6 Deterministic equivalents

6.1 Introduction to deterministic equivalents

The first applications of random matrix theory to the field of wireless communications, e.g., [Tse and Hanly, 1999; Tse and Verdú, 2000; Verdú and Shamai, 1999], originally dealt with the limiting behavior of some simple random matrix models. In particular, these results are attractive as these limiting behaviors only depend on the limiting eigenvalue distribution of the deterministic matrices of the model. This is in fact the case of all the results we have derived and introduced so far; for instance, Theorem 3.13 unveils the limiting behavior of the e.s.d. of $\mathbf{B}_N = \mathbf{A}_N + \mathbf{X}_N^{\mathsf{H}}\mathbf{T}_N\mathbf{X}_N$ when both e.s.d. of $\mathbf{A}_N$ and $\mathbf{T}_N$ converge toward given deterministic distribution functions and $\mathbf{X}_N$ is random with i.i.d. entries. However, for practical applications, it might turn out that:

(i) the e.s.d. of $\mathbf{A}_N$ or $\mathbf{T}_N$ do not necessarily converge to a limiting distribution;
(ii) even if the e.s.d. of the deterministic matrices in the model do all converge to their respective l.s.d., the e.s.d. of the output matrix $\mathbf{B}_N$ might not converge. This is of course not the case in Theorem 3.13, but we will show that this may happen for more involved models, e.g. the models treated by [Couillet *et al.*, 2011a] and [Hachem *et al.*, 2007].

Let us introduce a simple scenario for which the e.s.d. of the random matrix does not converge. This example is borrowed from [Hachem *et al.*, 2007]. Define $\mathbf{X}_N \in \mathbb{C}^{2N\times 2N}$ as

$$\mathbf{X}_N = \begin{pmatrix} \mathbf{X}'_N & 0 \\ 0 & 0 \end{pmatrix} \tag{6.1}$$

with the entries of $\mathbf{X}'_N$ being i.i.d. with zero mean and variance $\frac{1}{N}$. Consider in addition the matrix $\mathbf{T}_N \in \mathbb{C}^{2N\times 2N}$ defined as

$$\mathbf{T}_N = \begin{cases} \begin{pmatrix} \mathbf{I}_N & 0 \\ 0 & 0 \end{pmatrix}, & N \text{ even} \\ \begin{pmatrix} 0 & 0 \\ 0 & \mathbf{I}_N \end{pmatrix}, & N \text{ odd.} \end{cases} \tag{6.2}$$

Then, taking $\mathbf{B}_N = (\mathbf{T}_N + \mathbf{X}_N)(\mathbf{T}_N + \mathbf{X}_N)^{\mathsf{H}}$, $F^{\mathbf{B}_{2N}}$ and $F^{\mathbf{B}_{2N+1}}$ both converge weakly towards limit distributions, as $N \to \infty$, but those distributions

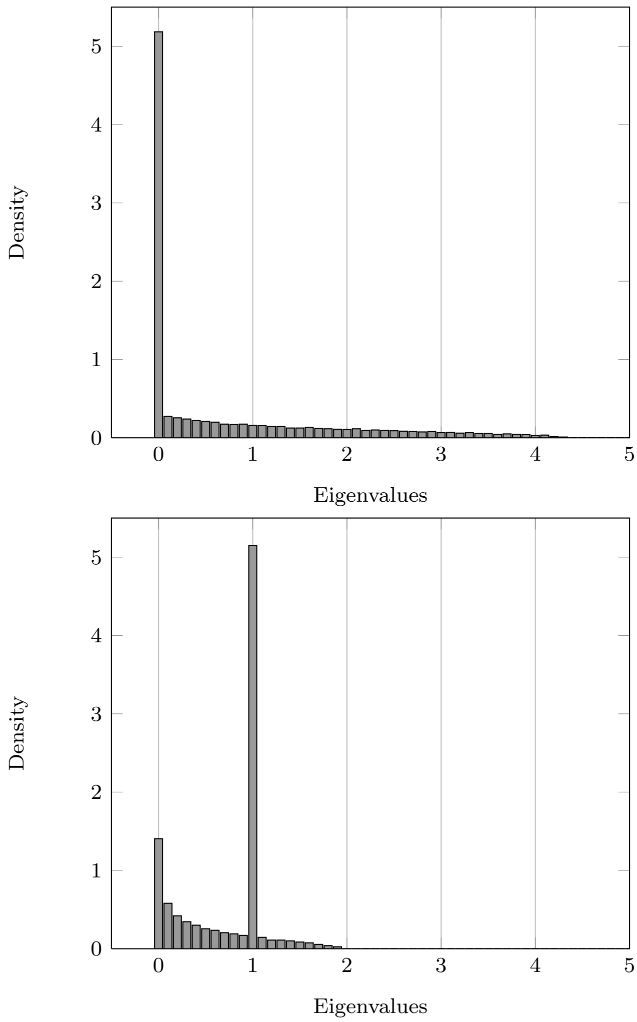

Figure 6.1 Histogram of the eigenvalues of $\mathbf{B}_N = (\mathbf{T}_N + \mathbf{X}_N)(\mathbf{T}_N + \mathbf{X}_N)^{\mathsf{H}}$ modeled in (6.1)–(6.2), for $N = 1000$ (top) and $N = 1001$ (bottom).

differ. Indeed, for N even, half of the spectrum of $\mathbf{B}_N$ is formed of zeros, while for N odd, half of the spectrum of $\mathbf{B}_N$ is formed of ones, the rest of the spectrum being a weighted version of the Marčenko–Pastur law. And therefore there does not exist a limit to $F^{\mathbf{B}_N}$, while $F^{\mathbf{X}_N\mathbf{X}_N^{\mathsf{H}}}$ tends to the uniformly weighted sum of the Marčenko–Pastur law and a mass in zero, and $F^{\mathbf{T}_N\mathbf{T}_N^{\mathsf{H}}}$ tends to the uniformly weighted sum of two masses in zero and one. This is depicted in Figure 6.1.

In such situations, there is therefore no longer any interest in looking at the asymptotic behavior of e.s.d. Instead, we will be interested in finding *deterministic equivalents* for the underlying model.

Definition 6.1. Consider a series of Hermitian random matrices $\mathbf{B}_1, \mathbf{B}_2, \ldots$, with $\mathbf{B}_N \in \mathbb{C}^{N\times N}$ and a series $f_1, f_2, \ldots$ of functionals of $1\times 1, 2\times 2, \ldots$ matrices. A *deterministic equivalent of* $\mathbf{B}_N$ *for the functional* f_N is a series $\mathbf{B}_1^\circ, \mathbf{B}_2^\circ, \ldots$ where $\mathbf{B}_N^\circ \in \mathbb{C}^{N\times N}$, of *deterministic* matrices, such that

$$\lim_{N\to\infty} f_N(\mathbf{B}_N) - f_N(\mathbf{B}_N^\circ) \to 0$$

where the convergence will often be with probability one. Note that $f_N(\mathbf{B}_N^\circ)$ does not need to have a limit as $N \to \infty$. We will similarly call $g_N \triangleq f_N(\mathbf{B}_N^\circ)$ the *deterministic equivalent of* $f_N(\mathbf{B}_N)$, i.e. the deterministic series $g_1, g_2, \ldots$ such that $f_N(\mathbf{B}_N) - g_N \to 0$ in some sense.

We will often take f_N to be the normalized trace of $(\mathbf{B}_N - z\mathbf{I}_N)^{-1}$, i.e. the Stieltjes transform of $F^{\mathbf{B}_N}$. When $f_N(\mathbf{B}_N^\circ)$ does not have a limit, the Marčenko–Pastur method, developed in Section 3.2, will fail. This is because, at some point, all the entries of the underlying matrices will have to be taken into account and not only the diagonal entries, as in the proof we provided in Section 3.2. However, the Marčenko–Pastur method can be tweaked adequately into a technique that can cope with deterministic equivalents. In the following, we first introduce this technique, which we will call the *Bai and Silverstein technique*, and then discuss an alternative technique, known as the *Gaussian method*, which is particularly suited to random matrix models with Gaussian entries. Hereafter, we detail these methods by successively proving two (similar) results of importance in wireless communications, see further Chapters 13–14.

6.2 Techniques for deterministic equivalents

6.2.1 Bai and Silverstein method

We first introduce a deterministic equivalent for the model

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

where the K matrices $\mathbf{X}_k$ have i.i.d. entries for each k, mutually independent for different k, and the matrices $\mathbf{T}_1, \ldots, \mathbf{T}_K$, $\mathbf{R}_1, \ldots, \mathbf{R}_K$ and $\mathbf{A}$ are 'bounded' in some sense to be defined later. This is more general than the model of Theorem 3.13 in several respects:

(i) left product matrices $\mathbf{R}_k$, $1 \leq k \leq K$, have been introduced. As an exercise, it can already be verified that a l.s.d. for the model $\mathbf{R}_1^{\frac{1}{2}} \mathbf{X}_1 \mathbf{T}_1 \mathbf{X}_1^{\mathsf{H}} \mathbf{R}_1^{\frac{1}{2}} + \mathbf{A}$ may not exist even if $F^{\mathbf{R}_1}$ and $F^{\mathbf{A}}$ both converge vaguely to deterministic limits, unless some severe additional constraint is put on the eigenvectors of $\mathbf{R}_1$ and $\mathbf{A}$, e.g. $\mathbf{R}_1$ and $\mathbf{A}$ are codiagonalizable. This suggests that the Marčenko–Pastur method will fail to treat this model;

(ii) a sum of K such models is considered (K does not grow along with N here);
(iii) the e.s.d. of the (possibly random) matrices $\mathbf{T}_k$ and $\mathbf{R}_k$ are not required to converge.

While the result to be introduced hereafter is very likely to hold for $\mathbf{X}_1, \ldots, \mathbf{X}_K$ with non-identically distributed entries (as long as they have common mean and variance and some higher order moment condition), we only present here the result where these entries are identically distributed, which is less general than the conditions of Theorem 3.13.

Theorem 6.1 ([Couillet *et al.*, 2011a]). *Let K be some positive integer. For some integer N, let*

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

be an $N \times N$ matrix with the following hypotheses, for all $k \in \{1, \ldots, K\}$

1. $\mathbf{X}_k = \left(\frac{1}{\sqrt{n_k}} X_{k,ij}\right) \in \mathbb{C}^{N \times n_k}$ *is such that the $X_{k,ij}$ are identically distributed for all N, i, j, independent for each fixed N, and* $\mathrm{E}|X_{k,11} - \mathrm{E}X_{k,11}|^2 = 1$;
2. $\mathbf{R}_k^{\frac{1}{2}} \in \mathbb{C}^{N \times N}$ *is a Hermitian non-negative definite square root of the non-negative definite Hermitian matrix* $\mathbf{R}_k$;
3. $\mathbf{T}_k = \mathrm{diag}(\tau_{k,1}, \ldots, \tau_{k,n_k}) \in \mathbb{C}^{n_k \times n_k}$, $n_k \in \mathbb{N}^*$, *is diagonal with* $\tau_{k,i} \geq 0$;
4. *the sequences $F^{\mathbf{T}_1}, F^{\mathbf{T}_2}, \ldots$ and $F^{\mathbf{R}_1}, F^{\mathbf{R}_2}, \ldots$ are tight, i.e. for all $\varepsilon > 0$, there exists $M > 0$ such that $1 - F^{\mathbf{T}_k}(M) < \varepsilon$ and $1 - F^{\mathbf{R}_k}(M) < \varepsilon$ for all n_k, N;*
5. $\mathbf{A} \in \mathbb{C}^{N \times N}$ *is Hermitian non-negative definite;*
6. *denoting $c_k = N/n_k$, for all k, there exist $0 < a < b < \infty$ for which*

$$a \leq \liminf_N c_k \leq \limsup_N c_k \leq b. \tag{6.3}$$

Then, as all N and n_k grow large, with ratio c_k, for $z \in \mathbb{C} \setminus \mathbb{R}^+$, the Stieltjes transform $m_{\mathbf{B}_N}(z)$ of $\mathbf{B}_N$ satisfies

$$m_{\mathbf{B}_N}(z) - m_N(z) \xrightarrow{\text{a.s.}} 0 \tag{6.4}$$

where

$$m_N(z) = \frac{1}{N} \operatorname{tr} \left(\mathbf{A} + \sum_{k=1}^{K} \int \frac{\tau_k dF^{\mathbf{T}_k}(\tau_k)}{1 + c_k \tau_k e_{N,k}(z)} \mathbf{R}_k - z \mathbf{I}_N \right)^{-1} \tag{6.5}$$

and the set of functions $e_{N,1}(z), \ldots, e_{N,K}(z)$ forms the unique solution to the K equations

$$e_{N,i}(z) = \frac{1}{N} \operatorname{tr} \mathbf{R}_i \left(\mathbf{A} + \sum_{k=1}^{K} \int \frac{\tau_k dF^{\mathbf{T}_k}(\tau_k)}{1 + c_k \tau_k e_{N,k}(z)} \mathbf{R}_k - z \mathbf{I}_N \right)^{-1} \tag{6.6}$$

such that $\mathrm{sgn}(\Im[e_{N,i}(z)]) = \mathrm{sgn}(\Im[z])$, *if $z \in \mathbb{C} \setminus \mathbb{R}$, and $e_{N,i}(z) > 0$ if z is real negative.*

Moreover, for any $\varepsilon > 0$*, the convergence of Equation* (6.4) *is uniform over any region of* $\mathbb{C}$ *bounded by a contour interior to*

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

For all N*, the function* m_N *is the Stieltjes transform of a distribution function* F_N*, and*

$$F^{\mathbf{B}_N} - F_N \Rightarrow 0$$

almost surely as $N \to \infty$*.*

In [Couillet *et al.*, 2011a], Theorem 6.1 is completed by the following result.

Theorem 6.2. *Under the conditions of Theorem 6.1, the scalars* $e_{N,1}(z), \ldots, e_{N,K}(z)$ *are also explicitly given by:*

$$e_{N,i}(z) = \lim_{t \to \infty} e_{N,i}^t(z)$$

where, for all i*,* $e_{N,i}^0(z) = -1/z$ *and, for* $t \geq 1$

$$e_{N,i}^t(z) = \frac{1}{N} \operatorname{tr} \mathbf{R}_i \left(\mathbf{A} + \sum_{j=1}^{K} \int \frac{\tau_j dF^{\mathbf{T}_j}(\tau_j)}{1 + c_j \tau_j e_{N,j}^{t-1}(z)} \mathbf{R}_j - z \mathbf{I}_N \right)^{-1} .$$

This result, which ensures the convergence of the classical fixed-point algorithm for an adequate initial condition, is of fundamental importance for practical purposes as it ensures that the $e_{N,1}(z), \ldots, e_{N,K}(z)$ can be determined numerically in a deterministic way. Since the proof of Theorem 6.2 relies heavily on the proof of Theorem 6.1, we will prove Theorem 6.2 later.

Several remarks are in order before we prove Theorem 6.1. We have given much detail on the conditions for Theorem 6.1 to hold. We hereafter discuss the implications of these conditions. Condition 1 requires that the $X_{k,ij}$ be identically distributed across N, i, j, but not necessarily across k. Note that the identical distribution condition could be further released under additional mild conditions (such as all entries must have a moment of order $2 + \varepsilon$, for some $\varepsilon > 0$), see Theorem 3.13. Condition 4 introduces tightness requirements on the e.s.d. of $\mathbf{R}_k$ and $\mathbf{T}_k$. Tightness can be seen as the probabilistic equivalent to boundedness for deterministic variables. Tightness ensures here that no mass of the $F^{\mathbf{R}_k}$ and $F^{\mathbf{T}_k}$ escapes to infinity as n grows large. Condition 6 is more general than the requirement that c_k has a limit as it allows c_k, for all k, to wander between two positive values.

From a practical point of view, $\mathbf{R}_K^{\frac{1}{2}} \mathbf{X}_k \mathbf{T}_k^{\frac{1}{2}}$ will often be used to model a multiple antenna $N \times n_k$ channel with i.i.d. entries with transmit and receive correlations. From the assumptions of Theorem 6.1, the correlation matrices $\mathbf{R}_k$ and $\mathbf{T}_k$ are only required to be 'bounded' in the sense of tightness of their e.s.d. This means that, as the number of antennas grows, the eigenvalues of $\mathbf{R}_k$ and $\mathbf{T}_k$

can only blow up with increasingly low probability. If we increase the number N of antennas on a bounded three-dimensional space, then the rough tendency is for the eigenvalues of $\mathbf{T}_k$ and $\mathbf{R}_k$ to be all small except for a few of them, which grow large but have a probability of order $O(1/N)$, see, e.g., [Pollock *et al.*, 2003]. In that context, Theorem 6.1 holds, i.e. for $N \to \infty$, $F^{\mathbf{B}_N} - F_N \Rightarrow 0$.

It is also important to remark that the matrices $\mathbf{T}_k$ are constrained to be diagonal. This is unimportant when the matrices $\mathbf{X}_k$ are assumed Gaussian in practical applications, as the $\mathbf{X}_k$, being bi-unitarily invariant, can be multiplied on the right by any deterministic unitary matrix without altering the final result. This limitation is linked to the technique used for proving Theorem 6.1. For mathematical completion, though, it would be convenient for the matrices $\mathbf{T}_k$ to be unconstrained. We mention that Zhang and Bai [Zhang, 2006] derive the limiting spectral distribution of the model $\mathbf{B}_N = \mathbf{R}_1^{\frac{1}{2}}\mathbf{X}_1\mathbf{T}_1\mathbf{X}_1^{\mathsf{H}}\mathbf{R}_1^{\frac{1}{2}}$ for unconstrained Hermitian $\mathbf{T}_1$, using a different approach than that presented below.

For practical applications, it will be easier in the following to write (6.6) in a more symmetric way. This is discussed in the following remark.

Remark 6.1. In the particular case where $\mathbf{A} = 0$, the K implicit Equations (6.6) can be developed into the $2K$ linked equations

$$e_{N,i}(z) = \frac{1}{N}\operatorname{tr}\mathbf{R}_i\left(-z\left[\mathbf{I}_N + \sum_{k=1}^{K}\bar{e}_k(z)\mathbf{R}_k\right]\right)^{-1}$$
$$\bar{e}_{N,i}(z) = \frac{1}{n_i}\operatorname{tr}\mathbf{T}_i\left(-z\left[\mathbf{I}_{n_i} + c_i e_{N,i}(z)\mathbf{T}_i\right]\right)^{-1} \tag{6.7}$$

whose symmetric aspect is both more readable and more useful for practical reasons that will be evidenced later in Chapters 13–14. As a consequence, $m_N(z)$ in (6.5) becomes

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

In the literature and, as a matter of fact, in some deterministic equivalents presented later in this chapter, the variables $e_{N,i}(z)$ may be normalized by $\frac{1}{n_i}$ instead of $\frac{1}{N}$ in order to avoid carrying the factor c_i in front of $e_{N,i}(z)$ in the second fixed-point equation of (6.7). In the application chapters, Chapters 12–15, depending on the situation, either one or the other convention will be taken.

We present hereafter the general techniques, based on the Stieltjes transform, to prove Theorem 6.1 and other similar results introduced in this section. As opposed to the proof of the Marčenko–Pastur law, we cannot prove that that there exists a space of probability one over which $m_{\mathbf{B}_N}(z) \to m(z)$ for all $z \in \mathbb{C} \setminus \mathbb{R}^+$, for a certain limiting function m. Instead, we prove that there exists a space of probability one over which $m_{\mathbf{B}_N}(z) - m_N(z) \to 0$ for all z, for a certain series of Stieltjes transforms $m_1(z), m_2(z), \ldots$. There are in general two main approaches to prove this convergence. The first option is a point-wise approach that consists in proving the convergence for all z in a compact subspace of $\mathbb{C} \setminus \mathbb{R}^+$ having a limit point. Invoking Vitali's convergence theorem, similar to the proof of the Marčenko–Pastur law, we then prove the convergence for all $z \in \mathbb{C} \setminus \mathbb{R}^+$. In the coming proof, we will take $z \in \mathbb{C}^+$. In the proof of Theorem 6.17, we will take z real negative. The second option is a functional approach in which the objects under study are not $m_{\mathbf{B}_N}(z)$ and $m_N(z)$ taken at a precise point $z \in \mathbb{C} \setminus \mathbb{R}^+$ but rather $m_{\mathbf{B}_N}(z)$ and $m_N(z)$ seen as functions lying in the space of Stieltjes transforms of distribution functions with support on $\mathbb{R}^+$. The convergence $m_{\mathbf{B}_N}(z) - m_N(z) \xrightarrow{\text{a.s.}} 0$ is in this case functional and Vitali's convergence theorem is not called for. This is the approach followed in, e.g., [Hachem *et al.*, 2007]. The latter is not detailed in this book.

The first step of the general proof, for either option, consists in determining $m_N(z)$. For this, similar to the Marčenko–Pastur proof, we develop the expression of $m_{\mathbf{B}_N}(z)$, seeking for a limiting result of the kind

$$m_{\mathbf{B}_N}(z) - h_N(m_{\mathbf{B}_N}(z); z) \xrightarrow{\text{a.s.}} 0$$

for some deterministic function h_N, possibly depending on N. Such an expression allows us to infer the nature of a deterministic approximation $m_N(z)$ of $m_{\mathbf{B}_N}(z)$ as a particular solution of the equation in m

$$m - h_N(m; z) = 0. \tag{6.8}$$

This equation rarely has a unique point-wise solution, i.e. for every z, but often has a unique functional solution $z \to m_N(z)$ that is the Stieltjes transform of a distribution function. If the point-wise approach is followed, a unique point-wise solution of (6.8) can often be narrowed down to a certain subspace of $\mathbb{C}$ for z lying in some other subspace of $\mathbb{C}$. In Theorem 6.1, there exists a single solution in $\mathbb{C}^+$ when $z \in \mathbb{C}^+$, a single solution in $\mathbb{C}^-$ when $z \in \mathbb{C}^-$, and a single positive solution when z is real negative. Standard holomorphicity arguments on the function $m_N(z)$ then ensure that $z \to m_N(z)$ is the unique Stieltjes transform satisfying $h_N(m_N(z); z) = m_N(z)$. When using the functional approach, this fact tends to be proved more directly. In the coming proof of Theorem 6.1, we will prove point-wise uniqueness by assuming, as per standard techniques, the alleged existence of two distinct solutions and prove a contradiction. An alternative approach is to prove that the fixed-point algorithm

$$\begin{aligned} &m_0 \in \mathcal{D} \\ &m_{t+1} = h_N(m_t; z),\ t \geq 0 \end{aligned}$$

always converges to $m_N(z)$, where $\mathcal{D}$ is taken to be either $\mathbb{R}^-$, $\mathbb{C}^+$ or $\mathbb{C}^-$. This approach, when valid (in some involved cases, convergence may not always arise), is doubly interesting as it allows both (i) to prove point-wise uniqueness for z taken in some subset of $\mathbb{C} \setminus \mathbb{R}^+$, leading to uniqueness of the Stieltjes transform using again holomorphicity arguments, and (ii) to provide an explicit algorithm to compute $m_N(z)$ for $z \in \mathcal{D}$, which is in particular of interest for practical applications when $z = -\sigma^2 < 0$. In the proof of Theorem 6.1, we will introduce both results for completion. In the proof of Theorem 6.17, we will directly proceed to proving the convergence of the fixed-point algorithm for z real negative.

When the uniqueness of the Stieltjes transform $m_N(z)$ has been made clear, the last step is to prove that, in the large N limit

$$m_{\mathbf{B}_N}(z) - m_N(z) \xrightarrow{\text{a.s.}} 0.$$

This step is not so immediate. To this point, we indeed only know that $m_{\mathbf{B}_N}(z) - h_N(m_{\mathbf{B}_N}(z); z) \xrightarrow{\text{a.s.}} 0$ and $m_N(z) - h_N(m_N(z); z) = 0$. This does not imply immediately that $m_{\mathbf{B}_N}(z) - m_N(z) \xrightarrow{\text{a.s.}} 0$. If there are several point-wise solutions to $m - h_N(m; z) = 0$, we need to verify that $m_N(z)$ was chosen to be the one that will eventually satisfy $m_{\mathbf{B}_N}(z) - m_N(z) \xrightarrow{\text{a.s.}} 0$. This will conclude the proof.

We now provide the specific proof of Theorem 6.1. In order to determine the above function h_N, we first develop the Marčenko–Pastur method (for simplicity for $K = 2$ and $\mathbf{A} = 0$). We will realize that this method fails unless all $\mathbf{R}_k$ and $\mathbf{A}$ are constrained to be co-diagonalizable. To cope with this limitation, we will introduce the more powerful Bai and Silverstein method, whose idea is to *guess* along the derivations the suitable form of h_N. In fact, as we will shortly realize, the problem is slightly more difficult here as we will not be able to find such a function h_N (which may actually not exist at all in the first place). We will however be able to find functions $f_{N,i}$ such that, for each i

$$e_{\mathbf{B}_N,i}(z) - f_{N,i}(e_{\mathbf{B}_N,1}(z), \ldots, e_{\mathbf{B}_N,K}(z); z) \xrightarrow{\text{a.s.}} 0$$

where $e_{\mathbf{B}_N,i}(z) \triangleq \frac{1}{N} \operatorname{tr} \mathbf{R}_i(\mathbf{B}_N - z\mathbf{I}_N)^{-1}$. We will then look for a function $e_{N,i}(z)$ that satisfies

$$e_{N,i}(z) = f_{N,i}(e_{N,1}(z), \ldots, e_{N,K}(z); z).$$

From there, it will be easy to determine a further function g_N such that

$$m_{\mathbf{B}_N}(z) - g_N(e_{\mathbf{B}_N,1}(z), \ldots, e_{\mathbf{B}_N,K}(z); z) \xrightarrow{\text{a.s.}} 0$$

and

$$m_N(z) - g_N(e_{N,1}(z), \ldots, e_{N,K}(z); z) = 0.$$

We will therefore have finally

$$m_{\mathbf{B}_N}(z) - m_N(z) \xrightarrow{\text{a.s.}} 0.$$

Proof of Theorem 6.1. In order to have a first insight on what the deterministic equivalent m_N of $m_{\mathbf{B}_N}$ may look like, the Marčenko–Pastur method will be applied with the (strong) additional assumption that $\mathbf{A}$ and all $\mathbf{R}_k$, $1 \leq k \leq K$, are diagonal and that the e.s.d. $F^{\mathbf{T}_k}$, $F^{\mathbf{R}_k}$ converge for all k as N grows large. In this scenario, $m_{\mathbf{B}_N}$ has a limit when $N \to \infty$ and the method, however more tedious than in the proof of the Marčenko–Pastur law, leads naturally to m_N.

Consider the case when $K = 2$, $\mathbf{A} = 0$ for simplicity and denote $\mathbf{H}_k = \mathbf{R}_k^{\frac{1}{2}}\mathbf{X}_k\mathbf{T}_k^{\frac{1}{2}}$. Following similar steps as in the proof of the Marčenko–Pastur law, we start with matrix inversion lemmas

$$\begin{aligned}&\left(\mathbf{H}_1\mathbf{H}_1^{\mathsf{H}} + \mathbf{H}_2\mathbf{H}_2^{\mathsf{H}} - z\mathbf{I}_N\right)_{11}^{-1}\\ &= \left[-z - z[\mathbf{h}_1^{\mathsf{H}}\mathbf{h}_2^{\mathsf{H}}]\left(\begin{bmatrix}\mathbf{U}_1^{\mathsf{H}}\\ \mathbf{U}_2^{\mathsf{H}}\end{bmatrix}[\mathbf{U}_1\mathbf{U}_2] - z\mathbf{I}_{n_1+n_2}\right)^{-1}\begin{bmatrix}\mathbf{h}_1\\ \mathbf{h}_2\end{bmatrix}\right]^{-1}\end{aligned}$$

with the definition $\mathbf{H}_i^{\mathsf{H}} = [\mathbf{h}_i\mathbf{U}_i^{\mathsf{H}}]$. Using the block matrix inversion lemma, the inner inversed matrix in this expression can be decomposed into four submatrices. The upper-left $n_1 \times n_1$ submatrix reads:

$$\left(-z\mathbf{U}_1^{\mathsf{H}}(\mathbf{U}_2\mathbf{U}_2^{\mathsf{H}} - z\mathbf{I}_{N-1})^{-1}\mathbf{U}_1 - z\mathbf{I}_{n_1}\right)^{-1}$$

while, for the second block diagonal entry, it suffices to revert all ones in twos and vice-versa. Taking the limits, using Theorem 3.4 and Theorem 3.9, we observe that the two off-diagonal submatrices will not play a role, and we finally have

$$\begin{aligned}&\left(\mathbf{H}_1\mathbf{H}_1^{\mathsf{H}} + \mathbf{H}_2\mathbf{H}_2^{\mathsf{H}} - z\mathbf{I}_N\right)_{11}^{-1}\\ &\simeq \left[-z - zr_{11}\frac{1}{n_1}\operatorname{tr}\mathbf{T}_1\left(-z\mathbf{H}_1^{\mathsf{H}}(\mathbf{H}_2\mathbf{H}_2^{\mathsf{H}} - z\mathbf{I}_N)^{-1}\mathbf{H}_1 - z\mathbf{I}_{n_1}\right)^{-1}\right.\\ &\qquad \left. - zr_{21}\frac{1}{n_2}\operatorname{tr}\mathbf{T}_2\left(-z\mathbf{H}_2^{\mathsf{H}}(\mathbf{H}_1\mathbf{H}_1^{\mathsf{H}} - z\mathbf{I}_N)^{-1}\mathbf{H}_1 - z\mathbf{I}_{n_2}\right)^{-1}\right]^{-1}\end{aligned}$$

where the symbol "$\simeq$" denotes some kind of yet unknown large N convergence and where we denoted r_{ij} the jth diagonal entry of $\mathbf{R}_i$. Observe that we can proceed to a similar derivation for the matrix $\mathbf{T}_1\left(-z\mathbf{H}_1^{\mathsf{H}}(\mathbf{H}_2\mathbf{H}_2^{\mathsf{H}} - z\mathbf{I}_N)^{-1}\mathbf{H}_1 - z\mathbf{I}_{n_1}\right)^{-1}$ that now appears. Denoting now $\mathbf{H}_i = [\tilde{\mathbf{h}}_i\tilde{\mathbf{U}}_i]$, we have indeed

$$\begin{aligned}&\left[\mathbf{T}_1\left(-z\mathbf{H}_1^{\mathsf{H}}(\mathbf{H}_2\mathbf{H}_2^{\mathsf{H}} - z\mathbf{I}_N)^{-1}\mathbf{H}_1 - z\mathbf{I}_{n_1}\right)^{-1}\right]_{11}\\ &= \tau_{11}\left[-z - z\tilde{\mathbf{h}}_1^{\mathsf{H}}\left(\tilde{\mathbf{U}}_1\tilde{\mathbf{U}}_1^{\mathsf{H}} + \mathbf{H}_2\mathbf{H}_2^{\mathsf{H}} - z\mathbf{I}_N\right)^{-1}\tilde{\mathbf{h}}_1\right]^{-1}\\ &\simeq \tau_{11}\left[-z - zc_1\tau_{11}\frac{1}{N}\operatorname{tr}\mathbf{R}_1\left(\mathbf{H}_1\mathbf{H}_1^{\mathsf{H}} + \mathbf{H}_2\mathbf{H}_2^{\mathsf{H}} - z\mathbf{I}_N\right)^{-1}\right]^{-1}\end{aligned}$$

with τ_{ij} the jth diagonal entry of $\mathbf{T}_i$. The limiting result here arises from the trace lemma, Theorem 3.4 along with the rank-1 perturbation lemma, Theorem 3.9. The same result holds when changing ones in twos.

We now denote by e_i and $\bar{e}_i$ the (almost sure) limits of the random quantities

$$e_{\mathbf{B}_N,i} = \frac{1}{N}\operatorname{tr}\mathbf{R}_i\left(\mathbf{H}_1\mathbf{H}_1^{\mathsf{H}} + \mathbf{H}_2\mathbf{H}_2^{\mathsf{H}} - z\mathbf{I}_N\right)^{-1}$$

and

$$\bar{e}_{\mathbf{B}_N,i} = \frac{1}{N}\operatorname{tr}\mathbf{T}_i\left(-z\mathbf{H}_1^{\mathsf{H}}(\mathbf{H}_2\mathbf{H}_2^{\mathsf{H}} - z\mathbf{I}_N)^{-1}\mathbf{H}_1 - z\mathbf{I}_{n_1}\right)^{-1}$$

respectively, as $F^{\mathbf{T}_i}$ and $F^{\mathbf{R}_i}$ converge in the large N limit. These limits exist here since we forced $\mathbf{R}_1$ and $\mathbf{R}_2$ to be co-diagonalizable. We find

$$e_i = \lim_{N\to\infty} \frac{1}{N} \operatorname{tr} \mathbf{R}_i \left(-z\bar{e}_{\mathbf{B}_N,i}\mathbf{R}_1 - z\bar{e}_{\mathbf{B}_N,i}\mathbf{R}_2 - z\mathbf{I}_N\right)^{-1}$$
$$\bar{e}_i = \lim_{N\to\infty} \frac{1}{N} \operatorname{tr} \mathbf{T}_i \left(-zc_i e_{\mathbf{B}_N,i}\mathbf{T}_i - z\mathbf{I}_{n_i}\right)^{-1}$$

where the type of convergence is left to be determined. From this short calculus, we can infer the form of (6.7).

