EXTREME EIGENVALUES OF SPARSE, HEAVY TAILED RANDOM MATRICES

ANTONIO AUFFINGER AND SI TANG

Abstract. We study the statistics of the largest eigenvalues of $p \times p$ sample covariance matrices $\Sigma_{p,n} = M_{p,n}M_{p,n}^*$ when the entries of the $p \times n$ matrix $M_{p,n}$ are sparse and have a distribution with tail $t^{-\alpha}$, $\alpha > 0$. On average the number of nonzero entries of $M_{p,n}$ is of order $n^{\mu+1}$, $0 \leq \mu \leq 1$. We prove that in the large $n$ limit, the largest eigenvalues are Poissonian if $\alpha < 2(1 + \mu - 1)$ and converge to a constant in the case $\alpha > 2(1 + \mu - 1)$. We also extend the results of [7] in the Hermitian case, removing restrictions on the number of nonzero entries of the matrix.

1. Introduction

We study the statistics of the largest eigenvalues of sample covariance matrices of the form $\Sigma = MM^*$ when the entries of $M$ are heavy tailed and sparse. Let $x$ be a complex-valued random variable. We say $x$ has a heavy tailed distribution with parameter $\alpha$ if the (two-sided) tail probability

$$G_\alpha(t) := P(|x| > t) = L(t)t^{-\alpha}, \quad t > 0$$

where $\alpha > 0$ and $L$ is a slowly varying function, i.e.,

$$\lim_{t \to \infty} \frac{L(st)}{L(t)} = 1, \quad \forall s > 0.$$

For each $n \geq 1$, let $y = y(n)$ be a Bernoulli random variable, independent of $x$, with $P(y = 1) = n^{\mu-1} = 1 - P(y = 0)$, where $0 \leq \mu \leq 1$ is a constant. The ensemble of random sample covariance matrices that we study here is defined as follows. For each $n \geq 1$, let $p = p(n) \in \mathbb{Z}_+$ be a function of $n$ such that

$$\frac{p}{n} \to \rho, \quad 0 < \rho \leq 1,$$

as $n \to \infty$. Let $A_{p,n} = [a_{ij}]_{i,j=1}^{p,n}$ and $B_{p,n} = [b_{ij}]_{i,j=1}^{p,n}$ be $p \times n$ random matrices whose entries are i.i.d. copies of $x$ and $y$, respectively. Form the $p \times n$ sparse matrix $M_{p,n} = A_{p,n} \cdot B_{p,n} = [m_{ij}]_{i,j=1}^{p,n}$ by setting $m_{ij} = a_{ij}b_{ij}$. Then

$$\Sigma_{p,n} := M_{p,n}M_{p,n}^*$$

is the $p \times p$ heavy tailed random sample covariance matrix with parameters $\alpha$ and $\mu$. Note that $\Sigma_{p,n}$ is positive semi-definite so all its eigenvalues are non-negative.

© 2016. This manuscript version is made available under the Elsevier user license
http://www.elsevier.com/open-access/userlicense/1.0/
The extreme eigenvalues of $\Sigma_{p,n}$ are the main subject of this paper. We will see that, depending on the tail exponent $\alpha$ and the sparsity exponent $\mu$, when properly rescaled, the top eigenvalues will either converge to the points of a Poisson point process or to the right edge of the Marchenko-Pastur law.

To put our theorems in context, we briefly review past results. The study of extreme eigenvalues of heavy tailed random matrices started with the work of Soshnikov. In [26], he proved that if $0 < \alpha < 2$, the asymptotic behavior of the top eigenvalues of a heavy tailed Hermitian matrix is determined by the behavior of the largest entries of the matrix, i.e., the point process of the largest eigenvalues (properly normalized) converges to a Poisson point process, as in the usual extreme value theory for i.i.d. random variables. This result was extended to sample covariance matrices and for all values of $\alpha \in (0, 4)$ in the work of Auffinger, Ben Arous and Péché [2]. The upper bound on the tail exponent $\alpha$ is optimal as for i.i.d. entries with finite fourth moment, the largest eigenvalues converge to the right edge of the bulk distribution and have Tracy-Widom fluctuations [4, 5, 17, 29]. Eigenvector localization and delocalization were studied in [6]. In the physics literature, many of these results were predicted in the seminal paper of Bouchaud and Cizeau [13].

The largest eigenvalues of sparse Hermitian random matrices with bounded moments were investigated by Benaych-Georges and Péché [8] under the assumptions of at least $\omega (\log n)$ nonzero entries in each row. They extended the results of [16, 25], establishing the convergence of the largest eigenvalue to the edge and also obtained results on localization/delocalization of eigenvectors. For bulk statistics in the sparse setting, readers are invited to see Erdős, Knowles, Yau, and Yin [14] and the references therein. In [7], Benaych-Georges and Péché considered a class of $n \times n$ Hermitian, heavy tailed, sparse matrices. In their work, the authors looked at matrices, where in $n - o(n)$ rows, the number of nonzero entries was asymptotically equal to $n^\mu$ for $\mu \in (0, 1]$. For the remaining $o(n)$ rows, the number of nonzero entries was no more than $n^\mu$. This assumption is well-suited to treat the case of heavy-tailed band matrices. In the last section, we will extend the work of [7] by removing all restrictions on the number of nonzero entries in each row, allowing, for instance, the sparsity to come from the adjacency matrix of an Erdős-Rényi random graph.

Although we extend the results of [7], the main objective of this paper is to treat the spectrum of sample covariance matrices $\Sigma_{p,n}$ constructed from a sparse matrix $M_{p,n}$. These matrices naturally appear in applications such as models of complex networks with two species of nodes [19] and also in information theory as channel capacity of wideband CDMA schemes [30]. For more applications and predictions one can look at [18, 20, 23, 24] and the references therein. In the mathematical literature, as far as we know, there are no results dealing with the top eigenvalues of sample covariance matrices coming from sparse matrices. The main purpose of this paper is to provide such results. Compared to [2], [7], the main obstacle here is to control the trace of diverging powers of the matrix (see Theorem 4.4) via the method of moments. Here the approach of [2], [7] does not suffice and the analysis requires subtler combinatorial estimates (see Remark 8).
Throughout the paper, we will use $\lambda_l(A)$ to denote the $l$-th largest eigenvalue of a Hermitian matrix $A$, $\vec{v}_l(A)$ the corresponding eigenvector. For a matrix $A = [a_{ij}]$, either Hermitian or rectangular, $a_{njl}$ denotes its $l$-th largest entry in absolute value in the upper-triangular part (if $A$ is Hermitian) or of all entries (if $A$ is rectangular), and $\theta_l(A)$ be its argument, i.e., $\theta_l(A) = \arg(a_{njl})$. Let $\vec{e}_1, \ldots, \vec{e}_n$ represent the canonical basis vectors for $\mathbb{R}^n$. The notation $f(x) \sim g(x)$ means that there exists some slowly varying function $l(x)$ such that $f(x) = l(x)g(x)$. A sequence of events $(E_n)_{n \geq 1}$ is said to occur with exponentially high probability (w.e.h.p.) if there exist $C, \theta > 0$ and $n_0 \in \mathbb{N}$ such that for $n \geq n_0$, $\mathbb{P}(E_n) \geq 1 - e^{-Cn^\theta}$. We will also use the following matrix norms:

$$||A||_\infty := \max_i \sum_j |a_{ij}|, \quad ||A||_1 := \max_j \sum_i |a_{ij}|, \quad ||A|| := \max_{\vec{v}:||\vec{v}||_2 = 1} ||A\vec{v}||_2.$$ 

The rest of the sections will be organized as follows. In Section 2, we state our main results. Section 3 will be devoted to the proof of the main theorems while we defer most of the technical key lemmas to Section 4.

1.1. **Acknowledgements.** The authors would like to thank an anonymous referee. His valuable comments greatly improved the presentation of the paper.

2. **Main results**

Our main results are about the eigenvalues of the sample covariance matrix $\Sigma_{p,n} = M_{p,n}^* M_{p,n}$. In our setting, there are approximately $p \cdot n \cdot n^{\mu - 1} \approx p n^{1+\mu}$ nonzero entries in $M_{p,n}$. We know from extreme value theory [22, Section 1.2] that the scaling factor for the largest entries of the matrix $M_{p,n}$ should be

$$c_{np} := \inf \left\{ t : G_\alpha(t) \leq \frac{1}{p n^\mu} \right\}.$$ 

Moreover, $c_{np} \sim n^{(1+\mu)/\alpha}$ and

$$\lim_{n \to \infty} \mathbb{P} \left( \max_{i,j} c_{np}^{-1} |m_{ij}| \leq t \right) = e^{-t^{-\alpha}}.$$ 

Our first theorem says that when $0 < \alpha < 2(1 + \mu^{-1})$, the extreme eigenvalues of $\Sigma_{p,n}$ behave like the square of the top entries of $M_{p,n}$.

**Theorem 2.1.** Suppose $0 \leq \mu \leq 1$ and $0 < \alpha < 2(1 + \mu^{-1})$. For $(1 + \mu^{-1}) \leq \alpha < 2(1 + \mu^{-1})$, we also assume that $x$ is centered. Then as $n \to \infty$, we have for each $l \geq 1$

$$\frac{\lambda_l(\Sigma_{p,n})}{|m_{i,j}|^2} P \to 1.$$ 

The eigenvectors are localized: for each $l \geq 1$, $||\vec{v}_i(\Sigma_{p,n}) - \vec{e}_i||_2 P \to 0$. 
It follows from Theorem 2.1 and a routine computation (see [2, Page 593]) that the random point processes
\begin{equation}
\mathcal{Q}_n = \sum_{i=1}^{p} \sum_{j=1}^{n} \delta_{c_{n}^{2}|m_{ij}|^{2}}, \quad \hat{\mathcal{Q}}_n = \sum_{i=1}^{p} \delta_{c_{n}^{2} \lambda_i(\Sigma_{p,n})}
\end{equation}
converge in distribution to the same Poisson point process on \((0, +\infty)\) with intensity \(\frac{\alpha}{2^{1+\alpha/2}}\).

Remark 1. As mentioned in the introduction, the conclusion of the Theorem above holds in the non-sparse case if and only if \(0 < \alpha < 4\). Roughly speaking, when we introduce sparseness, we increase the localization of the eigenvectors towards the position of the largest entry and the Poissonian limit holds with lighter tails \(2(1 + \mu - 1) > 4\). Note that, when \(\mu = 0\), we interpret \(\mu^{-1} = 1/0 = \infty\). In this case, \(\alpha < \infty\) always holds and thus any polynomial tail is allowed. This was also observed in [7]. One should also note that although \(M_{p,n}\) is sparse, \(\Sigma_{p,n}\) is, in general, not.

