Bernoulli **23**(4A), 2017, 2466–2532
DOI: 10.3150/16-BEJ815

Spectral analysis of high-dimensional sample covariance matrices with missing observations

KAMIL JURCZAK[*] and ANGELIKA ROHDE[**]

Fakultät für Mathematik, Ruhr-Universität Bochum, 44780 Bochum, Germany.
E-mail: [*]*kamil.jurczak@ruhr-uni-bochum.de;* [**]*angelika.rohde@ruhr-uni-bochum.de*

We study high-dimensional sample covariance matrices based on independent random vectors with missing coordinates. The presence of missing observations is common in modern applications such as climate studies or gene expression micro-arrays. A weak approximation on the spectral distribution in the "large dimension d and large sample size n" asymptotics is derived for possibly different observation probabilities in the coordinates. The spectral distribution turns out to be strongly influenced by the missingness mechanism. In the null case under the missing at random scenario where each component is observed with the same probability p, the limiting spectral distribution is a Marčenko–Pastur law shifted by $(1-p)/p$ to the left. As $d/n \to y \in (0,1)$, the almost sure convergence of the extremal eigenvalues to the respective boundary points of the support of the limiting spectral distribution is proved, which are explicitly given in terms of y and p. Eventually, the sample covariance matrix is positive definite if p is larger than

$$1-(1-\sqrt{y})^2,$$

whereas this is not true any longer if p is smaller than this quantity.

Keywords: almost sure convergence of extremal eigenvalues; characterization of positive definiteness; limiting spectral distribution; sample covariance matrix with missing observations; Stieltjes transform

1. Introduction

In many modern applications, high-dimensional data suffers from missing observations. As pointed out in [24], "The data from microarray experiments is usually in the form of large matrices of expression levels of genes (rows) under different experimental conditions (columns) and frequently with some values missing. Missing values occur for diverse reasons, including insufficient resolution, image corruption, or simply due to dust or scratches on the slide. Missing data may also occur systematically as a result of the robotic methods used to create them." "Data available for climate research typically suffer from uneven sampling due to… sporadic instrument failure; or other interruptions during the period of interest," [20]. Further, missing observations in telescope data may be caused by a cloudy sky, [18]. In the statistical literature, high-dimensional low-rank covariance matrix estimation with missing observations has been investigated in [16], where sparsity oracle inequalities for a matrix-Lasso estimator are derived. Very recently, an adaptive test for large covariance matrices with missing observations has been proposed in [5]. While in view of inference statements asymptotic properties of the eigenvalues and eigenvectors for high-dimensional sample covariance matrices based on complete data are exhaustively investigated in random matrix theory, the statistically equally important case of missing observations has not been studied so far. Concerning spectral based dimension reduction techniques and statistics such as the log-determinant, a profound spectral analysis is inevitable. The aim of this article is to get this development underway. We study asymptotic spectral properties of high-dimensional sample covariance matrices with missing observations. Let

$$Y = (Y_1, \dots, Y_n) \in \mathbb{R}^{d\times n}, \qquad Y_k = (Y_{1k}, \dots, Y_{dk})^* \in \mathbb{R}^d, \qquad k = 1, \dots, n,$$

be a sample of independent identically distributed (i.i.d.) random vectors with covariance matrix

$$T = \mathbb{E}\big((Y_1 - \mathbb{E}Y_1) \otimes (Y_1 - \mathbb{E}Y_1)\big).$$

In examples as described above, we do not observe the whole random vector Y_k but some of its components. This missingness is represented by a random matrix $\varepsilon \in \mathbb{R}^{d\times n}$ with entries

$$\varepsilon_{ik} = \begin{cases} 1, & \text{if } Y_{ik} \text{ is observed,} \\ 0, & \text{if } Y_{ik} \text{ is missing.} \end{cases}$$

Under the assumption that the matrices Y and ε are independent, the estimator

$$\hat{T}_{ij} = \frac{1}{N_{ij}} \sum_{k\in\mathcal{N}_{ij}} (Y_{ik} - \bar{Y}_i)(Y_{jk} - \bar{Y}_j)$$

is the analogue of the sample covariance and hence the natural estimator for T_{ij}, where

$$\mathcal{N}_{ij} = \big\{k \in \{1, \dots, n\} : \varepsilon_{ik}\varepsilon_{jk} = 1\big\}, \qquad N_{ij} = 1 \vee \#\mathcal{N}_{ij} \tag{1.1}$$

and

$$\bar{Y}_i = \frac{1}{N_{ii}} \sum_{k\in\mathcal{N}_{ii}} Y_{ik}.$$

Subsequently, $\hat{T} = (\hat{T}_{ij}) \in \mathbb{R}^{d\times d}$ is referred to as sample covariance matrix with missing observations. If $\mathbb{E}Y_k = 0$ is known in advance one typically uses the estimator

$$\hat{\Sigma} = (\hat{\Sigma}_{ij}) \in \mathbb{R}^{d\times d}, \qquad \hat{\Sigma}_{ij} = \frac{1}{N_{ij}} \sum_{k\in\mathcal{N}_{ij}} Y_{ik}Y_{jk}.$$

In what follows, we write $\hat{\Xi}$ for $\hat{T}$ and $\hat{\Sigma}$ if a statement holds for both estimators. The distribution of the missingness matrix ε substantially influences the spectrum of $\hat{\Xi}$ (see Figure 1). In the high-dimensional scenario, $\hat{\Xi}$ may be asymptotically indefinite even if the smallest eigenvalue of T stays uniformly bounded away from zero. Heuristically, it is not clear at all how the high dimensionality affects the spectral properties in the situation of missing observations, and whether

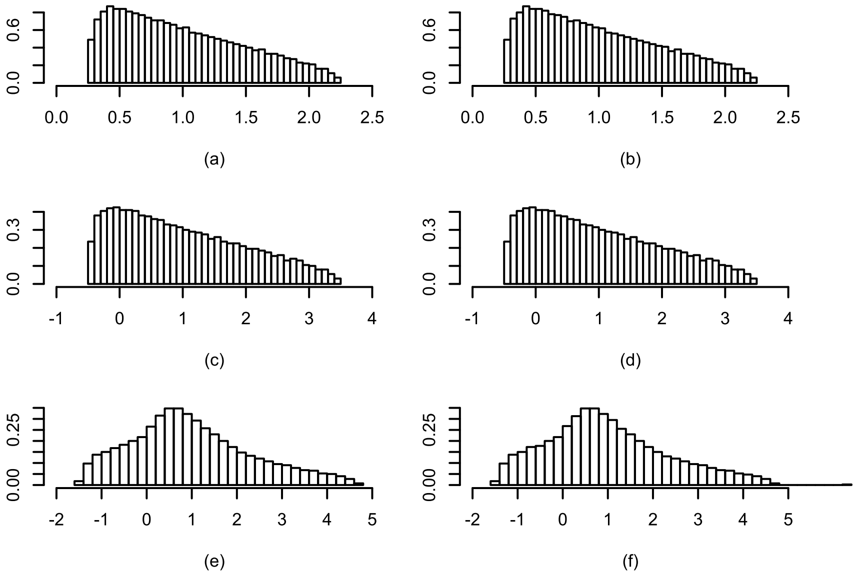

Figure 1. The left column shows histograms of the eigenvalues of the estimator $\hat{\Sigma}$ and the right column of the estimator $\hat{T}$ from a centered Gaussian sample. The underlying population covariance matrix in each histogram is the identity. The dimension of the observations in the first row is 2000, the sample size 8000 and all coordinates are observed. In the second row each coordinate is observed with probability $1/2$. In the last row, the probabilities of observation are changed to $1/4$ for the first 1000 coordinates and to $3/4$ for the other half of the coordinates.

well-known phenomena occur in a possibly modified way. In this article we investigate asymptotic spectral properties of $\hat{\Xi}$ under the classical missing (completely) at random (MAR) setting. Here, the variables ε_{ik}, $i = 1, \dots, d$, $k = 1, \dots, n$, are independent random Bernoulli variables with

$$\mathbb{P}(\varepsilon_{ik} = 1) = p_i \quad \text{and} \quad \mathbb{P}(\varepsilon_{ik} = 0) = 1 - p_i,$$

and they are jointly independent of $Y_1, \dots, Y_n$. The latter are assumed to be of the form

$$Y_k = T^{1/2} X_k + \mathbb{E} Y_k, \qquad k = 1, \dots, n,$$

where $X_1, \dots, X_n$ are i.i.d. centered random vectors with independent coordinates of variance 1. This representation is common in literature on random matrix theory. Without missing observations, that is, for completely observed random vectors $Y_1, \dots Y_n$, the classical sample covariance matrix is a well-studied object in the large dimension d and large sample size n asymptotics. The first result on its spectral distribution is due to [17]. They established in particular weak convergence in probability of the empirical spectral distribution for diagonal T under the assumption of finite fourth moment on the entries of $X_1, \dots, X_n$ and some dependency condition reflected in their mixed second and fourth moments. The most general version of this statement has been proved in [22], where weak convergence (almost surely) is established under the finite second moment assumption for rather general matrices T. The almost sure convergence of the largest eigenvalue in the null case $T = I_{d\times d}$ (identity matrix) has been proved in [29] under the assumption of the existence of the fourth moment, which generalizes a first result in this direction due to [8]. Bai, Silverstein and Yin [3] have shown that the existence of the fourth moment is in fact necessary. As concerns the smallest eigenvalue in the null case, the most current theorem on its almost sure convergence has been derived by [4]. Under quite general regularity conditions on T, the convergence of the extremal eigenvalues to the respective boundaries of the support of the limiting spectral distribution follows from [2]. Our contributions in this article are the following.

(i) We establish a weak approximation of the empirical spectral distribution of the sample covariance matrix with missing observations $\hat{\Xi}$ by a non-random sequence of probability measures expressed in terms of their Stieltjes transforms, which holds true for possibly different observation probabilities in the coordinates. In the null case under the missing at random scenario where each component is observed with the same probability p, the limiting spectral distribution is shown to be a Marčenko–Pastur law shifted by $(1-p)/p$ to the left.

(ii) As $d/n \to y \in (0, 1)$ and under the missing at random scenario where each component is observed with the same probability, we prove almost sure convergence of the extremal eigenvalues of $\hat{\Sigma}$ to the respective boundary points of the support of the limiting spectral distribution in the null case. A statistically important consequence is the characterization of positive definiteness for the sample covariance matrix with missing observations.

Understanding the empirical spectral distribution of sample covariance matrices with missing observations is of great importance to develop improved estimators for the population covariance matrix and the precision matrix. Such estimators have been already established for completely observed data by [7] and [12] based on non-linear shrinkage of the eigenvalues. However, if some data is missing, the situation is more intricate since the analysis in our article reveals that the limiting behavior of the empirical spectral distribution does not only depend on the eigenvalues of the population covariance matrix but also on its eigenvectors. Nevertheless, we expect that adjusting the diagonal of the sample covariance matrix with missing observations yields a more suitable matrix for spectrum estimation.

Very recently, various authors studied asymptotic spectral properties of sample autocovariance matrices of high-dimensional time series which is another statistically relevant scenario. Jin *et al.* [10] derived the limiting spectral distribution of the symmetrized autocovariance matrix in the i.i.d. case. Liu, Aue and Paul [15] established a Marčenko–Pastur-type law for the empirical spectral distribution in case of general high-dimensional linear time series. They investigated the moderately high-dimensional case of this problem in [27]. Li, Pan and Yao [14] developed the limiting singular value distribution of the sample autocovariance matrix by means of the Stieltjes transform for an independent sequence with elements possessing finite fourth moments. Wang and Yao [28] proved the same result by the method of moments, and additionally the almost sure convergence of the spectral norm. The strong limit of the extreme eigenvalues of symmetrized autocovariance matrices is established in [26].

The article is organized as follows. First, we introduce the essential notation and the model assumptions in the next section. Section 3 is devoted to our main results. The proof of Theorem 3.1 is quite long and therefore decomposed into Section 4, Section 5 and Appendix A. The proof of Theorem 3.3 is deferred to Section 6 and Appendix B. Some auxiliary results which are used throughout the proofs are collected in Appendix C.

2. Notation and preliminaries

2.1. Notation

For any bounded function $f : \mathbb{R} \to \mathbb{R}$

$$\|f\| = \sup_{x \in \mathbb{R}} \big|f(x)\big|$$

denotes its supremum norm. If f is Lipschitz in addition, then the bounded Lipschitz norm is defined as

$$\|f\|_{\mathrm{BL}} = \max\big(\|f\|_L, \|f\|\big),$$

where $\|f\|_L$ denotes is the best Lipschitz constant of f. We write

$$\mathbb{C}^+ = \{z \in \mathbb{C} : \Im z > 0\}$$

for the upper complex half plane. For any Hermitian matrix $A \in \mathbb{C}^{d \times d}$ denote the (normalized) spectral measure by

$$\mu^A = \frac{1}{d} \sum_{i=1}^{d} \delta_{\lambda_i(A)},$$

where $\lambda_1(A) \geq \cdots \geq \lambda_d(A)$ are the eigenvalues of A and δ_x denotes the Dirac measure in x. If it is clear that we refer to a matrix A, we use the shortened notation $\lambda_1 \geq \cdots \geq \lambda_d$. We write A^* for the adjoint of A. Let us introduce the Schatten norms for matrices

$$\|A\|_{S_p} = \left(\sum_{i=1}^{d} \lambda_i \big(AA^*\big)^{p/2} \right)^{1/p}, \qquad p \geq 1.$$

Furthermore, $\mathrm{tr}(A)$ denotes the trace of A and $\mathrm{rank}(A)$ its rank. For two matrices $A, B \in \mathbb{R}^{d \times n}$ we write $A \circ B = (A_{ik} B_{ik})_{i,k}$ for the Hadamard product. For any vector $v \in \mathbb{R}^d$, $\mathrm{diag}(v) \in \mathbb{R}^{d \times d}$ is the diagonal matrix with the ith diagonal entry equal to v_i. With slight abuse of notation we also write $\mathrm{diag}(A)$ for $\mathrm{diag}(A_{11}, \ldots, A_{dd})$, $A \in \mathbb{R}^{d \times d}$. The Stieltjes transform of a measure μ on the real line is defined by

$$m_\mu(z) = \int_{\mathbb{R}} \frac{1}{\lambda - z} \, \mathrm{d}\mu(\lambda), \qquad z \in \mathbb{C}^+.$$

On the space of probability measures on $\mathbb{R}$ recall the following distance measures

$$\text{Kolmogorov metric: } d_K(\mu, \nu) = \big\| \mu\big((-\infty, \cdot]\big) - \nu\big((-\infty, \cdot]\big) \big\|,$$

$$\text{Dual bounded Lipschitz metric: } d_{\mathrm{BL}}(\mu, \nu) = \sup_{\|f\|_{\mathrm{BL}} \leq 1} \int_{\mathbb{R}} f \, \mathrm{d}(\mu - \nu),$$

$$\text{Lévy metric: } d_L(\mu,\nu) = \inf\bigl\{\varepsilon > 0 \,|\, \mu\bigl((-\infty, x-\varepsilon]\bigr) - \varepsilon \le \nu\bigl((-\infty, x]\bigr) \le \mu\bigl((-\infty, x+\varepsilon]\bigr) + \varepsilon \text{ for all } x \in \mathbb{R}\bigr\}.$$

We will frequently make use of the well-known relation $d_L(\mu,\nu) \le d_K(\mu,\nu)$ for any two probability measures μ and ν on the real line, cf. [19], page 43. For any measures μ and ν, $\mu \star \nu$ denotes their convolution. As usual, $\Longrightarrow$ stands for weak convergence. The Marčenko–Pastur distribution with parameters $y, \sigma^2 > 0$ is given by

$$\mu^{\mathrm{MP}}_{y,\sigma^2} = \left(1 - \frac{1}{y}\right)_+ \delta_0 + \frac{1}{2\pi\sigma^2} \frac{\sqrt{(b-x)(x-a)}}{yx} \mathbb{1}\{a \le x \le b\}\,\mathrm{d}x \tag{2.1}$$

with $a = \sigma^2(1-\sqrt{y})^2$ and $b = \sigma^2(1+\sqrt{y})^2$. Moreover, for $\sigma^2 > 0$ let $\mu^{\mathrm{MP}}_{0,\sigma^2} = \delta_{\sigma^2}$. The notation $\lesssim$ means less or equal up to some positive multiplicative constant which does not depend on the variable parameters in the expression.

2.2. Preliminaries

Let $(X(i,k))_{i,k\in\mathbb{N}}$ be a double array of i.i.d. centered random variables with unit variance. The left upper $d \times n$ submatrix is denoted by $X_{d,n}$. Then the random vectors $Y_{1,d,n}, \dots, Y_{n,d,n} \in \mathbb{R}^d$ are the columns of the matrix

$$Y_{d,n} - \mathbb{E}Y_{d,n} = T_{d,n}^{1/2} X_{d,n}$$

with

$$T_{d,n} = \operatorname{diag}(T_{11,d,n}, \dots, T_{dd,d,n}) \in \mathbb{R}^{d\times d}.$$

This structure on the population covariance matrix is the simplest one which allows to visualize the effects of missing observations on the spectrum of the sample covariance matrix. The non-diagonal case is discussed at the end of Section 4. Its treatment requires some technical modification of the arguments presented here but not substantially new ideas and is beyond the scope of the article. $(\varepsilon_{d,n})_{d,n}$ is a triangular array of random matrices $\varepsilon_{d,n} \in \mathbb{R}^{d\times n}$ independent of $(X(i,k))_{i,k\in\mathbb{N}}$, where the entries $\varepsilon_{ik,d,n}$ are independent Bernoulli variables with observation probabilities

$$\mathbb{P}(\varepsilon_{ik,d,n} = 1) = p_{i,d,n}, \qquad i = 1,\dots,d, k = 1,\dots n.$$

The dependence of the set $\mathcal{N}_{ij}$ and the number N_{ij} in (1.1) on the sequence $(\varepsilon_{d,n})$ is indicated by an additional subscript d, n. Throughout this article we impose that the family of spectral measures of the population covariance matrices $(T_{d,n})$ as well as the family of empirical distributions

$$\bigl(\mu^{w_{d,n}}\bigr)_{d,n}, \qquad \text{with } \mu^{w_{d,n}} = \frac{1}{d}\sum_{i=1}^{d} \delta_{w_{i,d,n}} \quad \text{and} \quad w_{d,n} = \bigl(p_{1,d,n}^{-1}, \dots, p_{d,d,n}^{-1}\bigr),$$

are tight. This assumption ensures that there are not too many probabilities of observation $p_{i,d,n}$ in the vector $p_{d,n}$ that are very close to zero, in the sense that for most coordinates $i = 1,\dots,n$

the number of observations remains proportional to n, while a few degeneracies may occur. Asymptotic statements refer to

$$d \to \infty \qquad \text{while } n = n(d) \text{ satisfies } \limsup_{d\to\infty}(d/n) < \infty. \tag{2.2}$$

The sequence of sample covariance matrices with missing observations is denoted by

$$(\hat{\Xi}_{d,n})_{d,n},$$

the corresponding sequence of spectral measures by $(\mu_{d,n})_{d,n}$ and their Stieltjes transforms by $(m_{d,n})_{d,n}$.

3. Results

The main results of the article are the weak approximation of the spectral measure $\mu_{d,n}$ of $\hat{\Xi}_{d,n}$ by a non-random sequence of probability measures, and, in the null case, the almost sure convergence of the extremal eigenvalues of $\hat{\Sigma}_{d,n}$. Thereto, define the matrices

$$S_{d,n} = \operatorname{diag}\left(\frac{1-p_{1,d,n}}{p_{1,d,n}}T_{11,d,n}, \dots, \frac{1-p_{d,d,n}}{p_{d,d,n}}T_{dd,d,n}\right) \quad \text{and}$$
$$R_{d,n} = \operatorname{diag}\left(\frac{1}{p_{1,d,n}}T_{11,d,n}, \dots, \frac{1}{p_{d,d,n}}T_{dd,d,n}\right).$$

Theorem 3.1. *Suppose that the assumptions stated in Section 2.2 hold, and*

$$\sup_d \|R_{d,n}\|_{S_\infty} < \infty.$$

Then for any $z \in \mathbb{C}^+$,

$$\left|m_{d,n}(z) - m^\circ_{d,n}(z)\right| \to 0 \qquad a.s.,$$

where $m^\circ_{d,n}(z)$ satisfies

$$m^\circ_{d,n}(z) = \frac{1}{d}\operatorname{tr}\left\{\left(\frac{1}{1+(d/n)e^\circ_{d,n}(z)}R_{d,n} - S_{d,n} - zI_{d\times d}\right)^{-1}\right\} \tag{3.1}$$

and $e^\circ_{d,n}$ is the (unique) solution of the fixed point equation

$$e^\circ_{d,n}(z) = \frac{1}{d}\operatorname{tr}\left\{R_{d,n}\left(\frac{1}{1+(d/n)e^\circ_{d,n}(z)}R_{d,n} - S_{d,n} - zI_{d\times d}\right)^{-1}\right\}.$$

Moreover, $m^\circ_{d,n}$ is the Stieltjes transform of a probability measure $\mu^\circ_{d,n}$ on the real line and

$$\mu^\circ_{d,n} - \mu_{d,n} \Longrightarrow 0 \qquad a.s.$$

Remark. Note that the theorem covers in particular the case $d/n \to 0$. It follows from the proof that

$$\left|e^{\circ}_{d,n}(z)\right| \le \frac{\|R_{d,n}\|_{S_\infty}}{\Im z}, \qquad z \in \mathbb{C}^+.$$

Due to $R_{d,n} - S_{d,n} = T_{d,n}$, this implies that the Stieltjes transforms $m^{\circ}_{d,n}$ approach those of the spectral measures of $T_{d,n}$ as $d/n \to 0$. That is, an effect caused by missing observations appears asymptotically only in the high-dimensional scenario $\liminf_d d/n > 0$.

The equation (3.1) characterizes uniquely the approximating spectral measure via its Stieltjes transform. Without missing observation, that is, $p_{i,d,n} = 1$, the solution of (3.1) coincides with the solution to the Marčenko–Pastur equation

$$m^{\circ}_{d,n}(z) = \frac{1}{d}\sum_{i=1}^{d} \frac{1}{T_{ii,d,n}(1 - (d/n) - (d/n)z \cdot m^{\circ}_{d,n}(z)) - z}.$$

The difference in the representation results from the fact that the spectra of

$$T^{1/2}_{d,n} X_{d,n} X^*_{d,n} T^{1/2}_{d,n} \quad \text{and} \quad X^*_{d,n} T_{d,n} X_{d,n}$$

are identical up to $|d - n|$ zero eigenvalues, which is used in the classical analysis. Except for special cases, this simplification is not possible in the missing at random scenario.

It is well known that the Stieltjes transform of the Marčenko–Pastur law with parameters $(y, \sigma^2/p_0)$ is the unique solution to

$$s(z) = \left(\frac{\sigma^2}{p_0} \cdot \frac{1}{1 + (\sigma^2/p_0) y s(z)} - z\right)^{-1}$$

from $\mathbb{C}^+ \to \mathbb{C}^+$. In the special case $T_{d,n} = \sigma^2 I_{d\times d}$ and $p_{d,n} = (p_0, \dots, p_0) \in (0,1)^d$, we have

$$m^{\circ}_{d,n}\left(z - \sigma^2 \frac{1 - p_0}{p_0}\right) = \left(\frac{\sigma^2}{p_0} \frac{1}{1 + (d/n)(\sigma^2/p_0) m^{\circ}_{d,n}(z - \sigma^2(1 - p_0)/p_0)} - z\right)^{-1}.$$

Hence, $\mu^{\circ}_{d,n}$ is the Marčenko–Pastur law $\mu^{\mathrm{MP}}_{d/n,\sigma^2/p_0}$ shifted by $\sigma^2 \frac{1-p_0}{p_0}$ to the left.

Corollary 3.2. *Grant the conditions of Theorem 3.1. If $p_{i,d,n} = p_0 > 0$ for $i = 1, \dots, d$ and $d, n \in \mathbb{N}$ and $T_{d,n} = \sigma^2 I_{d\times d}$, $\sigma^2 > 0$, we obtain*

$$\mu_{d,n} \Longrightarrow \mu^{\mathrm{MP}}_{y,\sigma^2/p_0} \star \delta_{-\sigma^2(1-p_0)/p_0} \qquad a.s.$$

as $d \to \infty$ and $d/n \to y > 0$. Eventually, as $y < 1$,

$$\limsup_d \lambda_{\min}(\hat{\Xi}_{d,n}) < 0 \qquad a.s.\ \text{if}\ p_0 < 1 - (1 - \sqrt{y})^2.$$

In other words, under the missing at random scenario where each component is observed with the same probability p_0, the limiting spectral distribution is a Marčenko–Pastur law shifted by

$\sigma^2(1-p_0)/p_0$ to the left. Eventually, the sample covariance matrix is not positive definite if p_0 is smaller than

$$1-(1-\sqrt{y})^2.$$

For the estimator $\hat{\Sigma}_{d,n}$ we even determine the almost sure limit of the extremal eigenvalues.

Theorem 3.3. *Grant the conditions of Corollary 3.2 let additionally $\mathbb{E}X_{11}^4 < \infty$ and $\varepsilon_{d,n} \in \mathbb{R}^{d\times n}$ be the upper left corner of a double array $(\varepsilon(i,k))_{i,k\in\mathbb{N}}$ of i.i.d. Bernoulli variables with parameter p_0. Assume that $\mathbb{E}Y_{d,n}=0$. Then, if $0<y<1$,*

$$\lim_{d\to\infty}\lambda_{\min}(\hat{\Sigma}_{d,n}) = \frac{\sigma^2}{p_0}(1-\sqrt{y})^2 - \frac{1-p_0}{p_0}\sigma^2 \qquad a.s., \quad and$$

$$\lim_{d\to\infty}\lambda_{\max}(\hat{\Sigma}_{d,n}) = \frac{\sigma^2}{p_0}(1+\sqrt{y})^2 - \frac{1-p_0}{p_0}\sigma^2 \qquad a.s.$$

The limit of the smallest eigenvalue is always smaller than in the completely observed case $p_0=1$, whereas the largest eigenvalue is always larger. In the limiting case $y\to 0$ both expressions on the right-hand side reduce to σ^2 as in the completely observed classical case, independently of p_0.

As in Theorem 1 of [4], the existence of the fourth moment is necessary for the above Theorem to hold. The proof of the necessity is a straightforward adaption of the arguments in [29].

The characterization of positive definiteness in the null case under the missing at random scenario is an immediate corollary of Theorem 3.3.

Corollary 3.4. *Under the condition of Theorem 3.3,*

$$\lim_{d\to\infty}\lambda_{\min}(\hat{\Sigma}_{d,n}) < 0 \qquad a.s. \text{ if } p_0 < 1-(1-\sqrt{y})^2, \quad and$$

$$\lim_{d\to\infty}\lambda_{\min}(\hat{\Sigma}_{d,n}) > 0 \qquad a.s. \text{ if } p_0 > 1-(1-\sqrt{y})^2.$$

4. Proof of Theorem 3.1, Part I

Reduction to the form $\frac{1}{n}R_{d,n}^{1/2}Z_{d,n}Z_{d,n}^*R_{d,n}^{1/2} - S_{d,n}$

With the notation

$$\bar{T}_{d,n} = \frac{1}{n}R_{d,n}^{1/2}Z_{d,n}Z_{d,n}^*R_{d,n}^{1/2} - S_{d,n}$$

and

$$Z_{d,n}\in\mathbb{R}^{d\times n}, \qquad Z_{ik,d,n} = \frac{X_{ik,d,n}\varepsilon_{ik,d,n}}{p_{i,d,n}^{1/2}}, \qquad i=1,\dots,d, k=1,\dots,n,$$

let $\bar{\mu}_{d,n}$ be the spectral measure of $\bar{T}_{d,n}$. The aim of this section is to show that the spectral distributions $\mu_{d,n}$ of $\hat{\Xi}_{d,n}$ may be approximated by $\bar{\mu}_{d,n}$.

Proposition 4.1. *Grant the conditions of Section 2.2. Then*

$$d_L(\bar{\mu}_{d,n}, \mu_{d,n}) \longrightarrow 0 \qquad a.s.$$

Remark. Corollary 3.2 can be equally deduced from Proposition 4.1. Since in that case $S_{d,n}$ is a multiple of identity, the eigenvalues satisfy

$$\lambda_i(\bar{T}_{d,n}) = \lambda_i\left(\frac{1}{n} R_{d,n}^{1/2} Z_{d,n} Z_{d,n}^* R_{d,n}^{1/2}\right) - \frac{1-p_0}{p_0}\sigma^2, \qquad i = 1, \dots, d.$$

For the matrix

$$\frac{1}{n} R_{d,n}^{1/2} Z_{d,n} Z_{d,n}^* R_{d,n}^{1/2}$$

it is well known (see, e.g., [22]) that the spectral distribution converges weakly to $\mu^{\mathrm{MP}}_{y,\sigma^2/p_0}$ almost surely as $d/n \to y > 0$.

The proof of Proposition 4.1 is postponed to Appendix A. At this place we give a sketch of the proof. Subsequently, we restrict our attention to the estimator $\hat{T}_{d,n}$. The proof for $\hat{\Sigma}_{d,n}$ is just a simplified version.

