On the Capacity Achieving Covariance Matrix for Rician MIMO Channels: An Asymptotic Approach

Julien Dumont, Walid Hachem, Samson Lasaulce, Philippe Loubaton, *Fellow, IEEE*, and Jamal Najim

***Abstract*—In this paper, the capacity-achieving input covariance matrices for coherent block-fading correlated multiple input multiple output (MIMO) Rician channels are determined. In contrast with the Rayleigh and uncorrelated Rician cases, no closed-form expressions for the eigenvectors of the optimum input covariance matrix are available. Classically, both the eigenvectors and eigenvalues are computed numerically and the corresponding optimization algorithms remain computationally very demanding. In the asymptotic regime where the number of transmit and receive antennas converge to infinity at the same rate, new results related to the accuracy of the approximation of the average mutual information are provided. Based on the accuracy of this approximation, an attractive optimization algorithm is proposed and analyzed. This algorithm is shown to yield an effective way to compute the capacity achieving matrix for the average mutual information and numerical simulation results show that, even for a moderate number of transmit and receive antennas, the new approach provides the same results as direct maximization approaches of the average mutual information.**

***Index Terms*—Multiple input multiple output (MIMO) Rician channels, ergodic capacity, large random matrices, capacity achieving covariance matrices, iterative waterfilling.**

I. Introduction

SINCE the seminal work of Telatar [38], the advantage of considering multiple antennas at the transmitter and the receiver in terms of capacity, for Gaussian and fast Rayleigh fading single-user channels, is well understood. In that paper, the figure of merit chosen for characterizing the performance of a coherent[1] communication over a fading multiple input multiple output (MIMO) channel is the ergodic mutual information (EMI). Assuming the knowledge of the channel statistics at the transmitter, an important issue is to maximize the EMI with respect to the channel input distribution. Without loss of optimality, the search for the optimal input distribution can be restricted to circularly Gaussian inputs. The problem then amounts to finding the optimum covariance matrix.

This optimization problem has been addressed extensively in the case of certain Rayleigh channels. In the context of the so-called Kronecker model, it has been shown by various authors (see, e.g., [16] for a review) that the eigenvectors of the optimal input covariance matrix must coincide with the eigenvectors of the transmit correlation matrix. It is therefore sufficient to evaluate the eigenvalues of the optimal matrix, a problem which can be solved by using standard optimization algorithms. Note that [39] extended this result to more general (non Kronecker) Rayleigh channels.

Rician channels have been comparatively less studied from this point of view. Let us mention the work [20] devoted to the case of uncorrelated Rician channels, where the authors proved that the eigenvectors of the optimal input covariance matrix are the right-singular vectors of the line of sight component of the channel. As in the Rayleigh case, the eigenvalues can then be evaluated by standard routines. The case of correlated Rician channels is more complicated as the eigenvectors of the optimum matrix have no closed form expressions. Moreover, the exact expression of the EMI being complicated (see, e.g., [23]), both the eigenvalues and the eigenvectors have to be evaluated numerically. In [41], a barrier interior-point method is proposed and implemented to directly evaluate the EMI as an expectation. The corresponding algorithms are however computationally very demanding as they heavily rely on intensive Monte Carlo simulations.

In this paper, we address the optimization of the input covariance of Rician channels with a two-sided (Kronecker) correlation. As the exact expression of the EMI is very complicated, we propose to optimize the approximation of the EMI, first presented in [35], valid when the number of transmit and receive antennas converge to infinity at the same rate.

This will turn out to be a simpler problem. The results of the present contribution have been presented in part in the short conference paper [13].

The asymptotic approximation of the mutual information has been obtained by various authors in the case of MIMO Rayleigh channels, and has shown to be quite reliable even for a moderate number of antennas, see [10], [40], and [29]. The case of Rician channels has been considered more recently. Using the replica method, [30] obtained the asymptotic expression of the ergodic

Manuscript received October 12, 2007; revised July 17, 2009. Current version published March 10, 2010. This work was supported in part by the "Fonds National de la Science" via the ACI program "Nouvelles Interfaces des Mathématiques," project MALCOM number 205, and by the IST Network of Excellence NEWCOM, project number 507325. The material in the paper was presented in part at Globecom Communication Theory Symposium, San Francisco, CA, December 2006.

J. Dumont was with France Telecom and Université Paris-Est, France (e-mail: dumont@crans.org).

W. Hachem and J. Najim are with CNRS and Télécom Paris, Paris, France (e-mail: hachem@enst.fr; najim@enst.fr).

S. Lasaulce is with CNRS and SUPELEC, France (e-mail: lasaulce@lss.supelec.fr).

Ph. Loubaton is with the Université Paris Est, IGM LabInfo, UMR CNRS 8049, France (e-mail: loubaton@univ-mlv.fr).

Communicated by A. J. Goldsmith, Associate Editor for Communications.

Color versions of Figures 1 and 2 in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIT.2009.2039063

[1]Instantaneous channel state information is assumed at the receiver but not necessarily at the transmitter.

0018-9448/$26.00 © 2010 IEEE

mutual information together with the variance of the mutual information in the case of uncorrelated Rician channels. These results were generalized to the context of general bicorrelated Rician channels in [35] and [37]. Using large random matrix techniques, an asymptotic approximation of the EMI is provided in [18] in the case of a Rician channel whose random entries are independent random variables with nonseparable variance profile. If the variance profile is separable, this channel is equivalent (up to unitary invariance) to a bicorrelated Rician channels, and one can recover the expression of the EMI given in [35], [37]. Finally, [36] generalizes the results of [35] and [37] to the case of a Rician channel with interference and proposes to optimize the approximation of the EMI in order to obtain a capacity achieving covariance matrix in the context of a Rician channel with interference. The optimization algorithm of the large system approximant of the EMI proposed in [36] is however different from the algorithm studied here.

In this paper, we consider the closed-form asymptotic approximation for the mutual information as it appeared in [35], [37], and [18] and present new results concerning its accuracy. We prove in particular that the relative error decreases at rate t^{-2} where t represents the number of transmit antennas. Such an analysis is new in the context of a Rician channel with two-sided correlation.

We then address the optimization of the large system approximation with respect to the input covariance matrix and propose a simple iterative maximization algorithm which, in some sense, can be seen as a generalization to the Rician case of [43] devoted to the Rayleigh context: Each iteration will be devoted to solve a system of two nonlinear equations as well as a standard waterfilling problem. Among the convergence results that we provide: It is proved that the asymptotic equivalent of the average mutual information is concave with respect to the input covariance matrix. This garantees good convergence properties if any, and also a good speed of convergence. It is also proved that the algorithm converges towards the optimum input covariance matrix as long as it converges[2]. Concavity and convergence issues are not addressed in [43] and [36]. Finally, we also prove that the matrix which optimizes the large system approximation asymptotically achieves the capacity. This result, which has not been previously established for any approximation results, has an important practical range as it asserts that the optimization algorithm yields a procedure that asymptotically achieves the *true* capacity. Finally, simulation results confirm the relevance of our approach.

The paper is organized as follows. Section II is devoted to the presentation of the channel model and the underlying assumptions. The asymptotic approximation of the ergodic mutual information is given in Section III. In Section IV, the strict concavity of the asymptotic approximation as a function of the covariance matrix of the input signal is established; it is also proved that the resulting optimal argument asymptotically achieves the true capacity. The maximization problem of the EMI approximation is studied in Section V. Numerical results are provided in Section VI.

[2]Note however that we have been unable to prove formally its convergence.

II. Problem Statement

A. General Notations

In this paper, the notations $s, \mathbf{x}, \mathbf{M}$ stand for scalars, vectors and matrices, respectively. As usual, $\|\mathbf{x}\|$ represents the Euclidian norm of vector $\mathbf{x}$ and $\|\mathbf{M}\|$ stands for the spectral norm of matrix $\mathbf{M}$. The superscripts $(\cdot)^T$ and $(\cdot)^H$ represent, respectively, the transpose and transpose conjugate. The trace of $\mathbf{M}$ is denoted by $\mathrm{Tr}(\mathbf{M})$. The mathematical expectation operator is denoted by $\mathbb{E}(\cdot)$ and the symbols $\Re$ and $\Im$ denote, respectively, the real and imaginary parts of a given complex number. If x is a possibly complex-valued random variable, $\mathrm{Var}(x) = \mathbb{E}|x|^2 - |\mathbb{E}(x)|^2$ represents the variance of x.

All along this paper, t and r stand for the number of transmit and receive antennas. Certain quantities will be studied in the asymptotic regime $t \to \infty, r \to \infty$ in such a way that $\frac{t}{r} \to c \in (0, \infty)$. In order to simplify the notations, $t \to \infty$ should be understood from now on as $t \to \infty, r \to \infty$ and $\frac{t}{r} \to c \in (0, \infty)$. A matrix $\mathbf{M}_t$ whose size depends on t is said to be uniformly bounded if $\sup_t \|\mathbf{M}_t\| < \infty$.

Several variables used throughout this paper depend on various parameters, e.g., the number of antennas, the noise level, the covariance matrix of the transmitter, etc. In order to simplify the notations, we may not always mention all these dependencies.

B. Channel Model

We consider a wireless MIMO link with t transmit and r receive antennas. In our analysis, the channel matrix can possibly vary from symbol vector (or space-time codeword) to symbol vector. The channel matrix is assumed to be perfectly known at the receiver whereas the transmitter has only access to the statistics of the channel. The received signal can be written as

$$\mathbf{y}(\tau) = \mathbf{H}(\tau)\mathbf{x}(\tau) + \mathbf{z}(\tau) \tag{1}$$

where $\mathbf{x}(\tau)$ is the $t \times 1$ vector of transmitted symbols at time τ, $\mathbf{H}(\tau)$ is the $r \times t$ channel matrix (stationary and ergodic process) and $\mathbf{z}(\tau)$ is a complex white Gaussian noise distributed as $N(0, \sigma^2 \mathbf{I}_r)$. For the sake of simplicity, we omit the time index τ from our notations. The channel input is subject to a power constraint $\mathrm{Tr}[\mathbb{E}(\mathbf{x}\mathbf{x}^H)] \leq t$. Matrix $\mathbf{H}$ has the following structure:

$$\mathbf{H} = \sqrt{\frac{K}{K+1}}\mathbf{A} + \frac{1}{\sqrt{K+1}}\mathbf{V} \tag{2}$$

where matrix $\mathbf{A}$ is deterministic, $\mathbf{V}$ is a random matrix and constant $K \geq 0$ is the so-called Rician factor which expresses the relative strength of the direct and scattered components of the received signal. Matrix $\mathbf{A}$ satisfies $\frac{1}{r}\mathrm{Tr}(\mathbf{A}\mathbf{A}^H) = 1$ while $\mathbf{V}$ is given by

$$\mathbf{V} = \frac{1}{\sqrt{t}}\mathbf{C}_R^{1/2}\mathbf{W}\mathbf{C}_T^{1/2} \tag{3}$$

where $\mathbf{W} = (W_{ij})$ is a $r \times t$ matrix whose entries are independent and identically distributed (i.i.d.) complex circular Gaussian random variables $\mathcal{CN}(0,1)$, i.e., $W_{ij} = \Re W_{ij} + \mathbf{i}\Im W_{ij}$ where $\Re W_{ij}$ and $\Im W_{ij}$ are independent centered real

Gaussian random variables with variance $\frac{1}{2}$. The matrices $\mathbf{C}_T > 0$ and $\mathbf{C}_R > 0$ account for the transmit and receive antenna correlation effects, respectively, and satisfy $\frac{1}{t}\mathrm{Tr}(\mathbf{C}_T) = 1$ and $\frac{1}{r}\mathrm{Tr}(\mathbf{C}_R) = 1$. This correlation structure is often referred to as a separable or Kronecker correlation model.

C. Maximum Ergodic Mutual Information

We denote by $\mathcal{C}$ the cone of nonnegative Hermitian $t \times t$ matrices and by $\mathcal{C}_1$ the subset of all matrices $\mathbf{Q}$ of $\mathcal{C}$ for which $\frac{1}{t}\mathrm{Tr}(\mathbf{Q}) = 1$. Let $\mathbf{Q}$ be an element of $\mathcal{C}_1$ and denote by $I(\mathbf{Q})$ the ergodic mutual information (EMI) defined by

$$I(\mathbf{Q}) = \mathbb{E}_{\mathbf{H}}\left[\log\det\left(\mathbf{I}_r + \frac{1}{\sigma^2}\mathbf{H}\mathbf{Q}\mathbf{H}^H\right)\right]. \quad (4)$$

Maximizing the EMI with respect to the input covariance matrix $\mathbf{Q} = \mathbb{E}(\mathbf{x}\mathbf{x}^H)$ leads to the channel Shannon capacity for *fast* fading MIMO channels, i.e., when the channel vary from symbol to symbol. This capacity is achieved by averaging over channel variations over time.

We will denote by C_E the maximum value of the EMI over the set $\mathcal{C}_1$

$$C_E = \sup_{\mathbf{Q}\in\mathcal{C}_1} I(\mathbf{Q}). \quad (5)$$

The optimal input covariance matrix thus coincides with the argument of the above maximization problem. Note that $I : \mathbf{Q} \mapsto I(\mathbf{Q})$ is a strictly concave function on the convex set $\mathcal{C}_1$, which guarantees the existence of a unique maximum $\mathbf{Q}_*$ (see [27]). When $\mathbf{C}_T = \mathbf{I}_t$, $\mathbf{C}_R = \mathbf{I}_r$, [20] shows that the eigenvectors of the optimal input covariance matrix coincide with the right-singular vectors of $\mathbf{A}$. By adapting the proof of [20], one can easily check that this result also holds when $\mathbf{C}_T = \mathbf{I}_t$ and $\mathbf{C}_R$ and $\mathbf{A}\mathbf{A}^H$ share a common eigenvector basis. Apart from these two simple cases, it seems difficult to find a closed-form expression for the eigenvectors of the optimal covariance matrix. Therefore the evaluation of C_E requires the use of numerical techniques (see, e.g., [41]) which are very demanding since they rely on computationally intensive Monte Carlo simulations. This problem can be circumvented in the following way: The EMI $I(\mathbf{Q})$ can be approximated by a simple expression denoted by $\bar{I}(\mathbf{Q})$ (see Section III) as $t \to \infty$, this expression is in turn optimized with respect to $\mathbf{Q}$ (see Section V).

D. Summary of the Main Results.

The main contributions of this paper can be summarized as follows:

1) The approximation $\bar{I}(\mathbf{Q})$ of $I(\mathbf{Q})$ as $t \to \infty$ presented in [35], [37] can be written as

$$\bar{I}(\mathbf{Q}) = \log\det\left[\mathbf{I}_t + \mathbf{G}(\delta_R(\mathbf{Q}), \delta_T(\mathbf{Q}))\mathbf{Q}\right] + i(\delta_R(\mathbf{Q}), \delta_T(\mathbf{Q})) \quad (6)$$

where $\delta_R(\mathbf{Q})$ and $\delta_T(\mathbf{Q})$ are two positive terms defined as the solutions of a system of two equations [see (28)]. Functions $\mathbf{G}$ and i are given in closed form and depend on $(\delta_R(\mathbf{Q}), \delta_T(\mathbf{Q})), K, \mathbf{A}, \mathbf{C}_R, \mathbf{C}_T$, and on the noise variance σ^2.

We prove that the error term $I(\mathbf{Q}) - \bar{I}(\mathbf{Q})$ is of order $O(t^{-1})$. As $I(\mathbf{Q})$ is known to increase linearly with t, the relative error $\frac{I(\mathbf{Q})-\bar{I}(\mathbf{Q})}{I(\mathbf{Q})}$ is of order $O(t^{-2})$. This supports the fact that $\bar{I}(\mathbf{Q})$ is an accurate approximation of $I(\mathbf{Q})$, and that it is relevant to study $\bar{I}(\mathbf{Q})$ in order to obtain some insight on $I(\mathbf{Q})$.

2) We prove that the function $\mathbf{Q} \mapsto \bar{I}(\mathbf{Q})$ is strictly concave on $\mathcal{C}_1$. As a consequence, the maximum of $\bar{I}$ over $\mathcal{C}_1$ is reached for a unique matrix $\bar{\mathbf{Q}}_*$. We also show that $I(\bar{\mathbf{Q}}_*) - I(\mathbf{Q}_*) = O(t^{-1})$ where we recall that $\mathbf{Q}_*$ is the capacity achieving covariance matrix. Otherwise stated, the computation of $\bar{\mathbf{Q}}_*$ (see below) allows one to (asymptotically) achieve the capacity $I(\mathbf{Q}_*)$.

3) We study the structure of $\bar{\mathbf{Q}}_*$ and establish that $\bar{\mathbf{Q}}_*$ is solution of the standard waterfilling problem

$$\max_{\mathbf{Q}\in\mathcal{C}_1} \log\det(\mathbf{I} + \mathbf{G}(\delta_{R,*}, \delta_{T,*})\mathbf{Q})$$

where $\delta_{R,*} = \delta_R(\bar{\mathbf{Q}}_*), \delta_{T,*} = \delta_T(\bar{\mathbf{Q}}_*)$ and

$$\mathbf{G}(\delta_{R,*}, \delta_{T,*}) = \frac{\delta_{R,*}}{K+1}\mathbf{C}_T + \frac{1}{\sigma^2}\frac{K}{K+1}\mathbf{A}^H\left(\mathbf{I}_r + \frac{\delta_{T,*}}{K+1}\mathbf{C}_R\right)^{-1}\mathbf{A}.$$

This result provides insights on the structure of the approximating capacity achieving covariance matrix, but cannot be used to evaluate $\bar{\mathbf{Q}}_*$ since the parameters $\delta_{R,*}$ and $\delta_{T,*}$ depend on the optimum matrix $\bar{\mathbf{Q}}_*$. We therefore propose an attractive iterative maximization algorithm of $\bar{I}(\mathbf{Q})$ where each iteration consists in solving a standard waterfilling problem and a 2×2 system characterizing the parameters (δ_R, δ_T).

III. Asymptotic Behavior of the Ergodic Mutual Information

In this section, the input covariance matrix $\mathbf{Q} \in \mathcal{C}_1$ is fixed and the purpose is to evaluate the asymptotic behavior of the ergodic mutual information $I(\mathbf{Q})$ as $t \to \infty$ (recall that $t \to \infty$ means $t \to \infty, r \to \infty$ and $t/r \to c \in (0, \infty)$).

As we shall see, it is possible to study the accuracy of approximation $\bar{I}(\mathbf{Q})$ of $I(\mathbf{Q})$. The starting point of our approach is partly based on the results of [18] devoted to the study of the asymptotic behavior of the eigenvalue distribution of matrix $\mathbf{\Sigma}\mathbf{\Sigma}^H$ where $\mathbf{\Sigma}$ is given by

$$\mathbf{\Sigma} = \mathbf{B} + \mathbf{Y} \quad (7)$$

matrix $\mathbf{B}$ being a deterministic $r \times t$ matrix, and $\mathbf{Y}$ being a $r \times t$ zero mean (possibly complex circular Gaussian) random matrix with independent entries whose variances are given by $\mathrm{E}|Y_{ij}|^2 = \frac{\sigma_{ij}^2}{t}$. Notice in particular that the variables $(Y_{ij};\ 1 \leq i \leq r,\ 1 \leq j \leq t)$ are not necessarily identically distributed. We shall refer to the triangular array $(\sigma_{ij}^2;\ 1 \leq i \leq r,\ 1 \leq j \leq t)$ as the variance profile of $\mathbf{\Sigma}$; we shall say that it is separable if $\sigma_{ij}^2 = d_i\tilde{d}_j$ where $d_i \geq 0$ for $1 \leq i \leq r$ and $\tilde{d}_j \geq 0$ for $1 \leq j \leq t$. Due to the unitary invariance of the EMI of Gaussian channels, the study of $I(\mathbf{Q})$ will turn out to be equivalent to the study of the EMI of model (7) in the complex circular Gaussian case with a separable variance profile. We however stress that the mathematical technics used in the present paper completely differ from the tools used in [18] (see Remark 2 below).

A. Introduction of the Virtual Channel $\mathbf{HQ}^{\frac{1}{2}}$

The purpose of this section is to establish a link between the simplified model (7): $\mathbf{\Sigma} = \mathbf{B} + \mathbf{Y}$ where $\mathbf{Y} = \frac{1}{\sqrt{t}}\mathbf{D}^{\frac{1}{2}}\mathbf{X}\tilde{\mathbf{D}}^{\frac{1}{2}}, \mathbf{X}$ being a matrix with i.i.d $\mathcal{CN}(0,1)$ entries, $\mathbf{D}$ and $\tilde{\mathbf{D}}$ being diagonal matrices, and the Rician model (2) under investigation. As we shall see, the key point is the unitary invariance of the EMI of Gaussian channels together with a well-chosen eigenvalue/eigenvector decomposition.

Proposition 1: Let $\mathbf{X}$ be a $r \times t$ matrix whose individual entries are i.i.d. $\mathcal{CN}(0,1)$ random variables. The two ergodic mutual informations

$$I(\mathbf{Q}) = \mathbb{E} \log\det\left(\mathbf{I} + \frac{\mathbf{HQH}^H}{\sigma^2}\right)$$

and

$$J(\sigma^2) = \mathbb{E} \log\det\left(\mathbf{I} + \frac{\mathbf{\Sigma\Sigma}^H}{\sigma^2}\right)$$

are equal provided that

- Channel $\mathbf{\Sigma}$ is given by $\mathbf{\Sigma} = \mathbf{B} + \mathbf{Y}$ with $\mathbf{Y} = \frac{1}{\sqrt{t}}\mathbf{D}^{\frac{1}{2}}\mathbf{X}\tilde{\mathbf{D}}^{\frac{1}{2}}$
- The following eigenvalue/eigenvector decompositions hold true:

$$\frac{\mathbf{C}_R}{\sqrt{K+1}} = \mathbf{UDU}^H \quad \text{and} \quad \frac{\mathbf{Q}^{\frac{1}{2}}\mathbf{C}_T\mathbf{Q}^{\frac{1}{2}}}{\sqrt{K+1}} = \tilde{\mathbf{U}}\tilde{\mathbf{D}}\tilde{\mathbf{U}}^H \quad (8)$$

  where $\mathbf{U}$ and $\tilde{\mathbf{U}}$ are the eigenvectors matrices while $\mathbf{D}$ and $\tilde{\mathbf{D}}$ are the eigenvalues diagonal matrices.
- Matrices $\mathbf{A}$ and $\mathbf{B}$ are related via the identity

$$\mathbf{B} = \sqrt{\frac{K}{K+1}}\mathbf{U}^H\mathbf{A}\mathbf{Q}^{\frac{1}{2}}\tilde{\mathbf{U}}. \quad (9)$$

Proof: We introduce the virtual channel $\mathbf{HQ}^{1/2}$

$$\mathbf{HQ}^{\frac{1}{2}} = \sqrt{\frac{K}{K+1}}\mathbf{AQ}^{\frac{1}{2}} + \frac{1}{\sqrt{K+1}}\mathbf{C}_R^{\frac{1}{2}}\frac{\mathbf{W}}{\sqrt{t}}\mathbf{\Theta}\left(\mathbf{Q}^{\frac{1}{2}}\mathbf{C}_T\mathbf{Q}^{\frac{1}{2}}\right)^{\frac{1}{2}} \quad (10)$$

where $\mathbf{\Theta}$ is the deterministic unitary $t \times t$ matrix defined by $\mathbf{\Theta} = \mathbf{C}_T^{\frac{1}{2}}\mathbf{Q}^{\frac{1}{2}}(\mathbf{Q}^{\frac{1}{2}}\mathbf{C}_T\mathbf{Q}^{\frac{1}{2}})^{-\frac{1}{2}}$. The virtual channel $\mathbf{HQ}^{\frac{1}{2}}$ has thus a structure similar to $\mathbf{H}$, with $(\mathbf{A}, \mathbf{C}_R, \mathbf{C}_T, \mathbf{W})$, respectively, replaced $(\mathbf{AQ}^{\frac{1}{2}}, \mathbf{C}_R, \mathbf{Q}^{\frac{1}{2}}\mathbf{C}_T\mathbf{Q}^{\frac{1}{2}}, \mathbf{W\Theta})$. Consider now the decomposition (8). It is then clear that the ergodic mutual information of channel $\mathbf{HQ}^{\frac{1}{2}}$ coincides with the EMI of $\mathbf{\Sigma} = \mathbf{U}^H\mathbf{HQ}^{1/2}\tilde{\mathbf{U}}$. Matrix $\mathbf{\Sigma}$ can be written as $\mathbf{\Sigma} = \mathbf{B} + \mathbf{Y}$ where $\mathbf{B}$ is given by (9) and $\mathbf{Y} = \frac{1}{\sqrt{t}}\mathbf{D}^{\frac{1}{2}}\mathbf{X}\tilde{\mathbf{D}}^{\frac{1}{2}}$ with $\mathbf{X} = \mathbf{U}^H\mathbf{W\Theta}\tilde{\mathbf{U}}$. As matrix $\mathbf{W}$ has i.i.d. $\mathcal{CN}(0,1)$ entries, so has matrix $\mathbf{X} = \mathbf{U}^H\mathbf{W\Theta}\tilde{\mathbf{U}}$ due to the unitary invariance (note that the entries of $\mathbf{Y}$ are independent since $\mathbf{D}$ and $\tilde{\mathbf{D}}$ are diagonal). Proposition 1 is proved. □

B. Study of the EMI of the Equivalent Model (7)

We first introduce the resolvent and the Stieltjes transform associated with $\mathbf{\Sigma\Sigma}^H$ (Section III-B-1); we then introduce auxiliary quantities (Section III-B-2) and their main properties, together with the approximation of the EMI.

1) The Resolvent, the Stieltjes Transform: Denote by $\mathbf{S}(\sigma^2)$ and $\tilde{\mathbf{S}}(\sigma^2)$ the resolvents of matrices $\mathbf{\Sigma\Sigma}^H$ and $\mathbf{\Sigma}^H\mathbf{\Sigma}$ defined by

$$\mathbf{S}(\sigma^2) = [\mathbf{\Sigma\Sigma}^H + \sigma^2\mathbf{I}_r]^{-1}$$
$$\tilde{\mathbf{S}}(\sigma^2) = [\mathbf{\Sigma}^H\mathbf{\Sigma} + \sigma^2\mathbf{I}_t]^{-1}. \quad (11)$$

These resolvents satisfy the obvious, but useful property

$$\mathbf{S}(\sigma^2) \le \frac{\mathbf{I}_r}{\sigma^2}, \qquad \tilde{\mathbf{S}}(\sigma^2) \le \frac{\mathbf{I}_t}{\sigma^2}. \quad (12)$$

Recall that the Stieltjes transform of a nonnegative measure μ is defined by $\int \frac{\mu(d\lambda)}{\lambda - z}$. The quantity $s(\sigma^2) = \frac{1}{r}\mathrm{Tr}(\mathbf{S}(\sigma^2))$ coincides with the Stieltjes transform of the eigenvalue distribution of matrix $\mathbf{\Sigma\Sigma}^H$ evaluated at point $z = -\sigma^2$. In fact, denote by $(\lambda_i)_{1\le i\le r}$ its eigenvalues, then

$$s(\sigma^2) = \frac{1}{r}\sum_{i=1}^{r}\frac{1}{\lambda_i + \sigma^2} = \int_{\mathrm{R}^+}\frac{\nu(d\lambda)}{\lambda + \sigma^2}$$

where ν represents the empirical distribution of the eigenvalues of $\mathbf{\Sigma\Sigma}^H$, that is the probability distribution $\frac{1}{r}\sum_{i=1}^{r}\delta_{\lambda_i}$ where δ_x represents the Dirac distribution at point x. The Stieltjes transform $s(\sigma^2)$ is important as the characterization of the asymptotic behavior of the eigenvalue distribution of $\mathbf{\Sigma\Sigma}^H$ is equivalent to the study of $s(\sigma^2)$ when $t \to \infty$ for each σ^2. This observation is the starting point of the approaches developed by Pastur [28], Girko [14], Bai and Silverstein [1], etc.

