Random matrices with exchangeable entries

Werner Kirsch
Fakultät für Mathematik und Informatik
FernUniversität in Hagen, Germany

Thomas Kriecherbauer
Mathematisches Institut
Universität Bayreuth, Germany

Abstract

We consider ensembles of real symmetric band matrices with entries drawn from an infinite sequence of exchangeable random variables, as far as the symmetry of the matrices permits. In general the entries of the upper triangular parts of these matrices are correlated and no smallness or sparseness of these correlations is assumed. It is shown that the eigenvalue distribution measures still converge to a semicircle but with random scaling. We also investigate the asymptotic behavior of the corresponding $\ell_2$-operator norms. The key to our analysis is a generalisation of a classic result by de Finetti that allows to represent the underlying probability spaces as averages of Wigner band ensembles with entries that are not necessarily centred. Some of our results appear to be new even for such Wigner band matrices.

1 Introduction

In this paper we consider (full) real symmetric random matrices of the form

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

for certain random schemes $X_N(i,j)$ (with $X_N(i,j) = X_N(j,i) \in \mathbb{R}$), as well as real symmetric band random matrices where $X_N(i,j)$ is random in a strip of size $w_N \to \infty$ centred around the diagonal $i = j$ and $X_N(i,j) \equiv 0$ otherwise. We use
the word full to distinguish from the case of band matrices. The parantheses above indicate that we will omit this specification if there is no danger of confusion.

Let us denote the underlying probability space by \((\Omega, \mathcal{F}, \mathbb{P})\) and the expectation with respect to \(\mathbb{P}\) by \(\mathbb{E}\).

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

\[
\lambda_1(M) \leq \lambda_2(M) \leq \cdots \leq \lambda_N(M)
\]

where degenerate eigenvalues are repeated according to their multiplicity.

The eigenvalue distribution measure \(\nu_M\) of \(M\) is defined by

\[
\nu_M(A) = \frac{1}{N^2} \sum_{j=1}^{N} \delta_{\lambda_j(M)}(A)
\]

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 \(\delta_a\) is the Dirac measure in \(a\), i.e.

\[
\delta_a(A) = \begin{cases} 
1 & \text{if } a \in A \\
0 & \text{otherwise}
\end{cases}
\]

In this paper we study the limiting behavior of both the eigenvalue distribution measures and the \(\ell^2\)-operator norms as the matrix dimension \(N\) becomes large. Within the theory of random matrices the first results on these quantities were obtained for full Wigner ensembles (see Theorem 3 below) using the method of moments.

The distinctive feature of Wigner ensembles is that for each fixed matrix size \(N\) the entries \(X_N(i, j), 1 \leq i \leq j \leq N\), of the upper triangular part are independent, identically distributed random variables. For this class of matrix ensembles it has been shown in great generality that the eigenvalue distribution measures converge to the famous semicircle law. Having obtained such a universal limiting law it is natural to test its range of validity. For example, one might ask whether the assumption of independence of the matrix entries in the upper triangular part can be relaxed. Indeed, a number of matrix ensembles with correlated entries have been introduced in the literature and their limiting spectral distributions have been analysed. We refer the reader to the survey \[13\] for a detailed description of these results (see also the paper \[2\] for recent developments). Most of these ensembles are defined with some kind of smallness of the correlations built in: They are sparse or they decay. The decay can be with respect to the distance of the corresponding matrix entries or with respect to the matrix dimension \(N\).
The main focus of our paper lies on matrix ensembles with entries that are drawn from an exchangeable sequence of random variables (see Definition 13). For such models the correlations may neither be sparse nor decaying. In addition, we do not only consider full matrices but also band random matrices. One of our main results is Theorem 23 where we show for a large class of such ensembles that the eigenvalue distribution measures still converge to a semicircle, but its radius may now be random.

A key element in our proof is the fact that an exchangeable sequence of real-valued random variables can be represented as an average of i.i.d. sequences. This classic result is due to de Finetti [10, 11] in the special case of spin random variables that only assume values ±1 and was later generalized by Hewitt-Savage [15, Theorem 7.4] to a setting that includes in particular real-valued variables. Therefore we can relate matrix ensembles with exchangeable entries to ensembles with i.i.d entries and this brings us back to the realm of Wigner ensembles. Accordingly, we begin the more precise discussion of our results by a definition of Wigner ensembles that is suitable for the analysis of ensembles with exchangeable entries.

**Definition 1**  By a (full) Wigner ensemble we understand a probability measure on sequences \((X_N)_N\) of real symmetric \(N \times N\) matrices \(X_N\) such that for each fixed \(N\) the random variables \(X_N(i, j), i \leq j,\) are independent. Moreover, we require that the \(X_N(i, j)\) for all \(N, i, j\) have a common distribution \(\rho\) with finite moments of all orders. We call \(m = \int x \, d\rho\) the mean and \(v = \int x^2 \, d\rho(x) - m^2\) the variance of the Wigner ensemble. In case the mean vanishes, \(m = 0\), we say that the Wigner ensemble is centred.

A few remarks are in order. First, note that no assumptions are made on how the entries of \(X_N\) and \(X_M\) are correlated for \(N \neq M\). The reason is that these correlations play no role for results on spectral limits relevant for this paper. Secondly, the assumption of identically distributed entries is often relaxed for Wigner ensembles by conditions that only require agreement of some moments. As explained above ensembles with exchangeable entries are related to the i.i.d. case and therefore we do not strive for more generality in this respect. Thirdly, the condition that all moments of the law \(\rho\) exist could be downgraded as well. For Wigner ensembles it is well known how to adapt the arguments in the situation that only a few moments exist by a truncation procedure. In order to avoid the associated substantial technicalities we restrict ourselves to the case that all moments exist. Fourthly, and this is the most important point, we do not require that the entries are centred random variables. One motivation for this is our recent work on random matrices with Curie-Weiss distributed entries [19] where non-zero means are generated by magnetisation at low temperatures. The results we obtain for means \(m \neq 0\) appear to be new, some of them even in the case of Wigner band
ensembles (see Definition 7). In comparison, the influence of the variance $v$ on the spectrum is simple, because it translates to a linear scaling by the factor $\sqrt{v}$.

Based on the work of Wigner and others (see [25, 26], [14], [4]) it is well known that for centred Wigner ensembles $X_N$ the eigenvalue distribution measures $\mu_N$ of the matrix $\frac{1}{\sqrt{N}}X_N$ converge in the case $v > 0$ to the (scaled) semicircle distribution, i.e. to the measure $\sigma_v$ with density (with respect to the Lebesgue measure)

$$s_v(x) = \frac{1}{2\pi v} \sqrt{4v - x^2}.$$  \hfill (2)

Above we use the notation $g_+(x) := \max(g(x), 0)$. The classical Wigner case corresponds to $v = 1$, the above slightly more general case follows through scaling.

For our purposes it will be convenient to include the trivial case of variance $v = 0$. In the centred case the entries of the Wigner matrices are then equal to zero almost surely. Thus the corresponding eigenvalue distribution measures of the matrices $\frac{1}{\sqrt{N}}X_N$ are all given by the Dirac measure $\delta_0$. We therefore extend definition (2) for $\sigma_v$ by

$$\sigma_0 := \delta_0.$$  \hfill (3)

Despite its degeneracy we call $\delta_0$ a semicircle distribution throughout the paper.

There are various forms of convergence for sequences of random measures:

**Definition 2** Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $\mu_N^\omega$ and $\mu^\omega$, $\omega \in \Omega$, be random probability measures on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

1) We say that $\mu_N^\omega$ converges to $\mu^\omega$ weakly in expectation, if for every $f \in C_b(\mathbb{R})$, the set of bounded continuous functions on $\mathbb{R}$,

$$\mathbb{E}\left(\int f(x) \ d\mu_N^\omega(x)\right) \to \mathbb{E}\left(\int f(x) \ d\mu^\omega(x)\right)$$  \hfill (4)

as $N \to \infty$.

2) We say that $\mu_N^\omega$ converges to $\mu^\omega$ weakly in probability, if for every $f \in C_b(\mathbb{R})$ and any $\epsilon > 0$

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

as $N \to \infty$. 

4
3) We say that \( \mu_N^{\omega} \) converges to \( \mu^{\omega} \) 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 \( \omega \in \Omega_0 \). Here \( \mu_N^{\omega} \Rightarrow \mu^{\omega} \) means weak convergence, i.e. for every \( f \in C_b(\mathbb{R}) \)

\[
\int f(x) \, d\mu_N^{\omega}(x) \to \int f(x) \, d\mu^{\omega}(x) \quad \text{as} \quad N \to \infty.
\]

We can now formulate: For centred Wigner matrices \( X_N \) with variance \( v \) the eigenvalue distribution measures \( \mu_N^{\omega} \) of \( \frac{1}{\sqrt{N}}X_N \) converge weakly \( \mathbb{P} \)-almost surely to \( \sigma_v \) [4].

Besides the limiting spectral distribution we also want to understand the behavior of the \( \ell_2 \)-operator norms \( ||X_N||_{op} \) as the matrix size \( N \) becomes large. Recall that for real symmetric matrices the operator norm is given by the largest eigenvalue in modulus,

\[
||X_N||_{op} = \max(|\lambda_1(X_N)|, |\lambda_N(X_N)|).
\]

Guided by the semicircle law one might expect that \( ||X_N||_{op} \approx 2\sqrt{vN} \). Indeed, the semicircle law can be used to show that the limes inferior of \( ||X_N||_{op}/\sqrt{N} \) is bounded below by \( 2\sqrt{v} \) (see proof of Parts II and III of Theorem 28). However, an upper bound for the operator norm cannot possibly be extracted from the semicircle law since, for example, a single outlier of the spectrum has no effect on the limiting spectral distribution but may determine the operator norm. Therefore additional arguments are needed. These were provided in [6], see also [13], where it was shown for centred Wigner ensembles that \( ||X_N||_{op}/\sqrt{N} \) converges to \( 2\sqrt{v} \) almost surely.

The classical results that we have discussed so far can be summarized as follows.

**Theorem 3** Let \( (X_N)_N \) be a (full) centred Wigner ensemble with variance \( v \) in the sense of Definition 7. Then:

a) The eigenvalue distribution measures \( \mu_N^{\omega} \) of \( \frac{1}{\sqrt{N}}X_N \) converge weakly \( \mathbb{P} \)-almost surely to \( \sigma_v \).

b) The scaled operator norms \( ||X_N||_{op}/\sqrt{N} \) converge \( \mathbb{P} \)-almost surely to \( 2\sqrt{v} \).

Our main goal is to generalise Theorem 3 to ensembles of full or banded matrices with exchangeable entries. In the remainder of the Introduction we outline the plan of the paper highlighting our main results along the way.

As a first step we formulate in Section 2 the semicircle law for Wigner band matrices (see Definition 7). The corresponding result is stated in Theorem 10. Note that the scaling \( X_N/\sqrt{N} \) of Theorem 3 needs to be replaced by \( X_N/\sqrt{w_N} \).
where $w_N$ is a measure for the bandwidth. The case of centred entries is essentially known \[7, 20\] (see also \[8, 12\]) and this is the starting point for our proof. In order to analyse arbitrary means $m$ we write

$$\frac{1}{\sqrt{w_N}}X_N = \frac{1}{\sqrt{w_N}}(X_N - \mathbb{E}(X_N)) + \frac{1}{\sqrt{w_N}}\mathbb{E}(X_N). \quad (5)$$

The first summand has centred entries and its eigenvalue distribution measure therefore obeys the semicircle law. The main work in the proof consists of showing that the deterministic matrix $\mathbb{E}(X_N)/\sqrt{w_N}$ can be decomposed into two parts such that one of them has small enough norm and the other one has small enough rank to allow the semicircle law to persist.

As can be seen from the statement of Theorem 10 the result for Wigner band matrices is a little more involved than for full matrices. For example, even in the case of centred entries one needs (at least for our proof) an additional (mild) condition on the bandwidths to improve from convergence in probability to almost sure convergence. This subtlety and its proof seem to be somewhat buried in the literature. We refer the reader to \[12\] for a proof in a more general setting that also includes the case of correlated entries. Band matrices with linearly growing bandwidths appear as a special case of ‘general Wigner-type matrices’ in \[1\], where a ‘local law’ for eigenvalue statistics is proved. In particular, their results imply almost sure convergence of the eigenvalue distribution measures for these band matrices.

For the convenience of the reader we sketch a proof of almost sure convergence for centred Wigner band ensembles for our simpler situation in Subsection A.1 of the Appendix.

Section 3 is devoted to generalising Theorem 10 to band ensembles with exchangeable entries. We call them de Finetti band ensembles in reference to the remarkable work of de Finetti \[10, 11\] on which our analysis is based (see Definition 20). Observe that we use a definition of band matrices that includes full matrices by choosing the bandwidth sufficiently large. Theorem 23 states our main result for these ensembles. As it was already mentioned above, the only difference between the results in the Wigner and in the de Finetti case is that the limiting law is given by the semicircle $\sigma_V$ rather than $\sigma_v$. I.e. the variance $v$ of the Wigner ensemble needs to be replaced by a real-valued random variable $V$ that we call the limiting empirical variance of the ensemble (see Definition 19).

We show in Subsection 3.3 that a non-random limit law for the eigenvalue distribution measures can be achieved but, except for trivial cases, one needs to settle for the weaker notion of convergence in expectation (cf. Definition 2). In addition we derive some properties of the deterministic limit law including a characterisation of all cases in which it is a semicircle.
So far we have generalised part a) of Theorem 3. In the final section of this paper we study the corresponding operator norms. As explained above the statement of Theorem 3 the main task is to obtain upper bounds once the limit law for the eigenvalue distribution measures is established. Observe that for centred Wigner band matrices it was shown in [7] that $\|X_N\|_{op}/\sqrt{w_N}$ is unbounded if the bandwidth grows slowly enough with matrix dimension $N$. With Theorem 28 we provide a result in the opposite direction. We prove for centred Wigner band ensembles with bandwidths $w_N$ growing at least of order $N^{\alpha}$ for some arbitrarily small $\alpha > 0$ that $\|X_N\|_{op}/\sqrt{w_N}$ converges to $2\sqrt{\nu}$ almost surely in all cases where the eigenvalue distribution measures converge to $\sigma_v$.