In the second regime, \(\alpha > 2(1 + \mu^{-1})\), the Poissonian limit no longer holds. The largest eigenvalues, when normalized by \(n^{\mu}\), converge to the edge of the bulk distribution. We also need the following definition. For \(L \in \mathbb{N}\) and \(\eta \in (0, 1]\), we say that a unit vector \(\vec{v} = (v_1, \ldots, v_n) \in \mathbb{C}^n\) is \((L, \eta)-localized\) if there exists a set \(S \subseteq \{1, \ldots, n\}\) with cardinality \(L\) such that
\[\sum_{j \in S} |v_j|^2 > 1 - \eta.\]

**Theorem 2.2.** Suppose \(0 < \mu \leq 1, \alpha > 2(1 + \mu^{-1})\), and \(x\) has mean zero and variance one. Then for each \(l \geq 1\), as \(n \to \infty\), we have
\[\frac{\lambda_l(\Sigma_{p,n})}{n^{\mu}} \xrightarrow{p} (1 + \sqrt{\rho})^2.\]
The eigenvectors of \(\Sigma_{p,n}\) are delocalized, namely, there exists \(\beta, \eta_0 > 0\) such that for each \(l \geq 1\), \(0 < \eta < \eta_0\) we have
\begin{equation}
\mathbb{P}\left(\vec{v}_l(\Sigma_{p,n}) \text{ is } (\lfloor p^{\beta} \rfloor, \eta)-localized\right) \to 0
\end{equation}
as \(n \to \infty\).

Remark 2. In the regime of Theorem 2.2, when \(\mu = 0\), we are in the critical case of an Erdős-Rényi adjacency matrix. To have \(\alpha > 2(1 + \mu^{-1})\), we are forced to take \(\alpha = \infty\), which is not allowed. Again, we interpret \(1/0 = \infty\). In this case, it is still an open question to obtain explicit formulas for the limiting spectral distribution. To see more in this direction, the reader is invited to check [14] and the references therein.

Remark 3. The form of delocalization in (3) is relatively simple compared to the results obtained in [6, 12] when considering non-heavy tailed distributions for Wigner matrices. In words, (3) says that if \(\alpha > 2(1 + \mu^{-1})\) eigenvectors must have nonzero coordinates spread over at least \(p^{\beta}\) coordinates, different from the case \(\alpha < 2(1 + \mu^{-1})\) where the number of nonzero coordinates does not diverge with \(n\).
Remark 4. One can also take $P(y = 1) = f(n)n^{\mu - 1}$ for a slowly varying function $f \neq 0$. The results of the above theorems remain true, with an additional slowly varying function in the normalization of the entries.

Remark 5. Our results also hold for Hermitian matrices when the positions of the nonzero entries are specified using i.i.d. Bernoulli random variables. This shows that the results of [7, 8] are true if the positions of the nonzero entries are random, in which case the deterministic asymptotic constraint on the number of nonzeros in a row is not needed.

We can also obtain the convergence of the corresponding empirical spectral measures. We assume $\mu > 0$ for the next proposition.

**Proposition 2.3.** Suppose $0 < \mu \leq 1$, $\alpha > 2$, and $x$ has variance one. Then the empirical spectral distribution of $\Sigma_{p,n}/n^\mu$ converges almost surely to the Marchenko-Pastur law with density

$$\frac{\sqrt{(\lambda_+ - x)(x - \lambda_-)}}{2\pi \rho x} 1_{[\lambda_-, \lambda_+]}(x)$$

with $\lambda_\pm = (1 \pm \sqrt{\rho})^2$.

Remark 6. The proof of this Proposition follows from the classic truncation and moment method for random matrices (See for instance [1, Exercise 2.1.18]). Normalizing the entries $m_{ij}$ by $n^{\mu/2}$ gives the desired variance:

$$\text{Var}(m_{ij}/n^{\mu/2}) = n^{-\mu} \text{Var}(m_{ij}) = n^{-\mu} \cdot n^{\mu - 1} = \frac{1}{n}.$$ 

3. Proof of the main theorems

3.1. Proof of Theorem 2.1.

3.1.1. Proof strategy. We will use the strategy that was first proposed by Soshnikov [26] when proving the heavy tailed Hermitian matrix case with $0 < \alpha < 2$. This idea was later developed in [2] for proving the Hermitian case and the sample covariance matrix case when $0 < \alpha < 4$ and used in [7] for proving the band Hermitian matrix case when $\alpha > 0$.

The strategy is as follows. We first show that the convergence holds when $l = 1$, i.e.,

$$\frac{\lambda_1(\Sigma_{p,n})}{|m_{11}|^2} \xrightarrow{P} 1.$$ 

Then, we remove the $i_1$-th row from $M_{p,n}$. Lemma 4.1 guarantees that, with high probability, the second largest entry will not be removed. The convergence for the second largest eigenvalue and the second largest entry follows from Theorem 4.5 and the same argument for the $l = 1$ case. Iterating this process, one proves $\frac{\lambda_l(\Sigma_{p,n})}{|m_{l1}|^2} \xrightarrow{P} 1$ for each $l$ fixed.
3.1.2. Eigenvalues. We begin by computing the two-sided tail probability of $|m_{ij}|^2$. For any $t > 0$,
\[
  \mathbb{P}(|m_{ij}|^2 \geq t) = \mathbb{P}(|m_{ij}| > \sqrt{t}) = L(\sqrt{t})n^{\mu-1}(\sqrt{t})^{-\alpha} = n^{\mu-1}L(\sqrt{t})t^{-\alpha/2}.
\]
Since $L(\sqrt{t})$ is also a slowly varying function in $t$, $[|m_{ij}|^2]_{i,j=1}^{p,n}$ is a sparse heavy tailed random matrix of $p \times n$ independent entries with parameter $\mu$ and $\alpha/2$. Classic extreme value theory tells us that the random point process $Q_n$, defined in (2), converges to the desired Poisson point process with intensity $\alpha/(2x^{1+\alpha/2})$. In particular, $c_n^{-\alpha/2}|m_{i_1j_1}|^2$ converges to a Fréchet distribution with parameter $\alpha/2$.

We next show that the largest eigenvalue of $\Sigma_{p,n}$ behaves like the square of the largest entry of $M_{p,n}$ ($l = 1$ case), i.e.,
\[
  \lambda_1(\Sigma_{p,n}) \geq \mathbf{v}^*M_{p,n}M_{p,n}^*\mathbf{v} = \sum_{j=1}^n |m_{i_1j}|^2 + \sum_{j \neq j_1}^n |m_{i_1j}|^2 \geq |m_{i_1j_1}|^2,
\]
and it suffices to prove the reverse direction, i.e., $\forall \epsilon > 0$,
\[
  (4) \quad \mathbb{P}(\lambda_1(\Sigma_{p,n}) > |m_{i_1j_1}|^2(1 + \epsilon)) \rightarrow 0, \quad \text{as } n \rightarrow \infty.
\]
We use the infinity norm of $\Sigma_{p,n}$ to bound $\lambda_1(\Sigma_{p,n})$ and truncate the matrix $M_{p,n}$, when necessary.

- **Case I:** $0 < \alpha < 1 + \mu^{-1}$.

In this case, we can directly show (4). Observing that
\[
  \lambda_1(\Sigma_{p,n}) = \|\Sigma_{p,n}\| = \|M_{p,n}M_{p,n}^*\| \leq \|M_{p,n}\|^2 \leq \|M_{p,n}\|_{\infty}\|M_{p,n}\|_1
\]
it suffices to show that
\[
  (5) \quad \mathbb{P}(\|M_{p,n}\|_{\infty} \leq |m_{i_1j_1}|(1 + o(1))) \rightarrow 1,
\]
\[
  (6) \quad \mathbb{P}(\|M_{p,n}\|_1 \leq |m_{i_1j_1}|(1 + o(1))) \rightarrow 1.
\]

The proof of (6) will be almost identical to (5), by switching the role of $p$ and $n$. We hence show (5) only. We will assume that $\mu > 0$ and mention the modifications needed to treat the case $\mu = 0$ in Remark 7, which is easier. Lemma 4.1 (a) says, with probability going to 1, there is no row that has two entries with absolute value greater than $c_{np}^\kappa$, where $\kappa = \frac{1+2\mu}{2\gamma+2\mu} + \delta$, and $\delta > 0$ can be chosen arbitrarily small. Consider the
following summation and break it into three pieces,

\[ \tilde{S}_i := \sum_{j=1}^{n} |m_{ij}| 1_{\{|m_{ij}| \leq c_{np}^i\}} \]

\[ = \sum_{j=1}^{n} |m_{ij}| 1_{\{|m_{ij}| \leq n^{\frac{\alpha}{2}}-\eta\}} + \sum_{j=1}^{n} |m_{ij}| 1_{\{|n^{\frac{\alpha}{2}}-\eta} < |m_{ij}| \leq n^{\frac{\alpha}{2}}+\eta\}} + \sum_{j=1}^{n} |m_{ij}| 1_{\{|n^{\frac{\alpha}{2}}+\eta} < |m_{ij}| \leq c_{np}^i\}} \]

\[ = \tilde{S}_{i,1} + \tilde{S}_{i,2} + \tilde{S}_{i,3}, \]

where we choose \( \eta \in (0, \min\{\frac{1}{2\alpha(\alpha+1)}, \frac{\mu}{\alpha}\}) \).

By Lemma 4.2 (a), w.e.h.p., for any \( \epsilon > 0 \), \( \tilde{S}_{i,1} \leq n^{\mu+1}(\frac{\mu}{\alpha}-\eta)(1-\alpha)^+ + \epsilon \), which is \( o(n^{\frac{\mu+1}{\alpha}}) \) if we choose \( \epsilon \) small enough. To see this, if \( \alpha < 1 \), set \( \epsilon = \eta(1-\alpha) > 0 \), and w.e.h.p.,

\[ \tilde{S}_{i,1} \leq n^{\mu+1}(\frac{\mu}{\alpha}-\eta)(1-\alpha)^+ = n^{\frac{\mu}{\alpha}}. \]

If \( 1 \leq \alpha < 1 + \mu^{-1} \), then \( (1-\alpha)^+ = 0 \) and hence \( \tilde{S}_{i,1} \leq n^{\mu+\epsilon} \). But \( \alpha < 1 + \mu^{-1} \) implies \( \mu < \frac{\mu+1}{\alpha} \), so for \( \epsilon = \frac{1}{2}(\frac{\mu+1}{\alpha}-\mu) > 0 \), we have \( \tilde{S}_{i,1} \leq n^{\frac{\mu+1}{\alpha}-\epsilon} \).

By Lemma 4.2 (c), w.e.h.p., \( \tilde{S}_{i,2} \leq n^{\mu/\alpha+\epsilon}, \forall \epsilon > \eta(\alpha+1) \). Since \( \eta < \frac{1}{2\alpha(\alpha+1)} \), we can choose \( \epsilon = \frac{1}{2\alpha}, \) and this gives w.e.h.p., \( \tilde{S}_{i,2} \leq n^{\frac{\mu+1}{\alpha}-\frac{1}{2\alpha}} \).

Finally, since \( \frac{1}{2} + \delta \leq \kappa \leq \frac{3}{4} + \delta < 1 \), \( c_{np} \sim n^{\frac{\mu+1}{\alpha}} \), and by Lemma 4.2 (d), we can choose \( \frac{(\mu+1)\kappa}{\alpha} < \gamma < \frac{\mu+1}{\alpha} \) such that \( \tilde{S}_{i,3} \leq n^{\gamma} = o(n^{\frac{\mu+1}{\alpha}-\epsilon_0}) \), for small \( \epsilon_0 > 0 \).

