

Eigenvalue distribution of large dilute random matrices

A. Khorunzhy

Institute for Low Temperature Physics, Kharkov 310164, Ukraine

G. J. Rodgers

*Department of Physics, Brunel University, Uxbridge,
Middlesex UB8 3PH, United Kingdom*

(Received 28 March 1996; accepted for publication 17 January 1997)

We study the eigenvalue distribution of dilute $N \times N$ random matrices H_N that in the pure (undiluted) case describe the Hopfield model. We prove that for the fixed dilution parameter a the normalized counting function (NCF) of H_N converges as $N \rightarrow \infty$ to a unique $\sigma_a(\lambda)$. We find the moments of this distribution explicitly, analyze the $1/a$ correction, and study the asymptotic properties of $\sigma_a(\lambda)$ for large $|\lambda|$. We prove that $\sigma_a(\lambda)$ converges as $a \rightarrow \infty$ to the Wigner semicircle distribution (SCD). We show that the SCD is the limit of the NCF of other ensembles of dilute random matrices. This could be regarded as evidence of stability of the SCD to dilution, or more generally, to random modulations of large random matrices.

© 1997 American Institute of Physics. [S0022-2488(97)03106-X]

I. INTRODUCTION

Large random ($N \times N$) matrices are currently of considerable interest, mainly because of their applications in a number of different branches of theoretical physics. By having all entries of the same order, they represent an approximation to real systems and lead to exactly solvable models in the limit $N \rightarrow \infty$. Dilute random matrices, with an average of p nonzero elements per row, frequently provide an improved physical description of a real system and are often tractable in the limit of large dimension.

In this paper we study the eigenvalue distribution of dilute random matrices, which in the pure, undiluted case can be written as

$$A_N(x, y) = \frac{1}{N} \sum_{\mu=1}^m \xi^\mu(x) \xi^\mu(y), \quad x, y = \overline{1, N}, \quad (1.1)$$

where $\xi^\mu(x)$, $\mu = \overline{1, m}$, $x = \overline{1, N}$ are real independent identically distributed (i.i.d.) random variables with zero average and variance v^2 .

The matrix $A_N(x, y)$ was used in the statistical mechanics of disordered systems, where it was suggested as an interaction matrix of a simplified mean field model of a random spin system.¹ Later it was reintroduced in the neural network theory of autoassociative memory,² where the random N -dimensional vectors $\tilde{\xi}^\mu(x) = \xi^\mu(x)/N^{1/2}$ are interpreted as patterns to be memorized by the system and where the model is known as the Hopfield model.

This new field of applications created by the neural network theory has motivated a number of studies of matrices like (1) and their modifications (see, e.g., the monographs, Refs. 3–5, and references therein). Of special interest are randomly diluted versions of (1.1), which can be defined as

$$\hat{A}_N(x, y) = \sum_{\mu=1}^m \xi^\mu(x) \xi^\mu(y) d_N(\mu; x, y), \quad (1.2)$$

where d_N are independent random variables (also independent from $\{\xi^\mu(x)\}$) and take the nonzero values with probability p_N vanishing as $N \rightarrow \infty$.

Such ensembles are well known in statistical mechanics and a number of results have been obtained for disordered spin systems with (1.2) as the matrix of interactions in the Hamiltonian; see, e.g., Ref. 6. Several important particular cases of (1.2) have also been studied in neural network theory.⁷⁻¹⁰

However, the spectral characteristics of (1.2) are poorly understood. Even the simplest quantity in spectral theory of random matrices, the normalized eigenvalue counting function (NCF), has not been studied for the dilute ensemble (1.2).

For an $N \times N$ symmetric matrix A_N , the NCF can be defined as

$$\sigma(\lambda; A_N) \equiv \#\{\lambda_i^{(N)} \leq \lambda\} N^{-1}, \tag{1.3}$$

where $\lambda_i^{(N)}$ are the eigenvalues of A_N . The limit $N \rightarrow \infty$ of the NCF, if it exists, is called the integrated density of states (IDS) of matrix ensemble $\{A_N\}$.

The IDS of ensemble (1.1) was first studied by Marchenko and Pastur.¹¹ It follows from the results of Ref. 11 that the NCF of A_N (1) with i.i.d. $\xi^\mu(x)$ having zero mean and variance v^2 , converges in probability as $N, m \rightarrow \infty, m/N \rightarrow c > 0$ to a nonrandom function $\sigma(\lambda)$ of the form

$$\sigma(\lambda) = [1 - c]_+ \theta(\lambda) + \int_{-\infty}^{\lambda} \frac{[4cv^4 - (x - (c + 1)v^2)^2]_+^{1/2}}{2\pi v^2 x} dx, \tag{1.4}$$

where $[x]_+ = \max(0, x)$. In the case of $m = N$ the ensemble (1.1) represents the square of the more widely known Wigner ensemble of random symmetric matrices (see, e.g., Refs. 12, 13),

$$W_N(x, y) = \frac{1}{\sqrt{N}} w(x, y), \quad x, y = \overline{1, N}, \tag{1.5}$$

with independent (apart from a symmetry condition) identically distributed random variables with properties

$$Ew(x, y) = 0, \quad E[w(x, y)]^2 = u^2. \tag{1.6}$$

Thus, the results of Ref. 11 can be regarded as a generalization of the famous semicircle (or Wigner) law;¹⁴

$$\lim_{N \rightarrow \infty} \sigma(\lambda, W_N) = \sigma_{sc}(\lambda), \tag{1.7}$$

where

$$\sigma'_{sc}(\lambda) = \begin{cases} \frac{1}{2\pi u^2} \sqrt{4u^2 - \lambda^2}, & |\lambda| \leq 2u, \\ 0, & |\lambda| > 2u. \end{cases} \tag{1.8}$$

Spectral properties of the dilute Wigner ensemble \hat{W}_N with entries

$$\hat{W}_N(x, y) = w(x, y) d_N(x, y), \quad d_N(x, y) = d_N(y, x), \tag{1.9}$$

where $d_N(x, y)$ $x \leq y$ are independent random variables taking nonzero values with vanishing probability as $N \rightarrow \infty$, are well understood.¹⁵⁻¹⁸ In particular, it follows from the results of Refs. 17 and 18, obtained using the replica trick¹⁷ and supersymmetric methods,¹⁸ that if $w(x, y) = \pm 1$ with equal probability and

$$d_N(x,y) = \begin{cases} 1/\sqrt{p}, & \text{with probability } p/N, \\ 0, & \text{with probability } 1-p/N, \end{cases} \quad (1.10)$$

then

$$\lim_{N \rightarrow \infty} E\{\sigma(\lambda; \hat{W}_N)\} = \sigma_p(\lambda), \quad (1.11)$$

such that

$$\lim_{p \rightarrow \infty} \sigma_p(\lambda) = \sigma_{sc}(\lambda). \quad (1.12)$$

with $\sigma_{sc}(\lambda)$ given by (1.8) with $u^2 = 1$. In addition, both the $1/p$ correction to the density of states $\sigma'_p(\lambda)$ and an asymptotic estimate of the density of states for large $|\lambda|$ were found. In Ref. 16 the weak convergence of $E\{\sigma(\lambda; \hat{W}_N)\}$ to $\sigma_{sc}(\lambda)$ is proved rigorously for the ensemble (1.9), (1.10) with i.i.d. $w(x,y)$ satisfying (1.6) and having the third moment finite.

Starting from the square of the diluted matrix (1.9), we arrive at the ensemble (1.2) with

$$d_N(\mu, x, y) = d_N^\mu(x) d_N^\mu(y). \quad (1.13)$$

The IDS of this ensemble can be studied by the replica trick as in Ref. 17 or by the resolvent approach used in Ref. 16. This ensemble is discussed further in Sec. IV. However, more interesting for applications in dilution phenomena is the ensemble

$$H_N(x, y) = \sum_{\mu=1}^m \xi^\mu(x) \xi^\mu(y) a_N(x, y), \quad a_N(x, y) = a_N(y, x), \quad (1.14)$$

which cannot be related to the square of the Wigner ensemble and does not admit the direct use of the methods in Refs. 16–18.

In the present paper we study the NCF $\sigma(\lambda; H_N)$ of ensemble (1.14) with jointly independent $\{\xi^\mu(x)\}$ and $\{a_N(x, y)\}$. We refer to this ensemble as the dilute MPH (Marchenko–Pastur–Hopfield) ensemble. We assume $\xi^\mu(x)$ has zero average and finite variance v^2 and define the dilution matrix $a_N(x, y)$, in analogy with (1.10), as

$$a_N(x, y) = \frac{1}{N^\nu \sqrt{a}} \begin{cases} 1, & \text{with probability } aN^{-\alpha}, \\ 0, & \text{with probability } 1 - aN^{-\alpha}, \end{cases} \quad (1.15)$$

with some α, ν , such that $\alpha \geq 0$ and $0 \leq \nu \leq 1$. We show that if

$$\alpha + 2\nu = 2, \quad \frac{1}{2} \leq \nu < 1, \quad (1.16)$$

and

$$a_N(x, x) \equiv 0, \quad x = \overline{1, N}, \quad (1.17)$$

then $\sigma(\lambda; H_N)$ converges in probability to the semicircle distribution (1.8) (a) in the limit of infinite m and N when $\frac{1}{2} < \nu < 1$, and (b) in the limit of infinite m, N and infinite a when $\nu = 1/2$.

