

Sixty Years of Moments for Random Matrices

Werner Kirsch and Thomas Kriecherbauer

Abstract.

This is an elementary review, aimed at non-specialists, of results that have been obtained for the limiting distribution of eigenvalues and for the operator norms of real symmetric random matrices via the method of moments. This method goes back to a remarkable argument of Eugene Wigner some sixty years ago which works best for independent matrix entries, as far as symmetry permits, that are all centered and have the same variance. We then discuss variations of this classical result for ensembles for which the variance may depend on the distance of the matrix entry to the diagonal, including in particular the case of band random matrices, and/or for which the required independence of the matrix entries is replaced by some weaker condition. This includes results on ensembles with entries from Curie-Weiss random variables or from sequences of exchangeable random variables that have been obtained quite recently.

2010 Mathematics Subject Classification. Primary 60-B20; Secondary 82-B44.

Keywords. Random Matrices, Moment Method, Wigner Matrices, Curie-Weiss Ensembles, Semicircle Law

Dedicated to Helge Holden on the occasion of his 60th birthday

1. Introduction

Approximately at the time when Helge Holden was born the physicist Eugene Wigner presented a result in [45] that may be considered to be the starting signal for an extremely fruitful line of investigations creating the now ample realm of random matrices. The reader may consult the handbook [3] to obtain an impression of the richness of the field. Its ongoing briskness is well documented by over 700 publications listed in MathSciNet after the print of [3] in 2011.

In view of later developments that often use heavy machinery to provide very detailed knowledge about specific spectral statistics, Wigner's observation impresses by its simplicity and fine combinatorics. For certain matrix ensembles, which in various generalizations are nowadays called *Wigner ensembles*, he was able to determine the limiting density of eigenvalues by the *moment method*. More precisely, he computed the expectations of all moments of the *empirical eigenvalue distribution measures* in the limit of matrix dimensions tending to infinity. Furthermore, he observed that these limits agree with the moments of the *semicircle distribution* thus proving the semicircle law that bears his name (see Sections 2 and 3 for definitions of the phrases in italics).

It is quite remarkable that the moment method continues to provide new insights into the distribution of random eigenvalues. With this article we take the reader on a tour that starts with Wigner's discovery and ends with the description of recent results, some yet unpublished. Along the way we try to explain a few developments in more detail while briefly pointing at others.

The first application of the moment method to the analysis of random eigenvalues appears almost accidental. In an effort to understand "the wave functions of quantum mechanical systems which are assumed to be so complicated that statistical considerations can be applied to them" Wigner introduces in [45] three types of ensembles of "real symmetric matrices of high dimensionality". Although he considers his results not satisfactory from a physical point of view, he expresses the hope that "the calculation which follows may have some independent interest". Moreover, the reader learns that one of the three models considered just serves "as an intermediate step". And it is only this auxiliary ensemble that we would now call a Wigner ensemble. Wigner names it the "random sign symmetric matrix" by which he understands $(2N + 1) \times (2N + 1)$ matrices for which the diagonal elements are zero and "non diagonal elements $v_{ik} = v_{ki} = \pm v$ have all the same absolute value but random signs".

In the short note [46] that appeared a few years later, Wigner remarks that the arguments of [45] show the semicircle law for a much larger class of real symmetric ensembles. He observes that, except for technical assumptions, two features of the model were essential for his proof: Firstly, stochastic independence of the matrix entries (as far as the symmetry permits) and, secondly, that all (or at least most) matrix entries are centered and have the same variance.

In Section 3 we present Wigner's proof with enough detail to make the significance of these two assumptions apparent. The remaining sections are then devoted to the discussion of results where at least one of these essential assumptions are weakened.

In Section 4 independence and centeredness of the matrix entries are kept. However, we allow the variances to vary as a function of the distance to the diagonal. The most prominent examples in this class are band matrices and we discuss them in detail.

A first step to loosen the assumption of independence is presented in Section 5. Its central result provides conditions on the number and location of matrix entries that may be dependent without affecting the validity of Wigner's reasoning. We call such dependence structures *sparse*. Sparse dependence structures appear for example in certain types of block random matrices that are used in modelling disordered systems in mesoscopic physics (see e. g. [1]).

In the last three sections 6–8 we report on results for ensembles with a dependence structure that is not sparse. This is largely uncharted territory. However, in recent years a number of special cases were analyzed using the method of moments. They show interesting phenomena that should be explored further. We divide the models into three groups. In the first group the correlations decay to 0 as the distance of the matrix entries becomes large in some prescribed metric. Then we look at those

ensembles for which the entries are drawn from Curie-Weiss random variables. Here the correlations have no spatial decay but decay for supercritical temperatures as the matrix dimension becomes large. Finally, we pick the matrix entries from an infinite sequence of exchangeable random variables. Here the correlations between matrix entries depend neither on their locations nor on the size of the matrix.

We close this introduction by stating what is not contained in this survey. One of the striking features of random matrix theory is the observation that local statistics of the eigenvalues obey universal laws that, somewhat surprisingly, have also arisen in certain combinatorial problems, in some models from statistical mechanics and even in the distribution of the non-trivial zeros of zeta-functions. By local statistics we mean statistics after local but deterministic rescaling so that the spacings between neighboring eigenvalues are of order 1. Examples are the statistics of spacings or the distribution of extremal eigenvalues. Such results were first obtained for Gaussian ensembles, i.e. Wigner ensembles with normally distributed entries. In this special case it is possible to derive an explicit formula for the joint distribution of eigenvalues that can then be analyzed using the method of orthogonal polynomials. In the Gaussian case this requires detailed asymptotic formulas for Hermite polynomials of large degree that had already been derived in the beginning of the twentieth century. The first step to prove universality beyond Gaussian ensembles was then taken about twenty years ago within the class of ensembles that are invariant under change of orthonormal bases. For such ensembles the eigenvectors are distributed according to Haar measure, the joint distribution of eigenvalues is still explicit, and the method of orthogonal polynomials works, albeit they generally do not belong to the well studied families of classical orthogonal polynomials (see for example [11, 12, 37] and references therein). It is only seven years ago that universality results for local statistics became available for Wigner matrices (see e. g. [14, 15, 43, 44, 22] and references therein). Since all of these results do not use the moment method, we will not discuss them in this paper.

There is one notable exception to what has just been said. The distribution of extremal eigenvalues (and consequently of the operator norm) can and has been investigated for Wigner ensembles on the local level, using the method of moments [40], see also [39] and references therein. However, this requires quite substantial extensions of the ideas that we explain and goes way beyond the scope of this paper. We therefore only state weaker results that might be considered as laws of large numbers for the operator norm and that can be proved with much less effort. Nevertheless, we do not discuss their proofs either and refer the reader to Section 2.3 of the textbook [42].

Finally, we mention that the moment method can also be applied to complex Hermitian matrices and to sample covariance matrices (also known as Wishart ensembles), but in the present article we always restrict ourselves to the case of real symmetric matrices to keep the presentation as elementary as possible.

Acknowledgment: The authors would like to thank the referee for an exceptionally careful reading of the manuscript and for a number of valuable suggestions.

2. Setup

We begin by setting the scene and fixing some notation.

Definition 2.1. A (real symmetric) *matrix ensemble* is a family

$X_N(i, j), i, j = 1, \dots, N, N \in \mathbb{N}$ of real valued random variables on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ such that $X_N(i, j) = X_N(j, i)$. We then denote by X_N the corresponding $N \times N$ -matrix, i.e.

$$X_N = \begin{pmatrix} X_N(1, 1) & X_N(1, 2) & \cdots & X_N(1, N) \\ X_N(2, 1) & X_N(2, 2) & \cdots & X_N(2, N) \\ \vdots & \vdots & & \vdots \\ X_N(N, 1) & X_N(N, 2) & \cdots & X_N(N, N) \end{pmatrix} \quad (1)$$

Since we deal exclusively with real symmetric matrices by 'matrix ensemble' we always mean a real symmetric one.

Definition 2.2. A (real symmetric) matrix ensemble is called *independent* if for each $N \in \mathbb{N}$ the random variables $X_N(i, j), 1 \leq i \leq j \leq N$ are independent. It is called *identically distributed*, if all $X_N(i, j)$ have the same distribution.

An independent and identically distributed matrix ensemble X_N is called a *Wigner ensemble* if $\mathbb{E}(X_N(i, j)) = 0$ and $\mathbb{E}(X_N(i, j)^2) = 1$.

By a slight abuse of language we use the phrase ' X_N is a Wigner matrix' to indicate that the family X_N of random (symmetric) matrices form a Wigner ensemble. Some authors allow for Wigner ensembles a probability distribution for the diagonal elements which differs from the distribution for the nondiagonal entries.

Definition 2.3. The k^{th} *moment* of a random variable X is the expectation $\mathbb{E}(X^k)$. We say that all moments of X exist, if $\mathbb{E}(|X|^k) < \infty$ for all $k \in \mathbb{N}$.

Unless stated otherwise we always assume that all random variables occurring in this text have all moments existing.

For any symmetric $N \times N$ -matrix M we denote the eigenvalues of M by $\lambda_j(M)$. We order these eigenvalues such that

$$\lambda_1(M) \leq \lambda_2(M) \leq \dots \leq \lambda_N(M)$$

where degenerate eigenvalues are repeated according to their multiplicity.

The empirical eigenvalue distribution measure ν_N of M is defined by

$$\begin{aligned} \nu_N(A) &= \frac{1}{N} \left| \{j \mid \lambda_j(M) \in A\} \right| \\ &= \frac{1}{N} \sum_{j=1}^N \delta_{\lambda_j(M)}(A) \end{aligned}$$

where $|B|$ denotes the number of points in B , N - as above - is the dimension of the matrix M , A is a Borel-subset of \mathbb{R} and δ_a is the Dirac measure in a , i.e.

$$\delta_a(A) = \begin{cases} 1 & \text{if } a \in A \\ 0 & \text{otherwise.} \end{cases} \quad (2)$$

It turns out that for a Wigner matrix X_N the empirical eigenvalue distribution measure ν_N of X_N has no chance to converge as $N \rightarrow \infty$ as the following back-of-the-envelope calculations show.

