

A Deterministic Equivalent for the Analysis of Correlated MIMO Multiple Access Channels

Romain Couillet^{†,‡}, Mérouane Debbah[‡], Jack W. Silverstein[⊥]

Abstract

This article provides novel deterministic equivalents for the Stieltjes transform and the Shannon transform of a class of large dimensional random matrices. These results are used to characterise the ergodic rate region of multiple antenna multiple access channels, when each point-to-point propagation channel is modelled according to the Kronecker model. We specifically provide an approximation of all rates achieved within the ergodic rate region and we provide an approximation of the linear precoders that achieve the boundary of the rate region as well as an iterative water-filling algorithm to obtain these precoders. An original feature of this work is that the proposed deterministic equivalents are proved valid even for strong correlation patterns at both communication sides. The above results are validated by Monte Carlo simulations.

I. INTRODUCTION

When mobile networks were expected to run out of power and frequency resources while being simultaneously subject to a demand for higher transmission rates, Foschini [1] introduced the idea of multiple input multiple output (MIMO) communication schemes. Telatar [2] then predicted a growth of the channel capacity by a factor $\min(N, n)$ for an $N \times n$ MIMO system compared to the single-antenna case when the matrix-valued channel is modelled with independent and identically distributed (i.i.d.) standard Gaussian entries. In practical systems though, this linear gain can only be achieved for high signal-to-noise ratios (SNR) and for uncorrelated transmit and receive antenna arrays at both communication sides. Nevertheless, the current scarcity of available frequency resources has led to a widespread incentive for MIMO communications. Mobile terminal engineers now embed numerous antennas in small devices. Due to space limitations, this inevitably induces channel correlation and thus reduced transmission rates. An implication of the results introduced in this paper is the ability to study the performance of MIMO systems subject to strong correlation effects in multi-user and multi-cellular contexts, a question which is paramount to cellular service providers.

Although alternative communication models could be treated using similar mathematical expressions, such as cooperative and non-cooperative multi-cell communications with users equipped with multiple antennas, the present

[†]ST-Ericsson, 635 Route des Lucioles, 06560 Sophia Antipolis

[‡]Alcatel-Lucent Chair on Flexible Radio, Supélec, 3 rue Joliot-Curie, 91192 Gif sur Yvette

[⊥]Department of Mathematics, North Carolina State University, Raleigh, North Carolina 27695-8205

article investigates the MIMO multiple access channel (MIMO-MAC), where K multi-antenna mobile terminals transmit information to a single receiver. Under perfect channel state information at the transmitters (CSIT), the boundaries of the achievable rate region of the MIMO-MAC have been characterised by Yu *et al.* [3] who provide an iterative water-filling algorithm to obtain the sum rate maximising precoders. However, to achieve perfect CSIT, the channel must be quasi-static during a sufficiently long period to allow feedback or pilot signalling from the receiver to the transmitters. For high mobility wireless services, this is often unacceptable. In this situation, the transmitters are often assumed to have statistical information about the random fast varying channels, such as first order moments of their distribution. The achievable rates are in this case the points lying in the *ergodic* rate region. It is however difficult to characterise the boundary of the ergodic rate region because it is difficult to compute the optimal precoders that reach the boundaries. In the single-user context, an algorithm was provided by Vu *et al.* in [4] to solve this problem. However, the technique of [4] is rather involved as it requires nested Monte Carlo simulations and does not provide any insight on the nature of the optimal precoders.

In the present article, we provide a parallel approach that consists in approximating the ergodic sum rate by *deterministic equivalents*. That is, for all finite system dimensions, we provide an approximation of the ergodic rates, which is accurate as the system dimensions grow asymptotically large. Furthermore, we provide an efficient way to derive an asymptotically accurate approximation of the optimal precoders. The mathematical field of large dimensional random matrices is particularly suited for this purpose, as it can provide deterministic equivalents of the achievable rates depending only on the relevant channel parameters, e.g., the long-term transmit and receive channel covariance matrices in the present situation, the deterministic line of sight components in Rician models as in [5] etc. The earliest notable result in line with the present study is due to Tulino *et al.* [6], who provide an expression of the asymptotic ergodic capacity of point-to-point MIMO communications when the random channel matrix is composed of i.i.d. Gaussian entries. In [7], Peacock *et al.* extend the asymptotic result of [6] in the context of multi-user communications by considering a K -user MAC with channels $\mathbf{H}_1, \dots, \mathbf{H}_K$ modelled as Gaussian with a separable variance profile. This is, the entries of \mathbf{H}_k are Gaussian independent with (i, j) -th entry of zero mean and variance $\sigma_{k,ij}^2$ that can be written as a product $\sigma_{k,ij}^2 = r_{k,i}t_{k,j}$ of a term depending on i and a term depending on j . The asymptotic eigenvalue distribution of this matrix model is derived, but neither any explicit expression of the sum rate is provided as in [6], nor is any ergodic capacity maximising policy derived. In [8], Soysal *et al.* derive the sum rate maximising precoder policy in the case of a MAC channel with K users whose channels $\mathbf{H}_1, \dots, \mathbf{H}_k$ are one-side correlated zero mean Gaussian, in the sense that all rows of \mathbf{H}_k have a common covariance matrix, different for each k .

In this article, we concentrate on the more general *Kronecker* channel model. This is, we assume a K -user MIMO-MAC, with channels $\mathbf{H}_1, \dots, \mathbf{H}_K$, where each \mathbf{H}_k can be written in the form of a product $\mathbf{R}_k^{\frac{1}{2}} \mathbf{X}_k \mathbf{T}_k^{\frac{1}{2}}$ where \mathbf{X}_k has i.i.d. zero mean Gaussian entries and the left and right correlation matrices \mathbf{R}_k and \mathbf{T}_k are deterministic nonnegative definite Hermitian matrices. This model clearly covers the aforementioned channel models of [6], [7] and [8] as special cases. The Kronecker model is particularly suited to model communication channels that show transmit and receive correlations, different from one user to the next, in a rich scattering environment. Nonetheless,

the Kronecker model is only valid in the absence of a line-of-sight component in the channel, when a sufficiently large number of scatterers is present in the communication medium to justify the i.i.d. aspect of the inner Gaussian matrix and when the channel is frequency flat over the transmission bandwidth. Using similar tools as those used in this article, many works have studied these channel models, mostly in a single-user context. We remind the main contributions, from which the present work borrows several ideas. In [5], [9], [10], Hachem *et al.* study the point-to-point multi-antenna Rician channel model, i.e., non-central Gaussian matrices with a variance profile, for which they provide a deterministic equivalent of the ergodic capacity [5], the corresponding ergodic capacity-achieving input covariance matrix [9] and a central limit theorem for the ergodic capacity [10]. In [11], Moustakas *et al.* provide an expression of the mutual information in time varying frequency selective Kronecker channels, using the replica method [12]. This result has been recently proved by Dupuy *et al.* in a yet unpublished work. Dupuy *et al.* then derived the expression of the capacity maximising precoding matrix for the frequency selective channel [13]. A more general frequency selective channel model with non-separable variance profile is studied in [14] by Rashibi *et al.* using alternative tools from free probability theory. Of practical interest is also the theoretical work of Tse [15] on MIMO point-to-point capacity in both uncorrelated and correlated channels, which are validated by ray-tracing simulations.

The main contribution of this paper is summarized in two theorems contributing to the field of random matrix theory and enabling the evaluation of the ergodic rate region of the MIMO-MAC with Kronecker channels. We subsequently derive an iterative water-filling algorithm enabling the description of the boundaries of the rate region by providing an expression of the asymptotically optimal precoders. The remainder of this paper is structured as follows: in Section II, we provide a short summary of the main results and how they apply to multi-user wireless communications. In Section III, the two theorems are introduced, the complete proofs being left to the appendices. In Section IV, the ergodic rate region of the MIMO-MAC is studied. In this section, we introduce our third main result: an iterative water-filling algorithm to describe the boundary of the ergodic rate region of the MIMO-MAC. In Section V, we provide simulation results of the previously derived theoretical formulas. Finally, in Section VI, we give our conclusions.

Notation: In the following, boldface lower-case characters represent vectors, capital boldface characters denote matrices (\mathbf{I}_N is the $N \times N$ identity matrix). X_{ij} denotes the (i, j) entry of \mathbf{X} . The Hermitian transpose is denoted $(\cdot)^H$. The operators $\text{tr } \mathbf{X}$, $|\mathbf{X}|$ and $\|\mathbf{X}\|$ represent the trace, determinant and spectral norm of matrix \mathbf{X} , respectively. The symbol $\mathbb{E}[\cdot]$ denotes expectation. The notation $F^{\mathbf{Y}}$ stands for the (cumulative) distribution function of the eigenvalues of the Hermitian matrix \mathbf{Y} . The function $(x)^+$ equals $\max(x, 0)$ for real x . For F, G two distribution functions, we denote $F \Rightarrow G$ the vague convergence of F to G . The notation $x_n \xrightarrow{\text{a.s.}} x$ denotes the almost sure convergence of the sequence x_n to x . The notation $\|F\|$ for the distribution function F is the supremum norm defined as $\|F\| = \sup_x F(x)$. The symbol $\mathbf{X} \geq 0$ for a square matrix \mathbf{X} means that \mathbf{X} is Hermitian nonnegative definite.

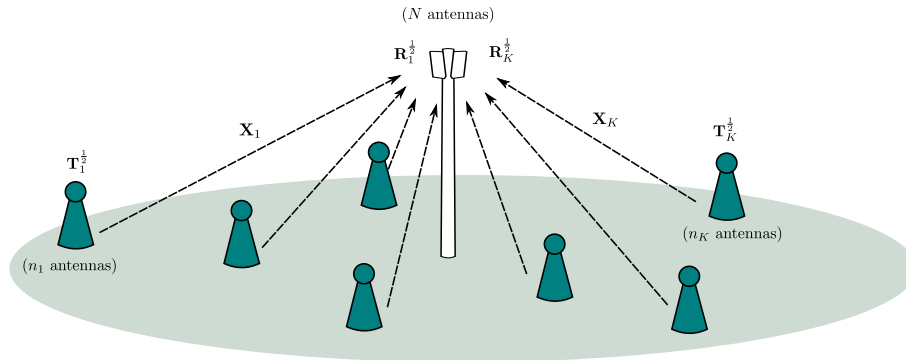


Fig. 1. Multi-antenna multiple access scenario with Kronecker channels.

II. SCOPE AND SUMMARY OF MAIN RESULTS

In this section, we summarise the main results of this article and explain their impact on the study of the effects of channel correlation on the achievable communication rates in the present multi-user framework.

A. General Model

Consider a set of K wireless terminals, equipped with n_1, \dots, n_K antennas respectively, which we refer to as the transmitters, and another device equipped with N antennas, which we call the receiver or the base station. We consider the uplink communication from the terminals to the base station. Denote $\mathbf{H}_k \in \mathbb{C}^{N \times n_k}$ the channel matrix between transmitter k and the receiver. Let \mathbf{H}_k be defined as

$$\mathbf{H}_k = \mathbf{R}_k^{\frac{1}{2}} \mathbf{X}_k \mathbf{T}_k^{\frac{1}{2}}, \quad (1)$$

where $\mathbf{R}_k^{\frac{1}{2}} \in \mathbb{C}^{N \times N}$ and $\mathbf{T}_k^{\frac{1}{2}} \in \mathbb{C}^{n_k \times n_k}$ are nonnegative Hermitian matrices and $\mathbf{X}_k \in \mathbb{C}^{N \times n_k}$ is a realisation of a random matrix with independent Gaussian entries of zero mean and variance $1/n_k$. In this scenario, the matrices \mathbf{T}_k and \mathbf{R}_k model the correlation present in the channel at transmitter k and at the receiver, respectively. This setup is depicted in Figure 1.

It is important to underline that the correlation patterns emerge both from the inter-antenna spacings on the volume-limited transmit and receive radio devices and from the solid angles of transmitted and received signal energy. Even though the transmit antennas emit signals in an isotropic manner, only a limited solid angle of emission is effectively received, and the same holds for the receiver which captures signal energy in a non-isotropic manner. Given this propagation factor, it is clear that the transmit covariance matrices \mathbf{R}_k matrices are not equal for all users. We nonetheless assume physically identical and interchangeable antennas on each device. We therefore claim that the diagonal entries of \mathbf{R}_k and \mathbf{T}_k , i.e., the variance of the channel fading on every antenna, are identical and equal to one, which, along with the normalisation of the Gaussian matrix \mathbf{X}_k , allows for a consistent definition of the SNR. As a consequence, $\text{tr} \mathbf{R}_k = N$ and $\text{tr} \mathbf{T}_k = n_k$. We will see that under these trace constraints the hypotheses made in Theorem 1 are always satisfied, therefore making Theorem 1 valid for all possible figures of

correlation, including strongly correlated patterns. The hypotheses of Theorem 2, used to characterise the ergodic rate region of the MIMO-MAC, require additional mild assumptions, making Theorem 2 valid for most practical models of \mathbf{R}_k and \mathbf{T}_k . These statements are of major importance and rather new since in other contributions, e.g., [5], [13], it is usually assumed that the correlation matrices have uniformly bounded spectral norms across N . Physically this means that only low correlation patterns are allowed, excluding short distances between antennas and small solid angles of energy propagation. The counterpart of this interesting property is a theoretical reduction of the convergence rates of the derived deterministic equivalents, compared to those proposed in [5] and [13].

The rate performance of multi-cell or multi-user communication schemes is connected to the so-called Stieltjes transform and Shannon transform of matrices \mathbf{B}_N of the type

$$\mathbf{B}_N = \sum_{k=1}^K \mathbf{R}_k^{\frac{1}{2}} \mathbf{X}_k \mathbf{T}_k \mathbf{X}_k^H \mathbf{R}_k^{\frac{1}{2}}. \quad (2)$$

We study these matrices using tools from the field of large dimensional random matrix theory [16]. Among these tools, we define the Stieltjes transform $m_N(z)$ of the Hermitian nonnegative definite matrix $\mathbf{B}_N \in \mathbb{C}^{N \times N}$, for $z \in \mathbb{C} \setminus \mathbb{R}^+$, as

$$m_N(z) = \int \frac{1}{\lambda - z} dF_N(\lambda) = \frac{1}{N} \text{tr} (\mathbf{B}_N - z\mathbf{I}_N)^{-1},$$

where F_N denotes the (cumulative) distribution function of the eigenvalues of \mathbf{B}_N . The Stieltjes transform was originally used to characterise the asymptotic distribution of the eigenvalues of large dimensional random matrices [17]. From a wireless communications viewpoint, it can be used to characterise the signal-to-interference plus noise ratio of certain communication models, e.g., [18], [19]. A second interest of the Stieltjes transform in wireless communications is its link to the so-called Shannon transform $\mathcal{V}_N(x)$ of \mathbf{B}_N , that we define for $x > 0$ as

$$\mathcal{V}_N(x) = \frac{1}{N} \log \det \left(\mathbf{I}_N + \frac{1}{x} \mathbf{B}_N \right) = \int_0^{+\infty} \log \left(1 + \frac{\lambda}{x} \right) dF_N(\lambda) = \int_x^{+\infty} \left(\frac{1}{w} - m_N(-w) \right) dw.$$

The Shannon transform is commonly used to provide approximations of capacity expressions in large dimensional systems, e.g., [6]. In the present work, the Shannon transform of \mathbf{B}_N will be used to provide a deterministic approximation of the ergodic achievable rate of the MIMO-MAC.

Before introducing our main results, namely Theorem 1 and Theorem 2, which are rather technical and difficult to fathom without a preliminary explanation, we briefly describe these results in telecommunication terms and their consequences to the multi-user multi-cell communication models at hand.

B. Main results

The main results of this work unfold as follows:

- Theorem 1 provides a deterministic equivalent $m_N^{\circ}(z)$ for the Stieltjes transform $m_N(z)$ of \mathbf{B}_N , under the assumption that N and n_k grow large with the same order of magnitude and the sequences of distribution functions $\{F^{\mathbf{T}_k}\}_{n_k}$ and $\{F^{\mathbf{R}_k}\}_N$ form *tight sequences* [20]. This is, we provide an approximation $m_N^{\circ}(z)$ of $m_N(z)$ which can be expressed without reference to the random \mathbf{X}_k matrices and which is almost surely

asymptotically exact when $N \rightarrow \infty$. The tightness hypothesis is the key assumption that allows degenerated \mathbf{R}_k and \mathbf{T}_k matrices to be valid in our framework, and that therefore allows us to study strongly correlated channel models;

- Theorem 2 provides a deterministic equivalent $\mathcal{V}_N^{\circ}(x)$ for the Shannon transform $\mathcal{V}_N(x)$ of \mathbf{B}_N . In this theorem, the assumptions on the \mathbf{R}_k and \mathbf{T}_k matrices are only slightly more constraining and of marginal importance for practical purposes. In particular, Theorem 2 theoretically allows the largest eigenvalues of \mathbf{T}_k or \mathbf{R}_k to grow linearly with N , as the number of antennas increases, as long as the number of these large eigenvalues is of order $o(N)$;
- based on Theorem 2, the precoders that maximise the deterministic equivalent of the ergodic sum rate of the MIMO-MAC are computed. Those precoders have the following properties:
 - the eigenspace of the precoder for user k coincides with the eigenspace of the transmit channel correlation matrix at user k ;
 - the eigenvalues of the precoder for user k are the solution of a water-filling algorithm;
 - as the system dimensions grow large, the mutual information achieved using these precoders becomes asymptotically close to the channel capacity.

The major practical interest of Theorems 1 and 2 lies in the possibility to analyze mutual information expressions for multi-dimensional channels, not as the averaged of stochastic variables depending on the random matrices \mathbf{X}_k but as approximated deterministic expressions which do no longer feature the matrices \mathbf{X}_k . The study of those quantities is in general simpler than the study of the averaged stochastic expressions, which leads here to a simple derivation of the approximated rate optimal precoders.

In the next section, we introduce our theoretical results, whose proofs are left to the appendices.

III. MATHEMATICAL PRELIMINARIES

In this section, we first introduce Theorem 1, which provides a deterministic equivalent for the Stieltjes transform of matrices \mathbf{B}_N defined in (2). A deterministic equivalent for the Shannon transform of \mathbf{B}_N is then provided in Theorem 2, before we discuss in detail how this last result can be used to characterise the performance of the MIMO-MAC with strong channel correlation patterns.

A. Main results

Theorem 1: Let K, N, n_1, \dots, n_K be positive integers and let

$$\mathbf{B}_N = \sum_{k=1}^K \mathbf{R}_k^{\frac{1}{2}} \mathbf{X}_k \mathbf{T}_k \mathbf{X}_k^H \mathbf{R}_k^{\frac{1}{2}} + \mathbf{S} \quad (3)$$

be an $N \times N$ matrix with the following hypothesis for all $k \in \{1, \dots, K\}$:

- 1) $\mathbf{X}_k \in \mathbb{C}^{N \times n_k}$ has i.i.d. entries $\frac{1}{\sqrt{n_k}} X_{ij}^k$, such that $\mathbb{E}[|X_{11}^k - \mathbb{E}X_{11}^k|^2] = 1$;
- 2) $\mathbf{R}_k^{\frac{1}{2}} \in \mathbb{C}^{N \times N}$ is a Hermitian nonnegative square root of the nonnegative definite Hermitian matrix \mathbf{R}_k ;

- 3) $\mathbf{T}_k = \text{diag}(\tau_1, \dots, \tau_{n_k})$ with $\tau_i \geq 0$ for all i ;
- 4) the sequences $\{F^{\mathbf{T}_k}\}_{n_k}$ and $\{F^{\mathbf{R}_k}\}_N$ are tight;¹
- 5) $\mathbf{S} \in \mathbb{C}^{N \times N}$ is Hermitian nonnegative definite;
- 6) there exist $b > a > 0$ for which

$$a < \min_k \liminf_N c_k \leq \max_k \limsup_N c_k < b, \quad (4)$$

with $c_k = N/n_k$.

Also denote, for $z \in \mathbb{C} \setminus \mathbb{R}^+$, $m_N(z) = \int \frac{1}{\lambda - z} dF_N(\lambda)$, the Stieltjes transform of \mathbf{B}_N . Then, as all n_k and N grow large, with ratio c_k ,

$$m_N(z) - m_N^\circ(z) \xrightarrow{\text{a.s.}} 0, \quad (5)$$

where

$$m_N^\circ(z) = \frac{1}{N} \text{tr} \left(\mathbf{S} + \sum_{k=1}^K \int \frac{\tau_k dF^{\mathbf{T}_k}(\tau_k)}{1 + c_k \tau_k e_k(z)} \mathbf{R}_k - z \mathbf{I}_N \right)^{-1}$$

and the functions $e_i(z)$, $i \in \{1, \dots, K\}$, form the unique solution to the K equations

$$e_i(z) = \frac{1}{N} \text{tr} \mathbf{R}_i \left(\mathbf{S} + \sum_{k=1}^K \int \frac{\tau_k dF^{\mathbf{T}_k}(\tau_k)}{1 + c_k \tau_k e_k(z)} \mathbf{R}_k - z \mathbf{I}_N \right)^{-1} \quad (6)$$

such that $\text{sgn}(\Im[e_i(z)]) = \text{sgn}(\Im[z])$ when $\Im[z] \neq 0$ and such that $e_i(z) > 0$ when z is real and negative.