Combining the three terms, we have w.e.h.p., \( \tilde{S}_i = o(n^{\frac{\mu+1}{\alpha}-\epsilon_0}) \), for some small \( \epsilon_0 > 0 \).

The sum of absolute values in row \( i \) of \( M_{p,n} \) can be written as

\[ S_i := \sum_{j=1}^{n} |m_{ij}| = \tilde{S}_i + \sum_{j=1}^{n} |m_{ij}| 1_{\{|m_{ij}| \geq c_{np}^i\}}. \]

Moreover, a crude union bound and Lemma 4.1(a) give us

\[ \mathbb{P} \left( \exists i, S_i - \max_{1 \leq j \leq n} |m_{ij}| > c_{np}^{1-\epsilon_0} \right) \rightarrow 0 \]

for some \( \epsilon_0 > 0 \) sufficiently small, which implies (5). Hence we have proved that

\[ \frac{\lambda_1(\Sigma_{p,n})}{|m_{i,j}|^2} \rightarrow 1. \]

Remark 7. In the case \( \mu = 0 \), Lemma 4.2 (a) still holds, and we can deal with the terms in the summation \( \tilde{S}_i \) all at once. To see this, note, in this case, \( c_{np} \sim n^{1/\alpha} \). Since \( \frac{1}{2} + \delta \leq \kappa \leq \frac{3}{4} + \delta \), we can choose \( \frac{\kappa}{\alpha} < \gamma < \frac{1}{\alpha} - 2\epsilon \), for some \( \epsilon > 0 \). Then, w.e.h.p.,

\[ \tilde{S}_i = \sum_{j=1}^{n} |m_{ij}| 1_{\{0 < |m_{ij}| \leq c_{np}^i\}} \leq \sum_{j=1}^{n} |m_{ij}| 1_{\{0 < |m_{ij}| \leq n^{\gamma}\}} \leq n^{\gamma+\epsilon} = o(n^{1/\alpha}), \]

where the last inequality is due to Lemma 4.2(e).
On the other hand, by the interlacing of eigenvalues, Theorem 4.6, we know that

\[ \lambda_l(\Sigma_{p,n}) \geq |m_{i,j,l}|^2 + o(c_{np}^2). \]

Next, we show that, with probability tending to one, \( \Sigma_{p,n} \) has eigenvalues at \( |m_{i,j,l}|^2 + o(c_{np}^2) \). We compute the \( l \)-th residual \( \vec{r}_l \), for \( l \geq 1 \), i.e.,

\[ \Sigma_{p,n} \vec{e}_{i_l} = |m_{i,j,l}|^2 \vec{e}_{i_l} + \vec{r}_l, \]

and hence

\[ \vec{r}_l = \left( \sum_{k=1}^{n} m_{1,k}m_{i,k}, \ldots, \sum_{k=1}^{n} |m_{i,k}|^2, \ldots, \sum_{k=1}^{n} m_{p,k}m_{i,k} \right)^T. \]

The norm of \( \vec{r}_l \) is \( o(c_{np}^2) \). To see this, we compute the \( ||\vec{r}_l|| \) explicitly:

\[
||\vec{r}_l||^2 = \left( \sum_{s=1}^{p} \left[ \sum_{k=1}^{n} m_{s,k}m_{i,k} \right]^2 \right) + \left[ \sum_{k=1}^{n} |m_{i,k}|^2 \right]^{1/2} \leq \sum_{k=1}^{n} |m_{i,k}|^2 + \sum_{k=1}^{n} |m_{i,k}|^2.
\]

By Lemma 4.1 and 4.2 (a), (c) and (d), and a calculation similar to that when we bound the row sum of \( M_{p,n} \), one can see that for some \( \epsilon_0 > 0 \), \( ||\vec{r}_l|| \leq c_{np}^{2-\epsilon_0} \), with high probability. Hence, we have

\[ c_{np}^{-2} \Sigma_{p,n} \vec{e}_{i_l} = c_{np}^{-2} |m_{i,j,l}|^2 \vec{e}_{i_l} + c_{np}^{-2} \vec{r}_l, \]

where \( ||c_{np}^{-2} \vec{r}_l|| \to 0 \). It then follows from Theorem 4.6 that \( \Sigma_{p,n} \) has eigenvalues \( |m_{i,j,l}|^2 + o(c_{np}^2) \). Hence, with probability tending to one,

\[ \lambda_l(\Sigma_{p,n}) \geq |m_{i,j,l}|^2 + o(c_{np}^2). \]

To show that these are exactly the largest eigenvalues (where the case \( l = 1 \) is proved), we use Theorem 4.5. When \( l = 2 \), let \( M_{p,n,-i} \) be the submatrix of \( M_{p,n} \) removing the \( i \)-th row and let \( \Sigma_{p,n}^{(i)} := M_{p,n,-i} \) \( M_{p,n,-i}^* \). By Lemma 4.1(a), with probability going to one, the second largest entry of \( M_{p,n} \) (in absolute value), \( m_{i,j,2} \), will remain in \( M_{p,n,-i} \). Repeating the computation above for \( \Sigma_{p,n}^{(i)} \), we have, with probability tending to one,

\[ \lambda_1(\Sigma_{p,n}^{(i)}) = |m_{i,j,2}|^2 (1 + o(1)), \]

On the other hand, by the interlacing of eigenvalues, Theorem 4.6, we know that \( \lambda_2(\Sigma_{p,n}) \leq \lambda_1(\Sigma_{p,n}^{(i)}) \), which establishes

\[ \lambda_2(\Sigma_{p,n}) = |m_{i,j,2}|^2 (1 + o(1)). \]

The claim for general \( \lambda_l(\Sigma_{p,n}) \) then follows from iterating the above argument.
• Case II: \(1 + \mu^{-1} \leq \alpha < 2(1 + \mu^{-1})\).

For this case, in order to show (4), we choose \(\gamma, \gamma' > 0\) such that

\[
0 \leq \frac{\mu}{\alpha} - \frac{1}{\alpha(\alpha - 1)} < \gamma < \frac{\mu + 1}{\alpha}, \quad \max \left( \gamma, \frac{\mu}{2} \right) < \gamma' < \frac{\mu + 1}{\alpha},
\]

which is always possible if \(1 + \mu^{-1} \leq \alpha < 2(1 + \mu^{-1})\). We truncate the entries of \(M_{p,n}\) at \(n^\gamma\). Let \(\hat{M}_{p,n} := \hat{m}_{ij}^{p,n}_{i,j=1} = [m_{ij}1\{m_{ij} \leq n^\gamma\}]_{i,j=1}^{p,n}\), and \(\tilde{M}_{p,n} = M_{p,n} - \hat{M}_{p,n}\) be the truncated part and the remaining part of \(M_{p,n}\), respectively.

We decompose \(\Sigma_{p,n}\) as below:

\[
\Sigma_{p,n} = (\hat{M}_{p,n} + \tilde{M}_{p,n})(\hat{M}_{p,n} + \tilde{M}_{p,n})^* = (\hat{M}_{p,n}\hat{M}_{p,n}^*) + (\hat{M}_{p,n}\tilde{M}_{p,n}^* + \tilde{M}_{p,n}\hat{M}_{p,n}^* + \hat{M}_{p,n}\tilde{M}_{p,n}^*) \quad := \hat{\Sigma}_{p,n} + \Sigma'_{p,n}.
\]

Using triangular inequality, we have

\[
\lambda_1(\Sigma_{p,n}) = ||\Sigma_{p,n}|| \leq ||\hat{M}_{p,n} + \tilde{M}_{p,n}||^2 \leq ||\hat{M}_{p,n}|| + ||\tilde{M}_{p,n}||^2 \leq ||\hat{\Sigma}_{p,n}|| + 2||\Sigma'_{p,n}||^{1/2}(||\hat{M}_{p,n}||_1||\tilde{M}_{p,n}||_{\infty})^{1/2} + ||\tilde{M}_{p,n}||_1||\tilde{M}_{p,n}||_{\infty}.
\]

Hence, we will prove, with probability tending to one,

(8) \[||\hat{\Sigma}_{p,n}|| = o(c_{np}^2)\]

(9) \[||\tilde{M}_{p,n}||_{\infty} \leq |m_{i_1j_1}|(1 + o(1)), \quad ||\hat{M}_{p,n}||_1 \leq |m_{i_1j_1}|(1 + o(1))\]

which gives (4). For (8), first, one can deduce from the lower bound of \(\gamma\) that \(\mu + \gamma(1 - \alpha) < \frac{\mu + 1}{\alpha}\), and hence,

\[
||\hat{M}_{p,n}|| - ||\hat{M}_{p,n} - \hat{m}_{ij}|| \leq \sqrt{np} \hat{m}_{ij} \leq CL(n^\gamma)n^{\mu + \gamma(1 - \alpha)} = o(c_{np}).
\]

So we may assume that the truncated entries are centered. Here, the first inequality is a consequence of [3, Theorem A.46] and the second inequality is due to [2, Lemma 13]. Theorem 4.4 indicates that

\[
\mathbb{P}(||\hat{\Sigma}_{p,n}|| \geq Cn^{2\gamma'}) \to 0,
\]

Now with \(\gamma'\) chosen such that \(\frac{\gamma'}{2} < \gamma' < \frac{\mu + 1}{\alpha}\), (8) is proved.

For (9), again, we only show the upper bound for the infinity norm. As in the previous case, it is enough to show that for any \(1 \leq i \leq n\) fixed, w.e.h.p.,

\[
\tilde{S}_i := \sum_{j=1}^{n} |m_{ij}|1\{n^\gamma < |m_{ij}| \leq c_{np}^*\} = o(n^{\frac{\mu + 1}{\alpha}}).
\]
We treat $\tilde{S}_i$ similarly:

$$\tilde{S}_i = \sum_{j=1}^{n} |m_{ij}|1_{\{n^{-1}|m_{ij}|\leq n^{-1}\gamma - \eta\}} + \sum_{j=1}^{n} |m_{ij}|1_{\{n^{-1}\gamma - \eta < |m_{ij}|\leq n^{-1}\gamma + \eta\}}$$

$$+ \sum_{j=1}^{n} |m_{ij}|1_{\{n^{-1}\gamma + \eta \leq |m_{ij}| < \epsilon_{np}\}} : = \tilde{S}_{i,1} + \tilde{S}_{i,2} + \tilde{S}_{i,3}.$$ 

Here, the only difference from the previous case is the $\tilde{S}_{i,1}$ term. By Lemma 4.2 (b), for arbitrarily small $\delta > 0$, w.e.h.p., $\tilde{S}_{i,1} \leq n^{\mu - \gamma(\alpha - 1) + \delta}$. The choice of $\gamma > \frac{\mu}{\alpha} - \frac{1}{\alpha(\alpha - 1)}$ guarantees $\tilde{S}_{i,1} = o(n^{\frac{\mu - 1}{\alpha}})$.