We prove these statements by studying the moments

$$M_{2k}^{(N)} = \int \lambda^{2k} d\sigma(\lambda; H_N) = \frac{1}{N} \sum_{x=1}^N (H^{2k})(x, x), \tag{1.18}$$

in the asymptotic limit of large m and N (and a for $\nu = \frac{1}{2}$). Using the independence of matrices $a_N(x, y)$ and $\sum_{\mu} \xi^{\mu}(x) \xi^{\mu}(y)$, we compute the mathematical expectation of $M_{2k}^{(N)}$. To do this, we combine Wigner’s original approach to the matrices $a_N(x, y)$ with a diagrammatic technique developed for dealing with matrices with the structure of $\sum_{\mu} \xi^{\mu}(x) \xi^{\mu}(y)$.

Our results show that the dilution $d_N(\mu; x, y) = a_N(x, y)$ of the MPH ensemble (1.2) makes those properties of matrices (1.1) that differ from the Wigner matrices (1.5) irrelevant. We investigate the role of the dilution parameter a in this property of the MPH ensemble. The technique we use allows us to study the NCF of (1.14) for finite a . We prove that for each fixed $a > 1$ there exists $\sigma_a(\lambda)$, which is the weak limit of $E\{\sigma(\lambda; H_N)\}$ when $m, N \rightarrow \infty$. We study the support, asymptotics for large $|\lambda|$ and $1/a$ correction of $\sigma_a(\lambda)$ and compare the results with those derived in Ref. 17 for $\sigma_p(\lambda)$ (1.11) for the dilute Wigner ensemble. We show that the difference between the dilute MPH ensemble and the dilute Wigner ensemble vanishes in the limit $N \rightarrow \infty$.

This paper is organized as follows. The remainder of this section is devoted to an explanation of conditions (1.16) and (1.17). In Sec. II we prove our main result concerning the convergence of the NCF to the semicircle distribution. In Sec. III we are concerned with the case of $\nu = \frac{1}{2}$, with finite dilution parameter a . In Sec. IV we describe different diluted random matrix ensembles and their possible generalizations. Section V is devoted to a discussion of the origin of the semicircle distribution in the ensemble (1.13).

Now let us turn to conditions (1.16) and (1.17). We can show that these conditions are necessary by considering $E H_N^2$ and $E H_N^4$, where $E(\dots) = \langle \langle \dots \rangle \rangle_{\xi}^a$. For the case of i.i.d. $\xi^{\mu}(x)$, we have

$$\begin{aligned} E H_N^2 &= \sum_{s=1}^N \sum_{\mu_1, \mu_2=1}^m \langle \xi^{\mu_1}(x) \xi^{\mu_1}(s) \xi^{\mu_2}(x) \xi^{\mu_2}(s) \rangle_{\xi} \langle a_N(x, s) a_N(s, x) \rangle_a \\ &= \frac{(N-1)m}{N^{2\nu+\alpha}} v^4 + \frac{m(m-1)}{N^{2\nu+\alpha}} v^4 + \frac{m}{N^{2\nu+\alpha}} \langle [\xi^{\mu}(x)]^4 \rangle_{\xi}, \end{aligned} \tag{1.19}$$

where $v^2 = \langle \xi^2 \rangle_{\xi}$. Thus, the first nontrivial moment,

$$E \int \lambda^2 d\sigma_N(\lambda) = \frac{1}{N} \sum_{x=1}^N E H_N^2(x, x), \tag{1.20}$$

is finite and nonzero if and only if $\alpha + 2\nu = 2$. The fourth moment $E H_4(x, x)$ includes averages

$$\begin{aligned} \sum_{\mu_j=1}^m \sum_{s=1}^N \langle \xi^{\mu_1}(x) \xi^{\mu_1}(s) \xi^{\mu_2}(s) \xi^{\mu_2}(x) \xi^{\mu_3}(x) \xi^{\mu_3}(s) \xi^{\mu_4}(s) \xi^{\mu_4}(x) \rangle_{\xi} \langle a_N(x, s) a_N(s, x) a_N(x, s) \\ \times a_N(s, x) \rangle_a \geq m(m-1)(v^2)^4 N \langle [a_N(x, s)]^4 \rangle_a = \frac{m(m-1)Nv^8}{N^{2+2\nu} a}, \end{aligned} \tag{1.21}$$

and we arrive at the conclusion that $\nu \geq \frac{1}{2}$. On the other hand, when the term

$$E [H_N(x, x)]^2 \geq m(m-1)(m-2)(m-3)v^8 \langle [a_N(x, x)]^4 \rangle_a. \tag{1.22}$$

is finite, either $\nu \geq 1$ or $a_N(x, x) = 0$. The first possibility contradicts (1.16) and so we have shown that (1.17) holds.

II. MAIN RESULT AND PROOF

Let us consider the ensemble of random matrices with entries,

$$H_N(x,y) = \sum_{\mu=1}^m \xi_N^\mu(x) \xi_N^\mu(y) a_N(x,y), \quad x,y = \overline{1,N}, \quad (2.1)$$

where $\xi_N^\mu(x)$ and $a_N(x,y) = a_N(y,x)$, $\mu = \overline{1,m}$, $x = \overline{1,N}$ are jointly independent random variables. For each fixed N we denote the average over the measure generated by $\{\xi_N^\mu(x)\}$ as $\langle \cdots \rangle_\xi$ and the average over the measure generated by $\{a_N(x,y)\}$ as $\langle \cdots \rangle_a$. Let us assume

$$\langle \xi_N^\mu(x) \rangle_\xi = 0, \quad \langle [\xi_N^\mu(x)]^2 \rangle_\xi = v^2, \quad (2.2)$$

and

$$a_N(x,y) = \frac{1 - \delta_{xy}}{N^v \sqrt{a}} \begin{cases} 1, & \text{with probability } aN^{-2(1-\nu)}, \\ 0, & \text{with probability } 1 - aN^{-2(1-\nu)}, \end{cases} \quad (2.3)$$

where $1 > \nu \geq \frac{1}{2}$ and

$$\delta_{xy} = \begin{cases} 1 & x=y, \\ 0 & x \neq y. \end{cases}$$

We study the NCF $\sigma(\lambda; H_N)$ in the limit $m, N \rightarrow \infty, m/N \rightarrow c > 0$.

Theorem 2.1: *Let each of the random variables $\{\xi_N^\mu(x)\}$ have a symmetric distribution and let, for any fixed $\tau > 0$,*

$$\lim_{N,m \rightarrow \infty} \frac{1}{Nm} \sum_{x=1}^N \sum_{\mu=1}^m \int_{|t| > \tau \sqrt{N}} |t|^{2+\beta} dP_{\mu,x}^{(N)}(t) = 0, \quad (2.4)$$

where $\beta = 4(2\nu - 1)$ and $P_{\mu,x}^{(N)}(t) = \text{Prob}\{\xi_N^\mu(x) \leq t\}$. Then

$$p\text{-Lim } \sigma(\lambda; H_N) = \sigma_{\text{sc}}(\lambda), \quad (2.5)$$

where *Lim* denotes the limiting transitions

$$m, N \rightarrow \infty, \quad m/N \rightarrow c > 0 \quad \text{and fixed } a \text{ when } \nu > \frac{1}{2}, \quad (2.6a)$$

and

$$m, N, a \rightarrow \infty, \quad m/N \rightarrow c > 0 \quad \text{and } a < N \text{ when } \nu = \frac{1}{2}, \quad (2.6b)$$

and $\sigma_{\text{sc}}(\lambda)$ is given by Eq. (1.8) with $u^2 = cv^4$.

From now on we use ‘‘Lim’’ to denote this limit (2.6) and ‘‘lim’’ to denote other limits that are defined as required.

Remarks:

(1) By p -Lim in (2.5) we mean weak convergence in probability of measures associated with $\sigma(\lambda; H_N)$. In other words, (2.5) means that for any smooth $\psi(\lambda)$ with finite support,

$$p\text{-Lim } \int \psi(\lambda) d\sigma(\lambda; H_N) = \int \psi(\lambda) d\sigma_{\text{sc}}(\lambda). \quad (2.7)$$

(2) Condition (2.4) is the analog of the well-known Lindberg condition from probability theory. In random matrix theory this condition has been proved to be sufficient¹⁹ and necessary¹² for the semicircle law (1.7)–(1.8) to be the IDS of the Wigner ensemble (1.5).

Let us stress that if $\{\xi_N^\mu(x)\}$ are identically distributed and $\langle |\xi_N^\mu(1)|^{2+\beta} \rangle < \infty$, then (2.4) holds. It should also be noted that β is always less than 4, so in the case of i.i.d. $\{\xi_N^\mu(x)\}$, $\langle |\xi_N^\mu(1)|^6 \rangle < \infty$ is sufficient for Theorem 2.1 to be true for all $\nu \geq \frac{1}{2}$. On the other hand, if $\nu = \frac{1}{2}$ then $\beta = 0$ and (2.4) takes the form closest to the Lindberg condition. The only difference is that the latter has $\tau N^{1/2}$ instead of $\tau N^{1/4}$ in (2.4). This difference is due to the quadratic character of the $\{\xi_N^\mu(x)\}$ terms in $H_N(x, y)$.