2) Important Auxiliary Quantities and Asymptotic Approximation of the EMI: We gather in this section many results of [18] that will be of help in the sequel.

Assumption 1: Let $(\mathbf{B}_t)$ be a family of $r \times t$ deterministic matrices such that: $\sup_{t,i}\sum_{j=1}^{t}|B_{ij}|^2 < \infty, \sup_{t,j}\sum_{i=1}^{r}|B_{ij}|^2 < \infty$.

Theorem 1: Consider the random matrix $\mathbf{\Sigma} = \mathbf{B} + \mathbf{Y}$, where $\mathbf{Y} = \frac{1}{\sqrt{t}}\mathbf{D}^{\frac{1}{2}}\mathbf{X}\,\tilde{\mathbf{D}}^{\frac{1}{2}}, \mathbf{D}$ and $\tilde{\mathbf{D}}$ represent the diagonal matrices $\mathbf{D} = \mathrm{diag}(d_i,\, 1 \le i \le r)$ and $\tilde{\mathbf{D}} = \mathrm{diag}(\tilde{d}_j,\, 1 \le j \le t)$, respectively, and where $\mathbf{X}$ is a matrix whose entries are i.i.d. complex centered with variance one. The following facts hold true:

i) (*Existence and uniqueness of auxiliary quantities*) For σ^2 fixed, consider the system of equations

$$\begin{cases} \beta = \dfrac{1}{t}\mathrm{Tr}\left[\mathbf{D}(\sigma^2(\mathbf{I}_r + \mathbf{D}\tilde{\beta}) + \mathbf{B}(\mathbf{I}_t + \tilde{\mathbf{D}}\beta)^{-1}\mathbf{B}^H)^{-1}\right] \\ \tilde{\beta} = \dfrac{1}{t}\,\mathrm{Tr}\,\left[\tilde{\mathbf{D}}(\sigma^2(\mathbf{I}_t + \tilde{\mathbf{D}}\beta) + \mathbf{B}^H(\mathbf{I}_r + \mathbf{D}\tilde{\beta})^{-1}\mathbf{B})^{-1}\right]. \end{cases} \quad (13)$$

Then, among the solutions of (13), there is a unique couple of strictly positive solutions $(\beta(\sigma^2), \tilde{\beta}(\sigma^2))$. Denote by $\mathbf{T}(\sigma^2)$ and $\tilde{\mathbf{T}}(\sigma^2)$ the following matrix-valued functions:

$$\begin{cases} \mathbf{T}(\sigma^2) = \left[\sigma^2(\mathbf{I} + \tilde{\beta}(\sigma^2)\mathbf{D}) + \mathbf{B}(\mathbf{I} + \beta(\sigma^2)\tilde{\mathbf{D}})^{-1}\mathbf{B}^H\right]^{-1} \\ \tilde{\mathbf{T}}(\sigma^2) = \left[\sigma^2(\mathbf{I} + \beta(\sigma^2)\tilde{\mathbf{D}}) + \mathbf{B}^H(\mathbf{I} + \tilde{\beta}(\sigma^2)\mathbf{D})^{-1}\mathbf{B}\right]^{-1}. \end{cases} \tag{14}$$

Matrices $\mathbf{T}(\sigma^2)$ and $\tilde{\mathbf{T}}(\sigma^2)$ satisfy

$$\mathbf{T}(\sigma^2) \le \frac{\mathbf{I}_r}{\sigma^2}, \qquad \tilde{\mathbf{T}}(\sigma^2) \le \frac{\mathbf{I}_t}{\sigma^2}. \tag{15}$$

ii) (*Representation of the auxiliary quantities*) The solutions $\beta(\sigma^2)$ and $\tilde{\beta}(\sigma^2)$ of system (13) are given by

$$\beta(\sigma^2) = \frac{1}{t}\,\mathrm{Tr}\,\mathbf{D}\mathbf{T}(\sigma^2) \qquad \tilde{\beta}(\sigma^2) = \frac{1}{t}\,\mathrm{Tr}\,\tilde{\mathbf{D}}\tilde{\mathbf{T}}(\sigma^2) \tag{16}$$

and can be written as

$$\beta(\sigma^2) = \int_{\mathrm{R}^+} \frac{\mu_b(d\lambda)}{\lambda + \sigma^2} \qquad \tilde{\beta}(\sigma^2) = \int_{\mathrm{R}^+} \frac{\tilde{\mu}_b(d\lambda)}{\lambda + \sigma^2} \tag{17}$$

where μ_b and $\tilde{\mu}_b$ are nonnegative scalar measures with respective total mass $\frac{1}{t}\mathrm{Tr}(\mathbf{D})$ and $\frac{1}{t}\mathrm{Tr}(\tilde{\mathbf{D}})$. Similarly, there exist probability measures μ and $\tilde{\mu}$ such that

$$\begin{aligned} \frac{1}{t}\,\mathrm{Tr}\,\mathbf{T}(\sigma^2) &= \int_{\mathrm{R}^+} \frac{\mu(d\lambda)}{\lambda + \sigma^2}, \\ \frac{1}{t}\,\mathrm{Tr}\,\tilde{\mathbf{T}}(\sigma^2) &= \int_{\mathrm{R}^+} \frac{\tilde{\mu}(d\lambda)}{\lambda + \sigma^2}. \end{aligned} \tag{18}$$

iii) (*Asymptotic approximation of the EMI*) Assume that Assumption 1 holds and that

$$\sup_t \|\mathbf{D}\| < d_{\max} < \infty \quad \text{and} \quad \sup_t \|\tilde{\mathbf{D}}\| < \tilde{d}_{\max} < \infty.$$

For every deterministic matrices $\mathbf{M}$ and $\tilde{\mathbf{M}}$ satisfying $\sup_t \|\mathbf{M}\| < \infty$ and $\sup_t \|\tilde{\mathbf{M}}\| < \infty$, the following limits hold true almost surely:

$$\begin{cases} \lim_{t\to\infty} \frac{1}{r}\,\mathrm{Tr}\,[(\mathbf{S}(\sigma^2) - \mathbf{T}(\sigma^2))\mathbf{M}] = 0 \\ \lim_{t\to\infty} \frac{1}{t}\,\mathrm{Tr}\,\left[(\tilde{\mathbf{S}}(\sigma^2) - \tilde{\mathbf{T}}(\sigma^2))\tilde{\mathbf{M}}\right] = 0. \end{cases} \tag{19}$$

Denote by $J(\sigma^2) = \mathbb{E}\log\det(\mathbf{I}_r + \sigma^{-2}\mathbf{\Sigma}\mathbf{\Sigma}^H)$ the EMI associated with matrix $\mathbf{\Sigma}$. Then $J(\sigma^2)$ writes

$$J(\sigma^2) = r\mathbb{E}\int_{\sigma^2}^{\infty} \left(\frac{1}{\omega} - \frac{1}{r}\mathrm{Tr}\,\mathbf{S}(\omega)\right) d\omega. \tag{20}$$

Define by $\bar{J}(\sigma^2)$ the quantity

$$\bar{J}(\sigma^2) = r\int_{\sigma^2}^{\infty} \left(\frac{1}{\omega} - \frac{1}{r}\mathrm{Tr}\,\mathbf{T}(\omega)\right) d\omega. \tag{21}$$

Then $\bar{J}(\sigma^2)$ can be expressed as

$$\begin{aligned} \log\det&\left[\mathbf{I}_r + \tilde{\beta}(\sigma^2)\mathbf{D} + \frac{1}{\sigma^2}\mathbf{B}(\mathbf{I}_t + \beta(\sigma^2)\tilde{\mathbf{D}})^{-1}\mathbf{B}^H\right] \\ &+ \log\det\left[\mathbf{I}_t + \beta(\sigma^2)\tilde{\mathbf{D}}\right] - \sigma^2 t\beta(\sigma^2)\tilde{\beta}(\sigma^2) \end{aligned} \tag{22}$$

or equivalently

$$\begin{aligned} \log\det&\left[\mathbf{I}_t + \beta(\sigma^2)\tilde{\mathbf{D}} + \frac{1}{\sigma^2}\mathbf{B}^H(\mathbf{I}_r + \tilde{\beta}(\sigma^2)\mathbf{D})^{-1}\mathbf{B}\right] \\ &+ \log\det\left[\mathbf{I}_r + \tilde{\beta}(\sigma^2)\mathbf{D}\right] - \sigma^2 t\beta(\sigma^2)\tilde{\beta}(\sigma^2). \end{aligned} \tag{23}$$

Moreover, the following convergence holds true:

$$J(\sigma^2) = \bar{J}(\sigma^2) + o(t) \quad \text{as} \quad t \to \infty. \tag{24}$$

Proof of (i) is provided in Appendix I (note that in [18], the existence and uniqueness of solutions to (13) is proved in a certain class of analytic functions depending on σ^2; this implies the existence of a solution $(\beta, \tilde{\beta})$ when σ^2 is fixed, but not the uniqueness; we provide in Appendix I an elementary proof of the existence which immediately implies the uniqueness). The rest of the statements of Theorem 1 have been established in [18], and their proof is omitted here.

Remark 1: As shown in [18], the results in Theorem 1 do not require any Gaussian assumption for $\mathbf{\Sigma}$. Notice that (19) implies in some sense that the entries of $\mathbf{S}(\sigma^2)$ and $\tilde{\mathbf{S}}(\sigma^2)$ have the same behavior as the entries of the deterministic matrices $\mathbf{T}(\sigma^2)$ and $\tilde{\mathbf{T}}(\sigma^2)$ [which can be evaluated by solving (13)]. In particular, using (19) for $\mathbf{M} = \mathbf{I}$, it follows that the Stieltjes transform $s(\sigma^2)$ of the eigenvalue distribution of $\mathbf{\Sigma}\mathbf{\Sigma}^H$ behaves like $\frac{1}{r}\,\mathrm{Tr}\,\mathbf{T}(\sigma^2)$, which is itself the Stieltjes transform of a probability measure μ (see, for instance, [18]).

In order to evaluate the precision of the asymptotic approximation $\bar{J}$, we shall improve (24) and get the speed $J(\sigma^2) = \bar{J}(\sigma^2) + O(t^{-1})$ in the next theorem. This result completes those in [18] and in Theorem 1-(iii) but heavily relies on the Gaussian structure of $\mathbf{\Sigma}$. We first introduce very mild extra assumptions:

Assumption 2: Let $(\mathbf{B}_t)$ be a family of $r \times t$ deterministic matrices such that

$$\sup_t \|\mathbf{B}\| < b_{\max} < +\infty.$$

Assumption 3: Let $\mathbf{D}$ and $\tilde{\mathbf{D}}$ be, respectively, $r \times r$ and $t \times t$ diagonal matrices such that

$$\sup_t \|\mathbf{D}\| < d_{\max} < \infty \quad \text{and} \quad \sup_t \|\tilde{\mathbf{D}}\| < \tilde{d}_{\max} < \infty.$$

Assume moreover that

$$\inf_t \frac{1}{t}\,\mathrm{Tr}\,\mathbf{D} > 0 \quad \text{and} \quad \inf_t \frac{1}{t}\,\mathrm{Tr}\,\tilde{\mathbf{D}} > 0.$$

Theorem 2: Consider the simplified model as in Theorem 1: $\mathbf{\Sigma} = \mathbf{B} + \mathbf{Y}$, with$\mathbf{Y} = \frac{1}{\sqrt{t}}\mathbf{D}^{\frac{1}{2}}\mathbf{X}\,\tilde{\mathbf{D}}^{12}$. Assume moreover that Assumptions 2 and 3 hold true. Then, for every deterministic matrices $\mathbf{M}$ and $\tilde{\mathbf{M}}$ satisfying $\sup_t \|\mathbf{M}\| < \infty$ and $\sup_t \|\tilde{\mathbf{M}}\| < \infty$, the following facts hold true:

$$\begin{aligned} \mathrm{Var}\left(\frac{1}{r}\,\mathrm{Tr}\,[\mathbf{S}(\sigma^2)\mathbf{M}]\right) &= O\left(\frac{1}{t^2}\right) \\ \mathrm{Var}\left(\frac{1}{t}\,\mathrm{Tr}\,[\tilde{\mathbf{S}}(\sigma^2)\tilde{\mathbf{M}}]\right) &= O\left(\frac{1}{t^2}\right). \end{aligned} \tag{25}$$

Moreover

$$\frac{1}{r}\ \mathrm{Tr}\ [(\mathbb{E}(\mathbf{S}(\sigma^2)) - \mathbf{T}(\sigma^2))\mathbf{M}] = O(t^{-2})$$
$$\frac{1}{t}\ \mathrm{Tr}\ [(\mathbb{E}(\tilde{\mathbf{S}}(\sigma^2)) - \tilde{\mathbf{T}}(\sigma^2))\tilde{\mathbf{M}}] = O(t^{-2}) \qquad (26)$$

and

$$J(\sigma^2) = \bar{J}(\sigma^2) + O\left(t^{-1}\right). \qquad (27)$$

The proof is given in Appendix II. We provide here some comments.

Remark 2: The proof of Theorem 2 takes full advantage of the Gaussian structure of matrix $\mathbf{\Sigma}$ and relies on two simple ingredients: An integration by parts formula that provides an expression for the expectation of certain functionals of Gaussian vectors, widely used in Random Matrix Theory [26], [31] and Poincaré-Nash inequality that bounds the variance of functionals of Gaussian vectors. Although well known, the application of this inequality to random matrices is fairly recent [7], [32], and also [17].

Remark 3 (Gaussian versus Non-Gaussian): Equations (25) also hold in the non Gaussian case and can be established by using the so-called Resolvent FORmula Martingale (REFORM) method introduced by Girko [14].

Equations (26) and (27) are specific to the complex Gaussian structure of the channel matrix $\mathbf{\Sigma}$. In particular, in the non-Gaussian case, or in the real Gaussian case, one would get $J(\sigma^2) = \bar{J}(\sigma^2) + O(1)$. These two facts are in accordance with [2] in which a weaker result $(o(1))$ is proved in the simpler case where $\mathbf{B} = \mathbf{0}$, and the predictions of the replica method in [29] (respectively, [30]) in the case where $\mathbf{B} = \mathbf{0}$ (respectively, in the case where $\tilde{\mathbf{D}} = \mathbf{I}_t$ and $\mathbf{D} = \mathbf{I}_r$).

Remark 4 (Standard Deviation and Bias): Equation (25) implies that the standard deviation of $\frac{1}{r}\ \mathrm{Tr}\ [(\mathbf{S}(\sigma^2) - \mathbf{T}(\sigma^2))\mathbf{M}]$ and $\frac{1}{t}\mathrm{Tr}[(\tilde{\mathbf{S}}(\sigma^2) - \tilde{\mathbf{T}}(\sigma^2))\tilde{\mathbf{M}}]$ are of order $O(t^{-1})$ terms. However, their mathematical expectations (which correspond to the bias) converge much faster towards 0 as (26) shows (the order is $O(t^{-2})$).

Remark 5: Both $J(\sigma^2)$ and $\bar{J}(\sigma^2)$ increase linearly with t. Equation (27), thus, implies that the relative error $\frac{J(\sigma^2)-\bar{J}(\sigma^2)}{J(\sigma^2)}$ is of order $O(t^{-2})$. This remarkable convergence rate strongly supports the observed fact that approximations of the EMI remain reliable even for small numbers of antennas (see also the numerical results in Section VI). Note that similar observations have been done in other contexts where random matrices are used, see, e.g., [3], [29], [35], and [37].

C. Study of the EMI $I(\mathbf{Q})$

We now apply the previous results to the study of the EMI of channel $\mathbf{H}$. We first state the corresponding result.

Theorem 3: For $\mathbf{Q} \in \mathcal{C}_1$, consider

$$\begin{cases} \delta_R &= f_R(\delta_R, \delta_T, \mathbf{Q}) \\ \delta_T &= f_T(\delta_R, \delta_T, \mathbf{Q}) \end{cases} \qquad (28)$$

where $f_R(\delta_R, \delta_T, \mathbf{Q})$ and $f_T(\delta_R, \delta_T, \mathbf{Q})$ are given by

$$f_R(\delta_R, \delta_T, \mathbf{Q}) = \frac{1}{t}\ \mathrm{Tr}\left\{\mathbf{C}_R\left[\sigma^2\left(\mathbf{I}_r + \frac{\delta_T}{K+1}\mathbf{C}_R\right) + \frac{K}{K+1}\mathbf{A}\mathbf{Q}^{1/2} \times\left(\mathbf{I}_t + \frac{\delta_R}{K+1}\mathbf{Q}^{1/2}\mathbf{C}_T\mathbf{Q}^{1/2}\right)^{-1}\mathbf{Q}^{1/2}\mathbf{A}^H\right]^{-1}\right\} \qquad (29)$$

$$f_T(\delta_R, \delta_T, \mathbf{Q}) = \frac{1}{t}\ \mathrm{Tr}\left\{\mathbf{Q}^{1/2}\mathbf{C}_T\mathbf{Q}^{1/2}\left[\sigma^2\left(\mathbf{I}_t + \frac{\delta_R}{K+1}\mathbf{Q}^{1/2}\mathbf{C}_T\mathbf{Q}^{1/2}\right) + \frac{K}{K+1}\mathbf{Q}^{1/2}\mathbf{A}^H \times\left(\mathbf{I}_r + \frac{\delta_T}{K+1}\mathbf{C}_R\right)^{-1}\mathbf{A}\mathbf{Q}^{1/2}\right]^{-1}\right\}. \qquad (30)$$

Then (28) has a unique strictly positive solution $(\delta_R(\mathbf{Q}), \delta_T(\mathbf{Q}))$.

Furthermore, assume that $\sup_t \|\mathbf{Q}\| < \infty$, $\sup_t \|\mathbf{A}\| < \infty$, $\sup_t \|\mathbf{C}_R\| < \infty$, and $\sup_t \|\mathbf{C}_T\| < \infty$. Assume also that $\inf_t \lambda_{\min}(\mathbf{C}_T) > 0$ where $\lambda_{\min}(\mathbf{C}_T)$ represents the smallest eigenvalue of $\mathbf{C}_T$. Then, as $t \to \infty$

$$I(\mathbf{Q}) = \bar{I}(\mathbf{Q}) + O\left(\frac{1}{t}\right) \qquad (31)$$

where the asymptotic approximation $\bar{I}(\mathbf{Q})$ is given by

$$\bar{I}(\mathbf{Q}) = \log\det\left(\mathbf{I}_t + \frac{\delta_R(\mathbf{Q})}{K+1}\mathbf{Q}^{1/2}\mathbf{C}_T\mathbf{Q}^{1/2} + \frac{1}{\sigma^2}\frac{K}{K+1}\mathbf{Q}^{1/2}\mathbf{A}^H \times\left(\mathbf{I}_r + \frac{\delta_T(\mathbf{Q})}{K+1}\mathbf{C}_R\right)^{-1}\mathbf{A}\mathbf{Q}^{1/2}\right) + \log\det\left(\mathbf{I}_r + \frac{\delta_T(\mathbf{Q})}{K+1}\mathbf{C}_R\right) - \frac{t\sigma^2}{K+1}\delta_R(\mathbf{Q})\,\delta_T(\mathbf{Q}) \qquad (32)$$

or equivalently by

$$\bar{I}(\mathbf{Q}) = \log\det\left(\mathbf{I}_r + \frac{\delta_T(\mathbf{Q})}{K+1}\mathbf{C}_R + \frac{1}{\sigma^2}\frac{K}{K+1}\mathbf{A}\mathbf{Q}^{1/2} \times\left(\mathbf{I}_t + \frac{\delta_R(\mathbf{Q})}{K+1}\mathbf{Q}^{1/2}\mathbf{C}_T\mathbf{Q}^{1/2}\right)^{-1}\mathbf{Q}^{1/2}\mathbf{A}^H\right) + \log\det\left(\mathbf{I}_t + \frac{\delta_R(\mathbf{Q})}{K+1}\mathbf{Q}^{1/2}\mathbf{C}_T\mathbf{Q}^{1/2}\right) - \frac{t\sigma^2}{K+1}\delta_R(\mathbf{Q})\,\delta_T(\mathbf{Q}). \qquad (33)$$

Proof: We rely on the virtual channel introduced in Section III-A and on the eigenvalue/eigenvector decomposition performed there.

Matrices $\mathbf{B}, \mathbf{D}, \tilde{\mathbf{D}}$ as introduced in Proposition 1 are clearly uniformly bounded, while $(K+1)^{1/2} \inf_t \frac{1}{t} \operatorname{Tr} \mathbf{D} = \inf_t \frac{1}{t} \operatorname{Tr} \mathbf{C}_R = 1$ due to the model specifications and $\inf_t \frac{1}{t} \operatorname{Tr} \mathbf{Q}^{\frac{1}{2}} \mathbf{C}_T \mathbf{Q}^{\frac{1}{2}} \geq \inf_t \lambda_{\min}(\mathbf{C}_T) \frac{1}{t} \operatorname{Tr} \mathbf{Q} > 0$ as $\frac{1}{t} \operatorname{Tr} \mathbf{Q} = 1$. Therefore, matrices $\mathbf{B}, \mathbf{D}$ and $\tilde{\mathbf{D}}$ clearly satisfy the assumptions of Theorems 1 and 2.

We first apply the results of Theorem 1-(i) to matrix $\mathbf{\Sigma}$ (and use the same notations). Using the unitary invariance of the trace of a matrix, it is straightforward to check that:

$$\frac{f_R(\delta_R, \delta_T, \mathbf{Q})}{\sqrt{K+1}} = \frac{1}{t} \operatorname{Tr} \left[\mathbf{D} \left(\sigma^2 \left(\mathbf{I} + \mathbf{D} \frac{\delta_T}{\sqrt{K+1}} \right) + \mathbf{B} \left(\mathbf{I} + \tilde{\mathbf{D}} \frac{\delta_R}{\sqrt{K+1}} \right)^{-1} \mathbf{B}^H \right)^{-1} \right] \tag{34}$$

$$\frac{f_T(\delta_R, \delta_T, \mathbf{Q})}{\sqrt{K+1}} = \frac{1}{t} \operatorname{Tr} \left[\tilde{\mathbf{D}} \left(\sigma^2 \left(\mathbf{I} + \tilde{\mathbf{D}} \frac{\delta_R}{\sqrt{K+1}} \right) + \mathbf{B}^H \left(\mathbf{I} + \mathbf{D} \frac{\delta_T}{\sqrt{K+1}} \right)^{-1} \mathbf{B} \right)^{-1} \right]. \tag{35}$$

Therefore, (δ_R, δ_T) is solution of (28) if and only if $(\frac{\delta_R}{\sqrt{K+1}}, \frac{\delta_T}{\sqrt{K+1}})$ is solution of (13). As (13) admits a unique strictly positive pair of solutions, say $(\beta, \tilde{\beta})$, (28) satisfies the same property and the corresponding pair (δ_R, δ_T) is related to $(\beta, \tilde{\beta})$ by

$$\beta = \frac{\delta_R}{\sqrt{K+1}}, \tilde{\beta} = \frac{\delta_T}{\sqrt{K+1}}. \tag{36}$$

In order to justify (32) and (33), we note that $J(\sigma^2)$ coincides with the EMI $I(\mathbf{Q})$. Moreover, the unitary invariance of the determinant of a matrix together with (36) imply that $\bar{I}(\mathbf{Q})$ defined by (32) and (33) coincide with the approximation $\bar{J}$ given by (22) and (23). This proves (31) as well. □

In the following, we denote by $\mathbf{T}_R(\sigma^2)$ and $\mathbf{T}_T(\sigma^2)$ the following matrix-valued functions:

$$\mathbf{T}_R(\sigma^2) = \left[\sigma^2 \left(\mathbf{I} + \frac{\delta_T}{K+1} \mathbf{C}_R \right) + \frac{K}{K+1} \mathbf{A} \mathbf{Q}^{\frac{1}{2}} \times \left(\mathbf{I} + \frac{\delta_R}{K+1} \mathbf{Q}^{\frac{1}{2}} \mathbf{C}_T \mathbf{Q}^{\frac{1}{2}} \right)^{-1} \mathbf{Q}^{\frac{1}{2}} \mathbf{A}^H \right]^{-1} \tag{37}$$

$$\mathbf{T}_T(\sigma^2) = \left[\sigma^2 \left(\mathbf{I} + \frac{\delta_R}{K+1} \mathbf{Q}^{\frac{1}{2}} \mathbf{C}_T \mathbf{Q}^{\frac{1}{2}} \right) + \frac{K}{K+1} \mathbf{Q}^{\frac{1}{2}} \mathbf{A}^H \times \left(\mathbf{I} + \frac{\delta_T}{K+1} \mathbf{C}_R \right)^{-1} \mathbf{A} \mathbf{Q}^{\frac{1}{2}} \right]^{-1}. \tag{38}$$

They are related to matrices $\mathbf{T}$ and $\tilde{\mathbf{T}}$ defined by (14) by

$$\begin{cases} \mathbf{T}_R(\sigma^2) & = \mathbf{U} \mathbf{T}(\sigma^2) \mathbf{U}^H \\ \mathbf{T}_T(\sigma^2) & = \tilde{\mathbf{U}} \tilde{\mathbf{T}}(\sigma^2) \tilde{\mathbf{U}}^H \end{cases} \tag{39}$$

and their entries represent deterministic approximations of $(\mathbf{H}\mathbf{Q}\mathbf{H}^H + \sigma^2 \mathbf{I}_r)^{-1}$ and $(\mathbf{Q}^{\frac{1}{2}} \mathbf{H}^H \mathbf{H} \mathbf{Q}^{\frac{1}{2}} + \sigma^2 \mathbf{I}_t)^{-1}$.

As $\frac{1}{r} \operatorname{Tr} \mathbf{T}_R = \frac{1}{r} \operatorname{Tr} \mathbf{T}$ and $\frac{1}{t} \operatorname{Tr} \mathbf{T}_T = \frac{1}{t} \operatorname{Tr} \tilde{\mathbf{T}}$, the quantities $\frac{1}{r} \operatorname{Tr} \mathbf{T}_R$ and $\frac{1}{t} \operatorname{Tr} \mathbf{T}_T$ are the Stieltjes transforms of probability measures μ and $\tilde{\mu}$ introduced in Theorem 1-(ii). As matrices $\mathbf{H}\mathbf{Q}\mathbf{H}^H$ and $\mathbf{\Sigma}\mathbf{\Sigma}^H$ (resp. $\mathbf{Q}^{\frac{1}{2}} \mathbf{H}^H \mathbf{H} \mathbf{Q}^{\frac{1}{2}}$ and $\mathbf{\Sigma}^H \mathbf{\Sigma}$) have the same eigenvalues, one can notice that the eigenvalue distribution of $\mathbf{H}\mathbf{Q}\mathbf{H}^H$ (respectively, $\mathbf{Q}^{\frac{1}{2}} \mathbf{H}^H \mathbf{H} \mathbf{Q}^{\frac{1}{2}}$) behaves like μ (respectively, $\tilde{\mu}$).

