ALEA, Lat. Am. J. Probab. Math. Stat. 17, 857–876 (2020)
DOI: 10.30757/ALEA.v17-33

High-dimensional sample covariance matrices with Curie–Weiss entries

Michael Fleermann and Johannes Heiny

FernUniversität in Hagen
Fakultät für Mathematik und Informatik
Universitätsstraße 1
58084 Hagen, Germany
E-mail address: michael.fleermann@fernuni-hagen.de
URL: https://www.fernuni-hagen.de/stochastik/team/michael.fleermann.shtml

Ruhr-Universität Bochum
Fakultät für Mathematik
Universitätsstraße 150
44801 Bochum, Germany
E-mail address: johannes.heiny@rub.de
URL: https://www.ruhr-uni-bochum.de/ffm/Lehrstuehle/stochastik/heiny.html

Abstract. We study the limiting spectral distribution of sample covariance matrices XX^T, where X are $p \times n$ random matrices with correlated entries and $p/n \to y \in [0, \infty)$. If $y > 0$, we obtain the Marčenko–Pastur distribution and in the case $y = 0$ the semicircle distribution after appropriate rescaling. The entries we consider are Curie–Weiss spins, which are correlated random signs, where the degree of the correlation is governed by an inverse temperature $\beta > 0$. The model exhibits a phase transition at $\beta = 1$. The correlation between any two entries is of order $O((np)^{-1})$ for $\beta \in (0, 1)$, $O((np)^{-1/2})$ for $\beta = 1$, and for $\beta > 1$ the correlation does not vanish in the limit. In our proofs we use Stieltjes transforms and concentration of random quadratic forms.

1. Introduction and Preliminaries

In many contemporary applications, one is faced with large data sets where both the dimension of the observations and the sample size are large. In quantum mechanics, for example, the energy levels of particles in a large system can

Received by the editors December 12th, 2019; accepted August 13th, 2020.

2010 *Mathematics Subject Classification.* Primary 60B20; Secondary 60F05 60G55.

Key words and phrases. Curie–Weiss, random matrix, Marčenko–Pastur law, semicircle law, high dimension, dependent entries, full correlation.

M.F.'s research was supported by the FernUniversität in Hagen. J.H. was supported by the Deutsche Forschungsgemeinschaft (DFG) via RTG 2131 High-dimensional Phenomena in Probability – Fluctuations and Discontinuity.

be approximated by the eigenvalues of a large random matrix. Estimating the underlying covariance structure of high-dimensional data with the sample covariance matrix can be misleading Bai and Silverstein (2010); El Karoui (2009). Even in the case of independent covariates, it is well-known that the sample covariance matrix poorly estimates the population covariance matrix. The fluctuations of the off-diagonal entries of the sample covariance matrix aggregate, creating an estimation bias which was quantified in 1967 by the famous Marčenko–Pastur theorem Marčenko and Pastur (1967). Ever since, the classical setting of well-behaved i.i.d. ensembles was extended to investigate settings more aligned with reality. In many situations, it is reasonable to assume that entries in data sets are dependent. The dependence might span between different observations, but also between covariates of individual observations. In random matrix theory, one often considers models exhibiting linear dependence between the entries. Works that consider non-linear dependencies are sparse. The paper Bai and Zhou (2008), for example, incorporates non-linear dependence within the columns of the data matrix, but assumes these columns to be independent. In this paper, we consider a data matrix filled with Curie–Weiss spins. This model exhibits nonlinear dependence between all entries. For technical reasons, settings with correlated entries are harder to analyze, since many proof techniques break down in presence of correlations.

Another way to deviate from the classical setting is to assume that data might stem from heavy-tailed distributions. The theory for the eigenvalues and eigenvectors of the sample covariance matrices stemming from heavy-tailed time series with infinite fourth moment is quite different from the classical Marčenko–Pastur theory which applies in the light-tailed case. For detailed discussions about classical random matrix theory, we refer to the monographs Bai and Silverstein (2010); Yao et al. (2015), while the developments in the heavy-tailed case can be found in Davis et al. (2016b,a); Heiny and Mikosch (2017); Auffinger et al. (2009); Heiny and Mikosch (2018, 2019); Basrak et al. (2020) and the references therein.

The Marčenko–Pastur law gives insight into the spectrum of large dimensional sample covariance matrices. Assume we have n observations $x_1, \dots, x_n$, each with p real-valued covariates, where $n, p \in \mathbb{N}$, so that $x_i = (x_i(1), \dots, x_i(p))^T$ for all $i \in \{1, \dots, n\}$. Define the $p \times n$ data matrix $X_n := (x_1, x_2, \dots, x_n)$, that is, X_n has columns x_i. The (centered) sample covariance matrix is then defined by

$$\tilde{V}_n := \frac{1}{n-1} \sum_{k=1}^{n} (x_k - \bar{x})(x_k - \bar{x})^T,$$

which is of dimension $p \times p$. Here, the vector $\bar{x}$ denotes the arithmetic mean of the vectors x_k. Assuming that the data stems from n i.i.d. realizations of an $\mathbb{R}^p$-valued random vector x with $\mathcal{L}_2$-entries, $\tilde{V}_n$ is an unbiased estimator for its covariance matrix $\mathrm{var}(x)$.

The sample covariance matrix is of crucial importance in multivariate statistics, for instance in principal component analysis, canonical correlation analysis, multivariate regression, factor analysis, hypothesis testing and discriminant analysis. Many test statistics are based on the eigenvalues of the sample covariance matrix. Examples include independence tests (Bodnar et al., 2019) and likelihood ratio tests. For the latter it is essential that the log-determinant of $\tilde{V}_n$ can be written as $\log(\lambda_1) + \cdots + \log(\lambda_p)$, where (λ_i) are the eigenvalues of $\tilde{V}_n$.

When analyzing the limiting spectral distribution (LSD) of the eigenvalues, it suffices to consider the (non-centered) sample covariance matrix

$$V_n := \frac{1}{n}\sum_{k=1}^{n} x_k x_k^T = \frac{1}{n} X_n X_n^T, \tag{1.1}$$

since $\bar{x}\bar{x}^T$ is of rank 1, see Theorem A.44 in Bai and Silverstein (2010). From now on we will refer to V_n as the sample covariance matrix. Our object of interest in this paper will be the limit of the empirical spectral distributions (ESD) F_{V_n} defined as

$$F_{V_n}(x) = \frac{1}{p}\sum_{i=1}^{p} \mathbb{1}_{\{\lambda_i(V_n)\le x\}}, \qquad x \in \mathbb{R},$$

where $\lambda_1(V_n) \ge \cdots \ge \lambda_p(V_n)$ are the ordered eigenvalues of V_n. If such a limit exists in the sense of weak convergence almost surely, we call it the limiting spectral distribution of V_n.

Also, we will assume that the number of covariates p and the sample size n are large and tend to infinity together. In this paper, the sample size n is a function of the dimension p (cf. Remark 2.2) and the dimension increases at most proportionally to the sample size. To be precise, we assume

$$n = n_p \to \infty \quad \text{and} \quad \frac{p}{n_p} \to y \in [0,\infty), \quad \text{as } p \to \infty. \tag{1.2}$$

The constant y controls the growth of the dimension relative to the sample size. Most of the random matrix literature focuses exclusively on the case $y > 0$, while the case $y = 0$ plays only a minor role. In many fields, however, the wider range of possible growth rates arising in the $y = 0$ regime is desirable. The framework in this paper unifies these two lines of research.

1.1. *Background.* Before presenting our model, we provide some background. Assume that the entries of X_n are i.i.d. with unit variance and zero mean. Then if $p/n \to y \in (0,\infty)$ the limiting spectral distribution of (V_n) is the so-called Marčenko–Pastur (MP) distribution μ^y. The MP distribution with ratio index $y \in (0,\infty)$ is the probability measure μ^y on $(\mathbb{R},\mathcal{B})$ given by

$$\mu^y = \frac{1}{2\pi xy}\sqrt{(b-x)(x-a)}\,\mathbb{1}_{(a,b)}(x)\lambda(\mathrm{d}x) + \left(1-\frac{1}{y}\right)\delta_0 \mathbb{1}_{y>1},$$

where $a = (1-\sqrt{y})^2$, $b = (1+\sqrt{y})^2$ and λ denotes the Lebesgue measure on $(\mathbb{R},\mathcal{B})$ and δ_0 denotes the Dirac measure in 0.

It is well-known that measures on $\mathbb{R}$ are uniquely characterized by their Stieltjes transforms Yao et al. (2015). The Stieltjes transform of μ^y is given for $z \in \mathbb{C}_+ = \{c \in \mathbb{C} : \mathrm{Im}(c) > 0\}$ by

$$S_{\mu^y}(z) := \int_{\mathbb{R}} \frac{1}{x-z}\mu^y(dx) = \frac{1-y-z+\sqrt{(1-y-z)^2-4yz}}{2yz},$$

where throughout this paper, if $z \in \mathbb{R}_+$, $\sqrt{z}$ denotes the positive square root, while if $z \in \mathbb{C}\backslash\mathbb{R}_+$, then $\sqrt{z}$ denotes the complex square root with positive imaginary part; see for example Bai and Silverstein (2010). If $p/n \to \infty$, we observe δ_0 as LSD of V_n, as there are at most $\min(p,n)$ positive eigenvalues of V_n.