The proof of this result uses the strategy that was introduced in [6]. We follow the presentation of the monograph [24]. Both references deal with ensembles of full matrices. Although the generalisation to band matrices does not pose any difficulties we provide a proof in the second subsection of the Appendix. The reason is that we have improved on some of the inequalities (see in particular Lemma 34) in order to obtain weaker conditions on the required rate of growth for the bandwidths (see Remark 29 and Lemma 35), an issue that is not present in the case of full random matrices.

For ‘general Wigner-type matrices’ upper bounds for the operator norms are given in [9]. As mentioned above these matrices include band matrix ensembles with linearly growing bandwidths.

Finally we consider operator norms for de Finetti band ensembles and for Wigner band ensembles that are not centred in Subsection 4.2. This is a much less subtle question than in the centred case since the deterministic part $E(X_N)$ in the decomposition (5) has operator norm of order $w_N$ (see Lemma 9) that dominates the centred part that is only of order $\sqrt{w_N}$. Therefore the mean of the entries of the Wigner ensemble and the (random) empirical mean of the de Finetti ensemble (see Definition 19) respectively determine the asymptotic behavior of the operator norms. In the special case of full matrices one may use that the matrix of means $E(X_N)$ has rank one to show that the discrepancy between the $N$-scaling of the operator norms and the $\sqrt{N}$-scaling of the eigenvalue distribution measures is caused by a single outlier of the spectrum, see Proposition 6 and part 2 of Remark 32.

Acknowledgement Most of this work was done during the first author’s visit to the Institute for Mathematics of the University of Bayreuth, and the second author’s stay at the Lehrgebiet Stochastics at the FernUniversität in Hagen. We thank those institutions for their support and their great hospitality. We would also like to thank Michael Fleermann for many valuable discussions.
2 Wigner ensembles

This section is mainly concerned with the fate of the semicircle law as stated in Theorem 3b) if we consider Wigner band matrices with entries that are not necessarily centred. Our main result in this direction is Theorem 10 in the third subsection. A precise definition of Wigner band ensembles is provided in Subsection 2.2. There we will distinguish two different types of band matrices, strict and periodic, that differ in the way distance is measured on the index set \{1, \ldots, N\}. We begin our discussion with the case of full Wigner matrices and show what happens to both statements a) and b) of Theorem 3 if one removes the condition of centred entries.

2.1 Warm up: Full matrices

It was already observed by Füredi and Komlós in [13] that the semicircle law still holds for full Wigner ensembles with arbitrary means \(m\).

**Proposition 4** Let \((X_N)_N\) be a Wigner ensemble with arbitrary mean \(m\) and variance \(v\) (see Definition 7). Then the eigenvalue distribution measures \(\mu_N\) of \(\frac{1}{\sqrt{N}} X_N\) converge weakly \(\mathbb{P}\)-almost surely to \(\sigma_v\).

**Proof.** We give only a brief sketch here since we provide a detailed argument in the more general situation of band matrices in the proof of Theorem 10 below.

The crucial observation is the following. Denote by \(E_N\) the \(N \times N\)-matrix with \(E_N(i, j) = 1\) for all \(1 \leq i, j \leq N\). Use

\[
X_N = (X_N - m \cdot E_N) + m \cdot E_N
\]

and observe that the matrix \(E_N\) is a matrix of rank one, so it can change the number \(|\{j| \lambda_j \in A\}|\) of eigenvalues of \(\frac{1}{\sqrt{N}} (X_N - mE_N)\) inside any interval \(A\) by at most 2, which is negligible for the limiting empirical eigenvalue measure (see also Proposition 11).

On the other hand the matrix \(X_N - mE_N\) is a centred Wigner matrix with variance \(v\), hence the corresponding empirical eigenvalue measures converge weakly almost surely to the semicircle distribution \(\sigma_v\) by Theorem 3(a).

We have just argued that the limiting empirical eigenvalue measure is insensitive to perturbations of rank 1. The operator norm, however, may feel such a perturbation since it could create a single outlier of the spectrum. This happens e.g. in the situation of Proposition 4 if the mean \(m\) of the entries does not vanish. In fact, the operator norm of the matrix \(E_N\) equals \(N\) so that \(||mE_N||_{op} = |m| N\). Moreover, it follows from Theorem 3b) that \(||X_N - m \cdot E_N||_{op}\) is of the order \(2\sqrt{vN}\) for large values of \(N\). It is therefore asymptotically negligible when
compared to $||m\mathcal{E}_N||_{\text{op}}$. As

$$||m\mathcal{E}_N||_{\text{op}} - ||X_N - m \cdot \mathcal{E}_N||_{\text{op}} \leq ||X_N||_{\text{op}} \leq ||m\mathcal{E}_N||_{\text{op}} + ||X_N - m \cdot \mathcal{E}_N||_{\text{op}}$$

we have proved

**Proposition 5** For Wigner ensembles $(X_N)_N$ with arbitrary mean $m$ and variance $v$ (see Definition 1) the operator norm $||X_N||_{\text{op}}$ satisfies

$$\mathbb{P}\left(\lim_{N \to \infty} \frac{||X_N||_{\text{op}}}{\sqrt{N}} = |m|\right) = 1.$$ 

We repeat: In the case that the mean of the entries $m$ does not vanish there is a discrepancy between the $N$ scaling of the operator norm and the $\sqrt{N}$ scaling of the semicircle law.

We now formulate the fact that this discrepancy is due to only one outlier. To this end we introduce the singular values of $X_N$. By spectral calculus the singular values $s_j(X_N)$ of $X_N$ are given by the absolute values of the eigenvalues. We order them according to size taking their multiplicities into account

$$0 \leq s_1(X_N) \leq s_2(X_N) \leq \cdots \leq s_N(X_N).$$

The largest singular value $s_N(X_N) = \max\{||\lambda_1(M)||, ||\lambda_N(M)||\}$ is of particular interest as it agrees with the $\ell_2$-operator norm $||X_N||_{\text{op}}$. We can be sure that $X_N$ has only one outlier of order $N$ if we can prove that the second largest singular value $s_{N-1}(X_N)$ is of order $\sqrt{N}$. Let us introduce the notation

$$||X_N||_{\text{op}}^\prime := s_{N-1}(X_N).$$

**Proposition 6** For Wigner ensembles $(X_N)_N$ with arbitrary mean $m$ and variance $v$ the second largest singular value $||X_N||_{\text{op}}^\prime$ satisfies

$$\mathbb{P}\left(\lim_{N \to \infty} \frac{||X_N||_{\text{op}}^\prime}{\sqrt{N}} = 2\sqrt{v}\right) = 1.$$ 

**Proof.** Let us first consider the case where the mean $m$ of the Wigner ensemble is non-negative. As the matrix $\mathcal{E}_N$ has rank 1 and is positive definite the eigenvalues of $X_N$ interlace with the eigenvalues of $X_N - m\mathcal{E}_N$ in the following way.

$$\lambda_j(X_N - m\mathcal{E}_N) \leq \lambda_j(X_N) \leq \lambda_{j+1}(X_N - m\mathcal{E}_N)$$

(7) for all $1 \leq j \leq N - 1$ and the first inequality also holds true for $j = N$. The first step in obtaining upper and lower bounds for the second largest singular value of $X_N$ is the observation that

$$\max\{||\lambda_2(X_N)||, ||\lambda_{N-1}(X_N)||\} \leq ||X_N||_{\text{op}}^\prime \leq \max\{||\lambda_1(X_N)||, ||\lambda_{N-1}(X_N)||\}$$
Using the interlacing inequalities (7) we conclude further
\[
\max\{|\lambda_1(X_N)|, |\lambda_{N-1}(X_N)|\} \leq \max\{|\lambda_1(X_N - mE_N)|, |\lambda_{N-1}(X_N - mE_N)|\},
\]
\[
\max\{|\lambda_2(X_N)|, |\lambda_{N-1}(X_N)|\} \geq \max\{|\lambda_3(X_N - mE_N)|, |\lambda_{N-1}(X_N - mE_N)|\}.
\]
A moment’s thought then yields
\[
s_{N-3}(X_N - mE_N) \leq ||X_N||_{op}^\prime \leq ||X_N - mE_N||_{op}, \tag{8}
\]
A similar reasoning shows that the estimates (8) are also valid in the case of negative means \(m\). Clearly \((X_N - mE_N)_N\) is a centred Wigner ensemble with variance \(v\) so that Theorem 3(b) implies that the upper bound, divided by \(\sqrt{N}\) almost surely. As we show in Corollary 30 below (see also the discussion above the statement of Corollary 30) it is also true that the lower bound, divided by \(\sqrt{N}\), converges to \(2\sqrt{v}\) almost surely. This completes the proof up to the verification of Corollary 30.

2.2 Strict and periodic band matrices

Let us first define the notion of strict and periodic Wigner band matrices.

**Definition 7**  
a) Denote by \((b_N)_N\) a sequence of integers that is bounded by \(0 \leq b_N \leq N - 1\). Then the deterministic prototypes of \(N \times N\) strict and periodic band matrices, \(B_N\) and \(P_N\), are defined by
\[
B_N(i, j) := \begin{cases} 1, & \text{if } |i - j| \leq b_N, \\ 0, & \text{if } |i - j| > b_N, \end{cases}
\]
\[
P_N(i, j) := \begin{cases} 1, & \text{if } |i - j|_N \leq b_N, \\ 0, & \text{if } |i - j|_N > b_N, \end{cases}
\]
where \(|i - j|_N\) denotes the distance between \(i\) and \(j\) on the circle \(\mathbb{Z}/N\mathbb{Z}\), i.e. \(|i|_N := \min(|i|, N - |i|)\) for \(|i| \leq N\).

b) An ensemble of a family of \(N \times N\) real symmetric matrices \((W_N)_N / (W_N^{per})_N\) is called a strict Wigner band ensemble/ a periodic Wigner band ensemble with mean \(m\) and variance \(v\), if it can be generated from a Wigner ensemble \((X_N)_N\) with mean \(m\) and variance \(v\) (see Definition 7) via
\[
W_N(i, j) = \begin{cases} X_N(i, j), & \text{if } B_N(i, j) = 1 \\ 0, & \text{else} \end{cases},
\]
\[
W_N^{per}(i, j) = \begin{cases} X_N(i, j), & \text{if } P_N(i, j) = 1 \\ 0, & \text{else} \end{cases}.
\]
We call these ensembles centred if the mean $m$ vanishes.

c) We call $b_N$ the half-width and $w_N := \min(N, 2b_N + 1)$ the (maximal) band-
width of the band matrices defined above.

Remark 8 Observe that for periodic band matrices every row (and every column) of $P_N$ has the same number $w_N$ of non-zero entries. Therefore the case of band-
width $w_N = N$ makes $W_N^\text{per}$ a full Wigner matrix.

For strict band matrices, however, these ensembles may differ from each other
even if $w_N = N$, depending on the value of the half-width $b_N \geq (N - 1)/2$.
In this situation $w_N = N$ is the maximal number of non-zero entries that a row
(column) of $B_N$ may have. Therefore $w_N$ was named the maximal bandwidth.
The band matrix $W_N$ is a full Wigner matrix only in the case $b_N = N - 1$.

We end this subsection by noting a few spectral properties of the band matrices
$B_N$ and $P_N$ for later reference.

Lemma 9 Let $B_N$ and $P_N$ be defined as in Definition 7a) with bandwidth $w_N$ and
half-width $0 \leq b_N \leq N - 1$.
   a) For $0 \leq j \leq N - 1$ define $\omega_j := j\pi/N$ and $u^{(j)} \in \mathbb{C}^N$ by
   \[ u^{(j)}_k := e^{2i\omega_j k}, \quad 1 \leq k \leq N. \]
The vectors $(u^{(j)})$ form an orthogonal basis of eigenvectors of $P_N$ and the corresponding eigenvalues $\mu_j$ are given by $\mu_0 = w_N$ and
$\mu_j = \frac{\sin(\omega_j w_N)}{\sin \omega_j}$ for $j \geq 1$.
   b) The $\ell_2$-operator norms satisfy
   \[ ||P_N||_{op} = w_N \quad \text{and} \quad w_N(1 - \delta_N) \leq ||B_N||_{op} \leq w_N \quad \text{with} \]
   \[ \delta_N = \frac{w_N}{w_N^2} \quad \text{if} \quad 2b_N + 1 \leq N \quad \text{and} \quad \delta_N = \left(1 - \frac{b_N}{N}\right)^2 \quad \text{else}. \]

Proof. Statement a) can be verified by computation. In order to see claim b) recall
first that the modulus of any eigenvalue of a given matrix $(A(i,j))_{i,j}$ is bounded
above by $\max_{j} \sum_{j} |A(i,j)|$ (e.g. consider the eigenvalue equation for a component
for which the eigenvector has maximal modulus). Thus both the operator norms
of $B_N$ and $P_N$ are bounded above by $w_N$. Secondly, for real symmetric matrices
$A$ the operator norm is bounded below by $(v, Av)/(v, v)$ for any non-zero vector
$v$. Choose $v = (1, \ldots, 1)$. Then $(v, Av)$ is just the number of non-zero entries
for $A \in \{P_N, B_N\}$. In the case of $P_N$ this number is $w_N N$. For $B_N$ this number is
$w_N N - k_N(k_N + 1)$ with $k_N = \min(b_N, N - b_N - 1)$. Using $k_N(k_N + 1) \leq
(k_N + \frac{1}{2})^2 = w_N^2/4$ for $2b_N + 1 \leq N$ and $k_N(k_N + 1) \leq (k_N + 1)^2 = (N - b_N)^2$
for $2b_N + 1 > N$ completes the proof. \hfill \blacksquare
2.3 Semicircle for band matrices

For the ensembles that we have defined in the previous subsection we now formulate our main result on the limiting spectral distribution. Our proof starts from the special case of centred ensembles where the result is known. The extension to arbitrary means uses Proposition 11 which provides estimates on the effects on the spectral measure of adding matrices of small operator norm or of small rank.

**Theorem 10** Let $0 \leq b_N \leq N - 1$ be a given sequence with $b_N \to \infty$ for $N \to \infty$. Recall the notion of Wigner band matrices with half-width $b_N$ and bandwidth $w_N = \min(N, 2b_N + 1)$ from Definition 7.

a) We distinguish the two cases of periodic and strict band matrices.

1. Assume that the entries of the periodic Wigner band matrices $W_{N}^{\text{per}}$ have variance $v$ and arbitrary mean $m$. Then the empirical eigenvalue measures $\mu_{N}^{\omega}$ of $\frac{1}{\sqrt{w_N}}W_{N}^{\text{per}}$ converge weakly in probability to the semicircle law $\sigma_v$.

