{-1}(1-\Delta(2))^{-1}G_1(2)R.\qquad(3.58)$$

According to (3.54), (3.6) and (3.1), we have uniformly in n,

$$|k(z)|\le y_0^{-1}n^{-1}\,\mathrm{Tr}B^2+|n^{-1}\mathrm{Tr}B|\le y_0^{-1}m_4^{1/2}+m_4^{1/4}<\infty.\qquad(3.59)$$

This bound and the universal bound (3.6) imply that the matrix $(1-\langle k(z)\rangle(1-\Delta(z))^{-1}$ $G_1(z))$ is uniformly in n invertible if $|\mathrm{Im}\ z|\ge y_0$ and y_0 is large enough, and hence the matrix

$$Q=\Big(1-\langle k(z)\rangle(1-\Delta(z))^{-1}G_1(z)\Big)^{-1}(1-\Delta(z))^{-1}G_1(z)$$

admits the following bound for $|\text{Im } z| \geq y_0$ and sufficiently large y_0:

$$||Q|| \leq \frac{C}{y_0}, \tag{3.60}$$

where C is an absolute constant.

Setting now in (3.58) $z_1 = z, z_2 = \overline{z}$ and applying to this relation the operation $n^{-1}\text{Tr}$ we obtain in the l.h.s. the variance $v_1(z)$ because of (3.50). As for the r.h.s., its terms can be estimated as follows in view of (3.57):

(i)

$$|\langle g^\circ(1)\delta^\circ(2)n^{-1}\text{Tr}QH_2G(2)\rangle| \leq \alpha_{12}(y_0)v_1^{1/2}v_2^{1/2}, \tag{3.61}$$

where v_2 is defined in (3.41) and because, according to (3.1), (3.3), (3.6) and (3.60),

$$\begin{aligned}|n^{-1}\text{Tr}QH_2G(2)| &\leq (n^{-1}\text{Tr}Q^*Q)^{1/2}(n^{-1}\text{Tr}H_2^2G(2)G^*(2))^{1/2} \leq \\ &\leq Cy_0^{-2}m_4^{1/4} \equiv \alpha_{12}(y);\end{aligned} \tag{3.62}$$

(ii)

$$|\langle g^\circ(1)k^\circ(2)n^{-1}\text{Tr}QG(2)\rangle| \leq \alpha_{13}(y_0)v_1^{1/2}v_3^{1/2}, \tag{3.63}$$

where

$$v_3 = \langle |k^\circ(z)|^2\rangle, \tag{3.64}$$

because

$$\begin{aligned}&|n^{-1}\text{Tr}QG(2)| \leq (n^{-1}\text{Tr}Q^*Q)^{1/2}(n^{-1}\text{Tr}G(2)G^*(2))^{1/2} \\ &\leq Cy_0^{-2} \equiv \alpha_{13}(y_0);\end{aligned} \tag{3.65}$$

(iii)

$$|n^{-3}\text{Tr}(Q[G^2(1), H_2]H_2G(2))| \leq Cm_4^{1/2}y_0^{-4}n^{-2} \equiv \frac{\beta_1(y_0)}{n^2}.$$

Thus we obtain the inequality

$$v_1 \leq \alpha_{12}(y_0)v_1^{1/2}v_2^{1/2} + \alpha_{13}(y_0)v_1^{1/2}v_3^{1/2} + \frac{\beta_1(y_0)}{n^2}, \tag{3.66}$$

where α_{12}, α_{13} and β_1 are independent on n and vanish as $y_0 \to \infty$.

Now we are going to derive analogous inequalities for v_2 and v_3 defined in (3.41) and in (3.64) and to obtain the system

$$v_i \leq \sum_{j=1, j\neq i}^{3} \alpha_{ij}v_i^{1/2}v_j^{1/2} + \frac{\beta_i(y_0)}{n^2},\ i = 1, 2, 3. \tag{3.67}$$

To get the second inequality of the system we consider the matrix (cf. (3.49))

$$V_2 = \langle \delta^\circ(1)H_2G(2)\rangle. \tag{3.68}$$

Applying to V_2 the operation $n^{-1}\mathrm{Tr}$ and setting $z_1 = z$, $z_2 = \overline{z}$, we obtain the variance v_2 of (3.42). On the other hand, using Proposition 3.2 for the function

$$\Phi(M) = (MG(1))^\circ_{aa}(MG(2))_{cd},$$

we obtain, after performing in essence the same procedure as that used in the derivation of (3.53), in particular, choosing the Hermitian matrix X with only the $(c, j)^{\text{th}}$ and $(j, c)^{\text{th}}$ non-zero entries,

$$v_2 = -\langle g(2)\delta^\circ(1)k(2)\rangle + \langle \delta^\circ(1)\delta^2(2)\rangle - T_2, \tag{3.69}$$

where

$$T_2 = \langle n^{-3}\mathrm{Tr}([G_U(1), K(1)]BG(2))\rangle \tag{3.70}$$

and $K(z)$, $k(z)$ are defined in (3.54). Using again centralized quantities (3.26) and (3.55), we can write

$$\langle g(2)\delta^\circ(1)k(2)\rangle = \langle g^\circ(2)\delta^\circ(1)k(2)\rangle + \langle g(2)\rangle\langle \delta^\circ(1)k^\circ(2)\rangle$$

and

$$\langle \delta^\circ(1)\delta^2(2)\rangle = \langle \delta^\circ(1)\delta^\circ(2)\delta(2)\rangle + \langle \delta^\circ(1)\delta^\circ(2)\rangle\langle \delta(2)\rangle.$$

Thus, in view of (3.33), (3.34), (3.59), and the Schwarz inequality we have the bounds

$$\langle g(2)\delta^\circ(1)k(2)\rangle \le v_1^{1/2}v_2^{1/2}m_4^{1/4}(1 + m_4^{1/4}y_0^{-1}) + v_2^{1/2}v_3^{1/2}y_0^{-1},$$

and

$$\langle \delta^\circ(1)\delta^2(2)\rangle \le 2v_2 m_4^{1/4}y_0^{-1}.$$

These bounds and the analogously obtained bound for T_2 in (3.70) lead for $m_4^{1/4}y_0^{-1} \le 1/4$ to the second inequality (3.67), in which

$$\alpha_{21}(y_0) = 4m_4^{1/4},\ \alpha_{23}(y_0) = 2y_0^{-1},\ \beta_2 = 8m_4^{1/4}y_0^{-2}. \tag{3.71}$$

To obtain the third inequality of (3.67) we may use the same scheme as above applied to the matrix $V_3 = \langle k^\circ(1)K(2)\rangle$ (cf. (3.49) and (3.68)). However this requires rather tedious computations and the existence of the uniformly bounded in n sixth moment m_6 of the measure $N_{2,n}$. For this reason we consider the quantity

$$\langle n^{-1}\mathrm{Tr}(BG_U(1)B)^\circ G_U(2)B\rangle, \tag{3.72}$$

where $G_U(z)$ is defined in (3.54). As before we would like to obtain for this quantity a certain relation, based on the invariance of the Haar measure with respect to the group shifts. To this end we will introduce the following function of the unitary matrix U:

$$(BUG(1)U^*B)^\circ_{aa}(UG(2)U^*B)_{cd},$$

where $G(z) = (H_1 + U^*BU - z)^{-1}$ and we will use the analogue of (3.11) obtained from the left shift invariance of the Haar measure. This leads to the relation (cf. (3.53) and (3.69))

$$\langle k^\circ(1)g(2)K(2)\rangle + \langle k^\circ(1)\delta(2)G_U(2)B\rangle - \langle k^\circ(1)G_U(2)B\rangle - T_3 = 0, \tag{3.73}$$

where

$$T_3 = n^{-2}\langle G_U(1)BK(1)G_U(2)B - K(1)BG_U(1)G_U(2)B\rangle.$$

We multiply (3.73) by B from the left and introduce again the centralized quantities g°, δ° and k° defined in (3.26) and (3.55). We obtain

$$(1 - \Delta(2) - f(2)B)\langle k^\circ(1)K(2)\rangle \\ = -\langle k^\circ(1)g^\circ(2)BK(2)\rangle + \langle k^\circ(1)\delta^\circ(2)BG_U(2)B\rangle + BT_3.$$

In view of (3.32) and (3.36) the imaginary part of the function $1 - \Delta(z)$ is uniformly in n bounded away from zero if $|\text{Im}\ z|$ is large enough. Since B is a Hermitian matrix, the matrix

$$S = (1 - \Delta(2) - f(2)B)^{-1} \tag{3.74}$$

admits the bound

$$||S|| = |f(2)|^{-1} \cdot ||((1 - \Delta(2))f^{-1}(2) - B)^{-1}|| \le |f(2)|^{-1} \left|\text{Im}\frac{1 - \Delta(2)}{f(2)}\right|^{-1}.$$