Proof: Let us introduce truncated random variables,

$$\bar{\xi}_N^\mu(x) = \begin{cases} \xi_N^\mu(x), & \text{if } |\xi_N^\mu(x)| \leq \tau^4 \sqrt{N}, \\ 0, & \text{if } |\xi_N^\mu(x)| > \tau^4 \sqrt{N}, \end{cases} \tag{2.8}$$

with $\tau < 1$ and consider the ensemble \bar{H}_N given by (2.1) with $\xi_N^\mu(x)\xi_N^\mu(y)$ replaced by $\bar{\xi}_N^\mu(x)\bar{\xi}_N^\mu(y)$. In Lemmas 1 and 2 (at the end of this section) we prove that for any smooth function $\psi(\lambda)$ with finite support,

$$p\text{-Lim} \left[\int \psi(\lambda) d\sigma(\lambda; H_N) - \int \psi(\lambda) d\sigma_{sc}(\lambda) \right] = 0. \tag{2.9}$$

Consequently, our main goal is to prove that

$$p\text{-Lim} \sigma(\lambda; \bar{H}_N) = \sigma_{sc}(\lambda). \tag{2.10}$$

To achieve this we start with the moments of \bar{H}_N and show that for any fixed p ,

$$\lim_{\tau \rightarrow 0} \text{Lim} E \bar{H}_N^p(0,0) = \begin{cases} \frac{(2k)!}{k!(k+1)!} [c\nu^4]^k, & \text{if } p = k, \\ 0, & \text{if } p = 2k + 1, \end{cases} \tag{2.11}$$

and

$$\lim_{\tau \rightarrow 0} \text{Lim} E \bar{H}_N^p(0,0) \bar{H}_N^p(0,0) - [E \bar{H}_N^p(0,0)]^2 = 0. \tag{2.12}$$

Then in Lemma 3 we prove that (2.11) and (2.12) imply (2.10).

Our study of the average,

$$E\{H_N^p(0,0)\} = \sum_{\{s_i\}} \sum_{\{\mu_j\}} \langle X_p \rangle_\xi \langle Y_p \rangle_a, \tag{2.13}$$

where

$$\langle X_p \rangle_\xi = \langle \xi^{\mu_1}(0) \xi^{\mu_1}(s_2) \xi^{\mu_2}(s_2) \cdots \xi^{\mu_{p-1}}(s_p) \xi^{\mu_p}(s_p) \xi^{\mu_p}(0) \rangle_\xi \tag{2.14}$$

and

$$\langle Y_p \rangle_a \equiv \langle a_N(0, s_2) a_N(s_2, s_3) \cdots a_N(s_{p-1}, s_p) a_N(s_p, 0) \rangle_a, \tag{2.15}$$

is based on the separation of those sets of $S_p = (0, s_2, s_3, \dots, s_p)$ and $M_p = (\mu_1, \dots, \mu_p)$, which give a nonzero contribution in the limit of infinite m, N (and a for $\nu = \frac{1}{2}$).

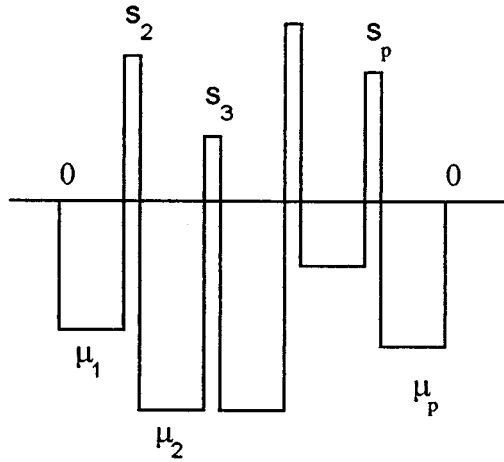


FIG. 1. Diagrammatic representation of a general $\langle X_p \rangle$.

Our main observation is that condition (1.17), $a_N(x, x) = 0$, together with the independence of $\{\xi^\mu(x)\}$ and the property

$$\langle [\bar{\xi}_N^\mu(x)]^{2k+1} \rangle_\xi = 0, \tag{2.16}$$

reduces the number of independently changing μ variables in the sum (2.12) while the properties of $\{a_N(x, y)\}$ (2.3) allow us to restrict the number of sets S_p . As we shall see, $\langle Y_p \rangle_a$ plays the same role in the selection of S_p as that played by the average $E\{W(0, s_2) \cdots W(s_p, 0)\}$ in the original Wigner proof of the semicircle law.¹⁴ This observation is crucial for counting the number of appropriate sets S_p .

When separating S_p and M_p and counting the number of nonzero contributions, we use the fact that all the moments of $\bar{\xi}_N^\mu$ are finite for fixed p and N . Then, calculating the averages $\langle X_p \rangle_\xi$ we show that, due to the independence of $\{\xi^\mu(x)\}$, the leading contribution includes only powers $\langle [\bar{\xi}_N^\mu]^2 \rangle$ while higher moments of ξ come with factors of $1/N$. This allows us to estimate terms including $\langle \xi^{2+t} \rangle$, where t is an integer, by $c_p \tau$, where c_p does not depend on N .

The role of the independence of $\{\xi^\mu(x)\}$ becomes clearer if we introduce a diagram for a fixed S_p and M_p , where each random variable $\{\xi^\mu(x)\}$ is given by a vertical interval. This interval consists of two parts; the upper part is of length s and the lower of length μ . Then the average (2.14) can be presented in the form of Fig. 1. Due to the independence of the $\{\xi^\mu(x)\}$ and the condition (2.16), the average of X_p is nonzero only when the corresponding diagram has an even number of each interval present.

For example, if we consider a fixed S_p where all numbers $0, s_2, s_3, \dots, s_p$ are different, then the average $\langle X_p \rangle$ is nonzero only when all $\{\mu_j\}$ are equal. Hence, such a sequence of $\{s_i\}$ produces m nonzero terms. It is clear that if one considers general sequences S_p , the more coincident points $\{s_i\}$, the more $\{\mu_j\}$ are allowed to vary independently, and vice versa.

Let us now consider the case of $p = 2k$. Due to (1.17), the maximal number of coincident points s_i is k and the only set that achieves this is given by $S_{2k}^* = (0, s, 0, s, \dots, 0, s)$. The corresponding diagram contains $2k$ vertical intervals with upper points s , and these need to be paired. Thus, among the $\{\mu_j\}$ only k variables are allowed to change independently and the number of nonzero terms in average (2.14) for S_{2k}^* is $c_{2k}^* m(m-1) \cdots (m-k+1) = c_{2k}^* m^k + o(m^k)$. Any change in S_{2k}^* can only diminish the number of independent μ variables. Thus, we come to the conclusion that any fixed S_{2k} produces $c_{2k} m^k + o(m^k)$ terms.

Now let us turn to the case $p = 2k + 1$. Here the S_{2k+1}^* with a maximal number of coincident s variables are of the form $(0, t, s, 0, s, \dots, 0, s, 0)$. Since $t \neq s$, only sums over $\{\mu_j\}$ with $\mu_1 = \mu_2$ provide a nonzero average. Thus, the point $s_2 = t$ can be omitted and we apply the rest of the arguments in the previous paragraph.

So, we have proved the following.

Proposition 1: Each fixed S_p produces $c_p m^{\lfloor p/2 \rfloor} + o(m^{\lfloor p/2 \rfloor})$ nonzero terms, where $[x]$ equals a maximal integer not greater than x and c_p is some constant independent of m .

Let us now prove the following.

Proposition 2: The number, L_p , of nonzero terms in (2.13) is of order $N^{2\lfloor p/2 \rfloor}$.

Proof: Let us consider the sum over those S_p in (2.2) that have d points $s_{i(t)}$, $t = \overline{1, d}$ that are unpaired, i.e., such that $s_{i(t)} \neq s_j$ for all t and $j \neq i(t)$. There are no more than $N^{d + \lfloor (p-1-d)/2 \rfloor}$ such S_p .

Since $\{\xi^\mu(x)\}$ are independent with zero mean, we have a nonzero average in (2.15) only when $\mu_{i(1)} = \mu_{i(1)+1}, \dots, \mu_{i(d)} = \mu_{i(d)+1}$. If the neighbors of $s_{i(t)}$ do not coincide, $s_{i(t)-1} \neq s_{i(t)+1}$, we have the diagram as given in Fig. 1. This diagram can be regarded as one with $p - d$ points $(0, s_2, \dots, s_{i(1)-1}, s_{i(1)+1}, \dots, s_{i(d)-1}, s_{i(d)+1}, \dots, s_p)$, which due to Proposition 1 produces no more than $m^{\lfloor (p-d)/2 \rfloor}$ terms. Thus the total number of terms in this case is no more than $N^{d + \lfloor (p-1-d)/2 \rfloor} m^{\lfloor (p-d)/2 \rfloor}$, which is of order N^p .

If some $s_{i(t)}$ has equal neighbors, $s_{i(t)-1} = s_{i(t)+1}$, then we cannot apply Proposition 1. However, such a diagram can be reduced to a new diagram corresponding to sets $S_{p'}$, where $p' = p - 2$ with $s_{i(t)}$ omitted and then $L_p = NmL_{p'}$. Consequently, Proposition 2 is proved.

Let us now study $\langle Y_{2k} \rangle$ for the case of even p . We consider the average (2.14) with $p = 2k$ and show that the leading contribution to (2.13) comes from sums over those S_{2k}^+ , where each step (s, s') is paired with its inverse (s', s) and the pairs obtained have no coincidence between them. This picture is exactly the same as the Wigner ensemble (5) and the number of such S_{2k}^+ is¹⁴

$$n_{2k} = \frac{(2k)!}{k!(k+1)!} \tag{2.17}$$

There are three ways in which general sets S_{2k} can differ from S_{2k}^+ .

- (I) There can be steps (s, s') having no inversion (s', s) or repetition (s, s') .
- (II) There exist steps (s, s') having repetition (s, s') .
- (III) There can be a coincidence between pairs of steps.

We consider these three possibilities separately because the general case can be trivially subdivided into these three scenarios.

