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

J. Dumont, W. Hachem, S. Lasaulce, Ph. Loubaton and J. Najim

12 octobre 2007

Abstract

In this contribution, the capacity-achieving input covariance matrices for coherent block-fading correlated 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 by numerical techniques. As the corresponding optimization algorithms are not very attractive, an approximation of the average mutual information is evaluated in this paper in the asymptotic regime where the number of transmit and receive antennas converge to $+\infty$ at the same rate. New results related to the accuracy of the corresponding large system approximation are provided. An attractive optimization algorithm of this approximation is proposed and we establish that it yields an effective way to compute the capacity achieving covariance matrix for the average mutual information. Finally, 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, while being much more computationally attractive.

Index Terms

This work was partially supported 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.

J. Dumont was with France Telecom and Université Paris-Est, France. dumont@crans.org ,

W. Hachem and J. Najim are with CNRS and Télécom Paris, Paris, France. {hachem,najim}@enst.fr ,

S. Lasaulce is with CNRS and SUPELEC, France. lasaulce@lss.supelec.fr ,

P. Loubaton is with Université Paris Est, IGM LabInfo, UMR CNRS 8049, France. loubaton@univ-mlv.fr
,

I. INTRODUCTION

Since the seminal work of Telatar [39], 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). This choice will be justified in section II-C. Assuming the knowledge of the channel statistics at the transmitter, one important issue is then 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. [15] 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 [40] 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 [19] 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 undoubtedly more complicated because the eigenvectors of the optimum matrix have no closed form expressions. Moreover, the exact expression of the EMI being complicated (see e.g. [22]), both the eigenvalues and the eigenvectors have to be evaluated numerically. In [42], a barrier interior-point method is proposed and implemented to directly evaluate the EMI as an expectation. The corresponding algorithms are however not very attractive because they rely on computationally-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 evaluate an approximation of the EMI, valid when the number of transmit and receive antennas converge to $+\infty$ at the same rate, and then to optimize this asymptotic approximation.

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

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 [12].

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. The general case of a Rician correlated channel has recently been established in [17] using large random matrix theory and completes a number of previous works among which [9], [41] and [30] (Rayleigh channels), [8] and [31] (Rician uncorrelated channels), [10] (Rician receive correlated channel) and [37] (Rician correlated channels). Notice that the latest work (together with [30] and [31]) relies on the powerful but non-rigorous replica method. It also gives an expression for the variance of the mutual information. We finally mention the recent paper [38] in which the authors generalize our approach sketched in [12] to the MIMO Rician channel with interference. The optimization algorithm of the large system approximant of the EMI proposed in [38] is however different from our proposal.

In this paper, we rely on the results of [17] in which a closed-form asymptotic approximation for the mutual information is provided, and present new results concerning its accuracy. We then address the optimization of the large system approximation w.r.t. 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 [44] 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 (and in contrast with [44]) : We prove that the algorithm converges towards the optimum input covariance matrix as long as it converges. We also prove that the matrix which optimizes the large system approximation asymptotically achieves the capacity. This result 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. Validations, interpretations and numerical results are provided in section VI.

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 $(.)^T$ and $(.)^H$ 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, r and t 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 $\mathrm{Tr}\left[\mathbb{E}(\mathbf{x}\mathbf{x}^H)\right] \leq t$. 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

$$\mathbf{V} = \frac{1}{\sqrt{t}}\mathbf{C}^{\frac{1}{2}}\mathbf{W}\tilde{\mathbf{C}}^{\frac{1}{2}} \ , \tag{3}$$

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 $\mathcal{C}N(0,1)$, 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 $\tilde{\mathbf{C}} > 0$ and $\mathbf{C} > 0$ account for the transmit and receive antenna correlation effects respectively and satisfy $\frac{1}{t}\mathrm{Tr}(\tilde{\mathbf{C}}) = 1$ and $\frac{1}{r}\mathrm{Tr}(\mathbf{C}) = 1$. This correlation structure is often referred to as a separable or Kronecker correlation model.

Remark 1: Note that no extra assumption related to the rank of the deterministic component $\mathbf{A}$ of the channel is done. Generally, it is often assumed that $\mathbf{A}$ has rank one ([15], [27], [18], [26], etc..) because of the relatively small path loss exponent of the direct path. Although the rank-one assumption is often relevant, it becomes questionable if one wants to address, for instance, a multi-user setup and determine the sum-capacity of a cooperative multiple access or broadcast channel in the high cooperation regime. Consider for example a macro-diversity situation in the downlink : Several base stations interconnected [2] through ideal wireline channels cooperate to maximize the performance of a given multi-antenna receiver. Here the matrix $\mathbf{A}$ is likely to have a rank higher than one or even to be of full rank : Assume that the receive array of antennas is linear and uniform. Then a typical structure for $\mathbf{A}$ is

$$\mathbf{A} = \frac{1}{\sqrt{t}}\left[\mathbf{a}(\theta_1), \dots, \mathbf{a}(\theta_t)\right]\mathbf{\Lambda} \ , \tag{4}$$

where $\mathbf{a}(\theta) = (1, e^{i\theta}, \dots, e^{i(r-1)\theta})^T$ and $\mathbf{\Lambda}$ is a diagonal matrix whose entries represent the complex amplitudes of the t line of sight (LOS) components.

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 :

$$I(\mathbf{Q}) = \mathbb{E}_{\boldsymbol{H}}\left[\log\det\left(\mathbf{I}_r + \frac{1}{\sigma^2}\mathbf{H}\mathbf{Q}\mathbf{H}^H\right)\right] . \tag{5}$$

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.

[2]For example in a cellular system the base stations are connected with one another via a radio network controller.

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

$$C_E = \sup_{\mathbf{Q}\in\mathcal{C}_1} I(\mathbf{Q}). \tag{6}$$

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 [28]). When $\tilde{\mathbf{C}} = \mathbf{I}_t$, $\mathbf{C} = \mathbf{I}_r$, [19] shows that the eigenvectors of the optimal input covariance matrix coincide with the right-singular vectors of $\mathbf{A}$. By adapting the proof of [19], one can easily check that this result also holds when $\tilde{\mathbf{C}} = \mathbf{I}_t$ and $\mathbf{C}$ 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. [42]) which are very demanding since they rely on computationally-intensive Monte-Carlo simulations. This problem can be circumvented as the EMI $I(\mathbf{Q})$ can be approximated by a simple expression denoted by $\bar{I}(\mathbf{Q})$ (see section III) as $t \to \infty$ which in turn will be optimized with respect to $\mathbf{Q}$ (see section V).

Remark 2: Finding the optimum covariance matrix is useful in practice, in particular if the channel input is assumed to be Gaussian. In fact, there exist many practical space-time encoders that produce near-Gaussian outputs (these outputs are used as inputs for the linear precoder $\mathbf{Q}^{1/2}$). See for instance [34].