This derivation obviously only provides a hint on the deterministic equivalent for $m_N(z)$. It also provides the aforementioned observation that $m_N(z)$ is not itself solution of a fixed-point equation, although $e_{N,1}(z), \ldots, e_{N,K}(z)$ are. To prove Theorem 6.1, irrespective of the conditions imposed on $\mathbf{R}_1, \ldots, \mathbf{R}_K$, $\mathbf{T}_1, \ldots, \mathbf{T}_K$ and $\mathbf{A}$, we will successively go through four steps, given below. For readability, we consider the case $K = 1$ and discard the useless indexes. The generalization to $K \geq 1$ is rather simple for most of the steps but requires cumbersome additional calculus for some particular aspects. These pieces of calculus are not interesting here, the reader being invited to refer to [Couillet *et al.*, 2011a] for more details. The four-step procedure is detailed below.

- *Step 1.* We first seek a function f_N, such that, for $z \in \mathbb{C}^+$

  $$e_{\mathbf{B}_N}(z) - f_N(e_{\mathbf{B}_N}(z); z) \xrightarrow{\text{a.s.}} 0$$

  as $N \to \infty$, where $e_{\mathbf{B}_N}(z) = \frac{1}{N} \operatorname{tr} \mathbf{R}(\mathbf{B}_N - z\mathbf{I}_N)^{-1}$. This function f_N was already inferred by the Marčenko–Pastur approach. Now, we will make this step rigorous by using the *Bai and Silverstein approach*, as is done in, e.g., [Dozier and Silverstein, 2007a; Silverstein and Bai, 1995]. Basically, the function f_N will be found using an inference procedure. That is, starting from a very general form of f_N, i.e. $f_N = \frac{1}{N} \operatorname{tr} \mathbf{R}\mathbf{D}^{-1}$ for some matrix $\mathbf{D} \in \mathbb{C}^{N\times N}$ (not yet written as a function of z or $e_{\mathbf{B}_N}(z)$), we will evaluate the difference $e_{\mathbf{B}_N}(z) - f_N$ and progressively discover which matrix $\mathbf{D}$ will make this difference increasingly small for large N.
- *Step 2.* For fixed N, we prove the existence of a solution to the implicit equation in the dummy variable e

  $$f_N(e; z) = e. \tag{6.9}$$

  This is often performed by proving the existence of a sequence $e_{N,1}, e_{N,2}, \ldots$, lying in a compact space such that $f_N(e_{N,k}; z) - e_{N,k}$ converges to zero, in which case there exists at least one converging subsequence of $e_{N,1}, e_{N,2}, \ldots$, whose limit e_N satisfies (6.9).
- *Step 3.* Still for fixed N, we prove the uniqueness of the solution of (6.9) lying in some specific space and we call this solution $e_N(z)$. This is classically performed by assuming the existence of a second distinct solution and by exhibiting a contradiction.

- *Step 4.* We finally prove that

$$e_{\mathbf{B}_N}(z) - e_N(z) \xrightarrow{\text{a.s.}} 0$$

and, similarly, that

$$m_{\mathbf{B}_N}(z) - m_N(z) \xrightarrow{\text{a.s.}} 0$$

as $N \to \infty$, with $m_N(z) \triangleq g_N(e_N(z); z)$ for some function g_N.

At first, following the works of Bai and Silverstein, a truncation, centralization, and rescaling step is required to replace the matrices $\mathbf{X}$, $\mathbf{R}$, and $\mathbf{T}$ by truncated versions $\hat{\mathbf{X}}$, $\hat{\mathbf{R}}$, and $\hat{\mathbf{T}}$, respectively, such that the entries of $\hat{\mathbf{X}}$ have zero mean, $\|\hat{\mathbf{X}}\| \leq k \log(N)$, for some constant k, $\|\hat{\mathbf{R}}\| \leq \log(N)$ and $\|\hat{\mathbf{T}}\| \leq \log(N)$. Similar to the truncation steps presented in Section 3.2.2, it is shown in [Couillet *et al.*, 2011a] that these truncations do not restrict the generality of the final result for $\{F^{\mathbf{T}}\}$ and $\{F^{\mathbf{R}}\}$ forming tight sequences, that is:

$$F^{\hat{\mathbf{R}}^{\frac{1}{2}}\hat{\mathbf{X}}\hat{\mathbf{T}}\hat{\mathbf{X}}^{\mathsf{H}}\hat{\mathbf{R}}^{\frac{1}{2}}} - F^{\mathbf{R}^{\frac{1}{2}}\mathbf{X}\mathbf{T}\mathbf{X}^{\mathsf{H}}\mathbf{R}^{\frac{1}{2}}} \Rightarrow 0$$

almost surely, as N grows large. Therefore, we can from now on work with these truncated matrices. We recall that the main interest of this procedure is to be able to derive a deterministic equivalent (or l.s.d.) of the underlying random matrix model without the need for any moment assumption on the entries of $\mathbf{X}$, by replacing the entries of $\mathbf{X}$ by truncated random variables that have moments of all orders. Here, the interest is in fact two-fold, since, in addition to truncating the entries of $\mathbf{X}$, also the entries of $\mathbf{T}$ and $\mathbf{R}$ are truncated in order to be able to prove results for matrices $\mathbf{T}$ and $\mathbf{R}$ that in reality have eigenvalues growing very large but that will be assumed to have entries bounded by $\log(N)$. For readability in the following, we rename $\mathbf{X}$, $\mathbf{T}$, and $\mathbf{R}$ the truncated matrices.

Remark 6.2. Alternatively, expected values can be used to discard the stochastic character. This introduces an additional convergence step, which is the approach followed by Hachem, Najim, and Loubaton in several publications, e.g., [Hachem *et al.*, 2007] and [Dupuy and Loubaton, 2009]. This additional step consists in first proving the almost sure weak convergence of $F^{\mathbf{B}_N} - G_N$ to zero, for G_N some auxiliary deterministic distribution (such as $G_N = \mathrm{E}[F^{\mathbf{B}_N}]$), before proving the convergence $G_N - F_N \Rightarrow 0$.

Step 1. First convergence step

We start with the introduction of two fundamental identities.

Lemma 6.1 (Resolvent identity)**.** *For invertible* $\mathbf{A}$ *and* $\mathbf{B}$ *matrices, we have the identity*

$$\mathbf{A}^{-1} - \mathbf{B}^{-1} = -\mathbf{A}^{-1}(\mathbf{A} - \mathbf{B})\mathbf{B}^{-1}.$$

This can be verified easily by multiplying both sides on the left by $\mathbf{A}$ and on the right by $\mathbf{B}$ (the resulting equality being equivalent to Lemma 6.1 for $\mathbf{A}$ and $\mathbf{B}$ invertible).

Lemma 6.2 (A matrix inversion lemma, (2.2) in [Silverstein and Bai, 1995]). *Let $\mathbf{A} \in \mathbb{C}^{N\times N}$ be Hermitian invertible, then, for any vector $\mathbf{x} \in \mathbb{C}^N$ and any scalar $\tau \in \mathbb{C}$, such that $\mathbf{A} + \tau\mathbf{x}\mathbf{x}^{\mathsf{H}}$ is invertible*

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

This is verified by multiplying both sides by $\mathbf{A} + \tau\mathbf{x}\mathbf{x}^{\mathsf{H}}$ from the right.

Lemma 6.1 is often referred to as the *resolvent identity*, since it will be mainly used to take the difference between matrices of type $(\mathbf{X} - z\mathbf{I}_N)^{-1}$ and $(\mathbf{Y} - z\mathbf{I}_N)^{-1}$, which we remind are called the *resolvent matrices* of $\mathbf{X}$ and $\mathbf{Y}$, respectively.

The fundamental idea of the approach by Bai and Silverstein is to *guess* the deterministic equivalent of $m_{\mathbf{B}_N}(z)$ by writing it under the form $\frac{1}{N}\operatorname{tr}\mathbf{D}^{-1}$ at first, where $\mathbf{D}$ needs to be determined. This will be performed by taking the difference $m_{\mathbf{B}_N}(z) - \frac{1}{N}\operatorname{tr}\mathbf{D}^{-1}$ and, along the lines of calculus, successively determining the good properties $\mathbf{D}$ must satisfy so that the difference tends to zero almost surely.

We then start by taking $z \in \mathbb{C}^+$ and $\mathbf{D} \in \mathbb{C}^{N\times N}$ some invertible matrix whose normalized trace would ideally be close to $m_{\mathbf{B}_N}(z) = \frac{1}{N}\operatorname{tr}(\mathbf{B}_N - z\mathbf{I}_N)^{-1}$. We then write

$$\mathbf{D}^{-1} - (\mathbf{B}_N - z\mathbf{I}_N)^{-1} = \mathbf{D}^{-1}(\mathbf{A} + \mathbf{R}^{\frac{1}{2}}\mathbf{X}\mathbf{T}\mathbf{X}^{\mathsf{H}}\mathbf{R}^{\frac{1}{2}} - z\mathbf{I}_N - \mathbf{D})(\mathbf{B}_N - z\mathbf{I}_N)^{-1} \tag{6.10}$$

using Lemma 6.1.

Notice here that, since $\mathbf{B}_N$ is Hermitian non-negative definite, and $z \in \mathbb{C}^+$, the term $(\mathbf{B}_N - z\mathbf{I}_N)^{-1}$ has uniformly bounded spectral norm (bounded by $1/\Im[z]$). Since $\mathbf{D}^{-1}$ is desired to be close to $(\mathbf{B}_N - z\mathbf{I}_N)^{-1}$, the same property should also hold for $\mathbf{D}^{-1}$. In order for the normalized trace of (6.10) to be small, we need therefore to focus exclusively on the inner difference on the right-hand side. It seems then interesting at this point to write $\mathbf{D} \triangleq \mathbf{A} - z\mathbf{I}_N + p_N\mathbf{R}$ for p_N left to be defined. This leads to

$$\begin{aligned}
&\mathbf{D}^{-1} - (\mathbf{B}_N - z\mathbf{I}_N)^{-1} \\
&= \mathbf{D}^{-1}\mathbf{R}^{\frac{1}{2}}\left(\mathbf{X}\mathbf{T}\mathbf{X}^{\mathsf{H}}\right)\mathbf{R}^{\frac{1}{2}}(\mathbf{B}_N - z\mathbf{I}_N)^{-1} - p_N\mathbf{D}^{-1}\mathbf{R}(\mathbf{B}_N - z\mathbf{I}_N)^{-1} \\
&= \mathbf{D}^{-1}\sum_{j=1}^{n}\tau_j\mathbf{R}^{\frac{1}{2}}\mathbf{x}_j\mathbf{x}_j^{\mathsf{H}}\mathbf{R}^{\frac{1}{2}}(\mathbf{B}_N - z\mathbf{I}_N)^{-1} - p_N\mathbf{D}^{-1}\mathbf{R}(\mathbf{B}_N - z\mathbf{I}_N)^{-1}
\end{aligned}$$

where in the second equality we used the fact that $\mathbf{X}\mathbf{T}\mathbf{X}^{\mathsf{H}} = \sum_{j=1}^{n}\tau_j\mathbf{x}_j\mathbf{x}_j^{\mathsf{H}}$, with $\mathbf{x}_j \in \mathbb{C}^N$ the jth column of $\mathbf{X}$ and τ_j the jth diagonal element of $\mathbf{T}$. Denoting $\mathbf{B}_{(j)} = \mathbf{B}_N - \tau_j\mathbf{R}^{\frac{1}{2}}\mathbf{x}_j\mathbf{x}_j^{\mathsf{H}}\mathbf{R}^{\frac{1}{2}}$, i.e. $\mathbf{B}_N$ with column j removed, and using

Lemma 6.2 for the matrix $\mathbf{B}_{(j)}$, we have:

$$\begin{aligned}&\mathbf{D}^{-1} - (\mathbf{B}_N - z\mathbf{I}_N)^{-1}\\&= \sum_{j=1}^{n} \tau_j \frac{\mathbf{D}^{-1}\mathbf{R}^{\frac{1}{2}}\mathbf{x}_j\mathbf{x}_j^{\mathsf{H}}\mathbf{R}^{\frac{1}{2}}(\mathbf{B}_{(j)} - z\mathbf{I}_N)^{-1}}{1 + \tau_j \mathbf{x}^{\mathsf{H}}\mathbf{R}^{\frac{1}{2}}(\mathbf{B}_{(j)} - z\mathbf{I}_N)^{-1}\mathbf{R}^{\frac{1}{2}}\mathbf{x}_j} - p_N \mathbf{D}^{-1}\mathbf{R}(\mathbf{B}_N - z\mathbf{I}_N)^{-1}.\end{aligned}$$

Taking the trace on each side, and recalling that, for a vector $\mathbf{x}$ and a matrix $\mathbf{A}$, $\operatorname{tr}(\mathbf{A}\mathbf{x}\mathbf{x}^{\mathsf{H}}) = \operatorname{tr}(\mathbf{x}^{\mathsf{H}}\mathbf{A}\mathbf{x}) = \mathbf{x}^{\mathsf{H}}\mathbf{A}\mathbf{x}$, this becomes

$$\begin{aligned}&\frac{1}{N}\operatorname{tr}\mathbf{D}^{-1} - \frac{1}{N}\operatorname{tr}(\mathbf{B}_N - z\mathbf{I}_N)^{-1}\\&= \frac{1}{N}\sum_{j=1}^{n} \tau_j \frac{\mathbf{x}_j^{\mathsf{H}}\mathbf{R}^{\frac{1}{2}}(\mathbf{B}_{(j)} - z\mathbf{I}_N)^{-1}\mathbf{D}^{-1}\mathbf{R}^{\frac{1}{2}}\mathbf{x}_j}{1 + \tau_j \mathbf{x}^{\mathsf{H}}\mathbf{R}^{\frac{1}{2}}(\mathbf{B}_{(j)} - z\mathbf{I}_N)^{-1}\mathbf{R}^{\frac{1}{2}}\mathbf{x}_j} - p_N \frac{1}{N}\operatorname{tr}\mathbf{R}(\mathbf{B}_N - z\mathbf{I}_N)^{-1}\mathbf{D}^{-1}\end{aligned} \tag{6.11}$$

where quadratic forms of the type $\mathbf{x}^{\mathsf{H}}\mathbf{A}\mathbf{x}$ appear.

Remembering the trace lemma, Theorem 3.4, which can a priori be applied to the terms $\mathbf{x}_j^{\mathsf{H}}\mathbf{R}^{\frac{1}{2}}(\mathbf{B}_{(j)} - z\mathbf{I}_N)^{-1}\mathbf{D}^{-1}\mathbf{R}^{\frac{1}{2}}\mathbf{x}_j$ since $\mathbf{x}_j$ is independent of the matrix $\mathbf{R}^{\frac{1}{2}}(\mathbf{B}_{(j)} - z\mathbf{I}_N)^{-1}\mathbf{D}^{-1}\mathbf{R}^{\frac{1}{2}}$, we notice that by setting

$$p_N = \frac{1}{n}\sum_{j=1}^{n} \frac{\tau_j}{1 + \tau_j c \frac{1}{N}\operatorname{tr}\mathbf{R}(\mathbf{B}_N - z\mathbf{I}_N)^{-1}}.$$

Equation (6.11) becomes

$$\begin{aligned}&\frac{1}{N}\operatorname{tr}\mathbf{D}^{-1} - \frac{1}{N}\operatorname{tr}(\mathbf{B}_N - z\mathbf{I}_N)^{-1}\\&= \frac{1}{N}\sum_{j=1}^{n} \tau_j \left[\frac{\mathbf{x}_j^{\mathsf{H}}\mathbf{R}^{\frac{1}{2}}(\mathbf{B}_{(j)} - z\mathbf{I}_N)^{-1}\mathbf{D}^{-1}\mathbf{R}^{\frac{1}{2}}\mathbf{x}_j}{1 + \tau_j \mathbf{x}^{\mathsf{H}}\mathbf{R}^{\frac{1}{2}}(\mathbf{B}_{(j)} - z\mathbf{I}_N)^{-1}\mathbf{R}^{\frac{1}{2}}\mathbf{x}_j} - \frac{\frac{1}{n}\operatorname{tr}\mathbf{R}(\mathbf{B}_N - z\mathbf{I}_N)^{-1}\mathbf{D}^{-1}}{1 + c\tau_j \frac{1}{N}\operatorname{tr}\mathbf{R}(\mathbf{B}_N - z\mathbf{I}_N)^{-1}}\right]\end{aligned} \tag{6.12}$$

which is suspected to converge to zero as N grows large, since both the numerators and the denominators converge to one another. Let us assume for the time being that the difference effectively goes to zero almost surely. Equation (6.12) implies

$$\frac{1}{N}\operatorname{tr}(\mathbf{B}_N - z\mathbf{I}_N)^{-1} - \frac{1}{N}\operatorname{tr}\left(\mathbf{A} + \frac{1}{n}\sum_{j=1}^{n} \frac{\tau_j \mathbf{R}}{1 + \tau_j c \frac{1}{N}\operatorname{tr}\mathbf{R}(\mathbf{B}_N - z\mathbf{I}_N)^{-1}} - z\mathbf{I}_N\right)^{-1} \xrightarrow{\text{a.s.}} 0$$

which determines $m_{\mathbf{B}_N}(z) = \frac{1}{N}\operatorname{tr}(\mathbf{B}_N - z\mathbf{I}_N)^{-1}$ as a function of the trace $\frac{1}{N}\operatorname{tr}\mathbf{R}(\mathbf{B}_N - z\mathbf{I}_N)^{-1}$, and not as a function of itself. This is the observation made earlier when we obtained a first hint on the form of $m_N(z)$ using the Marčenko–Pastur method, according to which we cannot find a function f_N such that $m_{\mathbf{B}_N}(z) - f_N(m_{\mathbf{B}_N}(z), z) \xrightarrow{\text{a.s.}} 0$. Instead, running the same steps as above, it is rather easy now to observe that

$$\frac{1}{N}\operatorname{tr}\mathbf{R}\mathbf{D}^{-1}-\frac{1}{N}\operatorname{tr}\mathbf{R}(\mathbf{B}_N-z\mathbf{I}_N)^{-1}$$
$$=\frac{1}{N}\sum_{j=1}^{n}\tau_j\left[\frac{\mathbf{x}_j^{\mathsf{H}}\mathbf{R}^{\frac{1}{2}}(\mathbf{B}_{(j)}-z\mathbf{I}_N)^{-1}\mathbf{R}\mathbf{D}^{-1}\mathbf{R}^{\frac{1}{2}}\mathbf{x}_j}{1+\tau_j\mathbf{x}_j^{\mathsf{H}}\mathbf{R}^{\frac{1}{2}}(\mathbf{B}_{(j)}-z\mathbf{I}_N)^{-1}\mathbf{R}^{\frac{1}{2}}\mathbf{x}_j}-\frac{\frac{1}{n}\operatorname{tr}\mathbf{R}(\mathbf{B}_N-z\mathbf{I}_N)^{-1}\mathbf{R}\mathbf{D}^{-1}}{1+\tau_j\frac{c}{N}\operatorname{tr}\mathbf{R}(\mathbf{B}_N-z\mathbf{I}_N)^{-1}}\right]$$

where $\|\mathbf{R}\|\leq\log N$. Then, denoting $e_{\mathbf{B}_N}(z)\triangleq\frac{1}{N}\operatorname{tr}\mathbf{R}(\mathbf{B}_N-z\mathbf{I}_N)^{-1}$, we suspect to have also

$$e_{\mathbf{B}_N}(z)-\frac{1}{N}\operatorname{tr}\mathbf{R}\left(\mathbf{A}+\frac{1}{n}\sum_{j=1}^{n}\frac{\tau_j}{1+\tau_jce_{\mathbf{B}_N}(z)}\mathbf{R}-z\mathbf{I}_N\right)^{-1}\xrightarrow{\text{a.s.}}0$$

and

$$m_{\mathbf{B}_N}(z)-\frac{1}{N}\operatorname{tr}\left(\mathbf{A}+\frac{1}{n}\sum_{j=1}^{n}\frac{\tau_j}{1+\tau_jce_{\mathbf{B}_N}(z)}\mathbf{R}-z\mathbf{I}_N\right)^{-1}\xrightarrow{\text{a.s.}}0$$

which is exactly what was required, i.e. $e_{\mathbf{B}_N}(z)-f_N(e_{\mathbf{B}_N}(z);z)\xrightarrow{\text{a.s.}}0$ with

$$f_N(e;z)=\frac{1}{N}\operatorname{tr}\mathbf{R}\left(\mathbf{A}+\frac{1}{n}\sum_{j=1}^{n}\frac{\tau_j}{1+\tau_jce}\mathbf{R}-z\mathbf{I}_N\right)^{-1}$$

and $m_{\mathbf{B}_N}(z)-g_N(e_{\mathbf{B}_N}(z);z)\xrightarrow{\text{a.s.}}0$ with

$$g_N(e;z)=\frac{1}{N}\operatorname{tr}\left(\mathbf{A}+\frac{1}{n}\sum_{j=1}^{n}\frac{\tau_j}{1+\tau_jce}\mathbf{R}-z\mathbf{I}_N\right)^{-1}.$$

We now prove that the right-hand side of (6.12) converges to zero almost surely. This rather technical part justifies the use of the truncation steps and is the major difference between the works of Bai and Silverstein [Dozier and Silverstein, 2007a; Silverstein and Bai, 1995] and the works of Hachem *et al.* [Hachem *et al.*, 2007]. We first define

$$w_N\triangleq\sum_{j=1}^{n}\frac{\tau_j}{N}\left[\frac{\mathbf{x}_j^{\mathsf{H}}\mathbf{R}^{\frac{1}{2}}(\mathbf{B}_{(j)}-z\mathbf{I}_N)^{-1}\mathbf{R}\mathbf{D}^{-1}\mathbf{R}^{\frac{1}{2}}\mathbf{x}_j}{1+\tau_j\mathbf{x}_j^{\mathsf{H}}\mathbf{R}^{\frac{1}{2}}(\mathbf{B}_{(j)}-z\mathbf{I}_N)^{-1}\mathbf{R}^{\frac{1}{2}}\mathbf{x}_j}-\frac{\frac{1}{n}\operatorname{tr}\mathbf{R}(\mathbf{B}_N-z\mathbf{I}_N)^{-1}\mathbf{R}\mathbf{D}^{-1}}{1+\tau_j\frac{c}{N}\operatorname{tr}\mathbf{R}(\mathbf{B}_N-z\mathbf{I}_N)^{-1}}\right]$$

which we then divide into four terms, in order to successively prove the convergence of the numerators and the denominators. Write

$$w_N=\frac{1}{N}\sum_{j=1}^{n}\tau_j\left(d_j^1+d_j^2+d_j^3+d_j^4\right)$$

where

$$d_j^1 = \frac{\mathbf{x}_j^{\mathsf{H}}\mathbf{R}^{\frac{1}{2}}(\mathbf{B}_{(j)} - z\mathbf{I}_N)^{-1}\mathbf{R}\mathbf{D}^{-1}\mathbf{R}^{\frac{1}{2}}\mathbf{x}_j}{1+\tau_j\mathbf{x}_j^{\mathsf{H}}\mathbf{R}^{\frac{1}{2}}(\mathbf{B}_{(j)} - z\mathbf{I}_N)^{-1}\mathbf{R}^{\frac{1}{2}}\mathbf{x}_j} - \frac{\mathbf{x}_j^{\mathsf{H}}\mathbf{R}^{\frac{1}{2}}(\mathbf{B}_{(j)} - z\mathbf{I}_N)^{-1}\mathbf{R}\mathbf{D}_{(j)}^{-1}\mathbf{R}^{\frac{1}{2}}\mathbf{x}_j}{1+\tau_j\mathbf{x}_j^{\mathsf{H}}\mathbf{R}^{\frac{1}{2}}(\mathbf{B}_{(j)} - z\mathbf{I}_N)^{-1}\mathbf{R}^{\frac{1}{2}}\mathbf{x}_j}$$

$$d_j^2 = \frac{\mathbf{x}_j^{\mathsf{H}}\mathbf{R}^{\frac{1}{2}}(\mathbf{B}_{(j)} - z\mathbf{I}_N)^{-1}\mathbf{R}\mathbf{D}_{(j)}^{-1}\mathbf{R}^{\frac{1}{2}}\mathbf{x}_j}{1+\tau_j\mathbf{x}_j^{\mathsf{H}}\mathbf{R}^{\frac{1}{2}}(\mathbf{B}_{(j)} - z\mathbf{I}_N)^{-1}\mathbf{R}^{\frac{1}{2}}\mathbf{x}_j} - \frac{\frac{1}{n}\operatorname{tr}\mathbf{R}(\mathbf{B}_{(j)} - z\mathbf{I}_N)^{-1}\mathbf{R}\mathbf{D}_{(j)}^{-1}}{1+\tau_j\mathbf{x}_j^{\mathsf{H}}\mathbf{R}^{\frac{1}{2}}(\mathbf{B}_{(j)} - z\mathbf{I}_N)^{-1}\mathbf{R}^{\frac{1}{2}}\mathbf{x}_j}$$

$$d_j^3 = \frac{\frac{1}{n}\operatorname{tr}\mathbf{R}(\mathbf{B}_{(j)} - z\mathbf{I}_N)^{-1}\mathbf{R}\mathbf{D}_{(j)}^{-1}}{1+\tau_j\mathbf{x}_j^{\mathsf{H}}\mathbf{R}^{\frac{1}{2}}(\mathbf{B}_{(j)} - z\mathbf{I}_N)^{-1}\mathbf{R}^{\frac{1}{2}}\mathbf{x}_j} - \frac{\frac{1}{n}\operatorname{tr}\mathbf{R}(\mathbf{B}_N - z\mathbf{I}_N)^{-1}\mathbf{R}\mathbf{D}^{-1}}{1+\tau_j\mathbf{x}_j^{\mathsf{H}}\mathbf{R}^{\frac{1}{2}}(\mathbf{B}_{(j)} - z\mathbf{I}_N)^{-1}\mathbf{R}^{\frac{1}{2}}\mathbf{x}_j}$$

$$d_j^4 = \frac{\frac{1}{n}\operatorname{tr}\mathbf{R}(\mathbf{B}_N - z\mathbf{I}_N)^{-1}\mathbf{R}\mathbf{D}^{-1}}{1+\tau_j\mathbf{x}_j^{\mathsf{H}}\mathbf{R}^{\frac{1}{2}}(\mathbf{B}_{(j)} - z\mathbf{I}_N)^{-1}\mathbf{R}^{\frac{1}{2}}\mathbf{x}_j} - \frac{\frac{1}{n}\operatorname{tr}\mathbf{R}(\mathbf{B}_N - z\mathbf{I}_N)^{-1}\mathbf{R}\mathbf{D}^{-1}}{1+c\tau_j e_{\mathbf{B}_N}}$$

where we introduced $\mathbf{D}_{(j)} = \mathbf{A} + \frac{1}{n}\sum_{k=1}^n \frac{\tau_k}{1+\tau_k c e_{\mathbf{B}_{(j)}}(z)}\mathbf{R} - z\mathbf{I}_N$, i.e. $\mathbf{D}$ with $e_{\mathbf{B}_N}(z)$ replaced by $e_{\mathbf{B}_{(j)}}(z)$. Under these notations, it is simple to show that $w_N \xrightarrow{\text{a.s.}} 0$ since every term d_j^k can be shown to go fast to zero.

One of the difficulties in proving that the d_j^k tends to zero at a sufficiently fast rate lies in providing inequalities for the quadratic terms of the type $\mathbf{y}^{\mathsf{H}}(\mathbf{A} - z\mathbf{I}_N)^{-1}\mathbf{y}$ present in the denominators. For this, we use Corollary 3.2, which states that, for any non-negative definite matrix $\mathbf{A}$, $\mathbf{y} \in \mathbb{C}^N$ and for $z \in \mathbb{C}^+$

$$\left|\frac{1}{1+\tau_j\mathbf{y}^{\mathsf{H}}(\mathbf{A} - z\mathbf{I}_N)^{-1}\mathbf{y}}\right| \leq \frac{|z|}{\Im[z]}. \tag{6.13}$$

Also, we need to ensure that $\mathbf{D}^{-1}$ and $\mathbf{D}_{(j)}^{-1}$ have uniformly bounded spectral norm. This unfolds from the following lemma.

Lemma 6.3 (Lemma 8 of [Couillet *et al.*, 2011a])**.** *Let* $\mathbf{D} = \mathbf{A} + i\mathbf{B} + iv\mathbf{I}_N$, *with* $\mathbf{A} \in \mathbb{C}^{N\times N}$ *Hermitian,* $\mathbf{B} \in \mathbb{C}^{N\times N}$ *Hermitian non-negative and* $v > 0$. *Then* $\|\mathbf{D}\| \leq v^{-1}$.

Proof. Noticing that $\mathbf{D}\mathbf{D}^{\mathsf{H}} = (\mathbf{A}+i\mathbf{B})(\mathbf{A}-i\mathbf{B}) + v^2\mathbf{I}_N + 2v\mathbf{B}$, the smallest eigenvalue of $\mathbf{D}\mathbf{D}^{\mathsf{H}}$ is greater than or equal to v^2 and therefore $\|\mathbf{D}^{-1}\| \leq v^{-1}$. □

At this step, we need to invoke the generalized trace lemma, Theorem 3.12. From Theorem 3.12, (6.13), Lemma 6.3 and the inequalities due to the truncation steps, we can then show that

$$\tau_j|d_j^1| \leq \|\mathbf{x}_j\|^2 \frac{c\log^7 N|z|^3}{N\Im[z]^7}$$

$$\tau_j|d_j^2| \leq \frac{\log N\left|\mathbf{x}_j^{\mathsf{H}}\mathbf{R}^{\frac{1}{2}}(\mathbf{B}_{(j)} - z\mathbf{I}_N)^{-1}\mathbf{R}\mathbf{D}_{(j)}^{-1}\mathbf{R}^{\frac{1}{2}}\mathbf{x}_j - \frac{1}{n}\operatorname{tr}\mathbf{R}(\mathbf{B}_{(j)} - z\mathbf{I}_N)^{-1}\mathbf{R}\mathbf{D}_{(j)}^{-1}\right|}{\Im[z]|z|^{-1}}$$

$$\tau_j|d_j^3| \leq \frac{|z|\log^3 N}{\Im[z]N}\left(\frac{1}{\Im[z]^2} + \frac{c|z|^2\log^3 N}{\Im[z]^6}\right)$$

$$\tau_j|d_j^4| \leq \frac{\log^4 N\left(\left|\mathbf{x}_j^{\mathsf{H}}\mathbf{R}^{\frac{1}{2}}(\mathbf{B}_{(j)} - z\mathbf{I}_N)^{-1}\mathbf{R}^{\frac{1}{2}}\mathbf{x}_j - \frac{1}{n}\operatorname{tr}\mathbf{R}(\mathbf{B}_{(j)} - z\mathbf{I}_N)^{-1}\right| + \frac{\log N}{N\Im[z]}\right)}{\Im[z]^3|z|^{-1}}.$$

Applying the trace lemma for truncated variables, Theorem 3.12, and classical inequalities, there exists $\bar{K} > 0$ such that we have simultaneously

$$\mathrm{E}|\|\mathbf{x}_j\|^2 - 1|^6 \leq \frac{\bar{K}\log^{12} N}{N^3}$$

and

$$\mathrm{E}|\mathbf{x}_j^{\mathsf{H}}\mathbf{R}^{\frac{1}{2}}(\mathbf{B}_{(j)} - z\mathbf{I}_N)^{-1}\mathbf{R}\mathbf{D}_{(j)}^{-1}\mathbf{R}^{\frac{1}{2}}\mathbf{x}_j - \frac{1}{n}\operatorname{tr}\mathbf{R}(\mathbf{B}_{(j)} - z\mathbf{I}_N)^{-1}\mathbf{R}\mathbf{D}_{(j)}^{-1}|^6$$
$$\leq \frac{\bar{K}\log^{24} N}{N^3\Im[z]^{12}}$$

and

$$\mathrm{E}|\mathbf{x}_j^{\mathsf{H}}\mathbf{R}^{\frac{1}{2}}(\mathbf{B}_{(j)} - z\mathbf{I}_N)^{-1}\mathbf{R}^{\frac{1}{2}}\mathbf{x}_j - \frac{1}{n}\operatorname{tr}\mathbf{R}^{\frac{1}{2}}(\mathbf{B}_{(j)} - z\mathbf{I}_N)^{-1}\mathbf{R}^{\frac{1}{2}}|^6$$
$$\leq \frac{\bar{K}\log^{18} N}{N^3\Im[z]^6}.$$

All three moments above, when summed over the n indexes j and multiplied by any power of $\log N$, are summable. Applying the Markov inequality, Theorem 3.5, the Borel–Cantelli lemma, Theorem 3.6, and the line of arguments used in the proof of the Marčenko–Pastur law, we conclude that, for any $k > 0$, $\log^k N \max_{j\leq n} \tau_j d_j \xrightarrow{\text{a.s.}} 0$ as $N \to \infty$, and therefore:

$$e_{\mathbf{B}_N}(z) - f_N(e_{\mathbf{B}_N}(z); z) \xrightarrow{\text{a.s.}} 0$$
$$m_{\mathbf{B}_N}(z) - g_N(e_{\mathbf{B}_N}(z); z) \xrightarrow{\text{a.s.}} 0.$$

This convergence result is similar to that of Theorem (3.22), although in the latter each side of the minus sign converges, when the eigenvalue distributions of the deterministic matrices in the model converge. In the present case, even if the series $\{F^{\mathbf{T}}\}$ and $\{F^{\mathbf{R}}\}$ converge, it is not necessarily true that either $e_{\mathbf{B}_N}(z)$ or $f_N(e_{\mathbf{B}_N}(z), z)$ converges.