By using (3.28) and (3.34) we find that for $z \in E(y_0)$, where $E(y_0)$ is defined in (3.17) with sufficiently big y_0, we have the uniform in n inequality $|f(2)\text{Im}(1 - \Delta(2))f^{-1}(2)| \ge 1/2$, i.e.

$$||S|| \le 2. \tag{3.75}$$

This leads to the relation

$$V_3 \equiv \langle k^\circ(1)K(2)\rangle = -\langle k^\circ(1)g^\circ(2)SBK(2)\rangle \\ + \langle k^\circ(1)\delta^\circ(2)SBG_U(2)B\rangle + SBT_3. \tag{3.76}$$

We apply to this relation the operation $n^{-1}\text{Tr}$, set $z_1 = z$, $z_2 = \overline{z}$ and estimate the contribution of the first two terms of the r.h.s. as (3.76) as above, using in addition (3.75). We obtain

$$|n^{-1}\text{Tr}SBK(2)| \le 4m_4^{1/2} \equiv \alpha_{31}(y_0), \\ |n^{-1}\text{Tr}SBG_U(2)B| \le 4m_4^{1/2}y_0^{-1} \equiv \alpha_{32}(y_0). \tag{3.77}$$

To estimate the third term of the r.h.s. of (3.76) we use the identity

$$SB = -f^{-1}(2) + (1 - \Delta(2))f^{-1}(2)S,$$

the asymptotic relations (3.32) and (3.34) and the bound (3.75). This yields the bound $||SB|| \le 4y_0$. By using this bound and the same reasoning as in obtaining other bounds above, we obtain

$$|n^{-1}\text{Tr}SBT_3| \le \frac{Cm_4}{y_0^2 n^2} \equiv \frac{\beta_3}{n^2},$$

where C is an absolute constant.

Let us introduce new variables

$$u_1 = y_0 v_1^{1/2},\ u_2 = v_2^{1/2},\ u_3 = v_3^{1/2}. \tag{3.78}$$

Then we obtain from (3.67) and (3.62), (3.65), (3.71), and (3.77) the system

$$u_i^2 \le \sum_{j=1, j\neq i}^{3} a_{ij} u_i u_j + \frac{\gamma_i}{n^2}, \tag{3.79}$$

in which the coefficients $\{a_{ij}, i \neq j\}$ have the form $a_{ij} = y_0^{-1} b_{ij}$, where b_{ij} are bounded in y_0 and in n as $y_0 \to \infty$ and $n \to \infty$. By choosing y_0 sufficiently big (and then fixing it) we can guarantee that $0 \le a_{ij} \le 1/4, i \neq j$. Thus summing the three relations (3.79) we can write the result in the form $(\hat{a}u, u) \le \gamma/n^2$, where $\gamma = \gamma_1 + \gamma_2 + \gamma_3$ and $(\hat{a})_{ij} = \delta_{ij} + (1 - \delta_{ij})/4,\ \ i, j = 1, 2, 3$. Since the minimum eigenvalue of the matrix $\hat{a}$ is $1/2$, we obtain from (3.78) bounds (3.47) and (3.48). □

4. Convergence in Probability

In this section we prove Theorem 2.1. Since, according to Theorem 3.2, the randomness of U_n in (2.1) (or (2.16)) already allows us to prove that the variance of the Stieltjes transform of the NCM (2.2) vanishes as $n \to \infty$, we have only to prove that the additional randomness due to the matrices A_n and B_n in (2.1) does not destroy this property. We will prove this fact first for A_n and B_n whose norms are uniformly bounded in n (see Lemma 4.1 below), and then we will treat the general case of Theorem 2.1 by using a certain truncating procedure.

Proposition 4.1. *Let $\{m_n\}$ be a sequence of random non-negative unit measures on the line and $\{s_n\}$ be the sequence of their Stieltjes transforms (2.11). Then the sequence $\{m_n\}$ converges weakly in probability to a nonrandom non-negative unit measure m if and only if the sequence $\{s_n\}$ converges in probability for any fixed z belonging to a compact $K \subset \{z \in C : \mathrm{Im} z > 0\}$ to the Stieltjes transform f of the measure m.*

Proof. Let us prove first the necessity. According to the hypothesis for any continuous function $\varphi(\lambda)$ having compact support we obtain

$$\lim_{n\to\infty} \mathbf{P}\left\{\left|\int \varphi(\lambda) m(d\lambda) - \int \varphi(\lambda) m_n(d\lambda)\right| > \varepsilon\right\} = 0. \tag{4.1}$$

Let $\chi(\lambda)$ be a continuous function that is equal to 1 if $|\lambda| < A$ and is equal to 0 if $|\lambda| > A + 1$ for some $A > 0$. Then

$$|s(z) - s_n(z)| \le \left|\int \frac{\chi(\lambda) m(d\lambda)}{\lambda - z} - \int \frac{\chi(\lambda) m_n(d\lambda)}{\lambda - z}\right| + \frac{2}{\min\{\mathrm{dist}\{z, \pm A\}\}}.$$

According to (4.1) the first term in the r.h.s. of this inequality converges in probability to zero. Since A is arbitrary, we obtain the required assertion.

To prove sufficiency we assume that for any $z \in K$,

$$\lim_{n\to\infty} \mathbf{P}\{|s(z) - s_n(z)| > \varepsilon\} = 0. \tag{4.2}$$

This relation and the inequality (cf. (2.12))

$$|s_n(z)| \leq \max_{z\in K} |\mathrm{Im} z|^{-1} \equiv y_0^{-1} < \infty \tag{4.3}$$

imply that

$$\lim_{n\to\infty} \mathbf{E}\{|s(z) - s_n(z)|\}=0, \tag{4.4}$$

i.e. the sequence $\{s_n(z)\}$ converges to zero in mean. We have also the inequality

$$|s_n^{'}(z)| \leq y_0^{-2} < \infty. \tag{4.5}$$

Inequalities (4.3) and (4.5) imply that the sequence $\{s_n\}_{n=1}^{\infty}$ of random analytic functions is uniformly bounded and equicontinuous. Thus, for any $\eta > 0$ we can construct in K a finite η-network, i.e. a set $\{z_l\}_{l=1}^{p(\eta)}$ such that for any $z \in K$ there exists z_l satisfying the inequality $|z - z_l| \leq \eta$. Then we have for $\phi_n(z) \equiv s_n(z) - s(z)$, $S_l = \{z : |z - z_l| \leq \eta\}$, and $\eta = y_0^2\varepsilon/2$, where ε is arbitrary

$$\sup_K |\phi_n(z)| = \max_{l=1\ldots p(\eta)} \sup_{z\in K\cap S_l} |\phi_n(z)| \leq \varepsilon + \sum_{l=1}^{p(\eta)} |\phi_n(z_l)|,$$

and hence

$$\mathbf{E}\{ \sup_K |\phi_n(z)|\} \leq \varepsilon + \sum_{l=1}^{p(\eta)} \mathbf{E}\{|\phi_n(z_l)|\}.$$

This inequality and (4.4) imply that

$$\lim_{n\to\infty} \mathbf{E}\{ \sup_{z\in K} |s(z) - s_n(z)|\} =0. \tag{4.6}$$

Assume now that the statement is false, i.e. the sequence $\{m_n\}$ does not converge weakly in probability to m. This means that there exists a continuous function φ of a compact support, a subsequence $\{n_k\}$ and some $\varepsilon > 0$ such that

$$\lim_{n_k\to\infty} \mathbf{P}\left\{\left|\int \varphi(\lambda)m(d\lambda) - \int \varphi(\lambda)m_{n_k}(d\lambda)\right| \geq \varepsilon\right\}=\xi > 0. \tag{4.7}$$

On the other hand, we have from (4.6) and the Tchebyshev inequality that for any r there exists an integer $n(r)$ such that for $n \geq n(r)$,

$$\mathbf{P}\left\{\sup_{z\in K} |\phi_n(z)| \leq r^{-1}\right\} \geq 1 - \xi/2. \tag{4.8}$$

Hence, one can select from the sequence $\{n_k\}$ a subsequence $\{n_{k'}\}$ such that inequalities (4.7) and (4.8) are both satisfied. Denote by $\mathcal{A}$ and by $\mathcal{B}$ the events whose probabilities are written in the l.h.s. of (4.7) and (4.8). Then $\mathbf{P}\{\mathcal{A} \cap \mathcal{B}\} \geq \mathbf{P}\{\mathcal{A}\}+ \mathbf{P}\{\mathcal{B}\} - 1 \geq \xi/2$.