The part applying Cauchy interlacing theorem is identical to Case 1, so we remain to show that with probability going to one, the norm of $\tilde{r}_l$, as defined in (7) is of smaller order with respect to $c_{np}^2$, as $n \to \infty$. We estimate $||\tilde{r}_l||$ using the triangular inequality and the decomposition of $\Sigma_{p,n}$ above:

$$||\tilde{r}_l|| \leq ||\tilde{\Sigma}_{p,n}|| + ||\tilde{M}_{p,n}\tilde{M}_{p,n}^*\tilde{e}_i|| + ||\tilde{M}_{p,n}\tilde{M}_{p,n}^*\tilde{e}_i - |m_{ij}|^2\tilde{e}_i||$$

$$\leq ||\tilde{\Sigma}_{p,n}|| + 2||\tilde{\Sigma}_{p,n}||^{1/2}(||\tilde{M}_{p,n}||_1||\tilde{M}_{p,n}||_\infty)^{1/2} + ||\tilde{M}_{p,n}\tilde{M}_{p,n}^*\tilde{e}_i - |m_{ij}|^2\tilde{e}_i||.$$ 

In view of (8) and (9), we remain to show that with probability going to one,

$$||\tilde{M}_{p,n}\tilde{M}_{p,n}^*\tilde{e}_i - |m_{ij}|^2\tilde{e}_i|| = o(c_{np}^2).$$

We compute the left side directly, which yields

$$||\tilde{M}_{p,n}\tilde{M}_{p,n}^*\tilde{e}_i - |m_{ij}|^2\tilde{e}_i|| \leq \sum_{k=1}^{n} |m_{ik}|1_{\{|m_{ik}| > n^{\gamma}\}} \sum_{s=1}^{p} |m_{sk}|1_{\{|m_{sk}| > n^{\gamma}\}}$$

$$+ |m_{ij}| \sum_{s=1}^{p} |m_{sj}|1_{\{|m_{sj}| > n^{\gamma}\}} + \left( \sum_{k=1}^{n} |m_{ik}|1_{\{|m_{ik}| > n^{\gamma}\}} \right)^2.$$ 

Using Lemma 4.1 and 4.2 (b), (c) and (d), we see that each summation above is $o(c_{np})$ with probability tending to one and hence (11) is proved. The proof is complete.

3.1.3. Eigenvectors.

In this section we show that for each $l \geq 1$,

$$||\tilde{r}_l(\Sigma_{p,n}) - \tilde{e}_i|| \xrightarrow{P} 0.$$

Let

$$\zeta := \langle \tilde{e}_i, c_{np}^{-2}\Sigma_{p,n}\tilde{e}_i \rangle = c_{np}^{-2} \sum_{j=1}^{n} |m_{ij}|^2,$$

and, for $z \in \mathbb{R}$, write $B(z, \rho) = \{ y \in \mathbb{R} : |y - z| < \rho \}$. 

Lemma 3.1. Fix \( l \geq 1 \). For any \( \epsilon > 0 \), there exists \( \delta > 0 \) and \( n_0 \in \mathbb{N} \) so that for any \( n \geq n_0 \)
\[
\mathbb{P} \left( \lambda_l(\Sigma_{p,n}) c_{np}^{-2} \text{ is the only eigenvalue of } c_{np}^{-2} \Sigma_{p,n} \text{ in } \mathbb{B}(\zeta, \delta) \right) > 1 - \epsilon.
\]

Proof. To prove the lemma, it suffices to show that for each \( k \geq 1 \), the spacing of the eigenvalues satisfies
\[
\lim_{\delta \to 0} \limsup_{n \to \infty} \mathbb{P} \left( \frac{\lambda_k(\Sigma_{p,n}) - \lambda_{k+1}(\Sigma_{p,n})}{c_{np}^2} < \delta \right) = 0.
\]
Since we have proved in the first part of Theorem 2.1 that \( \lambda_k(\Sigma_{p,n}) / |m_{ik,jk}|^2 \xrightarrow{P} 1 \), this is equivalent to
\[
\lim_{\delta \to 0} \limsup_{n \to \infty} \mathbb{P} \left( c_{np}^{-2} |m_{ik,jk}|^2 - c_{np}^{-2} |m_{ik+1,jk+1}|^2 < \delta \right) = 0.
\]
However, this follows from the fact that \( Q_n = \sum_{i=1}^{p} \sum_{j=1}^{n} \delta c_{np}^{-2} |m_{ij}|^2 \) converges to a Poisson point process on \((0, +\infty)\).

Proof of (12). The proof will follow from part (b) of Theorem 4.6. Define \( \epsilon_n = c_{np}^{-2} |(\Sigma_{p,n} - \zeta) \vec{v}| |. Then,
\[
\epsilon_n \leq c_{np}^{-2} |(\Sigma_{p,n} - |m_{ij}|^2) \vec{e}_i || + c_{np}^{-2} ||(\zeta - |m_{ij}|^2) \vec{e}_i ||
\leq c_{np}^{-2} |\vec{v}|^2 + c_{np}^{-2} \sum_{j=1,j \neq i}^{n} |m_{ij}|^2.
\]
By the calculations done in the last section (see (10) and (11)), we know \( \epsilon_n \xrightarrow{P} 0 \). Now, fix \( \epsilon > 0 \). Take \( \delta > 0 \) and \( n_0 \in \mathbb{N} \) coming from Lemma 3.1. If needed, make \( n_0 \) larger so that for any \( n > n_0 \),
\[
\mathbb{P}(\delta > \epsilon_n) > 1 - \epsilon \quad \text{and} \quad \mathbb{P}(2\epsilon_n/(\delta - \epsilon_n) < \epsilon) > 1 - \epsilon.
\]
If we let \( r = \epsilon_n \) in part (b) of Theorem 4.6, and if \( E_n \) is the event in the statement of Lemma 3.1, we have for any \( n > n_0 \),
\[
\mathbb{P} \left( \| \vec{v}_l(\Sigma_{p,n}) - \vec{e}_i \| \leq r \right) \geq \mathbb{P} \left( E_n, \frac{2\epsilon_n}{\delta - \epsilon_n} \leq \epsilon, \delta > \epsilon_n \right) > 1 - 3\epsilon,
\]
which leads to (12).

3.2. Proof of Theorem 2.2.
3.2.1. Eigenvalues. Proposition 2.3 implies that for any \( k \geq 1 \) fixed, and any \( \epsilon > 0 \)
\[
\mathbb{P}(\lambda_k(\Sigma_{p,n}) \geq (1 + \sqrt{\rho})^2(1 - \epsilon)n^\mu) \to 1, \quad \text{as } n \to \infty.
\]
It remains to prove the upper bound, i.e., \( \mathbb{P}(\lambda_k(\Sigma_{p,n}) \leq (1 + \sqrt{\rho})^2(1 + \epsilon)n^\mu) \to 1. \)
Since \( \lambda_k(\Sigma_{p,n}) \leq \lambda_1(\Sigma_{p,n}) \) and since we can decompose \( \Sigma_{p,n} \) in the same way as in the previous case, it is enough to show that
\[
\mathbb{P}\left(||\hat{\Sigma}_{p,n}||+2||\hat{\Sigma}_{p,n}||^{1/2}(||M_{p,n}||_1||M_{p,n}||_\infty^{1/2}+||\bar{M}_{p,n}||_1||\bar{M}_{p,n}||_\infty) > (1+\sqrt{\rho})^2(1+\epsilon)n^\mu\right)
\]
go to zero. In this regime, we choose \( \gamma' = \mu/2 \) and \( \gamma \in \left(\frac{\mu}{2(\alpha-1)}, \frac{\mu}{2}\right) \), which is always possible when \( \alpha > 2 \). Such \( \gamma \) and \( \gamma' \) satisfy the assumptions in Theorem 4.4, which gives the bound for the truncated part, i.e.,
\[
\mathbb{P}(||\hat{\Sigma}_{p,n}|| \geq (1 + \sqrt{\rho})^2(1 + \epsilon)n^\mu) \to 0.
\]
We remain to show that for some \( \epsilon_0 > 0 \), \( ||M_{p,n}||_\infty \leq n^{\mu/2-\epsilon_0} \) and \( ||\bar{M}_{p,n}||_1 \leq n^{\mu/2-\epsilon_0} \) with high probability. Again, we prove for the infinity norm only, i.e.,
\[
(13) \quad \mathbb{P}(S_i \leq n^{\mu/2-\epsilon_0}, \text{ for all } 1 \leq i \leq p) \to 1,
\]
where
\[
S_i := \sum_{j=1}^{n} |m_{ij}|1_{\{|m_{ij}| > n^\gamma\}}.
\]
Since \( c_{np}^\text{sl} \sim n^{(1+\mu)/\alpha} \) and \( c_{np}^{-2}|m_{i1}|^2 \) converges in distribution, then for any \( \theta > \frac{\mu+1}{\alpha} \),
\[
\max_{1 \leq i,j \leq n} |m_{ij}| \leq n^{\theta} \quad \text{and}
\]
\[
S_i = \sum_{j=1}^{n} |m_{ij}|1_{\{|m_{ij}| \leq n^{\mu}\}} + \sum_{j=1}^{n} |m_{ij}|1_{\{|m_{ij}| \leq n^{\mu+1}\}} + \sum_{j=1}^{n} |m_{ij}|1_{\{|m_{ij}| \leq n^{\mu+1}\}}
\]
\[
:= S_{i,1} + S_{i,2} + S_{i,3}
\]
with high probability. By Lemma 4.2 (b), (c) and (d), respectively, for any \( \epsilon > 0 \), we have w.e.h.p.,
\[
S_{i,1} \leq n^{\mu-\gamma(\alpha-1)+\epsilon}, \quad S_{i,2} \leq n^{\mu+1-\alpha+\epsilon}, \quad S_{i,3} \leq n^{\theta+\epsilon}.
\]
As \( \alpha > 2(1+\mu^{-1}) \) we have \( \frac{\mu+1}{\alpha} < \frac{\mu}{2} \). We can choose \( \theta \) arbitrarily close to \( \frac{\mu+1}{\alpha} \) and \( \epsilon > 0 \) small enough such that for some \( \epsilon_0 > 0 \) w.e.h.p., \( S_{i,2} + S_{i,3} \leq \frac{1}{2}n^{\mu/2-\epsilon_0} \). Moreover, the choice of \( \gamma > \frac{\mu}{2(\alpha-1)} \) implies \( \mu - \gamma(\alpha-1) < \frac{\mu}{2} \). By making \( \epsilon \) even small, we get w.e.h.p., \( S_{i,1} \leq \frac{1}{2}n^{\mu/2-\epsilon_0} \). Therefore, by a union bound, we obtain (13).
3.2.2. Eigenvectors. Last, we prove the localization result of the eigenvectors of $\Sigma_{p,n}$.
We use the following simple linear algebra lemma [7, Lemma 4.2] that we quote without proof.