2. Statement 1 also holds for the empirical eigenvalue measures $\mu_{N}^{\omega}$ of the strict band matrices $W_N/\sqrt{w_N}$ if we require in addition that the scaled half-widths $\frac{b_N}{N}$ converge either to 0 or to 1 for $N \to \infty$.

b) Let us add to the general assumption $b_N \to \infty$ above the summability condition $\sum_N (Nb_N)^{-1} < \infty$. Then both statements of part a) remain true if we strengthen the assertion of weak convergence in probability to weak convergence $\mathbb{P}$-almost surely.

Before we set out to prove the theorem for arbitrary values of the mean $m$, let us briefly describe what is known in the case of centred entries. In this case statement a) of this theorem is due to [7] and [20]. For periodic band matrices with centred entries statement b) has been observed in [12] as a special case of matrix ensembles with almost uncorrelated entries, see also [22, Problem 2.4.13]. For the convenience of the reader we sketch a proof of part b) for the centred case $m = 0$ in Section A.1 of the Appendix. Finally, we mention that for strict Wigner band matrices with centred entries [7, 8, 22] also treat the case where $\lim_{N \to \infty} \frac{b_N}{N}$ exists and where the limit lies in the open interval $(0, 1)$. In this situation the empirical eigenvalue measures of $\frac{1}{\sqrt{w_N}}W_N$ still converge but not to a semicircle law.

In order to derive Theorem 10 from its specialized version with centred entries we proceed as in the proof of Proposition 4 for full matrices. We split off a matrix $M_N$ containing the expectations of the matrix elements. In the case of full random matrices $M_N$ turned out to be a matrix of rank one, in fact $M_N = m E_N$ (see (6)) with $m$ being the mean of the entries. In the case of band matrices we obtain instead $M_N = m P_N$ or $M_N = m B_N$ for periodic or strict band matrices respectively.
The simple ‘rank-one’-argument of Proposition 4 cannot work in the case of band matrices, since in this case the corresponding matrices $M_N$ do not have bounded rank, they may even have full rank $N$. However, we will develop a more refined argument that is based on Lemma 9 and on the following observation.

**Proposition 11** Let $A$, $R$ be real symmetric $N \times N$ matrices and denote by $\rho$ and $\mu$ the eigenvalue distribution measures of $A$ and $B := A + R$ respectively. Then for every bounded function $f \in C^1(\mathbb{R})$ the following estimates hold.

\[ a) \quad |\int f d\rho - \int f d\mu| \leq \|R\|_{op}\|f'\|_{L_\infty}, \]

\[ b) \quad |\int f d\rho - \int f d\mu| \leq \frac{2\text{rank}(R)}{N}\|f'\|_{L_1}. \]

**Proof.** Denote by $\rho_1 \leq \cdots \leq \rho_N$ and $\mu_1 \leq \cdots \leq \mu_N$ the eigenvalues of $A$ and $B$ respectively. It is well known that the minmax principle allows to compare the spectra of the matrices $A$ and $B$ in terms of the operator norm of $R = B - A$, leading to statement a), and in terms of the rank of $R$ which is the basis for statement b).

a) Since $|\rho_j - \mu_j| \leq \|R\|_{op}$ for all $j$ we have $|f(\rho_j) - f(\mu_j)| \leq \|R\|_{op}\|f'\|_{L_\infty}$ and the claim follows by summation over $j$.

b) Denote $r := \text{rank}(R)$. Then for all $1 \leq j \leq N$ the eigenvalue $\mu_j$ lies in the interval $[\rho_j - r, \rho_j + r]$ where we set $\rho_i := -\infty$ for $i \leq 0$ and $\rho_i := \infty$ for $i \geq N + 1$. Hence

\[ |f(\rho_j) - f(\mu_j)| \leq \left| \int_{\rho_j}^{\rho_j + r} |f'(x)| dx \right| \leq \int_{\rho_j - r}^{\rho_j + r} |f'(x)| dx \]

and summation over $j$ yields statement b) via

\[ \sum_{j=1}^{N} |f(\rho_j) - f(\mu_j)| \leq \sum_{j=1}^{N} \sum_{k=-r}^{r-1} \int_{\rho_j + k}^{\rho_j + k + 1} |f'| \leq 2r \sum_{l=0}^{N} \int_{\rho_l}^{\rho_l + 1} |f'| = 2r\|f'\|_{L_1}. \]

In order to prove Theorem 10 we provide a second auxiliary result that shows how the estimates of Proposition 11 can be used to conclude persistence of weak convergence in probability.

**Lemma 12** Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $\mu_N^\omega$, $\rho_N^\omega$, $\mu$ be probability measures on $(\mathbb{R}, B(\mathbb{R}))$ for every $\omega \in \Omega$. Assume that $(\rho_N^\omega)_N$ converges weakly in probability to $\mu$. Moreover, suppose that there exists a real-valued sequence $(c_N)_N$ that converges to 0 such that for all $\omega \in \Omega$ and all functions $f \in \mathcal{S}$ with

\[ \mathcal{S} := \{ f \in C^1(\mathbb{R}) : \|f\|_{L_1} + \|f'\|_{L_\infty} < \infty \} \]

\[ \mathbb{E}[|f(\rho_N^\omega) - f(\mu)|] \rightarrow 0 \text{ as } N \rightarrow \infty. \]
we have
\[ \left| \int f d\rho_N^\omega - \int f d\mu_N^\omega \right| \leq c_N \left( \|f'\|_{L_1} + \|f'\|_{L_\infty} \right). \] (9)

Then $(\mu_N^\omega)_N$ also converges weakly in probability to $\mu$.

**Proof.** Fix $f \in C_b(\mathbb{R})$ and $\epsilon > 0$. Since the conditions of Definition 2 for weak convergence in probability are trivially satisfied in the case $f = 0$ one may assume $\|f\|_{L_\infty} > 0$. We first approximate $f$ by differentiable functions on suitable compact sets that depend on the given value of $\epsilon$.

Since $\mu$ is a probability measure a number $R > 0$ can be picked such that
\[ \mu(\mathbb{R} \setminus [-R, R]) \leq \frac{\epsilon}{8\|f\|_{L_\infty}}. \] (10)

Then choose $g \in C^1(\mathbb{R})$ with
\[ \sup \{ |f(x) - g(x)| : |x| \leq R + 1 \} \leq \frac{\epsilon}{8}, \] (11)
and a smooth cut-off function $\chi : \mathbb{R} \to [0, 1]$ that satisfies
\[ \chi(x) = \begin{cases} 1, & \text{if } |x| \leq R, \\ 0, & \text{if } |x| > R + 1. \end{cases} \] (12)

Write $\int f d\mu_N^\omega - \int f d\mu = \sum_{i=1}^4 \Delta_i^\omega$ with
\[ \Delta_1^\omega := \int (1 - \chi) d\mu_N^\omega - \int (1 - \chi) d\mu, \quad \Delta_2^\omega := \int g \chi d\mu_N^\omega - \int g \chi d\rho_N^\omega, \]
\[ \Delta_3^\omega := \int (f - g) \chi d\mu_N^\omega - \int (f - g) \chi d\rho_N^\omega, \quad \Delta_4^\omega := \int f \chi d\rho_N^\omega - \int f \chi d\mu. \]

We estimate
\[ |\Delta_i^\omega| \leq \|f\|_{L_\infty} (|\int (1 - \chi) d\mu_N^\omega| + |\int (1 - \chi) d\mu|) \leq \|f\|_{L_\infty} (\Gamma_1^\omega + \Gamma_2^\omega + 2\Gamma_3) \]
where
\[ \Gamma_1^\omega := \left| \int (1 - \chi) d\mu_N^\omega - \int (1 - \chi) d\rho_N^\omega \right|, \]
\[ \Gamma_2^\omega := \left| \int (1 - \chi) d\rho_N^\omega - \int (1 - \chi) d\mu \right|, \]
\[ \Gamma_3 := \left| \int (1 - \chi) d\mu \right|. \]

Use (10), (12) to bound $\Gamma_3$, (11), (12) for $|\Delta_3^\omega|$, and (9) for $|\Delta_2^\omega|$ and $\Gamma_4^\omega$. 

14
By the hypothesis of Lemma 12 there exists $N_0$ such that for all $N \geq N_0$
\[ c_N \left[ \| (g\chi)' \|_1 + \| (g\chi)' \|_\infty + \| f \|_\infty \left( \| (1 - \chi)' \|_1 + \| (1 - \chi)' \|_\infty \right) \right] \leq \frac{\epsilon}{4}. \]
Combining all these estimates we obtain for all $\omega \in \Omega$ and all $N \geq N_0$:
\[ \left| \int f \, d\mu^\omega_N - \int f \, d\mu \right| \leq \frac{3}{4} \epsilon + \| f \|_\infty \Gamma^\omega_{1/2}. \]

Since $(\rho^\omega_N)_N$ converges weakly in probability to $\mu$ one easily concludes from the last inequality that $\mathbb{P}(\int f \, d\mu^\omega_N - \int f \, d\mu > \epsilon) \to 0$ as $N \to \infty$.

Now we have gathered all the technical ingredients to derive the statement of Theorem 10 from its special version with centred entries.

**Proof of Theorem 10.** We begin with the periodic case. Let $\Pi_N$ be the orthogonal projection on the spectral subspace of $P_N$ with respect to the eigenvalues with absolute value $\leq 4 \sqrt{w_N}$ and set $S_N := P_N \Pi_N$ and $R_N := P_N (1 - \Pi_N)$. Then $\|S_N\|_{op} \leq \sqrt{w_N}$. Moreover, the rank of $R_N$ equals the number $r$ of eigenvalues of $P_N$ larger than $4 \sqrt{w_N}$. By Lemma 9 the number $r$ can be estimated:
\[ r \leq \left| \left\{ j \in \{0, 1, \ldots, N - 1\} \mid \sin\left(\frac{j\pi}{N}\right) < \frac{1}{\sqrt{w_N}} \right\} \right|. \]
Using $\sin(\pi x) \geq 2x$ for $x \in [0, 1/2]$ one may deduce that $r \leq 1 + N/\sqrt{w_N}$.

Denote by $\mu^\omega_N$ the eigenvalue distribution measure of $W_{per}^N / \sqrt{w_N}$ and by $\rho^\omega_N$ the eigenvalue distribution measure of $(W_{per}^N - mP_N) / \sqrt{w_N}$ where again $m$ denotes the mean of the matrix entries. Applying Proposition 11 twice we obtain for all bounded functions $f \in C^1(\mathbb{R})$ and for all $\omega \in \Omega$:
\[ \left| \int f \, d\rho^\omega_N - \int f \, d\mu^\omega_N \right| \leq \frac{m}{\sqrt{w_N}} \|S_N\|_{op} \|f'\|_L_\infty + \frac{2}{N} \text{rank} (R_N) \|f'\|_L_1 \leq \frac{m}{\sqrt{w_N}} \|f'\|_L_\infty + 2 \left( \frac{1}{\sqrt{w_N}} + \frac{1}{N} \right) \|f'\|_L_1. \]

Since $W_{per}^N - mP_N$ corresponds to the centred case for which we know Theorem 10 to hold, we have the desired convergence of $\rho^\omega_N$ to the semicircle $\sigma_v$. In order to transfer this result to the eigenvalue distribution measures $\mu^\omega_N$ of $W_{per}^N / \sqrt{w_N}$ we distinguish between statements a) and b) of the Theorem.

For a) the claim follows from Lemma 12 and estimates (14). In the situation of b) we proceed differently. Let $\omega \in \Omega_0$ be contained in the set of full measure
for which \( (\rho_N) \) converges weakly to the semicircle law. Using in addition inequalities (14) we deduce for all infinitely differentiable functions with compact support \( f \in C^\infty_0(\mathbb{R}) \) that

\[
\int f \, d\mu_N \to \int f \, d\sigma_v
\]
as \( N \to \infty \). Hence vague convergence of \( (\mu_N) \) is established. As the limiting measure \( \sigma_v \) is a probability measure, vague convergence implies weak convergence \( \mu_N \Rightarrow \sigma_v \) for all \( \omega \in \Omega_0 \).

Finally, we turn to the case of strict band matrices. Observe that

\[
\text{rank} \left( P_N - B_N \right) \leq 2 \min(b_N, N - b_N - 1).
\]

If \( \frac{b_N}{N} \) converges to either 0 or 1 we conclude that \( \frac{1}{N} \text{rank}(W_{\text{per}} - W) \to 0 \) as \( N \to \infty \). Thus statement 2 of Theorem 10 can be deduced from statement 1 via Proposition (11b) in the same way as statement 1 was inferred from the case of centred ensembles above.

\section{3 Exchangeable Random variables and de Finetti matrix ensembles}

The main result in this section is Theorem 23 that shows for large classes of band matrices with exchangeable entries that include in particular the case of full matrices that the empirical eigenvalue measures of the appropriately rescaled random matrices still converge to a semicircle \( \sigma_v \). In contrast to the Wigner case the scale \( V \) of the semicircle is now random.

In Subsection 3.3 we obtain a deterministic limit law by downgrading the quality of convergence from almost sure convergence to convergence in expectation. In addition, we can characterize all cases for which the deterministic law is a semicircle.

In order to get started we first explain de Finetti’s Theorem in a somewhat generalized form that links sequences of exchangeable real-valued random variables to i.i.d. sequences.

\subsection{3.1 Exchangeable random variables}

After the definition of exchangeable sequences of random variables we discuss de Finetti’s theorem on their representation as averages of i.i.d. sequences in a form that is suitable for the present paper. As a first application we prove a strong law of large numbers that differs from the classic result for i.i.d. variables only in the fact that the limit may be a random variable rather than a constant.
Definition 13 A finite sequence \((\xi_i)_{1 \leq i \leq N}\) of random variables with underlying probability space \((\Omega, \mathcal{F}, \mathbb{P})\) is called exchangeable, if for all permutations \(\pi\) on \(\{1, \ldots, N\}\), and all \(F \in \mathcal{F}\) it is true that
\[
\mathbb{P}\left( (\xi_1, \ldots, \xi_N) \in F \right) = \mathbb{P}\left( (\xi_{\pi(1)}, \ldots, \xi_{\pi(N)}) \in F \right).
\]

An infinite sequence \((\xi_i)_{i \in \mathbb{N}}\) is called exchangeable if the finite sequences \((\xi_i)_{1 \leq i \leq N}\) are exchangeable for all \(N\).