D. Summary of the main results.

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

1) We derive an accurate approximation of $I(\mathbf{Q})$ as $t \to +\infty$: $I(\mathbf{Q}) \simeq \bar{I}(\mathbf{Q})$ where

$$\bar{I}(\mathbf{Q}) = \log\det\left[\mathbf{I}_t + \mathbf{G}(\delta(\mathbf{Q}, \tilde{\delta}(\mathbf{Q}))\mathbf{Q}\right] + i(\delta(\mathbf{Q}), \tilde{\delta}(\mathbf{Q})) \tag{7}$$

where $\delta(\mathbf{Q})$ and $\tilde{\delta}(\mathbf{Q})$ are two positive terms defined as the solutions of a system of 2 equations (see Eq. (33)). The functions $\mathbf{G}$ and i depend on $(\delta(\mathbf{Q}), \tilde{\delta}(\mathbf{Q}))$, K, $\mathbf{A}$, $\mathbf{C}$, $\tilde{\mathbf{C}}$, and on the noise variance σ^2. They are given in closed form.

The derivation of $\bar{I}(\mathbf{Q})$ is based on the observation that the eigenvalue distribution of random matrix $\mathbf{HQH}^H$ becomes close to a deterministic distribution as $t \to +\infty$. This in particular implies that if $(\lambda_i)_{1\le i\le r}$ represent the eigenvalues of $\mathbf{HQH}^H$, then :

$$\frac{1}{r}\log\det\left[\mathbf{I}_r + \frac{1}{\sigma^2}\mathbf{HQH}^H\right] = \frac{1}{r}\sum_{i=1}^{r}\log\left(1 + \frac{\lambda_i}{\sigma^2}\right)$$

has the same behaviour as a deterministic term, which turns out to be equal to $\frac{\bar{I}(\mathbf{Q})}{r}$. Taking the mathematical expectation w.r.t. the distribution of the channel, and multiplying by r gives $I(\mathbf{Q}) \simeq \bar{I}(\mathbf{Q})$.

The error term $I(\mathbf{Q}) - \bar{I}(\mathbf{Q})$ is shown to be of order $O(\frac{1}{t})$. 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 $O(\frac{1}{t^2})$. 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 $\overline{\mathbf{Q}}_*$. We also show that $I(\overline{\mathbf{Q}}_*) - I(\mathbf{Q}_*) = O(1/t)$ where we recall that $\mathbf{Q}_*$ is the capacity achieving covariance matrix. Otherwise stated, the computation of $\overline{\mathbf{Q}}_*$ (see below) allows one to (asymptotically) achieve the capacity $I(\mathbf{Q}_*)$.

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

$$\max_{\mathbf{Q}\in\mathcal{C}_1} \log\det\left(\mathbf{I} + \mathbf{G}(\delta_*, \tilde{\delta}_*)\mathbf{Q}\right) \ ,$$

where $\delta_* = \delta(\overline{\mathbf{Q}}_*)$, $\tilde{\delta}_* = \tilde{\delta}(\overline{\mathbf{Q}}_*)$ and

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

This result provides insights on the structure of the approximating capacity achieving covariance matrix, but cannot be used to evaluate $\overline{\mathbf{Q}}_*$ since the parameters δ_* and $\tilde{\delta}_*$ depend on the optimum matrix $\overline{\mathbf{Q}}_*$. 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 $(\delta, \tilde{\delta})$.

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 behaviour 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 evaluate in closed form an accurate approximation $\bar{I}(\mathbf{Q})$ of $I(\mathbf{Q})$. The corresponding result is partly based on the results of [17] devoted to the study of the asymptotic behaviour of the eigenvalue distribution of matrix $\mathbf{\Sigma}\mathbf{\Sigma}^H$ where $\mathbf{\Sigma}$ is given by

$$\mathbf{\Sigma} = \mathbf{B} + \mathbf{Y} \ , \tag{8}$$

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 variance is 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 (8) in the complex circular Gaussian case with a separable variance profile.

A. Study of the EMI of the equivalent model (8).

We first introduce the resolvent and the Stieltjes transform associated with $\mathbf{\Sigma\Sigma}^H$ (Section III-A.1) ; we then introduce auxiliary quantities (Section III-A.2) and their main properties ; we finally introduce the approximation of the EMI in this case (Section III-A.3).

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 :

$$\mathbf{S}(\sigma^2) = \left[\mathbf{\Sigma\Sigma}^H + \sigma^2 \mathbf{I}_r\right]^{-1}, \qquad \tilde{\mathbf{S}}(\sigma^2) = \left[\mathbf{\Sigma}^H\mathbf{\Sigma} + \sigma^2 \mathbf{I}_t\right]^{-1} \ . \tag{9}$$

These resolvents satisfy the obvious, but useful property :

$$\mathbf{S}(\sigma^2) \leq \frac{\mathbf{I}_r}{\sigma^2}, \qquad \tilde{\mathbf{S}}(\sigma^2) \leq \frac{\mathbf{I}_t}{\sigma^2} \ . \tag{10}$$

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 $(\lambda_i)_{1 \leq i \leq r}$ 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 eigenvalue distribution of $\mathbf{\Sigma\Sigma}^H$ defined as the probability distribution :

$$\nu = \frac{1}{r}\sum_{i=1}^{r} \delta_{\lambda_i}$$

where δ_x represents the Dirac distribution at point x. The Stieltjes transform $s(\sigma^2)$ is important as the characterization of the asymptotic behaviour 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 [29], Girko [13], Bai and Silverstein [1], etc.

We finally recall that a positive $p \times p$ matrix-valued measure $\boldsymbol{\mu}$ is a function defined on the Borel subsets of $\mathbb{R}$ onto the set of all complex-valued $p \times p$ matrices satisfying :

(i) For each Borel set B, $\boldsymbol{\mu}(B)$ is a Hermitian nonnegative definite $p \times p$ matrix with complex entries ;

(ii) $\boldsymbol{\mu}(0) = \mathbf{0}$;

(iii) For each countable family $(B_n)_{n\in\mathrm{N}}$ of disjoint Borel subsets of R,

$$\boldsymbol{\mu}(\cup_n B_n) = \sum_n \boldsymbol{\mu}(B_n) \ .$$

Note that for any nonnegative Hermitian $p \times p$ matrix $\mathbf{M}$, then $\mathrm{Tr}(\mathbf{M}\boldsymbol{\mu})$ is a (scalar) positive measure. The matrix-valued measure $\boldsymbol{\mu}$ is said to be finite if $\mathrm{Tr}(\boldsymbol{\mu}(\mathbb{R})) < +\infty$.