Lemma 3.2. Let $H$ be a Hermitian matrix and $\rho_L(H)$ be the maximum spectral radius of its $L \times L$ principal sub-matrix. Let $\lambda$ be an eigenvalue of $H$ and $\vec{v}$ an associated unit eigenvector. If $\vec{v}$ is $(L, \eta)$-localized, then

$$|\lambda| \leq \frac{\rho_L(H) + \sqrt{\eta} \| H \|}{\sqrt{1 - \eta}}.$$ 

If $\alpha > 2(1 + \mu^{-1})$ then, by Theorem 4.4, we know that $\| \Sigma_{p,n} \|$ is of order $(1 + \sqrt{\rho})^2 n^\mu$.

In view of Lemma 3.2, it suffices to show that there exists $\eta > 0$ such that with probability going to one

$$\rho_{|p^\beta|} (\Sigma_{p,n}) < (\sqrt{1 - \eta} - \sqrt{\eta}) (1 + \sqrt{\rho})^2 n^\mu.$$ 

In other words, we must establish that there exist $\epsilon > 0$ such that with probability going to one any $[p^\beta] \times [p^\beta]$ principal sub-matrix $W$ of $\Sigma_{n,p}$ satisfies the inequality $\| W \| \leq (1 + \epsilon) (1 + \sqrt{\rho})^2 n^\mu$.

We proceed as follows. A principal sub-matrix $W$ is obtained by choosing $[p^\beta]$ rows $i_1, \ldots, i_{|p^\beta|}$ of the rectangular matrix $M_{p,n}$ and writing $W = M_I M_I^\dag$, where $I = \{i_1, \ldots, i_{|p^\beta|}\}$. Here, the notation $M_I$ stands for the $[p^\beta] \times n$ sub-matrix of $M$ formed by the rows with indices in $I$. As before, we write $\tilde{M}_I$ and $\tilde{M}_I^\dag$ for the truncation and remainder of the matrix $M_I$ at level $n^\gamma$, for $\gamma \in (\frac{\mu}{2(\alpha-1)}, \frac{\mu}{2})$. Then

$$\| W \| \leq \| \tilde{M}_I \tilde{M}_I^\dag \| + 2 \| \tilde{M}_I \tilde{M}_I^\dag \|^{1/2} (\| M_I \|_1 \| \tilde{M}_I \|_\infty)^{1/2} + \| M_I \tilde{M}_I^\dag \|.$$ 

For any choice of $I$, $\| M_I \tilde{M}_I^\dag \| \leq \| M_{p,n} \|_1 \| M_{p,n} \|_\infty = o(n^\mu)$ as in the proof in Section 3.2.1. On the other hand, for any choice of $I$, one can adapt the proof of Theorem 4.4 to deal with the case of $\hat{p} = [p^\beta]$ rows to show that for any $1 < c < (1 + \sqrt{\rho})^2$, there exists $\theta = \theta(c) > 0$ and $\gamma' = \gamma'(c) > 0$ so that

$$\mathbb{P} \left( \| \tilde{M}_I \tilde{M}_I^\dag \| \geq c n^\mu \right) \leq P(n)n^{-\theta[n^{\gamma'}]},$$ 

where $P$ is a polynomial in $n$. Indeed, in the case where $\hat{p} \to \infty$ and $\hat{p}/n \to 0$ one needs to control the appearance of odd and even innovations (or odd/even marked vertices as in [21, Section 2.2]). We leave the details to the reader. Since there are at most $n^{p^\beta}$ ways to choose the indices in $I$, the probability of the existence of such a principal sub-matrix is bounded above by:

$$P(n)n^{-\theta[n^{\gamma'}]} n^{(2\rho n)^\beta}.$$ 

Thus if we choose $\beta < \gamma'$, we obtain the desired result.
4. SOME USEFUL LEMMAS

In this section we collect some tools and lemmas that are used throughout the proofs of the main results. The proof of the first lemma is identical to the one given in [2, Lemma 5].

Lemma 4.1. Suppose $M_{p,n}$ is the $p \times n$ rectangular, sparse, heavy tailed matrix. Let $c_{np}$ be as given in (1). Then, for all values of $\alpha > 0$ and $\eta > 0$, we have:

(a) $P \left\{ \exists i, \exists j_1 \neq j_2, 1 \leq i \leq p, 1 \leq j_1, j_2 \leq n : \min(|m_{ij_1}|, |m_{ij_2}|) > \frac{c_{np}^{1+2\eta}}{2} \right\} \to 0.$

(b) $P \left\{ \exists j, \exists i_1 \neq i_2, 1 \leq i_1, i_2 \leq p, 1 \leq j \leq n : \min(|m_{i_1j}|, |m_{i_2j}|) > \frac{c_{np}^{1+2\eta}}{2} \right\} \to 0.$

The next lemma is used to control the sum of absolute values within a given row or a given column of $M_{p,n}$. It is modified from [7, Proposition A.6]. The main difference is that in [7, Proposition A.6], because each row has asymptotically (up to a slowly varying function) $n \mu$ nonzero entries (nonrandom), the summation in each part ran through only the $n \mu$ nonzero terms, whereas in our setting, the number of nonzero terms in each row and column and their positions are random and hence we include every term in a row or column.

Lemma 4.2. Let $M_{p,n}$ be the sparse, heavy tailed, rectangular random matrix.

(a) For any sequence $\beta_n \sim n^b$, where $0 \leq b \leq \frac{\mu}{\alpha}$, and any $\epsilon > 0$, then w.e.h.p.,

$$\sum_{j=1}^{n} |m_{ij}| 1_{\{|m_{ij}| < \beta_n\}} \leq n^{\mu+b(1-\alpha)^++\epsilon},$$

where $(1 - \alpha)^+ := \max\{1 - \alpha, 0\}$.

(b) If $\alpha > 1$ and $\mu > 0$, then for any sequences $\alpha_n \sim n^a$ and $\beta_n \sim n^b$ with $0 \leq a < b \leq \frac{\mu}{\alpha}$, and any $\epsilon > 0$, we have, w.e.h.p.,

$$\sum_{j=1}^{n} |m_{ij}| 1_{\{\alpha_n < |m_{ij}| \leq \beta_n\}} \leq n^{\mu-(a-1)a+\epsilon}.$$

(c) If $\mu > 0$, then for any sequences $\alpha_n \sim n^{\frac{\mu}{\alpha} - \eta}$ and $\beta_n \sim n^{\frac{\mu}{\alpha} + \eta'}$ with $\eta, \eta' \geq 0$, and any $\epsilon > \alpha \eta + \eta'$, we have, w.e.h.p.,

$$\sum_{j=1}^{n} |m_{ij}| 1_{\{\alpha_n < |m_{ij}| \leq \beta_n\}} \leq n^{\frac{\mu}{\alpha} + \epsilon}.$$

(d) For any sequences $\alpha_n \sim n^{(\mu/\alpha)+\eta}$ and $\beta_n \sim n^b$ with $\eta, b > 0$, and any $\gamma > b$, we have w.e.h.p.,

$$\sum_{j=1}^{n} |m_{ij}| 1_{\{\alpha_n < |m_{ij}| \leq \beta_n\}} \leq n^\gamma.$$
(e) If $\mu = 0$ then for any $0 \leq \eta < \epsilon$ and any sequence $\beta_n \sim n^\eta$, we have, w.e.h.p.,
\[ \sum_{j=1}^{n} |m_{ij}| \mathbf{1}_{0 < |m_{ij}| \leq \beta_n} \leq n^{\epsilon}. \]
Moreover, we have the same bounds for the column sums, that is, the same results in (a)–(e) hold if we replace $\sum_{j=1}^{n}$ by $\sum_{i=1}^{n}$ in each part.

The proof of Lemma 4.2 uses the following consequence of Bennett’s inequality [9] (see also [7, Lemma A.7]).

**Lemma 4.3.** For each $n \geq 1$, let $X_1, \ldots, X_m$ be independent Bernoulli random variables with parameter $p$, where $m$ and $p$ depend on the parameter $n$. If $mp > Cn^\theta$ for some constants $C, \theta > 0$, then for any $\eta > 0$, w.e.h.p.,
\[ \left| \frac{1}{m} \sum_{i=1}^{m} X_i - p \right| \leq \eta p. \]

**Proof of Lemma 4.2.**
(a) Assume first that $\mu > 0$. Choose $\epsilon_0 \in (0, \epsilon)$ such that $b/\epsilon_0 \notin \mathbb{Z}$. Let $T = \lfloor b/\epsilon \rfloor$, then $T\epsilon_0 < b$. Hence we may choose
\[ \theta = \frac{\mu - \alpha T\epsilon_0}{2} > \frac{\mu - \alpha b}{2} \geq \frac{\mu - \mu}{2} = 0. \]
For each $k = 0, 1, \ldots, T$, define $Y_i^{(k)} := \# \{ |m_{ij}| : n^{k\epsilon_0} < |m_{ij}| \leq n^{(k+1)\epsilon_0} \}$. The summation on the left side of (14) is bounded by
\[ \sum_{j=1}^{n} |m_{ij}| \mathbf{1}_{0 < |m_{ij}| \leq 1} + \sum_{k=0}^{T} Y_i^{(k)} n^{(k+1)\epsilon_0} = (I) + (II). \]
Note that (I) is bounded by $\sum_{j=1}^{n} \mathbf{1}_{|m_{ij}| > 0}$, and by Lemma 4.3, we have, w.e.h.p.,
\[ (I) \leq \sum_{j=1}^{n} \mathbf{1}_{|m_{ij}| > 0} \leq 2n^{\mu} = o(n^{\mu + \epsilon}). \]
Moreover, $Y_i^{(k)} \leq \sum_{j=1}^{n} \mathbf{1}_{n^{k\epsilon_0} < |m_{ij}|}$, where each $\mathbf{1}_{n^{k\epsilon_0} < |m_{ij}|}$ is an independent copy of a Ber($L(n^{k\epsilon_0}) n^{\mu - 1 - \alpha k \epsilon_0}$) random variable. Since $\mu - \alpha k \epsilon_0 > \theta > 0$ for all $k = 0, \ldots, T$, we know that w.e.h.p., by Lemma 4.3,
\[ Y_i^{(k)} \leq 2L(n^{k\epsilon_0}) n^{\mu - \alpha k \epsilon_0}. \]
Thus, w.e.h.p., (II) is no more than
\[ L(n^{k\epsilon_0}) \sum_{k=0}^{T} 2n^{\mu - \alpha k \epsilon_0} n^{(k+1)\epsilon_0} = 2L(n^{k\epsilon_0}) n^{\mu + \epsilon_0} \sum_{k=0}^{T} n^{(1-\alpha)k \epsilon_0}. \]
If $0 < \alpha \leq 1$, then $n^{(1-\alpha)k\epsilon_0} \leq n^{(1-\alpha)T\epsilon_0}$. Since $b > T\epsilon_0$ and $\epsilon_0 < \epsilon$, we have

$$ (\text{II}) \leq 2L(n^{k\epsilon_0})n^{\mu+\epsilon_0}(T+1)n^{(1-\alpha)T\epsilon_0} $$

$$ \leq 2L(n^{k\epsilon_0})(T+1)n^{\mu+\epsilon_0+(1-\alpha)b} < n^{\mu+b(1-\alpha)+\epsilon}. $$

If $\alpha > 1$, then $n^{(1-\alpha)k\epsilon_0} \leq 1$. Using again $\epsilon_0 < \epsilon$, we obtain,

$$ (\text{II}) \leq 2L(n^{k\epsilon_0})n^{\mu+\epsilon_0} \cdot (T+1) \leq n^{\mu+\epsilon}. $$

Combining (15), (16) and (17), we prove Part (a) for $\mu > 0$. In the case $\mu = 0$, then $b = 0$ and the number $V_n$ of nonzero terms in the sum $\sum_{0 < |m_{ij}|} m_{ij}$ converges in distribution to a Poisson random variable of mean 1. It suffices then to bound the sum in (14) by $\beta_n V_n$ to obtain the desired result.