First consider the simplest case (I) when S_{2k} contains $k + d$ different steps. Then at least $2d$ steps have no inverse and the other $2(k - d)$ steps are paired. Then

$$\langle Y_{2k} \rangle = \langle a_N^2 \rangle^{k-d} \langle a_N \rangle^{2d} = \frac{1}{N^{2(k-d)}} \left[\frac{\sqrt{a}}{NN^{1-\nu}} \right]^{2d} = \frac{1}{N^{2k}} \left[\frac{\sqrt{a}}{N^{1-\nu}} \right]^{2d}$$

Due to Proposition 2, all such terms give vanishing contributions to (2.13) as $N \rightarrow \infty$.

Before considering cases (II) and (III) let us first compute the contribution of sets S_{2k}^+ . The sum over each particular set can be obtained as follows: we first identify the steps (s_i, s_{i+1}) that are paired, and then allow s_i to run from 1 to N , but conserving this pairing. This pairing of $2k$ intervals $(0, s_2), (s_2, s_3), \dots, (s_{2k}, 0)$ splits the set of $2k + 1$ points $(0 = s_1, s_2, \dots, s_{2k}, s_{2k+1} = 0)$ into r groups; all equal points are put in the same group. Such a partition gives $N(N-1) \cdots (N-r+2) = N^{r-1} + o(N^{r-1})$ terms. Taking into account Proposition 1 and the equality $\langle Y_{2k} \rangle = N^{-2k}$, we conclude that nonvanishing contributions come from partitions into not less than $k + 1$ groups ($r \geq k + 1$). In this case at least one group consists of one point that we call the

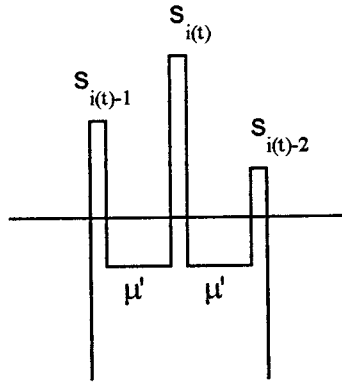


FIG. 2. Pairing of μ variables in a general diagram for $\langle X_{2k} \rangle$.

“peak” point. Thus, for each particular S_{2k}^+ there exists at least one peak point $s_{i(t)}$ such that $s_{i(t)-1} = s_{i(t)+1}$. One can then reduce S_{2k}^+ to $S_{2k'}^+$, $2k' = 2k - 2$ by removing point $s_{i(t)}$ and considering $s_{i(t)-1} = s_{i(t)+1}$ as a new point in $S_{2k'}^+$.

We repeat this reducing procedure until we are left with the two points at $(0,0)$. There are k steps in the reduction of S_{2k}^+ to $(0,0)$, in which k peak points removed. Due to Propositions 1 and 2, nonvanishing contributions to (2.13) come from S_{2k}^+ , where all these peak points vary independently, are nonzero and take different values.

Turning to the diagram for X_{2k} (Fig. 2), we see that in this sum over particular S_{2k}^+ two vertical cuts drawn down from the peak point $s_{i(t)}$ are independent from all other random variables for any M_{2k} . Then the μ variables corresponding to these s variables must be paired; $\mu_{i(t)-1} = \mu_{i(t)} = \mu'$. If in the sum considered μ' is not equal to the other μ variables, then the random variables $\xi^{\mu'}(s_{i(t)-1}) = \xi^{\mu'}(s_{i(t)+1})$ and $\xi^{\mu'}(s_{i(t)})$ are independent from the others, and the diagram for X_{2k} can also be reduced by removing four vertical cuts belonging to $\mu_{i(t)-1}$ and $\mu_{i(t)}$ and multiplying the average in (2.13) by

$$\sum_{\mu', i(t)} \langle [\bar{\xi}_N^{\mu'}(s_{i(t)-1})]^2 \rangle \frac{1}{N^2} \langle [\bar{\xi}_N^{\mu'}(s_{i(t)})]^2 \rangle.$$

Thus we have reduced the whole average $\langle X_{2k} \rangle \langle Y_{2k} \rangle$ to $\langle X_{2k-2} \rangle \langle Y_{2k-2} \rangle$. Repeating this procedure k times, and taking into account (2.4), we come to the conclusion that the sum over each particular S_{2k}^+ with noncoincident pairs of μ variables gives a contribution to (2.13) of $(cv^4)^k (1 + o(1))$, in the limit $m, N \rightarrow \infty$. Terms that come from coincident pairs of μ variables are of order τ and will be considered later.

Let us consider case (II) when each step in S_{2k} has its repetition or inverse and at least one step (s_i, s_{i+1}) has its repetition (s_j, s_{j+1}) , i.e. $s_i = s_j, s_{i+1} = s_{j+1}$. In this case $2k + 1$ points $(0, s_2, s_3, \dots, s_{2k}, 0)$ are split into r groups. If $r < k$ then such splitting gives a vanishing contribution. If $r \geq k + 1$, there is at least one peak point in S_{2k} and we reduce it as was done for S_{2k}^+ .

Repeating this reducing procedure, we come to the position where the peak point has either s_i, s_{i+1}, s_j or s_{j+1} as a neighbor. Supposing that this neighbor is $s_j \neq 0$, we obtain $s_{j-2} = s_j = s_i$. Thus, in the partition of the points of $S_{2k'}$ one group of equal points consists of three or more elements. This implies that the contribution from these sets is vanishingly small in the limit $N \rightarrow \infty$.

Now it remains to consider case (III) and show that sums over sets S_{2k} with paired steps and coincidences between pairs provide contributions vanishing in the limit $m, N \rightarrow \infty$.

In each sum over a particular set S_{2k}^+ a nonvanishing contribution is given by k peak points moving independently. To make a pair of steps coincident with another pair, one has to make at least one peak point equal to another one. It is easy to see that the contribution of such sums is

$$c_k^{(2)} N^{k-1} m^k \langle a^2 \rangle^{k-2} \langle a^4 \rangle (1 + o(1)) = O\left(\frac{1}{aN^{(2\nu-1)}}\right),$$

where $c_k^{(2)}$ counts the number of ways to make peak points coincident and $o(1)$ comes from the sums where more than two peaks are equal.

Sums over S_{2k} having exactly d pairs coincident give a contribution,

$$c_k^{(d)} N^{k-d+1} m^k \langle a^2 \rangle^{k-d} \langle a^{2d} \rangle (1 + o(1)) = O\left(\frac{1}{a^{d-1} N^{(2\nu-1)(d-1)}}\right),$$

which is also vanishing. Situations with more complex coincidences between pairs can be analyzed by generalization of the above arguments.

Let us turn now to the odd moments $E H^{2k+1}(0,0)$. In this case S_{2k+1} has at least one step (s_i, s_{i+1}) that is unpaired. In fact, due to condition (1.17), there are at least three unpaired steps. If the remaining $2k-2$ steps make a set S_{2k-2}^+ , then for such a set,

$$\langle Y_{2k+1} \rangle = \langle a_N \rangle^3 \langle a_N^2 \rangle^{k-1} = \frac{a\sqrt{a}}{N^{3(1-\nu)} N^3} \frac{1}{N^{2k-2}}.$$

But according to Proposition 2, there are no more than N^{2k+1} terms in the sum (2.13) and we obtain a contribution of order $O(a^{3/2} N^{3(\nu-1)})$ from the sum over the sets described. If the $2k-2$ steps do not form a set S_{2k-2}^+ , then the contribution is even smaller.

Stopping at this point, we see that in fact we have derived (2.11) for \bar{H}_N with bounded random variables $|\xi| < T$. Now we are going to prove that (2.11) holds for truncated random variables $\bar{\xi}_N^\mu(x)$ (2.7). Indeed, it is easy to understand that higher powers of ξ can only be obtained by coincidence between pairs of μ variables combined with the coincidence between pairs of s variables. Both of these conditions lead to extra factors m^{-1} or N^{-1} in the contributions from such sums.

We start with the sum over sets S_{2k}^+ having all steps paired with noncoincident pairs. First, consider the sums where all peak points $s_{i(i)}$ take different values. Then increased powers of ξ can be obtained just by making all the μ variables equal. The only case of interest is when around peak point i' , $\mu_{i'-2} = \mu_{i'-1} = \mu_{i'} = \mu_{i'+1}$. Then we obtain $\langle \xi^4 \rangle$ with a factor m^{-1} , which means that the contribution of these sums is $O(\tau^2 m^{-1/2})$.

If we consider sums over S_{2k}^+ with coincident peak points, then the increase in powers of ξ is followed, apart from the coincidence of pairs of μ variables, by extra powers of N^{-1} , which makes contributions from such terms even smaller than in the previous case.

Now consider sums over S_{2k} having all steps paired with d coincident pairs. In this case the maximal power of ξ is $2d$ and these sums give a contribution of order

$$\begin{aligned} N^{k-d+1} m^{k-d+1} \langle a_N^2 \rangle^{k-d} \langle a_N^{2d} \rangle \langle \xi^{2d} \rangle \langle \xi^{2d} \rangle &= \frac{1}{a^{d-1}} \frac{1}{N^{(d-1)(2\nu-1)}} \left\langle \xi^2 \frac{\xi^{2(d-1)}}{N^{(d-1)/2}} \right\rangle^2 \\ &= O\left(\frac{\tau}{a^{d-1} N^{(d-1)(2\nu-1)}}\right). \end{aligned}$$

It is obvious that the presence of unpaired steps does not lead to an increase in powers of ξ .

To complete the proof of (2.11), we just note that higher moments of ξ in odd moments $E H^{2k+1}(0,0)$ arise by the same mechanism and need not be studied separately.