A celebrated result of de Finetti \([10, 11]\) characterizes infinite exchangeable sequences with values in \(\{-1, +1\}\). For each such sequence \(\{\xi_i\}\) there is a probability measure \(\mu\) on \([-1, 1]\) such that
\[
\mathbb{P}\left( \{\xi_i\} \in F \right) = \int P_t(F) \, d\mu(t) \tag{16}
\]
where \(P_t\) is the infinite product \(\bigotimes_{i \in \mathbb{N}} \lambda_i\) on \(\{-1, +1\}^\mathbb{N}\) of the measures \(\lambda_i\) on \(\{-1, +1\}\) given by \(\lambda_i(\{1\}) = \frac{1}{2}(1 + t)\) and \(\lambda_i(\{-1\}) = \frac{1}{2}(1 - t)\).

Hewitt-Savage \([15\text{, Theorem 7.4}]\) extended de Finetti’s theorem to exchangeable sequences with values in rather general spaces. We will need here only the case of \(\mathbb{R}\)-valued random variables and formulate it in a form which we found convenient for our purpose.

For any probability measure \(\lambda\) on \(\mathbb{R}\) (as always equipped with the Borel \(\sigma\)-algebra) we denote by \(P_\lambda\) the product measure \(\bigotimes_{i \in \mathbb{N}} \lambda\) on \(\mathbb{R}^\mathbb{N}\).

We also denote by \(\mathcal{M}_1 = \mathcal{M}_1(\mathbb{R})\) the space of all probability measures on \(\mathbb{R}\) equipped with the topology of weak convergence.

Theorem 14 Let \((\xi_i)_{i \in \mathbb{N}}\) be an exchangeable sequence of \(\mathbb{R}\)-valued random variables on the probability space \((\Omega, \mathcal{F}, \mathbb{P})\).

Then there is a probability space \((T, \mathcal{T}, \mu)\) and a measurable mapping \(\Lambda : T \to \mathcal{M}_1(\mathbb{R})\) such that
\[
\mathbb{P}\left( \{\xi_i\} \in F \right) = \int P_{\Lambda(\tau)}(F) \, d\mu(\tau) \tag{17}
\]
We call the probability measure \(\mu\) the de Finetti measure associated with the sequence \(\{\xi_i\}\).

For details see e. g. \([3]\).

Remark 15 1. For \(T\) one can always choose \(T = \mathcal{M}_1\) and for \(\Lambda\) the identity. Usually the generalized de Finetti Theorem is formulated with this choice. For our purpose we prefer the above equivalent but somewhat more flexible version.
2. When the $\xi_i$ have values in $\{-1, 1\}$ we recover de Finetti’s original case and we may chose $T = [-1, 1] \cong M_1(\{-1, 1\})$. We refer to this as the spin case.

3. The members of an exchangeable sequence are identically distributed but in general not independent.

4. Observe that $\mathbb{E}(|\xi_1|^p) < \infty$ for some $0 < p < \infty$ implies that the $p$-th moment of $\Lambda_\tau$ exists for $\mu$-almost all $\tau$. This will be used in the following convention.

**Convention 16** When speaking of an exchangeable sequence of real-valued random variables $(\xi_i)_i$ in the following we will always tacitly suppose that all moments of $\xi_1$ (and therefore of all $\xi_i$ by Remark [15]) are finite. Due to the last observation in Remark [15] we may and will assume that the moments of the corresponding measures $\Lambda_\tau$ (as in (17)) are finite for all $\tau \in T$.

For $\tau \in T$ we introduce the moments and the variance of $\Lambda_\tau$:

$$m_k(\tau) := \int x^k \, d\Lambda_\tau(x),$$

$$v(\tau) := m_2(\tau) - m_1(\tau)^2.$$  \hspace{1cm} (18)  \hspace{1cm} (19)

According to Convention[16] the moments $m_k(\tau)$ are finite for all $\tau \in T$ and all $k$.

Furthermore, denote by $\mu_1$ resp. $\nu$ the push forwards of the measure $\mu$ on $T$ under the maps $\tau \mapsto m_1(\tau)$ resp. $\tau \mapsto v(\tau)$. Observe that $\mu_1, \nu$ are both probability measures on $\mathbb{R}$ with $\text{supp}(\nu) \subset [0, \infty)$. For the spin case defined in Remark [15] it is straightforward to compute $m_1(\tau) = \tau$, thus $\mu_1 = \mu$, and $v(\tau) = 1 - \tau^2$.

The following proposition formulates a strong law of large numbers for sequences of exchangeable $\mathbb{R}$-valued random variables. As we see below it is a simple consequence of the corresponding classic law for i.i.d. sequences. Nevertheless, there is a significant difference between these two cases. For exchangeable sequences the limit is generally not a number but a random variable.

**Proposition 17** Let $(\xi_i)_i$ be a sequence of exchangeable $\mathbb{R}$-valued random variables and recall (17) as well as Convention [16]. Define for $n \in \mathbb{N}$ the random variables

$$M_n := \frac{1}{n} \sum_{i=1}^{n} \xi_i.$$  \hspace{1cm} (18)

Then the sequence $(M_n)_n$ converges $\mathbb{P}$-almost surely to a random variable $M$. Moreover, the limit satisfies $M = m_1(\tau)$ almost surely with respect to $P_{\Lambda_\tau}$. Thus
the law for the random variable $M$ is given by the push forward $\mu_1$ of the first moment $m_1$.

**Proof.** Under the probability measure $P_{\Lambda}$, the random variables $M_n$ converge to $m_1(\tau)$ almost surely by the classic strong law of large numbers. Since

$$\mathbb{P}(M_N \to M) = \int P_{\Lambda}(M_N \to M) \, d\mu(\tau)$$

the claim follows.

Applying Proposition 17 in addition to the squares of the random variables, which also form an exchangeable sequence, we obtain

**Proposition 18** In the situation of Proposition 17 define for $n \in \mathbb{N}$ the random variables

$$V_n := \frac{1}{n} \sum_{i=1}^{n} \xi_i^2 - \left(\frac{1}{n} \sum_{i=1}^{n} \xi_i\right)^2.$$

Then the sequence $(V_n)_n$ converges $\mathbb{P}$-almost surely to a random variable $V$. Moreover, the limit satisfies $V = v(\tau)$ almost surely with respect to $P_{\Lambda}$ and the law for the random variable $V$ is given by the push forward $\nu$ of the variance $v$.

**Definition 19** For sequences $(\xi_i)_i$ of exchangeable $\mathbb{R}$-valued random variables we call the random variable $M$ defined in Proposition 17 the (limiting) empirical mean and the random variable $V$ defined in Proposition 18 the (limiting) empirical variance.

### 3.2 De Finetti band ensembles and the random semicircle law

In this subsection we transfer the assertion of Theorem 10 to band matrices with entries drawn from an exchangeable sequence that we call de Finetti band ensembles.

**Definition 20** Let $(b_N)_N$ be a sequence of integers with $0 \leq b_N \leq N - 1$ and denote the band matrices $B_N$ and $P_N$ as in Definition 7a). An ensemble of a family of $N \times N$ real symmetric matrices $(F_N)_N / (F_{N}^{\text{per}})_N$ is called a strict de Finetti band ensemble / a periodic de Finetti band ensemble if the entries of $F_N / F_{N}^{\text{per}}$ are zero whenever the corresponding entries of $B_N / P_N$ are zero and if all other entries of $F_N(i, j) / F_{N}^{\text{per}}(i, j)$ are filled for each $N$, $1 \leq i \leq j \leq N$, by the first entries of a fixed sequence of exchangeable $\mathbb{R}$-valued random variables. The remaining entries are then determined by the symmetry of the matrix.
Remark 21 We recall that we always assume that Convention 16 is satisfied.

Remark 22 Note that the exchangeability of the random variables \((\xi_i)_i\) implies that the ensemble depends neither on the set of \(\xi_i\) that is selected to fill the \(N\)-th matrix (as long as different \(\xi_i\)s are used for different entries of the upper triangular parts of the random matrices), nor on the specific order in which we fill the matrix. For example, we could fill it row-wise or sub-diagonal by sub-diagonal. For other sequences of random variables the way of filling the matrix may be crucial (see [21]).

We are now ready to state our main result on the limiting spectral density of de Finetti band matrices. Observe that this result also includes the case of full de Finetti ensembles by choosing the bandwidths sufficiently large.

Theorem 23 Let \((b_N)_N\) be a sequence of integers with \(0 \leq b_N \leq N-1\) and \(b_N \to \infty\). Recall from Definition 20 the meaning of the corresponding de Finetti band ensembles \((F_N)_N\) and \((F_{per}^N)_N\) with bandwidth \(w_N := \min(N, 2b_N + 1)\) (Definition 7). Denote by \((\xi_i)_i\) the sequence of exchangeable random variables from which the entries of the matrices are drawn and let \(V\) be its empirical variance (Definition 19).

a) We distinguish the two cases of periodic and strict band matrices.

1. The empirical eigenvalue measures \(\mu^N_{\omega}\) of the periodic band matrices \(F_{per}^N / \sqrt{w_N}\) converge weakly in probability to the (random semicircle) measure \(\sigma_V(\omega)\).

2. Statement 1 also holds for the empirical eigenvalue measures \(\mu^N_{\omega}\) of the strict band matrices \(F_N / \sqrt{w_N}\) if we require in addition that the scaled half-widths \(\frac{b_N}{N}\) converge either to 0 or to 1 for \(N \to \infty\).

b) Let us add to the general assumption \(b_N \to \infty\) above the summability condition \(\sum_N(Nb_N)^{-1} < \infty\). Then both statements of part a) remain true if we replace the assertion of weak convergence in probability by weak convergence \(\mathbb{P}\)-almost surely.

Proof. By assumption the measure \(\mathbb{P}\) associated with either the periodic or the strict de Finetti band ensembles have the form (17):

\[
\mathbb{P}\left((X_{i_1j_1}, \ldots, X_{i_kj_k}) \in A\right) = \int P_{\Lambda, \tau}(A) \, d\mu(\tau).
\]  

(20)

for \(\{i_1, j_1\}, \ldots, \{i_k, j_k\}\) pairwise distinct and for every Borel set \(A \subset \mathbb{R}^k\). Fix \(\tau \in T\) and consider the ensembles \((F_N)_N\) and \((F_{per}^N)_N\) with respect to the probability measure \(P_{\Lambda, \tau}\). Then they are strict and periodic Wigner band matrices respectively with variance \(v(\tau)\) (see Definition 7). We first consider part b). Then Theorem
b) implies that the empirical eigenvalue measures $\mu_N^\omega$ converge in both cases 1 and 2 to $\sigma_{v(\tau)}$ almost surely with respect to $P_{\Lambda,\tau}$. By Proposition 18 we have in addition $P_{\Lambda,\tau}$-almost surely that $v(\tau) = V(\omega)$. Hence

$$P \left( \mu_N^\omega \Rightarrow \sigma_{V(\omega)} \right) = \int P_{\Lambda,\tau} \left( \mu_N^\omega \Rightarrow \sigma_{V(\omega)} \right) d\mu(\tau) = 1. \quad (21)$$

In order to prove part a) fix $f \in C_b(\mathbb{R})$ and $\epsilon > 0$. Then part a) of Theorem 10 implies for all $\tau \in T$ that

$$P_{\Lambda,\tau} \left( \left| \int f(x) \, d\mu_N^\omega (x) - \int f(x) \, d\sigma_{v(\tau)}(x) \right| > \epsilon \right) \to 0$$

as $N \to \infty$. Integration of this relation over $T$ with respect to the de Finetti measure $\mu$ together with Lebesgue’s theorem of dominated convergence yield the claim.

The proof of Theorem 23 also implies the following conditional convergence to deterministic semicircle laws.

**Corollary 24** Recall the definition of the push forward measure $\nu$ below equation (19). Under the assumptions and with the notation of Theorem 23 b) we have for $\nu$-almost all $v \in [0, \infty)$:

$$P \left( \mu_N^\omega \Rightarrow \sigma_v \, \middle| \, V(\omega) = v \right) = 1.$$

### 3.3 Expected limiting spectral density

The weak almost sure convergence as well as the weak convergence in probability asserted in Theorem 23 both imply weak convergence in expectation (see Definition 2). Applying Fubini’s theorem to the right hand side of (4) one may choose the limiting measure to be deterministic by taking the expectation of the limiting measures $\sigma_{V(\omega)}$. The result of this averaging is the measure $\sigma_\mu$ on $\mathbb{R}$ that we define via the Riesz representation theorem through

$$\int f(x) \, d\sigma_\mu(x) := \mathbb{E} \left( \int f(x) \, d\sigma_{V(\omega)}(x) \right) \quad (22)$$

$$= \int_{\mathbb{T}} \int_{\mathbb{R}} f(x) \, d\sigma_{v(\tau)}(x) \, d\mu(\tau) \quad (23)$$

for each bounded continuous function $f \in C_b(\mathbb{R})$. The last equality follows from Proposition 18. We summarize:

**Theorem 25** Under the assumptions and with the notation of Theorem 23 the empirical eigenvalue measures $\mu_N^\omega$ converge weakly in expectation to the measure $\sigma_\mu$. 21
Next we derive a representation for the limiting measure $\sigma_{\mu}$.

**Proposition 26** The Borel probability measure $\sigma_{\mu}$ that Equation (23) defines on $\mathbb{R}$ is symmetric in the sense $\sigma_{\mu}(A) = \sigma_{\mu}(-A)$. Moreover,

$$\sigma_{\mu} = \mu(\{v(\tau) = 0\}) \delta_0 + \sigma^\text{abs}_{\mu}$$

with $\sigma^\text{abs}_{\mu}$ being absolutely continuous with Lebesgue density

$$\rho_{\mu}(x) := \frac{d\sigma^\text{abs}_{\mu}}{dx}(x) = \frac{1}{2\pi} \int_{x^2/4}^{\infty} \frac{\sqrt{4v-x^2}}{v} \, dv < \infty,$$  \hspace{1cm} (24)

for $0 < |x| < \infty$. The even function $\rho_{\mu}$ is decreasing with $|x|$.

The proof of this proposition is elementary, essentially an application of Fubini’s theorem to the integrals $\int_0^\infty \int_{-\infty}^s d\sigma_v(x) \, dv$. Note that the monotonicity of $\rho_{\mu}$ is obvious from the definition. It is also the reason why the finiteness stated in (24) can not only be shown to hold for almost all $x \in \mathbb{R} \setminus \{0\}$ (by Fubini), but for all of them.