We have

$$\begin{aligned} \int \lambda^2 d\nu_N(\lambda) &= \frac{1}{N} \sum_{\ell=1}^N \lambda_\ell(X_N)^2 \\ &= \frac{1}{N} \operatorname{tr} X_N^2 \end{aligned} \quad (3)$$

If the $N \times N$ -matrix X_N has entries ± 1 (random or not), then (3) shows

$$\int \lambda^2 d\nu_N(\lambda) = N \quad (4)$$

and if the X_N are random matrices with $\mathbb{E}(X_N(i, j)^2) = 1$ we get

$$\mathbb{E} \left(\int \lambda^2 d\nu_N(\lambda) \right) = N. \quad (5)$$

This shows that (at least the second moment of) the empirical eigenvalue distribution measure of X_N is divergent.

Moreover, the same calculation suggests that the empirical eigenvalue distribution measure of the normalized matrices

$$M_N = \frac{1}{\sqrt{N}} X_N$$

might converge as for M_N

$$\mathbb{E} \int \lambda^2 d\nu_N = \frac{1}{N} \operatorname{tr} M_N^2 = 1.$$

As we shall see below, this is indeed the case not only for Wigner ensembles, but for a huge class of random matrices.

A similar reasoning applies to the operator norm of a matrix ensemble X_N :

$$\|X_N\| = \max \left\{ |\lambda_1(X_N)|, |\lambda_N(X_N)| \right\}. \quad (6)$$

Since for any real symmetric $N \times N$ -matrix M :

$$\frac{1}{N} \operatorname{tr} M^2 = \frac{1}{N} \sum_{\ell=1}^N \lambda_{\ell}(M)^2 \leq \|M\|^2 \leq \operatorname{tr} M^2 \quad (7)$$

a matrix M with ± 1 -entries satisfies

$$\sqrt{N} \leq \|M\| \leq N$$

and similarly for $\mathbb{E}(X_N(i, j)^2) = 1$

$$\sqrt{N} \leq \mathbb{E}(\|X_N\|^2)^{\frac{1}{2}} \leq N.$$

Again, one is lead to look at the norm of $M_N = \frac{1}{\sqrt{N}} X_N$.

Indeed, for Wigner ensemble the norm of M_N will stay bounded as $N \rightarrow \infty$, in fact, it will converge to 2.

However, this fact is more subtle than the convergence of ν_N , and so is its proof (cf. Theorem 3.13 that was proved by Füredi and Komlós in [20], see also [5], and [42] for a textbook presentation).

To illustrate this, let us look at a particular example within the class considered in (4), namely the $N \times N$ -matrices

$$\mathcal{E}_N(i, j) = 1 \text{ for all } 1 \leq i, j \leq N. \quad (8)$$

The matrix \mathcal{E}_N can be written as

$$\mathcal{E}_N = N \cdot P_e$$

where P_e is the orthogonal projection onto the vector e , with $e(i) = \frac{1}{\sqrt{N}}$ for $i = 1, \dots, N$.

Consequently \mathcal{E}_N is of rank 1 and

$$\lambda_j(\mathcal{E}_N) = \begin{cases} N & \text{for } j = N \\ 0 & \text{otherwise.} \end{cases} \quad (9)$$

Thus we obtain

$$\left\| \frac{\mathcal{E}_N}{\sqrt{N}} \right\| = \sqrt{N} \rightarrow \infty$$

but the eigenvalue distribution function ν_N of \mathcal{E}_N/\sqrt{N} is given by

$$\nu_N = \frac{1}{N}(N-1)\delta_0 + \frac{1}{N}\delta_{\sqrt{N}} \implies \delta_0$$

where \implies means weak convergence (see definition 3.1).

3. Wigner's Semicircle Law

In this section we present and discuss the classical semicircle law for Wigner ensembles.

So, let X_N be a Wigner ensemble (see Definition 2.2), set $M_N = \frac{1}{\sqrt{N}}X_N$ and denote the empirical eigenvalue distribution measure of M_N by σ_N , thus

$$\sigma_N(A) = \frac{1}{N} \left| \{j \mid \lambda_j\left(\frac{1}{\sqrt{N}}X_N\right) \in A\} \right|.$$

The semicircle law, in its original form due to Wigner ([45], [46]), states that σ_N converges to the semicircle distribution σ given through its Lebesgue density

$$\sigma(x) = \begin{cases} \frac{1}{2\pi} \sqrt{4-x^2} & , \text{ for } |x| \leq 2 \\ 0 & , \text{ otherwise.} \end{cases} \quad (10)$$

σ describes a semicircle of radius 2 around the origin, hence the name.

So far, we have avoided to explain in which sense σ_N converges. This is what we do now.

Let us first look at the convergence of measures on \mathbb{R} .

Definition 3.1. Suppose μ_N and μ are probability measures on \mathbb{R} (equipped with the Borel σ -algebra $B(\mathbb{R})$).

We say that μ_N converges weakly to μ , in symbols

$$\mu_N \Rightarrow \mu$$

if

$$\int f(x) d\mu_N(x) \rightarrow \int f(x) d\mu(x)$$

for all $f \in C_b(\mathbb{R})$, the space of bounded continuous functions.

If the matrix X_N is random and

$$\sigma_N = \frac{1}{N} \sum \delta_{\lambda_j} \left(\frac{X_N}{\sqrt{N}} \right)$$

is the empirical eigenvalue distribution measure of $\frac{1}{\sqrt{N}}X_N$ then the measure σ_N itself is random.

Consequently, we have not only to define in which sense the measures converge (namely weakly), but also how this convergence is meant with respect to randomness, i.e. to the 'parameter' $\omega \in \Omega$. There are various ways to do this.

Definition 3.2. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let μ_N^ω and μ^ω be random probability measures on $(\mathbb{R}, B(\mathbb{R}))$.

- 1) We say that μ_N^ω converges to μ^ω *weakly in expectation*, if for every $f \in C_b(\mathbb{R})$

$$\mathbb{E}\left(\int f(x) d\mu_N^\omega(x)\right) \rightarrow \mathbb{E}\left(\int f(x) d\mu^\omega(x)\right) \quad (11)$$

as $N \rightarrow \infty$.

- 2) We say that μ_N^ω converges to μ^ω *weakly in probability*, if for every $f \in C_b(\mathbb{R})$ and any $\varepsilon > 0$

$$\mathbb{P}\left(\left|\int f(x) d\mu_N^\omega(x) - \int f(x) d\mu^\omega(x)\right| > \varepsilon\right) \rightarrow 0$$

as $N \rightarrow \infty$.

- 3) We say that μ_N^ω converges to μ^ω *weakly \mathbb{P} -almost surely* if there is a set $\Omega_0 \subset \Omega$ with $\mathbb{P}(\Omega_0) = 1$ such that $\mu_N^\omega \Rightarrow \mu^\omega$ for all $\omega \in \Omega_0$.

Theorem 3.3 (Semicircle Law). *Suppose X_N is a Wigner ensemble with*

$\mathbb{E}(|X_N(i, j)|^k) < \infty$ for all $k \in \mathbb{N}$ and let σ_N denote the empirical eigenvalue distribution measure of $M_N = \frac{1}{\sqrt{N}}X_N$ then σ_N converges to the semicircle distribution σ weakly \mathbb{P} -almost surely.

Remarks 3.4. 1. Wigner [45, 46] proved this theorem for weak convergence in expectation.

2. Grenander [23] showed under the same conditions that the convergence holds weakly in probability.
3. Arnold [4] proved that the convergence is weakly \mathbb{P} -almost surely. He also relaxed the moment condition to

$$\mathbb{E}\left(X_N(i, j)^6\right) < \infty$$

for \mathbb{P} -almost sure weak convergence and to

$$\mathbb{E}\left(X_N(i, j)^4\right) < \infty$$

for weak convergence in probability.

4. According to Definition 2.2 the entries in a Wigner ensemble are independent and identically distributed. Hence a condition of the form $\mathbb{E}(|X_N(i, j)|^k) < \infty$, as it appears for example in the previous remark, actually implies

$$\sup_{N, i, j} \mathbb{E}(|X_N(i, j)|^k) < \infty .$$

Besides the moment method we discuss in this article there is another important technique to prove the semicircle law. This is the Stieltjes transform method originating in [27], [35] and [36], see also [37] and references given there. Both methods are discussed in [2] and in [42].

The moment method is based on the observation that the following result is true.

Proposition 3.5. *If μ_N and μ are probability measures on \mathbb{R} such that all moments of μ_N exist and*

$$\int |x|^k d\mu(x) \leq AC^k k! \quad (12)$$

for all k and some constants A, C , then

$$\int x^k d\mu_N(x) \rightarrow \int x^k d\mu(x)$$

for all $k \in \mathbb{N}$ implies that

$$\mu_N \Rightarrow \mu.$$

For a proof see for example [8], [32] or [28].

Since the semicircle distribution σ has compact support, it obviously satisfies (12). The moments of σ are given by:

$$\int x^k d\sigma(x) = \begin{cases} C_{k/2}, & \text{if } k \text{ is even;} \\ 0, & \text{if } k \text{ is odd.} \end{cases} \quad (13)$$

where

$$C_\ell = \frac{1}{\ell + 1} \binom{2\ell}{\ell} \quad (14)$$

are the *Catalan numbers*. (For a concise introduction to Catalan numbers see e. g. [33] or [41].)

The moments of σ_N can be expressed through traces of the matrices X_N

$$\begin{aligned} \mathbb{E}\left(\int x^k d\sigma_N(x)\right) &= \frac{1}{N} \mathbb{E}\left(\sum_{j=1}^N \lambda_j \left(\frac{X_N}{\sqrt{N}}\right)^k\right) = \frac{1}{N} \mathbb{E}\left(\text{tr}\left(\frac{X_N}{\sqrt{N}}\right)^k\right) \\ &= \frac{1}{N^{1+\frac{k}{2}}} \sum_{i_1, \dots, i_k=1}^N \mathbb{E}\left(X_N(i_1, i_2) \cdot X_N(i_2, i_3) \cdot \dots \cdot X_N(i_k, i_1)\right). \end{aligned} \quad (15)$$

The sum in (15) contains N^k terms. So, at a first glance, the normalizing factor $N^{1+k/2}$ seems too small to compensate the growth of the sum. Fortunately, many of the summands are zero, as we shall see later.

