Spectral convergence for a general class of random matrices

Francisco Rubio, Xavier Mestre

HAL Id: hal-00725102
https://hal.archives-ouvertes.fr/hal-00725102

Submitted on 24 Aug 2012

Accepted Manuscript

Spectral convergence for a general class of random matrices

Francisco Rubio, Xavier Mestre

Please cite this article as: Rubio, F., Mestre, X., Spectral convergence for a general class of random matrices. *Statistics and Probability Letters* (2011), doi:10.1016/j.spl.2011.01.004

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Spectral Convergence for a General Class of Random Matrices

Francisco Rubio, Xavier Mestre

Centre Tecnològic de Telecomunicacions de Catalunya
Av. Carl Friedrich Gauss, 7, 08860 Castelldefels (Barcelona), Spain

Abstract

Let $\mathbf{X}$ be an $M \times N$ complex random matrix with i.i.d. entries having mean zero and variance $1/N$ and consider the class of matrices of the type $\mathbf{B} = \mathbf{A} + \mathbf{R}^{1/2}\mathbf{X}\mathbf{T}\mathbf{X}^H\mathbf{R}^{1/2}$, where $\mathbf{A}$, $\mathbf{R}$ and $\mathbf{T}$ are Hermitian nonnegative definite matrices, such that $\mathbf{R}$ and $\mathbf{T}$ have bounded spectral norm with $\mathbf{T}$ being diagonal, and $\mathbf{R}^{1/2}$ is the nonnegative definite square-root of $\mathbf{R}$. Under some assumptions on the moments of the entries of $\mathbf{X}$, it is proved in this paper that, for any matrix $\boldsymbol{\Theta}$ with bounded trace norm and for each complex z outside the positive real line, $\mathrm{Tr}\left[\boldsymbol{\Theta}\left(\mathbf{B} - z\mathbf{I}_M\right)^{-1}\right] - \delta_M\left(z\right) \to 0$ almost surely as $M, N \to \infty$ at the same rate, where $\delta_M\left(z\right)$ is deterministic and solely depends on $\boldsymbol{\Theta}, \mathbf{A}, \mathbf{R}$ and $\mathbf{T}$. The previous result can be particularized to the study of the limiting behavior of the Stieltjes transform as well as the eigenvectors of the random matrix model $\mathbf{B}$. The study is motivated by applications in the field of statistical signal processing.

Key words: random matrix theory, Stieltjes transform, multivariate statistics, sample covariance matrix, separable covariance model

Introduction

Consider an $M \times N$ random matrix $\mathbf{X}$ such that the entries of $N\mathbf{X}$ are i.i.d. standardized complex random variables with finite $8 + \varepsilon$ moment. Furthermore, consider an $M \times M$ Hermitian nonnegative definite matrix $\mathbf{R}$ and its nonnegative definite square-root $\mathbf{R}^{1/2}$. Then, the matrix $\mathbf{R}^{1/2}\mathbf{X}\mathbf{X}^H\mathbf{R}^{1/2}$ can be viewed as a sample covariance matrix constructed using the N columns of the data matrix $\mathbf{R}^{1/2}\mathbf{X}$, namely having population covariance matrix $\mathbf{R}$. Moreover, consider also an $N \times N$ diagonal matrix $\mathbf{T}$ with real nonnegative entries. The matrix $\mathbf{R}^{1/2}\mathbf{X}\mathbf{T}\mathbf{X}^H\mathbf{R}^{1/2}$ can be interpreted as a sample covariance matrix obtained by weighting the previous multivariate samples with the entries of $\mathbf{T}$.

In this paper, we are interested in the asymptotic behaviour of certain spectral functions of the random matrix model $\mathbf{B} = \mathbf{A} + \mathbf{R}^{1/2}\mathbf{X}\mathbf{T}\mathbf{X}^H\mathbf{R}^{1/2}$, where $\mathbf{A}$ is an $N \times N$ Hermitian nonnegative definite matrix, as the dimensions M and N grow large at the same rate. Consider the empirical distribution function of the eigenvalues of $\mathbf{B}$, i.e.,

$$F_{\mathbf{B}}^M (\lambda) = \frac{1}{M} \sum_{m=1}^M \mathcal{I}_{(\lambda_m(\mathbf{B}) \leq \lambda)}, \tag{1}$$

where $\lambda_m (\mathbf{B})$ stands for the mth eigenvalue of $\mathbf{B}$ and $\lambda_1 (\mathbf{B}) \geq \ldots \geq \lambda_M (\mathbf{B})$. From the connection between vague convergence of distributions and pointwise convergence of Stieltjes transforms (see, e.g., [3] and [7, Proposition 2.2]), almost sure convergence of the (random) distribution function $F_{\mathbf{B}}^M$ can be established by showing convergence of the associated Stieltjes transform, defined for each $z \in \mathbb{C}^+ = \{z \in \mathbb{C} : \mathrm{Im}\{z\} > 0\}$ as

$$m_F (z) = \int_{\mathbb{R}} \frac{dF_{\mathbf{B}}^M}{\lambda - z} = \frac{1}{M} \sum_{m=1}^M \frac{1}{\lambda_m (\mathbf{B}) - z} = \frac{1}{M} \mathrm{Tr} \left[(\mathbf{B} - z\mathbf{I}_M)^{-1} \right]. \tag{2}$$

Under the condition that M/N has a finite non-zero limit, convergence with probability one of the empirical spectral distribution (ESD), i.e., the empirical distribution of the eigenvalues, of some special cases of $\mathbf{B}$ towards a limit nonrandom distribution has been established in the random matrix theory literature under the assumption of the matrices $\mathbf{R}$, $\mathbf{T}$ and $\mathbf{A}$ having, in each case, an ESD which converges almost surely to a probability distribution (possibly defective in the case of $\mathbf{A}$). More specifically, using the fact that the limit of $F_{\mathbf{B}}^M$ is uniquely characterized by that of $m_F (z)$, vague convergence in the cases of $\mathbf{A} = \mathbf{0}_{M \times M}$ and $\mathbf{T} = \mathbf{I}_N$; $\mathbf{R} = \mathbf{I}_M$; and $\mathbf{A} = \mathbf{0}_{M \times M}$ is provided in [13], [14], and [12,9], respectively, by proving tightness of the sequence $\left\{ F_{\mathbf{B}}^M \right\}$ (the convergence is weakly to a proper probability measure if $\mathbf{A} = \mathbf{0}_{M \times M}$). The reader is also referred to [5] for similar results and other random matrix models.

In order to extend the previous spectral convergence results to the asymptotic behavior of the eigensubspaces of $\mathbf{B}$, we can consider the following empirical distribution function, namely,

$$H_{\mathbf{B}}^M (\lambda) = \sum_{m=1}^M |a_m|^2 \mathcal{I}_{(\lambda_m(\mathbf{B}) \leq \lambda)}, \tag{3}$$

where a_m is the mth entry of the vector $\mathbf{a} = \mathbf{U}^H v$, with $\mathbf{U} \in \mathbb{C}^{M \times M}$ being the matrix of orthonormal eigenvectors of $\mathbf{B}$, and v an arbitrary nonrandom vector in the unit sphere $\left\{ v \in \mathbb{C}^N : \|v\| = 1 \right\}$. Clearly, $H_{\mathbf{B}}^M$ is a random probability

2

distribution function with Stieltjes transform given by

$$m_H(z) = \int_{\mathbb{R}} \frac{dH_{\mathbf{B}}^M(\lambda)}{\lambda - z} = \upsilon^H \left(\mathbf{B} - z\mathbf{I}_M\right)^{-1} \upsilon. \tag{4}$$

In particular, spectral functions of the form of (3) or, equivalently, (4) were considered in [1] (see also references therein) to study the limiting eigenvectors of sample covariance matrices. Note further that, if $a_m = \frac{1}{M}$, $m = 1, \ldots, M$, then $H_{\mathbf{B}}^M$ and $F_{\mathbf{B}}^M$ coincide.

Motivated by practical applications, for instance in the fields of statistical signal processing, wireless communications and quantitative finance, where estimates of certain functions of the eigenvalues and eigenvectors of the random matrix model $\mathbf{B}$ are very often of interest, in this paper the convergence of (2) and (4) is extended to the convergence of more general spectral functions of $\mathbf{B}$. Moreover, for the purpose of practical applicability in realistic settings, the assumption of $\mathbf{R}, \mathbf{T}$ and $\mathbf{A}$ having convergent ESD is dropped in the following.

1 Main result

The remainder of the paper is devoted to the proof of the following theorem.

Theorem 1 *Assume the following:*

(a) $\mathbf{X}$ *is an* $M \times N$ *random matrix such that the entries of* $\sqrt{N}\mathbf{X}$ *are i.i.d. complex random variables with mean zero, variance one, and finite* $8 + \varepsilon$ *moment, for some* $\varepsilon > 0$.

(c) $\mathbf{A}$ *and* $\mathbf{R}$ *are* $M \times M$ *Hermitian nonnegative definite matrices, with the spectral norm (denoted by* $\|\cdot\|$*) of* $\mathbf{R}$ *being bounded uniformly in* M*, and* $\mathbf{T}$ *is an* $N \times N$ *diagonal matrix with real nonnegative entries uniformly bounded in* N.

(d) $\mathbf{B} = \mathbf{A} + \mathbf{R}^{1/2}\mathbf{X}\mathbf{T}\mathbf{X}^H\mathbf{R}^{1/2}$*, where* $\mathbf{R}^{1/2}$ *is the nonnegative definite square-root of* $\mathbf{R}$.