Moreover, for any $\varepsilon > 0$, the convergence of Equation (5) is uniform over any region of \mathbb{C} bounded by a contour interior to

$$\mathbb{C} \setminus (\{z : |z| \leq \varepsilon\} \cup \{z = x + iv : x > 0, |v| \leq \varepsilon\}). \quad (7)$$

For all N , the function m_N° is the Stieltjes transform of a distribution function F_N° . Denoting F_N the empirical eigenvalue distribution function of \mathbf{B}_N , we finally have

$$F_N - F_N^\circ \Rightarrow 0 \quad (8)$$

weakly and almost surely as $N \rightarrow \infty$.

Proof: The proof of Theorem 1 is deferred to Appendix A. ■

A few technical remarks are of order at his point.

Remark 1: In her PhD dissertation [21], Zhang derives an expression of the limiting eigenvalue distribution for the simpler case where $K = 1$ and $\mathbf{S} = 0$ but \mathbf{T}_1 is not constrained to be diagonal. Her work also uses a method based on the Stieltjes transform. Based on [21], it seems to the authors that Theorem 1 could well be extended to non-diagonal \mathbf{T}_k . However, proving so requires involved calculus, which we did not perform. Also, in [22], using the same techniques as in the proof provided in Appendix A, Silverstein *et al.* do not assume that the matrices \mathbf{T}_k are nonnegative definite. Our result could be extended to this less stringent requirement on the central \mathbf{T}_k

¹this is, for all $\varepsilon > 0$, there exists $M > 0$ such that $F^{\mathbf{T}_k}(M) > 1 - \varepsilon$ and $F^{\mathbf{R}_k}(M) > 1 - \varepsilon$ for all n_k, N . See e.g., [20] for more details.

Define $\varepsilon > 0$, the convergence threshold and $n \geq 0$, the iteration step. For all $k \in \{1, \dots, K\}$, set $e_j^0 = -1/z$ and $e_j^{-1} = \infty$.

while $\max_j \{|e_j^n - e_j^{n-1}|\} > \varepsilon$ **do**

for $k \in \{1, \dots, K\}$ **do**

Compute

$$e_k^{n+1} = \frac{1}{N} \operatorname{tr} \mathbf{R}_k \left(\mathbf{S} + \sum_{j=1}^K \int \frac{\tau_j dF^{\mathbf{T}_j}(\tau_j)}{1 + c_j \tau_j e_j^n} \mathbf{R}_j - z \mathbf{I}_N \right)^{-1} \quad (9)$$

end for

Assign $n \leftarrow n + 1$

end while

TABLE I

FIXED-POINT ALGORITHM CONVERGING TO THE SOLUTION OF (6)

matrices, although in this case Theorem 1 does not hold for z real negative. For application purposes though, it is fundamental that the Stieltjes transform of \mathbf{B}_N exist for $z \in \mathbb{R}^-$.

We now claim that, under proper initialisation, for $z \in \mathbb{C} \setminus \mathbb{R}^+$, a classical fixed-point algorithm converges surely to the solution of (6).

Proposition 1: For $z \in \mathbb{C} \setminus \mathbb{R}^+$, the output $\{e_1^n, \dots, e_K^n\}$ of fixed-point algorithm described in Table I converges surely to the unique solution $\{e_1(z), \dots, e_K(z)\}$ of (6), such that $\operatorname{sgn}(\Im[e_i(z)]) = \operatorname{sgn}(\Im[z])$ when $\Im[z] \neq 0$ and such that $e_i(z) > 0$ when $z < 0$, for all i .

Proof: The proof of Proposition 1, inspired by the work of Dupuy *et al.* [13] in the context of frequency selective channel models, is provided in Appendix E. ■

Different hypotheses will be used in the applications of Theorem 1 provided in Section IV. For practical reasons, we will in particular need the entries of \mathbf{X}_k will be Gaussian, the matrices \mathbf{T}_k to be non-diagonal and $\mathbf{S} = 0$. This entails the following corollary:

Corollary 1: Let K, N, n_1, \dots, n_K be positive integers and let

$$\mathbf{B}_N = \sum_{k=1}^K \mathbf{R}_k^{\frac{1}{2}} \mathbf{X}_k \mathbf{T}_k \mathbf{X}_k^H \mathbf{R}_k^{\frac{1}{2}} \quad (10)$$

be an $N \times N$ matrix with the following hypothesis for all $k \in \{1, \dots, K\}$:

- 1) $\mathbf{X}_k \in \mathbb{C}^{N \times n_k}$ has i.i.d. Gaussian entries $\frac{1}{\sqrt{n_k}} X_{ij}^k$, with $\mathbb{E}[X_{11}^k] = 0$ and $\mathbb{E}[|X_{11}^k|^2] = 1$;
- 2) $\mathbf{R}_k^{\frac{1}{2}} \in \mathbb{C}^{N \times N}$ is a Hermitian nonnegative square root of the nonnegative definite Hermitian matrix \mathbf{R}_k ;
- 3) $\mathbf{T}_k \in \mathbb{C}^{n_k \times n_k}$ is a nonnegative definite Hermitian matrix;
- 4) $\{F^{\mathbf{T}_k}\}_{n_k}$ and $\{F^{\mathbf{R}_k}\}_N$ form tight sequences;
- 5) there exist $b > a > 0$ for which

$$a < \min_k \liminf_N c_k \leq \max_k \limsup_N c_k < b \quad (11)$$

with $c_k = N/n_k$.

Also denote, for $x > 0$, $m_N(-x) = \frac{1}{N} \text{tr}(\mathbf{B}_N + x\mathbf{I}_N)^{-1}$. Then, as all N and n_k grow large (while K is fixed)

$$m_N(-x) - m_N^\circ(-x) \xrightarrow{\text{a.s.}} 0,$$

where

$$m_N^\circ(-x) = \frac{1}{N} \text{tr} \left(x \left[\mathbf{I}_N + \sum_{k=1}^K \delta_k(-x) \mathbf{R}_k \right] \right)^{-1}$$

and the set of functions $\{e_i(-x), \delta_i(-x)\}$, $i \in \{1, \dots, K\}$, form the unique solution to the equations

$$e_i(-x) = \frac{1}{N} \text{tr} \mathbf{R}_i \left(x \left[\mathbf{I}_N + \sum_{k=1}^K \delta_k(-x) \mathbf{R}_k \right] \right)^{-1}$$

$$\delta_i(-x) = \frac{1}{n_i} \text{tr} \mathbf{T}_i (x [\mathbf{I}_{n_i} + c_i e_i(-x) \mathbf{T}_i])^{-1},$$

such that $e_i(-x) > 0$ for all i .

Proof: Since the \mathbf{X}_k are Gaussian, the joint distribution of the entries of $\mathbf{X}_k \mathbf{U}$ coincides with that of \mathbf{X}_k , for \mathbf{U} any $n_k \times n_k$ unitary matrix. Therefore, $\mathbf{X}_k \mathbf{T}_k \mathbf{X}_k^H$ in Theorem 1 can be substituted by $\mathbf{X}_k (\mathbf{U} \mathbf{T}_k \mathbf{U}^H) \mathbf{X}_k^H$ without compromising the final result. As a consequence, the \mathbf{T}_k can be taken nonnegative definite Hermitian and the result of Theorem 1 holds. It then suffices to replace $\delta_i(-x)$ in the expression of $e_i(-x)$ to fall back on the result of Theorem 1. ■

The deterministic equivalent of the Stieltjes transform m_N of \mathbf{B}_N is then extended to a deterministic equivalent of the Shannon transform of \mathbf{B}_N in the following result:

Theorem 2: Let $x > 0$ and \mathbf{B}_N be a random Hermitian matrix as defined in Corollary 1 with the following additional assumptions:

- 1) there exists $\alpha > 0$ and a sequence r_1, r_2, \dots , such that, for all N ,

$$\max_{1 \leq k \leq K} \max(\lambda_{r_N+1}^{\mathbf{T}_k}, \lambda_{r_N+1}^{\mathbf{R}_k}) \leq \alpha,$$

where $\lambda_1^{\mathbf{Z}} \geq \dots \geq \lambda_N^{\mathbf{Z}}$ denote the ordered eigenvalues of the $N \times N$ matrix \mathbf{Z} .

- 2) denoting b_N an upper-bound on the spectral norm of the \mathbf{T}_k and \mathbf{R}_k , $k \in \{1, \dots, K\}$, and $\beta > 0$ a constant such that $\beta > \frac{Kb}{a}(1 + \sqrt{a})^2$, $a_N = b_N^2 \beta$ satisfies

$$r_N \log(1 + a_N/x) = o(N).$$

Then, for large N , n_k , the Shannon transform $\mathcal{V}_N(x) = \int \log(1 + \frac{1}{x}\lambda) dF_N(\lambda)$ of \mathbf{B}_N , satisfies

$$\mathcal{V}_N(x) - \mathcal{V}_N^\circ(x) \xrightarrow{\text{a.s.}} 0;$$

where

$$\mathcal{V}_N^\circ(x) = \sum_{k=1}^K \frac{1}{N} \log \det (\mathbf{I}_{n_k} + c_k e_k(-x) \mathbf{T}_k) + \frac{1}{N} \log \det \left(\mathbf{I}_N + \sum_{k=1}^K \delta_k(-x) \mathbf{R}_k \right) - x \sum_{k=1}^K \delta_k(-x) e_k(-x). \quad (12)$$

Proof: The proof of Theorem 2 is provided in Appendix B. ■

Note that this last result is consistent both with [6] when the transmission channels are i.i.d. Gaussian and with [23] when $K = 2$. This result is also similar in nature to the expressions obtained in [5] for the multi-antenna Rician channel model and with [11] in the case of frequency selective channels. We point out that the expressions obtained in [11], [13] and [9], when the entries of the \mathbf{X}_k matrices are Gaussian distributed, suggest a faster convergence rate of the deterministic equivalent of the Stieltjes and Shannon transforms than the one obtained here. Indeed, while we show here a convergence of order $o(1)$ (which is in fact refined to $o(1/\log^k N)$ for any k in Appendix A), in those works the convergence is proved to be of order $O(1/N^2)$.

However, contrary to the above contributions, we allow the \mathbf{R}_k and \mathbf{T}_k matrices to be more general than uniformly bounded in spectral norm. This is thoroughly discussed in the section below.

B. Kronecker channel with strong correlation patterns

Theorem 1 and Corollary 1 require $\{F^{\mathbf{R}_k}\}_N$ and $\{F^{\mathbf{T}_k}\}_{n_k}$ to form tight sequences. Remark that, because of the trace constraint $\frac{1}{N} \text{tr} \mathbf{R}_k = 1$, all sequences $\{F^{\mathbf{R}_k}\}_N$ are necessarily tight (the same reasoning naturally holds for \mathbf{T}_k). Indeed, given $\varepsilon > 0$, take $M = 2/\varepsilon$; $N[1 - F^{\mathbf{R}_k}(M)]$ is the number of eigenvalues in \mathbf{R}_k larger than $2/\varepsilon$, which is necessarily less than or equal to $N\varepsilon/2$ from the trace constraint, leading to $1 - F^{\mathbf{R}_k}(M) \leq \varepsilon/2$ and then $F^{\mathbf{R}_k}(M) \geq 1 - \varepsilon/2 > 1 - \varepsilon$. The same naturally holds for the \mathbf{T}_k matrices. Observe now that Condition 2) in Theorem 2 requires a stronger assumption on the correlation matrices. Under the trace constraint, a sufficient assumption for Condition 2) is that there exists $\alpha > 0$, such that the number of eigenvalues in \mathbf{R}_k greater than α is of order $o(N/\log N)$. This is a mild assumption, which may not be verified for some very specific choices of $\{F^{\mathbf{R}_k}\}_N$.² Nonetheless, most conventional models for the \mathbf{R}_k and \mathbf{T}_k , even when showing strong correlation properties, satisfy the assumptions of Theorem 2. We mention in particular the following examples:

- if all \mathbf{R}_k and \mathbf{T}_k have uniformly bounded spectral norm, then there exists $\alpha > 0$ such that all eigenvalues of \mathbf{R}_k and \mathbf{T}_k are less than α for all N . This implies $r_N = 0$ for all N and therefore the condition is trivially satisfied. Our model is therefore compatible with loosely correlated antenna structures;
- in contrast, when antennas are densely packed on a volume-limited device, the correlation matrices \mathbf{R}_k and \mathbf{T}_k tend to be asymptotically of finite rank, see e.g. [24] in the case of a dense circular array. That is, for any given $\alpha > 0$, for all large N , the number r_N of eigenvalues greater than α is finite, while a_N defined in Theorem 2 is of order N^2 . This implies $r_N \log(1 + a_N/x) = O(\log N) = o(N)$ and therefore volume-limited devices with densely packed antennas are consistent with our framework;

²As a counter-example, take $N = 2^p + 1$ and the eigenvalues of \mathbf{R}_k to be

$$2^{p-1}, \underbrace{p, \dots, p}_{\frac{2^{p-1}}{p}}, \underbrace{0, \dots, 0}_{2^p - \frac{2^{p-1}}{p}}.$$

The largest eigenvalue is of order N so that a_N is of order N^2 , and the number r_N of eigenvalues larger than any $\alpha > 0$ for N large is of order $\frac{2^{p-1}}{p} \sim \frac{N}{\log(N)}$. Therefore $r_N \log(1 + a_N/x) = O(N)$ here.

- for one, two or three dimensional antenna arrays with neighbors separated by half the wavelength as discussed by Moustakas *et al.* in [25], the correlation figures have a peculiar behaviour. In a linear array of antenna, $O(N)$ eigenvalues are of order of magnitude $O(1)$, the remaining eigenvalues being small. In a two-dimensional grid of antennas, $O(\sqrt{N})$ eigenvalues are of order $O(\sqrt{N})$, the remaining eigenvalues being close to zero. Finally, in a three-dimensional parallelepiped of antennas, $O(N^{\frac{2}{3}})$ eigenvalues are of order $O(N^{\frac{1}{3}})$, the remaining eigenvalues being close to 0 also. As such, in the p -dimensional scenario, we can approximate r_N by $N^{\frac{p-1}{p}}$, a_N by $N^{\frac{2}{p}}$ and we have

$$r_N \log(1 + a_N/x) \sim N^{\frac{p-1}{p}} \log N = o(N),$$

so that the multi-dimensional antenna arrays with close antennas separated by half the wavelength also satisfy the hypotheses of Theorem 2.

As a consequence, a wide scope of antenna correlation models enter our deterministic equivalent framework, which comes again at the price of a slower theoretical convergence of the difference $\mathcal{V}_N - \mathcal{V}_N^\circ$.

We now move to practical applications of the above results and more specifically to the determination of the ergodic rate region of the MIMO-MAC.

IV. RATE REGION OF THE MIMO-MAC

In this section, we successively apply Theorem 2 to approximate the ergodic mutual information for all deterministic precoders, and then we determine the precoders that maximise this approximated mutual information. This gives an approximation of all points on the boundary of the MIMO-MAC rate region. We also introduce an iterative power allocation algorithm to obtain explicitly the optimal precoders.

A. Deterministic equivalent of the mutual information

Consider the wireless multiple access channel as described in Section II and depicted in Figure 1. We denote $c_k = N/n_k$ the ratio between the number of antennas at the receive base station and the number of transmit antennas of user k . Denote $\mathbf{s}_k \in \mathbb{C}^{n_k}$ the Gaussian signal transmitted by user k , such that $\mathbb{E}[\mathbf{s}_k] = 0$ and $\mathbb{E}[\mathbf{s}_k \mathbf{s}_k^H] = \mathbf{P}_k$, with $\frac{1}{n_k} \text{tr} \mathbf{P}_k \leq P_k$ where P_k is the total power of transmitter k , $\mathbf{y} \in \mathbb{C}^N$ the signal received at the base station and \mathbf{n} the additive white Gaussian noise of variance $\mathbb{E}[\mathbf{n} \mathbf{n}^H] = \sigma^2 \mathbf{I}_N$. We recall that the Kronecker channel between user k and the base station is denoted $\mathbf{H}_k = \mathbf{R}_k^{\frac{1}{2}} \mathbf{X}_k \mathbf{T}_k^{\frac{1}{2}}$, with the entries of $\mathbf{X}_k \in \mathbb{C}^{N \times n_k}$ Gaussian independent of zero mean and variance $1/n_k$ and $\mathbf{R}_k, \mathbf{T}_k$ deterministic. The received vector \mathbf{y} is therefore given by

$$\mathbf{y} = \sum_{k=1}^K \mathbf{H}_k \mathbf{s}_k + \mathbf{n}.$$

Suppose that the \mathbf{H}_k channels are varying fast and that the transmitters in the MAC only have statistical channel state information about the \mathbf{H}_k in the sense that user k only knows the long term statistics $\mathbf{R}_1, \dots, \mathbf{R}_K$ and $\mathbf{T}_1, \dots, \mathbf{T}_K$. In this case, for a noise variance equal to σ^2 , the per-antenna ergodic MIMO-MAC rate region \mathcal{C}_{MAC}

is given by [26]

$$\mathcal{C}_{\text{MAC}} = \bigcup_{\substack{\frac{1}{n_i} \text{tr}(\mathbf{P}_i) \leq P_i \\ \mathbf{P}_i \geq 0 \\ i=1, \dots, K}} \left\{ \{R_i, 1 \leq i \leq K\} : \sum_{i \in \mathcal{S}} R_i \leq \text{E}\mathcal{V}_N(\mathbf{P}_{i_1}, \dots, \mathbf{P}_{i_{|\mathcal{S}|}}; \sigma^2), \forall \mathcal{S} \subset \{1, \dots, K\} \right\},$$

with the expectation taken over the joint random variable $(\mathbf{X}_1, \dots, \mathbf{X}_K)$, $\mathcal{S} = \{i_1, \dots, i_{|\mathcal{S}|}\}$, and where we introduced the notation

$$\mathcal{V}_N(\mathbf{P}_{i_1}, \dots, \mathbf{P}_{i_{|\mathcal{S}|}}; \sigma^2) \triangleq \frac{1}{N} \log \det \left(\mathbf{I}_N + \frac{1}{\sigma^2} \sum_{i \in \mathcal{S}} \mathbf{H}_i \mathbf{P}_i \mathbf{H}_i^H \right).$$

Now, assuming the \mathbf{T}_k , \mathbf{P}_k and \mathbf{R}_k satisfy the hypotheses of Theorem 2, we have

$$\mathcal{V}_N(\mathbf{P}_{i_1}, \dots, \mathbf{P}_{i_{|\mathcal{S}|}}; x) - \mathcal{V}_N^\circ(\mathbf{P}_{i_1}, \dots, \mathbf{P}_{i_{|\mathcal{S}|}}; x) \rightarrow 0,$$

as $N, n_{i_1}, \dots, n_{i_{|\mathcal{S}|}}$ grow large for some sequence $\{\mathcal{V}_N^\circ(\mathbf{P}_{i_1}, \dots, \mathbf{P}_{i_{|\mathcal{S}|}}; x)\}_N$, on a subset of measure 1 of the probability space Ω that engenders $(\mathbf{X}_{i_1}, \dots, \mathbf{X}_{i_{|\mathcal{S}|}})$. Integrating this expression over Ω therefore leads to

$$\text{E}\mathcal{V}_N^\circ(\mathbf{P}_{i_1}, \dots, \mathbf{P}_{i_{|\mathcal{S}|}}; x) - \mathcal{V}_N^\circ(\mathbf{P}_{i_1}, \dots, \mathbf{P}_{i_{|\mathcal{S}|}}; x) \rightarrow 0.$$

We can therefore apply Theorem 2 to determine the ergodic rate region \mathcal{C}_{MAC} of the MIMO-MAC. We specifically have

$$\begin{aligned} & \text{E}\mathcal{V}_N(\mathbf{P}_{i_1}, \dots, \mathbf{P}_{i_{|\mathcal{S}|}}; \sigma^2) - \\ & \left[\sum_{k \in \mathcal{S}} \frac{1}{N} \log \det (\mathbf{I}_{n_k} + c_k e_k(-\sigma^2) \mathbf{T}_k \mathbf{P}_k) + \frac{1}{N} \log \det \left(\mathbf{I}_N + \sum_{k \in \mathcal{S}} \delta_k(-\sigma^2) \mathbf{R}_k \right) - \sigma^2 \sum_{k \in \mathcal{S}} \delta_k(-\sigma^2) e_k(-\sigma^2) \right] \rightarrow 0, \end{aligned} \quad (13)$$

with $e_i(-\sigma^2)$ and $\delta_i(-\sigma^2)$ the unique positive solutions of

$$\begin{aligned} e_i(-\sigma^2) &= \frac{1}{N} \text{tr} \mathbf{R}_i \left(\sigma^2 \left[\mathbf{I}_N + \sum_{k \in \mathcal{S}} \delta_k(-\sigma^2) \mathbf{R}_k \right] \right)^{-1} \\ \delta_i(-\sigma^2) &= \frac{1}{n_i} \text{tr} \mathbf{T}_i^{\frac{1}{2}} \mathbf{P}_i \mathbf{T}_i^{\frac{1}{2}} \left(\sigma^2 \left[\mathbf{I}_{n_i} + c_i e_i(-\sigma^2) \mathbf{T}_i^{\frac{1}{2}} \mathbf{P}_i \mathbf{T}_i^{\frac{1}{2}} \right] \right)^{-1}. \end{aligned} \quad (14)$$

This provides a deterministic equivalent for all points in the MIMO-MAC rate region, i.e., for all $\mathbf{P}_1, \dots, \mathbf{P}_K$ precoders.