In the case $p/n \to 0$, the limiting spectral distribution of V_n is the Dirac measure at 1. After centering V_n by the identity matrix I and a subsequent appropriate rescaling, one can obtain a non-degenerate limiting spectral distribution. In Bai and Yin (1988) it is proved under the additional assumption $\mathbb{E}[X_n(1,1)^4] < \infty$ that the empirical spectral distribution of the matrices $\sqrt{n/p}\,(V_n - I)$ converges to the semicircle law G with Lebesgue density

$$g(x) = \tfrac{1}{2\pi}\sqrt{4-x^2}\mathbb{1}_{[-2,2]}(x)\,, \qquad x \in \mathbb{R},$$

and Stieltjes transform

$$s_G(z) = \frac{-z + \sqrt{z^2-4}}{2}\,, \qquad z \in \mathbb{C}_+. \tag{1.3}$$

The i.i.d. assumption on the entries of the data matrix X_n can be relaxed to linear dependence of the form $\Sigma_n^{1/2} X_n$ for symmetric positive definite deterministic matrices Σ_n with uniformly bounded spectral or operator norm $\|\Sigma_n\| := \sqrt{\lambda_1(\Sigma_n \Sigma_n^T)}$. For $p/n \to y > 0$, the Stieltjes transform of the LSD of $n^{-1}\Sigma_n^{1/2} X_n X_n^T \Sigma_n^{1/2}$ can then be characterized via the LSD of $\mathbb{E}[V_n] = \Sigma_n$; see Bai and Silverstein (2010) for details. The same holds in the case $p/n \to 0$ for the LSD of $\sqrt{n/p}\,(n^{-1}\Sigma_n^{1/2} X_n X_n^T \Sigma_n^{1/2} - \Sigma_n)$ as proved in Pan and Gao (2012) and Wang and Paul (2014).

It is important to note that the linear dependence between the entries of X_n was a crucial assumption for the above results. For nonlinear dependencies the situation becomes more delicate as the following examples will show. We present two examples of random matrices Y_n with dependent entries and $\mathbb{E}[n^{-1} Y_n Y_n^T] = I$ for which the LSD of $n^{-1} Y_n Y_n^T$ is not the Marčenko–Pastur distribution μ^y.

Example 1.1. Assume that the entries of X_n are i.i.d. continuous random variables and let $p/n \to y > 0$. Kendall's Tau is a U-statistic which measures the association of random variables. For higher dimensional observations, such as the columns $x_i = (x_i(1), \dots, x_i(p))^T$ of the data matrix X, the empirical Kendall's Tau matrix is defined as

$$\tau_n = \tfrac{2}{n(n-1)} \sum_{1 \le s < t \le n} \operatorname{sign}(x_s - x_t)(\operatorname{sign}(x_s - x_t))^T\,,$$

where sign of a vector is taken coordinatewise. In particular, one sees that $\tau_n(i,i) = 1$. Since X_n has i.i.d. continuous entries, we have $\mathbb{E}[\tau_n] = I$. Bandeira et al. (2017) proved that the empirical spectral distribution F_{τ_n} of τ_n converges, namely

$$F_{\tau_n} \xrightarrow{\mathbb{P}} \tfrac{2}{3}\,\xi + \tfrac{1}{3}\,, \qquad p \to \infty\,,$$

where the random variable ξ has the Marčenko–Pastur distribution μ^y. In Theorem 2.1 we will observe a similar scaling phenomenon. The exact formula for Y_n such that $n^{-1} Y_n Y_n^T = \tau_n$ can be found in Bao (2019).

Example 1.2. Assume that the entries of X_n are i.i.d. symmetric random variables with tails $\mathbb{P}(|X_n(1,1)| > t) = t^{-\alpha}\ell(t)$, where $\alpha \in (0,2)$ and ℓ is a slowly varying function at infinity. Under the regime $p/n \to y > 0$, set $Y_n = (\operatorname{diag}(X_n X_n^T))^{-1/2} X_n$ and consider the sample correlation matrices $R_n := Y_n Y_n^T$. It was shown in Heiny and Mikosch (2018) that $\mathbb{E}[R_n] = I$. In other words, the entries of Y_n are uncorrelated with variance $1/n$. Very recently, Heiny and Yao (2020) characterized the LSD of R_n, which they called α-heavy Marčenko–Pastur distribution $H_{\alpha,y}$. The moments of the α-heavy Marčenko–Pastur distribution are a sum of the moments of μ^y and a nonnegative heavy-tailed part which depends on the tail index α and y. They also showed that $\lim_{\alpha\to 0^+} H_{\alpha,y}$ is a modified Poisson distribution.

1.2. *Our model.* We will consider a data matrix X_n with correlated entries. To this end, we introduce the Curie–Weiss model which is an exactly solvable model of ferromagnetism. "Because of its simplicity and because of the correctness of at least of some of its predictions, the Curie–Weiss model occupies an important place in the statistical mechanics literature and its application to information theory Kochmański et al. (2013)." The first time that random matrices with Curie–Weiss spins were analyzed was in Friesen and Löwe (2013), with subsequent improvements in Hochstättler et al. (2016); Kirsch and Kriecherbauer (2018); Fleermann et al. (2021); Fleermann (2019b), where the last two publications are based on Fleermann (2019a). All of these texts were concerned with Wigner type matrices and convergence to the semicircle distribution.

Definition 1.3. Let $n \in \mathbb{N}$ be arbitrary and $Y_1, \ldots, Y_n$ be random variables defined on some probability space $(\Omega, \mathcal{A}, \mathbb{P})$. Let $\beta > 0$, then we say that $Y_1, \ldots, Y_n$ are Curie-Weiss(β,n) distributed, if for all $y_1, \ldots, y_n \in \{-1, 1\}$ we have that

$$\mathbb{P}(Y_1 = y_1, \ldots, Y_n = y_n) = \frac{1}{Z_{\beta,n}} \cdot \mathrm{e}^{\frac{\beta}{2n}\left(\sum_{i=1}^n y_i\right)^2},$$

where $Z_{\beta,n} = \sum_{y_1,\ldots,y_n\in\{-1,1\}} \mathrm{e}^{\frac{\beta}{2n}\left(\sum_{i=1}^n y_i\right)^2}$ is a normalization constant. The parameter β is called *inverse temperature.*

Note that in above definition, $(Y_1, \ldots, Y_n)$ is an exchangeable random vector, since the probability of any spin configuration $(y_1, \ldots, y_n)$ only depends on the sum of the spins. The Curie-Weiss(β, n) distribution is used to model the behavior of n ferromagnetic particles (spins) at the inverse temperature β. At low temperatures (if β is large), all magnetic spins are likely to have the same alignment, resembling a strong magnetic effect. On the contrary, at high temperatures (if β is small), spins can act almost independently, resembling a weak magnetic effect. The model exhibits a phase transition at $\beta = 1$, meaning that the behavior of the distribution varies significantly in the realms $\beta \in (0, 1)$, $\beta = 1$ and $\beta > 1$. To exemplify a manifestation of this phase transition, we formulate the following result; see Theorem 5.17 in Kirsch (2015).

Lemma 1.4. *Fix $l \in \mathbb{N}$ and let for all $n \geq l$, $(Y_1^{(n)}, \ldots, Y_l^{(n)})$ be part of a Curie-Weiss(β, n) distributed random vector. If l is even, the following statements hold:*

i) If $\beta < 1$, then for some constant $c(\beta, l) > 0$, $\mathbb{E}Y_1^{(n)} \cdots Y_l^{(n)} \sim c(\beta, l)n^{-l/2}$ as $n \to \infty$.

ii) If $\beta = 1$, then for some constant $c(l) > 0$, $\mathbb{E}Y_1^{(n)} \cdots Y_l^{(n)} \sim c(l)n^{-l/4}$ as $n \to \infty$.

iii) If $\beta > 1$, then $\mathbb{E}Y_1^{(n)} \cdots Y_l^{(n)} \sim m^l$ as $n \to \infty$, where $m \in (0, 1)$ is the unique positive number such that $\tanh(\beta m) = m$.

If l is odd, then for all $\beta > 0$ one has $\mathbb{E}Y_1^{(n)} \cdots Y_l^{(n)} = 0$.

Note that in the setting of Lemma 1.4, the correlation $\mathbb{E}Y_1^{(n)}Y_2^{(n)}$ is of a different order for the three regions of β. If $\beta < 1$, the correlation $\mathbb{E}Y_1^{(n)}Y_2^{(n)}$ decays at a rate of n^{-1}. For the critical temperature $\beta = 1$ the decay rate is $n^{-1/2}$, whereas if $\beta > 1$, the correlation $\mathbb{E}Y_1^{(n)}Y_2^{(n)}$ converges to m^2 and hence does not vanish as $n \to \infty$. In our main result Theorem 2.1, we will see that for $\beta > 1$ a different normalization of the sample covariance matrix is required to account for the correlation at level m^2.

Objective and structure of this paper. The aim of this paper is to characterize the LSD of the sample covariance matrices $V_n = n^{-1}X_nX_n^T$, where X_n follows a Curie–Weiss distribution. At the critical temperature $\beta = 1$ a phase transition occurs. In Section 2, we see that the LSD is a possibly rescaled Marčenko–Pastur or semicircle distribution. Section 3 contains some useful lemmas and the proof of our main result.

Notation. For simplicity of notation, we define for all $n \in \mathbb{N}$: $[n] := \{1, \ldots, n\}$. Further, whenever there is no ambiguity about the dimension we denote the identity matrix by I. The resolvent $(M - zI)^{-1}$ of a Hermitian matrix M will be denoted by $(M - z)^{-1}$.

2. Main result

Our main result characterizes the limiting spectral distributions of sample covariance matrices with Curie–Weiss entries with parameter $\beta > 0$ in the regimes $p/n \to y > 0$ and $p/n \to 0$.