We finally mention that $\delta_R(\sigma^2)$ and $\delta_T(\sigma^2)$ are given by

$$\begin{aligned} \delta_R(\sigma^2) &= \frac{1}{t} \operatorname{Tr} \mathbf{C}_R \mathbf{T}_R(\sigma^2) \\ \delta_T(\sigma^2) &= \frac{1}{t} \operatorname{Tr} \mathbf{Q}^{\frac{1}{2}} \mathbf{C}_T \mathbf{Q}^{1/2} \mathbf{T}_T(\sigma^2) \end{aligned} \tag{40}$$

and that the following representations hold true:

$$\delta_R(\sigma^2) = \int_{\mathbb{R}^+} \frac{\mu_R(d\lambda)}{\lambda + \sigma^2}$$

and

$$\delta_T(\sigma^2) = \int_{\mathbb{R}^+} \frac{\mu_T(d\lambda)}{\lambda + \sigma^2} \tag{41}$$

where μ_R and μ_T are positive measures on $\mathbb{R}^+$ satisfying $\mu_R(\mathbb{R}^+) = \frac{1}{t} \operatorname{Tr} \mathbf{C}_R$ and $\mu_T(\mathbb{R}^+) = \frac{1}{t} \operatorname{Tr} \mathbf{Q}^{1/2} \mathbf{C}_T \mathbf{Q}^{1/2}$.

IV. Strict Concavity of $\bar{I}(\mathbf{Q})$ and Approximation of the Capacity $I(\mathbf{Q}_*)$

A. Strict Concavity of $\bar{I}(\mathbf{Q})$

The strict concavity of $\bar{I}(\mathbf{Q})$ is an important issue for optimization purposes (see Section V). The main result of the section is as follows.

Theorem 4: The function $\mathbf{Q} \mapsto \bar{I}(\mathbf{Q})$ is strictly concave on $\mathcal{C}_1$.

As we shall see, the concavity of $\bar{I}$ can be established quite easily by relying on the concavity of the EMI $I(\mathbf{Q}) = \mathbb{E} \log \det(\mathbf{I} + \frac{\mathbf{H}\mathbf{Q}\mathbf{H}^H}{\sigma^2})$. The strict concavity is more demanding.

In the sequel, we shall rely on the following straightforward but useful result:

Proposition 2: Let $f : \mathcal{C}_1 \to \mathbb{R}$ be a real function. Then f is strictly concave if and only if for every matrices $\mathbf{Q}_1, \mathbf{Q}_2 (\mathbf{Q}_1 \neq \mathbf{Q}_2)$ of $\mathcal{C}_1$, the function $\phi(\lambda)$ defined on $[0, 1]$ by

$$\phi(\lambda) = f(\lambda \mathbf{Q}_1 + (1 - \lambda) \mathbf{Q}_2)$$

is strictly concave.

Recall that $I(\mathbf{Q}) = \mathbb{E} \log \det(\mathbf{I} + \frac{\mathbf{H}\mathbf{Q}\mathbf{H}^H}{\sigma^2})$ is concave on $\mathcal{C}_1$ (see, for instance, [27]).

1) Using Another Auxiliary Channel to Establish the Concavity of $\bar{I}(\mathbf{Q})$: Denote by $\otimes$ the Kronecker product of matrices. We introduce the following matrices:

$$\begin{aligned} \mathbf{\Delta}_R = \mathbf{I}_m \otimes \mathbf{C}_R, \quad \mathbf{\Delta}_T = \mathbf{I}_m \otimes \mathbf{C}_T \\ \check{\mathbf{A}} = \mathbf{I}_m \otimes \mathbf{A}, \quad \check{\mathbf{Q}} = \mathbf{I}_m \otimes \mathbf{Q}. \end{aligned} \tag{42}$$

Matrix $\mathbf{\Delta}_R$ is of size $rm \times rm$, matrices $\mathbf{\Delta}_T$ and $\check{\mathbf{Q}}$ are of size $tm \times tm$, and $\check{\mathbf{A}}$ is of size $rm \times tm$. Let us now introduce

$$\check{\mathbf{V}} = \frac{1}{\sqrt{mt}} \mathbf{\Delta}_R^{1/2} \check{\mathbf{W}} \mathbf{\Delta}_T^{1/2}$$

and

$$\check{\mathbf{H}} = \sqrt{\frac{K}{K+1}} \check{\mathbf{A}} + \frac{1}{\sqrt{K+1}} \check{\mathbf{V}}$$

where $\check{\mathbf{W}}$ is a $rm \times tm$ matrix whose entries are i.i.d $\mathcal{CN}(0,1)$-distributed random variables. Denote by $I_m(\check{\mathbf{Q}})$ the EMI associated with channel $\check{\mathbf{H}}$

$$I_m(\check{\mathbf{Q}}) = \mathbb{E} \log\det \left(\mathbf{I} + \frac{\check{\mathbf{H}}\check{\mathbf{Q}}\check{\mathbf{H}}^H}{\sigma^2} \right).$$

Applying Theorem 3 to the channel $\check{\mathbf{H}}$, we conclude that $I_m(\check{\mathbf{Q}})$ admits an asymptotic approximation $\bar{I}_m(\check{\mathbf{Q}})$ defined by (29)–(30) and (32), where one will substitute the quantities related to channel $\mathbf{H}$ by those related to channel $\check{\mathbf{H}}$, i.e.

$$\begin{aligned} t \leftrightarrow mt, \quad r \leftrightarrow mr, \quad \mathbf{A} \leftrightarrow \check{\mathbf{A}}, \\ \mathbf{Q} \leftrightarrow \check{\mathbf{Q}}, \quad \mathbf{C}_R \leftrightarrow \mathbf{\Delta}_R, \quad \mathbf{C}_T \leftrightarrow \mathbf{\Delta}_T. \end{aligned} \tag{43}$$

Due to the block-diagonal nature of matrices $\check{\mathbf{A}}, \check{\mathbf{Q}}, \mathbf{\Delta}_R$ and $\mathbf{\Delta}_T$, (28) associated with channel $\check{\mathbf{H}}$ is exactly the same as the one associated with channel $\mathbf{H}$. Moreover, a straightforward computation yields

$$\frac{1}{m} \bar{I}_m(\check{\mathbf{Q}}) = \bar{I}(\mathbf{Q}), \qquad \forall m \geq 1.$$

It remains to apply the convergence result (31) to conclude that

$$\lim_{m\to\infty} \frac{1}{m} I_m(\check{\mathbf{Q}}) = \bar{I}(\mathbf{Q}).$$

Since $\mathbf{Q} \mapsto I_m(\check{\mathbf{Q}}) = I_m(\mathbf{I}_m \otimes \mathbf{Q})$ is concave, $\bar{I}$ is concave as a pointwise limit of concave functions.

2) Uniform Strict Concavity of the EMI of the Auxiliary Channel—Strict Concavity of $\bar{I}(\mathbf{Q})$: In order to establish the strict concavity of $\bar{I}(\mathbf{Q})$, we shall rely on the following lemma.

Lemma 1: Let $\bar{\phi} : [0,1] \to \mathbb{R}$ be a real function such that there exists a family $(\phi_m)_{m\geq 1}$ of real functions satisfying:

i) The functions ϕ_m are twice differentiable and there exists $\kappa < 0$ such that

$$\forall m \geq 1, \quad \forall \lambda \in [0,1], \qquad \phi_m''(\lambda) \leq \kappa < 0. \tag{44}$$

ii) For every $\lambda \in [0,1], \phi_m(\lambda) \underset{m\to+\infty}{\longrightarrow} \bar{\phi}(\lambda)$.

Then $\bar{\phi}$ is a strictly concave real function.

Proof of Lemma 1 is straightforward and is therefore omitted. Let $\mathbf{Q}_1, \mathbf{Q}_2$ in $\mathcal{C}_1$; denote by $\mathbf{Q} = \lambda\mathbf{Q}_1 + (1-\lambda)\mathbf{Q}_2$, $\check{\mathbf{Q}}_1 = I_m \otimes \mathbf{Q}_1, \check{\mathbf{Q}}_2 = I_m \otimes \mathbf{Q}_2, \check{\mathbf{Q}} = I_m \otimes \mathbf{Q}$. Let $\check{\mathbf{H}}$ be the matrix associated with the auxiliary channel and denote by

$$\phi_m(\lambda) = \frac{1}{m} \mathbb{E} \log\det \left(\mathbf{I} + \frac{\check{\mathbf{H}}\check{\mathbf{Q}}\check{\mathbf{H}}^H}{\sigma^2} \right).$$

We have already proved that $\phi_m(\lambda) \underset{m\to+\infty}{\longrightarrow} \bar{\phi}(\lambda) \triangleq \bar{I}(\lambda\mathbf{Q}_1 + (1-\lambda)\mathbf{Q}_2)$. In order to fulfill assumptions of Lemma 1, it is sufficient to prove that there exists $\kappa < 0$ such that for every $\lambda \in [0,1]$,

$$\limsup_{m\to\infty} \phi_m''(\lambda) \leq \kappa < 0. \tag{45}$$

The proof of (45) is omitted, but available upon request (see also the extended version [44]).

B. Approximation of the Capacity $I(\mathbf{Q}_*)$

Since $\bar{I}$ is strictly concave over the compact set $\mathcal{C}_1$, it admits a unique argmax we shall denote by $\bar{\mathbf{Q}}_*$, i.e.

$$\bar{I}(\bar{\mathbf{Q}}_*) = \max_{\mathbf{Q}\in\mathcal{C}_1} \bar{I}(\mathbf{Q}).$$

As we shall see in Section V, matrix $\bar{\mathbf{Q}}_*$ can be obtained by a rather simple algorithm. Provided that $\sup_t \|\bar{\mathbf{Q}}_*\|$ is bounded, (31) in Theorem 3 yields $I(\bar{\mathbf{Q}}_*) - \bar{I}(\bar{\mathbf{Q}}_*) \to 0$ as $t \to \infty$. It remains to check that $I(\mathbf{Q}_*) - I(\bar{\mathbf{Q}}_*)$ goes asymptotically to zero to be able to approximate the capacity. This is the purpose of the next proposition.

Proposition 3: Assume that $\sup_t \|\mathbf{A}\| < \infty$, $\sup_t \|\mathbf{C}_T\| < \infty, \sup_t \|\mathbf{C}_R\| < \infty, \inf_t \lambda_{\min}(\mathbf{C}_T) > 0$, and $\inf_t \lambda_{\min}(\mathbf{C}_R) > 0$. Let $\bar{\mathbf{Q}}_*$ and $\mathbf{Q}_*$ be the maximizers over $\mathcal{C}_1$ of $\bar{I}$ and I, respectively. Then the following facts hold true:

i) $\sup_t \|\bar{\mathbf{Q}}_*\| < \infty$.

ii) $\sup_t \|\mathbf{Q}_*\| < \infty$.

iii) $I(\bar{\mathbf{Q}}_*) = I(\mathbf{Q}_*) + O(t^{-1})$.

Proof: The proof of items (i) and (ii) is postponed to Appendix III. Let us prove (iii). As

$$\begin{aligned} &\underbrace{(I(\mathbf{Q}_*) - I(\bar{\mathbf{Q}}_*))}_{\geq 0} + \underbrace{(\bar{I}(\bar{\mathbf{Q}}_*) - \bar{I}(\mathbf{Q}_*))}_{\geq 0} \\ &\quad = \underbrace{(I(\mathbf{Q}_*) - \bar{I}(\mathbf{Q}_*))}_{\substack{=O(t^{-1}) \\ \text{by (ii) and Th. 3 Eq. (31)}}} + \underbrace{(\bar{I}(\bar{\mathbf{Q}}_*) - I(\bar{\mathbf{Q}}_*))}_{\substack{=O(t^{-1}) \\ \text{by (i) and Th. 3 Eq. (31)}}} \end{aligned} \tag{46}$$

where the two terms of the LHS are nonnegative due to the fact that $\mathbf{Q}_*$ and $\bar{\mathbf{Q}}_*$ are the maximizers of I and $\bar{I}$, respectively. As a direct consequence of (46), we have $I(\mathbf{Q}_*) - I(\bar{\mathbf{Q}}_*) = O(t^{-1})$ and the proof is completed. □

V. Optimization of the Input Covariance Matrix

In the previous section, we have proved that matrix $\bar{\mathbf{Q}}_*$ asymptotically achieves the capacity. The purpose of this section is to propose an efficient way of maximizing the asymptotic approximation $\bar{I}(\mathbf{Q})$ without using complicated numerical optimization algorithms. In fact, we will show that our problem boils down to simple waterfilling algorithms.

A. Properties of the Maximum of $\bar{I}(\mathbf{Q})$

In this section, we shall establish some of $\bar{\mathbf{Q}}_*$'s properties. We first introduce a few notations. Let $V(\kappa, \tilde{\kappa}, \mathbf{Q})$ be the function defined by

$$V(\kappa,\tilde{\kappa},\mathbf{Q}) = \log\det\left(\mathbf{I}_t + \frac{\kappa}{K+1}\mathbf{Q}^{1/2}\mathbf{C}_T\mathbf{Q}^{1/2} + \frac{K}{\sigma^2(K+1)}\mathbf{Q}^{1/2}\mathbf{A}^H \times\left(\mathbf{I}_r + \frac{\tilde{\kappa}}{K+1}\mathbf{C}_R\right)^{-1}\mathbf{A}\mathbf{Q}^{1/2}\right) + \log\det\left(\mathbf{I}_r + \frac{\tilde{\kappa}}{K+1}\mathbf{C}_T\right) - \frac{t\sigma^2\kappa\tilde{\kappa}}{K+1}. \quad (47)$$

or equivalently by

$$V(\kappa,\tilde{\kappa},\mathbf{Q}) = \log\det\left(\mathbf{I}_r + \frac{\tilde{\kappa}}{K+1}\mathbf{C}_R + \frac{K}{\sigma^2(K+1)}\mathbf{A}\mathbf{Q}^{1/2} \times\left(\mathbf{I}_t + \frac{\kappa}{K+1}\mathbf{Q}^{1/2}\mathbf{C}_T\mathbf{Q}^{1/2}\right)^{-1}\mathbf{Q}^{1/2}\mathbf{A}^H\right) + \log\det\left(\mathbf{I}_t + \frac{\kappa}{K+1}\mathbf{Q}^{1/2}\mathbf{C}_T\mathbf{Q}^{1/2}\right) - \frac{t\sigma^2\kappa\tilde{\kappa}}{K+1}. \quad (48)$$

Note that if $(\delta_R(\mathbf{Q}), \delta_T(\mathbf{Q}))$ is the solution of system (28), then:

$$\bar{I}(\mathbf{Q}) = V(\delta_R(\mathbf{Q}), \delta_T(\mathbf{Q}), \mathbf{Q}).$$

Denote by $(\delta_{R,*}, \delta_{T,*})$ the solution $(\delta_R(\bar{\mathbf{Q}}_*), \delta_T(\bar{\mathbf{Q}}_*))$ of (28) associated with $\bar{\mathbf{Q}}_*$. The aim of the section is to prove that $\bar{\mathbf{Q}}_*$ is the solution of the following standard waterfilling problem:

$$\bar{I}(\bar{\mathbf{Q}}_*) = \max_{\mathbf{Q}\in\mathcal{C}_1} V(\delta_{R,*}, \delta_{T,*}, \mathbf{Q}).$$

Denote by $\mathbf{G}(\kappa, \tilde{\kappa})$ the $t \times t$ matrix given by

$$\mathbf{G}(\kappa,\tilde{\kappa}) = \frac{\kappa}{K+1}\mathbf{C}_T + \frac{K}{\sigma^2(K+1)}\mathbf{A}^H\left(\mathbf{I}_r + \frac{\tilde{\kappa}}{K+1}\mathbf{C}_R\right)^{-1}\mathbf{A}. \quad (49)$$

Then, $V(\kappa, \tilde{\kappa}, \mathbf{Q})$ also writes

$$V(\kappa,\tilde{\kappa},\mathbf{Q}) = \log\,\det\left(\mathbf{I} + \mathbf{Q}\mathbf{G}(\kappa,\tilde{\kappa})\right) + \log\det\left(\mathbf{I}_r + \frac{\tilde{\kappa}}{K+1}\mathbf{C}_R\right) - \frac{t\sigma^2\kappa\tilde{\kappa}}{K+1} \quad (50)$$

which readily implies the differentiability of $(\kappa, \tilde{\kappa}, \mathbf{Q}) \mapsto V(\kappa, \tilde{\kappa}, \mathbf{Q})$ and the strict concavity of $\mathbf{Q} \mapsto V(\kappa, \tilde{\kappa}, \mathbf{Q})$ (κ and $\tilde{\kappa}$ being frozen).

In the sequel, we will denote by $\nabla F(x)$ the derivative of the differentiable function F at point x and by $\langle \nabla F(x), y\rangle$ the value of this derivative at point y. The following proposition captures the main features needed in the sequel.

Proposition 4: Let $F : \mathcal{C}_1 \to \mathbb{R}$ be a concave and differentiable function. Then:

i) (*necessary condition*) If F attains its maximum for $\bar{\mathbf{Q}}_* \in \mathcal{C}_1$, then

$$\forall\mathbf{Q}\in\mathcal{C}_1, \quad \langle\nabla F(\bar{\mathbf{Q}}_*), \mathbf{Q} - \bar{\mathbf{Q}}_*\rangle \le 0. \quad (51)$$

ii) (*sufficient condition*) Assume that there exists $\bar{\mathbf{Q}}_* \in \mathcal{C}_1$ such that

$$\forall\mathbf{Q}\in\mathcal{C}_1, \quad \langle\nabla F(\bar{\mathbf{Q}}_*), \mathbf{Q} - \bar{\mathbf{Q}}_*\rangle \le 0. \quad (52)$$

Then F admits its maximum at $\bar{\mathbf{Q}}_*$ (i.e., $\bar{\mathbf{Q}}_*$ is an argmax of F over $\mathcal{C}_1$).

These results are standard (see for instance [5, Ch. 2]), therefore, the proof is omitted.

In the following proposition, we gather various properties related to $\bar{I}$.

Proposition 5: Consider the functions $\delta_R(\mathbf{Q}), \delta_T(\mathbf{Q})$ and $\bar{I}(\mathbf{Q})$ from $\mathcal{C}_1$ to $\mathbb{R}$. The following properties hold true:

i) Functions $\delta_R(\mathbf{Q}), \delta_T(\mathbf{Q})$ and $\bar{I}(\mathbf{Q})$ are differentiable (and in particular continuous) over $\mathcal{C}_1$.

ii) Let $\mathbf{Q} \in \mathcal{C}_1$. The following property:

$$\forall\mathbf{P}\in\mathcal{C}_1, \quad \langle\nabla\bar{I}(\mathbf{Q}), \mathbf{P} - \mathbf{Q}\rangle \le 0$$

holds true if and only if $\mathbf{Q} = \bar{\mathbf{Q}}_*$.

iii) Denote by $\delta_{R,*}$ and $\delta_{T,*}$ the quantities $\delta_R(\bar{\mathbf{Q}}_*)$ and $\delta_T(\bar{\mathbf{Q}}_*)$. Matrix $\bar{\mathbf{Q}}_*$ is the solution of the standard waterfilling problem: Maximize over $\mathbf{Q} \in \mathcal{C}_1$ the function $V(\delta_{R,*}, \delta_{T,*}, \mathbf{Q})$ or equivalently the function $\log\det(\mathbf{I} + \mathbf{Q}\mathbf{G}(\delta_{R,*}, \delta_{T,*}))$.

Proof: (i) is straightforward and its proof is therefore omitted. Let us establish (ii). Recall that $\bar{I}(\mathbf{Q})$ is strictly concave by Theorem 4 (and therefore its maximum is attained at most one point). On the other hand, $\bar{I}(\mathbf{Q})$ is continuous by (i) over $\mathcal{C}_1$ which is compact. Therefore, the maximum of $\bar{I}(\mathbf{Q})$ is uniquely attained at a point $\bar{\mathbf{Q}}_*$. Item (ii) follows then from Proposition 4.

Proof of item (iii) is based on the following identity, to be proved

$$\langle\nabla\bar{I}(\bar{\mathbf{Q}}_*), \mathbf{Q} - \bar{\mathbf{Q}}_*\rangle = \langle\nabla_{\mathbf{Q}}V(\delta_{R,*}, \delta_{T,*}, \bar{\mathbf{Q}}_*), \mathbf{Q} - \bar{\mathbf{Q}}_*\rangle \quad (53)$$

where $\nabla_{\mathbf{Q}}$ denote the derivative of $V(\kappa, \tilde{\kappa}, \mathbf{Q})$ with respect to V's third component, i.e., $\nabla_{\mathbf{Q}}V(\kappa, \tilde{\kappa}, \mathbf{Q}) = \nabla\Gamma(\mathbf{Q})$ with $\Gamma : \mathbf{Q} \mapsto V(\kappa, \tilde{\kappa}, \mathbf{Q})$. Assume that (53) holds true. Then item (ii) implies that $\langle\nabla_{\bar{\mathbf{Q}}_*}V(\delta_{R,*}, \delta_{T,*}, \bar{\mathbf{Q}}_*), \mathbf{Q} - \bar{\mathbf{Q}}_*\rangle \le 0$ for every $\mathbf{Q} \in \mathcal{C}_1$. As $\mathbf{Q} \mapsto V(\delta_{R,*}, \delta_{T,*}, \mathbf{Q})$ is strictly concave on $\mathcal{C}_1, \bar{\mathbf{Q}}_*$ is the argmax of $V(\delta_{R,*}, \delta_{T,*}, \cdot)$ by Proposition 4 and we are done.

It remains to prove (53). Consider $\mathbf{Q}$ and $\mathbf{P}$ in $\mathcal{C}_1$, and use the identity

$$\begin{aligned}\langle\nabla\bar{I}(\mathbf{P}), \mathbf{Q} - \mathbf{P}\rangle &= \langle\nabla_{\mathbf{Q}}V(\delta_R(\mathbf{P}), \delta_T(\mathbf{P}), \mathbf{P}), \mathbf{Q} - \mathbf{P})\rangle \\ &\quad + \left(\frac{\partial V}{\partial\kappa}\right)(\delta_R(\mathbf{P}), \delta_T(\mathbf{P}), \mathbf{P})\,\langle\nabla\delta_R(\mathbf{P}), \mathbf{Q} - \mathbf{P}\rangle \\ &\quad + \left(\frac{\partial V}{\partial\tilde{\kappa}}\right)(\delta_R(\mathbf{P}), \delta_T(\mathbf{P}), \mathbf{P})\,\langle\nabla\delta_T(\mathbf{P}), \mathbf{Q} - \mathbf{P}\rangle.\end{aligned}$$

We now compute the partial derivatives of V and obtain

$$\begin{cases}\dfrac{\partial V}{\partial\kappa} = -\dfrac{t\sigma^2}{K+1}(\tilde{\kappa} - f_T(\kappa,\tilde{\kappa},\mathbf{Q})) \\ \dfrac{\partial V}{\partial\tilde{\kappa}} = -\dfrac{t\sigma^2}{K+1}(\kappa - f_R(\kappa,\tilde{\kappa},\mathbf{Q}))\end{cases} \quad (54)$$

where f_R and f_T are defined by (29) and (30). The first relation follows from (47) and the second relation from (48). As $(\delta_R(\mathbf{Q}), \delta_T(\mathbf{Q}))$ is the solution of (28), (54) imply that

$$\frac{\partial V}{\partial \kappa}(\delta_R(\mathbf{Q}), \delta_T(\mathbf{Q}), \mathbf{Q}) = \frac{\partial V}{\partial \tilde{\kappa}}(\delta_R(\mathbf{Q}), \delta_T(\mathbf{Q}), \mathbf{Q}) = 0. \tag{55}$$

Letting $\mathbf{P} = \bar{\mathbf{Q}}_*$ and taking into account (55) yields:

$$\langle \nabla \bar{I}(\bar{\mathbf{Q}}_*), \mathbf{Q} - \bar{\mathbf{Q}}_* \rangle = \langle \nabla_{\mathbf{Q}} V(\delta_R(\bar{\mathbf{Q}}_*), \delta_T(\bar{\mathbf{Q}}_*), \bar{\mathbf{Q}}_*), \mathbf{Q} - \bar{\mathbf{Q}}_* \rangle$$

and (iii) is established. □

Remark 6: The quantities $\delta_{R,*}$ and $\delta_{T,*}$ depend on matrix $\bar{\mathbf{Q}}_*$. Therefore, Proposition 5 does not provide by itself any optimization algorithm. However, it gives valuable insights on the structure of $\bar{\mathbf{Q}}_*$. Consider first the case $\mathbf{C}_R = \mathbf{I}$ and $\mathbf{C}_T = \mathbf{I}$. Then, $\mathbf{G}(\delta_{R,*}, \delta_{T,*})$ is a linear combination of $\mathbf{I}$ and matrix $\mathbf{A}^H\mathbf{A}$. The eigenvectors of $\bar{\mathbf{Q}}_*$ thus coincide with the right singular vectors of matrix $\mathbf{A}$, a result consistent with the work [20] devoted to the maximization of the EMI $I(\mathbf{Q})$. If $\mathbf{C}_R = \mathbf{I}$ and $\mathbf{C}_T \neq \mathbf{I}, \mathbf{G}(\delta_{R,*}, \delta_{T,*})$ can be interpreted as a linear combination of matrices $\mathbf{C}_T$ and $\mathbf{A}^H\mathbf{A}$. Therefore, if the transmit antennas are correlated, the eigenvectors of the optimum matrix $\bar{\mathbf{Q}}_*$ coincide with the eigenvectors of some weighted sum of $\mathbf{C}_T$ and $\mathbf{A}^H\mathbf{A}$. This result provides a simple explanation of the impact of correlated transmit antennas on the structure of the optimal input covariance matrix. The impact of correlated receive antennas on $\bar{\mathbf{Q}}_*$ is however less intuitive because matrix $\mathbf{A}^H\mathbf{A}$ has to be replaced with $\mathbf{A}^H(\mathbf{I} + \delta_{T,*}\mathbf{C}_R)^{-1}\mathbf{A}$.

B. The Optimization Algorithm.

We are now in position to introduce our maximization algorithm of $\bar{I}$. It is mainly motivated by the simple observation that for each fixed $(\kappa, \tilde{\kappa})$, the maximization with respect to $\mathbf{Q}$ of function $V(\kappa, \tilde{\kappa}, \mathbf{Q})$ defined by (50) can be achieved by a standard waterfilling procedure, which, of course, does not need the use of numerical techniques. On the other hand, for $\mathbf{Q}$ fixed, the equation (28) have unique solutions that, in practice, can be obtained using a standard fixed-point algorithm. Our algorithm thus consists in adapting parameters $\mathbf{Q}$ and δ_R, δ_T separately by the following iterative scheme:

- Initialization: $\mathbf{Q}_0 = \mathbf{I}, (\delta_{R,1}, \delta_{T,1})$ are defined as the unique solutions of (28) in which $\mathbf{Q} = \mathbf{Q}_0 = \mathbf{I}$. Then, define $\mathbf{Q}_1$ as the maximum of function $\mathbf{Q} \to V(\delta_{R,1}, \delta_{T,1}, \mathbf{Q})$ on $\mathcal{C}_1$, which is obtained through a standard waterfilling procedure.
- Iteration k: assume $\mathbf{Q}_{k-1}, (\delta_{R,k-1}, \delta_{T,k-1})$ available. Then, $(\delta_{R,k}, \delta_{k,T})$ is defined as the unique solution of (28) in which $\mathbf{Q} = \mathbf{Q}_{k-1}$. Then, define $\mathbf{Q}_k$ as the maximum of function $\mathbf{Q} \to V(\delta_{R,k}, \delta_{T,k}, \mathbf{Q})$ on $\mathcal{C}_1$.

One can notice that this algorithm is the generalization of the procedure used by [43] for optimizing the input covariance matrix for correlated Rayleigh MIMO channels.

We now study the convergence properties of this algorithm, and state a result which implies that, if the algorithm converges, then it converges to the unique argmax $\bar{\mathbf{Q}}_*$ of $\bar{I}$.

Proposition 6: Assume that the two sequences $(\delta_{R,k})_{k\geq 0}$ and $(\delta_{T,k})_{k\geq 0}$ verify

$$\lim_{k\to\infty} \delta_{R,k} - \delta_{R,k-1} \to 0, \lim_{k\to\infty} \delta_{T,k} - \delta_{T,k-1} \to 0. \tag{56}$$

Then, the sequence $(\mathbf{Q}_k)_{k\geq 0}$ converges toward the maximum $\bar{\mathbf{Q}}_*$ of $\bar{I}$ on $\mathcal{C}_1$.