We wish to go further here by showing that, for all finite N, $f_N(e; z) = e$ has a solution (Step 2), that this solution is unique in some space (Step 3) and that, denoting $e_N(z)$ this solution, $e_N(z) - e_{\mathbf{B}_N}(z) \xrightarrow{\text{a.s.}} 0$ (Step 4). This will imply naturally that $m_N(z) \triangleq g_N(e_N(z); z)$ satisfies $m_{\mathbf{B}_N}(z) - m_N(z) \xrightarrow{\text{a.s.}} 0$, for all $z \in \mathbb{C}^+$. Vitali's convergence theorem, Theorem 3.11, will conclude the proof by showing that $m_{\mathbf{B}_N}(z) - m_N(z) \xrightarrow{\text{a.s.}} 0$ for all z outside the positive real half-line.

Step 2. Existence of a solution

We now show that the implicit equation $e = f_N(e; z)$ in the dummy variable e has a solution for each finite N. For this, we use a special trick that consists in growing the matrices dimensions asymptotically large while maintaining the deterministic components untouched, i.e. while maintaining $F^{\mathbf{R}}$ and $F^{\mathbf{T}}$ the same. The idea is to fix N and consider for all $j > 0$ the matrices $\mathbf{T}_{[j]} = \mathbf{T} \otimes \mathbf{I}_j \in \mathbb{C}^{jn\times jn}$, $\mathbf{R}_{[j]} =$

$\mathbf{R} \otimes \mathbf{I}_j \in \mathbb{C}^{jN \times jN}$ and $\mathbf{A}_{[j]} = \mathbf{A} \otimes \mathbf{I}_j \in \mathbb{C}^{jN \times jN}$. For a given x

$$f_{[j]}(x; z) \triangleq \frac{1}{jN} \operatorname{tr} \mathbf{R}_{[j]} \left(\mathbf{A}_{[j]} + \int \frac{\tau dF^{\mathbf{T}_{[j]}}(\tau)}{1 + c\tau x} \mathbf{R}_{[j]} - z \mathbf{I}_{Nj} \right)^{-1}$$

which is constant whatever j and equal to $f_N(x; z)$. Defining

$$\mathbf{B}_{[j]} = \mathbf{A}_{[j]} + \mathbf{R}_{[j]}^{\frac{1}{2}} \mathbf{X} \mathbf{T}_{[j]} \mathbf{X}^{\mathsf{H}} \mathbf{R}_{[j]}^{\frac{1}{2}}$$

for $\mathbf{X} \in \mathbb{C}^{Nj \times nj}$ with i.i.d. entries of zero mean and variance $1/(nj)$

$$e_{\mathbf{B}_{[j]}}(z) = \frac{1}{jN} \operatorname{tr} \mathbf{R}_{[j]} (\mathbf{A}_{[j]} + \mathbf{R}_{[j]}^{\frac{1}{2}} \mathbf{X} \mathbf{T}_{[j]} \mathbf{X}^{\mathsf{H}} \mathbf{R}_{[j]}^{\frac{1}{2}} - z \mathbf{I}_{Nj})^{-1}.$$

With the notations of Step 1, $w_{Nj} \to 0$ as $j \to \infty$, for all sequences $\mathbf{B}_{[1]}, \mathbf{B}_{[2]}, \ldots$ in a set of probability one. Take such a sequence. Noticing that both $e_{\mathbf{B}_{[j]}}(z)$ and the integrand $\frac{\tau}{1+c\tau e_{\mathbf{B}_{[j]}}(z)}$ of $f_{[j]}(x, z)$ are uniformly bounded for *fixed* N and growing j, there exists a subsequence of $e_{\mathbf{B}_{[1]}}, e_{\mathbf{B}_{[2]}}, \ldots$ over which they both converge, when $j \to \infty$, to some limits e and $\tau(1 + c\tau e)^{-1}$, respectively. But since $w_{jN} \to 0$ for this realization of $e_{\mathbf{B}_{[1]}}, e_{\mathbf{B}_{[2]}}, \ldots$, for growing j, we have that $e = \lim_j f_{[j]}(e, z)$. But we also have that, for all j, $f_{[j]}(e, z) = f_N(e, z)$. We therefore conclude that $e = f_N(e, z)$ and we have found a solution.

Step 3. Uniqueness of the solution

Uniqueness is shown classically by considering two hypothetical solutions $e \in \mathbb{C}^+$ and $\underline{e} \in \mathbb{C}^+$ to (6.6) and by showing then that $e - \underline{e} = \gamma(e - \underline{e})$, where $|\gamma|$ must be shown to be less than one. Indeed, taking the difference $e - \underline{e}$, we have with the resolvent identity

$$\begin{aligned} e - \underline{e} &= \frac{1}{N} \operatorname{tr} \mathbf{R} \mathbf{D}_e^{-1} - \frac{1}{N} \operatorname{tr} \mathbf{R} \mathbf{D}_{\underline{e}}^{-1} \\ &= \frac{1}{N} \operatorname{tr} \mathbf{R} \mathbf{D}_e^{-1} \left(\int \frac{c\tau^2 (e - \underline{e}) dF^{\mathbf{T}}(\tau)}{(1 + c\tau e)(1 + c\tau \underline{e})} \right) \mathbf{R} \mathbf{D}_{\underline{e}}^{-1} \end{aligned}$$

in which $\mathbf{D}_e$ and $\mathbf{D}_{\underline{e}}$ are the matrix $\mathbf{D}$ with $e_{\mathbf{B}_N}(z)$ replaced by e and $\underline{e}$, respectively. This leads to the expression of γ as follows.

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

Applying the Cauchy–Schwarz inequality to the diagonal elements of $\frac{1}{N} \mathbf{D}_e^{-1} \mathbf{R} \int \frac{\sqrt{c\tau}}{1+c\tau e} dF^{\mathbf{T}}(\tau)$ and of $\frac{1}{N} \mathbf{D}_{\underline{e}}^{-1} \mathbf{R} \int \frac{\sqrt{c\tau}}{1+c\tau \underline{e}} dF^{\mathbf{T}}(\tau)$, we then have

$$\begin{aligned} |\gamma| &\leq \sqrt{\int \frac{c\tau^2 dF^{\mathbf{T}}(\tau)}{|1 + c\tau e|^2 N} \operatorname{tr} \mathbf{D}_e^{-1} \mathbf{R} (\mathbf{D}_e^{\mathsf{H}})^{-1} \mathbf{R}} \sqrt{\int \frac{c\tau^2 dF^{\mathbf{T}}(\tau)}{|1 + c\tau \underline{e}|^2 N} \operatorname{tr} \mathbf{D}_{\underline{e}}^{-1} \mathbf{R} (\mathbf{D}_{\underline{e}}^{\mathsf{H}})^{-1} \mathbf{R}} \\ &\triangleq \sqrt{\alpha} \sqrt{\underline{\alpha}}. \end{aligned}$$

We now proceed to a parallel computation of $\Im[e]$ and $\Im[\underline{e}]$ in the hope of retrieving both expressions in the right-hand side of the above equation.

Introducing the product $(\mathbf{D}_e^{\mathsf{H}})^{-1}\mathbf{D}_e^{\mathsf{H}}$ in the trace, we first write e under the form

$$e = \frac{1}{N}\operatorname{tr}\left(\mathbf{D}_e^{-1}\mathbf{R}(\mathbf{D}_e^{\mathsf{H}})^{-1}\left(\mathbf{A} + \left[\int \frac{\tau}{1+c\tau e^*}dF^{\mathbf{T}}(\tau)\right]\mathbf{R} - z^*\mathbf{I}_N\right)\right). \quad (6.14)$$

Taking the imaginary part, this is:

$$\begin{aligned}\Im[e] &= \frac{1}{N}\operatorname{tr}\left(\mathbf{D}_e^{-1}\mathbf{R}(\mathbf{D}_e^{\mathsf{H}})^{-1}\left(\left[\int \frac{c\tau^2\Im[e]}{|1+c\tau e|^2}dF^{\mathbf{T}}(\tau)\right]\mathbf{R} + \Im[z]\mathbf{I}_N\right)\right)\\ &= \Im[e]\alpha + \Im[z]\beta\end{aligned}$$

where

$$\beta \triangleq \frac{1}{N}\operatorname{tr}\mathbf{D}_e^{-1}\mathbf{R}(\mathbf{D}_e^{\mathsf{H}})^{-1}$$

is positive whenever $\mathbf{R} \neq 0$, and similarly $\Im[\underline{e}] = \underline{\alpha}\Im[\underline{e}] + \Im[z]\underline{\beta}$, $\underline{\beta} > 0$ with

$$\underline{\beta} \triangleq \frac{1}{N}\operatorname{tr}\mathbf{D}_{\underline{e}}^{-1}\mathbf{R}(\mathbf{D}_{\underline{e}}^{\mathsf{H}})^{-1}.$$

Notice also that

$$\alpha = \frac{\alpha\Im[e]}{\Im[e]} = \frac{\alpha\Im[e]}{\alpha\Im[e] + \beta\Im[z]} < 1$$

and

$$\underline{\alpha} = \frac{\alpha\Im[\underline{e}]}{\Im[\underline{e}]} = \frac{\underline{\alpha}\Im[\underline{e}]}{\alpha\Im[\underline{e}] + \underline{\beta}\Im[z]} < 1.$$

As a consequence

$$|\gamma| \leq \sqrt{\alpha}\sqrt{\underline{\alpha}} = \sqrt{\frac{\Im[e]\alpha}{\Im[e]\alpha + \Im[z]\beta}}\sqrt{\frac{\Im[\underline{e}]\underline{\alpha}}{\Im[\underline{e}]\underline{\alpha} + \Im[z]\underline{\beta}}} < 1$$

as requested. The case $\mathbf{R} = 0$ is easy to verify.

Remark 6.3. Note that this uniqueness argument is slightly more technical when $K > 1$. In this case, uniqueness of the vector $e_1, \ldots, e_K$ (under the notations of Theorem 6.1) needs be proved. Denoting $\mathbf{e} \triangleq (e_1, \ldots, e_K)^{\mathsf{T}}$, this requires to show that, for two solutions $\mathbf{e}$ and $\underline{\mathbf{e}}$ of the implicit equation, $(\mathbf{e} - \underline{\mathbf{e}}) = \mathbf{\Gamma}(\mathbf{e} - \underline{\mathbf{e}})$, where $\mathbf{\Gamma}$ has spectral radius less than one. To this end, a possible approach is to show that $|\Gamma_{ij}| \leq \alpha_{ij}^{\frac{1}{2}}\underline{\alpha}_{ij}^{\frac{1}{2}}$, for α_{ij} and $\underline{\alpha}_{ij}$ defined similar as in Step 3. Then, applying some classical matrix lemmas (Theorem 8.1.18 of [Horn and Johnson, 1985] and Lemma 5.7.9 of [Horn and Johnson, 1991]), the previous inequality implies that

$$\|\mathbf{\Gamma}\| \leq \|(\alpha_{ij}^{\frac{1}{2}}\underline{\alpha}_{ij}^{\frac{1}{2}})_{ij}\|$$

where $(\alpha_{ij}^{\frac{1}{2}}\underline{\alpha}_{ij}^{\frac{1}{2}})_{ij}$ is the matrix with (i,j) entry $\alpha_{ij}^{\frac{1}{2}}\underline{\alpha}_{ij}^{\frac{1}{2}}$ and the norm is the matrix spectral norm. We further have that

$$\|(\alpha_{ij}^{\frac{1}{2}}\underline{\alpha}_{ij}^{\frac{1}{2}})_{ij}\| \leq \|\mathbf{A}\|^{\frac{1}{2}}\|\underline{\mathbf{A}}\|^{\frac{1}{2}}$$

where $\mathbf{A}$ and $\underline{\mathbf{A}}$ are now matrices with (i,j) entry α_{ij} and $\underline{\alpha}_{ij}$, respectively. The multi-dimensional problem therefore boils down to proving that $\|\mathbf{A}\| < 1$ and $\|\underline{\mathbf{A}}\| < 1$. This unfolds from yet another classical matrix lemma (Theorem 2.1 of [Seneta, 1981]), which states in our current situation that, if we have the vectorial relation

$$\Im[\mathbf{e}] = \mathbf{A}\Im[\mathbf{e}] + \Im[z]\mathbf{b}$$

with $\Im[\mathbf{e}]$ and $\mathbf{b}$ vectors of *positive* entries and $\Im[z] > 0$, then $\|\mathbf{A}\| < 1$. The above relation generalizes, without much difficulty, the relation $\Im[e] = \Im[e]\alpha + \Im[z]\beta$ obtained above.

Step 4. Final convergence step
We finally need to show that $e_N - e_{\mathbf{B}_N}(z) \xrightarrow{\text{a.s.}} 0$. This is performed using a similar argument as for uniqueness, i.e. $e_N - e_{\mathbf{B}_N}(z) = \gamma(e_N - e_{\mathbf{B}_N}(z)) + w_N$, where $w_N \to 0$ as $N \to \infty$ and $|\gamma| < 1$; this is true for any $e_{\mathbf{B}_N}(z)$ taken from a space of probability one such that $w_N \to 0$. The major difficulty compared to the previous proof is to control precisely w_N.

The details are as follows. We will show that, for any $\ell > 0$, almost surely

$$\lim_{N\to\infty} \log^{\ell} N(e_{\mathbf{B}_N} - e_N) = 0. \tag{6.15}$$

Let α_N, β_N be the values as above for which $\Im[e_N] = \Im[e_N]\alpha_N + \Im[z]\beta_N$. Using truncation inequalities

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

Therefore

$$\begin{aligned}\alpha_N &= \frac{\Im[e_N]\alpha_N}{\Im[e_N]\alpha_N + \Im[z]\beta_N}\\ &= \frac{\Im[e_N]\frac{\alpha_N}{\beta_N}}{\Im[z] + \Im[e_N]\frac{\alpha_N}{\beta_N}}\\ &\leq \frac{\log^2 N|z|}{\Im[z]^2 + \log^2 N|z|}.\end{aligned} \tag{6.16}$$

We also have

$$e_{\mathbf{B}_N}(z) = \frac{1}{N}\operatorname{tr}\mathbf{D}^{-1}\mathbf{R} - w_N.$$

We write as in Step 3

$$\begin{aligned}
&\Im[e_{\mathbf{B}_N}] \\
&= \frac{1}{N}\operatorname{tr}\left(\mathbf{D}^{-1}\mathbf{R}(\mathbf{D}^{\mathsf{H}})^{-1}\left(\left[\int \frac{c\tau^2\Im[e_{\mathbf{B}_N}]}{|1+c\tau e_{\mathbf{B}_N}|^2}dF^{\mathbf{T}}(\tau)\right]\mathbf{R}+\Im[z]\mathbf{I}_N\right)\right)-\Im[w_N] \\
&\triangleq \Im[e_{\mathbf{B}_N}]\alpha_{\mathbf{B}_N}+\Im[z]\beta_{\mathbf{B}_N}-\Im[w_N].
\end{aligned}$$

Similarly to Step 3, we have $e_{\mathbf{B}_N}-e_N=\gamma(e_{\mathbf{B}_N}-e_N)+w_N$, where now

$$|\gamma|\leq\sqrt{\alpha_{\mathbf{B}_N}}\sqrt{\alpha_N}.$$

Fix an $\ell>0$ and consider a realization of $\mathbf{B}_N$ for which $w_N\log^{\ell'}N\to 0$, where $\ell'=\max(\ell+1,4)$ and N large enough so that

$$|w_N|\leq\frac{\Im[z]^3}{4c|z|^2\log^3 N}. \tag{6.17}$$

As opposed to Step 2, the term $\Im[z]\beta_{\mathbf{B}_N}-\Im[w_N]$ can be negative. The idea is to verify that in both scenarios where $\Im[z]\beta_{\mathbf{B}_N}-\Im[w_N]$ is positive and uniformly away from zero, or is not, the conclusion $|\gamma|<1$ holds. First suppose $\beta_{\mathbf{B}_N}\leq\frac{\Im[z]^2}{4c|z|^2\log^3 N}$. Then by the truncation inequalities, we get

$$\alpha_{\mathbf{B}_N}\leq c\Im[z]^{-2}|z|^2\log^3 N\beta_{\mathbf{B}_N}\leq\frac{1}{4}$$

which implies $|\gamma|\leq\frac{1}{2}$. Otherwise we get from (6.16) and (6.17)

$$\begin{aligned}
|\gamma| &\leq \sqrt{\alpha_N}\sqrt{\frac{\Im[e_{\mathbf{B}_N}]\alpha_{\mathbf{B}_N}}{\Im[e_{\mathbf{B}_N}]\alpha_{\mathbf{B}_N}+\Im[z]\beta_{\mathbf{B}_N}-\Im[w_N]}} \\
&\leq \sqrt{\frac{\log N|z|}{\Im[z]^2+\log N|z|}}.
\end{aligned}$$

Therefore, for all N large

$$\begin{aligned}
\log^{\ell}N|e_{\mathbf{B}_N}-e_N| &\leq \frac{(\log^{\ell}N)w_N}{1-\left(\frac{\log^2 N|z|}{\Im[v]^2+\log^2 N|z|}\right)^{\frac{1}{2}}} \\
&\leq 2\Im[z]^{-2}(\Im[z]^2+\log^2 N|z|)(\log^{\ell}N)w_N \\
&\to 0
\end{aligned}$$

as $N\to\infty$, and (6.15) follows. Once more, the multi-dimensional case is much more technical; see [Couillet *et al.*, 2011a] for details.

We finally show

$$m_{\mathbf{B}_N}-m_N\xrightarrow{\text{a.s.}}0 \tag{6.18}$$

as $N\to\infty$. Since $m_{\mathbf{B}_N}=\frac{1}{N}\operatorname{tr}\mathbf{D}_N^{-1}-\tilde{w}_N$ (for some $\tilde{w}_N$ defined similar to w_N), we have

$$m_{\mathbf{B}_N}-m_N=\gamma(e_{\mathbf{B}_N}-e_N)-\tilde{w}_N$$

where now

$$\gamma = \int \frac{c\tau^2}{(1+c\tau e_{\mathbf{B}_N})(1+c\tau e_N)} dF^{\mathbf{T}}(\tau) \frac{1}{N} \operatorname{tr} \mathbf{D}^{-1}\mathbf{R}\mathbf{D}_N^{-1}.$$

From the truncation inequalities, we obtain $|\gamma| \leq c|z|^2\Im[z]^{-4}\log^3 N$. From (6.15) and the fact that $\log^\ell N\tilde{w}_N \xrightarrow{\text{a.s.}} 0$, we finally have (6.18).

In the proof of Theorem 6.17, we will use another technique for this last convergence part, which, instead of controlling precisely the behavior of w_N, consists in proving the convergence on a subset of $\mathbb{C} \setminus \mathbb{R}^+$ that does not meet strong difficulties. Using Vitali's convergence theorem, we then prove the convergence for all $z \in \mathbb{C} \setminus \mathbb{R}^+$. This approach is usually much simpler and is in general preferred.

Returning to the original non-truncated assumptions on $\mathbf{X}$, $\mathbf{T}$, and $\mathbf{R}$, for each of a countably infinite collection of z with positive imaginary part, possessing a limit point with positive imaginary part, we have (6.18). Therefore, by Vitali's convergence theorem, Theorem 3.11, and similar arguments as for the proof of the Marčenko–Pastur law, for any $\varepsilon > 0$, we have exactly that with probability one $m_{\mathbf{B}_N}(z) - m_N(z) \xrightarrow{\text{a.s.}} 0$ uniformly in any region of $\mathbb{C}$ bounded by a contour interior to

$$\mathbb{C} \setminus (\{z : |z| \leq \varepsilon\} \cup \{z = x + iv : x > 0, |v| \leq \varepsilon\}). \tag{6.19}$$

This completes the proof of Theorem 6.1. □

The previous proof is lengthy and technical, when it comes to precisely working out the inequalities based on the truncation steps. Nonetheless, in spite of these difficulties, the line of reasoning in this example can be generalized to more exotic models, which we will introduce also in this section. Moreover, we will briefly introduce alternative techniques of proof, such as the Gaussian method, which will turn out to be based on similar approaches, most particularly for Step 2 and Step 3.

We now prove Theorem 6.2, which we recall provides a deterministic way to recover the unique solution vector $e_{N,1}(z), \ldots, e_{N,K}(z)$ of the implicit Equation (6.6). The arguments of the proof are again very classical and can be reproduced for different random matrix models.

Proof of Theorem 6.2. The convergence of the fixed-point algorithm follows the same line of proof as the uniqueness (Step 2) of Theorem 6.1. For simplicity, we consider also here that $K = 1$. First assume $\Im[z] > 0$. If we consider the difference $e_N^{t+1} - e_N^t$, instead of $e - \underline{e}$, the same development as in the previous proof leads to

$$e_N^{t+1} - e_N^t = \gamma_t(e_N^t - e_N^{t-1}) \tag{6.20}$$

for $t \geq 1$, with γ_t defined by

$$\gamma_t = \int \frac{c\tau^2}{(1+c\tau e_N^{t-1})(1+c\tau e_N^t)} dF^{\mathbf{T}}(\tau) \frac{1}{N} \operatorname{tr} \mathbf{D}_{t-1}^{-1}\mathbf{R}\mathbf{D}_t^{-1}\mathbf{R} \tag{6.21}$$

where $\mathbf{D}_t$ is defined as $\mathbf{D}$ with $e_{\mathbf{B}_N}(z)$ replaced by $e_N^t(z)$. From the Cauchy–Schwarz inequality and the different truncation bounds on the $\mathbf{D}_t$, $\mathbf{R}$, and $\mathbf{T}$ matrices, we have:

$$\gamma_t \leq \frac{|z|^2 c}{\Im[z]^4} \frac{\log^4 N}{N}. \tag{6.22}$$

This entails

$$\left(e_N^{t+1} - e_N^t\right) < \bar{K} \frac{|z|^2 c}{\Im[z]^4} \frac{\log^4 N}{N} \left(e_N^t - e_N^{t-1}\right) \tag{6.23}$$

for some constant $\bar{K}$.

Let $0 < \varepsilon < 1$, and take now a countable set $z_1, z_2, \ldots$ possessing a limit point, such that

$$\bar{K} \frac{|z_k|^2 c}{\Im[z_k]^4} \frac{\log^4 N}{N} < 1 - \varepsilon$$

for all z_k (this is possible by letting $\Im[z_k] > 0$ be large enough). On this countable set, the sequences $e_N^1, e_N^2, \ldots$ are therefore Cauchy sequences on $\mathbb{C}^K$: they all converge. Since the e_N^t are holomorphic functions of z and bounded on every compact set included in $\mathbb{C} \setminus \mathbb{R}^+$, from Vitali's convergence theorem, Theorem 3.11, e_N^t converges on such compact sets.

From the fact that we forced the initialization step to be $e_N^0 = -1/z$, e_N^0 is the Stieltjes transform of a distribution function at point z. It now suffices to verify that, if $e_N^t = e_N^t(z)$ is the Stieltjes transform of a distribution function at point z, then so is e_N^{t+1}. From Theorem 3.2, this requires to ensure that: (i) $z \in \mathbb{C}^+$ and $e_N^t(z) \in \mathbb{C}^+$ implies $e_N^{t+1}(z) \in \mathbb{C}^+$, (ii) $z \in \mathbb{C}^+$ and $ze_N^t(z) \in \mathbb{C}^+$ implies $ze_N^{t+1}(z) \in \mathbb{C}^+$, and (iii) $\lim_{y\to\infty} -ye_N^t(iy) < \infty$ implies that $\lim_{y\to\infty} -ye_N^t(iy) < \infty$. These properties follow directly from the definition of e_N^t. It is not difficult to show also that the limit of e_N^t is a Stieltjes transform and that it is solution to (6.6) when $K = 1$. From the uniqueness of the Stieltjes transform, solution to (6.6) (this follows from the point-wise 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_N^t converges for all j and $z \in \mathbb{C} \setminus \mathbb{R}^+$, if e_N^0 is initialized at a Stieltjes transform. The choice $e_N^0 = -1/z$ follows this rule and the fixed-point algorithm converges to the correct solution.

This concludes the proof of Theorem 6.2. □

From Theorem 6.1, we now wish to provide deterministic equivalents for other functionals of the eigenvalues of $\mathbf{B}_N$ than the Stieltjes transform. In particular, we wish to prove that

$$\int f(x) d(F^{\mathbf{B}_N} - F_N)(x) \xrightarrow{\text{a.s.}} 0$$

for some function f. This is valid for all bounded continuous f from the dominated convergence theorem, which we recall presently.

Theorem 6.3 (Theorem 16.4 in [Billingsley, 1995]). *Let $f_N(x)$ be a sequence of real measurable functions converging point-wise to the measurable function $f(x)$, and such that $|f_N(x)| \leq g(x)$ for some measurable function $g(x)$ with $\int g(x)dx < \infty$. Then, as $N \to \infty$*

$$\int f_N(x)dx \to \int f(x)dx.$$

In particular, if $F_N \Rightarrow F$, the F_N and F being d.f., for any continuous bounded function $h(x)$

$$\int h(x)dF_N(x) \to \int h(x)dF(x).$$

However, for application purposes, such as the calculus of MIMO capacity, see Chapter 13, we would like in particular to take f to be the logarithm function. Proving such convergence results is not at all straightforward since f is here unbounded and because $F^{\mathbf{B}_N}$ may not have bounded support for all large N. This requires additional tools which will be briefly evoked here and which will be introduced in detail in Chapter 7.

We have the following result [Couillet *et al.*, 2011a].

Theorem 6.4. *Let x be some positive real number and f be some continuous function on the positive half-line. Let $\mathbf{B}_N$ be a random Hermitian matrix as defined in Theorem 6.1 with the following additional assumptions.*

1. *There exists $\alpha > 0$ and a sequence r_N, 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 $\lambda_1^{\mathbf{X}} \geq \ldots \geq \lambda_N^{\mathbf{X}}$ denote the ordered eigenvalues of the $N \times N$ matrix $\mathbf{X}$.
2. *Denoting b_N an upper-bound on the spectral norm of the $\mathbf{T}_k$ and $\mathbf{R}_k$, $k \in \{1, \ldots, K\}$, and β some real, such that $\beta > K(b/a)(1+\sqrt{a})^2$ (with a and b such that $a < \liminf_N c_k \leq \limsup_N c_k < b$ for all k), then $a_N = b_N^2 \beta$ satisfies*

$$r_N f(a_N) = o(N). \tag{6.24}$$

Then, for large N, n_k

$$\int f(x)dF^{\mathbf{B}_N}(x) - \int f(x)dF_N(x) \xrightarrow{\text{a.s.}} 0$$

with F_N defined in Theorem 6.1.

In particular, if $f(x) = \log(x)$, under the assumption that (6.24) is fulfilled, we have the following corollary.

Corollary 6.1. *For $\mathbf{A} = 0$, under the conditions of Theorem 6.4 with $f(t) = \log(1 + xt)$, the Shannon transform $\mathcal{V}_{\mathbf{B}_N}$ of $\mathbf{B}_N$, defined for positive x as*

$$\mathcal{V}_{\mathbf{B}_N}(x) = \int_0^\infty \log(1 + x\lambda) dF^{\mathbf{B}_N}(\lambda) = \frac{1}{N} \log\det\left(\mathbf{I}_N + x\mathbf{B}_N\right) \tag{6.25}$$

satisfies

$$\mathcal{V}_{\mathbf{B}_N}(x) - \mathcal{V}_N(x) \xrightarrow{\text{a.s.}} 0$$

where $\mathcal{V}_N(x)$ is defined as

$$\begin{aligned}\mathcal{V}_N(x) = & \frac{1}{N}\log\det\left(\mathbf{I}_N + x\sum_{k=1}^K \mathbf{R}_k \int \frac{\tau_k dF^{\mathbf{T}_k}(\tau_k)}{1 + c_k e_{N,k}(-1/x)\tau_k}\right) \\ & + \sum_{k=1}^K \frac{1}{c_k} \int \log\left(1 + c_k e_{N,k}(-1/x)\tau_k\right) dF^{\mathbf{T}_k}(\tau_k) \\ & + \frac{1}{x} m_N(-1/x) - 1\end{aligned}$$

with m_N and $e_{N,k}$ defined by (6.5) *and* (6.6)*, respectively.*

Again, it is more convenient, for readability and for the sake of practical applications in Chapters 12–15 to remark that

$$\begin{aligned}\mathcal{V}_N(x) = & \frac{1}{N}\log\det\left(\mathbf{I}_N + \sum_{k=1}^K \bar{e}_{N,k}(-1/x)\mathbf{R}_k\right) \\ & + \sum_{k=1}^K \frac{1}{N}\log\det\left(\mathbf{I}_{n_k} + c_k e_{N,k}(-1/x)\mathbf{T}_k\right) \\ & - \frac{1}{x}\sum_{k=1}^K \bar{e}_{N,k}(-1/x) e_{N,k}(-1/x)\end{aligned} \tag{6.26}$$

with $\bar{e}_{N,k}$ defined in (6.7).

Observe that the constraint

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

is in general not strong, as the $F^{\mathbf{T}_k}$ and the $F^{\mathbf{R}_k}$ are already known to form tight sequences as N grows large. Therefore, it is expected that only $o(N)$ largest eigenvalues of the $\mathbf{T}_k$ and $\mathbf{R}_k$ grow large. Here, we impose only a slightly stronger constraint that does not allow for the smallest eigenvalues to exceed a constant α. For practical applications, we will see in Chapter 13 that this constraint is met for all usual channel models, even those exhibiting strong correlation patterns (such as densely packed three-dimensional antenna arrays).

Proof of Theorem 6.4 and Corollary 6.1. The only problem in translating the weak convergence of the distribution function $F^{\mathbf{B}_N} - F_N$ in Theorem 6.1 to the convergence of $\int f d[F^{\mathbf{B}_N} - F_N]$ in Theorem 6.4 is that we must ensure that f behaves nicely. If f were bounded, no restriction in the hypothesis of Theorem 6.1 would be necessary and the weak convergence of $F^{\mathbf{B}_N} - F_N$ to zero gives the result. However, as we are particularly interested in the unbounded, though slowly increasing, logarithm function, this no longer holds. In essence, the proof consists first in taking a realization $\mathbf{B}_1, \mathbf{B}_2, \ldots$ for which the convergence $F^{\mathbf{B}_N} - F_N \Rightarrow 0$ is satisfied. Then we divide the real positive half-line in two sets $[0, d]$ and (d, ∞), with d an upper bound on the $2Kr_N$th largest eigenvalue of $\mathbf{B}_N$ for all large N, which we assume for the moment does exist. For any continuous f, the convergence result is ensured on the compact $[0, d]$; if the largest eigenvalue λ_1 of $\mathbf{B}_N$ is moreover such that $2Kr_N f(\lambda_1) = o(N)$, the integration over (d, ∞) for the measure $dF^{\mathbf{B}_N}$ is of order $o(1)$, which is negligible in the final result for large N. Moreover, since $F_N(d) - F^{\mathbf{B}_N}(d) \to 0$, we also have that, for all large N, $1 - F_N(d) = \int_d^\infty dF_N \leq 2Kr_N/N$, which tends to zero. This finally proves the convergence of $\int f d[F^{\mathbf{B}_N} - F_N]$. The major difficulty here lies in proving that there exists such a bound on the $2Kr_N$th largest eigenvalue of $\mathbf{B}_N$. The essential argument that validates the result is the *asymptotic absence of eigenvalues outside the support of the sample covariance matrix.* This is a result of utmost importance (here, we cannot do without it) which will be presented later in Section 7.1. It can be exactly proved that, almost surely, the largest eigenvalue of $\mathbf{X}_k\mathbf{X}_k^{\mathsf{H}}$ is uniformly bounded by any constant $C > (1 + \sqrt{b})^2$ for all large N, almost surely. In order to use the assumptions of Theorem 6.4, we finally need to introduce the following eigenvalue inequality lemma.