The proof of Proposition 4.1 is subdivided into eight steps. In each step, $\hat{T}_{d,n}$ is modified in a way which does not affect its spectral distribution asymptotically. In order to simplify the notation each modification of $\hat{T}_{d,n}$ from one step will be again denoted by $\hat{T}_{d,n}$ in the next step. Within the proof denote

$$\hat{W}_{d,n} \in \mathbb{R}^{d\times d}, \qquad \hat{W}_{ij,d,n} = \frac{n}{N_{ij,d,n}},$$
$$W_{d,n} \in \mathbb{R}^{d\times d}, \qquad W_{ij,d,n} = \frac{n}{\mathbb{E}\#\mathcal{N}_{ij,d,n}}.$$

Before we start with the description of the proof, we rearrange the entries $\hat{T}_{ij,d,n}$ as follows

$$\begin{aligned}
&\frac{1}{N_{ij,d,n}} \sum_{k\in\mathcal{N}_{ij,d,n}} (Y_{ik,d,n} - \bar{Y}_{i,d,n})(Y_{jk,d,n} - \bar{Y}_{j,d,n}) \\
&\quad = \frac{1}{N_{ij,d,n}} \sum_{k\in\mathcal{N}_{ij,d,n}} \big((Y_{ik,d,n} - \mathbb{E}Y_{ik,d,n}) - (\bar{Y}_{i,d,n} - \mathbb{E}Y_{ik,d,n})\big) \\
&\qquad \times \big((Y_{jk,d,n} - \mathbb{E}Y_{jk,d,n}) - (\bar{Y}_{j,d,n} - \mathbb{E}Y_{jk,d,n})\big) \\
&\quad = \frac{1}{N_{ij,d,n}} \sum_{k\in\mathcal{N}_{ij,d,n}} \left[(Y_{ik,d,n} - \mathbb{E}Y_{ik,d,n}) - \frac{1}{N_{ii,d,n}} \sum_{l\in\mathcal{N}_{ii,d,n}} (Y_{il,d,n} - \mathbb{E}Y_{il,d,n})\right] \\
&\qquad \times \left[(Y_{jk,d,n} - \mathbb{E}Y_{jk,d,n}) - \frac{1}{N_{jj,d,n}} \sum_{l\in\mathcal{N}_{jj,d,n}} (Y_{jl,d,n} - \mathbb{E}Y_{jl,d,n})\right].
\end{aligned}$$

Therefore, we may assume without loss of generality $Y_{d,n}$ to be centered. Rewrite $\hat{T}_{d,n}$ in the following way

$$\begin{aligned}\hat{T}_{d,n} &= \frac{1}{n}\hat{W}_{d,n} \circ \big((Y_{d,n} \circ \varepsilon_{d,n})(Y_{d,n} \circ \varepsilon_{d,n})^*\big) \\ &\quad - \frac{1}{n}\hat{W}_{d,n} \circ \big((\hat{M}_{d,n} \circ \varepsilon_{d,n})(Y_{d,n} \circ \varepsilon_{d,n})^*\big) \\ &\quad - \frac{1}{n}\hat{W}_{d,n} \circ \big((Y_{d,n} \circ \varepsilon_{d,n})(\hat{M}_{d,n} \circ \varepsilon_{d,n})^*\big) \\ &\quad + \frac{1}{n}\hat{W}_{d,n} \circ \big((\hat{M}_{d,n} \circ \varepsilon_{d,n})(\hat{M}_{d,n} \circ \varepsilon_{d,n})^*\big),\end{aligned}$$

where

$$\hat{M}_{d,n} = (\underbrace{\hat{m}_{d,n}, \dots, \hat{m}_{d,n}}_{n \text{ times}}) \in \mathbb{R}^{d\times n} \qquad \text{with } \hat{m}_{i,d,n} = \frac{1}{N_{ii,d,n}} \sum_{k\in\mathcal{N}_{ii,d,n}} Y_{ik,d,n}. \tag{4.1}$$

Let us briefly describe the separate steps of the proof. The first three steps use the inequality

$$d_K\big(\mu^A, \mu^B\big) \le \frac{1}{d}\operatorname{rank}(A - B)$$

for Hermitian matrices $A, B \in \mathbb{R}^{d\times d}$ in order to regularize certain rows of $\varepsilon_{d,n}$ for which the probability of observation $p_{i,d,n}$ is smaller than some given value $p_0 > 0$, to get rid of the additive term

$$\frac{1}{n}\hat{W}_{d,n} \circ \big((\hat{M}_{d,n} \circ \varepsilon_{d,n})(\hat{M}_{d,n} \circ \varepsilon_{d,n})^*\big),$$

and to truncate the diagonal entries of $T_{d,n}$. Thereafter, we want to make use of the inequality

$$d_L^3\big(\mu^A, \mu^B\big) \le \frac{1}{d}\operatorname{tr}\big((A - B)(A - B)^*\big), \tag{4.2}$$

where, in our case, A and B are two $d \times d$ random Hermitian matrices. In order to deduce almost sure convergence to 0 of the right-hand side by means of the Borel–Cantelli lemma, truncation of the random variables $X_{ik,d,n}$ is necessary to guarantee the existence of higher order moments of the empirical spectral distribution of $\hat{T}_{d,n}$. This is realized in Step IV. In Step V the matrix $\hat{W}_{d,n}$ is replaced by its deterministic counterpart $W_{d,n}$ the evaluation of which is based on a sophisticated combinatorial analysis of moments. In Step VI a combination of both inequalities displayed above is applied. More precisely, an entry $Y_{ik,d,n}$ is preserved depending on whether its absolute row sum $\sum_l |Y_{il,d,n}|$ exceeds a certain value or not. The number of removed rows is asymptotically negligible while the remaining matrix is suitable for an application of (4.2). Hereby, the matrices

$$-\frac{1}{n}W_{d,n} \circ \big((\hat{M}_{d,n} \circ \varepsilon_{d,n})(Y_{d,n} \circ \varepsilon_{d,n})^*\big) \quad \text{and} \quad -\frac{1}{n}W_{d,n} \circ \big((Y_{d,n} \circ \varepsilon_{d,n})(\hat{M}_{d,n} \circ \varepsilon_{d,n})^*\big)$$

are removed from $\hat{T}_{d,n}$. The form

$$W_{d,n} = w_{d,n} w_{d,n}^* + \operatorname{diag}\big(W_{d,n} - w_{d,n} w_{d,n}^*\big)$$

is the motivation for replacing

$$\frac{1}{n} \operatorname{diag}\big(W_{d,n} - w_{d,n} w_{d,n}^*\big) \circ \big((Y_{d,n} \circ \varepsilon_{d,n})(Y_{d,n} \circ \varepsilon_{d,n})^*\big)$$

by its expectation in Step VII. Reverting finally the truncation Steps II, III, IV yields the claim.

In the next section $\hat{\Xi}_{d,n}$ denotes the matrix

$$\frac{1}{n}\big(w_{d,n} w_{d,n}^*\big) \circ \big((Y_{d,n} \circ \varepsilon_{d,n})(Y_{d,n} \circ \varepsilon_{d,n})^*\big) - S_{d,n} = \frac{1}{n} R_{d,n}^{1/2} Z_{d,n} Z_{d,n}^* R_{d,n}^{1/2} - S_{d,n}$$

which is obtained in Step VIII. Correspondingly, we write $\mu_{d,n}$ and $m_{d,n}$ for its spectral measure and the Stieltjes transform.

Remark. In the case of non-diagonal $T_{d,n}$ we cannot reduce the sample covariance matrix with missing observations to the form

$$\frac{1}{n} R_{d,n}^{1/2} Z_{d,n} Z_{d,n}^* R_{d,n}^{1/2} - S_{d,n}$$

but instead have to analyze the spectrum of

$$\frac{1}{n}\big(w_{d,n} w_{d,n}^*\big) \circ \big((Y_{d,n} \circ \varepsilon_{d,n})(Y_{d,n} \circ \varepsilon_{d,n})^*\big) - S_{d,n} = \frac{1}{n}(\tilde{Y}_{d,n} \circ \varepsilon)(\tilde{Y}_{d,n} \circ \varepsilon)^* - S_{d,n}$$

with

$$\tilde{Y}_{d,n} = \operatorname{diag}(w) Y_{d,n}.$$

Nevertheless, the arguments of Section 5 can be modified at the cost of additional technical expenditure. We find that the ideas of the proof are much clearer for the diagonal special case and therefore omitted this extension due to length of the paper.

5. Proof of Theorem 3.1, Part II

Note that, in general, the spectral analysis and limiting behavior of $\hat{\Xi}_{d,n}$ significantly differ from those of the matrix analyzed in [23]. By Proposition 4.1 as well as Lemmas C.12 and C.13, we continue to show that

$$\big|m_{d,n}(z) - m_{d,n}^{\circ}(z)\big| \longrightarrow 0 \qquad \text{a.s.}$$

for all $z \in \mathbb{C}^+$. Such type of convergence has been established in [6] for

$$B_{d,n}^{1/2} X_{d,n} X_{d,n}^* B_{d,n}^{1/2} + A_{d,n}$$

for positive semidefinite Hermitian matrices $A_{d,n}, B_{d,n} \in \mathbb{C}^{d\times d}$. For the proof of Theorem 3.1, we establish the weak approximation in case of the negative semidefinite matrix $A_{d,n} = -S_{d,n}$. This requires several changes in the arguments of [6] due to the fact that the function

$$z \mapsto -\frac{1}{z(1+m(z))}$$

is a Stieltjes transform if m is a Stieltjes transform of a finite measure on $[0,\infty)$ but in general, this is not true any longer if m is just a Stieltjes transform of a finite measure on $\mathbb{R}$. Moreover, our proof includes also the case $d/n \to 0$.

The proof is structured as follows. In the first step, we truncate the entries of $X_{d,n}$ at the threshold level $K > 0$ which goes to infinity at the very end. Afterwards we start to analyze the Stieltjes transform of the empirical spectral distribution of $\hat{\Xi}_{d,n}$. With the resolvent

$$\hat{G}_{d,n}(z) = (\hat{\Xi}_{d,n} - zI_{d\times d})^{-1}$$

we prove that

$$e_{d,n}(z) = \frac{1}{d}\operatorname{tr}\{R_{d,n}\hat{G}_{d,n}(z)\}$$

is an approximate solution to the fixed point equation in Theorem 3.1 in Step II. Correspondingly, the Stieltjes transform $m_{d,n}$ is shown to be approximated by the expression (3.1) with $e_{d,n}$ in place of $e^{\circ}_{d,n}$. In the third step, existence and uniqueness of a solution to the system of equations for $m^{\circ}_{d,n}$ is established. The solution $m^{\circ}_{d,n}$ is identified as a Stieltjes transform in Step IV. In Steps V and VI, pointwise almost sure convergence of $e_{d,n} - e^{\circ}_{d,n}$ and $m_{d,n} - m^{\circ}_{d,n}$ to zero is derived. Finally, we deduce the weak convergence $\mu_{d,n} - \mu^{\circ}_{d,n} \Longrightarrow 0$ almost surely in Step VII.

5.1. Step I: Second truncation of $X_{d,n}$

For arbitrary $K > 0$, define matrices $\tilde{X}_{d,n}$, $\tilde{Z}_{d,n}$ and $\tilde{\Xi}_{d,n} = n^{-1}R^{1/2}_{d,n}\tilde{Z}_{d,n}\tilde{Z}^*_{d,n}R^{1/2}_{d,n} - S_{d,n}$, where

$$\tilde{X}_{ik} = X_{ik}\mathbb{1}\{|X_{ik}| \le K\} \quad \text{and} \quad \tilde{Z}_{ik,d,n} = \frac{\tilde{X}_{ik,d,n}\varepsilon_{ik,d,n}}{p^{1/2}_{i,d,n}}.$$

Moreover, define for arbitrary $\delta > 0$ the event

$$\Delta_{i,d,n} = \left\{\frac{1}{n}\left|\sum_{l=1}^{n} X^2_{il} - \mathbb{E}X^2_{il}\right| \vee \frac{1}{n}\left|\sum_{l=1}^{n} X^2_{il}\mathbb{1}\{|X_{il}| > K\} - \mathbb{E}X^2_{il}\mathbb{1}\{|X_{il}| > K\}\right| < \delta\right\}.$$

With this notation, let

$$\hat{\Xi}'_{d,n} = \frac{1}{n}R^{1/2}_{d,n}Z'_{d,n}\big(Z'_{d,n}\big)^*R^{1/2}_{d,n} - S_{d,n} \quad \text{and} \quad \tilde{\Xi}'_{d,n} = \frac{1}{n}R^{1/2}_{d,n}\tilde{Z}'_{d,n}\big(\tilde{Z}'_{d,n}\big)^*R^{1/2}_{d,n} - S_{d,n},$$

where

$$X'_{ik,d,n} = X_{ik}\mathbb{1}_{\Delta_{i,d,n}}, \qquad \tilde{X}'_{ik,d,n} = \tilde{X}_{ik}\mathbb{1}_{\Delta_{i,d,n}}$$

and

$$Z'_{ik,d,n} = \frac{X'_{ik,d,n}\varepsilon_{ik,d,n}}{p_{i,d,n}^{1/2}}, \qquad \tilde{Z}'_{ik,d,n} = \frac{\tilde{X}'_{ik,d,n}\varepsilon_{ik,d,n}}{p_{i,d,n}^{1/2}}.$$

Then,

$$\begin{aligned} &d_L\big(\mu^{\hat{\Xi}_{d,n}}, \mu^{\tilde{\Xi}_{d,n}}\big) \\ &\quad\le d_L\big(\mu^{\hat{\Xi}_{d,n}}, \mu^{\hat{\Xi}'_{d,n}}\big) + d_L\big(\mu^{\hat{\Xi}'_{d,n}}, \mu^{\tilde{\Xi}'_{d,n}}\big) + d_L\big(\mu^{\tilde{\Xi}'_{d,n}}, \mu^{\tilde{\Xi}_{d,n}}\big). \end{aligned} \tag{5.1}$$

First, we evaluate the second term $d_L(\mu^{\hat{\Xi}'_{d,n}}, \mu^{\tilde{\Xi}'_{d,n}})$ in (5.1). By Theorem C.10 for $\alpha = 1$, the Lidskii–Wielandt perturbation bound (1.2) in [13], and Hölder's inequality for Schatten norms,

$$\begin{aligned} &d_L^2\big(\mu^{\hat{\Xi}'_{d,n}}, \mu^{\tilde{\Xi}'_{d,n}}\big) \\ &\quad\le \frac{1}{d}\sum_{i=1}^{d}\big|\lambda_i\big(\hat{\Xi}'_{d,n}\big) - \lambda_i\big(\tilde{\Xi}'_{d,n}\big)\big| \\ &\quad\le \frac{1}{dn}\big\|R_{d,n}^{1/2}Z'_{d,n}\big(Z'_{d,n}\big)^* R_{d,n}^{1/2} - R_{d,n}^{1/2}\tilde{Z}'_{d,n}\big(\tilde{Z}'_{d,n}\big)^* R_{d,n}^{1/2}\big\|_{S_1} \\ &\quad\le \frac{1}{dn}\big\|Z'_{d,n}\big(Z'_{d,n}\big)^* - \tilde{Z}'_{d,n}\big(\tilde{Z}'_{d,n}\big)^*\big\|_{S_1}\|R_{d,n}\|_{S_\infty} \\ &\quad= \frac{1}{dn}\big\|\big(Z'_{d,n} - \tilde{Z}'_{d,n}\big)\big(Z'_{d,n} - \tilde{Z}'_{d,n}\big)^* + \big(Z'_{d,n} - \tilde{Z}'_{d,n}\big)\big(\tilde{Z}'_{d,n}\big)^* + \tilde{Z}'_{d,n}\big(Z'_{d,n} - \tilde{Z}'_{d,n}\big)^*\big\|_{S_1} \\ &\qquad\times\|R_{d,n}\|_{S_\infty} \\ &\quad\le \frac{1}{dn}\big(\big\|\big(Z'_{d,n} - \tilde{Z}'_{d,n}\big)\big(Z'_{d,n} - \tilde{Z}'_{d,n}\big)^*\big\|_{S_1} + 2\big\|\big(Z'_{d,n} - \tilde{Z}'_{d,n}\big)\big(\tilde{Z}'_{d,n}\big)^*\big\|_{S_1}\big)\|R_{d,n}\|_{S_\infty} \\ &\quad\le \frac{1}{dn}\big(\big\|\big(Z'_{d,n} - \tilde{Z}'_{d,n}\big)\big(Z'_{d,n} - \tilde{Z}'_{d,n}\big)^*\big\|_{S_1} + 2\big\|Z'_{d,n} - \tilde{Z}'_{d,n}\big\|_{S_2}\big\|\tilde{Z}'_{d,n}\big\|_{S_2}\big) \\ &\qquad\times\|R_{d,n}\|_{S_\infty} \\ &\quad\le \frac{1}{dn}\big\{\mathrm{tr}\big(\big(Z'_{d,n} - \tilde{Z}'_{d,n}\big)\big(Z'_{d,n} - \tilde{Z}'_{d,n}\big)^*\big) \\ &\qquad + 2\big(\mathrm{tr}\big(\big(Z'_{d,n} - \tilde{Z}'_{d,n}\big)\big(Z'_{d,n} - \tilde{Z}'_{d,n}\big)^*\big)\big)^{1/2}\big(\mathrm{tr}\big(\tilde{Z}'_{d,n}\big(\tilde{Z}'_{d,n}\big)^*\big)\big)^{1/2}\big\}\|R_{d,n}\|_{S_\infty}. \end{aligned}$$

As in Section A.1, let $p_0 > 0$ be the lower bound on $p_{i,d,n}$, $i = 1, \dots, d$ and $d \in \mathbb{N}$. With this notation, we show that

$$\sup_d \frac{1}{dn} \operatorname{tr}\bigl((Z'_{d,n} - \tilde{Z}'_{d,n})(Z'_{d,n} - \tilde{Z}'_{d,n})^*\bigr) \le \frac{\mathbb{E}X_{11}^2 \mathbb{1}\{|X_{11}| > K\} + \delta}{p_0},$$

while

$$\sup_d \frac{1}{dn} \operatorname{tr}\bigl(\tilde{Z}'_{d,n}(\tilde{Z}'_{d,n})^*\bigr) \le \frac{\delta + \mathbb{E}X_{11}^2}{p_0}.$$

We have

$$\begin{aligned}
\frac{1}{dn} \operatorname{tr}\bigl((Z'_{d,n} - \tilde{Z}'_{d,n})(Z'_{d,n} - \tilde{Z}'_{d,n})^*\bigr) &= \frac{1}{dn} \sum_{i=1}^{d} \sum_{k=1}^{n} (Z'_{ik,d,n} - \tilde{Z}'_{ik,d,n})^2 \\
&\le \frac{1}{dn} \max_{i=1,\dots,d} \left(\frac{1}{p_{i,d,n}}\right) \sum_{i=1}^{d} \mathbb{1}_{\Delta_{i,d,n}} \sum_{k=1}^{n} X_{ik}^2 \mathbb{1}\bigl\{|X_{ik}| > K\bigr\} \\
&\le \frac{\mathbb{E}X_{11}^2 \mathbb{1}\{|X_{11}| > K\} + \delta}{p_0}.
\end{aligned}$$

Moreover,

$$\begin{aligned}
\frac{1}{dn} \operatorname{tr}\bigl(\tilde{Z}'_{d,n}(\tilde{Z}'_{d,n})^*\bigr) &= \frac{1}{dn} \sum_{i=1}^{d} \sum_{k=1}^{n} (\tilde{Z}'_{ik,d,n})^2 \\
&\le \frac{1}{dn} \max_{i=1,\dots,d} \left(\frac{1}{p_{i,d,n}}\right) \sum_{i=1}^{d} \mathbb{1}_{\Delta_{i,d,n}} \sum_{k=1}^{n} X_{ik,d,n}^2 \mathbb{1}\bigl\{|X_{ik,d,n}| \le K\bigr\} \\
&\le \frac{\mathbb{E}X_{11}^2 + \delta}{p_0}.
\end{aligned}$$

As concerns the first summand in (5.1), it holds $\mathbb{P}(\Delta_{i,d,n}) \to 1$ as $d \to \infty$ by weak law of large numbers. Note that $\mathbb{P}(\Delta_{1,d,n}) = \mathbb{P}(\Delta_{2,d,n}) = \dots = \mathbb{P}(\Delta_{d,d,n})$. Then by Hoeffding's inequality for sufficiently large d,

$$\begin{aligned}
\mathbb{P}\left(\sum_{i=1}^{d} \mathbb{1}_{\Delta_{i,d,n}^c} \ge \delta d\right) &\le \mathbb{P}\left(\sum_{i=1}^{d} \bigl(\mathbb{1}_{\Delta_{i,d,n}^c} - \mathbb{P}(\Delta_{i,d,n}^c)\bigr) \ge \frac{1}{2}\delta d\right) \\
&\le \exp\left(-\frac{\delta^2 d}{2}\right).
\end{aligned}$$

Hence, by the Borel–Cantelli lemma

$$\limsup_{d\to\infty} \frac{1}{d} \sum_{i=1}^{d} \mathbb{1}_{\Delta^c_{i,d,n}} < \delta$$

almost surely. As in inequality (A.2) of Section A.4, we deduce

$$\begin{aligned}\limsup_{d\to\infty} d_L\left(\mu^{\hat{\Xi}_{d,n}}, \mu^{\hat{\Xi}'_{d,n}}\right) &\le \limsup_{d\to\infty} d_K\left(\mu^{\hat{\Xi}_{d,n}}, \mu^{\hat{\Xi}'_{d,n}}\right)\\ &\le \limsup_{d\to\infty} \frac{1}{d} \operatorname{rank}\left(\hat{\Xi}_{d,n} - \hat{\Xi}'_{d,n}\right)\\ &\le 2\delta\end{aligned}$$

almost surely. The third summand in (5.1) is bounded in the same way. Putting things together in right-hand side of (5.1),

$$\begin{aligned}&\limsup_{d\to\infty} d_L\left(\mu^{\hat{\Xi}_{d,n}}, \mu^{\tilde{\Xi}_{d,n}}\right)\\ &\quad\le 4\delta + \sup_d \|R_{d,n}\|^{1/2}_{S_\infty}\left[\frac{\mathbb{E}X_{11}^2 \mathbb{1}\{|X_{11}| > K\} + \delta}{p_0} + 2\frac{\sqrt{\mathbb{E}X_{11}^2 \mathbb{1}\{|X_{11}| > K\} + \delta}\sqrt{\delta + \mathbb{E}X_{11}^2}}{p_0}\right]^{1/2}\end{aligned}$$

almost surely. Since δ may be chosen arbitrarily small, we conclude

$$\begin{aligned}&\limsup_{d\to\infty} d_L\left(\mu^{\hat{\Xi}_{d,n}}, \mu^{\tilde{\Xi}_{d,n}}\right)\\ &\quad\le \sup_d \|R_{d,n}\|^{1/2}_{S_\infty}\left[\frac{\mathbb{E}X_{11}^2 \mathbb{1}\{|X_{11}| > K\}}{p_0} + 2\frac{\sqrt{\mathbb{E}X_{11}^2 \mathbb{1}\{|X_{11}| > K\}}\sqrt{\mathbb{E}X_{11}^2}}{p_0}\right]^{1/2}.\end{aligned}$$

In turn, the last expression can be made arbitrary small for K sufficiently large. Since the centralization of the truncated random variables $\tilde{X}_{ik}$ leads to a finite rank perturbation of $\tilde{\Xi}_{d,n}$ (uniformly in d), we may assume the entries of $\tilde{X}_{ik}$ to be centered. In the following, denote the centered truncated random matrix again by $X_{d,n}$. Then, analogously to the truncation step by replacing $\mathbb{1}\{|X_{ik}| \le K\}$ with $(\mathbb{E}X_{11}^2)^{-1/2}$ in the definition of $\tilde{X}$ we may assume the entries to be standardized since the variance of the truncated variables converges to one as the truncation level tends to infinity. Therefore, in the rest of the proof we analyze the matrix

$$\hat{\Xi}_{d,n} = \frac{1}{n} R_{d,n}^{1/2} Z_{d,n} Z_{d,n}^* R_{d,n}^{1/2} - S_{d,n},$$

where the entries of the matrix $Z_{d,n}$ are centered, standardized and bounded.

5.2. Step II: Approximate solution to the fixed point equation (3.1)

Subsequently, we assume that

$$\liminf_{d\to\infty} \frac{d^{3/2}}{n} > 0. \tag{5.2}$$

The general case is treated in Step VI. Recall that $\mu_{d,n}$ denotes the (normalized) spectral measure of $\hat{\Xi}_{d,n}$, and denote its Stieltjes transform by

$$m_{d,n}(z) = \int \frac{1}{\lambda - z} \, \mathrm{d}\mu_{d,n}(\lambda), \qquad z \in \mathbb{C}^+. \tag{5.3}$$

We use subsequently the following abbreviations for the resolvents

$$\hat{G}_{d,n}(z) = (\hat{\Xi}_{d,n} - zI_{d\times d})^{-1} \quad \text{and} \quad \hat{G}^{(k)}_{d,n}(z) = \big(\hat{\Xi}^{(k)}_{d,n} - zI_{d\times d}\big)^{-1}, \qquad k = 1, \dots, n.$$

For $z \in \mathbb{C}^+$, define

$$e_{d,n}(z) = \frac{1}{d} \operatorname{tr}\big\{R_{d,n}\hat{G}_{d,n}(z)\big\}.$$

Our goal in this step is to show that

$$\frac{1}{d} \operatorname{tr}\big(D^{-1}_{d,n}(z)\big) - m_{d,n}(z) \to 0 \qquad \text{a.s.,} \quad \text{and} \tag{5.4}$$

$$\frac{1}{d} \operatorname{tr}\big(R_{d,n}D^{-1}_{d,n}(z)\big) - e_{d,n}(z) \to 0 \qquad \text{a.s.} \tag{5.5}$$

with

$$D_{d,n}(z) = \frac{1}{1 + (d/n)e_{d,n}(z)} R_{d,n} - S_{d,n} - zI_{d\times d}. \tag{5.6}$$

Let $\hat{\Xi}_{d,n} = O_{d,n}\Lambda_{d,n}O^*_{d,n}$ denote a spectral decomposition, where

$$\Lambda_{d,n} = \operatorname{diag}(\lambda_{1,d,n}, \dots, \lambda_{d,d,n}),$$

and define $\underline{R}_{d,n} = O^*_{d,n}R_{d,n}O_{d,n}$. With this notation,

$$\begin{aligned} e_{d,n}(z) &= \frac{1}{d} \operatorname{tr}\big\{R_{d,n}\hat{G}_{d,n}(z)\big\} \\ &= \frac{1}{d} \operatorname{tr}\big\{R_{d,n}\big(O_{d,n}\Lambda_{d,n}O^*_{d,n} - zI_{d\times d}\big)^{-1}\big\} \\ &= \frac{1}{d} \operatorname{tr}\big\{R_{d,n}\big(O_{d,n}[\Lambda_{d,n} - zI_{d\times d}]O^*_{d,n}\big)^{-1}\big\} \\ &= \frac{1}{d} \operatorname{tr}\big\{R_{d,n}\big(O_{d,n}[\Lambda_{d,n} - zI_{d\times d}]^{-1}O^*_{d,n}\big)\big\} \end{aligned} \tag{5.7}$$

$$
\begin{aligned}
&= \frac{1}{d} \operatorname{tr}\bigl\{ O_{d,n}^* R_{d,n} O_{d,n} (\Lambda_{d,n} - z I_{d\times d})^{-1} \bigr\} \\
&= \frac{1}{d} \operatorname{tr}\bigl\{ \underline{R}_{d,n} (\Lambda_{d,n} - z I_{d\times d})^{-1} \bigr\} \\
&= \frac{1}{d} \sum_{i=1}^{d} \frac{\underline{R}_{ii,d,n}}{\lambda_{i,d,n} - z}.
\end{aligned}
$$

Since $R_{d,n}$ and therefore $\underline{R}_{d,n}$ are positive semidefinite, the diagonal entries $\underline{R}_{ii,d,n}$, $i = 1, \dots, d$, are non-negative. Hence, $e_{d,n}$ is the Stieltjes transform of a measure on $\mathbb{R}$ with at most d support points and total mass

$$
\frac{1}{d} \operatorname{tr} R_{d,n}.
$$

Note that $\hat{\Xi}_{d,n}$ is not necessarily positive semidefinite, hence the support points are not restricted to $[0, \infty)$. As a Stieltjes transform,

$$
e_{d,n} : \mathbb{C}^+ \to \mathbb{C}^+. \tag{5.8}
$$

This implies in particular that $D_{d,n}(z)$ as defined in (5.6) is in fact invertible by means of Lemma C.3. Moreover, since $\|R_{d,n}\|_{S_\infty} \le \kappa$ for some constant $\kappa > 0$, it follows by Hölder's inequality and the positive definiteness of $R_{d,n}$,

$$
\begin{aligned}
\bigl| e_{d,n}(z) \bigr| &\le \frac{1}{d} \|\underline{R}_{d,n}\|_{S_1} \bigl\| (\Lambda_{d,n} - z I_{d\times d})^{-1} \bigr\|_{S_\infty} \\
&= \left(\frac{1}{d} \operatorname{tr} R_{d,n} \right) \max_{1 \le i \le d} \frac{1}{|\lambda_{i,d,n} - z|} \\
&\le \frac{\kappa}{\Im z}.
\end{aligned} \tag{5.9}
$$

Let $Z_{k,d,n}$ be the kth column of the matrix $Z_{d,n}$, and define

$$
Y_{k,d,n} = \frac{1}{\sqrt{n}} R_{d,n}^{1/2} Z_{k,d,n} \quad \text{and} \quad \hat{\Xi}_{d,n}^{(k)} = \hat{\Xi}_{d,n} - Y_{k,d,n} Y_{k,d,n}^*, \qquad k = 1, \dots, n,
$$

which arises from $\hat{\Xi}_{d,n}$ by taking away the kth sample vector, and recall (5.6). Then,

$$
\hat{\Xi}_{d,n} - z I_{d\times d} - D_{d,n}(z) = \sum_{k=1}^{n} Y_{k,d,n} Y_{k,d,n}^* - \frac{1}{1 + (d/n) e_{d,n}(z)} R_{d,n},
$$

whence

$$
\begin{aligned}
D_{d,n}(z) \bigl\{ \hat{G}_{d,n}(z) - D_{d,n}^{-1}(z) \bigr\} (\hat{\Xi}_{d,n} - z I_{d\times d}) &= D_{d,n}(z) - (\hat{\Xi}_{d,n} - z I_{d\times d}) \\
&= \frac{1}{1 + (d/n) e_{d,n}(z)} R_{d,n} - \sum_{k=1}^{n} Y_{k,d,n} Y_{k,d,n}^*.
\end{aligned}
$$

Therefore,

$$\begin{aligned}\hat{G}_{d,n}(z) - D_{d,n}^{-1}(z) &= -\sum_{k=1}^{n} D_{d,n}^{-1}(z) Y_{k,d,n} Y_{k,d,n}^{*} \hat{G}_{d,n}(z) \\ &\quad + \frac{1}{1+(d/n)e_{d,n}(z)} D_{d,n}^{-1}(z) R_{d,n} \hat{G}_{d,n}(z) \\ &= -\sum_{k=1}^{n} \frac{D_{d,n}^{-1}(z) Y_{k,d,n} Y_{k,d,n}^{*} \hat{G}_{d,n}^{(k)}(z)}{1 + Y_{k,d,n}^{*} \hat{G}_{d,n}^{(k)}(z) Y_{k,d,n}} \\ &\quad + \frac{1}{1+(d/n)e_{d,n}(z)} D_{d,n}^{-1}(z) R_{d,n} \hat{G}_{d,n}(z),\end{aligned} \tag{5.10}$$

where (5.10) follows from Lemma C.1. Altogether,

$$\frac{1}{d} \operatorname{tr}\bigl(D_{d,n}^{-1}(z)\bigr) - m_{d,n}(z) = \frac{1}{n} \sum_{k=1}^{n} f_{k,m} \tag{5.11}$$

with

$$f_{k,m} = \frac{1}{d} \frac{Z_{k,d,n}^{*} R_{d,n}^{1/2} \hat{G}_{d,n}^{(k)}(z) D_{d,n}^{-1}(z) R_{d,n}^{1/2} Z_{k,d,n}}{1 + Y_{k,d,n}^{*} \hat{G}_{d,n}^{(k)}(z) Y_{k,d,n}} - \frac{1}{d} \frac{\operatorname{tr}(R_{d,n} \hat{G}_{d,n}(z) D_{d,n}^{-1}(z))}{1+(d/n)e_{d,n}(z)}.$$