Theorem 2.1. *Assume (1.2) and that the entries of the $p \times n$ matrix X_n are Curie-Weiss(β, np) distributed with $\beta > 0$, where we assume that $(X_n(i,j))_{i\in[p],j\in[n],n\in\mathbb{N}}$ are defined on a common probability space. Denote by F_n the ESD of $V_n := n^{-1}X_nX_n^T$.*

(i) Assume $\beta \in (0, 1]$. If $p/n \to y \in (0, \infty)$, then $(F_n)_n$ converges weakly almost surely to the Marčenko–Pastur distribution μ^y, as $p \to \infty$. If $p/n \to 0$, then the ESDs of $\sqrt{\frac{n}{p}}(V_n - I)$ converge weakly almost surely to the semicircle distribution G, as $p \to \infty$.

(ii) Assume $\beta \in (1, \infty)$ and let m be the unique number in $(0, 1)$ satisfying $\tanh(m\beta) = m$. If $p/n \to y \in (0, \infty)$, then the ESDs of $(1 - m^2)^{-1}V_n$ converge weakly almost surely to the Marčenko–Pastur distribution μ^y, as $p \to \infty$. If $p/n \to 0$, then the ESDs of $\sqrt{\frac{n}{p}}\left(\frac{1}{1-m^2}V_n - I\right)$ converge weakly almost surely to the semicircle distribution G, as $p \to \infty$.

By Lemma 1.4, the correlations between the entries of X_n increase with the value β. Theorem 2.1 shows that for $\beta \leq 1$ the correlation is still weak enough to not affect the LSD, in the sense that we obtain the same LSD as for a sample of i.i.d. random variables. For $\beta > 1$ the asymptotic behavior of the correlations changes drastically. Consequently, a different normalization of the sample covariance matrix is required to account for the correlation at constant level m^2.

Remark 2.2. The convergence in Theorem 2.1 is for $p \to \infty$, which is standard in the $p/n \to 0$ literature; see for example Bai and Yin (1988); Pan and Gao (2012); Wang and Paul (2014). If there exists a $\delta > 0$ such that $n^\delta/p \to 0$, the convergence also holds for $n \to \infty$; compare also with Corollary 2 in El Karoui (2009). Indeed, $n^\delta/p \to 0$ for some $\delta > 0$ is equivalent to p_n^{-a} being summable over n for some large

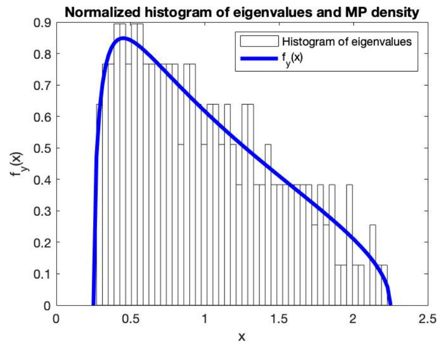

FIGURE 2.1. Simulation for $(p, n) = (200, 800)$ and $\beta = 0.5$. In blue: Density $f_{1/4}$. Histogram: Eigenvalues of $n^{-1}X_nX_n^T$.

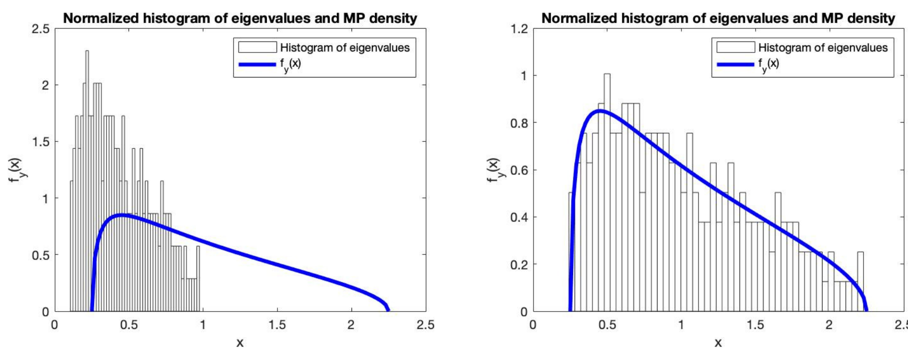

FIGURE 2.2. Simulation for $(p, n) = (200, 800)$ and $\beta = 1.29727$. In blue: Density $f_{1/4}$. Left histogram: Eigenvalues of $n^{-1}X_nX_n^T$. Right histogram: Eigenvalues of $(n(1-m^2))^{-1}X_nX_n^T$, where $m = 3/4$, so that $\tanh(\beta m) = m$.

$a > 0$, which is required for the Borel-Cantelli argument in the proof of Theorem 2.1. Our formulation with $p \to \infty$ is slightly more flexible because it also allows choices such as $p = \log n$.

In Figure 1 and Figure 2, we can see a simulation output where a 200×800 random matrix with Curie–Weiss entries was simulated, using the Metropolis algorithm with $16 \cdot 10^6$ steps. We compare the histogram of the eigenvalues with the Marčenko–Pastur density $f_{p/n}$,

$$f_y(x) = \frac{1}{2\pi xy}\sqrt{((1+\sqrt{y})^2 - x)(x-(1-\sqrt{y})^2)}\mathbb{1}_{((1-\sqrt{y})^2,(1+\sqrt{y})^2)}(x)\,,$$

where $x \in \mathbb{R}$ and $y \in (0, 1]$. While in Figure 2.1, the ensemble was simulated for $\beta = 0.5$, in Figure 2.2 we used $\beta = 1.29727$, so that $\tanh(\beta m) = m$ holds for $m = 3/4$. The largest eigenvalue of $n^{-1}X_nX_n^T$ resp. $(n(1-m^2))^{-1}X_nX_n^T$ was 112.51 resp. 257.16 and was excluded from the histogram in Figure 2.2. These large values in the case $\beta > 1$ are well explained by the fact that the largest eigenvalue of V_n is of order p as our next result shows.

Proposition 2.3 (Largest eigenvalue)**.** *Assume $p/n \to y \in (0,\infty)$ and let the entries of the $p \times n$ matrix X_n be Curie-Weiss(β, np) distributed with $\beta > 1$, where we assume that $(X_n(i,j))_{i\in[p],j\in[n],n\in\mathbb{N}}$ are defined on a common probability space. Then, as $p \to \infty$, the largest eigenvalue of V_n/p converges in probability to m^2, where m is the unique number in $(0,1)$ satisfying* $\tanh(m\beta) = m$.

Proof: Since m^2 is a constant, it suffices to show that $\lambda_1(V_n/p) \to m^2$ in distribution, and for this it suffices to show that the moments $\mathbb{E}\big[\lambda_1^k(V_n/p)\big]$ converge to m^{2k}, $k \geq 1$. Assume $\beta > 1$ and recall from Lemma 1.4 that $\mathbb{E}X_n(1,1)\cdots X_n(1,l) \sim m^l$ as $p \to \infty$. Set $w = p^{-1/2}(1,\dots,1)^T \in \mathbb{R}^p$ and consider the following lower bound on the second moment of $\lambda_1(V_n/p)$:

$$\mathbb{E}\big[\lambda_1^2(V_n/p)\big] \geq p^{-2}\mathbb{E}\big[\|V_n w\|_2^2\big] = \frac{1}{n^2p^3}\sum_{i=1}^{p}\mathbb{E}\Big[\Big(\sum_{j=1}^{p}\sum_{t=1}^{n}X_n(i,t)X_n(j,t)\Big)^2\Big] \sim m^4\,,$$

as $p \to \infty$, where $\|\cdot\|_2$ is the ℓ_2 norm. By Jensen's inequality, for $k \geq 1$ it follows that

$$\liminf_{p\to\infty}\mathbb{E}\big[\lambda_1^{2k}(V_n/p)\big] \geq \liminf_{p\to\infty}\mathbb{E}\big[\lambda_1^2(V_n/p)\big]^k \geq m^{4k}. \tag{2.1}$$

An upper bound is given by

$$\begin{aligned}&\mathbb{E}\big[\lambda_1^k(V_n/p)\big] \leq p^{-k}\mathbb{E}\big[\operatorname{tr}(V_n^k)\big]\\ &= \frac{1}{n^kp^k}\sum_{i_1,\dots,i_k=1}^{p}\sum_{t_1,\dots,t_k=1}^{n}\mathbb{E}[X_n(i_1,t_k)X_n(i_1,t_1)X_n(i_2,t_1)\cdots X_n(i_k,t_{k-1})X_n(i_k,t_k)]\\ &\to m^{2k}\,, \qquad p \to \infty\,.\end{aligned} \tag{2.2}$$

A combination of (2.1) and (2.2) implies that $\mathbb{E}\big[\lambda_1^{2k}(V_n/p)\big] \to m^{4k}$ for $k \geq 1$. This shows $\lambda_1^2(V_n/p) \to m^4$ weakly, so also in probability. By continuous mapping, $\lambda_1(V_n/p) = \sqrt{\lambda_1^2(V_n/p)} \to \sqrt{m^4} = m^2$ in probability. □

Remark 2.4. For $\beta \in (0,1)$ and $p/n \to y > 0$, a direct consequence of Theorem 2.1 is

$$\liminf_{p\to\infty}\lambda_1(V_n) \geq (1+\sqrt{y})^2 \quad \text{a.s.}$$

Using Lemma 1.4 and the arguments in Geman (1980), it can be shown that $(1+\sqrt{y})^2$ is the almost sure limit of $\lambda_1(V_n)$. To the best of our knowledge, the behavior of the largest eigenvalue at critical temperature $\beta = 1$ is still open and might yield an interesting phase transition.

In the proof of Theorem 2.1, we use Stieltjes transforms and concentration of random quadratic forms. Regarding the latter, we adapt techniques originally developed in Fleermann (2019b) and Fleermann et al. (2020) to our situation. An important tool is the fact that Curie-Weiss(β,n) spins $(Y_1,\dots,Y_n)$ are *conditionally i.i.d.* That is, without loss of generality we can assume that they are defined on the same probability space as a Lebesgue-continuous mixing variable M_n^β with support $[-1,1]$, such that conditioned on $M_n^\beta = t \in (-1,1)$, $(Y_1,\dots,Y_n)$ are i.i.d. P_t-distributed, where P_t is the probability measure on $\{\pm 1\}$ with

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

Next, we collect some properties of the mixing variable M_n^β in the following lemma which is taken from Kirsch (2015); see Theorem 5.6, Remark 5.7, Proposition 5.9 and Theorem 5.17 therein.