B. Rate maximisation

Now we wish to determine for which precoders the boundary of the MIMO-MAC rate region is reached. This requires here to determine the rate optimal precoding matrices $\mathbf{P}_{i_1}, \dots, \mathbf{P}_{i_{|\mathcal{S}|}}$, for all $\mathcal{S} \subset \{1, \dots, K\}$. To this end, we first need the following result:

Proposition 2: If at least one of the correlation matrices \mathbf{T}_k , $k \in \mathcal{S}$, is invertible, then \mathcal{V}_N° is a strictly concave function in $\mathbf{P}_{i_1}, \dots, \mathbf{P}_{i_{|\mathcal{S}|}}$.

Proof: The proof of Proposition 2 is provided in Appendix C. ■

Define $\eta > 0$ the convergence threshold and $l \geq 0$ the iteration step. At step $l = 0$, for all $k \in \mathcal{S}$, $i \in \{1, \dots, n_k\}$, set $p_{k,i}^0 = P_k$.

while $\max_{k,i} \{|p_{k,i}^l - p_{k,i}^{l-1}|\} > \eta$ **do**

For $k \in \mathcal{S}$, define e_k^{l+1} as the solution of (6) for $z = -\sigma^2$ and \mathbf{P}_k with eigenvalues $p_{k,1}^l, \dots, p_{k,n_k}^l$, obtained from the fixed-point algorithm of Table I.

for $k \in \mathcal{S}$ **do**

for $i = 1, \dots, n_k$ **do**

Set $p_{k,i}^{l+1} = \left(\mu_k - \frac{1}{c_k e_k^{l+1} t_{ki}} \right)^+$, with μ_k such that $\frac{1}{n_k} \text{tr } \mathbf{P}_k = P_k$.

end for

end for

Assign $l \leftarrow l + 1$

end while

TABLE II
ITERATIVE WATER-FILLING ALGORITHM

Without loss of generality, for any k , since the \mathbf{X}_k matrices are standard Gaussian, and therefore of unitarily invariant joint distribution, \mathbf{T}_k can be assumed diagonal. If \mathbf{T}_k is not of full rank then it can be reduced into a matrix of smaller size, such that the resulting matrix is invertible, without changing the problem at hand. We therefore assume all \mathbf{T}_k matrices to be of full rank from now on. From Proposition 2, we then immediately prove that the $|\mathcal{S}|$ -ary set of matrices $\{\mathbf{P}_{i_1}^\circ, \dots, \mathbf{P}_{i_{|\mathcal{S}|}}^\circ\}$ which maximises the deterministic equivalent of the ergodic sum rate over the set \mathcal{S} is unique. In a very similar way as in [9], we then show that the matrices \mathbf{P}_k° , $k \in \mathcal{S}$, have the following properties:

Proposition 3: For every $k \in \mathcal{S}$, denote $\mathbf{T}_k = \mathbf{U}_k \bar{\mathbf{T}}_k \mathbf{U}_k^H$ the spectral decomposition of \mathbf{T}_k with \mathbf{U}_k unitary and $\bar{\mathbf{T}}_k = \text{diag}(t_{k,1}, \dots, t_{k,n_k})$. Then the precoders $\mathbf{P}_{i_1}^\circ, \dots, \mathbf{P}_{i_{|\mathcal{S}|}}^\circ$ which maximise the right-hand side of (13) satisfy:

- 1) $\mathbf{P}_k^\circ = \mathbf{U}_k \bar{\mathbf{P}}_k^\circ \mathbf{U}_k^H$, with $\bar{\mathbf{P}}_k^\circ$ diagonal, i.e., the eigenspace of \mathbf{P}_k° is the same as the eigenspace of \mathbf{T}_k ;
- 2) denoting, for all k , $e_k^\circ = e_k(-\sigma^2)$ as in (14) for $\mathbf{P}_k = \mathbf{P}_k^\circ$, the i^{th} diagonal entry $p_{k,i}^\circ$ of $\bar{\mathbf{P}}_k^\circ$ satisfies:

$$p_{k,i}^\circ = \left(\mu_k - \frac{1}{c_k e_k^\circ t_{ki}} \right)^+, \quad (15)$$

where the μ_k are evaluated such that $\frac{1}{n_k} \text{tr } \mathbf{P}_k = P_k$.

In Table II, we provide an iterative water-filling algorithm to obtain the p_{ki}° .

Proof: The proof of Proposition 3 is provided in Appendix D. ■

Remark 2: In [9], it is proved that the convergence of this algorithm implies its convergence towards the correct limit. The line of reasoning in [9] can be directly adapted to the current situation so that, if the iterative water-filling algorithm of Table II converges, then $\mathbf{P}_{i_1}, \dots, \mathbf{P}_{i_{|\mathcal{S}|}}$ converge to the matrices $\mathbf{P}_{i_1}^\circ, \dots, \mathbf{P}_{i_{|\mathcal{S}|}}^\circ$. However, similar to [9], it is difficult to prove the sure convergence of the water-filling algorithm. Nonetheless, extensive simulations suggest that convergence is always attained.

For the set \mathcal{S} under consideration, denote now $\mathbf{P}_{i_1}^*, \dots, \mathbf{P}_{i_{|\mathcal{S}|}}^*$ the true sum rate maximising precoders. Then, if $\mathbf{P}_{i_1}^\circ, \dots, \mathbf{P}_{i_{|\mathcal{S}|}}^\circ$ and $\mathbf{P}_{i_1}^*, \dots, \mathbf{P}_{i_{|\mathcal{S}|}}^*$ are such that Condition 1) of Theorem 2 is satisfied with the sets $\{\mathbf{T}_1, \dots, \mathbf{T}_K\}$

and $\{\mathbf{R}_1, \dots, \mathbf{R}_K\}$ replaced by $\{\mathbf{T}_{i_1} \mathbf{P}_{i_1}^\circ, \dots, \mathbf{T}_{i_{|S|}} \mathbf{P}_{i_{|S|}}^\circ\}$ (or $\{\mathbf{T}_{i_1} \mathbf{P}_{i_1}^*, \dots, \mathbf{T}_{i_{|S|}} \mathbf{P}_{i_{|S|}}^*\}$) and $\{\mathbf{R}_{i_1}, \dots, \mathbf{R}_{i_{|S|}}\}$, respectively, we have from Theorem 2

$$\begin{aligned} \mathcal{V}_N(\mathbf{P}_{i_1}^*, \dots, \mathbf{P}_{i_{|S|}}^*; \sigma^2) - \mathcal{V}_N(\mathbf{P}_{i_1}^\circ, \dots, \mathbf{P}_{i_{|S|}}^\circ; \sigma^2) &= \left(\mathcal{V}_N(\mathbf{P}_{i_1}^*, \dots, \mathbf{P}_{i_{|S|}}^*; \sigma^2) - \mathcal{V}_N^\circ(\mathbf{P}_{i_1}^*, \dots, \mathbf{P}_{i_{|S|}}^*; \sigma^2) \right) \\ &+ \left(\mathcal{V}_N^\circ(\mathbf{P}_{i_1}^*, \dots, \mathbf{P}_{i_{|S|}}^*; \sigma^2) - \mathcal{V}_N^\circ(\mathbf{P}_{i_1}^\circ, \dots, \mathbf{P}_{i_{|S|}}^\circ; \sigma^2) \right) \\ &+ \left(\mathcal{V}_N^\circ(\mathbf{P}_{i_1}^\circ, \dots, \mathbf{P}_{i_{|S|}}^\circ; \sigma^2) - \mathcal{V}_N(\mathbf{P}_{i_1}^\circ, \dots, \mathbf{P}_{i_{|S|}}^\circ; \sigma^2) \right), \end{aligned}$$

where both right-hand side differences of the type $\mathcal{V}_N - \mathcal{V}_N^\circ$ tend to zero, while the left-hand side term is positive by definition of \mathbf{P}_k^* and the remaining right-hand side term is negative by definition of the \mathbf{P}_k° . This finally ensures that

$$\mathcal{V}_N(\mathbf{P}_{i_1}^*, \dots, \mathbf{P}_{i_{|S|}}^*; \sigma^2) - \mathcal{V}_N(\mathbf{P}_{i_1}^\circ, \dots, \mathbf{P}_{i_{|S|}}^\circ; \sigma^2) \rightarrow 0,$$

as $N, n_{i_1}, \dots, n_{i_{|S|}}$ grow large with uniformly bounded ratios. Therefore, the mutual information obtained based on the precoders $\mathbf{P}_{i_1}^\circ, \dots, \mathbf{P}_{i_{|S|}}^\circ$ is asymptotically close to the capacity achieved with the ideal precoders $\mathbf{P}_{i_1}^*, \dots, \mathbf{P}_{i_{|S|}}^*$. Finally, if, for all sets $S \subset \{1, \dots, K\}$, the matrices $\mathbf{T}_k, \mathbf{R}_k$ and the resulting $\mathbf{P}_k^*, \mathbf{P}_k^\circ$ satisfy the mild conditions of Theorem 2, then all points of the boundary of the MIMO-MAC rate region can be given a deterministic equivalent.

This concludes this application section. In the following section, we provide simulation results that confirm the accuracy of the deterministic equivalents as well as the validity of the hypotheses made on the \mathbf{T}_k and \mathbf{R}_k matrices.

V. SIMULATIONS AND RESULTS

In the following, we apply the results obtained in Section IV to provide comparative simulation results between ergodic rate regions, sum rates and their respective deterministic equivalents, for non negligible channel correlations on both communication sides. We provide simulation results in the context of a two-user MIMO-MAC, with N antennas at the base station and $n_1 = n_2$ antennas at the user terminals. The antennas are placed on a possibly multi-dimensional array, antenna i being located at $\mathbf{x}_i \in \mathbb{R}^3$. We further assume that both terminals are physically identical. To model the transmit and receive correlation matrices, we consider both the effect of the distances between adjacent antennas at the user terminals and at the base station, and the effect of the solid angles of effective energy transmission and reception. We assume a channel model where signals are transmitted and received isotropically in the vertical direction, but transmitted and received under an angle π in the horizontal direction. We then model the entries of the correlation matrices from a natural extension of Jakes model [27] with privileged direction of signal departure and arrival. Denoting λ the transmit signal wavelength, $T_{1_{ab}}$, the entry (a, b) of the matrix \mathbf{T}_1 , is

$$T_{1_{ab}} = \int_{\theta_{\min}^{(\mathbf{T}_1)}}^{\theta_{\max}^{(\mathbf{T}_1)}} \exp\left(\frac{2\pi i}{\lambda} \|\mathbf{x}_a - \mathbf{x}_b\| \cos(\theta)\right) d\theta$$

with $[\theta_{\min}^{(\mathbf{T}_1)}, \theta_{\max}^{(\mathbf{T}_1)}]$ the effective horizontal directions of signal propagation. With similar notations for the other correlation matrices, we choose $\theta_{\min}^{(\mathbf{T}_1)} = 0, \theta_{\max}^{(\mathbf{T}_1)} = \pi, \theta_{\min}^{(\mathbf{T}_2)} = \pi/3, \theta_{\max}^{(\mathbf{T}_2)} = 4\pi/3, \theta_{\min}^{(\mathbf{R}_1)} = 2\pi/3, \theta_{\max}^{(\mathbf{R}_1)} = -2\pi/3, \theta_{\min}^{(\mathbf{R}_2)} = \pi$ and $\theta_{\max}^{(\mathbf{R}_2)} = 0$.

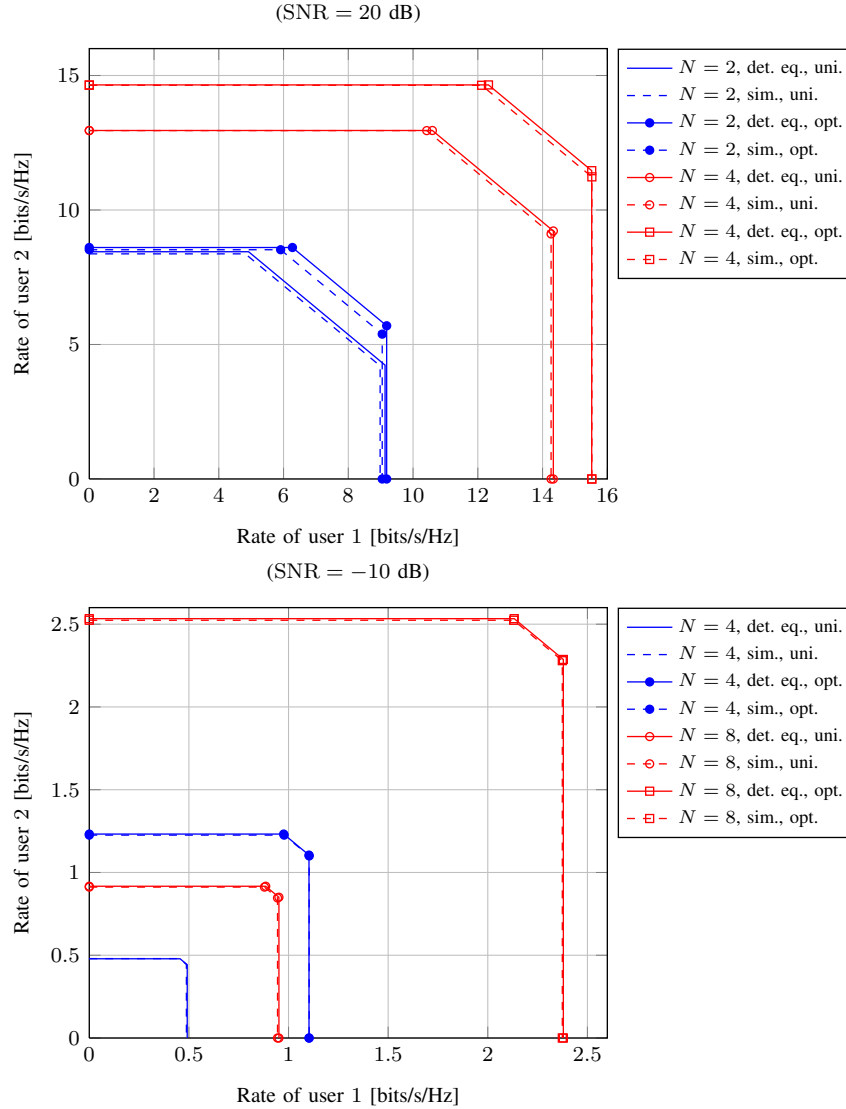


Fig. 2. Rate region for a two-user MIMO-MAC when the antennas are placed on a linear array with distance between close antennas is $\lambda/10$. The number of antennas $N = n_1 = n_2$ is taken equal to 2, 4 or 8. Simulations (sim.) are compared against deterministic equivalents (det. eq.). Uniform power allocation across transmit antennas (uni.) as well as optimal sum rate maximising precoding (opt.) are considered. The SNR is 20 dB in the top figure and -10 dB in the bottom figure.

We start by simulating the MIMO-MAC rate region obtained when $n_1 = n_2 = N$, either for $N = 2$, $N = 4$ or $N = 8$, for a linear antenna array with distance $\lambda/10$ between close antennas, under identity or optimal precoding policies and with signal to noise ratio of 20 dB or -10 dB. Simulation results, averaged over 10,000 channel realisations are compared with the deterministic equivalents. This is depicted in Figure 2. The deterministic equivalents of the rate regions appear to approximate the true rate regions extremely well, even for very small system dimensions. The case $N = 8$ shows in particular a perfect match. We note that increasing the number of antennas on both communication side provides a greater gain when using the optimal precoding policy. We also

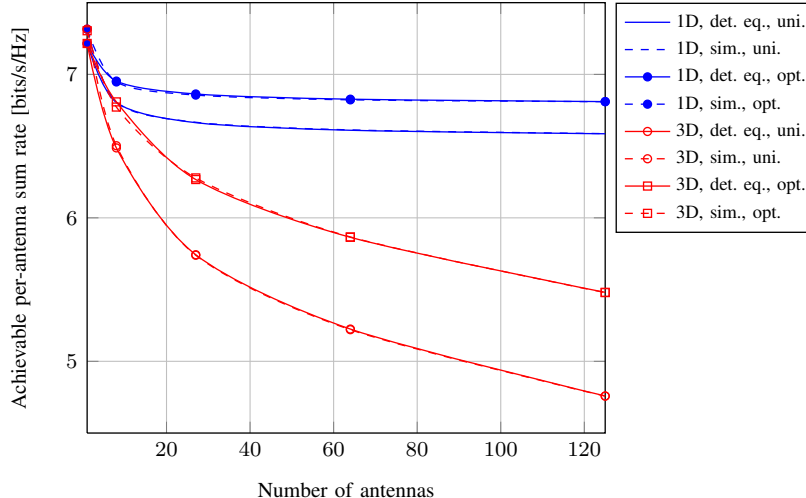


Fig. 3. Per-antenna sum rate for a two-user MIMO-MAC when the antennas are placed on a line (1D) or a cubic array (3D). The number of antennas satisfy $N = n_1 = n_2$ and range from 1 to 125. Simulations (sim.) are compared against deterministic equivalents (det. eq.). The distance between close antennas is $\lambda/2$. The SNR is 20 dB. Uniform power allocation (uni.) as well as optimal sum rate maximising precoding (opt.) are considered.

observe that, while increasing the number of antennas tends to reduce the individual per-antenna rate under uniform power allocation, using an optimal precoding policy significantly increases the per-antenna rate. This phenomenon is particularly accentuated in the low SNR regime. This confirms the observations made by Vishwanath *et al.* in [28], according to which the efficiency of individual antennas can grow as the correlation grows at low SNR. This is, for a given number of antennas, by increasing correlation and systematically applying optimal precoding, strong eigenmodes emerge over which data can be directed. This leads to higher rates than for an uncorrelated antenna array for which there are no such strong eigenmodes.

In order to test the robustness of the proposed deterministic equivalents to strongly correlated channel conditions, we then compare in Figure 3 the MIMO-MAC ergodic sum rate to the associated deterministic equivalents when precoder optimisation is performed or not and when the antenna grids are either one-dimensional regular arrays or three-dimensional regular cubes. In all situations, the distance between neighboring antennas is half the wavelength and the signal to noise ratio is taken to be 20 dB. We take the number of antennas on both sides to be successively 1, 8, 27, 64 and 125. We first observe that the deterministic equivalents are extremely accurate in this scenario. We confirm also the behaviour of the antenna efficiency, which saturates for the one-dimensional array and decreases fast for the three-dimensional antenna array, similar to what was observed in [25] for the single-user case. In terms of performance, a large improvement of the achievable sum rate is observed, especially in the three-dimensional case, when the optimal precoding policy is applied. Using optimal precoding policies can therefore significantly reduce the negative impact of antenna correlation even in this high SNR regime.

VI. CONCLUSION

In this article, we analyzed the performance of multi-antenna multi-user wireless communications and more particularly the multi-antenna multiple access channel, while taking into account the correlation effects due to close antennas and reduced solid angles of energy transmission and reception. The analytic approach is based on novel results, based on recent tools from the field of large dimensional random matrix theory. From these tools, we provide on the one hand a deterministic equivalent of the per-antenna mutual information of the MIMO-MAC for arbitrary precoders under possibly strongly correlated channel conditions and on the other hand an approximation of the rate maximising precoders along with an iterative water-filling algorithm to compute these precoders. In particular, while theoretical results prove the asymptotic accuracy of our model in the case of dense antenna packing or multi-dimensional antenna array on either communication side, simulations concur and suggest that the deterministic equivalents are moreover extremely accurate for very small system dimensions. These results can be used both from a practical side to easily derive optimal precoders and from a theoretical side to quantify the gains achieved by optimal power allocation policies in strongly correlated MIMO channels.

APPENDIX A
PROOF OF THEOREM 1

For ease of read, the proof will be divided into several sections.

We first consider the case $K = 1$, whose generalisation to $K \geq 1$ is given in Appendix A-E. Therefore, in the coming sections, we drop the useless indexes.