Hence, for any $n_{k'}$ there exists a realization $\omega_{n'_k}$ belonging to both sets $\mathcal{A}$ and $\mathcal{B}$, i.e. for which both inequalities

$$\left|\int \varphi(\lambda)m(d\lambda) - \int \varphi(\lambda)m_{n'_k}(d\lambda)\right| \geq \varepsilon, \quad \sup_{z\in K} |\phi_{n'_k}(z)| \leq r^{-1} \tag{4.9}$$

are valid. In view of the compactness of the family of the random analytic functions $\{s_n\}$ with respect to the uniform in K convergence and the weak compactness of the family of random measure $\{m_n\}$ there exists a subsequence $\{n''_k\}$ of $\{n'_k\}$ and a subsequence of realizations $\{\omega_{n''_k}\}$ such that the subsequence $\{m_{n''_k}\}$ corresponding to these realizations converges weakly to a certain measure $\widetilde{m}$ and we have in view of (4.7),

$$\left|\int \varphi(\lambda)m(d\lambda) - \int \varphi(\lambda)\widetilde{m}(d\lambda)\right| \geq \varepsilon > 0. \tag{4.10}$$

On the other hand, in view of (4.9) and the continuity of the correspondence between measures and their Stieltjes transforms (see Proposition 2.1(v)), the subsequence $\{s_{n''_k}\}$ converges uniformly on K to $s(z)$, the Stieltjes transform of the measure m. This is incompatible with (4.10), because of the one-to-one correspondence between measures and their Stieltjes transforms. □

Remark 1. Since the Stieltjes transforms of non-negative and normalized to unity measures are analytic and bounded for non-real z, we can replace the requirement of their convergence for any z belonging to a certain compact set of $\mathbb{C}_\pm$ by the convergence for any z belonging to any interval of the imaginary axis, i.e. for $z = iy,\ y \in [y_1, y_2]$, $y_1 > 0$.

Remark 2. The argument used in the proof of the proposition proves also that if $\{m_n\}$ is a sequence of random non-negative measures converging weakly in probability to a nonrandom non-negative measure m, then the Stieltjes transforms s_n of m_n and the Stieltjes transform s of m are related as follows:

$$\lim_{n\to\infty} \mathbf{E}\{\sup_{z\in K} |s_n(z) - s(z)|\} = 0 \tag{4.11}$$

for any compact set K of $\mathbb{C}_\pm$.

Lemma 4.1. *Let H_n be the random $n \times n$ matrix of the form (2.1) in which A_n and B_n are random Hermitian matrices, U_n and V_n are random unitary matrices distributed each according to the normalized to unity Haar measure on $\mathbf{U}(n)$ and A_n, B_n, U_n and V_n are mutually independent. Assume that the normalized counting measures $N_{r,n}$, $r = 1, 2$ of matrices A_n and B_n converge in probability as $n \to \infty$ to non-random non-negative unit measures N_r, $r = 1, 2$ respectively and that*

$$\sup_n ||A_n|| \leq T < \infty, \ \sup_n ||B_n|| \leq T < \infty. \tag{4.12}$$

Then the normalized counting measure of H_n converges in probability to a non-random unity measure N whose Stieltjes transform $f(z)$ is a unique solution of system (2.18) in the class of functions $f(z)$, $\Delta_r(z)$, $r = 1, 2$ analytic for $\operatorname{Im} z \neq 0$ *and satisfying conditions (2.12)–(2.14) and (2.19).*

Proof. In view of Proposition 4.1 it suffices to show that $\lim_{n\to\infty} \mathbf{E}\{|g_n(z) - f(z)|\} = 0$ for any z belonging to a certain compact set of $\mathbb{C}_\pm$. Moreover, according to Remark 1 after Proposition 4.1, we can restrict ourselves to a certain interval of the imaginary axis, i.e. to

$$z = iy,\ y \in [y_1, y_2],\ 0 < y_1 < y_2 < \infty. \tag{4.13}$$

Since condition (4.12) of the lemma implies evidently condition (3.1) of Theorem 3.1 and Theorem 3.2, all the results obtained in these theorems are valid in our case for any fixed realization of random matrices A_n and B_n. In addition, all n-independent estimating quantities entering various bounds in the proofs of these theorems and depending on the fourth moment m_4 in (3.1) and on y_0 will depend now on T and on y_1 and y_2 in (4.13), but not on particular realizations of random matrices A_n and B_n. We will denote below all these quantities simply by the unique symbol C that may have a different value in different formulas.

In particular, denoting as above by $\langle \dots \rangle$ the expectation with respect to the Haar measure and using (3.42), we can write that

$$\mathbf{E}\{|g_n(z) - \langle g_n(z)\rangle|\} \le \mathbf{E}\{|v_1^{1/2}(z)|\} \le \frac{C}{n}.$$

Thus, it suffices to show that

$$\lim_{n\to\infty} \mathbf{E}\{|\langle g_n(z)\rangle - f(z)|\} = 0,\ z = iy,\ \ y \in [y_1, y_2], \tag{4.14}$$

where y_1 is big enough. Introduce the quantities

$$\gamma_n(y) = iy(\langle g_n(iy)\rangle - f(iy)),\ \ \gamma_{r,n}(y) = \langle \delta_{r,n}(iy)\rangle - \Delta_r(iy),\ \ r = 1, 2. \tag{4.15}$$

By using the second equation of system (2.18) we can write the identity

$$\gamma_n(y) = iy[f_2(iy - t_{1,n}(y)) - f_2(iy - t_1(y))] + \varepsilon_{1,n}(y), \tag{4.16}$$

where

$$\varepsilon_{1,n}(y) = iy[\langle g_n(iy)\rangle - f_2(iy - t_{1,n}(y))], \tag{4.17}$$

$$t_{1,n}(y) = \frac{\langle \delta_{1,n}(iy)\rangle}{\langle g_n(iy)\rangle},\ \ t_1(y) = \frac{\Delta_1(iy)}{f(iy)}. \tag{4.18}$$

We have

$$\mathbf{E}\{|\varepsilon_{1,n}(y)|\} \le y_2 \mathbf{E}\{|\langle g_n(iy)\rangle - g_{2,n}(iy - t_{1,n}(y))|\} \\ \mathbf{E}\{|g_{2,n}(iy - t_{1,n}(y)) - f_2(iy - t_{1,n}(y))|\}. \tag{4.19}$$

The analogues of (3.38), (3.39) in our case are:

$$\langle g_n(z)\rangle = g_{2,n}(z - \langle \delta_{1,n}(z)\rangle \langle g_n(z)\rangle^{-1}) + \widehat{r}_{1,n}(z), \tag{4.20}$$

where

$$g_{2,n}(z) = n^{-1}\mathrm{Tr}G_2(z) = \int \frac{N_{2,n}(\mathrm{d}\lambda)}{\lambda - z},$$

is the Stieltjes transform of random NCM $N_{2,n}$ of $H_{2,n}$,

$$\widehat{r}_{1,n}(z) = \begin{array}{l} -\langle g_n^\circ(z) n^{-1} \mathrm{Tr} P^{-1} \langle g_n(z)\rangle^{-1} G(z)\rangle \\ -\langle \delta_{1,n}^\circ(z) n^{-1} \mathrm{Tr} P^{-1} \langle g_n(z)\rangle^{-1} G_2(z) G(z)\rangle, \end{array}$$

the symbol $\langle \dots \rangle$ denotes the expectation with respect to the Haar measure on $U(n)$, $P = 1 - G_2(z) t_{1,n}(z)$, and

$$g_n^\circ(z) = g_n(z) - \langle g_n(z)\rangle, \ \delta_{1,n}^\circ(z) = \delta_{1,n}(z) - \langle \delta_{1,n}(z)\rangle \tag{4.21}$$

are the respective random variables centralized by the partial expectations with respect to the Haar measure. In addition, we have the analogue of (3.43),

$$\left|\widehat{r}_{1,n}(z)\right| \le \frac{C}{n}.$$

This leads to the following bound for the first term in the r.h.s. of (4.19):

$$\mathbf{E}\{|\langle g_n(iy)\rangle - g_{1,n}(iy - t_{2,n}(y)|\} \le \mathbf{E}\{|\widehat{r}_{1,n}(iy)|\} \le \frac{C}{n}.$$

To show that the second term also vanishes as $n \to \infty$, we use the analogues of (3.32) and (3.36),

$$\left|\langle g_{1,n}(iy)\rangle + \frac{1}{iy}\right| \le \frac{T}{y^2}, \ |\delta_{2,n}(iy)| \le \frac{T}{y},$$

which imply that

$$|t_{1,n}(y)| \le 2T, \tag{4.22}$$

if y_1 is big enough. Thus

$$\mathbf{E}\{|g_{2,n}(iy - t_{1,n}(y)) - f_2(iy - t_{1,n}(y))|\} \le \sup_{|\zeta|\le T} \mathbf{E}\{|g_{2,n}(iy+\zeta) - f_1(iy+\zeta)|\}.$$