Lemma 2.5. *If $Y = (Y_1, \dots, Y_n)$ are Curie-Weiss(β,n) distributed for some $\beta > 0$ and $n \in \mathbb{N}$, then w.l.o.g. there exists a random variable M_n^β supported on $[-1,1]$ with the following properties:*

(1) The distribution of M_n^β has Lebesgue-density f_n^β,

$$f_n^\beta(t) := \frac{1}{\int_{(-1,1)} \frac{e^{-\frac{n}{2}F_\beta(s)}}{1-s^2}\lambda(ds)} \frac{e^{-\frac{n}{2}F_\beta(t)}}{1-t^2}\mathbb{1}_{(-1,1)}(t), \qquad t \in (-1,1),$$

where for all $s \in (-1,1)$ we define

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

(2) $\mathbb{P}^{M_n^\beta}$-almost surely, $\mathbb{P}^{Y|M_n^\beta=t} = \otimes_{i\in[n]}\mathbb{P}^{Y_i|M_n^\beta=t} = \otimes_{i\in[n]}P_t$. In words, conditionally on M_n^β the $Y_1, \dots, Y_n$ are i.i.d. P_t-distributed random variables.

(3) If $\beta < 1$, the mixing variable M_n^β satisfies the following moment decay:

$$\forall\ a \in 2\mathbb{N} : \int_{[-1,+1]} t^a \mathbb{P}^{M_n^\beta}(dt) \le \frac{K_{\beta,a}}{n^{\frac{a}{2}}}.$$

(4) If $\beta = 1$, the mixing variable M_n^β satisfies the following moment decay:

$$\forall\ a \in 2\mathbb{N} : \int_{[-1,+1]} t^a \mathbb{P}^{M_n^\beta}(dt) \le \frac{K_{\beta,a}}{n^{\frac{a}{4}}},$$

where $K_{\beta,a} \in \mathbb{R}_+$ are constants that depend on β and a only.

In the case $\beta > 1$ we will work with suitably restandardized Curie–Weiss spins in order to use the following lemma which can be found in Fleermann et al. (2020).

Lemma 2.6. *Let $(Y_1, \dots, Y_n)$ be Curie-Weiss(β, n) distributed with $\beta > 1$ and mixing variable M_n^β. Denote by $m \in (0,1)$ the unique positive number satisfying $\tanh(m\beta) = m$. For $i \in \{1, \dots, n\}$ define*

$$Z_i := \frac{1}{\sqrt{1-m^2}}\left(Y_i - m\mathbb{1}_{M_n^\beta > 0} + m\mathbb{1}_{M_n^\beta < 0}\right).$$

Then $(Y_1, \dots, Y_n)$ are conditionally i.i.d. given M_n^β and the following statements hold:

(1) Almost surely, $(Z_1, \dots, Z_n)$ takes values in $\{\frac{\pm 1+m}{\sqrt{1-m^2}}\}^n \cup \{\frac{\pm 1-m}{\sqrt{1-m^2}}\}^n$.

(2) For each $i \in \{1, \dots, n\}$,

$$\mathbb{E}(Z_i|M_n^\beta = t) = \zeta(t) := \begin{cases} \frac{1}{\sqrt{1-m^2}}(t-m), & t > 0, \\ \frac{1}{\sqrt{1-m^2}}(t+m), & t < 0, \end{cases}$$

$$\mathbb{E}(1 - Z_i^2|M_n^\beta = t) = \psi(t) := \begin{cases} \frac{2m}{1-m^2}(t-m), & t > 0, \\ \frac{2m}{1-m^2}(t+m), & t < 0. \end{cases}$$

(3) We obtain the following bounds on the moments of $\zeta(M_n^\beta)$ and $\psi(M_n^\beta)$:

$$\forall\ a \in 2\mathbb{N} : \int_{[-1,+1]} |\zeta(t)|^a \mathbb{P}^{M_n^\beta}(dt) \le \frac{K_{\beta,a}}{n^{\frac{a}{2}}},$$

$$\forall\ a \in 2\mathbb{N} : \int_{[-1,+1]} |\psi(t)|^a \mathbb{P}^{M_n^\beta}(dt) \leq \frac{K_{\beta,a}}{n^{\frac{a}{2}}}.$$

Here, the constants $K_{\beta,a} > 0$ depend only on β and a.

3. Proof of Theorem 2.1

We will prove the cases $\beta \leq 1$ and $\beta > 1$ separately, but before we begin, we will provide two lemmas which we will use throughout the proof.

Lemma 3.1. *Let $n \in \mathbb{N}$ be arbitrary, $(a_{i,j})_{i,j\in[n]}$ and $(b_i)_{i\in[n]}$ be deterministic complex numbers, $(Y_i)_{i\in[n]}$ be independent and complex-valued random variables with common expectation $t \in \mathbb{C}$. Further, we assume that for all $a \geq 2$ there exists a $\mu_a \in \mathbb{R}_+$ such that $\|Y_i - t\|_a := \mathbb{E}[|Y_i - t|^a]^{1/a} \leq \mu_a$ for all $i \in [n]$. Then we obtain for all $a \geq 2$:*

$$i) \qquad \left\| \sum_{i\in[n]} b_i Y_i \right\|_a \leq \left(A_a\mu_a + \sqrt{n}|t|\right) \sqrt{\sum_{i\in[n]} |b_i|^2}\,,$$

$$\begin{aligned} ii) \left\| \sum_{\substack{i,j\in[n]\\ i\neq j}} a_{i,j} Y_i Y_j \right\|_a &\leq A_a\mu_a|t| \sqrt{\sum_{j\in[n]} \left| \sum_{i\in[n]\setminus\{j\}} a_{i,j} \right|^2} + A_a\mu_a|t| \sqrt{\sum_{i\in[n]} \left| \sum_{j\in[n]\setminus\{i\}} a_{i,j} \right|^2} \\ &\quad + 4A_a^2\mu_a^2 \sqrt{\sum_{\substack{i,j\in[n]\\ i\neq j}} |a_{i,j}|^2} + |t|^2 \left| \sum_{\substack{i,j\in[n]\\ i\neq j}} a_{i,j} \right| \\ &\leq \left(4A_a^2\mu_a^2 + 2A_a\mu_a\sqrt{n}|t| + n|t|^2\right) \sqrt{\sum_{\substack{i,j\in[n]\\ i\neq j}} |a_{i,j}|^2}\,, \end{aligned}$$

where $A_a \in \mathbb{R}_+$ is a positive constant depending only on a.

Proof: This is straightforward refinement of the proof of Theorem 39 in Fleermann (2019b). □

The following lemma allows us to apply Lemma 3.1 to the setting we will encounter in our proof.

Lemma 3.2. *Let X be a $p \times n$ matrix with real-valued entries, $z \in \mathbb{C}_+$. Define*

$$F(X) := X^T \left(\frac{1}{n} X X^T - z\right)^{-1} X. \tag{3.1}$$

Then we obtain the following bounds:

$$i) \quad \sqrt{\sum_{i\neq j}^{n} |F_{ij}(X)|^2} \leq n\sqrt{p}\left(1 + \frac{|z|}{\mathrm{Im}(z)}\right), \qquad ii) \quad |\mathrm{tr}\, F(X)| \leq np\left(1 + \frac{|z|}{\mathrm{Im}(z)}\right).$$

Further, if the absolute values of all entries in X are uniformly bounded by some $b > 0$, it holds:

iii) $$\left|\sum_{i,j\in[n]} F_{ij}(X)\right| \le \frac{b^2 p}{\operatorname{Im}(z)} + \frac{1}{n}\left(1 + \frac{|z|}{\operatorname{Im}(z)}\right),$$

iv) $$\forall\, j \in [n] : \left|\sum_{i\in[n]\setminus\{j\}} F_{ij}(X)\right| \le \frac{b^2 p}{n\operatorname{Im}(z)}.$$

Proof: To prove i), we recall that

a) $\operatorname{Spectrum}(X^T(XX^T - z)^{-1}X) \cup \{0\} = \operatorname{Spectrum}((XX^T - z)^{-1}XX^T) \cup \{0\}$,
b) $(XX^T - z)^{-1}XX^T = I + z(XX^T - z)^{-1}$,

and that $\|\cdot\|_F \le \sqrt{m}\|\cdot\|$ for $m \times m$ matrices, where $\|\cdot\|_F$ denotes the Frobenius norm and $\|\cdot\|$ denotes the operator norm. Therefore,

$$\begin{aligned}
&\sqrt{\sum_{i\neq j}^{n} |F_{ij}(X)|^2} \le \sqrt{\sum_{i,j}^{n}\left|\left[X^T\left(\frac{1}{n}XX^T - z\right)^{-1}X\right](i,j)\right|^2} \\
&= n\left\|\frac{1}{n}X^T\left(\frac{1}{n}XX^T - z\right)^{-1}X\right\|_F = n\left\|\left(\frac{1}{n}XX^T - z\right)^{-1}\left(\frac{1}{n}XX^T\right)\right\|_F \\
&= n\left\|I_p + z\left(\frac{1}{n}XX^T - z\right)^{-1}\right\|_F \le n\sqrt{p}\left\|I_p + z\left(\frac{1}{n}XX^T - z\right)^{-1}\right\| \\
&\le n\sqrt{p}\left(1 + \frac{|z|}{\operatorname{Im}(z)}\right).
\end{aligned}$$