Multiplication of the matrix equality (5.10) with $R_{d,n}$ from the right, we deduce

$$\frac{1}{d} \operatorname{tr}\bigl(R_{d,n} D_{d,n}^{-1}(z)\bigr) - e_{d,n}(z) = \frac{1}{n} \sum_{k=1}^{n} f_{k,e} \tag{5.12}$$

with

$$\begin{aligned} f_{k,e} = \frac{1}{d} &\frac{Z_{k,d,n}^{*} R_{d,n}^{1/2} \hat{G}_{d,n}^{(k)}(z) R_{d,n} D_{d,n}^{-1}(z) R_{d,n}^{1/2} Z_{k,d,n}}{1 + Y_{k,d,n}^{*} \hat{G}_{d,n}^{(k)}(z) Y_{k,d,n}} \\ &- \frac{1}{d} \frac{\operatorname{tr}(R_{d,n} \hat{G}_{d,n}(z) R_{d,n} D_{d,n}^{-1}(z))}{1+(d/n)e_{d,n}(z)}. \end{aligned}$$

Subsequently, we show that

$$\lim_{d\to\infty} \frac{1}{n} \sum_{k=1}^{n} f_{k,x} = 0 \qquad \text{a.s., } x = e, m. \tag{5.13}$$

First, observe that

$$Y_{k,d,n}^{*} \hat{G}_{d,n}^{(k)}(z) Y_{k,d,n} = \operatorname{tr}\bigl(Y_{k,d,n} Y_{k,d,n}^{*} \hat{G}_{d,n}^{(k)}(z)\bigr)$$

is the Stieltjes transform of a measue on $\mathbb{R}$ with total mass $\|Y_{k,d,n}\|_2^2$, following the arguments in (5.7). Next, with $\lambda_{1,d,n}^{(k)}, \dots, \lambda_{d,d,n}^{(k)}$ denoting the eigenvalues of $\hat{\Xi}_{d,n}^{(k)}$,

$$\begin{aligned} \|\hat{G}_{d,n}^{(k)}(z)\|_{S_\infty} &= \max_{i=1,\dots,d} \frac{1}{\sqrt{(\lambda_{i,d,n}^{(k)} - \Re z)^2 + (\Im z)^2}} \\ &\le \frac{1}{\Im z}. \end{aligned} \tag{5.14}$$

The same holds true for $\hat{G}_{d,n}(z)$ in place of $\hat{G}_{d,n}^{(k)}(z)$. Therefore,

$$\left| Y_{k,d,n}^* \hat{G}_{d,n}^{(k)}(z) Y_{k,d,n} \right| \le \frac{\|Y_k\|_2^2}{\Im z},$$

which gives

$$\left| \frac{1}{1 + Y_{k,d,n}^* \hat{G}_{d,n}^{(k)}(z) Y_{k,d,n}} \right| \le \frac{1}{1 - \|Y_{k,d,n}\|_2^2/\Im z} \qquad \text{if } \frac{\|Y_{k,d,n}\|_2^2}{\Im z} < 1. \tag{5.15}$$

Denoting with $O \Lambda O^*$ a spectral decomposition of $\hat{\Xi}_{d,n}^{(k)}$ and $V_{ii}^{(k)} = (O^* Y_{k,d,n} Y_{k,d,n}^* O)_{ii}$ for the moment, we obtain for $\|Y_{k,d,n}\|_2 > 0$ the bound

$$\begin{aligned} \left| \frac{1}{1 + Y_{k,d,n}^* \hat{G}_{d,n}^{(k)}(z) Y_{k,d,n}} \right| &\le \frac{1}{\Im(Y_{k,d,n}^* \hat{G}_{d,n}^{(k)}(z) Y_{k,d,n})} \\ &= \frac{1}{\Im z \sum_{i=1}^d V_{ii}^{(k)}/((\lambda_{i,d,n}^{(k)} - \Re(z))^2 + (\Im z)^2)} \\ &\le \frac{1}{\Im z \sum_{i=1}^d V_{ii}^{(k)}/(2 \max_i |\lambda_{i,d,n}^{(k)}|^2 + 2|z|^2)} \\ &\le \frac{2 \max_i |\lambda_{i,d,n}^{(k)}|^2 + 2|z|^2}{\Im z \|Y_k\|_2^2}. \end{aligned} \tag{5.16}$$

Combining the first bound (5.15) in case $\|Y_{k,d,n}\|_2^2/\Im z \le 1/2$ with the second bound (5.16) if $\|Y_{k,d,n}\|_2^2/\Im z > 1/2$ yields

$$\begin{aligned} \left| \frac{1}{1 + Y_{k,d,n}^* \hat{G}_{d,n}^{(k)}(z) Y_{k,d,n}} \right| &\le 2 \left\{ \frac{\max_i |\lambda_{i,d,n}^{(k)}|^2 + |z|^2}{(\Im z)^2} + 1 \right\} \\ &\le \frac{2 \max_i |\lambda_{i,d,n}^{(k)}|^2 + 4|z|^2}{(\Im z)^2}. \end{aligned} \tag{5.17}$$

Finally, due to

$$\|\hat{\Xi}_{d,n}\|_{S_\infty} \le \|S_{d,n}\|_{S_\infty} + \Bigg\|\sum_{\substack{l=1\\ l\neq k}}^{n} Y_{l,d,n}Y_{l,d,n}^*\Bigg\|_{S_\infty}$$

and Lemma C.7,

$$\limsup_{d\to\infty}\left\{\left(\frac{2c+4|z|^2}{(\Im z)^2}\right)^{-1}\left|\frac{1}{1+Y_{k,d,n}^*\hat{G}_{d,n}^{(k)}(z)Y_{k,d,n}}\right|\right\} \le C < \infty \tag{5.18}$$

almost surely for some constants $C, c > 0$. Define

$$e_{d,n}^{(k)} = \frac{1}{d}\operatorname{tr}\big(R_{d,n}\hat{G}_{d,n}^{(k)}(z)\big), \qquad k \in \{1,\dots,n\}.$$

Note that analogously to (5.7), $e_{d,n}^{(k)}$ is a Stieltjes transform. Using (5.14) and the arguments of (5.15) for the case $n^{-1}\operatorname{tr}(R_{d,n})/\Im z \le 1/2$ as well as (5.16) for $n^{-1}\operatorname{tr}(R_{d,n})/\Im z > 1/2$, we obtain analogously

$$\left|\frac{1}{1+(d/n)e_{d,n}^{(k)}(z)}\right| \le \frac{2\max_i|\lambda_{i,d,n}^{(k)}|^2+4|z|^2}{(\Im z)^2} \tag{5.19}$$

and for some constants $C, c > 0$

$$\limsup_{d\to\infty}\left\{\left(\frac{2c+4|z|^2}{(\Im z)^2}\right)^{-1}\left|\frac{1}{1+(d/n)e_{d,n}^{(k)}(z)}\right|\right\} \le C < \infty. \tag{5.20}$$

The same bound holds true for $e_{d,n}$ instead of $e_{d,n}^{(k)}$, in which case $\lambda_{i,d,n}^{(k)}$ are to be replaced by the eigenvalues $\lambda_{i,d,n}$ of $\hat{\Xi}_{d,n}$. Therefore, with

$$\psi_{d,n}^{(k)} = \max_{i=1,\dots,d}\big\{\big(\lambda_{i,d,n}^{(k)}\big)^2, \lambda_{i,d,n}^2\big\},$$

$$\begin{aligned}
&\left|\frac{1}{1+(d/n)e_{d,n}(z)} - \frac{1}{1+(d/n)e_{d,n}^{(k)}(z)}\right| \\
&\quad = \frac{d}{n}\cdot\frac{|e_{d,n}^{(k)}(z)-e_{d,n}(z)|}{|(1+(d/n)e_{d,n}(z))(1+(d/n)e_{d,n}^{(k)}(z))|} \\
&\quad \le \frac{d}{n}\frac{\|R_{d,n}\|_{S_\infty}}{\Im z}\frac{1}{d}\frac{1}{|(1+(d/n)e_{d,n}(z))(1+(d/n)e_{d,n}^{(k)}(z))|} && (5.21)\\
&\quad \le \frac{1}{n}\frac{\|R_{d,n}\|_{S_\infty}}{\Im z}\left(\frac{2\psi_{d,n}^{(k)}+4|z|^2}{(\Im z)^2}\right)^2, && (5.22)
\end{aligned}$$

where inequality (5.21) follows from Lemma C.2 and (5.22) results from (5.19). Furthermore, with

$$D_{d,n}^{(k)}(z) = \frac{1}{1+(d/n)e_{d,n}^{(k)}(z)} R_{d,n} - S_{d,n} - zI_{d\times d}, \tag{5.23}$$

it follows from Lemma C.3 that

$$\left\| D_{d,n}^{-1}(z) \right\|_{S_\infty} \le \frac{1}{\Im z} \qquad \text{as well as } \left\| \left(D_{d,n}^{(k)}(z)\right)^{-1} \right\|_{S_\infty} \le \frac{1}{\Im z}. \tag{5.24}$$

We begin with establishing (5.13). To this aim, let

$$E_{x,d,n} = \begin{cases} I_{d\times d}, & \text{for } x = m, \\ R_{d,n}, & \text{for } x = e. \end{cases}$$

We decompose

$$f_{k,x} = f_{k,x}^{[1]} + f_{k,x}^{[2]} + f_{k,x}^{[3]} + f_{k,x}^{[4]},$$

where

$$\begin{aligned}
f_{k,x}^{[1]} &= \frac{1}{d} \frac{Z_{k,d,n}^{*} R_{d,n}^{1/2} \hat{G}_{d,n}^{(k)}(z) E_{x,d,n} D_{d,n}^{-1}(z) R_{d,n}^{1/2} Z_{k,d,n}}{1 + Y_{k,d,n}^{*} \hat{G}_{d,n}^{(k)}(z) Y_{k,d,n}} \\
&\quad - \frac{1}{d} \frac{Z_{k,d,n}^{*} R_{d,n}^{1/2} \hat{G}_{d,n}^{(k)}(z) E_{x,d,n} (D_{d,n}^{(k)}(z))^{-1} R_{d,n}^{1/2} Z_{k,d,n}}{1 + Y_{k,d,n}^{*} \hat{G}_{d,n}^{(k)}(z) Y_{k,d,n}}, \\
f_{k,x}^{[2]} &= \frac{1}{d} \frac{Z_{k,d,n}^{*} R_{d,n}^{1/2} \hat{G}_{d,n}^{(k)}(z) E_{x,d,n} (D_{d,n}^{(k)}(z))^{-1} R_{d,n}^{1/2} Z_{k,d,n}}{1 + Y_{k,d,n}^{*} \hat{G}_{d,n}^{(k)}(z) Y_{k,d,n}} \\
&\quad - \frac{1}{d} \frac{\operatorname{tr}(R_{d,n} \hat{G}_{d,n}^{(k)}(z) E_{x,d,n} (D_{d,n}^{(k)}(z))^{-1})}{1 + Y_{k,d,n}^{*} \hat{G}_{d,n}^{(k)}(z) Y_{k,d,n}}, \\
f_{k,x}^{[3]} &= \frac{1}{d} \frac{\operatorname{tr}(R_{d,n} \hat{G}_{d,n}^{(k)}(z) E_{x,d,n} (D_{d,n}^{(k)}(z))^{-1})}{1 + Y_{k,d,n}^{*} \hat{G}_{d,n}^{(k)}(z) Y_{k,d,n}} \\
&\quad - \frac{1}{d} \frac{\operatorname{tr}(R_{d,n} \hat{G}_{d,n}(z) E_{x,d,n} D_{d,n}^{-1}(z))}{1 + Y_{k,d,n}^{*} \hat{G}_{d,n}^{(k)}(z) Y_{k,d,n}}, \\
f_{k,x}^{[4]} &= \frac{1}{d} \frac{\operatorname{tr}(R_{d,n} \hat{G}_{d,n}(z) E_{x,d,n} D_{d,n}^{-1}(z))}{1 + Y_{k,d,n}^{*} \hat{G}_{d,n}^{(k)}(z) Y_{k,d,n}} \\
&\quad - \frac{1}{d} \frac{\operatorname{tr}(R_{d,n} \hat{G}_{d,n}(z) E_{x,d,n} D_{d,n}^{-1}(z))}{1 + (d/n) e_{d,n}(z)}.
\end{aligned}$$

Using Lemma C.1 in (5.25) as well as the spectral norm bounds (5.22), (5.14) and (5.24) in (5.27), we obtain

$$
\begin{aligned}
|f_{k,x}^{[1]}| &= \left| \frac{1}{d} \frac{Z_{k,d,n}^* R_{d,n}^{1/2} \hat{G}_{d,n}^{(k)}(z) E_{x,d,n} [D_{d,n}^{-1}(z) - (D_{d,n}^{(k)}(z))^{-1}] R_{d,n}^{1/2} Z_{k,d,n}}{1 + Y_{k,d,n}^* \hat{G}_{d,n}^{(k)}(z) Y_{k,d,n}} \right| \\
&= \left| \frac{n}{d} Y_{k,d,n}^* \hat{G}_{d,n}(z) E_{x,d,n} \big[D_{d,n}^{-1}(z) - \big(D_{d,n}^{(k)}(z)\big)^{-1} \big] Y_{k,d,n} \right| \qquad (5.25) \\
&= \left| \frac{n}{d} Y_{k,d,n}^* \hat{G}_{d,n}(z) E_{x,d,n} \big(D_{d,n}^{(k)}(z)\big)^{-1} \right. \\
&\qquad \left. \times \big[D_{d,n}^{(k)}(z) - D_{d,n}(z) \big] D_{d,n}^{-1}(z) Y_{k,d,n} \right| \\
&\le \frac{n}{d} \|Y_{k,d,n}\|_2^2 \big\| \hat{G}_{d,n}(z) \big\|_{S_\infty} \|E_{x,d,n}\|_{S_\infty} \qquad (5.26) \\
&\qquad \times \big\| D_{d,n}^{-1}(z) \big\|_{S_\infty} \big\| D_{d,n}^{(k)}(z) - D_{d,n} \big\|_{S_\infty} \big\| \big(D_{d,n}^{(k)}(z)\big)^{-1} \big\|_{S_\infty} \\
&\le \frac{1}{d} \|Y_{k,d,n}\|_2^2 \frac{(2\psi_{d,n}^{(k)} + 4|z|^2)^2 \|R_{d,n}\|_{S_\infty}^2 \|E_{x,d,n}\|_{S_\infty}}{(\Im z)^8} \qquad (5.27) \\
&\le \frac{1}{dn} \|Z_{k,d,n}\|_2^2 \frac{(2\psi_{d,n}^{(k)} + 4|z|^2)^2 \|R_{d,n}\|_{S_\infty}^3 \|E_{x,d,n}\|_{S_\infty}}{(\Im z)^8}.
\end{aligned}
$$

By (5.17),

$$
\begin{aligned}
|f_{k,x}^{[2]}| &= \left| \frac{1}{d} \frac{\operatorname{tr}[(R_{d,n}^{1/2} Z_{k,d,n} Z_{k,d,n}^* R_{d,n}^{1/2} - R_{d,n}) \hat{G}_{d,n}^{(k)}(z) E_{x,d,n} (D_{d,n}^{(k)}(z))^{-1}]}{1 + Y_{k,d,n}^* \hat{G}_{d,n}^{(k)}(z) Y_{k,d,n}} \right| \\
&\le \frac{2 \max_i |\lambda_{i,d,n}^{(k)}|^2 + 4|z|^2}{(\Im z)^2} \\
&\qquad \times \frac{1}{d} \big| \operatorname{tr}\big[\big(R_{d,n}^{1/2} Z_{k,d,n} Z_{k,d,n}^* R_{d,n}^{1/2} - R_{d,n} \big) \hat{G}_{d,n}^{(k)}(z) E_{x,d,n} \big(D_{d,n}^{(k)}(z)\big)^{-1} \big] \big|.
\end{aligned}
$$

Furthermore, using (5.17) in (5.28), the invariance of the trace under cyclic permutation and Lemma C.2 in (5.29) for the first term in the curly brackets and the spectral norm bounds (5.22), (5.14) and (5.24) in (5.30) yields the bound

$$
\begin{aligned}
|f_{k,x}^{[3]}| &= \left| \frac{1}{d} \frac{\operatorname{tr}[R_{d,n} (\hat{G}_{d,n}^{(k)}(z) E_{x,d,n} (D_{d,n}^{(k)}(z))^{-1} - \hat{G}_{d,n}(z) E_{x,d,n} D_{d,n}^{-1}(z))]}{1 + Y_{k,d,n}^* \hat{G}_{d,n}^{(k)}(z) Y_{k,d,n}} \right| \\
&\le \frac{2 \max_i |\lambda_{i,d,n}^{(k)}|^2 + 4|z|^2}{(\Im z)^2} \qquad (5.28)
\end{aligned}
$$

$$
\begin{aligned}
&\times \left\{ \frac{1}{d} \big| \mathrm{tr}\big[R_{d,n} \big(\hat{G}^{(k)}_{d,n}(z) - \hat{G}_{d,n}(z) \big) E_{x,d,n} \big(D^{(k)}_{d,n}(z) \big)^{-1} \big] \big| \right. \\
&\left. + \frac{1}{d} \big| \mathrm{tr}\big[R_{d,n} \hat{G}_{d,n}(z) E_{x,d,n} \big(\big(D^{(k)}_{d,n}(z) \big)^{-1} - D^{-1}_{d,n}(z) \big) \big] \big| \right\} \\
&\leq \frac{2 \max_i |\lambda^{(k)}_{i,d,n}|^2 + 4|z|^2}{(\Im z)^2} \left\{ \frac{1}{d} \frac{\| E_{x,d,n} (D^{(k)}_{d,n}(z))^{-1} R_{d,n} \|_{S_\infty}}{\Im z} \right. \qquad (5.29) \\
&\left. + \frac{1}{d} \big| \mathrm{tr}\big[R_{d,n} \hat{G}_{d,n}(z) E_{x,d,n} D^{-1}_{d,n}(z) \big[D_{d,n}(z) - D^{(k)}_{d,n}(z) \big] \big(D^{(k)}_{d,n}(z) \big)^{-1} \big] \big| \right\} \\
&\leq \frac{1}{d} \frac{2 \max_i |\lambda^{(k)}_{i,d,n}|^2 + 4|z|^2}{(\Im z)^2} \frac{1}{(\Im z)^2} \| R_{d,n} \|_{S_\infty} \| E_{x,d,n} \|_{S_\infty} \qquad (5.30) \\
&+ \frac{1}{n} \left(\frac{2\psi^{(k)}_{d,n} + 4|z|^2}{(\Im z)^2} \right)^3 \frac{\| R_{d,n} \|^3_{S_\infty}}{(\Im z)^4} \| E_{x,d,n} \|_{S_\infty}.
\end{aligned}
$$

Finally, using (5.14) and (5.24) in (5.31), (5.17) and (5.19) in (5.32) and Lemma C.2 in (5.33),

$$
\begin{aligned}
\big| f^{[4]}_{k,x} \big| &= \frac{1}{d} \big| \mathrm{tr}\big(R_{d,n} \hat{G}_{d,n}(z) E_{x,d,n} D^{-1}_{d,n}(z) \big) \big| \\
&\quad \times \left| \frac{n^{-1} \mathrm{tr}[R^{1/2}_{d,n} \hat{G}_{d,n}(z) R^{1/2}_{d,n}] - n^{-1} Z^*_{k,d,n} R^{1/2}_{d,n} \hat{G}^{(k)}_{d,n}(z) R^{1/2}_{d,n} Z_{k,d,n}}{(1 + Y^*_{k,d,n} \hat{G}^{(k)}_{d,n}(z) Y_{k,d,n})(1 + (d/n) e_{d,n}(z))} \right| \\
&\leq \frac{1}{d} \frac{\| R_{d,n} \|_{S_1} \| E_{x,d,n} \|_{S_\infty}}{(\Im z)^2} \qquad (5.31) \\
&\quad \times \left| \frac{n^{-1} \mathrm{tr}[R^{1/2}_{d,n} \hat{G}_{d,n}(z) R^{1/2}_{d,n}] - n^{-1} Z^*_{k,d,n} R^{1/2}_{d,n} \hat{G}^{(k)}_{d,n}(z) R^{1/2}_{d,n} Z_{k,d,n}}{(1 + Y^*_{k,d,n} \hat{G}^{(k)}_{d,n}(z) Y_{k,d,n})(1 + (d/n) e_{d,n}(z))} \right| \\
&\leq \frac{1}{d} \frac{\| R_{d,n} \|_{S_1} \| E_{x,d,n} \|_{S_\infty}}{(\Im z)^2} \left(\frac{2(\psi^{(k)}_{d,n})^2 + 4|z|^2}{(\Im z)^2} \right)^2 \qquad (5.32) \\
&\quad \times \left| \frac{1}{n} \mathrm{tr}\big[R^{1/2}_{d,n} \hat{G}_{d,n}(z) R^{1/2}_{d,n} \big] - \frac{1}{n} Z^*_{k,d,n} R^{1/2}_{d,n} \hat{G}^{(k)}_{d,n}(z) R^{1/2}_{d,n} Z_{k,d,n} \right| \\
&\leq \frac{1}{d} \frac{\| R_{d,n} \|_{S_1} \| E_{x,d,n} \|_{S_\infty}}{(\Im z)^2} \left(\frac{2(\psi^{(k)}_{d,n})^2 + 4|z|^2}{(\Im z)^2} \right)^2 \qquad (5.33) \\
&\quad \times \left\{ \left| \frac{1}{n} \mathrm{tr}\big[R^{1/2}_{d,n} \hat{G}^{(k)}_{d,n}(z) R^{1/2}_{d,n} \big] - \frac{1}{n} Z^*_{k,d,n} R^{1/2}_{d,n} \hat{G}^{(k)}_{d,n}(z) R^{1/2}_{d,n} Z_{k,d,n} \right| \right. \\
&\quad \left. + \frac{1}{n} \frac{\| R_{d,n} \|_{S_\infty}}{\Im z} \right\}.
\end{aligned}
$$

Based on these estimates on $f_{k,x}^{[l]}$, $l = 1, 2, 3, 4$, we are ready to prove (5.4) and (5.5). In the next display, $c > 0$ denotes a constant depending only on the support of Z_{11}, and may change from line to line. By means of Lemma C.4, Lemma C.5, Lemma C.6 and the spectral norm bounds (5.14) and (5.24),

$$\begin{aligned}
\mathbb{E}\big|f_{k,x}^{[1]}\big|^6 &\le \frac{\|R_{d,n}\|_{S_\infty}^{18}\|E_{x,d,n}\|_{S_\infty}^{6}}{n^6 d^6 (\Im z)^{48}}\mathbb{E}\big\{\|Z_{k,d,n}\|_2^{12}\big(2\psi_{d,n}^{(k)} + 4|z|^2\big)^6\big\} \\
&\le 2^{17}\frac{\|R_{d,n}\|_{S_\infty}^{18}\|E_{x,d,n}\|_{S_\infty}^{6}}{d^6 n^6 (\Im z)^{48}}\big(\mathbb{E}\|Z_{k,d,n}\|_2^{24}\big)^{1/2} \\
&\quad \times \left[\left(\mathbb{E}\left(\|S_{d,n}\|_{S_\infty}^2 + \max\left\{\left\|\sum_{l=1}^n Y_l Y_l^*\right\|_{S_\infty}^2, \left\|\sum_{\substack{l=1\\ l\neq k}}^n Y_l Y_l^*\right\|_{S_\infty}^2\right\}\right)^{12}\right)^{1/2} + |z|^{12}\right] \\
&\le c\frac{\|R_{d,n}\|_{S_\infty}^{18}\|E_{x,d,n}\|_{S_\infty}^{6}}{n^6 (\Im z)^{48}}\big(\|S_{d,n}\|_{S_\infty}^{12} + \|R_{d,n}\|_{S_\infty}^{12} + |z|^{12}\big), \\
\mathbb{E}\big|f_{k,x}^{[2]}\big|^6 &\le \frac{c}{d^3}\mathbb{E}\left\{\left(\frac{2\max_i|\lambda_{i,d,n}^{(k)}|^2 + 4|z|^2}{(\Im z)^2}\right)^6\right. \\
&\quad \left.\times \big\|R_{d,n}^{1/2}\hat{G}_{d,n}^{(k)}(z)E_{x,d,n}\big(D_{d,n}^{(k)}(z)\big)^{-1}R_{d,n}^{1/2}\big\|_{S_\infty}^6\right\} \\
&\le \frac{c}{d^3(\Im z)^{24}}\|R_{d,n}\|_{S_\infty}^6\|E_{x,d,n}\|_{S_\infty}^6\big(\|S_{d,n}\|_{S_\infty}^{12} + \|R_{d,n}\|_{S_\infty}^{12} + |z|^{12}\big), \\
\mathbb{E}\big|f_{k,x}^{[3]}\big|^6 &\le \frac{c}{(\Im z)^{24}d^6}\|R_{d,n}\|_{S_\infty}^6\|E_{x,d,n}\|_{S_\infty}^6\big(\|S_{d,n}\|_{S_\infty}^{12} + \|R_{d,n}\|_{S_\infty}^{12} + |z|^{12}\big) \\
&\quad + \frac{c}{(\Im z)^{60}n^6}\|R_{d,n}\|_{S_\infty}^{18}\|E_{x,d,n}\|_{S_\infty}^6\big(\|S_{d,n}\|_{S_\infty}^{36} + \|R_{d,n}\|_{S_\infty}^{36} + |z|^{36}\big), \\
\mathbb{E}\big|f_{k,x}^{[4]}\big|^6 &\le \frac{cd^3\|R_{d,n}\|_{S_\infty}^{12}\|E_{x,d,n}\|_{S_\infty}^{6}}{n^6(\Im z)^{42}}\big(\|S_{d,n}\|_{S_\infty}^{24} + \|R_{d,n}\|_{S_\infty}^{24} + |z|^{24}\big).
\end{aligned}$$

In order to show finally (5.13), it remains to note that for any $\varepsilon > 0$,

$$\begin{aligned}
\sum_{d=1}^\infty \mathbb{P}\left(\left|\frac{1}{n}\sum_{k=1}^n f_{k,x}\right| > \varepsilon\right) &\le \sum_{d=1}^\infty\sum_{k=1}^n\sum_{l=1}^4 \mathbb{P}\big(\big|f_{k,x}^{[l]}\big| > \varepsilon/4\big) \\
&\le \sum_{d=1}^\infty\sum_{k=1}^n\sum_{l=1}^4 \left(\frac{\varepsilon}{4}\right)^{-6}\mathbb{E}\big|f_{k,x}^{[l]}\big|^6 < \infty
\end{aligned}$$

by an application of the union bound, Markov's inequality, and (5.2). (5.13) is then a consequence of the Borel–Cantelli lemma.