The r.h.s of this inequality tends to zero as $n \to \infty$ in view of the hypothesis of Theorem 2.1 and Remark 2 after Proposition 4.1. Thus, there exist $0 < y_1 < y_2 < \infty$ such that for all $y \in [y_1, y_2]$, $\lim_{n\to\infty} \mathbf{E}\{|\varepsilon_{1,n}(y)|\} = 0$. Analogous arguments show that $\lim_{n\to\infty} \mathbf{E}\{|\varepsilon_{2,n}(y)|\} = 0$, where $\varepsilon_{2,n}(y)$ is defined in (4.17) and in (4.18) where the indices 1 and 2 are interchanged. Thus we have

$$\lim_{n\to\infty} \mathbf{E}\{|\varepsilon_{r,n}(y)|\} = 0, \ r = 1, 2. \tag{4.23}$$

Consider now the first term in the l.h.s. of (4.16). In view of (2.5) we can write this term in the form

$$[f_2(iy - t_{1,n}(y)) - f_2(iy - t_1(y))] = -\frac{\langle \delta_{1,n}\rangle}{f\langle g_n\rangle} I_2 \gamma_n + \frac{iy}{f} I_2 \gamma_{1,n} = -a_1\gamma_n + b_1\gamma_{1,n}, \tag{4.24}$$

where I_2, a_1 and b_1 are defined by formulas (3.15) and (3.16), in which we have to replace Δ_1', Δ_1'', f' and f'' by Δ_1, $\langle \delta_{1,n}\rangle$, f and $\langle g_n\rangle$ respectively. Denote by $\Phi = \{\Phi_{ij}\}_{i,j=1}^3$ the matrix defined by the l.h.s. of system (3.14) and by $\Gamma = \{\Gamma_i\}_{i=1}^3$ the vector with components $\Gamma_1 = \gamma_n$, $\Gamma_2 = \gamma_{1,n}$, $\Gamma_3 = \gamma_{2,n}$. Then we have from (4.16), (4.23) and (4.24),

$$\mathbf{E}\{|(\Phi\Gamma)_1|\} \le \mathbf{E}\{|\varepsilon_{1,n}|\}. \tag{4.25}$$

Interchanging in the above argument indices 1 and 2 we obtain also that

$$\mathbf{E}\{|(\Phi\Gamma)_2|\} \le \mathbf{E}\{|\varepsilon_{2,n}|\}. \tag{4.26}$$

Besides, applying to the identity $G(z)(H_1 + H_2 - z) = 1$ the operation $\langle n^{-1}Tr \dots \rangle$ and subtracting from the result the third equation of system (2.18), we obtain one more relation,

$$\mathbf{E}\{|(\Phi\Gamma)_3|\} = 0. \tag{4.27}$$

It follows from the proof of Proposition 3.3 that the matrix Φ is invertible if y_1 is big enough. Denote by $||\dots||_1$ the l^1-norm of $\mathbb{C}^3$ and by $||\dots||$ the induced matrix norm. Then we have

$$\mathbf{E}\{||\Gamma||_1\} \le \mathbf{E}\{||\Phi^{-1}\Phi\Gamma||_1\} \le \mathbf{E}^{1/2}\{||\Phi^{-1}||^2\}\mathbf{E}^{1/2}\{||\Phi\Gamma||_1^2\}. \tag{4.28}$$

It follows from our arguments above that all entries of the matrices Φ and Φ^{-1} and all components of the vector Γ are bounded uniformly in n and in realizations of random matrices A_n,B_n, U_n and V_n in (2.1). Thus we have

$$||\Phi^{-1}|| \le \sum_{i,j=1}^{3} |(\Phi^{-1})_{ij}| \le C, \;\; ||\Phi\Gamma||_1 \le \sum_{i,j=1}^{3} |\Phi_{ij}||\Gamma|_j \le C.$$

These bounds and (4.25)–(4.28) imply that

$$\mathbf{E}\{||\Gamma||_1\} \le C^{3/2}(\mathbf{E}\{|\varepsilon_{2,n}|\} + \mathbf{E}\{|\varepsilon_{2,n}|\})^{1/2}.$$

In view of (4.23) this inequality implies (4.14), i.e. the assertion of the lemma. □

Now we extend the result of Lemma 4.1 for the case of unbounded A_n and B_n, having the limiting NCM's with the finite first moments. We will apply the truncation technique standard in probability, whose random matrix version was used already in [16, 19].

Proof of Theorem 2.1. Without loss of generality we can assume that

$$\sup_n \int |\lambda|\mathbf{E}\{N_{1,n}(\mathrm{d}\lambda)\} \le m_1 < \infty. \tag{4.29}$$

For any $T > 0$ introduce the matrices A_n^T and B_n^T replacing eigenvalues A_n and B_n lying in $]T, \infty[$ by T and eigenvalues lying in $]-\infty, -T]$ by $-T$. Denote by $N_{r,n}^T$, $r = 1, 2$ the NCM of A_n^T and B_n^T. It is clear that for any $T > 0$ and $r = 1, 2$, the sequence $\{N_{r,n}^T\}_{n\ge1}$ converges weakly in probability to the measures N_r^T as $n \to \infty$, where N_r^T are analogously defined via N_r and have their supports in $[-T, T]$, and that for each $r = 1, 2$ the sequence $\{N_r^T\}_{T\ge1}$ converges weakly to N_r as $T \to \infty$. Denote by N_n^T, $r = 1, 2$ the NCM of $H_n^T = H_{1,n}^T + H_{2,n}^T = V_n^* A_n^T V_n + U_n^* B_n^T U_n$. According to linear algebra, if M_r, $r = 1, 2$ are two Hermitian $n \times n$ matrices, then

$$\mathrm{rank}(M_1 + M_2) \le \mathrm{rank} M_1 + \mathrm{rank} M_2, \tag{4.30}$$

and if $\{\mu_{r,l}\}_{l=1}^{n}$, $r=1,2$ are eigenvalues of M_r, $r=1,2$, then for any Borel set $\Delta\in\mathbb{R}$,

$$|\#\{\mu_{1,l}\in\Delta\}-\#\{\mu_{2,l}\in\Delta\}|\le \mathrm{rank}(M_1-M_2).$$

By using these facts we find that

$$\begin{aligned}|N_n(\Delta)-N_n^T(\Delta)| &\le \frac{1}{n}\mathrm{rank}(H_n-H_n^T)\le \frac{1}{n}\mathrm{rank}(A_n-A_n^T)\\ &+\frac{1}{n}\mathrm{rank}(B_n-B_n^T)\le N_{1,n}(\mathbb{R}\backslash]-T,T[)+N_{2,n}(\mathbb{R}\backslash]-T,T[),\end{aligned}\tag{4.31}$$

valid for any Borel set $\Delta\in\mathbb{R}$. As a result, the Stieltjes transform g_n^T of N_n^T and the Stieltjes transform g_n of N_n are related as follows:

$$|g_n^T(z)-g_n(z)|\le\frac{\pi}{|\mathrm{Im}z|}\left(N_{1,n}(\mathbb{R}\backslash]-T,T[)+N_{2,n}(\mathbb{R}\backslash]-T,T[)\right),$$

hence

$$\mathbf{E}\{|g_n^T(z)-g_n(z)|\}\le\frac{\pi}{|\mathrm{Im}z|}\left(\mathbf{E}\{N_{1,n}(\mathbb{R}\backslash]-T,T[)\}+\mathbf{E}\{N_{2,n}(\mathbb{R}\backslash]-T,T[)\}\right),\tag{4.32}$$

and

$$\lim_{n\to\infty}\mathbf{E}\{N_{r,n}(\mathbb{R}\backslash]-T,T[)\}\le 1-N_r(]-T,T[)=o(1),\quad T\to\infty.$$

Since the norms of matrices H_1^T and H_2^T are bounded, the results of Lemma 4.1 are applicable to the function $g_n^T(z)$, so that, in particular, for any non-real z it converges in probability as $n\to\infty$ to a function $f^T(z)$ satisfying the system

$$\begin{aligned}f^T(z)&=f_1^T\left(z-\frac{\Delta_2^T(z)}{f^T(z)}\right),\\ f^T(z)&=f_2^T\left(z-\frac{\Delta_1^T(z)}{f^T(z)}\right),\\ f^T(z)&=\frac{1-\Delta_1^T(z)-\Delta_2^T(z)}{-z}.\end{aligned}$$