For the purpose of bookkeeping it is useful to think of i_1, i_2, \dots, i_k in terms of a graph.

Definition 3.6. The multigraph \mathcal{G} with vertex set

$$\mathcal{V} := \{i_1, i_2, \dots, i_k\} \quad (16)$$

and ℓ (undirected) edges between i and j if $\{i, j\}$ occurs ℓ times in the sequence

$$\{i_1, i_2\}, \{i_2, i_3\}, \dots, \{i_k, i_1\} \quad (17)$$

is called the multigraph associated with (i_1, i_2, \dots, i_k) .

Remark 3.7. The sequence (i_1, i_2, \dots, i_k) defines a *multigraph* since there may be several edges between the vertices i_ν .

Definition 3.8. If \mathcal{G} is a multigraph we define the associated (simple) graph $\tilde{\mathcal{G}}$ in the following way. The set of vertices of $\tilde{\mathcal{G}}$ is the same as the vertex set of \mathcal{G} and $\tilde{\mathcal{G}}$ has a single edge between i and j whenever \mathcal{G} has at least one edge between i and j .

Remark 3.9. The sequence (17) describes not only a multigraph but in addition a closed path through the multigraph which uses each edge exactly once. Such paths are called *Eulerian circuits*. They occur for example in the famous problem of the ‘Seven Bridges of Königsberg’ (see e. g. [7]).

The existence of an Eulerian circuit implies in particular that the multigraph is connected.

Now, we order the sum in (15) according to the number $|\mathcal{V}| = |\{i_1, \dots, i_k\}|$ of *different* indices (=vertices) occurring in the sequence i_1, i_2, \dots, i_k .

$$\begin{aligned} & \sum_{i_1, \dots, i_k=1}^N \mathbb{E} \left(X_N(i_1, i_2) \cdot X_N(i_2, i_3) \cdot \dots \cdot X_N(i_k, i_1) \right) \\ &= \sum_{r=1}^k \sum_{|\{i_1, \dots, i_k\}|=r} \mathbb{E} \left(X_N(i_1, i_2) \cdot X_N(i_2, i_3) \cdot \dots \cdot X_N(i_k, i_1) \right) \end{aligned} \quad (18)$$

The number of index tuples (i_1, \dots, i_k) with $|\{i_1, \dots, i_k\}| = r$ is of order $\mathcal{O}(N^r)$ and can be bounded above by $r^k N^r$. In fact, to choose the r different numbers in $\{1, \dots, N\}$ we have less than N^r possibilities. Then, to choose which one to put at a given position we have at most r choices for each of the k positions.

Therefore the sum

$$\sum_{|\{i_1, \dots, i_k\}|=r} \mathbb{E} \left(X_N(i_1, i_2) \cdot X_N(i_2, i_3) \cdot \dots \cdot X_N(i_k, i_1) \right) \quad (19)$$

is of order $\mathcal{O}(N^r)$ as well. Thus the terms with $r = |\{i_1, \dots, i_k\}| < 1 + k/2$ in (15) can be neglected compared to prefactor $N^{-(1+k/2)}$. Consequently

$$\frac{1}{N^{1+k/2}} \sum_{|\{i_1, \dots, i_k\}| < 1+k/2} \mathbb{E} \left(X_N(i_1, i_2) \cdot \dots \cdot X_N(i_k, i_1) \right) \longrightarrow 0 \quad (20)$$

To handle those terms with $|\{i_1, \dots, i_k\}| > 1 + k/2$ we need the following two observations.

For comparison with results in Section 7 we formulate the first one as a lemma.

Lemma 3.10. *Whenever an edge $\{i, j\}$ occurs only once in (17) then*

$$\mathbb{E}\left(X_N(i_1, i_2) \cdot X_N(i_2, i_3) \cdot \dots \cdot X_N(i_k, i_1)\right) = 0. \quad (21)$$

This follows from independence and the assumption $\mathbb{E}(X_N(i, j)) = 0$.

The second observation is:

Proposition 3.11. *If $|\{i_1, \dots, i_k\}| > 1 + k/2$ there is an edge $\{i, j\}$ which occurs only once in $\{i_1, i_2\}, \{i_2, i_3\}, \dots, \{i_k, i_1\}$.*

Proof: Set $r = |\{i_1, \dots, i_k\}|$ and denote the distinct elements of $\{i_1, \dots, i_k\}$ by j_1, \dots, j_r .

To connect the vertices j_1, \dots, j_r we need at least $r - 1$ edges. To double each of these connections we need $2r - 2$ edges. So, if we have k edges we need that $k \geq 2r - 2$ to double each connection. Hence, if $r > 1 + k/2$, at least one edge occurs only once. \square

Remark 3.12. A similar reasoning as in the proof above shows: If a graph \mathcal{G} with k edges and $k + 1$ vertices is connected then \mathcal{G} is a *tree*, i. e. \mathcal{G} contains no loops. Indeed, if \mathcal{G} contained a loop we could remove an edge without destroying the connectedness of the graph. But the new graph would have $k - 1$ edges and $k + 1$ vertices, so it cannot be connected.

From Proposition 3.11 and (21) we learn that

$$\sum_{|\{i_1, \dots, i_k\}| > 1 + k/2} \mathbb{E}\left(X_N(i_1, i_2) \cdot X_N(i_2, i_3) \cdot \dots \cdot X_N(i_k, i_1)\right) = 0. \quad (22)$$

To summarize, what we proved so far is

$$\begin{aligned} & \frac{1}{N^{1+\frac{k}{2}}} \sum_{i_1, \dots, i_k=1}^N \mathbb{E}\left(X_N(i_1, i_2) \cdot X_N(i_2, i_3) \cdot \dots \cdot X_N(i_k, i_1)\right) \\ & \approx \frac{1}{N^{1+\frac{k}{2}}} \sum_{\substack{|\{i_1, \dots, i_k\}| = 1 + k/2 \\ \text{all } \{i, j\} \text{ occur exactly twice}}} \mathbb{E}\left(X_N(i_1, i_2) \cdot X_N(i_2, i_3) \cdot \dots \cdot X_N(i_k, i_1)\right) \end{aligned} \quad (23)$$

Let us set

$$\mathcal{I}_k^{(N)} = \left\{ (i_1, \dots, i_k) \in \{1, \dots, N\} \mid |\{i_1, \dots, i_k\}| = 1 + k/2 \right. \\ \left. \text{and all } \{i, j\} \text{ occur exactly twice.} \right\} \quad (24)$$

For odd k the set $\mathcal{I}_k^{(N)}$ is empty, so the sum (23) is obviously zero.

Due to independence and the assumptions $\mathbb{E}(X_N(i, j)) = 0$ and $\mathbb{E}(X_N(i, j)^2) = 1$ we have

$$\mathbb{E}\left(X_N(i_1, i_2) \cdot X_N(i_2, i_3) \cdot \dots \cdot X_N(i_k, i_1)\right) = 1$$

whenever all $\{i, j\}$ occur exactly twice.

Consequently,

$$\text{Right side of (23)} = \frac{1}{N^{1+\frac{k}{2}}} |\mathcal{I}_k^{(N)}|. \quad (25)$$

For even k , let us consider the multigraph \mathcal{G} associated with $(i_1, \dots, i_k) \in \mathcal{I}_k^{(N)}$. Since \mathcal{G} has $1 + k/2$ vertices and k double vertices, the corresponding simple graph $\tilde{\mathcal{G}}$ is a connected graph with $1 + k/2$ vertices and $k/2$ edges. Thus, this $\tilde{\mathcal{G}}$ is a tree by Remark 3.12. Moreover the path (i_1, \dots, i_k, i_1) defines an ordering on $\tilde{\mathcal{G}}$.

The number of ordered trees [33, 41] with ℓ edges (and hence $\ell + 1$ vertices) is known to be the Catalan number C_ℓ (see (14)).

Given an (abstract) ordered tree with $\ell = 1 + k/2$ vertices we find all corresponding paths $(i_1, i_2, \dots, i_k, i_1)$ with $i_j \in \{1, \dots, N\}$ by assigning $1 + k/2$ (different) numbers (=indices) from $\{1, \dots, N\}$ to the vertices of the tree. There are $\frac{N!}{(N-(1+k/2))!} \approx N^{1+k/2}$ ways to do this. Thus

$$\mathbb{E}\left(\int x^k d\sigma_N(x)\right) \approx \frac{1}{N^{1+\frac{k}{2}}} |\mathcal{I}_k^{(N)}| \quad (26)$$

$$\rightarrow \begin{cases} C_{\frac{k}{2}} & \text{for } k \text{ even} \\ 0 & \text{for } k \text{ odd.} \end{cases} \quad (27)$$

and these are the moments of the semicircle distribution σ (see (13)). In view of Proposition 3.5 this proves that σ_N converges to σ weakly in expectation (cf. Definition 3.2).

For more details on the semicircle law and its proof see [2], [42] or [28].

From Theorem 3.3 and (10) we conclude that $\liminf \|\frac{1}{\sqrt{N}}X_N\| \geq 2$ almost surely, since for symmetric $N \times N$ -matrices A the matrix norm $\|A\|$, as an operator on the Euclidian space \mathbb{R}^N , satisfies $\|A\| = \max\{|\lambda_1(A)|, |\lambda_N(A)|\}$.

However, Theorem 3.3 does *not* imply that $\liminf \|\frac{1}{\sqrt{N}}X_N\| \leq 2$! Wigner's result does imply that the *majority* of the eigenvalues will be less than $2 + \varepsilon$ finally, however some (in fact even $o(N)$) eigenvalues could be bigger and might even go to ∞ . In Sections 4, 7, and 8 we encounter ensembles for which exactly this happens.