(e) $\boldsymbol{\Theta}$ *is an arbitrary nonrandom* $M \times M$ *matrix, whose trace norm (i.e.,* $\mathrm{Tr}\left[\left(\boldsymbol{\Theta}^H\boldsymbol{\Theta}\right)^{1/2}\right] \equiv \|\boldsymbol{\Theta}\|_{\mathrm{tr}}$*) is bounded uniformly in* M.

Then, with probability one, for each $z \in \mathbb{C} - \mathbb{R}^+$*, as* $N = N(M) \to \infty$ *such that* $0 < \liminf c_M < \limsup c_M < \infty$*, with* $c_M = M/N$*,*

$$\mathrm{Tr}\left[\boldsymbol{\Theta}\left((\mathbf{B} - z\mathbf{I}_M)^{-1} - (\mathbf{A} + x_M(e_M)\mathbf{R} - z\mathbf{I}_M)^{-1}\right)\right] \to 0, \tag{5}$$

3

where $x_M(e_M)$ is defined as

$$x_M(e_M) = \frac{1}{N}\operatorname{Tr}\left[\mathbf{T}\left(\mathbf{I}_N + c_M e_M \mathbf{T}\right)^{-1}\right], \qquad (6)$$

and $e_M = e_M(z)$ is the Stieltjes transform of a certain positive measure on $\mathbb{R}^+$ with total mass $\frac{1}{M}\operatorname{Tr}[\mathbf{R}]$, obtained as the unique solution in $\mathbb{C}^+$ of the equation

$$e_M = \frac{1}{M}\operatorname{Tr}\left[\mathbf{R}\left(\mathbf{A} + x_M(e_M)\mathbf{R} - z\mathbf{I}_M\right)^{-1}\right]. \qquad (7)$$

Corollary 1 *(Limiting Stieltjes transform of ESD's) Let $\mathbf{\Theta} = \frac{1}{M}\mathbf{I}_M$. Then, Theorem 1 states the asymptotic convergence of the Stieltjes transform of the ESD of $\mathbf{B}$ defined in (2).*

Remark 1 *(Sample covariance matrices) If $\mathbf{A} = \mathbf{0}_{M\times M}$ and $\mathbf{T} = \mathbf{I}_N$, then Corollary 1 is equivalent to the convergence result for the Stieltjes transform of the ESD of sample covariance matrices provided in [16,4][13, Theorem 1.1].*

Remark 2 *(Marchenko-Pastur law) If $\mathbf{R} = \mathbf{I}_M$, then the result in Corollary 1 reduces to that obtained in [10][14, Theorem 1.1] under more general conditions.*

Remark 3 *(Separable covariance model) If $\mathbf{A} = \mathbf{0}_{M\times M}$, then Corollary 1 coincides with the result in [12, Theorem 1] (see also [9, Theorem 2]), which represents the special case of [7, Theorem 2.5] corresponding to a separable variance profile.*

Corollary 2 *(Asymptotic convergence of eigenvectors) Let $\mathbf{\Theta} = \upsilon\upsilon^H$, with $\upsilon \in \mathbb{C}^M$ being an arbitrary nonrandom unit vector. Then, Theorem 1 establishes the convergence for the class of Stieltjes transforms defined in (4).*

Remark 4 *The result in Corollary 2 for the cases $\mathbf{R} = \mathbf{I}_M, \mathbf{T} = \mathbf{I}_N$ and $\mathbf{A} = \mathbf{0}_{M\times M}, \mathbf{T} = \mathbf{I}_N$ has been previously reported in [11, Theorem 1 and Theorem 5], and [1, Theorem 1] under more generic assumptions.*

As a final remark, we notice that Theorem 1 holds verbatim if the matrices $\mathbf{A}$, $\mathbf{R}$, $\mathbf{T}$, $\mathbf{X}$ and $\mathbf{\Theta}$ have real-valued entries.

2 Applications to Array Signal Processing

The random matrix model introduced in previous sections is of fundamental interest in applications involving spatio-temporal statistics as well as separable covariance models. Consider for instance a collection of N narrowband signal observations that are obtained by sampling across a linear antenna array with

M sensors, and which can be modelled as $\mathbf{y}_n = s_n \mathbf{s} + \mathbf{n}_n$, $n = 1, \ldots, N$, where $s_n \in \mathbb{C}$ is the signal waveform, $\mathbf{s} \in \mathbb{C}^M$ is the associated signature vector, and $\mathbf{n}_n \in \mathbb{C}^M$ models the contribution of the interference and noise. Conventionally, s_n and $\mathbf{n}_n$ are assumed to be Gaussian, temporarily uncorrelated and mutually independent, with variance and covariance matrix given respectively by σ_s^2 and $\mathbf{R_n}$. Under these assumptions, the observations can be modelled as $\mathbf{y}_n = \mathbf{R}^{1/2} \xi_n$, $\mathbf{R} = \sigma_s^2 \mathbf{s} \mathbf{s}^H + \mathbf{R_n}$ is the covariance matrix of the array samples, and the vectors $\xi_n \in \mathbb{C}^M$ consist of i.i.d. standardized (i.e., with mean zero and variance one) Gaussian random entries. In the array processing literature, the minimum variance distortionless response (MVDR) spatial filter or beamformer is defined in terms of $\mathbf{R}$ as $\mathbf{w}_{\mathsf{MVDR}} = \left(\mathbf{s}^H \mathbf{R}^{-1} \mathbf{s} \right)^{-1} \mathbf{R}^{-1} \mathbf{s}$ (see, e.g., [15, Chapter 6]). In practice, the true covariance matrix $\mathbf{R}$ is unknown, and therefore a sample estimate, henceforth denoted by $\hat{\mathbf{R}}$, must be used instead for implementation purposes. A relevant measure of the filter performance is the output signal-to-interference-plus-noise ratio (SINR), which for a given beamformer $\hat{\mathbf{w}}$ is defined as

$$\mathsf{SINR}\left(\hat{\mathbf{w}}\right) = \left(\frac{\hat{\mathbf{w}}^H \mathbf{R} \hat{\mathbf{w}}}{\sigma_s^2 \left| \hat{\mathbf{w}}^H \mathbf{s} \right|^2} - 1 \right)^{-1}.$$

We note that $\mathsf{SINR}\left(\hat{\mathbf{w}}\right)$ is invariant under scaling and consider the filter $\hat{\mathbf{w}} = \hat{\mathbf{R}}^{-1} \mathbf{s}$, namely defined in terms of a generic covariance sample estimator given by $\hat{\mathbf{R}} = \alpha \mathbf{Y} \mathbf{W} \mathbf{Y}^H + \beta \mathbf{R}_0$, where $\mathbf{Y} = \begin{bmatrix} \mathbf{y}_1 \cdots \mathbf{y}_N \end{bmatrix} \in \mathbb{C}^{M \times N}$, $\mathbf{W}$ is an $N \times N$ diagonal nonnegative definite matrix, $\mathbf{R}_0$ is an $M \times M$ Hermitian nonnegative matrix, and $\alpha, \beta \in \mathbb{R}$. In particular, if $\mathbf{W} = \mathbf{I}_N$, $\alpha = 1$ and $\mathbf{R}_0 = \mathbf{0}_{M \times M}$, then $\hat{\mathbf{R}}$ is the sample covariance matrix or maximum likelihood estimator of $\mathbf{R}$; moreover, if $\mathbf{W}$ is a data windowing diagonal matrix with nth entry given by $w_n = \exp\left(N - n\right)$, $\alpha = 1$ and $\mathbf{R}_0 = \mathbf{I}_M$, then $\hat{\mathbf{w}}$ implements the recursive least-squares beamformer with exponential weighting and diagonal loading factor β (see, e.g., [15, Chapter 7]); finally, if $\mathbf{W} = \mathbf{I}_N$, $\hat{\mathbf{R}}$ has the form of a James-Stein shrinkage covariance estimator, where $\mathbf{R}_0$ is the shrinkage target or prior information about $\mathbf{R}$ and α, β are the shrinkage intensity parameters. Interestingly enough, a unified approach to the analysis of the SINR performance of the previous MVDR sample beamformers can be provided by applying Theorem 1 in order to find a deterministic asymptotic equivalent of $\mathsf{SINR}\left(\hat{\mathbf{w}}\right)$.

3 Preliminary results

In this section, we introduce some preparatory lemmas and auxiliary technical results that will be needed in the proof of Theorem 1. The first one has a

5

straightforward proof based on Markov's and Minkowski's inequalities (see, e.g., proof of Lemma 4.1 in [6] for details).