For ii) we calculate

$$\begin{aligned}
|\operatorname{tr} F(X)| &= n\left|\operatorname{tr}\left(\left(\frac{1}{n}XX^T - z\right)^{-1}\left(\frac{1}{n}XX^T\right)\right)\right| \\
&\le np\left\|\left(\frac{1}{n}XX^T - z\right)^{-1}\left(\frac{1}{n}XX^T\right)\right\| \le np\left(1 + \frac{|z|}{\operatorname{Im}(z)}\right),
\end{aligned}$$

where the last step follows as in the proof of i). This shows ii). For iii) let $1_n := (1, \ldots, 1)^T \in \mathbb{R}^n$ and $Y := n^{-1/2}X$. Then we see that

$$\begin{aligned}
n\left|\sum_{i,j\in[n]} F_{ij}(X)\right| &= \left|1_n^T Y^T (YY^T - z)^{-1} Y 1_n\right| \\
&\le \left|1_n^T Y^T\left[(YY^T - z)^{-1} - (YY^T + Y1_n 1_n^T Y^T - z)^{-1}\right] Y 1_n\right| \\
&\quad + \left|1_n^T Y^T (YY^T + Y1_n 1_n^T Y^T - z)^{-1} Y 1_n\right| =: P_1 + P_2 .
\end{aligned}$$

By Silverstein and Bai (1995, Lemma 2.6), one has

$$P_1 \le \frac{\|Y1_n 1_n^T Y^T\|}{\operatorname{Im}(z)} = \frac{|1_n^T Y^T Y 1_n|}{\operatorname{Im}(z)} = \frac{1}{\operatorname{Im}(z)}\left|\frac{1}{n}\sum_{i\in[n]}\sum_{j\in[p]}\sum_{s\in[n]} X_{ij}X_{js}\right| \le \frac{b^2 np}{\operatorname{Im}(z)} .$$

To bound P_2, recall from Yaskov (2016) that for a real, symmetric, positive semi-definite $m \times m$ matrix M; $x \in \mathbb{R}^m$, $z \in \mathbb{C}_+$ the following inequality holds:

$$\left|x^T(M + xx^T - z)^{-1}x\right| \leq 1 + \frac{|z|}{\operatorname{Im}(z)}\,. \tag{3.2}$$

So in particular P_2 is bounded by the right-hand side of (3.2). This yields the bound

$$\left|\sum_{i,j\in[n]} F_{ij}(X)\right| \leq \frac{P_1 + P_2}{n} \leq \frac{b^2 p}{\operatorname{Im}(z)} + \frac{1}{n}\left(1 + \frac{|z|}{\operatorname{Im}(z)}\right).$$

Lastly, to show $iv)$ let Y be defined as before and let $j \in [n]$ be arbitrary. Denote by y the j-th standard basis vector of $\mathbb{R}^n$ and let $x := 1_n - y$. Then, using that Yyx^TY^T has rank one,

$$\begin{aligned}
&n\left|\sum_{i\in[n]\setminus\{j\}} F_{ij}(X)\right| = \left|x^TY^T(YY^T - z)^{-1}Yy\right| = \left|\operatorname{tr}\left[(YY^T - z)^{-1}Yyx^TY^T\right]\right| \\
&\leq \left\|(YY^T - z)^{-1}Yyx^TY^T\right\| \leq \left\|(YY^T - z)^{-1}\right\| \left\|Yyx^TY^T\right\| \\
&\leq \frac{1}{\operatorname{Im}(z)}\left|x^TY^TYy\right| = \frac{1}{\operatorname{Im}(z)}\left|\frac{1}{n}\sum_{i\in[n]\setminus\{j\}}\sum_{s\in[p]} X_{js}X_{is}\right| \leq \frac{b^2 p}{\operatorname{Im}(z)}.
\end{aligned}$$

□

3.1. *The case $\beta \leq 1$.* We will show that the Stieltjes transforms converge. For a real symmetric matrix $M \in \mathbb{R}^{p\times p}$ we denote by S_M the Stieltjes transform of the ESD of M, that is:

$$\forall\, z \in \mathbb{C}_+ : S_M(z) = \frac{1}{p}\operatorname{tr}(M - z)^{-1} = \frac{1}{p}\sum_{i=1}^{p}\frac{1}{\lambda_i(M) - z}.$$

Further, we write $s_n = S_{V_n}$ for the Stieltjes transform of V_n.

Fix a $z \in \mathbb{C}_+$. Our starting point is the following identity, which is easy to verify using that $\lambda_i(V_n - I) = \lambda_i(V_n) - 1$:

$$S_{y_n^{-1/2}(V_n - I)}(z) = y_n^{1/2} s_n(1 + y_n^{1/2} z), \quad \text{where } y_n := \frac{p}{n}. \tag{3.3}$$

For simplicity of notation we write $\eta = \operatorname{Im}(z) > 0$ and $q = q_n = 1 + y_n^{1/2} z$. Note that $\operatorname{Im}(q) = \eta\sqrt{p/n}$. We know by equation (3.3.6) in Bai and Silverstein (2010) that

$$\begin{aligned}
s_n(q) &= \frac{1}{p}\sum_{k\in[p]}\frac{1}{\frac{1}{n}\alpha_k^T\alpha_k - q - \frac{1}{n^2}\alpha_k^T X_n^{(k)}\left(\frac{1}{n}X_n^{(k)}X_n^{(k)T} - q\right)^{-1}X_n^{(k)}\alpha_k} \\
&= \frac{1}{1 - q - y_n - y_n q s_n(q)} - \delta_n(q),
\end{aligned} \tag{3.4}$$

where α_k^T is the k-th row of X_n (note that α_k also depends on n, which we drop from the notation), $X_n^{(k)}$ is X_n with its k-th row removed (thus a $(p-1)\times n$-matrix).

Further,

$$\delta_n(q) = \frac{1}{p}\sum_{k\in[p]} \frac{\Omega_k^{(n)}(q)}{(1-q-y_n-y_nqs_n(q))(1-q-y_n-y_nqs_n(q)+\Omega_k^{(n)}(q))}, \quad (3.5)$$

where for all $k \in \{1,\dots,p\}$:

$$\Omega_k^{(n)}(q) = \underbrace{\frac{1}{n}\alpha_k^T\alpha_k - 1}_{=0} - \frac{1}{n^2}\alpha_k^T X_n^{(k)T}\left(\frac{1}{n}X_n^{(k)}X_n^{(k)T} - q\right)^{-1} X_n^{(k)}\alpha_k + y_n + y_nqs_n(q).$$

Solving (3.4), we obtain analogously to Bai and Silverstein (2010, pp. 55 and 56) that

$$s_n(q) = \frac{1}{2y_nq}\left(1-q-y_n-y_nq\delta_n(q) + \sqrt{(1-q-y_n+y_nq\delta_n(q))^2 - 4y_nq}\right). \quad (3.6)$$

If $y_n \to y > 0$, we see from (3.6) that $s_n(q)$ converges almost surely to $S_{\mu^y}(1+\sqrt{y}z)$ provided $\delta_n(q) \to 0$ almost surely as $n \to \infty$. Here S_{μ^y} is the Stieltjes transform of the Marčenko–Pastur law μ^y. Then also $s_n(1+\sqrt{y}z) \to S_{\mu^y}(1+\sqrt{y}z)$ almost surely for all $z \in \mathbb{C}_+$, since all s_n are $(\min_n \operatorname{Im}(\sqrt{y_n}z))^{-2} < \infty$ Lipschitz continuous on the relevant domain. Therefore, $\mu_n \to \mu^y$ weakly almost surely.

If $y_n \to 0$, a straightforward calculation using (3.6) and the definition of q yields as $p \to \infty$,

$$y_n^{1/2}\, s_n(q) = \frac{-z+y_n^{1/2}\delta_n(q)q+\sqrt{z^2-4+2zy_n^{1/2}\delta_n(q)+2y_n\delta_n(q)+y_n\delta_n^2(q)}}{2} + o(1).$$

We see that $S_{y_n^{-1/2}(V_n-I)}(z) = y_n^{1/2}s_n(q)$ converges almost surely to the Stieltjes transform $s_G(z)$ of the semicircle law (see (1.3)) provided

$$\lim_{p\to\infty} y_n^{1/2}\delta_n(q) = 0 \quad \text{a.s.} \quad (3.7)$$

Thus, condition (3.7) suffices in both cases $p/n \to y > 0$ and $p/n \to 0$. It remains to prove (3.7).