In addition, since $\mathbf{E}\{g_n^T(z)\}$ and $\mathbf{E}\{\delta_{1,n}^T(z)\}$ are bounded uniformly in n and T for $z\in E(y_0)$:

$$\begin{aligned}|\mathbf{E}\{g_n^T(z)\}|&\le\frac{1}{y_0},\\ |\mathbf{E}\{\delta_{1,n}^T(z)\}|&\le\frac{1}{y_0}\int|\lambda|\mathbf{E}\{N_{1,n}^T(d\lambda)\}\le\frac{1}{y_0}\int|\lambda|\mathbf{E}\{N_{1,n}(d\lambda)\}\le\frac{m_1}{y_0},\end{aligned}$$

we have

$$|f^T(z)|\le\frac{1}{y_0},\ |\Delta_1^T(z)|\le\frac{m_1}{y_0}.\tag{4.33}$$

Thus, there exists a sequence $T_k \to \infty$ such that sequences of analytic functions $\{f^{T_k}(z)\}$ and $\{\Delta_1^{T_k}(z)\}$ converge uniformly on any compact set of the $E(y_0)$ of (4.32). In addition, the measures $N_r^{T_k}$, $r = 1, 2$ converge weakly to the limiting measures N_r, $r = 1, 2$. Hence, there exist three analytic functions $f(z)$, $\Delta_1(z)$ and $\Delta_2(z) = zf(z) + 1 - \Delta_1(z)$ verifying (2.18). Besides, because of (4.33) and (3.1) for $z \in E(y_0)$ we have

$$|\Delta_1(z)| \leq \frac{m_1}{y_0}, \text{ and } \Delta_2(z) = o(1) \text{ as } y_0 \to \infty.$$

As a result of the relations above, $f(z)$ and $\Delta_r(z)$, $r = 1, 2$ satisfy the conditions of Proposition 3.3, hence they are defined uniquely.

Furthermore, we have

$$\mathbf{E}\{|g_n(z) - f(z)|\} \leq \mathbf{E}\{|g_n(z) - g_n^{T_k}(z)|\} + \mathbf{E}\{|g_n^{T_k}(z) - f^{T_k}(z)|\} + |f^{T_k}(z) - f(z)|.$$

Hence in view of (4.32), of the arguments above on the convergence of f^{T_k} to f, and of Lemma 4.1 we conclude that for each $z \in E(y_0)$,

$$\lim_{n\to\infty} \mathbf{E}\{|g_n(z) - f(z)|\} = 0.$$

In view of Proposition 4.1 this implies that the NCM (2.2) of random matrices (2.1) converges weakly in probability as $n \to \infty$ to the non-random measure, whose Stieltjes transform is a unique solution of system (2.18). □

5. Properties of the Solution

Here we will consider several simple properties of the limiting eigenvalue counting measure described by Theorem 2.1, i.e. the measure whose Stieltjes transform is a solution of (2.18) satisfying (2.12)–(2.14). We refer the reader to works [31,2,4,3] and references therein for a rather complete collection of results on properties of the measure, resulting from the binary operation in the space of the probability measures, defined by a version of system (2.18). This binary operation is called free additive convolution.

(i) *Assume that the supports of the limiting eigenvalue measures of the matrices* A_n *and* B_n *are bounded, i.e. there exist* $-\infty < a_r, b_r < \infty$, $r = 1, 2$, *such that*

$$\text{supp } N_r \subset [a_r, b_r], r = 1, 2. \tag{5.1}$$

Then

$$\text{supp } N \subset [a_1 + a_2, b_1 + b_2]. \tag{5.2}$$

Proof. Denote by $\{\lambda_l\}_{l=1}^n$ and by $\{\lambda_{r,l}\}_{l=1}^n$, $r = 1, 2$ eigenvalues of H_n and $H_{r,n}$ in (2.1) respectively. Then, according to the linear algebra (cf.(4.31)),

$$\#\{\lambda_l \in \mathbb{R}\backslash[a_1 + a_2, b_1 + b_2]\} \leq \#\{\lambda_{1,l} \in \mathbb{R}\backslash[a_1, b_1]\} + \#\{\lambda_{2,l} \in \mathbb{R}\backslash[a_2, b_2]\}.$$

In view of Theorem 2.1 and (5.1) this leads to the relation $N(\mathbb{R}\setminus\sigma) = 0$, i.e. to (5.2).

(ii) *Examples.* 1. Consider the case when $A_n = B_n$, i.e. $N_1 = N_2$. In this case system (2.18) will have the form

$$f(z) = f_1\left(\frac{z}{2} - \frac{2}{f(z)}\right). \tag{5.3}$$

Take $N_1 = N = \alpha\,\delta_0 + (1-\alpha)\,\delta_a$, where $0 \le \alpha \le 1$, $a > 0$ and δ_λ is the unit measure concentrated at $\lambda \in \mathbb{R}$. Then

$$f_1(z) = \frac{-\alpha}{z} + \frac{1-\alpha}{a-z}$$

and (2.18) reduces to the quadratic equation

$$z(z-2a)f^2 + 2a(1-2\alpha)f - 1 = 0,$$

whose solution satisfying (2.12) - (2.14) is

$$f(z) = \frac{-a(1-2\alpha) - \sqrt{(z-\lambda_+)(z-\lambda_-)}}{z(z-2a)},\ \ \lambda_\pm = a(1 \pm 2\sqrt{\alpha(1-\alpha)}).$$

By using (2.15) we find that the limiting measure in this case has the form

$$N = (2\alpha-1)_+\delta_0 + (1-2\alpha)_+\delta_{2a} + N^*, \tag{5.4}$$

where $x_+ = \max(0, x)$, and

$$N^*(\mathrm{d}\lambda) = \frac{1}{\pi}\frac{\sqrt{(\lambda_+ - \lambda)(\lambda - \lambda_-)}}{\lambda(\lambda - 2a)}\chi_{[0,2a]}(\lambda)\mathrm{d}\lambda \tag{5.5}$$

is the absolute continuous measure of the mass $1 - 2\alpha$. Here $\chi_\Delta(\lambda)$ is the indicator of the set $\Delta \subset \mathbb{R}$. In the cases $\alpha = 0, 1$ (5.4) is δ_{2a} and δ_0 respectively, and in the case $\alpha = 1/2$ (5.4) has no atoms, but only the square root singularities

$$N^*(\mathrm{d}\lambda) = \frac{1}{\pi\sqrt{\lambda(2a-\lambda)}}\chi_{[0,2a]}(\lambda)\mathrm{d}\lambda. \tag{5.6}$$

Formulas (5.3)–(5.6) show that:

- the result (5.2) is optimal with respect to the endpoints of the measures N_r, $r = 1, 2$ and N;
- in the case when $N_1 = N_2$ have atoms of the mass $\mu > 1/2$ at the same point then the measure N has also an atom of the mass $(2\mu - 1)$ (for general results of this type see [3]).

However, in general the support of N is strictly included in the sum of supports of measures N_r, $r = 1, 2$, i.e. the inclusion in the r.h.s part of (5.3) is strict. This can be illustrated by the following two examples.

2. Take again $N_1 = N_2$, where now

$$N_1(\mathrm{d}\lambda) = \frac{1}{\pi\sqrt{a^2 - \lambda^2)}}\chi_{[-a,a]}(\lambda)\mathrm{d}\lambda$$

is the arcsin law. This measure corresponds to the matrix ensemble (2.37) with

$$V(\lambda) = \begin{cases} 0, & |\lambda| < 1, \\ \infty, & |\lambda| > 1. \end{cases} \tag{5.7}$$

In this case the equation in (5.3) is again quadratic and leads to

$$N(\mathrm{d}\lambda) = \frac{\sqrt{3a^2 - \lambda^2)}}{\pi(4a^2 - \lambda^2)} \chi_{[-\sqrt{3}a, \sqrt{3}a]}(\lambda)\mathrm{d}\lambda.$$

3. In the next example we take

$$N_r(\mathrm{d}\lambda) = \frac{1}{4\pi a_r^2}\sqrt{8a_r^2 - \lambda^2}\chi_{[-2\sqrt{2}a_r, 2, \sqrt{2}a_r]}(\lambda)\mathrm{d}\lambda, r = 1, 2,$$

i.e. both measures are the semicircle laws (2.31). Then it is easy to find that N is also the semicircle measure with the parameter $a^2 = a_1^2 + a_2^2$. This case was indicated in [19]. It can be easily deduced from the law of addition of the R-transforms of Voiculescu [31], because in this case $R_r(f) = 2a_r^2 f$. For further properties of the measure N in the case when one of $N_r, r = 1, 2$ is the semicircle law see [14,4].