(b) Just like in Part (a), we choose $\epsilon_0 < \epsilon$ such that $\frac{b-a}{\epsilon_0} \notin \mathbb{Z}$, and take $T = \left[\frac{b-a}{\epsilon_0}\right]$.

Set $\theta = \frac{\mu-\alpha(a+T\epsilon_0)}{2} > \frac{\mu-a\beta}{2} \geq 0$. Define

$$ Y_i^{(k)} := \#\{|m_{ij}| : \alpha_n n^{k\epsilon_0} < |m_{ij}| \leq \alpha_n n^{(k+1)\epsilon_0}\}. $$

Then, for each $k$, $Y_i^{(k)}$ is bounded by $\#\{|m_{ij}| : \alpha_n n^{k\epsilon_0} < |m_{ij}|\}$, which is distributed as a $\text{Bin}(n, n^{\mu-aL(\alpha_n n^{k\epsilon_0})}(\alpha_n n^{k\epsilon_0})^{-\alpha})$ random variable. Again, for each $k \leq T$,

$$ n \cdot n^{\mu-1} L(\alpha_n n^{k\epsilon_0})(\alpha_n n^{k\epsilon_0})^{-\alpha} \sim n^{\mu-(a+k\epsilon_0)\alpha} > n^\theta. $$

From Lemma 4.3, w.e.h.p., $Y_i^{(k)} \leq 2n^{\mu-\alpha(a+k\epsilon_0)+\delta}$, for arbitrarily small $\delta > 0$. Then, we have w.e.h.p.,

$$ \sum_{j=1}^{n} |m_{ij}| 1_{\{\alpha_n < |m_{ij}| \leq \beta_n\}} \leq \sum_{k=0}^{T} Y_i^{(k)} \alpha_n n^{(k+1)\epsilon_0} \leq 2 \sum_{k=0}^{T} n^{\mu-\alpha(a+k\epsilon_0)+\delta} n^{a+(k+1)\epsilon_0+\delta'} $$

$$ \leq n^{\mu+\epsilon_0+\delta''-a(\alpha-1)} \sum_{k=0}^{T} n^{k\epsilon_0(1-\alpha)} \leq n^{\mu+\epsilon_0+\delta''-a(\alpha-1)(T+1)} \leq n^{\mu-a(\alpha-1)+\epsilon}. $$

(c) For any $\delta, \delta' > 0$, the left side is no larger than $n^{\mu+\eta'+\delta'} \sum_{j=1}^{n} 1_{\{n^{\mu-\eta-\delta} < |m_{ij}|\}} \sim \text{Bin}(n, n^{\mu-1} L'(n)n^{(\mu-\eta-\delta)(-\alpha)})$, for some slowly varying function $L'$. Since

$$ n \cdot n^{\mu-1} L'(n)n^{(\mu-\eta-\delta)(-\alpha)} = L'(n)n^{\alpha(\eta+\delta)} \geq n^\theta, \text{ for } \theta = \alpha\eta/2 > 0, $$

then w.e.h.p., $\sum_{j=1}^{n} 1_{\{n^{\mu-\eta-\delta} < |m_{ij}|\}} \leq 2L'(n)n^{\alpha(\eta+\delta)}$. But $\delta, \delta' > 0$ can be chosen arbitrarily small, so as long as $\epsilon > \alpha\eta + \eta'$, w.e.h.p., the left side is bounded by

$$ n^{\mu+\eta'+\delta'} \cdot 2L'(n)n^{\alpha(\eta+\delta)} = n^{\mu+\alpha\eta+\eta'}. (2L'(n)n^{\alpha\delta+\delta'}) \leq n^{\mu+\epsilon}. $$
(d) We compute this probability directly. Write $S_i := \sum_{j=1}^{n} m_{ij} \mathbf{1}_{\{m_{ij} < n\}}$. For any $\gamma > b$, and choose $\epsilon, \epsilon' > 0$ sufficiently small (e.g., $\epsilon < \alpha \eta/2$, $\epsilon' < (\gamma - b)/2$),

$$P(S_i > n^\gamma) \leq P(\text{at least } n^{\gamma}/\beta_n \text{ terms are nonzero among the } n \text{ } m_{ij} \text{'s})$$

$$\leq n^{(\gamma/\beta_n)} [P(|m_{ij}| > \alpha_n)]^{n^{\gamma}/\beta_n}$$

$$= [n \cdot n^{\mu-1} L(\alpha_n) \alpha_n^{-\alpha} n^{\gamma/\beta_n}] \leq [n^{\mu} n^{(\alpha+\eta) \gamma/\beta_n}]$$

$$\leq [\gamma^{\gamma-b-\epsilon'} n^{(\alpha \eta/2) n^{(\gamma-b)/2}}] = e^{-n^{(\gamma-b)/2}(\alpha \eta \log n)/2} \leq e^{-n^\delta}.$$

(e) Just note that

$$\sum_{j=1}^{n} |m_{ij}| \mathbf{1}_{\{0 < |m_{ij}| \leq \beta_n \}} \leq \beta_n Z_n,$$

where $Z_n := \#\{j : |m_{ij}| > 0\}$ is a Binomial random variable with parameters $(n, 1/n)$. The result of (e) then follows by estimating $P(Z_n > n^{\epsilon} \beta_n^{-1})$ using Poisson approximation.

We now prove an upper bound for the norm of the truncated sample covariance matrix. This section is the main technical step and novelty of the proof. The method used in [2] can not be adapted to our situation as the bounds would be suboptimal. See Remark 8 below.

If $1 + \mu^{-1} < \alpha < 2(1 + \mu^{-1})$, we consider $\gamma$ be any constant such that

$$\gamma > \frac{\mu}{\alpha} - \frac{1}{\alpha(\alpha - 1)}.$$

If $\alpha > 2(1 + \mu^{-1})$, we consider $\gamma$ be any constant such that

$$\gamma > \frac{\mu}{2(\alpha - 1)}.$$

**Theorem 4.4.** Suppose $\alpha > 2$ and $x$ has mean zero and variance one. Let $M_{p,n} = A_{p,n} \cdot B_{p,n}$ be the sparse $p \times n$ matrix with heavy tailed entries. Consider $\gamma$ that satisfies (18) or (19) and choose $\gamma' > \gamma, \gamma' \geq \mu/2$. Define

$$\tilde{M}_{p,n} = [\tilde{m}_{ij}]_{i,j=1}^{p,n} = [m_{ij} \mathbf{1}_{\{|m_{ij}| \leq \beta_n\}}]_{i,j=1}^{p,n}.$$

Then for any $\kappa > 1$,

$$P\left(\|\tilde{M}_{p,n} \tilde{M}_{p,n}^*\| \geq \kappa n^{2\gamma'} (1 + \sqrt{\rho})^2\right) \to 0.$$

**Remark 8.** In [2], when $\mu = 1$, the authors used bounds on a symmetrized version of $\tilde{M}_{p,n}$ to obtain upper bounds on $\|\tilde{M}_{p,n} \tilde{M}_{p,n}^*\|$. If we try to follow their method here, we would only obtain the bound $\|\tilde{M}_{p,n} \tilde{M}_{p,n}^*\| \leq 4n^{2\gamma'}$ with high probability. This does not suffice to prove our theorems. To obtain (20), we argue via the moment method, where the main combinatorial step requires careful counting and the introduction of innovations. It is similar in spirit to the work of Soshnikov [27].
Since the second term is bounded by $\epsilon > 0$. Thus random variables. With probability going to one this sum is at most $n$ as in [2] (see equations (36) and (37) therein). Let $\hat{m}_{ij}$, which goes to zero as the exponent $2\gamma$.

First we may assume that $\hat{m}_{ij}$’s are centered. To see this we proceed similarly as in [2] (see equations (36) and (37) therein). Let $\hat{M}_{p,n}^0 = \hat{M}_{p,n} - \mathbb{E}\hat{M}_{p,n}$, where the expectation is taken element-wise. The triangle inequality implies

$$
||\hat{M}_{p,n}\hat{M}_{p,n}^*|| \leq ||\hat{M}_{p,n}^0(\hat{M}_{p,n}^0)^*|| + 2||\mathbb{E}\hat{m}_{ij}|| ||\hat{M}_{p,n}^0 1_{n \times p}|| + (\mathbb{E}\hat{m}_{ij})^2 np,
$$

where $1_{n \times p}$ denotes the $n \times p$ matrix consisting of all 1’s. Note that by standard bounds on moments of heavy-tailed random variables (see Lemma 13 in [2] for instance)

$$
||\mathbb{E}\hat{m}_{ij}|| \leq \ell(n\gamma)n_{\gamma(1-\alpha)+\mu-1},
$$

where $\ell(\cdot)$ is a slowly varying function. Thus the third term in (21) is bounded by

$$
C\ell(n\gamma)^2n_{2\gamma(1-\alpha)+2\mu} = o(n^{2\gamma'}),
$$

by our choice of $\gamma$ and $\gamma'$. For the second term note that

$$
||\hat{M}_{p,n}^0 1_{n \times p}|| = \left|\sum_{ij}(\hat{m}_{ij} - \mathbb{E}\hat{m}_{ij})\right|
$$

and the right side is the absolute value of a sum of $np \sim Cn^2$ bounded, centered, i.i.d. random variables. With probability going to one this sum is at most $n^{1+\epsilon}$, for any $\epsilon > 0$. Thus

$$
\mathbb{P}\left(2||\mathbb{E}\hat{m}_{ij}|| ||\hat{M}_{p,n}^0 1_{n \times p}|| > n^{2\gamma'}\right) \leq \mathbb{P}\left(\left|\sum_{ij}(\hat{m}_{ij} - \mathbb{E}\hat{m}_{ij})\right| > \frac{n^{2\gamma'-\gamma(1-\alpha)-(\mu-1)}}{2}\right),
$$

which goes to zero as the exponent $2\gamma' - \gamma(1-\alpha) - (\mu-1) > 1$. Thus it suffices to show that (20) holds for the centered random matrix $M_{p,n}^0$. Note that when centering the random variables, we are working with the entries of the form $\hat{m}_{ij} - \mathbb{E}\hat{m}_{ij}$, and thus the upper bound $n^{\gamma}$ for $\hat{m}_{ij}$ given by the truncation may be violated. However, since

$$
||\mathbb{E}\hat{m}_{ij}|| \leq \ell(n\gamma)n_{\gamma(1-\alpha)+\mu-1} = o(n\gamma),
$$

we can slightly tune up $\gamma$ while still having $\gamma' > \gamma$. We thus proceed assuming the truncated entries $\hat{m}_{ij}$ are centered and bounded by $n\gamma$.