**Remark 27** Spin case. Observe first that in the spin case the support of the measure $\sigma_{\mu}$ is contained in $[-2,2]$ since the variances $v(\tau) = 1 - \tau^2$ never exceed the value 1. Moreover, relation (23) immediately leads to an expression for $\sigma_{\mu}$ directly in terms of the de Finetti measure $\mu$. Set $a(x) := \sqrt{4 - x^2}$, then

$$\sigma_{\mu} = \mu(\{-1,1\}) \delta_0 + \sigma^\text{abs}_{\mu}$$

with

$$\frac{d\sigma^\text{abs}_{\mu}}{dx}(x) = \frac{1}{2\pi} \int_{-a(x)/2}^{a(x)/2} \frac{\sqrt{a(x)^2 - 4t^2}}{1-t^2} \, d\mu(t), \quad \text{for } 0 < |x| \leq 2.$$  \hspace{1cm} (25)

One may evaluate (25) explicitly in special cases. For example, if the de Finetti measure $\mu$ equals the uniform distribution $\mu_{\text{uni}}$ on $[-1,1]$, i.e. if we have $d\mu_{\text{uni}}(t) := \frac{1}{2} \chi_{[-1,1]}(t) \, dt$, then the corresponding limiting spectral measure is given by $d\sigma_{\mu_{\text{uni}}}(x) = \frac{1}{4}(2 - |x|)\, dx$.

It is obvious from (23) that $\sigma_{\mu}$ is a semicircle if $v(\tau) = a$ for some $a \geq 0$ almost surely w.r.t. the measure $\mu$. We conclude the present section by showing that the converse is also true:

$$\sigma_{\mu} \text{ is a semicircle } \iff \text{there exists } a \in [0, \infty) \text{ such that } v = \delta_a$$  \hspace{1cm} (26)

We only need to consider ”$\Rightarrow$”. Suppose that $\sigma_{\mu} = \sigma_s$ for some $s \in [0, \infty)$. Then all moments of $\sigma_{\mu}$ and $\sigma_s$ agree. By linear scaling

$$m_v^{(k)} := \int x^k \, d\sigma_v(x) = v^{k/2} m_1^{(k)} \quad \text{for all } v \in [0, \infty), \ k > 0.$$
An application of Fubini’s theorem gives
\[ \int x^k d\sigma(x) = m_4^{(k)} \int_0^\infty v^{k/2} d\nu(v). \]
The equality of the fourth and second moments of \( \sigma_4 \) and \( \sigma_2 \) then yields
\[ \int_0^\infty v^2 d\nu(v) = s^2 = \left( \int_0^\infty v d\nu(v) \right)^2. \]
This implies that the Cauchy-Schwarz inequality \((\int f d\nu)^2 \leq \int f^2 d\nu\) is an equality for \( f(v) = v \). Hence the identity \( f \) is a constant function in \( L^2(d\nu) \) proving the claim.

## 4 The operator norm for band random matrices

The semicircle law for Wigner band ensembles suggests that in the case of centred entries the operator norm should asymptotically be of the order of the square root of the bandwidth \( w_N \). It was already observed in [7] that this cannot hold if the bandwidths do not grow at least at some logarithmic rate with the matrix size. In the first subsection we provide in Theorem 28 and in Remark 29 positive results in this direction that guarantee for centred Wigner band ensembles an almost sure upper bound on the operator norm that grows proportionally with \( \sqrt{w_N} \) if the bandwidth satisfies some growth condition. The second subsection considers the situation of Wigner band ensembles with arbitrary means and de Finetti band ensembles.

### 4.1 Centred Wigner band ensembles

The method of moment was used in [6] to obtain the almost sure limit of the appropriately rescaled operator norms for centred Wigner ensembles, see also [13]. We follow the basic strategies of [6] in the form presented in [24] and extend the arguments to band matrices. For the convenience of the reader we present all details but we refer her or him to [24, Section 2.3] for detailed motivation. A technical but essential lemma that provides bounds on the expected values of traces of matrix powers for exponents that might grow with the matrix dimension (see Lemma 34) is deferred to the Appendix.

**Theorem 28** Let \((W_N)_N\) and \((W_{\text{per}})_N\) be centred strict or periodic Wigner band ensembles with variance \( v \) as introduced in Definition 7. Suppose furthermore that there exist positive constants \( c \) and \( q \) such that the corresponding bandwidths \( w_N \) satisfy the growth condition \( w_N \geq cN^q \).

I. In both cases \( X_N = W_N \) and \( X_N = W_{\text{per}} \) we have
\[ \mathbb{P} \left( \limsup_{N \to \infty} \frac{\|X_N\|_{\text{op}}}{\sqrt{w_N}} \leq 2\sqrt{v} \right) = 1. \]
II. In the case of periodic ensembles $X_N = W_N^{\text{per}}$ we obtain the stronger result
\[
\mathbb{P} \left( \lim_{N \to \infty} \frac{||X_N||_{\text{op}}}{\sqrt{w_N}} = 2\sqrt{v} \right) = 1. \quad (28)
\]

III. Result (28) also holds for strict band matrices $X_N = W_N$ with half-widths $b_N$ satisfying $\lim_{N \to \infty} \frac{b_N}{N} \in \{0; 1\}$.

**Remark 29** It has already be shown in [7] that $||X_N||_{\text{op}}/\sqrt{w_N}$ is almost surely unbounded if the bandwidth $w_N$ does not grow at least at some logarithmic rate.

Theorem 28 provides a positive result by specifying a minimal growth rate for the sequence of bandwidths that guarantees almost surely that the scaled operator norms $||X_N||_{\text{op}}/\sqrt{w_N}$ remain bounded. Following the steps in the proof of Theorem 28 shows that the growth condition on the bandwidths is intimately connected to the decay of the tail of the law for the matrix entries. Recall that in the statement of Theorem 28 this decay is implicitly given by our general assumption that all moments of the matrix entries are finite. Less decay leads to stronger conditions on the growth of the bandwidths. Let us assume, in the opposite direction, that the distribution of the matrix entries has compact support. Then the truncation procedure in the proof of Theorem 28 is not needed and Lemma 35 immediately yields the growth condition $w_N \geq c (\log N)^{14+\epsilon}$ on the bandwidths where $c$ and $\epsilon$ can be any positive constants.

**Proof of Theorem 28** Multiplying the matrices of the ensemble by the factor $1/\sqrt{v}$ and treating the trivial case $v = 0$ separately, we may restrict ourselves to the case $v = 1$.

**Part I.** Fix $\delta > 0$. By the Borel-Cantelli Lemma it suffices to show
\[
\sum_{N=1}^{\infty} \mathbb{P} \left( ||X_N||_{\text{op}} \geq (2 + \delta)\sqrt{w_N} \right) < \infty. \quad (29)
\]

Set $\bar{K}_N := N^\alpha$ with any exponent $0 < \alpha < q/2$ and define a truncated version of the ensemble $(X_N)_N$ by
\[
\bar{Y}_N(i, j) := X_N(i, j) \cdot 1_{\{|X_N(i,j)| \leq \bar{K}_N\}}.
\]

Observe that $(\bar{Y}_N)_N$ might not be an auxiliary Wigner ensemble AWE (see Definition 33 in the Appendix), because its entries are not necessarily centred. A second modification is therefore needed.

The expectation $E_N := \mathbb{E}(\bar{Y}_N)$ is an $N \times N$ matrix with entries that take only the values 0 or $\mathbb{E}(\bar{Y}_N(1, 1))$. Then $e_N := ||\mathbb{E}(\bar{Y}_N(1, 1))||$ defines an upper bound
on the modulus of all entries of $E_N$. Finally, define $Y_N := \tilde{Y}_N - E_N$. Clearly all entries of $Y_N$ are centred and we have

$$\mathbb{E}(Y_N^2(i, j)) = \mathbb{V}(Y_N(i, j)) = \mathbb{V}(\tilde{Y}_N(i, j)) \leq \mathbb{E}(\tilde{Y}_N^2(i, j)) \leq \mathbb{E}(X_N^2(i, j)) = 1$$

for all $1 \leq i \leq j \leq N$ so that condition (C1) of Definition 33 is satisfied for $(Y_N)_N$. It is then clear that $(Y_N)_N$ is an AWE with support bounds $K_N := K_N + e_N$ and maximal row occupancies $n_N := w_N$.

Next we derive a bound on the entries of the matrix $E_N$. Here we use our assumption that all moments exist so that $C_p := \int |x|^p d\rho(x) < \infty$ for all $p > 0$ where $\rho$ denotes the common distribution of the matrix entries $X_N(i, j)$. We prove for all $p \geq 1$ that

$$e_N \leq C_p K_N^{-(p-1)} = C_p N^{-\alpha(p-1)}. \quad (30)$$

Indeed, since all entries $X_N(i, j)$ are centred we have

$$\mathbb{E} \left( X_N(1, 1) \cdot 1_{\{|X_N(1, 1)| \leq K_N\}} \right) = -\mathbb{E} \left( X_N(1, 1) \cdot 1_{\{|X_N(1, 1)| > K_N\}} \right)$$

and (30) follows by standard arguments. Using (30) together with our choice $0 < \alpha < q/2$ and together with the growth condition on the bandwidth, we may establish hypothesis (49) of Lemma 35 and we learn

$$\sum_{N=1}^{\infty} \mathbb{P}(\|Y_N\|_{op} \geq \left(2 + \frac{\delta}{2}\right) \sqrt{w_N}) < \infty.$$ 

Claim (29) then follows, provided we can show

$$\sum_{N=1}^{\infty} \mathbb{P}(\|Y_N - \tilde{Y}_N\|_{op} \geq \frac{\delta}{2} \sqrt{w_N}) < \infty \quad \text{and} \quad (31)$$

$$\sum_{N=1}^{\infty} \mathbb{P}(X_N \neq \tilde{Y}_N) < \infty. \quad (32)$$

The first estimate (31) follows from (30) in the case $\alpha(p-1) > 1/2$ via

$$\|E_N\|_{op} \leq \left( \sum_{i,j=1}^{N} E_N(i, j)^2 \right)^{1/2} \leq C_p N^{-\alpha(p-1)} (N w_N)^{1/2}$$

which implies that only finitely many terms in the sum in (31) do not vanish. Finally, we turn to statement (32). Markov’s inequality gives

$$\mathbb{P} \left( |X_N(i, j)| > K_N \right) \leq C_p K_N^{-p} = C_p N^{-\alpha p}.$$
Choosing \( p > 3/\alpha \) and using the independence of the entries of \( X_n \) we have
\[
P \left( \tilde{Y}_n \neq X_n \right) \leq \frac{N(N+1)}{2} C \rho N^{\alpha p} = O_p \left( N^{2-\alpha p} \right)
\]
and the summability claimed in (32) is proved. Observe that the conditions \( p \geq 1 \) and \( \alpha(p-1) > 1/2 \) that we assumed in our arguments above are weaker than \( p > 3/\alpha \) since \( 0 < \alpha < q/2 \leq 1/2 \).

**Parts II and III.** All the cases considered in parts II and III of Theorem 28 have the common feature that the empirical eigenvalue measures \( \mu^N_\omega \) of \( X_n/\sqrt{w_N} \) converge weakly almost surely to the semicircle \( \sigma_{v=1} \) (see Theorem [10b]) and recall that we have restricted ourselves to the case of unit variance \( v \). Thus for all bounded continuous functions \( f : \mathbb{R} \rightarrow \mathbb{R} \):
\[
P \left( \lim_{N \to \infty} \int f d\mu^N_\omega = \int f d\sigma_{v=1} \right) = 1 .
\]
(33)

Since we have already proved an upper bound in part I it suffices to show that
\[
P \left( \liminf_{N \to \infty} \frac{\|X_n\|_{op}}{\sqrt{w_N}} < 2 - \delta \right) = 0
\]
(34)
for all small but fixed \( \delta > 0 \). Choose a continuous functions \( f_\delta : \mathbb{R} \rightarrow [0, 1] \) that takes the value 1 on \( \mathbb{R} \setminus (-2; 2) \) and the value 0 on \([-2 + \delta; 2 - \delta] \). In order to verify (34) it is enough to convince ourselves that
\[
\left\{ \liminf_{N \to \infty} \frac{\|X_n\|_{op}}{\sqrt{w_N}} < 2 - \delta \right\} \subset \left\{ \lim_{N \to \infty} \int f_\delta d\mu^N_\omega \neq \int f_\delta d\sigma_{v=1} \right\}
\]
where the inequality on the right hand side could also mean that the limit does not exist. The above inclusion can be seen as follows:

If \( \liminf_{N \to \infty} \frac{\|X_n\|_{op}}{\sqrt{w_N}} < 2 - \delta \) then there exists a subsequence \( (N_k)_k \) such that the supports of the empirical measures \( \mu^N_{N_k} \) are all contained in the set \([-2 + \delta; 2 - \delta] \) and therefore \( \int f_\delta d\mu^N_{N_k} = 0 \). However, the explicit definition of the standard semicircle \( \sigma_{v=1} \) implies \( \int f_\delta(x) d\sigma_{v=1}(x) > 0 \) and we cannot have \( \int f_\delta d\mu^N_{N_k} \rightarrow \int f_\delta d\sigma_{v=1} \) as \( N \to \infty \).

Recall from the proof of Proposition [5] that we still have to prove that the forth largest singular value \( s_{N-3} \) of a centred full Wigner ensemble with variance \( v \) behaves asymptotically like the operator norm, i.e. like the largest singular value \( s_N \), and converges almost surely to \( 2\sqrt{v} \) when divided by \( \sqrt{N} \). We now prove this claim in the more general setting of centred Wigner band matrices for which the semicircle law holds. Therefore the ensembles listed in Parts II and III of Theorem [28] and in particular full matrices, are included.

More precisely, we show:
Corollary 30  
Parts II and III of Theorem 28 also holds true if we replace (28) by

\[ P \left( \lim_{N \to \infty} \frac{s_{N-m}(X_N)}{\sqrt{w_N}} = 2\sqrt{v} \right) = 1. \]

for any fixed \( m \in \mathbb{N} \).