However, for Wigner ensembles it is correct that the norm of $\frac{1}{\sqrt{N}}X_N$ goes to 2. This can be shown by a more sophisticated variant of the moment method.

Theorem 3.13. *Suppose X_N is a Wigner ensemble with $\mathbb{E}(|X_N(i, j)|^k) < \infty$ for all $k \in \mathbb{N}$ and let*

$$\lambda_N^* = \max \left\{ \left| \lambda_1 \left(\frac{1}{\sqrt{N}} X_N \right) \right|, \left| \lambda_N \left(\frac{1}{\sqrt{N}} X_N \right) \right| \right\} = \left\| \frac{1}{\sqrt{N}} X_N \right\|$$

be the operator norm of $M_N = \frac{1}{\sqrt{N}} X_N$, then

$$\lambda_N^* \rightarrow 2 \quad \text{as } N \rightarrow \infty \quad \mathbb{P}\text{-almost surely.}$$

This theorem was proved by Füredi and Komlós in [20], see also [5].

To prove the semicircle law we considered the k^{th} moment m_k of σ_N for fixed k as N goes to infinity. For the norm estimate we need bounds on m_k for $k = k_N$ for a sequence k_N which is growing with N . See [42, Section 2.3] for a pedagogical explanation.

4. Random Band Matrices

In a first variation of Wigner's semicircle law we abandon the assumption of identical distribution of the $X_N(i, j)$, by assuming that entries away from a band around the diagonal are zero, while the other entries are still iid, apart from the symmetry $X_N(i, j) = X_N(j, i)$.

More precisely, let $\tilde{X}_N(i, j)$ be a Wigner ensemble and set

$$X_N(i, j) = \begin{cases} \tilde{X}_N(i, j) & \text{for } |i - j| \leq b_N \\ 0 & \text{otherwise.} \end{cases} \quad (28)$$

where b_N is a sequence of integers with $b_N \rightarrow \infty$ and $2b_N + 1 \leq N$.

We call such matrices banded Wigner matrices with band width $\beta_N = 2b_N + 1$.

There is a 'Semicircle Law' for banded Wigner matrices due to Bogachev, Molchanov and Pastur [6].

Theorem 4.1. *Suppose X_N is a banded Wigner matrix with band width*

$\beta_N = 2b_N + 1 \leq N$ and assume that all moments of $X_N(i, j)$ exist.

Set $M_N = \frac{1}{\sqrt{\beta_N}} X_N$ and denote by σ_N the empirical eigenvalue distribution measure of M_N .

- 1) If $\beta_N \rightarrow \infty$ but $\frac{\beta_N}{N} \rightarrow 0$ then the σ_N converges to the semicircle distribution weakly in probability.
- 2) If $\beta_N \approx cN$ for some $c > 0$ then σ_N converges weakly in probability to a measure $\tilde{\sigma}$ which is not the semicircle distribution.

It turns out that the moment method used to prove Wigner's result can also be applied to banded random matrices.

Let us look at the products

$$X_N(i_1, i_2) \cdot X_N(i_2, i_3) \cdot \dots \cdot X_N(i_k, i_1)$$

which occur in evaluating traces as in (15).

We have N possibilities to choose i_1 . In principle, for i_2 we have again N possibilities. However, unlike to the Wigner case, at most β_N of these possibilities are not identically zero. This observation makes it plausible that

$$\sum_{i_1, \dots, i_k} \mathbb{E} \left(X_N(i_1, i_2) \cdot \dots \cdot X_N(i_k, i_1) \right) \approx N \beta_N^{k/2}$$

since - again - only those terms with each $\{i, j\}$ occurring exactly twice count in the limit. Note that our assumption $\beta_N \rightarrow \infty$ is needed here. Without this assumptions pairs $\{i, j\}$ occurring more than twice are not negligible.

Unfortunately, the above argument is not *quite* correct. It is true that most columns (and rows) contain β_N entries $X_N(i, j)$ which are not identically equal to zero.

However, this is wrong for the rows with row number j when

$$j \leq b_N \text{ or } j > N - b_N,$$

i.e. in the 'corners' of the matrix.

Thus for any $1 \leq \ell < k$ for which the vertex $i_{\ell+1}$ is new in the path, i.e. $i_{\ell+1} \notin \{i_1, \dots, i_\ell\}$ we have at least $\beta_N - \ell$ choices for $i_{\ell+1}$ *only* if

$$b_N < i_\ell \leq N - b_N.$$

If $\frac{b_N}{N} \rightarrow 0$ (as in case 1 of the theorem) the number of exceptions (i.e. $i_\ell \leq b_N$ or $i_\ell > N - b_N$) is negligible and the semicircle law is again valid.

However, if b_N grows proportional to N the 'exceptional' terms are not exceptional any more but rather contribute in the limit $N \rightarrow \infty$.

For details of the proof see [6] or [10].

The above argument suggests that for $b_N \approx cN$ the limit distribution might be again the semicircle distribution if we 'fill the corners' of the matrix appropriately. This can be achieved by the following modification of (28).

Definition 4.2. Set (for $i \in \mathbb{N}$)

$$|i|_N = \min \{|i|, |N - i|\} \tag{29}$$

and let \tilde{X}_N be a Wigner ensemble. Then we call the matrix

$$X_N(i, j) = \begin{cases} \tilde{X}_N(i, j) & \text{for } |i - j|_N \leq b_N \\ 0 & \text{otherwise.} \end{cases} \quad (30)$$

a periodic band matrix.

$|i - j|_N$ measures the distance of i and j on $\mathbb{Z}/N\mathbb{Z}$. The choice of $|\cdot|_N$ guarantees that each column (and each row) contains exactly $\beta_N = 2b_N + 1$ non zero (i.e. not identically zero) entries.

As we anticipated we have

Theorem 4.3. *If X_N is a periodic band random matrix with band width $\beta_N \leq N$ and $\beta_N \rightarrow \infty$, then the empirical eigenvalue distribution measure σ_N of $\frac{1}{\sqrt{\beta_N}}X_N$ converges weakly in probability to the semicircle distribution σ .*

A proof of this result due to Bogachev, Molchanov and Pastur can be found in [6] or in [10].

Catalano [10] has generalized the above result to matrices of the form

$$X_N(i, j) = \alpha\left(\frac{|i - j|}{N}\right) \tilde{X}_N(i, j) \quad (31)$$

where \tilde{X}_N is a Wigner matrix and $\alpha : [0, 1] \rightarrow \mathbb{R}$ a Riemann integrable function.

This class of matrices contains both random band matrices with $b_N \approx cN$ and periodic random band matrices, take either $\alpha(x) = \chi_{[0, c]}(x)$ or $\alpha(x) = \chi_{[0, c] \cup [1-c, 1]}(x)$,

where $\chi_A(x) = \begin{cases} 1, & \text{if } x \in A; \\ 0, & \text{otherwise.} \end{cases}$

Theorem 4.4. *Let X_N be a matrix ensemble as in (31), set*

$$\Phi := \int_0^1 \int_0^1 \alpha^2(|x - y|) dx dy$$

and let σ_N be the empirical eigenvalue distribution measure for $\frac{1}{\sqrt{\Phi N}}X_N$. Then σ_N converges weakly in probability to a limit measure τ .

The limit τ is the semicircle law if and only if

$$|\alpha(x)| = |\alpha(1 - x)| \quad (32)$$

for almost all $x \in \mathbb{R}$.

Note, that in the case of band matrices with bandwidth proportional to N condition (32) is fulfilled for the periodic case, but not for the non periodic case (28).

As for the Wigner case the question arises whether the norm of band matrices is bounded in the limit $N \rightarrow \infty$. In fact we have:

Theorem 4.5. *Let X_N be a banded Wigner ensemble (as in (28)) with band width $\beta_N \leq N$ and assume that all moments of $X_N(i, j)$ exist. If there are positive constants γ and C such that $\beta_N \geq C N^\gamma$ for all N then*

$$\limsup_{N \rightarrow \infty} \left\| \frac{1}{\sqrt{\beta_N}} X_N \right\| \leq 2 \quad (33)$$

\mathbb{P} -almost surely.

A proof of Theorem 4.5 is contained in the forthcoming paper [30]. This theorem applies to periodic band matrices as well.

Bogachev, Molchanov and Pastur [6] show that the norm of $\frac{1}{\sqrt{\beta_N}} X_N$ can go to infinity if β_N grows only on a logarithmic scale with N .

We mention that there are various other results about matrices with independent, but not identically distributed random variables. Already the papers [35] and [36] consider matrix entries with constant variances but not necessarily identical distribution. The identical distribution of the entries is replaced by a (far weaker) condition of Lindeberg type. In the paper [21] even the condition of constant variances is relaxed. Moreover these authors replace independence by a martingale condition.

5. Sparse Dependencies

Now we turn to attempts to weaken the assumption of independence between the $X_N(i, j)$ of a matrix ensemble. We start with what we call ‘sparse dependencies’.

This means that, while we don’t care *how* some of the $X_N(i, j)$ depend on each other, we restrict the *number* of dependencies in a way specified below. We follow Schenker and Schulz-Baldes [38] in this section.

We assume that for each N there is an equivalence relation \sim_N on \mathbb{N}_N^2 with $\mathbb{N}_N = \{1, 2, \dots, N\}$ and we suppose that the random variables $X_N(i, j)$ and $X_N(k, \ell)$ for $1 \leq j, k \leq \ell$ are independent unless (i, j) and (k, ℓ) belong to the same equivalence class with respect to \sim_N .

Definition 5.1. We call the equivalence relations \sim_N *sparse*, if the following conditions are fulfilled:

- 1) $\max_{i \in \mathbb{N}_N} |\{(j, k, \ell) \in \mathbb{N}_N^3 | (i, j) \sim_N (k, \ell)\}| = o(N^2)$
- 2) $|\{(i, j, \ell) \in \mathbb{N}_N^3 | (i, j) \sim_N (j, \ell) \text{ and } \ell \neq i\}| = o(N^2)$
- 3) $\max_{i, j, k \in \mathbb{N}_N} |\{\ell \in \mathbb{N}_N | (i, j) \sim_N (k, \ell)\}| \leq B$ for an N -independent constant B .