Proof: First note that the sequence $(\mathbf{Q}_k)$ belongs to the compact set $\mathcal{C}_1$. Therefore, in order to show that the sequence converges, it is sufficient to establish that the limits of all convergent subsequences coincide. We thus consider a convergent subsequence extracted from $(\mathbf{Q}_k)_{k\geq 0}$, say $(\mathbf{Q}_{\psi(k)})_{k\geq 0}$, where for each $k, \psi(k)$ is an integer, and denote by $\mathbf{Q}_*^{\psi}$ its limit. If we prove that

$$< \nabla \bar{I}(\mathbf{Q}_*^{\psi}), \mathbf{Q} - \mathbf{Q}_*^{\psi} > \leq 0 \tag{57}$$

for each $\mathbf{Q} \in \mathcal{C}_1$, Proposition 5–(ii) will imply that $\mathbf{Q}_*^{\psi}$ coincides with the argmax $\bar{\mathbf{Q}}_*$ of $\bar{I}$ over $\mathcal{C}_1$. This will prove that the limit of every convergent subsequence converges towards $\bar{\mathbf{Q}}_*$, which in turn will show that the whole sequence $(\mathbf{Q}_k)_{k\geq 0}$ converges to $\bar{\mathbf{Q}}_*$.

In order to prove (57), consider the iteration $\psi(k)$ of the algorithm. The matrix $\mathbf{Q}_{\psi(k)}$ maximizes the function $\mathbf{Q} \mapsto V(\delta_{R,\psi(k)}, \delta_{T,\psi(k)}, \mathbf{Q})$. As this function is strictly concave and differentiable, Proposition 4 implies that

$$< \nabla_{\mathbf{Q}} V(\delta_{R,\psi(k)}, \delta_{T,\psi(k)}, \mathbf{Q}_{\psi(k)}), \mathbf{Q} - \mathbf{Q}_{\psi(k)} > \leq 0 \tag{58}$$

for every $\mathbf{Q} \in \mathcal{C}_1$ (recall that $\nabla_{\mathbf{Q}}$ represents the derivative of $V(\kappa, \tilde{\kappa}, \mathbf{Q})$ with respect to V's third component). We now consider the pair of solutions $(\delta_{R,\psi(k)+1}, \delta_{T,\psi(k)+1})$ of the system (28) associated with matrix $\mathbf{Q}_{\psi(k)}$.

Due to the continuity of $\delta_R(\mathbf{Q})$ and $\delta_T(\mathbf{Q})$, the convergence of the subsequence $\mathbf{Q}_{\psi(k)}$ implies the convergence of the subsequences $(\delta_{R,\psi(k)+1}, \delta_{T,\psi(k)+1})$ towards a limit $(\delta_{R,*}^{\psi}, \delta_{T,*}^{\psi})$. The pair $(\delta_{R,*}^{\psi}, \delta_{T,*}^{\psi})$ is the solution of (28) associated with $\mathbf{Q}_*^{\psi}$ i.e., $\delta_{R,*}^{\psi} = \delta_R(\mathbf{Q}_*^{\psi})$ and $\delta_{T,*}^{\psi} = \delta_T(\mathbf{Q}_*^{\psi})$; in particular

$$\frac{\partial V}{\partial \kappa}\left(\delta_{R,*}^{\psi}, \delta_{T,*}^{\psi}, \mathbf{Q}_*^{\psi}\right) = \frac{\partial V}{\partial \tilde{\kappa}}\left(\delta_{R,*}^{\psi}, \delta_{T,*}^{\psi}, \mathbf{Q}_*^{\psi}\right) = 0$$

[see, for instance, (55)]. Using the same computation as in the proof of Proposition 5, we obtain

$$\left\langle \nabla \bar{I}(\mathbf{Q}_*^{\psi}), \mathbf{Q} - \mathbf{Q}_*^{\psi} \right\rangle = \left\langle \nabla V\left(\delta_{R,*}^{\psi}, \delta_{T,*}^{\psi}, \mathbf{Q}_*^{\psi}\right), \mathbf{Q} - \mathbf{Q}_*^{\psi} \right\rangle \tag{59}$$

for every $\mathbf{Q} \in \mathcal{C}_1$. Now (56) implies that the subsequence $(\delta_{R,\psi(k)}, \delta_{T,\psi(k)})$ also converges toward $(\delta_{R,*}^{\psi}, \delta_{T,*}^{\psi})$. As a consequence

$$\lim_{k\to\infty} \langle \nabla V(\delta_{R,\psi(k)}, \delta_{T,\psi(k)}, \mathbf{Q}_{\psi(k)}), \mathbf{Q} - \mathbf{Q}_{\psi(k)} \rangle = \left\langle \nabla V\left(\delta_{R,*}^{\psi}, \delta_{T,*}^{\psi}, \mathbf{Q}_*^{\psi}\right), \mathbf{Q} - \mathbf{Q}_*^{\psi} \right\rangle. \tag{60}$$

TABLE I
AVERAGE TIME PER ITERATION IN SECONDS

	$r=t=2$	$r=t=4$	$r=t=8$
Vu-Paulraj	0.75	8.2	138
New algorithm	10^{-2}	3.10^{-2}	7.10^{-2}

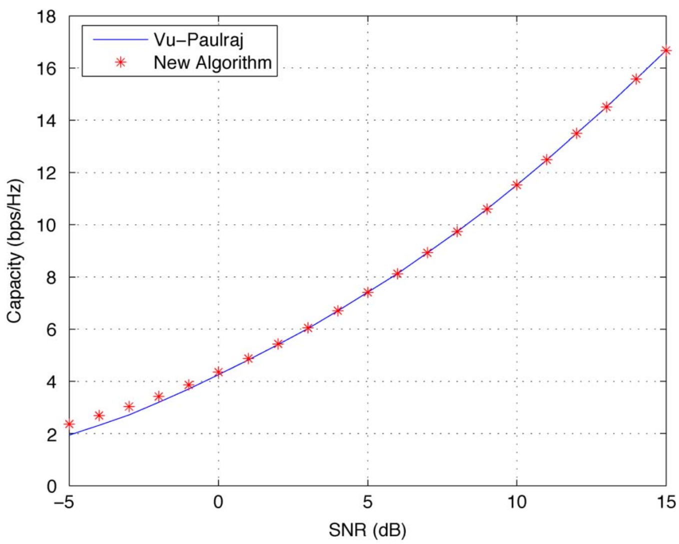

Fig. 1. Comparison with the Vu-Paulraj algorithm I.

Inequality (58), thus, implies that $\langle \nabla V(\delta_{R,*}^{\psi}, \delta_{T,*}^{\psi}, \mathbf{Q}_*^{\psi}), \mathbf{Q} - \mathbf{Q}_*^{\psi} \rangle \leq 0$ and (59) allows us to conclude the proof. □

Remark 7: If the algorithm is convergent, i.e., if sequence $(\mathbf{Q}_k)_{k\geq 0}$ converges towards a matrix $\mathbf{P}_*$, Proposition 6 implies that $\mathbf{P}_* = \bar{\mathbf{Q}}_*$. In fact, functions $\mathbf{Q} \mapsto \delta_R(\mathbf{Q})$ and $\mathbf{Q} \mapsto \delta_T(\mathbf{Q})$ are continuous by Proposition 5. As $\delta_{R,k} = \delta_R(\mathbf{Q}_{k-1})$ and $\delta_{T,k} = \delta_T(\mathbf{Q}_{k-1})$, the convergence of $(\mathbf{Q}_k)$ thus implies the convergence of $(\delta_{R,k})$ and $(\delta_{T,k})$, and (56) is fulfilled. Proposition 6 immediately yields $\mathbf{P}_* = \bar{\mathbf{Q}}_*$.

Remark 8: Although we have not been able to prove the convergence of the algorithm, we believe that it can be used in practice because its possible nonconvergence can be easily checked by evaluating $\delta_{R,k} - \delta_{R,k-1}$ and $\delta_{T,k} - \delta_{T,k-1}$ for each k. If one of the above sequences does not converge toward 0, Remark 7 implies that the algorithm does not converge. In this case, a simple solution consists in modifying the initialization point as many times as necessary. We, however, notice that all the numerical experiments we have conducted indicate that the algorithm converges if initialized at $\mathbf{Q}_0 = \mathbf{I}_t$.

VI. NUMERICAL EXPERIMENTS

In this section, we compare the proposed algorithm with Vu and Paulraj's algorithm as presented in [41], and based on the maximization of $I(\mathbf{Q})$.

Recall that Vu-Paulraj's algorithm is based on a Newton method and a barrier interior point method. Moreover, the average mutual informations and their first and second derivatives are evaluated by Monte Carlo simulations. In Fig. 1, we have evaluated $C_E = \max_{\mathbf{Q}\in\mathcal{C}_1} I(\mathbf{Q})$ versus the SNR for $r = t = 4$. Matrix $\mathbf{H}$ coincides with the example considered in [41]. The solid line corresponds to the results provided by the Vu-Paulraj's algorithm; the number of trials used to evaluate the mutual informations and its first and second derivatives is equal to 30.000, and the maximum number of iterations of the algorithm in [41] is fixed to 10. The dashed line corresponds to the results provided by our algorithm: Each point represents $I(\bar{\mathbf{Q}}_*)$ at the corresponding SNR, where $\bar{\mathbf{Q}}_*$ is the argmax of $\bar{I}$; the average mutual information at point $\bar{\mathbf{Q}}_*$ is evaluated by Monte Carlo simulation (30.000 trials are used). The number of iterations is also limited to 10. Fig. 1 shows that our asymptotic approach provides the same results than the Vu-Paulraj's algorithm. However, our algorithm is computationally much more efficient as Table I shows. The table gives the average executation time (in sec) of one iteration for both algorithms for $r = t = 2, r = t = 4, r = t = 8$. We finally notice that the algorithm proposed in [36] provides on the same channel matrix similar results (compare [36, Fig. 2 to Fig. 1]).

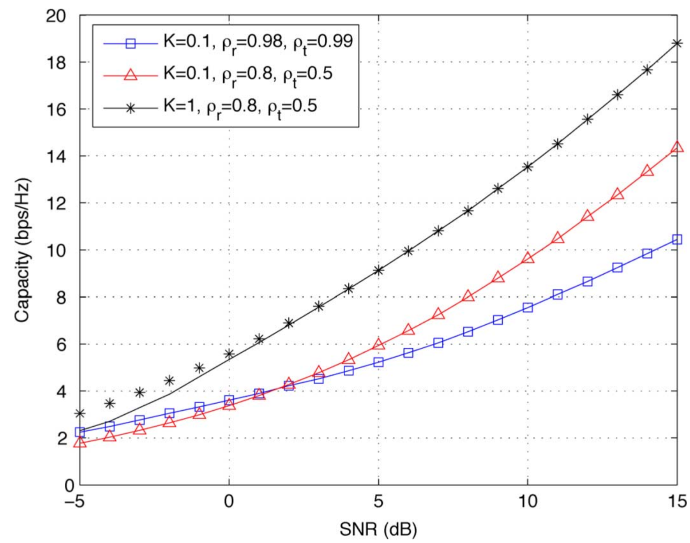

Fig. 2. Comparison with the Vu-Paulraj algorithm II.

In Fig. 2, we again compare Vu-Paulraj's algorithm and our proposal. Matrix $\mathbf{A}$ is generated according to the model

$$\mathbf{A} = \frac{1}{\sqrt{t}}[\mathbf{a}(\theta_1), \ldots, \mathbf{a}(\theta_t)]\mathbf{\Lambda} \tag{61}$$

where $\mathbf{a}(\theta) = (1, e^{i\theta}, \ldots, e^{i(r-1)\theta})^T$ and $\mathbf{\Lambda}$ is a diagonal matrix whose entries represent the complex amplitudes of the t line of sight (LOS) components. The angles of arrivals are chosen randomly according to a uniform distribution. The transmit and receive antennas correlations are exponential with parameter $0 < \rho_T < 1$ and $0 < \rho_R < 1$, respectively. In the experiments, $r = t = 4$, while various values of ρ_T, ρ_R and of the Rice factor K have been considered. As in the previous experiment, the maximum number of iterations for both algorithms is 10, while the number of trials generated to evaluate the average mutual informations and their derivatives is equal to 30.000. Our approach again provides the same results than Vu-Paulraj's algorithm, except for low SNRs for $K = 1, \rho_T = 0.5, \rho_R = 0.8$ where our method gives better results: at these points, the Vu-Paulraj's algorithm seems not to have converged at the 10th iteration.

VII. Conclusion

In this paper, the accuracy of the large system approximation of the ergodic mutual information for Rician MIMO channels with transmit and receive antenna correlation is studied. It is shown that the relative error of the approximation is a $O(\frac{1}{t^2})$ term. The approximation of the EMI is exploited to derive an efficient optimization algorithm providing an approximation of the optimum covariance matrix and of the capacity of the channel. The relative errors of these approximations are also $O(\frac{1}{t^2})$ terms.

Appendix I
Proof of the Existence and Uniqueness of (13)

We consider functions $g(\kappa,\tilde{\kappa})$ and $\tilde{g}(\kappa,\tilde{\kappa})$ defined by

$$\begin{aligned} g(\kappa,\tilde{\kappa}) &= \frac{1}{\kappa}\frac{1}{t}\,\mathrm{Tr}\,[\mathbf{D}(\sigma^2(\mathbf{I}_r+\mathbf{D}\tilde{\kappa})+\mathbf{B}(\mathbf{I}_t+\tilde{\mathbf{D}}\kappa)^{-1}\mathbf{B}^H)^{-1}] \\ \tilde{g}(\kappa,\tilde{\kappa}) &= \frac{1}{\tilde{\kappa}}\frac{1}{t}\,\mathrm{Tr}\,[\tilde{\mathbf{D}}(\sigma^2(\mathbf{I}_t+\tilde{\mathbf{D}}\kappa)+\mathbf{B}^H(\mathbf{I}_r+\mathbf{D}\tilde{\kappa})^{-1}\mathbf{B})^{-1}]. \end{aligned} \tag{62}$$

We have to establish that (62) has a pair of strictly positive solutions, and that this particular pair is unique [i.e., there is no other strictly positive pair satisfying (62)]. For this, we **construct** a strictly positive pair satisfying the equation. This shows the existence. The uniqueness is an easy consequence of the construction above, and is therefore omitted.

In order to construct the pair of strictly positive solutions, we first remark that for each $\tilde{\kappa} > 0$ fixed, function $\kappa \to g(\kappa,\tilde{\kappa})$ is clearly strictly decreasing, converges toward ∞ if $\kappa \to 0$ and converges to 0 if $\kappa \to \infty$. Therefore, there exists a unique $\kappa > 0$ satisfying $g(\kappa,\tilde{\kappa}) = 1$. As this solution depends on $\tilde{\kappa}$, it is denoted $h(\tilde{\kappa})$ in the following. We claim that

- (i) Function $\tilde{\kappa} \to h(\tilde{\kappa})$ is strictly decreasing for $\tilde{\kappa} > 0$,
- (ii) Function $\tilde{\kappa} \to \tilde{\kappa}h(\tilde{\kappa})$ is strictly increasing for $\tilde{\kappa} > 0$.

In fact, consider $\tilde{\kappa}_2 > \tilde{\kappa}_1 > 0$. It is easily checked that for each $\kappa > 0, g(\kappa,\tilde{\kappa}_1) > g(\kappa,\tilde{\kappa}_2)$. Hence, the solution $h(\tilde{\kappa}_1)$ and $h(\tilde{\kappa}_2)$ of the equations $g(\kappa,\tilde{\kappa}_1) = 1$ and $g(\kappa,\tilde{\kappa}_2) = 1$ satisfy $h(\tilde{\kappa}_1) > h(\tilde{\kappa}_2)$. This establishes (i). To prove (ii), we use the obvious relation $g(h(\tilde{\kappa}_1),\tilde{\kappa}_1) - g(h(\tilde{\kappa}_2),\tilde{\kappa}_2) = 0$. We denote by $(\mathbf{U}_i)_{i=1,2}$ the matrices

$$\mathbf{U}_i = \sigma^2\left(h(\tilde{\kappa}_i)\mathbf{I} + \tilde{\kappa}_i h(\tilde{\kappa}_i)\mathbf{D}\right) + \mathbf{B}\left(\frac{\mathbf{I}}{h(\tilde{\kappa}_i)} + \tilde{\mathbf{D}}\right)^{-1}\mathbf{B}^H$$

It is clear that $g(h(\tilde{\kappa}_i),\tilde{\kappa}_i) = \frac{1}{t}\,\mathrm{Tr}\,\mathbf{D}\mathbf{U}_i^{-1}$. We express $g(h(\tilde{\kappa}_1),\tilde{\kappa}_1) - g(h(\tilde{\kappa}_2),\tilde{\kappa}_2)$ as

$$g(h(\tilde{\kappa}_1),\tilde{\kappa}_1) - g(h(\tilde{\kappa}_2),\tilde{\kappa}_2) = \frac{1}{t}\,\mathrm{Tr}\,\mathbf{D}\left(\mathbf{U}_1^{-1} - \mathbf{U}_2^{-1}\right)$$

and use the identity

$$\mathbf{U}_1^{-1} - \mathbf{U}_2^{-1} = \mathbf{U}_1^{-1}(\mathbf{U}_2 - \mathbf{U}_1)\mathbf{U}_2^{-1}. \tag{63}$$

Using the form of matrices $(\mathbf{U}_i)_{i=1,2}$, we eventually obtain that

$$\begin{aligned} &g(h(\tilde{\kappa}_1),\tilde{\kappa}_1) - g(h(\tilde{\kappa}_2),\tilde{\kappa}_2) \\ &\quad = u(h(\tilde{\kappa}_2) - h(\tilde{\kappa}_1)) + v(\tilde{\kappa}_2 h(\tilde{\kappa}_2) - \tilde{\kappa}_1 h(\tilde{\kappa}_1)) \end{aligned}$$

where u and v are the strictly positive terms defined by

$$\begin{aligned} u = \frac{1}{t}\,\mathrm{Tr}\,\mathbf{D}\mathbf{U}_1^{-1}&(\sigma^2\mathbf{I} + \mathbf{B}(\mathbf{I} + h(\tilde{\kappa}_2)\tilde{\mathbf{D}})^{-1} \\ &\times (\mathbf{I} + h(\tilde{\kappa}_1)\tilde{\mathbf{D}})^{-1}\mathbf{B}^H)\mathbf{U}_2^{-1} \end{aligned} \tag{64}$$

and

$$v = \frac{1}{t}\,\mathrm{Tr}\,\mathbf{D}\mathbf{U}_1^{-1}\mathbf{D}\mathbf{U}_2^{-1}.$$

As $u(h(\tilde{\kappa}_2) - h(\tilde{\kappa}_1)) + v(\tilde{\kappa}_2 h(\tilde{\kappa}_2) - \tilde{\kappa}_1 h(\tilde{\kappa}_1)) = 0$, $(h(\tilde{\kappa}_2) - h(\tilde{\kappa}_1)) < 0$ implies that $\tilde{\kappa}_2 h(\tilde{\kappa}_2) - \tilde{\kappa}_1 h(\tilde{\kappa}_1) > 0$. Hence, $\tilde{\kappa}h(\tilde{\kappa})$ is a strictly increasing function as expected. From this, it follows that function $\tilde{\kappa} \to \tilde{g}(h(\tilde{\kappa}),\tilde{\kappa})$ is strictly decreasing for $\tilde{\kappa} > 0$. This function converges to ∞ if $\tilde{\kappa} \to 0$ and to 0 if $\tilde{\kappa} \to \infty$. Moreover, it is easily seen that function h is continuous. Therefore, function $\tilde{\kappa} \to \tilde{g}(h(\tilde{\kappa}),\tilde{\kappa})$ is itself continuous, so that the equation with respect to $\tilde{\kappa}$

$$\tilde{g}(h(\tilde{\kappa}),\tilde{\kappa}) = 1$$

has a unique strictly positive solution $\tilde{\beta}$. Denote by β the strictly positive term $\beta = h(\tilde{\beta})$. It is clear that $g(\beta,\tilde{\beta}) = \tilde{g}(\beta,\tilde{\beta}) = 1$ or equivalently that $\beta = f(\beta,\tilde{\beta})$ and $\tilde{\beta} = \tilde{f}(\beta,\tilde{\beta})$. We have therefore shown that $(\beta,\tilde{\beta})$ is a strictly positive pair solution of (13).

Appendix II
Proof of Theorem 2

This section is organized as follows. We first recall in Section II-A some useful mathematical tools. In Section II-B, we establish (25). In Section II-C, we prove (26) and (27).

We shall use the following notations. If u is a random variable, the zero mean random variable $u - \mathbb{E}(u)$ is denoted by $\overset{\circ}{u}$. If $z = x + \mathbf{i}y$ is a complex number, the differential operators $\frac{\partial}{\partial z}$ and $\frac{\partial}{\partial \bar{z}}$ are defined, respectively, by $\frac{1}{2}(\frac{\partial}{\partial x} - \mathbf{i}\frac{\partial}{\partial y})$ and $\frac{1}{2}(\frac{\partial}{\partial x} + \mathbf{i}\frac{\partial}{\partial y})$. Finally, if $\mathbf{\Sigma}, \mathbf{B}, \mathbf{Y}$ are given matrices, we denote, respectively, by $\boldsymbol{\xi}_j, \mathbf{b}_j, \mathbf{y}_j$ their columns.

A. Mathematical Tools

1) The Poincaré-Nash Inequality: (see, e.g., [8] and [22]). Let $\mathbf{x} = [x_1,\ldots,x_M]^T$ be a complex Gaussian random vector whose law is given by $\mathbb{E}[\mathbf{x}] = \mathbf{0}, \mathbb{E}[\mathbf{x}\mathbf{x}^T] = \mathbf{0}$, and $\mathbb{E}[\mathbf{x}\mathbf{x}^*] = \mathbf{\Xi}$.

Let $\Phi = \Phi(x_1,\ldots,x_M,\bar{x}_1,\ldots,\bar{x}_M)$ be a C^1 complex function polynomially bounded together with its partial derivatives. Then the following inequality holds true:

$$\begin{aligned} \mathrm{Var}(\Phi(\mathbf{x})) \le \mathbb{E}&\left[\nabla_z\Phi(\mathbf{x})^T\,\mathbf{\Xi}\,\overline{\nabla_z\Phi(\mathbf{x})}\right] \\ &+ \mathbb{E}\left[(\nabla_{\bar{z}}\Phi(\mathbf{x}))^H\,\mathbf{\Xi}\nabla_{\bar{z}}\Phi(\mathbf{x})\right] \end{aligned} \tag{65}$$

where $\nabla_z\Phi = [\partial\Phi/\partial z_1,\ldots,\partial\Phi/\partial z_M]^T$ and $\nabla_{\bar{z}}\Phi = [\partial\Phi/\partial\bar{z}_1,\ldots,\partial\Phi/\partial\bar{z}_M]^T$. Let $\mathbf{Y}$ be the $r \times t$ matrix $\mathbf{Y} = \frac{1}{\sqrt{t}}\mathbf{D}^{\frac{1}{2}}\mathbf{X}\tilde{\mathbf{D}}^{\frac{1}{2}}$, where $\mathbf{X}$ has i.i.d. $\mathcal{CN}(0,1)$ entries and consider the stacked $rt \times 1$ vector $\mathbf{x} = [Y_{11},\ldots,Y_{rt}]^T$. In this case, Poincaré-Nash inequality writes

$$\mathrm{Var}\,(\Phi(\mathbf{Y})) \le \frac{1}{t}\sum_{i=1}^{r}\sum_{j=1}^{t} d_i\tilde{d}_j\mathbb{E}\left[\left|\frac{\partial\Phi(\mathbf{Y})}{\partial Y_{i,j}}\right|^2 + \left|\frac{\partial\Phi(\mathbf{Y})}{\partial\bar{Y}_{i,j}}\right|^2\right]. \tag{66}$$

2) The Differentiation Formula for Functions of Gaussian Random Vectors: With $\mathbf{x}$ and Φ given as above, we have the following

$$\mathbb{E}[x_p \Phi(\mathbf{x})] = \sum_{m=1}^{M} [\mathbf{\Xi}]_{pm} \mathbb{E}\left[\frac{\partial \Phi(\mathbf{x})}{\partial \bar{x}_m}\right]. \tag{67}$$

This formula relies on an integration by parts, and is thus referred to as the Integration by parts formula for Gaussian vectors. It is widely used in Mathematical Physics ([15]) and has been used in Random Matrix Theory in [26] and [31].

If $\mathbf{x}$ coincides with the $rt \times 1$ vector $\mathbf{x} = [Y_{11}, \ldots, Y_{rt}]^T$, (67) becomes

$$\mathbb{E}[Y_{pq}\Phi(\mathbf{Y})] = \frac{d_p \tilde{d}_q}{t} \mathbb{E}\left[\frac{\partial \Phi(\mathbf{Y})}{\partial \bar{Y}_{pq}}\right]. \tag{68}$$

Replacing matrix $\mathbf{Y}$ by matrix $\bar{\mathbf{Y}}$ also provides

$$\mathbb{E}[\bar{Y}_{pq}\Phi(\mathbf{Y})] = \frac{d_p \tilde{d}_q}{t} \mathbb{E}\left[\frac{\partial \Phi(\mathbf{Y})}{\partial Y_{pq}}\right]. \tag{69}$$

3) Some Useful Differentiation Formulas: The following partial derivatives $\frac{\partial(\mathbf{S}_{pq})}{\partial Y_{ij}}$ and $\frac{\partial \mathbf{S}_{pq}}{\partial \bar{Y}_{ij}}$ for each $p, q \in \{1, \ldots, r\}$ and $1 \le i \le r, 1 \le j \le t$ will be of use in the sequel. Straightforward computations yield

$$\begin{cases} \frac{\partial S_{pq}}{\partial Y_{ij}} = -S_{p,i}\left(\boldsymbol{\xi}_j^H \mathbf{S}\right)_q \\ \frac{\partial S_{pq}}{\partial \bar{Y}_{ij}} = -S_{i,q}\left(\mathbf{S}\boldsymbol{\xi}\right)_p. \end{cases} \tag{70}$$

B. Proof of (25)

We just prove that the variance of $\frac{1}{t}\mathrm{Tr}(\mathbf{MS})$ is a $O(t^{-2})$ term. For this, we note that the random variable $\frac{1}{t}\mathrm{Tr}(\mathbf{MS})$ can be interpreted as a function $\Phi(\mathbf{Y})$ of the entries of matrix $\mathbf{Y}$, and use the Poincaré-Nash inequality (66) to $\Phi(\mathbf{Y})$. Function $\Phi(\mathbf{Y})$ is equal to

$$\Phi(\mathbf{Y}) = \frac{1}{t}\sum_{p,q} M_{q,p} S_{p,q}.$$

Therefore, the partial derivative of $\Phi(\mathbf{Y})$ with respect to Y_{ij} is given by $\frac{\partial \Phi(\mathbf{Y})}{\partial Y_{ij}} = \frac{1}{t}\sum_{p,q} \mathbf{M}_{q,p}\frac{\partial S_{pq}}{\partial Y_{ij}}$ which, by (70), coincides with

$$\frac{\partial \Phi(\mathbf{Y})}{\partial Y_{ij}} = -\frac{1}{t}\sum_{p,q} \mathbf{M}_{q,p}\mathbf{S}_{p,i}\left(\boldsymbol{\xi}_j^H \mathbf{S}\right)_q = -\frac{1}{t}\left(\boldsymbol{\xi}_j^H \mathbf{SMS}\right)_i.$$

As $d_i \le d_{\max}$ and $\tilde{d}_j \le \tilde{d}_{\max}$, it is clear that

$$\sum_{i=1}^{r}\sum_{j=1}^{t} d_i \tilde{d}_j \mathbb{E}\left|\frac{\partial \Phi(\mathbf{Y})}{\partial Y_{ij}}\right|^2 \le d_{\max}\tilde{d}_{\max}\sum_{i=1}^{r}\sum_{j=1}^{t}\mathbb{E}\left|\frac{\partial \Phi(\mathbf{Y})}{\partial Y_{ij}}\right|^2.$$

It is easily seen that

$$\sum_{i=1}^{r}\mathbb{E}\left|\frac{\partial \Phi(\mathbf{Y})}{\partial Y_{ij}}\right|^2 = \frac{1}{t^2}\mathbb{E}\left(\boldsymbol{\xi}_j^H \mathbf{SMS}^2\mathbf{M}^H\mathbf{S}\boldsymbol{\xi}_j^H\right).$$

As $\|\mathbf{S}\| \le \frac{1}{\sigma^2}$ and $\sup_t \|\mathbf{M}\| < \infty, \boldsymbol{\xi}_j^H \mathbf{SMS}^2\mathbf{M}^H\mathbf{S}\boldsymbol{\xi}_j^H$ is less than $\frac{1}{\sigma^8}\sup_t \|\mathbf{M}\|^2\|\boldsymbol{\xi}_j\|^2$. Moreover, $\mathbb{E}\|\boldsymbol{\xi}_j\|^2$ coincides with $\|\mathbf{b}_j\|^2 + \frac{1}{t}\tilde{d}_j\sum_{i=1}^{r} d_i$, which is itself less than $b_{\max}^2 + d_{\max}\tilde{d}_{\max}\frac{r}{t}$, a uniformly bounded term. Therefore, $\sum_{i=1}^{r}\mathbb{E}|\frac{\partial \Phi(\mathbf{Y})}{\partial Y_{ij}}|^2$ is a $O(t^{-2})$ term. This proves that

$$\frac{1}{t}\sum_{i=1}^{r}\sum_{j=1}^{t} d_i \tilde{d}_j \mathbb{E}\left|\frac{\partial \Phi(\mathbf{Y})}{\partial Y_{ij}}\right|^2 = O\left(\frac{1}{t^2}\right).$$

It can be shown similarly that $t^{-1}\sum_{i=1}^{r}\sum_{j=1}^{t} d_i\tilde{d}_j\mathbb{E}|\frac{\partial \Phi(\mathbf{Y})}{\partial \bar{Y}_{ij}}|^2 = O(t^{-2})$. The conclusion follows from Poincaré-Nash inequality (66).