Lemma 1 *Let $\left\{ y_n^{(N)}, 1 \leq n \leq N \right\}$ denote a collection of (possibly dependent) random variables such that*

$$\max_{1 \leq n \leq N} \mathbb{E}\left[\left| y_n^{(N)} \right|^p \right] \leq \frac{K_p}{N^{1+\delta}},$$

for some constants $p \geq 1$, $\delta > 0$ and K_p (in the sequel, constants K_p may take different values at each appearance) depending on p but not on N. Then, almost surely as $N \to \infty$,

$$\frac{1}{N} \sum_{n=1}^{N} \left| y_n^{(N)} \right| \to 0.$$

Lemma 2 *([2, Lemma 2.7]) Let $\xi \in \mathbb{C}^M$ denote a random vector with i.i.d. entries having mean zero and variance one, and $\mathbf{C} \in \mathbb{C}^{M \times M}$ an arbitrary nonrandom matrix. Then, for any $p \geq 2$,*

$$\mathbb{E}\left[\left| \xi^H \mathbf{C} \xi - \mathrm{Tr}\left[\mathbf{C}\right] \right|^p \right] \leq K_p \left\{ \left(\mathbb{E}\left[|\zeta|^4 \right] \mathrm{Tr}\left[\mathbf{C}\mathbf{C}^H \right] \right)^{p/2} + \mathbb{E}\left[|\zeta|^{2p} \right] \mathrm{Tr}\left[\left(\mathbf{C}\mathbf{C}^H\right)^{p/2} \right] \right\},$$

where ζ denotes a particular entry of ξ and the constant K_p does not depend on M, the entries of $\mathbf{C}$, nor the distribution of ζ.

Lemma 3 *Let ξ and $\mathbf{C}$ be defined as in Lemma 2. Then, for any $p \geq 2$,*

$$\mathbb{E}\left[\left| \xi^H \mathbf{C} \xi \right|^p \right] \leq \|\mathbf{C}\|_{\mathrm{tr}}^p \left(K_{1,p} + K_{2,p} \mathbb{E}\left[|\zeta|^{2p} \right] \right).$$

PROOF. Let μ and ν be two complex random variables with finite pth-order moment. Using Jensen's inequality, note that

$$\mathbb{E}\left[|\mu + \nu|^p \right] \leq 2^{p-1} \left\{ \mathbb{E}\left[|\mu|^p \right] + \mathbb{E}\left[|\nu|^p \right] \right\}. \tag{8}$$

Furthermore, consider the following two useful inequalities:

$$\mathrm{Tr}\left[\left(\mathbf{C}\mathbf{C}^H\right)^{p/2} \right] \leq \left(\mathrm{Tr}\left[\mathbf{C}\mathbf{C}^H \right] \right)^{p/2} \leq \left(\mathrm{Tr}\left[\left(\mathbf{C}\mathbf{C}^H\right)^{1/2} \right] \right)^p = \|\mathbf{C}\|_{\mathrm{tr}}^p, \tag{9}$$

and, for $\mathbf{B}_1$ and $\mathbf{B}_2$ two arbitrary square complex matrices [8, Chapter 3]

$$\left| \mathrm{Tr}\left[\mathbf{B}_1 \mathbf{B}_2 \right] \right| \leq \|\mathbf{B}_1 \mathbf{B}_2\|_{\mathrm{tr}} \leq \|\mathbf{B}_1\|_{\mathrm{tr}} \|\mathbf{B}_2\|, \tag{10}$$

which will be used repeatedly in the sequel. Now, using (8) we write

$$\mathbb{E}\left[\left| \xi^H \mathbf{C} \xi \right|^p \right] \leq 2^{p-1} \left\{ \mathbb{E}\left[\left| \xi^H \mathbf{C} \xi - \mathrm{Tr}\left[\mathbf{C}\right] \right|^p \right] + \left| \mathrm{Tr}\left[\mathbf{C}\right] \right|^p \right\}. \tag{11}$$

Moreover, using the first inequality in (9) and Jensen's inequality as $\mathbb{E}^{p/2}\left[|\zeta|^4\right] \leq \mathbb{E}\left[|\zeta|^{2p}\right]$, from Lemma 2 we get

$$\mathbb{E}\left[\left|\xi^H \mathbf{C}\xi - \mathrm{Tr}\left[\mathbf{C}\right]\right|^p\right] \leq K_p \left(\mathrm{Tr}\left[\mathbf{C}\mathbf{C}^H\right]\right)^{p/2} \mathbb{E}\left[|\zeta|^{2p}\right]. \tag{12}$$

Finally, the result of the lemma follows by applying to (11) the inequality (12) along with the first inequality in (10) with $\mathbf{B}_1 = \mathbf{C}$ and $\mathbf{B}_2 = \mathbf{I}_M$.

Lemma 4 *Let* $\mathcal{U} = \left\{\xi_n \in \mathbb{C}^M, 1 \leq n \leq N\right\}$ *denote a collection of i.i.d. random vectors defined as in Lemma 2, and whose entries are assumed to have finite* $4 + \varepsilon$ *moment,* $\varepsilon > 0$. *Furthermore, consider a collection of random matrices* $\left\{\mathbf{C}_{(n)} \in \mathbb{C}^{M \times M}, 1 \leq n \leq N\right\}$ *such that, for each* n, $\mathbf{C}_{(n)}$ *may depend on all the elements of* $\mathcal{U}$ *except for* ξ_n, *and* $\left\|\mathbf{C}_{(n)}\right\|_{\mathrm{tr}}$ *is uniformly bounded for all* M. *Then, almost surely as* $N \to \infty$,

$$\left|\frac{1}{N}\sum_{n=1}^{N} \xi_n^H \mathbf{C}_{(n)} \xi_n - \mathrm{Tr}\left[\mathbf{C}_{(n)}\right]\right| \to 0. \tag{13}$$

PROOF. Write $\eta_n = \xi_n^H \mathbf{C}_{(n)} \xi_n - \mathrm{Tr}\left[\mathbf{C}_{(n)}\right]$ and let $\mathcal{F}_n$ denote the σ-field generated by the random vectors $\{\xi_1, \xi_2, \ldots, \xi_n\}$. Now, notice that $\{\eta_n, 1 \leq n \leq N\}$ forms a martingale difference array with respect to the filtration $\{\mathcal{F}_n, 1 \leq n \leq N\}$. Indeed, since $\mathbb{E}\left[\xi_n \xi_n^H | \mathcal{F}_{n-1}\right] = \mathbf{I}_M$, observe that

$$\mathbb{E}\left[\eta_n | \mathcal{F}_{n-1}\right] = \mathbb{E}\left[\xi_n^H \mathbf{C}_{(n)} \xi_n | \mathcal{F}_{n-1}\right] - \mathbb{E}\left[\mathrm{Tr}\left[\mathbf{C}_{(n)}\right] | \mathcal{F}_{n-1}\right] = 0.$$

Consequently, Burkholder's inequality (see, e.g., [2, Lemma 2.1]) implies

$$\mathbb{E}\left[\left|\sum_{n=1}^{N} \eta_n\right|^p\right] \leq K_p \left\{\mathbb{E}\left[\left(\sum_{n=1}^{N} \mathbb{E}\left[|\eta_n|^2 | \mathcal{F}_{n-1}\right]\right)^{p/2}\right] + \mathbb{E}\left[\sum_{n=1}^{N} |\eta_n|^p\right]\right\},$$

for any $p \geq 2$. Then, since ξ_n and $\mathbf{C}_{(n)}$ are independent for each n, using (12) and the properties of conditional expectation, it is easy to check that

$$\mathbb{E}\left[|\eta_n|^2 | \mathcal{F}_{n-1}\right] = \mathbb{E}\left[\left|\xi_n^H \mathbf{C}_{(n)} \xi_n - \mathrm{Tr}\left[\mathbf{C}_{(n)}\right]\right|^2 | \mathcal{F}_{n-1}\right] \leq K_p \mathbb{E}\left[|\zeta|^4\right],$$

where we have used the fact that the random variable $\mathrm{Tr}\left[\mathbf{C}_{(n)} \mathbf{C}_{(n)}^H\right] \leq \left\|\mathbf{C}_{(n)}\right\|_{\mathrm{tr}}^2$ is bounded uniformly in M by assumption. Similarly, we also get

$$\mathbb{E}\left[|\eta_n|^p\right] = \mathbb{E}\left[\left|\xi_n^H \mathbf{C}_{(n)} \xi_n - \mathrm{Tr}\left[\mathbf{C}_{(n)}\right]\right|^p\right] \leq K_p \mathbb{E}\left[|\zeta|^{2p}\right].$$

7

Hence, we can finally write

$$\mathbb{E}\left[\left|\frac{1}{N}\sum_{n=1}^{N}\eta_n\right|^p\right] \leq \frac{1}{N^p}\left(K_{1,p}\left(N\mathbb{E}\left[|\zeta|^4\right]\right)^{p/2} + K_{2,p}N\mathbb{E}\left[|\zeta|^{2p}\right]\right) \leq \mathbb{E}\left[|\zeta|^{2p}\right]\left(\frac{K_{1,p}}{N^{p/2}} + \frac{K_{2,p}}{N^{p-1}}\right),$$

and the result follows from the Borel-Cantelli lemma with $p = 2 + \varepsilon/2$, $\varepsilon > 0$.