A. Truncation and centralisation

We begin with the truncation and centralisation steps which will replace \mathbf{X} , \mathbf{R} and \mathbf{T} by matrices with bounded entries, more suitable for analysis; the difference of the Stieltjes transforms of the original and new \mathbf{B}_N converging to zero. Since vague convergence of distribution functions is equivalent to the convergence of their Stieltjes transforms, it is sufficient to show the original and new empirical distribution functions of the eigenvalues approach each other almost surely in the space of subprobability measures on \mathbb{R} with respect to the topology which yields vague convergence.

Let $\tilde{X}_{ij} = X_{ij} \mathbf{1}_{\{|X_{ij}| < \sqrt{N}\}} - \mathbb{E}(X_{ij} \mathbf{1}_{\{|X_{ij}| < \sqrt{N}\}})$ and $\tilde{\mathbf{X}} = \left(\frac{1}{\sqrt{n}} \tilde{X}_{ij}\right)$. Then, from c), Lemma 1 and a), Lemma 3, it follows exactly as in the initial truncation and centralisation steps in [22] and [29] (which provide more details in their appendices), that

$$\|F_N - F^{\mathbf{S}+\mathbf{R}^{\frac{1}{2}} \tilde{\mathbf{X}} \mathbf{T} \tilde{\mathbf{X}}^{\mathbf{H}} \mathbf{R}^{\frac{1}{2}}}\| \xrightarrow{\text{a.s.}} 0,$$

as $N \rightarrow \infty$.

Let now $\bar{X}_{ij} = \tilde{X}_{ij} \cdot \mathbf{1}_{\{|X_{ij}| < \ln N\}} - \mathbb{E}(\tilde{X}_{ij} \mathbf{1}_{\{|X_{ij}| < \ln N\}})$ and $\bar{\mathbf{X}} = \left(\frac{1}{\sqrt{n}} \bar{X}_{ij}\right)$. This is the final truncation and centralisation step, which will be practically handled the same way as in [22], which some minor modifications, given presently.

For any Hermitian non-negative definite $r \times r$ matrix \mathbf{A} , let $\lambda_i^{\mathbf{A}}$ denote its i -th smallest eigenvalue of \mathbf{A} . With $\mathbf{A} = \mathbf{U} \text{diag}(\lambda_1^{\mathbf{A}}, \dots, \lambda_r^{\mathbf{A}}) \mathbf{U}^{\mathbf{H}}$ its spectral decomposition, let for any $\alpha > 0$

$$\mathbf{A}^{\alpha} = \mathbf{U} \text{diag}(\lambda_1^{\mathbf{A}} \mathbf{1}_{\{\lambda_1^{\mathbf{A}} \leq \alpha\}}, \dots, \lambda_r^{\mathbf{A}} \mathbf{1}_{\{\lambda_r^{\mathbf{A}} \leq \alpha\}}) \mathbf{U}^{\mathbf{H}}.$$

Then for any $N \times N$ matrix \mathbf{Q} , we get from 1) and 2), Lemma 3,

$$\begin{aligned} \|F^{\mathbf{S}+\mathbf{R}^{\frac{1}{2}} \mathbf{Q} \mathbf{T} \mathbf{Q}^{\mathbf{H}} \mathbf{R}^{\frac{1}{2}}} - F^{\mathbf{S}+\mathbf{R}^{\frac{1}{2} \alpha} \mathbf{Q} \mathbf{T}^{\alpha} \mathbf{Q}^{\mathbf{H}} \mathbf{R}^{\frac{1}{2} \alpha}}\| &\leq \frac{2}{N} \text{rank}(\mathbf{R}^{\frac{1}{2}} - \mathbf{R}^{\frac{1}{2} \alpha}) + \frac{1}{N} \text{rank}(\mathbf{T} - \mathbf{T}^{\alpha}) \\ &= \frac{2}{N} \sum_{i=1}^N \mathbf{1}_{\{\lambda_i^{\mathbf{R}} > \alpha\}} + \frac{1}{N} \sum_{i=1}^n \mathbf{1}_{\{\lambda_i^{\mathbf{T}} > \alpha\}} \\ &= 2F^{\mathbf{R}}((\alpha, \infty)) + \frac{1}{c_N} F^{\mathbf{T}}((\alpha, \infty)). \end{aligned}$$

Therefore, from the assumptions 4) and 6) in Theorem 1, we have for any sequence $\{\alpha_N\}$ with $\alpha_N \rightarrow \infty$

$$\|F^{\mathbf{S}+\mathbf{R}^{\frac{1}{2}} \mathbf{Q} \mathbf{T} \mathbf{Q}^{\mathbf{H}} \mathbf{R}^{\frac{1}{2}}} - F^{\mathbf{S}+\mathbf{R}^{\frac{1}{2} \alpha_N} \mathbf{Q} \mathbf{T}^{\alpha_N} \mathbf{Q}^{\mathbf{H}} \mathbf{R}^{\frac{1}{2} \alpha_N}}\| \rightarrow 0, \quad (16)$$

as $N \rightarrow \infty$.

A metric D on probability measures defined on \mathbb{R} , which induces the topology of vague convergence, is introduced in [22] to handle the last truncation step. The matrices studied in [22] are essentially \mathbf{B}_N with $\mathbf{R} = \mathbf{I}_N$. Following the steps beginning at (3.4) in [22], we see in our case that when α_N is chosen so that as $N \rightarrow \infty$, $\alpha_N \uparrow \infty$,

$$\alpha_N^8 (\mathbb{E}|X_{11}|^2 \mathbf{1}_{\{|X_{11}| \geq \ln N\}} + N^{-1}) \rightarrow 0$$

and

$$\sum_{N=1}^{\infty} \frac{\alpha_N^{16}}{N^2} \left(\mathbb{E}|X_{11}|^4 \mathbf{1}_{\{|X_{11}| < \sqrt{N}\}} + 1 \right) < \infty.$$

We will get

$$D(F^{\mathbf{S}+\mathbf{R}^{\frac{1}{2}\alpha_N} \tilde{\mathbf{X}} \mathbf{T}^{\alpha_N} \tilde{\mathbf{X}}^{\mathbf{H}} \mathbf{R}^{\frac{1}{2}\alpha_N}, F^{\mathbf{S}+\mathbf{R}^{\frac{1}{2}\alpha_N} \bar{\mathbf{X}} \mathbf{T}^{\alpha_N} \bar{\mathbf{X}}^{\mathbf{H}} \mathbf{R}^{\frac{1}{2}\alpha_N}}) \xrightarrow{\text{a.s.}} 0 \quad (17)$$

as $N \rightarrow \infty$.

Since $\mathbb{E}|\bar{X}_{11}|^2 \rightarrow 1$ as $N \rightarrow \infty$ we can rescale and replace $\bar{\mathbf{X}}$ with $\bar{\mathbf{X}}/\sqrt{\mathbb{E}|\bar{X}_{11}|^2}$, whose components are bounded by $k \ln N$ for some $k > 2$. Let $\log N$ denote logarithm of N with base $e^{1/k}$ (so that $k \ln N = \log N$). Therefore, from (16) and (17) we can assume that for each N the X_{ij} are i.i.d., $\mathbb{E}X_{11} = 0$, $\mathbb{E}|X_{11}|^2 = 1$, and $|X_{ij}| \leq \log N$.

Later on the proof will require a restricted growth rate on both $\|\mathbf{R}\|$ and $\|\mathbf{T}\|$. We see from (16) that we can also assume

$$\max(\|\mathbf{R}\|, \|\mathbf{T}\|) \leq \log N. \quad (18)$$

B. Deterministic approximation of $m_N(z)$

Write $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_n]$, $\mathbf{x}_i \in \mathbb{C}^N$ and let $\mathbf{y}_j = (1/\sqrt{n})\mathbf{R}^{\frac{1}{2}}\mathbf{x}_j$. Then we can write

$$\mathbf{B}_N = \mathbf{S} + \sum_{j=1}^n \tau_j \mathbf{y}_j \mathbf{y}_j^{\mathbf{H}}.$$

We assume $z \in \mathbb{C}^+$ and let $v = \Im[z]$. Define

$$e_N = e_N(z) = (1/N) \text{tr} \mathbf{R}(\mathbf{B}_N - z\mathbf{I}_N)^{-1}$$

and

$$p_N = -\frac{1}{nz} \sum_{j=1}^n \frac{\tau_j}{1 + c_N \tau_j e_N} = \int \frac{-\tau}{z(1 + c_N \tau e_N)} dF^{\mathbf{T}}(\tau)$$

Write $\mathbf{B}_N = \mathbf{O}\mathbf{\Lambda}\mathbf{O}^{\mathbf{H}}$, $\mathbf{\Lambda} = \text{diag}(\lambda_1, \dots, \lambda_N)$, its spectral decomposition. Let $\underline{\mathbf{R}} = \{\underline{R}_{ij}\} = \mathbf{O}^{\mathbf{H}}\mathbf{R}\mathbf{O}$. Then

$$e_N = (1/N) \text{tr} \underline{\mathbf{R}}(\mathbf{\Lambda} - z\mathbf{I}_N)^{-1} = (1/N) \sum_{i=1}^N \frac{\underline{R}_{ii}}{\lambda_i - z}.$$

We therefore see that e_N is the Stieltjes transform of a measure on the nonnegative reals with total mass $(1/N) \text{tr} \mathbf{R}$. It follows that both $e_N(z)$ and $ze_N(z)$ map \mathbb{C}^+ into \mathbb{C}^+ . This implies that $p_N(z)$ and $zp_N(z)$ map \mathbb{C}^+ into \mathbb{C}^+ and, as $z \rightarrow \infty$, $zp_N(z) \rightarrow -(1/n) \text{tr} \mathbf{T}$. Therefore, from Lemma 6, we also have p_N the Stieltjes transform of a measure on the nonnegative reals with total mass $(1/n) \text{tr} \mathbf{T}$. From (18), it follows that

$$|e_N| \leq v^{-1} \log N \quad (19)$$

and

$$\left| \int \frac{\tau}{(1 + c_N \tau e_N)} dF^{\mathbf{T}}(\tau) \right| = |z p_N(z)| \leq |z| v^{-1} \log N. \quad (20)$$

More generally, from Lemma 6, any function of the form

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

where $\tau \geq 0$ and $m(z)$ is the Stieltjes transform of a finite measure on \mathbb{R}^+ , is the Stieltjes transform of a measure on the nonnegative reals with total mass τ . It follows that

$$\left| \frac{\tau}{1 + m(z)} \right| \leq \tau |z| v^{-1}. \quad (21)$$

Fix now $z \in \mathbb{C}^+$. Let $\mathbf{B}_{(j)} = \mathbf{B}_N - \tau_j \mathbf{y}_j \mathbf{y}_j^H$. Define $\mathbf{D} = -z \mathbf{I}_N + \mathbf{S} - z p_N(z) \mathbf{R}$. We write

$$\mathbf{B}_N - z \mathbf{I}_N - \mathbf{D} = \sum_{j=1}^n \tau_j \mathbf{y}_j \mathbf{y}_j^H + z p_N \mathbf{R}.$$

Taking inverses and using Lemma 4 we have

$$\begin{aligned} (\mathbf{B}_N - z \mathbf{I}_N)^{-1} - \mathbf{D}^{-1} &= \sum_{j=1}^n \tau_j \mathbf{D}^{-1} \mathbf{y}_j \mathbf{y}_j^H (\mathbf{B}_N - z \mathbf{I}_N)^{-1} + z p_N \mathbf{D}^{-1} \mathbf{R} (\mathbf{B}_N - z \mathbf{I}_N)^{-1} \\ &= \sum_{j=1}^n \tau_j \frac{\mathbf{D}^{-1} \mathbf{y}_j \mathbf{y}_j^H (\mathbf{B}_{(j)} - z \mathbf{I}_N)^{-1}}{1 + \tau_j \mathbf{y}_j^H (\mathbf{B}_{(j)} - z \mathbf{I}_N)^{-1} \mathbf{y}_j} + z p_N \mathbf{D}^{-1} \mathbf{R} (\mathbf{B}_N - z \mathbf{I}_N)^{-1}. \end{aligned}$$

Taking traces and dividing by N , we have

$$\frac{1}{N} \operatorname{tr} \mathbf{D}^{-1} - m_N(z) = \frac{1}{n} \sum_{j=1}^n \tau_j d_j \equiv w_N^m,$$

where

$$d_j = \frac{(1/N) \mathbf{x}_j^H \mathbf{R}^{\frac{1}{2}} (\mathbf{B}_{(j)} - z \mathbf{I}_N)^{-1} \mathbf{D}^{-1} \mathbf{R}^{\frac{1}{2}} \mathbf{x}_j}{1 + \tau_j \mathbf{y}_j^H (\mathbf{B}_{(j)} - z \mathbf{I}_N)^{-1} \mathbf{y}_j} - \frac{(1/N) \operatorname{tr} \mathbf{R} (\mathbf{B}_N - z \mathbf{I}_N)^{-1} \mathbf{D}^{-1}}{1 + c_N \tau_j e_N}.$$

Multiplying both sides of the above matrix identity by \mathbf{R} , and then taking traces and dividing by N , we find

$$\frac{1}{N} \operatorname{tr} \mathbf{D}^{-1} \mathbf{R} - e_N(z) = \frac{1}{n} \sum_{j=1}^n \tau_j d_j^e \equiv w_N^e,$$

where

$$d_j^e = \frac{(1/N) \mathbf{x}_j^H \mathbf{R}^{\frac{1}{2}} (\mathbf{B}_{(j)} - z \mathbf{I}_N)^{-1} \mathbf{R} \mathbf{D}^{-1} \mathbf{R}^{\frac{1}{2}} \mathbf{x}_j}{1 + \tau_j \mathbf{y}_j^H (\mathbf{B}_{(j)} - z \mathbf{I}_N)^{-1} \mathbf{y}_j} - \frac{(1/N) \operatorname{tr} \mathbf{R} (\mathbf{B}_N - z \mathbf{I}_N)^{-1} \mathbf{R} \mathbf{D}^{-1}}{1 + c_N \tau_j e_N}.$$

We then show that, for any $k > 0$, almost surely

$$\lim_{N \rightarrow \infty} (\log^k N) w_N^m = 0 \quad (22)$$

and

$$\lim_{n \rightarrow \infty} (\log^k N) w_N^e = 0. \quad (23)$$

Notice that for each j , $\mathbf{y}_j^H (\mathbf{B}_{(j)} - z \mathbf{I}_N)^{-1} \mathbf{y}_j$ can be viewed as the Stieltjes transform of a measure on \mathbb{R}^+ . Therefore from (21) we have

$$\left| \frac{1}{1 + \tau_j \mathbf{y}_j^H (\mathbf{B}_{(j)} - z \mathbf{I}_N)^{-1} \mathbf{y}_j} \right| \leq \frac{|z|}{v}. \quad (24)$$

For each j , let $e_{(j)} = e_{(j)}(z) = (1/N) \operatorname{tr} \mathbf{R}(\mathbf{B}_{(j)} - z\mathbf{I}_N)^{-1}$, and

$$p_{(j)} = p_{(j)}(z) = \int \frac{-\tau}{z(1 + c_N \tau e_{(j)})} dF^{\mathbf{T}}(\tau),$$

both being Stieltjes transforms of measures on \mathbb{R}^+ , along with the integrand for each τ .

Using Lemma 4, Equations (18) and (21), we have

$$|zp_N - zp_{(j)}| = |e_N - e_{(j)}|c_N \left| \int \frac{\tau^2}{(1 + c_N \tau e_N)(1 + c_N \tau e_{(j)})} dF^{\mathbf{T}}(\tau) \right| \leq \frac{c_N |z|^2 \log^3 N}{Nv^3}. \quad (25)$$

Let $\mathbf{D}_{(j)} = -z\mathbf{I}_N + \mathbf{S} - zp_{(j)}(z)\mathbf{R}$. Notice that $(\mathbf{B}_N - z\mathbf{I}_N)^{-1}$ and $(\mathbf{B}_{(j)} - z\mathbf{I}_N)^{-1}$ are bounded in spectral norm by v^{-1} and, from Lemma 8, the same holds true for \mathbf{D}^{-1} and $\mathbf{D}_{(j)}^{-1}$.

In order to handle both w_N^m , d_j and w_N^e , d_j^e at the same time, we shall denote by \mathbf{E} either \mathbf{T} or \mathbf{I}_N , and w_N , d_j for now will denote either the original w_N^m , d_j or w_N^e , d_j^e . Write $d_j = d_j^1 + d_j^2 + d_j^3 + d_j^4$, where

$$\begin{aligned} d_j^1 &= \frac{(1/N)\mathbf{x}_j^H \mathbf{R}^{\frac{1}{2}}(\mathbf{B}_{(j)} - z\mathbf{I}_N)^{-1} \mathbf{E} \mathbf{D}^{-1} \mathbf{R}^{\frac{1}{2}} \mathbf{x}_j}{1 + \tau_j \mathbf{y}_j^H (\mathbf{B}_{(j)} - z\mathbf{I}_N)^{-1} \mathbf{y}_j} - \frac{(1/N)\mathbf{x}_j^H \mathbf{R}^{\frac{1}{2}}(\mathbf{B}_{(j)} - z\mathbf{I}_N)^{-1} \mathbf{E} \mathbf{D}_{(j)}^{-1} \mathbf{R}^{\frac{1}{2}} \mathbf{x}_j}{1 + \tau_j \mathbf{y}_j^H (\mathbf{B}_{(j)} - z\mathbf{I}_N)^{-1} \mathbf{y}_j} \\ d_j^2 &= \frac{(1/N)\mathbf{x}_j^H \mathbf{R}^{\frac{1}{2}}(\mathbf{B}_{(j)} - z\mathbf{I}_N)^{-1} \mathbf{E} \mathbf{D}_{(j)}^{-1} \mathbf{R}^{\frac{1}{2}} \mathbf{x}_j}{1 + \tau_j \mathbf{y}_j^H (\mathbf{B}_{(j)} - z\mathbf{I}_N)^{-1} \mathbf{y}_j} - \frac{(1/N) \operatorname{tr} \mathbf{R}(\mathbf{B}_{(j)} - z\mathbf{I}_N)^{-1} \mathbf{E} \mathbf{D}_{(j)}^{-1}}{1 + \tau_j \mathbf{y}_j^H (\mathbf{B}_{(j)} - z\mathbf{I}_N)^{-1} \mathbf{y}_j} \\ d_j^3 &= \frac{(1/N) \operatorname{tr} \mathbf{R}(\mathbf{B}_{(j)} - z\mathbf{I}_N)^{-1} \mathbf{E} \mathbf{D}_{(j)}^{-1}}{1 + \tau_j \mathbf{y}_j^H (\mathbf{B}_{(j)} - z\mathbf{I}_N)^{-1} \mathbf{y}_j} - \frac{(1/N) \operatorname{tr} \mathbf{R}(\mathbf{B}_N - z\mathbf{I}_N)^{-1} \mathbf{E} \mathbf{D}^{-1}}{1 + \tau_j \mathbf{y}_j^H (\mathbf{B}_{(j)} - z\mathbf{I}_N)^{-1} \mathbf{y}_j} \\ d_j^4 &= \frac{(1/N) \operatorname{tr} \mathbf{R}(\mathbf{B}_N - z\mathbf{I}_N)^{-1} \mathbf{E} \mathbf{D}^{-1}}{1 + \tau_j \mathbf{y}_j^H (\mathbf{B}_{(j)} - z\mathbf{I}_N)^{-1} \mathbf{y}_j} - \frac{(1/N) \operatorname{tr} \mathbf{R}(\mathbf{B}_N - z\mathbf{I}_N)^{-1} \mathbf{E} \mathbf{D}^{-1}}{1 + c_N \tau_j e_N}. \end{aligned}$$

From Lemma 4, Equations (18), (24) and (25), we have

$$\begin{aligned} \tau_j |d_j^1| &\leq \frac{1}{N} \|\mathbf{x}_j\|^2 \frac{c_N \log^7 N |z|^3}{Nv^7} \\ \tau_j |d_j^2| &\leq |z|v^{-1} \frac{\log N}{N} \left| \mathbf{x}_j^H \mathbf{R}^{\frac{1}{2}}(\mathbf{B}_{(j)} - z\mathbf{I}_N)^{-1} \mathbf{E} \mathbf{D}_{(j)}^{-1} \mathbf{R}^{\frac{1}{2}} \mathbf{x}_j - \operatorname{tr} \mathbf{R}(\mathbf{B}_{(j)} - z\mathbf{I}_N)^{-1} \mathbf{E} \mathbf{D}_{(j)}^{-1} \right| \\ \tau_j |d_j^3| &\leq \frac{|z| \log^3 N}{vN} \left(\frac{1}{v^2} + \frac{c_N |z|^2 \log^3 N}{v^6} \right) \rightarrow 0, \text{ as } n \rightarrow \infty \\ \tau_j |d_j^4| &\leq \frac{|z|c_N \log^4 N}{Nv^3} \left(\left| \mathbf{x}_j^H \mathbf{R}^{\frac{1}{2}}(\mathbf{B}_{(j)} - z\mathbf{I}_N)^{-1} \mathbf{R}^{\frac{1}{2}} \mathbf{x}_j - \operatorname{tr} \mathbf{R}^{\frac{1}{2}}(\mathbf{B}_{(j)} - z\mathbf{I}_N)^{-1} \mathbf{R}^{\frac{1}{2}} \right| + \frac{\log N}{v} \right). \end{aligned}$$

From Lemma 7, there exists $\bar{K} > 0$ such that,

$$\begin{aligned} \mathbb{E} \left| \frac{1}{N} \|\mathbf{x}_j\|^2 - 1 \right|^6 &\leq KN^{-3} \log^{12} N \\ \mathbb{E} \frac{1}{N^6} \left| \mathbf{x}_j^H \mathbf{R}^{\frac{1}{2}}(\mathbf{B}_{(j)} - z\mathbf{I}_N)^{-1} \mathbf{E} \mathbf{D}_{(j)}^{-1} \mathbf{R}^{\frac{1}{2}} \mathbf{x}_j - \operatorname{tr} \mathbf{R}(\mathbf{B}_{(j)} - z\mathbf{I}_N)^{-1} \mathbf{E} \mathbf{D}_{(j)}^{-1} \right|^6 &\leq KN^{-3} v^{-12} \log^{24} N \\ \mathbb{E} \frac{1}{N^6} \left| \mathbf{x}_j^H \mathbf{R}^{\frac{1}{2}}(\mathbf{B}_{(j)} - z\mathbf{I}_N)^{-1} \mathbf{R}^{\frac{1}{2}} \mathbf{x}_j - \operatorname{tr} \mathbf{R}^{\frac{1}{2}}(\mathbf{B}_{(j)} - z\mathbf{I}_N)^{-1} \mathbf{R}^{\frac{1}{2}} \right|^6 &\leq KN^{-3} v^{-6} \log^{18} N. \end{aligned}$$

All three moments when multiplied by n times any power of $\log N$, are summable. Applying standard arguments using the Borel-Cantelli lemma and Boole's inequality (on $4n$ events), we conclude that, for any $k > 0$ $\log^k N \max_{j \leq n} \tau_j d_j \xrightarrow{\text{a.s.}} 0$ as $N \rightarrow \infty$. Hence Equations (22) and (23).