5.3. Step III: Existence and uniqueness of $e^\circ_{d,n}$

We show that for any d, n and $R_{d,n}$, there exists a unique $e(z) \in \mathbb{C}^+$ which solves the fixed point equation

$$e^\circ_{d,n}(z) = \frac{1}{d}\operatorname{tr}\left\{R_{d,n}\left(\frac{1}{1+(d/n)e^\circ_{d,n}(z)}R_{d,n} - S_{d,n} - zI_{d\times d}\right)^{-1}\right\}, \qquad z \in \mathbb{C}^+. \tag{5.34}$$

To this end, define for any fixed d, n the subsequences $(d_l)_{l\in\mathbb{N}}$ and $(n_l)_{l\in\mathbb{N}}$, where $d_l = ld$ and $n_l = ln$, $l \in \mathbb{N}$, and correspondingly the l-block diagonal matrices

$$R_{(d,n)_l} = \operatorname{diag}(R_{d,n}, \dots, R_{d,n}) \quad \text{and} \quad S_{(d,n)_l} = \operatorname{diag}(S_{d,n}, \dots, S_{d,n})$$

of size $dl \times dl$. Note that the right-hand side of (5.34) remains unchanged when replacing $d, n, R_{d,n}$ and $I_{d\times d}$ by $d_l, n_l, R_{(d,n)_l}$ and $I_{d_l\times d_l}$. By (5.5) of the previous section,

$$e_{(d,n)_l}(z) - \frac{1}{d_l}\operatorname{tr}\left\{R_{(d,n)_l}\left(\frac{1}{1+(d_l/n_l)e_{(d,n)_l}(z)}R_{(d,n)_l} - S_{(d,n)_l} - zI_{d_l\times d_l}\right)^{-1}\right\} \to 0$$

a.s. as $l \to \infty$ with

$$e_{(d,n)_l}(z) = \frac{1}{d_l}\operatorname{tr}\left\{R_{(d,n)_l}(\hat{\Xi}_{(d,n)_l} - zI_{d_l\times d_l})^{-1}\right\},$$

where

$$\hat{\Xi}_{(d,n)_l} = \frac{1}{nl}\sum_{k=1}^{nl} R^{1/2}_{(d,n)_l}\check{Z}_{k,d_l,n_l}\check{Z}^*_{k,d_l,n_l}R^{1/2}_{(d,n)_l} - S_{(d,n)_l},$$

$\check{Z} = (\check{Z}_{ik})_{i,k\in\mathbb{N}}$ is a double array of i.i.d. Rademacher variables, and $\check{Z}_{k,d,n}$ is the kth column of the submatrix $\check{Z}_{d,n} = (\check{Z}_{ik,d,n})_{i\le d,k\le n}$. Consider a realization of these random variables where this convergence occurs. First note by (5.9),

$$\left|e_{(d,n)_l}(z)\right| \le \frac{\kappa}{\Im z} \qquad \forall l \in \mathbb{N}.$$

By Bolzano–Weierstraß, there exists a convergent subsequence of $(e_{(d,n)_l})$ with limit $e(z)$, say, such that in particular

$$\frac{1}{1+(d_l/n_l)e_{(d,n)_l}(z)} \to \frac{1}{1+(d/n)e(z)} \tag{5.35}$$

along this subsequence due to (5.20) for $e_{(d,n)_l}(z)$. By (5.5), $e(z)$ solves the fixed point equation (5.34). As $\Im(e_{(d,n)_l}(z)) > 0$ for any $l \in \mathbb{N}$ and $z \in \mathbb{C}^+$, it follows that its limit satisfies $\Im(e(z)) \ge 0$ and therefore $\Im(e(z)) > 0$, because $\Im(e(z)) = 0$ contradicts with $e(z)$ being a solution of the fixed point equation. Consequently, any such solution e of (5.34) enjoys the following two properties:

$$e : \mathbb{C}^+ \to \mathbb{C}^+ \tag{5.36}$$

and

$$\left|e(z)\right| \le \frac{\kappa}{\Im z} \qquad \forall z \in \mathbb{C}^+. \tag{5.37}$$

It remains to show uniqueness. Denoting

$$\tilde{D}_{d,n}(z) = \tilde{D}_{d,n}\big(z, e(z)\big) = \frac{1}{1+(d/n)e(z)} R_{d,n} - S_{d,n} - zI_{d\times d}, \tag{5.38}$$

we obtain the representation

$$\begin{aligned} e(z) &= \frac{1}{d}\operatorname{tr}\big(\tilde{D}_{d,n}^{-1}(z)R_{d,n}\big) \\ &= \frac{1}{d}\operatorname{tr}\bigg(\tilde{D}_{d,n}^{-1}(z)R_{d,n}\big(\tilde{D}_{d,n}^{*}(z)\big)^{-1}\bigg[\frac{1}{1+(d/n)e(z)^*}R_{d,n} - S_{d,n} - z^* I_{d\times d}\bigg]\bigg). \end{aligned}$$

Note that $(A^*)^{-1} = (A^{-1})^*$. Now, the expression

$$\operatorname{tr}\big(\tilde{D}_{d,n}^{-1}(z)R_{d,n}\big(\tilde{D}_{d,n}^{*}(z)\big)^{-1}S_{d,n}\big) \ge 0 \tag{5.39}$$

is in particular real because the trace of the product of two positive semidefinite Hermitian matrices is non-negative. Hence,

$$\begin{aligned} \Im\big(e(z)\big) &= \frac{1}{d}\Im\bigg(\operatorname{tr}\bigg\{\tilde{D}_{d,n}^{-1}(z)R_{d,n}\big(\tilde{D}_{d,n}^{*}(z)\big)^{-1}\bigg(\frac{1}{1+(d/n)e(z)^*}R_{d,n} - z^* I_{d\times d}\bigg)\bigg\}\bigg) \\ &= \frac{1}{d}\operatorname{tr}\bigg\{\tilde{D}_{d,n}^{-1}(z)R_{d,n}\big(\tilde{D}_{d,n}^{*}(z)\big)^{-1}\bigg(\Im\bigg(\frac{1}{1+(d/n)e(z)^*}\bigg)R_{d,n} - \big(\Im z^*\big)I_{d\times d}\bigg)\bigg\} \\ &= \frac{1}{d}\operatorname{tr}\bigg\{\tilde{D}_{d,n}^{-1}(z)R_{d,n}\big(\tilde{D}_{d,n}^{*}(z)\big)^{-1}\bigg(\frac{(d/n)\Im(e(z))}{|1+(d/n)e(z)|^2}R_{d,n} + (\Im z)I_{d\times d}\bigg)\bigg\} \\ &= \alpha\big(e(z)\big)\Im\big(e(z)\big) + \beta\big(e(z)\big)\Im z \end{aligned}$$

with

$$\begin{aligned} \alpha\big(e(z)\big) &= \frac{1}{n}\left|1+\frac{d}{n}e(z)\right|^{-2}\operatorname{tr}\big\{\tilde{D}_{d,n}^{-1}(z)R_{d,n}\big(\tilde{D}_{d,n}^{*}(z)\big)^{-1}R_{d,n}\big\}, \\ \beta\big(e(z)\big) &= \frac{1}{d}\operatorname{tr}\big\{\tilde{D}_{d,n}^{-1}(z)R_{d,n}\big(\tilde{D}_{d,n}^{*}(z)\big)^{-1}\big\}. \end{aligned}$$

Note that both, α and β, are non-negative, and $\alpha(e(z)) > 0$ implies $\beta(e(z)) > 0$ since the trace of a positive semidefinite Hermitian matrix equals zero only for the null matrix. If $\bar{e}(z)$ is another solution of (5.34), we obtain the analogous identity

$$\Im\big(\bar{e}(z)\big) = \alpha\big(\bar{e}(z)\big)\Im\big(\bar{e}(z)\big) + \beta\big(\bar{e}(z)\big)\Im(z).$$

We denote by $\bar{D}_{d,n}(z)$ the matrix $\tilde{D}_{d,n}(z)$ as defined in (5.38) with $\bar{e}(z)$ in place of $e(z)$, and define $\alpha(\bar{e}(z))$ and $\beta(\bar{e}(z))$ correspondingly. Then

$$\begin{aligned}
e(z) - \bar{e}(z) &= \frac{1}{d}\operatorname{tr}\bigl\{\bigl(\tilde{D}_{d,n}^{-1}(z) - \bar{D}_{d,n}^{-1}(z)\bigr)R_{d,n}\bigr\} \\
&= \frac{1}{d}\operatorname{tr}\bigl\{\tilde{D}_{d,n}^{-1}(z)\bigl(\bar{D}_{d,n}(z) - \tilde{D}_{d,n}(z)\bigr)\bar{D}_{d,n}^{-1}(z)R_{d,n}\bigr\} \\
&= \frac{1}{d}\operatorname{tr}\left\{\tilde{D}_{d,n}^{-1}(z)\frac{(1+(d/n)e(z)) - (1+(d/n)\bar{e}(z))}{(1+(d/n)\bar{e}(z))(1+(d/n)e(z))}R_{d,n}\bar{D}_{d,n}^{-1}(z)R_{d,n}\right\} \qquad (5.40) \\
&= \bigl(e(z) - \bar{e}(z)\bigr)\frac{d/n}{(1+(d/n)\bar{e}(z))(1+(d/n)e(z))}\frac{1}{d}\operatorname{tr}\bigl\{\tilde{D}_{d,n}^{-1}(z)R_{d,n}\bar{D}_{d,n}^{-1}(z)R_{d,n}\bigr\} \\
&=: \bigl(e(z) - \bar{e}(z)\bigr)\gamma.
\end{aligned}$$

If $\gamma = 0$, uniqueness of $e(z)$ follows immediately. In case $\gamma \neq 0$, we deduce the inequality

$$\begin{aligned}
|\gamma| &\le \left[\frac{d/n}{|1+(d/n)e(z)|^2}\frac{1}{d}\operatorname{tr}\bigl\{\tilde{D}_{d,n}^{-1}(z)R_{d,n}\bigl(\tilde{D}_{d,n}^{*}(z)\bigr)^{-1}R_{d,n}\bigr\}\right]^{1/2} \\
&\quad\times\left[\frac{d/n}{|1+(d/n)\bar{e}(z)|^2}\frac{1}{d}\operatorname{tr}\bigl\{\bar{D}_{d,n}^{-1}(z)R_{d,n}\bigl(\bar{D}_{d,n}^{*}(z)\bigr)^{-1}R_{d,n}\bigr\}\right]^{1/2} \\
&= \sqrt{\alpha\bigl(e(z)\bigr)}\cdot\sqrt{\alpha\bigl(\bar{e}(z)\bigr)} \qquad (5.41) \\
&= \left(\frac{\Im(e(z))\alpha(e(z))}{\Im(e(z))\alpha(e(z)) + (\Im z)\beta(e(z))}\right)^{1/2} \\
&\quad\times\left(\frac{\Im(\bar{e}(z))\alpha(\bar{e}(z))}{\Im(\bar{e}(z))\alpha(\bar{e}(z)) + (\Im z)\beta(\bar{e}(z))}\right)^{1/2}.
\end{aligned}$$

But $\beta(e(z)), \beta(\bar{e}(z)) > 0$ for $\alpha(e(z)), \alpha(\bar{e}(z)) > 0$ which implies $|\gamma| < 1$ and therefore, $e = \bar{e}$.

5.4. Step IV: Identification of $e^{\circ}_{d,n}$ and $m^{\circ}_{d,n}$ as Stieltjes transforms

As concerns $e^{\circ}_{d,n}$, we know already that $e^{\circ}_{d,n} : \mathbb{C}^+ \to \mathbb{C}^+$. Its analyticity follows by the analyticity of the pointwise approximating sequence $e_{(d,n)_l}$ and the local boundedness of $(e_{(d,n)_l})$ on $\mathbb{C}^+$. Note that the pointwise convergence occurs simultaneously on a countable set with a accumulation point in $\mathbb{C}^+$ with probability 1. Using on the right-hand side of (5.34), the fact that $e^{\circ}_{d,n}(z) \to 0$ as $\Im z \to \infty$ which follows from (5.37), we also have

$$z \cdot e^{\circ}_{d,n}(z) \to -\frac{1}{d}\operatorname{tr}(R_{d,n}) \qquad \text{as } \Im z, \Re z \to \infty.$$

Hence, Lemma 2.2 in [21] implies that $e^{\circ}_{d,n}$ is the Stieltjes transform of a measure on the real line with total mass $d^{-1}\operatorname{tr}(R_{d,n})$.

Define

$$D^{\circ}_{d,n}(z)=\frac{1}{1+(d/n)e^{\circ}_{d,n}(z)}R_{d,n}-S_{d,n}-zI_{d\times d}. \tag{5.42}$$

Finally, observe that for any $z\in\mathbb{C}^+$,

$$\begin{aligned}\Im\big(m^{\circ}_{d,n}(z)\big)&=\frac{1}{d}\Im\operatorname{tr}\big\{D^{\circ}_{d,n}(z)^{-1}\big(\big(D^{\circ}_{d,n}(z)\big)^*\big)^{-1}\big(D^{\circ}_{d,n}(z)\big)^*\big\}\\&=\frac{1}{d}\Im\operatorname{tr}\bigg\{D^{\circ}_{d,n}(z)^{-1}\big(\big(D^{\circ}_{d,n}(z)\big)^*\big)^{-1}\\&\quad\times\bigg(\frac{1}{1+(d/n)(e^{\circ}_{d,n}(z))^*}R_{d,n}-z^*I_{d\times d}\bigg)\bigg\}\\&=\frac{1}{n}\frac{\Im(e^{\circ}_{d,n}(z))}{|1+(d/n)e^{\circ}_{d,n}(z)|^2}\operatorname{tr}\big(D^{\circ}_{d,n}(z)^{-1}\big(\big(D^{\circ}_{d,n}(z)\big)^*\big)^{-1}R_{d,n}\big)\\&\quad+\frac{1}{d}(\Im z)\operatorname{tr}\big(D^{\circ}_{d,n}(z)^{-1}\big(\big(D^{\circ}_{d,n}(z)\big)^*\big)^{-1}\big)\\&>0\end{aligned} \tag{5.43}$$

since both $\Im z$ and $\Im(e^{\circ}_{d,n}(z))$ are strictly positive. Furthermore, since $e^{\circ}_{d,n}(z)\to 0$ as $\Im z\to\infty$ by (5.37), we conclude

$$z\cdot m^{\circ}_{d,n}(z)\to -1 \qquad \text{as } \Im z,\Re z\to\infty.$$

As above, $m^{\circ}_{d,n}$ is the Stieltjes transform of a measure on the real line with total mass 1.

5.5. Step V: Approximation of $e_{d,n}$ by $e^{\circ}_{d,n}$

Let $e^{\circ}_{d,n}$ denote the solution of (5.34). We will show that for any $z\in\mathbb{C}^+$,

$$e_{d,n}(z)-e^{\circ}_{d,n}(z)\to 0 \qquad \text{a.s. as } d\to\infty. \tag{5.44}$$

Define

$$\alpha^{\circ}(z)=\alpha\big(e^{\circ}_{d,n}(z)\big) \quad\text{and}\quad \beta^{\circ}(z)=\beta\big(e^{\circ}_{d,n}(z)\big)$$

such that

$$\Im\big(e^{\circ}_{d,n}(z)\big)=\alpha^{\circ}(z)\Im\big(e^{\circ}_{d,n}\big)+\beta^{\circ}(z)\Im z. \tag{5.45}$$

Noting that

$$\frac{\alpha^{\circ}(z)}{\beta^{\circ}(z)}\le\|R_{d,n}\|_{S_\infty}\frac{d}{n}\bigg|1+\frac{d}{n}e^{\circ}_{d,n}(z)\bigg|^{-2},$$

we deduce

$$\Im\big(e^{\circ}_{d,n}(z)\big)\frac{\alpha^{\circ}(z)}{\beta^{\circ}(z)} \le \Im\big(e^{\circ}_{d,n}(z)\big)\|R_{d,n}\|_{S_\infty}\frac{d}{n}\left|1+\frac{d}{n}e^{\circ}_{d,n}(z)\right|^{-2} \tag{5.46}$$

$$= -\|R_{d,n}\|_{S_\infty}\Im\left(\frac{1}{1+(d/n)e^{\circ}_{d,n}(z)}\right)$$

$$\le \|R_{d,n}\|_{S_\infty}\left|\frac{1}{1+(d/n)e^{\circ}_{d,n}(z)}\right|$$

$$\le \|R_{d,n}\|_{S_\infty}\limsup_{l\to\infty}\frac{2\max_i|\lambda_i(\hat{\Xi}_{(d,n)_l})|^2+4|z|^2}{(\Im z)^2}, \tag{5.47}$$

where the last inequality follows by convergence (5.35) and bound (5.19) (in the latter the eigenvalues corresponding to $\hat{\Xi}_{(d,n)_l}$ have to be inserted). As a consequence,

$$\alpha^{\circ}(z) = \left(\frac{\Im(e^{\circ}_{d,n}(z))\alpha^{\circ}(z)}{(\Im z)\beta^{\circ}(z)+\Im(e^{\circ}_{d,n}(z))\alpha^{\circ}(z)}\right) \tag{5.48}$$

$$\le \frac{2\|R_{d,n}\|_{S_\infty}\|\hat{\Xi}_{d,n}\|^2_{S_\infty}+4|z|^2}{(\Im z)^3+2\|R_{d,n}\|_{S_\infty}\|\hat{\Xi}_{d,n}\|^2_{S_\infty}+4|z|^2}, \tag{5.49}$$

where the first identity (5.48) follows by rearrangement of (5.45), and after expanding the fraction by $(\beta^{\circ}(z))^{-1}$ we used the elementary inequality

$$\frac{x}{y+x} \le \frac{z}{y+z} \qquad \text{for } x, y, z > 0 \text{ and } x \le z$$

and (5.47) in (5.49). By (5.12),

$$e_{d,n}(z) = \frac{1}{d}\operatorname{tr}\big(R_{d,n}D^{-1}_{d,n}(z)\big) - \frac{1}{n}\sum_{k=1}^{n} f_{k,e}.$$

Then as previously in (5.39) and the subsequent display, we obtain the representation

$$\Im\big(e_{d,n}(z)\big) = \frac{1}{d}\Im\left(\frac{1}{1+(d/n)e^{*}_{d,n}(z)}\right)\operatorname{tr}\big\{D^{-1}_{d,n}(z)R_{d,n}\big(D^{*}_{d,n}(z)\big)^{-1}R_{d,n}\big\}$$

$$-\frac{1}{d}\big(\Im z^{*}\big)\operatorname{tr}\big\{D^{-1}_{d,n}(z)R_{d,n}\big(D^{*}_{d,n}(z)\big)^{-1}\big\} - \frac{1}{n}\sum_{k=1}^{n}\Im(f_{k,e}) \tag{5.50}$$

$$= \Im\big(e_{d,n}(z)\big)\alpha\big(e_{d,n}(z)\big) + (\Im z)\beta\big(e_{d,n}(z)\big) - \frac{1}{n}\sum_{k=1}^{n}\Im(f_{k,e}),$$

and as in (5.40),

$$e_{d,n}(z) - e^{\circ}_{d,n}(z) = \gamma\big(e_{d,n}(z) - e^{\circ}_{d,n}(z)\big) - \frac{1}{n}\sum_{k=1}^{n} f_{k,e} \tag{5.51}$$

with

$$|\gamma| \le \sqrt{\alpha^{\circ}(z)\alpha\big(e_{d,n}(z)\big)}. \tag{5.52}$$

Consider a realization for which the convergence

$$\frac{1}{n}\sum_{k=1}^{n} f_{k,e} \to 0$$

occurs. Then in particular,

$$\left|\frac{1}{n}\sum_{k=1}^{n} f_{k,e}\right| \le \Im z \frac{n}{4d(\|R_{d,n}\|_{S_\infty} \vee 1)}\left(\frac{2\|\hat{\Xi}_{d,n}\|^2_{S_\infty} + 4|z|^2}{(\Im z)^2}\right)^{-2} \tag{5.53}$$

for sufficiently large d. Recall that by definition of $\alpha(e_{d,n}(z))$ and $\beta(e_{d,n}(z))$,

$$\frac{\alpha(e_{d,n}(z))}{\beta(e_{d,n}(z))} \le \|R_{d,n}\|_{S_\infty}\frac{d}{n}\left|1 + \frac{d}{n}e_{d,n}(z)\right|^{-2}. \tag{5.54}$$

Hence, if

$$\beta\big(e_{d,n}(z)\big) \le \frac{n}{4d(\|R_{d,n}\|_{S_\infty} \vee 1)}\left(\frac{2\|\hat{\Xi}_{d,n}\|^2_{S_\infty} + 4|z|^2}{(\Im z)^2}\right)^{-2},$$

then inserting (5.19) into (5.54) yields

$$\alpha\big(e_{d,n}(z)\big) \le \|R_{d,n}\|_{S_\infty}\frac{d}{n}\left(\frac{2\|\hat{\Xi}_{d,n}\|^2_{S_\infty} + 4|z|^2}{(\Im z)^2}\right)^{2}\beta\big(e_{d,n}(z)\big) \le \frac{1}{4},$$

in which case (5.52) implies $|\gamma| \le 1/2$ since $\alpha^{\circ}(z) \le 1$ by (5.48) and the non-negativity of $\alpha^{\circ}(z)$, $\beta^{\circ}(z)$ and $\Im(e^{\circ}_{d,n}(z))$. Otherwise, if

$$\beta\big(e^{\circ}_{d,n}(z)\big) > \frac{n}{4d(\|R_{d,n}\|_{S_\infty} \vee 1)}\left(\frac{2\|\hat{\Xi}_{d,n}\|^2_{S_\infty} + 4|z|^2}{(\Im z)^2}\right)^{-2},$$

(5.52), (5.50), (5.53), and (5.49) imply

$$\begin{aligned}|\gamma| &\le \sqrt{\alpha^{\circ}(z)}\left(\frac{\Im(e_{d,n}(z))\alpha(e_{d,n}(z))}{\Im(e_{d,n}(z))\alpha(e_{d,n}(z)) + (\Im z)\beta(e_{d,n}(z)) - (1/n)\sum_{k=1}^{n}\Im(f_{k,e})}\right)^{1/2}\\ &\le \left(\frac{2\|R_{d,n}\|_{S_\infty}\|\hat{\Xi}_{d,n}\|^2_{S_\infty} + 4|z|^2}{(\Im z)^3 + 2\|R_{d,n}\|_{S_\infty}\|\hat{\Xi}_{d,n}\|^2_{S_\infty} + 4|z|^2}\right)^{1/2}.\end{aligned}$$

As $d \to \infty$ the limes superior of the last expression is bounded by some positive constant $\tilde{\gamma}(z) < 1$ almost surely. Finally, solving the equation (5.51) for $e_{d,n} - e^{\circ}_{d,n}$ and using the upper bounds on $|\gamma|$, we obtain

$$\begin{aligned} \left|e_{d,n}(z) - e^{\circ}_{d,n}(z)\right| &\le \frac{|(1/n)\sum_{k=1}^{n} f_{k,e}|}{1 - ((1/4) \vee \tilde{\gamma}(z))} \\ &\to 0 \qquad \text{a.s.} \end{aligned} \tag{5.55}$$

as $d \to \infty$, by (5.13). This proves (5.44).

5.6. Step VI: Approximation of $m_{d,n}$ by $m^{\circ}_{d,n}$

Without loss of generality, we may assume that either

$$\frac{d^{3/2}}{n} > 1 \quad \text{or} \quad \frac{d^{3/2}}{n} \le 1$$

holds on the whole sequence. We start with the first case. Recall the definition (3.1) of $m^{\circ}_{d,n}$ and (5.42) of $D^{\circ}_{d,n}(z)$, and note that

$$m^{\circ}_{d,n}(z) = \frac{1}{d}\operatorname{tr}\bigl(\bigl(D^{\circ}_{d,n}(z)\bigr)^{-1}\bigr),$$

while by (5.11),

$$m_{d,n}(z) = \frac{1}{d}\operatorname{tr}\bigl(\bigl(D_{d,n}(z)\bigr)^{-1}\bigr) - \frac{1}{n}\sum_{k=1}^{n} f_{k,m}$$

with

$$\frac{1}{n}\sum_{k=1}^{n} f_{k,m} \to 0 \qquad \text{a.s. as } d \to \infty.$$

Then,

$$\begin{aligned} m_{d,n}(z) - m^{\circ}_{d,n}(z) &= \frac{1}{d}\operatorname{tr}\bigl\{D^{-1}_{d,n}(z) - \bigl(D^{\circ}_{d,n}(z)\bigr)^{-1}\bigr\} - \frac{1}{n}\sum_{k=1}^{n} f_{k,m} \\ &= \frac{1}{d}\operatorname{tr}\bigl\{D^{-1}_{d,n}(z)\bigl(D^{\circ}_{d,n}(z) - D_{d,n}(z)\bigr)\bigl(D^{\circ}_{d,n}(z)\bigr)^{-1}\bigr\} - \frac{1}{n}\sum_{k=1}^{n} f_{k,m} \\ &= \frac{1}{n}\frac{e_{d,n}(z) - e^{\circ}_{d,n}(z)}{(1 + (d/n)e_{d,n}(z))(1 + (d/n)e^{\circ}_{d,n}(z))}\operatorname{tr}\bigl\{D^{-1}_{d,n}(z)R_{d,n}\bigl(D^{\circ}_{d,n}(z)\bigr)^{-1}\bigr\} \\ &\quad - \frac{1}{n}\sum_{k=1}^{n} f_{k,m}. \end{aligned}$$

So, almost surely by (5.24), (5.19) and (5.55),

$$\begin{aligned}&\limsup_{d\to\infty}\left|m_{d,n}(z)-m^{\circ}_{d,n}(z)\right|\\&\quad\leq\limsup_{d\to\infty}\left|e_{d,n}(z)-e^{\circ}_{d,n}(z)\right|\limsup_{d\to\infty}\frac{d}{n}\|R_{d,n}\|_{S_\infty}\frac{(2\|\hat{\Xi}_{d,n}\|_{S_\infty}+4|z|^4)^2}{(\Im z)^6}\\&\quad=0.\end{aligned}$$

Now, consider the case

$$\frac{d^{3/2}}{n}\leq 1.$$

Due to

$$\frac{d}{n}\left|e^{\circ}_{d,n}(z)\right|\leq\frac{d}{n}\frac{\sup_d\|R_{d,n}\|_{S_\infty}}{\Im z}\longrightarrow 0$$

for any $z\in\mathbb{C}^+$ and by reasons of continuity, we conclude

$$\left|m^{\circ}_{d,n}(z)-m_{\mu^{T_{d,n}}}(z)\right|\to 0$$

for any $z\in\mathbb{C}^+$, where $\mu^{T_{d,n}}$ is the spectral measure of the matrix $T_{d,n}$. Therefore, it remains to show that

$$\left|m_{d,n}(z)-m_{\mu^{T_{d,n}}}(z)\right|\to 0\qquad\text{a.s.}$$

By Lemma C.12 and Lemma C.13, this convergence holds true if $d_L(\mu_{d,n},\mu^{T_{d,n}})\to 0$ almost surely. Theorem C.10 for $\alpha=1$ and inequality (1.2) of [13] yield

$$d_L^2\big(\mu_{d,n},\mu^{T_{d,n}}\big)\leq\frac{1}{d}\sum_{i=1}^{d}\left|\lambda_i(\Xi_{d,n})-\lambda_i(T_{d,n})\right|\leq\left\|\frac{1}{n}R^{1/2}_{d,n}X_{d,n}X^*_{d,n}R^{1/2}_{d,n}-R_{d,n}\right\|_{S_\infty}.$$

Finally, for arbitrary $\varepsilon>0$ and d sufficiently large we apply Corollary 5.50 of [25] with $t=1$ so that

$$\left\|\frac{1}{n}R^{1/2}_{d,n}X_{d,n}X^*_{d,n}R^{1/2}_{d,n}-R_{d,n}\right\|_{S_\infty}\leq\varepsilon$$

with probability at least $1-2\exp(-d)$. Again, by the Borel–Cantelli lemma,

$$d_L\big(\mu_{d,n},\mu^{T_{d,n}}\big)\leq\left\|\frac{1}{n}R^{1/2}_{d,n}X_{d,n}X^*_{d,n}R^{1/2}_{d,n}-R_{d,n}\right\|^{1/2}_{S_\infty}\to 0$$

almost surely as $d\to\infty$.

5.7. Step VII: Weak approximation of the spectral measures

First, we show that the measure $\mu^{\circ}_{d,n}$ has compact support. Thereto, define similarly to the definition of $e_{(d,n)_l}$, $l \in \mathbb{N}$, in Step III,

$$m_{(d,n)_l}(z) = \frac{1}{d_l} \operatorname{tr}\{(\hat{\Xi}_{(d,n)_l} - zI_{d_l \times d_l})^{-1}\}.$$

By (5.4),

$$m_{(d,n)_l}(z) - \frac{1}{d_l} \operatorname{tr}\left\{\left(\frac{1}{1+(d/n)e_{(d,n)_l}(z)} R_{(d,n)_l} - S_{(d,n)_l} - zI_{d_l \times d_l}\right)^{-1}\right\} \to 0 \qquad \text{as } l \to \infty$$

almost surely. Note that

$$\begin{aligned} &\frac{1}{d_l} \operatorname{tr}\left\{\left(\frac{1}{1+(d/n)e_{(d,n)_l}(z)} R_{(d,n)_l} - S_{(d,n)_l} - zI_{d_l \times d_l}\right)^{-1}\right\} \\ &\quad = \frac{1}{d} \operatorname{tr}\left\{\left(\frac{1}{1+(d/n)e_{(d,n)_l}(z)} R_{d,n} - S_{d,n} - zI_{d \times d}\right)^{-1}\right\}, \end{aligned}$$

and therefore by reasons of continuity

$$m_{(d,n)_l}(z) - \frac{1}{d} \operatorname{tr}\left\{\left(\frac{1}{1+(d/n)e^{\circ}_{d,n}(z)} R_{d,n} - S_{d,n} - zI_{d \times d}\right)^{-1}\right\} \to 0 \qquad \text{as } l \to \infty \text{ a.s.}$$

because of (5.44). By Vitali's convergence theorem, the exceptional set on which convergence does not occur can be chosen uniformly in $z \in \mathbb{C}^+$. This implies that $\mu^{\circ}_{d,n}$ is the weak limit of $\mu_{(d,n)_l}$ almost surely. In particular, the support of $\mu^{\circ}_{d,n}$ is bounded since

$$\left|\inf\{x : \mu^{\circ}_{d,n}((-\infty, x]) > 0\}\right| \geq \liminf_{l \to \infty} \lambda_{dl}(\hat{\Xi}_{(d,n)_l}) \geq -\|S_{d,n}\|_{S_\infty}$$

and

$$\left|\sup\{x : \mu^{\circ}_{d,n}((-\infty, x]) < 1\}\right| \leq \limsup_{l \to \infty} \|\hat{\Xi}_{(d,n)_l}\|_{S_\infty} \leq \|S_{d,n}\|_{S_\infty} + c',$$

where $c' > 0$ is a constant satisfying inequality (C.6) of Lemma C.7 applied to

$$\frac{1}{nl} \sum_{k=1}^{nl} R^{1/2}_{(d,n)_l} \check{Z}_{k,d_l,n_l} \check{Z}^*_{k,d_l,n_l} R^{1/2}_{(d,n)_l},$$

and is chosen uniformly over $d \in \mathbb{N}$. Subsequently, we assume that d (in dependence on the specific realization) is sufficiently large such that

$$\left|\inf\{x : \mu_{d,n}((-\infty, x]) > 0\}\right| \geq -\|S_{d,n}\|_{S_\infty} - c''$$

and

$$\left|\sup\left\{x : \mu_{d,n}\left((-\infty, x]\right) < 1\right\}\right| \le \|S_{d,n}\|_{S_\infty} + c''$$

with an appropriate contant $c'' > 0$ from (C.6). Now, define $c = c' \vee c''$. For fixed $0 < v < 1$, define the closed interval $K = [u_0, u_{\lfloor v^{-3}\rfloor+1}]$ with

$$u_l = -v^{-1/4}\left(\|S_{d,n}\|_{S_\infty} + c\right) + \frac{2l}{\lfloor v^{-3}\rfloor + 1} v^{-1/4}\left(\|S_{d,n}\|_{S_\infty} + c\right)$$

for $l = 1, \dots, \lfloor v^{-3}\rfloor + 1$. By Step VI, we have

$$\left|m_{d,n}(u_l + iv) - m^\circ_{d,n}(u_l + iv)\right| < v$$

simultaneously at all points u_l, $l = 0, \dots, \lfloor v^{-3}\rfloor + 1$, almost surely for all d sufficiently large. Furthermore, for any inner point u of K, pick l such that $u \in [u_l, u_{l+1})$. Then,

$$\begin{aligned}
&\left|m_{d,n}(u + iv) - m^\circ_{d,n}(u + iv)\right| \\
&\quad \le \left|m_{d,n}(u + iv) - m_{d,n}(u_l + iv)\right| + \left|m^\circ_{d,n}(u + iv) - m^\circ_{d,n}(u_l + iv)\right| \\
&\qquad + \left|m_{d,n}(u_l + iv) - m^\circ_{d,n}(u_l + iv)\right| \\
&\quad \le \int \left|\frac{1}{x - u - iv} - \frac{1}{x - u_l - iv}\right| \mathrm{d}\mu_{d,n}(x) \\
&\qquad + \int \left|\frac{1}{x - u - iv} - \frac{1}{x - u_l - iv}\right| \mathrm{d}\mu^\circ_{d,n}(x) + v \\
&\quad \le \int \frac{u - u_l}{v^2} \,\mathrm{d}\left(\mu_{d,n} + \mu^\circ_{d,n}\right)(x) + v \\
&\quad \le v\left(4|u_0| + 1\right).
\end{aligned}$$