Lemma 5 *([2, Lemma 2.7]) Consider two $N \times N$ matrices $\mathbf{B}_1$ and $\mathbf{B}_2$, with $\mathbf{B}_2$ being Hermitian, and $\tau \in \mathbb{R}$, $\mathbf{q} \in \mathbb{C}^N$. Then, for each $z \in \mathbb{C}^+$,*

$$\left|\text{Tr}\left[\mathbf{B}_1\left(\left(\mathbf{B}_2 - z\mathbf{I}_M\right)^{-1} - \left(\mathbf{B}_2 + \tau\mathbf{q}\mathbf{q}^H - z\mathbf{I}_M\right)^{-1}\right)\right]\right|$$
$$= \left|\frac{\tau\mathbf{q}^H\left(\mathbf{B}_2 - z\mathbf{I}_M\right)^{-1}\mathbf{B}_1\left(\mathbf{B}_2 - z\mathbf{I}_M\right)^{-1}\mathbf{q}}{1 + \tau\mathbf{q}^H\left(\mathbf{B}_2 - z\mathbf{I}_M\right)^{-1}\mathbf{q}}\right| \leq \frac{\|\mathbf{B}_1\|}{\text{Im}\{z\}}.$$

Lemma 6 *Let $m(z)$ be the Stieltjes transform of a certain measure on $\mathbb{R}^+$, and $\tau \in \mathbb{R}^+$. Then, for each $z \in \mathbb{C}^+$,*

$$\left|\frac{\tau}{1 + \tau m(z)}\right| \leq \tau\frac{|z|}{\text{Im}\{z\}}.$$

PROOF. The bound follows from the fact that, for any $\tau \in \mathbb{R}^+$, $-\frac{\tau}{z(1+\tau m(z))}$ is the Stieltjes transform of a measure on $\mathbb{R}^+$ with total mass τ [12], whose absolute value is therefore bounded by $\tau/\text{Im}\{z\}$ [7, Proposition 2.2].

4 Proof of Theorem 1

In this section, we give a proof of Theorem 1. Let $\boldsymbol{\Xi} = \sqrt{N}\mathbf{X}$ and write

$$\frac{1}{N}\mathbf{R}^{1/2}\boldsymbol{\Xi}\mathbf{T}\boldsymbol{\Xi}^H\mathbf{R}^{1/2} = \frac{1}{N}\sum_{n=1}^{N}t_n\mathbf{R}^{1/2}\xi_n\xi_n^H\mathbf{R}^{1/2} \equiv \frac{1}{N}\sum_{n=1}^{N}\mathbf{y}_n\mathbf{y}_n^H,$$

where t_n is the nth diagonal entry of $\mathbf{T}$, and $\xi_n \in \mathbb{C}^M$ is the nth column vector of $\boldsymbol{\Xi}$. Moreover, let $\mathbf{B}_{(n)} = \mathbf{B} - \frac{1}{N}\mathbf{y}_n\mathbf{y}_n^H$, and consider the map $x_M(e) = \frac{1}{N}\text{Tr}\left[\mathbf{T}\left(\mathbf{I}_N + c_M e\mathbf{T}\right)^{-1}\right] : \mathbb{C}^+ \rightarrow \mathbb{C}^- = \{z \in \mathbb{C} : \text{Im}\{z\} < 0\}$. Indeed, note that

$$\text{Im}\{x_M(e)\} = -\text{Im}\{e\}\frac{1}{N}\sum_{n=1}^{N}\frac{c_M t_n^2}{|1 + t_n c_M e|^2}. \tag{14}$$

For the sake of notational convenience, we will use the following definitions:

$$\mathbf{Q}(z) = \left(\mathbf{B} - z\mathbf{I}_M\right)^{-1}, \quad \mathbf{Q}_{(n)}(z) = \left(\mathbf{B}_{(n)} - z\mathbf{I}_M\right)^{-1}, \quad \mathbf{P}(e) = \left(\mathbf{A} + x(e)\mathbf{R} - z\mathbf{I}_M\right)^{-1}.$$

8

It is easy to see that both $\|\mathbf{Q}(z)\|$ and $\left\|\mathbf{Q}_{(n)}(z)\right\|$ are upper-bounded by $1/\operatorname{Im}\{z\}$. Equivalently, define $\boldsymbol{\Upsilon} = \mathbf{A} + \operatorname{Re}\{x(e)\}\mathbf{R} - \operatorname{Re}\{z\}\mathbf{I}_M$ and $\boldsymbol{\Phi} = \operatorname{Im}\{z\}\mathbf{I}_M - \operatorname{Im}\{x(e)\}\mathbf{R}$, which are, respectively, Hermitian and Hermitian positive definite, and note that

$$\|\mathbf{P}(e)\| = \left\|\left(\boldsymbol{\Phi}^{-1/2}\boldsymbol{\Upsilon}\boldsymbol{\Phi}^{-1/2} - \mathrm{i}\,\mathbf{I}_M\right)^{-1}\boldsymbol{\Phi}^{-1}\right\| \le \left\|\boldsymbol{\Phi}^{-1}\right\| \le \frac{1}{\operatorname{Im}\{z\}}. \tag{15}$$

Furthermore, let $\hat{e}_M(z) = \frac{1}{M}\operatorname{Tr}[\mathbf{R}\mathbf{Q}(z)]$, and notice (see, e.g., [7,12]) that $\hat{e}_M = \hat{e}_M(z)$ is the Stieltjes transform of a certain measure on $\mathbb{R}^+$ with total mass $M^{-1}\operatorname{Tr}[\mathbf{R}] \le \|\mathbf{R}\|_{\text{sup}}$, where with the subscript sup we denote the supremum of the sequence, i.e., here, $\|\mathbf{R}\|_{\text{sup}} = \sup_{M\ge 1}\|\mathbf{R}\|$ (in the sequel, we will similarly use inf for the infimum).

Now, consider the equality

$$\mathbf{Q}^{-1}(z) = \mathbf{B} - z\mathbf{I}_M = \mathbf{A} + x_M(\hat{e}_M)\mathbf{R} - z\mathbf{I}_M + \frac{1}{N}\mathbf{R}^{1/2}\boldsymbol{\Xi}\mathbf{T}\boldsymbol{\Xi}^H\mathbf{R}^{1/2} - x_M(\hat{e}_M)\mathbf{R}.$$

We proceed by factoring the difference of inverses as

$$\mathbf{P}(\hat{e}_M) - \mathbf{Q}(z) = \mathbf{P}(\hat{e}_M)\left(\frac{1}{N}\mathbf{R}^{1/2}\boldsymbol{\Xi}\mathbf{T}\boldsymbol{\Xi}^H\mathbf{R}^{1/2} - x_M(\hat{e}_M)\mathbf{R}\right)\mathbf{Q}(z), \tag{16}$$

where we have used the resolvent identity, i.e., $\mathbf{B}_1 - \mathbf{B}_2 = \mathbf{B}_1\left(\mathbf{B}_2^{-1} - \mathbf{B}_1^{-1}\right)\mathbf{B}_2$. Furthermore, the middle factor on the RHS of (16) can be expanded as

$$\left(\frac{1}{N}\mathbf{R}^{1/2}\boldsymbol{\Xi}\mathbf{T}\boldsymbol{\Xi}^H\mathbf{R}^{1/2} - x_M(\hat{e}_M)\mathbf{R}\right)\mathbf{Q}(z) = \frac{1}{N}\sum_{n=1}^{N}t_n\mathbf{R}^{1/2}\xi_n\xi_n^H\mathbf{R}^{1/2}\mathbf{Q}(z) - x_M(\hat{e}_M)\mathbf{R}\mathbf{Q}(z)$$

$$= \frac{1}{N}\sum_{n=1}^{N}t_n\mathbf{R}^{1/2}\xi_n\xi_n^H\mathbf{R}^{1/2}\mathbf{Q}(z) - \frac{t_n\mathbf{R}\mathbf{Q}(z)}{1 + t_n c_M\hat{e}_M(z)}$$

$$= \frac{1}{N}\sum_{n=1}^{N}\frac{t_n\mathbf{R}^{1/2}\xi_n\xi_n^H\mathbf{R}^{1/2}\mathbf{Q}_{(n)}(z)}{1 + t_n\frac{1}{N}\xi_n^H\mathbf{R}^{1/2}\mathbf{Q}_{(n)}(z)\mathbf{R}^{1/2}\xi_n} - \frac{t_n\mathbf{R}\mathbf{Q}(z)}{1 + t_n c_M\hat{e}_M(z)},$$

where, in the last equality, we have used the Sherman-Morrison-Woodbury identity for rank augmenting matrices in order to write

$$\mathbf{Q}(z) = \mathbf{Q}_{(n)}(z) - \frac{t_n\frac{1}{N}\mathbf{Q}_{(n)}(z)\mathbf{R}^{1/2}\xi_n\xi_n^H\mathbf{R}^{1/2}\mathbf{Q}_{(n)}(z)}{1 + t_n\frac{1}{N}\xi_n^H\mathbf{R}^{1/2}\mathbf{Q}_{(n)}(z)\mathbf{R}^{1/2}\xi_n}, \tag{17}$$

along with the following useful inequality (see [14, Eq. (2.2)])

$$\xi_n^H\mathbf{R}^{1/2}\left(\mathbf{Q}_{(n)}^{-1}(z) + t_n\frac{1}{N}\mathbf{R}^{1/2}\xi_n\xi_n^H\mathbf{R}^{1/2}\right)^{-1} = \frac{1}{1 + t_n\frac{1}{N}\xi_n^H\mathbf{R}^{1/2}\mathbf{Q}_{(n)}(z)\mathbf{R}^{1/2}\xi_n}\xi_n^H\mathbf{R}^{1/2}\mathbf{Q}_{(n)}(z).$$