C. Existence and uniqueness of $m_N^\circ(z)$

We show now that for any $N, n, \mathbf{S}, \mathbf{R}$, $N \times N$ nonnegative definite and $\mathbf{T} = \text{diag}(\tau_1, \dots, \tau_N)$, $\tau_k \geq 0$ for all $1 \leq k \leq N$, there exists a unique e with positive imaginary part for which

$$e = \frac{1}{N} \text{tr} \left(\mathbf{S} + \left[\int \frac{\tau}{1 + c_N \tau e} dF^{\mathbf{T}}(\tau) \right] \mathbf{R} - z \mathbf{I}_N \right)^{-1} \mathbf{R}. \quad (26)$$

For existence we consider the subsequences $\{N_j\}, \{n_j\}$ with $N_j = jN, n_j = jn$, so that c_{N_j} remains c_N , form the block diagonal matrices

$$\mathbf{R}_{N_j} = \text{diag}(\mathbf{R}, \mathbf{R}, \dots, \mathbf{R}), \quad \mathbf{S}_{N_j} = \text{diag}(\mathbf{S}, \mathbf{S}, \dots, \mathbf{S}) \quad (27)$$

both $jN \times jN$ and

$$\mathbf{T}_{N_j} = \text{diag}(\mathbf{T}, \mathbf{T}, \dots, \mathbf{T}) \quad (28)$$

of size $jn \times jn$. We see that $F^{\mathbf{T}_{N_j}} = F^{\mathbf{T}}$ and the right side of (26) remains unchanged for all N_j . Consider a realisation where $w_{N_j}^e \rightarrow 0$ as $j \rightarrow \infty$. We have $|e_{N_j}(z)| = |(jN)^{-1} \text{tr} \mathbf{R}(\mathbf{B}_{jN} - z \mathbf{I}_N)^{-1}| \leq v^{-1} \log N$, remaining bounded as $j \rightarrow \infty$. Consider then a subsequence for which e_{N_j} converges to, say, e . From (21), we see that

$$\left| \frac{\tau}{1 + c_N \tau e_{N_j}} \right| \leq \tau |z| v^{-1}$$

so that from the dominated convergence theorem we have

$$\int \frac{\tau}{1 + c_N \tau e_{N_j}(z)} dF^{\mathbf{T}}(\tau) \rightarrow \int \frac{\tau}{1 + c_N \tau e} dF^{\mathbf{T}}(\tau)$$

along this subsequence. Therefore e solves (26).

We now show uniqueness. Let e be a solution to (26) and let $e_2 = \Im[e]$. Recalling the definition of \mathbf{D} we write

$$e = \frac{1}{N} \text{tr} \left(\mathbf{D}^{-1} \mathbf{R} \mathbf{D}^{-\text{H}} \left(\mathbf{S} + \left[\int \frac{\tau}{1 + c_N \tau e^*} dF^{\mathbf{T}}(\tau) \right] \mathbf{R} - z^* \mathbf{I} \right) \right).$$

We see that since both \mathbf{R} and \mathbf{S} are Hermitian nonnegative definite, $\text{tr}(\mathbf{D}^{-1} \mathbf{R} \mathbf{D}^{-\text{H}} \mathbf{S})$ is real and nonnegative. Therefore we can write

$$e_2 = \frac{1}{N} \text{tr} \left(\mathbf{D}^{-1} \mathbf{R} (\mathbf{D}^{\text{H}})^{-1} \left(\left[\int \frac{c_N \tau^2 e_2}{|1 + c_N \tau e|^2} dF^{\mathbf{T}}(\tau) \right] \mathbf{R} + v \mathbf{I}_N \right) \right) = e_2 \alpha + v \beta, \quad (29)$$

where we denoted

$$\begin{aligned} \alpha &= \frac{1}{N} \text{tr} \left(\mathbf{D}^{-1} \mathbf{R} (\mathbf{D}^{\text{H}})^{-1} \left[\int \frac{c_N \tau^2}{|1 + c_N \tau e|^2} dF^{\mathbf{T}}(\tau) \right] \mathbf{R} \right) \\ \beta &= \frac{1}{N} \text{tr} (\mathbf{D}^{-1} \mathbf{R} (\mathbf{D}^{\text{H}})^{-1}). \end{aligned}$$

Let \underline{e} be another solution to (26), with $\underline{e}_2 = \Im[\underline{e}]$, and analogously we can write $\underline{e}_2 = \underline{e}_2 \alpha + v \beta$. Let $\underline{\mathbf{D}}$ denote \mathbf{D} with e replaced by \underline{e} . Then we have $e - \underline{e} = \gamma(e - \underline{e})$ where

$$\gamma = \int \frac{c_N \tau^2}{(1 + c_N \tau e)(1 + c_N \tau \underline{e})} dF^{\mathbf{T}}(\tau) \frac{\text{tr} \mathbf{D}^{-1} \mathbf{R} \underline{\mathbf{D}}^{-1} \mathbf{R}}{N}.$$

If \mathbf{R} is the zero matrix, then $\gamma = 0$, and $e = \underline{e}$ would follow. For $\mathbf{R} \neq 0$ we use Cauchy-Schwarz to find

$$\begin{aligned} |\gamma| &\leq \left(\int \frac{c_N \tau^2}{|1 + c_N \tau e|^2} dF^{\mathbf{T}}(\tau) \frac{\text{tr } \mathbf{D}^{-1} \mathbf{R} (\mathbf{D}^{\mathbf{H}})^{-1} \mathbf{R}}{N} \right)^{\frac{1}{2}} \left(\int \frac{c_N \tau^2}{|1 + c_N \tau \underline{e}|^2} dF^{\mathbf{T}}(\tau) \frac{\text{tr } \underline{\mathbf{D}}^{-1} \mathbf{R} (\underline{\mathbf{D}}^{\mathbf{H}})^{-1} \mathbf{R}}{N} \right)^{\frac{1}{2}} \\ &= \alpha^{\frac{1}{2}} \underline{\alpha}^{\frac{1}{2}} \\ &= \left(\frac{e_2 \alpha}{e_2 \alpha + v \beta} \right)^{\frac{1}{2}} \left(\frac{\underline{e}_2 \underline{\alpha}}{\underline{e}_2 \underline{\alpha} + v \underline{\beta}} \right)^{\frac{1}{2}}. \end{aligned}$$

Necessarily β and $\underline{\beta}$ are positive since $\mathbf{R} \neq 0$. Therefore $|\gamma| < 1$ so we must have $e = \underline{e}$. For $z < 0$ and $e > 0$, the same calculus can be performed, with γ remaining the same. The step (29) is changed by evaluating e , instead of e_2 , using the same technique. We obtain the same α while β is replaced by another positive scalar. We therefore still have that $\gamma < 1$.

D. Termination of the proof

Let e_N° denote the solution to (26). We show now for any $\ell > 0$, almost surely

$$\lim_{N \rightarrow \infty} \log^\ell N (e_N - e_N^\circ) = 0. \quad (30)$$

Let $e_2^\circ = \Im[e_N^\circ]$, and $\alpha^\circ = \alpha_N^\circ$, $\beta^\circ = \beta_N^\circ$ be the values as above for which $e_2^\circ = e_2^\circ \alpha^\circ + v \beta^\circ$. We have, using (18) and (21),

$$\begin{aligned} e_2^\circ \alpha_N^\circ / \beta_N^\circ &\leq e_2^\circ c_N \log N \int \frac{\tau^2}{|1 + c_N \tau e_N^\circ|^2} dF^{\mathbf{T}}(\tau) \\ &= -\log N \Im \left[\int \frac{\tau}{1 + c_N \tau e_N^\circ} dF^{\mathbf{T}}(\tau) \right] \\ &\leq \log^2 N |z| v^{-1}. \end{aligned}$$

Therefore

$$\begin{aligned} \alpha^\circ &= \left(\frac{e_2^\circ \alpha^\circ}{e_2^\circ \alpha^\circ + v \beta^\circ} \right) \\ &= \left(\frac{e_2^\circ \alpha^\circ / \beta^\circ}{v + e_2^\circ \alpha^\circ / \beta^\circ} \right) \\ &\leq \left(\frac{\log^2 N |z|}{v^2 + \log^2 N |z|} \right) \end{aligned} \quad (31)$$

Let \mathbf{D}° , \mathbf{D} denote \mathbf{D} as above with e replaced by, respectively e_N° and e_N . We have

$$e_N = \frac{1}{N} \text{tr } \mathbf{D}^{-1} \mathbf{R} - w_N^e.$$

With $e_2 = \Im[e_N]$ we write as above

$$\begin{aligned} e_2 &= \frac{1}{N} \text{tr} \left(\mathbf{D}^{-1} \mathbf{R} \mathbf{D}^{-\mathbf{H}} \left(\left[\int \frac{c_N \tau^2 e_2}{|1 + c_N \tau e_N|^2} dF^{\mathbf{T}}(\tau) \right] \mathbf{R} + v \mathbf{I}_N \right) \right) - \Im[w_N^e] \\ &= e_2 \alpha + v \beta - \Im w_N^e \end{aligned}$$

We have as above $e_N - e_N^\circ = \gamma(e_N - e_N^\circ) + w_N^e$ where now

$$|\gamma| \leq \alpha^{\circ\frac{1}{2}} \alpha^{\frac{1}{2}}.$$

Fix an $\ell > 0$ and consider a realisation for which $\log^{\ell'} N w_N^e \rightarrow 0$, where $\ell' = \max(\ell + 1, 4)$ and n large enough so that

$$|w_N^e| \leq \frac{v^3}{4c_N |z|^2 \log^3 N}. \quad (32)$$

Suppose $\beta \leq \frac{v^2}{4c_N |z|^2 \log^3 N}$. Then by Equations (18) and (21) we get

$$\alpha \leq c_N v^{-2} |z|^2 \log^3 N \beta \leq 1/4,$$

which implies $|\gamma| \leq 1/2$. Otherwise we get from (31) and (32)

$$\begin{aligned} |\gamma| &\leq \alpha^{\circ\frac{1}{2}} \left(\frac{e_2 \alpha}{e_2 \alpha + v \beta - \Im[w_N^e]} \right)^{\frac{1}{2}} \\ &\leq \left(\frac{\log N |z|}{v^2 + \log N |z|} \right)^{\frac{1}{2}}. \end{aligned}$$

Therefore for all N large

$$\begin{aligned} \log^\ell N |e_N - e_N^\circ| &\leq \frac{(\log^\ell N) w_N^e}{1 - \left(\frac{\log^2 N |z|}{v^2 + \log^2 N |z|} \right)^{\frac{1}{2}}} \\ &\leq 2v^{-2} (v^2 + \log^2 N |z|) (\log^\ell N) w_N^e \\ &\rightarrow 0, \end{aligned}$$

as $n \rightarrow \infty$. Therefore (30) follows.

Let $m_N^\circ = N^{-1} \text{tr } \mathbf{D}^\circ$. We finally show

$$m_N - m_N^\circ \xrightarrow{\text{a.s.}} 0, \quad (33)$$

as $n \rightarrow \infty$. Since $m_N = N^{-1} \text{tr } \mathbf{D}^{-1} - w_N^m$, we have

$$m_N - m_N^\circ = \gamma(e_N - e_N^\circ) - w_N^m,$$

where now

$$\gamma = \int \frac{c_N \tau^2}{(1 + c_N \tau e_N)(1 + c_N \tau e_N^\circ)} dF^{\mathbf{T}}(\tau) \frac{\text{tr } \mathbf{D}^{-1} \mathbf{R} \mathbf{D}^{\circ-1}}{N}.$$

From (18) and (21) we get $|\gamma| \leq c_N |z|^2 v^{-4} \log^3 N$. Therefore, from (22) and (30), we get (33).

Returning to the original assumptions on X_{11} , \mathbf{T} , and \mathbf{R} , for each of a countably infinite collection of z with positive imaginary part, possessing a cluster point with positive imaginary part, we have (33). Therefore, by Vitali's convergence theorem, page 168 of [30], for any $\varepsilon > 0$ we have with probability one $m_N(z) - m_N^\circ(z) \rightarrow 0$ uniformly in any region of \mathbb{C} bounded by a contour interior to

$$\mathbb{C} \setminus (\{z : |z| \leq \varepsilon\} \cup \{z = x + iv : x > 0, |v| \leq \varepsilon\}).$$

If $\mathbf{S} = f(\mathbf{R})$, meaning the eigenvalues of \mathbf{R} are changed via f in the spectral decomposition of \mathbf{R} , then we have

$$\begin{aligned} m_N^\circ(z) &= \int \frac{1}{f(r) + r \int \frac{\tau}{1+c_N \tau e_N^\circ(z)} dF^{\mathbf{T}}(\tau) - z} dF^{\mathbf{R}}(r) \\ e_N^\circ(z) &= \int \frac{r}{f(r) + r \int \frac{\tau}{1+c_N \tau e_N^\circ(z)} dF^{\mathbf{T}}(\tau) - z} dF^{\mathbf{R}}(r). \end{aligned}$$

E. Extension to $K \geq 1$

Suppose now

$$\mathbf{B}_N = \mathbf{S} + \sum_{k=1}^K \mathbf{R}_k^{\frac{1}{2}} \mathbf{X}_k \mathbf{T}_k \mathbf{X}_k^H \mathbf{R}_k^{\frac{1}{2}}$$

where K remains fixed, \mathbf{X}_k is $N \times n_k$ satisfying 1, the \mathbf{X}_k 's are independent, \mathbf{R}_k satisfies 2) and 4), \mathbf{T}_k is $n_k \times n_k$ satisfying 3) and 4), $c_k = N/n_k$ satisfies 6), and \mathbf{S} satisfies 5). After truncation and centralisation we may assume the same condition on the entries of the \mathbf{X}_k 's, and the spectral norms of the \mathbf{R}_k 's and the \mathbf{T}_k 's. Write $\mathbf{y}_{k,j} = (1/\sqrt{n_k}) \mathbf{R}_k^{\frac{1}{2}} \mathbf{x}_{k,j}$, with $\mathbf{x}_{k,j}$ denoting the j -th column of \mathbf{X}_k , and let $\tau_{k,j}$ denote the j -th diagonal element of \mathbf{T}_k . Then we can write

$$\mathbf{B}_N = \mathbf{S} + \sum_{k=1}^K \sum_{j=1}^{n_k} \tau_{k,j} \mathbf{y}_{k,j} \mathbf{y}_{k,j}^H.$$

Define

$$e_{N,k} = e_{N,k}(z) = (1/N) \text{tr} \mathbf{R}_k (\mathbf{B}_N - z \mathbf{I}_N)^{-1}$$

and

$$\begin{aligned} p_k &= -\frac{1}{n_k z} \sum_{j=1}^{n_k} \frac{\tau_{k,j}}{1 + c_k \tau_{k,j} e_{N,k}} \\ &= \int \frac{-\tau_k}{1 + c_k \tau_k e_{N,k}} dF^{\mathbf{T}_k}(\tau_k). \end{aligned}$$

We see $e_{N,k}$ and p_k have the same properties as e_N and p_N . Let $\mathbf{B}_{k,(j)} = \mathbf{B}_N - \tau_{k,j} \mathbf{y}_{k,j} \mathbf{y}_{k,j}^H$. Define $\mathbf{D} = -z \mathbf{I}_N + \mathbf{S} - \sum_{k=1}^K z p_k(z) \mathbf{R}_k$. We write

$$\mathbf{B}_N - z \mathbf{I}_N - \mathbf{D} = \sum_{k=1}^K \left(\sum_{j=1}^{n_k} \tau_{k,j} \mathbf{y}_{k,j} \mathbf{y}_{k,j}^H + z p_k(z) \mathbf{R}_k \right).$$

Taking inverses and using Lemma 4, we have

$$\begin{aligned} \mathbf{D}^{-1} - (\mathbf{B}_N - z \mathbf{I}_N)^{-1} &= \sum_{k=1}^K \left(\sum_{j=1}^{n_k} \tau_{k,j} \mathbf{D}^{-1} \mathbf{y}_{k,j} \mathbf{y}_{k,j}^H (\mathbf{B}_N - z \mathbf{I}_N)^{-1} + z p_k \mathbf{D}^{-1} \mathbf{R}_k (\mathbf{B}_N - z \mathbf{I}_N)^{-1} \right) \\ &= \sum_{k=1}^K \left(\sum_{j=1}^{n_k} \tau_{k,j} \frac{\mathbf{D}^{-1} \mathbf{y}_{k,j} \mathbf{y}_{k,j}^H (\mathbf{B}_{k,(j)} - z \mathbf{I}_N)^{-1}}{1 + \tau_{k,j} \mathbf{y}_{k,j}^H (\mathbf{B}_{k,(j)} - z \mathbf{I}_N)^{-1} \mathbf{y}_{k,j}} + z p_k \mathbf{D}^{-1} \mathbf{R}_k (\mathbf{B}_N - z \mathbf{I}_N)^{-1} \right). \end{aligned}$$

Taking traces and dividing by N , we have

$$(1/N) \text{tr} \mathbf{D}^{-1} - m_N(z) = \sum_{k=1}^K \frac{1}{n_k} \sum_{j=1}^{n_k} \tau_{k,j} d_{k,j} \equiv w_N^m,$$

where

$$d_{k,j} = \frac{(1/N)\mathbf{x}_{k,j}^H \mathbf{R}_k^{\frac{1}{2}} (\mathbf{B}_{k,(j)} - z\mathbf{I}_N)^{-1} \mathbf{D}^{-1} \mathbf{R}_k^{\frac{1}{2}} \mathbf{x}_{k,j}}{1 + \tau_{k,j} \mathbf{y}_{k,j}^H (\mathbf{B}_{k,(j)} - z\mathbf{I}_N)^{-1} \mathbf{y}_{k,j}} - \frac{(1/N) \operatorname{tr} \mathbf{R}_k (\mathbf{B}_N - z\mathbf{I}_N)^{-1} \mathbf{D}^{-1}}{1 + c_k \tau_{k,j} e_{N,k}}.$$

For a fixed $\underline{k} \in \{1, \dots, K\}$, we multiply the above matrix identity by $\mathbf{R}_{\underline{k}}$, take traces and divide by N . Thus we get

$$(1/N) \operatorname{tr} \mathbf{D}^{-1} \mathbf{R}_{\underline{k}} - e_{\underline{k}}(z) = \sum_{k=1}^K \frac{1}{n_k} \sum_{j=1}^{n_k} \tau_{k,j} d_{k\underline{k}j}^e \equiv w_{\underline{k}}^e,$$

where

$$d_{k\underline{k}j}^e = \frac{(1/N)\mathbf{x}_{k,j}^H \mathbf{R}_k^{\frac{1}{2}} (\mathbf{B}_{k,(j)} - z\mathbf{I}_N)^{-1} \mathbf{R}_{\underline{k}} \mathbf{D}^{-1} \mathbf{R}_k^{\frac{1}{2}} \mathbf{x}_{k,j}}{1 + \tau_{k,j} \mathbf{y}_{k,j}^H (\mathbf{B}_{k,(j)} - z\mathbf{I}_N)^{-1} \mathbf{y}_{k,j}} - \frac{(1/N) \operatorname{tr} \mathbf{R}_k (\mathbf{B}_N - z\mathbf{I}_N)^{-1} \mathbf{R}_{\underline{k}} \mathbf{D}^{-1}}{1 + c_k \tau_{k,j} e_{N,k}}.$$