Lemma 6.4 ([Fan, 1951])**.** *Consider a rectangular matrix* $\mathbf{A}$ *and let* $s_i^{\mathbf{A}}$ *denote the* ith *largest singular value of* $\mathbf{A}$*, with* $s_i^{\mathbf{A}} = 0$ *whenever* $i > \operatorname{rank}(A)$*. 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* $\mathbf{A}_1, \ldots, \mathbf{A}_K$*, all of the same size*

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

Since $\lambda_i^{\mathbf{T}_k}$ and $\lambda_i^{\mathbf{R}_k}$ are bounded by α for $i \geq r_N + 1$ and that $\|\mathbf{X}_k\mathbf{X}_k^{\mathsf{H}}\|$ is bounded by C, we have from Lemma 6.4 that the $2Kr_N$th largest eigenvalue of $\mathbf{B}_N$ is uniformly bounded by $CK\alpha^2$. We can then take d any positive real, such that $d > CK\alpha^2$, which is what we needed to show, up to some fine tuning on the final bound.

As for the explicit form of $\int \log(1+xt)dF_N(t)$ given in (6.26), it results from a similar calculus as in Theorem 4.10. Precisely, we expect the Shannon transform to be somehow linked to $\frac{1}{N}\log\det\left(\mathbf{I}_N + \sum_{k=1}^K \bar{e}_{N,k}(-z)\mathbf{R}_k\right)$ and $\frac{1}{N}\log\det\left(\mathbf{I}_{n_k} + c_k e_{N,k}(-z)\mathbf{T}_k\right)$. We then need to find a connection between the derivatives of these functions along z and $\frac{1}{z} - m_N(-z)$, i.e. the derivative of the Shannon transform. Notice that

$$\begin{aligned}\frac{1}{z} - m_N(-z) &= \frac{1}{N}\left((z\mathbf{I}_N)^{-1} - \left(z\left[\mathbf{I}_N + \sum_{k=1}^K \bar{e}_{N,k}\mathbf{R}_k\right]\right)^{-1}\right) \\ &= \sum_{k=1}^K \bar{e}_{N,k}(-z)e_{N,k}(-z).\end{aligned}$$

Since the Shannon transform $\mathcal{V}_N(x)$ satisfies $\mathcal{V}_N(x) = \int_{1/x}^{\infty}[w^{-1} - m_N(-w)]dw$, we need to find an integral form for $\sum_{k=1}^K \bar{e}_{N,k}(-z)e_{N,k}(-z)$. Notice now that

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

and

$$\begin{aligned}\frac{d}{dz}\left(z\sum_{k=1}^K \bar{e}_{N,k}(-z)e_{N,k}(-z)\right) &= \sum_{k=1}^K \bar{e}_{N,k}(-z)e_{N,k}(-z) \\ &\quad - z\sum_{k=1}^K \left(\bar{e}'_{N,k}(-z)e_{N,k}(-z) + \bar{e}_{N,k}(-z)e'_{N,k}(-z)\right).\end{aligned}$$

Combining the last three equations, we have:

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

which after integration leads to

$$\begin{aligned}&\int_z^{\infty}\left(\frac{1}{w} - m_N(-w)\right)dw \\ &= \frac{1}{N}\log\det\left(\mathbf{I}_N + \sum_{k=1}^K \bar{e}_{N,k}(-z)\mathbf{R}_k\right)\end{aligned}$$

$$+\sum_{k=1}^{K}\frac{1}{N}\log\det\left(\mathbf{I}_{n_k}+c_k e_{N,k}(-z)\mathbf{T}_k\right)-z\sum_{k=1}^{K}\bar{e}_{N,k}(-z)e_{N,k}(-z)$$

which is exactly the right-hand side of (6.26) for $z=-1/x$. □

Theorem 6.4 and Corollary 6.1 have obvious direct applications in wireless communications since the Shannon transform $\mathcal{V}_{\mathbf{B}_N}$ defined above is the per-dimension capacity of the multi-dimensional channel, whose model is given by $\sum_{k=1}^{K}\mathbf{R}_k^{\frac{1}{2}}\mathbf{X}_k\mathbf{T}_k^{\frac{1}{2}}$. This is the typical model used for evaluating the rate region of a narrowband multiple antenna multiple access channel. This topic is discussed and extended in Chapter 14, e.g. to the question of finding the transmit covariance matrix that maximizes the deterministic equivalent (hence the asymptotic capacity).

6.2.2 Gaussian method

The second result that we present is very similar in nature to Theorem 6.1 but instead of considering sums of matrices of the type

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

we treat the question of matrices of the type

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

To obtain a deterministic equivalent for this model, the same technique as before could be used. Instead, we develop an alternative method, known as the *Gaussian method*, when the $\mathbf{X}_k$ have Gaussian i.i.d. entries, for which fast convergence rates of the functional of the mean e.s.d. can be proved.

Theorem 6.5 ([Dupuy and Loubaton, 2009]). *Let K be some positive integer. For two positive integers N,n, denote*

$$\mathbf{B}_N=\left(\sum_{k=1}\mathbf{R}_k^{\frac{1}{2}}\mathbf{X}_k\mathbf{T}_k^{\frac{1}{2}}\right)\left(\sum_{k=1}\mathbf{R}_k^{\frac{1}{2}}\mathbf{X}_k\mathbf{T}_k^{\frac{1}{2}}\right)^{\mathsf{H}}$$

where the notations are the same as in Theorem 6.1, with the additional assumptions that $n_1=\ldots=n_K=n$, the random matrix $\mathbf{X}_k\in\mathbb{C}^{N\times n_k}$ has independent Gaussian entries (of zero mean and variance $1/n$) and the spectral norms $\|\mathbf{R}_k\|$ and $\|\mathbf{T}_k\|$ are uniformly bounded with N. Note additionally that, from the unitarily invariance of $\mathbf{X}_k$, $\mathbf{T}_k$ is not restricted to be diagonal. Then, denoting as above $m_{\mathbf{B}_N}$ the Stieltjes transform of $\mathbf{B}_N$, we have

$$N\left(\mathrm{E}[m_{\mathbf{B}_N}(z)]-m_N(z)\right)=O\left(1/N\right)$$

with m_N defined, for $z \in \mathbb{C} \setminus \mathbb{R}^+$, as

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

where $(\bar{e}_{N,1}, \ldots, \bar{e}_{N,K})$ is the unique solution of

$$e_{N,i}(z) = \frac{1}{n} \operatorname{tr} \mathbf{R}_i \left(-z \left[\mathbf{I}_N + \sum_{k=1}^{K} \bar{e}_{N,k}(z) \mathbf{R}_k \right] \right)^{-1}$$

$$\bar{e}_{N,i}(z) = \frac{1}{n} \operatorname{tr} \mathbf{T}_i \left(-z \left[\mathbf{I}_n + \sum_{k=1}^{K} e_{N,k}(z) \mathbf{T}_k \right] \right)^{-1} \tag{6.27}$$

all with positive imaginary part if $z \in \mathbb{C}^+$, negative imaginary part if $z \in \mathbb{C}^-$, and positive if $z < 0$.

Remark 6.4. Note that, due to the Gaussian assumption on the entries of $\mathbf{X}_k$, the convergence result $N\left(\mathrm{E}[m_{\mathbf{B}_N}(z)] - m_N(z)\right) \to 0$ is both (i) looser than the convergence result $m_{\mathbf{B}_N}(z) - m_N(z) \xrightarrow{\text{a.s.}} 0$ of Theorem 6.1 in that it is only shown to converge in expectation, and (ii) stronger in the sense that a convergence rate of $O(1/N)$ of the Stieltjes transform is ensured. Obviously, Theorem 6.1 also implies $\mathrm{E}[m_{\mathbf{B}_N}(z)] - m_N(z) \to 0$. In fact, while this was not explicitly mentioned, a convergence rate of $1/(\log(N)^p)$, for all $p > 0$, is ensured in the proof of Theorem 6.1. The main applicative consequence is that, while the conditions of Theorem 6.1 allow us to deal with instantaneous or quasi-static channel models $\mathbf{H}_k = \mathbf{R}_k^{\frac{1}{2}} \mathbf{X}_k \mathbf{T}_k^{\frac{1}{2}}$, the conditions of Theorem 6.5 are only valid from an ergodic point of view. However, while Theorem 6.1 can only deal with the per-antenna capacity of a quasi-static (or ergodic) MIMO channel, Theorem 6.5 can deal with the total ergodic capacity of MIMO channels, see further Theorem 6.8.

Of course, while this has not been explicitly proved in the literature, it is to be expected that Theorem 6.5 holds also under the looser assumptions and conclusions of Theorem 6.1 and conversely.

The proof of Theorem 6.5 needs the introduction of new tools, gathered together into the so-called *Gaussian method*. Basically, the Gaussian method relies on two main ingredients:

- an integration by parts formula, borrowed from mathematical physics [Glimm and Jaffe, 1981]

  Theorem 6.6. *Let $\mathbf{x} = [x_1, \ldots, x_N]^\mathsf{T} \sim \mathcal{CN}(0, \mathbf{R})$ be a complex Gaussian random vector and $f(\mathbf{x}) \triangleq f(x_1, \ldots, x_N, x_i^*, \ldots, x_N^*)$ be a continuously differentiable functional, the derivatives of which are all polynomially bounded.*

We then have the integration by parts formula

$$\mathrm{E}[x_k f(\mathbf{x})] = \sum_{i=1}^{N} r_{ki} \mathrm{E}\left[\frac{\partial f(\mathbf{x})}{\partial x_i^*}\right]$$

with r_{ki} the entry (k,i) of $\mathbf{R}$.

This relation will be used to *derive directly* the deterministic equivalent, which substitutes to the 'guess-work' step of the proof of Theorem 6.1. Note in particular that it requires us to use all entries of $\mathbf{R}$ here and not simply its eigenvalues. This generalizes the Marčenko–Pastur method that only handled diagonal entries. However, as already mentioned, the introduction of the expectation in front of $x_k f(\mathbf{x})$ cannot be avoided;

- the Nash–Poincaré inequality

Theorem 6.7 ([Pastur, 1999])**.** *Let $\mathbf{x}$ and f be as in Theorem 6.6, and let $\nabla_{\mathbf{z}} f = [\partial f/\partial z_1, \ldots, \partial f/\partial z_N]^{\mathsf{T}}$. Then, we have the following Nash–Poincaré inequality*

$$\mathrm{var}(f(\mathbf{x})) \leq \mathrm{E}\left[\nabla_{\mathbf{x}} f(\mathbf{x})^{\mathsf{T}} \mathbf{R} (\nabla_{\mathbf{x}} f(\mathbf{x}))^*\right] + \mathrm{E}\left[(\nabla_{\mathbf{x}^*} f(\mathbf{x}))^{\mathsf{H}} \mathbf{R} \nabla_{\mathbf{x}^*} f(\mathbf{x})\right].$$

This result will be used to bound the deviations of the random matrices under consideration.

For more details on Gaussian methods, see [Hachem *et al.*, 2008a]. We now give the main steps of the proof of Theorem 6.5.

Proof of Theorem 6.5. We first consider $\mathrm{E}(\mathbf{B}_N - z\mathbf{I}_N)^{-1}$. Noting that $-z(\mathbf{B}_N - z\mathbf{I}_N)^{-1} = \mathbf{I}_N - (\mathbf{B}_N - z\mathbf{I}_N)^{-1}\mathbf{B}_N$, we apply the integration by parts, Theorem 6.6, in order to evaluate the matrix

$$\mathrm{E}\left[(\mathbf{B}_N - z\mathbf{I}_N)^{-1}\mathbf{B}_N\right].$$

To this end, we wish to characterize every entry

$$\begin{aligned}&\mathrm{E}\left[\left((\mathbf{B}_N - z\mathbf{I}_N)^{-1}\mathbf{B}_N\right)_{aa'}\right]\\ &= \sum_{1\leq k,\bar{k}\leq K} \mathrm{E}\left[\left((\mathbf{B}_N - z\mathbf{I}_N)^{-1}\mathbf{R}_k^{\frac{1}{2}}(\mathbf{X}_k\mathbf{T}_k^{\frac{1}{2}}\mathbf{R}_{\bar{k}}^{\frac{1}{2}})(\mathbf{X}_{\bar{k}}\mathbf{T}_{\bar{k}}^{\frac{1}{2}})^{\mathsf{H}}\right)_{aa'}\right].\end{aligned}$$

This is however not so simple and does not lead immediately to a nice form enabling us to use the Gaussian entries of the $\mathbf{X}_k$ as the inputs of Theorem 6.6. Instead, we will consider the multivariate expression

$$\mathrm{E}\left[(\mathbf{B}_N - z\mathbf{I}_N)^{-1}_{ab}(\mathbf{R}_k^{\frac{1}{2}}\mathbf{X}_k\mathbf{T}_k^{\frac{1}{2}})_{cd}(\mathbf{R}_{\bar{k}}^{\frac{1}{2}}\mathbf{X}_{\bar{k}}\mathbf{T}_{\bar{k}}^{\frac{1}{2}})^{\mathsf{H}}_{ea'}\right]$$

for some $k, \bar{k} \in \{1, \ldots, K\}$ and given a, a', b, c, d, e. This enables us to somehow unfold easily the matrix products before we set $b = c$ and $d = e$, and simplify the management of the Gaussian variables. This being said, we take the vector $\mathbf{x}$ of Theorem 6.6 to be the vector whose entries are denoted

$$x_{k,c,d} \triangleq x_{(k-1)Nn+(c-1)N+d} = (\mathbf{R}_k^{\frac{1}{2}}\mathbf{X}_k\mathbf{T}_k^{\frac{1}{2}})_{cd}$$

for all k, c, d. This is therefore a vector of total dimension KNn that collects the entries of all $\mathbf{X}_k$ and accounts for the (Kronecker-type) correlation profile due to $\mathbf{R}_k$ and $\mathbf{T}_k$. The functional $f(\mathbf{x}) = f_{a,b}(\mathbf{x})$ of Theorem 6.6 is taken to be the KNn-dimensional vector $\mathbf{y}^{(a,b)}$ with entry

$$y^{(a,b)}_{\bar{k},a',e} \triangleq y^{(a,b)}_{(\bar{k}-1)Nn+(a'-1)N+e} = (\mathbf{B}_N - z\mathbf{I}_N)^{-1}_{ab}(\mathbf{R}^{\frac{1}{2}}_{\bar{k}}\mathbf{X}_{\bar{k}}\mathbf{T}^{\frac{1}{2}}_{\bar{k}})^{\mathsf{H}}_{ea'}$$

for all $\bar{k}, e, a'$. This expression depends on $\mathbf{x}$ through $(\mathbf{B}_N - z\mathbf{I}_N)^{-1}_{ab}$ and through $x^*_{\bar{k},a',e} = (\mathbf{R}^{\frac{1}{2}}_{\bar{k}}\mathbf{X}_{\bar{k}}\mathbf{T}^{\frac{1}{2}}_{\bar{k}})^{\mathsf{H}}_{ea'}$.

We therefore no longer take $b = c$ or $d = e$ as matrix products would require. This trick allows us to apply seamlessly the integration by parts formula. Applying Theorem 6.6, we have that the entry $(\bar{k}-1)Nn + (a'-1)N + e$ of $\mathrm{E}[x_{k,c,d}f_{a,b}(\mathbf{x})]$, i.e. $\mathrm{E}[x_{k,c,d}y^{(a,b)}_{\bar{k},a',e}]$, is given by:

$$\begin{aligned}&\mathrm{E}[(\mathbf{B}_N - z\mathbf{I}_N)^{-1}_{ab}(\mathbf{R}^{\frac{1}{2}}_{k}\mathbf{X}_{k}\mathbf{T}^{\frac{1}{2}}_{k})_{cd}(\mathbf{R}^{\frac{1}{2}}_{\bar{k}}\mathbf{X}_{\bar{k}}\mathbf{T}^{\frac{1}{2}}_{\bar{k}})^{\mathsf{H}}_{ea'}] \\ &= \sum_{k',c',d'} \mathrm{E}\left[x_{k,c,d}x^*_{k',c',d'}\right]\mathrm{E}\left[\frac{\partial\left((\mathbf{B}_N - z\mathbf{I}_N)^{-1}_{ab}x^*_{\bar{k},a',e}\right)}{\partial x^*_{k',c',d'}}\right]\end{aligned}$$

for all choices of a, b, c, d, e, a'. At this point, we need to proceed to cumbersome calculus, that eventually leads to a nice form when setting $b = c$ and $d = e$.

This gives an expression of $\mathrm{E}\left[\left((\mathbf{B}_N - z\mathbf{I}_N)^{-1}\mathbf{R}^{\frac{1}{2}}_{k}\mathbf{X}_{k}\mathbf{T}^{\frac{1}{2}}_{k}\mathbf{R}^{\frac{1}{2}}_{\bar{k}}\mathbf{X}_{\bar{k}}\mathbf{T}^{\frac{1}{2}}_{\bar{k}}\right)_{aa'}\right]$, which is then summed over all couples $k, \bar{k}$ to obtain

$$\mathrm{E}\left[\left((\mathbf{B}_N - z\mathbf{I}_N)^{-1}\mathbf{B}_N\right)_{aa'}\right] = -z\sum_{k=1}^{K}\bar{e}_{\mathbf{B}_N,k}(z)\mathrm{E}\left[\left((\mathbf{B}_N - z\mathbf{I}_N)^{-1}\mathbf{R}_k\right)_{aa'}\right] + w_{N,aa'}$$

where we defined

$$\begin{aligned}\bar{e}_{\mathbf{B}_N,k}(z) &\triangleq \frac{1}{n}\operatorname{tr}\mathbf{T}_k\left(-z\left[\mathbf{I}_n + \sum_{k=1}^{K}e_{\mathbf{B}_N,k}\mathbf{T}_k\right]\right)^{-1} \\ e_{\mathbf{B}_N,k}(z) &\triangleq \mathrm{E}\left[\frac{1}{N}\operatorname{tr}\mathbf{R}_k(\mathbf{B}_N - z\mathbf{I}_N)^{-1}\right]\end{aligned}$$

and $w_{N,aa'}$ is a residual term that must be shown to be going to zero at a certain rate for increasing N. Using again the formula $-z(\mathbf{B}_N - z\mathbf{I}_N)^{-1} = \mathbf{I}_N - (\mathbf{B}_N - z\mathbf{I}_N)^{-1}\mathbf{B}_N$, this entails

$$\mathrm{E}\left[(\mathbf{B}_N - z\mathbf{I}_N)^{-1}\right] = -\frac{1}{z}\left(\mathbf{I}_N + \sum_{k=1}^{K}\bar{e}_{\mathbf{B}_N,k}(z)\mathbf{R}_k\right)^{-1}[\mathbf{I}_N + \mathbf{W}_N]$$

with $\mathbf{W}_N$ the matrix of (a, a') entry $w_{N,aa'}$. Showing that $\mathbf{W}_N$ is negligible with summable entries as $N \to \infty$ is then solved using the Nash–Poincaré inequality, Theorem 6.7, which again leads to cumbersome but doable calculus.

The second main step consists in considering the system (6.27) (the uniqueness of the solution of which is treated as for Theorem 6.1) and showing that, for any uniformly bounded matrix $\mathbf{E}$

$$\mathrm{E}\left[\operatorname{tr}\mathbf{E}(\mathbf{B}_N - z\mathbf{I}_N)^{-1}\right] = \operatorname{tr}\mathbf{E}(-z[\mathbf{I}_N + \sum_{k=1}^{K}\bar{e}_{N,k}(z)\mathbf{R}_k])^{-1} + O\left(\frac{1}{N}\right)$$

from which $N(\mathrm{E}[e_{\mathbf{B}_N,k}(z)] - e_{N,k}(z)) = O(1/N)$ (for $\mathbf{E} = \mathbf{R}_k$) and finally $N(\mathrm{E}[m_{\mathbf{B}_N}(z)] - m_N(z)) = O(1/N)$ (for $\mathbf{E} = \mathbf{I}_N$). This is performed in a similar way as in the proof for Theorem 6.1, with the additional results coming from the Nash–Poincaré inequality. □

The Gaussian method, while requiring more intensive calculus, allows us to unfold naturally the deterministic equivalent under study for all types of matrix combinations involving Gaussian matrices. It might as well be used as a tool to infer the deterministic equivalent of more involved models for which such deterministic equivalents are not obvious to 'guess' or for which the Marčenko–Pastur method for diagonal matrices cannot be used. For the latest results derived from this technique, refer to, e.g., [Hachem *et al.*, 2008a; Khorunzhy *et al.*, 1996; Pastur, 1999]. It is believed that Haar matrices can be treated using the same tools, to the effort of more involved computations but, to the best of our knowledge, there exists no reference of such a work, yet.

In the same way as we derived the expression of the Shannon transform of the model $\mathbf{B}_N$ of Theorem 6.1 in Corollary 6.1, we have the following result for $\mathbf{B}_N$ in Theorem 6.5.

Theorem 6.8 ([Dupuy and Loubaton, 2010]). *Let $\mathbf{B}_N \in \mathbb{C}^{N\times N}$ be defined as in Theorem 6.5. Then the Shannon transform $\mathcal{V}_{\mathbf{B}_N}$ of $\mathbf{B}_N$ satisfies*

$$N(\mathrm{E}[\mathcal{V}_{\mathbf{B}_N}(x)] - \mathcal{V}_N(x)) = O(1/N)$$

where $\mathcal{V}_N(x)$ is defined, for $x > 0$, as

$$\begin{aligned}\mathcal{V}_N(x) &= \frac{1}{N}\log\det\left(\mathbf{I}_N + \sum_{k=1}^{K}\bar{e}_{N,k}(-1/x)\mathbf{R}_k\right)\\ &+ \frac{1}{N}\log\det\left(\mathbf{I}_n + \sum_{k=1}^{K}e_{N,k}(-1/x)\mathbf{T}_k\right)\\ &- \frac{n}{N}\frac{1}{x}\sum_{k=1}^{K}\bar{e}_{N,k}(-1/x)e_{N,k}(-1/x).\end{aligned} \tag{6.28}$$

Note that the expressions of (6.26) and (6.28) are very similar, apart from the position of a summation symbol.

Both Theorem 6.1 and Theorem 6.5 can then be compiled into an even more general result, as follows. This is however not a corollary of Theorem 6.1 and Theorem 6.5, since the complete proof must be derived from the beginning.

Theorem 6.9. *For $k = 1, \ldots, K$, denote $\mathbf{H}_k \in \mathbb{C}^{N \times n_k}$ the random matrix such that, for a given positive L_k*

$$\mathbf{H}_k = \sum_{l=1}^{L_k} \mathbf{R}_{k,l}^{\frac{1}{2}} \mathbf{X}_{k,l} \mathbf{T}_{k,l}^{\frac{1}{2}}$$

for $\mathbf{R}_{k,l}^{\frac{1}{2}}$ a Hermitian non-negative square root of the Hermitian non-negative $\mathbf{R}_{k,l} \in \mathbb{C}^{N \times N}$, $\mathbf{T}_{k,l}^{\frac{1}{2}}$ a Hermitian non-negative square root of the Hermitian non-negative $\mathbf{T}_{k,l} \in \mathbb{C}^{n_k \times n_k}$ and $\mathbf{X}_{k,l} \in \mathbb{C}^{N \times n_k}$ with Gaussian i.i.d. entries of zero mean and variance $1/n_k$. All $\mathbf{R}_{k,l}$ and $\mathbf{T}_{k,l}$ are uniformly bounded with respect to N, n_k. Denote also for all k, $c_k = N/n_k$.

Call $m_{\mathbf{B}_N}(z)$ the Stieltjes transform of $\mathbf{B}_N = \sum_{k=1}^{K} \mathbf{H}_k \mathbf{H}_k^{\mathsf{H}}$, i.e. for $z \in \mathbb{C} \setminus \mathbb{R}^+$

$$m_{\mathbf{B}_N}(z) = \frac{1}{N} \operatorname{tr} \left(\sum_{k=1}^{K} \mathbf{H}_k \mathbf{H}_k^{\mathsf{H}} - z \mathbf{I}_N \right)^{-1}.$$

We then have

$$N \left(\mathrm{E}[m_{\mathbf{B}_N}(z)] - m_N(z) \right) \to 0$$

where $m_N(z)$ is defined as

$$m_N(z) = \frac{1}{N} \operatorname{tr} \left(-z \left[\sum_{k=1}^{K} \sum_{l=1}^{L_k} e_{N;k,l}(z) \mathbf{R}_{k,l} + \mathbf{I}_N \right] \right)^{-1}$$

and $e_{N;k,l}$ solves the fixed-point equations

$$e_{N;k,l}(z) = \frac{1}{n_k} \operatorname{tr} \mathbf{T}_{k,l} \left(-z \left[\sum_{l'=1}^{L_k} \bar{e}_{N;k,l'}(z) \mathbf{T}_{k,l'} + \mathbf{I}_{n_k} \right] \right)^{-1}$$

$$\bar{e}_{N;k,l}(z) = \frac{1}{n_k} \operatorname{tr} \mathbf{R}_{k,l} \left(-z \left[\sum_{k'=1}^{K} \sum_{l'=1}^{L_{k'}} e_{N;k',l'}(z) \mathbf{R}_{k',l'} + \mathbf{I}_N \right] \right)^{-1}.$$

We also have that the Shannon transform $\mathcal{V}_{\mathbf{B}_N}(x)$ of $\mathbf{B}_N$ satisfies

$$N \left(\mathrm{E}[\mathcal{V}_{\mathbf{B}_N}(x)] - \mathcal{V}_N(x) \right) \to 0$$

where

$$\mathcal{V}_N(x) = \frac{1}{N} \log\det \left(\sum_{k=1}^{K} \sum_{l=1}^{L_k} e_{N;k,l}(-1/x) \mathbf{R}_{k,l} + \mathbf{I}_N \right) + \sum_{k=1}^{K} \frac{1}{N} \log\det \left(\sum_{l=1}^{L_k} \bar{e}_{N;k,l}(-1/x) \mathbf{T}_{k,l} + \mathbf{I}_{n_k} \right) - \frac{1}{x} \sum_{k=1}^{K} \frac{n_k}{N} \sum_{l=1}^{L_k} e_{N;k,l}(-1/x) \bar{e}_{N;k,l}(-1/x).$$

For practical applications, this formula provides the whole picture for the *ergodic* rate region of large MIMO multiple access channels, with K multiple antenna users, user k being equipped with n_k antennas, when the different channels into consideration are frequency selective with L_k taps for user k, slow fading in time, and for each tap modeled as Kronecker with receive and transmit correlation $\mathbf{R}_{k,l}$ and $\mathbf{T}_{k,l}$, respectively.

We now move to another type of deterministic equivalents, when the entries of the matrix $\mathbf{X}$ are not necessarily of zero mean and have possibly different variances.

6.2.3 Information plus noise models

In Section 3.2, we introduced an important limiting Stieltjes transform result, Theorem 3.14, for the Gram matrix of a random i.i.d. matrix $\mathbf{X} \in \mathbb{C}^{N\times n}$ with a variance profile $\{\sigma_{ij}^2/n\}$, $1 \leq i \leq N$ and $1 \leq j \leq n$. One hypothesis of Girkos's law is that the profile $\{\sigma_{ij}\}$ converges to a density $\sigma(x,y)$ in the sense that

$$\sigma_{ij} - \int_{\frac{i-1}{N}}^{\frac{i}{N}} \int_{\frac{j-1}{n}}^{\frac{j}{n}} \sigma(x,y)dxdy \to 0.$$

It will turn out in practical applications that such an assumption is in general unusable. Typically, suppose that σ_{ij} is the channel fading between antenna i and antenna j, respectively, at the transmitter and receiver of a multiple antenna channel. As one grows N and n simultaneously, there is no reason for the σ_{ij} to converge in any sense to a density $\sigma(x,y)$. In the following, we therefore rewrite Theorem 3.14 in terms of deterministic equivalents without the need for any assumption of convergence. This result is in fact a corollary of the very general Theorem 6.14, presented later in this section, although the deterministic equivalent is written in a slightly different form. A sketch of the proof using the Bai and Silverstein approach is also provided.

Theorem 6.10. *Let $\mathbf{X}_N \in \mathbb{C}^{N\times n}$ have independent entries x_{ij} with zero mean, variance σ_{ij}^2/n and $4+\varepsilon$ moment of order $O(1/N^{2+\varepsilon/2})$, for some ε. Assume that the σ_{ij} are deterministic and uniformly bounded, over n, N. Then, as N, n grow large with ratio $c_n \triangleq N/n$ such that $0 < \liminf_n c_n \leq \limsup_n c_n < \infty$, the e.s.d. $F^{\mathbf{B}_N}$ of $\mathbf{B}_N = \mathbf{X}_N\mathbf{X}_N^{\mathsf{H}}$ satisfies*

$$F^{\mathbf{B}_N} - F_N \Rightarrow 0$$

almost surely, where F_N is the distribution function of Stieltjes transform $m_N(z)$, $z \in \mathbb{C} \setminus \mathbb{R}^+$, given by:

$$m_N(z) = \frac{1}{N}\sum_{k=1}^{N} \frac{1}{\frac{1}{n}\sum_{i=1}^{n} \sigma_{ki}^2 \frac{1}{1+e_{N,i}(z)} - z}$$

where $e_{N,1}(z), \ldots, e_{N,n}(z)$ *form the unique solution of*

$$e_{N,j}(z) = \frac{1}{n} \sum_{k=1}^{N} \frac{\sigma_{kj}^2}{\frac{1}{n} \sum_{i=1}^{n} \sigma_{ki}^2 \frac{1}{1+e_{N,i}(z)} - z} \tag{6.29}$$

such that all $e_{N,j}(z)$ *are Stieltjes transforms of a distribution function.*

The reason why point-wise uniqueness of the $e_{N,j}(z)$ is not provided here is due to the approach of the proof of uniqueness followed by Hachem *et al.* [Hachem *et al.*, 2007] which is a functional proof of uniqueness of the Stieltjes transforms that the applications $z \mapsto e_{N,i}(z)$ define. This does not mean that point-wise uniqueness does not hold but this is as far as this theorem goes.

Theorem 6.10 can then be written is a more compact and symmetric form by rewriting $e_{N,j}(z)$ in (6.29) as

$$\begin{aligned} e_{N,j}(z) &= -\frac{1}{z}\frac{1}{n} \sum_{k=1}^{N} \frac{\sigma_{kj}^2}{1+\bar{e}_{N,k}(z)} \\ \bar{e}_{N,k}(z) &= -\frac{1}{z}\frac{1}{n} \sum_{i=1}^{n} \frac{\sigma_{ki}^2}{1+e_{N,i}(z)}. \end{aligned} \tag{6.30}$$

In this case, $m_N(z)$ is simply

$$m_N(z) = -\frac{1}{z}\frac{1}{N} \sum_{k=1}^{N} \frac{1}{1+\bar{e}_{N,k}(z)}.$$

Note that this version of Girko's law, Theorem 3.14, is both more general in the assumptions made, and more explicit. We readily see in this result that fixed-point algorithms, if they converge at all, allow us to recover the $2n$ coupled Equations (6.30), from which $m_N(z)$ is then explicit.

For the sake of understanding and to further justify the strength of the techniques introduced so far, we provide hereafter the first steps of the proof using the Bai and Silverstein technique. A complete proof can be found as a particular case of [Hachem *et al.*, 2007; Wagner *et al.*, 2011].