2) The auxiliary quantities $\beta, \tilde{\beta}$, $\mathbf{T}$ and $\tilde{\mathbf{T}}$: We gather in this section many results of [17] 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: Recall that $\boldsymbol{\Sigma} = \mathbf{B} + \mathbf{Y}$ and assume that $\mathbf{Y} = \frac{1}{\sqrt{t}} \mathbf{D}^{\frac{1}{2}} \mathbf{X} \tilde{\mathbf{D}}^{\frac{1}{2}}$, where $\mathbf{D}$ 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 :

$$\left\{ \begin{array}{l} \beta = \frac{1}{t}\mathrm{Tr}\left[\mathbf{D}\left(\sigma^2(\mathbf{I}_r + \mathbf{D}\tilde{\beta}) + \mathbf{B}(\mathbf{I}_t + \tilde{\mathbf{D}}\beta)^{-1}\mathbf{B}^H\right)^{-1}\right] \\ \tilde{\beta} = \frac{1}{t}\mathrm{Tr}\left[\tilde{\mathbf{D}}\left(\sigma^2(\mathbf{I}_t + \tilde{\mathbf{D}}\beta) + \mathbf{B}^H(\mathbf{I}_r + \mathbf{D}\tilde{\beta})^{-1}\mathbf{B}\right)^{-1}\right] \end{array} \right. . \tag{11}$$

Then, the system (11) admits a unique couple of 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 :

$$\left\{ \begin{array}{lcl} \mathbf{T}(\sigma^2) & = & \left[\sigma^2(\mathbf{I} + \tilde{\beta}(\sigma^2)\mathbf{D}) + \mathbf{B}(\mathbf{I} + \beta(\sigma^2)\tilde{\mathbf{D}})^{-1}\mathbf{B}^H\right]^{-1} \\ \tilde{\mathbf{T}}(\sigma^2) & = & \left[\sigma^2(\mathbf{I} + \beta(\sigma^2)\tilde{\mathbf{D}}) + \mathbf{B}^H(\mathbf{I} + \tilde{\beta}(\sigma^2)\mathbf{D})^{-1}\mathbf{B}\right]^{-1} \end{array} \right. . \tag{12}$$

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

$$\mathbf{T}(\sigma^2) \le \frac{\mathbf{I}_r}{\sigma^2}, \qquad \tilde{\mathbf{T}}(\sigma^2) \le \frac{\mathbf{I}_t}{\sigma^2} \ . \tag{13}$$

(ii) (*Representation of the auxiliary quantities*) There exist two uniquely defined positive matrix-valued measures $\boldsymbol{\mu}$ and $\tilde{\boldsymbol{\mu}}$ such that $\boldsymbol{\mu}(\mathbb{R}^+) = \mathbf{I}_r$, $\tilde{\boldsymbol{\mu}}(\mathbb{R}^+) = \mathbf{I}_t$ and

$$\mathbf{T}(\sigma^2) = \int_{\mathrm{R}^+} \frac{\boldsymbol{\mu}(d\lambda)}{\lambda + \sigma^2}, \qquad \tilde{\mathbf{T}}(\sigma^2) = \int_{\mathrm{R}^+} \frac{\tilde{\boldsymbol{\mu}}(d\lambda)}{\lambda + \sigma^2} \ . \tag{14}$$

The solutions $\beta(\sigma^2)$ and $\tilde{\beta}(\sigma^2)$ of system (11) are given by :

$$\beta(\sigma^2) = \frac{1}{t}\mathrm{Tr}\mathbf{D}\mathbf{T}(\sigma^2) \ , \qquad \tilde{\beta}(\sigma^2) = \frac{1}{t}\mathrm{Tr}\tilde{\mathbf{D}}\tilde{\mathbf{T}}(\sigma^2) \ , \tag{15}$$

and can thus be written as

$$\beta(\sigma^2) = \int_{\mathrm{R}^+} \frac{\mu_b(d\lambda)}{\lambda + \sigma^2} \ , \qquad \tilde{\beta}(\sigma^2) = \int_{\mathrm{R}^+} \frac{\tilde{\mu}_b(d\lambda)}{\lambda + \sigma^2} \tag{16}$$

where μ_b and $\tilde{\mu}_b$ are nonnegative scalar measures defined by

$$\mu_b(d\lambda) = \frac{1}{t}\mathrm{Tr}(\mathbf{D}\boldsymbol{\mu}(d\lambda)) \quad \text{and} \quad \tilde{\mu}_b(d\lambda) = \frac{1}{t}\mathrm{Tr}(\tilde{\mathbf{D}}\tilde{\boldsymbol{\mu}}(d\lambda)).$$

(iii) (*Asymptotic approximation*) 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 :

$$\begin{cases} \lim_{t\to+\infty} \frac{1}{r}\mathrm{Tr}\left[(\mathbf{S}(\sigma^2) - \mathbf{T}(\sigma^2))\mathbf{M}\right] & = & 0 \\ \lim_{t\to+\infty} \frac{1}{t}\mathrm{Tr}\left[(\tilde{\mathbf{S}}(\sigma^2) - \tilde{\mathbf{T}}(\sigma^2))\tilde{\mathbf{M}}\right] & = & 0 \end{cases} . \tag{17}$$

Denote by μ and $\tilde{\mu}$ the (scalar) probability measures $\mu = \frac{1}{r}\mathrm{Tr}\boldsymbol{\mu}$ and $\tilde{\mu} = \frac{1}{t}\mathrm{Tr}\tilde{\boldsymbol{\mu}}$, by (λ_i) (resp. $(\tilde{\lambda}_j)$) the eigenvalues of $\mathbf{\Sigma}\mathbf{\Sigma}^H$ (resp. of $\mathbf{\Sigma}^H\mathbf{\Sigma}$). The following limits hold true almost surely :

$$\begin{cases} \lim_{t\to+\infty} \frac{1}{r}\sum_{i=1}^r \phi(\lambda_i) - \int_0^{+\infty} \phi(\lambda)\ \mu(d\lambda) & = & 0 \\ \lim_{t\to+\infty} \frac{1}{t}\sum_{j=1}^t \tilde{\phi}(\lambda_j) - \int_0^{+\infty} \tilde{\phi}(\lambda)\ \tilde{\mu}(d\lambda) & = & 0 \end{cases} , \tag{18}$$

for continuous bounded functions ϕ and $\tilde{\phi}$ defined on $\mathbb{R}^+$.

The proof of (i) is provided in Appendix I (note that in [17], the existence and uniqueness of solutions to the system (11) is proved in a certain class of analytic functions depending on σ^2 but this does not imply the existence of a unique solution $(\beta, \tilde{\beta})$ when σ^2 is fixed). The rest of the statements of Theorem 1 have been established in [17], and their proof is omitted here.