In exactly the same way as in the case with $K = 1$ we find that for any nonnegative ℓ , $\log^\ell N w_N^m$ and the $\log^\ell w_{\underline{i}}^e$'s converge almost surely to zero. By considering block diagonal matrices as before with N , n_i 's, \mathbf{S} , \mathbf{R}_i 's and \mathbf{T}_i 's all fixed we find that there exist $e_1^\circ, \dots, e_K^\circ$ with positive imaginary parts for which for each i

$$e_i^\circ = \frac{1}{N} \operatorname{tr} \mathbf{R}_i \left(\mathbf{S} + \sum_{k=1}^K \left[\int \frac{\tau}{1 + c_k \tau e_k^\circ} dF^{\mathbf{T}_k}(\tau) \right] \mathbf{R}_k - z\mathbf{I}_N \right)^{-1}. \quad (34)$$

Let us verify uniqueness. Let $\mathbf{e}^\circ = (e_1^\circ, \dots, e_K^\circ)^\top$, and let \mathbf{D}° denote the matrix in (34) whose inverse is taken (essentially \mathbf{D} after the $e_{N,i}$'s are replaced by the e_i° 's). Let for each j , $e_{j,2}^\circ = \Im e_j^\circ$, and $\mathbf{e}_2^\circ = (e_{1,2}^\circ, \dots, e_{K,2}^\circ)^\top$. Then, noticing that for each i , $\operatorname{tr} \mathbf{S} \mathbf{D}^{\circ-H} \mathbf{R}_i \mathbf{D}^{\circ-1}$ is real and nonnegative (positive whenever $\mathbf{S} \neq 0$) and $\operatorname{tr} \mathbf{D}^{\circ-H} \mathbf{R}_i \mathbf{D}^{\circ-1}$ and $\operatorname{tr} \mathbf{R}_j \mathbf{D}^{\circ-H} \mathbf{R}_i \mathbf{D}^{\circ-1}$ are real and positive for all i, j , we have

$$\begin{aligned} e_{i,2}^\circ &= \Im \left[\frac{1}{N} \operatorname{tr} \left(\mathbf{S} + \sum_{j=1}^K \left[\int \frac{\tau}{1 + c_j \tau e_j^\circ} dF^{\mathbf{T}_j}(\tau) \right] \mathbf{R}_j - z^* \mathbf{I} \right) \mathbf{D}^{\circ-H} \mathbf{R}_i \mathbf{D}^{\circ-1} \right] \\ &= \sum_{j=1}^K e_{j,2}^\circ \frac{1}{N} \operatorname{tr} \mathbf{R}_j \mathbf{D}^{\circ-H} \mathbf{R}_i \mathbf{D}^{\circ-1} c_j \int \frac{\tau^2}{|1 + c_j \tau e_j^\circ|^2} dF^{\mathbf{T}_j}(\tau) + \frac{v}{N} \operatorname{tr} \mathbf{D}^{\circ-H} \mathbf{R}_i \mathbf{D}^{\circ-1}. \end{aligned} \quad (35)$$

Let $\mathbf{C}^\circ = (c_{ij}^\circ)$, $\mathbf{b}^\circ = (b_1^\circ, \dots, b_N^\circ)^\top$, where

$$c_{ij}^\circ = \frac{1}{N} \operatorname{tr} \mathbf{R}_j \mathbf{D}^{\circ-H} \mathbf{R}_i \mathbf{D}^{\circ-1} c_j \int \frac{\tau^2}{|1 + c_j \tau e_j^\circ|^2} dF^{\mathbf{T}_j}(\tau)$$

and

$$b_i^\circ = \frac{1}{N} \operatorname{tr} \mathbf{D}^{\circ-H} \mathbf{R}_i \mathbf{D}^{\circ-1}.$$

Therefore we have that \mathbf{e}_2° satisfies

$$\mathbf{e}_2^\circ = \mathbf{C}^\circ \mathbf{e}_2^\circ + v \mathbf{b}^\circ. \quad (36)$$

We see that each $e_{j,2}^\circ$, c_{ij}° , and b_j° are positive. Therefore, from Lemma 9 we have $\rho(\mathbf{C}^\circ) < 1$.

Let $\underline{\mathbf{e}}^\circ = (\underline{e}_1^\circ, \dots, \underline{e}_K^\circ)^\top$ be another solution to (34), with $\underline{\mathbf{e}}_2^\circ$, $\underline{\mathbf{D}}^\circ$, $\underline{\mathbf{C}}^\circ = (\underline{c}_{ij}^\circ)$, $\underline{\mathbf{b}}^\circ$ defined analogously, so that (36) holds and $\rho(\underline{\mathbf{C}}^\circ) < 1$. We have for each i ,

$$e_i^\circ - \underline{e}_i^\circ = \frac{1}{N} \operatorname{tr} \mathbf{R}_i \mathbf{D}^{\circ-1} \sum_{j=1}^K (e_j^\circ - \underline{e}_j^\circ) c_j \int \frac{\tau^2}{(1 + c_j \tau e_j^\circ)(1 + c_j \tau \underline{e}_j^\circ)} dF^{\mathbf{T}_j}(\tau) \mathbf{R}_j \underline{\mathbf{D}}^{\circ-1}.$$

Thus with $\mathbf{A} = (a_{ij})$ where

$$a_{ij} = \frac{1}{N} \text{tr} \mathbf{R}_i \mathbf{D}^{\circ-1} \mathbf{R}_j \underline{\mathbf{D}}^{\circ-1} c_j \int \frac{\tau^2}{(1 + c_j \tau e_j^\circ)(1 + c_j \tau \underline{e}_j^\circ)} dF^{\mathbf{T}_j}(\tau), \quad (37)$$

we have

$$\mathbf{e}^\circ - \underline{\mathbf{e}}^\circ = \mathbf{A}(\mathbf{e}^\circ - \underline{\mathbf{e}}^\circ), \quad (38)$$

which means, if $\mathbf{e}^\circ \neq \underline{\mathbf{e}}^\circ$, then \mathbf{A} has an eigenvalue equal to 1.

Applying Cauchy-Schwarz we have

$$\begin{aligned} |a_{ij}| &\leq \left(\frac{1}{N} \mathbf{R}_i \mathbf{D}^{\circ-1} \mathbf{R}_j \mathbf{D}^{\circ-H} \int \frac{\tau^2}{|1 + c_j \tau e_j^\circ|^2} dF^{\mathbf{T}_j}(\tau) \right)^{\frac{1}{2}} \left(\frac{1}{N} \mathbf{R}_i \underline{\mathbf{D}}^{\circ-1} \mathbf{R}_j \underline{\mathbf{D}}^{\circ-H} \int \frac{\tau^2}{|1 + c_j \tau \underline{e}_j^\circ|^2} dF^{\mathbf{T}_j}(\tau) \right)^{\frac{1}{2}} \\ &= c_{ij}^{\circ 1/2} \underline{c}_{ij}^{\circ 1/2}. \end{aligned}$$

Therefore from Lemmas 10 and 11 we get

$$\rho(\mathbf{A}) \leq \rho(c_{ij}^{\circ \frac{1}{2}} \underline{c}_{ij}^{\circ \frac{1}{2}}) \leq \rho(\mathbf{C}^\circ)^{\frac{1}{2}} \rho(\underline{\mathbf{C}}^\circ)^{\frac{1}{2}} < 1;$$

a contradiction to the statement \mathbf{A} has an eigenvalue equal to 1. Consequently we have $\mathbf{e} = \underline{\mathbf{e}}$.

The same reasoning can be applied to $z < 0$, with $e_i^\circ > 0$. In this case matrix \mathbf{A} remains the same. The step (35) is now replaced by taking e_i° , instead of its imaginary part, using the same line of reasoning. This leads to the same matrix \mathbf{C}° with (36) remaining true with \mathbf{b}° replaced by another positive vector. The conclusion $\rho(\mathbf{A}) < 1$ therefore remains.

Let $\mathbf{e}_N = (e_{N,1}, \dots, e_{N,K})^\top$ and $\mathbf{e}_N^\circ = (e_{N,1}^\circ, \dots, e_{N,K}^\circ)$ denote the vector solution to (34) for each N . We will show for any $\ell > 0$, almost surely

$$\lim_{N \rightarrow \infty} \log^\ell N (\mathbf{e}_N - \mathbf{e}_N^\circ) \rightarrow \mathbf{0}. \quad (39)$$

We have

$$\mathbf{e}_N^\circ = \left(\frac{1}{N} \text{tr} \mathbf{R}_1 \mathbf{D}^{\circ-1}, \dots, \frac{1}{N} \text{tr} \mathbf{R}_K \mathbf{D}^{\circ-1} \right)^\top.$$

Let $\mathbf{w}^e = \mathbf{w}_N^e = -(w_1^e, \dots, w_K^e)^\top$. Then we can write

$$\mathbf{e}_N = \left(\frac{1}{N} \text{tr} \mathbf{R}_1 \mathbf{D}^{-1}, \dots, \frac{1}{N} \text{tr} \mathbf{R}_K \mathbf{D}^{-1} \right)^\top + \mathbf{w}^e.$$

Therefore

$$\mathbf{e}_N - \mathbf{e}_N^\circ = \mathbf{A}(N)(\mathbf{e}_N - \mathbf{e}_N^\circ) + \mathbf{w}^e,$$

where $\mathbf{A}(N) = (a_{ij}(N))$ with

$$a_{ij}(N) = \frac{1}{N} \text{tr} \mathbf{R}_i \mathbf{D}^{-1} \mathbf{R}_j \mathbf{D}^{\circ-1} c_j \int \frac{\tau^2}{(1 + c_j \tau e_{N,j})(1 + c_j \tau e_{N,j}^\circ)} dF^{\mathbf{T}_j}(\tau).$$

We let $e_{N,2}^\circ$, $b_{ij}^\circ(N)$, $\mathbf{C}^\circ(N)$, $b_{N,i}^\circ$, and \mathbf{b}_N° , denote the quantities from above, reflecting now their dependence on N . Let $\mathbf{C}(N) = (c_{ij}(N))$ be $K \times K$ with

$$c_{ij}(N) = \frac{1}{N} \text{tr} \mathbf{R}_j \mathbf{D}^{-H} \mathbf{R}_i \mathbf{D}^{-1} c_j \int \frac{\tau^2}{|1 + c_j \tau e_{N,j}|^2} dF^{\mathbf{T}_j}(\tau).$$

Let $\mathbf{e}_{N,2} = \Im[\mathbf{e}_N]$ and $\mathbf{w}_2^e = \Im[\mathbf{w}^e]$. Define $\mathbf{b}_N = (b_{N,1}, \dots, b_{N,K})^\top$ with

$$b_{N,i} = \frac{1}{N} \text{tr} \mathbf{D}^{-\text{H}} \mathbf{R}_i \mathbf{D}^{-1}.$$

Then, as above we find that

$$\mathbf{e}_{N,2} = \mathbf{C}(N) \mathbf{e}_{N,2} + v \mathbf{b}_N + \mathbf{w}_2^e. \quad (40)$$

Using (18) and (21) we see there exists a constant $K_1 > 0$ for which

$$c_{ij}^\circ(N) \leq K_1 \log^3 N b_{N,i}^\circ$$

and

$$c_{ij}(N) \leq K_1 \log^3 N b_{N,i}$$

$$c_{ij}(N) \leq K_1 \log^4 N$$

for each i, j . Therefore, from (36) we see there exists $\hat{K} > 0$ for which

$$e_{N,i}^\circ \leq \hat{K} \log^4 N v b_{N,i}^\circ. \quad (41)$$

Let \mathbf{x} be such that \mathbf{x}^\top is a left eigenvector of $\mathbf{C}^\circ(N)$ corresponding to eigenvalue $\rho(\mathbf{C}^\circ(N))$, guaranteed by Lemma 12. Then from (40) we have

$$\mathbf{x}^\top \mathbf{e}_{N,2}^\circ = \rho(\mathbf{C}^\circ(N)) \mathbf{x}^\top \mathbf{e}_{N,2}^\circ + v \mathbf{x}^\top \mathbf{b}_N^\circ. \quad (42)$$

Using (42) we have

$$1 - \rho(\mathbf{C}^\circ(N)) = \frac{v \mathbf{x}^\top \mathbf{b}_N^\circ}{\mathbf{x}^\top \mathbf{e}_{N,2}^\circ} \geq (\hat{K} \log^4 N)^{-1}. \quad (43)$$

Fix an $\ell > 0$ and consider a realisation for which $\log^{\ell+3+p} N \mathbf{w}_N^e \rightarrow 0$, as $N \rightarrow \infty$, where $p \geq 12K - 7$. We will show for all N large

$$\rho(\mathbf{C}(N)) \leq 1 + (\hat{K} \log^4 N)^{-1}. \quad (44)$$

For each N we rearrange the entries of $\mathbf{e}_{N,2}$, $v \mathbf{b}_m + \mathbf{w}_2^e$, and $\mathbf{C}(n)$ depending on whether the i^{th} entry of $v \mathbf{b}_m + \mathbf{w}_2^e$ is greater than, or less than or equal to zero. We can therefore assume

$$\mathbf{C} = \begin{pmatrix} \mathbf{C}_{11}(N) & \mathbf{C}_{12}(N) \\ \mathbf{C}_{21}(N) & \mathbf{C}_{22}(N) \end{pmatrix},$$

where $\mathbf{C}_{11}(N)$ is $k_1 \times k_1$, $\mathbf{C}_{22}(N)$ is $k_2 \times k_2$, $\mathbf{C}_{12}(N)$ is $k_1 \times k_2$, and $\mathbf{C}_{21}(N)$ is $k_2 \times k_1$. From Lemma 9 we have $\rho(\mathbf{C}_{11}(N)) < 1$. If $v b_{N,i} + \mathbf{w}_{2,i}^e \leq 0$, then necessarily $v b_{N,i} \leq |\mathbf{w}_N^e| \leq K_1 (\log n)^{-(3+p)}$, and so from (41) we have the entries of $\mathbf{C}_{21}(N)$ and $\mathbf{C}_{22}(N)$ bounded by $K_1 (\log N)^{-p}$. We may assume for all N large $0 < k_1 < K$, since otherwise we would have $\rho(\mathbf{C}(N)) < 1$.

We seek an expression for $\det(\mathbf{C}(N) - \lambda \mathbf{I}_N)$ in which Lemma 14 can be used. We consider N large enough so that, for $|\lambda| \geq 1/2$, we have $(\mathbf{C}_{22}(N) - \lambda \mathbf{I}_N)^{-1}$ existing with entries uniformly bounded. We have

$$\begin{aligned} \det(\mathbf{C}(N) - \lambda \mathbf{I}) &= \det \left[\begin{pmatrix} \mathbf{I} & -\mathbf{C}_{12}(N)(\mathbf{C}_{22}(N) - \lambda \mathbf{I})^{-1} \\ 0 & \mathbf{I} \end{pmatrix} \begin{pmatrix} \mathbf{C}_{11}(N) - \lambda \mathbf{I} & \mathbf{C}_{12}(N) \\ \mathbf{C}_{21}(N) & \mathbf{C}_{22}(N) - \lambda \mathbf{I} \end{pmatrix} \right] \\ &= \det \begin{pmatrix} \mathbf{C}_{11}(N) - \lambda \mathbf{I} - \mathbf{C}_{12}(N)(\mathbf{C}_{22}(N) - \lambda \mathbf{I})^{-1} \mathbf{C}_{21}(N) & 0 \\ \mathbf{C}_{21}(N) & \mathbf{C}_{22}(N) - \lambda \mathbf{I} \end{pmatrix} \\ &= \det(\mathbf{C}_{11}(N) - \lambda \mathbf{I} - \mathbf{C}_{12}(N)(\mathbf{C}_{22}(N) - \lambda \mathbf{I})^{-1} \mathbf{C}_{21}(N)) \det(\mathbf{C}_{22}(N) - \lambda \mathbf{I}). \end{aligned}$$

We see then that for $\lambda = \rho(\mathbf{C}(N))$ real and greater than 1,

$$\det(\mathbf{C}_{11}(N) - \lambda \mathbf{I} - \mathbf{C}_{12}(N)(\mathbf{C}_{22}(N) - \lambda \mathbf{I})^{-1} \mathbf{C}_{21}(N)) \quad (45)$$

must be zero.

Notice that from (41), the entries of $\mathbf{C}_{12}(N)(\mathbf{C}_{22}(N) - \lambda \mathbf{I})^{-1} \mathbf{C}_{21}(N)$ can be made smaller than any negative power of $\log N$ for p sufficiently large. Notice also that the diagonal elements of $\mathbf{C}_{11}(N)$ are all less than 1. From this, Lemma 13 and (41), we see that $\rho(\mathbf{C}(N)) \leq K_1 \log^4 N$. The determinant in (45) can be written as

$$\det(\mathbf{C}_{11}(N) - \lambda \mathbf{I}) + g(\lambda),$$

where $g(\lambda)$ is a sum of products, each containing at least one entry from $\mathbf{C}_{12}(N)(\mathbf{C}_{22}(N) - \lambda \mathbf{I})^{-1} \mathbf{C}_{21}(N)$. Again, from (41) we see that for all $|\lambda| \geq 1/2$, $g(\lambda)$ can be made smaller than any negative power of $\log N$ by making p sufficiently large. Choose p so that $|g(\lambda)| < (\widehat{K} \log N)^{-4k_1}$ for these λ . It is clear that any $p > 8k_1 + 4$ will suffice. Let $\lambda_1, \dots, \lambda_{k_1}$ denote the eigenvalues of \mathbf{C}_{11} . Since $\rho(\mathbf{C}_{11}) < 1$, we see that for $|\lambda| \geq (\widehat{K} \log N)^{-4}$, we have

$$\begin{aligned} |\det(\mathbf{C}_{11}(N) - \lambda \mathbf{I})| &= \left| \prod_{i=1}^{k_1} (\lambda_i - \lambda) \right| \\ &> (\widehat{K} \log N)^{-4k_1}. \end{aligned}$$

Thus with $f(\lambda) = \det(\mathbf{C}_{11}(N) - \lambda \mathbf{I})$, a polynomial, and $g(\lambda)$ being a rational function, we have the conditions of Lemma 14 being met on any rectangle C , with vertical lines going through $((\widehat{K} \log N)^{-4}, 0)$ and $(K_1(\log N)^4, 0)$. Therefore, since $f(\lambda)$ has no zeros inside C , neither does $\det(\mathbf{C}(N) - \lambda \mathbf{I})$. Thus we get (44). As before we see that

$$|a_{ij}(N)| \leq c_{ij}^{1/2}(N) c_{ij}^{\circ 1/2}(N).$$

Therefore, from (43), (44), and Lemmas 10 and 11, we have for all N large

$$\rho(\mathbf{A}(N)) \leq \left(\frac{\widehat{K}^2 \log^8 N - 1}{\widehat{K}^2 \log^8 N} \right)^{\frac{1}{2}}. \quad (46)$$

For these N we have then $\mathbf{I} - \mathbf{A}(N)$ invertible, and so

$$\mathbf{e}_N - \mathbf{e}_N^0 = (\mathbf{I} - \mathbf{A}(N))^{-1} \mathbf{w}^e.$$

By (18) and (21) we have the entries of $\mathbf{A}(N)$ bounded by $K_1 \log^4 N$. Notice also, from (46)

$$|\det(\mathbf{I} - \mathbf{A}(N))| \geq (1 - \rho(\mathbf{A}(N)))^K \geq \left(\widehat{K}^2 \log^8 N \left(1 + \frac{\widehat{K}^2 \log^8 N - 1}{\widehat{K}^2 \log^8 N} \right)^{\frac{1}{2}} \right)^{-K} \geq (2\widehat{K}^2 \log^8 N)^{-K}.$$

When considering the inverse of a square matrix in terms of its adjoint divided by its determinant, we see that the entries of $(\mathbf{I} - \mathbf{A}(N))^{-1}$ are bounded by

$$\frac{(K-1)!K_1(\log N)^{4(K-1)}}{|\det(\mathbf{I} - \mathbf{A}(N))|} \leq K_3(\log N)^{12K-4}.$$

Therefore, since $p \geq 12K - 7 (> 8k_1 + 4)$, (39) follows on this realisation, an event which occurs with probability one.