Definition 5.2. A symmetric random matrix ensemble $X_N(i, j)$ with $\mathbb{E}(X_N(i, j)) = 0$, $\mathbb{E}(X_N(i, j)^2) = 1$ and $\sup_{N, i, j} \mathbb{E}(X_N(i, j)^k) < \infty$ for all $k \in \mathbb{N}$ is called a *generalized Wigner ensemble with sparse dependence structure* if there are sparse equivalence relations \sim_N , such that $X_N(i, j)$ and $X_N(k, \ell)$ are independent if $(i, j) \sim_N (k, \ell)$.

Examples 5.3. If A_N and B_N are Wigner matrices, then the $2N \times 2N$ -matrices

$$X_N = \begin{pmatrix} A_N & B_N \\ B_N & -A_N \end{pmatrix}$$

and

$$X'_N = \begin{pmatrix} A_N & B_N \\ B_N & A_N \end{pmatrix}$$

are generalized Wigner ensembles with sparse dependence structure.

Many more example classes can be found in [26].

Theorem 5.4. *If X_N is a generalized Wigner ensemble with sparse dependence structure and σ_N is the empirical eigenvalue distribution measure of $M_N = \frac{1}{\sqrt{N}}X_N$ then σ_N converges to the semicircle distribution weakly in probability.*

This theorem is due to Schenker and Schulz-Baldes [38] who proved weak convergence in expectation, for convergence in probability see [10].

Catalano [10] combines sparse dependence structures with generalized band structures as in (31).

6. Decaying Correlations

In this section we discuss some matrix ensembles for which the random variables $X_N(i, j)$ have decaying correlations. We begin by what we call ‘diagonal’ ensembles. By this we mean that the random variables $X_N(i, j)$ and $X_N(i', j')$ are independent if the index pairs (i, j) and (i', j') belong to different diagonals, i. e. if $i - j \neq i' - j'$ (for $i \leq j$ and $i' \leq j'$).

Definition 6.1. Suppose Y_n is a sequence of random variables and $Y_n^{(\ell)}$, $\ell \in \mathbb{N}$ are independent copies of Y_n , then the matrix ensemble

$$X_N(i, j) = Y_i^{(|i-j|)} \quad \text{for } 1 \leq i \leq j \leq N \quad (34)$$

is called the *matrix ensemble with independent diagonals* generated by Y_n .

Of course, if the random variables Y_n themselves are independent then we obtain an independent matrix ensemble. If, on the other hand, $Y_n = Y_1$ we get a matrix with constant entries along each diagonal, which vary randomly from diagonal to diagonal. Such a matrix is thus a *random Toeplitz matrix*. Random Toeplitz matrices were considered by Bryc, Dembo and Jiang in [9]. They prove:

Theorem 6.2. *Suppose that $X_N(i, j)$ is the random Toeplitz matrix ensemble associated with $Y_n = Y$ with $\mathbb{E}(Y) = 0$, $\mathbb{E}(Y^2) = 1$ and $\mathbb{E}(Y^K) < \infty$ for all K , then the empirical eigenvalue distribution measures σ_N of $\frac{1}{\sqrt{N}} X_N$ converge weakly almost surely to a nonrandom measure γ which is independent of the distribution of Y and has unbounded support.*

In particular, γ is not the semicircle distribution.

Friesen and Löwe [18] consider matrix ensembles with independent diagonals generated by a sequence Y_n of weakly correlated random variables. In their case the limit distribution is the semicircle law again.

Theorem 6.3. *Let Y_n be a stationary sequence of random variables with $\mathbb{E}(Y_1) = 0$, $\mathbb{E}(Y_1^2) = 1$ and $\mathbb{E}(Y_1^K) < \infty$ for all K . Assume*

$$\sum_{\ell=1}^{\infty} |\mathbb{E}(Y_1 Y_{1+\ell})| < \infty. \quad (35)$$

Let X_N be the matrix ensemble with independent diagonals generated by Y_n . Then the empirical eigenvalue distribution measures σ_N of $\frac{1}{\sqrt{N}} X_N$ converge to the semicircle distribution \mathbb{P} -almost surely.

The next step away from independence is to start with a sequence $\{Z_n\}_{n \in \mathbb{N}}$ of random variables and to distribute them in some prescribed way on the matrix entries $X_N(i, j)$. It turns out (see [34]) that the validness of the semicircle law depends on the way we fill the matrix with the random number Z_n .

One main example of a filling is the ‘diagonal’ one, resulting in:

$$X_N = \begin{pmatrix} Z_1 & Z_{N+1} & Z_{2N} & \cdots & \cdots & \cdots & Z_{\frac{N(N+1)}{2}} \\ Z_{N+1} & Z_2 & Z_{N+2} & \cdots & \cdots & \cdots & \cdots \\ Z_{2N} & Z_{N+2} & Z_3 & Z_{N+3} & \cdots & \cdots & \cdots \\ \cdots & Z_{2N+1} & Z_{N+3} & Z_4 & \cdots & \cdots & \cdots \\ \vdots & \vdots & & & \vdots & & \\ Z_{\frac{N(N+1)}{2}} & \cdots & \cdots & \cdots & \cdots & Z_{2N-1} & Z_N \end{pmatrix} \quad (36)$$

Löwe and Schubert define abstractly:

Definition 6.4. A *filling* is a sequence of bijective mappings

$$\varphi_N : \left\{1, 2, \dots, \frac{N(N+1)}{2}\right\} \longrightarrow \{(i, j) \in \{1, 2, \dots, N\}^2 \mid i \leq j\} \quad (37)$$

If Z_n is a stochastic process and $\{\varphi_N\}$ is a filling we say that

$$X_N(i, j) = Z_{\varphi_N^{-1}(i, j)} \quad \text{for } 1 \leq i \leq j \leq N \quad (38)$$

is the matrix ensemble corresponding to $\{Z_n\}$ with filling $\{\varphi_N\}$

Another example of a filling, besides the ‘diagonal’ one, is the (symmetric) ‘row by row’ filling:

$$X_N = \begin{pmatrix} Z_1 & Z_2 & Z_3 & \cdots & \cdots & \cdots & Z_N \\ Z_2 & Z_{N+1} & Z_{N+2} & Z_{N+3} & \cdots & \cdots & Z_{2N-1} \\ Z_3 & Z_{N+2} & Z_{2N} & Z_{2N+1} & \cdots & \cdots & Z_{3N-3} \\ \cdots & Z_{N+3} & Z_{2N+1} & Z_{3N-1} & \cdots & \cdots & \cdots \\ \vdots & \vdots & & & \vdots & & \\ Z_N & Z_{2N-1} & \cdots & \cdots & \cdots & \cdots & Z_{\frac{N(N+1)}{2}} \end{pmatrix} \quad (39)$$

Among other results, Löwe and Schubert prove:

Theorem 6.5. *Suppose Z_n is an ergodic Markov chain with finite state space $S \subset \mathbb{R}$ started in its stationary measure and assume*

$$\mathbb{E}(Z_{n_1} Z_{n_2} \cdots Z_{n_k}) = 0 \quad (40)$$

$$\mathbb{E}(Z_n^2) = 1 \quad (41)$$

for any n and any n_1, \dots, n_k with k odd.

If X_N is the matrix ensemble corresponding to $\{Z_n\}$ with diagonal filling then the empirical eigenvalue distribution measures σ_N of $\frac{1}{\sqrt{N}} X_N$ converge to the semicircle distribution \mathbb{P} -almost surely.

The assumptions we made in Theorem 6.5 both on Z_n and on the filling are but an example of the abstract assumptions given in [34]. These authors also show:

Theorem 6.6. *There is an ergodic Markov chain $\{Z_n\}$ with finite state space $S \subset \mathbb{R}$ started in its stationary measure satisfying (40) and (41) such that for the matrix ensemble X_N corresponding to $\{Z_n\}$ with row by row filling the empirical eigenvalue distribution measures σ_N of $\frac{1}{\sqrt{N}} X_N$ do not converge to the semicircle distribution.*

Consequently, the convergence behavior of σ_N depends not only on the process $\{Z_n\}$ but also on the way we fill the matrices with this process. For details we refer to [34].

7. Curie-Weiss Ensembles

In section 6 we discussed matrix ensembles $X_N(i, j)$ which are generated through stochastic processes with decaying correlations. Thus, for fixed N , the correlations

$\mathbb{E} \left(X_N(i, j) X_N(k, \ell) \right)$ become small for (i, j) and (k, ℓ) far apart, in some appropriate sense.

In the present section we investigate matrix ensembles $X_N(i, j)$ with $\mathbb{E}(X_N(i, j)) = 0$ for which the correlations $\mathbb{E} \left(X_N(i, j) X_N(k, \ell) \right)$ do *not* depend on i, j, k, ℓ for most (or at least many) choices of i, j, k and ℓ , but the correlations depend on N instead.

More precisely, we will have that for given (i, j)

$$\mathbb{E} \left(X_N(i, j) X_N(k, \ell) \right) \sim C_N \geq 0$$

for $(k, \ell) \in B_N$ with $|B_N| \sim N$ or even $|B_N| \sim N^2$, and, as a rule, $C_N \rightarrow 0$. However, in Theorem 7.13 we will encounter an example for which C_N does not decay.

The main example we discuss comes from statistical physics, more precisely from the Curie-Weiss model.

Definition 7.1. Curie-Weiss random variables ξ_1, \dots, ξ_M take values in $\{-1, 1\}^M$ with probability

$$\mathbb{P}_\beta^M(\xi_1 = x_1, \dots, \xi_M = x_M) = Z^{-1} e^{\frac{\beta}{2M} (\sum_{i=1}^M x_i)^2} \quad (42)$$

where $Z = Z_{\beta, M}$ is a normalization constant (to make \mathbb{P}_β^M a probability measure) and $\beta \geq 0$ is a parameter which is interpreted in physics as ‘inverse temperature’, $\beta = \frac{1}{T}$.