9

Accordingly, observe that we can further write

$$
\left(\frac{1}{N} \mathbf{R}^{1/2} \mathbf{\Xi T \Xi}^H \mathbf{R}^{1/2} - x_M \left(\hat{e}_M \right) \mathbf{R} \right) \mathbf{Q} \left(z \right) =
$$
$$
= \frac{1}{N} \sum_{n=1}^N \frac{t_n}{1 + t_n c_M \hat{e}_M \left(z \right)} \mathbf{R} \left(\mathbf{Q}_{(n)} \left(z \right) - \mathbf{Q} \left(z \right) \right)
$$
$$
+ \frac{1}{N} \sum_{n=1}^N \frac{t_n}{1 + t_n c_M \hat{e}_M \left(z \right)} \frac{t_n \left(\frac{1}{N} \operatorname{Tr} \left[\mathbf{R Q} \left(z \right) \right] - \frac{1}{N} \xi_n^H \mathbf{R}^{1/2} \mathbf{Q}_{(n)} \left(z \right) \mathbf{R}^{1/2} \xi_n \right)}{1 + t_n \frac{1}{N} \xi_n^H \mathbf{R}^{1/2} \mathbf{Q}_{(n)} \left(z \right) \mathbf{R}^{1/2} \xi_n} \mathbf{R}^{1/2} \xi_n \xi_n^H \mathbf{R}^{1/2} \mathbf{Q}_{(n)} \left(z \right)
$$
$$
+ \frac{1}{N} \sum_{n=1}^N \frac{t_n}{1 + t_n c_M \hat{e}_M \left(z \right)} \left(\mathbf{R}^{1/2} \xi_n \xi_n^H \mathbf{R}^{1/2} \mathbf{Q}_{(n)} \left(z \right) - \mathbf{R Q}_{(n)} \left(z \right) \right).
$$

Let $\phi_M \left(e \right) = \frac{1}{M} \operatorname{Tr} \left[\mathbf{R P} \left(e \right) \right]$ be a function mapping $\mathbb{C}^+$ into $\mathbb{C}^+$ and $f_M \left(e \right) = e - \phi_M \left(e \right)$. Moreover, we define $\mathcal{D} = \left\{ z \in \mathbb{C}^+ : \frac{|z|}{\operatorname{Im} \left\{ z \right\}} \leq K \right\}$, for some finite, positive K large enough. Next, we prove that, almost surely as $M, N \to \infty$ with $0 < \liminf c_M < \limsup c_M < \infty$, for each $z \in \mathcal{D}$,

$$
\operatorname{Tr} \left[\mathbf{\Theta} \left(\mathbf{Q} \left(z \right) - \mathbf{P} \left(\hat{e}_M \right) \right) \right] \to 0, \tag{18}
$$

and

$$
f_M \left(\hat{e}_M \right) \to 0. \tag{19}
$$

To that effect, note that it is enough to show the following almost sure convergence to zero on $\mathcal{D}$ of the following quantities:

$$
\left| \frac{1}{N} \sum_{n=1}^N \frac{t_n}{1 + t_n c_M \hat{e}_M \left(z \right)} \operatorname{Tr} \left[\tilde{\mathbf{\Theta}} \mathbf{R} \left(\mathbf{Q}_{(n)} \left(z \right) - \mathbf{Q} \left(z \right) \right) \right] \right|, \tag{20}
$$

$$
\left| \frac{1}{N} \sum_{n=1}^N \frac{t_n}{1 + t_n c_M \hat{e}_M \left(z \right)} \left(\frac{1}{N} \operatorname{Tr} \left[\mathbf{R Q} \left(z \right) \right] - \frac{1}{N} \xi_n^H \mathbf{R}^{1/2} \mathbf{Q}_{(n)} \left(z \right) \mathbf{R}^{1/2} \xi_n \right) \frac{t_n \xi_n^H \mathbf{R}^{1/2} \mathbf{Q}_{(n)} \left(z \right) \tilde{\mathbf{\Theta}} \mathbf{R}^{1/2} \xi_n}{1 + t_n \frac{1}{N} \xi_n^H \mathbf{R}^{1/2} \mathbf{Q}_{(n)} \left(z \right) \mathbf{R}^{1/2} \xi_n} \right|, \tag{21}
$$

$$
\left| \frac{1}{N} \sum_{n=1}^N \frac{t_n}{1 + t_n c_M \hat{e}_M \left(z \right)} \operatorname{Tr} \left[\tilde{\mathbf{\Theta}} \left(\mathbf{R}^{1/2} \xi_n \xi_n^H \mathbf{R}^{1/2} \mathbf{Q}_{(n)} \left(z \right) - \mathbf{R Q}_{(n)} \left(z \right) \right) \right] \right|, \tag{22}
$$

where $\tilde{\mathbf{\Theta}} = \mathbf{\Psi P} \left(\hat{e}_M \right)$, with $\mathbf{\Psi}$ being an $M \times M$ matrix with uniformly bounded trace norm, which can either be $\mathbf{\Psi} = \mathbf{\Theta}$ or $\mathbf{\Psi} = \frac{1}{M} \mathbf{R}$ (note that $\| M^{-1} \mathbf{R} \|_{\mathrm{tr}} = M^{-1} \operatorname{Tr} \left[\mathbf{R} \right] \leq \| \mathbf{R} \|_{\mathrm{sup}}$). Before proceeding, we note that Lemma 6 yields

$$
\max_{1 \leq n \leq N} \left| \frac{t_n}{1 + t_n c_M \hat{e}_M \left(z \right)} \right| \leq \frac{|z| \, \| \mathbf{T} \|_{\mathrm{sup}}}{\operatorname{Im} \left\{ z \right\}}. \tag{23}
$$

Consider the convergence of (20). First, notice that, for each n, $\frac{1}{N} \xi_n^H \mathbf{R}^{1/2} \mathbf{Q}_{(n)} \left(z \right) \mathbf{R}^{1/2} \xi_n$ can be viewed as the Stieltjes transform of a certain measure on $\mathbb{R}^+$ (see, e.g.,

10

[12]). Hence, from Lemma 6 we have

$$\max_{1 \leq n \leq N} \left| \frac{t_n}{1 + t_n \frac{1}{N} \xi_n^H \mathbf{R}^{1/2} \mathbf{Q}_{(n)}(z) \mathbf{R}^{1/2} \xi_n} \right| \leq \frac{|z| \, \|\mathbf{T}\|_{\sup}}{\text{Im}\{z\}}. \tag{24}$$

Now, applying (17) to expand the term $\mathbf{Q}_{(n)}(z) - \mathbf{Q}(z)$ and using (23) and (24), the result follows readily by Lemma 1 with $p \geq 2$ together with the fact that, for each $z \in \mathcal{D}$,

$$\max_{1 \leq n \leq N} \mathbb{E}\left[\left|\xi_n^H \mathbf{R}^{1/2} \mathbf{Q}_{(n)}(z) \tilde{\mathbf{\Theta}} \mathbf{Z}_{(n)} \mathbf{R}^{1/2} \xi_n\right|^p\right] \leq K_p, \tag{25}$$

for any $M \times M$ matrix $\mathbf{Z}_{(n)}$ independent of ξ_n such that $\left\|\mathbf{Z}_{(n)}\right\|_{\sup}^p \leq K_p$, in particular for $\mathbf{Z}_{(n)} = \mathbf{R}\mathbf{Q}_{(n)}(z)$. Let us now prove (25). Let $\hat{e}_M^{(n)}(z) = \frac{1}{M} \text{Tr}\left[\mathbf{R}\mathbf{Q}_{(n)}(z)\right]$, which is a Stieltjes transform with same properties as $\hat{e}_M(z)$, and notice by applying twice the resolvent identity to $\mathbf{P}(\hat{e}_M) - \mathbf{P}\left(\hat{e}_M^{(n)}\right)$ that

$$\mathbf{P}(\hat{e}_M) = \mathbf{P}\left(\hat{e}_M^{(n)}\right) + \theta_M\left(\hat{e}_M, \hat{e}_M^{(n)}\right) \frac{c_M}{M} \text{Tr}\left[\mathbf{R}\left(\mathbf{Q}(z) - \mathbf{Q}_{(n)}(z)\right)\right] \mathbf{P}(\hat{e}_M) \mathbf{R} \mathbf{P}\left(\hat{e}_M^{(n)}\right), \tag{26}$$

where we have defined

$$\theta_M(e_1, e_2) = \frac{1}{N} \text{Tr}\left[\mathbf{T}^2\left(\mathbf{I}_N + c_M e_1 \mathbf{T}\right)^{-1}\left(\mathbf{I}_N + c_M e_2 \mathbf{T}\right)^{-1}\right],$$

and note from Lemma 6 that

$$\left|\theta_M\left(\hat{e}_M, \hat{e}_M^{(n)}\right)\right| \leq \max_{1 \leq n \leq N}\left|\frac{t_n}{1 + t_n c_M \hat{e}_M^{(n)}(z)}\right|\left|\frac{t_n}{1 + t_n c_M \hat{e}_M(z)}\right| \leq \left(\frac{|z| \, \|\mathbf{T}\|_{\sup}}{\text{Im}\{z\}}\right)^2. \tag{27}$$

Now, we write $\tilde{\mathbf{\Theta}} = \mathbf{\Psi} \mathbf{P}(\hat{e}_M)$ using the two terms on the RHS of (26). From the first term, by Lemma 3 with $\mathbf{C} = \mathbf{R}^{1/2} \mathbf{Q}_{(n)}(z) \mathbf{\Psi} \mathbf{P}\left(\hat{e}_M^{(n)}\right) \mathbf{Z}_{(n)} \mathbf{R}^{1/2}$, we obtain

$$\max_{1 \leq n \leq N} \mathbb{E}\left[\left|\xi_n^H \mathbf{R}^{1/2} \mathbf{Q}_{(n)}(z) \mathbf{\Psi} \mathbf{P}\left(\hat{e}_M^{(n)}\right) \mathbf{Z}_{(n)} \mathbf{R}^{1/2} \xi_n\right|^p\right] \leq \frac{K_p \, \|\mathbf{R}\|_{\sup}^p \, \|\mathbf{\Psi}\|_{\text{tr,sup}}^p}{(\text{Im}\{z\})^{2p}}. \tag{28}$$