Letting $m_N^\circ = \frac{1}{N} \text{tr} \mathbf{D}^{\circ-1}$, we have

$$m_N - m_N^\circ = \vec{\gamma}^\top (\mathbf{e}_N - \mathbf{e}_N^\circ),$$

where $\vec{\gamma} = (\gamma_1, \dots, \gamma_K)^\top$ with

$$\gamma_j = \int \frac{c_N \tau^2}{(1 + c_N \tau e_{N,j})(1 + c_n \tau e_{N,j}^\circ)} dF^{\mathbf{T}_N}(\tau) \frac{\text{tr} \mathbf{D}^{-1} \mathbf{R}_j \mathbf{D}^{\circ-1}}{N}.$$

From (18) and (21) we get each $|\gamma_j| \leq c_N |z|^2 v^{-4} \log^3 N$. Therefore from (39) and the fact that $w_N^m \rightarrow 0$, we have

$$m_N - m_N^\circ \rightarrow 0,$$

almost surely, as $N \rightarrow \infty$.

This completes the proof.

APPENDIX B PROOF OF THEOREM 2

We first prove that $\mathcal{V}_N^\circ(x)$ as defined in Equation (12) verifies

$$\mathcal{V}_N^\circ(x) = \int_x^\infty \left(\frac{1}{w} - m_N^\circ(-w) \right) dw \quad (47)$$

and then we prove that, under the conditions of Theorem 2, $\mathcal{V}^\circ(x)$ defined as such verifies

$$\mathcal{V}_N^\circ(x) - \mathcal{V}_N(x) \xrightarrow{\text{a.s.}} 0. \quad (48)$$

A. Proof of (47)

First, write $e_i(z)$ under the symmetric form

$$e_i(z) = \frac{1}{N} \text{tr} \mathbf{R}_i \left(-z \left[\mathbf{I}_N + \sum_{k=1}^K \delta_k \mathbf{R}_k \right] \right)^{-1}$$

$$\delta_i(z) = \frac{1}{n_i} \text{tr} \mathbf{T}_i (-z [\mathbf{I}_{n_i} + c_i e_i(z) \mathbf{T}_i])^{-1}$$

and then for $m_N^\circ(z)$,

$$m_N^\circ(z) = \frac{1}{N} \operatorname{tr} \left(-z \left[\mathbf{I}_N + \sum_{k=1}^K \delta_k \mathbf{R}_k \right] \right)^{-1}.$$

Now, notice that

$$\begin{aligned} \frac{1}{z} - m_N^\circ(-z) &= \frac{1}{N} \left((z\mathbf{I})^{-1} - \left(z \left[\mathbf{I}_N + \sum_{k=1}^K \delta_k \mathbf{R}_k \right] \right)^{-1} \right) \\ &= \sum_{k=1}^K \delta_k(-z) \cdot e_k(-z). \end{aligned}$$

Since the Shannon transform $\mathcal{V}(x)$ satisfies $\mathcal{V}(x) = \int_x^{+\infty} [w^{-1} - m_N(-w)] dw$, we need to find an integral form for $\sum_{k=1}^K \delta_k(-z) \cdot e_k(-z)$. Notice now that

$$\begin{aligned} \frac{d}{dz} \frac{1}{N} \log \det \left(\mathbf{I}_N + \sum_{k=1}^K \delta_k(-z) \mathbf{R}_k \right) &= -z \sum_{k=1}^K e_k(-z) \cdot \delta_k'(-z) \\ \frac{d}{dz} \frac{1}{N} \log \det (\mathbf{I}_{n_k} + c_k e_k(-z) \mathbf{T}_k) &= -z \cdot e_k'(-z) \cdot \delta_k(-z) \\ \frac{d}{dz} \left(z \sum_{k=1}^K \delta_k(-z) e_k(-z) \right) &= \sum_{k=1}^K \delta_k(-z) e_k(-z) - z \sum_{k=1}^K \delta_k'(-z) \cdot e_k(-z) + \delta_k(-z) \cdot e_k'(-z). \end{aligned}$$

Combining the last three lines, we have

$$\begin{aligned} \sum_{k=1}^K \delta_k(-z) e_k(-z) &= \\ \frac{d}{dz} \left[-\frac{1}{N} \log \det \left(\mathbf{I}_N + \sum_{k=1}^K \delta_k(-z) \mathbf{R}_k \right) - \sum_{k=1}^K \frac{1}{N} \log \det (\mathbf{I}_{n_k} + c_k e_k(-z) \mathbf{T}_k) + z \sum_{k=1}^K \delta_k(-z) e_k(-z) \right], \end{aligned}$$

which after integration leads to

$$\begin{aligned} \int_z^{+\infty} \left(\frac{1}{w} - m_N^\circ(-w) \right) dw &= \\ \frac{1}{N} \log \det \left(\mathbf{I}_N + \sum_{k=1}^K \delta_k(-z) \mathbf{R}_k \right) + \sum_{k=1}^K \frac{1}{N} \log \det (\mathbf{I}_{n_k} + c_k e_k(-z) \mathbf{T}_k) - z \sum_{k=1}^K \delta_k(-z) e_k(-z), \end{aligned} \quad (49)$$

which is exactly the right-hand side of (12).

B. Proof of (48)

Consider now the existence of a nonrandom α and for each N a non-negative integer r_N for which

$$\max_{i \leq K} \max(\lambda_{r_N+1}^{\mathbf{T}_i}, \lambda_{r_N+1}^{\mathbf{R}_i}) \leq \alpha$$

(eigenvalues also arranged in non-increasing order). Then for each i

$$\begin{aligned} \lambda_{2r_N+1}^{\mathbf{R}_i^{\frac{1}{2}} \mathbf{X}_i \mathbf{T}_i \mathbf{X}_i^H \mathbf{R}_i^{\frac{1}{2}}} &= (s_{2r_N+1}^{\mathbf{R}_i^{\frac{1}{2}} \mathbf{X}_i \mathbf{T}_i^{\frac{1}{2}}})^2 \\ &\leq \alpha^2 \|\mathbf{X}_i \mathbf{X}_i^H\| \end{aligned}$$

and then we have, from Lemma 15,

$$\lambda_{2Kr_N+1}^{\mathbf{B}_N} \leq \alpha^2(\|\mathbf{X}_1\mathbf{X}_1^H\| + \cdots + \|\mathbf{X}_K\mathbf{X}_K^H\|).$$

We can in fact consider that the spectral norms of the \mathbf{X}_i are bounded in the limit. Either Gaussian assumptions on the components, or finite fourth moment, but all coming from doubly infinite arrays (remember though that we need the right-unitary invariance structure of \mathbf{X}_i). Because of assumption 5 in Corollary 1, we can, by enlarging the sample space, assume each \mathbf{X}_i is embedded in an $N \times n'_i$ matrix \mathbf{X}'_i , where $N/n'_i \rightarrow a$ as $N \rightarrow \infty$. Then, with probability one (see e.g. [16]),

$$\begin{aligned} \limsup_N \lambda_{2Kr_N+1}^{\mathbf{B}_N} &\leq \limsup_N \alpha^2(\|\mathbf{X}'_1\mathbf{X}'_1{}^H\| + \cdots + \|\mathbf{X}'_K\mathbf{X}'_K{}^H\|) \\ &\leq \alpha^2 \frac{Kb}{a} (1 + \sqrt{a})^2. \end{aligned} \quad (50)$$

Let a° be any real greater than $\alpha^2 K \frac{b}{a} (1 + \sqrt{a})^2$.

Since $\mathbf{S} = 0$ here, it follows as in [22] that $\{F^{\mathbf{B}_n}\}$ is almost surely tight. Let F_N° denote the distribution function having Stieltjes transform m_N° , and let f on $[0, \infty)$ be a continuous function. Then the function

$$f_{a^\circ}(x) = \begin{cases} f(x) & , x \leq a^\circ \\ f(a^\circ) & , x > a^\circ \end{cases}$$

is bounded and continuous. Therefore, with probability 1,

$$\int f_{a^\circ}(x) dF_N(x) - \int f_{a^\circ}(x) dF_N^\circ(x) \rightarrow 0,$$

as $N \rightarrow \infty$.

Suppose now $r_N = o(N)$. Then, since almost surely there are at most $2Kr_N$ eigenvalues greater than a° for all N large, any converging subsequence of $\{F_N^\circ\}$ must have some mass lying on $[0, a^\circ]$. This implies, with probability 1,

$$\frac{1}{N} \sum_{\lambda_i \leq a^\circ} f(\lambda_i) - \int_{[0, a^\circ]} f(x) dF_N^\circ(x) \rightarrow 0,$$

as $N \rightarrow \infty$.

Let b_N be a bound on the spectral norms of the \mathbf{T}_i and \mathbf{R}_i . Then

$$\|\mathbf{B}_n\| \leq b_N^2(\|\mathbf{X}'_1\mathbf{X}'_1{}^H\| + \cdots + \|\mathbf{X}'_K\mathbf{X}'_K{}^H\|). \quad (51)$$

Fix a number $\beta > \frac{Kb}{a}(1 + \sqrt{a})^2$, and let $a_N = b_N^2\beta$. Suppose also that f is increasing and that $f(a_N)r_N = o(N)$.

Then

$$\int f(x) dF^{\mathbf{B}_n}(x) - \frac{1}{N} \sum_{\lambda_i \leq a^\circ} f(\lambda_i) \rightarrow 0,$$

almost surely, as $N \rightarrow \infty$. Therefore, with probability 1,

$$\int f(x) dF_N(x) - \int_{[0, a^\circ]} f(x) dF_N^\circ(x) \rightarrow 0,$$

as $N \rightarrow \infty$.

For any N we consider, for $j = 1, 2, \dots$, the $jN \times jN$ matrix $\mathbf{B}_{N,j}$ formed, as before, from block diagonal matrices and $jN \times jn_i$ matrices of i.i.d. variables. Then with probability 1, $F^{\mathbf{B}_{N,j}}$ converges weakly to F_N° as $j \rightarrow \infty$. Properties on the eigenvalues of $\mathbf{B}_{N,j}$ will thus yield properties of F_N° .

By considering the bound on $\|\mathbf{B}_{n,j}\|$ analogous to (51), we must have $F_N^\circ(a_N) = 1$ for all N large.

Similar to (50) we see that, with probability 1

$$\limsup_j \lambda_{2Kj r_N + 1}^{\mathbf{B}_{N,j}} \leq a^2((1 + \sqrt{c_1})^2 + \dots + (1 + \sqrt{c_K})^2),$$

this latter number being less than a° for all N large.

At this point we will use the fact that for probability measures P_N, P on \mathbb{R} with P_N converging weakly to P , we have (see e.g. [31])

$$\liminf_N P_N(G) \geq P(G)$$

for any open set G . Thus, with $G = (a^\circ, \infty)$ we see that, with probability 1, for all N large

$$\begin{aligned} F_N^\circ((a^\circ, \infty)) &= 1 - F_N^\circ(a^\circ) \leq \liminf_j F^{\mathbf{B}_{N,j}}((a^\circ, \infty)) \\ &\leq 2K r_N / N. \end{aligned}$$

Therefore, for all N large

$$\int_{(a^\circ, \infty)} f(x) dF_N^\circ(x) \leq f(a_N) 2K r_N / N \rightarrow 0,$$

as $N \rightarrow \infty$.

Therefore, we conclude that, $\int f(x) dF_N^\circ(x)$ is bounded, and with probability 1

$$\int f(x) dF_N(x) - \int f(x) dF_N^\circ(x) \rightarrow 0,$$

as $N \rightarrow \infty$. This concludes the proof.

APPENDIX C

PROOF OF PROPOSITION 2

The proof stems from the following result,

Proposition 4: $f(\mathbf{P}_1, \dots, \mathbf{P}_K)$ is a strictly concave matrix in the Hermitian nonnegative definite matrices $\mathbf{P}_1, \dots, \mathbf{P}_K$, if and only if, for any couples $(\mathbf{P}_{1_a}, \mathbf{P}_{1_b}), \dots, (\mathbf{P}_{K_a}, \mathbf{P}_{K_b})$ of Hermitian nonnegative definite matrices, the function

$$\phi(\lambda) = f(\lambda \mathbf{P}_{1_a} + (1 - \lambda) \mathbf{P}_{1_b}, \dots, \lambda \mathbf{P}_{K_a} + (1 - \lambda) \mathbf{P}_{K_b})$$

is strictly concave.

Denote

$$\bar{\mathcal{V}}_N^\circ(\lambda) = \mathcal{V}_N^\circ(\lambda \mathbf{P}_{1_a} + (1 - \lambda) \mathbf{P}_{1_b}, \dots, \lambda \mathbf{P}_{|S|_a} + (1 - \lambda) \mathbf{P}_{|S|_b})$$

and consider a set $(\delta_k, e_k, \mathbf{P}_{i_1}, \dots, \mathbf{P}_{i_{|S|}})$ which satisfies the system of equations (53)-(55). Then, from remark (56) and (57),

$$\begin{aligned} \frac{d\bar{\mathcal{V}}_N^\circ}{d\lambda} &= \sum_{k \in S} \frac{\partial \bar{\mathcal{V}}}{\partial \delta_k} \frac{\partial \delta_k}{\partial \lambda} + \frac{\partial \bar{\mathcal{V}}}{\partial e_k} \frac{\partial e_k}{\partial \lambda} + \frac{\partial \bar{\mathcal{V}}}{\partial \lambda} \\ &= \frac{\partial \bar{\mathcal{V}}}{\partial \lambda}, \end{aligned}$$

where

$$\bar{V} : (\delta_1, \dots, \delta_{|S|}, e_1, \dots, e_{|S|}, \lambda) \mapsto \bar{\mathcal{V}}_N^\circ(\lambda). \quad (52)$$

Mere derivations of \bar{V} lead then to

$$\frac{\partial^2 \bar{V}}{\partial \lambda^2} = - \sum_{i \in S} (c_i^2 e_i^2) \frac{1}{N} \text{tr} (\mathbf{I} + c_i e_i \mathbf{R}_i \mathbf{P}_i)^{-2} (\mathbf{R}_i (\mathbf{P}_{i_a} - \mathbf{P}_{i_b}))^2.$$

Since $e_i > 0$ on the strictly negative real axis, if any of the \mathbf{R}_i 's is positive definite, then, for all nonnegative definite couples $(\mathbf{P}_{i_a}, \mathbf{P}_{i_b})$, such that $\mathbf{P}_{i_a} \neq \mathbf{P}_{i_b}$, $\bar{\mathcal{V}}_N^\circ < 0$. Then, from Proposition 4, the deterministic approximate \mathcal{V}_N° is strictly concave in $\mathbf{P}_1, \dots, \mathbf{P}_{|S|}$ if any of the \mathbf{R}_i matrices is invertible.

APPENDIX D

PROOF OF PROPOSITION 3

The proof of Proposition 3 recalls the proof from [9], Proposition 5. Let us define the functions

$$\begin{aligned} \mathcal{V}_N^\circ(\mathbf{P}_1, \dots, \mathbf{P}_{|S|}) &= \sum_{k \in S} \frac{1}{N} \log \det (\mathbf{I}_{n_k} + c_k e_k \mathbf{R}_k \mathbf{P}_k) \\ &\quad + \frac{1}{N} \log \det \left(\mathbf{I}_N + \sum_{k \in S} \delta_k \mathbf{T}_k \right) \\ &\quad - \sigma^2 \sum_{k=1}^K \delta_k (-\sigma^2) e_k (-\sigma^2), \end{aligned} \quad (53)$$

where

$$e_i = e_i(\mathbf{P}_1, \dots, \mathbf{P}_{|S|}) = \frac{1}{N} \text{tr} \mathbf{T}_i \left(\sigma^2 \left[\mathbf{I}_N + \sum_{k \in S} \delta_k \mathbf{T}_k \right] \right)^{-1} \quad (54)$$

$$\delta_i = \delta_i(\mathbf{P}_1, \dots, \mathbf{P}_{|S|}) = \frac{1}{n_i} \text{tr} \mathbf{R}_i \mathbf{P}_i (\sigma^2 [\mathbf{I}_{n_i} + c_i e_i(z) \mathbf{R}_i \mathbf{P}_i])^{-1} \quad (55)$$

and $V : (\mathbf{P}_1, \dots, \mathbf{P}_{|S|}, \delta_1, \dots, \delta_{|S|}, e_1, \dots, e_{|S|}) \mapsto \mathcal{V}_N^\circ(\mathbf{P}_1, \dots, \mathbf{P}_{|S|})$. Then we need only prove that, for all $k \in S$,

$$\begin{aligned} \frac{\partial V}{\partial \delta_k}(\mathbf{P}_1, \dots, \mathbf{P}_{|S|}, \delta_1^\circ, \dots, \delta_{|S|}^\circ, e_1^\circ, \dots, e_{|S|}^\circ) &= 0 \\ \frac{\partial V}{\partial e_k}(\mathbf{P}_1, \dots, \mathbf{P}_{|S|}, \delta_1^\circ, \dots, \delta_{|S|}^\circ, e_1^\circ, \dots, e_{|S|}^\circ) &= 0. \end{aligned}$$

Remark then that

$$\frac{\partial V}{\partial \delta_k}(\mathbf{P}_1, \dots, \mathbf{P}_{|S|}, \delta_1, \dots, \delta_{|S|}, e_1, \dots, e_{|S|}) = \frac{1}{N} \text{tr} \left[\left(\mathbf{I} + \sum_{i \in S} \delta_i \mathbf{T}_i \right)^{-1} \mathbf{T}_k \right] - \sigma^2 e_k \quad (56)$$

$$\frac{\partial V}{\partial e_k}(\mathbf{P}_1, \dots, \mathbf{P}_{|S|}, \delta_1, \dots, \delta_{|S|}, e_1, \dots, e_{|S|}) = c_k \frac{1}{N} \text{tr} \left[(\mathbf{I} + c_k e_k \mathbf{R}_i \mathbf{P}_i)^{-1} \mathbf{R}_k \mathbf{P}_k \right] - \sigma^2 \delta_k, \quad (57)$$

both being null whenever, for all k , $e_k = e_k(-\sigma^2, \mathbf{P}_1, \dots, \mathbf{P}_{|\mathcal{S}|})$ and $\delta_k = \delta_k(-\sigma^2, \mathbf{P}_1, \dots, \mathbf{P}_{|\mathcal{S}|})$, which is true in particular for the unique power optimal solution $\mathbf{P}_1^\circ, \dots, \mathbf{P}_{|\mathcal{S}|}^\circ$ whenever $e_k = e_k^\circ$ and $\delta_k = \delta_k^\circ$.

When, for all k , $e_k = e_k^\circ$, $\delta_k = \delta_k^\circ$, the maximum of V over the \mathbf{P}_k is then obtained by maximising the expressions $\log \det(\mathbf{I}_{n_k} + c_k e_k^\circ \mathbf{R}_k \mathbf{P}_k)$ over \mathbf{P}_k . From the inequality (see e.g. [2])

$$\det(\mathbf{I}_{n_k} + c_k e_k^\circ \mathbf{R}_k \mathbf{P}_k) \leq \prod_{i=1}^{n_k} (\mathbf{I}_{n_k} + c_k e_k^\circ \mathbf{R}_k \mathbf{P}_k)_{ii},$$

where, only here, we denote $(\mathbf{X})_{ii}$ the entry (i, i) of matrix \mathbf{X} . The equality is obtained if and only if $\mathbf{I}_{n_k} + c_k e_k^\circ \mathbf{R}_k \mathbf{P}_k$ is diagonal. The equality case arises for \mathbf{P}_k and $\mathbf{R}_k = \mathbf{U}_k \mathbf{D}_k \mathbf{U}_k^H$ co-diagonalizable. In this case, denoting $\mathbf{P}_k = \mathbf{U}_k \mathbf{Q}_k \mathbf{U}_k^H$, the entries of \mathbf{Q}_k , constrained by $\frac{1}{n_k} \text{tr}(\mathbf{Q}_k) = P_k$ are solutions of the classical optimisation problem under constraint,

$$\sup_{\substack{\mathbf{Q}_k \\ \frac{1}{n_k} \text{tr}(\mathbf{Q}_k) \leq P_k}} \log \det(\mathbf{I}_{n_k} + c_k e_k^\circ \mathbf{Q}_k \mathbf{D}_k),$$

whose solution is given by the classical water-filling algorithm. Hence (15).

APPENDIX E

PROOF OF PROPOSITION 1

The convergence of the fixed-point algorithm follows the same line of proof as the uniqueness in Section A-E. We prove the convergence for $z \in \mathbb{C}^+$, although this can be easily generalised. If one considers the difference $\mathbf{e}^{n+1} - \mathbf{e}^n$, where $\mathbf{e}^n = (e_1^n, \dots, e_K^n)$, instead of $\mathbf{e}^\circ - \underline{\mathbf{e}}^\circ$, the same development as in Section A-E leads to

$$\mathbf{e}^{n+1} - \mathbf{e}^n = \mathbf{A}_n (\mathbf{e}^n - \mathbf{e}^{n-1})$$

for $n \geq 1$, where \mathbf{A}_n is defined, similarly as in (37), as $\mathbf{A}_n = (a_{ij}^n)$, with a_{ij}^n defined by

$$a_{ij}^n = \frac{1}{N} \text{tr} \mathbf{R}_i \mathbf{D}_{n-1}^{-1} \mathbf{R}_j \mathbf{D}_n^{-1} c_j \int \frac{\tau^2}{(1 + c_j \tau e_j^{n-1})(1 + c_j \tau e_j^n)} dF^{\mathbf{T}_j}(\tau),$$

where \mathbf{D}_n is \mathbf{D} for $e_j(z)$ replaced by $e_j^n(z)$.