C. Proof of (26) and (27)

As we shall see, proofs of (26) and (27) are demanding. We first introduce the following notations: Define scalar parameters $\eta(\sigma^2), \alpha(\sigma^2), \tilde{\alpha}(\sigma^2)$ as

$$\begin{aligned} \eta(\sigma^2) &= \frac{1}{t}\,\mathrm{Tr}\,(\mathbf{DS}(\sigma^2)) \\ \alpha(\sigma^2) &= \mathbb{E}\left[\frac{1}{t}\,\mathrm{Tr}\,(\mathbf{DS}(\sigma^2))\right] \\ \tilde{\alpha}(\sigma^2) &= \mathbb{E}\left[\frac{1}{t}\,\mathrm{Tr}\,(\tilde{\mathbf{D}}\tilde{\mathbf{S}}(\sigma^2))\right] \end{aligned} \tag{71}$$

and matrices $\mathbf{R}(\sigma^2), \tilde{\mathbf{R}}(\sigma^2)$ as

$$\begin{aligned} \mathbf{R}(\sigma^2) &= [\sigma^2(\mathbf{I} + \tilde{\alpha}\mathbf{D}) + \mathbf{B}(\mathbf{I} + \alpha\tilde{\mathbf{D}})^{-1}\mathbf{B}^H]^{-1} \\ \tilde{\mathbf{R}}(\sigma^2) &= [\sigma^2(\mathbf{I} + \alpha\tilde{\mathbf{D}}) + \mathbf{B}^H(\mathbf{I} + \tilde{\alpha}\mathbf{D})^{-1}\mathbf{B}]^{-1}. \end{aligned} \tag{72}$$

We note that, as $\alpha(\sigma^2) \ge 0$ and $\tilde{\alpha}(\sigma^2) \ge 0$, then

$$0 < \mathbf{R}(\sigma^2) < \frac{\mathbf{I}_r}{\sigma^2}, 0 < \tilde{\mathbf{R}}(\sigma^2) < \frac{\mathbf{I}_t}{\sigma^2}. \tag{73}$$

It is difficult to study directly the term $\frac{1}{r}\mathrm{Tr}\mathbf{M}(\mathbb{E}(\mathbf{S}) - \mathbf{T})$. In some sense, matrix $\mathbf{R}$ can be seen as an intermediate quantity between $\mathbb{E}(\mathbf{S})$ and $\mathbf{T}$. Thus the proof consists into two steps: 1) for each uniformly bounded matrix $\mathbf{M}$, we first prove that $\frac{1}{r}\mathrm{Tr}\mathbf{M}(\mathbb{E}(\mathbf{S}) - \mathbf{R})$ and $\frac{1}{r}\,\mathrm{Tr}\,\mathbf{M}(\mathbf{R} - \mathbf{T})$ converge to 0 as $t \to \infty$; 2) we then refine the previous result and establish in fact that $\frac{1}{r}\mathrm{Tr}\mathbf{M}(\mathbb{E}(\mathbf{S}) - \mathbf{R})$ and $\frac{1}{r}\,\mathrm{Tr}\,\mathbf{M}(\mathbf{R} - \mathbf{T})$ are $O(t^{-2})$ terms. This, of course, imply (26). Equation (27) eventually follows from (26), the integral representation

$$\bar{J}(\sigma^2) - J(\sigma^2) = \int_{\sigma^2}^{\infty} \mathrm{Tr}\,(\mathrm{E}(\mathbf{S}(\omega)) - \mathbf{T}(\omega))\,d\omega \tag{74}$$

which follows from (20) and (21), as well as a dominated convergence argument that is omitted.

1) First Step: Convergence of $\frac{1}{r}\mathrm{Tr}\mathbf{M}(\mathbb{E}(\mathbf{S}) - \mathbf{R})$ and $\frac{1}{r}\,\mathrm{Tr}\,\mathbf{M}(\mathbf{R} - \mathbf{T})$ to Zero: The first step consists in showing the following Proposition.

Proposition 7: For each deterministic $r \times r$ matrix $\mathbf{M}$, uniformly bounded (for the spectral norm) as $t \to \infty$, we have

$$\lim_{t \to +\infty} \frac{1}{t}\mathrm{Tr}[\mathbf{M}(\mathbb{E}(\mathbf{S}) - \mathbf{R})] = 0 \tag{75}$$

$$\lim_{t \to +\infty} \frac{1}{t}\mathrm{Tr}[\mathbf{M}(\mathbf{R}) - \mathbf{T}] = 0. \tag{76}$$

Proof: We first prove (75). For this, we state the following useful Lemma.

Lemma 2: Let $\mathbf{P}, \mathbf{P}_1$ and $\mathbf{P}_2$ be deterministic $r\times t, t\times t, t\times r$ matrices, respectively, uniformly bounded with respect to the spectral norm as $t \to \infty$. Consider the following functions of $\mathbf{Y}$:

$$\begin{aligned}\mathbf{\Phi}(\mathbf{Y}) &= \frac{1}{t}\ \mathrm{Tr}\ [\mathbf{S}\mathbf{P}\mathbf{\Sigma}^H] \\ \mathbf{\Psi}(\mathbf{Y}) &= \frac{1}{t}\ \mathrm{Tr}\ [\mathbf{S}\mathbf{\Sigma}\mathbf{P}_1\mathbf{\Sigma}^H\mathbf{P}_2] \\ \mathbf{\Psi}'(\mathbf{Y}) &= \frac{1}{t}\ \mathrm{Tr}\ [\mathbf{S}\mathbf{\Sigma}\mathbf{P}_1\mathbf{Y}^H\mathbf{P}_2]. \end{aligned} \tag{77}$$

Then, the following estimates hold true:

$$\mathrm{Var}\ (\mathbf{\Phi}) = O\left(\frac{1}{t^2}\right), \mathrm{Var}\ (\mathbf{\Psi}) = O\left(\frac{1}{t^2}\right)$$
$$\mathrm{Var}\ (\mathbf{\Psi}') = O\left(\frac{1}{t^2}\right)$$

The proof, based on the Poincaré-Nash inequality (66), is omitted.

In order to use the Integration by parts formula (68), notice that

$$\sigma^2\mathbf{S}(\sigma^2) + \mathbf{S}(\sigma^2)\mathbf{\Sigma}\mathbf{\Sigma}^H = \mathbf{I}. \tag{78}$$

Taking the mathematical expectation, we have for each $p,q \in \{1,\dots,r\}$

$$\sigma^2\mathbb{E}(S_{pq}) + \mathbb{E}[(\mathbf{S}\mathbf{\Sigma}\mathbf{\Sigma}^H)_{pq}] = \delta(p-q). \tag{79}$$

A convenient use of the Integration by parts formula allows to express $\mathbb{E}[(\mathbf{S}\mathbf{\Sigma}\mathbf{\Sigma}^H)_{pq}]$ in terms of the entries of $\mathbb{E}(\mathbf{S})$. To see this, note that

$$\mathbb{E}[(\mathbf{S}\mathbf{\Sigma}\mathbf{\Sigma}^H)_{pq}] = \sum_{j=1}^{t}\sum_{i=1}^{r}\mathbb{E}(S_{pi}\Sigma_{ij}\overline{\Sigma_{qj}}).$$

For each $i, \mathbb{E}(S_{pi}\Sigma_{ij}\overline{\Sigma_{qj}})$ can be written as

$$\mathbb{E}(S_{pi}\Sigma_{ij}\overline{\Sigma_{qj}}) = \mathbb{E}(S_{pi})B_{ij}\overline{B_{qj}} + \mathbb{E}(S_{pi}\overline{Y_{qj}})B_{ij} + \mathbb{E}(S_{pi}Y_{ij}\overline{\Sigma_{qj}}). \tag{80}$$

Using (68) with function $\Phi(\mathbf{Y}) = S_{pi}\overline{\Sigma_{qj}}$ and (69) with $\Phi(\mathbf{Y}) = S_{pi}$, and summing over index i yields

$$\begin{aligned}\mathbb{E}[(\mathbf{S}\boldsymbol{\xi}_j)_p\overline{\Sigma_{q,j}}] &= \frac{d_q\tilde{d}_j}{t}\mathbb{E}(S_{pq}) - \tilde{d}_j\mathbb{E}[\eta(\mathbf{S}\boldsymbol{\xi}_j)_p\overline{\Sigma_{q,j}}] \\ &\quad -\frac{d_q\tilde{d}_j}{t}\mathbb{E}\left[S_{pq}\boldsymbol{\xi}_j^H\mathbf{S}\mathbf{b}_j\right] + \mathbb{E}\left[(\mathbf{S}\mathbf{b}_j)_p\right]\overline{B_{qj}}. \end{aligned} \tag{81}$$

Equation (25) for $\mathbf{M} = \mathbf{D}$ implies that $\mathrm{Var}(\eta) = O(t^{-2})$, or equivalently that $\mathbb{E}(\overset{\circ}{\eta}{}^2) = O(t^{-2})$. We now complete proof of (75). We take (81) as a starting point, and write η as $\eta = \mathbb{E}(\eta) + \overset{\circ}{\eta} = \alpha + \overset{\circ}{\eta}$. Therefore,

$$\mathbb{E}[\eta\,(\mathbf{S}\boldsymbol{\xi}_j)_p\,\overline{\Sigma_{q,j}}] = \alpha\mathbb{E}[(\mathbf{S}\boldsymbol{\xi}_j)_p\,\overline{\Sigma_{q,j}}] + \mathbb{E}[\overset{\circ}{\eta}\ (\mathbf{S}\boldsymbol{\xi}_j)_p\,\Sigma_{q,j}].$$

Plugging this relation into (81), and solving with respect to $\mathbb{E}[(\mathbf{S}\boldsymbol{\xi}_j)_p\,\overline{\Sigma_{q,j}}]$ yields

$$\begin{aligned}\mathbb{E}[(\mathbf{S}\boldsymbol{\xi}_j)_p\overline{\Sigma_{q,j}}] &= \frac{1}{t}\frac{d_q\tilde{d}_j}{1+\alpha\tilde{d}_j}\mathbb{E}(S_{pq}) + \frac{1}{1+\alpha\tilde{d}_j}\mathbb{E}[(\mathbf{S}\mathbf{b}_j)_p]\overline{B_{qj}} \\ &\quad - \frac{1}{t}\frac{d_q\tilde{d}_j}{1+\alpha\tilde{d}_j}\mathbb{E}\left[S_{pq}\boldsymbol{\xi}_j^H\mathbf{S}\mathbf{b}_j\right] \\ &\quad - \frac{\tilde{d}_j}{1+\alpha\tilde{d}_j}\mathbb{E}[\overset{\circ}{\eta}\ (\mathbf{S}\boldsymbol{\xi}_j)_p\,\overline{\Sigma_{q,j}}]. \end{aligned} \tag{82}$$

Writing $\boldsymbol{\xi}_j = \mathbf{b}_j + \mathbf{y}_j$, and summing over j provides the following expression of $\mathbb{E}[(\mathbf{S}\mathbf{\Sigma}\mathbf{\Sigma}^H)_{pq}]$:

$$\begin{aligned}&\mathbb{E}[(\mathbf{S}\mathbf{\Sigma}\mathbf{\Sigma}^H)_{pq}] \\ &\quad= d_q\frac{1}{t}\ \mathrm{Tr}\ [\tilde{\mathbf{D}}(\mathbf{I}+\alpha\tilde{\mathbf{D}})^{-1}]\mathbb{E}(S_{pq}) \\ &\qquad + \mathbb{E}[(\mathbf{S}\mathbf{B}(\mathbf{I}+\alpha\tilde{\mathbf{D}})^{-1}\mathbf{B}^H)_{pq}] \\ &\qquad - d_q\,\mathbb{E}\left[S_{pq}\frac{1}{t}\ \mathrm{Tr}\ (\mathbf{S}\mathbf{B}\tilde{\mathbf{D}}(\mathbf{I}+\alpha\tilde{\mathbf{D}})^{-1}\mathbf{B}^H)\right] \\ &\qquad - d_q\,\mathbb{E}\left[S_{pq}\frac{1}{t}\ \mathrm{Tr}\ (\mathbf{S}\mathbf{B}\tilde{\mathbf{D}}(\mathbf{I}+\alpha\tilde{\mathbf{D}})^{-1}\mathbf{Y}^H)\right] \\ &\qquad - \mathbb{E}\left[\overset{\circ}{\eta}\ (\mathbf{S}\mathbf{\Sigma}\tilde{\mathbf{D}}(\mathbf{I}+\alpha\tilde{\mathbf{D}})^{-1}\mathbf{\Sigma}^H)_{p,q}\right]. \end{aligned} \tag{83}$$

The resolvent identity (78) thus implies that

$$\begin{aligned}\delta(p-q) &= \sigma^2\mathbb{E}(S_{pq}) + \frac{d_q}{t}\ \mathrm{Tr}\ [\tilde{\mathbf{D}}(\mathbf{I}+\alpha\tilde{\mathbf{D}})^{-1}]\mathbb{E}(S_{pq}) \\ &\quad + \mathbb{E}[(\mathbf{S}\mathbf{B}(\mathbf{I}+\alpha\tilde{\mathbf{D}})^{-1}\mathbf{B}^H)_{pq}] \\ &\quad - d_q\,\mathbb{E}\left[S_{pq}\frac{1}{t}\ \mathrm{Tr}\ (\mathbf{S}\mathbf{B}\tilde{\mathbf{D}}(\mathbf{I}+\alpha\tilde{\mathbf{D}})^{-1}\mathbf{B}^H)\right] \\ &\quad - d_q\,\mathbb{E}\left[S_{pq}\frac{1}{t}\ \mathrm{Tr}\ (\mathbf{S}\mathbf{B}\tilde{\mathbf{D}}(\mathbf{I}+\alpha\tilde{\mathbf{D}})^{-1}\mathbf{Y}^H)\right] \\ &\quad - \mathbb{E}\left[\overset{\circ}{\eta}\ (\mathbf{S}\mathbf{\Sigma}\tilde{\mathbf{D}}(\mathbf{I}+\alpha\tilde{\mathbf{D}})^{-1}\mathbf{\Sigma}^H)_{p,q}\right]. \end{aligned} \tag{84}$$

In order to simplify the notations, we define ρ_1 and ρ_2 by

$$\begin{aligned}\rho_1 &= \frac{1}{t}\ \mathrm{Tr}\ (\mathbf{S}\mathbf{B}\tilde{\mathbf{D}}(\mathbf{I}+\alpha\tilde{\mathbf{D}})^{-1}\mathbf{B}^H) \\ \rho_2 &= \frac{1}{t}\ \mathrm{Tr}\ (\mathbf{S}\mathbf{B}\tilde{\mathbf{D}}(\mathbf{I}+\alpha\tilde{\mathbf{D}})^{-1}\mathbf{Y}^H)\end{aligned}$$

For $i = 1, 2$, we write $\mathbb{E}(S_{pq}\rho_i)$ as

$$\mathbb{E}(S_{pq}\rho_i) = \mathbb{E}(S_{pq})\,\mathbb{E}(\rho_i) + \mathbb{E}(\overset{\circ}{S}_{pq}\overset{\circ}{\rho}_i).$$

Thus, (84) can be written as

$$\begin{aligned}&\delta(p-q) \\ &\quad= \sigma^2\mathbb{E}(S_{pq}) + d_q\,\frac{1}{t}\ \mathrm{Tr}\ [\tilde{\mathbf{D}}(\mathbf{I}+\alpha\tilde{\mathbf{D}})^{-1}]\mathbb{E}(S_{pq}) \\ &\qquad + (\mathbb{E}(\mathbf{S})\mathbf{B}(\mathbf{I}+\alpha\tilde{\mathbf{D}})^{-1}\mathbf{B}^H)_{pq} \\ &\qquad - d_q\,\mathbb{E}(S_{pq})\frac{1}{t}\ \mathrm{Tr}\ (\mathbb{E}(\mathbf{S})\mathbf{B}\tilde{\mathbf{D}}(\mathbf{I}+\alpha\tilde{\mathbf{D}})^{-1}\mathbf{B}^H) \\ &\qquad - d_q\,\mathbb{E}(S_{pq})\mathbb{E}\left[\frac{1}{t}\ \mathrm{Tr}\ (\mathbf{S}\mathbf{B}\tilde{\mathbf{D}}(\mathbf{I}+\alpha\tilde{\mathbf{D}})^{-1}\mathbf{Y}^H)\right] \\ &\qquad - d_q\,\mathbb{E}(\overset{\circ}{S}_{pq}\overset{\circ}{\rho}_1) \\ &\qquad - d_q\,\mathbb{E}(\overset{\circ}{S}_{pq}\overset{\circ}{\rho}_2) - \mathbb{E}\left[\overset{\circ}{\eta}\ (\mathbf{S}\mathbf{\Sigma}\tilde{\mathbf{D}}(\mathbf{I}+\alpha\tilde{\mathbf{D}})^{-1}\mathbf{\Sigma}^H)_{p,q}\right]. \end{aligned} \tag{85}$$

We now establish the following lemma.

Lemma 3:

$$\begin{aligned}\mathbb{E}\rho_2 &= \mathbb{E}\left[\frac{1}{t}\ \mathrm{Tr}\ (\mathbf{S}\mathbf{B}\tilde{\mathbf{D}}(\mathbf{I}+\alpha\tilde{\mathbf{D}})^{-1}\mathbf{Y}^H)\right] \\ &= -\alpha\,\frac{1}{t}\ \mathrm{Tr}\ (\mathbb{E}(\mathbf{S})\mathbf{B}\tilde{\mathbf{D}}^2(\mathbf{I}+\alpha\tilde{\mathbf{D}})^{-2}\mathbf{B}^H) - \mathbb{E}(\overset{\circ}{\eta}\overset{\circ}{\rho}_3)\end{aligned} \tag{86}$$

where ρ_3 is defined by

$$\rho_3 = \frac{1}{t}\,\mathrm{Tr}\,(\mathbf{SB}\tilde{\mathbf{D}}^2(\mathbf{I}+\alpha\tilde{\mathbf{D}})^{-2}\mathbf{\Sigma}^H).$$

Proof: We express $\mathbb{E}(\rho_2)$ as

$$\begin{aligned}\mathbb{E}(\rho_2) &= \frac{1}{t}\sum_{j=1}^{t}\frac{\tilde{d}_j}{1+\alpha\tilde{d}_j}\mathbb{E}\left(\mathbf{y}_j^H\mathbf{S}\mathbf{b}_j\right)\\ &= \frac{1}{t}\sum_{j=1}^{t}\frac{\tilde{d}_j}{1+\alpha\tilde{d}_j}\sum_{i=1}^{r}\mathbb{E}((\mathbf{S}\mathbf{b}_j)_i\overline{Y_{ij}}) \qquad (87)\end{aligned}$$

and evaluate $\mathbb{E}((\mathbf{S}\mathbf{b}_j)_i\overline{Y_{ij}})$ using formula (67) for $\Phi(\mathbf{Y}) = (\mathbf{S}\mathbf{b}_j)_i$. This gives

$$\mathbb{E}((\mathbf{S}\mathbf{b}_j)_i\overline{Y_{ij}}) = \frac{1}{t}d_i\tilde{d}_j\sum_{k=1}^{r}\mathbb{E}\left(\frac{\partial S_{ik}}{\partial Y_{ij}}\right)B_{kj}.$$

By (70)

$$\mathbb{E}\left(\frac{\partial S_{ik}}{\partial Y_{ij}}\right) = -\mathbb{E}\left(S_{ii}(\mathbf{b}_j^H\mathbf{S})_k\right) - \mathbb{E}\left(S_{ii}(\mathbf{y}_j^H\mathbf{S})_k\right).$$

Therefore

$$\mathbb{E}\left(\mathbf{y}_j^H\mathbf{S}\mathbf{b}_j\right) = -\tilde{d}_j\mathbb{E}\left(\eta\,\mathbf{b}_j^H\mathbf{S}\mathbf{b}_j\right) - \tilde{d}_j\mathbb{E}\left(\eta\,\mathbf{y}_j^H\mathbf{S}\mathbf{b}_j\right).$$

Writing again $\eta = \mathbb{E}(\eta) + \overset{\circ}{\eta} = \alpha + \overset{\circ}{\eta}$, we get that

$$\begin{aligned}\mathbb{E}\left(\mathbf{y}_j^H\mathbf{S}\mathbf{b}_j\right) &= -\alpha\tilde{d}_j\mathbb{E}\left(\mathbf{b}_j^H\mathbf{S}\mathbf{b}_j\right) - \alpha\tilde{d}_j\mathbb{E}\left(\mathbf{y}_j^H\mathbf{S}\mathbf{b}_j\right)\\ &\quad - \tilde{d}_j\mathbb{E}\left(\overset{\circ}{\eta}\ \mathbf{b}_j^H\mathbf{S}\mathbf{b}_j\right) - \tilde{d}_j\mathbb{E}\left(\overset{\circ}{\eta}\ \mathbf{y}_j^H\mathbf{S}\mathbf{b}_j\right). \qquad (88)\end{aligned}$$

Solving this equation with respect to $\mathbb{E}(\mathbf{y}_j^H\mathbf{S}\mathbf{b}_j)$ yields

$$\begin{aligned}\mathbb{E}\left(\mathbf{y}_j^H\mathbf{S}\mathbf{b}_j\right) &= -\frac{\alpha\tilde{d}_j}{1+\alpha\tilde{d}_j}\mathbb{E}\left(\mathbf{b}_j^H\mathbf{S}\mathbf{b}_j\right) - \frac{\tilde{d}_j}{1+\alpha\tilde{d}_j}\mathbb{E}\left(\overset{\circ}{\eta}\ \mathbf{b}_j^H\mathbf{S}\mathbf{b}_j\right)\\ &\quad - \frac{\tilde{d}_j}{1+\alpha\tilde{d}_j}\mathbb{E}\left(\overset{\circ}{\eta}\ \mathbf{y}_j^H\mathbf{S}\mathbf{b}_j\right) \qquad (89)\end{aligned}$$

or equivalently

$$\mathbb{E}\left(\mathbf{y}_j^H\mathbf{S}\mathbf{b}_j\right) = -\frac{\alpha\tilde{d}_j}{1+\alpha\tilde{d}_j}\mathbb{E}\left(\mathbf{b}_j^H\mathbf{S}\mathbf{b}_j\right) - \frac{\tilde{d}_j}{1+\alpha\tilde{d}_j}\mathbb{E}\left(\overset{\circ}{\eta}\ \boldsymbol{\xi}_j^H\mathbf{S}\mathbf{b}_j\right). \qquad (90)$$

Equation (86) immediately follows from (87), (90), and $\mathbb{E}(\overset{\circ}{\eta}\,\rho_3) = \mathbb{E}(\overset{\circ}{\eta}\,\overset{\circ}{\rho}_3)$. □

Plugging (86) into (85) yields

$$\begin{aligned}&\delta(p-q) + \Delta_{pq}\\ &= \mathbb{E}(S_{pq})\left[\sigma^2 + d_q\left(\left(\frac{1}{t}\,\mathrm{Tr}\,\tilde{\mathbf{D}}(\mathbf{I}+\alpha\tilde{\mathbf{D}}\right)^{-1} - \mathbb{E}(\rho_1)\right.\right.\\ &\qquad\left.\left. + \alpha\frac{1}{t}\,\mathrm{Tr}\,\mathbb{E}(\mathbf{S})\mathbf{B}\tilde{\mathbf{D}}^2(\mathbf{I}+\alpha\tilde{\mathbf{D}})^{-2}\mathbf{B}^H\right)\right]\\ &\quad + [\mathbb{E}(\mathbf{S})\mathbf{B}(\mathbf{I}+\alpha\tilde{\mathbf{D}})^{-1}\mathbf{B}^H]_{pq} \qquad (91)\end{aligned}$$

where $\mathbf{\Delta}$ is the $r\times r$ matrix defined by

$$\begin{aligned}\Delta_{pq} &= \mathbb{E}\left[\overset{\circ}{\eta}\left(\mathbf{S\Sigma}\tilde{\mathbf{D}}(\mathbf{I}+\alpha\tilde{\mathbf{D}})^{-1}\mathbf{\Sigma}^H\right)_{pq}\right]\\ &\quad + d_q\mathbb{E}\left(\overset{\circ}{S}_{pq}\,(\overset{\circ}{\rho}_1 + \overset{\circ}{\rho}_2)\right) - d_q\mathbb{E}(S_{pq})\,\mathbb{E}\left(\overset{\circ}{\eta}\,\overset{\circ}{\rho}_3\right) \qquad (92)\end{aligned}$$

for each p, q or equivalently by

$$\begin{aligned}\mathbf{\Delta} &= \mathbb{E}[\overset{\circ}{\eta}\,(\mathbf{S\Sigma}\tilde{\mathbf{D}}(\mathbf{I}+\alpha\tilde{\mathbf{D}})^{-1}\mathbf{\Sigma}^H)]\\ &\quad + \mathbb{E}((\overset{\circ}{\rho}_1 + \overset{\circ}{\rho}_2)\ \overset{\circ}{\mathbf{S}})\,\mathbf{D} - \mathbb{E}\left(\overset{\circ}{\eta}\,\overset{\circ}{\rho}_3\right)\mathbb{E}(\mathbf{S})\,\mathbf{D}. \qquad (93)\end{aligned}$$