Regarding the second term in (26), i.e., $\mathbf{P}(\hat{e}_M) - \mathbf{P}\left(\hat{e}_M^{(n)}\right)$, from (27) and Lemma 5 we have

$$\left\|\left(\mathbf{P}(\hat{e}_M) - \mathbf{P}\left(\hat{e}_M^{(n)}\right)\right) \mathbf{Z}_{(n)} \mathbf{Q}_{(n)}(z)\right\|^p \leq \frac{K_p\left(|z| \, \|\mathbf{R}\|_{\sup} \, \|\mathbf{T}\|_{\sup}\right)^{2p}}{N^p\left(\text{Im}\{z\}\right)^{6p}}. \tag{29}$$

11

Hence, from the second term we get

$$\max_{1 \leq n \leq N} \mathbb{E}\left[\left|\xi_n^H \mathbf{R}^{1/2} \mathbf{Q}_{(n)}(z) \mathbf{\Psi}\left(\mathbf{P}(\hat{e}_M) - \mathbf{P}(\hat{e}_M^{(n)})\right)\mathbf{Z}_{(n)}\mathbf{R}^{1/2}\xi_n\right|^p\right]$$

$$\overset{(a)}{\leq} \frac{K_p\left(|z|\,\|\mathbf{R}\|_{\mathrm{sup}}\,\|\mathbf{T}\|_{\mathrm{sup}}\right)^{2p}\|\mathbf{\Psi}\|_{\mathrm{tr,sup}}}{(\mathrm{Im}\{z\})^{6p}}\max_{1 \leq n \leq N}\mathbb{E}\left[\left(\frac{1}{N}\xi_n^H \mathbf{R}\xi_n\right)^p\right]$$

$$\overset{(b)}{\leq} \frac{K_p\,\|\mathbf{R}\|_{\mathrm{sup}}^{3p}\left(|z|\,\|\mathbf{T}\|_{\mathrm{sup}}\right)^{2p}\|\mathbf{\Psi}\|_{\mathrm{tr,sup}}^p}{(\mathrm{Im}\{z\})^{6p}}, \tag{30}$$

where (a) in (30) follows by the Cauchy-Schwarz inequality along with (29) and the second inequality in (10), and (b) follows by Lemma 3 with $\mathbf{C} = N^{-1}\mathbf{R}$. Finally, (25) follows by applying (8) together with (28) and (30).

We now prove the convergence of (21). Using the Cauchy-Schwartz inequality, we write

$$\mathbb{E}\left[\left|\left(\frac{1}{N}\mathrm{Tr}\left[\mathbf{R}\mathbf{Q}(z)\right] - \frac{1}{N}\xi_n^H \mathbf{R}^{1/2}\mathbf{Q}_{(n)}(z)\mathbf{R}^{1/2}\xi_n\right)\frac{t_n\xi_n^H \mathbf{R}^{1/2}\mathbf{Q}_{(n)}(z)\tilde{\mathbf{\Theta}}\mathbf{R}^{1/2}\xi_n}{1 + t_n\frac{1}{N}\xi_n^H \mathbf{R}^{1/2}\mathbf{Q}_{(n)}(z)\mathbf{R}^{1/2}\xi_n}\right|^p\right]$$

$$\leq \mathbb{E}^{1/2}\left[\left|\frac{1}{N}\mathrm{Tr}\left[\mathbf{R}\mathbf{Q}(z)\right] - \frac{1}{N}\xi_n^H \mathbf{R}^{1/2}\mathbf{Q}_{(n)}(z)\mathbf{R}^{1/2}\xi_n\right|^{2p}\right]\mathbb{E}^{1/2}\left[\left|\frac{t_n\xi_n^H \mathbf{R}^{1/2}\mathbf{Q}_{(n)}(z)\tilde{\mathbf{\Theta}}\mathbf{R}^{1/2}\xi_n}{1 + t_n\frac{1}{N}\xi_n^H \mathbf{R}^{1/2}\mathbf{Q}_{(n)}(z)\mathbf{R}^{1/2}\xi_n}\right|^{2p}\right].$$

Now, notice that from (24) and (25) with $\mathbf{Z}_{(n)} = \mathbf{I}_M$ the second factor of the RHS of the previous inequality is uniformly bounded on $\mathcal{D}$ for any $p \geq 1$. Furthermore, regarding the first factor, we rewrite $\mathbf{Q}(z)$ by using the two terms on the RHS of (17) and get, on the one hand (cf. Lemma 5),

$$\max_{1 \leq n \leq N}\mathbb{E}\left[\left|\frac{1}{N}\frac{t_n\frac{1}{N}\xi_n^H \mathbf{R}^{1/2}\mathbf{Q}_{(n)}(z)\mathbf{R}\mathbf{Q}_n(z)\mathbf{R}^{1/2}\xi_n}{1 + t_n\frac{1}{N}\xi_n^H \mathbf{R}^{1/2}\mathbf{Q}_{(n)}(z)\mathbf{R}^{1/2}\xi_n}\right|^{2p}\right] \leq \frac{1}{N^{2p}}\left(\frac{\|\mathbf{R}\|_{\mathrm{sup}}}{\mathrm{Im}\{z\}}\right)^{2p}, \tag{31}$$

and on the other hand, using (12) with $\mathbf{C} = N^{-1/2}\mathbf{R}^{1/2}\mathbf{Q}_{(n)}(z)\mathbf{R}^{1/2}$ we get

$$\max_{1 \leq n \leq N}\mathbb{E}\left[\left|\frac{1}{N}\mathrm{Tr}\left[\mathbf{R}\mathbf{Q}_{(n)}(z)\right] - \frac{1}{N}\xi_n^H \mathbf{R}^{1/2}\mathbf{Q}_{(n)}(z)\mathbf{R}^{1/2}\xi_n\right|^{2p}\right] \leq \frac{1}{N^p}\left(\frac{K_p\,\|\mathbf{R}\|_{\mathrm{sup}}}{\mathrm{Im}\{z\}}\right)^{2p}. \tag{32}$$

Finally, the result follows from Lemma 1 with $p \geq 2$ by applying (8) along with (31) and (32).

In order to prove the convergence of (22), by the triangular inequality, it is enough to show the almost sure convergence to zero on $\mathcal{D}$ of the following

12

quantities:

$$\left| \frac{1}{N} \sum_{n=1}^{N} \left(\frac{t_n}{1 + t_n c_M \hat{e}_M(z)} - \frac{t_n}{1 + t_n c_M \hat{e}_M^{(n)}(z)} \right) \left(\xi_n^H \mathbf{R}^{1/2} \mathbf{Q}_{(n)}(z) \tilde{\mathbf{\Theta}} \mathbf{R}^{1/2} \xi_n - \text{Tr} \left[\mathbf{R} \mathbf{Q}_{(n)}(z) \tilde{\mathbf{\Theta}} \right] \right) \right|,$$
(33)

$$\left| \frac{1}{N} \sum_{n=1}^{N} \frac{t_n}{1 + t_n c_M \hat{e}_M^{(n)}(z)} \left(\xi_n^H \mathbf{R}^{1/2} \mathbf{Q}_{(n)}(z) \mathbf{\Psi} \mathbf{P} \left(\hat{e}_M^{(n)} \right) \mathbf{R}^{1/2} \xi_n - \text{Tr} \left[\mathbf{R} \mathbf{Q}_{(n)}(z) \mathbf{\Psi} \mathbf{P} \left(\hat{e}_M^{(n)} \right) \right] \right) \right|,$$
(34)

$$\left| \frac{1}{N} \sum_{n=1}^{N} \frac{t_n}{1 + t_n c_M \hat{e}_M^{(n)}(z)} \xi_n^H \mathbf{R}^{1/2} \mathbf{Q}_{(n)}(z) \mathbf{\Psi} \left(\mathbf{P}(\hat{e}_M) - \mathbf{P} \left(\hat{e}_M^{(n)} \right) \right) \mathbf{R}^{1/2} \xi_n \right|, \quad (35)$$

$$\left| \frac{1}{N} \sum_{n=1}^{N} \frac{t_n}{1 + t_n c_M \hat{e}_M^{(n)}(z)} \text{Tr} \left[\mathbf{R} \mathbf{Q}_{(n)}(z) \mathbf{\Psi} \left(\mathbf{P} \left(\hat{e}_M^{(n)} \right) - \mathbf{P}(\hat{e}_M) \right) \right] \right|. \quad (36)$$

Regarding (33), using (27) along with Lemma 5, note first that

$$\frac{t_n}{1 + t_n c_M \hat{e}_M(z)} - \frac{t_n}{1 + t_n c_M \hat{e}_M^{(n)}(z)} = \frac{1}{N} \frac{c_M t_n^2 \text{Tr} \left[\mathbf{R} \left(\mathbf{Q}_{(n)}(z) - \mathbf{Q}(z) \right) \right]}{(1 + t_n c_M \hat{e}_M(z)) \left(1 + t_n c_M \hat{e}_M^{(n)}(z) \right)} \leq \frac{\|\mathbf{R}\|_{\text{sup}} \left(|z| \|\mathbf{T}\|_{\text{sup}} \right)^2}{N (\text{Im}\{z\})^3}.$$
(37)