Next, we derive an upper bound on the integral

$$\int_{K^c} \left|m_{d,n}(u + iv) - m^\circ_{d,n}(u + iv)\right| \mathrm{d}u$$

which tends to zero for $v \to 0$. For this aim, we decompose the integral into

$$\begin{aligned}
&\int_{K^c} \left|m_{d,n}(u + iv) - m^\circ_{d,n}(u + iv)\right| \mathrm{d}u \\
&\quad = \int_{(-\infty, u_0)} \left|m_{d,n}(u + iv) - m^\circ_{d,n}(u + iv)\right| \mathrm{d}u \\
&\qquad + \int_{(u_{\lfloor v^{-3}\rfloor+1}, \infty)} \left|m_{d,n}(u + iv) - m^\circ_{d,n}(u + iv)\right| \mathrm{d}u.
\end{aligned}$$

We can use the same arguments for both integrals and therefore only consider the first one. By Fubini's theorem and the bounds on the support of $\mu_{d,n}$ and $\mu^{\circ}_{d,n}$,

$$\begin{aligned}
&\int_{(-\infty,u_0)} \left|m_{d,n}(u+iv) - m^{\circ}_{d,n}(u+iv)\right| \mathrm{d}u \\
&\quad \le \int\int\int_{(-\infty,u_0)} \left|\frac{1}{x-u-iv} - \frac{1}{y-u-iv}\right| \mathrm{d}u\, \mathrm{d}\mu_{d,n}(x)\, \mathrm{d}\mu^{\circ}_{d,n}(y) \\
&\quad \le \int\int\int_{(-\infty,u_0)} \frac{|x-y|}{(u-v^{1/4}u_0)^2} \mathrm{d}u\, \mathrm{d}\mu_{d,n}(x)\, \mathrm{d}\mu^{\circ}_{d,n}(y) \\
&\quad \le \frac{1}{(1-v^{1/4})|u_0|} \int\int |x-y|\, \mathrm{d}\mu_{d,n}(x)\, \mathrm{d}\mu^{\circ}_{d,n}(y) \\
&\quad \le \frac{1}{(1-v^{1/4})|u_0|} \left(\int |x|\, \mathrm{d}\mu_{d,n}(x) + \int |y|\, \mathrm{d}\mu^{\circ}_{d,n}(y)\right) \\
&\quad \le 2\frac{v^{1/4}}{1-v^{1/4}}.
\end{aligned}$$

Now, by Lemma C.11 we conclude

$$\begin{aligned}
d_L\big(\mu_{d,n}, \mu^{\circ}_{d,n}\big) &\le 2\sqrt{\frac{v}{\pi}} + \frac{1}{2\pi}\int \left|m_{d,n}(u+iv) - m^{\circ}_{d,n}(u+iv)\right| \mathrm{d}u \\
&\le 2\sqrt{\frac{v}{\pi}} + \frac{1}{\pi}|u_0|\big(4|u_0|+1\big)v + \frac{2}{\pi}\frac{v^{1/4}}{1-v^{1/4}},
\end{aligned}$$

where the inequalities hold almost surely for all d sufficiently large. Hence,

$$d_L\big(\mu_{d,n}, \mu^{\circ}_{d,n}\big) \longrightarrow 0$$

almost surely as $d \to \infty$. Lemma C.13 yields finally $\mu_{d,n} - \mu^{\circ}_{d,n} \Longrightarrow 0$ a.s. □

5.8. Proof of Corollary 3.2

As afore-mentioned to the corollary,

$$\mu^{\circ}_{d,n} = \mu^{\mathrm{MP}}_{d/n, p_0/\sigma^2} \star \delta_{-\sigma^2(1-p_0)/p_0}.$$

Therefore, by the representation (2.1) of the Marčenko–Pastur distribution we deduce

$$\mu^{\circ}_{d,n} \Longrightarrow \mu^{\mathrm{MP}}_{y,\sigma^2/p_0} \star \delta_{-\sigma^2(1-p_0)/p_0},$$

such that

$$\mu_{d,n} \Longrightarrow \mu^{\mathrm{MP}}_{y,\sigma^2/p_0} \star \delta_{-\sigma^2(1-p_0)/p_0}.$$

Furthermore, if the left edge of the limiting distribution

$$\mu^{\mathrm{MP}}_{y,\sigma^2/p_0} \star \delta_{-\sigma^2(1-p_0)/p_0}$$

is smaller than zero, then almost surely

$$\limsup_{d\to\infty} \lambda_{\min}(\hat{\Xi}_{d,n}) < 0.$$

For $y < 1$ the left edge of the limiting distribution is smaller than zero if and only if $p_0 < 1-(1-\sqrt{y})^2$. □

6. Proof of Theorem 3.3

We will show Theorem 3.3 by means of the next proposition. The proof of the proposition is postponed to Appendix B.

Proposition 6.1. *Let $(X(i,k))_{i,k\in\mathbb{N}}$ be a double array of i.i.d. centered random variables with unit variance and finite fourth moment, and denote by $X_{d,n} \in \mathbb{R}^{d\times n}$ its $d\times n$ submatrix in the upper left corner. Moreover, let $(A_{d,n})_{d,n}$, $A_{d,n} \in \mathbb{R}^{d\times d}$, be a sequence of symmetric random matrices and $(B_{d,n})_{d,n}$, $B_{d,n} \in \mathbb{R}^{d\times n}$ be another sequence of random matrices such that $(A_{d,n}, B_{d,n})$ and $X_{d,n}$ are independent. Let $d,n\to\infty$ and $d/n \to y > 0$. If*

$$\limsup_{d\to\infty} \max_{i,j} |A_{ij,d,n}| \max_{i,k} B^2_{ik,d,n} < \alpha \qquad a.s. \tag{6.1}$$

for some absolute constant $\alpha > 0$, then

$$\limsup_{d\to\infty} \left\| \frac{1}{n} A_{d,n} \circ \big((X_{d,n}\circ B_{d,n})(X_{d,n}\circ B_{d,n})^*\big)\right\|_{S_\infty} \le \alpha(1+\sqrt{y})^2 \qquad a.s. \tag{6.2}$$

Proof of Theorem 3.3. By Weyl's inequality, we obtain

$$\begin{aligned}
&\lambda_{\max}\left(\frac{1}{n} W_{d,n} \circ \big((X_{d,n}\circ \varepsilon_{d,n})(X_{d,n}\circ \varepsilon_{d,n})^*\big)\right) \\
&\qquad + \lambda_{\min}\left(\frac{1}{n} (\hat{W}_{d,n} - W_{d,n}) \circ \big((X_{d,n}\circ \varepsilon_{d,n})(X_{d,n}\circ \varepsilon_{d,n})^*\big)\right) \\
&\le \lambda_{\max}\left(\frac{1}{n} \hat{W}_{d,n} \circ \big((X_{d,n}\circ \varepsilon_{d,n})(X_{d,n}\circ \varepsilon_{d,n})^*\big)\right) \\
&\le \lambda_{\max}\left(\frac{1}{n} W_{d,n} \circ \big((X_{d,n}\circ \varepsilon_{d,n})(X_{d,n}\circ \varepsilon_{d,n})^*\big)\right) \\
&\qquad + \lambda_{\max}\left(\frac{1}{n} (\hat{W}_{d,n} - W_{d,n}) \circ \big((X_{d,n}\circ \varepsilon_{d,n})(X_{d,n}\circ \varepsilon_{d,n})^*\big)\right)
\end{aligned}$$

and

$$\begin{aligned}
&\lambda_{\min}\left(\frac{1}{n} W_{d,n} \circ \big((X_{d,n} \circ \varepsilon_{d,n})(X_{d,n} \circ \varepsilon_{d,n})^*\big)\right) \\
&\qquad + \lambda_{\min}\left(\frac{1}{n}(\hat{W}_{d,n} - W_{d,n}) \circ \big((X_{d,n} \circ \varepsilon_{d,n})(X_{d,n} \circ \varepsilon_{d,n})^*\big)\right) \\
&\quad \le \lambda_{\min}\left(\frac{1}{n} \hat{W}_{d,n} \circ \big((X_{d,n} \circ \varepsilon_{d,n})(X_{d,n} \circ \varepsilon_{d,n})^*\big)\right) \\
&\quad \le \lambda_{\min}\left(\frac{1}{n} W_{d,n} \circ \big((X_{d,n} \circ \varepsilon_{d,n})(X_{d,n} \circ \varepsilon_{d,n})^*\big)\right) \\
&\qquad + \lambda_{\max}\left(\frac{1}{n}(\hat{W}_{d,n} - W_{d,n}) \circ \big((X_{d,n} \circ \varepsilon_{d,n})(X_{d,n} \circ \varepsilon_{d,n})^*\big)\right).
\end{aligned}$$

Because of

$$\begin{aligned}
&\lambda_{\max}\left(\frac{1}{n} W_{d,n} \circ \big((X_{d,n} \circ \varepsilon_{d,n})(X_{d,n} \circ \varepsilon_{d,n})^*\big)\right) \\
&\quad = \lambda_{\max}\left(\frac{1}{n}\big(w_{d,n} w_{d,n}^*\big) \circ \big((X_{d,n} \circ \varepsilon_{d,n})(X_{d,n} \circ \varepsilon_{d,n})\big) - \frac{1-p_0}{p_0}\sigma^2 I_{d\times d}\right. \\
&\qquad \left. + \operatorname{diag}\left[\frac{1}{n}\big(W_{d,n} - w_{d,n} w_{d,n}^*\big) \circ \big((X_{d,n} \circ \varepsilon_{d,n})(X_{d,n} \circ \varepsilon_{d,n})^*\big)\right] + \frac{1-p_0}{p_0}\sigma^2 I_{d\times d}\right),
\end{aligned}$$

and

$$\begin{aligned}
&\left\| \operatorname{diag}\left[\frac{1}{n}\big(W_{d,n} - w_{d,n} w_{d,n}^*\big) \circ \big((X_{d,n} \circ \varepsilon_{d,n})(X_{d,n} \circ \varepsilon_{d,n})^*\big)\right] + \frac{1-p_0}{p_0}\sigma^2 I_{d\times d} \right\|_{S_\infty} \\
&\quad \longrightarrow 0 \qquad \text{a.s. as } d \to \infty
\end{aligned}$$

by the Marcinkiewicz–Zygmund strong law of large numbers (cf. Lemma B.25 in [1]), we obtain again by Weyl's inequality and Theorem 1 of [4]

$$\lambda_{\max}\left(\frac{1}{n} W_{d,n} \circ \big((X_{d,n} \circ \varepsilon_{d,n})(X_{d,n} \circ \varepsilon_{d,n})^*\big)\right) \xrightarrow{\text{a.s.}} \frac{\sigma^2}{p_0}(1+\sqrt{y})^2 - \frac{1-p_0}{p_0}\sigma^2.$$

With same argument,

$$\lambda_{\min}\left(\frac{1}{n} W_{d,n} \circ \big((X_{d,n} \circ \varepsilon_{d,n})(X_{d,n} \circ \varepsilon_{d,n})^*\big)\right) \xrightarrow{\text{a.s.}} \frac{\sigma^2}{p_0}(1-\sqrt{y})^2 - \frac{1-p_0}{p_0}\sigma^2.$$

In order to finish the proof, it suffices to show that

$$\left\| \frac{1}{n}(\hat{W}_{d,n} - W_{d,n}) \circ \big((X_{d,n} \circ \varepsilon_{d,n})(X_{d,n} \circ \varepsilon_{d,n})^*\big) \right\|_{S_\infty} \xrightarrow{\text{a.s.}} 0.$$

But this is an easy consequence of Proposition 6.1 since by (A.12),

$$\limsup_{n\to\infty} \max_{i,j} |\hat{W}_{ij,d,n} - W_{ij,d,n}| \xrightarrow{\text{a.s.}} 0.$$

□

Appendix A: Proof of Proposition 4.1

A.1. Step I: Modifying $\boldsymbol{\varepsilon_{d,n}}$

By tightness of $(\mu^{w_{d,n}})$, we have for any $\delta > 0$ a constant $p_0 > 0$ such that for sufficiently large $d \in \mathbb{N}$

$$\#\{p_{i,d,n} < p_0\} \le d\delta.$$

We replace the matrix $\varepsilon_{d,n}$ by $\tilde{\varepsilon}_{d,n}$, where $\tilde{\varepsilon}_{ik,d,n} = \varepsilon_{ik,d,n}$ if $p_i \ge p_0$ and otherwise $\tilde{\varepsilon}_{ik,d,n}$ is a Bernoulli random variable with $\mathbb{P}(\tilde{\varepsilon}_{ik} = 1) = p_0$ such that the entries of $\tilde{\varepsilon}_{d,n}$ are independent and jointly independent of $Y_{1,d,n}, \dots, Y_{n,d,n}$. $\tilde{T}_{d,n}$ be the matrix as $\hat{T}_{d,n}$ but relying on the missingness matrix $\tilde{\varepsilon}_{d,n}$ in place of $\varepsilon_{d,n}$. Since by Theorem C.8

$$d_K\left(\mu^{\tilde{T}_{d,n}}, \mu^{\hat{T}_{d,n}}\right) \le \frac{1}{d}\operatorname{rank}(\tilde{T}_{d,n} - \hat{T}_{d,n}) \le \delta,$$

we may assume subsequently $p_{i,d,n} \ge p_0$.

A.2. Step II: Removing $\boldsymbol{\frac{1}{n}\hat{W}_{d,n} \circ ((\hat{M}_{d,n} \circ \varepsilon_{d,n})(\hat{M}_{d,n} \circ \varepsilon_{d,n})^*)}$

Let

$$\tilde{T}_{d,n} = \hat{T}_{d,n} - \frac{1}{n}\hat{W}_{d,n} \circ \left((\hat{M}_{d,n} \circ \varepsilon_{d,n})(\hat{M}_{d,n} \circ \varepsilon_{d,n})^*\right).$$

First, note that

$$\mathbb{P}\Big(\min_{i,j} \#\mathcal{N}_{ij} = 0\Big) \le d^2 \max_{i,j} P(\#\mathcal{N}_{ij} = 0) \le d^2\left(1 - p_0^2\right)^n.$$

Hence, by the Borel–Cantelli lemma we have almost surely for all but finitely many indices d

$$\frac{1}{n}\hat{W}_{d,n} \circ \left((\hat{M}_{d,n} \circ \varepsilon_{d,n})(\hat{M}_{d,n} \circ \varepsilon_{d,n})^*\right) = \hat{m}_{d,n}\hat{m}_{d,n}^*.$$

Now, by Theorem C.8 we have

$$\limsup_{d\to\infty} d_K\left(\mu^{\hat{T}_{d,n}}, \mu^{\tilde{T}_{d,n}}\right) = 0 \qquad \text{a.s.}$$

Therefore, it is sufficient to prove $d_L(\mu^{\tilde{T}_{d,n}}, \mu^{\bar{T}_{d,n}}) \to 0$. In the next subsection, we refer to $\tilde{T}_{d,n}$ as $\hat{T}_{d,n}$.

A.3. Step III: Truncation of $T_{d,n}$

By the tightness of the sequence $(\mu^{T_{d,n}})$ we have for any $\delta > 0$ a constant $\tau_0 > 0$ such that for sufficiently large $d \in \mathbb{N}$

$$\#\{T_{kk,d,n} > \tau_0\} \leq d\delta.$$

Therefore, let $\check{T}_{d,n} = \operatorname{diag}(\mathbb{1}\{T_{11,d,n} \leq \tau_0\}T_{11,d,n}, \dots, \mathbb{1}\{T_{dd,d,n} \leq \tau_0\}T_{dd,d,n})$ and $\tilde{T}_{d,n}$ be the sample covariance matrix with missing observations built from the random variables

$$\tilde{Y}_{i,d,n} = \check{T}_{d,n}X_{i,d,n}, \qquad i = 1, \dots, n,$$

while $\varepsilon_{d,n}$ remains the same. Since again by Theorem C.8

$$d_K\big(\mu^{\check{T}_{d,n}}, \mu^{\hat{T}_{d,n}}\big) \leq \frac{1}{d}\operatorname{rank}(\check{T}_{d,n} - \hat{T}_{d,n}) \leq \delta,$$

it is sufficient to assume subsequently that the spectral measures of the sequence $(T_{d,n})$ have uniformly bounded support.

A.4. Step IV: Truncation of $X_{d,n}$

For $0 < \delta < \frac{1}{2}$ we truncate the variables $X_{ik,d,n}$ at the threshold level $n^{1/2}d^{\alpha-1/2}$, $\alpha > \frac{1+\delta}{2}$. Hence, let

$$\tilde{X}_{ik,d,n} = X_{ik,d,n}\mathbb{1}\big(|X_{ik,d,n}| \leq n^{1/2}d^{\alpha-1/2}\big) \tag{A.1}$$

and $\tilde{T}_{d,n}$, $\tilde{Y}_{d,n}$ and $\tilde{M}_{d,n}$ be the matrices constructed by replacing $X_{d,n}$ with $\tilde{X}_{d,n} = (\tilde{X}_{ik,d,n})$ in $\hat{T}_{d,n}$, $Y_{d,n}$, and $\hat{M}_{d,n}$. We have

$$\begin{aligned}
&d_K\big(\mu^{\tilde{T}_{d,n}}, \mu^{\hat{T}_{d,n}}\big)\\
&\quad\leq \frac{1}{d}\operatorname{rank}(\tilde{T}_{d,n} - \hat{T}_{d,n})\\
&\quad= \frac{1}{d}\operatorname{rank}\bigg[\frac{1}{n}\hat{W}_{d,n} \circ \big((Y_{d,n} \circ \varepsilon_{d,n})(Y_{d,n} \circ \varepsilon_{d,n})^* - (\tilde{Y}_{d,n} \circ \varepsilon_{d,n})(\tilde{Y}_{d,n} \circ \varepsilon_{d,n})^*\\
&\qquad - (\hat{M}_{d,n} \circ \varepsilon_{d,n})(Y_{d,n} \circ \varepsilon_{d,n})^* + (\tilde{M}_{d,n} \circ \varepsilon_{d,n})(\tilde{Y}_{d,n} \circ \varepsilon_{d,n})^*\\
&\qquad - (Y_{d,n} \circ \varepsilon_{d,n})(\hat{M}_{d,n} \circ \varepsilon_{d,n})^* + (\tilde{Y}_{d,n} \circ \varepsilon_{d,n})(\tilde{M}_{d,n} \circ \varepsilon_{d,n})^*\big)\bigg]\\
&\quad= \frac{1}{d}\operatorname{rank}\bigg[\frac{1}{n}\hat{W}_{d,n} \circ \Big(\big((Y_{d,n} - \tilde{Y}_{d,n}) \circ \varepsilon_{d,n}\big)(Y_{d,n} \circ \varepsilon_{d,n})^*\\
&\qquad + (\tilde{Y}_{d,n} \circ \varepsilon_{d,n})\big((Y_{d,n} - \tilde{Y}_{d,n}) \circ \varepsilon_{d,n}\big)^*\\
&\qquad - (\hat{M}_{d,n} \circ \varepsilon_{d,n})\big((Y_{d,n} - \tilde{Y}_{d,n}) \circ \varepsilon_{d,n}\big)^*
\end{aligned}$$

$$
\begin{aligned}
&- \big((\hat{M}_{d,n} - \tilde{M}_{d,n}) \circ \varepsilon_{d,n}\big)(\tilde{Y}_{d,n} \circ \varepsilon_{d,n})^* \\
&- \big((Y_{d,n} - \tilde{Y}_{d,n}) \circ \varepsilon_{d,n}\big)(\hat{M}_{d,n} \circ \varepsilon_{d,n})^* \\
&- (\tilde{Y}_{d,n} \circ \varepsilon_{d,n})\big((\hat{M}_{d,n} - \tilde{M}_{d,n}) \circ \varepsilon_{d,n}\big)^*\big)\Big] \\
\le{}& \frac{1}{d}\operatorname{rank}\Big[\frac{1}{n}\hat{W}_{d,n} \circ \Big(\big((Y_{d,n} - \tilde{Y}_{d,n}) \circ \varepsilon_{d,n}\big)\big((Y_{d,n} - \hat{M}_{d,n}) \circ \varepsilon_{d,n}\big)^* \\
&- \big((\hat{M}_{d,n} - \tilde{M}_{d,n}) \circ \varepsilon_{d,n}\big)(\tilde{Y}_{d,n} \circ \varepsilon_{d,n})^*\Big)\Big] \\
&+ \frac{1}{d}\operatorname{rank}\Big[\frac{1}{n}\hat{W}_{d,n} \circ \Big(\big((\tilde{Y}_{d,n} - \hat{M}_{d,n}) \circ \varepsilon_{d,n}\big)\big((Y_{d,n} - \tilde{Y}_{d,n}) \circ \varepsilon_{d,n}\big)^* \\
&- (\tilde{Y}_{d,n} \circ \varepsilon_{d,n})\big((\hat{M}_{d,n} - \tilde{M}_{d,n}) \circ \varepsilon_{d,n}\big)^*\Big)\Big] \\
\le{}& \frac{2}{d}\#\left\{i \in \{1,\dots,d\} : \sum_{k=1}^{n} \mathbb{1}\big(|X_{ik,d,n}| > n^{1/2}d^{\alpha-1/2}\big) > 0\right\} \qquad \text{(A.2)} \\
\le{}& \frac{2}{d}\sum_{i,k} \mathbb{1}\big(|X_{ik,d,n}| > n^{1/2}d^{\alpha-1/2}\big), \qquad \text{(A.3)}
\end{aligned}
$$

where inequality (A.2) follows by the simple observation that the ith row respectively the ith column of the matrices

$$\big((Y_{d,n} - \tilde{Y}_{d,n}) \circ \varepsilon_{d,n}\big)\big((Y_{d,n} - \hat{M}_{d,n}) \circ \varepsilon_{d,n}\big)^*$$

and

$$\big((\hat{M}_{d,n} - \tilde{M}_{d,n}) \circ \varepsilon_{d,n}\big)(\tilde{Y}_{d,n} \circ \varepsilon_{d,n})^*$$

respectively

$$\big((\tilde{Y}_{d,n} - \hat{M}_{d,n}) \circ \varepsilon_{d,n}\big)\big((Y_{d,n} - \tilde{Y}_{d,n}) \circ \varepsilon_{d,n}\big)^*$$

and

$$(\tilde{Y}_{d,n} \circ \varepsilon_{d,n})\big((\hat{M}_{d,n} - \tilde{M}_{d,n}) \circ \varepsilon_{d,n}\big)^*$$

is the null vector if

$$\sum_{k=1}^{n} \mathbb{1}\big(|X_{ik,d,n}| > n^{1/2}d^{\alpha-1/2}\big) = 0.$$

Next, we prove that

$$\frac{2}{d}\sum_{i,k} \mathbb{1}\big(|X_{ik,d,n}| > n^{1/2}d^{\alpha-1/2}\big) \xrightarrow{\text{a.s.}} 0$$

as $d \to \infty$. Note first that by Markov's inequality

$$\operatorname{Var}\big(\mathbb{1}\big\{|X_{11,d,n}| > n^{1/2}d^{\alpha-1/2}\big\}\big) \le \mathbb{E}\mathbb{1}\big\{|X_{ik,d,n}| > n^{1/2}d^{\alpha-1/2}\big\} \le n^{-1}d^{1-2\alpha}. \tag{A.4}$$

Using (A.4) in (A.5), and (A.4) in Bernstein's inequality in (A.7), we conclude for sufficiently large d and some constant $\beta > 0$

$$\begin{aligned}
&\mathbb{P}\Big(\sum_{k,i}\mathbb{1}\big\{|X_{ik,d,n}| > n^{1/2}d^{\alpha-1/2}\big\} \ge d^{1-\delta}\Big)\\
&\quad= \mathbb{P}\Big(\sum_{k,i}\big(\mathbb{1}\big\{|X_{ik,d,n}| > n^{1/2}d^{\alpha-1/2}\big\} - \mathbb{E}\mathbb{1}\big\{|X_{ik,d,n}| > n^{1/2}d^{\alpha-1/2}\big\}\big)\\
&\qquad\quad \ge d^{1-\delta} - nd\mathbb{E}\mathbb{1}\big\{|X_{11,d,n}| > n^{1/2}d^{\alpha-1/2}\big\}\Big)\\
&\quad\le \mathbb{P}\Big(\sum_{k,i}\big(\mathbb{1}\big\{|X_{ik,d,n}| > n^{1/2}d^{\alpha-1/2}\big\} - \mathbb{E}\mathbb{1}\big\{|X_{ik,d,n}| > n^{1/2}d^{\alpha-1/2}\big\}\big)\\
&\qquad\quad \ge d^{1-\delta} - d^{2(1-\alpha)}\Big) \qquad\qquad (\text{A.5})\\
&\quad\le \mathbb{P}\Big(\sum_{k,i}\big(\mathbb{1}\big\{|X_{ik,d,n}| > n^{1/2}d^{\alpha-1/2}\big\} - \mathbb{E}\mathbb{1}\big\{|X_{ik,d,n}| > n^{1/2}d^{\alpha-1/2}\big\}\big)\\
&\qquad\quad \ge \frac{1}{2}d^{1-\delta}\Big) \qquad\qquad (\text{A.6})\\
&\quad\le \exp\big(-\beta d^{1-\delta}\big), \qquad\qquad (\text{A.7})
\end{aligned}$$

where inequality (A.6) holds since $\alpha > (1+\delta)/2$. So, by inequality (A.3) follows

$$d_K\big(\mu^{\tilde{T}_{d,n}}, \mu^{\hat{T}_{d,n}}\big) \xrightarrow{\text{a.s.}} 0 \qquad \text{for } d \to \infty.$$

Note that $\tilde{X}_{d,n}$ is not centered and standardized, but by Cauchy–Schwarz inequality and Markov inequality,

$$\begin{aligned}
|\mathbb{E}\tilde{X}_{ik,d,n}| &= |\mathbb{E}X_{ik,d,n} - \mathbb{E}\tilde{X}_{ik,d,n}|\\
&= \big|\mathbb{E}X_{ik,d,n}\mathbb{1}\big(|X_{ik,d,n}| > n^{1/2}d^{\alpha-1/2}\big)\big|\\
&\le \sqrt{\mathbb{P}\big(|X_{ik,d,n}| > n^{1/2}d^{\alpha-1/2}\big)}\\
&\le n^{-1/2}d^{1/2-\alpha}
\end{aligned} \tag{A.8}$$

and moreover, $\operatorname{Var}(\tilde{X}_{ik,d,n}) \uparrow 1$ as $d \to \infty$. In the subsequent section we redefine the matrix $X_{d,n}$ by $\tilde{X}_{d,n}$ and keep the initial notations.