Using $\alpha\tilde{\mathbf{D}}(\mathbf{I}+\alpha\tilde{\mathbf{D}})^{-1} = \mathbf{I} - (\mathbf{I}+\alpha\tilde{\mathbf{D}})^{-1}$, we obtain that

$$\begin{aligned}&\alpha\frac{1}{t}\,\mathrm{Tr}\,(\mathbb{E}(\mathbf{S})\mathbf{B}\tilde{\mathbf{D}}^2(\mathbf{I}+\alpha\tilde{\mathbf{D}})^{-2}\mathbf{B}^H)\\ &\quad = \frac{1}{t}\,\mathrm{Tr}\,(\mathbb{E}(\mathbf{S})\mathbf{B}\tilde{\mathbf{D}}(\mathbf{I}+\alpha\tilde{\mathbf{D}})^{-1}\mathbf{B}^H)\\ &\qquad - \frac{1}{t}\,\mathrm{Tr}\,(\mathbb{E}(\mathbf{S})\mathbf{B}\tilde{\mathbf{D}}(\mathbf{I}+\alpha\tilde{\mathbf{D}})^{-2}\mathbf{B}^H)\\ &\quad = \mathbb{E}(\rho_1) - \frac{1}{t}\,\mathrm{Tr}\,(\mathbb{E}(\mathbf{S})\mathbf{B}\tilde{\mathbf{D}}(\mathbf{I}+\alpha\tilde{\mathbf{D}})^{-2}\mathbf{B}^H). \qquad (94)\end{aligned}$$

Therefore, the term

$$\frac{1}{t}\,\mathrm{Tr}\,\tilde{\mathbf{D}}(\mathbf{I}+\alpha\tilde{\mathbf{D}})^{-1} - \mathbb{E}(\rho_1) + \alpha\frac{1}{t}\,\mathrm{Tr}\,(\mathbb{E}(\mathbf{S})\mathbf{B}\tilde{\mathbf{D}}^2(\mathbf{I}+\alpha\tilde{\mathbf{D}})^{-2}\mathbf{B}^H)$$

is equal to

$$\begin{aligned}&\frac{1}{t}\,\mathrm{Tr}\,\tilde{\mathbf{D}}(\mathbf{I}+\alpha\tilde{\mathbf{D}})^{-1} - \frac{1}{t}\,\mathrm{Tr}\,(\mathbb{E}(\mathbf{S})\mathbf{B}\tilde{\mathbf{D}}(\mathbf{I}+\alpha\tilde{\mathbf{D}})^{-2}\mathbf{B}^H)\\ &\quad = \frac{1}{t}\,\mathrm{Tr}\,[\tilde{\mathbf{D}}(\mathbf{I}+\alpha\tilde{\mathbf{D}})^{-1}(\mathbf{I} - \mathbf{B}^H\mathbb{E}(\mathbf{S})\mathbf{B}(\mathbf{I}+\alpha\tilde{\mathbf{D}})^{-1})]\end{aligned}$$

which, in turn, coincides with $\sigma^2\,\tilde{\tau}$, where $\tilde{\tau}$ is defined by

$$\begin{aligned}\tilde{\tau}(\sigma^2) &= \frac{1}{t}\,\mathrm{Tr}\,[\tilde{\mathbf{D}}(\sigma^2(\mathbf{I}+\alpha\tilde{\mathbf{D}}))^{-1}\\ &\qquad \times(\mathbf{I} - \mathbf{B}^H\,\mathbb{E}(\mathbf{S}(\sigma^2))\,\mathbf{B}(\mathbf{I}+\alpha\tilde{\mathbf{D}})^{-1})] \qquad (95)\end{aligned}$$

Equation (91) is, thus, equivalent to

$$(\mathbb{E}(\mathbf{S})[\sigma^2(\mathbf{I}+\tilde{\tau}\mathbf{D}) + \mathbf{B}(\mathbf{I}+\alpha\tilde{\mathbf{D}})^{-1}\mathbf{B}^H]) = \mathbf{I} + \mathbf{\Delta} \qquad (96)$$

or equivalently to

$$\begin{aligned}&(\mathbb{E}[\mathbf{S}(\sigma^2(\mathbf{I}+\tilde{\alpha}\mathbf{D}) + \mathbf{B}(\mathbf{I}+\alpha\tilde{\mathbf{D}})^{-1}\mathbf{B}^H)])\\ &\qquad = \mathbf{I} + \sigma^2(\tilde{\alpha} - \tilde{\tau})\mathbb{E}(\mathbf{S})\mathbf{D} + \mathbf{\Delta} \qquad (97)\end{aligned}$$

or to

$$\mathbb{E}(\mathbf{S}) = \mathbf{R} + \sigma^2(\tilde{\alpha} - \tilde{\tau})\mathbb{E}(\mathbf{S})\mathbf{DR} + \mathbf{\Delta R}. \qquad (98)$$

We now verify that if $\mathbf{M}$ is a deterministic, uniformly bounded matrix for the spectral norm as $t\to\infty$, then $t^{-1}\,\mathrm{Tr}\,\mathbf{\Delta RM} = O(t^{-2})$. For this, we write $\frac{1}{t}\,\mathrm{Tr}\,\mathbf{\Delta RM}$ as $\frac{1}{t}\,\mathrm{Tr}\,\mathbf{\Delta RM} = T_1 + T_2 - T_3$ where

$$\begin{aligned}T_1 &= \mathbb{E}\left[\overset{\circ}{\eta}\,\frac{1}{t}\,\mathrm{Tr}\,(\mathbf{S\Sigma}\tilde{\mathbf{D}}(\mathbf{I}+\alpha\tilde{\mathbf{D}})^{-1}\mathbf{\Sigma}^H\mathbf{RM})\right]\\ T_2 &= \mathbb{E}\left(\left(\overset{\circ}{\rho}_1 + \overset{\circ}{\rho}_2\right)\frac{1}{t}\,\mathrm{Tr}\,(\overset{\circ}{\mathbf{S}}\,\mathbf{DRM})\right)\\ T_3 &= \mathbb{E}\left(\overset{\circ}{\eta}\,\overset{\circ}{\rho}_3\right)\frac{1}{t}\,\mathrm{Tr}\,(\mathbb{E}(\mathbf{S})\mathbf{DRM}).\end{aligned}$$

We denote by ρ_4 the term

$$\rho_4 = \frac{1}{t}\,\mathrm{Tr}\,(\mathbf{S}\boldsymbol{\Sigma}\tilde{\mathbf{D}}(\mathbf{I}+\alpha\tilde{\mathbf{D}})^{-1}\boldsymbol{\Sigma}^H\mathbf{R}\mathbf{M})$$

and notice that $T_1 = \mathbb{E}(\overset{\circ}{\eta}\overset{\circ}{\rho}_4)$. Equation (25) implies that $\mathbb{E}(\overset{\circ}{\eta}{}^2)$ and $\mathbb{E}[\frac{1}{t}\,\mathrm{Tr}(\overset{\circ}{\mathbf{S}}\mathbf{D}\mathbf{R}\mathbf{M})]^2$ are $O(t^{-2})$ terms. Moreover, matrix $\mathbf{R}$ is uniformly bounded for the spectral norm as $t\to\infty$ [see (73)]. Lemma 2 immediately shows that for each $i=1,2,3, \mathbb{E}(\overset{\circ}{\rho}_i{}^2)$ is a $O(t^{-2})$ term. The Cauchy-Schwarz inequality eventually provides $\frac{1}{t}\mathrm{Tr}\boldsymbol{\Delta}\mathbf{R}\mathbf{M} = O(t^{-2})$.

In order to establish (75), it remains to show that $\tilde{\alpha}-\tilde{\tau}\to 0$. For this, we remark that exchanging the roles of matrices $\boldsymbol{\Sigma}$ and $\boldsymbol{\Sigma}^H$ leads to the following relation:

$$\mathbb{E}(\tilde{\mathbf{S}}) = \tilde{\mathbf{R}} + \sigma^2(\alpha-\tau)\mathbb{E}(\tilde{\mathbf{S}}\tilde{\mathbf{D}}\tilde{\mathbf{R}}) + \tilde{\boldsymbol{\Delta}}\tilde{\mathbf{R}} \tag{99}$$

where $\tau(\sigma^2)$ is defined by

$$\tau(\sigma^2) = \frac{1}{t}\,\mathrm{Tr}\,[\mathbf{D}(\sigma^2(\mathbf{I}+\tilde{\alpha}\mathbf{D}))^{-1} \times(\mathbf{I}-\mathbf{B}\,\mathbb{E}(\tilde{\mathbf{S}}(\sigma^2))\,\mathbf{B}^H(\mathbf{I}+\tilde{\alpha}\mathbf{D})^{-1})] \tag{100}$$

and where $\tilde{\boldsymbol{\Delta}}$, the analogue of $\boldsymbol{\Delta}$, satisfies

$$\frac{1}{t}\mathrm{Tr}(\tilde{\boldsymbol{\Delta}}\tilde{\mathbf{M}}) = O\left(\frac{1}{t^2}\right) \tag{101}$$

for every matrix $\tilde{\mathbf{M}}$ uniformly bounded for the spectral norm.

Equations (98) and (99) allow to evaluate $\tilde{\alpha}$ and $\tilde{\tau}$. More precisely, writing $\tilde{\alpha} = \frac{1}{t}\mathrm{Tr}(\tilde{\mathbf{D}}\mathbb{E}(\tilde{\mathbf{S}}))$ and using (99) of $\mathbb{E}(\tilde{\mathbf{S}})$, we obtain that

$$\tilde{\alpha} = \frac{1}{t}\mathrm{Tr}(\tilde{\mathbf{D}}\tilde{\mathbf{R}}) + \sigma^2(\alpha-\tau)\frac{1}{t}\mathrm{Tr}(\tilde{\mathbf{D}}\mathbb{E}(\tilde{\mathbf{S}})\tilde{\mathbf{D}}\tilde{\mathbf{R}}) + \frac{1}{t}\,\mathrm{Tr}\,(\tilde{\mathbf{D}}\tilde{\boldsymbol{\Delta}}\tilde{\mathbf{R}}). \tag{102}$$

Similarly, replacing $\mathbb{E}(\mathbf{S})$ by (98) into the expression (95) of $\tilde{\tau}$, we get that

$$\begin{aligned}\tilde{\tau} = {} & \frac{1}{t}\,\mathrm{Tr}\,[\tilde{\mathbf{D}}(\sigma^2(\mathbf{I}+\alpha\tilde{\mathbf{D}})^{-1}(\mathbf{I}-\mathbf{B}^H\mathbf{R}\mathbf{B}(\mathbf{I}+\alpha\tilde{\mathbf{D}})^{-1}] \\ & -(\tilde{\alpha}-\tilde{\tau})\frac{1}{t}\,\mathrm{Tr}\,[\tilde{\mathbf{D}}(\mathbf{I}+\alpha\tilde{\mathbf{D}})^{-1}\mathbf{B}^H\mathbb{E}(\mathbf{S})\mathbf{D}\mathbf{R}\mathbf{B}(\mathbf{I}+\alpha\tilde{\mathbf{D}})^{-1}] \\ & -\frac{1}{t}\,\mathrm{Tr}\,[\tilde{\mathbf{D}}(\sigma^2(\mathbf{I}+\alpha\tilde{\mathbf{D}})^{-1}\mathbf{B}^H\boldsymbol{\Delta}\mathbf{R}\mathbf{B}(\mathbf{I}+\alpha\tilde{\mathbf{D}})^{-1}].\end{aligned} \tag{103}$$

Using standard algebra, it is easy to check that the first term of the right-hand side (RHS) of (103) coincides with $\frac{1}{t}\,\mathrm{Tr}\,(\tilde{\mathbf{D}}\tilde{\mathbf{R}})$. Substracting (103) from (102), we get that

$$(\alpha-\tau)\tilde{u}_0 + (\tilde{\alpha}-\tilde{\tau})\tilde{v}_0 = \tilde{\epsilon} \tag{104}$$

where

$$\begin{aligned}\tilde{u}_0 &= \sigma^2\frac{1}{t}\mathrm{Tr}(\tilde{\mathbf{D}}\mathbb{E}(\tilde{\mathbf{S}})\tilde{\mathbf{D}}\tilde{\mathbf{R}}) \\ \tilde{v}_0 &= 1-\frac{1}{t}\,\mathrm{Tr}\,[\tilde{\mathbf{D}}(\mathbf{I}+\alpha\tilde{\mathbf{D}})^{-1}\mathbf{B}^H\mathbb{E}(\mathbf{S})\mathbf{D}\mathbf{R}\mathbf{B}(\mathbf{I}+\alpha\tilde{\mathbf{D}})^{-1}] \\ \tilde{\epsilon} &= \frac{1}{t}\,\mathrm{Tr}\,(\tilde{\mathbf{D}}\tilde{\boldsymbol{\Delta}}\tilde{\mathbf{R}}) \\ &\quad + \frac{1}{t}\,\mathrm{Tr}\,[\tilde{\mathbf{D}}(\sigma^2(\mathbf{I}+\alpha\tilde{\mathbf{D}})^{-1}\mathbf{B}^H\boldsymbol{\Delta}\mathbf{R}\mathbf{B}(\mathbf{I}+\alpha\tilde{\mathbf{D}})^{-1}].\end{aligned} \tag{105}$$

Using the properties of $\boldsymbol{\Delta}$ and $\tilde{\boldsymbol{\Delta}}$, we get that $\tilde{\epsilon} = 0(t^{-2})$.

Similar calculations allow to evaluate α and τ, and to obtain

$$(\alpha-\tau)u_0 + (\tilde{\alpha}-\tilde{\tau})v_0 = \epsilon \tag{106}$$

where

$$\begin{aligned}u_0 &= 1-\frac{1}{t}\,\mathrm{Tr}\,\left[\mathbf{D}(\mathbf{I}+\tilde{\alpha}\mathbf{D})^{-1}\mathbf{B}\mathbb{E}(\tilde{\mathbf{S}})\tilde{\mathbf{D}}\tilde{\mathbf{R}}\mathbf{B}^H(\mathbf{I}+\tilde{\alpha}\mathbf{D})^{-1}\right] \\ v_0 &= \sigma^2\frac{1}{t}\mathrm{Tr}(\mathbf{D}\mathbb{E}(\mathbf{S})\mathbf{D}\mathbf{R})\end{aligned} \tag{107}$$

and where $\epsilon = O(t^{-2})$. (106), (104) can be written as

$$\begin{pmatrix} u_0 & v_0 \\ \tilde{u}_0 & \tilde{v}_0\end{pmatrix}\begin{pmatrix}\alpha-\tau \\ \tilde{\alpha}-\tilde{\tau}\end{pmatrix} = \begin{pmatrix}\epsilon \\ \tilde{\epsilon}\end{pmatrix}. \tag{108}$$

If the determinant $u_0\tilde{v}_0 - \tilde{u}_0 v_0$ of the 2×2 matrix governing the system is nonzero, $\alpha-\tau$ and $\tilde{\alpha}-\tilde{\tau}$ are given by

$$\alpha-\tau = \frac{\tilde{v}_0\epsilon - v_0\tilde{\epsilon}}{u_0\tilde{v}_0-\tilde{u}_0v_0}, \qquad \tilde{\alpha}-\tilde{\tau} = \frac{u_0\tilde{\epsilon}-\tilde{u}_0\epsilon}{u_0\tilde{v}_0-\tilde{u}_0v_0} \tag{109}$$

As matrices $\mathbf{R}$ and $\mathbb{E}(\mathbf{S})$ are less than $\frac{1}{\sigma^2}\mathbf{I}_r$ and matrices $\tilde{\mathbf{R}}$ and $\mathbb{E}(\tilde{\mathbf{S}})$ are less than $\frac{1}{\sigma^2}\mathbf{I}_t$, it is easy to check that $u_0, v_0, \tilde{u}_0, \tilde{v}_0$ are uniformly bounded. As ϵ and $\tilde{\epsilon}$ are $O(t^{-2})$ terms, $(\alpha-\tau)$ and $(\tilde{\alpha}-\tilde{\tau})$ will converge to 0 as long as the inverse $(u_0\tilde{v}_0-\tilde{u}_0v_0)^{-1}$ of the determinant is uniformly bounded. For the moment, we show this property for σ^2 large enough. For this, we study the behavior of coefficients $u_0, \tilde{u}_0, v_0, \tilde{v}_0$ for large enough values of σ^2. It is easy to check that

$$\begin{aligned}u_0 &\geq 1-\frac{1}{\sigma^4}\frac{r}{t}d_{\max}\tilde{d}_{\max}b_{\max}^2 \\ \tilde{v}_0 &\geq 1-\frac{1}{\sigma^4}d_{\max}\tilde{d}_{\max}b_{\max}^2 \\ \tilde{u}_0 &\leq \frac{\tilde{d}_{\max}^2}{\sigma^2} \\ v_0 &\leq \frac{r}{t}\frac{d_{\max}^2}{\sigma^2}.\end{aligned} \tag{110}$$

As $\frac{t}{r}\to c$, it is clear that there exists σ_0^2 and an integer t_0 for which $u_0\geq 1/2, \tilde{v}_0\geq 1/2, \tilde{u}_0\leq 1/4, v_0\leq 1/4$ for $t\geq t_0$ and $\sigma^2\geq\sigma_0^2$. Therefore, $u_0\tilde{v}_0-\tilde{u}_0v_0 > \frac{3}{16}$ for $t\geq t_0$ and $\sigma^2\geq\sigma_0^2$. Equation (109), thus, implies that if $\sigma^2\geq\sigma_0^2$, then $\alpha-\tau$ and $\tilde{\alpha}-\tilde{\tau}$ are of the same order of magnitude as $\epsilon = O(t^{-2})$, and therefore converge to 0 when $t\to\infty$. It remains to prove that this convergence still holds for $0<\sigma^2<\sigma_0^2$. For this, we shall rely on Montel's theorem (see, e.g., [6]), a tool frequently used in the context of large random matrices. It is based on the observation that, considered as functions of parameter $\sigma^2, \alpha(\sigma^2)-\tau(\sigma^2)$ and $\tilde{\alpha}(\sigma^2)-\tilde{\tau}(\sigma^2)$ can be extended to holomorphic functions on $\mathbb{C}-\mathbb{R}^-$ by replacing σ^2 by a complex number z. Moreover, it can be shown that these holomorphic functions are uniformly bounded on each compact subset K of $\mathbb{C}-\mathbb{R}^-$, in the sense that $\sup_t\sup_{z\in K}|\alpha(z)-\tau(z)|<\infty$ and $\sup_t\sup_{z\in K}|\tilde{\alpha}(z)-\tilde{\tau}(z)|<\infty$. Using Montel's theorem, it can thus be shown that if $\alpha(\sigma^2)-\tau(\sigma^2)$ and $\tilde{\alpha}(\sigma^2)-\tilde{\tau}(\sigma^2)$ converge towards zero for each $\sigma^2>\sigma_0^2$, then for each $z\in\mathbb{C}-\mathbb{R}^-$, $\alpha(z)-\tau(z)$ and $\tilde{\alpha}(z)-\tilde{\tau}(z)$ converge as well towards 0. This in particular implies that $\alpha(\sigma^2)-\tau(\sigma^2)$ and $\tilde{\alpha}(\sigma^2)-\tilde{\tau}(\sigma^2)$ converge towards 0 for each $\sigma^2>0$. For more details, the reader may, e.g., refer to [18]. This completes the proof of (75).

We note that Montel's theorem does not guarantee that $\alpha - \tau$ and $\tilde{\alpha} - \tilde{\tau}$ are still $O(t^{-2})$ terms for $\sigma^2 < \sigma_0^2$. This is one of the purpose of the proof of Step 2 below.

In order to finish the proof of Proposition 7, it remains to check that (76) holds. We first observe that $\mathbf{R} - \mathbf{T} = \mathbf{R}(\mathbf{T}^{-1} - \mathbf{R}^{-1})\mathbf{T}$. Using the expressions of $\mathbf{R}^{-1}$ and $\mathbf{T}^{-1}$, multiplying by $\mathbf{M}$, and taking the trace yields

$$\frac{1}{t}\operatorname{Tr}[\mathbf{M}(\mathbf{R}-\mathbf{T})] = (\tilde{\beta}-\tilde{\alpha})\sigma^2\frac{1}{t}\operatorname{Tr}(\mathbf{MRDT}) + (\alpha-\beta)\frac{1}{t}\operatorname{Tr}[\mathbf{MRB}(\mathbf{I}+\beta\tilde{\mathbf{D}})^{-1}\tilde{\mathbf{D}}(\mathbf{I}+\beta\tilde{\mathbf{D}})^{-1}\mathbf{B}^H\mathbf{T}].$$

As the terms $\frac{\sigma^2}{t}\operatorname{Tr}(\mathbf{MRDT})$ and $\frac{1}{t}\operatorname{Tr}[\mathbf{MRB}(\mathbf{I}+\beta\tilde{\mathbf{D}})^{-1}\tilde{\mathbf{D}}(\mathbf{I}+\beta\tilde{\mathbf{D}})^{-1}\mathbf{B}^H\mathbf{T}]$ are uniformly bounded, it is sufficient to establish that $(\alpha-\beta)$ and $(\tilde{\alpha}-\tilde{\beta})$ converge towards 0. For this, we note that (75) implies that

$$\alpha = \frac{1}{t}\operatorname{Tr}(\mathbf{DR}) + \epsilon', \qquad \tilde{\alpha} = \frac{1}{t}\operatorname{Tr}(\tilde{\mathbf{D}}\tilde{\mathbf{R}}) + \tilde{\epsilon}' \tag{111}$$

where ϵ' and $\tilde{\epsilon}'$ converge towards 0. We express $(\alpha-\beta) = \frac{1}{t}\operatorname{Tr}\mathbf{D}(\mathbf{R}-\mathbf{T}) + \epsilon$. Using $\mathbf{R}-\mathbf{T} = \mathbf{R}(\mathbf{T}^{-1}-\mathbf{R}^{-1})\mathbf{T}$, multiplying by $\mathbf{D}$ from both sides, and taking the trace yields

$$(\alpha-\beta)\left(1-\frac{1}{t}\operatorname{Tr}[\mathbf{DRB}(\mathbf{I}+\beta\tilde{\mathbf{D}})^{-1}\tilde{\mathbf{D}}(\mathbf{I}+\beta\tilde{\mathbf{D}})^{-1}\mathbf{B}^H\mathbf{T}]\right) + (\tilde{\alpha}-\tilde{\beta})\sigma^2\frac{1}{t}\operatorname{Tr}(\mathbf{DRDT}) = \epsilon'. \tag{112}$$

Similarly, we obtain that

$$(\alpha-\beta)\sigma^2\frac{1}{t}\operatorname{Tr}(\tilde{\mathbf{D}}\tilde{\mathbf{R}}\tilde{\mathbf{D}}\tilde{\mathbf{T}}) + (\tilde{\alpha}-\tilde{\beta}) \times\left(1-\frac{1}{t}\operatorname{Tr}[\tilde{\mathbf{D}}\tilde{\mathbf{R}}\mathbf{B}^H(\mathbf{I}+\tilde{\beta}\mathbf{D})^{-1}\mathbf{D}(\mathbf{I}+\tilde{\beta}\mathbf{D})^{-1}\mathbf{D}\tilde{\mathbf{T}}]\right) = \tilde{\epsilon}'. \tag{113}$$

Equations (112) and (113) can be interpreted as a linear systems with respect to $(\alpha-\beta)$ and $(\tilde{\alpha}-\tilde{\beta})$. Using the same approach as in the proof of (75), we prove that $(\alpha-\beta)$ and $(\tilde{\alpha}-\tilde{\beta})$ converge towards 0. This establishes (76) and completes the proof of Proposition (7). ■

2) Second Step: $\frac{1}{r}\operatorname{Tr}\mathbf{M}(\mathbb{E}(\mathbf{S})-\mathbf{R})$ and $\frac{1}{r}\operatorname{Tr}\mathbf{M}(\mathbf{R}-\mathbf{T})$ are $O(t^{-2})$ Terms: This section is devoted to the proof of the following proposition.

Proposition 8: For each deterministic $r\times r$ matrix $\mathbf{M}$, uniformly bounded (for the spectral norm) as $t\to\infty$, we have:

$$\frac{1}{t}\operatorname{Tr}[\mathbf{M}(\mathbb{E}(\mathbf{S})-\mathbf{R})] = O(t^{-2}) \tag{114}$$

$$\frac{1}{t}\operatorname{Tr}[\mathbf{M}(\mathbf{R})-\mathbf{T}] = O(t^{-2}) \tag{115}$$

Proof: We first establish (114). For this, we prove that the inverse of the determinant $u_0\tilde{v}_0 - \tilde{u}_0 v_0$ of linear system (108) is uniformly bounded for each $\sigma^2 > 0$. In order to state the corresponding result, we define $(u, v, \tilde{u}, \tilde{v})$ by

$$\begin{aligned} u &= 1-\frac{1}{t}\operatorname{Tr}(\tilde{\mathbf{D}}\tilde{\mathbf{T}}\mathbf{B}^H(\mathbf{I}+\tilde{\beta}\mathbf{D})^{-1}\mathbf{D}(\mathbf{I}+\tilde{\beta}\mathbf{D})^{-1}\mathbf{B}\tilde{\mathbf{T}})\\ \tilde{v} &= 1-\frac{1}{t}\operatorname{Tr}(\mathbf{DTB}(\mathbf{I}+\beta\tilde{\mathbf{D}})^{-1}\tilde{\mathbf{D}}(\mathbf{I}+\beta\tilde{\mathbf{D}})^{-1}\mathbf{B}^H\mathbf{T})\\ v &= \sigma^2\frac{1}{t}\operatorname{Tr}(\mathbf{DTDT})\\ \tilde{u} &= \sigma^2\frac{1}{t}\operatorname{Tr}(\tilde{\mathbf{D}}\tilde{\mathbf{T}}\tilde{\mathbf{D}}\tilde{\mathbf{T}}). \end{aligned} \tag{116}$$

The expressions of $(u, v, \tilde{u}, \tilde{v})$ nearly coincide with the expressions of coefficients $(u_0, v_0, \tilde{u}_0, \tilde{v}_0)$, the only difference being that, in the definition of $(u, v, \tilde{u}, \tilde{v})$, matrices $(\mathbb{E}(\mathbf{S}), \mathbf{R})$ are both replaced by matrix $\mathbf{T}$, matrices $(\mathbb{E}(\tilde{\mathbf{S}}), \tilde{\mathbf{R}})$ are both replaced by matrix $\tilde{\mathbf{T}}$ and scalars $(\alpha, \tilde{\alpha})$ are replaced by scalars $(\beta, \tilde{\beta})$. (75) and (76) immediately imply that $(u_0, v_0, \tilde{u}_0, \tilde{v}_0)$ can be written as

$$\begin{aligned} u_0 &= u + \epsilon_u\\ \tilde{v}_0 &= \tilde{v} + \tilde{\epsilon}_v\\ v_0 &= v + \epsilon_v\\ \tilde{u}_0 &= \tilde{u} + \tilde{\epsilon}_u \end{aligned} \tag{117}$$

where $\epsilon_u, \tilde{\epsilon}_v, \tilde{\epsilon}_u, \epsilon_v$ converge to 0 when $t\to\infty$. The behavior of $u\tilde{v} - \tilde{u}v$ is provided in the following Lemma, whose proof is given in Appendix II-C3 below.

Lemma 4: Coefficients $(u, v, \tilde{u}, \tilde{v})$ satisfy: (i) $u = \tilde{v}$, (ii) $0 < u < 1$ and $\inf_t u > 0$; (iii) $0 < u\tilde{v} - \tilde{u}v < 1$ and $\sup_t \frac{1}{u\tilde{v}-\tilde{u}v} < \infty$.