If $\beta = 0$ ($T = \infty$) the random variables ξ_i are independent while for $\beta > 0$ there is a positive correlation between the ξ_i , so the ξ_i tend to have the same value $+1$ or -1 . This tendency is growing as $\beta \rightarrow \infty$. The Curie-Weiss model is used in physics as an easy model to describe magnetism. The ξ_i represent small magnets (‘spins’) which can be directed upwards (‘ $\xi_i = 1$ ’) or downwards (‘ $\xi_i = -1$ ’). At low temperature (high β) such systems tend to be aligned, i. e. a majority of the spins have the same direction (either upwards or downwards). For high temperature they behave almost like independent spins. These different types of behavior are described in the following theorem.

Theorem 7.2. Suppose ξ_1, \dots, ξ_M are \mathbb{P}_β^M -distributed Curie-Weiss random variables. Then the mean $\frac{1}{M} \sum_{i=1}^M \xi_i$ converges in distribution, namely

$$\frac{1}{M} \sum_{i=1}^M \xi_i \xrightarrow{\mathcal{D}} \begin{cases} \delta_0 & \text{if } \beta \leq 1 \\ \frac{1}{2}(\delta_{-m(\beta)} + \delta_{m(\beta)}) & \text{if } \beta > 1. \end{cases} \quad (43)$$

where $m = m(\beta)$ is the (unique) strictly positive solution of

$$\tanh(\beta m) = m \quad (44)$$

Above we used $\xrightarrow{\mathcal{D}}$ to indicate convergence in distribution: Random variables ζ_i converge in distribution to a measure μ if the distributions of ζ_i converge weakly to μ . Also, δ_x denotes the Dirac measure (see (2)).

For a proof of the above theorem see e. g. [13] or [28].

Theorem 7.2 makes the intuition from physics precise: The ξ_i satisfy a law of large numbers, like independent random variables do, if $\beta \leq 1$, in the sense that the distribution of $m_M = \frac{1}{M} \sum_{i=1}^M \xi_i$ converges weakly to zero, while m_M , the ‘mean magnetization’, equals $\pm m(\beta) \neq 0$ in the limit, with probability $\frac{1}{2}$ each, for $\beta > 1$.

In physics jargon, there is a phase transition for the Curie-Weiss model at $\beta = 1$, the ‘critical inverse temperature’.

We now discuss two matrix ensembles connected with Curie-Weiss random variables. The first one, which we call the *diagonal Curie-Weiss ensemble*, was introduced in [19]. It has independent ‘diagonals’ and the matrix entries within the same diagonal are Curie-Weiss distributed. Thus, it is closely related to the diagonal filling as defined in (36).

Definition 7.3. Let the random variables $\xi_1, \xi_2, \dots, \xi_N$ be \mathbb{P}_β^N -distributed Curie-Weiss random variables and take N independent copies of the ξ_i , which we call $\xi_1^1, \xi_2^1, \dots, \xi_N^1, \xi_1^2, \xi_2^2, \dots, \xi_N^2, \dots, \xi_1^N, \xi_2^N, \dots, \xi_N^N$.

Then we call the random matrix

$$X_N(i, i + \ell) := \xi_i^\ell \quad \text{for } \ell = 0, \dots, N - 1 \text{ and } i = 1, \dots, N - \ell \quad (45)$$

$$X_N(i, j) := X_N(j, i) \quad \text{for } i > j \quad (46)$$

the *diagonal Curie-Weiss ensemble* (with diagonal distribution \mathbb{P}_β^N).

For the diagonal Curie-Weiss ensemble Friesen and Löwe [19] prove the following result.

Theorem 7.4. Suppose X_N is a diagonal Curie-Weiss ensemble with diagonal distribution \mathbb{P}_β^N .

Then the empirical eigenvalue distribution measure σ_N of $\frac{1}{\sqrt{N}}X_N$ converges weakly almost surely to a measure σ_β . σ_β is the semicircle law σ if and only if $\beta \leq 1$.

Remarks 7.5. 1. The theorem shows that there is a phase transition for the eigenvalue distribution of the diagonal Curie-Weiss ensemble at $\beta = 1$.

2. The proof in [19] uses the moment method. It allows the authors to give an expression for the moments of σ_β in terms of $m(\beta)$ (see (43)).

For large β the empirical eigenvalue distribution measure of the diagonal Curie-Weiss ensemble approaches the eigenvalue distribution measure of random Toeplitz matrices we discussed in Theorem 6.2 (see Bryc, Dembo and Jiang [9]).

The second Curie-Weiss-type matrix ensemble, which we call the ‘full Curie-Weiss ensemble’, is defined as follows.

Definition 7.6. Take N^2 Curie-Weiss random variables $\tilde{X}_N(i, j)$ with distribution $\mathbb{P}_\beta^{N^2}$ and set

$$X_N(i, j) = \begin{cases} \tilde{X}_N(i, j) & \text{for } i \leq j \\ \tilde{X}_N(j, i) & \text{otherwise.} \end{cases} \quad (47)$$

We call the random matrix X_N defined above the *full Curie-Weiss ensemble*.

To our knowledge this ensemble was first considered in [25], where the following result was proved.

Theorem 7.7. *Let X_N be the full Curie-Weiss matrix ensemble with inverse temperature $\beta \leq 1$. Then the empirical eigenvalue distribution measure σ_N of $\frac{1}{\sqrt{N}}X_N$ converges weakly in probability to the semicircle distribution σ .*

The proof is based on the moment method we discussed in section 3. In [25] the authors prove this result just using assumptions on correlations of the $X_N(i, j)$ which are in particular satisfied by the full Curie-Weiss model if $\beta \leq 1$. Here, we only discuss this special case and refer to [25] for the more general case.

The main difficulty in this proof is the fact that for the Curie-Weiss ensemble it is *not* true that

$$\mathbb{E}_\beta^{N^2} \left(X_N(i_1, i_2) \cdot X_N(i_2, i_3) \cdot \dots \cdot X_N(i_k, i_1) \right) \quad (48)$$

is zero if an edge $\{i, j\}$ occurs only once in (48) (cf. (21) for the independent case). In other words, we need an appropriate substitute for Lemma 3.10.

So, we need a way to handle expectations as in (48) when there are edges (=index pairs, see Definition 3.6) which occur only once. Let us call such index pairs ‘single edges’.

Correlation estimates as we need them can be obtained from a special way of writing expectations \mathbb{E}_β^M with respect to the measure \mathbb{P}_β^M .

Definition 7.8. For $t \in [-1, 1]$ we denote by $P_t^{(1)}$ the probability measure on $\{-1, 1\}$ given by

$$P_t^{(1)}(1) = \frac{1}{2}(1+t) \quad \text{and} \quad P_t^{(1)}(-1) = \frac{1}{2}(1-t).$$

$P_t^{(M)}$ denotes the M -fold product of $P_t^{(1)}$ on $\{-1, 1\}^M$. If M is clear from the context we write P_t instead of $P_t^{(M)}$.

By E_t resp. $E_t^{(M)}$ we denote the corresponding expectation.

Proposition 7.9. *For any function ϕ on $\{-1, 1\}^M$ we have*

$$\mathbb{E}_\beta^M \left(\phi(X_1, \dots, X_M) \right) = \int_{-1}^1 E_t \left(\phi(X_1, \dots, X_M) \right) \frac{e^{-M \frac{1}{2} F_\beta(t)}}{1-t^2} dt \quad (49)$$

where $F_\beta(t) = \frac{1}{\beta} \left(\frac{1}{2} \ln \frac{1+t}{1-t} \right)^2 + \ln(1-t^2)$.

This proposition can be proved using the so called Hubbard-Stratonovich transformation. For a proof see [25] or [28]. The way to write expectations with respect to \mathbb{P}_β^M as a combination of independent measure is typical for exchangeable random variables and is known as de Finetti representation [16]. We will discuss this issue in detail in Section 8 and in particular in [29, 31].

The advantage of the representation (49) comes from the observation that under the probability measure P_t the random variables X_1, \dots, X_M are *independent* and the fact that the integral is in a form which is immediately accessible to the Laplace method for the asymptotic evaluation of integrals.

The Laplace method and Proposition 7.9 yield the required correlation estimates.

Proposition 7.10. *Suppose X_1, \dots, X_M are \mathbb{P}_β^M -distributed Curie-Weiss random variables.*

If ℓ is even, then as $M \rightarrow \infty$

1. *if $\beta < 1$*

$$\mathbb{E}_\beta^{(M)}(X_1 \cdot X_2 \cdot \dots \cdot X_\ell) \approx (l-1)!! \left(\frac{\beta}{1-\beta} \right)^{\frac{\ell}{2}} \frac{1}{M^{\frac{\ell}{2}}}$$

2. *if $\beta = 1$ there is a constant c_ℓ such that*

$$\mathbb{E}_\beta^{(M)}(X_1 \cdot X_2 \cdot \dots \cdot X_\ell) \approx c_\ell \frac{1}{M^{\frac{\ell}{4}}}$$

3. *if $\beta > 1$*

$$\mathbb{E}_\beta^{(M)}(X_1 \cdot X_2 \cdot \dots \cdot X_\ell) \approx m(\beta)^\ell$$

where $t = m(\beta)$, as in (44), is the strictly positive solution of $\tanh \beta t = t$.

If ℓ is odd then $\mathbb{E}_\beta^{(M)}(X_1 \cdot X_2 \cdot \dots \cdot X_\ell) = 0$ for all β .

We remind the reader that for an odd number k we set $k!! = k \cdot (k-2) \cdot \dots \cdot 3 \cdot 1$.

For proof of Proposition 7.10 see again [25] or [28].

From Proposition 7.10 we get immediately the following Corollary, which substitutes Lemma 3.10.