(iii) *Suppose that one of the measures* $N_r(\mathrm{d}\lambda), r = 1, 2$ *is absolutely continuous with respect to the Lebesgue measure, i.e., say,* $N_1(\mathrm{d}\lambda) = \rho_1(\lambda)\mathrm{d}\lambda,$ *and*

$$\overline{\rho}_1 = \operatorname*{ess\,sup}_{\lambda\in\mathbb{R}} |\rho_1(\lambda)| < \infty.$$

Then N is also absolutely continuous with respect to the Lebesgue measure, i.e.$N(\mathrm{d}\lambda) = \rho(\lambda)\mathrm{d}\lambda,$ *and*

$$\operatorname*{ess\,sup}_{\lambda\in\mathbb{R}} |\rho_1(\lambda)| = \overline{\rho}_1 < \infty. \tag{5.8}$$

Proof. Indeed, since the function $z_1^* = z - \Delta_{2,1}/f(z)$ is analytic for non-real z, the number of its zeros in any compact set of $\mathbb{C}\backslash\mathbb{R}$ is finite. Thus, for any $\lambda \in \mathbb{R}$ there exists a sequence $\{z_n\}$ of non-real numbers such that $z_n \to \lambda$ as $n \to \infty$ and Im $z_n^* \neq 0$. Hence, we have from the first equation of system (2.18) for $z_n^* = \lambda_n^* + i\varepsilon_n^*$,

$$\frac{1}{\pi}\mathrm{Im} f(z) = \frac{1}{\pi}\int \frac{\varepsilon_r^*\rho_r(\mu)\mathrm{d}\mu}{(\mu - \lambda_r^*)^2 + (\varepsilon_r^*)^2} \leq \overline{\rho}_1 \frac{1}{\pi}\int \frac{\varepsilon_r^*\mathrm{d}\mu}{(\mu - \lambda_r^*)^2 + (\varepsilon_r^*)^2} = \overline{\rho}_1.$$

This relation and the inversion formula (2.15) yield (5.8). For more general results in this direction see the recent paper [3]. □

6. Discussion

In this section we comment on several topics related to those studied above.

1. In this paper we deal with Hermitian and unitary matrices, i.e. we assume that the matrices A_n and B_n in (2.1) are Hermitian and U_n and V_n are unitary. It is natural also to consider the case of real symmetric A_n and B_n and orthogonal U_n and V_n. This case can be handled by using the analogue of formula (3.11) of the orthogonal group $O(n)$. Indeed, it is easy to see that this analogue has the form

$$\int_{O(n)} \Phi^{'}(O^T MO) \cdot [X, O^T MO]\mathrm{d}O = 0,$$

where O^T is transposed to O and X is a real symmetric matrix. By using this formula we obtain instead of (3.23),

$$\langle G_{aa}(H_2G)_{bc}\rangle + \langle G_{ab}(H_2G)_{ac}\rangle = \langle (GH_2)_{aa}G_{bc}\rangle + \langle (GH_2)_{ab}G_{bc}\rangle.$$

The second terms in both sides of this formula give two additional terms

$$-n^{-1}G^T H_2G + n^{-1}H_2G^T G$$

(cf. (3.40)). These terms, however, produce the asymptotically vanishing contribution because, in view of (3.3), (3.6) and (3.37), we have

$$\left| n^{-2}\langle \mathrm{Tr}(1 + \Delta_{2,n} f_n^{-1}G_1)^{-1}G_1(-G^T H_2G + H_2G^T G)\rangle \right| \leq \frac{2}{ny_0^3} m_4^{1/4}.$$

Similar and also negligible as $n \to \infty$ terms appear in analogues of formulas (3.53), (3.69) and (3.73) of the proof of Theorem 3.2. As the result, we obtain in this case the same system (2.18), defining the Stieltjes transform of the limiting eigenvalue counting measure of the analogue of (2.1) with the real symmetric A_n and B_n and orthogonal Haar-distributed U_n and V_n.

2. As was mentioned in the Introduction, our main result, Theorem 2.1, can be viewed as an extension of the result by Speicher [26], obtained by the moment method and valid for uniformly in n bounded matrices A_n and B_n in (2.1). Both results are analogues for randomly rotated matrices of old results of [16,19] (see (2.24) and (2.33)) on the form of the limiting eigenvalue counting measure of the sum of an arbitrary matrix and certain random matrices (see (2.20) and (2.26)), in particular, Gaussian random matrices (2.28). In this case, however, there exists another model, proposed by Wegner [32] that combines properties of random matrices, having all entries roughly of the same order, and of random operators, whose entries decay sufficiently fast in the distance from the principal diagonal (see e.g. [22]). A simple, but rather non-trivial version of the Wegner model corresponds to the selfadjoint operator H acting in $l^2(\mathbb{Z}^d) \times \mathbf{C}^n$ and defined by the matrix

$$H(x, j; y, k) = v(x - y)\delta_{jk} + \delta(x - y) f_{jk}(x), \tag{6.1}$$

where $x, y \in \mathbb{Z}^d$, $j, k = 1, \dots, n$, $\delta(x)$ is the d-dimensional Kronecker symbol,

$$v(-x) = \overline{v}(x), \quad \sum_{x\in\mathbb{Z}^d} |v(x)| < \infty, \tag{6.2}$$

and $f(x) = \{f_{jk}(x)\}_{j,k=1}^n$, $x \in \mathbb{Z}^d$ are independent for different x and identically distributed $n \times n$ Hermitian matrices, whose distribution for any x is given by (2.28). According to [32] (see also [14]) asymptotic for $n \to \infty$ properties of operator (6.1) resemble, in many aspects, asymptotic properties of matrices (2.28). The "free" analogue of the Wegner model was proposed in [18]. In this case i.i.d. matrices $f(x)$ have the form

$$f(x) = U_n^*(x) B_n U_n(x), \tag{6.3}$$

where B_n is as in (2.1) and $U_n(x)$, $x \in \mathbb{Z}^d$ are i.i.d. unitary $n \times n$ matrices whose distribution is given by the Haar measure on $U(n)$. By using a version of the moment method, similar to that of paper [26], or, rather, its formal scheme, the authors derived the limiting form of

$$\mathbf{E}\left\{n^{-1}\sum_{j=1}^{n} G(x, j; y, j)\right\},$$

where $G(x, j; y, k)$ is the matrix (the Green function) of the resolvent $(H - z)^{-1}$ of (6.1)–(6.3). The authors also found a certain second moment of the Green function. This moment is necessary to compute the a.c. conductivity via the Kubo formula. Because of the moment method results of [18] are valid for uniformly bounded in n matrices B_n in (6.3), similar to results for matrices (2.1) obtained in [26]. By using a natural extension of the differentiation formula (3.11) and the technique developed in [14] to analyze the Wegner model, the results of paper [18] can be extended to the case of arbitrary matrices B_n in (6.3), because in this case the role of condition (2.17) of Theorem 2.1) plays condition (6.2).

3. As was mentioned before asymptotic properties of random matrices are of considerable interest in certain branches of operator algebra theory and in the related branch of non-commutative probability theory, known as free probability (see [28, 31, 30] and references therein). Here large random matrices are an important example of the asymptotically free non-commutative random variables, providing a sufficiently rich analytic model of the abstract notion of freeness of elements of an operator algebra. The most widely used examples of asymptotically free families of non-commutative random variables are Gaussian random matrices and unitary Haar-distributed random matrices. The proof of asymptotic freeness of unitary matrices given in [28, 31] reduces to that for complex Gaussian matrices and is based on the observation that the unitary part of the polar decomposition of the complex Gaussian matrix with independent entries is the Haar-distributed unitary matrix. This method requires certain technicalities because of the singularity of the polar decomposition at zero. On the other hand, the differentiation formula (3.11) allows one to prove directly similar statements. Here is an example of results of this type (related results are proved in [35]).

Theorem 6.1. *Let k be a positive integer,* $\{T_{r,n}\}_{r=1}^k$ *be a set of* $n \times n$*matrices, such that*

$$\sup_{r \le k;\ k,l,n \in \mathbb{N}} n^{-1}\mathrm{Tr}(T_{r,n}^* T_{r,n})^l < \infty, \tag{6.4}$$

and let U_n *be the unitary and Haar-distributed random matrix. If for any* $k \in \mathbb{N}$,

$$\lim_{n\to\infty} n^{-1}\mathrm{Tr}T_{r,n} = 0, \quad r = 1, \dots, k, \tag{6.5}$$

then for any set of non-zero integers such that $\{m_r\}_{r=1}^k$, $\sum_{r=1}^k m_r = 0$,

$$\lim_{n\to\infty} \langle n^{-1}\mathrm{Tr}U_n^{m_1}T_{1,n}\dots U_n^{m_k}T_{k,n}\rangle = 0, \tag{6.6}$$

where $\langle\cdot\rangle$ *denotes the integration with respect to the Haar measure over* $U(n)$.