Proof. Since \( s_{N-m} \leq s_N \) and \( s_N \) equals the \( \ell_2 \)-operator norm the upper bound follows from Theorem 28. For the lower bound we proceed exactly as in the proof of Parts II and III of Theorem 28. Clearly, we can again restrict ourselves to the case of unit variance \( v \). Then the arguments there show for any fixed \( \delta > 0 \) that

\[ \left\{ \liminf_{N \to \infty} \frac{s_{N-m}(X_N)}{\sqrt{w_N}} < 2 - \delta \right\} \subset \left\{ \lim_{N \to \infty} \int f_\delta d\mu_N \neq \int f_\delta d\sigma_{v=1} \right\}. \]

Thus, by the semicircle law,

\[ P \left( \liminf_{N \to \infty} \frac{s_{N-m}(X_N)}{\sqrt{w_N}} < 2 - \delta \right) = 0 \]

and the claim follows. \( \blacksquare \)

4.2  Arbitrary means and the de Finetti case

The proof of Proposition 5 and statement b) of Lemma 9 suggest that the operator norm of non-centred Wigner band ensembles is asymptotically proportional to the bandwidth \( w_N \) rather than to its square root as in the centred case. We formulate this in Theorem 31 immediately for de Finetti band ensembles and remark thereafter that this includes the case of Wigner band ensembles with arbitrary means.

**Theorem 31** Let \((F_N)_N\) and \((F^\text{per}_N)_N\) be strict and periodic de Finetti band ensembles respectively (see Definition 20). Suppose furthermore that there exist positive constants \( c \) and \( q \) such that the corresponding bandwidths \( w_N \) satisfy the growth condition \( w_N \geq cN^q \). Denote by \((\xi_i)_i\) the sequence of exchangeable random variables from which the entries of the matrices are drawn. Recall also the definition of the empirical mean \( M \) in Definition 19.

I. In both cases \( X_N = F_N \) and \( X_N = F^\text{per}_N \) we have

\[ P \left( \limsup_{N \to \infty} \frac{||X_N||_{op}}{w_N} \leq |M(\omega)| \right) = 1 \quad \text{and} \quad P \left( \liminf_{N \to \infty} \frac{||X_N||_{op}}{w_N} \geq \frac{3}{4} |M(\omega)| \right) = 1. \]

(35)

II. In the case of periodic ensembles \( X_N = F^\text{per}_N \) we obtain the stronger result

\[ P \left( \lim_{N \to \infty} \frac{||X_N||_{op}}{w_N} = |M(\omega)| \right) = 1. \]

(36)
III. Result (36) also holds for strict band matrices $X_N = F_N$ with half-widths $b_N$ satisfying $\lim_{N \to \infty} \frac{b_N}{N} \in \{0; 1\}$.

Proof. Part I. Let $\mathbb{P}$ be given as in (17). With respect to the probability measure $P_{\Lambda_\tau}$ the ensembles $(F_N - m_1(\tau)B_N)_N$ and $(F_N^\per - m_1(\tau)P_N)_N$ are centred strict/periodic Wigner band matrices with variance $v(\tau)$ (Definition 7). We conclude from Theorem 28 and Proposition 17 that

$$P_{\Lambda_\tau}\left( \limsup_{N \to \infty} \frac{||F_N - m_1(\tau)B_N||_{op}}{\sqrt{w_N}} \leq 2\sqrt{v(\tau)} \right) = 1$$

and

$$P_{\Lambda_\tau}\left( \limsup_{N \to \infty} \frac{||F_N^\per - m_1(\tau)P_N||_{op}}{\sqrt{w_N}} \leq 2\sqrt{v(\tau)} \right) = 1$$

We now argue that this suffices to show (35). Let us begin with the upper bound. Using inequality $||X_N||_{op} \leq ||X_N - m_1(\tau)A||_{op} + |m_1(\tau)||A||_{op}$ for $A \in \{B_N, P_N\}$, $||A||_{op} \leq w_N$ by Lemma 9(b), and $w_N \to \infty$ for $N \to \infty$ gives the first statement of (35) with $\mathbb{P}$ being replaced by $P_{\Lambda_\tau}$. Integration over $\tau$ with respect to the de Finetti measure then yields the first relation of (35).

For the lower bound we proceed in a similar fashion. Lemma 9(b) implies $||A||_{op} \geq \frac{3}{4}w_N$ for $A \in \{B_N, P_N\}$. Together with the general inequality $||X_N||_{op} \geq -||X_N - m_1(\tau)A||_{op} + |m_1(\tau)||A||_{op}$ and with $w_N \to \infty$ integration with respect to the de Finetti measure completes the proof of Part I.

Parts II and III. The additional assumptions of Parts II and III imply that $||A||_{op}/w_N$ tends to 1 as $N \to \infty$ for $A \in \{B_N, P_N\}$ by Lemma 9(b). This allows to remove the factor $\frac{3}{4}$ in the second statement of (35) and claim (36) follows.

Remark 32. 1. Strict and periodic Wigner band ensembles are also strict and periodic de Finetti ensembles respectively, and Theorem 31 applies with the empirical mean $M$ in the relations of (35) and (36) being replaced by the mean $m$ of the Wigner ensemble. Indeed, let $\rho$ denote the law of the entries of the Wigner ensemble. Then $T = \{0\}$, $\Lambda(0) = \rho$, and $\mu = \delta_0$ provide a suitable representation (17). Moreover, the mean $m$ of the Wigner ensemble agrees with the empirical mean $M$ of the corresponding de Finetti ensemble that is deterministic in this trivial case.

2. Full de Finetti ensembles $(X_N)_N$ may be viewed as periodic de Finetti band ensembles with bandwidth $w_N = N$. In this case one can show for the second largest singular value (cf. Proposition 6)

$$\mathbb{P}\left( \lim_{N \to \infty} \frac{||X_N||_{op}'}{\sqrt{N}} = 2\sqrt{V(\omega)} \right) = 1$$

(37)
where \( V \) denotes again the empirical variance. Thus, whenever the empirical mean \( M \) does not vanish on a set \( \Omega' \subset \Omega \) of positive probability the discrepancy on \( \Omega' \) between the growth of the operator norm (order \( N \), see Theorem 31) and the growth given by the semicircle law (order \( \sqrt{N} \), see Theorem 23) is caused by a single outlier.

In order to prove (37) it suffices to show this relation with \( P \) being replaced by \( P_\Lambda \) and \( V(\omega) \) being replaced by \( v(\tau) \) for any \( \tau \in T \). This, however, is exactly the assertion of Proposition 6.

## A Moment method for Wigner ensembles

In the main text we make use of two results, Theorem 10 b) in the centred case and Lemma 35, that are essentially known and can be proved using the method of moments. Since these proofs are not readily available in the literature we provide them in the subsequent subsections for the convenience of the reader.

The method of moments is based on the fact that a large class of probability measures \( \sigma \), including in particular all such measures with compact support, are determined by their sequence of moments \( \left( \int x^k d\sigma(x) \right)_{k \in \mathbb{N}} \). Since the sum of the \( k \)-th powers of all eigenvalues of some \( N \times N \) matrix \( X \) is given by the trace of \( X^k \) it is clear that the \( k \)-th moment of the eigenvalue distribution measure of \( X \) is given by \( \frac{1}{N} \text{tr}(X^k) \). Moreover, it follows for matrices \( X \) with real spectrum that for even positive integers \( k \) the moduli of all eigenvalues are bounded above by the \( k \)-th root of \( \text{tr}(X^k) \). This shows in a nutshell how the objects studied in this paper are related to traces of matrix powers.

Wigner introduced in [25, 26] a method to analyse the large \( N \) asymptotics of the expectations of traces of matrix powers for ensembles that now bear his name. We begin by deriving a useful representation for these expectations.

Let \( (X_N)_N \) be a Wigner ensemble that may be full or banded. Then

\[
\text{tr}(X_N^k) = \sum_{\gamma_1, \ldots, \gamma_k=1}^N X_N(\gamma_1, \gamma_2) \cdot \ldots \cdot X_N(\gamma_k, \gamma_1) = \sum_{\gamma \in \mathcal{P}} X_\gamma \quad (38)
\]

with \( X_\gamma := \prod_{s=1}^k X_N(\gamma_s, \gamma_{s+1}) \) and

\[
\mathcal{P} := \{ \gamma \in \{1, \ldots, N\}^{k+1} \mid \gamma_1 = \gamma_{k+1} \text{ and } X_\gamma \neq 0 \} . \quad (39)
\]

By the condition \( X_\gamma \neq 0 \) we mean that no factor \( X_N(\gamma_s, \gamma_{s+1}) \) in the definition of \( X_\gamma \) is identically equal to 0 due to the prescribed band structure of the matrix (see Definition 7). Thus the set \( \mathcal{P} \) not only depends on \( N \) and \( k \) but also on the half-width \( b_N \) and on the question whether strict or periodic band matrices are
considered. We would like to alert the reader that none of this required information is recorded in the notation. For each path $\gamma \in \mathcal{P}$ we denote
\[
\mathcal{E}(\gamma) := \{\{\gamma_s, \gamma_{s+1}\} \mid s \in \{1, \ldots, k\}\}
\]
the set of (undirected) edges,
\[
\eta(\gamma) := \#\mathcal{E}(\gamma),
\]
and by $e_1(\gamma), \ldots, e_{\eta(\gamma)}(\gamma)$ the elements of $\mathcal{E}(\gamma)$ in their order of appearance as one travels along the path $\gamma$.

Furthermore, for each $1 \leq i \leq \eta(\gamma)$, let $a_i(\gamma) \in \mathbb{N}$ be the multiplicity with which edge $e_i(\gamma)$ occurs and set $\xi_i(\gamma) := X_N(e_i(\gamma))$. Here we use $X_N(e) := X_N(p, q)$ for edges $e = \{p, q\}$ which is well defined by the symmetry of $X_N$. The assumed independence of matrix entries in the upper triangular part gives for all $\gamma \in \mathcal{P}$:
\[
\mathbb{E}(X_\gamma) = \mathbb{E}\left(\prod_{i=1}^{\eta(\gamma)} \xi_i(\gamma)^{a_i(\gamma)}\right) = \prod_{i=1}^{\eta(\gamma)} \mathbb{E}(\xi_i^{a_i}),
\]
where we have omitted the $\gamma$-dependency in the last term for notational simplicity. For centred Wigner ensembles we have in addition that $\mathbb{E}(X_\gamma) = 0$ if there exists an $i \in \{1, \ldots, \eta(\gamma)\}$ with $a_i = 1$. This, by the way, is the reason why the method of moments works so well in the case of centred entries and cannot be applied directly for non-vanishing means. Set
\[
\mathcal{P}_0 := \{\gamma \in \mathcal{P} \mid a_i \geq 2 \text{ for all } 1 \leq i \leq \eta(\gamma)\}.
\]
Then, for centred Wigner ensembles
\[
\mathbb{E}(\text{tr}(X_N^k)) = \sum_{\gamma \in \mathcal{P}_0} \mathbb{E}(X_\gamma).
\]
Together with formula (40), we have a representation for the expectation of the $k$-th moments of the eigenvalue distribution measures related to $X_N$ that is used in both of the following subsections.

### A.1 Proof of Theorem 10 b) in the centred case

Denote by $X_N$ a centred Wigner band ensemble, periodic or strict, that satisfies the corresponding assumptions on the sequence $(b_N)_N$ of half-widths stated in Theorem 10. Recall the notation introduced at the beginning of this appendix and denote for any path $\gamma \in \mathcal{P}_0$ by $r(\gamma)$ the number of different vertices contained
in the path. Since every edge occurs at least twice the number of different edges \( \eta(\gamma) \) is bounded above by \( \frac{k}{2} \) and consequently \( r(\gamma) \leq 1 + \frac{k}{2} \). The following asymptotics on the number of elements in \( \mathcal{P}_0 \) with a prescribed number of vertices \( 1 \leq r \leq 1 + \frac{k}{2} \) use all of the assumptions of Theorem 10 on the bandwidths (recall also the remark after definition (39)):

\[
n_{r,k}(N) := \# \{ \gamma \in \mathcal{P}_0 \mid r(\gamma) = r \} \sim c_{r,k}N^{r-1} \tag{42}
\]

as \( N \to \infty \). Here \( \sim \) means that the ratio of left hand side and right hand side converges to 1. The number \( c_{r,k} \) in (42) can be defined as the number of prototypes in \( \{ \gamma \in \mathcal{P}_0 \mid r(\gamma) = r \} \). By a prototype we understand a path \( \gamma \) that has the additional property that the vertices are numbered in the order of their occurrence in the path. For example, \( \gamma = (1, 2, 3, 2, 4, 2, 3, 2, 1) \) is a prototype in \( \mathcal{P}_0 \) for \( k = 8 \) with \( r(\gamma) = 4 \).