For $\kappa > 1$ given, we find $C \in (1, \kappa)$ and let $E_n$ be the event

$$
E_n = \{L \leq Cn^\mu, \bar{L} \leq Cpn^{\mu-1}\},
$$

where $L$ (resp. $\bar{L}$) is the maximum number of nonzero entries in a row (resp. a column) among the $p$ rows (resp. $n$ columns) of $M_{p,n}$. We break the desired probability into two parts:

$$
\mathbb{P}\left(||\hat{M}_{p,n}\hat{M}_{p,n}^*|| \geq \kappa n^{2\gamma'}(1 + \sqrt{p})^2, E_n\right) + \mathbb{P}\left(||\hat{M}_{p,n}\hat{M}_{p,n}^*|| \geq \kappa n^{2\gamma'}(1 + \sqrt{p})^2, E_n^c\right).
$$

Since the second term is bounded by $\mathbb{P}(E_n^c)$ which vanishes as $n \to \infty$ by Chernoff inequality, it suffices to prove that the first term vanishes as well. To do this, we
choose \( \gamma'' > 0 \) such that \( \gamma' > \gamma + 6\gamma'' \) and set \( k = k_n = \lceil n^{\gamma''} \rceil \). We will prove that for any \( \delta > 0 \) small,
\[
\mathbb{E}(\text{Tr}(\tilde{M}_{p,n} \tilde{M}_{p,n}^*)^k \mathbf{1}_{E_n}) \leq p \left[ (1 + \sqrt{\delta})^2 C n^{2\gamma'} (1 + \sqrt{\rho})^2 \right]^k.
\]

It will then follow from (22) that
\[
\mathbb{P}\left( ||\tilde{M}_{p,n} \tilde{M}_{p,n}^*|| \geq \kappa n^{2\gamma'} (1 + \sqrt{\rho})^2, E_n \right) \leq \frac{\mathbb{E}(\text{Tr}(\tilde{M}_{p,n} \tilde{M}_{p,n}^*)^k \mathbf{1}_{E_n})}{\kappa^{kn^{2\gamma'}(1 + \sqrt{\rho})^{2k}}} \leq p \left[ C(1 + \sqrt{\delta})^2 \right]^k.
\]

The right side goes to zero as \( n \to \infty \), if we choose \( \delta > 0 \) such that \( C(1 + \sqrt{\delta})^2 < \kappa \). To show (22), we will make use of the combinatorics that was invented in [29] to prove the convergence of the largest eigenvalue of random sample covariance matrices.

We first expand the left side of (22):
\[
\mathbb{E}(\text{Tr}(\tilde{M}_{p,n} \tilde{M}_{p,n}^*)^k \mathbf{1}_{E_n}) = \sum_{i_1, i_2, \ldots, i_{2k} = 1}^{p} \sum_{i_1, i_2, \ldots, i_{2k} = 1}^{n} \mathbb{E}(\hat{m}_{i_1 i_2} \hat{m}_{i_3 i_2} \cdots \hat{m}_{i_{2k-1} i_{2k}} \hat{m}_{i_1 i_{2k}} \mathbf{1}_{E_n}).
\]

Then, we associate each summand on the right side with an undirected graph \( G \) that has vertices \( \{i_1, \ldots, i_{2k}\} \) and edges \( \{(i_1, i_2), (i_2, i_3), \ldots, (i_{2k-1}, i_{2k}), (i_{2k}, i_1)\} \). We read the vertices sequentially, \( i_1, i_2, i_3, \ldots, i_{2k}, \) one at a time, and classify the edges into four different types. We call an edge \( (i_{s-1}, i_s), s \geq 2 \), an innovation if \( i_s \) does not occur in \( i_1, \ldots, i_{s-1} \). An innovation \( (i_{s-1}, i_s) \) is called a row innovation if \( s \) is odd and a column innovation if \( s \) is even. If two edges \( (i_{a-1}, i_a) \) and \( (i_{b-1}, i_b) \) have the same set of vertices, i.e., \( \{i_{a-1}, i_a\} = \{i_{b-1}, i_b\} \), we say that they coincide. And an edge \( (i_{a-1}, i_a) \) is said to be single up to \( i_b \), with \( b \geq a \), if there is no other edge \( (i_{c-1}, i_c) \) with \( 2 \leq c \leq b \) that coincides with \( (i_{a-1}, i_a) \). For \( b \geq 3 \), we call \( (i_{b-1}, i_b) \) a \( T_3 \)-edge if there is an innovation \( (i_{a-1}, i_a) \), \( a < b \), that is single up to \( i_{b-1} \) and coincides with \( (i_{b-1}, i_b) \). And finally, an edge is called a \( T_4 \)-edge if it is neither a \( T_3 \)-edge nor an innovation. Hence, observing \( \hat{m}_{ij} = \hat{a}_{ij} \mathbf{1}_{\{\{i_j\} \leq n\gamma\}}, \hat{b}_{ij} = \hat{a}_{ij} \) and using independence, the expectation can be rewritten as
\[
\mathbb{E}(\text{Tr}(\tilde{M}_{p,n} \tilde{M}_{p,n}^*)^k \mathbf{1}_{E_n}) = \sum' \sum'' \sum''' \mathbb{E}(\hat{m}_{i_1 i_2} \hat{m}_{i_3 i_2} \cdots \hat{m}_{i_{2k-1} i_{2k}} \hat{m}_{i_1 i_{2k}} \mathbf{1}_{E_n})
\]

where \( \sum' \) sums over all possible arrangements of the four types of the edges, \( \sum'' \) is to count the total number of different canonical graphs given the arrangements of the four types of edges, and \( \sum''' \) runs through all graphs that are isomorphic to the given canonical graph.

Let \( l \) be the number of \( T_3 \)-edges. Note \( l \) is also the number of innovations since every edge must be visited at least twice, and hence \( (2k - 2l) \) is the number of \( T_4 \)-edges. Let \( r \) be the number of row innovations. We see that \( \sum' \) is bounded by \( \sum_{l=1}^{k} \sum_{r=0}^{l} \binom{k}{l-r} \binom{2k-l}{r} \). Since every row innovation \( (i_{2s-2}, i_{2s-1}) \) leads to a new vertex \( i_{2s-1} \in \{1, 2, \ldots, p\} \) and every column innovation \( (i_{2s-1}, i_{2s}) \) leads to a new vertex \( i_{2s} \in \{1, 2, \ldots, n\} \), except for the first innovation \( (i_1, i_2) \), which leads to both a
new vertex $i_1 \in \{1, 2, \ldots, p\}$ and a new vertex $i_2 \in \{1, 2, \ldots, n\}$, then, on the event $E_n$, there are at most $p(Cpn^{\mu - 1})^r(Cn^{\mu})^{l-r}$ terms that have nonzero contributions to $\sum^{\mu}_t$. Let $q$ be the number of distinct $T_4$-edges. It was shown in [29, page 519] that $\sum^{\mu}$ is bounded by $k^{2q}(q + 1)^{6k - 6l}$.

Finally, let $b$ be the number of $T_4$-edges among the $q$ distinct ones that coincide with some innovations, and let $n_s$, $s = 1, 2, \ldots, b$, be the multiplicity of the $T_4$-edges of the $s$-th such coincidence. Then, $(q - b)$ distinct $T_4$-edges do not coincide with any innovations but only among the $T_4$-edges, and we denote by $m_t, t = 1, 2, \ldots, (q - b)$ the multiplicity of the $t$-th such coincidence. These numbers have to satisfy the relation $2k - 2l = \sum_{s=1}^{b} n_s + \sum_{t=1}^{q-b} m_t$. Hence, for such composition of four types of edges, we can write

$$\mathbb{E}(\hat{a}_{i_1i_2}\hat{a}_{i_3i_4}\cdots\hat{a}_{i_{2k-1}i_{2k}}\hat{a}_{i_1i_2})$$

$$= (\mathbb{E}^{2}_{11})^{l-b} \prod_{s=1}^{b} (\mathbb{E}^{n_s+2}_{11})^{q-b} \prod_{t=1}^{q-b} (\mathbb{E}^{m_t}_{11})$$

$$\leq L_0(n^\gamma)^q \prod_{s=1}^{b} \left(\frac{n_s + 2}{n_s + 2 - \alpha}\right) n^{\gamma(n_s+2-\alpha)} \prod_{t=1}^{q-b} \left(\frac{m_t}{m_t - \alpha}\right) n^{\gamma(m_t-\alpha)},$$

where $L_0$ is a slowly varying function. Here, in the last inequality, we have used the following classic fact for the moments of truncated, heavy tailed random variables (see, e.g., [10, Proposition 1.5.8] or [7, Lemma A.8]),

$$\mathbb{E}|a_{11}|^s 1_{|a_{11}| \leq x} = \begin{cases} L_0(x), & \text{if } s \leq \alpha \\ L_0(x) \frac{s}{s-\alpha} x^{s-\alpha}, & \text{if } s > \alpha \end{cases}.$$  

Observing that (i) $n_s \leq 2k - 2$, $2 \leq m_t \leq 2k - 2$, (ii) $\frac{1}{m-\alpha}$ is bounded above, say, by some $C_\alpha$, which only depends on $\alpha$, for all $m > \alpha$, and $m \in \mathbb{Z}$, and (iii) $\alpha > 2$, we have

$$\mathbb{E}(\hat{a}_{i_1i_2}\hat{a}_{i_3i_4}\cdots\hat{a}_{i_{2k-1}i_{2k}}\hat{a}_{i_1i_2}) \leq L_0(n^\gamma)^q (2C_\alpha k)^q n^{\gamma(\sum_{s=1}^{b}(n_s+2-\alpha)+\sum_{t=1}^{q-b}(m_t-\alpha)+\sum_{s=1}^{b}(n_s+2-\alpha)+\sum_{t=1}^{q-b}(m_t-2)+\sum_{s=1}^{b}(2k-2t-2q-\alpha)})$$

$$\leq L_0(n^\gamma)^q (2C_\alpha k)^q n^{\gamma(2k-2t-2q-\alpha)}$$

where $f^+ = \max(f, 0)$. After reorganizing and combining the terms, the expectation of the trace is then bounded by