Convergence of (33) follows by Lemma 1 with $p \geq 2$ by applying (37) and (8) along with (25) with $\mathbf{Z}_{(n)} = \mathbf{I}_M$ and $\left| \text{Tr} \left[\mathbf{R} \mathbf{Q}_{(n)}(z) \tilde{\mathbf{\Theta}} \right] \right| \leq \|\mathbf{R}\| |z| \|\mathbf{\Psi}\|_{\text{tr,sup}} / (\text{Im}\{z\})^2$ (cf. inequality (10)). On the other hand, convergence of (34) follows directly by Lemma 4 with $\mathbf{C}_{(n)} = t_n \left(1 + t_n c_M \hat{e}_M^{(n)}(z) \right)^{-1} \mathbf{R}^{1/2} \mathbf{Q}_{(n)}(z) \mathbf{\Psi} \mathbf{P} \left(\hat{e}_M^{(n)} \right) \mathbf{R}^{1/2}$. Convergence of (35) follows from Lemma 1 with $p \geq 2$ by applying (26) and using (23) and (27) along with Lemma 5, and (25) with $\mathbf{Z}_{(n)} = \mathbf{R} \mathbf{P} \left(\hat{e}_M^{(n)} \right)$. Finally, (36) vanishes almost surely by Lemma 1 with $p \geq 2$ using (23) along with (cf. inequality (10))

$$\left| \text{Tr} \left[\mathbf{R} \mathbf{Q}_{(n)}(z) \mathbf{\Psi} \left(\mathbf{P} \left(\hat{e}_M^{(n)} \right) - \mathbf{P}(\hat{e}_M) \right) \right] \right|^p \leq \frac{c_M^p \|\mathbf{R}\|_{\text{sup}}^{3p} \left(|z| \|\mathbf{T}\|_{\text{sup}} \right)^{2p} \|\mathbf{\Psi}\|_{\text{tr,sup}}^p}{N^p (\text{Im}\{z\})^{6p}},$$

where we have used (29) with $\mathbf{Z}_{(n)} = \mathbf{R}$. This proves (18) and (19).

We next show that $\mathbf{P}(\hat{e}_M)$ in (18) and (19) can be replaced by $\mathbf{P}(e_M)$, where e_M is the unique deterministic equivalent in the statement of Theorem 1. From (19) and $f_M(e_M) = 0$ we clearly have that $f_M(\hat{e}_M) - f_M(e_M) \to 0$, for each $z \in \mathcal{D}$, and so we only need to show that this implies $\hat{e}_M - e_M \to 0$. Indeed, notice that $f_M(\hat{e}_M) - f_M(e_M) = (\hat{e}_M - e_M)(1 - \rho_M(\hat{e}_M, e_M))$, where we have defined

$$\rho_M(e_1, e_2) = \theta_M(e_1, e_2) \frac{1}{M} \text{Tr} [\mathbf{R} \mathbf{P}(e_1) \mathbf{R} \mathbf{P}(e_2)]. \quad (38)$$

Observe that $\rho_M(e_M, e_M) = 1$. Moreover, after some algebraic manipulations

13

we get $f_M(\hat{e}_M) - f_M(e_M) = (\hat{e}_M - e_M)^2 \gamma_M(\hat{e}_M, e_M)$, where

$$\gamma_M(\hat{e}_M, e_M) = \frac{1}{N} \sum_{n=1}^{N} \frac{c_M^2 t_n^3}{(1 + t_n c_M e_M)^2 (1 + t_n c_M \hat{e}_M)} \frac{1}{M} \operatorname{Tr}\left[\mathbf{R}\mathbf{P}(e_M) \mathbf{R}\mathbf{P}(\hat{e}_M)\right]$$
$$- \theta_M(\hat{e}_M, e_M) \frac{1}{N} \sum_{n=1}^{N} \left(\frac{c_M t_n}{1 + t_n c_M e_M}\right)^2 \frac{1}{M} \operatorname{Tr}\left[\left(\mathbf{R}\mathbf{P}(e_M)\right)^2 \mathbf{R}\mathbf{P}(\hat{e}_M)\right],$$

such that, using the fact that e_M is the Stieltjes transform of a certain measure on $\mathbb{R}^+$ (cf. Section 4.2), by the triangular inequality and the application of Lemma 6 as in (23) along with inequality (10),

$$|\gamma_M(\hat{e}_M, e_M)| \leq \frac{\left(c_M \|\mathbf{R}\|_{\sup}\right)^2 \left(|z| \|\mathbf{T}\|_{\sup}\right)^3}{(\operatorname{Im}\{z\})^5} \left(1 + \frac{|z| \|\mathbf{R}\|_{\sup} \|\mathbf{T}\|_{\sup}}{(\operatorname{Im}\{z\})^2}\right).$$

This implies that $\hat{e}_M - e_M \to 0$ almost surely for each z of a countable family with an accumulation point in a compact subset of $\mathcal{D}$. Now, as a consequence of being Stieltjes transforms of bounded measures on $\mathbb{R}^+$, we have that both $\hat{e}_M(z)$ and $e_M(z)$ are analytic on $\mathbb{C} - \mathbb{R}^+$ (see [7, Proposition 2.2]), and so is $h_M(z) = \hat{e}_M(z) - e_M(z)$. Moreover, we have that $|h_M(z)| \leq \frac{2}{\operatorname{dist}(z, \mathbb{R}^+)}$, where dist stands for the Euclidean distance (see again [7, Proposition 2.2]). Thus, $\{h_M\}$ is a normal family and by Montel's theorem there exists a subsequence which converges uniformly on each compact subset of $\mathbb{C} - \mathbb{R}^+$ to an analytic function which, from above, vanishes almost surely on $\mathbb{C} - \mathbb{R}^+$. Thus, the entire sequence converges uniformly to zero on each compact subset of $\mathbb{C} - \mathbb{R}^+$, and so $|h_M(z)| \to 0$ for each $z \in \mathbb{C} - \mathbb{R}^+$.

Define $g_M(e) = \operatorname{Tr}[\boldsymbol{\Theta}\mathbf{P}(e)]$. The fact that $g_M(\hat{e}_M) - g_M(e_M) \to 0$ for each $z \in \mathbb{C} - \mathbb{R}^+$ follows finally by observing that

$$g_M(\hat{e}_M) - g_M(e_M) = (e_M - \hat{e}_M) \theta_M(\hat{e}_M, e_M) \operatorname{Tr}\left[\boldsymbol{\Theta}\mathbf{P}(\hat{e}_M) \mathbf{R}\mathbf{P}(e_M)\right],$$

and noting from (27) and $|\operatorname{Tr}[\boldsymbol{\Theta}\mathbf{P}(\hat{e}_M) \mathbf{R}\mathbf{P}(e_M)]| \leq \|\mathbf{R}\|_{\sup} \|\boldsymbol{\Theta}\|_{\mathrm{tr,sup}} / (\operatorname{Im}\{z\})^2$ (cf. inequality (10)) that the factor multiplying $e_M - \hat{e}_M$ in the RHS is bounded in absolute value uniformly in M.

4.1 Uniqueness of the limit

In order to prove uniqueness in $\mathbb{C}^+$ of the solution of equation (7), assume that we have two solutions $e_1, e_2 \in \mathbb{C}^+$ and, using the resolvent identity, write

$$e_1 - e_2 = -(x_M(e_1) - x_M(e_2)) \frac{1}{M} \operatorname{Tr}\left[\mathbf{R}\mathbf{P}(e_1) \mathbf{R}\mathbf{P}(e_2)\right] = (e_1 - e_2) \rho_M(e_1, e_2).$$

14

Assume $e_1 \neq e_2$, so that, necessarily, $\rho_M = \rho_M(e_1, e_2) = 1$. Using the Cauchy-Schwarz inequality, we have $|\rho_M|^2 \leq \varphi_M(e_1)\varphi_M(e_2)$, where

$$\varphi_M(e_i) = \frac{1}{N}\sum_{n=1}^{N}\frac{c_M t_n^2}{|1+t_n c_M e_i|^2}\frac{1}{M}\operatorname{Tr}\left[\mathbf{P}^H(e_i)\mathbf{R}\mathbf{P}(e_i)\mathbf{R}\right], \quad i = 1, 2. \quad (39)$$

Then, from (14) we note that

$$\frac{1}{N}\sum_{n=1}^{N}\frac{c_M t_n^2}{|1+t_n c_M e_i|^2} = -\frac{\operatorname{Im}\{x(e_i)\}}{\operatorname{Im}\{e_i\}}. \quad (40)$$

On the other hand, using the fact that $x(e) \in \mathbb{C}^-$ for each $z \in \mathbb{C}^+$ along with

$$\frac{\operatorname{Im}\{e_i\}}{\operatorname{Im}\{x_M(e_i)\}} = \frac{\operatorname{Im}\{z\}}{\operatorname{Im}\{x_M(e_i)\}}\frac{1}{M}\operatorname{Tr}\left[\mathbf{P}^H(e_1)\mathbf{R}\mathbf{P}(e_1)\right] - \frac{1}{M}\operatorname{Tr}\left[\mathbf{P}^H(e_1)\mathbf{R}\mathbf{P}(e_1)\mathbf{R}\right],$$

we can conclude that

$$\frac{1}{M}\operatorname{Tr}\left[\mathbf{P}^H(e_1)\mathbf{R}\mathbf{P}(e_1)\mathbf{R}\right] < -\frac{\operatorname{Im}\{e_i\}}{\operatorname{Im}\{x(e_i)\}}. \quad (41)$$

Hence, it is clear from (39) by using (40) and (41) that $\varphi_M(e_i) < 1$, $i = 1, 2$, and so $\rho_M < 1$, contradicting the fact that $\rho_M = 1$. Therefore we must necessarily have $e_1 = e_2$.