Equation (117) and Lemma 4 immediately imply that it exists t_0 such that $0 < u_0\tilde{v}_0 - \tilde{u}_0 v_0 \le 1$ for each $t \ge t_0$ and

$$\sup_{t\ge t_0}\frac{1}{u_0\tilde{v}_0 - \tilde{u}_0 v_0} < \infty. \tag{118}$$

This eventually shows $\alpha-\tau$ and $\tilde{\alpha}-\tilde{\tau}$ are of the same order of magnitude than ϵ and $\tilde{\epsilon}$, i.e., are $O(t^{-2})$ terms.

In order to prove (115), we first remark that, by (114), ϵ' and $\tilde{\epsilon}'$ defined by (111) are $O(t^{-2})$ terms. It is thus sufficient to establish that the inverse of the determinant of the linear system associated to (112) and (113) is uniformly bounded. Equation (76) implies that the behavior of this determinant is equivalent to the study of $u\tilde{v} - \tilde{u}v$. Equation (115) thus follows from Lemma 4. This completes the proof of Proposition 8. □

3) Proof of Lemma 4: In order to establish item (i), we notice that a direct application of the matrix inversion Lemma yields:

$$\tilde{\mathbf{T}}\mathbf{B}^H(\mathbf{I}+\tilde{\beta}\mathbf{D})^{-1} = (\mathbf{I}+\beta\tilde{\mathbf{D}})^{-1}\mathbf{B}^H\mathbf{T}. \tag{119}$$

The equality $u = \tilde{v}$ immediately follows from (119).

The proofs of (ii) and (iii) are based on the observation that function $\sigma^2 \to \sigma^2\beta(\sigma^2)$ is increasing while function $\sigma^2 \to \tilde{\beta}(\sigma^2)$ is decreasing. This claim is a consequence of (17) that we recall below

$$\beta(\sigma^2) = \int_{\mathbb{R}^+}\frac{d\mu_b(\lambda)}{\lambda+\sigma^2}, \quad \tilde{\beta}(\sigma^2) = \int_{\mathbb{R}^+}\frac{d\tilde{\mu}_b(\lambda)}{\lambda+\sigma^2}$$

where $\mu_b(\mathbb{R}^+) = \frac{1}{t}\mathrm{Tr}(\mathbf{D})$ and $\tilde{\mu}_b(\mathbb{R}^+) = \frac{1}{t}\mathrm{Tr}(\tilde{\mathbf{D}})$. Note that $\tilde{\beta}$ is decreasing because $\sigma^2 \mapsto \frac{1}{\lambda+\sigma^2}$ is decreasing and $\sigma^2\beta(\sigma^2)$ is increasing because $\sigma^2 \mapsto \frac{\sigma^2}{\lambda+\sigma^2}$ is increasing. Denote by $'$ the differentiation operator with respect to σ^2. Then, $(\sigma^2\beta)' > 0$ and $\tilde{\beta}' < 0$ for each σ^2. We now differentiate (16) with respect to σ^2. After some algebra, we obtain

$$\tilde{v}(\sigma^2\beta)' + \sigma^2 v\tilde{\beta}' = \frac{1}{t}\mathrm{Tr}(\mathbf{DTB}(\mathbf{I}+\beta\tilde{\mathbf{D}})^{-1}(\mathbf{I}+\beta\tilde{\mathbf{D}})^{-1}\mathbf{B}^H\mathbf{T})$$
$$\frac{\tilde{u}}{\sigma^2}(\sigma^2\beta)' + u\tilde{\beta}' = -\frac{1}{t}\ \mathrm{Tr}\ \tilde{\mathbf{T}}\tilde{\mathbf{D}}\tilde{\mathbf{T}}. \quad (120)$$

As $\tilde{\beta}' < 0$, the first equation of (120) implies that $\tilde{v}(\sigma^2\beta)' > 0$. As $(\sigma^2\beta)' > 0$, this yields $\tilde{v} > 0$. As $\tilde{v} < 1$ clearly holds, the first part of (ii) is proved.

We now prove that $\inf_t \tilde{v} > 0$. The first equation of (120) yields

$$\tilde{v} > -\sigma^2 v\tilde{\beta}'\frac{1}{(\sigma^2\beta)'}. \quad (121)$$

In the following, we show that $\inf_t \frac{1}{(\sigma^2\beta)'} > 0, \inf_t|\tilde{\beta}'| > 0$ and that $\inf_t\ v > 0$.

By (17)

$$-\tilde{\beta}' = \int_{\mathrm{R}^+}\frac{d\tilde{\mu}_b(\lambda)}{(\lambda+\sigma^2)^2}$$

and

$$(\sigma^2\beta(\sigma^2))' = \int_{\mathrm{R}^+}\frac{\lambda d\mu_b(\lambda)}{(\lambda+\sigma^2)^2}.$$

As $\frac{\lambda}{(\lambda+\sigma^2)^2} \le \frac{1}{\sigma^2}$ for $\lambda \ge 0, (\sigma^2\beta)' \le \frac{1}{\sigma^2}\mu_b(\mathbb{R}^+) = \frac{1}{t}\mathrm{Tr}\mathbf{D}$. Therefore, the term $\frac{1}{(\sigma^2\beta)'}$ is lowerbounded by $\sigma^2(\frac{1}{t}\mathrm{Tr}\mathbf{D})^{-1}$. As $\frac{1}{t}\ \mathrm{Tr}\ \mathbf{D} \le \frac{r}{t}d_{\max}$, we have $\inf_t \frac{1}{(\sigma^2\beta)'} > 0$.

We now establish that $\inf_t|\tilde{\beta}'| > 0$. We first use Jensen's inequality: As measure $(\frac{1}{t}\ \mathrm{Tr}\ \tilde{\mathbf{D}})^{-1}d\tilde{\mu}_b(\lambda)$ is a probability distribution:

$$\left[\int_{\mathrm{R}^+}\frac{1}{\lambda+\sigma^2}\left(\frac{1}{t}\ \mathrm{Tr}\ \tilde{\mathbf{D}}\right)^{-1}d\tilde{\mu}_b(\lambda)\right]^2 \le \int_{\mathrm{R}^+}\frac{1}{(\lambda+\sigma^2)^2}\left(\frac{1}{t}\ \mathrm{Tr}\ \tilde{\mathbf{D}}\right)^{-1}d\tilde{\mu}_b(\lambda). \quad (122)$$

In other words, $|\tilde{\beta}'| = \int_{\mathrm{R}^+}\frac{1}{(\lambda+\sigma^2)^2}d\tilde{\mu}_b(\lambda)$ satisfies

$$|\tilde{\beta}'| \ge \frac{1}{\frac{1}{t}\ \mathrm{Tr}\ \tilde{\mathbf{D}}}\left[\int_{\mathrm{R}^+}\frac{1}{\lambda+\sigma^2}d\tilde{\mu}_b(\lambda)\right]^2 = \frac{1}{\frac{1}{t}\ \mathrm{Tr}\ \tilde{\mathbf{D}}}\tilde{\beta}^2.$$

As aforementioned, $(\frac{1}{t}\ \mathrm{Tr}\ \tilde{\mathbf{D}})^{-1}$ is lower-bounded by $(d_{\max})^{-1}$. Therefore, it remains to establish that $\inf_t \tilde{\beta}^2 > 0$, or equivalently that $\inf_t \tilde{\beta} > 0$. For this, we assume that $\inf_t \tilde{\beta}_t(\sigma^2) = 0$ (we indicate that $\tilde{\beta}$ depends both on σ^2 and t). Therefore, there exists an increasing sequence of integers $(t_k)_{k\ge0}$ for which $\lim_{k\to\infty}\tilde{\beta}_{t_k}(\sigma^2) = 0$, i.e., $\lim_{k\to\infty}\int_{\mathrm{R}^+}\frac{1}{\lambda+\sigma^2}d\tilde{\mu}_b^{(t_k)}(\lambda) = 0$ where $\tilde{\mu}_b^{(t_k)}$ is the positive measure associated with $\tilde{\beta}_{t_k}(\sigma^2)$. As $\tilde{\mathbf{D}}$ is uniformly bounded, the sequence $(\tilde{\mu}_b^{(t_k)})_{k\ge0}$ is tight. One can, therefore, extract from $(\tilde{\mu}_b^{(t_k)})_{k\ge0}$ a subsequence $(\tilde{\mu}_b^{(t'_l)})_{l\ge0}$ that converges weakly to a certain measure $\tilde{\mu}_b^*$ which of course satisfies

$$\int_{\mathrm{R}^+}\frac{1}{\lambda+\sigma^2}d\tilde{\mu}_b^*(\lambda) = 0.$$

This implies that $\tilde{\mu}_b^* = 0$, and thus $\tilde{\mu}_b^*(\mathrm{R}^+) = 0$, while the convergence of $(\tilde{\mu}_b^{(t'_l)})_{l\ge0}$ gives

$$\tilde{\mu}_b^*(\mathrm{R}^+) = \lim_{l\to\infty}\tilde{\mu}_b^{(t'_l)}(\mathrm{R}^+) = \lim_{l\to\infty}\frac{1}{t'_l}\mathrm{Tr}\tilde{\mathbf{D}}_{t'_l} > 0$$

by assumption (3). Therefore, the assumption $\inf_t \tilde{\beta}_t(\sigma^2) = 0$ leads to a contradiction. Thus, $\inf_t \tilde{\beta}_t(\sigma^2) > 0$ and $\inf_t|\tilde{\beta}'| > 0$ is proved.

We finally establish that v is lower-bounded, i.e., that $\inf_t \frac{1}{t}\ \mathrm{Tr}\ \mathbf{DTDT} > 0$. For any Hermitian positive matrix $\mathbf{M}$,

$$\frac{1}{t}\mathrm{Tr}(\mathbf{M}^2) \ge \left[\frac{1}{t}\mathrm{Tr}(\mathbf{M})\right]^2.$$

We use this inequality for $\mathbf{M} = \mathbf{T}^{1/2}\mathbf{D}\mathbf{T}^{1/2}$. This leads to

$$\frac{1}{t}\ \mathrm{Tr}\ \mathbf{DTDT} = \frac{1}{t}\ \mathrm{Tr}\ \mathbf{M}^2 > \left[\frac{1}{t}\mathrm{Tr}(\mathbf{M})\right]^2 = \left[\frac{1}{t}\mathrm{Tr}(\mathbf{DT})\right]^2 = \beta^2.$$

Therefore, $\inf_t \frac{1}{t}\ \mathrm{Tr}\ \mathbf{DTDT} \ge \inf_t \beta^2$. Using the same approach as above, we can prove that $\inf_t \beta^2 > 0$. Proof of (ii) is completed.

In order to establish (iii), we use the first equation of (120) to express $(\sigma^2\beta)'$ in terms of $\tilde{\beta}'$, and plug this relation into the second equation of (120). This gives

$$\left(u - \frac{1}{\tilde{v}}\tilde{u}v\right)\tilde{\beta}' = -\frac{1}{t}\ \mathrm{Tr}\ \tilde{\mathbf{T}}\tilde{\mathbf{D}}\tilde{\mathbf{T}} - \frac{\tilde{u}}{\sigma^2\tilde{v}}\frac{1}{t}\mathrm{Tr}(\mathbf{DTB}(\mathbf{I}+\beta\tilde{\mathbf{D}})^{-1}(\mathbf{I}+\beta\tilde{\mathbf{D}})^{-1}\mathbf{B}^H\mathbf{T}). \quad (123)$$

The RHS of (123) is negative as well as $\tilde{\beta}'$. Therefore, $u - \frac{1}{\tilde{v}}\tilde{u}v > 0$. As $\tilde{v}$ is positive, $u\tilde{v} - \tilde{u}v$ is also positive. Moreover, u et $\tilde{v}$ are strictly less than 1. As $\tilde{u}$ and v are both strictly positive, $u\tilde{v} - \tilde{u}v$ is strictly less than 1. To complete the proof of (iii), we notice that by (123)

$$\frac{1}{u\tilde{v} - \tilde{u}v} \le \frac{|\tilde{\beta}'|}{\tilde{v}\frac{1}{t}\ \mathrm{Tr}\ \tilde{\mathbf{T}}\tilde{\mathbf{D}}\tilde{\mathbf{T}}}.$$

$|\tilde{\beta}'|$ clearly satisfies $|\tilde{\beta}'| \le \frac{1}{\sigma^4}\frac{1}{t}\ \mathrm{Tr}\ \tilde{\mathbf{D}}$ and is thus upper bounded by $\frac{d_{\max}}{\sigma^4}$. (ii) implies that $\sup_t \frac{1}{\tilde{v}} < +\infty$. It remains to verify that $\inf_t \frac{1}{t}\ \mathrm{Tr}\ \tilde{\mathbf{T}}\tilde{\mathbf{D}}\tilde{\mathbf{T}} > 0$. Denote by $x = \frac{1}{t}\ \mathrm{Tr}\ \tilde{\mathbf{T}}\tilde{\mathbf{D}}\tilde{\mathbf{T}}$.

$$x = \frac{1}{t}\sum_{i=1}^{t}\tilde{d}_i\sum_{j=1}^{t}|\tilde{T}_{i,j}|^2.$$

In order to use Jensen's inequality, we consider $\tilde{\kappa}_i = \frac{\tilde{d}_i}{\frac{1}{t}\,\mathrm{Tr}\,\tilde{\mathbf{D}}}$, and notice that $\frac{1}{t}\sum_{i=1}^t \tilde{\kappa}_i = 1$. x can be written as

$$x = \frac{1}{t}\,\mathrm{Tr}\,\tilde{\mathbf{D}}\frac{1}{t}\sum_{i=1}^t \tilde{\kappa}_i \left[\left(\sum_{j=1}^t |\tilde{T}_{i,j}|^2\right)^{1/2}\right]^2.$$

By Jensen's inequality

$$\frac{1}{t}\sum_{i=1}^t \tilde{\kappa}_i \left[\left(\sum_{j=1}^t |T_{i,j}|^2\right)^{1/2}\right]^2 \geq \left[\frac{1}{t}\sum_{i=1}^t \tilde{\kappa}_i \left(\sum_{j=1}^t |\tilde{T}_{i,j}|^2\right)^{1/2}\right]^2.$$

Moreover

$$\left[\frac{1}{t}\sum_{i=1}^t \tilde{\kappa}_i \left(\sum_{j=1}^t |\tilde{T}_{i,j}|^2\right)^{1/2}\right]^2 \geq \left[\frac{1}{t}\sum_{i=1}^t \tilde{\kappa}_i\, \tilde{T}_{i,i}\right]^2 = \left[\left(\frac{1}{t}\,\mathrm{Tr}\,\tilde{\mathbf{D}}\right)^{-1}\tilde{\beta}\right]^2.$$

Finally

$$x = \frac{1}{t}\,\mathrm{Tr}\,\tilde{\mathbf{T}}\tilde{\mathbf{D}}\tilde{\mathbf{T}} \geq \left(\frac{1}{t}\,\mathrm{Tr}\,\tilde{\mathbf{D}}\right)^{-1}\tilde{\beta}^2.$$

Since $\inf_t \tilde{\beta}^2 > 0$, we have $\inf_t \frac{1}{t}\,\mathrm{Tr}\,\tilde{\mathbf{T}}\tilde{\mathbf{D}}\tilde{\mathbf{T}} > 0$ and the proof of (iii) is completed.

Appendix III
End of proof of Proposition 3

Proof of Proposition 3 relies on properties of $\bar{\mathbf{Q}}_*$ established in Proposition 5–(iii). Denote by

$$A = \max(\sup_t \|\mathbf{A}\|, \sup_t \|\mathbf{C}_T\|, \sup_t \|\mathbf{C}_R\|) < \infty$$

and

$$a = \min\left(\inf_t \lambda_{\min}(\mathbf{C}_T), \inf_t \lambda_{\min}(\mathbf{C}_R)\right) > 0. \quad (124)$$

Proof of (i): Recall that by Proposition 5–(iii), $\bar{\mathbf{Q}}_*$ maximizes $\log\det(\mathbf{I}+\mathbf{Q}\mathbf{G}(\delta_{R,*}, \delta_{T,*}))$. This implies that the eigenvalues $(\lambda_j(\bar{\mathbf{Q}}_*))$ are the solutions of the waterfilling equation

$$\lambda_j(\bar{\mathbf{Q}}_*) = \max\left(\gamma - \frac{1}{\lambda_j(\mathbf{G})}, 0\right), \quad 1 \leq j \leq t$$

where γ is tuned in such a way that $\sum_j \lambda_j(\bar{\mathbf{Q}}_*) = t$. It is clear from this equation that $\|\bar{\mathbf{Q}}_*\| \leq \gamma$. If $\gamma \leq \lambda_{\min}(\mathbf{G})^{-1}$ then $\|\bar{\mathbf{Q}}_*\| \leq \lambda_{\min}(\mathbf{G})^{-1}$. If $\gamma \geq \lambda_{\min}(\mathbf{G})^{-1}$ then $\gamma \geq \lambda_j(\mathbf{G})^{-1}$ and we have $t = \sum_j \lambda_j(\bar{\mathbf{Q}}_*) = \gamma t - \sum_j \frac{1}{\lambda_j(\mathbf{G})}$. Hence, $\gamma = 1 + \frac{1}{t}\sum_j \frac{1}{\lambda_j(\mathbf{G})} \leq 1 + \frac{1}{\lambda_{\min}(\mathbf{G})}$. In both cases, we have

$$\|\bar{\mathbf{Q}}_*\| \leq 1 + \frac{1}{\lambda_{\min}(\mathbf{G})}. \quad (125)$$

It remains to prove

$$\forall\, \mathbf{Q} \in \mathcal{C}_1, \quad \inf_t \lambda_{\min}(\mathbf{G}(\delta_R(\mathbf{Q}), \delta_T(\mathbf{Q}))) > 0 \quad (126)$$

and we are done. To this end, we first show that $\inf_t \delta_R(\mathbf{Q}) > 0$ for all $\mathbf{Q} \in \mathcal{C}_1$. From (37) and (40), we have

$$\begin{aligned}\delta_R(\mathbf{Q}) &= \frac{1}{t}\,\mathrm{tr}\,\mathbf{C}_R\mathbf{T}_R(\sigma^2)\\ &\geq \lambda_{\min}(\mathbf{C}_R)\frac{1}{t}\,\mathrm{tr}\,\mathbf{T}_R(\sigma^2)\\ &\overset{(a)}{\geq} \lambda_{\min}(\mathbf{C}_R)\left[\frac{1}{t}\,\mathrm{tr}\left(\sigma^2\mathbf{I}_r + \frac{\sigma^2}{K+1}\delta_T\mathbf{C}_R + \frac{K}{K+1}\mathbf{A}\mathbf{Q}^{1/2}\right.\right.\\ &\qquad \left.\left.\times\left(\mathbf{I}_t + \frac{\delta_R}{K+1}\mathbf{Q}^{1/2}\mathbf{C}_T\mathbf{Q}^{1/2}\right)^{-1}\mathbf{Q}^{1/2}\mathbf{A}^H\right)\right]^{-1}\\ &\overset{(b)}{\geq} \lambda_{\min}(\mathbf{C}_R)\left(\frac{1}{t}\,\mathrm{tr}\,\left(\sigma^2\mathbf{I}_r + \frac{\sigma^2}{K+1}\delta_T\mathbf{C}_R\right.\right.\\ &\qquad \left.\left. + \frac{K}{K+1}\mathbf{A}\mathbf{Q}\mathbf{A}^H\right)\right)^{-1} \quad (127)\end{aligned}$$

where (a) follows from Jensen's Inequality and (b) is due to the facts that $\|(\mathbf{I}_t + \mathbf{Y})^{-1}\| \leq 1$ and $\mathrm{tr}\,(\mathbf{X}\mathbf{Y}) \leq \|\mathbf{X}\|\mathrm{tr}\,(\mathbf{Y})$ when $\mathbf{Y}$ is a nonnegative matrix. We now find an upper bound for δ_T. From (39) and (15), we have $\|\mathbf{T}_T(\sigma^2)\| \leq 1/\sigma^2$. Using (40) we then have

$$\delta_T \leq \|\mathbf{T}_T\|\frac{1}{t}\mathrm{tr}\,\mathbf{C}_T\mathbf{Q} \leq \|\mathbf{T}_T\|\|\mathbf{C}_T\|\frac{1}{t}\mathrm{tr}\,\mathbf{Q} \leq \frac{A}{\sigma^2}$$

(recall that $\frac{1}{t}\mathrm{tr}\,\mathbf{Q} = 1$). Getting back to (127), we easily obtain

$$\begin{aligned}&\frac{1}{t}\mathrm{tr}\,\left(\sigma^2\mathbf{I}_r + \frac{\sigma^2}{K+1}\delta_T\mathbf{C}_R + \frac{K}{K+1}\mathbf{A}\mathbf{Q}\mathbf{A}^H\right)\\ &\leq \frac{r}{t}\left(\sigma^2 + \frac{A}{K+1}\right) + \frac{A^2K}{K+1} \leq C_0 \quad \forall (t,r),\ \frac{t}{r} \to c \quad (128)\end{aligned}$$

where C_0 is a certain constant term. Hence we have $\delta_R(\mathbf{Q}) \geq aC_0^{-1}$. By inspecting the expression (49) of $\mathbf{G}(\delta_R, \delta_T)$, we then obtain

$$\lambda_{\min}(\mathbf{G}) \geq \frac{aC_0^{-1}}{K+1}\lambda_{\min}(\mathbf{C}_T) \geq \frac{a^2C_0^{-1}}{K+1} = C_1 > 0$$

and (126) is proven. It remains to plug this estimate into (125) and (i) is proved.

Proof of (ii): We begin by restricting the maximization of $I(\mathbf{Q})$ to the set $\mathcal{C}_1^{\mathrm{d}} = \{\mathbf{Q} : \mathbf{Q} = \mathrm{diag}\,(q_1, \ldots, q_t) \geq \mathbf{0}, \mathrm{tr}\,(\mathbf{Q}) = t\}$ of the diagonal matrices within $\mathcal{C}_1$, and show that $\mathbf{Q}_*^{\mathrm{d}} = \arg\max_{\mathbf{Q}\in\mathcal{C}_1^{\mathrm{d}}} I(\mathbf{Q})$ satisfies $\sup_t \|\mathbf{Q}_*^{\mathrm{d}}\| < \infty$ where the bound is a function of (a, A, σ^2, c, K) only. The set $\mathcal{C}_1^{\mathrm{d}}$ is clearly convex and the solution $\mathbf{Q}_*^{\mathrm{d}}$ is given by the Lagrange Karush-Kuhn-Tucker (KKT) conditions

$$\frac{\partial I(\mathbf{Q})}{\partial q_j} = \frac{\partial}{\partial q_j}\mathbb{E}[\mathcal{I}(\mathbf{Q})] = \eta - \beta_j \quad (129)$$

where $\mathcal{I}(\mathbf{Q}) = \log\det(\mathbf{I}_r + \frac{1}{\sigma^2}\mathbf{H}\mathbf{Q}\mathbf{H}^H)$ and the Lagrange multipliers η and the β_i are associated with the power constraint and with the positivity constraints, respectively. More specifically, η is the unique real positive number for which $\sum_{j=1}^t q_j = t$, and the β_j satisfy $\beta_j = 0$ if $q_j > 0$ and $\beta_j \geq 0$ if $q_j = 0$. We have

$$\frac{\partial \mathcal{I}(\mathbf{Q})}{\partial q_j} = \frac{1}{\sigma^2}\mathbf{h}_j^H\left(\mathbf{I}_r + \frac{1}{\sigma^2}\mathbf{H}\mathbf{Q}\mathbf{H}^H\right)^{-1}\mathbf{h}_j$$

where $\mathbf{h}_j$ the jth column of $\mathbf{H}$. By consequence, $\mathbb{E}[\partial\mathcal{I}(\mathbf{Q})/\partial q_j] \leq \frac{1}{\sigma^2}\mathbb{E}[\|\mathbf{h}_j\|^2]$. As $\mathbf{h}_j$ is a Gaussian vector, the RHS of this inequality is defined and, therefore, by the Dominated Convergence Theorem, we can exchange $\partial/\partial q_j$ with $\mathbb{E}$ in (129) and write

$$\frac{\partial I(\mathbf{Q})}{\partial q_j} = \frac{1}{\sigma^2}\mathbb{E}\left[\mathbf{h}_j^H\left(\mathbf{I}_r + \frac{1}{\sigma^2}\mathbf{HQH}^H\right)^{-1}\mathbf{h}_j\right]. \quad (130)$$

Let us denote by $\mathbf{H}_j$ the $r \times (t-1)$ matrix that remains after extracting $\mathbf{h}_j$ from $\mathbf{H}$. Similarly, we denote by $\mathbf{Q}_j$ the $(t-1) \times (t-1)$ diagonal matrix that remains after deleting row and column j from $\mathbf{Q}$. Writing $\mathbf{R}_j = (\mathbf{I}_r + \frac{1}{\sigma^2}\mathbf{H}_j\mathbf{Q}_j\mathbf{H}_j^H)^{-1}$, we have by the Matrix Inversion Lemma ([21, Sec. 0.7.4])

$$\left(\mathbf{I}_r + \frac{1}{\sigma^2}\mathbf{HQH}^H\right)^{-1} = \mathbf{R}_j - \frac{q_j}{\sigma^2 + q_j\mathbf{h}_j^H\mathbf{R}_j\mathbf{h}_j}\mathbf{R}_j\mathbf{h}_j\mathbf{h}_j^H\mathbf{R}_j.$$

By plugging this expression into the RHS of (130), the Lagrange-KKT conditions become

$$\mathbb{E}\left[\frac{X_j}{\sigma^2 + q_jX_j}\right] = \eta - \beta_j \quad (131)$$

where $X_j = \mathbf{h}_j^H\mathbf{R}_j\mathbf{h}_j$. A consequence of this last equation is that $q_j \leq 1/\eta$ for every j. Indeed, assume that $q_j > 1/\eta$ for some j. Then $\sigma^2 + q_jX_j > X_j/\eta$ hence $\mathbb{E}[\frac{X_j}{\sigma^2+q_jX_j}] < \eta$, therefore, $\beta_j > 0$ (131), which implies that $q_j = 0$, a contradiction. As a result, in order to prove that $\sup_t \|\mathbf{Q}_*^{\mathrm{d}}\| < \infty$, it will be enough to prove that $\sup_t 1/\eta < \infty$. To this end, we shall prove that there exists a constant $C > 0$ such that

$$\max_{j=1,\ldots,t} \mathbb{P}(X_j \leq C) \underset{t\to\infty}{\longrightarrow} 0. \quad (132)$$

Indeed, let us admit (132) temporarily. We have

$$\begin{aligned}
&\mathbb{E}\left[\frac{X_j}{\sigma^2 + q_jX_j}\right] - \frac{C}{\sigma^2 + q_jC} \\
&\quad = \mathbb{E}\left[\frac{X_j}{\sigma^2 + q_jX_j}\mathbf{1}_{X_j>C}\right] - \frac{C}{\sigma^2 + q_jC} \\
&\qquad + \mathbb{E}\left[\frac{X_j}{\sigma^2 + q_jX_j}\mathbf{1}_{X_j\leq C}\right] \\
&\quad \geq \frac{C}{\sigma^2 + q_jC}\mathbb{P}(X_j > C) - \frac{C}{\sigma^2 + q_jC} \\
&\quad = \varepsilon_j
\end{aligned}$$

where $\varepsilon_j = -\frac{C}{\sigma^2+q_jC}\mathbb{P}(X_j \leq C)$, and the inequality is due to the fact that the function $f(x) = \frac{x}{\sigma^2+q_jx}$ is increasing. As

$$\max_{j=1,\ldots,t} |\varepsilon_j| \leq \frac{C}{\sigma^2}\max_{j=1,\ldots,t}\mathbb{P}(X_j \leq C) \underset{t\to\infty}{\longrightarrow} 0$$

by (132), we have

$$\liminf_t \min_j \left(\mathbb{E}\left[\frac{X_j}{\sigma^2 + q_jX_j}\right] - \frac{C}{\sigma^2 + q_jC}\right) \geq 0.$$