Remark 3: As shown in [17], the results in Theorem 1 do not require any Gaussian assumption for $\mathbf{\Sigma}$. Remark that (17) implies in some sense that the entries of $\mathbf{S}(\sigma^2)$ and $\tilde{\mathbf{S}}(\sigma^2)$ have the same behaviour as the entries of the deterministic matrices $\mathbf{T}(\sigma^2)$ and $\tilde{\mathbf{T}}(\sigma^2)$ (which can be evaluated by solving the system (11)). In particular, using (17) 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 measure $\mu = \frac{1}{r}\mathrm{Tr}\boldsymbol{\mu}$. The convergence statement (18) which states that the eigenvalue distribution of $\mathbf{\Sigma}\mathbf{\Sigma}^H$ (resp. $\mathbf{\Sigma}^H\mathbf{\Sigma}$) has the same behavior as μ (resp. $\tilde{\mu}$) directly follows from this observation.

3) The asymptotic approximation of the EMI: Denote by $J(\sigma^2) = \mathbb{E} \log\det\left(\mathbf{I}_r + \sigma^{-2}\mathbf{\Sigma}\mathbf{\Sigma}^H\right)$ the EMI associated with matrix $\mathbf{\Sigma}$. First notice that

$$\log\det\left(\mathbf{I} + \frac{\mathbf{\Sigma}\mathbf{\Sigma}^H}{\sigma^2}\right) = \sum_{i=1}^{r} \log\left(1 + \frac{\lambda_i}{\sigma^2}\right) \ ,$$

where the λ_i's stand for the eigenvalues of $\mathbf{\Sigma}\mathbf{\Sigma}^H$. Applying (18) to function $\phi(\lambda) = \log(\lambda+\sigma^2)$ (plus some extra work since ϕ is not bounded), we obtain :

$$\lim_{t\to+\infty}\left(\frac{1}{r}\log\det\left(\mathbf{I} + \frac{\mathbf{\Sigma}\mathbf{\Sigma}^H}{\sigma^2}\right) - \int_0^{+\infty} \log(\lambda+\sigma^2)\ d\mu(\lambda)\right) = 0 \ . \tag{19}$$

Using the well known relation :

$$\begin{aligned} \frac{1}{r}\log\det\left(\mathbf{I} + \frac{\mathbf{\Sigma}\mathbf{\Sigma}^H}{\sigma^2}\right) &= \int_{\sigma^2}^{+\infty}\left(\frac{1}{\omega} - \frac{1}{r}\mathrm{Tr}(\mathbf{\Sigma}\mathbf{\Sigma}^H + \omega\mathbf{I})^{-1}\right) d\omega \\ &= \int_{\sigma^2}^{+\infty}\left(\frac{1}{\omega} - \frac{1}{r}\mathrm{Tr}\,\mathbf{S}(\omega)\right) d\omega \ , \end{aligned} \tag{20}$$

together with the fact that $\boldsymbol{S}(\omega) \approx \boldsymbol{T}(\omega)$ (which follows from Theorem 1), it is proved in [17] that :

$$\lim_{t\to+\infty}\left[\frac{1}{r}\log\det\left(\mathbf{I} + \frac{\mathbf{\Sigma}\mathbf{\Sigma}^H}{\sigma^2}\right) - \int_{\sigma^2}^{+\infty}\left(\frac{1}{\omega} - \frac{1}{r}\mathrm{Tr}\mathbf{T}(\omega)\right) d\omega\right] = 0 \tag{21}$$

almost surely. Define by $\bar{J}(\sigma^2)$ the quantity :

$$\bar{J}(\sigma^2) = r\int_{\sigma^2}^{+\infty}\left(\frac{1}{\omega} - \frac{1}{r}\mathrm{Tr}\mathbf{T}(\omega)\right) d\omega \ . \tag{22}$$

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

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

or equivalently as

$$\begin{aligned} \bar{J}(\sigma^2) = \log\det&\left[\mathbf{I}_t + \beta(\sigma^2)\tilde{\mathbf{D}} + \frac{1}{\sigma^2}\mathbf{B}^H(\mathbf{I}_r + \tilde{\beta}(\sigma^2)\mathbf{D})^{-1}\mathbf{B}\right] \\ &+ \log\det\left[\mathbf{I}_r + \tilde{\beta}(\sigma^2)\mathbf{D}\right] - \sigma^2 t\beta(\sigma^2)\tilde{\beta}(\sigma^2) \ . \end{aligned} \tag{24}$$

Taking the expectation with respect to the channel $\mathbf{\Sigma}$ in (21), the EMI $J(\sigma^2) = \mathbb{E}\log\det\left(\mathbf{I}_r + \sigma^{-2}\mathbf{\Sigma}\mathbf{\Sigma}^H\right)$ can be approximated by $\bar{J}(\sigma^2)$:

$$J(\sigma^2) = \bar{J}(\sigma^2) + o(t) \tag{25}$$

as $t \to +\infty$. This result is fully proved in [17] and is of potential interest since the numerical evaluation of $\bar{J}(\sigma^2)$ only requires to solve the 2×2 system (11) while the calculation of $J(\sigma^2)$ either rely on Monte-Carlo simulations or on the implementation of rather complicated explicit formulas (see for instance [22]).

In order to evaluate the precision of the asymptotic approximation $\bar{J}$, we shall improve (25) and get the speed $J(\sigma^2) = \bar{J}(\sigma^2) + O(t^{-1})$ in the next theorem. This result completes those in [17] and on the contrary of the rest of Theorem 1 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: Recall that $\mathbf{\Sigma} = \mathbf{B} + \mathbf{Y}$ and assume that $\mathbf{Y} = \frac{1}{\sqrt{t}}\mathbf{D}^{\frac{1}{2}}\mathbf{X}\,\tilde{\mathbf{D}}^{\frac{1}{2}}$, where $\mathbf{D} = \mathrm{diag}(d_i)$ and $\tilde{\mathbf{D}} = \mathrm{diag}(\tilde{d}_j)$ are $r \times r$ and $t \times t$ diagonal matrices and where $\mathbf{X}$ is a matrix whose entries are i.i.d. complex circular Gaussian variables $\mathcal{C}N(0,1)$. 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 :

$$\mathrm{Var}\left(\frac{1}{r}\mathrm{Tr}\left[\mathbf{S}(\sigma^2)\mathbf{M}\right]\right) = O\left(\frac{1}{t^2}\right) \quad \text{and} \quad \mathrm{Var}\left(\frac{1}{t}\mathrm{Tr}\left[\tilde{\mathbf{S}}(\sigma^2)\tilde{\mathbf{M}}\right]\right) = O\left(\frac{1}{t^2}\right) \tag{26}$$

where $\mathrm{Var}(.)$ stands for the variance. Moreover,

$$\begin{array}{rcl} \frac{1}{r}\mathrm{Tr}\left[(\mathbb{E}(\mathbf{S}(\sigma^2)) - \mathbf{T}(\sigma^2))\mathbf{M}\right] & = & O\left(\frac{1}{t^2}\right) \\ \frac{1}{t}\mathrm{Tr}\left[(\mathbb{E}(\tilde{\mathbf{S}}(\sigma^2)) - \tilde{\mathbf{T}}(\sigma^2))\tilde{\mathbf{M}}\right] & = & O\left(\frac{1}{t^2}\right) \end{array} \tag{27}$$

and

$$J(\sigma^2) = \bar{J}(\sigma^2) + O\left(\frac{1}{t}\right) \ . \tag{28}$$

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

Remark 4: The proof of Theorem 2 takes full advantage of the Gaussian structure of matrix $\mathbf{\Sigma}$ and relies on two simple ingredients :