3.2. *Proof of* (3.7). Recall the definition of $\delta_n(q)$ in (3.5). First, we lower bound the denominator. By (3.3.13) in Bai and Silverstein (2010) and p. 57 below (3.3.15), we have

$$\begin{aligned} \operatorname{Im}(1-q-y_n-y_nqs_n(q)) &\le -\operatorname{Im}(q), \\ \operatorname{Im}(1-q-y_n-y_nqs_n(q)+\Omega_k^{(n)}(q)) &\le -\operatorname{Im}(q). \end{aligned} \quad (3.8)$$

Using (3.8) we see that

$$|y_n^{1/2}\delta_n(q)| \le y_n^{1/2}|\operatorname{Im}(q)|^{-2}\frac{1}{p}\sum_{k\in[p]}|\Omega_k^{(n)}(q)| = \eta^{-2}y_n^{-1/2}\frac{1}{p}\sum_{k\in[p]}|\Omega_k^{(n)}(q)|.$$

Thus, it suffices to show that

$$\lim_{p\to\infty} y_n^{-1/2}\max_{k=1,\dots,p}|\Omega_k^{(n)}(q)| = 0 \quad \text{a.s.} \quad (3.9)$$

Now we prove (3.9). Note that

$$\Omega_k^{(n)}(q) = -\frac{1}{n^2}\alpha_k^T X_n^{(k)T}\left(\frac{1}{n}X_n^{(k)}X_n^{(k)T} - q\right)^{-1} X_n^{(k)}\alpha_k + y_n + y_nqs_n(q)$$

$$= \left(-\frac{1}{n^2} \sum_{i \neq j}^{n} \alpha_k(i) \left[X_n^{(k)T} \left(\frac{1}{n} X_n^{(k)} X_n^{(k)T} - q \right)^{-1} X_n^{(k)} \right] (i,j) \alpha_k(j) \right)$$
$$+ \left(-\frac{1}{n^2} \operatorname{tr} X_n^{(k)T} \left(\frac{1}{n} X_n^{(k)} X_n^{(k)T} - q \right)^{-1} X_n^{(k)} + y_n + y_n q s_n(q) \right)$$
$$=: A(n,k,q) + B(n,k,q).$$

We analyse $B(n,k,q)$ first. We have

$$- \frac{1}{n^2} \operatorname{tr} X_n^{(k)T} \left(\frac{1}{n} X_n^{(k)} X_n^{(k)T} - q \right)^{-1} X_n^{(k)}$$
$$= -\frac{1}{n} \operatorname{tr} \left[I_{p-1} + q \left(\frac{1}{n} X_n^{(k)} X_n^{(k)T} - q \right)^{-1} \right]$$
$$= -\frac{p}{n} + \frac{1}{n} - \frac{q}{n} \operatorname{tr} \left(\frac{1}{n} X_n^{(k)} X_n^{(k)T} - q \right)^{-1}.$$

Hence, using $y_n = p/n$, $\operatorname{Im}(q) = \sqrt{y_n}\eta$, and (A.1.12) in Bai and Silverstein (2010), we find

$$|B(n,k,q)|$$
$$= \left| -\frac{p}{n} + \frac{1}{n} - \frac{q}{n} \operatorname{tr} \left(\frac{1}{n} X_n^{(k)} X_n^{(k)T} - q \right)^{-1} + y_n + y_n q \frac{1}{p} \operatorname{tr} \left(\frac{1}{n} X_n X_n^T - q \right)^{-1} \right|$$
$$\leq \frac{1}{n} + \frac{|q|}{n \operatorname{Im}(q)} = \frac{1}{n} + \frac{|1 + \sqrt{y_n} z|}{n \sqrt{y_n} \eta}.$$

Since this bound holds uniformly for all $k \in \{1, \ldots, p\}$, it follows that

$$y_n^{-1/2} \max_{k=1,\ldots,p} |B(n,k,q)| \leq \frac{1}{n\sqrt{p/n}} + \frac{|1 + z\sqrt{p/n}|}{p\eta} \xrightarrow[p \to \infty]{} 0 \quad \text{a.s.}$$

It is left to show that $y_n^{-1/2} \max_{k=1,\ldots,p} |A(n,k,q)| \to 0$ almost surely. We do so by bounding the terms

$$S(n,k,q) := n^2 A(n,k,q) = \sum_{i \neq j}^{n} \alpha_k(i) \left[X_n^{(k)T} \left(\frac{1}{n} X_n^{(k)} X_n^{(k)T} - q \right)^{-1} X_n^{(k)} \right] (i,j) \alpha_k(j)$$

using Lemma 3.1 and Lemma 3.2. In accordance with Lemma 3.2, we consider the *symmetric* matrix

$$F\left(X_n^{(k)}\right) = X_n^{(k)T} \left(\frac{1}{n} X_n^{(k)} X_n^{(k)T} - q \right)^{-1} X_n^{(k)}.$$

Using $q = 1 + \sqrt{y_n} z$, we draw the following corollary of Lemma 3.2:

Corollary 3.3. *For any $a \in \mathbb{N}$ there exists a constant $C_a > 0$ independent of n and p such that for any $k \in [p]$ and any realization X of $X_n^{(k)}$, it holds*

i) $\left(\sum_{i \neq j} |F_{ij}(X)|^2 \right)^{\frac{a}{2}} \leq C_a n^{3a/2}$

ii) $\left| \sum_{i \neq j} F_{ij}(X) \right|^a \leq C_a n^{3a/2} p^{a/2}$

iii) $\left(\sum_{j\in[n]}\left|\sum_{i\in[n]\setminus\{j\}} F_{ij}(X)\right|^2\right)^{\frac{a}{2}} \leq C_a p^{a/2}$.

Proof: We will use Lemma 3.2 throughout the proof. For i) we obtain

$$\left(\sum_{i\neq j}|F_{ij}(X)|^2\right)^{\frac{1}{2}} \leq n\sqrt{p-1}\left(1+\frac{|q|}{\eta}\sqrt{\frac{n}{p}}\right) \leq C_1 n^{3/2}$$

for some constant C_1 independent of n and p. For ii) we note that

$$\begin{aligned}\left|\sum_{i\neq j}F_{ij}(X)\right| &\leq \left|\sum_{i,j}F_{ij}(X)\right| + |\operatorname{tr} F(X)| \\ &\leq \frac{p}{\operatorname{Im}(q)} + \frac{1}{n}\left(1+\frac{|q|}{\operatorname{Im}(q)}\right) + np\left(1+\frac{|q|}{\operatorname{Im}(q)}\right) \\ &\leq \frac{\sqrt{pn}}{\eta} + \frac{1}{n} + \frac{|q|}{\eta\sqrt{pn}} + np + \frac{|q|n^{3/2}p^{1/2}}{\eta} \leq C_2 n^{3/2}p^{1/2}\end{aligned}$$

for some constant C_2 independent of n and p. Now for iii) we calculate

$$\left(\sum_{j\in[n]}\left|\sum_{i\in[n]\setminus\{j\}}F_{ij}(X)\right|^2\right)^{\frac{1}{2}} \leq \left(n\frac{p}{n\eta^2}\right)^{\frac{1}{2}} \leq C_3 p^{1/2}$$

for a constant C_3 independent of n and p. This shows the statement with $C_a := \max\{C_1^a, C_2^a, C_3^a\}$. □

Throughout this section the random variable M_{np}^{β} satisifies the properties listed in Lemma 2.5 if $\beta\leq 1$ or those in Lemma 2.6 if $\beta>1$.

Note that the matrix X_n consists of np Curie-Weiss(β, np) spins, and that for any $k\in[p]$, α_k is the k-th row of X_n and thus contains variables disjoint from the variables in $X_n^{(k)}$. In what follows, we will use that for $r,s,t\geq 0$ and $a\in\mathbb{N}$ we have $(s+t)^a \leq 2^a(s^a+t^a)$ and $(r+s+t)^a\leq 4^a(r^a+s^a+t^a)$. We calculate for $a\in 2\mathbb{N}$ and $k\in[p]$ arbitrary, where sums over $i\neq j$ are for $i,j\in[n]$, and further explanations can be found beneath the calculation:

$$\begin{aligned}&\mathbb{E}|S(n,k,q)|^a = \mathbb{E}\mathbb{E}[|S(n,k,q)|^a|M_{np}^{\beta}] = \mathbb{E}\mathbb{E}\left[\sum_{i\neq j}\alpha_k(i)F_{ij}\left(X_n^{(k)}\right)\alpha_k(j)\,\middle|\, M_{np}^{\beta}\right] \\ &= \int_{[-1,1]}\int_{\{\pm1\}^{(p-1)\times n}}\int_{\{\pm1\}^n}\left|\sum_{i\neq j}x_iF_{ij}(X)x_j\right|^a \mathbb{P}^{\alpha_k|M_{np}^{\beta}=t}(\mathrm{d}x)\mathbb{P}^{X_n^{(k)}|M_{np}^{\beta}=t}(\mathrm{d}X)\mathbb{P}^{M_{np}^{\beta}}(\mathrm{d}t) \\ &\leq \int_{[-1,1]}\int_{\{\pm1\}^{(p-1)\times n}} 4^a(4A_a^2\mu_a^2)^a\left(\sum_{i\neq j}|F_{ij}(X)|^2\right)^{\frac{a}{2}}\mathbb{P}^{X_n^{(k)}|M_{np}^{\beta}=t}(\mathrm{d}X)\mathbb{P}^{M_{np}^{\beta}}(\mathrm{d}t) \\ &+ \int_{[-1,1]}\int_{\{\pm1\}^{(p-1)\times n}} 4^a(2A_a\mu_a)^a|t|^a\left(\sum_{j\in[n]}\left|\sum_{i\in[n]\setminus\{j\}}F_{ij}(X)\right|^2\right)^{\frac{a}{2}}\mathbb{P}^{X_n^{(k)}|M_{np}^{\beta}=t}(\mathrm{d}X)\mathbb{P}^{M_{np}^{\beta}}(\mathrm{d}t)\end{aligned}$$

$$+ \int_{[-1,1]} \int_{\{\pm 1\}^{(p-1)\times n}} 4^a |t|^{2a} \left| \sum_{i\neq j} F_{ij}(X) \right|^a \mathbb{P}^{X_n^{(k)}|M_{np}^\beta = t}(\mathrm{d}X) \mathbb{P}^{M_{np}^\beta}(\mathrm{d}t)$$
$$\leq K \left(n^{3a/2} + p^{a/2} \int_{[-1,1]} |t|^a \mathbb{P}^{M_{np}^\beta}(\mathrm{d}t) + n^{3a/2} p^{a/2} \int_{[-1,1]} |t|^{2a} \mathbb{P}^{M_{np}^\beta}(\mathrm{d}t) \right)$$
$$\leq K \left(n^{3a/2} + p^{a/2}(np)^{-a/4} + n^{3a/2} p^{a/2} (np)^{-a/2} \right) \leq K n^{3a/2}.$$

For the fourth step we used Lemma 3.1 and the constants A_a and μ_a therein (note that $F(X)$ is symmetric). In the fifth step we used Corollary 3.3 and from here on out, K denotes a constant not depending on n and p, but only on a, β and η, and K may change its value from one occurrence to the next. In the sixth step we applied Lemma 2.5. Hence, if $\epsilon > 0$ is arbitrary, we get with a union bound and Markov's inequality that

$$\mathbb{P}\left(\frac{\max_{k\in[p]} |A(n,k,q)|}{\sqrt{y_n}} > \epsilon \right) \leq p \max_{k\in[p]} \mathbb{P}(S(n,k,q)^a > \epsilon n^{3a/2} p^{a/2}) \leq \frac{pKn^{3a/2}}{\epsilon^a n^{3a/2} p^{a/2}},$$

which is summable in p for (say) $a = 6$. By the Borel-Cantelli lemma, it follows that $y_n^{-1/2} \max_{k=1,\dots,p} |A(n,k,q)| \to 0$ almost surely.