$$\mathbb{E}(\text{Tr}(\hat{M}_{p,n} \hat{M}_{p,n}^*)^k 1_{E_n}) \leq p \sum_{l=1}^{k} \sum_{r=0}^{l} \binom{k}{r} \binom{k}{l-r} (2k-l) (Cpn^{\mu - 1})^r(Cn^{\mu})^{l-r} \cdot \left(\sum_{q=0}^{2k-2l} (q + 1)^{6k-6l} (2L_0(n^\gamma)C_\alpha k^3)^q n^{\gamma(2k-2l-2q)} \sum_{b=0}^{q} n^b \right).$$
We now consider the terms inside the bracket. Firstly, for $\gamma > 0$,
\[
\sum_{b=0}^{q} n^{2b} = n^{2q} - 1 \leq \frac{n^{2q+1} - 1}{n^{2q}/2} = 2n^{2q}.
\]
Next, we use the elementary inequality $(q + 1)^2 \leq w^{q+1}(z/w)^z$ for any $w > 1, z > 0, q > 0$. In what follows, we apply this inequality, substituting $w = 2$ and $z = 6k - 6l$, and $L_j(\cdot)$’s are all slowly varying functions.
\[
\sum_{q=0}^{2k-2l} (q + 1)^{6k-6l} (2L_0(n^{\gamma})C\alpha k^3)^q n^{\gamma(2k-2l-2q)} \sum_{b=0}^{q} n^{2b} \\
\leq 2(6n^{\gamma/3}k/\log 2)^{6k-6l} \sum_{q=0}^{2k-2l} (2L_1(n^{\gamma})k^3)^q \\
\leq 2(6n^{\gamma/3}k/\log 2)^{6k-6l} (2L_2(n^{\gamma})k^3)^{2k-2l} \leq (k^2(n^{\gamma}L_3(n^{\gamma}))^{1/3})^{6k-6l}.
\]
Next, using the combinatorial inequality that for any $\delta > 0$ (see [29, Lemma 2.1]),
\[
\binom{k}{r} \binom{k}{l-r} \binom{2k-l}{l} \leq (1 + \sqrt{\delta})^{2k-l} \binom{k}{l} \binom{2l}{2r}.
\]
we get
\[
\mathbb{E}(\text{Tr}(\hat{M}_{p,n} \hat{M}_{p,n}^*)^k 1_{E_n}) \\
\leq p(1 + \sqrt{\delta})^{2k} \sum_{l=1}^{k} \binom{k}{l} \left( \frac{2l}{2r} \right) (p/n)^r (Cn^\mu)^l \left( k^2(n^{\gamma}L_3(n^{\gamma}))^{1/3}^{\delta-1/6} \right)^{6k-6l} \\
\leq p(1 + \sqrt{\delta})^{2k} \sum_{l=1}^{k} \binom{k}{l} \left( 1 + \sqrt{\delta} \right)^{2l} (Cn^\mu)^l \left( k^2(n^{\gamma}L_3(n^{\gamma}))^{1/3}^{\delta-1/6} \right)^{6k-6l} \\
\leq p \left[ (1 + \sqrt{\delta})^2 Cn^\mu (1 + \sqrt{p/n})^2 + \left( k^2(n^{\gamma}L_4(n^{\gamma}))^{1/3}^{\delta-1/6} \right)^6 \right]^k.
\]
Since $\mu \leq 2\gamma'$, and for any $\delta > 0, k = \lfloor n^{\gamma''} \rfloor$,
\[
\left( k^2(n^{\gamma}L_4(n^{\gamma}))^{1/3}^{\delta-1/6} \right)^6 \sim n^{2(\gamma+6\gamma'')} = o(n^{2\gamma'}),
\]
we get (22).
\[\square\]

Remark 9. Note that all the proofs in this subsection only used sparseness to determine the number of nonzero entries in each row and column. The exact location of the nonzero entries played no role in the proof as long as $\mu > 0$. 

The next two theorems are classical tools in the perturbation theory of eigenvalues and can be found in several books and papers. We only state them here for convenience. We write $P_{\vec{v}}(\cdot)$ for the orthogonal projection on the vector $\vec{v}$.

**Theorem 4.5** (Cauchy interlacing theorem - Lemma 22 [28]). Let $1 \leq p \leq n$.

(a) Let $A_n$ be an $n \times n$ Hermitian matrix and $A_{n-1}$ be its $(n-1) \times (n-1)$ minor, then $\lambda_1(A_n) \geq \lambda_1(A_{n-1}) \geq \lambda_2(A_n) \geq \cdots \geq \lambda_{n-1}(A_{n-1}) \geq \lambda_n(A_n)$;

(b) Let $A_{p,n}$ be a $p \times n$ matrix and $A_{(p-1),n}$ be its $(p-1) \times n$ minor, then $\sigma_1(A_{p,n}) \geq \sigma_1(A_{(p-1),n}) \geq \sigma_2(A_{p,n}) \geq \cdots \geq \sigma_{p-1}(A_{(p-1),n}) \geq \sigma_p(A_{p,n})$;

(c) If $p < n$ and $A_{p,n}$ is a $p \times n$ matrix and $A_{p,(n-1)}$ is its $p \times (n-1)$ minor, then $\sigma_1(A_{p,n}) \geq \sigma_1(A_{p,(n-1)}) \geq \sigma_2(A_{p,n}) \geq \cdots \geq \sigma_{p-1}(A_{p,(n-1)}) \geq \sigma_p(A_{p,n})$,

where in (b) and (c), $\sigma_i(\cdot)$ denotes the $i$-th largest singular value.

**Theorem 4.6** (Perturbation of eigenvalues and eigenvectors [11]). Let $A$ be a Hermitian matrix and $\vec{v}$ be a unit vector. Let $\zeta = \langle \vec{v}, A\vec{v} \rangle$ and $r = ||(A - \zeta)\vec{v}||$.

(a) There exists an eigenvalue $\lambda_r$ of $A$ in the closed ball $B(\zeta, r)$.

(b) If $\lambda_r$ is the only eigenvalue in $B(\zeta, r)$ with corresponding eigenvector $\vec{v}_r$, and all other eigenvalues are at distance at least $\delta > r$ of $\zeta$ then $||\vec{v}_r - P_{\vec{v}}(\vec{v}_r)|| \leq \frac{2r}{\delta - r}$.

**References**

[1] Greg W. Anderson, Alice Guionnet, and Ofer Zeitouni, *An Introduction to Random Matrices*, Cambridge University Press, 2010.

[2] Antonio Auffinger, Gérard Ben Arous, and Sandrine Péché, *Poisson convergence for the largest eigenvalues of Heavy Tailed Random Matrices*, Annales de l’Institut Henri Poincaré 45 (2009), 589–610.

[3] Zhidong Bai and Jack W. Silverstein, *Spectral Analysis of Large Dimensional Random Matrices*, Springer Science & Business Media, 2009.

[4] Zhidong Bai, Jack W. Silverstein, and Y. Q. Yin, *A note on the largest eigenvalue of a large dimensional sample covariance matrix*, Journal of Multivariate Analysis 26 (1988), no. 2, 166–168.

[5] Zhidong Bai and Y. Q. Yin, *Necessary and sufficient conditions for almost sure convergence of the largest eigenvalue of a Wigner matrix*, The Annals of Probability (1988), 1729–1741.

[6] Florent Benaych-Georges and Alice Guionnet, *Central limit theorem for eigenvectors of heavy tailed matrices*, Electronic Journal of Probability 19 (2014), 1–27.

[7] Florent Benaych-Georges and Sandrine Péché, *Localization and delocalization for heavy tailed band matrices*, Annales de l’Institut Henri Poincaré, Probabilités et Statistiques 50 (2014), no. 4, 1385–1403.

[8] Florent Benaych-Georges and Sandrine Péché, *Largest eigenvalues and eigenvectors of band or sparse random matrices*, Electron. Commun. Probab. 19 (2014), no. 4, 1–9.

[9] George Bennett, *Probability Inequalities for the Sum of Independent Random Variables*, Journal of the American Statistical Association 57 (1962), no. 297, 33–45.

[10] Nicholas H. Bingham, Charles M. Goldie and Jozef L. Teugels, *Regular variation*, Cambridge University Press, 1987.

[11] Rajendra Bhatia, *Matrix analysis*, Graduate texts in mathematics. Springer, 1997.

[12] Paul Bourgade and Horng-Tzer Yau, *The Eigenvector Moment Flow and local Quantum Unique Ergodicity*, arXiv:1312.1301 (2013).
[13] Pierre Cizeau and Jean-Philippe Bouchaud, Theory of Lévy matrices, Phys. Rev. E 50 (1994), 1810–1822.
[14] László Erdős, Antti Knowles, Horng-Tzer Yau, and Jun Yin, Spectral statistics of Erdős-Rényi Graphs II: Eigenvalue spacing and the extreme eigenvalues, Communications in Mathematical Physics 314 (2012), 3, 587–640.
[15] William Feller, An introduction to probability theory and its applications, Vol. II, 1971.
[16] Oleksiy Khorunzhiy, Estimates for moments of random matrices with gaussian elements, Lecture Notes in Mathematics 1934 (2008), 51–92 (English).
[17] Ji Oon Lee and Jun Yin, A necessary and sufficient condition for edge universality of wigner matrices, Duke Math. J. 163 (2014), no. 1, 117–173.
[18] Alexander Mirlin and Yan V. Fyodorov, On the density of states of sparse random matrices, J. Phys. A: Math. Gen. 24 (1991), no. 10, 2219.
[19] Taro Nagao, Spectral density of complex networks with two species of nodes, Journal of Physics A: Mathematical and Theoretical 46 (2013), no. 6, 065003.
[20] Taro Nagao and Toshiyuki Tanaka, Spectral density of sparse sample covariance matrices, Journal of Physics A: Mathematical and Theoretical 40 (2007), no. 19, 4973.
[21] Sandrine Péché, Universality results for the largest eigenvalues of some sample covariance matrix ensembles, Probab. Theory Relat. Fields 143, (2009) 481–516.
[22] Sidney I. Resnick, Extreme Values, Regular Variation, and Point Processes, Springer Science & Business Media, November 2007.
[23] Geoff Rodgers and Cyrano De Dominicis, Density of states of sparse random matrices, J. Phys. A: Math. Gen. 23 (1990), no. 9, 1567.
[24] Guilhem Semerjian and Leticia F Cugliandolo, Sparse random matrices: the eigenvalue spectrum revisited, Journal of Physics A: Mathematical and General 35 (2002), no. 23, 4837.
[25] Sasha Sodin, The spectral edge of some random band matrices, Ann. of Math (2) 172 (2010), 2223–2251 (English).
[26] Alexander Soshnikov, Poisson statistics for the largest eigenvalues of Wigner random matrices with heavy tails, Electronic Communications in Probability 9 (2004), 82–91.
[27] Alexander Soshnikov, Universality at the edge of the spectrum in Wigner random matrices, Comm. Math. Phys. 207 (1999), 697–733.
[28] Terence Tao and Van Vu, Random covariance matrices: universality of local statistics of eigenvalues, The Annals of Probability 40 (2012), no. 3, 1285–1315.
[29] Y. Q. Yin, Zhidong Bai, and Pathak R. Krishnaiah, On the limit of the largest eigenvalue of the large-dimensional sample covariance matrix, Probability Theory and Related Fields 78 (1988), no. 4, 599–521.
[30] Mika Yoshida and Toshiyuki Tanaka, Analysis of Sparsely-Spread CDMA via Statistical Mechanics, Proc. 2006 IEEE Int. Symp. Info. Theory (2006), 2278–2382.

Antonio Auffinger - Northwestern University, Department of Mathematics, 2033 Sheridan Road, Evanston, IL 60208, USA.
E-mail address: tuca@northwestern.edu

Si Tang - University of Chicago, Department of Statistics, 5734 S. University Avenue, Chicago, IL 60637, USA.
E-mail address: sitang@galton.uchicago.edu