(i) An integration by parts formula that provides an expression for the expectation of certain functionals of Gaussian vectors, already well-known and widely used in Random Matrix Theory [25], [32].

(ii) An inequality known as Poincaré-Nash inequality that bounds the variance of functionals of Gaussian vectors. Although well known, its application to random matrices is fairly recent ([6], [33], see also [16]).

Remark 5: Equations (26) also hold in the non Gaussian case and can be established by using the so-called REFORM (Resolvent FORmula Martingale) method introduced by Girko ([13]).

Equations (27) and (28) 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 :

(i) The work of [2] in which a weaker result ($o(1)$ instead of $O(t^{-1})$) is proved in the simpler case where $\mathbf{B} = \mathbf{0}$;

(ii) The predictions of the replica method in [30] (resp. [31]) in the case where $\mathbf{B} = \mathbf{0}$ (resp. in the case where $\tilde{\mathbf{D}} = \mathbf{I}_t$ and $\mathbf{D} = \mathbf{I}_r$) ;

Remark 6 (Standard deviation and bias): Eq. (26) implies that the standard deviation of $\frac{1}{r}\mathrm{Tr}\left[(\mathbf{S}(\sigma^2) - \mathbf{T}(\sigma^2))\mathbf{M}\right]$ and $\frac{1}{t}\mathrm{Tr}\left[(\tilde{\mathbf{S}}(\sigma^2) - \tilde{\mathbf{T}}(\sigma^2))\tilde{\mathbf{M}}\right]$ are of order $O(t^{-1})$ terms. However, their mathematical expectations (which correspond to the bias) converge much faster towards 0 as (27) shows (the order is $O(t^{-2})$).

Remark 7: By adapting the techniques developed in the course of the proof of Theorem 2, one may establish that $\mathbf{u}^H \mathbb{E}\mathbf{S}(\sigma^2)\mathbf{v} - \mathbf{u}^H \mathbf{T}(\sigma^2)\mathbf{v} = O\left(\frac{1}{t}\right)$, where $\mathbf{u}$ and $\mathbf{v}$ are uniformly bounded r-dimensional vectors.

Remark 8: Both $J(\sigma^2)$ and $\bar{J}(\sigma^2)$ increase linearly with t. Equation (28) 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], [30], [**?**].

B. Introduction of the virtual channel $\mathbf{HQ}^{\frac{1}{2}}$

The purpose of this section is to establish a link between the simplified model (8) : $\mathbf{\Sigma} = \mathbf{B} + \mathbf{Y}$ 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 $\mathcal{C}N(0,1)$ entries, $\mathbf{D}$ and $\tilde{\mathbf{D}}$ being diagonal matrices, and the Rician model (2) under investigation : $\mathbf{H} = \sqrt{\frac{K}{K+1}}\mathbf{A} + \frac{1}{\sqrt{K+1}}\mathbf{V}$ where $\mathbf{V} = \frac{1}{\sqrt{t}}\mathbf{C}^{\frac{1}{2}}\mathbf{W}\tilde{\mathbf{C}}^{\frac{1}{2}}$. As we shall see, the key point is the unitary invariance of the EMI of Gaussian channels together with a well-chosen eingenvalue/eigenvector decomposition.

We introduce the virtual channel $\mathbf{HQ}^{\frac{1}{2}}$ which can be written as :

$$\mathbf{HQ}^{\frac{1}{2}} = \sqrt{\frac{K}{K+1}}\mathbf{AQ}^{\frac{1}{2}} + \frac{1}{\sqrt{K+1}}\mathbf{C}^{\frac{1}{2}}\frac{\mathbf{W}}{\sqrt{t}}\mathbf{\Theta}(\mathbf{Q}^{\frac{1}{2}}\tilde{\mathbf{C}}\mathbf{Q}^{\frac{1}{2}})^{\frac{1}{2}} \ , \tag{29}$$

where $\mathbf{\Theta}$ is the deterministic unitary $t \times t$ matrix defined by

$$\mathbf{\Theta} = \tilde{\mathbf{C}}^{\frac{1}{2}}\mathbf{Q}^{\frac{1}{2}}(\mathbf{Q}^{\frac{1}{2}}\tilde{\mathbf{C}}\mathbf{Q}^{\frac{1}{2}})^{-\frac{1}{2}} \ . \tag{30}$$

The virtual channel $\mathbf{HQ}^{\frac{1}{2}}$ has thus a structure similar to $\mathbf{H}$, where $(\mathbf{A}, \mathbf{C}, \tilde{\mathbf{C}}, \mathbf{W})$ are respectively replaced with $(\mathbf{AQ}^{\frac{1}{2}}, \mathbf{C}, \mathbf{Q}^{\frac{1}{2}}\tilde{\mathbf{C}}\mathbf{Q}^{\frac{1}{2}}, \mathbf{W\Theta})$.

Consider now the eigenvalue/eigenvector decompositions of matrices $\frac{\mathbf{C}}{\sqrt{K+1}}$ and $\frac{\mathbf{Q}^{\frac{1}{2}}\tilde{\mathbf{C}}\mathbf{Q}^{\frac{1}{2}}}{\sqrt{K+1}}$:

$$\frac{\mathbf{C}}{\sqrt{K+1}} = \mathbf{UDU}^H \qquad \text{and} \qquad \frac{\mathbf{Q}^{\frac{1}{2}}\tilde{\mathbf{C}}\mathbf{Q}^{\frac{1}{2}}}{\sqrt{K+1}} = \tilde{\mathbf{U}}\tilde{\mathbf{D}}\tilde{\mathbf{U}}^H \ . \tag{31}$$

Matrices $\mathbf{U}$ and $\tilde{\mathbf{U}}$ are the eigenvectors matrices while $\mathbf{D}$ and $\tilde{\mathbf{D}}$ are the eigenvalues diagonal matrices. 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

$$\mathbf{B} = \sqrt{\frac{K}{K+1}}\mathbf{U}^H\mathbf{AQ}^{\frac{1}{2}}\tilde{\mathbf{U}} \qquad \text{and} \qquad \mathbf{Y} = \frac{1}{\sqrt{t}}\mathbf{D}^{\frac{1}{2}}\mathbf{X}\tilde{\mathbf{D}}^{\frac{1}{2}} \quad \text{with} \quad \mathbf{X} = \mathbf{U}^H\mathbf{W\Theta}\tilde{\mathbf{U}} \ . \tag{32}$$

As matrix $\mathbf{W}$ has i.i.d. $\mathcal{C}N(0,1)$ 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. We sum up the previous discussion in the following proposition.