3.3. *The case* $\beta > 1$. To prove part (ii) of Theorem 2.1, let $\beta > 1$. Instead of the matrix X_n we consider

$$\tilde{X}_n := \frac{1}{\sqrt{1-m^2}} \left(X_n(i,j) - m\mathbb{1}_{M_{np}^\beta > 0} + m\mathbb{1}_{M_{np}^\beta < 0} \right)_{i\in[p], j\in[n]},$$

which for every realization is just a rank 1 perturbation of $(1-m^2)^{-1/2} X_n$. As a consequence, it suffices to prove Theorem 2.1 (ii) for $\tilde{V}_n := n^{-1} \tilde{X}_n \tilde{X}_n^T$ instead of $(1-m^2)^{-1} V_n$. Using the terminology as above, but substituting $\tilde{X}_n$ for X_n and $\tilde{V}_n$ for V_n, we obtain new terms $\tilde{s}_n$, $\tilde{\Omega}_n^{(k)}$, $\tilde{\alpha}_k$, $\tilde{\delta}_n$ and $\tilde{S}(n,k,z)$. Inspecting above proof for the case $\beta \leq 1$ and observing (3.9), it will suffice to show

$$\lim_{p\to\infty} y_n^{-1/2} \max_{k=1,\dots,p} |\tilde{\Omega}_k^{(n)}(q)| = 0 \quad \text{a.s.}$$

Here,

$$\begin{aligned}
&\tilde{\Omega}_k^{(n)}(q) \\
&= -\frac{1}{n^2} \tilde{\alpha}_k^T \tilde{X}_n^{(k)T} \left(\frac{1}{n} \tilde{X}_n^{(k)} \tilde{X}_n^{(k)T} - q \right)^{-1} \tilde{X}_n^{(k)} \tilde{\alpha}_k + y_n + y_n q \tilde{s}_n(q) + \frac{1}{n} \tilde{\alpha}_k^T \tilde{\alpha}_k - 1 \\
&= \left(-\frac{1}{n^2} \sum_{i\neq j}^n \tilde{\alpha}_k(i) \left[\tilde{X}_n^{(k)T} \left(\frac{1}{n} \tilde{X}_n^{(k)} \tilde{X}_n^{(k)T} - q \right)^{-1} \tilde{X}_n^{(k)} \right] (i,j) \tilde{\alpha}_k(j) \right) \\
&\quad + \left(-\frac{1}{n^2} \operatorname{tr} \tilde{X}_n^{(k)T} \left(\frac{1}{n} \tilde{X}_n^{(k)} \tilde{X}_n^{(k)T} - q \right)^{-1} \tilde{X}_n^{(k)} + y_n + y_n q \tilde{s}_n(q) \right) \\
&\quad + \left(\frac{1}{n} \tilde{\alpha}_k^T \tilde{\alpha}_k - 1 \right) \\
&=: \tilde{A}(n,k,q) + \tilde{B}(n,k,q) + \tilde{C}(n,k,q).
\end{aligned}$$

The term $\tilde{B}(n,k,q)$ can be treated analogously to the term $B(n,k,q)$ above, so we obtain

$$y_n^{-1/2} \max_{k=1,\dots,p} |\tilde{B}(n,k,q)| \leq \frac{1}{n\sqrt{p/n}} + \frac{|1+z\sqrt{p/n}|}{p\eta} \xrightarrow{p\to\infty} 0 \quad \text{a.s.}$$

To handle $\tilde{A}(n,k,q)$, we use Lemma 2.6 and the definitions therein. $\tilde{X}_n$ is a matrix of np perturbed Curie-Weiss(β, np) spins. Now $\tilde{\alpha}_k$ is the k-th row of $\tilde{X}_n$ and thus contains variables disjoint from those in $\tilde{X}_n^{(k)}$. Analogously to the case above we consider the term

$$F(\tilde{X}_n^{(k)}) = \tilde{X}_n^{(k)T} \left(\frac{1}{n} \tilde{X}_n^{(k)} \tilde{X}_n^{(k)T} - q \right)^{-1} \tilde{X}_n^{(k)}.$$

We use a slightly faster calculation than for the case $\beta \leq 1$, where a finer analysis was necessary due to the slowly decaying correlations when $\beta = 1$. In the following, we will directly compare $|\tilde{S}(n,k,q)|$ to

$$\tilde{R}(n,k,q) := \left(\sum_{i\neq j} |F_{ij}(\tilde{X}_n^{(k)})|^2 \right)^{\frac{1}{2}}.$$

Note that $\tilde{R}(n,k,q)$ never vanishes, so we may divide by it. Now for $T > 0$ and $a \in 2\mathbb{N}$ (and where sums over $i \neq j$ are for $i,j \in [n]$) we calculate for $k \in [p]$:

$$\begin{aligned}
&\mathbb{P}\left(|\tilde{S}(n,k,q)| > T\tilde{R}(n,k,q) \right) \\
&= \mathbb{P}\left(\left| \sum_{i\neq j} \tilde{\alpha}_k(i) F_{ij}(\tilde{X}_n^{(k)}) \tilde{\alpha}_k(j) \right| > T \left(\sum_{i\neq j} |F_{ij}(\tilde{X}_n^{(k)})|^2 \right)^{\frac{1}{2}} \right) \\
&\leq \frac{1}{T^a} \int_{[-1,1]} \int_{\mathcal{Z}_2} \int_{\mathcal{Z}_1} \left| \frac{\sum_{i\neq j} \tilde{x}_i F_{ij}(\tilde{X}) \tilde{x}_j}{\left(\sum_{i\neq j} |F_{ij}[\tilde{X}]|^2 \right)^{\frac{1}{2}}} \right|^a \mathbb{P}^{\tilde{\alpha}_k | M_{np}^\beta = t}(\mathrm{d}\tilde{x}) \mathbb{P}^{\tilde{X}_n^{(k)} | M_{np}^\beta = t}(\mathrm{d}\tilde{X}) \mathbb{P}^{M_{np}^\beta}(\mathrm{d}t) \\
&\leq \frac{1}{T^a} \int_{[-1,1]} \int_{\mathcal{Z}_2} \left[4A_a^2 \mu_a^2 + 2A_a \mu_a \sqrt{n} |\zeta(t)| + n|\zeta(t)|^2 \right]^a \mathbb{P}^{\tilde{X}_n^{(k)} | M_{np}^\beta = t}(\mathrm{d}\tilde{X}) \mathbb{P}^{M_{np}^\beta}(\mathrm{d}t) \\
&\leq \frac{1}{T^a} \int_{[-1,1]} 4^a (4A_a^2 \mu_a^2)^a \mathbb{P}^{M_{np}^\beta}(\mathrm{d}t) \;+\; \frac{1}{T^a} \int_{[-1,1]} 4^a (2A_a \mu_a)^a n^{a/2} |\zeta(t)|^a \mathbb{P}^{M_{np}^\beta}(\mathrm{d}t) \\
&\quad + \frac{1}{T^a} \int_{[-1,1]} 4^a n^a |\zeta(t)|^{2a} \mathbb{P}^{M_{np}^\beta}(\mathrm{d}t) \\
&\leq \frac{1}{T^a} \left(K + Kn^{a/2} \int_{[-1,1]} |\zeta(t)|^a \mathbb{P}^{M_{np}^\beta}(\mathrm{d}t) + Kn^a \int_{[-1,1]} |\zeta(t)|^{2a} \mathbb{P}^{M_{np}^\beta}(\mathrm{d}t) \right) \\
&\leq \frac{K}{T^a},
\end{aligned}$$

where $\mathcal{Z}_1$ and $\mathcal{Z}_2$ denote the ranges of $\tilde{\alpha}_k$ and $\tilde{X}_n^{(k)}$, respectively (cf. Lemma 2.6). Further, in the third step we used Lemma 3.1 and in the last step Lemma 2.6. Again, K denotes a floating constant which may change its value from one occurrence to the next, but remains independent of k, n and p. Therefore we have

$$\frac{K}{T^a} \geq \mathbb{P}\left(|\tilde{S}(n,k,q)| > T\tilde{R}(n,k,q) \right) \geq \mathbb{P}\left(n^2 |\tilde{A}(n,k,q)| > Tn\sqrt{p} \left(1 + \frac{|q|}{\mathrm{Im}(q)} \right) \right)$$

$$= \mathbb{P}\left(\frac{|\tilde{A}(n,k,q)|}{\sqrt{p/n}} > \frac{T}{\sqrt{n}}\left(1+\frac{|q|\sqrt{n}}{\eta\sqrt{p}}\right)\right) = \mathbb{P}\left(\frac{|\tilde{A}(n,k,q)|}{\sqrt{y_n}} > \frac{T}{\sqrt{n}} + \frac{T|q|}{\eta\sqrt{p}}\right),$$

where in the second step we used the bound on $\tilde{R}(n,k,q)$ given by Lemma 3.2. Choosing $T = p^{1/4}$, $a \in 2\mathbb{N}$ such that $a > 8$ and using the union bound shows by Borel-Cantelli that

$$y_n^{-1/2} \max_{k=1,\dots,p} |\tilde{A}(n,k,q)| \xrightarrow[p\to\infty]{} 0 \quad \text{almost surely.}$$