A.5. Step V: Replacing the normalizing matrix $n^{-1}\hat{W}_{d,n}$

Let

$$\begin{aligned}\tilde{T}_{d,n} = \frac{1}{n} W_{d,n} \circ \big((Y_{d,n} \circ \varepsilon_{d,n})(Y_{d,n} \circ \varepsilon_{d,n})^*\big) \\ - \frac{1}{n} W_{d,n} \circ \big((\hat{M}_{d,n} \circ \varepsilon_{d,n})(Y_{d,n} \circ \varepsilon_{d,n})^*\big) \\ - \frac{1}{n} W_{d,n} \circ \big((Y_{d,n} \circ \varepsilon_{d,n})(\hat{M}_{d,n} \circ \varepsilon_{d,n})^*\big).\end{aligned}$$

By Theorem C.9, the elementary inequality

$$\operatorname{tr}\big((C+D)^2\big) \le 2 \operatorname{tr}\big(C^2 + D^2\big)$$

for real, symmetric $d \times d$ matrices C and D, applied to

$$\begin{aligned}C &= \frac{1}{n}(\hat{W}_{d,n} - W_{d,n}) \circ \Big(\big((Y_{d,n} \circ \varepsilon_{d,n})(Y_{d,n} \circ \varepsilon_{d,n})^*\big)\Big), \\ D &= -\frac{1}{n}(\hat{W}_{d,n} - W_{d,n}) \circ \Big(\big((\hat{M}_{d,n} \circ \varepsilon_{d,n})(Y_{d,n} \circ \varepsilon_{d,n})^*\big) + \big((Y_{d,n} \circ \varepsilon_{d,n})(\hat{M}_{d,n} \circ \varepsilon_{d,n})^*\big)\Big),\end{aligned}$$

as well as the inequality

$$\operatorname{tr}\big[\big(A + A^*\big)^2\big] \le 4 \operatorname{tr}\big(AA^*\big)$$

for any $d \times d$ matrix A with real entries, we deduce

$$\begin{aligned}&d_L^3\big(\mu^{\tilde{T}_{d,n}}, \mu^{\hat{T}_{d,n}}\big) \\ &\quad \le \frac{1}{d} \operatorname{tr}\bigg[\bigg(\frac{1}{n}(\hat{W}_{d,n} - W_{d,n}) \circ \Big(\big((Y_{d,n} \circ \varepsilon_{d,n})(Y_{d,n} \circ \varepsilon_{d,n})^*\big) \\ &\qquad - \big((\hat{M}_{d,n} \circ \varepsilon_{d,n})(Y_{d,n} \circ \varepsilon_{d,n})^*\big) - \big((Y_{d,n} \circ \varepsilon_{d,n})(\hat{M}_{d,n} \circ \varepsilon_{d,n})^*\big)\Big)\bigg)^2\bigg] \\ &\quad \le \frac{2}{d} \operatorname{tr}\bigg[\bigg(\frac{1}{n}(\hat{W}_{d,n} - W_{d,n}) \circ \Big(\big((Y_{d,n} \circ \varepsilon_{d,n})(Y_{d,n} \circ \varepsilon_{d,n})^*\big)\Big)\bigg)^2\bigg] \\ &\qquad + \frac{8}{d} \operatorname{tr}\bigg[\frac{1}{n^2}(\hat{W}_{d,n} - W_{d,n})^2 \circ \Big(\big((\hat{M}_{d,n} \circ \varepsilon_{d,n})(Y_{d,n} \circ \varepsilon_{d,n})^*\big) \\ &\qquad \times \big((Y_{d,n} \circ \varepsilon_{d,n})(\hat{M}_{d,n} \circ \varepsilon_{d,n})^*\big)\Big)\bigg] \\ &\quad =: h_{d,n}.\end{aligned} \tag{A.9}$$

We prove that $h_{d,n} \to 0$ a.s. as $d \to \infty$. Thereto, define for an arbitrary constant

$$\gamma > \sqrt{4\alpha + 7} \tag{A.10}$$

the event

$$A_{d,n} = \left\{ \forall 1 \le i, j \le d : \left| (\hat{W}_{ij,d,n})^{-1} - (W_{ij,d,n})^{-1} \right| \le \gamma \sqrt{\frac{\log n}{n}} \right\}. \tag{A.11}$$

Then, for sufficiently large d the union bound and Hoeffding's inequality yield

$$\begin{aligned}
\mathbb{P}(A_{d,n}) &= 1 - \mathbb{P}\bigl(A_{d,n}^c\bigr) \\
&\ge 1 - d^2 \max_{i,j} \mathbb{P}\left(\left| (\hat{W}_{ij,d,n})^{-1} - (W_{ij,d,n})^{-1} \right| > \gamma \sqrt{\frac{\log n}{n}} \right) \\
&\ge 1 - 2d^2 \exp\left(-\frac{\gamma^2 \log n}{2} \right) \\
&= 1 - 2d^2 n^{-2\gamma^2}.
\end{aligned}$$

By the Borel–Cantelli lemma all but finitely many events $A_{d,n}$ almost surely occur. Hence, if $\mathbb{1}_{A_{d,n}} h_{d,n} \to 0$ a.s. for $d \to \infty$ then $h_{d,n} \to 0$ a.s. Note furthermore that on the event $A_{d,n}$,

$$\begin{aligned}
|\hat{W}_{ij,d,n} - W_{ij,d,n}| &= \left| \frac{1}{(\hat{W}_{ij,d,n})^{-1}} - \frac{1}{(W_{ij,d,n})^{-1}} \right| \\
&= \frac{|(\hat{W}_{ij,d,n})^{-1} - W_{ij,d,n}^{-1}|}{|(\hat{W}_{ij,d,n})^{-1}(W_{ij,d,n})^{-1}|} \\
&\le \frac{\gamma \sqrt{(\log n)/n}}{\min_i p_{i,d,n}^2 ((W_{ij,d,n})^{-1} - |(\hat{W}_{ij,d,n})^{-1} - (W_{ij,d,n})^{-1}|)} \qquad \text{(A.12)} \\
&\le \frac{\gamma \sqrt{(\log n)/n}}{\min_i p_{i,d,n}^2 (\min_i p_{i,d,n}^2 - \gamma \sqrt{(\log n)/n})} \\
&\le \frac{2\gamma}{\min_i p_{i,d,n}^4} \sqrt{\frac{\log n}{n}}
\end{aligned}$$

for d sufficiently large. Now we prove that $\mathbb{E}\mathbb{1}_{A_{d,n}} h_{d,n} \to 0$. In order to save space, the explicit dependence on d and n is suppressed in the displays until the end of the section. By inequality (A.12), we have

$$\begin{aligned}
\mathbb{E} h \mathbb{1}_A &\le \frac{8\gamma^2 \log n}{\min_i p_i^8 d n^3} \sum_{i,j=1}^{d} \mathbb{E}\left(\left(\sum_{k=1}^{n} Y_{ik} Y_{jk} \varepsilon_{ik} \varepsilon_{jk} \right)^2 + 4 \left(\sum_{k=1}^{n} \hat{M}_{ik} Y_{jk} \varepsilon_{ik} \varepsilon_{jk} \right)^2 \right) \mathbb{1}_A \\
&\le \frac{8\gamma^2 \log n}{\min_i p_i^8 d n^3} \left(\sum_{i,j=1}^{d} \sum_{k,l=1}^{n} |\mathbb{E} Y_{ik} Y_{jk} Y_{il} Y_{jl}| + 4 \sum_{k,l=1}^{n} \mathbb{E} \hat{M}_{ik} Y_{jk} \hat{M}_{il} Y_{jl} \varepsilon_{ik} \varepsilon_{jk} \varepsilon_{il} \varepsilon_{jl} \mathbb{1}_A \right) \\
&= I_1 + I_2,
\end{aligned}$$

where

$$I_1 = \frac{8\gamma^2 \log n}{\min_i p_i^8 dn^3} \sum_{i,j=1}^{d} \sum_{k,l=1}^{n} |\mathbb{E} Y_{ik} Y_{jk} Y_{il} Y_{jl}|$$

and

$$I_2 = \frac{32\gamma^2 \log n}{\min_i p_i^8 dn^3} \sum_{i,j=1}^{d} \sum_{k,l=1}^{n} \mathbb{E} \hat{M}_{ik} Y_{jk} \hat{M}_{il} Y_{jl} \varepsilon_{ik} \varepsilon_{jk} \varepsilon_{il} \varepsilon_{jl} \mathbb{1}_A.$$

For the first term, we obtain by (A.8), (2.2), uniform boundedness of the entries of $T_{d,n}$, and (A.1)

$$\begin{aligned} I_1 &= \frac{8\gamma^2 \log n}{\min_i p_i^8 dn^3} \Bigg(\sum_{\substack{i,j=1 \\ i \neq j}}^{d} \sum_{\substack{k,l=1 \\ k \neq l}}^{n} |\mathbb{E} Y_{ik} Y_{jk} Y_{il} Y_{jl}| + \sum_{i=1}^{d} \sum_{\substack{k,l=1 \\ k \neq l}}^{n} \left|\mathbb{E} Y_{ik}^2 Y_{il}^2\right| \\ &\quad + \sum_{\substack{i,j=1 \\ i \neq j}}^{d} \sum_{k=1}^{n} \left|\mathbb{E} Y_{ik}^2 Y_{jk}^2\right| + \sum_{i=1}^{d} \sum_{k=1}^{n} \mathbb{E} Y_{ik}^4 \Bigg) \\ &\lesssim \frac{\log n}{nd^{4\alpha-1}} + \frac{\log n}{n} + \frac{d \log n}{n^2} + \frac{\log n}{nd^{1-2\alpha}} \\ &\lesssim \frac{\log n}{nd^{1-2\alpha}} \longrightarrow 0. \end{aligned}$$

Recall the definition (4.1) of $\hat{M}_{d,n}$. Using again the bound

$$|\hat{W}_{ii}| \leq \frac{1}{(W_{ii})^{-1} - |(\hat{W}_{ii})^{-1} - (W_{ii})^{-1}|} \leq \frac{2}{\min_i p_i^2} \qquad \text{on the event } A \tag{A.13}$$

for d sufficiently large, we get for the second term with the same type of arguments

$$\begin{aligned} I_2 &= \frac{24\gamma^2 \log n}{\min_i p_i^8 dn^3} \sum_{i,j=1}^{d} \sum_{k_1,k_2,k_3,k_4=1}^{n} \mathbb{E} \frac{1}{n^2} \hat{W}_{ii}^2 Y_{jk_1} Y_{jk_2} Y_{ik_3} Y_{ik_4} \varepsilon_{ik_1} \varepsilon_{jk_1} \varepsilon_{ik_2} \varepsilon_{jk_2} \varepsilon_{ik_3} \varepsilon_{ik_4} \mathbb{1}_A \\ &\leq \frac{96\gamma^2 \log n}{\min_i p_i^{12} dn^5} \sum_{i,j=1}^{d} \sum_{k_1,k_2,k_3 k_4=1}^{n} |\mathbb{E} Y_{jk_1} Y_{jk_2} Y_{ik_3} Y_{ik_4}| \\ &\lesssim \frac{\log n}{dn^5} \Bigg[\sum_{i=1}^{d} \Bigg(\sum_{\substack{k_1,k_2,k_3,k_4=1 \\ k_1 \neq k_2 \neq k_3 \neq k_4}}^{n} + \sum_{\substack{k_1,k_2,k_3,k_4=1 \\ \neg(k_1 \neq k_2 \neq k_3 \neq k_4)}}^{n} \Bigg) |\mathbb{E} Y_{ik_1} Y_{ik_2} Y_{ik_3} Y_{ik_4}| \\ &\quad + \sum_{\substack{i,j=1 \\ i \neq j}}^{d} \Bigg(\sum_{\substack{k_1,k_2,k_3,k_4=1 \\ k_1 \neq k_2 \neq k_3 \neq k_4}}^{n} + \sum_{\substack{k_1,k_2,k_3,k_4=1 \\ \neg(k_1 \neq k_2 \neq k_3 \neq k_4)}}^{n} \Bigg) |\mathbb{E} Y_{jk_1} Y_{jk_2} Y_{ik_3} Y_{ik_4}| \Bigg] \end{aligned}$$

$$
\begin{aligned}
&\lesssim \frac{\log n}{dn^5}\big(d^{3-4\alpha}n^2 + d^{2\alpha}n^4 + d^{4-4\alpha}n^2 + d^2n^3\big) \\
&\lesssim \frac{d^{2\alpha-1}\log n}{n} \longrightarrow 0.
\end{aligned}
$$

We need a sufficiently tight bound on the variance of $h_{d,n}\mathbb{1}_{A_{d,n}}$ in order to conclude by the Borel–Cantelli lemma that in addition $h_{d,n}\mathbb{1}_{A_{d,n}} \to 0$ almost surely. Thereto, define

$$
\hat{G}_{ij,d,n} = \frac{1}{n}(\hat{W}_{ij,d,n} - W_{ij,d,n}), \qquad i, j = 1, \dots, d.
$$

Using (A.12) in (A.15) and dropping those summands of (A.14) whose indices satisfy $\{i_1, j_1\} \cap \{i_2, j_2\} \neq \varnothing$, we get

$$
\begin{aligned}
&\operatorname{Var} h\mathbb{1}_A \\
&\quad = \frac{1}{d^2}\sum_{i_1,i_2,j_1,j_2=1}^{d} \mathbb{E}\Bigg\{\hat{G}_{i_1j_1}^2\Bigg(2\Bigg(\sum_{k\in\mathcal{N}_{i_1j_1}} Y_{i_1k}Y_{j_1k}\Bigg)^2 + 8\Bigg(\sum_{k\in\mathcal{N}_{i_1j_1}} \hat{M}_{i_1k}Y_{j_1k}\Bigg)^2\Bigg) \\
&\qquad \times \hat{G}_{i_2j_2}^2\Bigg(2\Bigg(\sum_{k\in\mathcal{N}_{i_2j_2}} Y_{i_2k}Y_{j_2k}\Bigg)^2 + 8\Bigg(\sum_{k\in\mathcal{N}_{i_2j_2}} \hat{M}_{i_2k}Y_{j_2k}\Bigg)^2\Bigg)\mathbb{1}_A\Bigg\} \\
&\qquad - \frac{1}{d^2}\sum_{i_1,i_2,j_1,j_2=1}^{d} \mathbb{E}\Bigg\{\hat{G}_{i_1j_1}^2\Bigg(2\Bigg(\sum_{k\in\mathcal{N}_{i_1j_1}} Y_{i_1k}Y_{j_1k}\Bigg)^2 + 8\Bigg(\sum_{k\in\mathcal{N}_{i_1j_1}} \hat{M}_{i_1k}Y_{j_1k}\Bigg)^2\Bigg)\mathbb{1}_A\Bigg\} \qquad \text{(A.14)} \\
&\qquad \times \mathbb{E}\Bigg\{\hat{G}_{i_2j_2}^2\Bigg(2\Bigg(\sum_{k\in\mathcal{N}_{i_2j_2}} Y_{i_2k}Y_{j_2k}\Bigg)^2 + 8\Bigg(\sum_{k\in\mathcal{N}_{i_2j_2}} \hat{M}_{i_2k}Y_{j_2k}\Bigg)^2\Bigg)\mathbb{1}_A\Bigg\} \\
&\quad \le \frac{2^{10}\gamma^4(\log n)^2}{\min p_i^{16}d^2n^6} \sum_{\substack{i_1,i_2,j_1,j_2=1 \\ \{i_1,j_1\}\cap\{i_2,j_2\}\neq\varnothing}}^{d} \mathbb{E}\Bigg\{\Bigg(\Bigg(\sum_{k\in\mathcal{N}_{i_1j_1}} Y_{i_1k}Y_{j_1k}\Bigg)^2 + \Bigg(\sum_{k\in\mathcal{N}_{i_1j_1}} \hat{M}_{i_1k}Y_{j_1k}\Bigg)^2\mathbb{1}_A\Bigg) \\
&\qquad \times \Bigg(\Bigg(\sum_{k\in\mathcal{N}_{i_2j_2}}^{n} Y_{i_2k}Y_{j_2k}\Bigg)^2 + \Bigg(\sum_{k\in\mathcal{N}_{i_2j_2}} \hat{M}_{i_2k}Y_{j_2k}\Bigg)^2\mathbb{1}_A\Bigg)\Bigg\} \qquad \text{(A.15)} \\
&\qquad + \frac{1}{d^2}\sum_{\substack{i_1,i_2,j_1,j_2=1 \\ \{i_1,j_1\}\cap\{i_2,j_2\}=\varnothing}}^{d} \mathbb{E}\Bigg\{\hat{G}_{i_1j_1}^2\Bigg(\Bigg(\sum_{k\in\mathcal{N}_{i_1j_1}} Y_{i_1k}Y_{j_1k}\Bigg)^2 + \Bigg(\sum_{k\in\mathcal{N}_{i_1j_1}} \hat{M}_{i_1k}Y_{j_1k}\Bigg)^2\Bigg) \\
&\qquad \times \hat{G}_{i_2j_2}^2\Bigg(\Bigg(\sum_{k\in\mathcal{N}_{i_2j_2}} Y_{i_2k}Y_{j_2k}\Bigg)^2 + \Bigg(\sum_{k\in\mathcal{N}_{i_2j_2}} \hat{M}_{i_2k}Y_{j_2k}\Bigg)^2\Bigg)\mathbb{1}_A\Bigg\} \qquad \text{(A.16)}
\end{aligned}
$$

$$
\begin{aligned}
&- \frac{1}{d^2} \sum_{\substack{i_1,i_2,j_1,j_2=1\\ \{i_1,j_1\}\cap\{i_2,j_2\}=\varnothing}}^{d} \mathbb{E}\Big\{\hat{G}^2_{i_1 j_1}\Big(\Big(\sum_{k\in\mathcal{N}_{i_1 j_1}} Y_{i_1 k} Y_{j_1 k}\Big)^2 + \Big(\sum_{k\in\mathcal{N}_{i_1 j_1}} \hat{M}_{i_1 k} Y_{j_1 k}\Big)^2\Big)\mathbb{1}_A\Big\} \\
&\times \mathbb{E}\Big\{\hat{G}^2_{i_2 j_2}\Big(\Big(\sum_{k=\in\mathcal{N}_{i_2 j_2}} Y_{i_2 k} Y_{j_2 k}\Big)^2 + \Big(\sum_{k\in\mathcal{N}_{i_2 j_2}} \hat{M}_{i_2 k} Y_{j_2 k}\Big)^2\Big)\mathbb{1}_A\Big\} \qquad \text{(A.17)}\\
&= I_1 + I_2,
\end{aligned}
$$

where I_1 consists of the term (A.15) and I_2 of (A.16) and (A.17). The term I_1 yields

$$
I_1 \lesssim I_{1,1} + I_{1,2} + I_{1,3},
$$

with

$$
\begin{aligned}
I_{1,1} &= \frac{(\log n)^2}{d^2 n^6} \sum_{\substack{i_1,i_2,j_1,j_2=1\\ \{i_1,j_1\}\cap\{i_2,j_2\}\neq\varnothing}}^{d} \sum_{k_1,k_2,k_3,k_4=1}^{n} |\mathbb{E} Y_{i_1 k_1} Y_{j_1 k_1} Y_{i_1 k_2} Y_{j_1 k_2} Y_{i_2 k_3} Y_{j_2 k_3} Y_{i_2 k_4} Y_{j_2 k_4}|, \\
I_{1,2} &= \frac{(\log n)^2}{d^2 n^6} \sum_{\substack{i_1,i_2,j_1,j_2=1\\ \{i_1,j_1\}\cap\{i_2,j_2\}\neq\varnothing}}^{d} \sum_{k_1,k_2,k_3,k_4=1}^{n} \big|\mathbb{E}(Y_{i_1 k_1} Y_{j_1 k_1} Y_{i_1 k_2} Y_{j_1 k_2} \hat{M}_{i_2 k_3} Y_{j_2 k_3} \hat{M}_{i_2 k_4} Y_{j_2 k_4} \\
&\quad \times \varepsilon_{i_1 k_1}\varepsilon_{j_1 k_1}\varepsilon_{i_1 k_2}\varepsilon_{j_1 k_2}\varepsilon_{i_2 k_3}\varepsilon_{j_2 k_3}\varepsilon_{i_2 k_4}\varepsilon_{j_2 k_4}\mathbb{1}_A)\big|, \\
I_{1,3} &= \frac{(\log n)^2}{d^2 n^6} \sum_{\substack{i_1,i_2,j_1,j_2=1\\ \{i_1,j_1\}\cap\{i_2,j_2\}\neq\varnothing}}^{d} \sum_{k_1,k_2,k_3,k_4=1}^{n} \big|\mathbb{E}(\hat{M}_{i_1 k_1} Y_{j_1 k_1} \hat{M}_{i_1 k_2} Y_{j_1 k_2} \hat{M}_{i_2 k_3} Y_{j_2 k_3} \hat{M}_{i_2 k_4} Y_{j_2 k_4} \\
&\quad \times \varepsilon_{i_1 k_1}\varepsilon_{j_1 k_1}\varepsilon_{i_1 k_2}\varepsilon_{j_1 k_2}\varepsilon_{i_2 k_3}\varepsilon_{j_2 k_3}\varepsilon_{i_2 k_4}\varepsilon_{j_2 k_4}\mathbb{1}_A)\big|.
\end{aligned}
$$

For $I_{1,1}$ we have

$$
\begin{aligned}
I_{1,1} &= \frac{(\log n)^2}{d^2 n^6} \sum_{\substack{i_1,i_2,j_1,j_2=1\\ \{i_1,j_1\}\cap\{i_2,j_2\}\neq\varnothing\\ i_1\neq j_1 \vee i_2\neq j_2}}^{d} \sum_{k_1,k_2,k_3,k_4=1}^{n} |\mathbb{E} Y_{i_1 k_1} Y_{j_1 k_1} Y_{i_1 k_2} Y_{j_1 k_2} Y_{i_2 k_3} Y_{j_2 k_3} Y_{i_2 k_4} Y_{j_2 k_4}| \\
&\quad + \frac{(\log n)^2}{d^2 n^6} \sum_{i=1}^{d} \sum_{k_1,k_2,k_3,k_4=1}^{n} \mathbb{E} Y^2_{ik_1} Y^2_{ik_2} Y^2_{ik_3} Y^2_{ik_4} \\
&\lesssim \frac{n^{4\alpha}(\log n)^2}{n^4},
\end{aligned}
$$

where we used for i_1, j_1, i_2, j_2 with $\{i_1, j_1\} \cap \{i_2, j_2\} \neq \varnothing$ and $i_1 \neq j_1$ or $i_2 \neq j_2$ the bounds

$$|\mathbb{E} Y_{i_1k_1} Y_{j_1k_1} Y_{i_1k_2} Y_{j_1k_2} Y_{i_2k_3} Y_{j_2k_3} Y_{i_2k_4} Y_{j_2k_4}| \lesssim \begin{cases} n^2 d^{4\alpha-2}, & \text{for } \#\{k_1, k_2, k_3, k_4\} = 1, \\ n d^{2\alpha-1}, & \text{for } \#\{k_1, k_2, k_3, k_4\} = 2, \\ 1, & \text{for } \#\{k_1, k_2, k_3, k_4\} = 3, \\ n^{-2} d^{2-4\alpha}, & \text{for } \#\{k_1, k_2, k_3, k_4\} = 4 \end{cases}$$

and for $i = i_1 = j_1 = i_2 = j_2$ the estimates

$$\mathbb{E} Y_{ik_1}^2 Y_{ik_2}^2 Y_{ik_3}^2 Y_{ik_4}^2 \lesssim \begin{cases} n^3 d^{6\alpha-3}, & \text{for } \#\{k_1, k_2, k_3, k_4\} = 1, \\ n^2 d^{4\alpha-2}, & \text{for } \#\{k_1, k_2, k_3, k_4\} = 2, \\ n d^{2\alpha-1}, & \text{for } \#\{k_1, k_2, k_3, k_4\} = 3, \\ 1, & \text{for } \#\{k_1, k_2, k_3, k_4\} = 4. \end{cases}$$

These estimates are deduced by the following consideration. First, the expectation is factorized by independence into a product of moments of the Y_{ik}'s. Then applying (A.1) and (A.8), the lth moment is bounded by

$$\left|\mathbb{E} Y_{ik}^l\right| \lesssim \left(n^{1/2} d^{\alpha-1/2}\right)^{l-2}, \qquad l \in \mathbb{N}.$$

Now we evaluate $I_{1,2}$. Using (A.13),

$$\begin{aligned} I_{1,2} &= \frac{(\log n)^2}{d^2 n^6} \sum_{\substack{i_1, i_2, j_1, j_2 = 1 \\ \{i_1, j_1\} \cap \{i_2, j_2\} \neq \varnothing}}^{d} \sum_{k_1, \dots, k_6 = 1}^{n} |\mathbb{E} Y_{i_1k_1} Y_{j_1k_1} Y_{i_1k_2} Y_{j_1k_2} Y_{j_2k_3} Y_{j_2k_4} Y_{i_2k_5} Y_{i_2k_6}| \\ &\quad \times \mathbb{E}\left(\frac{1}{n^2} \hat{W}_{i_2i_2}^2 \varepsilon_{i_1k_1} \varepsilon_{j_1k_1} \varepsilon_{i_1k_2} \varepsilon_{j_1k_2} \varepsilon_{i_2k_3} \varepsilon_{j_2k_3} \varepsilon_{i_2k_4} \varepsilon_{j_2k_4} \varepsilon_{i_2k_5} \varepsilon_{i_2k_6} \mathbb{1}_A\right) \\ &\lesssim \frac{(\log n)^2}{d^2 n^8} \sum_{\substack{i_1, i_2, j_1, j_2 = 1 \\ \{i_1, j_1\} \cap \{i_2, j_2\} \neq \varnothing}}^{d} \sum_{k_1, \dots, k_6 = 1}^{n} |\mathbb{E} Y_{i_1k_1} Y_{j_1k_1} Y_{i_1k_2} Y_{j_1k_2} Y_{i_2k_5} Y_{j_2k_3} Y_{i_2k_6} Y_{j_2k_4}| \\ &\lesssim \frac{(\log n)^2 d^{6\alpha}}{d^2 n^4}, \end{aligned}$$

where we used for the bound

$$\begin{aligned} &|\mathbb{E} Y_{i_1k_1} Y_{j_1k_1} Y_{i_1k_2} Y_{j_1k_2} Y_{i_2k_5} Y_{j_2k_3} Y_{i_2k_6} Y_{j_2k_4}| \\ &\quad \lesssim \left(\frac{d}{n}\right)^{i-4} d^{2\alpha(4-i)} \qquad \text{for } i = \#\{k_1, k_2, k_3, k_4, k_5, k_6\}. \end{aligned}$$

Again by (A.13), we obtain with the same argument as for $I_{1,2}$

$$I_{1,3} \lesssim \frac{(\log n)^2}{d^2 n^{10}} \sum_{\substack{i_1,i_2,j_1,j_2=1\\ \{i_1,j_1\}\cap\{i_2,j_2\}\neq\varnothing}}^{d} \sum_{k_1,\dots,k_8=1}^{n} |\mathbb{E} Y_{i_1k_5} Y_{i_1k_6} Y_{j_1k_1} Y_{j_1k_2} Y_{i_2k_7} Y_{i_2k_8} Y_{j_2k_3} Y_{j_2k_4}|$$
$$\lesssim \frac{(\log n)^2 d^{6\alpha}}{d^2 n^6}$$

with

$$|\mathbb{E} Y_{i_1k_5} Y_{i_1k_6} Y_{j_1k_1} Y_{j_1k_2} Y_{i_2k_7} Y_{i_2k_8} Y_{j_2k_3} Y_{j_2k_4}|$$
$$\lesssim \left(\frac{d}{n}\right)^{i-4} d^{2\alpha(4-i)} \qquad \text{for } i = \#\{k_1,k_2,k_3,k_4,k_5,k_6,k_7,k_8\}.$$

As concerns I_2, define

$$U_{ij,d,n} = \hat{G}^2_{ij,d,n}\left\{\left(\sum_{k\in\mathcal{N}_{ij,d,n}} Y_{ik,d,n} Y_{jk,d,n}\right)^2 + \left(\sum_{k\in\mathcal{N}_{ij,d,n}} \hat{M}_{ik,d,n} Y_{jk,d,n}\right)^2\right\},$$

and note that $U_{ij,d,n}$ is bounded by a constant multiple of $n^6 d^{4\alpha-2}$ because $\mathcal{N}_{ij,d,n}$ contains at most n elements, $\hat{G}^2_{ij,d,n} \lesssim 1$ since by Section A.1 $\min_i p_{i,d,n}$ is uniformly bounded away from zero, $|Y_{ik,d,n}| \lesssim n^{1/2} d^{\alpha-1/2}$ by Section A.3 and Section A.4,

$$|\hat{M}_{ik,d,n}| = \left|\frac{1}{N_{ii,d,n}} \sum_{l\in\mathcal{N}_{ii,d,n}} Y_{il,d,n}\right| \lesssim n^{1/2} d^{\alpha-1/2}.$$

Hence,

$$I_2 = \frac{1}{d^2} \sum_{\substack{i_1,i_2,j_1,j_2=1\\ \{i_1,j_1\}\cap\{i_2,j_2\}=\varnothing}}^{d} \mathbb{E}(U_{i_1j_1} U_{i_2j_2} \mathbb{1}_A) - \mathbb{E}(U_{i_1j_1}\mathbb{1}_A)\mathbb{E}(U_{i_2j_2}\mathbb{1}_A)$$
$$= \frac{1}{d^2} \sum_{\substack{i_1,i_2,j_1,j_2=1\\ \{i_1,j_1\}\cap\{i_2,j_2\}=\varnothing}}^{d} \big\{-\mathbb{E}(U_{i_1j_1} U_{i_2j_2} \mathbb{1}_{A^c}) + \mathbb{E}(U_{i_1j_1}\mathbb{1}_{A^c})\mathbb{E}(U_{i_2j_2})$$
$$+ \mathbb{E}(U_{i_1j_1})\mathbb{E}(U_{i_2j_2}\mathbb{1}_{A^c}) - \mathbb{E}(U_{i_1j_1}\mathbb{1}_{A^c})\mathbb{E}(U_{i_2j_2}\mathbb{1}_{A^c})\big\}$$
$$\leq \frac{2}{d^2} \sum_{\substack{i_1,i_2,j_1,j_2=1\\ \{i_1,j_1\}\cap\{i_2,j_2\}=\varnothing}}^{d} \mathbb{E}(U_{i_1j_1})\mathbb{E}(U_{i_2j_2}\mathbb{1}_{A^c})$$
$$\lesssim d^{8\alpha-2} n^{12} \mathbb{P}(A^c) \lesssim n^{12+8\alpha-2\gamma^2}.$$

Note that by choice of γ in (5.40) the exponent in the last line is strictly smaller than -1. Therefore by the lemma of Borel–Cantelli $h_{d,n}\mathbb{1}_{A_{d,n}} \to 0$ almost surely ($d \to \infty$). In the following subsection we redefine the matrix $\hat{T}_{d,n}$ by $\tilde{T}_{d,n}$.

A.6. Step VI: Removing $n^{-1}W \circ ((Y \circ \varepsilon)(\hat{M} \circ \varepsilon)^* + (\hat{M} \circ \varepsilon)(Y \circ \varepsilon)^*)$

By the same arguments as in Section A.4, we return to the original centered and standardized matrix $X_{d,n}$. Define

$$\tilde{T}_{d,n} = \frac{1}{n} W_{d,n} \circ \big((Y_{d,n} \circ \varepsilon_{d,n})(Y_{d,n} \circ \varepsilon_{d,n})^*\big).$$

We prove that

$$d_L\big(\mu^{T_{d,n}}, \mu^{\tilde{T}_{d,n}}\big) \to 0$$

almost surely. For $\gamma > 1$, define the event

$$\tilde{A}_{d,n} = \Big\{\max_i |N_{ii,d,n} - np_{i,d,n}| < \gamma\sqrt{n\log n}\Big\}.$$

Note that

$$\Big\{\max_i |N_{ii,d,n} - np_{i,d,n}| < \gamma\sqrt{n\log n}\Big\} = \bigg\{\max_i \bigg|\sum_{k=1}^n (\varepsilon_{ik,d,n} - p_{i,d,n})\bigg| < \gamma\sqrt{n\log n}\bigg\}$$

for d sufficiently large. The union bound and Hoeffding's inequality yield

$$\mathbb{P}\big(\tilde{A}^c_{d,n}\big) \le 2dn^{-2\gamma^2}. \tag{A.18}$$

By the Borel–Cantelli lemma, all but finitely many of the events (A_d) occur. Moreover, for $\frac{1}{2} < \eta < 1$ define the event

$$B_{d,n} = \left\{\sum_{i=1}^d \mathbb{1}\left\{|\hat{m}_{i,d,n}| > \sqrt{\frac{d^{2(1-\eta)}}{n}}\right\} < d^\eta\right\}.$$

First, observe that by the same type of argument as used in (A.13) and by Markov's inequality

$$\begin{aligned}
\max_i \mathbb{P}\left(|\hat{m}_{i,d,n}| > \sqrt{\frac{d^{2(1-\eta)}}{n}}, \tilde{A}_{d,n}\right) &\le \max_i \mathbb{P}\left(\frac{2}{n\min p_{i,d,n}}\bigg|\sum_k \varepsilon_{ik,d,n}Y_{ik,d,n}\bigg| > \sqrt{\frac{d^{2(1-\eta)}}{n}}\right) \\
&\le \frac{4\mathbb{E}(\sum_k \varepsilon_{ik,d,n}Y_{ik,d,n})^2}{n^2\min_i p^2_{i,d,n}(d^{2(1-\eta)}/n)} \\
&\lesssim d^{2\eta-2},
\end{aligned}$$

where we have used

$$\frac{1}{N_{ii,d,n}} \leq \frac{2}{n \min p_{i,d,n}}$$

for d sufficiently large in the first inequality. In particular,

$$\begin{aligned}\mathbb{E}\mathbb{1}\left\{|\hat{m}_{i,d,n}| > \sqrt{\frac{d^{2(1-\eta)}}{n}}\right\} &= \mathbb{E}\mathbb{1}\left\{|\hat{m}_{i,d,n}| > \sqrt{\frac{d^{2(1-\eta)}}{n}}\right\}(\mathbb{1}_{\tilde{A}_{d,n}} + \mathbb{1}_{\tilde{A}^c_{d,n}}) \\ &\leq \kappa\left(d^{2\eta-2} + dn^{-2\gamma^2}\right)\end{aligned}$$

for some suitably chosen constant $\kappa > 0$. We conclude for d sufficiently large by Hoeffding's inequality

$$\begin{aligned}\mathbb{P}\left(B^c_{d,n}\right) &\leq \mathbb{P}\left(\sum_{i=1}^{d}\mathbb{1}\left\{\hat{m}_{i,d,n} > \sqrt{\frac{d^{2(1-\eta)}}{n}}\right\} - \mathbb{E}\mathbb{1}\left\{\hat{m}_{i,d,n} > \sqrt{\frac{d^{2(1-\eta)}}{n}}\right\}\right. \\ &\qquad \left. > d^{\eta} - \kappa\left(d^{2\eta-1} - d^2 n^{-2\gamma^2}\right)\right) \\ &\leq \mathbb{P}\left(\sum_{i=1}^{d}\mathbb{1}\left\{\hat{m}_{i,d,n} > \sqrt{\frac{d^{2(1-\eta)}}{n}}\right\} - \mathbb{E}\mathbb{1}\left\{\hat{m}_{i,d,n} > \sqrt{\frac{d^{2(1-\eta)}}{n}}\right\} > \frac{1}{2}d^{\eta}\right) \\ &\leq \exp\left(-\frac{d^{2\eta-1}}{2}\right).\end{aligned}$$

By the Borel–Cantelli lemma, all but finitely many of the events $(B_{d,n})$ occur.