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

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

are equal provided that channel $\mathbf{H}$ is given by :

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

with $\mathbf{V} = \frac{1}{\sqrt{t}}\mathbf{C}^{\frac{1}{2}}\mathbf{W}\tilde{\mathbf{C}}^{\frac{1}{2}}$; channel $\mathbf{\Sigma}$ by $\mathbf{\Sigma} = \mathbf{B} + \mathbf{Y}$ with $\mathbf{Y} = \frac{1}{\sqrt{t}}\mathbf{D}^{\frac{1}{2}}\mathbf{X}\tilde{\mathbf{D}}^{\frac{1}{2}}$ and that (30), (31) and (32) hold true.

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 the system of equations

$$\begin{cases} \delta &= f(\delta, \tilde{\delta}, \mathbf{Q}) \\ \tilde{\delta} &= \tilde{f}(\delta, \tilde{\delta}, \mathbf{Q}) \end{cases} , \tag{33}$$

where $f(\delta, \tilde{\delta}, \mathbf{Q})$ and $\tilde{f}(\delta, \tilde{\delta}, \mathbf{Q})$ are given by :

$$f(\delta, \tilde{\delta}, \mathbf{Q}) = \frac{1}{t}\operatorname{Tr}\Bigg\{\mathbf{C}\Big[\sigma^2\left(\mathbf{I}_r + \frac{\tilde{\delta}}{K+1}\mathbf{C}\right) + \frac{K}{K+1}\mathbf{AQ}^{\frac{1}{2}}\left(\mathbf{I}_t + \frac{\delta}{K+1}\mathbf{Q}^{\frac{1}{2}}\tilde{\mathbf{C}}\mathbf{Q}^{\frac{1}{2}}\right)^{-1}\mathbf{Q}^{\frac{1}{2}}\mathbf{A}^H\Big]^{-1}\Bigg\} , \tag{34}$$

$$\tilde{f}(\delta,\tilde{\delta},\mathbf{Q}) = \frac{1}{t}\,\mathrm{Tr}\Big\{\mathbf{Q}^{\frac{1}{2}}\tilde{\mathbf{C}}\mathbf{Q}^{\frac{1}{2}}\Big[\sigma^2\left(\mathbf{I}_t + \frac{\delta}{K+1}\,\mathbf{Q}^{\frac{1}{2}}\tilde{\mathbf{C}}\mathbf{Q}^{\frac{1}{2}}\,\right) + \frac{K}{K+1}\mathbf{Q}^{\frac{1}{2}}\mathbf{A}^H\left(\mathbf{I}_r + \frac{\tilde{\delta}}{K+1}\,\mathbf{C}\right)^{-1}\mathbf{A}\mathbf{Q}^{\frac{1}{2}}\Big]^{-1}\Big\}\ . \quad (35)$$

Then the system of equations (33) has a unique strictly positive solution $(\delta(\mathbf{Q}),\tilde{\delta}(\mathbf{Q}))$. Furthermore, assume that $\sup_t\|\mathbf{Q}\| < +\infty$, $\sup_t\|\mathbf{A}\| < +\infty$, $\sup_t\|\mathbf{C}\| < +\infty$, and $\sup_t\|\tilde{\mathbf{C}}\| < +\infty$. Assume also that $\inf_t \lambda_{\min}(\tilde{\mathbf{C}}) > 0$ where $\lambda_{\min}(\tilde{\mathbf{C}})$ represents the smallest eigenvalue of $\tilde{\mathbf{C}}$. Then, as $t \to +\infty$,

$$I(\mathbf{Q}) = \bar{I}(\mathbf{Q}) + O\left(\frac{1}{t}\right) \quad (36)$$

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

$$\bar{I}(\mathbf{Q}) = \log\,\det\left(\mathbf{I}_t + \frac{\delta(\mathbf{Q})}{K+1}\,\mathbf{Q}^{\frac{1}{2}}\tilde{\mathbf{C}}\mathbf{Q}^{\frac{1}{2}} + \frac{1}{\sigma^2}\,\frac{K}{K+1}\,\mathbf{Q}^{\frac{1}{2}}\mathbf{A}^H\left(\mathbf{I}_r + \frac{\tilde{\delta}(\mathbf{Q})}{K+1}\,\mathbf{C}\right)^{-1}\mathbf{A}\mathbf{Q}^{\frac{1}{2}}\right) + \log\det\left(\mathbf{I}_r + \frac{\tilde{\delta}(\mathbf{Q})}{K+1}\mathbf{C}\right) - \frac{t\sigma^2}{K+1}\delta(\mathbf{Q})\,\tilde{\delta}(\mathbf{Q})\ , \quad (37)$$

or equivalently by

$$\bar{I}(\mathbf{Q}) = \log\det\left(\mathbf{I}_r + \frac{\tilde{\delta}(\mathbf{Q})}{K+1}\mathbf{C} + \frac{1}{\sigma^2}\,\frac{K}{K+1}\,\mathbf{A}\mathbf{Q}^{\frac{1}{2}}\left(\mathbf{I}_t + \frac{\delta(\mathbf{Q})}{K+1}\,\mathbf{Q}^{\frac{1}{2}}\tilde{\mathbf{C}}\mathbf{Q}^{\frac{1}{2}}\right)^{-1}\mathbf{Q}^{\frac{1}{2}}\mathbf{A}^H\right) + \log\det\left(\mathbf{I}_t + \frac{\delta(\mathbf{Q})}{K+1}\mathbf{Q}^{1/2}\tilde{\mathbf{C}}\mathbf{Q}^{1/2}\right) - \frac{t\sigma^2}{K+1}\delta(\mathbf{Q})\,\tilde{\delta}(\mathbf{Q}). \quad (38)$$