Getting back to the Lagrange KKT condition (131) we therefore have for t large enough $\eta - \beta_j > \frac{C/2}{\sigma^2+q_jC/2}$ for every $j = 1,\ldots,t$. By consequence

$$\frac{1}{\eta} \leq \frac{1}{\eta - \beta_j} < \frac{2\sigma^2}{C} + q_j$$

for large t. Summing over j and taking into account the power constraint $\sum_j q_j = t$, we obtain $\frac{t}{\eta} < \frac{2\sigma^2 t}{C} + t$, i.e., $\frac{1}{\eta} < \frac{2\sigma^2}{C} + 1$ and

$$\sup_t \|\mathbf{Q}_*^{\mathrm{d}}\| < \frac{2\sigma^2}{C} + 1 \quad (133)$$

which is the desired result. To prove (132), we make use of MMSE estimation theory. Recall that $\mathbf{H} = \sqrt{\frac{K}{K+1}}\mathbf{A} + \frac{1}{\sqrt{K+1}}\frac{1}{\sqrt{t}}\mathbf{C}_R^{1/2}\mathbf{W}\mathbf{C}_T^{1/2}$. Denoting by $\mathbf{a}_j$ and $\mathbf{z}_j$ the j^{th} columns of the matrices $\mathbf{A}$ and $\mathbf{W}\mathbf{C}_T^{1/2}$, respectively, we have

$$\begin{aligned}
X_j = &\left(\sqrt{\frac{K}{K+1}}\mathbf{a}_j^H + \frac{1}{\sqrt{K+1}}\frac{1}{\sqrt{t}}\mathbf{z}_j^H\mathbf{C}_R^{1/2}\right)\mathbf{R}_j \\
&\times\left(\sqrt{\frac{K}{K+1}}\mathbf{a}_j + \frac{1}{\sqrt{K+1}}\frac{1}{\sqrt{t}}\mathbf{C}_R^{1/2}\mathbf{z}_j\right).
\end{aligned}$$

We decompose $\mathbf{z}_j$ as $\mathbf{z}_j = \mathbf{u}_j + \mathbf{u}_j^\perp$ where $\mathbf{u}_j$ is the conditional expectation $\mathbf{u}_j = \mathbb{E}[\mathbf{z}_j\|\mathbf{z}_1,\ldots,\mathbf{z}_{j-1},\mathbf{z}_{j+1},\ldots,\mathbf{z}_t]$, in other words, $\mathbf{u}_j$ is the MMSE estimate of $\mathbf{z}_j$ drawn from the other columns of $\mathbf{W}\mathbf{C}_T^{1/2}$. Put

$$\begin{aligned}
S_j = 2\Re&\left(\frac{1}{\sqrt{K+1}}\frac{1}{\sqrt{t}}\mathbf{u}_j^{\perp H}\mathbf{C}_R^{1/2}\mathbf{R}_j\right. \\
&\left.\times\left(\sqrt{\frac{K}{K+1}}\mathbf{a}_j + \frac{1}{\sqrt{K+1}}\frac{1}{\sqrt{t}}\mathbf{C}_R^{1/2}\mathbf{u}_j\right)\right) \\
&+ \frac{1}{t(K+1)}\mathbf{u}_j^{\perp H}\mathbf{C}_R^{1/2}\mathbf{R}_j\mathbf{C}_R^{1/2}\mathbf{u}_j^\perp. \quad (134)
\end{aligned}$$

Then

$$\begin{aligned}
X_j = S_j + &\left(\sqrt{\frac{K}{K+1}}\mathbf{a}_j^H + \frac{1}{\sqrt{K+1}}\frac{1}{\sqrt{t}}\mathbf{u}_j^H\mathbf{C}_R^{1/2}\right)\mathbf{R}_j \\
&\times\left(\sqrt{\frac{K}{K+1}}\mathbf{a}_j + \frac{1}{\sqrt{K+1}}\frac{1}{\sqrt{t}}\mathbf{C}_R^{1/2}\mathbf{u}_j\right) \\
\geq S_j.& \quad (135)
\end{aligned}$$

Let us study the asymptotic behavior of S_j. First, we note that due to the fact that the joint distribution of the elements of $\mathbf{W}\mathbf{C}_T^{1/2}$ is the Gaussian distribution, $\mathbf{u}_j^\perp$ and $\mathbf{v}_j = [\mathbf{z}_1^T,\ldots,\mathbf{z}_{j-1}^T,\mathbf{z}_{j+1}^T,\ldots,\mathbf{z}_t^T]^T$ are independent. By consequence, $\mathbf{u}_j^\perp$ and $(\mathbf{R}_j,\mathbf{u}_j)$ are independent. Let us derive the expression of the covariance matrix $\mathbf{R}_\mathbf{u} = \mathbb{E}[\mathbf{u}_j^\perp\mathbf{u}_j^{\perp H}]$. From the well-known formulas for MMSE estimation ([33]), we have $\mathbf{R}_\mathbf{u} = \mathbb{E}[\mathbf{z}_j\mathbf{z}_j^H] - \mathbb{E}[\mathbf{z}_j\mathbf{v}_j^H]\ (\mathbb{E}[\mathbf{v}_j\mathbf{v}_j^H])^{-1}\mathbb{E}[\mathbf{v}_j\mathbf{z}_j^H]$. To obtain $\mathbf{R}_\mathbf{u}$, we note that the covariance matrix of the vector $\mathbf{z} = [\mathbf{z}_1^T,\ldots,\mathbf{z}_t^T]^T$ is $\mathbb{E}[\mathbf{z}\mathbf{z}^H] = \mathbf{C}_T^T \otimes \mathbf{I}_r$ (just check that $\mathbb{E}[[\mathbf{W}\mathbf{C}_T^{1/2}]_{ij}[\mathbf{W}\mathbf{C}_T^{1/2}]_{kl}] = \delta(i-k)[\mathbf{C}_T]_{lj}$). Let us denote by $\tilde{c}_j, \tilde{\mathbf{c}}_j$ and $\mathbf{C}_{T,j}$ the scalar $\tilde{c}_j = [\mathbf{C}_T]_{jj}$, the jth vector column of $\mathbf{C}_T$ without element $\tilde{c}_j$, and the $(t-1)\times(t-1)$ matrix that remains after extracting row and column j from $\mathbf{C}_T$, respectively. With these notations we have $\mathbf{R}_\mathbf{u} = (\tilde{c}_j - \tilde{\mathbf{c}}_j^H\mathbf{C}_{T,j}^{-1}\tilde{\mathbf{c}}_j)\mathbf{I}_r$. Recalling that $\mathbf{u}_j^\perp$ and $(\mathbf{R}_j,\mathbf{u}_j)$ are independent, one may see that the first term of the RHS of (134) is negligible while the second is close to $\rho_j = \frac{1}{t}\frac{\tilde{c}_j - \tilde{\mathbf{c}}_j^H\mathbf{C}_{T,j}^{-1}\tilde{\mathbf{c}}_j}{K+1}\mathrm{tr}\ (\mathbf{R}_j\mathbf{C}_R)$.

More rigorously, using this independence in addition to $A = \max(\|\mathbf{A}\|, \|\mathbf{C}_R\|, \|\mathbf{C}_T\|) < \infty$ and $\|\mathbf{R}_j\| \leq 1$, we can prove with the help of [1, Lemma 2.7] or by direct calculation that there exists a constant C_1 such that

$$\mathbb{E}[(S_j - \rho_j)^2] \leq \frac{C_1}{t}. \tag{136}$$

In order to prove (132), we will prove that the ρ_j are bounded away from zero in some sense. First, we have

$$\tilde{c}_j - \tilde{\mathbf{c}}_j^H \mathbf{C}_{T,j}^{-1} \tilde{\mathbf{c}}_j \overset{(a)}{=} \left[\mathbf{C}_T^{-1}\right]_{jj}^{-1} \overset{(b)}{\geq} \|\mathbf{C}_T^{-1}\|^{-1} = \lambda_{\min}(\mathbf{C}_T) \geq a$$

(for (a) see [21, Sec. 0.7.3] and for (b), use the fact that $|[\mathbf{X}]_{kl}| \leq \|\mathbf{X}\|$ for any element (k,l) of a matrix $\mathbf{X}$). By consequence

$$\begin{aligned}\rho_j &\geq \frac{a\lambda_{\min}(\mathbf{C}_R)}{K+1}\frac{1}{t}\mathrm{tr}\left(\mathbf{I}_r + \frac{1}{\sigma^2}\mathbf{H}_j\mathbf{Q}_j\mathbf{H}_j^H\right)^{-1}\\ &\overset{(a)}{\geq} \frac{a\lambda_{\min}(\mathbf{C}_R)}{K+1}\left(\frac{1}{t}\mathrm{tr}\left(\mathbf{I}_r + \frac{1}{\sigma^2}\mathbf{H}_j\mathbf{Q}_j\mathbf{H}_j^H\right)\right)^{-1}\\ &\overset{(b)}{\geq} \frac{a^2}{K+1}\left(\frac{r}{t} + \frac{1}{\sigma^2}\left(\|\mathbf{A}\| + \|\mathbf{C}_R\|^{1/2}\|\mathbf{C}_T\|^{1/2}\|\frac{1}{\sqrt{t}}\mathbf{W}\|\right)^2\right.\\ &\qquad\left.\times\frac{1}{t}\mathrm{tr}\,(\mathbf{Q})\right)^{-1}\end{aligned}$$

where (a) is Jensen Inequality and (b) is due to $\mathrm{tr}\,(\mathbf{XY}) \leq \|\mathbf{X}\|\mathrm{tr}\,(\mathbf{Y})$ when $\mathbf{Y}$ is a nonnegative matrix. As $\lim_t \|\frac{1}{\sqrt{t}}\mathbf{W}\| = 1 + \sqrt{1/c}$ with probability one ([1]), and furthermore, $\mathrm{tr}\,(\mathbf{Q}) = t$, we have with probability one

$$\liminf_t \min_{j=1,\ldots,t} \rho_j \geq \frac{a^2}{K+1}\left(c^{-1} + \frac{A^2}{\sigma^2}(2 + c^{-1/2})^2\right)^{-1} = C_2. \tag{137}$$

Choose the constant C in the left-hand side (LHS) of (132) as $C = C_2/4$. From (132) we have

$$\begin{aligned}&\max_j \mathbb{P}(X_j \leq C)\\ &\quad\leq \max_j \mathbb{P}(S_j \leq C)\\ &\quad= \max_j \mathbb{P}(S_j \leq C, |S_j - \rho_j| \geq C)\\ &\qquad+ \max_j \mathbb{P}(S_j \leq C, |S_j - \rho_j| < C)\\ &\quad\leq \max_j \mathbb{P}(|S_j - \rho_j| \geq C) + \max_j \mathbb{P}(\rho_j \leq 2C)\\ &\quad\overset{(a)}{\leq} \frac{1}{C^2}\max_j \mathbb{E}[(S_j - \rho_j)^2] + \max_j \mathbb{P}(\rho_j \leq 2C)\\ &\quad\overset{(b)}{\leq} \frac{1}{C^2}\max_j \mathbb{E}[(S_j - \rho_j)^2] + \mathbb{P}(\min_j \rho_j \leq 2C)\\ &\quad\overset{(c)}{=} o(1)\end{aligned}$$

where (a) is Tchebychev's Inequality, (b) is due to $\max_j \mathbb{P}(\mathcal{E}_j) \leq \mathbb{P}(\cup_j \mathcal{E}_j)$, and (c) is due to (136) and to (137).

We have proven (132) and hence that $\mathbf{Q}_*^{\mathrm{d}} = \arg\max_{\mathbf{Q}\in\mathcal{C}_1^{\mathrm{d}}} I(\mathbf{Q})$ satisfies $\sup_t \|\mathbf{Q}_*^{\mathrm{d}}\| < \infty$. \cr In order to prove that $\mathbf{Q}_* = \arg\max_{\mathbf{Q}\in\mathcal{C}_1} I(\mathbf{Q})$ satisfies $\sup_t \|\mathbf{Q}_*\| < \infty$, we begin by noticing that

$$\max_{\mathbf{Q}\in\mathcal{C}_1} I(\mathbf{Q}) = \max_{\mathbf{U}\in\mathcal{U}_t}\max_{\mathbf{\Lambda}\in\mathcal{C}_1^{\mathrm{d}}} \mathbb{E}\left[\log\det\left(\mathbf{I}_r + \frac{1}{\sigma^2}\mathbf{HU\Lambda U}^H\mathbf{H}^H\right)\right] \tag{138}$$

where $\mathcal{U}_t$ is the group of unitary $t \times t$ matrices. For a given matrix $\mathbf{U} \in \mathcal{U}_t$, the inner maximization in (138) is equivalent to the problem of maximizing the mutual information over $\mathcal{C}_1^{\mathrm{d}}$ when the channel matrix $\mathbf{H}$ is replaced with $\mathbf{H}' = \mathbf{HU} = \sqrt{\frac{K}{K+1}}\mathbf{A}' + \frac{1}{\sqrt{K+1}}\frac{1}{\sqrt{t}}\mathbf{C}_R^{1/2}\mathbf{W}'(\mathbf{C}_T')^{1/2}$. Here, matrix $\mathbf{C}_T'$ is defined by $\mathbf{C}_T' = \mathbf{U}^H\mathbf{C}_T\mathbf{U}$, $\mathbf{A}' = \mathbf{AU}$, $\mathbf{W}' = \mathbf{W\Theta}$ where $\mathbf{\Theta}$ is the unitary matrix $\mathbf{\Theta} = \mathbf{C}_T^{1/2}\mathbf{U}\mathbf{C}_T^{'-1/2}$. As $\mathbf{U} \in \mathcal{U}_t$, we clearly have $\|\mathbf{A}'\| = \|\mathbf{A}\|$, $\|\mathbf{C}_T'\| = \|\mathbf{C}_T\|$, and $\|\mathbf{C}_T^{'-1}\| = \|\mathbf{C}_T^{-1}\|$. By consequence, the bounds a and A, and hence the constant C in the left hand member of (132) (which depends only on (a, A, σ^2, c, K)) remain unchanged when we replace $\mathbf{H}$ with $\mathbf{H}'$. By consequence, for every $\mathbf{U} \in \mathcal{U}_t$ the matrix $\mathbf{\Lambda}_*(\mathbf{U})$ that maximizes $\mathbb{E}[\log\det(\mathbf{I}_r + \frac{1}{\sigma^2}\mathbf{HU\Lambda U}H\mathbf{H}^H)]$ satisfies $\|\mathbf{\Lambda}_*(\mathbf{U})\| < 2\sigma^2/C + 1$ [see (132) which is independent of $\mathbf{U}$]. Hence, $\|\mathbf{Q}_*\| < 2\sigma^2/C + 1$ which terminates the proof of (ii).

REFERENCES

[1] Z. D. Bai and J. W. Silverstein, "No eigenvalues outside the support of the limiting spectral distribution of large dimensional sample covariance matrices," *Ann. Prob.*, vol. 26, no. 1, pp. 316–345, 1998.

[2] Z. Bai and J. Silverstein, "CLT for linear statistics of large-dimensional sample covariance matrices," *Ann. Probab.*, vol. 32, no. 1A, pp. 553–605, 2004.

[3] E. Biglieri, G. Taricco, and A. Tulino, "How far is infinite? using asymptotic analyses in multiple-antennas systems," *Proc. ISSTA-02*, vol. 1, pp. 1–6, 2002.

[4] H. Bolcskei, D. Gesbert, and A. Paulraj, "On the capacity of OFDM-based spatial multiplexing systems," *IEEE Trans. Commun.*, vol. 50, no. 2, pp. 225–234, Feb. 2002.

[5] J. Borwein and A. Lewis, *Convex Analysis and Nonlinear Optimization*. New York: Springer-Verlag, 2000.

[6] H. Cartan, "Théorie elementaire des fonctions analytiques d'une ou plusieurs variables complexes," *Hermann*, 1978.

[7] S. Chatterjee and A. Bose, "A new method for bounding rates of convergence of empirical spectral distributions," *J. Theoret. Probab.*, vol. 17, no. 4, pp. 1003–1019, Oct. 2004.

[8] L. H. Chen, "An inequality for the multivariate normal distribution," *J. Multiv. Anal.*, vol. 12, no. 2, pp. 306–315, 1982.

[9] L. Cottatellucci and M. Debbah, "The effect of line of sight components on the asymptotic capacity of MIMO systems," in *Proc. ISIT 04*, Chicago, IL, Jun.–Jul. 2004.

[10] C. N. Chuah, D. N. C. Tse, J. M. Kahn, and R. A. Valenzuela, "Capacity scaling in MIMO wireless systems under correlated fading," *IEEE Trans. Inf. Theory*, vol. 48, no. 3, pp. 637–650, Mar. 2002.

[11] J. Dumont, Ph. Loubaton, S. Lasaulce, and M. Debbah, "On the asymptotic performance of MIMO correlated rician channels," *Proc. ICASSP05*, vol. 5, pp. 813–816, Mar. 2005.

[12] J. Dumont, W. Hachem, P. Loubaton, and J. Najim, "On the asymptotic analysis of mutual information of MIMO Rician correlated channels," in *Proc. Int. Symp. on Commun. Control Signal Process.*, Marrakech, Mar. 2006.

[13] J. Dumont, P. Loubaton, and S. Lasaulce, "On the capacity achieving covariance matrix for Rician MIMO channels: An asymptotic approach," in *Proc. Globecom 2006*, San Francisco, CA, Dec. 2006.

[14] V. L. Girko, *Theory of Stochastic Canonical Equations, Volume I Chap. 7*. Dordrecht, The Netherlands: Kluwer Academic, 2001.

[15] J. Glimm and A. Jaffe, *Quantum Physics: A Functional Point of View*, 2nd ed. New York: Springer-Verlag, 1987.

[16] A. Goldsmith, S. A. Jafar, N. Jindal, and S. Vishwanath, "Capacity limits of MIMO channels," *IEEE J. Sel. Areas Commun.*, vol. 21, no. 5, Jun. 2003.

[17] W. Hachem, O. Khorunzhiy, P. Loubaton, J. Najim, and L. Pastur, "A new approach for capacity analysis of large dimensional multi-antenna channels," *IEEE Trans. Inf. Theory*, vol. 54, no. 9, pp. 3987–4004, Sep. 2008.

[18] W. Hachem, P. Loubaton, and J. Najim, “Deterministic equivalents for certain functionals of large random matrices,” *Ann. Appl. Probab.*, vol. 17, no. 3, pp. 875–930, 2007.
[19] J. Hansen and H. Bölcskei, “A geometrical investigation of the rank-1 Rician MIMO channel at high SNR,” in *Proc. IEEE Int. Symp. Inf. Theory*, Chicago, IL, Jun./Jul. 2004.
[20] D. Hoesli, Y. H. Kim, and A. Lapidoth, “Monotonicity results for coherent MIMO Rician channels,” *IEEE Trans. Inf. Theory*, vol. 51, no. 12, pp. 4334–4339, Dec. 2005.
[21] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1994.
[22] C. Houdré, V. Pérez-Abreu, and D. Surgailis, “Interpolation, correlation identities, and inequalities for infinitely divisible variables,” *J. Fourier Anal. Appl.*, vol. 4, no. 6, pp. 651–668, 1998.
[23] M. Kang and M. S. Alouini, “Capacity of MIMO Rician channels,” *IEEE Trans. Wireless Commun.*, vol. 5, no. 1, pp. 112–122, Jan. 2006.
[24] M. Kang and M. S. Alouini, “Capacity of correlated MIMO Rayleigh channels,” *IEEE Trans. Wireless Commun.*, vol. 5, no. 1, pp. 143–155, Jan. 2006.
[25] M. G. Krein and A. A. Nudelman, *The Markov Moment Problem and Extremal Problems*. Providence, RI: Amer. Math. Society, 1997.
[26] A. M. Khorunzhy and L. A. Pastur, “Limits of infinite interaction radius, dimensionality and the number of components for random operators with off-diagonal randomness,” *Commun. Math. Phys.*, vol. 153, no. 3, pp. 605–646, 1993.
[27] D. G. Luenberger, *Optimization by Vector Space Methods*. New York: Wiley, 1969.
[28] V. A. Marcenko and L. A. Pastur, “Distribution of eigenvalues in a certain set of random matrices,” *Mat. Sb.*, vol. 72, no. 114, pp. 507–526, 1967.
[29] A. L. Moustakas, S. H. Simon, and A. M. Sengupta, “MIMO capacity through correlated channels in the presence of correlated interference and noise : A (not so) large N analysis,” *Trans. Inf. Theory*, vol. 49, no. 10, pp. 2545–2561, Oct. 03.
[30] A. L. Moustakas and S. H. Simon, “Random matrix theory of multi-antenna communications: The Rician case,” *J. Phys. A: Math. Gen.*, vol. 38, pp. 10859–10872, 2005.
[31] L. A. Pastur, A. M. Khorunzhy, and V. Y. Vasil’chuck, “On the asymptotic property of the spectrum of the sum of one-dimensional independent random operators,” *Dopov. Nats. Akad. Nauk Ukraini*, vol. 2, pp. 27–30, 1995.
[32] L. A. Pastur, “A simple approach for the study of global regime of large random matrices,” *Ukrainian Math. J.*, vol. 57, no. 6, pp. 936–966, Jun. 2005.
[33] L. L. Scharf, *Statistical Signal Processing*. Reading, MA: Addison-Wesley, 1991.
[34] J. Silverstein, “The smallest eigenvalue of a large-dimensional wishart matrix,” *Ann. Probab.*, vol. 13, no. 4, pp. 1364–1368, 1985.
[35] G. Taricco, “On the capacity of separately-correlated MIMO rician fading channels,” in *Proc. Globecom 2006*, San Francisco, CA, Dec. 2006.
[36] G. Taricco and E. Riegler, “On the ergodic capacity of the asymptotic separately correlated Rician fading MIMO channel with interference,” in *Proc. ISIT 2007*, Nice, France, 24-June 29 2007, pp. 531–535.
[37] G. Taricco, “Asymptotic mutual information statistics of separately correlated Rician fading MIMO channels,” *Trans. Inf. Theory*, vol. 54, no. 8, pp. 3490–3504, Aug. 08.
[38] E. Telatar, “Capacity of multi-antenna Gaussian channels,” *Eur. Trans. Telecom.*, vol. 10, pp. 585–595, Nov. 99.
[39] A. M. Tulino, A. Lozano, and S. Verdú, “Capacity-achieving input covariance for correlated multi-antenna channels,” in *Proc. 41st Ann. Allerton Conf. Comm., Control Comput.*, Monticello, IL, Oct. 03.
[40] A. M. Tulino and S. Verdu, “Random matrix theory and wireless communications,” *Found. Trends Commun. Inf. Theory*, vol. 1, pp. 1–182, Jun. 2004, Now Publishers.
[41] M. Vu and A. Paulraj, “Capacity optimization for Rician correlated MIMO wireless channels,” in *Proc. Asilomar Conf.*, Asilomar, Nov. 2005, pp. 133–138.
[42] C.-K. Wen, K.-K. Wong, P. Ting, and C.-L. I, “Optimal power-saving input covariance for MIMO wireless systems exploiting only channel spatial correlations,” *Proc. IEEE ICC 2005*, vol. 3, pp. 2011–2015, May 2005.
[43] C.-K. Wen, P. Ting, and J.-T. Chen, “Asymptotic analysis of MIMO wireless systems with spatial correlation at the receiver,” *IEEE Trans. Commun.*, vol. 54, no. 2, pp. 349–363, Feb. 2006.
[44] J. Dumont, W. Hachem, S. Lasaulce, P. Loubaton, and J. Najim, *On the Capacity Achieving Covariance Matrix for Rician MIMO Channels: An Asymptotic Approach*, (extended version) arXiv:0710.4051.

Julien Dumont was born in Flers, France, in 1979. He entered the Ecole Normale Supérieure of Cachan, France, in 1999, and received the “Agrégation” degree in applied physics in 2002, the Master’s degree in control in 2003 from ENS Cachan and University of Paris Sud-Orsay, and the Ph.D. degree from University of Marne-La-Vallée in 2006.

From 2003 to 2006, he worked with France Telecom Labs and the University of Marne-La-Vallée as a Research Engineer and a Ph.D. student. In 2007, he became a Professor of physics in the preparatory classes to the national entrance competitive exam to French engineering schools. His current research interests are random matrix theory and game theory.

Walid Hachem was born in Bhamdoun, Lebanon, in 1967. He received the Engineering degree in telecommunications from St. Joseph University (ESIB), Beirut, Lebanon, in 1989, the Masters degree from Telecom ParisTech (then Ecole Nationale Supérieure des Télécommunications, ENST), France, in 1990, the Ph.D. degree in signal processing from Université de Marne-La-Vallée in 2000, and the Habilitation à Diriger des Recherches from the Université de Paris-Sud (Orsay), in 2006.

Between 1990 and 2000, he worked with Philips, and then with FERMA, as a signal processing engineer. In 2001, he joined Ecole Supérieure d’Electricité (Supélec). Since 2007, he has been a Centre national de la Recherche Scientifique (CNRS) researcher with Telecom ParisTech. His research interests concern the analysis of communication systems based on the study of large random matrices, detection and estimation, and cooperative wireless networks.

Dr. Hachem currently serves as an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING.

Samson Lasaulce received the B.Sc. and Agrégation degree in applied physics from École Normale Supérieure (Cachan) and the M.Sc. and Ph.D. degrees in signal processing from the École Nationale Supérieure des Téécommunications, Paris.

He has been with Motorola Labs (1999, 2000, 2001) and France Télécom Labs (2002, 2003). Since 2004, he has been with the CNRS and Supélec and is Chargé d’Enseignement at École Polytechnique. His broad interests lie in the areas of communications, signal processing, information theory, and game theory for wireless communications.

Dr. Lasaulce is the recipient of the 2007 ACM/ICST Valuetools conference and the 2009 IEEE Crowncom Conference Best Student Paper awards.

Philippe Loubaton (M’88–SM’06–F’08) was born in 1958 in Villers Semeuse, France. He received the M.Sc. and the Ph.D. degrees from Ecole Nationale Supérieure des Télécommunications, Paris, France, in 1981 and 1988, respectively.

From 1982 to 1986, he was a member of the Technical Staff of Thomson-CSF/RGS, where he worked in digital communications. From 1986 to 1988, he was with the Institut National des Télécommunications as an Assistant Professor of Electrical Engineering. In 1988, he joined the Signal Processing Department, Ecole Nationale Supérieure des Télécommunications, Paris. Since 1995, he has been a Professor of Electrical Engineering at Marne la Vallée University, Champs sur Marne, France. After some works on two-parameter stationary stochastic processes, his present research interests are in statistical signal processing, and digital communications with a special emphasis on blind equalization, multiuser communication systems and multicarrier modulations, and performance analysis of large communication systems via large random matrices methods.

Dr. Loubaton served as an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING (1999–2001, 2005–2007) and for the IEEE COMMUNICATIONS LETTERS (2001–2003). He is currently an Associate Editor for EURASIP *Signal Processing* and the French journal *Traitement du Signal*. He was a member of the Technical Committee of the IEEE Signal Processing for Communications (1998–2004).

Jamal Najim received the *diplôme d’ingénieur* from Télécom ParisTech, France, in 1998, and the Ph.D. degree in probability from Université Nanterre Paris-10 in 2001.

Since 2002, he is with the Centre Nationale de la Recherche Scientifique (CNRS) and Télécom ParisTech. His research interests include wireless networks, the theory of random matrices, and probability at large.