Proof. Instead of studying $m_N(z)$, let us consider the more general $e_{\mathbf{A}_N}(z)$, a deterministic equivalent for

$$\frac{1}{N} \operatorname{tr} \mathbf{A}_N \left(\mathbf{X}_N \mathbf{X}_N^{\mathsf{H}} - z\mathbf{I}_N\right)^{-1}.$$

Using Bai and Silverstein approach, we introduce $\mathbf{F} \in \mathbb{C}^{N \times N}$ some matrix yet to be defined, and compute

$$e_{\mathbf{A}_N}(z) = \frac{1}{N} \operatorname{tr} \mathbf{A}_N \left(\mathbf{F} - z\mathbf{I}_N\right)^{-1}.$$

Using the resolvent identity, Lemma 6.1, and writing $\mathbf{X}_N\mathbf{X}_N^{\mathsf{H}} = \sum_{i=1}^n \mathbf{x}_i\mathbf{x}_i^{\mathsf{H}}$, we have:

$$\begin{aligned}
&\frac{1}{N}\operatorname{tr}\mathbf{A}_N\left(\mathbf{X}_N\mathbf{X}_N^{\mathsf{H}} - z\mathbf{I}_N\right)^{-1} - \frac{1}{N}\operatorname{tr}\mathbf{A}_N\left(\mathbf{F} - z\mathbf{I}_N\right)^{-1}\\
&= \frac{1}{N}\operatorname{tr}\mathbf{A}_N\left(\mathbf{X}_N\mathbf{X}_N^{\mathsf{H}} - z\mathbf{I}_N\right)^{-1}\mathbf{F}\left(\mathbf{F} - z\mathbf{I}_N\right)^{-1}\\
&- \frac{1}{N}\sum_{i=1}^n \operatorname{tr}\mathbf{A}_N\left(\mathbf{X}_N\mathbf{X}_N^{\mathsf{H}} - z\mathbf{I}_N\right)^{-1}\mathbf{x}_i\mathbf{x}_i^{\mathsf{H}}\left(\mathbf{F} - z\mathbf{I}_N\right)^{-1}
\end{aligned}$$

from which we then express the second term on the right-hand side under the form of sums for $i \in \{1, \ldots, N\}$ of $\mathbf{x}_i^{\mathsf{H}}\left(\mathbf{F} - z\mathbf{I}_N\right)^{-1}\mathbf{A}_N\left(\mathbf{X}_N\mathbf{X}_N^{\mathsf{H}} - z\mathbf{I}_N\right)^{-1}\mathbf{x}_i$ and we use Lemma 6.2 on the matrix $\left(\mathbf{X}_N\mathbf{X}_N^{\mathsf{H}} - z\mathbf{I}_N\right)^{-1}$ to obtain

$$\begin{aligned}
&\frac{1}{N}\operatorname{tr}\mathbf{A}_N\left(\mathbf{X}_N\mathbf{X}_N^{\mathsf{H}} - z\mathbf{I}_N\right)^{-1} - \frac{1}{N}\operatorname{tr}\mathbf{A}_N\left(\mathbf{F} - z\mathbf{I}_N\right)^{-1}\\
&= \frac{1}{N}\operatorname{tr}\mathbf{A}_N\left(\mathbf{X}_N\mathbf{X}_N^{\mathsf{H}} - z\mathbf{I}_N\right)^{-1}\mathbf{F}\left(\mathbf{F} - z\mathbf{I}_N\right)^{-1}\\
&- \frac{1}{N}\sum_{i=1}^n \frac{\mathbf{x}_i^{\mathsf{H}}\left(\mathbf{F} - z\mathbf{I}_N\right)^{-1}\mathbf{A}_N\left(\mathbf{X}_{(i)}\mathbf{X}_{(i)}^{\mathsf{H}} - z\mathbf{I}_N\right)^{-1}\mathbf{x}_i}{1 + \mathbf{x}_i^{\mathsf{H}}\left(\mathbf{X}_{(i)}\mathbf{X}_{(i)}^{\mathsf{H}} - z\mathbf{I}_N\right)^{-1}\mathbf{x}_i}
\end{aligned} \tag{6.31}$$

with $\mathbf{X}_{(i)} = [\mathbf{x}_1, \ldots, \mathbf{x}_{i-1}, \mathbf{x}_{i+1}, \ldots, \mathbf{x}_n]$.

Under this form, $\mathbf{x}_i$ and $\left(\mathbf{X}_{(i)}\mathbf{X}_{(i)}^{\mathsf{H}} - z\mathbf{I}_N\right)^{-1}$ have independent entries. However, $\mathbf{x}_i$ does not have identically distributed entries, so that Theorem 3.4 cannot be straightforwardly applied. We therefore define $\mathbf{y}_i \in \mathbb{C}^N$ as

$$\mathbf{x}_i = \mathbf{\Sigma}_i\mathbf{y}_i$$

with $\mathbf{\Sigma}_i \in \mathbb{C}^{N\times N}$ a diagonal matrix with kth diagonal entry equal to σ_{ki}, and $\mathbf{y}_i$ has identically distributed entries of zero mean and variance $1/n$. Replacing all occurrences of $\mathbf{x}_i$ in (6.31) by $\mathbf{\Sigma}_i\mathbf{y}_i$, we have:

$$\begin{aligned}
&\frac{1}{N}\operatorname{tr}\mathbf{A}_N\left(\mathbf{X}_N\mathbf{X}_N^{\mathsf{H}} - z\mathbf{I}_N\right)^{-1} - \frac{1}{N}\operatorname{tr}\mathbf{A}_N\left(\mathbf{F} - z\mathbf{I}_N\right)^{-1}\\
&= \frac{1}{N}\operatorname{tr}\mathbf{A}_N\left(\mathbf{X}_N\mathbf{X}_N^{\mathsf{H}} - z\mathbf{I}_N\right)^{-1}\mathbf{F}\left(\mathbf{F} - z\mathbf{I}_N\right)^{-1}\\
&- \frac{1}{N}\sum_{i=1}^n \frac{\mathbf{y}_i^{\mathsf{H}}\mathbf{\Sigma}_i\left(\mathbf{F} - z\mathbf{I}_N\right)^{-1}\mathbf{A}_N\left(\mathbf{X}_{(i)}\mathbf{X}_{(i)}^{\mathsf{H}} - z\mathbf{I}_N\right)^{-1}\mathbf{\Sigma}_i\mathbf{y}_i}{1 + \mathbf{y}_i^{\mathsf{H}}\mathbf{\Sigma}_i\left(\mathbf{X}_{(i)}\mathbf{X}_{(i)}^{\mathsf{H}} - z\mathbf{I}_N\right)^{-1}\mathbf{\Sigma}_i\mathbf{y}_i}.
\end{aligned} \tag{6.32}$$

Applying the trace lemma, Theorem 3.4, the quadratic terms of the form $\mathbf{y}_i^{\mathsf{H}}\mathbf{Y}\mathbf{y}_i$ are close to $\frac{1}{n}\operatorname{tr}\mathbf{Y}$. Therefore, in order for (6.32) to converge to zero, $\mathbf{F}$ *ought to* take the form

$$\mathbf{F} = \frac{1}{n}\sum_{i=1}^n \frac{1}{1 + e_{\mathbf{B}_N,i}(z)}\mathbf{\Sigma}_i^2$$

with

$$e_{\mathbf{B}_N,i}(z) = \frac{1}{n} \operatorname{tr} \mathbf{\Sigma}_i^2 \left(\mathbf{X}_N \mathbf{X}_N^{\mathsf{H}} - z\mathbf{I}_N\right)^{-1}.$$

We therefore infer that $e_{N,i}(z)$ takes the form

$$e_{N,i}(z) = \frac{1}{n} \sum_{k=1}^{N} \frac{\sigma_{ki}^2}{\frac{1}{n} \sum_{i=1}^{n} \sigma_{ki}^2 \frac{1}{1+e_{N,i}(z)} - z}$$

by setting $\mathbf{A}_N = \mathbf{\Sigma}_i^2$.

From this point on, the result unfolds by showing the almost sure convergence towards zero of the difference $e_{N,i}(z) - \frac{1}{n} \operatorname{tr} \mathbf{\Sigma}_i^2 \left(\mathbf{X}_N \mathbf{X}_N^{\mathsf{H}} - z\mathbf{I}_N\right)^{-1}$ and the functional uniqueness of the implicit equation for the $e_{N,i}(z)$. □

The symmetric expressions (6.30) make it easy to derive also a deterministic equivalent of the Shannon transform.

Theorem 6.11. *Let $\mathbf{B}_N$ be defined as in Theorem 6.10 and let $x > 0$. Then, as N, n grow large with uniformly bounded ratio $c_n = N/n$, the Shannon transform $\mathcal{V}_{\mathbf{B}_N}(x)$ of $\mathbf{B}_N$, defined as*

$$\mathcal{V}_{\mathbf{B}_N}(x) \triangleq \frac{1}{N} \log\det\left(\mathbf{I}_N + x\mathbf{B}_N\right)$$

satisfies

$$\mathrm{E}[\mathcal{V}_{\mathbf{B}_N}(x)] - \mathcal{V}_N(x) \to 0$$

where $\mathcal{V}_N(x)$ is given by:

$$\begin{aligned}\mathcal{V}_N(x) &= \frac{1}{N} \sum_{k=1}^{N} \log\left(1 + \bar{e}_{N,k}(-\frac{1}{x})\right) + \frac{1}{N} \sum_{i=1}^{n} \log\left(1 + e_{N,i}(-\frac{1}{x})\right) \\ &- \frac{x}{nN} \sum_{\substack{1\leq k\leq N \\ 1\leq i\leq n}} \frac{\sigma_{ki}^2}{\left(1 + \bar{e}_{N,k}(-\frac{1}{x})\right)\left(1 + e_{N,i}(-\frac{1}{x})\right)}.\end{aligned}$$

It is worth pointing out here that the Shannon transform convergence result is only stated in the mean sense and not, as was the case in Theorem 6.4, in the almost sure sense. Remember indeed that the convergence result of Theorem 6.4 depends strongly on the fact that the empirical matrix $\mathbf{B}_N$ can be proved to have bounded spectral norm for all large N, almost surely. This is a consequence of spectral norm inequalities and of Theorem 7.1. However, it is not known whether Theorem 7.1 holds true for matrices with a variance profile and the derivation of Theorem 6.4 can therefore not be reproduced straightforwardly.

It is in fact not difficult to show the convergence of the Shannon transform in the mean via a simple dominated convergence argument. Indeed, remembering the Shannon transform definition, Definition 3.2, we have:

$$\mathrm{E}[\mathcal{V}_{\mathbf{B}_N}(x)] - \mathcal{V}_N(x) = \int_{\frac{1}{x}}^{\infty} \left(\frac{1}{t} - \mathrm{E}[m_{\mathbf{B}_N}(-t)]\right) dt - \int_{\frac{1}{x}}^{\infty} \left(\frac{1}{t} - m_N(-t)\right) dt \tag{6.33}$$

for which we in particular have

$$\begin{aligned}
&\left|\left(\frac{1}{t} - \mathrm{E}[m_{\mathbf{B}_N}(-t)]\right) - \left(\frac{1}{t} - m_N(-t)\right)\right| \\
&\leq \left|\frac{1}{t} - \mathrm{E}[m_{\mathbf{B}_N}(-t)]\right| + \left|\frac{1}{t} - m_N(-t)\right| \\
&= \left|\int \left(\frac{1}{t} - \frac{1}{\lambda + t}\right) \mathrm{E}[dF^{\mathbf{B}_N}(\lambda)]\right| + \left|\int \left(\frac{1}{t} - \frac{1}{\lambda + t}\right) dF_N(\lambda)\right| \\
&\leq \frac{1}{t^2} \int \lambda \mathrm{E}[dF^{\mathbf{B}_N}(\lambda)] + \frac{1}{t^2} \int \lambda dF_N(\lambda).
\end{aligned}$$

It is now easy to prove from standard expectation calculus that both integrals above are upper-bound by $\limsup_N \sup_i \|\mathbf{R}_i\| < \infty$. Writing Equation (6.33) under the form of a single integral, we have that the integrand tends to zero as $N \to \infty$ and is summable over the integration parameter t. Therefore, from the dominated convergence theorem, Theorem 6.3, $\mathrm{E}[\mathcal{V}_{\mathbf{B}_N}(x)] - \mathcal{V}_N(x) \to 0$.

Note now that, in the proof of Theorem 6.10, there is no actual need for the matrices $\mathbf{\Sigma}_k$ to be diagonal. Also, there is no huge difficulty added by considering the matrix $\mathbf{X}_N\mathbf{X}_N^{\mathsf{H}} + \mathbf{A}_N$, instead of $\mathbf{X}_N\mathbf{X}_N^{\mathsf{H}}$ for any deterministic $\mathbf{A}_N$. As such, Theorem 6.10 can be further generalized as follows.

Theorem 6.12 ([Wagner *et al.*, 2011])**.** *Let $\mathbf{X}_N \in \mathbb{C}^{N\times n}$ have independent columns $\mathbf{x}_i = \mathbf{H}_i\mathbf{y}_i$, where $\mathbf{y}_i \in \mathbb{C}^{N_i}$ has i.i.d. entries of zero mean, variance $1/n$, and $4+\varepsilon$ moment of order $O(1/n^{2+\varepsilon/2})$, and $\mathbf{H}_i \in \mathbb{C}^{N\times N_i}$ are such that $\mathbf{R}_i \triangleq \mathbf{H}_i\mathbf{H}_i^{\mathsf{H}}$ has uniformly bounded spectral norm over n, N. Let also $\mathbf{A}_N \in \mathbb{C}^{N\times N}$ be Hermitian non-negative and denote $\mathbf{B}_N = \mathbf{X}_N\mathbf{X}_N^{\mathsf{H}} + \mathbf{A}_N$. Then, as N, $N_1, \ldots, N_n$, and n grow large with ratios $c_i \triangleq N_i/n$ and $c_0 \triangleq N/n$ satisfying $0 < \liminf_n c_i \leq \limsup_n c_i < \infty$ for $0 \leq i \leq n$, we have that, for all non-negative Hermitian matrix $\mathbf{C}_N \in \mathbb{C}^{N\times N}$ with uniformly bounded spectral norm*

$$\frac{1}{n}\operatorname{tr}\mathbf{C}_N\left(\mathbf{B}_N - z\mathbf{I}_N\right)^{-1} - \frac{1}{n}\operatorname{tr}\mathbf{C}_N\left(\frac{1}{n}\sum_{i=1}^{n}\frac{1}{1+e_{N,i}(z)}\mathbf{R}_i + \mathbf{A}_N - z\mathbf{I}_N\right)^{-1} \xrightarrow{\text{a.s.}} 0$$

where $e_{N,1}(z), \ldots, e_{N,n}(z)$ form the unique functional solution of

$$e_{N,j}(z) = \frac{1}{n}\operatorname{tr}\mathbf{R}_j\left(\frac{1}{n}\sum_{i=1}^{n}\frac{1}{1+e_{N,i}(z)}\mathbf{R}_i + \mathbf{A}_N - z\mathbf{I}_N\right)^{-1} \tag{6.34}$$

such that all $e_{N,j}(z)$ are Stieltjes transforms of a non-negative finite measure on $\mathbb{R}^+$. Moreover, $(e_{N,1}(z), \ldots, e_{N,n}(z))$ is given by $e_{N,i}(z) = lim_{k\to\infty} e_{N,i}^{(k)}(z)$, where

$e_{N,i}^{(0)} = -1/z$ *and, for* $k \geq 0$

$$e_{N,j}^{(k+1)}(z) = \frac{1}{n} \operatorname{tr} \mathbf{R}_j \left(\frac{1}{n} \sum_{i=1}^{n} \frac{1}{1+e_{N,i}^{(k)}(z)} \mathbf{R}_i + \mathbf{A}_N - z\mathbf{I}_N \right)^{-1}.$$

Also, for $x > 0$*, the Shannon transform* $\mathcal{V}_{\mathbf{B}_N}(x)$ *of* $\mathbf{B}_N$*, defined as*

$$\mathcal{V}_{\mathbf{B}_N}(x) \triangleq \frac{1}{N} \log\det\left(\mathbf{I}_N + x\mathbf{B}_N\right)$$

satisfies

$$\mathrm{E}[\mathcal{V}_{\mathbf{B}_N}(x)] - \mathcal{V}_N(x) \to 0$$

where $\mathcal{V}_N(x)$ *is given by:*

$$\mathcal{V}_N(x) = \frac{1}{N} \log\det\left(\mathbf{I}_N + x\left[\frac{1}{n}\sum_{i=1}^{n} \frac{1}{1+e_{N,i}(-\frac{1}{x})}\mathbf{R}_i + \mathbf{A}_N \right] \right)$$
$$+ \frac{1}{N}\sum_{i=1}^{n} \log\left(1 + e_{N,i}(-\frac{1}{x})\right) - \frac{1}{N}\sum_{i=1}^{n} \frac{e_{N,i}(-\frac{1}{x})}{1+e_{N,i}(-\frac{1}{x})}.$$

Remark 6.5. Consider the identically distributed entries $\mathbf{x}_1, \ldots, \mathbf{x}_n$ in Theorem 6.12, and take $n_1, \ldots, n_K$ to be K integers such that $\sum_i n_i = n$. Define $\tilde{\mathbf{R}}_1, \ldots, \tilde{\mathbf{R}}_K \in \mathbb{C}^{N\times N}$ to be K non-negative definite matrices with uniformly bounded spectral norm and $\mathbf{T}_1 \in \mathbb{C}^{n_1\times n_1}, \ldots, \mathbf{T}_K \in \mathbb{C}^{n_K\times n_K}$ to be K diagonal matrices with positive entries, $\mathbf{T}_k = \operatorname{diag}(t_{k1}, \ldots, t_{kn_k})$. Denote $\mathbf{R}_k = \tilde{\mathbf{R}}_j t_{ji}$, $k \in \{1, \ldots, n\}$, with j the smallest integer such that $k - (n_1 + \ldots + n_{j-1}) > 0$, $n_0 = 0$, and $i = k - (n_1 + \ldots + n_{j-1})$. Under these conditions and notations, up to some hypothesis restrictions, Theorem 6.12 with $\mathbf{H}_i = \tilde{\mathbf{R}}_i^{\frac{1}{2}}$ also generalizes Theorem 6.1 applied to the sum of K Gram matrices with left correlation matrix $\tilde{\mathbf{R}}_1, \ldots, \tilde{\mathbf{R}}_K$ and right correlation matrices $\mathbf{T}_1, \ldots, \mathbf{T}_K$.

From Theorem 6.12, taking $\mathbf{A}_N = 0$, we also immediately have that the distribution function F_N with Stieltjes transform

$$m_N(z) = \frac{1}{N} \operatorname{tr} \left(\frac{1}{n}\sum_{i=1}^{n} \frac{1}{1+e_{N,i}(z)}\mathbf{R}_i - z\mathbf{I}_N \right)^{-1} \tag{6.35}$$

where

$$e_{N,j}(z) = \frac{1}{n} \operatorname{tr} \mathbf{R}_j \left(\frac{1}{n}\sum_{i=1}^{n} \frac{1}{1+e_{N,i}(z)}\mathbf{R}_i - z\mathbf{I}_N \right)^{-1} \tag{6.36}$$

is a deterministic equivalent for $F^{\mathbf{X}_N\mathbf{X}_N^{\mathsf{H}}}$. An interesting result with application in low complex filter design, see Section 13.6 of Chapter 13, is the description in closed-form of the successive moments of the distribution function F_N.

Theorem 6.13 ([Hoydis *et al.*, 2011c]). *Let F_N be the d.f. associated with the Stieltjes transform $m_N(z)$ defined by (6.35) with $e_{N,i}(z)$ given by (6.36). Further denote $M_{N,0}, M_{N,1}, \ldots$ the successive moments of F_N*

$$M_{N,p} = \int x^p dF_N(x).$$

Then, these moments are explicitly given by:

$$M_{N,p} = \frac{(-1)^p}{p!} \frac{1}{N} \operatorname{tr} \mathbf{T}_p$$

with $\mathbf{T}_0, \mathbf{T}_1, \ldots$ defined iteratively from the following set of recursive equations initialized with $\mathbf{T}_0 = \mathbf{I}_N$, $f_{k,0} = -1$ and $\delta_{k,0} = \frac{1}{n} \operatorname{tr} \mathbf{R}_k$ for $k \in \{1, \ldots, n\}$

$$\begin{aligned}
\mathbf{T}_{p+1} &= \sum_{i=0}^{p} \sum_{j=0}^{i} \binom{p}{i} \binom{i}{j} \mathbf{T}_{p-i} \mathbf{Q}_{i-j+1} \mathbf{T}_j \\
\mathbf{Q}_{p+1} &= \frac{p+1}{n} \sum_{k=1}^{n} f_{k,p} \mathbf{R}_k \\
f_{k,p+1} &= \sum_{i=0}^{p} \sum_{j=0}^{i} \binom{p}{i} \binom{i}{j} (p-i+1) f_{k,j} f_{k,i-j} \delta_{k,p-i} \\
\delta_{k,p+1} &= \frac{1}{n} \operatorname{tr} \mathbf{R}_k \mathbf{T}_{p+1}.
\end{aligned}$$

Moreover, with $\mathbf{B}_N = \mathbf{X}_N \mathbf{X}_N^{\mathsf{H}}$, $\mathbf{X}_N$ being defined in Theorem 6.12, we have for all integer p

$$\int x^p \mathrm{E}[dF^{\mathbf{B}_N}(x)] - M_{N,p} \to 0$$

as $N, n \to \infty$.

Note that a similar result was established from a combinatorics approach in [Li *et al.*, 2004] which took the form of involved sums over non-crossing partitions, when all $\mathbf{R}_k$ matrices are Toeplitz and of Wiener class [Gray, 2006]. The proof of the almost sure convergence of $\int x^p dF^{\mathbf{B}_N}(x)$ to $M_{N,p}$, claimed in [Li *et al.*, 2004], would require proving that the support $\mathbf{B}_N$ is almost surely uniformly bounded from above for all large N. However, this fact is unknown to this day so that convergence in the mean can be ensured, while almost sure convergence can only be conjectured. It holds true in particular when the family $\{\mathbf{R}_1, \ldots, \mathbf{R}_n\}$ is extracted from a finite set.

Proof. Note that F_N is necessarily compactly supported as the $\|\mathbf{R}_i\|$ are uniformly bounded and that the $e_{N,i}(z)$ are non-negative for $z < 0$. Reminding then that the Stieltjes transform m_N of F_N can be written in that case under the form of a moment generating function by (3.6), the expression of the successive moments unfolds from successive differentiations of $-z^{-1} m_N(-z^{-1})$, taken in

$z = 0$. The convergence of the difference of moments is then a direct consequence of the dominated convergence theorem, Theorem 6.3. □

Another generalization of Theorem 6.10 is found in [Hachem *et al.*, 2007], where $\mathbf{X}_N$ still has a variance profile but has non-zero mean. The result in the latter is more involved and expresses as follows.

Theorem 6.14. *Let $\mathbf{X}_N = \mathbf{Y}_N + \mathbf{A}_N \in \mathbb{C}^{N\times n}$ be a random matrix where $\mathbf{Y}_N$ has independent entries y_{ij} with zero mean, variance σ_{ij}^2/n and finite $4+\varepsilon$ moment of order $O(1/N^{2+\varepsilon/2})$, and $\mathbf{A}_N$ is a deterministic matrix. Denote $\mathbf{\Sigma}_j \in \mathbb{C}^{N\times N}$ the diagonal matrix with ith diagonal entry σ_{ij} and $\bar{\mathbf{\Sigma}}_i \in \mathbb{C}^{n\times n}$ the diagonal matrix with jth diagonal entry σ_{ij}. Suppose moreover that the columns of $\mathbf{A}_N$ have uniformly bounded Euclidean norm and that the σ_{ij} are uniformly bounded, with respect to N and n. Then, as N, n grow large with ratio $c_N = N/n$, such that $0 < \liminf_N c_N \leq \limsup_N c_N < \infty$, the e.s.d. $F^{\mathbf{B}_N}$ of $\mathbf{B}_N \triangleq \mathbf{X}_N\mathbf{X}_N^{\mathsf{H}}$ satisfies*

$$F^{\mathbf{B}_N} - F_N \Rightarrow 0$$

almost surely, with F_N the distribution function with Stieltjes transform $m_N(z)$, $z \in \mathbb{C}\setminus\mathbb{R}^+$, given by:

$$m_N(z) = \frac{1}{N}\operatorname{tr}\left(\mathbf{\Psi}^{-1} - z\mathbf{A}_N\bar{\mathbf{\Psi}}\mathbf{A}_N^{\mathsf{T}}\right)^{-1}$$

where $\mathbf{\Psi} \in \mathbb{C}^{N\times N}$ is diagonal with ith entry $\psi_i(z)$, $\bar{\mathbf{\Psi}} \in \mathbb{C}^{n\times n}$ is diagonal with jth entry $\bar{\psi}_j(z)$, with $\psi_i(z)$ and $\bar{\psi}_j(z)$, $1 \leq i \leq N$, $1 \leq j \leq n$, the unique solutions of

$$\psi_i(z) = -\frac{1}{z}\left[1 + \frac{1}{n}\operatorname{tr}\bar{\mathbf{\Sigma}}_i^2\left(\bar{\mathbf{\Psi}}^{-1} - z\mathbf{A}_N^{\mathsf{T}}\mathbf{\Psi}\mathbf{A}_N\right)^{-1}\right]^{-1}$$

$$\bar{\psi}_j(z) = -\frac{1}{z}\left[1 + \frac{1}{n}\operatorname{tr}\mathbf{\Sigma}_j^2\left(\mathbf{\Psi}^{-1} - z\mathbf{A}_N\bar{\mathbf{\Psi}}\mathbf{A}_N^{\mathsf{T}}\right)^{-1}\right]^{-1}$$

which are Stieltjes transforms of distribution functions.

Besides, for $x = -\frac{1}{z} > 0$, let $\mathcal{V}_{\mathbf{B}_N}(x) = \frac{1}{N}\log\det\left(\mathbf{I}_N + x\mathbf{X}_N\mathbf{X}_N^{\mathsf{H}}\right)$ be the Shannon transform of $\mathbf{B}_N$. Then

$$\mathrm{E}[\mathcal{V}_{\mathbf{B}_N}(x)] - \mathcal{V}_N(x) \to 0$$

as N, n grow large, where $\mathcal{V}_N(x)$ is defined by

$$\mathcal{V}_N(x) = \frac{1}{N}\log\det\left[x\mathbf{\Psi}^{-1} + \mathbf{A}_N\bar{\mathbf{\Psi}}\mathbf{A}_N^{\mathsf{T}}\right] + \frac{1}{N}\log\det\left(x\bar{\mathbf{\Psi}}^{-1}\right) - \frac{1}{x}\frac{1}{nN}\sum_{i,j}\sigma_{ij}^2 t_i\bar{t}_j$$

with t_i the ith diagonal entry of the diagonal matrix $\left(\mathbf{\Psi}^{-1} + x\mathbf{A}_N\bar{\mathbf{\Psi}}\mathbf{A}_N^{\mathsf{T}}\right)^{-1}$ and $\bar{t}_j$ the jth diagonal entry of the diagonal matrix $\left(\bar{\mathbf{\Psi}}^{-1} + x\mathbf{A}_N^{\mathsf{T}}\mathbf{\Psi}\mathbf{A}_N\right)^{-1}$.

Remark 6.6. In [Hachem *et al.*, 2008b], it is shown in particular that, if the entries of $\mathbf{Y}_N$ are Gaussian distributed, then the difference between the Stieltjes transform of $\mathrm{E}[F^{\mathbf{B}_N}]$ and its deterministic equivalent, as well as the difference between the Shannon transform of $\mathrm{E}[F^{\mathbf{B}_N}]$ and its deterministic equivalent converge to zero at rate $O(1/N^2)$.

6.2.4 Models involving Haar matrices

As evidenced in the previous section, Hermitian random matrices with i.i.d. entries or originating from general sums or products of such matrices are convenient to study using Stieltjes transform-based methods. This is essentially due to the trace lemma, Theorem 3.4, which provides an almost sure limit to $\mathbf{x}^{\mathsf{H}}(\mathbf{X}\mathbf{X}^{\mathsf{H}} - \mathbf{x}\mathbf{x}^{\mathsf{H}} - z\mathbf{I}_N)^{-1}\mathbf{x}$ with $\mathbf{x}$ one of the independent columns of the random matrix $\mathbf{X}$. Such results can actually be found for more structured random matrices, such as the random bi-unitarily invariant unitary $N \times N$ matrices. We recall from Definition 4.6 that these random matrices are often referred to as *Haar matrices* or *isometric matrices*. Among the known properties of interest here of Haar matrices [Petz and Réffy, 2004], we have the following trace lemma [Chaufray *et al.*, 2004; Debbah *et al.*, 2003a], equivalent to Theorem 3.4 for i.i.d. random matrices.

Theorem 6.15. *Let* $\mathbf{W}$ *be* $n < N$ *columns of an* $N \times N$ *Haar matrix and suppose* $\mathbf{w}$ *is a column of* $\mathbf{W}$. *Let* $\mathbf{B}_N$ *be an* $N \times N$ *random matrix, which is a function of all columns of* $\mathbf{W}$ *except* $\mathbf{w}$. *Then, assuming that, for growing* N, $c = \sup_n n/N < 1$ *and* $B = \sup_N \|\mathbf{B}_N\| < \infty$, *we have:*

$$\mathrm{E}\left[\left|\mathbf{w}^{\mathsf{H}}\mathbf{B}_N\mathbf{w} - \frac{1}{N-n}\operatorname{tr}(\mathbf{\Pi}\mathbf{B}_N)\right|^4\right] \leq \frac{C}{N^2} \tag{6.37}$$

where $\mathbf{\Pi} = \mathbf{I}_N - \mathbf{W}\mathbf{W}^{\mathsf{H}} + \mathbf{w}\mathbf{w}^{\mathsf{H}}$ *and* C *is a constant which depends only on* B *and* c. *If* $\sup_N \|\mathbf{B}_N\| < \infty$, *by the Markov inequality, Theorem 3.5, and the Borel–Cantelli lemma, Theorem 3.6, this entails*

$$\mathbf{w}^{\mathsf{H}}\mathbf{B}_N\mathbf{w} - \frac{1}{N-n}\operatorname{tr}(\mathbf{\Pi}\mathbf{B}_N) \xrightarrow{\text{a.s.}} 0. \tag{6.38}$$

Proof. We provide here an intuitive, yet non-rigorous, sketch of the proof. Let $\mathbf{U} \in \mathbb{C}^{N\times(n-1)}$ be $n-1$ columns of a unitary matrix. We can write all unit-norm vectors $\mathbf{w}$ in the space orthogonal to the space spanned by the columns of $\mathbf{U}$ as $\mathbf{w} = \frac{\mathbf{\Pi}\mathbf{x}}{\|\mathbf{\Pi}\mathbf{x}\|}$, where $\mathbf{\Pi} = \mathbf{I}_N - \mathbf{U}\mathbf{U}^{\mathsf{H}}$ is the projector on the space orthogonal to $\mathbf{U}\mathbf{U}^{\mathsf{H}}$ (and thus $\mathbf{\Pi}\mathbf{\Pi} = \mathbf{\Pi}$) and $\mathbf{x}$ is a Gaussian vector with zero mean and covariance matrix $\mathrm{E}[\mathbf{x}\mathbf{x}^{\mathsf{H}}] = \mathbf{I}_N$ independent of $\mathbf{U}$. This makes $\mathbf{w}$ uniformly distributed in its space. Also, the vector $\mathbf{x}$ is independent of $\mathbf{\Pi}$ by construction. We therefore have from Theorem 3.4 and for N large

$$\mathbf{w}^{\mathsf{H}}\mathbf{B}_N\mathbf{w} = \frac{1}{N}\mathbf{x}^{\mathsf{H}}\mathbf{\Pi}\mathbf{B}_N\mathbf{\Pi}\mathbf{x}\frac{N}{\|\mathbf{\Pi}\mathbf{x}\|^2} \simeq \frac{1}{N}\operatorname{tr}\left(\mathbf{\Pi}\mathbf{B}_N\right)\frac{N}{\|\mathbf{\Pi}\mathbf{x}\|^2}.$$

where the symbol "$\simeq$" stands for some approximation in the large N limit. Notice then that $\mathbf{\Pi x}$ is, up to a basis change, a vector composed of $N-n+1$ i.i.d. standard Gaussian entries and $n-1$ zeros. Hence $\frac{\|\mathbf{\Pi x}\|^2}{N-n} \to 1$. Defining now $\mathbf{W}$ such that $\mathbf{W}\mathbf{W}^\mathsf{H} - \mathbf{w}\mathbf{w}^\mathsf{H} = \mathbf{U}\mathbf{U}^\mathsf{H}$, the reasoning remains valid, and this entails (6.38). □

Since $\mathbf{B}_N$ in Theorem 6.15 is assumed of uniformly bounded spectral norm, $\mathbf{w}^\mathsf{H}\mathbf{B}_N\mathbf{w}$ is uniformly bounded also. Hence, if N, n grow large with ratio n/N uniformly away from one, the term $\frac{1}{N-n}\mathbf{w}^\mathsf{H}\mathbf{B}_N\mathbf{w}$ tends to zero. This therefore entails the following corollary, which can be seen as a rank-1 perturbation of Theorem 6.15.