Corollary 7.11. *Let X_N be the full Curie-Weiss matrix ensemble with inverse temperature β and let the graph corresponding to the sequence i_1, i_2, \dots, i_k contain ℓ single edges.*

1. *If $\beta < 1$ then*

$$\left| \mathbb{E} \left(X_N(i_1, i_2) \cdot X_N(i_2, i_3) \cdot \dots \cdot X_N(i_k, i_1) \right) \right| \leq C N^{-\ell}. \quad (50)$$

2. If $\beta = 1$ then

$$\left| \mathbb{E} \left(X_N(i_1, i_2) \cdot X_N(i_2, i_3) \cdot \dots \cdot X_N(i_k, i_1) \right) \right| \leq C N^{-\ell/2}. \quad (51)$$

In the next step we have to prove a quantitative version of Proposition 3.11.

Proposition 7.12. *If $|\{i_1, \dots, i_k\}| \geq 1 + k/2 + s$ for some $s > 0$, then there are at least $2s + 2$ single edges in $\{i_1, i_2\}, \{i_2, i_3\}, \dots, \{i_k, i_1\}$.*

Proof: The proof is a refinement of the proof of Proposition 3.11.

Suppose \mathcal{G} is a multigraph with r vertices and k edges. Then, as we saw already, $k \geq r - 1$, if \mathcal{G} is connected. So, there are at most $k - r + 1$ edges left for ‘double’ connections. This means that there are at least $\ell = r - 1 - (k - r + 1)$ single edges and

$$\begin{aligned} \ell &= r - 1 - (k - r + 1) = 2r - k - 2 \\ &\geq (k + 2 + 2s) - k - 2 \\ &= 2s \end{aligned} \quad (52)$$

by assumption on r . So, by the above simple argument we are off the assertion by two only.

Now, we take into account that the sequence (i_1, \dots, i_k, i_1) defines a closed path through the graph. Since $|\{i_1, \dots, i_k\}| > 1 + k/2$ there is at least one single edge. If we remove one of the single edges from the graph, this new graph \mathcal{G}' is still connected. \mathcal{G}' has r vertices and $k - 1$ edges.

We redo the above argument with the graph \mathcal{G}' and get for the minimal number ℓ' of single edges in \mathcal{G}' equation (52) with k replaced by $k - 1$ and thus obtain

$$\begin{aligned} \ell' &= r - 1 - (k - 1 - r + 1) = 2r - k - 1 \\ &\geq k + 2 + 2s - k - 1 \\ &= 2s + 1 \end{aligned} \quad (53)$$

Since we have removed a single edge from \mathcal{G} , the graph \mathcal{G} has at least $2s + 2$ single edges. \square

Corollary 7.11 and Proposition 7.12 together allow us to do the moment argument as in Section 2.2.

We turn to the case $\beta > 1$ for the full Curie-Weiss model. Part 3 of Proposition 7.10 shows that there are strong correlations in this case, so one is tempted to believe that there is no semicircle law for $\beta > 1$.

In fact, it is easy to see that for $\beta > 1$ the expectations of $\frac{1}{N^{1+k}} \text{tr}(X_N^{2k})$ cannot

converge for $k \geq 2$ as $N \rightarrow \infty$. For example, for $k = 2$ we have

$$\begin{aligned} & \mathbb{E} \left(\frac{1}{N^3} \operatorname{tr}(X_N^4) \right) \\ &= \frac{1}{N^3} \sum_{\substack{i_1, i_2, i_3, i_4 \\ \text{all different}}} \mathbb{E} \left(X_N(i_1, i_2) X_N(i_2, i_3) X_N(i_3, i_4) X_N(i_4, i_1) \right) + \mathcal{O}(1) \\ &\approx \frac{N(N-1)(N-2)(N-3)}{N^3} m(\beta)^4 \rightarrow \infty \end{aligned} \quad (54)$$

so the moment method will not work here.

A closer analysis of the problem shows that the divergence of the moments of traces is due to a *single* eigenvalue of $\frac{1}{\sqrt{N}} X_N$ which goes to infinity. All the other eigenvalues behave ‘nicely’. Informally speaking, for $\beta > 1$ the matrices X_N fluctuate around the matrices $\pm m(\beta) \mathcal{E}_N$ (see (8)) with probability 1/2 each. As we saw in Section 2 these matrices have rank one. So, one may hope that they do not change the empirical eigenvalue distribution measure in the limit.

Analyzing the fluctuations around $\pm m(\beta) \mathcal{E}_N$ one can apply the moment method to $X_N \mp m(\beta) \mathcal{E}_N$. The variance of the matrix entries is $v(\beta) = 1 - m(\beta)^2$, so this has a chance to converge to the semicircle distribution, but scaled due to the variance $v(\beta) < 1$. In fact we have:

Theorem 7.13. *Let X_N be the full Curie-Weiss matrix ensemble with arbitrary inverse temperature $\beta \geq 0$. Then the empirical eigenvalue distribution measure σ_N of $\frac{1}{\sqrt{N}} X_N$ converges weakly in probability to the rescaled semicircle distribution $\sigma_{v(\beta)}$, given by:*

$$\sigma_{v(\beta)}(x) = \begin{cases} \frac{1}{2\pi v(\beta)} \sqrt{4v(\beta) - x^2} & , \text{ for } |x| \leq 2\sqrt{v(\beta)} \\ 0 & , \text{ otherwise.} \end{cases} \quad (55)$$

Here, $v(\beta) = 1 - m(\beta)^2$ with $m(\beta) = 0$ for $\beta \leq 1$ and $m = m(\beta)$ is the unique positive solution of $\tanh(\beta m) = m$ for $\beta > 1$ (cf. (44)).

A detailed proof will be contained in [29].

Already in (54) we saw that the norm of $\frac{1}{\sqrt{N}} X_N$ does not converge for the full Curie-Weiss ensemble if $\beta > 1$. This is made precise in the following theorem.

Theorem 7.14. *Suppose X_N is a full Curie-Weiss ensemble.*

1. *If $\beta < 1$ then*

$$\left\| \frac{1}{\sqrt{N}} X_N \right\| \rightarrow 2 \text{ as } N \rightarrow \infty$$

\mathbb{P} -almost surely.

2. If $\beta = 1$ then

$$\left\| \frac{1}{N^\gamma} X_N \right\| \rightarrow 0 \text{ as } N \rightarrow \infty$$

for every $\gamma > 1/2$ \mathbb{P} -almost surely.

3. If $\beta > 1$ then

$$\left\| \frac{1}{N} X_N \right\| \rightarrow m(\beta) \text{ as } N \rightarrow \infty$$

\mathbb{P} -almost surely.

Theorem 7.14.3 was proved in [25], 1 and 2 can be found in [30].

8. Ensembles with Exchangeable Entries

The results presented in the previous section for Curie-Weiss ensembles with subcritical temperatures ($\beta > 1$) suggest that models with correlations that do not decay sufficiently fast as N tends to infinity (e.g. in the sense of Corollary 7.11) may display a wealth of spectral phenomena depending on the specific features of the model. This is largely uncharted territory. One step into this world is to consider matrix ensembles with entries chosen from a sequence of exchangeable random variables.

A sequence $(\xi_i)_{i \in \mathbb{N}}$ of real valued random variables with underlying probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is called *exchangeable*, if for all integers $N \in \mathbb{N}$, all permutations π on $\{1, \dots, N\}$, and $F \in \mathcal{B}(\mathbb{R}^N)$ it is true that

$$\mathbb{P}((\xi_1, \dots, \xi_N) \in F) = \mathbb{P}((\xi_{\pi(1)}, \dots, \xi_{\pi(N)}) \in F).$$

Generalizing a result of de Finetti [16, 17] for random variables that only take on two values, Hewitt and Savage [24, Theorem 7.4] showed in a very general setting that such probability measures \mathbb{P} may be represented as averages of i.i.d. sequences with respect to some probability measure μ . In our context we impose the additional condition that all moments of the random variables ξ_i exist (cf. Definition 2.3). This leads us to the following general definition of ensembles of real symmetric matrices with exchangeable entries.

Definition 8.1. Let μ denote a probability measure on some measurable space (T, \mathcal{T}) and let $\Lambda : T \rightarrow \mathcal{M}_1^{(0)}(\mathbb{R})$ be a measurable map that assigns every element τ of T to a Borel probability measure Λ_τ on \mathbb{R} for which all moments exist (we call $\mathcal{M}_1^{(0)}(\mathbb{R})$ the set of all such probability measures on \mathbb{R}). Define

$$\mathbb{P}^{\mu, \Lambda} := \int_T P_\tau d\mu(\tau), \quad \text{with } P_\tau := \bigotimes_{i=1}^{\infty} \Lambda_\tau, \quad (56)$$

as the μ -average of i.i.d. sequences of real random variables with distributions Λ_τ . The corresponding *matrix ensemble with exchangeable entries* consists of matrices

X_N with entries $X_N(i, j)$, $1 \leq i \leq j \leq N$, given by the first $N(N+1)/2$ members of the sequence $(\xi_i)_i$ of exchangeable random variables that is distributed according to $\mathbb{P}^{\mu, \Lambda}$ of (56). The remaining entries $X_N(i, j)$, $1 \leq j < i \leq N$ are then fixed by symmetry $X_N(i, j) = X_N(j, i)$. Observe that due to the exchangeability of $(\xi_i)_i$ it is of no relevance in which order the upper triangular part of X_N is filled by $\xi_1, \dots, \xi_{N(N+1)/2}$. Moreover, one could have chosen any $N(N+1)/2$ distinct members of $(\xi_i)_i$ to fill the entries of X_N without changing the ensemble.

It is instructive to consider the special case of ensembles that allow only for matrix entries $X_N(i, j) \in \{1, -1\}$. We refer to it as the *spin case*. Observe that the probability measures with support contained in $\{1, -1\}$ are all represented by the family $\Lambda_\tau = \frac{1}{2} [(1 + \tau)\delta_1 + (1 - \tau)\delta_{-1}]$, $\tau \in T := [-1, 1]$. Hence all ensembles of the spin case are given by (56) with the just mentioned choices for T and Λ_τ . They are parameterized by the probability measures μ on $[-1, 1]$. Recall that Λ_τ already appeared in Definition 7.8 as the building block $P_t^{(1)}$ for Curie-Weiss ensembles. What is different from Section 7 is that there the averaging measure μ depends on the matrix size N and is of a special form.