The asymptotic formula (42) indicates that only paths \( \gamma \in \mathcal{P}_0 \) with the maximal number of vertices \( r(\gamma) = 1 + \frac{k}{2} \) matter. For such paths, which can only occur for even values of \( k \), the number of different edges is also maximal \( \eta(\gamma) = \frac{k}{2} \) and all edges \( e_i \) have multiplicity \( a_i = 2 \). Since the second moments of all matrix entries are assumed to equal the same constant \( v \) we obtain

\[
E(\text{tr}(X_N^k)) = \begin{cases} c_{1+\frac{k}{2},k}N(vw_N)^{k/2}, & \text{if } k \text{ is even,} \\ O(Nw_N^{(k-1)/2}), & \text{if } k \text{ is odd.} \end{cases}
\]

These asymptotics show in particular that one needs to divide \( X_N \) by \( \sqrt{w_N} \) for the expected moments of the eigenvalue distribution measures to converge to some nontrivial measure. This justifies the definition of the corresponding measures \( \mu_{\omega,N}^k \) in the statement of Theorem 10. Indeed, we may conclude for every positive integer \( k \) that

\[
\lim_{N \to \infty} \mathbb{E} \left( \int x^k d\mu_{\omega,N}^k(x) \right) = \lim_{N \to \infty} \frac{E(\text{tr}(X_N^k))}{Nw_N^{k/2}} = \begin{cases} c_{1+\frac{k}{2},k}v^{k/2}, & \text{for even } k, \\ 0, & \text{for odd } k. \end{cases}
\]

It is a classical result in combinatorics (see e.g. [23], Theorem 1.51) that the numbers \( c_{1+\frac{k}{2},k} \) (\( k \) even) are given by Catalan numbers \( C_{\frac{k}{2}} = \frac{(k/2)!}{(1/2)!} \) and that

\[
\lim_{N \to \infty} \mathbb{E} \left( \int x^k d\mu_{\omega,N}^k(x) \right) = \int x^k d\sigma_v(x) \tag{43}
\]

holds for all \( k \). In order to prove part b) of Theorem 10 it suffices to enhance (43) to \( \mathbb{P} \)-almost sure convergence which then implies that \( \mu_{\omega,N}^k \) converges weakly \( \mathbb{P} \)-almost surely to \( \sigma_v \). Standard applications of the Chebyshev inequality and of the
Borel-Cantelli lemma show that this can be achieved by proving for every positive integer \( k \) the summability of the variances

\[
\sum_{N=1}^{\infty} \mathbb{V} \left( \int x^k \, d\mu_N^\omega(x) \right) < \infty. \tag{44}
\]

Using the notation introduced in (38) and (39) we obtain

\[
\mathbb{V} \left( \int x^k \, d\mu_N^\omega(x) \right) = \frac{1}{N^2 w_N^k} \sum_{\gamma, \gamma' \in \mathcal{P}} \mathbb{E}(X_{\gamma}X_{\gamma'}) - \mathbb{E}(X_{\gamma})\mathbb{E}(X_{\gamma'}) \tag{45}
\]

Observe that pairs \((\gamma, \gamma') \in \mathcal{P}^2\) that do not share a common edge do not contribute to the sum, because \(\mathbb{E}(X_{\gamma}X_{\gamma'}) = \mathbb{E}(X_{\gamma})\mathbb{E}(X_{\gamma'})\) holds in this case by the assumed independence of the matrix entries in the upper triangular part of \(X_N\). For pairs \((\gamma, \gamma')\) that do share a common edge we construct a path \(\hat{\gamma} = \hat{\gamma}(\gamma, \gamma') \in \mathcal{P}\) with \(2k - 2\) edges (!) in the following way. Choose \(i_0\) such that \(e_{i_0}\) is the first edge in \(E(\gamma)\) that also appears in \(E(\gamma')\) and let \(i'_0\) be minimal with \(e_{i_0} = e_{i'_0}\). Furthermore, denote \(P := \gamma_{i_0}\) and \(Q := \gamma_{i_0+1}\). By construction we have \(P = \gamma'_{i'_0+1}\) or \(P = \gamma'_{i'_0}\) in the first case we set

\[
\hat{\gamma} := (\gamma_{i_0+1}, \gamma_{i_0+2}, \ldots, \gamma_k, \gamma_1, \ldots, \gamma_{i_0}, \gamma'_{i'_0+2}, \gamma'_{i'_0+3}, \ldots, \gamma'_k, \gamma'_1, \ldots, \gamma'_{i'_0})
\]

and

\[
\hat{\gamma} := (\gamma_{i_0+1}, \gamma_{i_0+2}, \ldots, \gamma_k, \gamma_1, \ldots, \gamma_{i_0}, \gamma'_{i'_0-1}, \gamma'_{i'_0-2}, \ldots, \gamma'_1, \gamma'_k, \gamma'_{i_0+1})
\]

in the latter case. Loosely speaking \(\hat{\gamma}\) is the path that starts and ends in \(Q\) by first connecting \(Q\) with \(P\) following the cyclic extension of \(\gamma\) and then connecting \(P\) back to \(Q\) along the cyclic extension of \(\gamma'\), adapting the direction if necessary. Since we are dealing with a centred ensemble only those pairs \((\gamma, \gamma')\) with a common edge have a non-zero contribution to the sum in (45) for which \(\hat{\gamma}(\gamma, \gamma') \in \mathcal{P}_0\). From this we conclude that the number of non-zero terms in the sum in (45) is bounded above by

\[
k^2 \sum_{r=1}^{k} n_{r,2k-2}(N) = \mathcal{O}(N w_N^{k-1}) \tag{46}
\]

(see (42) for a definition of \(n_{r,k}(N)\)). The prefactor \(k^2\) takes into account that the values of \(i_0\) and \(i'_0\) are lost in the construction and this is the only reason why the map \((\gamma, \gamma') \mapsto \hat{\gamma}\) is not injective.

Inserting the upper bound (46) into the representation (45) for the variance, we obtain

\[
\mathbb{V} \left( \int x^k \, d\mu_N^\omega(x) \right) = \mathcal{O} \left( \frac{1}{N w_N} \right). \tag{47}
\]
The assumption \( \sum_N (Nb_N)^{-1} < \infty \) of Theorem 10 b) therefore implies the desired summability (44). We mention in passing that for centred ensembles part a) of Theorem 10 follows from (43) and

\[
\lim_{N \to \infty} \mathbb{V} \left( \int x^k \, d\mu_N^\omega(x) \right) = 0
\]

which is also a consequence of (47).

A.2 Statement and proof of Lemma 35

The proof of Lemma 35 relies on an estimate on the expected trace of matrix powers as stated in Lemma 34. The essential difference from the analysis of the previous subsection is that we need to allow the exponent to grow with the matrix dimension \( N \).

Both, Lemma 35 and Lemma 34, are formulated for an auxiliary type of Wigner ensembles that is convenient for the analysis of Wigner band matrices treated in Subsection 4.1. These auxiliary Wigner ensembles arise due to the truncation procedure in the proof of Theorem 28. The truncation depends on the size \( N \) of the matrices. This is why we cannot insist that all entries of the ensemble are identically distributed. Consequently, the auxiliary Wigner ensembles do not fall into the class of Wigner ensembles described in Definitions 1 and 7, albeit they are still well embedded in a more general framework of Wigner ensembles that is well-known in the literature.

As it turns out the results in this subsection do not use the spatial structure of band matrices. They only require a bound on the maximal number of entries in each row that do not vanish identically. In order to bring this to the fore we do not require the band structure for the auxiliary ensembles. This generalisation, however, is not used in the present paper.

The arguments used in the proofs are taken from the monograph [24, Section 2.3] where full matrices are discussed. The adaption to the case of band matrices does not pose additional difficulties. However, we are somewhat more careful in the formulation of Lemma 34 since the inequalities there determine the growth conditions on the bandwidths as discussed in Remark 29. This is also our motivation to improve on inequality (57) for which we present a detailed proof.

**Definition 33** By an Auxiliary Wigner Ensemble AWE we understand a probability measure \( \mathbb{P} \) on families \((X_N)_{N}\) of real symmetric \(N \times N\)-matrices such that sequences \((K_N)_{N}, (n_N)_{N}\) in \([1, \infty)\) exist for which conditions (C1) - (C3) hold for all \( N \in \mathbb{N} \).

(C1) The entries \(X_N(i, j), 1 \leq i \leq j \leq N\), are independent with \(\mathbb{E}(X_N(i, j)) = 0\) and \(\mathbb{E}(X_N(i, j)^2) \leq 1\).
(C2) $\mathbb{P}(|X_N(i, j)| \geq K_N) = 0$ for all $1 \leq i, j \leq N$.

(C3) For all $1 \leq i \leq N$ : $\#\{j \in \{1, \ldots, N\}|\mathbb{P}(X_N(i, j) = 0) < 1\} \leq n_N$.

We call $K_N$ the support bound and $n_N$ the maximal row occupancy of $X_N$.

Observe that the band matrices introduced in Definition 7, strict or periodic, satisfy Condition (C3) with $n_N = w_N$ (cf. Remark 8). We are now ready to state the main results of this subsection.

Lemma 34 Let $(X_N)_N$ be a AWE with support bounds $(K_N)_N$ and maximal row occupancies $(n_N)_N$. Then for all integers $k, N \in \mathbb{N}$ with $2K_N^2k^{14} \leq n_N$ we have

$$|\mathbb{E}(\text{tr}(X_N^k))| \leq 4N(2\sqrt{n_N})^k.$$  

Before proving Lemma 34 we apply it to the operator norm. As mentioned in the introduction to the Appendix the connection is based on the observation that for all even $k \in \mathbb{N}$ we have

$$||X_N||_{op} \leq \text{tr}(X_N^k).$$  

(48)

Lemma 35 Let $(X_N)_N$ be a AWE with support bounds $(K_N)_N$ and maximal row occupancies $(n_N)_N$. Assume furthermore that

$$\sup_{N \in \mathbb{N}} \frac{K_N^2(\log N)^{14+\epsilon}}{n_N} < \infty$$  

for some $\epsilon > 0$. (49)

Then

$$\sum_{N=1}^{\infty} \mathbb{P}(||X_N||_{op} \geq (2 + \delta)\sqrt{n_N}) < \infty$$  

(50)

for any $\delta > 0$ and it follows from the Borel-Cantelli Lemma that

$$\mathbb{P}\left(\limsup_{N \to \infty} \frac{||X_N||_{op}}{\sqrt{n_N}} \leq 2\right) = 1.$$  

Proof. Fix $\delta > 0$. For even $k \in \mathbb{N}$ relation (48) and Markov’s inequality yield the estimate

$$\mathbb{P}(||X_N||_{op} > (2 + \delta)\sqrt{n_N}) \leq \frac{\mathbb{E}(\text{tr}(X_N^k))}{[(2 + \delta)\sqrt{n_N}]^k}.\quad (51)$$

Assumption (49) implies the existence of a number $C > 0$ such that for all $N \in \mathbb{N}$:

$$2^{15}K_N^2(\log N)^{14+\epsilon} \leq Cn_N.$$  

(52)

Since $\log N > 1$ for all $N \geq 3$ we may choose even integers $k_N$ satisfying

$$(\log N)^{1+\frac{\epsilon}{2}} \leq k_N \leq 2(\log N)^{1+\frac{\epsilon}{2}}$$  

(53)
for all \( N \geq 3 \). The upper bound in \((53)\) together with \((52)\) yield the inequality 
\[
2 K_N^2 k_N^{14} \leq C (\log N)^{-\frac{1}{4}} n_N.
\]
Thus there exists \( N_0 \geq 3 \) such that the hypothesis 
\[
2 K_N^2 k_N^{14} \leq n_N
\]
of Lemma 34 holds for all \( N \geq N_0 \). Hence the right hand side of \((51)\) with \( k = k_N \) can be bounded above by
\[
4 N(1 + \delta/2)^{-k_N} \leq 4 N^{1-(\log N)^{-\frac{1}{4}}} \log(1+\delta/2)
\]
for all \( N \geq N_0 \), where we have also used the lower bound in \((53)\). This proves \((50)\).

\[\blacksquare\]

\textbf{Proof of Lemma 34.} Recall the notation introduced at the beginning of the Appendix. We begin by estimating \(|\mathbb{E}(\xi_{a_i}^j)|\) that appears in \((40)\). As we are dealing with centred ensembles we only need to consider the case of \( a_i \geq 2 \). It follows from conditions (C1), (C2) of Definition 33 that
\[
|\mathbb{E}(\xi_{a_i}^j)| \leq K a_i^{-2} n_N \mathbb{E}(\xi_{a_i}^j) \leq K a_i^{-2} n_N.
\]

As
\[
\sum_{i=1}^{\eta(\gamma)} a_i = k
\]
equals the total number of steps of the path \( \gamma \) we obtain for each \( \gamma \in \mathcal{P}_0 \) the bound \(|\mathbb{E}(X_\gamma)| \leq K_N^{k/2-j} n_N\) from \((40)\) (where \( \eta = \eta(\gamma) \)). Formula \((41)\) then proves Proposition 36.

\textbf{Proposition 36} Let \((X_N)_N\) be a AWE with support bounds \((K_N)_N\). Then for positive integers \( k \):
\[
|\mathbb{E}(tr(X_N^k))| \leq \sum_{j=1}^{[k/2]} K_N^{k-2j} M_j,
\]
where \( M_j := \#\{\gamma \in \mathcal{P}_0 \mid \eta(\gamma) = j\} \).

The main effort of proving Lemma 34 is to obtain combinatorial bounds on the numbers \( M_j \).

\textbf{Lemma 37} Let \((X_N)_N\) be a AWE with maximal row occupancies \((n_N)_N\). For all integers \( j, k, N \in \mathbb{N} \), \( 1 \leq j \leq \frac{k}{2} \), and \( n_N \geq 2k^3 \) we have (cf. Proposition 36)
\[
M_j \leq 2 N (2\sqrt{n_N})^k \left( \frac{k^7}{\sqrt{n_N}} \right)^{k-2j}.
\]

Assuming the validity of Lemma 37 we may deduce the claim of Lemma 34. Indeed, since the assumption \( 2 K_N^{2} k_N^{14} \leq n_N \) of Lemma 34 implies \( n_N \geq 2k^3 \) (recall \( K_N \geq 1 \) from Definition 33) we may apply Lemma 37. In addition we also have \( (k^7 K_N / \sqrt{n_N})^2 \leq \frac{1}{2} \) and the sum in the statement of Proposition 36 is dominated by a geometric series, implying the bound of Lemma 34.

We are left to derive the combinatorial estimate of Lemma 37.

\textbf{Proof of Lemma 37.} Fix \( j \in \{1, \ldots, \lfloor \frac{k}{2} \rfloor \} \) and integers \( a_1, \ldots, a_j \geq 2 \) with \( \sum_{i=1}^{j} a_i = k \). The main work goes into proving
\[
M_{a_1, \ldots, a_j} \leq 2 N^2 n_N^2 k^{6(k-2j)}, \quad \text{with} \quad (54)
\]
\[ M_{a_1, \ldots, a_j} := \#\{ \gamma \in P_0 \mid \eta(\gamma) = j \text{ and } a_i(\gamma) = a_i \text{ for all } 1 \leq i \leq j \}. \]

As the bound in (54) is independent of the values of \( a_1, \ldots, a_j \) one obtains the statement of Lemma 37 by showing that there are at most \( k^{k-2j} \) choices for \( a_1, \ldots, a_j \) with \( a_i \geq 2 \) and \( \sum_{i=1}^j a_i = k \) is mapped bijectively via

\[
(a_1, \ldots, a_j) \mapsto \left( \sum_{p=1}^q (a_p - 1) \right)_{1 \leq q \leq j-1}
\]

to a selection of \( j - 1 \) different elements out of \( \{1, \ldots, k - j - 1\} \).

For that we have

\[
\binom{k - j - 1}{j - 1} = \binom{k - j - 1}{k - 2j} \leq k^{k-2j}
\]

possibilities.