Remark 1. The theorem is trivially true in the case when $\sum_{r=1}^k m_r \neq 0$.

In the two subsequent lemmas we omit the subindex n.

Lemma 6.1. *Let* $\{T_i\}_{i=1}^k$ *be a set of* $n\times n$ *matrices and* U *is the Haar-distributed unitary matrix. Then for any set of non-zero integers* $\{m_i\}_{i=1}^k$, $\sum_{i=1}^k m_i = 0$ *the following identity holds:*

$$\begin{aligned} n^{-1}\mathrm{Tr}\langle U^{m_1}T_1\dots U^{m_k}T_k\rangle &= -\sum_{l_1=2}^{m_1}\langle n^{-1}\mathrm{Tr}U^{l_1-1}n^{-1}\mathrm{Tr}(U^{m_1-l_1+1}T_1\dots U^{m_k}T_k)\rangle \\ &- \sum_{r\in\{2,\dots,k\},m_r>0}\sum_{l_r=1}^{m_r}\langle n^{-1}\mathrm{Tr}(U^{m_1}T_1\dots T_{r-1}U^{l_r-1})n^{-1}\mathrm{Tr}(U^{m_r-l_r+1}T_r\dots U^{m_k}T_k)\rangle \\ &+ \sum_{r\in\{2,\dots,k\},m_r<0}\sum_{l_r=1}^{-m_r}\langle n^{-1}\mathrm{Tr}(U^{m_1}T_1\dots T_{r-1}U^{-l_r})n^{-1}\mathrm{Tr}(U^{m_r+l_r}T_r\dots U^{m_k}T_k)\rangle. \end{aligned} \tag{6.7}$$

Proof. Without loss of generality assume that $m_1 > 0$. Then, using the analogue of formula (3.11) for the average $\langle[U^{m_1}T_1\dots U^{m_k}T_k]_{ab}\rangle$, we obtain for any Hermitian X,

$$\begin{aligned} &\sum_{r\in\{1,\dots,k\},m_r>0}\sum_{l_r=1}^{m_r}\langle[U^{m_1}T_1\dots T_{r-1}U^{l_1-1}XU^{m_r-l_r+1}T_r\dots U^{m_k}T_k]_{ab}\rangle \\ &\qquad \sum_{r\in\{2,\dots,k\},m_r<0}\sum_{l_r=1}^{-m_r}\langle[U^{m_1}T_1\dots T_{r-1}U^{-l_r}XU^{m_r+l_r}T_r\dots U^{m_k}T_k]_{ab}\rangle = 0. \end{aligned} \tag{6.8}$$

Choosing as X the Hermitian matrix having only (c,d)th and (d,c)th non-zero entries, setting then $a = c$ and $b = d$ and applying to the result the operation $n^{-2}\sum_{a,b}$, we obtain (6.7). □

Lemma 6.2. *Under the conditions (6.4) and (6.5) the variance* $D = \langle|\xi^\circ|^2\rangle$ *of the random variable*

$$\xi = n^{-1}\mathrm{TrL},\ \mathrm{L} = U^{m_1}T_1\dots U^{m_k}T_k \tag{6.9}$$

is of the order n^{-2} *as* $n\to\infty$.

Proof. Using the same technique as that in Lemma 6.1 for $\langle \overline{L_{ab}} L_{cd}\rangle$ we obtain the relation

$$\begin{aligned}
D = &-\sum_{l_1=2}^{m_1} \langle \overline{\xi}^{\circ} n^{-1}\mathrm{Tr} U^{l_1-1} n^{-1}\mathrm{Tr}(U^{m_1-l_1+1}T_1\dots U^{m_k}T_k)\rangle \\
&- \sum_{r\in\{2,\dots,k\}, m_r>0}\sum_{l_r=1}^{m_r} \langle \overline{\xi}^{\circ} n^{-1}\mathrm{Tr}(U^{m_1}T_1\dots T_{r-1}U^{l_r-1}) n^{-1}\mathrm{Tr}(U^{m_r-l_r+1}T_r\dots U^{m_k}T_k)\rangle \\
&+ \sum_{r\in\{2,\dots,k\}, m_r<0}\sum_{l_r=1}^{-m_r} \langle \overline{\xi}^{\circ} n^{-1}\mathrm{Tr}(U^{m_1}T_1\dots T_{r-1}U^{-l_r}) n^{-1}\mathrm{Tr}(U^{m_r+l_r}T_r\dots U^{m_k}T_k)\rangle \\
&+ n^{-2}\Phi,
\end{aligned} \tag{6.10}$$

where

$$\begin{aligned}
\Phi = &-\sum_{r\in\{1,\dots,k\}, m_r>0}\sum_{l_r=1}^{m_r} n^{-1}\mathrm{Tr}\langle (U^{m_r-l_r+1}T_r\dots T_k U^{m_1}T_1\dots T_{r-1}U^{l_r-1})^* L\rangle \\
&+ \sum_{r\in\{2,\dots,k\}, m_r<0}\sum_{l_r=1}^{-m_r} n^{-1}\mathrm{Tr}\langle (U^{m_r+l_r}T_r\dots T_k U^{m_1}T_1\dots T_{r-1}U^{-l_r})^* L\rangle.
\end{aligned}$$

We have obviously for $k = m = 1$,

$$\langle n^{-1}\overline{\mathrm{Tr}(UT)^{\circ}} n^{-1}\mathrm{Tr}(UT)\rangle \le \frac{1}{n^2} n^{-1}\mathrm{Tr}(TT^*).$$

We proceed further by induction. In view of condition (6.4) and Proposition 3.1 we have the bound

$$|n^{-1}\mathrm{Tr}(U^{m_1}T_{r_1}\dots U^{m_p}T_{r_p})| \le C^2_, \tag{6.11}$$

where C may depend only on p. Now, since $n^{-1}\mathrm{Tr}\langle U^l\rangle = 0, l \neq 0$, the summands of the first term in r.h.s. of (6.10) can be estimated as follows:

$$\left|\langle \overline{\xi}^{\circ} n^{-1}\mathrm{Tr}(U^{l_1}) n^{-1}\mathrm{Tr}(U^{m_1-l_1+1}T_1\dots U^{m_k}T_k)\rangle\right| \le C\sqrt{D}\sqrt{\langle |n^{-1}\mathrm{Tr}(U^{l_1})^{\circ}|^2\rangle}. \tag{6.12}$$

Likewise, by using the cyclic property of the trace, the identity $\langle a^{\circ}bc\rangle = \langle a^{\circ}b^{\circ}c\rangle + \langle a^{\circ}c^{\circ}\rangle\langle b\rangle$, Schwarz inequality, and (6.11), we obtain for the second term in the right-hand side of (6.10) the following estimates for $r \ge 2$:

$$\begin{aligned}
&\left|\langle \overline{\xi}^{\circ} n^{-1}\mathrm{Tr}(U^{m_1}T_1\dots T_{r-1}U^{l_r-1}) n^{-1}\mathrm{Tr}(U^{m_r-l_r+1}T_r\dots U^{m_k}T_k)\rangle\right| \\
&\quad \le C\sqrt{D}\left\{\sqrt{\langle |n^{-1}\mathrm{Tr}(U^{m_1+l_r-1}T_1\dots U^{m_{r-1}}T_{r-1})^{\circ}|^2\rangle}\right. \\
&\qquad \left. +\sqrt{\langle |n^{-1}\mathrm{Tr}(U^{m_r-l_r+1}T_r\dots U^{m_k}T_k)^{\circ}|^2\rangle}\right\}.
\end{aligned} \tag{6.13}$$

The third term in the right-hand side of (6.10) can be estimated analogously. The fourth term is of the order $1/n^2$ in view of (6.9). By the induction hypothesis the expectations under square roots in the r.h.s. of (6.13) and (6.12) are of the order n^{-2}. This leads to the inequality

$$D \leq \frac{C_1}{n}\sqrt{D} + \frac{C_2}{n^2},$$

where C_1 and C_2 are independent of n. This implies the bound $D = O(n^{-2})$. □

Proof of Theorem 6.1. We use Lemma 6.1 and again the induction. We have first

$$n^{-1}\mathrm{Tr}\langle U^m T_1 U^{-m} T_2\rangle = n^{-1}\mathrm{Tr}T_1 n^{-1}\mathrm{Tr}T_2 = 0.$$