Let $\gamma' > 0$ be an appropriate constant such that for all n

$$2\sum_{k=1}^{n}\mathbb{E}|Y_{ik,d,n}| \leq \gamma' n.$$

Then, define the event

$$D_{d,n} = \left\{\sum_{i=1}^{d}\mathbb{1}\left\{\sum_{k=1}^{n}|Y_{ik,d,n}| > \gamma' n\right\} \leq \frac{d}{\log d}\right\}.$$

In the next step, we shall prove that $\mathbb{P}(\limsup_d D^c_{d,n}) = 0$ in order to remove the corresponding rows from the matrix Y. By Chebyshev's inequality we have

$$\begin{aligned}&\max_i \mathbb{P}\left(\sum_{k=1}^{n}|Y_{ik,d,n}| > \gamma' n\right) \\ &\quad \leq \max_i \mathbb{P}\left(\sum_{k=1}^{n}|Y_{ik,d,n}| - \mathbb{E}|Y_{ik,d,n}| > \gamma' n - \sum_{k=1}^{n}\mathbb{E}|Y_{ik,d,n}|\right)\end{aligned}$$

$$\leq \max_i \mathbb{P}\left(\sum_{k=1}^{n} |Y_{ik,d,n}| - \mathbb{E}|Y_{ik,d,n}| > \frac{1}{2}\gamma' n\right)$$
$$\leq \frac{\kappa'}{n}$$

for an appropriate constant $\kappa' > 0$. Again, by the Hoeffding inequality for sufficiently large d,

$$\begin{aligned}
\mathbb{P}\big(D_{d,n}^c\big) &\leq \mathbb{P}\left(\sum_{i=1}^{d} \mathbb{1}\left\{\sum_{k=1}^{n} |Y_{ik,d,n}| > \gamma' n\right\} - \mathbb{E}\mathbb{1}\left\{\sum_{k=1}^{n} |Y_{ik,d,n}| > \gamma' n\right\} > \frac{d}{\log d} - \frac{\kappa' d}{n}\right) \\
&\leq \mathbb{P}\left(\sum_{i=1}^{d} \mathbb{1}\left\{\sum_{k=1}^{n} |Y_{ik,d,n}| > \gamma' n\right\} - \mathbb{E}\mathbb{1}\left\{\sum_{k=1}^{n} |Y_{ik,d,n}| > \gamma' n\right\} > \frac{1}{2}\frac{d}{\log d}\right) \\
&\leq \exp\left(-\frac{d}{2(\log d)^2}\right),
\end{aligned}$$

and therefore $\mathbb{P}(\limsup_d D_{d,n}^c) = 0$. Now let

$$\begin{aligned}
\check{T}_{d,n} = \frac{1}{n} W_{d,n} \circ \big(&(Y_{d,n} \circ \varepsilon_{d,n})(Y_{d,n} \circ \varepsilon_{d,n})^* - (\tilde{Y}_{d,n} \circ \varepsilon_{d,n})(\tilde{M}_{d,n} \circ \varepsilon_{d,n})^* \\
&- (\tilde{M}_{d,n} \circ \varepsilon_{d,n})(\tilde{Y}_{d,n} \circ \varepsilon_{d,n})^*\big),
\end{aligned}$$

where

$$\tilde{M}_{ik,d,n} = \hat{M}_{ik,d,n} \mathbb{1}\left\{|\hat{M}_{ik,d,n}| \leq \sqrt{\frac{d^{2(1-\eta)}}{n}}\right\}$$

and

$$\tilde{Y}_{ik,d,n} = Y_{ik,d,n} \mathbb{1}\left\{\sum_{l=1}^{n} |Y_{il,d,n}| \leq \gamma' n\right\}.$$

By Theorem C.8 and due to $\mathbb{P}(\limsup_d (D_{d,n}^c \cup B_{d,n}^c)) = 0$ we conclude by the same type of arguments as in Section A.4

$$\begin{aligned}
&d_L\big(\mu^{\hat{T}_{d,n}}, \mu^{\check{T}_{d,n}}\big) \\
&\leq \frac{1}{d} \operatorname{rank}\Bigg(\frac{1}{n} W_{d,n} \circ \big((Y_{d,n} \circ \varepsilon_{d,n})(\hat{M}_{d,n} \circ \varepsilon_{d,n})^* + (\hat{M}_{d,n} \circ \varepsilon_{d,n})(Y_{d,n} \circ \varepsilon_{d,n})^* \\
&\quad - (\tilde{Y}_{d,n} \circ \varepsilon_{d,n})(\tilde{M}_{d,n} \circ \varepsilon_{d,n})^* - (\tilde{M}_{d,n} \circ \varepsilon_{d,n})(\tilde{Y}_{d,n} \circ \varepsilon_{d,n})^*\big)\Bigg) \\
&\xrightarrow{\text{a.s.}} 0 \qquad \text{as } d \to \infty.
\end{aligned}$$

In order to save space the explicit dependence on d and n is suppressed in the displays until the end of the section. By Theorem C.9,

$$\begin{aligned}
&d_L^3\big(\mu^{\check{T}}, \mu^{\tilde{T}}\big) &\text{(A.19)}\\
&\le \frac{1}{d}\operatorname{tr}\bigg(\bigg(\frac{1}{n}W \circ \big((\tilde{Y}\circ\varepsilon)(\tilde{M}\circ\varepsilon)^* + (\tilde{M}\circ\varepsilon)(\tilde{Y}\circ\varepsilon)^*\big)\bigg)\\
&\qquad \times \bigg(\frac{1}{n}W \circ \big((\tilde{Y}\circ\varepsilon)(\tilde{M}\circ\varepsilon)^* + (\tilde{M}\circ\varepsilon)(\tilde{Y}\circ\varepsilon)^*\big)\bigg)\bigg)\\
&= \frac{2}{d}\operatorname{tr}\bigg(\frac{1}{n^2}W^2 \circ \big((\tilde{M}\circ\varepsilon)(\tilde{Y}\circ\varepsilon)^*(\tilde{Y}\circ\varepsilon)(\tilde{M}\circ\varepsilon)^* + (\tilde{Y}\circ\varepsilon)(\tilde{M}\circ\varepsilon)^*(\tilde{Y}\circ\varepsilon)(\tilde{M}\circ\varepsilon)^*\big)\bigg)\\
&\le \frac{4}{d}\operatorname{tr}\bigg(\frac{1}{n^2}W^2 \circ \big((\tilde{M}\circ\varepsilon)(\tilde{Y}\circ\varepsilon)^*(\tilde{Y}\circ\varepsilon)(\tilde{M}\circ\varepsilon)^*\big)\bigg) &\text{(A.20)}\\
&\le \frac{4}{d}\sum_{i=1}^d \hat{m}_i^2 \mathbb{1}\bigg\{|\hat{m}_i| \le \sqrt{\frac{d^{2(1-\eta)}}{n}}\bigg\}\sum_{j=1}^d \frac{1}{n^2}W_{ij}^2\bigg(\sum_{k=1}^n \varepsilon_{ik}\varepsilon_{jk}Y_{jk}\bigg)^2 \mathbb{1}\bigg\{\sum_{l=1}^n |Y_{jl}| \le \gamma' n\bigg\}\\
&\lesssim \frac{d^{2(\eta-1)}}{dn^3}\sum_{i=1}^d\sum_{j=1}^d\bigg(\sum_{k=1}^n \varepsilon_{ik}\varepsilon_{jk}Y_{jk}\bigg)^2 \mathbb{1}\bigg\{\sum_{l=1}^n |Y_{jl}| \le \gamma' n\bigg\}, &\text{(A.21)}
\end{aligned}$$

where we have used the elementary inequality

$$\operatorname{tr}\big(C^2\big) \le \operatorname{tr}\big(CC^*\big) \qquad \text{for any } C \in \mathbb{R}^{d\times d}$$

in (A.20). It remains to prove that the last line (A.21) converges to zero almost surely. Let $\eta < \eta' < 1$, and rewrite

$$\begin{aligned}
&\max_i \mathbb{P}\bigg(\sum_{j=1}^d\bigg(\sum_{k=1}^n \varepsilon_{ik}\varepsilon_{jk}Y_{jk}\bigg)^2 \mathbb{1}\bigg\{\sum_{l=1}^n |Y_{jl}| \le \gamma' n\bigg\} \ge \frac{n^3}{d^{2(\eta'-1)}}\bigg)\\
&\quad = \max_i \mathbb{E}\bigg\{\mathbb{P}\bigg(\sum_{j=1}^d\bigg(\sum_{k=1}^n \varepsilon_{ik}\varepsilon_{jk}Y_{jk}\bigg)^2 \mathbb{1}\bigg\{\sum_{l=1}^n |Y_{jl}| \le \gamma' n\bigg\} \ge \frac{n^3}{d^{2(\eta'-1)}}\bigg|\varepsilon\bigg)\bigg\}.
\end{aligned} \tag{A.22}$$

Define for $\eta' < \eta'' < 1$ the random variables

$$I_{ij,d,n} = \mathbb{1}\bigg\{\bigg|\sum_{l=1}^n \varepsilon_{il,d,n}\varepsilon_{jl,d,n}Y_{jl,d,n}\bigg| \ge \sqrt{n}d^{(\eta''-1)}\bigg\}, \qquad 1 \le i, j \le d.$$

Then by Markov's inequality for the conditional probability and an appropriate constant $\kappa'' > 0$,

$$\mathbb{E}(I_{ij}|\varepsilon) = \mathbb{P}\bigg(\bigg|\sum_{l=1}^n \varepsilon_{il}\varepsilon_{jl}Y_{jl}\bigg| \ge \sqrt{nd^{2(\eta''-1)}}\bigg|\varepsilon\bigg) \le \sum_{l=1}^n \frac{\varepsilon_{il}\varepsilon_{jl}\mathbb{E}Y_{jl}^2}{nd^{2(\eta''-1)}} \le \frac{\kappa''}{d^{2(\eta''-1)}}.$$

The inner conditional probability in line (A.22) can be further estimated by

$$\mathbb{P}\left(\sum_{j=1}^{d}\left(\sum_{k=1}^{n}\varepsilon_{ik}\varepsilon_{jk}Y_{jk}\right)^{2}\mathbb{1}\left\{\sum_{l=1}^{n}|Y_{jl}|\le\gamma' n\right\}\ge\frac{n^{3}}{d^{2(\eta'-1)}}\middle|\varepsilon\right)$$

$$\le\mathbb{P}\left(\left(\gamma' n\right)^{2}\sum_{j=1}^{d}\mathbb{1}\left\{\sqrt{n}d^{(\eta''-1)}\le\left|\sum_{l=1}^{n}\varepsilon_{il}\varepsilon_{jl}Y_{jl}\right|\le\gamma' n\right\}\ge\frac{n^{3}}{2d^{2(\eta'-1)}}\middle|\varepsilon\right)$$

$$+P\left(nd^{2(\eta''-1)}\sum_{j=1}^{d}\mathbb{1}\left\{\left|\sum_{l=1}^{n}\varepsilon_{il}\varepsilon_{jl}Y_{jl}\right|\le\sqrt{n}d^{(\eta''-1)}\right\}\ge\frac{n^{3}}{2d^{2(\eta'-1)}}\middle|\varepsilon\right),$$

where the last conditional probability disappears for d sufficiently large. For the first probability on the right-hand side, we obtain

$$\mathbb{P}\left(\left(\gamma' n\right)^{2}\sum_{j=1}^{d}\mathbb{1}\left\{\sqrt{n}d^{(\eta''-1)}\le\left|\sum_{l=1}^{n}\varepsilon_{il}\varepsilon_{jl}Y_{jl}\right|\le\gamma' n\right\}\ge\frac{n^{3}}{2d^{2(\eta'-1)}}\middle|\varepsilon\right)$$

$$\le\mathbb{P}\left(\left(\gamma' n\right)^{2}\sum_{j=1}^{d}\left(I_{ij}-\mathbb{E}(I_{ij}|\varepsilon)\right)\ge\frac{n^{3}}{2d^{2(\eta'-1)}}-\kappa''\frac{(\gamma' n)^{2}d}{d^{2(\eta''-1)}}\middle|\varepsilon\right)$$

$$\le\mathbb{P}\left(\sum_{j=1}^{d}\left(I_{ij}-\mathbb{E}(I_{ij}|\varepsilon)\right)\ge\frac{n}{4\gamma'^{2}d^{2(\eta'-1)}}\middle|\varepsilon\right)$$

for d sufficiently large. Finally, by Hoeffding's inequality the last line is bounded by

$$\exp\left(-\frac{n^{2}}{8\gamma'^{4}d^{4\eta'-3}}\right).$$

Altogether, (A.21) is bounded by $d^{2(\eta-\eta')}$ with probability

$$1-\mathbb{P}\left(\frac{d^{2(\eta-1)}}{dn^{3}}\sum_{i=1}^{d}\sum_{j=1}^{d}\left(\sum_{k=1}^{n}\varepsilon_{ik}\varepsilon_{jk}Y_{jk}\right)^{2}\mathbb{1}\left\{\sum_{l=1}^{n}|Y_{jl}|\le\gamma'\right\}\ge d^{2(\eta-\eta')}\right)$$

$$\ge 1-d\max_{i}\mathbb{P}\left(\sum_{j=1}^{d}\left(\sum_{k=1}^{n}\varepsilon_{ik}\varepsilon_{jk}Y_{jk}\right)^{2}\mathbb{1}\left\{\sum_{l=1}^{n}|Y_{jl}|\le\gamma' n\right\}\ge\frac{n^{3}}{d^{2(\eta'-1)}}\right)$$

$$\ge 1-d\exp\left(-\frac{n^{2}}{8\gamma'^{4}d^{4\eta'-3}}\right).$$

By the lemma of Borel–Cantelli,

$$d_{L}^{3}\left(\mu^{\check{T}_{d,n}},\mu^{\tilde{T}_{d,n}}\right)\to 0$$

almost surely. Consequently,

$$d_L\big(\mu^{\hat{T}_{d,n}},\mu^{\tilde{T}_{d,n}}\big) \le d_L\big(\mu^{\hat{T}_{d,n}},\mu^{\check{T}_{d,n}}\big) + d_L\big(\mu^{\check{T}_{d,n}},\mu^{\tilde{T}_{d,n}}\big) \xrightarrow{\text{a.s.}} 0 \qquad \text{as } d\to\infty.$$

Subsequently, we denote $\tilde{T}_{d,n}$ by $\hat{T}_{d,n}$.

A.7. Step VII: Diagonal manipulation

Rewrite the matrix $\hat{T}_{d,n}$ in the following way

$$\begin{aligned}\hat{T}_{d,n} = {} & \frac{1}{n}\big(w_{d,n}w_{d,n}^*\big)\circ\big((Y_{d,n}\circ\varepsilon_{d,n})(Y_{d,n}\circ\varepsilon_{d,n})^*\big)\\ & - \operatorname{diag}\bigg[\frac{1}{n}\big(w_{d,n}w_{d,n}^*\big)\circ\big((Y_{d,n}\circ\varepsilon_{d,n})(Y_{d,n}\circ\varepsilon_{d,n})^*\big)\\ & - \frac{1}{n}W_{d,n}\circ\big((Y_{d,n}\circ\varepsilon_{d,n})(Y_{d,n}\circ\varepsilon_{d,n})^*\big)\bigg].\end{aligned}$$

In this step, we replace the diagonal matrix

$$\begin{aligned}\hat{S}_{d,n} := \operatorname{diag}\bigg[&\frac{1}{n}\big(w_{d,n}w_{d,n}^*\big)\circ\big((Y_{d,n}\circ\varepsilon_{d,n})(Y_{d,n}\circ\varepsilon_{d,n})^*\big)\\ & - \frac{1}{n}W_{d,n}\circ\big((Y_{d,n}\circ\varepsilon_{d,n})(Y_{d,n}\circ\varepsilon_{d,n})^*\big)\bigg]\end{aligned}$$

by its diagonal deterministic counterpart $S_{d,n}$ with

$$S_{ii,d,n} = \frac{1-p_{i,d,n}}{p_{i,d,n}}T_{ii,d,n}, \qquad i=1,\dots,d.$$

Thereto, we use similar arguments as in the last subsection. In contrast to the last subsection, we cannot simply rely on Markov's inequality since $Y_{ik,d,n}$ is assumed to possess only two moments. In order to save space, the explicit dependence on d and n is suppressed in the displays until the end of the section. Note that for any $u>0$,

$$\begin{aligned}\alpha_{\max} &= \max_{i=1,\dots,d}\mathbb{P}\big(|\hat{S}_{ii}-S_{ii}|>u\big)\\ &= \max_{i=1,\dots,d}\mathbb{P}\left(\left|\frac{1-p_i}{np_i}\sum_{k=1}^n\left(Y_{ik}^2\frac{\varepsilon_{ik}}{p_i}-T_{ii}\right)\right|>u\right)\\ &\le \max_{i=1,\dots,d}\mathbb{P}\left(\left|\frac{1-p_i}{p_i}T_{ii}-\frac{1-p_i}{np_i}\sum_{k=1}^n Y_{ik}^2\frac{\varepsilon_{ik}}{p_i}\right|>u, \left|\sum_{k=1}^n(\varepsilon_{ik}-p_i)\right|>\sqrt{n\log n}\right)\\ &\quad + \max_{i=1,\dots,d}\mathbb{P}\left(\left|\frac{1-p_i}{p_i}T_{ii}-\frac{1-p_i}{np_i}\sum_{k=1}^n Y_{ik}^2\frac{\varepsilon_{ik}}{p_i}\right|>u, \left|\sum_{k=1}^n(\varepsilon_{ik}-p_i)\right|\le\sqrt{n\log n}\right).\end{aligned}$$

As concerns the first term in this last inequality, Hoeffding's inequality yields

$$
\begin{aligned}
&\max_{i=1,\dots,d} \mathbb{P}\left(\left|\frac{1-p_i}{p_i}T_{ii} - \frac{1-p_i}{np_i}\sum_{k=1}^n Y_{ik}^2 \frac{\varepsilon_{ik}}{p_i}\right| > u, \left|\sum_{k=1}^n (\varepsilon_{ik} - p_i)\right| > \sqrt{n\log n}\right) \\
&\quad \le \max_{i=1,\dots,d} \mathbb{P}\left(\left|\sum_{k=1}^n (\varepsilon_{ik} - p_i)\right| > \sqrt{n\log n}\right) \\
&\quad \le 2n^{-2}.
\end{aligned}
$$

In order to bound the second term, note that

$$
\begin{aligned}
&\max_{i=1,\dots,d} \mathbb{P}\left(\left|\frac{1-p_i}{p_i}T_{ii} - \frac{1-p_i}{np_i}\sum_{k=1}^n Y_{ik}^2 \frac{\varepsilon_{ik}}{p_i}\right| > u, \left|\sum_{k=1}^n (\varepsilon_{ik} - p_i)\right| \le \sqrt{n\log n}\right) \\
&\quad = \max_{i=1,\dots,d} \sum_{l=\lceil np_i - \sqrt{n\log n}\rceil}^{\lfloor np_i + \sqrt{n\log n}\rfloor} \mathbb{P}\left(\left|\frac{1-p_i}{p_i}T_{ii} - \frac{1-p_i}{np_i}\sum_{k=1}^n Y_{ik}^2 \frac{\varepsilon_{ik}}{p_i}\right| > u, \sum_{k=1}^n \varepsilon_{ik} = l\right) \\
&\quad = \max_{i=1,\dots,d} \sum_{l=\lceil np_i - \sqrt{n\log n}\rceil}^{\lfloor np_i + \sqrt{n\log n}\rfloor} \mathbb{P}\left(\left|\frac{1-p_i}{p_i}T_{ii} - \frac{1-p_i}{np_i}\sum_{k=1}^n Y_{ik}^2 \frac{\varepsilon_{ik}}{p_i}\right| > u \,\middle|\, \sum_{k=1}^n \varepsilon_{ik} = l\right) \qquad \text{(A.23)} \\
&\qquad \times \mathbb{P}\left(\sum_{k=1}^n \varepsilon_{ik} = l\right) \\
&\quad = \max_{i=1,\dots,d} \sum_{l=\lceil np_i - \sqrt{n\log n}\rceil}^{\lfloor np_i + \sqrt{n\log n}\rfloor} \mathbb{P}\left(\left|\frac{1-p_i}{p_i}T_{ii} - \frac{1-p_i}{np_i}\sum_{k=1}^l \frac{Y_{ik}^2}{p_i}\right| > u\right)\mathbb{P}\left(\sum_{k=1}^n \varepsilon_{ik} = l\right),
\end{aligned}
$$

where the last identity holds true because $Y_{i1,d,n}, \dots, Y_{in,d,n}$ are i.i.d. and jointly independent of $\varepsilon_{d,n}$. By the elementary inequality

$$
\left|T_{ii} - \frac{1}{n}\sum_{k=1}^l \frac{Y_{ik}^2}{p_i}\right| \le \left|T_{ii} - \frac{1}{n}\sum_{k=1}^{\lceil np_i - \sqrt{n\log n}\rceil} \frac{Y_{ik}^2}{p_i}\right| \vee \left|T_{ii} - \frac{1}{n}\sum_{k=1}^{\lfloor np_i + \sqrt{n\log n}\rfloor} \frac{Y_{ik}^2}{p_i}\right|,
$$

we conclude

$$
\begin{aligned}
\text{(A.23)} \le \max_{i=1,\dots,d} &\mathbb{P}\left(\left|\frac{1-p_i}{p_i}T_{ii} - \frac{1-p_i}{np_i}\sum_{k=1}^{\lceil np_i - \sqrt{n\log n}\rceil} \frac{Y_{ik}^2}{p_i}\right| > u\right) \\
&+ \max_{i=1,\dots,d} \mathbb{P}\left(\left|\frac{1-p_i}{p_i}T_{ii} - \frac{1-p_i}{np_i}\sum_{k=1}^{\lfloor np_i + \sqrt{n\log n}\rfloor} \frac{Y_{ik}^2}{p_i}\right| > u\right)
\end{aligned}
$$

$$
\begin{aligned}
&\le 2 \max_{i=1,\dots,d}\Bigg[\mathbb{P}\Bigg(\bigg|\frac{1-p_i}{p_i}T_{ii} - \frac{1-p_i}{np_i^2}\sum_{k=1}^{\lfloor np_i\rfloor} Y_{ik}^2\bigg| > \frac{u}{2}\Bigg) \\
&\quad + \mathbb{P}\Bigg(\frac{1-p_i}{np_i^2}\sum_{k=1}^{\lceil\sqrt{n\log n}\rceil+1} Y_{ik}^2 > \frac{u}{2}\Bigg)\Bigg] \\
&\le 2 \max_{i=1,\dots,d}\Bigg[\mathbb{P}\Bigg(\bigg|\frac{1-p_i}{p_i}T_{ii} - \frac{1-p_i}{np_i^2}\sum_{k=1}^{\lfloor np_i\rfloor} T_{ii}X_{ik}^2\bigg| > \frac{u}{2}\Bigg) \\
&\quad + \frac{2T_{ii}(1-p_i)}{up_i^2}\bigg(\sqrt{\frac{\log n}{n}} + \frac{2}{n}\bigg)\Bigg].
\end{aligned}
$$

For n sufficiently large, the last expression is bounded by

$$
2 \max_{i=1,\dots,d}\Bigg[\mathbb{P}\Bigg(\bigg|\frac{1}{\lfloor np_i\rfloor}\sum_{k=1}^{\lfloor np_i\rfloor}\big(X_{ik}^2 - 1\big)\bigg| > \frac{up_i}{4(T_{ii}\vee 1)}\Bigg) + \frac{4T_{ii}(1-p_i)}{up_i^2}\sqrt{\frac{\log n}{n}}\Bigg]. \quad \text{(A.24)}
$$

Note that by Section A.3 and Section A.1

$$
\liminf_{d\to\infty}\min_{i=1,\dots,d}\frac{p_{i,d,n}}{T_{ii,d,n}\vee 1} > 0 \quad \text{and} \quad \liminf_{d\to\infty}\min_{i=1,\dots,d}\lfloor np_{i,d,n}\rfloor = \infty.
$$

Hence, by the weak law of large numbers (A.24) converges to zero as $d \to \infty$ which implies $\alpha_{\max} \to 0$. Now, with $\alpha_i = \mathbb{P}(|\hat{S}_{ii} - S_{ii}| > u)$, $i = 1, \dots, d$,

$$
\begin{aligned}
&\mathbb{P}\Bigg(\sum_{i=1}^{d}\mathbb{1}\big\{|\hat{S}_{ii} - S_{ii}| > u\big\} > 2d\sqrt{\alpha_{\max}\vee\sqrt{\frac{1}{d}}}\Bigg) \\
&\le \mathbb{P}\Bigg(\sum_{i=1}^{d}\mathbb{1}\big\{|\hat{S}_{ii} - S_{ii}| > u\big\} - \alpha_i > 2d\sqrt{\alpha_{\max}\vee\sqrt{\frac{1}{d}}} - d\alpha_{\max}\Bigg) \\
&\le \mathbb{P}\Bigg(\sum_{i=1}^{d}\mathbb{1}\big\{|\hat{S}_{ii} - S_{ii}| > u\big\} - \alpha_i > d^{3/4}\Bigg) \\
&\le \exp(-2\sqrt{d}),
\end{aligned}
$$

where we used Hoeffding's inequality in the last line. Therefore,

$$
\frac{1}{d}\sum_{i=1}^{d}\mathbb{1}\big\{|\hat{S}_{ii,d,n} - S_{ii,d,n}| > u\big\} \xrightarrow{\text{a.s.}} 0
$$

as $d \to \infty$. Let $\tilde{S}_{d,n}$ be the diagonal matrix with entries

$$\tilde{S}_{ii,d,n} = \hat{S}_{ii,d,n} \mathbb{1}\big\{ |\hat{S}_{ii,d,n} - S_{ii,d,n}| \le u \big\}.$$

We conclude by Theorem C.9 and Theorem C.8 that almost surely for sufficiently large d

$$\begin{aligned} & d_L\big(\mu^{\hat{T}_{d,n}}, \mu^{\hat{T}_{d,n} - S_{d,n} + \hat{S}_{d,n}}\big) \\ & \le d_L\big(\mu^{\hat{T}_{d,n}}, \mu^{\hat{T}_{d,n} - \tilde{S}_{d,n} + \hat{S}_{d,n}}\big) + d_L\big(\mu^{\hat{T}_{d,n} - \tilde{S}_{d,n} + \hat{S}_{d,n}}, \mu^{\hat{T}_{d,n} - S_{d,n} + \hat{S}_{d,n}}\big) \\ & \le \frac{1}{d} \operatorname{rank}(\hat{S}_{d,n} - \tilde{S}_{d,n}) + \left(\frac{1}{d} \sum_{i=1}^{d} (S_{ii,d,n} - \tilde{S}_{ii,d,n})^2 \right)^{1/3} \\ & \le \frac{1}{d} \sum_{i=1}^{d} \mathbb{1}\big\{ |\hat{S}_{ii,d,n} - S_{ii,d,n}| > u \big\} + u^{2/3} \le 2u^{2/3}. \end{aligned}$$

Since the constant $u > 0$ is chosen arbitrarily, we have

$$d_L\big(\mu^{\hat{T}_{d,n}}, \mu^{\hat{T}_{d,n} - S_{d,n} + \hat{S}_{d,n}}\big) \xrightarrow{\text{a.s.}} 0$$

for $d \to \infty$.

A.8. Step VIII: Reverting the truncation

Reverting finally the truncation Steps I, III, IV yields the claim.