The proof of (54) requires a somewhat involved enumeration procedure. Let us first introduce some terminology: A path \( \gamma \in P_0 \) consists of \( k \) steps \((s, s+1)\) with \( s = 1, \ldots, k \). For each such \( s \) there exists \( i \in \{1, \ldots, j\} \) such that \( \{\gamma_s, \gamma_{s+1}\} = e_i \). We then say that step \( s \) is taken along edge \( e_i \) and that this step departs from vertex \( \gamma_s \) and arrives at \( \gamma_{s+1} \). We call edges \( e_i \) to be of higher multiplicity iff \( a_i \geq 3 \) and we denote their number by \( l = l(\gamma) \). For the total number of steps along edges of higher multiplicities \( L := \sum_{a_i \geq 3} a_i \) the following relations hold:

\[
k = \sum_{i=1}^j a_i = 2(j - l) + L, \quad \text{and} \quad L \geq 3l.
\]

These imply the useful estimates

\[
l \leq k - 2j, \quad L = k - 2j + 2l \leq 3(k - 2j). \quad (55)
\]

Steps are called opening steps or closing steps iff they are taken along an edge of multiplicity 2 for the first or for the second time respectively. Opening steps are called innovative iff they arrive at a vertex that hasn’t appeared in the path before. We denote by \( m = m(\gamma) \) the number of non-innovative opening steps. Unlike \( l \) and \( L \) the number \( m \) is not determined by \( a_1, \ldots, a_j \) and may take values \( 0 \leq m \leq j - l \).

To derive (54) we proceed as follows: Besides \( a_1, \ldots, a_j \) fix also the integer \( m \). We want to estimate the number of \( \gamma \in P_0 \) with \( \eta(\gamma) = j \), \( a_i(\gamma) = a_i \) for all \( 1 \leq i \leq j \), and \( m(\gamma) = m \). In order to obtain our bounds we divide the set of all such paths \( \gamma \) into different types.

A type determines at which step
1. the edge $e_i$ occurs, for all $i$ with $a_i \geq 3$ (higher multiplicity),

2. a non-innovative opening step occurs,

3. an innovative opening step occurs.

A crude upper bound on the number of different types is given by (see also (55))

$$k^L \cdot k^m \cdot \binom{k}{j-l-m} \leq k^{3(k-2j)+m} \cdot 2^k$$

Below we argue that each type contains at most

$$N \cdot k^{3(k-2j)} \cdot n_N^{j-m} \cdot k^{2m}$$

(56)

different paths, so that

$$M_{a_1,\ldots,a_j} \leq N2^k k^{6(k-2j)} n_N^{j-1} \sum_{m=0}^{j-l} \left( \frac{k^3}{n_N} \right)^m$$

and (54) follows due to the assumption $2k^3 \leq n_N$.

Thus we are left to establish the upper bound (56) on the number of paths of any given type, i.e. on the number of possibilities to choose vertices. Let us proceed inductively along the path so that only the starting point and the vertices of arrival need to be selected for each step. The number of possibilities at every step depends on the kind of step being taken. We distinguish five cases:

Case A: Starting point.

There are $N$ possibilities.

Case B: Steps along edges of higher multiplicity.

If an edge $e_i$, with $a_i \geq 3$ is taken for the first time then there are at most $n_N$ possibilities to choose the vertex of arrival. Otherwise this vertex is already determined uniquely. In total we have at most $n_N$ possibilities from all steps of this kind.

Case C: Non-innovative opening steps.

Since we may only select a vertex of arrival that has already appeared in the path, the number of choices for each such step is crudely bounded by $k$, in total by $k^m$.

Case D: Innovative opening steps.

At each of these steps we have at most $n_N$ possibilities to select the vertex of arrival, in total at most $n_N^{j-l-m}$ choices.

Case E: Closing steps.

Here a subtle difficulty occurs. Note that the type of the path does not determine at which step a particular edge $e_i$ of multiplicity 2 is being closed (unlike
in the case of edges of higher multiplicity). The type prescribes only which steps are closing steps. However, given the choice of vertices up to the closing step we may only perform a step along an open edge, i.e. along an edge of multiplicity 2 which has got a previous opening step but not a previous closing step. In case there is at most one such open edge that contains our vertex of departure then there is at most one possible choice for picking the vertex of arrival. Denote by $f$ the number of instances out of the $j - l$ closing steps for which more than one choice exists for selecting the vertex of arrival. For each such instance that we call a free closing step we use the crude bound $k$ on the number of choices. Moreover, we show below that the number of free closing steps is bounded by $f \leq L + m \leq 3(k - 2j) + m$ (see (55)) so that the total bound reads $k^{3(k-2j)+m}$.

Multiplication of the estimates from all five Cases A-E leads to (56).

The proof of Lemma 37 is thus concluded by showing

$$f(\gamma) \leq L(\gamma) + m(\gamma) \text{ for all } \gamma \in \mathcal{P}_0. \quad (57)$$

In order to derive (57) we count for each vertex $b$ on the path the number $f(b)$ of closing steps that depart from $b$ and for which more than one possibility exists to do so. Denote furthermore by $L(b)$ the total number of steps arriving at $b$ along an edge of higher multiplicity and by $m(b)$ the total number of steps arriving at $b$ along a non-innovative opening step. It suffices to show

$$f(b) \leq L(b) + m(b), \quad (58)$$

because summation over all vertices $b$ that appear in the path $\gamma$ then proves (57).

Our proof of (58) requires to distinguish 8 cases. The reasoning in all these cases is quite similar and we begin by presenting it in the simplest case that the loop $\{b\}$ is not an edge of the path $\gamma \in \mathcal{P}_0$ and that $b$ is not the starting point of $\gamma$. We denote by $\Lambda_s(b, \gamma)$ the number of edges containing $b$ that are open after step $s$ has been completed, i.e. the number of edges of $\gamma$ of multiplicity 2 that contain $b$, that are opened at one of the steps $1, \ldots, s$, and that are closed at one of the steps $s + 1, \ldots, k$. Observe that (58) is trivially satisfied if $f(b) = 0$. Otherwise pick $t_0 \in \{1, \ldots, k\}$ so that step $t_0$ is the last free closing step departing from $b$. We argue below that

$$1 \leq \Lambda_{t_0}(b, \gamma) \leq 2(m(b) + 1) + L(b) - 2f(b) \quad (59)$$

from which (58) follows using $\lfloor \frac{1}{2}(L(b) + 1) \rfloor \leq L(b)$.

The first inequality of (59) is a consequence of our definition that step $t_0$ is a free closing step. The second inequality is based on the observation that $\Lambda_s(b, \gamma)$ may only change its value at steps $s$ that are connected to visits of the path $\gamma$ at the vertex $b$. Since we assumed that the loop $\{b\}$ is not an edge of $\gamma$ each visit consists
of a step of arrival at \(b\) and a subsequent step of departure. For each visit we obtain an upper bound on the change of the value of \(\Lambda_s(b, \gamma)\) by assuming that the step of departure is an opening step. In case the step of arrival is an opening step/ a step along an edge of higher multiplicity/ a closing step this upper bound is given by 2/1/0 respectively. Using the assumption that \(b\) is not the starting point of \(\gamma\), the fact that there are at most \(m(b) + 1\) visits to \(b\) that arrive by an opening step, and the fact that there are at most \(L(b)\) visits to \(b\) that arrive along an edge of higher multiplicity we have derived the second inequality of (59) except for the term \(-2f(b)\). This term is explained by the observation that we have overestimated the change of the value of \(\Lambda_s(b, \gamma)\) by 2 whenever the departure from \(b\) is realized by a closing step. In addition, we know that there must be at least \(f(b)\) such instances up to step \(t_0\).

For a complete proof of (58) we distinguish

**Case 1:** \(b\) is not the starting point of \(\gamma\)

**Case 2:** \(b\) is the starting point of \(\gamma\)

Both cases are divided into four subcases each:

**Case A:** \(\{b\} \notin \mathcal{E}(\gamma)\) or multiplicity of \(\{b\} \geq 3\)

**Case B:** Multiplicity of \(\{b\} = 2\) and closing step of \(\{b\}\) is not free

**Case C:** Closing step of \(\{b\}\) is free but not the last free closing step

**Case D:** Closing step of \(\{b\}\) is the last free closing step

In all four subcases of Case 1 we derive below the estimate (cf. (59))

\[
0 \leq 1 + 2m(b) + L(b) - 2f(b)
\]

(or better) from which we already know that (58) follows. In order to transfer the above reasoning to all cases conveniently we make precise what we mean by a visit of \(\gamma\) at \(b\): It is a collection of consecutive steps \(s\) up to \(s + p, p \geq 1\), with \(\gamma_s \neq b, \gamma_{s+p+1} \neq b\) and \(\gamma_{i} = b\) for all \(s + 1 \leq i \leq s + p\).

**Case 1A:** The reasoning presented above holds also in the case that \(\{b\}\) is an edge of higher multiplicity because \(\{b\}\) cannot be an open edge then.

**Case 1B:**

\[
1 \leq \Lambda_{t_0}(b, \gamma) \leq 1 + 2m(b) + L(b) - 2f(b) .
\]

Indeed, in comparison with (59), the additional first summand 1 of the right hand side accounts for the edge \(\{b\}\) that may be open after step \(t_0\) is completed. The term \(m(b) + 1\) is replaced by \(m(b)\) since the opening step of edge \(\{b\}\) does not initiate a visit at \(b\).

**Case 1C:**

\[
1 \leq \Lambda_{t_0}(b, \gamma) \leq 2m(b) + L(b) - 2(f(b) - 1) .
\]

In comparison with (61) the first summand 1 of the right hand side has vanished, because edge \(\{b\}\) is already closed at step \(t_0\). The term \(f(b)\) must be replaced by \(f(b) - 1\) since the free closing step of edge \(\{b\}\) does not end a visit at \(b\).
**Case 1D:** Observe that in this case the visit at \( b \) is not completed after step \( t_0 \) but after step \( t_0 + 1 \). Taking into account whether step \( t_0 + 1 \) is an opening step, a closing step, or a step along an edge of higher multiplicity, we obtain in all three cases

\[
2 \leq \Lambda_{t_0+1}(b, \gamma) \leq 2m(b) + L(b) - 2(f(b) - 1). \tag{63}
\]

For each of the subcases of Case 2 we may derive an inequality that improves on the estimate for the corresponding subcase of Case 1 by 1, i.e. we have in the case that \( b \) is the starting point of \( \gamma \) always (cf. (60))

\[
0 \leq 2m(b) + L(b) - 2f(b) \tag{64}
\]

which implies again (58). The reason for this is the same in all four subcases. On the one hand one must increase the right hand side of the inequalities (60) - (63) by 1 to account for the possibility that the starting point \( b \) is left for the first time by an opening step (recall that the first part of \( \gamma \) staying at \( b \) is not considered a visit because the step of arrival at \( b \) is missing). On the other hand Case 2 does not allow for the possibility to start a visit at \( b \) by an innovative opening step which reduces the right hand side of the inequalities (60) - (63) by \( 2 \cdot 1 = 2. \)

**References**

[1] O. Ajanki, L. Erdős, T. Krüger: Universality for general Wigner-type matrices, *Probab. Theory Relat. Fields* **169**, 667–727 (2017).

[2] J. Alt, L. Erdős, D. Schröder: Correlated Random Matrices: Band Rigidity and Edge Universality, arXiv: 1804.07744v3.

[3] D. Aldous: *Exchangeability and related topics*, pp. 1-198 in: Lecture Notes in Mathematics 117, Springer (1985).

[4] L. Arnold: On the Asymptotic Distribution of the Eigenvalues of Random Matrices, *J. Math. Anal. Appl.* **20**, 262–268 (1967).

[5] L. Arnold: On Wigner’s semicircle law for the eigenvalues of random matrices, *Z. Wahrscheinlichkeitstheorie verw. Geb.* **19**, 191–198 (1971).

[6] Z. Bai, Y. Yin: Necessary and sufficient conditions for almost sure convergence of the largest eigenvalue of a Wigner matrix, *Ann. Prob.* **16**, 1729–1741, (1988).

[7] L. Bogachev, S. Molchanov, L. Pastur: On the density of states of random band matrices *Math. Notes* **50**, 1232–1242 (1992).
[8] R. Catalano: On weighted random band-matrices with dependencies, PhD thesis, FernUniversität Hagen, (2016).

[9] L. Erdős, P. Mühlbacher: Bounds on the norm of Wigner-type random matrices, arXiv: 1802.05175v1.

[10] 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).

[11] B. de Finetti: La prevision: ses lois logiques, ses sources subjectives, Annales de l’Institut Henri Poincare, 7, 1–68 (1937).

[12] M. Fleermann: The almost sure semicircle law for random band matrices with dependent entries, arXiv:1711.10196.

[13] Z. Füredi, J. Komlós: The eigenvalues of random symmetric matrices, *Combinatorica* 1 no. 3, 233–241 (1981).

[14] U. Grenander: *Probabilities on algebraic structures*, Wiley (1968).

[15] E. Hewitt, L. J. Savage: Symmetric measures on Cartesian products; Trans. Amer. Math. Soc. 80, 470–501 (1955).

[16] W. Hochstättler, W. Kirsch, S. Warzel: Semicircle law for a matrix ensemble with dependent entries, *J. Theoret. Probab.* 29, 1047–1068 (2016).

[17] W. Kirsch: *Moments in Probability*, book in preparation, to appear at DeGruyter.

[18] W. Kirsch, T. Kriecherbauer: Sixty years of moments for random matrices, F. Gesztesy et al (Edts.), *Non-Linear Partial Differential Equations, Mathematical Physics, and Stochastic Analysis. The Helge Holden Anniversary Volume*, EMS Congress Reports, 349–379 (2018).

[19] W. Kirsch, T. Kriecherbauer: Semicircle Law for Generalized Curie-Weiss Matrix Ensembles at Subcritical Temperature, to appear in: *J. Theoret. Probab.*, arXiv:1703.05183.

[20] S. Molchanov, L. Pastur, A. Khorunzhii: Distribution of the eigenvalues of random band matrices in the limit of their infinite order *Theoret. and Math. Phys.* 90, 108–118 (1992).

[21] 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.
[22] L. Pastur, M. Sherbina: *Eigenvalue distribution of large random matrices*, Mathematical Surveys and Monographs 171, AMS (2011).

[23] R. Stanley: *Catalan Numbers*, Cambridge University Press (2015).

[24] T. Tao: *Topics in random matrix theory*, AMS (2012).

[25] E. Wigner: Characteristic vectors of bordered matrices with infinite dimension, *Ann. Math.* 62, 548–564 (1955).

[26] E. Wigner: On the distribution of the roots of certain symmetric matrices, *Ann. Math.* 67, 325–328 (1958).

Werner Kirsch  
werner.kirsch@fernuni-hagen.de

Thomas Kriecherbauer  
thomas.kriecherbauer@uni-bayreuth.de