Proof: We rely on the virtual channel introduced in Section III-B 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 $\inf_t \frac{1}{t}\mathrm{Tr}\mathbf{D} = \inf_t \frac{1}{t}\mathrm{Tr}\mathbf{C} = 1$ due to the model specifications and $\inf_t \frac{1}{t}\mathrm{Tr}\mathbf{Q}^{\frac{1}{2}}\tilde{\mathbf{C}}\mathbf{Q}^{\frac{1}{2}} \geq \inf_t \lambda_{\min}(\tilde{\mathbf{C}})\frac{1}{t}\mathrm{Tr}\mathbf{Q} > 0$ as $\frac{1}{t}\mathrm{Tr}\mathbf{Q} = 1$. 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 to matrix $\mathbf{\Sigma}$, and use the same notations as in the statement of Theorem 1. Using the unitary invariance of the trace of a matrix, it is straightforward to check that :

$$\begin{aligned}
\frac{f(\delta,\tilde{\delta},\mathbf{Q})}{\sqrt{K+1}} &= \frac{1}{t}\mathrm{Tr}\left[\mathbf{D}\left(\sigma^2\left(\mathbf{I}+\mathbf{D}\frac{\tilde{\delta}}{\sqrt{K+1}}\right)+\mathbf{B}\left(\mathbf{I}+\tilde{\mathbf{D}}\frac{\delta}{\sqrt{K+1}}\right)^{-1}\mathbf{B}^H\right)^{-1}\right]\ , \\
\frac{\tilde{f}(\delta,\tilde{\delta},\mathbf{Q})}{\sqrt{K+1}} &= \frac{1}{t}\mathrm{Tr}\left[\tilde{\mathbf{D}}\left(\sigma^2\left(\mathbf{I}+\tilde{\mathbf{D}}\frac{\delta}{\sqrt{K+1}}\right)+\mathbf{B}^H\left(\mathbf{I}+\mathbf{D}\frac{\tilde{\delta}}{\sqrt{K+1}}\right)^{-1}\mathbf{B}\right)^{-1}\right]\ .
\end{aligned}$$

Therefore, $(\delta, \tilde{\delta})$ is solution of (33) if and only if $(\frac{\delta}{\sqrt{K+1}}, \frac{\tilde{\delta}}{\sqrt{K+1}})$ is solution of (11). As the system (11) admits a unique solution, say $(\beta, \tilde{\beta})$, the solution $(\delta, \tilde{\delta})$ to (33) exists, is unique and is related to $(\beta, \tilde{\beta})$ by the relations :

$$\beta = \frac{\delta}{\sqrt{K+1}}, \quad \tilde{\beta} = \frac{\tilde{\delta}}{\sqrt{K+1}} \ . \tag{39}$$

In order to justify (37) and (38), 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 (39) imply that $\bar{I}(\mathbf{Q})$ defined by (37) and (38) coincide with the approximation $\bar{J}$ given by (23) and (24). This proves (36) as well. ■

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

$$\begin{cases} \mathbf{T}_K(\sigma^2) &= \left[\sigma^2(\mathbf{I} + \frac{\tilde{\delta}}{K+1}\mathbf{C}) + \frac{K}{K+1}\mathbf{A}\mathbf{Q}^{\frac{1}{2}}(\mathbf{I} + \frac{\delta}{K+1}\mathbf{Q}^{\frac{1}{2}}\tilde{\mathbf{C}}\mathbf{Q}^{\frac{1}{2}})^{-1}\mathbf{Q}^{\frac{1}{2}}\mathbf{A}^H\right]^{-1} \\ \tilde{\mathbf{T}}_K(\sigma^2) &= \left[\sigma^2(\mathbf{I} + \frac{\delta}{K+1}\mathbf{Q}^{\frac{1}{2}}\tilde{\mathbf{C}}\mathbf{Q}^{\frac{1}{2}}) + \frac{K}{K+1}\mathbf{Q}^{\frac{1}{2}}\mathbf{A}^H(\mathbf{I} + \frac{\tilde{\delta}}{K+1}\mathbf{C})^{-1}\mathbf{A}\mathbf{Q}^{\frac{1}{2}}\right]^{-1} \end{cases} . \tag{40}$$

They are related to matrices $\mathbf{T}$ and $\tilde{\mathbf{T}}$ defined by (12) by the relations :

$$\begin{cases} \mathbf{T}_K(\sigma^2) &= \mathbf{U}\mathbf{T}(\sigma^2)\mathbf{U}^H \\ \tilde{\mathbf{T}}_K(\sigma^2) &= \tilde{\mathbf{U}}\tilde{\mathbf{T}}(\sigma^2)\tilde{\mathbf{U}}^H \end{cases} , \tag{41}$$

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

As $\frac{1}{r}\mathrm{Tr}\mathbf{T}_K = \frac{1}{r}\mathrm{Tr}\mathbf{T}$ and $\frac{1}{t}\mathrm{Tr}\tilde{\mathbf{T}}_K = \frac{1}{t}\mathrm{Tr}\tilde{\mathbf{T}}$, the quantities $\frac{1}{r}\mathrm{Tr}\mathbf{T}_K$ and $\frac{1}{t}\mathrm{Tr}\tilde{\mathbf{T}}_K$ are the Stieltjes transforms of probability measures μ and $\tilde{\mu}$ introduced in Theorem 1. As matrices $\mathbf{HQH}^H$ and $\mathbf{\Sigma\Sigma}^H$ (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, (18) implies that the eigenvalue distribution of $\mathbf{HQH}^H$ (resp. $\mathbf{Q}^{\frac{1}{2}}\mathbf{H}^H\mathbf{H}\mathbf{Q}^{\frac{1}{2}}$) behaves like μ (resp. $\tilde{\mu}$).

We finally mention that $\delta(\sigma^2)$ and $\tilde{\delta}(\sigma^2)$ are given by

$$\delta(\sigma^2) = \frac{1}{t}\mathrm{Tr}\mathbf{C}\mathbf{T}_K(\sigma^2) \qquad \text{and} \qquad \tilde{\delta}(\sigma^2) = \frac{1}{t}\mathrm{Tr}\mathbf{Q}^{\frac{1}{2}}\tilde{\mathbf{C}}\mathbf{Q}^{1/2}\tilde{\mathbf{T}}_K(\sigma^2) \ , \tag{42}$$

and that the following representations hold true :

$$\delta(\sigma^2) = \int_{\mathrm{R}^+} \frac{\mu_d(d\,\lambda)}{\lambda + \sigma^2} \qquad \text{and} \qquad \tilde{\delta}(\sigma^2) = \int_{\mathrm{R}^+} \frac{\tilde{\mu}_d(d\,\lambda)}{\lambda + \sigma^2} \ , \tag{43}$$

where μ_d and $\tilde{\mu}_d$ are positive measures on $\mathbb{R}^+$ satisfying $\mu_d(\mathbb{R}^+) = \frac{1}{t}\mathrm{Tr}\mathbf{C}$ and $\tilde{\mu}_d(\mathbb{R}^+) = \frac{1}{t}\mathrm{Tr}\mathbf{Q}^{1/2}\tilde{\mathbf{C}}\mathbf{Q}^{1/2}$.