Now let us describe the proof of (2.12). We rewrite the average in (2.12) for $p = 2k$ as

$$\sum_{s,\mu} \sum_{s',\mu'} [\langle X_{2k} X'_{2k} \rangle \langle Y_{2k} Y'_{2k} \rangle - \langle X_{2k} \rangle \langle X'_{2k} \rangle \langle Y_{2k} \rangle \langle Y'_{2k} \rangle], \tag{2.18}$$

and note that the difference is nonzero only in the case when X_{2k} contains random variables common with X'_{2k} or when $\langle Y_{2k} Y'_{2k} \rangle \neq \langle Y_{2k} \rangle \langle Y'_{2k} \rangle$.

We consider these two possibilities separately. The latter inequality is possible if some step from S_{2k} has its inverse only in S'_{2k} and if pairs from S_{2k} coincide with pairs from S'_{2k} . In the first case the set $(0 = s_1, s_2, s_3, \dots, s_{2k}, 0 = s'_1, s'_2, s'_3, \dots, s'_{2k})$ can be regarded as a new set S_{4k} . Reducing this set by eliminating peak points, we easily come to the conclusion that the central point, $s'_1 = 0$, is a peak point. Since it is fixed, then this sum is of order $1/N$. Averages $\langle Y_{2k} \rangle \langle Y'_{2k} \rangle$, apparently having unpaired steps, give a vanishing contribution as $N \rightarrow \infty$.

Let us consider the case when pairs S_{2k} coincide with a pair from S'_{2k} . Then the sum

$$\sum_{s,\mu}^* \sum_{s',\mu'}^* \langle X_{2k} \rangle \langle X'_{2k} \rangle \langle Y_{2k} \rangle \langle Y'_{2k} \rangle \tag{2.19}$$

over such sets is of order $1/N$ because it corresponds to the case when some of the peak points are fixed. On the other hand, the sum over sets,

$$\sum_{s,\mu}^* \sum_{s',\mu'}^* \langle X_{2k} X'_{2k} \rangle \langle Y_{2k} Y'_{2k} \rangle, \tag{2.20}$$

can be regarded as a sum for $\langle X_{4k} \rangle \langle Y_{4k} \rangle$, where the corresponding S_{4k} has coincident pairs of steps. Thus, according to arguments presented above, (2.20) contributes to (2.18) as a variable of order $(1 + O(\tau))/aN^{2\nu-1}$. It remains to check the sums where $\langle Y_{2k} Y'_{2k} \rangle = \langle Y_{2k} \rangle \langle Y'_{2k} \rangle$ but

$$\langle X_{2k} X'_{2k} \rangle \neq \langle X_{2k} \rangle \langle X'_{2k} \rangle. \tag{2.21}$$

Obviously, it is sufficient to study sums over S_{2k}^+ and $(S')_{2k}^+$ such that there is no coincidence between pairs. The one way to obtain (2.21) is to make peak points in S_{2k}^+ equal to some peak points in $(S')_{2k}^+$ and to make corresponding pairs of μ variables coincident. Another way is to make equal pairs of μ variables that correspond to bottom points of S_{2k}^+ and $(S')_{2k}^+$. It is easy to see that these ways lead to terms with contribution $O(N^{-1/2})$.

Similar reasoning shows that (2.13) holds for odd moments $p = 2k + 1$. Thus (2.11) and (2.12) are proved.

Lemma 1: Let $\{\sigma_N(\lambda; \omega)\}$ be a sequence of random nondecreasing non-negative bounded functions, and let $\{f_N(\lambda; \omega)\}$ be the sequence of their Stieltjes transforms,

$$f_N(\lambda) = \int (\lambda - z)^{-1} d\sigma_N(\lambda),$$

where ω is a point (realization) of the corresponding probability space Ω_N . Suppose that there exists a nonrandom function $f(z)$ that is analytic for $\text{Im } z \neq 0$ satisfying inequalities

$$\sup_{\eta > 0} \eta f(\eta) \leq 1, \quad \text{Im } f(z) \text{Im } z > 0,$$

and that

$$\text{Lim sup}_{z \in U_0} E |f_N(z) - f(z)|^2 = 0, \tag{2.22}$$

where $U_0 = \{z \in \mathbb{C}, |\operatorname{Im} z| \geq \eta_0\}$ and $\eta_0 > 0$. If $\sigma(\lambda), \sigma(-\infty) = 0$ is the nondecreasing function that corresponds to $f(z)$, then at each continuity point of $\sigma(\lambda)$ we have

$$p\text{-}\lim \sigma_N(\lambda) = \sigma(\lambda), \tag{2.23}$$

or, in other words, the measures $\sigma_N(d\lambda; \omega)$ weakly converge in probability to $\sigma(d\lambda)$ [cf. (2.7)].

The proof of this lemma can be found, for example, in Ref. 20. The key point is that the family $f_N(z, \omega) - f(z)$ is analytic and uniformly bounded on any compact set T belonging to U_0 .²¹ This allows one to derive from (2.22) the relation

$$\lim E \sup_{z \in T} |f_N(z) - f(z)|^2 = 0,$$

which together with the compactness of the family $\sigma_N(d\lambda; \omega) - \sigma(d\lambda)$ ²¹ implies (2.23).

Let us define

$$\bar{G}_N = \frac{1}{\bar{H}_N - z} \quad \text{and} \quad G_N = \frac{1}{H_N - z},$$

where H_N is given by (2.1)–(2.4) and \bar{H}_N is obtained from H_N by truncation (2.8). Then according to the definition of $\sigma(\lambda)$ (1.3),

$$\frac{1}{N} \operatorname{Tr} \bar{G}_N = \int (\lambda - z)^{-1} d\sigma(\lambda; \bar{H}_N)$$

and

$$\frac{1}{N} \operatorname{Tr} G_N = \int (\lambda - z)^{-1} d\sigma(\lambda; H_N).$$

Lemma 2: For $z \in U_0$,

$$p\text{-}\lim_{m, N \rightarrow \infty} \left| \frac{1}{N} \operatorname{Tr} G_N(z) - \frac{1}{N} \operatorname{Tr} \bar{G}_N(z) \right| = 0.$$

Proof: Let us consider the resolvent identity $G' - G = -G'(H' - H)G$, where $G = (H - z)^{-1}$ and $G' = (H' - z)^{-1}$, $|\operatorname{Im} z| > 0$ and H, H' are symmetric matrices of the same dimension. Then

$$\Delta_N(z) = \frac{1}{N} \operatorname{Tr}(G_N(z) - \bar{G}_N(z)) = \frac{1}{N} \sum_{s,t} (\bar{G}_N G_N)(s,t) \sum_{\mu} [\hat{\xi}_N^{\mu}(s) \xi_N^{\mu}(t) + \hat{\xi}_N^{\mu}(t) \bar{\xi}_N^{\mu}(s)] a_N(s,t), \tag{2.24}$$

where $\hat{\xi} = \xi - \bar{\xi}$.

We denote $\sum_{\mu} \hat{\xi}_N^{\mu}(s) \xi_N^{\mu}(t)$ by $\gamma_N(s,t)$ and, using the inequality $\|G_N\| < |\operatorname{Im} z|^{-1}$ and $|G(s,t)| < \|G_N\|$, derive from (2.24) the relation

$$E\{|\Delta_N(z)|\} \leq \frac{2}{N|\operatorname{Im} z|^2} \sum_{s,t} \langle |\gamma_N(s,t)| \rangle \langle a_N(s,t) \rangle \leq \frac{1}{N^2} \frac{\sqrt{a}}{N^{1-\nu}} \sum_{s \neq t} \langle (\gamma_N(s,t))^2 \rangle^{1/2}. \tag{2.25}$$

It is easy to see that if $s \neq t$ then

$$\langle (\gamma_N(s,t))^2 \rangle = \left\langle \sum_{\mu,\lambda} \hat{\xi}^\mu(s) \xi^\mu(t) \hat{\xi}^\lambda(s) \xi^\lambda(t) \right\rangle = v^2 \left\langle \sum_{\mu} [\hat{\xi}^\mu(s)]^2 \right\rangle.$$

Then we derive from (2.25) that

$$\begin{aligned} E\{|\Delta_N(z)|\} &\leq \frac{v\sqrt{a}}{NN^{1-\nu}} \sum_s \left\langle \sum_{\mu} |\hat{\xi}_N^\mu(s)|^2 \right\rangle^{1/2} \\ &\leq \frac{mv\sqrt{a}}{N} \left[\frac{1}{mN} \sum_{s,\mu} \frac{1}{N^{1-2\nu}} \langle |\hat{\xi}_N^\mu(s)|^2 \rangle \right]^{1/2} \\ &\leq \frac{mv\sqrt{a}}{N} \left[\frac{1}{mN} \sum_{s,\mu} \int_{|t|>\tau^4/N} |t|^{2+4(2\nu-1)} \frac{N^{2\nu-1}}{|t|^{4(2\nu-1)}} dP_{\mu,s}^{(N)}(s) \right]^{1/2}. \end{aligned}$$

Using (2.4), we complete the proof of Lemma 2.

Lemma 3: Relations (2.10) and (2.11) imply that for $z \in U_0$, with $\eta_0 = (2cv^4 + 2)^2$ and $|\operatorname{Re} z| < 1$,

$$\operatorname{Lim} E \left\{ \left| \frac{1}{N} \operatorname{Tr} \bar{G}_N(z) - f(z) \right|^2 \right\} = 0, \tag{2.26}$$

where

$$f(z) = \int (\lambda - z)^{-1} d\sigma_{\text{sc}}(\lambda),$$

with $\sigma_{\text{sc}}(\lambda)$ given by (1.8) with $u^2 = cv^4$.