4.2 Existence of a deterministic asymptotic equivalent

We use the fixed point theorem to show that the equation $e = \phi_M(e)$ presents at least one solution in $\mathbb{C}^+$. Let $\chi_M = \max_{1 \leq m \leq M}|\lambda_m(\mathbf{A}) - z|$, which is strictly positive by assumption for all M and each $z \in \mathbb{C} - \mathbb{R}^+$. Consider the open set:

$$\Omega = \left\{ e \in \mathbb{C}^+ : \operatorname{Im}\{e\} > \frac{\|\mathbf{R}\|_{\sup}}{2c_{\inf}\chi_{\inf}}\left[1 + \sqrt{1 + \frac{4c_{\inf}\chi_{\inf}}{\|\mathbf{R}\|_{\sup}}}\right]\right\}.$$

We show that, for any two values $e_1, e_2 \in \Omega$, $|\phi_M(e_1) - \phi_M(e_2)| < |e_1 - e_2|$, so that the mapping ϕ_M is contractive when constrained to Ω. The fixed point theorem guarantees that ϕ_M presents a fixed point in Ω, which guarantees the existence of a solution of the equation $e = \phi_M(e)$ in $\Omega \subset \mathbb{C}^+$.

Observe first that we can write, using the resolvent identity, $\phi_M(e_1) - \phi_M(e_2) = (e_1 - e_2)\rho_M(e_1, e_2)$. Hence, it is sufficient to see that $|\rho_M(e_1, e_2)| < 1$, or alternatively, from the Cauchy-Schwarz inequality, that $\varphi_M(e) < 1$ for any $e \in \Omega$. Note that we can write

$$\frac{1}{M}\operatorname{Tr}\left[\mathbf{P}^H(e)\mathbf{R}\mathbf{P}(e)\mathbf{R}\right] \leq \left\|\mathbf{R}^{1/2}\mathbf{P}(e)\mathbf{R}^{1/2}\right\| = \|\mathbf{R}\|\left\|\mathbf{A} + x(e)\mathbf{R} - z\mathbf{I}_M\right\|^{-1}. \quad (42)$$

15

Now, using $\|\mathbf{B}_1\| - \|\mathbf{B}_2\| \leq \|\mathbf{B}_1 + \mathbf{B}_2\| \leq \|\mathbf{B}_1\| + \|\mathbf{B}_2\|$, together with

$$|\mathrm{Im}\,\{x\}| \leq |x\,(e)| \leq \frac{1}{N} \sum_{n=1}^{N} \frac{t_n}{|1 + t_n c_M e|} \leq \frac{1}{N} \sum_{n=1}^{N} \frac{t_n}{t_n c_M \mathrm{Im}\,\{e\}} \leq \frac{1}{c_{\mathrm{inf}} \mathrm{Im}\,\{e\}}, \tag{43}$$

and the fact that $\|\mathbf{A} - z\mathbf{I}_M\| \geq \chi_{\mathrm{inf}}$, we have

$$\|\mathbf{A} + x\,(e)\,\mathbf{R} - z\mathbf{I}_M\|^{-1} \geq \chi_{\mathrm{inf}} - \frac{1}{c_{\mathrm{inf}} \mathrm{Im}\,\{e\}}, \tag{44}$$

which is clearly positive on Ω by assumption. Now, using (14) and the above inequalities, in order to show that $\varphi_M\,(e) < 1$, it is sufficient to prove that

$$\frac{|\mathrm{Im}\,\{x\,(e)\}|}{\mathrm{Im}\,\{e\}} \frac{\|\mathbf{R}\|_{\mathrm{sup}}}{\chi_{\mathrm{inf}} - \frac{1}{c_{\mathrm{inf}} \mathrm{Im}\{e\}}} < 1,$$

or, equivalently, $c_{\mathrm{inf}} \chi_{\mathrm{inf}} \mathrm{Im}^2\,\{e\} - \mathrm{Im}\,\{e\}\,\|\mathbf{R}\|_{\mathrm{sup}} - \|\mathbf{R}\|_{\mathrm{sup}} > 0$, which is always the case for each $e \in \Omega$, and this concludes the proof of the existence of e_M.

Finally, we show that $e_M\,(z)$ is the Stieltjes transform of a certain bounded measure on $\mathbb{R}^+$ with total mass $\frac{1}{M} \mathrm{Tr}\,[\mathbf{R}]$. Indeed, notice that $e_M\,(z)$ is analytic on $\mathbb{C}^+$ and, additionally,

$$\mathrm{Im}\,\{e_M\} = \frac{\mathrm{Im}\,\{z\} \frac{1}{M} \mathrm{Tr}\left[\mathbf{P}^H\,(e_M)\,\mathbf{R}\mathbf{P}\,(e_M)\right]}{1 - \varphi_M\,(e_M)},$$

and

$$\mathrm{Im}\,\{ze_M\} = \frac{\mathrm{Im}\,\{z\}\left(\frac{1}{M} \mathrm{Tr}\left[\mathbf{P}^H\,(e_M)\,\mathbf{R}\mathbf{P}\,(e_M)\,\mathbf{A}\right] + \frac{1}{N}\sum_{n=1}^{N}\frac{t_n}{|1 + t_n c_M e_M|^2}\frac{1}{M} \mathrm{Tr}\left[\mathbf{P}^H\,(e_M)\,\mathbf{R}\mathbf{P}\,(e_M)\,\mathbf{R}\right]\right)}{1 - \varphi_M\,(e_M)}.$$

Now, since $\varphi_M\,(e_M) < 1$ on $\mathbb{C}^+$ (cf. Section 4.1) and since $\mathbf{A}$ is assumed to be nonnegative definite, we conclude that both $e_M\,(z)$ and $ze_M\,(z)$ map $\mathbb{C}^+$ into $\mathbb{C}^+$, and the claim follows from [7, Proposition 2.2] using the fact that $(y = \mathrm{Im}\,\{z\})\,\lim_{y\to+\infty} -\mathrm{i}\,y e_M\,(\mathrm{i}\,y) = \frac{1}{M} \mathrm{Tr}\,[\mathbf{R}]$.

References

[1] Z.D. Bai, G.M. Pan, and B.Q. Miao. On asymptotics of eigenvectors of large sample covariance matrix. *Annals of Probability*, 35(4):1532–1572, 2007.

[2] Z.D. Bai and J. Silverstein. No eigenvalues outside the support of the limiting spectral distribution of large dimensional sample covariance matrices. *Annals of Probability*, 26(1):316–345, 1998.

16

[3] J.S. Geronimo and T.P. Hill. Necessary and sufficient condition that the limit of Stieltjes transforms is a Stieltjes transform. *J. Approx. Theory*, 121(1):54–60, March 2003.

[4] V.L. Girko. Asymptotic behavior of eigenvalues of empirical covariance matrices I. *Theory of Probability and Mathematical Statistics*, 44:37–44, 1992.

[5] V.L. Girko. *Theory of Stochastic Canonical Equations, Vol. 1 & 2*. Kluwer, 2001.

[6] W. Hachem, P. Loubaton, and J. Najim. The empirical distribution of the eigenvalues of a gram matrix with a given variance profile. *Annales de l'Institut Henri Poincaré, Probab. Statist.*, 42(6):649–670, 2006.

[7] W. Hachem, P. Loubaton, and J. Najim. Deterministic equivalents for certain functionals of large random matrices. *Annals of Applied Probability*, 17(3):875–930, 2007.

[8] R.A. Horn and C.R. Johnson. *Topics in Matrix Analysis*. Cambridge University Press, 1991.

[9] N. El Karoui. On spectral properties of large dimensional correlation matrices and covariance matrices computed from elliptically distributed data. *Accepted for publication in Annals of Applied Probability*, 2008.

[10] V.A. Marçenko and L.A. Pastur. Distribution of eigenvalues for some sets of random matrices. *Mathematics of the USSR - Sbornik*, 1(4):457–483, 1967.

[11] X. Mestre. On the asymptotic behavior of quadratic forms of the resolvent of certain covariance-type matrices. Tech. Rep. CTTC/RC/2006-01, Centre Tecnològic de Telecomunicacions de Catalunya, July 2006.

[12] D. Paul and J.W. Silverstein. No eigenvalues outside the support of limiting empirical spectral distribution of a separable covariance matrix. *Journal of Multivariate Analysis*, 100(1):37–57, January 2009.

[13] J.W. Silverstein. Strong convergence of the empirical distribution of eigenvalues of large dimensional random matrices. *Journal of Multivariate Analysis*, 55:331–339, August 1995.

[14] J.W. Silverstein and Z.D. Bai. On the empirical distribution of eigenvalues of a class of large dimensional random matrices. *Journal of Multivariate Analysis*, 54(2):175–192, February 1995.

[15] H.L. Van Trees. *Optimum Array Processing*. John Wiley & Sons, NY, USA, 2002.

[16] Y.Q. Yin. Limiting spectral distribution for a class of random matrices. *Journal of Multivariate Analysis*, 20(1):50–68, October 1986.