It is left to analyze $\tilde{C}(n,k,q)$. Note that this term was zero in the case $\beta \leq 1$. Note also that it suffices to show

$$\frac{1}{\sqrt{y_n}} \max_{k\in[p]} \left| \frac{1}{n}\sum_{l=1}^{n}(\tilde{X}(k,l)^2 - 1)\right| = \frac{1}{\sqrt{y_n}} \max_{k\in[p]} |\tilde{C}(n,k,q)| \xrightarrow[p\to\infty]{} 0 \qquad \text{a.s.}$$

To this end, for $T > 0$, $k \in [p]$ and $a \in 2\mathbb{N}$ arbitrary it holds

$$\begin{aligned}
&\mathbb{P}\left(\left|\frac{1}{n}\sum_{l=1}^{n}(\tilde{X}_n(k,l)^2-1)\right| > T\right) \leq \frac{1}{(Tn)^a}\mathbb{E}\left|\sum_{l=1}^{n}(\tilde{X}_n(k,l)^2-1)\right|^a \\
&= \frac{1}{(Tn)^a}\int_{[-1,1]}\int_{\mathcal{Z}_1}\left|\sum_{l=1}^{n}(x_{kl}^2-1)\right|^a \mathbb{P}^{\tilde{\alpha}_k|M_{np}^\beta=t}(\mathrm{d}x)\mathbb{P}^{M_{np}^\beta}(\mathrm{d}t) \\
&\leq \frac{1}{(Tn)^a}\int_{[-1,1]}\left[(A_a\mu_a+\sqrt{n}\psi(t))\sqrt{n}\right]^a \mathbb{P}^{M_{np}^\beta}(\mathrm{d}t) \\
&= \frac{1}{(Tn)^a}\int_{[-1,1]}\left[A_a\mu_a\sqrt{n}+n\psi(t)\right]^a \mathbb{P}^{M_{np}^\beta}(\mathrm{d}t) \\
&\leq \frac{1}{(Tn)^a}\left(2^a(A_a\mu_a)^a n^{a/2} + 2^a n^a\int_{[-1,1]}\psi(t)^a\mathbb{P}^{M_{np}^\beta}(\mathrm{d}t)\right) \\
&\leq \frac{1}{(Tn)^a}Kn^{a/2} + \frac{1}{(Tn)^a}K\frac{n^a}{n^{a/2}} \leq \frac{K}{T^a n^{a/2}},
\end{aligned}$$

where in the first step we used Markov's inequality, in the second step conditional expectations, in the third step Lemma 3.1 i) with $b_i = 1$, and in the last line Lemma 2.6.

Choosing $\epsilon > 0$ arbitrarily and setting $T := \epsilon\sqrt{p/n}$ we obtain for $a \in \mathbb{N}$ with $a \in 2\mathbb{N}$ arbitrary that

$$\mathbb{P}\left(\frac{1}{\sqrt{y_n}}\max_{k\in[p]}|\tilde{C}(n,k,q)| > \epsilon\right) \leq \sum_{k\in[p]}\mathbb{P}\left(|\tilde{C}(n,k,q)| > \epsilon\sqrt{y_n}\right) \leq \frac{pK}{\epsilon^a p^{a/2}},$$

which is summable over p for $a > 4$. This ends the proof via Borel-Cantelli.

Acknowledgements

We thank Nestor Parolya for proofreading the manuscript. The authors gratefully acknowledge the work of an anonymous referee and an associate editor.

References

Auffinger, A., Ben Arous, G., and Péché, S. Poisson convergence for the largest eigenvalues of heavy tailed random matrices. *Ann. Inst. Henri Poincaré Probab. Stat.*, **45** (3), 589–610 (2009). MR2548495.

Bai, Z. and Silverstein, J. W. *Spectral analysis of large dimensional random matrices*. Springer Series in Statistics. Springer, New York, second edition (2010). ISBN 978-1-4419-0660-1. MR2567175.

Bai, Z. and Zhou, W. Large sample covariance matrices without independence structures in columns. *Statist. Sinica*, **18** (2), 425–442 (2008). MR2411613.

Bai, Z. D. and Yin, Y. Q. Convergence to the semicircle law. *Ann. Probab.*, **16** (2), 863–875 (1988). MR929083.

Bandeira, A. S., Lodhia, A., and Rigollet, P. Marčenko-Pastur law for Kendall's tau. *Electron. Commun. Probab.*, **22**, Paper No. 32, 7 (2017). MR3663103.

Bao, Z. Tracy-Widom limit for Kendall's tau. *Ann. Statist.*, **47** (6), 3504–3532 (2019). MR4025750.

Basrak, B., Cho, Y., Heiny, J., and Jung, P. Extreme eigenvalue statistics of m-dependent heavy-tailed matrices. *ArXiv Mathematics e-prints* (2020). arXiv: 1910.08511.

Bodnar, T., Dette, H., and Parolya, N. Testing for independence of large dimensional vectors. *Ann. Statist.*, **47** (5), 2977–3008 (2019). MR3988779.

Davis, R. A., Heiny, J., Mikosch, T., and Xie, X. Extreme value analysis for the sample autocovariance matrices of heavy-tailed multivariate time series. *Extremes*, **19** (3), 517–547 (2016a). MR3535965.

Davis, R. A., Mikosch, T., and Pfaffel, O. Asymptotic theory for the sample covariance matrix of a heavy-tailed multivariate time series. *Stochastic Process. Appl.*, **126** (3), 767–799 (2016b). MR3452812.

El Karoui, N. Concentration of measure and spectra of random matrices: applications to correlation matrices, elliptical distributions and beyond. *Ann. Appl. Probab.*, **19** (6), 2362–2405 (2009). MR2588248.

Fleermann, M. *Global and Local Semicircle Laws for Random Matrices with Correlated Entries*. Ph.D. thesis, FernUniversität in Hagen, Germany (2019a).

Fleermann, M. Local semicircle laws for curie-weiss type ensembles. *ArXiv Mathematics e-prints* (2019b). arXiv: 1907.08782.

Fleermann, M., Kirsch, W., and Kriecherbauer, T. Local laws for Curie-Weiss matrix ensembles at subcritical temperatures (in preparation) (2020).

Fleermann, M., Kirsch, W., and Kriecherbauer, T. The almost sure semicircle law for random band matrices with dependent entries. *Stochastic Process. Appl.*, **131**, 172–200 (2021). MR4153882.

Friesen, O. and Löwe, M. A phase transition for the limiting spectral density of random matrices. *Electron. J. Probab.*, **18**, no. 17, 17 (2013). MR3035745.

Geman, S. A limit theorem for the norm of random matrices. *Ann. Probab.*, **8** (2), 252–261 (1980). MR566592.

Heiny, J. and Mikosch, T. Eigenvalues and eigenvectors of heavy-tailed sample covariance matrices with general growth rates: the iid case. *Stochastic Process. Appl.*, **127** (7), 2179–2207 (2017). MR3652410.

Heiny, J. and Mikosch, T. Almost sure convergence of the largest and smallest eigenvalues of high-dimensional sample correlation matrices. *Stochastic Process. Appl.*, **128** (8), 2779–2815 (2018). MR3811704.

Heiny, J. and Mikosch, T. The eigenstructure of the sample covariance matrices of high-dimensional stochastic volatility models with heavy tails. *Bernoulli*, **25** (4B), 3590–3622 (2019). MR4010966.

Heiny, J. and Yao, J. Limiting distributions for eigenvalues of sample correlation matrices from heavy-tailed populations. *ArXiv Mathematics e-prints* (2020). arXiv: 2003.03857.

Hochstättler, W., Kirsch, W., and Warzel, S. Semicircle law for a matrix ensemble with dependent entries. *J. Theoret. Probab.*, **29** (3), 1047–1068 (2016). MR3540489.

Kirsch, W. A Survey on the Method of Moments (2015). URL `https://www.fernuni-hagen.de/stochastik/downloads/momente.pdf`.

Kirsch, W. and Kriecherbauer, T. Semicircle law for generalized Curie-Weiss matrix ensembles at subcritical temperature. *J. Theoret. Probab.*, **31** (4), 2446–2458 (2018). MR3866621.

Kochmański, M., Paszkiewicz, T., and Wolski, S. Curie–Weiss magnet—a simple model of phase transition. *European Journal of Physics*, **34** (6), 1555 (2013). DOI: 10.1088/0143-0807/34/6/1555.

Marčenko, V. A. and Pastur, L. A. Distribution of eigenvalues in certain sets of random matrices. *Mat. Sb. (N.S.)*, **72 (114)**, 507–536 (1967). MR0208649.

Pan, G. and Gao, J. Asymptotic theory for sample covariance matrix under cross-sectional dependence (2012). Preprint.

Silverstein, J. W. and Bai, Z. D. On the empirical distribution of eigenvalues of a class of large-dimensional random matrices. *J. Multivariate Anal.*, **54** (2), 175–192 (1995). MR1345534.

Wang, L. and Paul, D. Limiting spectral distribution of renormalized separable sample covariance matrices when $p/n \to 0$. *J. Multivariate Anal.*, **126**, 25–52 (2014). MR3173080.

Yao, J., Zheng, S., and Bai, Z. *Large sample covariance matrices and high-dimensional data analysis*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, New York (2015). ISBN 978-1-107-06517-8. MR3468554.

Yaskov, P. A short proof of the Marchenko-Pastur theorem. *C. R. Math. Acad. Sci. Paris*, **354** (3), 319–322 (2016). MR3463031.