Proof: We prove (2.26) by showing that

$$\operatorname{Lim} E\{g_N(z)\} = f(z), \tag{2.27}$$

where $g_N(z) = N^{-1} \operatorname{Tr} \bar{G}_N(z)$, and that

$$\operatorname{Lim} E\{g_N(z)g_N(z)\} - E\{g_N(z)\}E\{g_N(z)\} = 0. \tag{2.28}$$

For given $\epsilon > 0$, we choose $2q$ such that

$$\frac{1}{(2cv^4 + 2)^{2q}} < \frac{\epsilon}{4},$$

and expand $g_N(z)$ into the series

$$g_N(z) = -\frac{1}{N} \sum_{p=0}^{2q} \frac{H^p(x,x)}{z^{p+1}} - \frac{1}{z^{2q+1}} R_N^{(q)}(z),$$

where

$$R_N^{(q)}(z) = \frac{1}{N} \operatorname{Tr} H_N^{2q+1} G_N.$$

Let us note that

$$|R_N^{(q)}(z)| = \left| \int_{-\infty}^{\infty} \frac{\lambda^{2q+1}}{\lambda - z} d\sigma(\lambda; \bar{H}_N) \right| \leq \left| \int_{-\infty}^{\infty} \lambda^{2q} \frac{\lambda - \operatorname{Re} z + i \operatorname{Im} z}{(\lambda - \operatorname{Re} z)^2 + (\operatorname{Im} z)^2} d\sigma(\lambda; \bar{H}_N) \right| \leq \frac{2}{N} \operatorname{Tr} H_N^{2q}.$$

Then we expand $f(z)$ into the series

$$f(z) = - \sum_{p=0}^{2q} \frac{M_p}{z^{p+1}} - \frac{r_q(z)}{z^{2q+1}},$$

where M_p are given by the right-hand side of (2.11) and $r_q(z) \leq 2M_{2q}$.

Then taking into account that $EH_N^p(x,x) = EH_N^p(0,0)$, we can write the inequality

$$|Eg_N(z) - f(z)| \leq \sum_{p=0}^{2q} \frac{|EH_N^p(0,0) - M_p|}{4^p} + \frac{1}{(2cv^4 + 2)^{2q}} \frac{4M_{2q} + 2|EH_N^{2q}(0,0) - M_{2q}|}{(2cv^4 + 2)^{2q+2}}.$$

The trivial inequality $M_{2q} \leq (cv^4)^{2q}$ together with (2.11) implies (2.27). The relation (2.28) can be derived from (2.12) using the same procedure. Theorem 2.1 is proved.

III. FINITE DILUTION PARAMETER

In this section we study the moments

$$\int \lambda^p d\sigma(\lambda; H_N)$$

of the ensemble (2.1)–(2.3) in the case $\nu = 1/2$ and finite $a \geq 1$. We prove that there exist numbers $h^{(a)}_p$ such that

$$\lim E\{H_N^p(0,0)\} = h^{(a)}_p, \quad p \in N \cup \{0\}. \tag{3.1}$$

We derive estimates for $h^{(a)}_p$, which imply that

$$\sum_{k=0}^{\infty} [h^{(a)}_{2k}]^{-1/2k} = \infty. \tag{3.2}$$

This Carleman’s condition provides existence²² and uniqueness²³ of a non-negative nondecreasing function $\sigma_a(\lambda)$ satisfying the relation

$$h^{(a)}_p = \int_{-\infty}^{\infty} \lambda^p d\sigma_a(\lambda).$$

We prove that the support of the measure $d\sigma_a(\lambda)$ is unbounded and study the asymptotic behavior of $\sigma_a(\lambda)$ for large $|\lambda|$.

Finally, we show that if a function $\sigma^{(1)}(\lambda)$ exists, such that

$$\sigma_a(\lambda) = \sigma_{sc}(\lambda) + \frac{1}{a} \sigma^{(1)}(\lambda) + O\left(\frac{1}{a^2}\right), \tag{3.3}$$

then $\sigma^{(1)}(\lambda)$ can be written in the form

$$\sigma^{(1)}(\lambda) = 3\theta(\lambda) + \frac{3}{2\pi u^2} \int_{-2u}^{\lambda} \left(\frac{\tau^2}{u^2} - 2 \right) \sqrt{4u^2 - \tau^2} d\tau, \tag{3.4}$$

where

$$\theta(\lambda) = \begin{cases} 1, & \lambda \geq 0, \\ 0, & \lambda < 0, \end{cases}$$

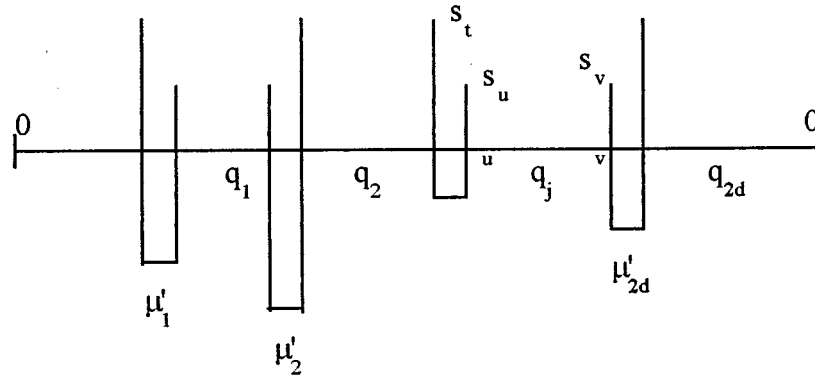


FIG. 3. Division into intervals for the case of d coincident pairs.

and $u^2 = cv^4$. This can be compared with the results of Ref. 17, where the $1/p$ correction for the diluted Wigner ensemble (1.9) was calculated. The corrections are slightly different but have the same structure of the semicircle distribution multiplied by a quadratic function of λ . The $\theta(\lambda)$ term, which does appear in the results of Ref. 17, probably arises in our problem from the condition of zero diagonal (1.17).

Let us first note that it follows from the proof of Theorem 2.1 that $h_{2k+1}^{(a)} = 0$ and $h_0^{(a)} = 1$. The next observation is that in the average

$$E\{H_N^{2k}(0,0)\} = \sum_{\{s_i\}} \sum_{\{\mu_j\}} \langle X_{2k} \rangle \langle Y_{2k} \rangle,$$

the nonvanishing contribution in the limit $N \rightarrow \infty$ comes from sums over those sets $S_{2k} = (0, s_2, s_3, \dots, s_{2k})$, where each step (s_i, s_{i+1}) has an inverse (s_{i+1}, s_i) . Since $\nu = 1/2$ and a is finite, sums over S_{2k} that have coincident pairs of steps, as well as over S_{2k}^+ with no coincident pairs, give a nonvanishing contribution to (3.5).

Let us consider sums over $S_{2k}^{(d)}$ with exactly d equal pairs. The remaining $2(k-d)$ steps are paired and $\{s_i\}$ run from one to N such that these pairs are not equal. Let us calculate the number $L_{2k}^{(d)}$ sequences S_{2k} of this type. Having marked $2d$ steps, we obtain $2d$ intervals between them of lengths q_1, \dots, q_{2d} , $q_j \geq 0$. Note that the last interval q_{2d} consists of two parts because we consider two edge points 0 as one (see Fig. 3).

Let us consider a particular interval number j with left end u and right end v . Due to the independence of pairs given by $S_{q_j}^+$ from other pairs we can sum over $S_{q_j}^+$ and the corresponding μ variables and obtain (to leading order) the factor $E\{H_N^{q_j}(S_u, S_v)\}$ in (3.1). Thus, we conclude that each interval is of even length $q_j = 2p_j$, $p_j \geq 0$ and $s_u = s_v$.

The latter is because we can consider s_u and s_v fixed (we sum over them at the last stage), and use the fact that for each fixed t ,

$$EH_n^{2k}(0,t) = O(1/N + \tau).$$

This can easily be proved from the observations that

$$EH_N^{2k}(0,t) \leq [EH_N^{2k}(0,t)H_N^{2k}(t,0)]^{1/2}$$

and that in the last average there is one fixed peak point.

So, the leading contribution comes from the diagrams of the type in Fig. 3, where steps (s_t, s_u) are separated by steps (s_u, s_t) . It is easy to see that the number of different sets $S_{2k}^{(d)}$ is given by the formula

$$L_{2k}^{(d)} = \sum_{\substack{p_i \geq 0 \\ \sum p_i = k-d}} n_{p_1} n_{p_2} \cdots n_{p_{2d}}, \tag{3.6}$$

where n_p is defined by (2.17).

Now let us compute the contributions of sums over $S_{2k}^{(d)}$. The sums over $S_{q_1}^+, \dots, S_{q_{2k}}^+$ give the leading terms as $N \rightarrow \infty, \tau \rightarrow 0$,

$$(c v^4)^{k-d} N^{k-d} m^{k-d} \langle a^2 \rangle^{k-d} (1/N), \tag{3.7}$$

where the factor $1/N$ comes from the fixed peak point 0 in the interval q_{2d} .