In the general case we use Lemma 6.2 to factorize asymptotically the moments in the r.h.s. of (6.7). In the resulting relation the expressions $n^{-1}\mathrm{Tr}\langle U^{m_{r_1}} T_{r_1} \ldots U^{m_{r_s}} T_{r_s}\rangle$ are zero for any collection $(T_{r_1}, \ldots, T_{r_s})$ and any n, if $\sum_{i=1}^{s} m_{r_i} \neq 0$, and tend to zero as $n \to \infty$ if $\sum_{i=1}^{s} m_{r_i} = 0$ in view of the induction hypothesis and condition (6.5). This leads to (6.6). □

Remark 2. A simple version of the above arguments allows us to prove that the normalized counting measure of the Haar distributed unitary matrices converges with probability one to the uniform distribution on the unit circle. Indeed, consider again the Stieltjes transform g_n of this measure, supported now on the unit circle. By the spectral theorem for unitary matrices we have

$$g_n(z) = n^{-1}\mathrm{Tr}G(z),\ G(z) = (U - z)^{-1},\ |z| \neq 1. \tag{6.14}$$

We can then obtain the following identities:

$$\langle \mathrm{Tr}G^2(z)U\rangle = 0,\ \langle g_n(z)n^{-1}\mathrm{Tr}G(u)U\rangle = 0, \tag{6.15}$$

$$\langle g_n(z_1)n^{-1}\mathrm{Tr}G(z_1)Ug(z_2)\rangle + \langle n^{-3}\mathrm{Tr}G(z_1)G(z_2)UG(z_2)\rangle = 0. \tag{6.16}$$

By using the obvious relations

$$G'(z) = G^2(z),\ G(0) = U^{-1},\ G(\infty) = 0,$$

we obtain from the first of identities (6.15)

$$f_n(z) \equiv \langle g_n(z)\rangle = \begin{cases} 0, & |z| < 1 \\ -z^{-1}, & |z| > 1. \end{cases}$$

This relation shows that the expectation of the normalized counting measure of U is the uniform distribution on the unit circle, the fact that follows easily from the shift invariance of the Haar measure. Now the second identity (6.15) and (6.16) lead to the bound

$$|\langle g_n(z)\rangle|^2 \leq \frac{C(r_0)}{n^2},\ |z| \leq r_0,$$

where $C(r_0)$ is independent of n and finite if r_0 is small enough. This bound and arguments analogous to those used in the proof of Theorem 3.1 imply that the normalized eigenvalue counting measure of unitary Haar distributed random matrices converges with probability one to the uniform distribution on the unit circle. This fact as well as the analogous fact for the orthogonal group can be deduced from the works by Dyson (see e.g. [17]), where the joint probability distribution of all n eigenvalues of the Haar distributed unitary or orthogonal matrices was found and studied. This technique is more powerful but also more complex than that used above and based on rather elementary means.

Acknowledgements. V. Vasilchuk is thankful to the Laboratoire de Physique Mathématique et Géométrie de l'Université, Paris-7, for hospitality and to the Ministère des Affaires Etrangères de France for the financial support.

References

1. Akhiezer, N.I., Glazman, I.M.: *Theory of Linear Operators in Hilbert Space*. New York: Frederick Ungar, 1963
2. Bercovici, H., Voiculescu, D.: Free convolution of measures with unbounded support. Indiana University Math. J. **42**, 733–773 (1993)
3. Bercovici, H.,Voiculescu, D.: Regularity questions for free convolution of measures. In: Bercovici, H., et al. (eds.) *Nonselfadjoint Operator Algebras, Operator Theory, and Related Topics*. Basel: Birkhäuser, 1998, pp. 37–47
4. Biane, P.: On the free convolution with a semicircular distribution. Indiana Univ. Math. J. **46**, 705–717 (1997)
5. Bessis, B., Itzykson, C., Zuber, J.-B.: Quantum field theory technique in graphical enumeration. Adv. Appl. Math. **1**, 109–157 (1980)
6. Boutet de Monvel, A., Pastur, L., Shcherbina, M.: On the statistical mechanics approach in the random matrix theory: Integrated density of states. J. Stat. Phys. **79**, 585–611 (1995)
7. Brody, T.A., French, J., Melo, P., Pandy, A., Wong S.: Random matrix physics: Spectrum and strength fluctuations. Rev. Mod. Phys. **53**, 385–480 (1981)
8. Fulton, W.: Eigenvalues, invariant factors, highest weights, and Schubert calculus. Bull. Math. Soc. **37**, 200–249 (2000)
9. Girko, V.L.: *Random Matrices (Sluchainye matricy)*. Kiev: Vyshcha Shkola, 1975 (in Russian)
10. Girko, V.L.: *Theory of Random Determinants*. Dordrecht: Kluwer Academic Publishers, 1990
11. Khorunzhenko, B., Khorunzhy, A., Pastur, L.: Asymptotic properties of large random matrices with independent entries. J. Math. Phys. **37**, 5033–5060 (1996)
12. Khoruzhenko, B., Khorunzhy, A., Pastur, L., Shcherbina, M.: Large-n limit in the statistical mechanics and the spectral theory of disordered systems. In: C. Domb, C., Lebowitz, J. (eds.), *Phase Transitions and Critical Phenomena*. **15**, New-York: Academic Press, 1992, pp. 74–239
13. Khorunzhy, A.: Eigenvalue distribution of large random matrices with correlated entries. Mat. Phyz. Anal. Geom. **3**, 80–101 (1996)
14. Khorunzhy, A., Pastur, L.: Limits of infinite interaction radius, dimensionality and the number of components for random operators with off-diagonal randomness. Commun. Math. Phys. **153**, 605–646 (1993)
15. Loeve, M.: *Probability Theory*. Berlin: Springer, 1977
16. Marchenko, V.A., Pastur, L.A.: Distribution of eigenvalues for some sets of random matrices. Math. USSR, Sb. **1**, 457–483 (1967)
17. Mehta, M.: *Random Matrices*. Boston: Academic Press, 1991
18. Neu, P., Speicher, R.: Rigorous mean field model for CPA: Anderson model with free random variables. J. Stat. Phys. **80**, 1279–1308 (1995)
19. Pastur, L.A.: On the spectrum of random matrices. Teor. Math. Phys. **10**, 67–74 (1972)
20. Pastur, L.: Eigenvalue distribution of random matrices. Annales l'Inst. H.Poincare **64**, 325–337 (1996)
21. Pastur, L.: A simple approach to global regime of the random matrix theory. In: Miracle-Sole, S., Ruiz, J., Zagrebnov V. (eds.). *Mathematical Results in Statistical Mechanics*. Singapore: World Scientific, 1999, pp. 429–454
22. Pastur, L., Figotin, A.: *Spectra of Random and Almost Periodic Operators*. Berlin: Springer, 1992
23. Pastur, L., Shcherbina, M.: Universality of the local eigenvalue statistics for a class of unitary invariant random matrix ensembles. J. Stat. Phys. **86**, 109–147 (1997)
24. Silverstein, J.: Strong convergence of the empirical distribution of eigenvalues of large dimensional random matrices. J. Multivariate Anal. **55**, 331–339 (1995)
25. Silverstein, J.W., Choi, S.I.: Analysis of the limiting spectral distribution of large dimensional random matrices. J. Multivariate Anal. **54**, 295–309 (1995)

26. Speicher, R.: Free convolution and the random sum of matrices. Publ. Res. Inst. Math. Sci. **29**, 731–744 (1993)
27. Vasilchuk, V.: On the law of multiplication of random matrices. Submitted to the Math. Physics, Analysis and Geometry
28. Voiculescu, D.V.: Limit laws for random matrices and free products. Invent. Math. **104**, 201–220 (1991)
29. Voiculescu, D.V.(ed.).: *Free Probability Theory*. Providence: American Mathematical Society, 1997
30. Voiculescu, D.: A strengthened asymptotic freeness result for random matrices with applications to free entropy. Int. Math. Res. Not. **1**, 41–62 (1998)
31. Voiculescu, D.V., Dykema, K.J., Nica, A.: *Free Probability Theory. A Noncommutative Probability Approach to Free Products with Applications to Random Matrices, Operator Algebras and Harmonic Analysis on Free Groups*. Providence: American Mathematical Society, 1992
32. Wegner, F.: Disordered systems with n-orbitals per site: $n \to \infty$ limit. Phys. Rev. **19**, 783–792 (1979)
33. Weyl, H.: Das asymptotische Verteilungsgesetz der Eigenwerte lineare partieller differential Gleichungen. Math. Ann. **71**, 441–479 (1912)
34. Wigner, E.: On the distribution of the roots of certain symmetric matrices. Ann. of Math. **67**, 325–327 (1958)
35. Xu, F.: A random matrix model from two dimensional Yang-Mills theory. Commun. Math. Phys. **190**, 287–307 (1997)
36. Zee, A.: Law of addition in random matrix theory. Nucl. Phys. **B474**, 726–744 (1996)

Communicated by Ya. G. Sinai