From Cauchy-Schwarz inequality, and the different bounds on the \mathbf{D}_n , \mathbf{R}_k and \mathbf{T}_k matrices used so far, we have

$$a_{ij}^n \leq \frac{|z|^2 c_j \log N^4}{v^4 N},$$

with $v = \Im[z]$. Denoting $c_0 = \max(c_j)$, we then have that

$$\max_j (e_j^{n+1} - e_j^n) < K \frac{|z|^2 c_0}{v^4} \leq \frac{\log N^4}{N} \max_j (e_j^n - e_j^{n-1}).$$

Let $0 < \varepsilon < 1$, and take now a countable set $\{z_1, z_2, \dots\}$, $v_k = \Im[z_k]$, such that $K \frac{|z_k|^2 c_0 \log N^4}{v_k^4} < 1 - \varepsilon$ for all z_k (this is possible by letting $v_k > 0$ be large enough). On this countable set, the sequences $\{\mathbf{e}^n\}$ are therefore Cauchy sequences on \mathbb{C}^K : they all converge. Since the e_j^n are holomorphic and bounded on every compact set included in $\mathbb{C} \setminus \mathbb{R}^+$, from Vitali's convergence theorem [30], the function $e_j^n(z)$ converges on such compact sets. Now, from the fact that we forced the initialisation step to be $e_j^0 = -1/z$, e_j^0 is the Stieltjes transform of a distribution function at

point z . It now suffices to verify that, if e_j^n is the Stieltjes transform of a distribution function at point z , then so is e_j^{n+1} . This requires to verify that $z \in \mathbb{C}^+$, $e_j^n \in \mathbb{C}^+$ implies $e_j^{n+1} \in \mathbb{C}^+$, $z \in \mathbb{C}^+$, $ze_j^n \in \mathbb{C}^+$ implies $ze_j^{n+1} \in \mathbb{C}^+$, and $\lim_{y \rightarrow \infty} -ye_j^n(iy) < \infty$ implies that $\lim_{y \rightarrow \infty} -ye_j^{n+1}(iy) < \infty$. This follows directly from the definition of e_j^n . From the dominated convergence theorem, we then also have that the limit of e_j^n is a Stieltjes transform that is solution to (6). From the uniqueness of the Stieltjes transform, solution to (6) (this follows from the pointwise uniqueness on \mathbb{C}^+ and the fact that the Stieltjes transform is holomorphic on all compact sets of $\mathbb{C} \setminus \mathbb{R}^+$), we then have that e_j^n converges for all j and $z \in \mathbb{C} \setminus \mathbb{R}^+$, if e_j^0 is initialised at a Stieltjes transform.

APPENDIX F USEFUL LEMMAS

In this section, we gather most of the known or new lemmas which are needed in various places in the proof of Appendices A-E.

The statements in the following Lemma are well-known

Lemma 1: 1) For rectangular matrices \mathbf{A} , \mathbf{B} of the same size,

$$\text{rank}(\mathbf{A} + \mathbf{B}) \leq \text{rank}(\mathbf{A}) + \text{rank}(\mathbf{B});$$

2) For rectangular matrices \mathbf{A} , \mathbf{B} for which \mathbf{AB} is defined,

$$\text{rank}(\mathbf{AB}) \leq \min(\text{rank}(\mathbf{A}), \text{rank}(\mathbf{B}));$$

3) For rectangular \mathbf{A} , $\text{rank}(\mathbf{A})$ is less than the number of non-zero entries of \mathbf{A} .

Lemma 2: (Lemma 2.4 of [22]) For $N \times N$ Hermitian matrices \mathbf{A} and \mathbf{B} ,

$$\|F^{\mathbf{A}} - F^{\mathbf{B}}\| \leq \frac{1}{N} \text{rank}(\mathbf{A} - \mathbf{B}).$$

From these two lemmas we get the following.

Lemma 3: Let \mathbf{S} , \mathbf{A} , $\overline{\mathbf{A}}$, be Hermitian $N \times N$, \mathbf{Q} , $\overline{\mathbf{Q}}$ both $N \times n$, and \mathbf{B} , $\overline{\mathbf{B}}$ both Hermitian $n \times n$. Then

1)

$$\|F^{\mathbf{S} + \mathbf{AQBQ}^H \mathbf{A}} - F^{\mathbf{S} + \mathbf{A}\overline{\mathbf{Q}}\overline{\mathbf{B}}\overline{\mathbf{Q}}^H \mathbf{A}}\| \leq \frac{2}{N} \text{rank}(\mathbf{Q} - \overline{\mathbf{Q}});$$

2)

$$\|F^{\mathbf{S} + \mathbf{AQBQ}^H \mathbf{A}} - F^{\mathbf{S} + \overline{\mathbf{A}}\mathbf{QBQ}^H \overline{\mathbf{A}}}\| \leq \frac{2}{N} \text{rank}(\mathbf{A} - \overline{\mathbf{A}});$$

3)

$$\|F^{\mathbf{S} + \mathbf{AQBQ}^H \mathbf{A}} - F^{\mathbf{S} + \mathbf{AQ}\overline{\mathbf{B}}\mathbf{Q}^H \mathbf{A}}\| \leq \frac{1}{N} \text{rank}(\mathbf{B} - \overline{\mathbf{B}}).$$

Lemma 4: For $N \times N$ \mathbf{A} , $\tau \in \mathbb{C}$ and $\mathbf{r} \in \mathbb{C}^N$ for which \mathbf{A} and $\mathbf{A} + \tau \mathbf{r} \mathbf{r}^H$ are invertible,

$$\mathbf{r}^H (\mathbf{A} + \tau \mathbf{r} \mathbf{r}^H)^{-1} = \frac{1}{1 + \tau \mathbf{r}^H \mathbf{A}^{-1} \mathbf{r}} \mathbf{r}^H \mathbf{A}^{-1}.$$

This result follows from $\mathbf{r}^H \mathbf{A}^{-1} (\mathbf{A} + \tau \mathbf{r} \mathbf{r}^H) = (1 + \tau \mathbf{r}^H \mathbf{A}^{-1} \mathbf{r}) \mathbf{r}^H$.

Moreover, we recall Lemma 2.6 of [22]

Lemma 5: Let $z \in \mathbb{C}^+$ with $v = \Im[z]$, \mathbf{A} and \mathbf{B} $N \times N$ with \mathbf{B} Hermitian, and $\mathbf{r} \in \mathbb{C}^N$. Then

$$\left| \text{tr} \left((\mathbf{B} - z\mathbf{I}_N)^{-1} - (\mathbf{B} + \mathbf{r}\mathbf{r}^H - z\mathbf{I}_N)^{-1} \right) \mathbf{A} \right| = \left| \frac{\mathbf{r}^H (\mathbf{B} - z\mathbf{I}_N)^{-1} \mathbf{A} (\mathbf{B} - z\mathbf{I}_N)^{-1} \mathbf{r}}{1 + \mathbf{r}^H (\mathbf{B} - z\mathbf{I}_N)^{-1} \mathbf{r}} \right| \leq \frac{\|\mathbf{A}\|}{v}.$$

If $z < 0$, we also have

$$\left| \text{tr} \left((\mathbf{B} - z\mathbf{I}_N)^{-1} - (\mathbf{B} + \mathbf{r}\mathbf{r}^H - z\mathbf{I}_N)^{-1} \right) \mathbf{A} \right| \leq \frac{\|\mathbf{A}\|}{|z|}.$$

From Lemma 2.2 of [32], and Theorems A.2, A.4, A.5 of [33], we have the following

Lemma 6: If f is analytic on \mathbb{C}^+ , both $f(z)$ and $zf(z)$ map \mathbb{C}^+ into \mathbb{C}^+ , and there exists a $\theta \in (0, \pi/2)$ for which $zf(z) \rightarrow c$, finite, as $z \rightarrow \infty$ restricted to $\{w \in \mathbb{C} : \theta < \arg w < \pi - \theta\}$, then $c < 0$ and f is the Stieltjes transform of a measure on the nonnegative reals with total mass $-c$.

Also, from [22], we need

Lemma 7: Let $\mathbf{y} = (y_1, \dots, y_N)^T$ with the y_i 's i.i.d. such that $E y_1 = 0$, $E |y_1|^2 = 1$ and $y_1 \leq \log N$, and \mathbf{A} an $N \times N$ matrix independent of \mathbf{y} , then

$$E |\mathbf{y}^H \mathbf{A} \mathbf{y} - \text{tr} \mathbf{A}|^6 \leq K \|\mathbf{A}\|^6 N^3 \log^{12} N,$$

where K does not depend on N , \mathbf{A} , nor on the distribution of y_1 .

Additionally, we need

Lemma 8: Let $\mathbf{D} = \mathbf{A} + i\mathbf{B} + iv\mathbf{I}$, where \mathbf{A} , \mathbf{B} are $N \times N$ Hermitian, \mathbf{B} is also positive semi-definite, and $v > 0$. Then $\|\mathbf{D}^{-1}\| \leq v^{-1}$.

Proof: We have $\mathbf{D}\mathbf{D}^H = (\mathbf{A} + i\mathbf{B})(\mathbf{A} - i\mathbf{B}) + v^2\mathbf{I} + 2v\mathbf{B}$. Therefore the eigenvalues of $\mathbf{D}\mathbf{D}^H$ are greater or equal to v^2 , which implies the singular values of \mathbf{D} are greater or equal to v , so that the singular values of \mathbf{D}^{-1} are less or equal to v^{-1} . We therefore get our result. \blacksquare

From Theorem 2.1 of [34],

Lemma 9: Let $\rho(\mathbf{C})$ denote the spectral radius of the $N \times N$ matrix \mathbf{C} (the largest of the absolute values of the eigenvalues of \mathbf{C}). If $\mathbf{x}, \mathbf{b} \in \mathbb{R}^N$ with the components of \mathbf{C} , \mathbf{x} , and \mathbf{b} all positive, then the equation $\mathbf{x} = \mathbf{C}\mathbf{x} + \mathbf{b}$ implies $\rho(\mathbf{C}) < 1$.

From Theorem 8.1.18 of [35],

Lemma 10: Suppose $\mathbf{A} = (a_{ij})$ and $\mathbf{B} = (b_{ij})$ are $N \times N$ with b_{ij} nonnegative and $|a_{ij}| \leq b_{ij}$. Then

$$\rho(\mathbf{A}) \leq \rho(|a_{ij}|) \leq \rho(\mathbf{B}).$$

Also, from Lemma 5.7.9 of [36],

Lemma 11: Let $\mathbf{A} = (a_{ij})$ and $\mathbf{B} = (b_{ij})$ be $N \times N$ with a_{ij}, b_{ij} nonnegative. Then

$$\rho((a_{ij}^{\frac{1}{2}} b_{ij}^{\frac{1}{2}})) \leq (\rho(\mathbf{A}))^{\frac{1}{2}} (\rho(\mathbf{B}))^{\frac{1}{2}}.$$

And, Theorems 8.2.2 and 8.3.1 of [35],

Lemma 12: If \mathbf{C} is a square matrix with nonnegative entries, then $\rho(\mathbf{C})$ is an eigenvalue of \mathbf{C} having an eigenvector \mathbf{x} with nonnegative entries. Moreover, if the entries of \mathbf{C} are all positive, then $\rho(\mathbf{C}) > 0$ and the entries of \mathbf{x} are all positive.

From [36], we also need Theorem 6.1.1,

Lemma 13: Gersgorin's Theorem All the eigenvalues of an $N \times N$ matrix $\mathbf{A} = (a_{ij})$ lie in the union of the N disks in the complex plane, the i^{th} disk having center a_{ii} and radius $\sum_{j \neq i} |a_{ij}|$.

Theorem 3.42 of [30],

Lemma 14: Rouché's Theorem If $f(z)$ and $g(z)$ are analytic inside and on a closed contour C of the complex plane, and $|g(z)| < |f(z)|$ on C , then $f(z)$ and $f(z) + g(z)$ have the same number of zeros inside C .

In order to prove Theorem 2, we also need, from [37]

Lemma 15: Consider a rectangular matrix \mathbf{A} and let $s_i^{\mathbf{A}}$ denote the i^{th} largest singular value of \mathbf{A} , with $s_i^{\mathbf{A}} = 0$ whenever $i > \text{rank}(\mathbf{A})$. Let m, n be arbitrary non-negative integers. Then for \mathbf{A}, \mathbf{B} rectangular of the same size

$$s_{m+n+1}^{\mathbf{A}+\mathbf{B}} \leq s_{m+1}^{\mathbf{A}} + s_{n+1}^{\mathbf{B}},$$

and for \mathbf{A}, \mathbf{B} rectangular for which \mathbf{AB} is defined

$$s_{m+n+1}^{\mathbf{AB}} \leq s_{m+1}^{\mathbf{A}} s_{n+1}^{\mathbf{B}}.$$

As a corollary, for any integer $r \geq 0$ and rectangular matrices $\mathbf{A}_1, \dots, \mathbf{A}_K$, all of the same size,

$$s_{Kr+1}^{\mathbf{A}_1+\dots+\mathbf{A}_K} \leq s_{r+1}^{\mathbf{A}_1} + \dots + s_{r+1}^{\mathbf{A}_K}.$$

ACKNOWLEDGMENT

Silverstein's work is supported by the U.S. Army Research Office under Grant W911NF-05-1-0244. Debbah's work is partially supported by the European Commission in the framework of the FP7 Network of Excellence in Wireless Communications NEWCOM++. The authors would like to thank Walid Hachem for the fruitful discussions we shared concerning some results of this paper.

REFERENCES

- [1] G. J. Foschini and M. J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Personal Communications*, vol. 6, no. 3, pp. 311–335, Mar. 1998.
- [2] I. E. Telatar, "Capacity of multi-antenna gaussian channels," *European Transactions on Telecommunications*, vol. 10, no. 6, pp. 585–595, Feb. 1999.
- [3] W. Yu, W. Rhee, S. Boyd, and J. M. Cioffi, "Iterative water-filling for gaussian vector multiple-access channels," *IEEE Trans. Inf. Theory*, vol. 50, no. 1, pp. 145–152, 2004.
- [4] M. Vu and A. Paulraj, "Capacity optimization for rician correlated mimo wireless channels," in *Proc. IEEE Conference Record of the Asilomar Conference on Signals, Systems, and Computers (Asilomar'05)*, Pacific Grove, CA, USA, 2005, pp. 133–138.
- [5] W. Hachem, P. Loubaton, and J. Najim, "Deterministic equivalents for certain functionals of large random matrices," *Annals of Applied Probability*, vol. 17, no. 3, pp. 875–930, 2007.
- [6] A. M. Tulino and S. Verdú, "Impact of antenna correlation on the capacity of multiantenna channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 7, pp. 2491–2509, 2005.
- [7] M. J. M. Peacock, I. B. Collings, and M. L. Honig, "Eigenvalue distributions of sums and products of large random matrices via incremental matrix expansions," *IEEE Trans. Inf. Theory*, vol. 54, no. 5, pp. 2123–2138, 2008.
- [8] A. Soysal and S. Ulukus, "Optimality of beamforming in fading mimo multiple access channels," *IEEE Trans. Commun.*, vol. 57, no. 4, pp. 1171–1183, Apr. 2009.

- [9] J. Dumont, W. Hachem, P. Loubaton, and J. Najim, "On the capacity achieving covariance matrix for rician mimo channels: An asymptotic approach," *IEEE Trans. Inf. Theory*, vol. 56, no. 3, pp. 1048–1069, 2010.
- [10] W. Hachem, P. Loubaton, and J. Najim, "A clt for information theoretic statistics of gram random matrices with a given variance profile," *Annals of Probability*, vol. 18, no. 6, pp. 2071–2130, Dec. 2008.
- [11] A. L. Moustakas and S. H. Simon, "On the outage capacity of correlated multiple-path mimo channels," *IEEE Trans. Inf. Theory*, vol. 53, no. 11, pp. 3887–3903, 2007.
- [12] M. Mézard, G. Parisi, and M. Virasoro, "Spin glass theory and beyond," *Physics Today*, vol. 41, no. 12, 1988.
- [13] F. Dupuy and P. Loubaton, "On the capacity achieving covariance matrix for frequency selective mimo channels using the asymptotic approach," *IEEE Trans. Inf. Theory*, submitted for publication. [Online]. Available: <http://arxiv.org/abs/1001.3102>
- [14] R. R. Far, T. Oraby, W. Bryc, and R. Speicher, "On slow-fading mimo systems with nonseparable correlation," *IEEE Trans. Inf. Theory*, vol. 54, no. 2, pp. 544–553, 2008.
- [15] C. N. Chuah, D. N. C. Tse, J. M. Kahn, and R. A. Valenzuela, "Capacity scaling in mimo wireless systems under correlated fading," *IEEE Trans. Inf. Theory*, vol. 48, no. 3, pp. 637–650, Mar. 2002.
- [16] Z. Bai and J. W. Silverstein, "Spectral analysis of large dimensional random matrices," *Springer Series in Statistics*, 2009.
- [17] V. A. Marčenko and L. A. Pastur, "Distributions of eigenvalues for some sets of random matrices," *Math USSR-Sbornik*, vol. 1, no. 4, pp. 457–483, Apr. 1967.
- [18] D. N. C. Tse and S. V. Hanly, "Linear multiuser receivers: Effective interference, effective bandwidth and user capacity," *IEEE Trans. Inf. Theory*, vol. 45, no. 2, pp. 641–657, Feb. 1999.
- [19] M. Debbah, W. Hachem, P. Loubaton, and M. de Courville, "Mmse analysis of certain large isometric random precoded systems," *IEEE Trans. Inf. Theory*, vol. 49, no. 5, pp. 1293–1311, 2003.
- [20] P. Billingsley, *Probability and measure*, 3rd ed. Hoboken, NJ: John Wiley & Sons, Inc., 1995.
- [21] L. Zhang, "Spectral analysis of large dimensional random matrices," Ph.D. dissertation, National University of Singapore, 2006.
- [22] 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*, vol. 54, no. 2, pp. 175–192, 1995.
- [23] A. L. Moustakas, S. H. Simon, and A. M. Sengupta, "Mimo capacity through correlated channels in the presence of correlated interferers and noise: A (not so) large n analysis," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2545–2561, 2003.
- [24] T. Pollock, T. Abhayapala, and R. Kennedy, "Antenna saturation effects on dense array mimo capacity," in *Proc. IEEE International Conference on Communications (ICC'03)*, Anchorage, Alaska, 2003, pp. 2301–2305.
- [25] A. L. Moustakas, H. U. Baranger, L. Balents, A. M. Sengupta, and S. Simon, "Communication through a diffusive medium: Coherence and capacity," *Science*, vol. 287, pp. 287–290, 2000.
- [26] A. Goldsmith, S. A. Jafar, N. Jindal, and S. Vishwanath, "Capacity limits of mimo channels," *IEEE J. Sel. Areas Commun.*, vol. 21, no. 5, pp. 684–702, 2003.
- [27] R. Clarke, "A statistical theory of mobile-radio reception," *Bell System Technical Journal*, vol. 47, no. 6, pp. 957–1000, 1968.
- [28] S. Viswanath, N. Jindal, and A. Goldsmith, "Duality, achievable rates, and sum-rate capacity of gaussian mimo broadcast channels," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2658–2668, 2003.
- [29] R. B. Dozier and J. W. Silverstein, "On the empirical distribution of eigenvalues of large dimensional information plus noise-type matrices," *Journal of Multivariate Analysis*, vol. 98, no. 4, pp. 678–694, 2007.
- [30] E. C. Titchmarsh, *The Theory of Functions*. New York, NY, USA: Oxford University Press, 1939.
- [31] P. Billingsley, *Convergence of probability measures*. Hoboken, NJ: John Wiley & Sons, Inc., 1968.
- [32] J. A. Shohat and J. D. Tamarkin, *The Problem of Moments*. Providence, RI, USA: American Mathematical Society, 1970.
- [33] M. K. Krein and A. A. Nudelman, *The Markov Moment Problem and Extremal Problems*. Providence, RI, USA: American Mathematical Society, 1997.
- [34] E. Seneta, *Non-negative Matrices and Markov Chains, Second Edition*. Springer Verlag New York, 1981.
- [35] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge University Press, 1985.
- [36] —, *Topics in Matrix Analysis*. Cambridge University Press, 1991.
- [37] K. Fan, "Maximum properties and inequalities for the eigenvalues of completely continuous operators," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 37, no. 11, pp. 760–766, 1951.