We now compute averages over random variables belonging to coincident pairs. There are d upper points and d lower points and, hence, variables μ'_1, \dots, μ'_{2d} should be paired to obtain a nonzero average in the limit $N \rightarrow \infty, \tau \rightarrow 0$. Then we obtain for the sum expression,

$$T_{2d} N^2 m^d a^{2d} v^{2d} (1 + O(\tau)), \tag{3.8}$$

where $O(\tau)$ comes from the sums where more than two μ variables are equal and T_{2d} is the number of ways of splitting $2d$ points, (i_1, \dots, i_{2d}) , into pairs. T_{2d} has the property that $T_{2d+2} = (2d+1)T_{2d}$, because i_{2d+2} can make a pair with $2d+1$ points and the remaining $2d$ points produce T_{2d} possibilities. Thus $T_{2d} = (2d-1)!!$

Collecting (3.6), (3.7), and (3.8), we find that $S_{2k}^{(d)}$ gives a contribution,

$$(1/a^{d-1}) L_{2k}^{(d)} (c v^4)^k (2d-1)!!, \tag{3.9}$$

to (3.1). Let us stress that $S_{2k}^{(d)}$ are the only sources of terms of order $1/a^{d-1}$. It should be noted that (3.6) with $d=1$ results in the recurrence relation

$$n_k = \sum_{p, k-p-1 \geq 0} n_p n_{k-p-1}, \quad n_0 = 1. \tag{3.10}$$

This relation, leading to the exact form of n_{2k} (2.17), was first derived by Wigner.¹⁴

It follows from (3.10) that the moments M_k of the semicircle distribution given by (8) satisfy recurrence relation

$$M_k = u^2 \sum_{p, k-p-1 \geq 0} M_p M_{k-p-1}. \tag{3.11}$$

Taking into account previous considerations, we obtain finally for $h_{2k}^{(a)}$,

$$h_{2k}^{(a)} = \sum_{d=1}^k \frac{(2d-1)!!}{a^{d-1}} u^{2d} \sum_{k-d}^* M_{p_1} M_{p_2} \cdots M_{p_{2d}}, \quad h_{2k+1}^{(a)} = 0 \tag{3.12}$$

where a summation \sum_K^* denotes a sum over $\{p_i\}$ such that $p_i \geq 0$ for all i and $\sum p_i = K$.

Now let us show that (3.2) holds. Since $M_k = u^{2k} n_k$, we derive from (2.17) the trivial estimate,

$$M_k \leq (2u)^{2k}, \quad k \in \mathbb{N}.$$

Then each term in the sum over p_i in (3.12) is less than $(2u)^{2k-2d}$ and the number of terms in this sum is

$$\binom{2k}{2d} = \frac{(2k)!}{(2d)!(2k-2d)!}.$$

The latter fact can easily be understood if one remembers that (3.6) was obtained by choosing $2d$ from $2k$ steps. Thus we derive from (3.12) that

$$h_{2k}^{(a)} \leq a(2u)^{2k} \sum_{d=1}^k \binom{2k}{2d} \frac{(2d-1)!!}{(2\sqrt{a})^{2d}} = a \frac{(2u)^{2k}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \left(\left[\frac{x}{2\sqrt{a}} + 1 \right]^{2k} - 1 \right) \exp\left\{ \frac{-x^2}{2} \right\} dx. \quad (3.13)$$

Integrating by parts, we obtain the recurrence relation

$$\beta_p = \beta_{p-1} + \frac{p-1}{4a} \beta_{p-2}, \quad \beta_0 = 1, \quad \beta_1 = 1,$$

for the moments

$$\beta_p = \left\langle \left(1 + \frac{\gamma}{2\sqrt{a}} \right)^p \right\rangle_{\gamma},$$

where γ is a Gaussian distributed random variable with 0 mean and variance 1. This relation provides the elementary estimate for $a > 1$,

$$h_{2k}^{(a)} \leq 2a \left[2u \left(1 + \frac{1}{4a} \right) \right]^{2k} (2k-1)!!; \quad (3.14)$$

then (3.2) is shown to be true.

Now it is easy to see that $\sigma_a(\lambda)$ cannot have a bounded support. In the latter case the moments $h_{2k}^{(a)}$ admit an exponential estimate for all k , but it follows from (3.12) that

$$h_{2k}^{(a)} > \frac{(2k-1)!!}{a^{k-1}} u^{2k}.$$

Inequality (3.14) provides that

$$\frac{1}{k!} \int \left(\frac{\lambda}{T} \right)^{2k} d\sigma_a(\lambda) \leq \frac{a}{2^{k-1}},$$

is true for all $k \in \mathbb{N}$, where

$$T = 4u \left(1 + \frac{1}{4a} \right).$$

Then

$$\int_{-\infty}^{\infty} \exp\left\{ \frac{\lambda^2}{T^2} \right\} d\sigma_a(\lambda) \leq 4a$$

and

$$\int_{|\lambda|>xT} d\sigma_a(\lambda) \leq 4a \exp\{-x^2\}. \tag{3.15}$$

This gives the estimate for the asymptotic behavior of $\sigma_a(\lambda)$ for large $|\lambda|$.

Now let us derive (3.4). Considering the term $d=2$ in (3.12) and applying (3.11) twice, we obtain that

$$M_{2k}^{(1)} = 3u^4 \sum_{2k-4}^* M_{p_1} M_{p_2} M_{p_3} M_{p_4} = 3 \sum_{q, 2k-4-q \geq 0} M_{q+2} M_{2k-2-q} = \frac{3}{u^2} [M_{2k+2} - 2u^2 M_{2k}]. \tag{3.16}$$

Thus, if expression (3.3) holds, we have to find a function $\sigma^{(1)}(\lambda)$ such that

$$\int \lambda^{2k+1} d\sigma^{(1)}(\lambda) = 0$$

and

$$M_0^{(1)} = \int d\sigma^{(1)}(\lambda) = 0, \quad M_2^{(1)} = \int \lambda^2 d\sigma^{(1)}(\lambda) = 0, \tag{3.17}$$

and

$$M_{2k}^{(1)} = \int \lambda^{2k} d\sigma^{(1)}(\lambda) = 3 \int \left[\frac{\lambda^2}{u^2} - 2 \right] \lambda^{2k} d\sigma_{sc}(\lambda).$$

It is a simple matter to check that (3.4) satisfies these conditions. Let us note that all terms with higher powers of $1/a$ from (3.12) can be treated by the same technique and subsequent $1/a$ corrections $\sigma^{(k)}(\lambda)$, $k=2,3,\dots$, to the function $\sigma_a(\lambda)$ can be found. However, one needs some additional arguments to prove the existence of these corrections. This is because all functions $\sigma^{(k)}(\lambda)$ cannot be nondecreasing due to the condition $\int d\sigma^{(k)}(\lambda) = 0$ [c.f. (3.17)]. Hence, classical moment problem theory cannot be applied to prove the existence and uniqueness of corrections $\sigma^{(k)}(\lambda)$.

IV. OTHER ENSEMBLES OF DILUTE RANDOM MATRICES

In the two previous sections we studied the dilution of the Marchenko–Pastur–Hopfield (MPH) matrices (1.1) with

$$d_N(\mu, x, y) = a_N(x, y). \tag{4.1}$$

This dilution is known as a ‘‘spatial’’ dilution in neural network theory. We observe that it changes the IDS of the MPH ensemble and leads to the semicircle distribution. This can be interpreted as the spatial dilution (4.1) destroying the dependence between the entries in the MPH ensemble.

If one were to introduce a dilution of the form (1.13), then the matrix obtained,

$$\hat{A}_N(x, y) = \sum_{\mu=1}^m \xi_N^\mu(x) \xi_N^\mu(y) d_N^\mu(x) d_N^\mu(y), \quad (4.2)$$

is more closely related to the structure of the pure, undiluted MPH ensemble (1.1) than the dilute ensemble we considered. This observation is supported by the following theorem.

Theorem 4.1: *Let independent random variables $\{\xi^\mu(x)\}$ satisfy the conditions of Theorem 2.1. If random variables $a_N^\mu(x)$ are jointly independent and independent from $\{\xi^\mu(x)\}$ and*

$$d_N^\mu(x, y) = \begin{cases} \frac{1}{N^\alpha \sqrt{a}}, & \text{with probability } \frac{a}{N^{1-2\alpha}} \\ 0, & \text{with probability } 1 - \frac{a}{N^{1-2\alpha}}, \end{cases} \quad (4.3)$$

with $0 \leq \alpha \leq 1/2$, then

$$p\text{-Lim } \sigma(\lambda; \hat{A}_N) = \sigma(\lambda), \quad (4.4)$$

where $\sigma(\lambda)$ is given by (1.4).

Under Lim in (4.4) we mean the limiting transitions [c.f. (2.6)]

- (a) $m, N \rightarrow \infty, m/N \rightarrow c > 0$ when $\alpha > 0$ and
- (b) $m, N, a \rightarrow \infty, m/N \rightarrow c > 0$ and $a < N$ when $\alpha = 0$.

One can prove this theorem by using, for example, some modification of the resolvent technique developed in Ref. 16.

We see that the Marchenko–Pastur distribution can also be a limiting distribution for certain dilute random matrix ensembles. However, the following results show that this situation is quite unusual. Namely, applying the technique used in Sec. II, we prove the following.

Theorem 4.2: *Let i.i.d. random variables $w(x, y)$, $x < y$ have zero average and variance w^2 [c.f. (1.6)] and let $w(x, x) = 0$. Then the NCF of the random matrices,*

$$\tilde{W}_N(x, y) = w(x, y) \sum_{\mu=1}^m d_N^\mu(x) d_N^\mu(y), \quad x, y = \overline{1, N}, \quad (4.5)$$

where $d_N^\mu(x)$ are defined by (4.3) and $0 \leq \alpha \leq 1/4$, converge in probability to the semicircle distribution (1.8) with $v^2 = cw^2$ in the limit described in Theorem 4.1.

Remark: As we noted earlier, the technique of eliminating diagrams with vanishing contributions used in Sec. II is appropriate here. However, for the case of a finite dilution parameter (a fixed and $\alpha = 0$), the diagrams giving nonzero contributions to the IDS of (4.5) are different from those of the spatially dilute MPH ensemble (2.1). This results in different $1/a$ corrections to the semicircle distribution. We plan to study this problem in a separate publication.

Taking into account that the semicircle distribution is the IDS of a spatially diluted Wigner ensemble, we can conclude that it is the more natural eigenvalue distribution for dilute random matrices than the Marchenko–Pastur distribution.