Appendix B: Proof of Proposition 6.1

Define $\hat{X}_{d,n} \in \mathbb{R}^{d \times n}$ by $\hat{X}_{ik,d,n} = X_{ik} \mathbb{1}\{|X_{ik}| < \delta_{d,n}\sqrt{n}\}$. By Lemma 2.2 (truncation lemma) of [29] for $r = 1/2$, given any preassigned decay rate to zero, there exists a sequence $(\delta_{d,n})$, $\delta_{d,n} \to 0$, with lower speed of convergence than that decay rate such that

$$\mathbb{P}(X_{d,n} \ne \hat{X}_{d,n} \text{ infinitely often}) = 0.$$

Let $(\delta_{d,n})$ be a sequence satisfying the truncation lemma with

$$\frac{1}{\sqrt{n}\delta_{d,n}^3} = o(1). \tag{B.1}$$

Therefore,

$$\begin{aligned} \limsup_{d \to \infty} \Bigg| & \left\| \frac{1}{n} A_{d,n} \circ \big((X_{d,n} \circ B_{d,n})(X_{d,n} \circ B_{d,n})^* \big) \right\|_{S_\infty} \\ & - \left\| \frac{1}{n} A_{d,n} \circ \big((\hat{X}_{d,n} \circ B_{d,n})(\hat{X}_{d,n} \circ B_{d,n})^* \big) \right\|_{S_\infty} \Bigg| = 0 \qquad \text{a.s.} \end{aligned}$$

Now let $\tilde{X}_{d,n}$ be the random matrix with entries $\tilde{X}_{ik,d,n} = \hat{X}_{ik,d,n} - \mathbb{E}\hat{X}_{ik,d,n}$. We prove

$$\limsup_{d\to\infty}\bigg|\bigg\|\frac{1}{n}A_{d,n}\circ\big((\tilde{X}_{d,n}\circ B_{d,n})(\tilde{X}_{d,n}\circ B_{d,n})^*\big)\bigg\|_{S_\infty} - \bigg\|\frac{1}{n}A_{d,n}\circ\big((\hat{X}_{d,n}\circ B_{d,n})(\hat{X}_{d,n}\circ B_{d,n})^*\big)\bigg\|_{S_\infty}\bigg| = 0 \qquad \text{a.s.}$$

As $\mathbb{E}X_{11} = 0$, note first that

$$\begin{aligned}|\mathbb{E}\hat{X}_{11,d,n}| &= \big|\mathbb{E}X_{11} - \mathbb{E}X_{11}\mathbb{1}\{|X_{11}| \geq \delta_n\sqrt{n}\}\big| \\ &= \big|\mathbb{E}X_{11}\mathbb{1}\{|X_{11}| \geq \delta_n\sqrt{n}\}\big| \\ &\leq \mathbb{E}X_{11}^4 n^{-3/2}\delta_{d,n}^{-3}.\end{aligned} \tag{B.2}$$

Using the triangle inequality, the bound $\|\cdot\|_{S_\infty} \leq \|\cdot\|_{S_2}$ as well as the inequality

$$\|C\|_{S_\infty} \leq \max_{j=1,\dots,d}\sum_{i=1}^d |C_{ij}| \qquad \text{for symmetric } C \in \mathbb{R}^{d\times d}$$

in (B.3), we conclude

$$\begin{aligned}&\bigg|\bigg\|\frac{1}{n}A_{d,n}\circ\big((\tilde{X}_{d,n}\circ B_{d,n})(\tilde{X}_{d,n}\circ B_{d,n})^*\big)\bigg\|_{S_\infty} \\ &\quad - \bigg\|\frac{1}{n}A_{d,n}\circ\big((\hat{X}_{d,n}\circ B_{d,n})(\hat{X}_{d,n}\circ B_{d,n})^*\big)\bigg\|_{S_\infty}\bigg| \\ &\leq \bigg\|\frac{1}{n}A_{d,n}\circ\big(-(\hat{X}_{d,n}\circ B_{d,n})(\mathbb{E}\hat{X}_{d,n}\circ B_{d,n})^* - (B_{d,n}\circ\mathbb{E}\hat{X}_{d,n}) \\ &\quad \times (\hat{X}_{d,n}\circ B_{d,n})^* + (\mathbb{E}\hat{X}_{d,n}\circ B_{d,n})(\mathbb{E}\hat{X}_{d,n}\circ B_{d,n})^*\big)\bigg\|_{S_\infty} \\ &\leq \frac{2}{n}\sqrt{\sum_{i,j=1}^d A_{ij,d,n}^2\bigg(\sum_{k=1}^n \hat{X}_{ik,d,n}B_{ik,d,n}B_{jk,d,n}\mathbb{E}\hat{X}_{jk}\bigg)^2} \qquad\qquad (\text{B.3}) \\ &\quad + d\max_{i,j}|A_{ij,d,n}|\Big(\max_{ik}B_{ik,d,n}^2\Big)(\mathbb{E}\hat{X}_{11,d,n})^2 \\ &\leq 2\sqrt{\frac{d}{n}}|\mathbb{E}\hat{X}_{11,d,n}|\max_{i,j}|A_{ij,d,n}|\Big(\max_{ik}B_{ik,d,n}^2\Big)\sqrt{d\max_{i=1,\dots,d}\sum_{k=1}^n X_{ik}^2} \qquad\qquad (\text{B.4}) \\ &\quad + d\max_{i,j}|A_{ij,d,n}|\Big(\max_{ik}B_{ik,d,n}^2\Big)(\mathbb{E}\hat{X}_{11,d,n})^2 \\ &\longrightarrow 0 \qquad \text{a.s.,}\end{aligned}$$

where the first summand in inequality (B.4) tends to 0 by (6.1), (B.1), (B.2) and the Marcinkiewicz–Zygmund strong law of large numbers (cf. Lemma B.25 in [1] with $\beta = 1$ and $\alpha = 3/4$). Since the entries of $\tilde{X}_{d,n}$ have all the same finite variance and $\mathbb{E}\tilde{X}_{11,d,n}^2 \to 1$, we may assume for convergence statements about

$$\left\| \frac{1}{n} A_{d,n} \circ \big((\tilde{X}_{d,n} \circ B_{d,n})(\tilde{X}_{d,n} \circ B_{d,n})^* \big) \right\|_{S_\infty}$$

that the entries of $\tilde{X}_{d,n}$ to have unit variance. In order to apply the Lemma of Borel–Cantelli, we need to show that the probabilities

$$\mathbb{P}\left(\left\| \frac{1}{n} A_{d,n} \circ \big((\tilde{X}_{d,n} \circ B_{d,n})(\tilde{X}_{d,n} \circ B_{d,n})^* \big) \right\|_{S_\infty} > z\alpha \right)$$

are summable over $d \in \mathbb{N}$ for any $z > (1 + \sqrt{y})^2$. By Markov's inequality and because of $\|S\|_\infty^{2l} \leq \operatorname{tr}(S^{2l})$ for any symmetric matrix S and $l \in \mathbb{N}$, it is sufficient to show that for any sequence $(l_{d,n})$ of even integers with

$$l_{d,n} / \log n \to \infty \quad \text{and} \quad \delta_{d,n}^{1/6} l_{d,n} / \log n \to 0,$$

we get

$$m_{d,n,l_{d,n}} = \mathbb{E} \operatorname{tr}\left[\mathbb{1}_{E_{d,n}} \left(\frac{1}{n} A_{d,n} \circ \big((\tilde{X}_{d,n} \circ B_{d,n})(\tilde{X}_{d,n} \circ B_{d,n})^* \big) \right)^{l_{d,n}} \right] \leq (\alpha\eta)^{l_{d,n}},$$

where $(1 + \sqrt{y})^2 < \eta < z$ is an absolute constant and $E_{d,n}$ is the event

$$E_{d,n} = \left\{ \max_{i,j} |A_{ij,d,n}| \Big(\max_{i,k} B_{ik,d,n}^2 \Big) < \alpha \right\}.$$

We have by independence of $\tilde{X}_{d,n}$ and $(A_{d,n}, B_{d,n})$,

$$\begin{aligned}
m_{d,n,l_{d,n}} = n^{-l_{d,n}} & \sum_{i_1,\dots,i_{l_{d,n}}=1}^{d} \sum_{k_1,\dots,k_{l_{d,n}}=1}^{n} \mathbb{E}[\mathbb{1}_{E_{d,n}} A_{i_1 i_2} A_{i_2 i_3} \cdots A_{i_{l_{d,n}-1} i_{l_{d,n}}} A_{i_{l_{d,n}} i_1} \\
& \times B_{i_1 k_1} B_{i_2 k_1} \cdots B_{i_{l_{d,n}} k_{l_{d,n}}} B_{i_1 k_{l_{d,n}}}] \\
& \times \mathbb{E}[\tilde{X}_{i_1 k_1} \tilde{X}_{i_2 k_1} \cdots \tilde{X}_{i_{l_{d,n}} k_{l_{d,n}}} \tilde{X}_{i_1 k_{l_{d,n}}}] \\
\leq \alpha^{l_{d,n}} n^{-l_{d,n}} & \sum_{i_1,\dots,i_{l_{d,n}}=1}^{d} \sum_{k_1,\dots,k_{l_{d,n}}=1}^{n} \left| \mathbb{E}[\tilde{X}_{i_1 k_1} \tilde{X}_{i_2 k_1} \cdots \tilde{X}_{i_{l_{d,n}} k_{l_{d,n}}} \tilde{X}_{i_1 k_{l_{d,n}}}] \right| \\
\leq \alpha^{l_{d,n}} \eta^{l_{d,n}}, &
\end{aligned}$$

for d sufficiently large in which case the inequality

$$n^{-l_{d,n}} \sum_{i_1,\dots,i_{l_{d,n}}=1}^{d} \sum_{k_1,\dots,k_{l_{d,n}}=1}^{n} \left|\mathbb{E}[\tilde{X}_{i_1k_1}\tilde{X}_{i_2k_1}\cdots\tilde{X}_{i_{l_{d,n}}k_{l_{d,n}}}\tilde{X}_{i_1k_{l_{d,n}}}]\right| \le \eta^{l_{d,n}}$$

has been shown in the proof of Theorem 3.1 in [29].

Appendix C: Auxiliary results

Lemma C.1 (Lemma 4 in [6]). *Let $A \in \mathbb{C}^{d\times d}$, $\tau \in \mathbb{C}$ and $r \in \mathbb{R}^d$ such that A and $A + \tau rr^*$ are invertable. Then*

$$r^*\big(A+\tau rr^*\big)^{-1} = \frac{1}{1+\tau r^* A^{-1} r} r^* A^{-1}. \tag{C.1}$$

Lemma C.2 (Lemma 2.6 in [23]). *Let $z \in \mathbb{C}^+$, $A, B \in \mathbb{C}^{d\times d}$, B Hermitian, $\tau \in \mathbb{R}$ and $q \in \mathbb{C}^d$. Then*

$$\left|\operatorname{tr}\left[\left((B-zI_{d\times d})^{-1} - \left(B+\tau qq^* - zI_{d\times d}\right)^{-1}\right)A\right]\right| \le \frac{\|A\|_{S_\infty}}{\Im z}. \tag{C.2}$$

Lemma C.3 (Lemma 8 in [6]). *Let $C = A + iB + ivI_{d\times d}$, with $A, B \in \mathbb{R}^{d\times d}$ symmetric and B positive semidefinite, $v > 0$. Then*

$$\left\|C^{-1}\right\|_{S_\infty} \le v^{-1}. \tag{C.3}$$

Lemma C.4. *Let $Z = (Z_1, \dots, Z_d) \in \mathbb{R}^d$ be a centered random vector with components bounded in absolute value by some constant $c > 0$. Then for any $p \ge 1$,*

$$\mathbb{E}\left|\|Z\|_2^2 - \mathbb{E}\|Z\|_2^2\right|^p \le C^p p^{p/2} d^{p/2}, \tag{C.4}$$

$$\mathbb{E}\|Z\|_2^{2p} \le C^p p^{p/2} d^p, \tag{C.5}$$

where the constant $C > 0$ depends on c only.

Proof. The lemma is an easy consequence of Lemma 5.9 of [25] together with the Definition 5.7 of the sub-Gaussian norm of [25], since

$$\begin{aligned}\left\|\frac{1}{d}\left(\|Z\|_2^2 - \mathbb{E}\|Z\|_2^2\right)\right\|_{\psi_2}^2 &\le \frac{\Delta}{d^2}\sum_{i=1}^{d}\|Z_i^2 - \mathbb{E}Z_i^2\|_{\psi_2}^2 \\ &\le \frac{8\Delta}{d}c^4,\end{aligned}$$

where Δ corresponds to the absolute constant of Lemma 5.9 of [25], and

$$\begin{aligned}\left\|\frac{1}{d}\|Z\|_2^2\right\|_{\psi_2}^2 &= \left\|\frac{1}{d}\mathbb{E}\|Z\|_2^2 + \frac{1}{d}\big(\|Z\|_2^2 - \mathbb{E}\|Z\|_2^2\big)\right\|_{\psi_2}^2 \\ &\le 2\left(\left\|\frac{1}{d}\mathbb{E}\|Z\|_2^2\right\|_{\psi_2}^2 + \left\|\frac{1}{d}\big(\|Z\|_2^2 - \mathbb{E}\|Z\|_2^2\big)\right\|_{\psi_2}^2\right) \\ &\le \left(2 + \frac{16\Delta}{d}\right)c^4. \end{aligned}$$

□

Lemma C.5. *Let $d/n < c_1$ and $Z_1, \dots, Z_n \in \mathbb{R}^d$ be a sample of i.i.d. random vectors with centered and independent components of variance* 1 *and bounded in absolute value by some constant $c_2 > 0$. Denote the largest eigenvalue of the matrix $n^{-1}\sum_k Z_k Z_k^*$ by λ_1. Then for any $p \ge 1$,*

$$\mathbb{E}\lambda_1^p \le C,$$

where C depends on c_1, c_2 and p only.

Proof. Since

$$\frac{1}{n}\sum_{k=1}^n Z_k Z_k^* = \frac{1}{n} Z Z^*,$$

where the kth column of the matrix $Z \in \mathbb{R}^{d\times n}$ is given by Z_k, $\lambda_1 = s_1^2$ with s_1 the largest singular value of $n^{-1/2}Z$. Dividing the right-hand side of inequality (5.22) of [25] by $\sqrt{n}$ yields

$$s_1 \le \sqrt{c_1} + \Delta_1 + \frac{t}{\sqrt{n}}$$

with probability at least $1 - 2\exp(-\Delta_2 t^2)$ for some constant $\Delta_1, \Delta_2 > 0$ depending on c_2 only. Therefore,

$$\begin{aligned}\mathbb{E}\lambda_1^p &= \mathbb{E}s_1^{2p} \\ &= \int_0^\infty x^{2p}\mathbb{P}(s_1 > x)\,\mathrm{d}x \\ &\le (\sqrt{c_1} + \Delta_1)^{2p} + 2\int_{\sqrt{c_1}+\Delta_1}^\infty x^{2p}\exp\big(-\Delta_2 n\big(x - (\sqrt{c_1} + \Delta_1)\big)^2\big)\,\mathrm{d}x \\ &\le (\sqrt{c_1} + \Delta_1)^{2p} + 2\int_0^\infty (x + \sqrt{c_1} + \Delta_1)^{2p}\exp\big(-\Delta_2 n x^2\big)\,\mathrm{d}x \\ &\le C,\end{aligned}$$

where C can be chosen independently of n. □

Lemma C.6. *Let $U_1, \dots, U_d$ i.i.d. random $\mathbb{C}$-valued random variables with $\mathbb{E}U_i = 0$, $\mathbb{E}|U_i|^2 = 1$, $|U_i| \le C$ for some constant $C > 0$ and $A \in \mathbb{C}^{d\times d}$. Denote $U = (U_1, \dots, U_d)^*$. Then*

$$\mathbb{E}\big|U^*AU - \operatorname{tr} A\big|^6 \le c\|A\|_{S_\infty}^6 d^3 C^{12}$$

with a constant $c > 0$ which does not depend on d, A and the distribution of U_i.

Proof. The proof follows the lines of Lemma 3.1 in [23] by replacing the logarithmic bound on the entries of U with C. □

Lemma C.7. *For $d \in \mathbb{N}$ and $n = n_d \in \mathbb{N}$ with $\limsup_d d/n \le c_1 < \infty$ let $X_{1,d}, \dots, X_{n,d}$ be i.i.d. d-dimensional, centered random vectors with variance 1 such that*

$$\limsup_{d\to\infty} \max_{i=1,\dots,d} \max_{k=1,\dots,n} |X_{i,k,d}| \le c_2$$

almost surely and $R_d \in \mathbb{R}^{d\times d}$ be a positive definite diagonal matrix with

$$\limsup_{d\to\infty} \max_{i=1,\dots,d} |R_{i,i,d}| \le c_3.$$

Then,

$$\limsup_{d\to\infty} \lambda_{\max}\left(\frac{1}{n}\sum_{k=1}^n R_d^{1/2} X_{k,d} X_{k,d}^* R_d^{1/2}\right) < c \qquad a.s. \tag{C.6}$$

for some constant $c > 0$ depending on c_1, c_2 and c_3 only.

Proof. Since the random variables are uniformly bounded which implies uniform sub-Gaussian tails, Theorem 5.39 of [25] applies. The particular choice $t = \log d$ yields

$$\lambda_{\max}\left(\frac{1}{n}\sum_{k=1}^n R_d^{1/2} X_{k,d} X_{k,d}^* R_d^{1/2}\right) \le \frac{d}{n} + C + \frac{(\log d)^2}{n}$$

with probability at least $1 - 2\exp(-C'(\log d)^2)$ for two positive constants C, C' which depend only on c_1 and c_2. Hence, the claim follows by the lemma of Borel–Cantelli. □

Theorem C.8 (Theorem A.43, [1]). *Let A and B be two $d \times d$ Hermitian matrices. Then,*

$$d_K\big(\mu^A, \mu^B\big) \le \frac{1}{d}\operatorname{rank}(A - B), \tag{C.7}$$

where μ^A and μ^B denote the spectral distributions of A and B, respectively.

Theorem C.9 (Corollary A.41, [1]). *Let A and B be two $d \times d$ Hermitian matrices with spectral distribution μ^A and μ^B. Then,*

$$d_L^3\big(\mu^A, \mu^B\big) \le \frac{1}{d}\operatorname{tr}\big((A - B)(A - B)^*\big). \tag{C.8}$$

Theorem C.10 (Theorem A.38, [1]). *Let $\lambda_1, \dots, \lambda_d$ and $\delta_1, \dots, \delta_d$ be two families of real numbers and their empirical distributions be denoted by μ and $\bar{\mu}$. Then, for any $\alpha > 0$, we have*

$$d_L^{\alpha+1}(\mu, \bar{\mu}) \le \min_{\pi} \frac{1}{d} \sum_{k=1}^{d} |\lambda_k - \delta_{\pi(k)}|^{\alpha}, \tag{C.9}$$

where the minimum is running over all permutations π on $\{1, \dots, d\}$.

The next lemma and its proof are essentially taken from [11], Lemma 34. Since the necessary dependence of (in his notation) δ on y is neither mentioned in his statement nor its proof, we include a proof for completeness.

Lemma C.11. *Let μ and ν be two probability measures on the real line and m_μ and m_ν their Stieltjes transforms. Then for any $v > 0$ we have*

$$d_L(\mu, \nu) \le 2\sqrt{\frac{v}{\pi}} + \frac{1}{2\pi} \int \bigl|\Im\bigl(m_\mu(u + iv)\bigr) - \Im\bigl(m_\nu(u + iv)\bigr)\bigr| \,\mathrm{d}u.$$

Proof. Let C_v denote the Cauchy distribution with scale parameter $v > 0$. Recall that its Lebesgue density f_v is given by

$$f_v(x) = \frac{1}{\pi} \frac{v}{v^2 + x^2}, \qquad x \in \mathbb{R}.$$

By the triangle inequality,

$$d_L(\mu, \nu) \le d_L(\mu, \mu \star C_v) + d_L(\mu \star C_v, \nu \star C_v) + d_L(\nu, \nu \star C_v). \tag{C.10}$$

Now observe that for $\eta = \mu, \nu$ and any $z = u + iv \in \mathbb{C}^+$,

$$-\frac{1}{\pi}\Im\bigl(m_\eta(u + iv)\bigr) = \int \frac{1}{\pi} \frac{v}{(u - \lambda)^2 + v^2} \,\mathrm{d}\eta(\lambda) = f_{\eta \star C_v}(u),$$

where $f_{\eta\star C_v}$ is the Lebesgue density of the convolution $\eta \star C_v$. Therefore,

$$\begin{aligned} d_L(\mu \star C_v, \nu \star C_v) &\le d_K(\mu \star C_v, \nu \star C_v) \\ &\le \frac{1}{2} \int \bigl| f_{\mu\star C_v}(u) - f_{\nu\star C_v}(u)\bigr| \,\mathrm{d}u \\ &= \frac{1}{2\pi} \int \bigl|\Im\bigl(m_\mu(u + iv)\bigr) - \Im\bigl(m_\nu(u + iv)\bigr)\bigr| \,\mathrm{d}u. \end{aligned} \tag{C.11}$$

As concerns $d_L(\eta, \eta \star C_v)$, let $X \sim \eta$ and $Z \sim C_1$ be two independent random variables on a common probability space, whence $X + vZ \sim \eta \star C_v$ for any $v > 0$. Using the elementary tail inequalities

$$\mathbb{P}(Z < -t) = \mathbb{P}(Z > t) \le \int_t^\infty \frac{1}{\pi t^2} \,\mathrm{d}t = \frac{1}{\pi t} \qquad \text{for any } t > 0,$$

we obtain for any $\delta > 0$ and $x \in \mathbb{R}$,

$$\mathbb{P}(X \le x - \delta) \le \mathbb{P}(X + vZ \le x) + \mathbb{P}\left(Z > \frac{\delta}{v}\right) \le \mathbb{P}(X + vZ \le x) + \frac{1}{\pi}\frac{v}{\delta}.$$

That is,

$$\mathbb{P}(X \le x - \delta) - \delta \le \mathbb{P}(X + vZ \le x) \tag{C.12}$$

whenever $\delta \ge \sqrt{v/\pi}$, in which case we also have

$$\mathbb{P}(X + vZ \le x) \le \mathbb{P}(X \le x + \delta) + \mathbb{P}\left(Z < -\frac{\delta}{v}\right) \le \mathbb{P}(X \le x + \delta) + \delta. \tag{C.13}$$

(C.12) and (C.13) imply

$$d_L(\eta, \eta \star C_v) \le \sqrt{\frac{v}{\pi}}, \qquad \eta = \mu, \nu. \tag{C.14}$$

Plugging (C.14) and (C.11) into (C.10) yields the claim. □

Lemma C.12. *Let μ, ν be two probability measures on the real line and m_μ, m_ν the corresponding Stieltjes transforms. Then for any $z \in \mathbb{C}^+$,*

$$\left|m_\mu(z) - m_\nu(z)\right| \le 2\frac{d_{\mathrm{BL}}(\mu, \nu)}{(\Im z)^2 \wedge \Im z}. \tag{C.15}$$

Proof. Note that

$$\left|\frac{1}{\lambda - z} - \frac{1}{\lambda' - z}\right| = \frac{|\lambda - \lambda'|}{|(\lambda - z)(\lambda' - z)|} \le \frac{|\lambda - \lambda'|}{(\Im z)^2},$$

that is,

$$\lambda \mapsto \Re\left(\frac{(\Im z)^2 \wedge \Im z}{\lambda - z}\right) \quad \text{and} \quad \lambda \mapsto \Im\left(\frac{(\Im z)^2 \wedge \Im z}{\lambda - z}\right)$$

are bounded by 1 in absolute value and 1-Lipschitz. This proves (C.15). □

Lemma C.13. *Let $(\mu_n)_{n\in\mathbb{N}}$ and $(\nu_n)_{n\in N}$ be two sequences of probability measures on the Borel σ-algebra on $\mathbb{R}$. Assume that $(\mu_n)_{n\in\mathbb{N}}$ is tight. Then*

$$d_L(\mu_n, \nu_n) \to 0 \quad \Leftrightarrow \quad d_{\mathrm{BL}}(\mu_n, \nu_n) \to 0. \tag{C.16}$$

Moreover, tightness of $(\mu_n)_{n\in\mathbb{N}}$ and (C.16) imply weak convergence $\mu_n - \nu_n \Rightarrow 0$ on the space of finite signed measures on $\mathbb{R}$.

Proof. As concerns the equivalence relation, we need only to verify that

$$d_L(\mu_n, \nu_n) \to 0 \quad \Rightarrow \quad d_{\mathrm{BL}}(\mu_n, \nu_n) \to 0, \tag{C.17}$$

because $d_L^2 \leq d_{\mathrm{BL}}$ (see, e.g. [9]). Assume that $d_L(\mu_n, \nu_n) \to 0$. Tightness of $(\mu_n)_n$ implies that any subsequence $(\mu_{n_k})_k$ possesses a subsubsequence $(\mu_{n_{k_l}})_l$ which converges weakly to a limiting probability measure μ, say. Consequently, as both, d_{BL} and d_L metrize weak convergence on the space of probability measures on $\mathbb{R}$,

$$d_L(\mu_{n_{k_l}}, \mu) \to 0 \quad \Leftrightarrow \quad d_{\mathrm{BL}}(\mu_{n_{k_l}}, \mu) \to 0. \tag{C.18}$$

By the triangle inequality,

$$d_L(\nu_{n_{k_l}}, \mu) \leq d_L(\mu_{n_{k_l}}, \mu) + d_L(\mu_{n_{k_l}}, \nu_{n_{k_l}}) \to 0,$$

which in turn is equivalent to $d_{\mathrm{BL}}(\nu_{n_{k_l}}, \mu) \to 0$. Again by the triangle inequality, $d_{\mathrm{BL}}(\mu_{n_{kl}}, \nu_{n_{k_l}}) \to 0$. This proves (C.17) and therefore the equivalence relation (C.16).

As concerns the second statement, it is sufficient to show that any subsequence $(n_k)_k$ possesses a subsubsequence $(n_{k_l})_l$ with $\mu_{n_{k_l}} - \nu_{n_{k_l}} \Rightarrow 0$. But this follows immediately from the above arguments, because for any subsequence $(n_k)_k$, there exist a subsubsequence $(n_{k_l})_l$ and a measure μ such that both, $\mu_{n_{k_l}} \Rightarrow \mu$ and $\nu_{n_{k_l}} \Rightarrow \mu$, hence $\mu_{n_{k_l}} - \nu_{n_{k_l}} \Rightarrow 0$. □

Acknowledgements

This work has been supported by the DFG Research Unit 1735, RO 3766/3-1 and DE 502/26-2.

We are grateful to Jack Silverstein for kindly answering numerous questions and for pointing us to the reference [6].

References

[1] Bai, Z. and Silverstein, J.W. (2010). *Spectral Analysis of Large Dimensional Random Matrices*, 2nd ed. *Springer Series in Statistics*. New York: Springer. MR2567175

[2] Bai, Z.D. and Silverstein, J.W. (1998). No eigenvalues outside the support of the limiting spectral distribution of large-dimensional sample covariance matrices. *Ann. Probab.* **26** 316–345. MR1617051

[3] Bai, Z.D., Silverstein, J.W. and Yin, Y.Q. (1988). A note on the largest eigenvalue of a large-dimensional sample covariance matrix. *J. Multivariate Anal.* **26** 166–168. MR0963829

[4] Bai, Z.D. and Yin, Y.Q. (1993). Limit of the smallest eigenvalue of a large-dimensional sample covariance matrix. *Ann. Probab.* **21** 1275–1294. MR1235416

[5] Butucea, C. and Zgheib, R. (2016). Adaptive test for large covariance matrices with missing observations. Preprint. Available at arXiv:1602.04310.

[6] Couillet, R., Debbah, M. and Silverstein, J.W. (2011). A deterministic equivalent for the analysis of correlated MIMO multiple access channels. *IEEE Trans. Inform. Theory* **57** 3493–3514. MR2817033

[7] El Karoui, N. (2008). Spectrum estimation for large dimensional covariance matrices using random matrix theory. *Ann. Statist.* **36** 2757–2790. MR2485012

[8] Geman, S. (1980). A limit theorem for the norm of random matrices. *Ann. Probab.* **8** 252–261. MR0566592

[9] Huber, P. (1974). Some mathematical problems arising in robust statistics. In *Proceedings of the International Congress of Mathematicians*. Vancouver.

[10] Jin, B., Wang, C., Bai, Z.D., Nair, K.K. and Harding, M. (2014). Limiting spectral distribution of a symmetrized auto-cross covariance matrix. *Ann. Appl. Probab.* **24** 1199–1225. MR3199984
[11] Krishnapur, M. (2012). Random matrix theory. *Lecture notes*.
[12] Ledoit, O. and Wolf, M. (2012). Nonlinear shrinkage estimation of large-dimensional covariance matrices. *Ann. Statist.* **40** 1024–1060. MR2985942
[13] Li, C.-K. and Mathias, R. (1999). The Lidskii–Mirsky–Wielandt theorem—additive and multiplicative versions. *Numer. Math.* **81** 377–413. MR1668087
[14] Li, Z., Pan, G. and Yao, J. (2015). On singular value distribution of large-dimensional autocovariance matrices. *J. Multivariate Anal.* **137** 119–140. MR3332802
[15] Liu, H., Aue, A. and Paul, D. (2015). On the Marčenko–Pastur law for linear time series. *Ann. Statist.* **43** 675–712. MR3319140
[16] Lounici, K. (2014). High-dimensional covariance matrix estimation with missing observations. *Bernoulli* **20** 1029–1058. MR3217437
[17] Marčenko, V. and Pastur, L. (1967). Distribution of eigenvalues for some sets of random matrices. *Math. USSR—Sb.* **81** 377–413.
[18] Nishizawa, A.J. and Inoue, K.T. (2013). Reconstruction of missing data in the sky using iterative harmonic expansion. Preprint. Available at arXiv:1305.0116.
[19] Petrov, V.V. (1995). *Limit Theorems of Probability Theory*. *Oxford Studies in Probability* **4**. New York: The Clarendon Press, Oxford Univ. Press. MR1353441
[20] Sherwood, S.C. (2001). Climate signals from station arrays with missing data, and an application to winds. *J. Geophys. Res.* **105** 29489–29500.
[21] Shohat, J.A. and Tamarkin, J.D. (1943). *The Problem of Moments*. *American Mathematical Society Mathematical Surveys* **1** New York: Amer. Math. Soc. MR0008438
[22] Silverstein, J.W. (1995). Strong convergence of the empirical distribution of eigenvalues of large-dimensional random matrices. *J. Multivariate Anal.* **55** 331–339. MR1370408
[23] Silverstein, J.W. and Bai, Z.D. (1995). On the empirical distribution of eigenvalues of a class of large-dimensional random matrices. *J. Multivariate Anal.* **54** 175–192. MR1345534
[24] Troyanskaya, O., Cantor, M., Sherlock, G., Brown, P., Hastie, T., Tibshirani, R., Botstein, D. and Altman, R.B. (2001). Missing value estimation methods for DNA microarrays. *Bioinformatics* **17** 520–525.
[25] Vershynin, R. (2012). Introduction to the non-asymptotic analysis of random matrices. In *Compressed Sensing* 210–268. Cambridge: Cambridge Univ. Press. MR2963170
[26] Wang, C., Jin, B., Bai, Z.D., Nair, K.K. and Harding, M. (2015). Strong limit of the extreme eigenvalues of a symmetrized auto-cross covariance matrix. *Ann. Appl. Probab.* **25** 3624–3683. MR3404646
[27] Wang, L., Aue, A. and Paul, D. (2015). Spectral analysis of linear time series in moderately high dimensions. *Bernoulli*. To appear.
[28] Wang, Q. and Yao, J. (2015). Moment approach for singular values distribution of a large auto-covariance matrix. *Ann. Inst. Henri Poincaré Probab. Stat.* To appear.
[29] Yin, Y.Q., Bai, Z.D. and Krishnaiah, P.R. (1988). On the limit of the largest eigenvalue of the large-dimensional sample covariance matrix. *Probab. Theory Related Fields* **78** 509–521. MR0950344

Received October 2015