Corollary 6.2. *Let $\mathbf{W}$ and $\mathbf{B}_N$ be defined as in Theorem 6.15, with N and n such that* $\limsup_n \frac{n}{N} < 1$. *Then, as N, n grow large, for $\mathbf{w}$ any column of $\mathbf{W}$*

$$\mathbf{w}^\mathsf{H}\mathbf{B}_N\mathbf{w} - \frac{1}{N-n}\operatorname{tr}\mathbf{B}_N\left(\mathbf{I}_N - \mathbf{W}\mathbf{W}^\mathsf{H}\right) \xrightarrow{\text{a.s.}} 0.$$

Corollary 6.2 only differs from Theorem 6.15 by the fact that the projector $\mathbf{\Pi}$ is changed into $\mathbf{I}_N - \mathbf{W}\mathbf{W}^\mathsf{H}$.

Also, when $\mathbf{B}_N$ is independent of $\mathbf{W}$, we fall back on the same result as for the i.i.d. case.

Corollary 6.3. *Let $\mathbf{W}$ be defined as in Theorem 6.15, and let $\mathbf{A} \in \mathbb{C}^{N\times N}$ be independent of $\mathbf{W}$ and have uniformly bounded spectral norm. Then, as N grows large, for $\mathbf{w}$ any column of $\mathbf{W}$, we have:*

$$\mathbf{w}^\mathsf{H}\mathbf{A}\mathbf{w} - \frac{1}{N}\operatorname{tr}\mathbf{A} \xrightarrow{\text{a.s.}} 0.$$

Theorem 6.15 is the basis for establishing deterministic equivalents involving isometric matrices. In the following, we introduce a result, based on Silverstein and Bai's approach, which generalizes Theorems 4.10, 4.11, and 4.12 to the case when the $\mathbf{W}_i$ matrices are multiplied on the left by different non-necessarily co-diagonalizable matrices. These models are the basis for studying the properties of multi-user or multi-cellular communications both involving unitary precoders and taking into account the frequency selectivity of the channel. From a mathematical point of view, there exists no simple way to study such models using tools extracted solely from free probability theory. In particular, it is interesting to note that in [Peacock *et al.*, 2008], the authors already generalized Theorem 4.12 to the case where the left-product matrices are different but co-diagonalizable. To do so, the authors relied on tools from free probability as the basic instruments and then need some extra matrix manipulation to derive their limiting result, in a sort of hybrid method between free probability and analytical approach. In the results to come, though, no mention will be made to free probability theory, as the result can be derived autonomously from the tools developed in this section.

The following results are taken from [Couillet *et al.*, 2011b], where detailed proofs can be found. We start by introducing the fundamental equations.

Theorem 6.16 ([Couillet *et al.*, 2011b]). *For $i \in \{1,\ldots,K\}$, let $\mathbf{T}_i \in \mathbb{C}^{n_i \times n_i}$ be nonnegative diagonal and let $\mathbf{H}_i \in \mathbb{C}^{N \times N_i}$. Define $\mathbf{R}_i \triangleq \mathbf{H}_i\mathbf{H}_i^{\mathsf{H}} \in \mathbb{C}^{N\times N}$, $c_i = \frac{n_i}{N_i} < 1$ and $\bar{c}_i = \frac{N_i}{N}$. Let $z < 0$. Then the following system of equations in $\bar{e}_1(z),\ldots,\bar{e}_K(z)$:*

$$\bar{e}_i(z) = \frac{1}{N}\operatorname{tr}\mathbf{T}_i\left(e_i(z)\mathbf{T}_i + [\bar{c}_i - e_i(z)\bar{e}_i(z)]\mathbf{I}_{n_i}\right)^{-1}$$

$$e_i(z) = \frac{1}{N}\operatorname{tr}\mathbf{R}_i\left(\sum_{j=1}^{K}\bar{e}_j(z)\mathbf{R}_j - z\mathbf{I}_N\right)^{-1} \tag{6.39}$$

has a unique solution $\bar{e}_1(z),\ldots,\bar{e}_K(z)$ satisfying $0 \le e_i(z)\bar{e}_i(z) < c_i\bar{c}_i$ for all i. Moreover,

$$\bar{e}_i(z) = \lim_{t\to\infty}\bar{e}_i^{(t)}(z)$$

where $\bar{e}_i^{(t)}(z)$ is the unique solution of

$$\bar{e}_i^{(t)}(z) = \frac{1}{N}\operatorname{tr}\mathbf{T}_i\left(e_i^{(t)}(z)\mathbf{T}_i + [\bar{c}_i - e_i^{(t)}(z)\bar{e}_i^{(t)}(z)]\mathbf{I}_{n_i}\right)^{-1}$$

within the interval $[0, c_i\bar{c}_i/e_i^{(t)}(z))$, $e_i^{(0)}(z)$ can take any positive value and $e_i^{(t)}(z)$ is recursively defined by

$$e_i^{(t)}(z) = \frac{1}{N}\operatorname{tr}\mathbf{R}_i\left(\sum_{j=1}^{K}\bar{e}_j^{(t-1)}(z)\mathbf{R}_j - z\mathbf{I}_N\right)^{-1}.$$

It is important in this result to note the condition $n_i < N_i$. Counter examples can be found with $n_i = N_i$ for which the result no longer holds. The main reason is that, for $n_i = N_i$, $\bar{e}_i(z) = \bar{c}_i/e_i(z)$ is a second fixed-point solution of the first equation in (6.39), which lives in the closure of $[0, \bar{c}_i/e_i(z))$, and may become an attractor of the fixed-point algorithm. We then have the following theorem on a deterministic equivalent for the e.s.d. of the model $\mathbf{B}_N = \sum_{k=1}^{K}\mathbf{H}_i\mathbf{W}_i\mathbf{T}_i\mathbf{W}_i^{\mathsf{H}}\mathbf{H}_i^{\mathsf{H}}$.

Theorem 6.17 ([Couillet *et al.*, 2011b]). *For $i \in \{1,\ldots,K\}$, let $\mathbf{T}_i \in \mathbb{C}^{n_i\times n_i}$ be a Hermitian non-negative matrix with spectral norm bounded uniformly along n_i and $\mathbf{W}_i \in \mathbb{C}^{N_i\times n_i}$ be $n_i < N_i$ columns of a unitary Haar distributed random matrix. Consider $\mathbf{H}_i \in \mathbb{C}^{N\times N_i}$ a random matrix such that $\mathbf{R}_i \triangleq \mathbf{H}_i\mathbf{H}_i^{\mathsf{H}} \in \mathbb{C}^{N\times N}$ has uniformly bounded spectral norm along N, almost surely. Define $c_i = \frac{n_i}{N_i}$ and*

$\bar{c}_i = \frac{N_i}{N}$ *and denote*

$$\mathbf{B}_N = \sum_{i=1}^{K} \mathbf{H}_i \mathbf{W}_i \mathbf{T}_i \mathbf{W}_i^{\mathsf{H}} \mathbf{H}_i^{\mathsf{H}}.$$

Let $z < 0$*. Then, as* N*,* $N_1, \ldots, N_K, n_1, \ldots, n_K$ *grow to infinity with ratios* $\bar{c}_i$ *satisfying* $0 < \liminf \bar{c}_i \leq \limsup \bar{c}_i < \infty$ *and* $0 < \liminf c_i \leq \limsup 1 <$ *for all* i*,*

$$e_{\mathbf{B}_N,i}(z) - e_i(z) \xrightarrow{\text{a.s.}} 0$$

where

$$e_{\mathbf{B}_N,i}(z) = \frac{1}{N} \operatorname{tr} \mathbf{R}_i \left(\mathbf{B}_N - z\mathbf{I}_N\right)^{-1}$$

and $e_i(z)$ *are given by Theorem 6.16.*

Note that this result does not prove that $\frac{1}{N}\operatorname{tr}(\sum_{k=1}^{K} \bar{e}_j(z)\mathbf{R}_j - z\mathbf{I}_N)^{-1}$ is a deterministic equivalent of $m_{\mathbf{B}_N}(z)$ for z complex, and therefore we do not have the classical deterministic equivalent for the eigenvalue distribution of $\mathbf{B}_N$. The main reason is that, in the Haar case, the quantities $e_i(z)$ are not easily extensible as holomorphic function in $\mathbb{C} \setminus \mathbb{R}^+$, which is fundamental to ensure the convergence in law of the eigenvalue distribution (recall Theorem 3.10). In [Couillet *et al.*, 2011b], the holomorphic extension, at the core of the proof of Theorem 6.17, is only ensured on a cone of $\mathbb{C} \cap \{z \in \mathbb{C},\ \Re[z] < 0\}$ including the negative axis. This is sufficient for Theorem 6.17 for not for the convergence in law.

Consider the case when, for each i, $\bar{c}_i = 1$ and $\mathbf{H}_i = \mathbf{R}_i^{\frac{1}{2}}$ for some square Hermitian non-negative square root $\mathbf{R}_i^{\frac{1}{2}}$ of $\mathbf{R}_i$. We observe that the system of Equations (6.39) is very similar to the system of Equations (6.7) established for the case of i.i.d. random matrices. The noticeable difference here is the addition of the extra term $-e_i\bar{e}_i$ in the expression of $\bar{e}_i$. Without this term, we fall back on the i.i.d. case. Notice also that the case $K = 1$ corresponds exactly to Theorem 4.11, which was treated for $c_1 = 1$ and for which simplifications help discarding the limitation $\limsup c_i < 1$.

The above limitation may appear particularly constraining for application purposes as one traditionally considers SDMA or CDMA precoders with full rank Haar matrices. Nonetheless, it is not so common to consider $\mathbf{T}_i$ different from the identity matrix (see the applications in [Couillet *et al.*, 2011b]). For $c_i = 1$, $\mathbf{T}_i = \mathbf{I}_{N_i}$, the result is in fact trivial as all randomness disappears from the identity $\mathbf{W}_i\mathbf{W}_i^{\mathsf{H}} = \mathbf{I}_{N_i}$. One therefore does not loose much in assuming $\limsup c_i < 1$ in general. Nonetheless, it is clear that the extension to $c_i = 1$ must be possible although this is still an open question so far.

We hereafter provide both a sketch of the proof and a rather extensive derivation, which explains how (6.39) is derived and how uniqueness is proved. For readability, we take $\bar{c}_i = 1$. The main steps of the proof are similar to those developed for the proof of Theorem 6.1. In order to propose different approaches than in previous derivations, and because this is more complicated here, we will work almost exclusively with real negative z, instead of z with positive imaginary part. We will also provide a shorter proof of the final convergence step $e_{\mathbf{B}_N,i}(z) - e_i(z) \xrightarrow{\text{a.s.}} 0$, relying on restrictions of the domain of z along with arguments from Vitali's convergence theorem. These approaches are valid here because upper bounds on the spectral norms of $\mathbf{R}_i$ and $\mathbf{T}_i$ are considered, which was not the case for Theorem 6.1. Apart from these technical considerations, the main noticeable difference between the deterministic equivalent approaches proposed for matrices with independent entries and for Haar matrices lies in the first convergence step, which is much more intricate.

Proof. We first provide a sketch of the proof for better understanding, which will enhance the aforementioned main novelty. As usual, we wish to prove that there exists a matrix $\mathbf{F} = \sum_{i=1}^{K} \bar{f}_i \mathbf{R}_i$, such that, for all non-negative $\mathbf{A}$ with $\|\mathbf{A}\| < \infty$

$$\frac{1}{N} \operatorname{tr} \mathbf{A} \left(\mathbf{B}_N - z\mathbf{I}_N\right)^{-1} - \frac{1}{N} \operatorname{tr} \mathbf{A} \left(\mathbf{F} - z\mathbf{I}_N\right)^{-1} \xrightarrow{\text{a.s.}} 0.$$

Contrary to classical deterministic equivalent approaches for random matrices with i.i.d. entries, finding a deterministic equivalent for $\frac{1}{N} \operatorname{tr} \mathbf{A} \left(\mathbf{B}_N - z\mathbf{I}_N\right)^{-1}$ is not straightforward. The reason is that, during the derivation, terms such as $\frac{1}{N-n_i} \operatorname{tr} \left(\mathbf{I}_N - \mathbf{W}_i\mathbf{W}_i^{\mathsf{H}}\right) \mathbf{A}^{\frac{1}{2}} \left(\mathbf{B}_N - z\mathbf{I}_N\right)^{-1} \mathbf{A}^{\frac{1}{2}}$, with the $\left(\mathbf{I}_N - \mathbf{W}_i\mathbf{W}_i^{\mathsf{H}}\right)$ prefix will naturally appear, as a result of applying the trace lemma, Theorem 6.15, that will be required to be controlled. We proceed as follows.

- We first denote for all i, $\delta_i \triangleq \frac{1}{N-n_i} \operatorname{tr} \left(\mathbf{I}_N - \mathbf{W}_i\mathbf{W}_i^{\mathsf{H}}\right) \mathbf{R}_i^{\frac{1}{2}} \left(\mathbf{B}_N - z\mathbf{I}_N\right)^{-1} \mathbf{R}_i^{\frac{1}{2}}$ some auxiliary variable. Then, using the same techniques as in the proof of Theorem 6.1, denoting further $f_i \triangleq \frac{1}{N} \operatorname{tr} \mathbf{R}_i \left(\mathbf{B}_N - z\mathbf{I}_N\right)^{-1}$, we prove

$$f_i - \frac{1}{N} \operatorname{tr} \mathbf{R}_i \left(\mathbf{G} - z\mathbf{I}_N\right)^{-1} \xrightarrow{\text{a.s.}} 0$$

with $\mathbf{G} = \sum_{j=1}^{K} \bar{g}_j \mathbf{R}_j$ and

$$\bar{g}_i = \frac{1}{1 - c_i + \frac{1}{N} \sum_{l=1}^{n_i} \frac{1}{1+t_{il}\delta_i}} \frac{1}{N} \sum_{l=1}^{n_i} \frac{t_{il}}{1 + t_{il}\delta_i}$$

where $t_{i1}, \ldots, t_{in_i}$ are the eigenvalues of $\mathbf{T}_i$. Noticing additionally that

$$(1 - c_i)\delta_i - f_i + \frac{1}{N} \sum_{l=1}^{n_i} \frac{\delta_i}{1 + t_{il}\delta_i} \xrightarrow{\text{a.s.}} 0$$

we have a first hint on a first deterministic equivalent for f_i. Precisely, we expect to obtain the set of fundamental equations

$$\Delta_i = \frac{1}{1-c_i}\left[e_i - \frac{1}{N}\sum_{l=1}^{n_i}\frac{\Delta_i}{1+t_{il}\Delta_i}\right]$$
$$e_i = \frac{1}{N}\operatorname{tr}\mathbf{R}_i\left(\sum_{j=1}^{K}\frac{1}{1-c_j+\frac{1}{N}\sum_{l=1}^{n_j}\frac{1}{1+t_{jl}\Delta_j}}\frac{1}{N}\sum_{l=1}^{n_j}\frac{t_{jl}}{1+t_{jl}\Delta_j}\mathbf{R}_j - z\mathbf{I}_N\right)^{-1}.$$

- The expressions of $\bar{g}_i$ and their deterministic equivalents are however not very convenient under this form. It is then shown that

$$\bar{g}_i - \frac{1}{N}\sum_{l=1}^{n_i}\frac{t_{il}}{1+t_{il}f_i - f_i\bar{g}_i} = \bar{g}_i - \frac{1}{N}\operatorname{tr}\mathbf{T}_i\left(f_i\mathbf{T}_i + [1-f_i\bar{g}_i]\mathbf{I}_{n_i}\right)^{-1} \xrightarrow{\text{a.s.}} 0$$

which induces the $2K$-equation system

$$f_i - \frac{1}{N}\operatorname{tr}\mathbf{R}_i\left(\sum_{j=1}^{K}\bar{g}_j\mathbf{R}_j - z\mathbf{I}_N\right)^{-1} \xrightarrow{\text{a.s.}} 0$$
$$\bar{g}_i - \frac{1}{N}\operatorname{tr}\mathbf{T}_i\left(\bar{g}_i\mathbf{T}_i + [1-f_i\bar{g}_i]\right)^{-1} \xrightarrow{\text{a.s.}} 0.$$

- These relations are sufficient to infer the deterministic equivalent but will be made more attractive for further considerations by introducing $\mathbf{F} = \sum_{i=1}^{K}\bar{f}_i\mathbf{R}_i$ and proving that

$$f_i - \frac{1}{N}\operatorname{tr}\mathbf{R}_i\left(\sum_{j=1}^{K}\bar{f}_j\mathbf{R}_j - z\mathbf{I}_N\right)^{-1} \xrightarrow{\text{a.s.}} 0$$
$$\bar{f}_i - \frac{1}{N}\operatorname{tr}\mathbf{T}_i\left(\bar{f}_i\mathbf{T}_i + [1-f_i\bar{f}_i]\right)^{-1} = 0$$

where, for $z<0$, $\bar{f}_i$ lies in $[0, c_i/f_i)$ and is now uniquely determined by f_i. In particular, this step provides an explicit expression $\bar{f}_i$ as a function of f_i, which will be translated into an explicit expression of $\bar{e}_i$ as a function of e_i.

This is the very technical part of the proof. We then prove the existence and uniqueness of a solution to the fixed-point equation

$$e_i - \frac{1}{N}\operatorname{tr}\mathbf{R}_i\left(\sum_{j=1}^{K}\bar{e}_j\mathbf{R}_j - z\mathbf{I}_N\right)^{-1} = 0$$
$$\bar{e}_i - \frac{1}{N}\operatorname{tr}\mathbf{T}_i\left(\bar{e}_i\mathbf{T}_i + [1-e_i\bar{e}_i]\right)^{-1} = 0$$

for all finite N, z real negative, and for $\bar{e}_i \in [0, c_i/f_i)$. Here, instead of following the approach of the proof of uniqueness for the fundamental equations of Theorem 6.1, we use a property of so-called *standard functions*. We will show precisely that the vector application $\mathbf{h} = (h_1, \ldots, h_K)$ with

$$h_i : (x_1, \ldots, x_K) \mapsto \frac{1}{N} \operatorname{tr} \mathbf{R}_i \left(\sum_{j=1}^{K} \bar{x}_j \mathbf{R}_j - z\mathbf{I}_N \right)^{-1}$$

where $\bar{x}_i$ is the unique solution to

$$\bar{x}_i = \frac{1}{N} \operatorname{tr} \mathbf{T}_i \left(\bar{x}_i \mathbf{T}_i + [1 - x_i \bar{x}_i] \right)^{-1}$$

lying in $[0, c_i/x_i)$, is a standard function. It will unfold that the fixed-point equation in $(e_1, \ldots, e_K)$ has a unique solution with positive entries and that this solution can be determined as the limiting iteration of a classical fixed-point algorithm.

The last step proves that the unique solution $(e_1, \ldots, e_N)$ is such that

$$e_i - f_i \xrightarrow{\text{a.s.}} 0$$

which is solved by arguments borrowed from the work of Hachem *et al.* [Hachem *et al.*, 2007], using a restriction on the definition domain of z, which simplifies greatly the calculus.

We now turn to the precise proof. We use again the Bai and Silverstein steps: the convergence $f_i - \frac{1}{N} \operatorname{tr} \mathbf{R}_i (\sum_{j=1}^K \bar{f}_j \mathbf{R}_j - z\mathbf{I}_N)^{-1} \xrightarrow{\text{a.s.}} 0$ in a first step, the existence and uniqueness of a solution to $e_i = \frac{1}{N} \operatorname{tr} \mathbf{R}_i (\sum_{j=1}^K \bar{e}_j \mathbf{R}_j - z\mathbf{I}_N)^{-1}$ in a second, and the convergence $e_i - f_i \xrightarrow{\text{a.s.}} 0$ in a third. Although precise control of the random variables involved needs be carried out, as is detailed in [Couillet *et al.*, 2011b], we hereafter elude most technical parts for simplicity and understanding.

Step 1: First convergence step

In this section, we take $z < 0$, until further notice. Let us first introduce the following parameters. We will denote $T = \max_i\{\limsup \|\mathbf{T}_i\|\}$, $R = \max_i\{\limsup \|\mathbf{R}_i\|\}$ and $c = \max_i\{\limsup c_i\} < 1$.

We start with classical deterministic equivalent techniques. Let $\mathbf{A} \in \mathbb{C}^{N \times N}$ be a Hermitian non-negative definite matrix with spectral norm uniformly bounded by A. Taking $\mathbf{G} = \sum_{j=1}^K \bar{g}_j \mathbf{R}_j$, with $\bar{g}_1, \ldots, \bar{g}_K$ left undefined for the moment, we have:

$$\begin{aligned}
&\frac{1}{N} \operatorname{tr} \mathbf{A}(\mathbf{B}_N - z\mathbf{I}_N)^{-1} - \frac{1}{N} \operatorname{tr} \mathbf{A}(\mathbf{G} - z\mathbf{I}_N)^{-1} \\
&= \frac{1}{N} \operatorname{tr} \left[\mathbf{A}(\mathbf{B}_N - z\mathbf{I}_N)^{-1} \sum_{i=1}^{K} \mathbf{R}_i^{\frac{1}{2}} \left(-\mathbf{W}_i \mathbf{T}_i \mathbf{W}_i^{\mathsf{H}} + \bar{g}_i \mathbf{I}_N \right) \mathbf{R}_i^{\frac{1}{2}} (\mathbf{G} - z\mathbf{I}_N)^{-1} \right] \\
&= \sum_{i=1}^{K} \bar{g}_i \frac{1}{N} \operatorname{tr} \mathbf{A}(\mathbf{B}_N - z\mathbf{I}_N)^{-1} \mathbf{R}_i (\mathbf{G} - z\mathbf{I}_N)^{-1}
\end{aligned}$$

$$
\begin{aligned}
&- \frac{1}{N}\sum_{i=1}^{K}\sum_{l=1}^{n_i} t_{il}\mathbf{w}_{il}^{\mathsf{H}}\mathbf{R}_i^{\frac{1}{2}}(\mathbf{G}-z\mathbf{I}_N)^{-1}\mathbf{A}(\mathbf{B}_N-z\mathbf{I}_N)^{-1}\mathbf{R}_i^{\frac{1}{2}}\mathbf{w}_{il} \\
&= \sum_{i=1}^{K}\bar{g}_i\frac{1}{N}\operatorname{tr}\mathbf{A}(\mathbf{B}_N-z\mathbf{I}_N)^{-1}\mathbf{R}_i(\mathbf{G}-z\mathbf{I}_N)^{-1} \\
&- \frac{1}{N}\sum_{i=1}^{K}\sum_{l=1}^{n_i}\frac{t_{il}\mathbf{w}_{il}^{\mathsf{H}}\mathbf{R}_i^{\frac{1}{2}}(\mathbf{G}-z\mathbf{I}_N)^{-1}\mathbf{A}(\mathbf{B}_{(i,l)}-z\mathbf{I}_N)^{-1}\mathbf{R}_i^{\frac{1}{2}}\mathbf{w}_{il}}{1+t_{il}\mathbf{w}_{il}^{\mathsf{H}}\mathbf{R}_i^{\frac{1}{2}}(\mathbf{B}_{(i,l)}-z\mathbf{I}_N)^{-1}\mathbf{R}_i^{\frac{1}{2}}\mathbf{w}_{il}},
\end{aligned} \tag{6.40}
$$

with $t_{i1},\ldots,t_{in_i}$ the eigenvalues of $\mathbf{T}_i$.

The quadratic forms $\mathbf{w}_{il}^{\mathsf{H}}\mathbf{R}_i^{\frac{1}{2}}(\mathbf{G}-z\mathbf{I}_N)^{-1}\mathbf{A}(\mathbf{B}_{(i,l)}-z\mathbf{I}_N)^{-1}\mathbf{R}_i^{\frac{1}{2}}\mathbf{w}_{il}$ and $\mathbf{w}_{il}^{\mathsf{H}}\mathbf{R}_i^{\frac{1}{2}}(\mathbf{B}_{(i,l)}-z\mathbf{I}_N)^{-1}\mathbf{R}_i^{\frac{1}{2}}\mathbf{w}_{il}$ are not asymptotically close to the trace of the inner matrix, as in the i.i.d. case, but to the trace of the inner matrix multiplied by $(\mathbf{I}_N-\mathbf{W}_i\mathbf{W}_i^{\mathsf{H}})$, as suggested by Theorem 6.15. This complicates the calculus. In the following, we will therefore study the following stochastic quantities, namely the random variables δ_i, β_i and f_i, introduced below.

For every $i\in\{1,\ldots,K\}$, denote

$$
\begin{aligned}
\delta_i &\triangleq \frac{1}{N-n_i}\operatorname{tr}\left(\mathbf{I}_N-\mathbf{W}_i\mathbf{W}_i^{\mathsf{H}}\right)\mathbf{R}_i^{\frac{1}{2}}\left(\mathbf{B}_N-z\mathbf{I}_N\right)^{-1}\mathbf{R}_i^{\frac{1}{2}} \\
f_i &\triangleq \frac{1}{N}\operatorname{tr}\mathbf{R}_i\left(\mathbf{B}_N-z\mathbf{I}_N\right)^{-1}
\end{aligned}
$$

both being clearly non-negative.

Writing $\mathbf{W}_i=[\mathbf{w}_{i,1},\ldots,\mathbf{w}_{i,n_i}]$ and $\mathbf{W}_i\mathbf{W}_i^{\mathsf{H}}=\sum_{l=1}^{n_i}\mathbf{w}_{il}\mathbf{w}_{il}^{\mathsf{H}}$, we have from standard calculus and the matrix inversion lemma, Lemma 6.2, that

$$
\begin{aligned}
(1-c_i)\delta_i &= f_i-\frac{1}{N}\sum_{l=1}^{n_i}\mathbf{w}_{il}^{\mathsf{H}}\mathbf{R}_i^{\frac{1}{2}}\left(\mathbf{B}_N-z\mathbf{I}_N\right)^{-1}\mathbf{R}_i^{\frac{1}{2}}\mathbf{w}_{il} \\
&= f_i-\frac{1}{N}\sum_{l=1}^{n_i}\frac{\mathbf{w}_{il}^{\mathsf{H}}\mathbf{R}_i^{\frac{1}{2}}\left(\mathbf{B}_{(i,l)}-z\mathbf{I}_N\right)^{-1}\mathbf{R}_i^{\frac{1}{2}}\mathbf{w}_{il}}{1+t_{il}\mathbf{w}_{il}^{\mathsf{H}}\mathbf{R}_i^{\frac{1}{2}}\left(\mathbf{B}_{(i,l)}-z\mathbf{I}_N\right)^{-1}\mathbf{R}_i^{\frac{1}{2}}\mathbf{w}_{il}}
\end{aligned} \tag{6.41}
$$

with $\mathbf{B}_{(i,l)}=\mathbf{B}_N-t_{il}\mathbf{R}_i^{\frac{1}{2}}\mathbf{w}_{il}\mathbf{w}_{il}^{\mathsf{H}}\mathbf{R}_i^{\frac{1}{2}}$.

Since $z<0$, $\delta_i\geq 0$, so that $\frac{1}{1+t_{il}\delta_i}$ is well defined. We recognize already from Theorem 6.15 that each quadratic term $\mathbf{w}_{il}^{\mathsf{H}}\mathbf{R}_i^{\frac{1}{2}}(\mathbf{B}_{(i,l)}-z\mathbf{I}_N)^{-1}\mathbf{R}_i^{\frac{1}{2}}\mathbf{w}_{il}$ is asymptotically close to δ_i. By adding the term $\frac{1}{N}\sum_{l=1}^{n_i}\frac{\delta_i}{1+t_{il}\delta_i}$ on both sides, (6.41) can further be rewritten

$$
\begin{aligned}
&(1-c_i)\delta_i-f_i+\frac{1}{N}\sum_{l=1}^{n_i}\frac{\delta_i}{1+t_{il}\delta_i} \\
&= \frac{1}{N}\sum_{l=1}^{n_i}\left[\frac{\delta_i}{1+t_{il}\delta_i}-\frac{\mathbf{w}_{il}^{\mathsf{H}}\mathbf{R}_i^{\frac{1}{2}}\left(\mathbf{B}_{(i,l)}-z\mathbf{I}_N\right)^{-1}\mathbf{R}_i^{\frac{1}{2}}\mathbf{w}_{il}}{1+t_{il}\mathbf{w}_{il}^{\mathsf{H}}\mathbf{R}_i^{\frac{1}{2}}\left(\mathbf{B}_{(i,l)}-z\mathbf{I}_N\right)^{-1}\mathbf{R}_i^{\frac{1}{2}}\mathbf{w}_{il}}\right].
\end{aligned}
$$

We now apply the trace lemma, Theorem 6.15, which ensures that

$$\mathrm{E}\left[\left|(1-c_i)\delta_i - f_i + \frac{1}{N}\sum_{l=1}^{n_i}\frac{\delta_i}{1+t_{il}\delta_i}\right|^4\right] = O\left(\frac{1}{N^2}\right). \tag{6.42}$$

We do not provide the precise derivations of the fourth order moment inequalities here and in all the equations that follow, our main purpose being concentrated on the fundamental steps of the proof. Precise calculus and upper bounds can be found in [Couillet *et al.*, 2011b]. This is our first relation that links δ_i to $f_i = \frac{1}{N}\operatorname{tr}\mathbf{R}_i\left(\mathbf{B}_N - z\mathbf{I}_N\right)^{-1}$.

Introducing now an additional $\mathbf{A}(\mathbf{G}-z\mathbf{I}_N)^{-1}$ matrix in the argument of the trace of δ_i, with $\mathbf{G},\mathbf{A}\in\mathbb{C}^{N\times N}$ any non-negative definite matrices, $\|\mathbf{A}\|\leq A$, we denote

$$\beta_i \triangleq \frac{1}{N-n_i}\operatorname{tr}\left(\mathbf{I}_N - \mathbf{W}_i\mathbf{W}_i^{\mathsf{H}}\right)\mathbf{R}_i^{\frac{1}{2}}\left(\mathbf{G}-z\mathbf{I}_N\right)^{-1}\mathbf{A}\left(\mathbf{B}_N - z\mathbf{I}_N\right)^{-1}\mathbf{R}_i^{\frac{1}{2}}.$$

We then proceed similarly as for δ_i by showing

$$\begin{aligned}\beta_i &= \frac{1}{N-n_i}\operatorname{tr}\mathbf{R}_i^{\frac{1}{2}}\left(\mathbf{G}-z\mathbf{I}_N\right)^{-1}\mathbf{A}\left(\mathbf{B}_N-z\mathbf{I}_N\right)^{-1}\mathbf{R}_i^{\frac{1}{2}} \\ &\quad - \frac{1}{N-n_i}\sum_{l=1}^{n_i}\frac{\mathbf{w}_{il}^{\mathsf{H}}\mathbf{R}_i^{\frac{1}{2}}\left(\mathbf{G}-z\mathbf{I}_N\right)^{-1}\mathbf{A}\left(\mathbf{B}_{(i,l)}-z\mathbf{I}_N\right)^{-1}\mathbf{R}_i^{\frac{1}{2}}\mathbf{w}_{il}}{1+t_{il}\mathbf{w}_{il}^{\mathsf{H}}\mathbf{R}_i^{\frac{1}{2}}\left(\mathbf{B}_{(i,l)}-z\mathbf{I}_N\right)^{-1}\mathbf{R}_i^{\frac{1}{2}}\mathbf{w}_{il}}\end{aligned}$$

from which we have:

$$\begin{aligned}&\frac{1}{N-n_i}\operatorname{tr}\mathbf{R}_i^{\frac{1}{2}}\left(\mathbf{G}-z\mathbf{I}_N\right)^{-1}\mathbf{A}\left(\mathbf{B}_N-z\mathbf{I}_N\right)^{-1}\mathbf{R}_i^{\frac{1}{2}} - \frac{1}{N-n_i}\sum_{l=1}^{n_i}\frac{\beta_i}{1+t_{il}\delta_i} - \beta_i \\ &= \frac{1}{N-n_i}\sum_{l=1}^{n_i}\left[\frac{\mathbf{w}_{il}^{\mathsf{H}}\mathbf{R}_i^{\frac{1}{2}}\left(\mathbf{G}-z\mathbf{I}_N\right)^{-1}\mathbf{A}\left(\mathbf{B}_{(i,l)}-z\mathbf{I}_N\right)^{-1}\mathbf{R}_i^{\frac{1}{2}}\mathbf{w}_{il}}{1+t_{il}\mathbf{w}_{il}^{\mathsf{H}}\mathbf{R}_i^{\frac{1}{2}}\left(\mathbf{B}_{(i,l)}-z\mathbf{I}_N\right)^{-1}\mathbf{R}_i^{\frac{1}{2}}\mathbf{w}_{il}} - \frac{\beta_i}{1+t_{il}\delta_i}\right].\end{aligned}$$

Since numerators and denominators converge again to one another assuming $\mathbf{G}$ independent of $\mathbf{w}_{il}$, we can show from Theorem 6.15 again that

$$\mathrm{E}\left[\left|\frac{\mathbf{w}_{il}^{\mathsf{H}}\mathbf{R}_i^{\frac{1}{2}}\left(\mathbf{G}-z\mathbf{I}_N\right)^{-1}\mathbf{A}\left(\mathbf{B}_{(i,l)}-z\mathbf{I}_N\right)^{-1}\mathbf{R}_i^{\frac{1}{2}}\mathbf{w}_{il}}{1+t_{il}\mathbf{w}_{il}^{\mathsf{H}}\mathbf{R}_i^{\frac{1}{2}}\left(\mathbf{B}_{(i,l)}-z\mathbf{I}_N\right)^{-1}\mathbf{R}_i^{\frac{1}{2}}\mathbf{w}_{il}} - \frac{\beta_i}{1+t_{il}\delta_i}\right|^4\right] = O\left(\frac{1}{N^2}\right). \tag{6.43}$$

Hence

$$\begin{aligned}&\mathrm{E}\left[\left|\frac{1}{N}\operatorname{tr}\mathbf{R}_i\left(\mathbf{G}-z\mathbf{I}_N\right)^{-1}\mathbf{A}\left(\mathbf{B}_N-z\mathbf{I}_N\right)^{-1} - \beta_i\left(1-c_i+\frac{1}{N}\sum_{l=1}^{n_i}\frac{1}{1+t_{il}\delta_i}\right)\right|^4\right] \\ &= O\left(\frac{1}{N^2}\right).\end{aligned} \tag{6.44}$$

This provides us with the second relation that links β_i to $\frac{1}{N}\operatorname{tr}\mathbf{R}_i^{\frac{1}{2}}\left(\mathbf{G}-z\mathbf{I}_N\right)^{-1}\mathbf{A}\left(\mathbf{B}_N-z\mathbf{I}_N\right)^{-1}\mathbf{R}_i^{\frac{1}{2}}$. That is, we have expressed both δ_i and β_i as a function of the traces $\frac{1}{N}\operatorname{tr}\mathbf{R}_i^{\frac{1}{2}}(\mathbf{B}_N - z\mathbf{I}_N)^{-1}\mathbf{R}_i^{\frac{1}{2}}$ and $\frac{1}{N}\operatorname{tr}\mathbf{R}_i^{\frac{1}{2}}(\mathbf{G} - z\mathbf{I}_N)^{-1}\mathbf{A}(\mathbf{B}_N - z\mathbf{I}_N)^{-1}\mathbf{R}_i^{\frac{1}{2}}$, which are more conventional to work with.