The dilution could be regarded as a particular case of a more general problem in the random modulation of matrices,

$$\hat{H}_N(x, y) = A_N(x, y) D_N(x, y).$$

One could, for instance, ask about the stability of the semicircle distribution under modulation of the Wigner random matrices $W_N (= A_N)$ by some random perturbation D_N .

As a particular answer to this question we can present the result about the IDS of the curious ensemble of random matrices,

$$\hat{H}_N(x,y) = \frac{1}{\sqrt{N}} w(x,y) \frac{1}{\sqrt{N}} \sum_{\mu=1}^m \xi_N^\mu(x) \xi_N^\mu(y), \quad (4.5)$$

which can also be regarded as the modulation of MPH random matrices (1.1) by independent random variables $w(x,y)$, $x \leq y$, $w(x,y) = w(y,x)$ satisfying (1.6).

By slightly changing the reasoning presented in Sec. II, we can prove that the IDS of \hat{H}_N is also the semicircle distribution.

V. DISCUSSION

We have considered the IDS of an ensemble of dilute random matrices H_N in the limit $N \rightarrow \infty$. Our main tool was the moments $E\{H_N^K\}$, $k \in N$. To study their asymptotic behavior as $N \rightarrow \infty$, we modified the original technique used by Wigner to prove the semicircle law. Using this technique we obtained an exact expression for the moments in the limit $N \rightarrow \infty$, for both infinite and finite dilution parameter a .

Our main result, Theorem 2.1, is that the spatial dilution of the Marchenko–Pastur–Hopfield (MPH) ensemble leads to the semicircle distribution, and not an analog of the distribution for the pure, undiluted, MPH ensemble. In Secs. III and IV we showed that the IDS of the dilute MPH ensemble is similar to the IDS of the dilute Wigner ensemble, even for finite dilution parameter a , and that the semicircle distribution is stable with respect to several other types of dilution.

The nature of the similarity between the dilute MPH and dilute Wigner ensembles for fixed a becomes especially clear in the case of $\nu = 1/2$. Then (2.1) can be redefined as an $N \times N$ matrix with entries $\gamma_N(x,y)c_N(x,y)$, where

$$\gamma_N(x,y) = \frac{1}{\sqrt{N}} \sum_{\mu=1}^m \xi_N^\mu(x) \xi_N^\mu(y), \quad x \neq y,$$

and $c_N(x,y)$ is 1 with probability a/N and 0 with probability $1 - a/N$. In this case H_N for each N contains approximately $a^2/2$ nonzero entries and they converge when $N \rightarrow \infty$ to jointly independent random variables. This explains the convergence of the dilute MPH and Wigner ensembles.

The difference between the MPH and Wigner ensembles is that the entries in the MPH matrices are slightly dependent on one another. However, this dependence is enough to shift the IDS of the pure MPH ensemble from the semicircle distribution. The spatial dilution eliminates, in the limit $N \rightarrow \infty$, the dependence between entries in the MPH ensemble.

This conclusion suggests that it would be interesting to study the spatial dilution of random matrices with more strongly dependent entries. For example, one could consider (1.1) with Gaussian $\xi^\mu(x)$, such that²⁴

$$\langle \xi^\mu(x) \xi^\tau(y) \rangle = V_{\mu-\tau}(x-y).$$

We assume that the spatial dilution will break this dependence between entries in the limit $N \rightarrow \infty$.

The same phenomenon of breaking the dependence between the matrix elements with spatial dilution was observed in studies of the dilute MPH ensemble in neural network theory.^{4,8} These studies considered the case of strong dilution that corresponds to our problem when $\nu = 1/2$. Note that these works treated the case of an infinite dilution parameter ($a \rightarrow \infty$), while we observe breaking for finite values of a .

Another type of dilution, called weak dilution in the literature on neural network theory,^{3,5} corresponds to the case $\nu = 1$ in definition (2.3). Using this terminology, we have studied the IDS of the MPH ensemble with moderate and strong dilution.

It seems to be difficult to use our technique to study the weak dilution case directly. This is because the ensemble (2.1)–(2.3) with $\nu=1$ differs essentially from those with $1/2 \leq \nu < 1$. Preliminary studies show that the IDS of the weak dilution MPH ensemble cannot be equal to the semicircle or the Marchenko–Pastur distribution. We plan to study this ensemble separately.

Another of our observations concerns random matrices,

$$\hat{H}_N(x,y) = w(x,y)D_N(x,y),$$

where $w(x,y)$ are as in Wigner random matrices and $D_N(x,y)$ represents dilution independent from $w(x,y)$ or, more generally, a random modulation of the Wigner ensemble.

In Sec. IV we showed that if $D_N(x,y)$ is the proper dilution of the MPH ensemble and even if $D_N(x,y)$ are entries of MPH matrices by themselves, then the IDS of \hat{H}_N is again a semicircle distribution.

These facts, together with our main conclusion, suggest that the semicircle law is quite stable to dilution (or modulation). It would be interesting to develop a more precise formulation of this observation.

ACKNOWLEDGMENTS

A. K. would like to thank the Department of Physics at Brunel University for their hospitality while this work was performed and both authors would like to thank the Royal Society, London for financial support.

- ¹L. Pastur and A. Figotin, ‘‘Exactly solvable model of a spin glass,’’ *Sov. J. Low. Temp. Phys.* **3**, 378–383 (1977).
- ²J. J. Hopfield, ‘‘Neural networks and physical systems with emergent collective computational abilities,’’ *Proc. Natl. Acad. Sci. USA* **79**, 2554–2558 (1982).
- ³D. J. Amit, *Modeling Brain Function* (Cambridge University Press, Cambridge, 1989), p. 504.
- ⁴*Handbook of Brain Theory and Networks*, edited by M. Arbib (MIT Press, London, 1995) p. 118.
- ⁵J. A. Hertz, A. Krogh, and R. G. Palmer, *Introduction to The Theory of Neural Computations* (Addison-Wesley, Reading, MA, 1991) p. 327.
- ⁶A. Bovier and V. Gayrard, ‘‘Rigorous results on the thermodynamics of the dilute Hopfield Model,’’ *J. Stat. Phys.* **72**, 79–112 (1993).
- ⁷A. Bovier and V. Gayrard, ‘‘Rigorous bounds on the storage capacity of the dilute Hopfield Model,’’ *J. Stat. Phys.* **69**, 597–627 (1992).
- ⁸B. Derrida, E. Gardner, and A. Zippelius, ‘‘An exactly solvable asymmetric neural network model,’’ *Europhys. Lett.* **4**, 167–173 (1987).
- ⁹P. de Felice, C. Marangi, G. Nardulli, and G. Pasquarielli, ‘‘Dynamical phase transitions in the Little-Hopfield model,’’ *J. Phys. A* **27**, 4115–4125 (1994).
- ¹⁰H. Sompolinsky, ‘‘Neural networks with non-linear synapses and static noise,’’ *Phys. Rev. A* **34**, 2571–2574 (1986).
- ¹¹V. Marchenko and L. Pastur, ‘‘The distribution of eigenvalues in certain sets of random matrices,’’ *Mat. Sb.* **72**, 507–536 (1967) [*Math. USSR Sb.* **1**, 457–483 (1967)].
- ¹²V. L. Girko, *Spectral Theory Random Matrices* (Nauka, Moscow, 1988), p. 325 (in Russian).
- ¹³M. L. Mehta, *Random Matrices* (Academic, New York, 1991), p. 421.
- ¹⁴E. Wigner, ‘‘Characteristic vectors of bordered matrices with infinite dimensions,’’ *Ann. Math.* **62**, 548–564 (1955).
- ¹⁵Y. V. Fyodorov and A. D. Mirlin, ‘‘Universality of level correlation function of sparse random matrices,’’ *J. Phys. A* **24**, 2273–2286 (1991).
- ¹⁶A. Khorunzhy, B. Khoruzhenko, L. Pastur, and M. Shcherbina, ‘‘Large- n limit in statistical mechanics and the spectral theory of disordered systems,’’ in *Phase Transitions and Critical Phenomena*, edited by C. Domb and J. L. Lebowitz (Academic, London, 1992), Vol. 15, pp. 73–245.
- ¹⁷G. J. Rodgers and A. J. Bray, ‘‘Density of states of a sparse random matrix,’’ *Phys. Rev. B* **37**, 3557–3762 (1988).
- ¹⁸G. J. Rodgers and C. De Dominicis, ‘‘Density of states of sparse random matrices,’’ *J. Phys. A* **23**, 1567–1573 (1990).
- ¹⁹L. Pastur, ‘‘On the spectrum of random matrices,’’ *T. Mat. Fiz.* **10**, 102–112 (1972) (*Theor. Math. Phys.* **10**, 67–74 (1972)).
- ²⁰L. Pastur and A. Khorunzhy, ‘‘On the eigenvalue distribution of the deformed Wigner ensemble of random matrices,’’ in *Advances in Soviet Mathematics*, edited by V. Marchenko (AMS, Providence, 1994), Vol. 19, pp. 97–127.
- ²¹W. Donoghue, *Monotone Matrix Functions and Analytic Continuation* (Springer, Berlin, 1974), p. 182.
- ²²L. Breiman, *Probability* (Addison-Wesley, Reading, MA, 1968), p. 421.
- ²³N. Akhiezer, *The Classical Moment Problem and Some Related Questions of Analysis* (Oliver and Boyd, Edinburgh, 1965), p. 253.
- ²⁴A. Boutet de Monvel, A. Khorunzhy, and V. Vasilchuk, ‘‘Eigenvalue distribution of large random matrices with correlated entries,’’ *Markov Process. Relat. Fields* **21**, 607–636 (1996).