Let us return to the general ensembles with exchangeable entries of Definition 8.1. The key for analyzing both the empirical eigenvalue distribution measure and the operator norm is that for every $\tau \in T$ the measure P_τ generates i.i.d. entries for X_N . For the latter ensembles P_τ the following observations that can already be found in [20] are useful: Subtracting the mean of the entries yields a Wigner ensemble (multiplied by the standard deviation of Λ_τ) for which Theorem 3.3 is applicable. Considering first the empirical eigenvalue distribution measure we note that the mean is some multiple of the matrix \mathcal{E}_N defined in (8). Since \mathcal{E}_N has rank 1 the subtraction of the mean will not have an influence on the limiting spectral measure. As $\mathbb{P}^{\mu, \Lambda}$ is the μ -average over all measures P_τ it is plausible that the limit of the empirical eigenvalue distribution measures is an average of scaled semicircles w.r.t. the measure μ , where the scaling factors are given by the standard deviation of Λ_τ . Accordingly, we define

$$\sigma_\mu := \int_T \sigma_{v(\tau)} d\mu(\tau), \quad (57)$$

where $v(\tau)$ denotes the variance of Λ_τ and σ_v is the semicircle distribution with support $[-2\sqrt{v}, 2\sqrt{v}]$ (cf. Definition (55)). We prove in [31]

Theorem 8.2. *Denote by $\mathbb{P}^{\mu, \Lambda}$, σ_μ the measures introduced in Definition 8.1 and in (57). Then the empirical eigenvalue distribution measures σ_N of X_N/\sqrt{N} converge weakly in expectation to σ_μ w.r.t. to the measure $\mathbb{P}^{\mu, \Lambda}$.*

Moreover, it is shown in [31] that σ_μ is a semi-circle if and only if the function $\tau \mapsto v(\tau)$ is constant μ -almost surely.

For the operator norm the situation is quite different. Since $\|\mathcal{E}_N\| = N$ the operator norm of X_N w.r.t. the measure P_τ it is determined to leading order by the mean of

X_N , if the mean does not vanish. Therefore the operator norm scales with N , except for the special case that the matrix entries are $\mathbb{P}^{\mu, \Lambda}$ -almost surely centered. We prove in addition in [30, 31] that the N -scaling of the norm is due to a single outlier of the spectrum by showing that the second largest eigenvalue (in modulus) possesses a \sqrt{N} -scaling that is consistent with the law for the limiting spectral measure.

In [30, 31] we also generalize the just mentioned results to band matrices. Here an additional difficulty arises, because the mean of X_N is no longer a multiple of \mathcal{E}_N and will have large rank. Nevertheless it is shown that all results obtained for full matrices can be saved, except for the result on the second largest eigenvalue (in modulus).

References

- [1] A. Altland, M. Zirnbauer: Nonstandard symmetry classes in mesoscopic normal/superconducting hybrid structures, *Physical Review B* **55** no. 2, 114 (1997).
- [2] G. Anderson, A. Guionnet, O. Zeitouni: *An introduction to random matrices*, Cambridge University Press 2010.
- [3] G. Akemann, J. Baik, P. Di Francesco (Eds.): *The Oxford handbook of random matrix theory*, Oxford University Press, Oxford, 2011.
- [4] L. Arnold: On the Asymptotic Distribution of the Eigenvalues of Random Matrices, *J. Math. Anal. Appl.* **20**, 262–268 (1967).
- [5] Z. Bai, Y. Yin: Necessary and sufficient conditions for almost sure convergence of the largest eigenvalue of a Wigner matrix, *Ann. Probab.* **16**, 1729–1741 (1988).
- [6] L. Bogachev, S. Molchanov, L. Pastur: On the level density of random band matrices, *Math. Notes* **50** no. 5–6, 1232–1242 (1991).
- [7] B. Bollabás: *Modern Graph Theory*, Springer 1998.
- [8] L. Breiman: *Probability*, Addison-Wesley 1968.
- [9] W. Bryc, A. Dembo, T. Jiang: Spectral measure of large random Hankel, Markov and Toeplitz matrices, *Ann. Probab.* **34**(1), 1–38 (2006).
- [10] R. Catalano: On weighted random band-matrices with dependencies, PhD thesis, FernUniversität Hagen, 2016.
- [11] P. Deift: *Orthogonal polynomials and random matrices: a Riemann-Hilbert approach* Courant Lecture Notes in Mathematics **3**, Courant Institute of Mathematical Sciences, New York 1999.
- [12] P. Deift, D. Gioev: *Random matrix theory: invariant ensembles and universality*, Courant Lecture Notes in Mathematics 18, Courant Institute of Mathematical Sciences, New York 2009.
- [13] R. Ellis: *Entropy, large deviations, and statistical mechanics*, Springer 2006.
- [14] L. Erdős: Random matrices, log-gases and Hölder regularity, in: *Proceedings of ICM 2014*, Seoul, Vol. III, 213–236 (2015).
- [15] L. Erdős, H.-T. Yau, J. Yin: Bulk universality for generalized Wigner matrices. *Probab. Theory Related Fields* **154** no. 1-2, 341–407 (2012).

- [16] B. de Finetti: Funzione caratteristica di un fenomeno aleatorio, *Atti della R. Accademia Nazionale dei Lincei, Ser. 6, Memorie, Classe di Scienze Fisiche, Matematiche e Naturali* **4**, 251–299 (1931).
- [17] B. de Finetti: La prevision: ses lois logiques, ses sources subjectives, *Annales de l'Institut Henri Poincare* **7**, 1–68 (1937).
- [18] O. Friesen, M. Löwe: The Semicircle Law for Matrices with Independent Diagonals, *J. Theoret. Probab.* **26**, 1084–1096 (2013).
- [19] O. Friesen, M. Löwe: A phase transition for the limiting spectral density of random matrices, *Electron. J. Probab.* **18**, 1–17 (2013).
- [20] Z. Füredi, J. Komlós: The eigenvalues of random symmetric matrices, *Combinatorica* **1** no. 3, 233–241 (1981).
- [21] F. Götze, A. Naumov, A. Tikhomirov: Semicircle law for a class of random matrices with dependent entries, arXiv:1211.0389.
- [22] F. Götze, A. Naumov, D. Timushev, A. Tikhomirov: On the local semicircular law for Wigner ensembles arXiv:1602.03073.
- [23] U. Grenander: *Probabilities on algebraic structures*, Wiley 1968.
- [24] E. Hewitt, L. J. Savage: Symmetric measures on Cartesian products, *Trans. Amer. Math. Soc.* **80**, 470–501 (1955).
- [25] W. Hochstättler, W. Kirsch, S. Warzel: Semicircle Law for a Matrix Ensemble with Dependent Entries, *J. Theoret. Probab.* **29** no. 3, 1047–1068 (2016).
- [26] K. Hofmann-Credner, M. Stolz: Wigner theorems for random matrices with dependent entries: ensembles associated to symmetric spaces and sample covariance matrices; *Electron. Commun. Probab.* **13**, 401–414 (2008).
- [27] V. Marchenko, L. Pastur: Distribution of eigenvalues in certain sets of random matrices. *Math. USSR-Sbornik* **1**, 457–483 (1967).
- [28] W. Kirsch: *Moments in Probability*, book in preparation, to appear at DeGruyter.
- [29] W. Kirsch, T. Kriecherbauer: Semicircle law for generalized Curie-Weiss matrix ensembles at subcritical temperature; in preparation
- [30] W. Kirsch, T. Kriecherbauer: Largest and second largest singular values of de Finetti random matrices; in preparation
- [31] W. Kirsch, T. Kriecherbauer: in preparation
- [32] A. Klenke: *Probability*, Springer 2014.
- [33] T. Koshy: *Catalan Numbers with Applications*, Oxford University Press 2009.
- [34] M. Löwe, K. Schubert: On the limiting spectral density of random matrices filled with stochastic processes, to appear in: *Random Operators and Stochastic Equations*, arXiv:1512.02498.
- [35] L. Pastur: On the spectrum of random matrices, *Theoret. and Math. Phys.* **10** no. 1, 67–74 (1972).
- [36] L. Pastur: Spectra of random self adjoint operators, *Russian Math. Surveys* **28** no. 1, 1–67 (1973).
- [37] L. Pastur, M. Shcherbina: *Eigenvalue distribution of large random matrices*, AMS 2011.

- [38] J. Schenker, H. Schulz-Baldes: Semicircle law and freeness for random matrices with symmetries or correlations, *Mathematical Research Letters* **12**, 531–542 (2005).
- [39] S. Sodin: The spectral edge of some random band matrices, *Ann. of Math. (2)* **172** no. 3, 2223–2251 (2010).
- [40] A. Soshnikov: Universality at the edge of the spectrum in Wigner random matrices, *Commun. Math. Phys.* **207**, 697–733 (1999).
- [41] R. Stanley: *Catalan Numbers*, Cambridge University Press 2015.
- [42] T. Tao: *Topics in random matrix theory*, AMS 2012.
- [43] T. Tao, V. Vu: Random matrices: Universality of local eigenvalue statistics, *Acta Math* **206** no. 1, 127–204 (2011).
- [44] T. Tao, V. Vu: Random Matrices: The Universality Phenomenon for Wigner Ensembles, in: *Modern aspects of random matrix theory*, Proc. Sympos. Appl. Math. **72**, 121–172, AMS 2014.
- [45] E. Wigner: Characteristic vectors of bordered matrices with infinite dimension, *Ann. Math.* **62**, 548–564 (1955).
- [46] E. Wigner: On the distribution of the roots of certain symmetric matrices, *Ann. Math.* **67**, 325–328 (1958).

Fakultät für Mathematik und Informatik, FernUniversität Hagen, Germany
E-mail: werner.kirsch@fernuni-hagen.de

Mathematisches Institut, Universität Bayreuth, Germany
E-mail: thomas.kriecherbauer@uni-bayreuth.de