We are now in position to determine adequate expressions for $\bar{g}_1, \ldots, \bar{g}_K$. Anticipating on the coming equations, we choose

$$\bar{g}_i = \frac{1}{1 - c_i + \frac{1}{N}\sum_{l=1}^{n_i}\frac{1}{1+t_{il}\delta_i}}\frac{1}{N}\sum_{l=1}^{n_i}\frac{t_{il}}{1+t_{il}\delta_i}.$$

Note however that $\bar{g}_i$ is not independent of $\mathbf{w}_{il}$ as it depends on δ_i, so that the convergence (6.43) is not valid. However, it is easy to see that $\bar{g}_i$ so defined is within $O(1/N)$ of the same quantity defined with column $\mathbf{w}_{il}$ and t_{il} removed from the expression of $\mathbf{B}_N$ (and then of δ_i). The above convergence steps are therefore still valid.

Going back to our original problem with the inferred value for $\bar{g}_i$, we have

$$\begin{aligned}
&\frac{1}{N}\operatorname{tr}\mathbf{A}(\mathbf{B}_N - z\mathbf{I}_N)^{-1} - \frac{1}{N}\operatorname{tr}\mathbf{A}(\mathbf{G} - z\mathbf{I}_N)^{-1} \\
&= \sum_{i=1}^{K}\frac{\frac{1}{N}\sum_{l=1}^{n_i}\frac{t_{il}}{1+t_{il}\delta_i}\frac{1}{N}\operatorname{tr}\mathbf{R}_i(\mathbf{G} - z\mathbf{I}_N)^{-1}\mathbf{A}(\mathbf{B}_N - z\mathbf{I}_N)^{-1}}{1 - c_i + \frac{1}{N}\sum_{l=1}^{n_i}\frac{1}{1+t_{il}\delta_i}} \\
&- \frac{1}{N}\sum_{i=1}^{K}\sum_{l=1}^{n_i}\frac{t_{il}\mathbf{w}_{il}^{\mathsf{H}}\mathbf{R}_i^{\frac{1}{2}}(\mathbf{G} - z\mathbf{I}_N)^{-1}\mathbf{A}(\mathbf{B}_{(i,l)} - z\mathbf{I}_N)^{-1}\mathbf{R}_i^{\frac{1}{2}}\mathbf{w}_{il}}{1 + t_{il}\mathbf{w}_{il}^{\mathsf{H}}\mathbf{R}_i^{\frac{1}{2}}(\mathbf{B}_{(i,l)} - z\mathbf{I}_N)^{-1}\mathbf{R}_i^{\frac{1}{2}}\mathbf{w}_{il}} \\
&= \sum_{i=1}^{K}\frac{1}{N}\sum_{l=1}^{n_i}t_{il}\left[\frac{\frac{1}{N}\operatorname{tr}\mathbf{R}_i(\mathbf{G} - z\mathbf{I}_N)^{-1}\mathbf{A}(\mathbf{B}_N - z\mathbf{I}_N)^{-1}}{(1 - c_i + \frac{1}{N}\sum_{l'=1}^{n_i}\frac{1}{1+t_{i,l'}\delta_i})(1 + t_{il}\delta_i)}\right. \\
&\qquad\left. - \frac{\mathbf{w}_{il}^{\mathsf{H}}\mathbf{R}_i^{\frac{1}{2}}(\mathbf{G} - z\mathbf{I}_N)^{-1}\mathbf{A}(\mathbf{B}_{(i,l)} - z\mathbf{I}_N)^{-1}\mathbf{R}_i^{\frac{1}{2}}\mathbf{w}_{il}}{1 + t_{il}\mathbf{w}_{il}^{\mathsf{H}}\mathbf{R}_i^{\frac{1}{2}}(\mathbf{B}_{(i,l)} - z\mathbf{I}_N)^{-1}\mathbf{R}_i^{\frac{1}{2}}\mathbf{w}_{il}}\right].
\end{aligned}$$

To show that this last difference tends to zero, notice that $1 + t_{il}\delta_i \geq 1$ and

$$1 - c_i \leq 1 - c_i + \frac{1}{N}\sum_{l=1}^{n_i}\frac{1}{1 + t_{il}\delta_i} \leq 1$$

which ensure that we can divide the term in the expectation in the left-hand side of (6.44) by $1 + t_{il}\delta_i$ and $1 - c_i + \frac{1}{N}\sum_{l=1}^{n_i}\frac{1}{1+t_{il}\delta_i}$ without risking altering the order of convergence. This results in

$$\mathrm{E}\left[\left|\frac{\beta_i}{1 + t_{il}\delta_i} - \frac{\frac{1}{N}\operatorname{tr}\mathbf{R}_i^{\frac{1}{2}}(\mathbf{G} - z\mathbf{I}_N)^{-1}\mathbf{A}(\mathbf{B}_N - z\mathbf{I}_N)^{-1}\mathbf{R}_i^{\frac{1}{2}}}{\left(1 - c_i + \frac{1}{N}\sum_{l=1}^{n_i}\frac{1}{1+t_{il}\delta_i}\right)(1 + t_{il}\delta_i)}\right|^4\right] = O\left(\frac{1}{N^2}\right). \tag{6.45}$$

From (6.43) and (6.45), we finally have that

$$\mathrm{E}\left[\left|\frac{\frac{1}{N}\operatorname{tr}\mathbf{R}_i\left(\mathbf{G}-z\mathbf{I}_N\right)^{-1}\mathbf{A}\left(\mathbf{B}_N-z\mathbf{I}_N\right)^{-1}}{\left(1-c_i+\frac{1}{N}\sum_{l=1}^{n_i}\frac{1}{1+t_{il}\delta_i}\right)(1+t_{il}\delta_i)}\right.\right.$$

$$\left.\left. - \frac{\mathbf{w}_{il}^{\mathsf{H}}\mathbf{R}_i^{\frac{1}{2}}\left(\mathbf{G}-z\mathbf{I}_N\right)^{-1}\mathbf{A}\left(\mathbf{B}_{(i,l)}-z\mathbf{I}_N\right)^{-1}\mathbf{R}_i^{\frac{1}{2}}\mathbf{w}_{il}}{1+t_{il}\mathbf{w}_{il}^{\mathsf{H}}\mathbf{R}_i^{\frac{1}{2}}\left(\mathbf{B}_{(i,l)}-z\mathbf{I}_N\right)^{-1}\mathbf{R}_i^{\frac{1}{2}}\mathbf{w}_{il}}\right|^4\right] = O\left(\frac{1}{N^2}\right) \qquad (6.46)$$

from which we obtain finally

$$\mathrm{E}\left[\left|\frac{1}{N}\operatorname{tr}\mathbf{A}(\mathbf{B}_N-z\mathbf{I}_N)^{-1}-\frac{1}{N}\operatorname{tr}\mathbf{A}(\mathbf{G}-z\mathbf{I}_N)^{-1}\right|^4\right] = O\left(\frac{1}{N^2}\right). \qquad (6.47)$$

This provides us with a first interesting result, from which we could infer a deterministic equivalent of $e_{\mathbf{B}_N,j}(z)$, which would be written as a function of deterministic equivalents of the δ_i and deterministic equivalents of the f_i, for $i=\{1,\ldots,K\}$. However this form is impractical to work with and we need to go further in the study of $\bar{g}_i$.

Observe that $\bar{g}_i$ can be written under the form

$$\bar{g}_i = \frac{1}{N}\sum_{l=1}^{n_i}\frac{t_{il}}{(1-c_i+\frac{1}{N}\sum_{l=1}^{n_i}\frac{1}{1+t_{il}\delta_i})+t_{il}\delta_i(1-c_i+\frac{1}{N}\sum_{l=1}^{n_i}\frac{1}{1+t_{il}\delta_i})}.$$

We will study the denominator of the above expression and show that it can be synthesized into a much more attractive form.

From (6.42), we first have

$$\mathrm{E}\left[\left|f_i-\delta_i\left(1-c_i+\frac{1}{N}\sum_{l=1}^{n_i}\frac{1}{1+t_{il}\delta_i}\right)\right|^4\right] = O\left(\frac{1}{N^2}\right).$$

Noticing that

$$1-\bar{g}_i\delta_i\left(1-c_i+\frac{1}{N}\sum_{l=1}^{n_i}\frac{1}{1+t_{il}\delta_i}\right) = 1-c_i+\frac{1}{N}\sum_{l=1}^{n_i}\frac{1}{1+t_{il}\delta_i}$$

we therefore also have

$$\mathrm{E}\left[\left|(1-\bar{g}_if_i)-\left(1-c_i+\frac{1}{N}\sum_{l=1}^{n_i}\frac{1}{1+t_{il}\delta_i}\right)\right|^4\right] = O\left(\frac{1}{N^2}\right).$$

The two relations above lead to

$$\mathrm{E}\left[\left|\bar{g}_i-\frac{1}{N}\sum_{l=1}^{n_i}\frac{t_{il}}{t_{il}f_i+1-f_i\bar{g}_i}\right|^4\right]$$

$$= \mathrm{E}\left[\left|\frac{1}{N}\sum_{l=1}^{n_i}t_{il}\frac{t_{il}\left[f_i-\delta_i\kappa_i\right]+\left[1-f_i\bar{g}_i-\kappa_i\right]}{\left[\kappa_i+t_{il}\delta_i\kappa_i\right]\left[t_{il}f_i+1-f_i\bar{g}_i\right]}\right|^4\right] \qquad (6.48)$$

where we denoted $\kappa_i \triangleq 1 - c_i + \frac{1}{N}\sum_{l=1}^{n_i} \frac{1}{1+t_{il}\delta_i}$.

Again, all differences in the numerator converge to zero at a rate $O(1/N^2)$. However, the denominator presents now the term $t_{il} f_i + 1 - f_i \bar{g}_i$, which must be controlled and ensured to be away from zero. For this, we can notice that $\bar{g}_i \leq T/(1-c)$ by definition, while $f_i \leq R/|z|$, also by definition. It is therefore possible, by taking $z < 0$ sufficiently small, to ensure that $1 - f_i \bar{g}_i > 0$. We therefore from now on assume that such z are considered.

Equation (6.48) becomes in this case

$$\mathrm{E}\left[\left|\bar{g}_i - \frac{1}{N}\sum_{l=1}^{n_i} \frac{t_{il}}{t_{il} f_i + 1 - f_i \bar{g}_i}\right|^4\right] = O\left(\frac{1}{N^2}\right).$$

We are now ready to introduce the matrix $\mathbf{F}$. Consider

$$\mathbf{F} = \sum_{i=1}^{K} \bar{f}_i \mathbf{R}_i,$$

with $\bar{f}_i$ defined as the unique solution to the equation in x

$$x = \frac{1}{N}\sum_{l=1}^{n_i} \frac{t_{il}}{1 - f_i x + f_i t_{il}} \tag{6.49}$$

within the interval $0 \leq x < c_i/f_i$. To prove the uniqueness of the solution within this interval, note simply that

$$\begin{aligned} \frac{c_i}{f_i} &\geq \frac{1}{N}\sum_{l=1}^{n_i} \frac{t_{il}}{1 - f_i(c_i/f_i) + f_i t_{il}} \\ 0 &\leq \frac{1}{N}\sum_{l=1}^{n_i} \frac{t_{il}}{1 - f_i \cdot 0 + f_i t_{il}} \end{aligned}$$

and that the function $x \mapsto \frac{1}{N}\sum_{l=1}^{n_i} \frac{t_{il}}{1 - f_i x + f_i t_{il}}$ is strictly increasing. Hence the uniqueness of the solution in $[0, c_i/f_i)$. We also show that this solution is an attractor of the fixed-point algorithm, when correctly initialized. Indeed, let $x_0, x_1, \ldots$ be defined by

$$x_{n+1} = \frac{1}{N}\sum_{l=1}^{n_i} \frac{t_{il}}{1 - f_i x_n + f_i t_{il}}$$

with $x_0 \in [0, c_i/f_i)$. Then, $x_n \in [0, c_i/f_i)$ implies $1 - f_i x_n + f_i t_{il} \geq 1 - c_i + f_i t_{il} > f_i t_{il}$ and therefore $f_i x_{n+1} \leq c_i$, so $x_0, x_1, \ldots$ are all contained in $[0, c_i/f_i)$. Now observe that

$$x_{n+1} - x_n = \frac{1}{N}\sum_{l=1}^{n_i} \frac{f_i(x_n - x_{n-1})}{(1 + t_{il} f_i - f_i x_n)(1 + t_{il} f_i - f_i x_{n-1})}$$

so that the differences $x_{n+1} - x_n$ and $x_n - x_{n-1}$ have the same sign. The sequence $x_0, x_1, \ldots$ is therefore monotonic and bounded: it converges. Calling

x_∞ this limit, we have:

$$x_\infty = \frac{1}{N}\sum_{l=1}^{n_i} \frac{t_{il}}{1 + t_{il}f_i - f_i x_\infty}$$

as required.

To finally prove that $\frac{1}{N}\operatorname{tr}\mathbf{A}(\mathbf{B}_N - z\mathbf{I}_N)^{-1} - \frac{1}{N}\operatorname{tr}\mathbf{A}(\mathbf{F} - z\mathbf{I}_N)^{-1} \xrightarrow{\text{a.s.}} 0$, we want now to show that $\bar{g}_i - \bar{f}_i$ tends to zero at a sufficiently fast rate. For this, we write

$$\begin{aligned}
\mathrm{E}\left[\left|\bar{g}_i - \bar{f}_i\right|^4\right] &\leq 8\mathrm{E}\left[\left|\bar{g}_i - \frac{1}{N}\sum_{l=1}^{n_i}\frac{t_{il}}{t_{il}f_i + 1 - f_i\bar{g}_i}\right|^4\right] \\
&+ 8\mathrm{E}\left[\left|\frac{1}{N}\sum_{l=1}^{n_i}\frac{t_{il}}{t_{il}f_i + 1 - f_i\bar{g}_i} - \frac{1}{N}\sum_{l=1}^{n_i}\frac{t_{il}}{t_{il}f_i + 1 - f_i\bar{f}_i}\right|^4\right] \\
&= 8\mathrm{E}\left[\left|\bar{g}_i - \frac{1}{N}\sum_{l=1}^{n_i}\frac{t_{il}}{t_{il}f_i + 1 - f_i\bar{g}_i}\right|^4\right] \\
&+ \mathrm{E}\left[\left|\bar{g}_i - \bar{f}_i\right|^4\left|\frac{1}{N}\sum_{l=1}^{n_i}\frac{t_{il}f_i}{(t_{il}f_i + 1 - f_i\bar{f}_i)(t_{il}f_i + 1 - f_i\bar{g}_i)}\right|^4\right].
\end{aligned} \tag{6.50}$$

We only need to ensure now that the coefficient multiplying $\left|\bar{g}_i - \bar{f}_i\right|$ in the right-hand side term is uniformly smaller than one. This unfolds again from noticing that the numerator can be made very small, with the denominator kept away from zero, for sufficiently small $z < 0$. For these z, we can therefore prove that

$$\mathrm{E}\left[\left|\bar{g}_i - \bar{f}_i\right|^4\right] = O\left(\frac{1}{N^2}\right).$$

It is important to notice that this holds essentially because we took $\bar{f}_i$ to be the unique solution of (6.49) lying in the interval $[0, c_i/f_i)$. For $c_i = 1$, another solution happens to be equal to $1/f_i$, which does not satisfy this fourth moment inequality.

Finally, we can proceed to proving the deterministic equivalent relations.

$$\begin{aligned}
&\frac{1}{N}\operatorname{tr}\mathbf{A}\left(\mathbf{G} - z\mathbf{I}_N\right)^{-1} - \frac{1}{N}\operatorname{tr}\mathbf{A}\left(\mathbf{F} - z\mathbf{I}_N\right)^{-1} \\
&= \sum_{i=1}^{K}\frac{1}{N}\sum_{l=1}^{n_i} t_{il}\left[\frac{\frac{1}{N}\operatorname{tr}\mathbf{R}_i\mathbf{A}\left(\mathbf{G} - z\mathbf{I}_N\right)^{-1}\left(\mathbf{F} - z\mathbf{I}_N\right)^{-1}}{(1 - c_i + \frac{1}{N}\sum_{l'=1}^{n_i}\frac{1}{1 + t_{i,l'}\delta_i})(1 + t_{il}\delta_i)}\right. \\
&\qquad\left. - \frac{\frac{1}{N}\operatorname{tr}\mathbf{R}_i\mathbf{A}\left(\mathbf{G} - z\mathbf{I}_N\right)^{-1}\left(\mathbf{F} - z\mathbf{I}_N\right)^{-1}}{1 - f_i\bar{f}_i + t_{il}f_i}\right]
\end{aligned}$$

$$= \sum_{i=1}^{K} \frac{1}{N} \sum_{l=1}^{n_i} t_{il} \left[\left(\frac{1}{(1 - c_i + \frac{1}{N} \sum_{l'=1}^{n_i} \frac{1}{1+t_{i,l'}\delta_i})(1 + t_{il}\delta_i)} - \frac{1}{1 - f_i \bar{g}_i + t_{il} f_i} \right) \right.$$

$$\left. + \left(\frac{1}{1 - f_i \bar{g}_i + t_{il} f_i} - \frac{1}{1 - f_i \bar{f}_i + t_{il} f_i} \right) \right] \frac{1}{N} \operatorname{tr} \mathbf{R}_i \mathbf{A} \left(\mathbf{G} - z\mathbf{I}_N \right)^{-1} \left(\mathbf{F} - z\mathbf{I}_N \right)^{-1}.$$

The first difference in brackets is already known to be small from previous considerations on the relations between $\bar{g}_i$ and δ_i. As for the second difference, it also goes to zero fast as $\mathrm{E}[|\bar{g}_i - \bar{f}_i|^4]$ is summable. We therefore have

$$\mathrm{E}\left[\left| \frac{1}{N} \operatorname{tr} \mathbf{A} \left(\mathbf{G} - z\mathbf{I}_N \right)^{-1} - \frac{1}{N} \operatorname{tr} \mathbf{A} \left(\mathbf{F} - z\mathbf{I}_N \right)^{-1} \right|^4 \right] = O\left(\frac{1}{N^2} \right).$$

Together with (6.47), we finally have

$$\mathrm{E}\left[\left| \frac{1}{N} \operatorname{tr} \mathbf{A} \left(\mathbf{B}_N - z\mathbf{I}_N \right)^{-1} - \frac{1}{N} \operatorname{tr} \mathbf{A} \left(\mathbf{F} - z\mathbf{I}_N \right)^{-1} \right|^4 \right] = O\left(\frac{1}{N^2} \right).$$

Applying the Markov inequality, Theorem 3.5, and the Borel–Cantelli lemma, Theorem 3.6, this entails

$$\frac{1}{N} \operatorname{tr} \mathbf{A} \left(\mathbf{B}_N - z\mathbf{I}_N \right)^{-1} - \frac{1}{N} \operatorname{tr} \mathbf{A} \left(\mathbf{F} - z\mathbf{I}_N \right)^{-1} \xrightarrow{\text{a.s.}} 0 \tag{6.51}$$

as N grows large. This holds however to this point for a restricted set of negative z. In order to extend the convergence to all $\mathbb{R}^-$, we need to show that $\bar{f}_i(z)$ can be extended to an holomorphic function in a neighborhood of any segment of $\mathbb{R}^-$. Proving the holomorphic extension is performed in [Couillet *et al.*, 2011b] by showing that, for z in the cone $\mathcal{D} = \{z \in \mathbb{C},\ \Re[z] < 0, |\Im[z]| < \Re[z](1 - c_i)/c_i\}$, the algorithm defining $\bar{f}_i(z)$ keeps its successive iterations holomorphic and bounded. Since these iterations are known to converge for $z \in \mathbb{R}^-$, the Vitali convergence theorem, Theorem 3.11, ensures that the algorithm converges to a holomorphic function on all $\mathcal{D}$. This holomorphic function obviously extends $\bar{f}_i(z)$ to $\mathcal{D}$, and we have the desired result.

The proof of the above statement is very technical and not of interest for this book. Nonetheless, it is important to stress the fact that $\mathcal{D}$ is the "best" region where the successive iterations defining $\bar{f}_i(z)$ are maintained bounded, due to the instability of the denominators defining the algorithm. Simulations suggest that, outside $\mathcal{D}$, the algorithm sometimes does no no longer converge. It therefore remains an open problem to define an holomorphic extension outside $\mathcal{D}$, and this is the main reason why we cannot possibly state a result on the convergence in law of $F^{\mathbf{B}_N}$.

Applying the result for $\mathbf{A} = \mathbf{R}_j$, this is in particular

$$f_j - \frac{1}{N} \operatorname{tr} \mathbf{R}_j \left(\sum_{i=1}^{K} \bar{f}_i \mathbf{R}_i - z\mathbf{I}_N \right)^{-1} \xrightarrow{\text{a.s.}} 0$$

where we recall that $\bar{f}_i$ is the unique solution to

$$x = \frac{1}{N} \sum_{i=1}^{n_i} \frac{t_{il}}{1 - f_i x + t_{il} f_i}$$

within the set $[0, c_i/f_i)$.

Step 2: Existence and uniqueness

The existence step unfolds similarly as in the proof of Theorem 6.1. It suffices to consider the matrices $\mathbf{T}_{[p],i} \in \mathbb{C}^{n_i p}$ and $\mathbf{R}_{[p],i} \in \mathbb{C}^{Np}$ for all i defined as the Kronecker products $\mathbf{T}_{[p],i} \triangleq \mathbf{T}_i \otimes \mathbf{I}_p$, $\mathbf{R}_{[p],i} \triangleq \mathbf{R}_i \otimes \mathbf{I}_p$, which have, respectively, the d.f. $F^{\mathbf{T}_i}$ and $F^{\mathbf{R}_i}$ for all p. Similar to the i.i.d. case, it is easy to see that e_i is unchanged by substituting the $\mathbf{T}_{[p],i}$ and $\mathbf{R}_{[p],i}$ to the $\mathbf{T}_i$ and $\mathbf{R}_i$, respectively. Denoting in the same way $f_{[p],i}$ the equivalent of f_i for $\mathbf{T}_{[p],i}$ and $\mathbf{R}_{[p],i}$, from the convergence result of Step 1, we can choose $f_{[1],i}, f_{[2],i}, \ldots$ a sequence of the set of probability one where convergence is ensured as p grows large (N and the n_i are kept fixed). The limit over this sequence satisfies the fixed-point equation, which therefore proves existence. It is easy to see that the limit is also the Stieltjes transform of a finite measure on $\mathbb{R}^+$ by verifying the conditions of Theorem 3.2.

We will prove uniqueness of positive solutions $e_1, \ldots, e_K > 0$ for $z < 0$ and the convergence of the classical fixed-point algorithm to these values. We first introduce some notations and useful identities. Note that, similar to Step 1 with the δ_i terms, we can define, for any pair of variables x_i and $\bar{x}_i$, with $\bar{x}_i$ defined as the solution y to $y = \frac{1}{N} \sum_{l=1}^{n_i} \frac{t_{il}}{1 + x_j t_{il} - x_j y}$ such that $0 \leq y < c_j/x_j$, the auxiliary variables $\Delta_1, \ldots, \Delta_K$, with the properties

$$x_i = \Delta_i \left(1 - c_i + \frac{1}{N} \sum_{l=1}^{n_i} \frac{1}{1 + t_{il}\Delta_i}\right) = \Delta_i \left(1 - \frac{1}{N} \sum_{l=1}^{n_i} \frac{t_{il}\Delta_i}{1 + t_{il}\Delta_i}\right)$$

and

$$1 - x_i \bar{x}_i = 1 - c_i + \frac{1}{N} \sum_{l=1}^{n_i} \frac{1}{1 + t_{il}\Delta_i} = 1 - \frac{1}{N} \sum_{l=1}^{n_i} \frac{t_{il}\Delta_i}{1 + t_{il}\Delta_i}.$$

The uniqueness of the mapping between the x_i and Δ_i can be proved. In fact, it turns out that Δ_i is a monotonically increasing function of x_i with $\Delta_i = 0$ for $x_i = 0$. We take the opportunity of the above definitions to notice that, for $x_i > x_i'$ and $\bar{x}_i'$, Δ_i' defined similarly as $\bar{x}_i$ and Δ_i

$$x_i \bar{x}_i - x_i' \bar{x}_i' = \frac{1}{N} \sum_{l=1}^{n_i} \frac{t_{il}(\Delta_i - \Delta_i')}{(1 + t_{il}\Delta_i)(1 + t_{il}\Delta_i')} > 0 \qquad (6.52)$$

whenever $\mathbf{T}_i \neq 0$. Therefore $x_i \bar{x}_i$ is a growing function of x_i (or equivalently of Δ_i). This will turn out a useful remark later.

We are now in position to prove the step of uniqueness. Define, for $i \in \{1, \ldots, K\}$, the functions

$$h_i : (x_1, \ldots, x_K) \mapsto \frac{1}{N} \operatorname{tr} \mathbf{R}_i \left(\sum_{j=1}^{K} \bar{x}_j \mathbf{R}_j - z\mathbf{I}_N \right)^{-1}$$

with $\bar{x}_j$ the unique solution of the equation in y

$$y = \frac{1}{N} \sum_{l=1}^{n_j} \frac{t_{jl}}{1 + x_j t_{jl} - x_j y} \tag{6.53}$$

such that $0 \leq y \leq c_j / x_j$.

We will prove in the following that the multivariate function $\mathbf{h} = (h_1, \ldots, h_K)$ is a *standard function*, defined in [Yates, 1995], as follows.

Definition 6.2. A function $\mathbf{h}(x_1, \ldots, x_K) \in \mathbb{R}^K$, $\mathbf{h} = (h_1, \ldots, h_K)$, with arguments $x_1, \ldots, x_K \in \mathbb{R}^+$, is said to be a standard function or a standard interference function if it fulfills the following conditions

1. *Positivity:* for all j and all $x_1, \ldots, x_K \geq 0$, $h_j(x_1, \ldots, x_K) > 0$,
2. *Monotonicity:* if $x_1 \geq x'_1, \ldots, x_K \geq x'_K$, then $h_j(x_1, \ldots, x_K) \geq h_j(x'_1, \ldots, x'_K)$, for all j,
3. *Scalability:* for all $\alpha > 1$ and j, $\alpha h_j(x_1, \ldots, x_K) > h_j(\alpha x_1, \ldots, \alpha x_K)$.

The important result regarding standard functions [Yates, 1995] is given as follows.

Theorem 6.18. *If a K-variate function $\mathbf{h}(x_1, \ldots, x_K)$ is standard and there exists $(x_1, \ldots, x_K)$ such that, for all j, $x_j \geq h_j(x_1, \ldots, x_K)$, then the fixed-point algorithm that consists in setting*

$$x_j^{(t+1)} = h_j(x_1^{(t)}, \ldots, x_K^{(t)})$$

for $t \geq 1$ and for any initial values $x_1^{(0)}, \ldots, x_K^{(0)} \geq 0$ converges to the unique (nonnegative) solution of the system of K equations

$$x_j = h_j(x_1, \ldots, x_K)$$

with $j \in \{1, \ldots, K\}$.

Proof. The existence is based on the monotonicity condition and the boundedness $x_j \geq h_j(x_1, \ldots, x_K)$ for some $x_1, \ldots, x_K$. It is easily shown that, starting from $x_1^{(0)} = \ldots = x_K^{(0)} = 0$, the sequences of $x_i^{(l)}$ is monotonous. From the boundedness, it results that the sequence converges. The proof of uniqueness unfolds easily from the standard function assumptions. Take $(x_1, \ldots, x_K)$ and $(x'_1, \ldots, x'_K)$ two sets of supposedly distinct all positive solutions. Then there exists j such that $x_j < x'_j$, $\alpha x_j = x'_j$, and $\alpha x_i \geq x'_i$ for $i \neq j$. From monotonicity and scalability, it follows that

$$x'_j = h_j(x'_1, \ldots, x'_K) \leq h_j(\alpha x_1, \ldots, \alpha x_K) < \alpha h_j(x_1, \ldots, x_K) = \alpha x_j$$

a contradiction. The convergence of the fixed-point algorithm from any point $(x_1, \ldots, x_K)$ unfolds from similar arguments, see [Yates, 1995] for more details. $\square$

Therefore, by showing that $\mathbf{h} \triangleq (h_1, \ldots, h_K)$ is standard, we will prove that the classical fixed-point algorithm converges to the unique set of positive solutions $e_1, \ldots, e_K$, when $z < 0$.

The positivity condition is straightforward as $\bar{x}_i$ is positive for x_i positive and therefore $h_j(x_1, \ldots, x_K)$ is always positive whenever $x_1, \ldots, x_K$ are.

The scalability is also rather direct. Let $\alpha > 1$, then:

$$\begin{aligned}
&\alpha h_j(x_1, \ldots, x_K) - h_j(\alpha x_1, \ldots, \alpha x_K) \\
&= \frac{1}{N} \operatorname{tr} \mathbf{R}_j \left(\sum_{k=1}^{K} \frac{\bar{x}_k}{\alpha} \mathbf{R}_k - \frac{z}{\alpha} \mathbf{I}_N \right)^{-1} - \frac{1}{N} \operatorname{tr} \mathbf{R}_j \left(\sum_{k=1}^{K} \bar{x}_k^{(a)} \mathbf{R}_k - z \mathbf{I}_N \right)^{-1}
\end{aligned}$$

where we denoted $\bar{x}_j^{(a)}$ the unique solution to (6.53) with x_j replaced by αx_j, within the set $[0, c_j/(\alpha x_j))$. Since $\alpha x_i > x_i$, from the property (6.52), we have $\alpha x_k \bar{x}_k^{(\alpha)} > x_k \bar{x}_k$ or equivalently $\bar{x}_k^{(a)} - \frac{\bar{x}_k}{\alpha} > 0$. We now define the two matrices $\mathbf{A} \triangleq \sum_{k=1}^{k} \frac{\bar{x}_k}{\alpha} \mathbf{R}_k - \frac{z}{\alpha} \mathbf{I}_N$ and $\mathbf{A}^{(\alpha)} \triangleq \sum_{k=1}^{k} \bar{x}_k^{(\alpha)} \mathbf{R}_k - z \mathbf{I}_N$. For any vector $\mathbf{a} \in \mathbb{C}^N$

$$\mathbf{a}^{\mathsf{H}} \left(\mathbf{A} - \mathbf{A}^{(\alpha)} \right) \mathbf{a} = \sum_{k=1}^{K} \left(\frac{\bar{x}_k}{\alpha} - \bar{x}_k^{(\alpha)} \right) \mathbf{a}^{\mathsf{H}} \mathbf{R}_k \mathbf{a} + z \left(1 - \frac{1}{\alpha} \right) \mathbf{a}^{\mathsf{H}} \mathbf{a} \leq 0$$

since $z < 0$, $1 - \frac{1}{\alpha} > 0$ and $\frac{\bar{x}_k}{\alpha} - \bar{x}_k^{(\alpha)} < 0$. Therefore $\mathbf{A} - \mathbf{A}^{(\alpha)}$ is non-positive definite. Now, from [Horn and Johnson, 1985, Corollary 7.7.4], this implies that $\mathbf{A}^{-1} - (\mathbf{A}^{(\alpha)})^{-1}$ is non-negative definite. Writing

$$\frac{1}{N} \operatorname{tr} \mathbf{R}_j \left(\mathbf{A}^{-1} - (\mathbf{A}^{(\alpha)})^{-1} \right) = \frac{1}{N} \sum_{i=1}^{N} \mathbf{r}_{j,i}^{\mathsf{H}} \left(\mathbf{A}^{-1} - (\mathbf{A}^{(\alpha)})^{-1} \right) \mathbf{r}_{j,i}$$

with $\mathbf{r}_{j,i}$ the ith column of $\mathbf{R}_j$, this ensures $\alpha h_j(x_1, \ldots, x_K) > h_j(\alpha x_1, \ldots, \alpha x_K)$.

The monotonicity requires some more lines of calculus. This unfolds from considering $\bar{x}_i$ as a function of Δ_i, by verifying that $\frac{d}{d\Delta_i} \bar{x}_i$ is negative.

$$\begin{aligned}
\frac{d}{d\Delta_i} \bar{x}_i &= \frac{1}{\Delta_i^2} \left(1 - \frac{1}{1 - \frac{1}{N} \sum_{l=1}^{n_i} \frac{t_{il}\Delta_i}{1+t_{il}\Delta_i}} \right) + \frac{1}{\Delta_i^2} \left(\frac{\frac{1}{N} \sum_{l=1}^{n_i} \frac{t_{il}\Delta_i}{(1+t_{il}\Delta_i)^2}}{\left(1 - \frac{1}{N} \sum_{l=1}^{n_i} \frac{t_{il}\Delta_i}{1+t_{il}\Delta_i} \right)^2} \right) \\
&= \frac{-\frac{1}{N} \sum_{l=1}^{n_i} \frac{t_{il}\Delta_i}{1+t_{il}\Delta_i} \left(1 - \frac{1}{N} \sum_{l=1}^{n_i} \frac{t_{il}\Delta_i}{1+t_{il}\Delta_i} \right) + \frac{1}{N} \sum_{l=1}^{n_i} \frac{t_{il}\Delta_i}{(1+t_{il}\Delta_i)^2}}{\Delta_i^2 \left(1 - \frac{1}{N} \sum_{l=1}^{n_i} \frac{t_{il}\Delta_i}{1+t_{il}\Delta_i} \right)^2} \\
&= \frac{\left(\frac{1}{N} \sum_{l=1}^{n_i} \frac{t_{il}\Delta_i}{1+t_{il}\Delta_i} \right)^2 - \frac{1}{N} \sum_{l=1}^{n_i} \frac{t_{il}\Delta_i}{1+t_{il}\Delta_i} + \frac{1}{N} \sum_{l=1}^{n_i} \frac{t_{il}\Delta_i}{(1+t_{il}\Delta_i)^2}}{\Delta_i^2 \left(1 - \frac{1}{N} \sum_{l=1}^{n_i} \frac{t_{il}\Delta_i}{1+t_{il}\Delta_i} \right)^2}
\end{aligned}$$

$$= \frac{\left(\frac{1}{N}\sum_{l=1}^{n_i}\frac{t_{il}\Delta_i}{1+t_{il}\Delta_i}\right)^2 - \frac{1}{N}\sum_{l=1}^{n_i}\frac{(t_{il}\Delta_i)^2}{(1+t_{il}\Delta_i)^2}}{\Delta_i^2\left(1-\frac{1}{N}\sum_{l=1}^{n_i}\frac{t_{il}\Delta_i}{1+t_{il}\Delta_i}\right)^2}.$$

From the Cauchy–Schwarz inequality, we have:

$$\begin{aligned}\left(\sum_{l=1}^{n_i}\frac{1}{N}\frac{t_{il}\Delta_i}{1+t_{il}\Delta_i}\right)^2 &\leq \sum_{l=1}^{n_i}\frac{1}{N^2}\sum_{l=1}^{n_i}\frac{(t_{il}\Delta_i)^2}{(1+t_{il}\Delta_i)^2}\\ &= c_i\frac{1}{N}\sum_{l=1}^{n_i}\frac{(t_{il}\Delta_i)^2}{(1+t_{il}\Delta_i)^2}\\ &< \frac{1}{N}\sum_{l=1}^{n_i}\frac{(t_{il}\Delta_i)^2}{(1+t_{il}\Delta_i)^2}\end{aligned}$$

which is sufficient to conclude that $\frac{d}{d\Delta_i}\bar{x}_i < 0$. Since Δ_i is an increasing function of x_i, we have that $\bar{x}_i$ is a decreasing function of x_i, i.e. $\frac{d}{dx_i}\bar{x}_i < 0$. This being said, using the same line of reasoning as for scalability, we finally have that, for two sets $x_1, \ldots, x_K$ and $x'_1, \ldots, x'_K$ of positive values such that $x_j > x'_j$

$$\begin{aligned}&h_j(x_1,\ldots,x_K) - h(x'_1,\ldots,x'_K)\\ &= \frac{1}{N}\operatorname{tr}\mathbf{R}_j\left[\left(\sum_{k=1}^{K}\bar{x}_k\mathbf{R}_k - z\mathbf{I}_N\right)^{-1} - \left(\sum_{k=1}^{K}\bar{x}'_k\mathbf{R}_k - z\mathbf{I}_N\right)^{-1}\right] > 0\end{aligned}$$

with $\bar{x}'_j$ defined equivalently as $\bar{x}_j$, and where the terms $(\bar{x}'_k - \bar{x}_k)$ are all positive due to negativity of $\frac{d}{dx_i}\bar{x}_i$. This proves the monotonicity condition.

We finally have from Theorem 6.18 that $(e_1, \ldots, e_K)$ is uniquely defined and that the classical fixed-point algorithm converges to this solution from any initialization point (remember that, at each step of the algorithm, the set $\bar{e}_1, \ldots, \bar{e}_K$ must be evaluated, possibly thanks to a further fixed-point algorithm).

We finally complete the proof by showing that the stochastic f_i and the deterministic e_i are asymptotically close to one another as N grows large.

Step 3: Convergence of $e_i - f_i$

For this step, we follow the approach in [Hachem *et al.*, 2007]. Denote

$$\varepsilon_N^i \triangleq f_i - \frac{1}{N}\operatorname{tr}\mathbf{R}_i\left(\sum_{k=1}^{K}\bar{f}_k\mathbf{R}_k - z\mathbf{I}_N\right)^{-1}$$

and recall the definitions of f_i, e_i, $\bar{f}_i$ and $\bar{e}_i$:

$$\begin{aligned}f_i &= \frac{1}{N}\operatorname{tr}\mathbf{R}_i\left(\mathbf{B}_N - z\mathbf{I}_N\right)^{-1}\\ e_i &= \frac{1}{N}\operatorname{tr}\mathbf{R}_i\left(\sum_{j-1}^{K}\bar{e}_j\mathbf{R}_j - z\mathbf{I}_N\right)^{-1}\end{aligned}$$

$$\bar{f}_i = \frac{1}{N}\sum_{l=1}^{n_i} \frac{t_{i,l}}{1 - f_i\bar{f}_i + t_{i,l}f_i}, \qquad \bar{f}_i \in [0, c_i/f_i]$$
$$\bar{e}_i = \frac{1}{N}\sum_{l=1}^{n_i} \frac{t_{i,l}}{1 - e_i\bar{e}_i + t_{i,l}e_i}, \qquad \bar{e}_i \in [0, c_i/e_i].$$

From the definitions above, we have the following set of inequalities

$$f_i \leq \frac{R}{|z|}, \quad e_i \leq \frac{R}{|z|}, \quad \bar{f}_i \leq \frac{T}{1-c_i}, \quad \bar{e}_i \leq \frac{T}{1-c_i}. \tag{6.54}$$

We will show in the sequel that

$$e_i - f_i \xrightarrow{\text{a.s.}} 0 \tag{6.55}$$

for all $i \in \{1, \ldots, N\}$. Write the following differences

$$f_i - e_i = \sum_{j=1}^{K}(\bar{e}_j - \bar{f}_j)\frac{1}{N}\operatorname{tr}\mathbf{R}_i\left[\sum_{k=1}^{K}\bar{e}_k\mathbf{R}_k - z\mathbf{I}_N\right]^{-1}\mathbf{R}_j\left[\sum_{k=1}^{K}\bar{f}_k\mathbf{R}_k - z\mathbf{I}_N\right]^{-1} + \varepsilon_N^i$$
$$\bar{e}_i - \bar{f}_i = \frac{1}{N}\sum_{l=1}^{n_i}\frac{t_{i,l}^2(f_i - e_i) - t_{i,l}\left[f_i\bar{f}_i - e_i\bar{e}_i\right]}{(1 + t_{i,l}e_i - \bar{e}_ie_i)(1 + t_{i,l}f_i - \bar{f}_if_i)}$$

and

$$f_i\bar{f}_i - e_i\bar{e}_i = \bar{f}_i(f_i - e_i) + e_i(\bar{f}_i - \bar{e}_i).$$

For notational convenience, we define the following values

$$\alpha \triangleq \sup_i \mathrm{E}\left[|f_i - e_i|^4\right]$$
$$\bar{\alpha} \triangleq \sup_i \mathrm{E}\left[|\bar{f}_i - \bar{e}_i|^4\right].$$

It is thus sufficient to show that α is summable to prove (6.55). By applying (6.54) to the absolute of the first difference, we obtain

$$|f_i - e_i| \leq \frac{KR^2}{|z|^2}\sup_i|\bar{f}_i - \bar{e}_i| + \sup_i|\varepsilon_N^i|$$

and hence

$$\alpha \leq \frac{8K^4R^8}{|z|^8}\bar{\alpha} + \frac{8C}{N^2} \tag{6.56}$$

for some constant $C > 0$ such that $\mathrm{E}[|\sup_i \varepsilon_N^i|^4] \leq C/N^2$. This is possible since $\mathrm{E}[|\sup_i \varepsilon_N^i|^4] \leq 8K\sup_i \mathrm{E}[|\varepsilon_N^i|^4]$ and $\mathrm{E}[|\varepsilon_N^i|^4]$ has been proved to be of order $O(1/N^2)$. Similarly, we have for the third difference

$$|f_i\bar{f}_i - e_i\bar{e}_i| \leq |\bar{f}_i||f_i - e_i| + |e_i||\bar{f}_i - \bar{e}_i|$$
$$\leq \frac{T}{1-c}\sup_i|f_i - e_i| + \frac{R}{|z|}\sup_i|\bar{f}_i - \bar{e}_i|$$

with c an upper bound on $\max_i \limsup_n c_i$, known to be inferior to one. This result can be used to upper bound the second difference term, which writes

$$\begin{aligned}
|\bar{f}_i - \bar{e}_i| &\leq \frac{1}{(1-c)^2}\left(T^2 \sup_i |f_i - e_i| + T|f_i\bar{f}_i - e_i\bar{e}_i|\right) \\
&\leq \frac{1}{(1-c)^2}\left(T^2 \sup_i |f_i - e_i| + T\left[\frac{T}{1-c}\sup_i |f_i - e_i| + \frac{R}{|z|}\sup_i |\bar{f}_i - \bar{e}_i|\right]\right) \\
&= \frac{T^2(2-c)}{(1-c)^3}\sup_i |f_i - e_i| + \frac{RT}{|z|(1-c)^2}\sup_i |\bar{f}_i - \bar{e}_i|.
\end{aligned}$$

Hence

$$\bar{\alpha} \leq \frac{8T^8(2-c)^4}{(1-c)^{12}}\alpha + \frac{8R^4T^4}{|z|^4(1-c)^8}\bar{\alpha}. \tag{6.57}$$

For a suitable z, satisfying $|z| > \frac{2RT}{(1-c)^2}$, we have $\frac{8R^4T^4}{|z|^4(1-c)^8} < 1/2$ and, thus, moving all terms proportional to α on the left

$$\bar{\alpha} < \frac{16T^8(2-c)^4}{(1-c)^{12}}\alpha.$$

Plugging this result into (6.56) yields

$$\alpha \leq \frac{128K^4R^8T^8(2-c)^4}{|z|^8(1-c)^{12}}\alpha + \frac{8C}{N^2}.$$

Take $0 < \varepsilon < 1$. It is easy to check that, for $|z| > \frac{128^{1/8}RT\sqrt{K(2-c)}}{(1-c)^{3/2}(1-\varepsilon)^{1/8}}$, $\frac{128K^4R^8T^8(2-c)^4}{|z|^8(1-c)^{12}} < 1-\varepsilon$ and thus

$$\alpha < \frac{8C}{\varepsilon N^2}. \tag{6.58}$$

Since C does not depend on N, α is clearly summable which, along with the Markov inequality and the Borel–Cantelli lemma, concludes the proof, for these small values of z. To extend the proof to all $z < 0$, we need to prove that $e_i(z)$ can be extended to an holomorphic function in a neighborhood of $\mathbb{R}^-$. For this, we use the fact that $f_{[p],i}$ defined as f_i but with $\mathbf{T}_i$ and $\mathbf{H}_i$ replaced by $\mathbf{T}_{[p],i} = \mathbf{T}_i \otimes \mathbf{I}_p$ and $\mathbf{H}_{[p],i} = \mathbf{H}_i \otimes \mathbf{I}_p$, respectively, $f_{[p],i}(z)$ converges for $z \in \mathcal{D}$ almost surely. Taking a sequence from this probability one space, since the convergence is towards $e_i(z)$ for $z < 0$, it entails from Vitali's convergence theorem that $e_i(z)$ can be extended to an holomorphic function on $\mathcal{D}$. Therefore the convergence $f_i - e_i \xrightarrow{\text{a.s.}} 0$ is valid for all $z < 0$. This is our final result. □

As a (not immediate) corollary of the proof above, we have the following result, important for application purposes, see Section 12.2.

Theorem 6.19. *Under the assumptions of Theorem 6.17 with* $\mathbf{T}_i$ *diagonal for all* i*, denoting* $\mathbf{w}_{ij}$ *the* jth *column of* $\mathbf{W}_i$*,* t_{ij} *the* jth *diagonal entry of* $\mathbf{T}_i$*, and*

$z \in \mathbb{C} \setminus \mathbb{R}^+$

$$\mathbf{w}_{ij}^{\mathsf{H}}\mathbf{H}_i^{\mathsf{H}}\left(\mathbf{B}_N - t_{ij}\mathbf{H}_i\mathbf{w}_{ij}\mathbf{w}_{ij}^{\mathsf{H}}\mathbf{H}_i^{\mathsf{H}} - z\mathbf{I}_N\right)^{-1}\mathbf{H}_i\mathbf{w}_{ij} - \frac{e_i(z)}{\bar{c}_i - e_i(z)\bar{e}_i(z)} \xrightarrow{\text{a.s.}} 0. \quad (6.59)$$

where $e_i(z)$ *and* $\bar{e}_i(z)$ *are defined in Theorem 6.17.*

Similar to the i.i.d. case, a deterministic equivalent for the Shannon transform can be derived. This is given by the following proposition.

Theorem 6.20. *Under the assumptions of Theorem 6.17 with* $z = -1/x$*, for* $x > 0$*, denoting*

$$\mathcal{V}_{\mathbf{B}_N}(x) = \frac{1}{N}\log\det\left(x\mathbf{B}_N + \mathbf{I}_N\right)$$

the Shannon transform of $\mathbf{B}_N$*, we have:*

$$\mathcal{V}_{\mathbf{B}_N}(x) - \mathcal{V}_N(x) \xrightarrow{\text{a.s.}} 0$$

where

$$\mathcal{V}_N(x) = \frac{1}{N}\log\det\left(\mathbf{I}_N + x\sum_{i=1}^K \bar{e}_i\mathbf{R}_i\right) + \sum_{i=1}^K \frac{1}{N}\log\det\left([\bar{c}_i - e_i\bar{e}_i]\mathbf{I}_{n_i} + e_i\mathbf{T}_i\right)$$
$$+ \sum_{i=1}^K \left[(1-c_i)\log(\bar{c}_i - e_i\bar{e}_i) - \bar{c}_i\log(\bar{c}_i)\right]. \quad (6.60)$$

The proof for the deterministic equivalent of the Shannon transform follows from similar considerations as for the i.i.d. case, see Theorem 6.4 and Corollary 6.1, and is detailed below.

Proof. For the proof of Theorem 6.20, we again take $\bar{c}_i = 1$, $\mathbf{R}_i$ deterministic of bounded spectral norm for simplicity. First note that the system of Equations (6.39) is unchanged if we extend the $\mathbf{T}_i$ matrices into $N \times N$ diagonal matrices filled with $N - n_i$ zero eigenvalues. Therefore, we can assume $c_i = 1$ and that all $\mathbf{T}_i$ have size $N \times N$, although we restrict the $F^{\mathbf{T}_i}$ to have a mass in zero. Since this does not alter the Equations (6.39), we have in particular $\bar{e}_i < 1/e_i$. This being said, (6.60) now needs to be rewritten

$$\mathcal{V}_N(x) = \frac{1}{N}\log\det\left(\mathbf{I}_N + x\sum_{i=1}^K \bar{e}_i\mathbf{R}_i\right) + \sum_{i=1}^K \frac{1}{N}\log\det\left([1 - e_i\bar{e}_i]\mathbf{I}_N + e_i\mathbf{T}_i\right).$$

Calling V the function

$$V : (x_1, \ldots, x_K, \bar{x}_1, \ldots, \bar{x}_K, x) \mapsto \frac{1}{N}\log\det\left(\mathbf{I}_N + x\sum_{i=1}^K \bar{x}_i\mathbf{R}_i\right)$$
$$+ \sum_{i=1}^K \frac{1}{N}\log\det\left([1 - x_i\bar{x}_i]\mathbf{I}_N + x_i\mathbf{T}_i\right)$$

we have:

$$\frac{\partial V}{\partial x_i}(e_1,\ldots,e_K,\bar{e}_1,\ldots,\bar{e}_K,x) = \bar{e}_i - \bar{e}_i\frac{1}{N}\sum_{l=1}^{N}\frac{1}{1-e_i\bar{e}_i+e_it_{il}}$$
$$\frac{\partial V}{\partial \bar{x}_i}(e_1,\ldots,e_K,\bar{e}_1,\ldots,\bar{e}_K,x) = e_i - e_i\frac{1}{N}\sum_{l=1}^{N}\frac{1}{1-e_i\bar{e}_i+e_it_{il}}.$$

Noticing now that

$$1 = \frac{1}{N}\sum_{l=1}^{N}\frac{1-e_i\bar{e}_i+e_it_{il}}{1-e_i\bar{e}_i+e_it_{il}} = (1-e_i\bar{e}_i)\frac{1}{N}\sum_{l=1}^{N}\frac{1}{1-e_i\bar{e}_i+e_it_{il}} + e_i\bar{e}_i$$

we have:

$$(1-e_i\bar{e}_i)\left(1-\frac{1}{N}\sum_{l=1}^{N}\frac{1}{1-e_i\bar{e}_i+e_it_{il}}\right) = 0.$$

But we also know that $0 \leq \bar{e}_i < 1/e_i$ and therefore $1 - e_i\bar{e}_i > 0$. This entails

$$\frac{1}{N}\sum_{l=1}^{N}\frac{1}{1-e_i\bar{e}_i+e_it_{il}} = 1. \tag{6.61}$$

From (6.61), we conclude that

$$\frac{\partial V}{\partial x_i}(e_1,\ldots,e_K,\bar{e}_1,\ldots,\bar{e}_K,x) = 0$$
$$\frac{\partial V}{\partial \bar{x}_i}(e_1,\ldots,e_K,\bar{e}_1,\ldots,\bar{e}_K,x) = 0.$$

We therefore have that

$$\begin{aligned}
\frac{d}{dx}\mathcal{V}_N(x) &= \sum_{i=1}^{K}\left[\frac{\partial V}{\partial e_i}\frac{\partial e_i}{\partial x} + \frac{\partial V}{\partial \bar{e}_i}\frac{\partial \bar{e}_i}{\partial x}\right] + \frac{\partial V}{\partial x} \\
&= \frac{\partial V}{\partial x} \\
&= \sum_{i=1}^{K}\bar{e}_i\frac{1}{N}\operatorname{tr}\mathbf{R}_i\left(\mathbf{I}_N + x\sum_{j=1}^{K}\bar{e}_j\mathbf{R}_j\right)^{-1} \\
&= \frac{1}{x} - \frac{1}{x^2}\frac{1}{N}\operatorname{tr}\left(\frac{1}{x}\mathbf{I}_N + \sum_{j=1}^{K}\bar{e}_j\mathbf{R}_j\right)^{-1}.
\end{aligned}$$

Therefore, along with the fact that $\mathcal{V}_N(0) = 0$, we have:

$$\mathcal{V}_N(x) = \int_0^x \left(\frac{1}{t} - \frac{1}{t^2}m_N\left(-\frac{1}{t}\right)\right)dt$$

and therefore $\mathcal{V}_N(x)$ is the Shannon transform of F_N, according to Definition 3.2.

In order to prove the almost sure convergence $\mathcal{V}_{\mathbf{B}_N}(x) - \mathcal{V}_N(x) \xrightarrow{\text{a.s.}} 0$, we need simply to notice that the support of the eigenvalues of $\mathbf{B}_N$ is bounded. Indeed, the non-zero eigenvalues of $\mathbf{W}_i\mathbf{W}_i^{\mathsf{H}}$ have unit modulus and therefore $\|\mathbf{B}_N\| \leq KTR$. Similarly, the support of F_N is the support of the eigenvalues of $\sum_{i=1}^{K} \bar{e}_i\mathbf{R}_i$, which are bounded by KTR as well.

As a consequence, for $\mathbf{B}_1, \mathbf{B}_2, \ldots$ a realization for which $F^{\mathbf{B}_N} - F_N \Rightarrow 0$, we have, from the dominated convergence theorem, Theorem 6.3

$$\int_0^\infty \log(1 + xt)\, d[F^{\mathbf{B}_N} - F_N](t) \to 0.$$

Hence the almost sure convergence. □

Applications of the above results are found in various telecommunication systems employing random isometric precoders, such as random CDMA, SDMA [Couillet *et al.*, 2011b]. A specific application to assess the optimal number of stream transmissions in multi-antenna interference channels is in particular provided in [Hoydis *et al.*, 2011a], where an extension of Theorem 6.17 to correlated i.i.d. channel matrices $\mathbf{H}_i$ is provided. It is worth mentioning that the approach followed in [Hoydis *et al.*, 2011a] to prove this extension relies on an "inclusion" of the deterministic equivalent of Theorem 6.12 into the deterministic equivalent of Theorem 6.17. The final result takes a surprisingly simple expression and the proof of existence, uniqueness, and convergence of the implicit equations obtained do not require much effort. This "deterministic equivalent of a deterministic equivalent" approach is very natural and is expected to lead to very simple results even for intricate communication models; recall e.g. Theorem 6.9.

We conclude this chapter on deterministic equivalents by a central limit theorem for the Shannon transform of the non-centered random matrix with variance profile of Theorem 6.14.

6.3 A central limit theorem

Central limit theorems are also demanded for more general models than the sample covariance matrix of Theorem 3.17. In wireless communications, it is particularly interesting to study the limiting distribution of the Shannon transform of doubly correlated random matrices, e.g. to mimic Kronecker models, or even more generally matrices of i.i.d. entries with a variance profile. Indeed, the later allows us to study, in addition to the large dimensional ergodic capacity of Rician MIMO channels, as provided by Theorem 6.14, the large dimensional *outage* mutual information of such channels. In [Hachem *et al.*, 2008b], Hachem *et al.* provide the central limit theorem for the Shannon transform of this model.

Theorem 6.21 ([Hachem *et al.*, 2008b]). *Let* $\mathbf{Y}_N$ *be* $N \times n$ *whose* (i,j)th *entry is given by:*

$$Y_{N,ij} = \frac{\sigma_{ij}(n)}{\sqrt{n}} X_{N,ij}$$

with $\{\sigma_{ij}(n)\}_{ij}$ *uniformly bounded with respect to* n, *and* $X_{N,ij}$ *is the* (i,j)th *entry of an* $N \times n$ *matrix* $\mathbf{X}_N$ *with i.i.d. entries of zero mean, unit variance, and finite eighth order moment. Denote* $\mathbf{B}_N = \mathbf{Y}_N \mathbf{Y}_N^{\mathsf{H}}$. *We then have, as* $N, n \to \infty$ *with limit ratio* $c = \lim_N N/n$, *that the Shannon transform*

$$\mathcal{V}_{\mathbf{B}_N}(x) \triangleq \frac{1}{N} \log \det(\mathbf{I}_N + x\mathbf{B}_N)$$

of $\mathbf{B}_N$ *satisfies*

$$\frac{N}{\theta_n}(\mathcal{V}_{\mathbf{B}_N}(x) - \mathrm{E}[\mathcal{V}_{\mathbf{B}_N}(x)]) \Rightarrow X \sim \mathcal{N}(0,1)$$

with

$$\theta_n^2 = -\log\det(\mathbf{I}_n - \mathbf{J}_n) + \kappa \operatorname{tr}(\mathbf{J}_n)$$

$\kappa = \mathrm{E}[(X_{N,11})^4] - 3\mathrm{E}[(X_{N,11})^3]$ *for real* $X_{N,11}$, $\kappa = \mathrm{E}[|X_{N,11}|^4] - 2\mathrm{E}[|X_{N,11}|^2]$ *for complex* $X_{N,11}$, *and* $\mathbf{J}_n$ *the matrix with* (i,j)th *entry*

$$J_{n,ij} = \frac{1}{n} \frac{\frac{1}{n}\sum_{k=1}^N \sigma_{ki}^2(n)\sigma_{kj}^2(n) t_k(-1/x)^2}{\left(1 + \frac{1}{n}\sum_{k=1}^N \sigma_{ki}^2(n) t_k(-1/x)\right)^2}$$

with $t_i(z)$ *such that* $(t_1(z), \ldots, t_N(z))$ *is the unique Stieltjes transform vector solution of*

$$t_i(z) = \left(-z + \frac{1}{n}\sum_{j=1}^n \frac{\sigma_{ij}^2(n)}{1 + \frac{1}{n}\sum_{l=1}^N \sigma_{lj}^2(n) t_l(z)} \right)^{-1}.$$

Observe that the matrix $\mathbf{J}_n$ is in fact the Jacobian matrix associated with the fundamental equations in the $e_{N,i}(z)$, defined in the implicit relations (6.29) of Theorem 6.10 as

$$e_{N,i}(z) = \frac{1}{n}\sum_{k=1}^N \sigma_{ki}^2(n) t_k(z) = \frac{1}{n}\sum_{k=1}^N \sigma_{ki}^2(n) \frac{1}{-z + \frac{1}{n}\sum_{l=1}^n \frac{\sigma_{kl}^2(n)}{1+e_{N,l}(z)}}.$$

Indeed, for all $e_{N,k}(-1/x)$ fixed but $e_{N,j}(-1/x)$, we have:

$$\frac{\partial}{\partial e_{N,j}(-1/x)}\left[\frac{1}{n}\sum_{k=1}^{N}\sigma_{ki}^2(n)\frac{1}{\frac{1}{x}+\frac{1}{n}\sum_{l=1}^{n}\frac{\sigma_{kl}^2(n)}{1+e_{N,l}(-1/x)}}\right]$$
$$=\frac{1}{n}\sum_{k=1}^{N}\sigma_{ki}^2(n)\frac{\frac{1}{n}\sigma_{kj}^2(n)}{(1+e_{N,j}(-1/x))^2}\frac{1}{\left(\frac{1}{x}+\frac{1}{n}\sum_{l=1}^{n}\frac{\sigma_{kl}^2(n)}{1+e_{N,l}(-1/x)}\right)^2}$$
$$=\frac{1}{n}\sum_{k=1}^{N}\frac{\frac{1}{n}\sigma_{ki}^2(n)\sigma_{kj}^2(n)t_k(-1/x)^2}{(1+e_{N,j}(-1/x))^2}$$
$$=J_{n,ji}.$$

So far, this observation seems to generalize to all central limits derived for random matrix models with independent entries. This is however only an intriguing but yet unproven fact.

Similar to Theorem 3.17, [Hachem *et al.*, 2008b] provides more than an asymptotic central limit theorem for the Shannon transform of the information plus noise model $\mathcal{V}_{\mathbf{B}_N} - \mathrm{E}[\mathcal{V}_{\mathbf{B}_N}]$, but also the fluctuations for N large of the difference between $\mathcal{V}_{\mathbf{B}_N}$ and its deterministic equivalent $\mathcal{V}_N$, provided in [Hachem *et al.*, 2007]. In the case where $\mathbf{X}_N$ has Gaussian entries, this takes a very compact expression.

Theorem 6.22. *Under the conditions of Theorem 6.21 with the additional assumption that the entries of $\mathbf{X}_N$ are complex Gaussian, we have:*

$$\frac{N}{\sqrt{-\log\det(\mathbf{I}_n-\mathbf{J}_n)}}(\mathcal{V}_{\mathbf{B}_N}(x)-\mathcal{V}_N(x)) \Rightarrow X \sim \mathcal{N}(0,1)$$

where $\mathcal{V}_N$ is defined as

$$\mathcal{V}_N(x)=\frac{1}{N}\sum_{i=1}^{N}\log\left(\frac{x}{t_i(-1/x)}\right)+\frac{1}{N}\sum_{j=1}^{n}\log\left(1+\frac{1}{n}\sum_{l=1}^{N}\sigma_{lj}^2(n)t_l(-1/x)\right)$$
$$-\frac{1}{Nn}\sum_{\substack{1\le i\le N\\ 1\le j\le n}}\frac{\sigma_{ij}^2(n)t_i(-1/x)}{1+\frac{1}{n}\sum_{l=1}^{N}\sigma_{lj}^2(n)t_l(-1/x)}$$

with $t_1,\ldots,t_N$ and $\mathbf{J}_n$ defined as in Theorem 6.21.

The generalization to distributions of the entries of $\mathbf{X}_N$ with a non-zero kurtosis κ introduces an additional bias term corresponding to the limiting variations of $N(\mathrm{E}[\mathcal{V}_{\mathbf{B}_N}(x)] - \mathcal{V}_N(x))$. This converges instead to zero in the Gaussian case or, as a matter of fact, in the case of any distribution with null kurtosis.

This concludes this short section on central limit theorems for deterministic equivalents.

This also closes this chapter on the classical techniques used for deterministic equivalents, when there exists no limit to the e.s.d. of the random matrix under study. Those deterministic equivalents are seen today as one of the most powerful tools to evaluate the performance of large wireless communication systems encompassing multiple antennas, multiple users, multiple cells, random codes, fast fading channels, etc. which are studied with scrutiny in Part II. In order to study complicated system models involving e.g. doubly-scattering channels, multi-hop channels, random precoders in random channels, etc., the current trend is to study nested deterministic equivalents; that is, deterministic equivalents that account for the stochasticity of multiple independent random matrices, see e.g. [Hoydis *et al.*, 2011a,b].

In the following, we turn to a rather different subject and study more deeply the limiting spectra of the sample covariance matrix model and of the information plus noise model. For these, much more than limiting spectral densities is known. It has especially been proved that, under some mild conditions, the extreme eigenvalues for both models do not escape the support of the l.s.d. and that a precise characterization of the position of some eigenvalues can be determined. Some additional study will characterize precisely the links between the population covariance matrix (or the information matrix) and the sample covariance matrix (or the information plus noise matrix), which are fundamental to address the questions of inverse problems and more precisely statistical eigen-inference for large dimensional random matrix models. These questions are at the core of the very recent signal processing tools, which enable novel signal sensing techniques and (N, n)-consistent estimation procedures adapted to large dimensional networks.