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 the following :

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 $I(\mathbf{Q}) = \mathbb{E}\log\det\left(\mathbf{I} + \frac{\mathbf{HQH}^H}{\sigma^2}\right)$. The strict concavity is more demanding and its proof is mainly postponed to Appendix III.

Recall that we denote by $\mathcal{C}_1$ the set of nonnegative Hermitian $t \times t$ matrices whose normalized trace is equal to one (i.e. $t^{-1}\mathrm{Tr}\,\mathbf{Q} = 1$). 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

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

is strictly concave.

1) Concavity of the EMI: We first recall that $I(\mathbf{Q}) = \mathbb{E}\log\det\left(\mathbf{I} + \frac{\mathbf{HQH}^H}{\sigma^2}\right)$ is concave on $\mathcal{C}_1$, and provide a proof for the sake of completeness. Denote by $\mathbf{Q} = \lambda\mathbf{Q}_1 + (1-\lambda)\mathbf{Q}_2$ and let $\phi(\lambda) = I(\lambda\mathbf{Q}_1 + (1-\lambda)\mathbf{Q}_2)$. Following Proposition 2, it is sufficient to prove that ϕ is concave. As $\log\det\left(\mathbf{I} + \frac{\mathbf{HQH}^H}{\sigma^2}\right) = \log\det\left(\mathbf{I} + \frac{\mathbf{H}^H\mathbf{HQ}}{\sigma^2}\right)$, we have :

$$\begin{aligned}
\phi(\lambda) &= \mathbb{E}\log\det\left(\mathbf{I} + \frac{\mathbf{HQH}^H}{\sigma^2}\right) \ , \\
\phi'(\lambda) &= \mathbb{E}\,\mathrm{Tr}\left(\mathbf{I} + \frac{\mathbf{H}^H\mathbf{HQ}}{\sigma^2}\right)^{-1}\frac{\mathbf{H}^H\mathbf{H}}{\sigma^2}(\mathbf{Q_1} - \mathbf{Q_2}) \ , \\
\phi''(\lambda) &= -\mathbb{E}\,\mathrm{Tr}\left[\left(\mathbf{I} + \frac{\mathbf{H}^H\mathbf{HQ}}{\sigma^2}\right)^{-1}\frac{\mathbf{H}^H\mathbf{H}}{\sigma^2}(\mathbf{Q_1} - \mathbf{Q_2})\left(\mathbf{I} + \frac{\mathbf{H}^H\mathbf{HQ}}{\sigma^2}\right)^{-1}\frac{\mathbf{H}^H\mathbf{H}}{\sigma^2}(\mathbf{Q_1} - \mathbf{Q_2})\right] \ .
\end{aligned}$$

In order to conclude that $\phi''(\lambda) \leq 0$, we notice that $\left(\mathbf{I} + \frac{\mathbf{H}^H\mathbf{HQ}}{\sigma^2}\right)^{-1}\frac{\mathbf{H}^H\mathbf{H}}{\sigma^2}$ coincides with

$$\mathbf{H}^H\left(\mathbf{I} + \frac{\mathbf{HQH}^H}{\sigma^2}\right)^{-1}\frac{\mathbf{H}}{\sigma^2}$$

(use the well-known inequality $(\mathbf{I}+\mathbf{UV})^{-1}\mathbf{U} = \mathbf{U}(\mathbf{I}+\mathbf{VU})^{-1}$ for $\mathbf{U} = \mathbf{H}^H$ and $\mathbf{V} = \frac{\mathbf{HQ}}{\sigma^2}$). We denote by $\mathbf{M}$ the non negative matrix

$$\mathbf{M} = \mathbf{H}^H\left(\mathbf{I} + \frac{\mathbf{HQH}^H}{\sigma^2}\right)^{-1}\frac{\mathbf{H}}{\sigma^2}$$

and remark that

$$\phi''(\lambda) = -\mathbb{E}\mathrm{Tr}\left[\mathbf{M}(\mathbf{Q}_1 - \mathbf{Q}_2)\mathbf{M}(\mathbf{Q}_1 - \mathbf{Q}_2)\right] \tag{44}$$

or equivalently that

$$\phi''(\lambda) = -\mathbb{E}\mathrm{Tr}\left[\mathbf{M}^{1/2}(\mathbf{Q}_1 - \mathbf{Q}_2)\mathbf{M}^{1/2}\mathbf{M}^{1/2}(\mathbf{Q}_1 - \mathbf{Q}_2)\mathbf{M}^{1/2}\right] \ .$$

As matrix $\mathbf{M}^{1/2}(\mathbf{Q}_1 - \mathbf{Q}_2)\mathbf{M}^{1/2}$ is Hermitian, this of course implies that $\phi''(\lambda) \leq 0$. The concavity of ϕ and of I are established.

2) Using an auxiliary channel to establish concavity of $\bar{I}(\mathbf{Q})$: Denote by $\otimes$ the Kronecker product of matrices. We introduce the following matrices :

$$\boldsymbol{\Delta} = \mathbf{I}_m \otimes \mathbf{C}, \quad \tilde{\boldsymbol{\Delta}} = \mathbf{I}_m \otimes \tilde{\mathbf{C}}, \quad \check{\mathbf{A}} = \mathbf{I}_m \otimes \mathbf{A}, \quad \check{\mathbf{Q}} = \mathbf{I}_m \otimes \mathbf{Q} \ .$$

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

$$\check{\mathbf{V}} = \frac{1}{\sqrt{mt}} \boldsymbol{\Delta}^{\frac{1}{2}} \check{\mathbf{W}} \tilde{\boldsymbol{\Delta}}^{\frac{1}{2}} \quad \text{and} \quad \check{\mathbf{H}} = \sqrt{\frac{K}{K+1}} \check{\mathbf{A}} + \frac{1}{\sqrt{K+1}} \check{\mathbf{V}} \ ,$$

where $\check{\mathbf{W}}$ is a $rm \times tm$ matrix whose entries are i.i.d $\mathcal{C}N(0,1)$-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 the system (34)-(35) and formula (37), where one will substitute the quantities related to channel $\mathbf{H}$ by those related to channel $\check{\mathbf{H}}$, i.e. :

$$t \leftrightarrow mt, \quad r \leftrightarrow mr, \quad \mathbf{A} \leftrightarrow \check{\mathbf{A}}, \quad \mathbf{Q} \leftrightarrow \check{\mathbf{Q}}, \quad \mathbf{C} \leftrightarrow \boldsymbol{\Delta}, \quad \tilde{\mathbf{C}} \leftrightarrow \tilde{\boldsymbol{\Delta}} \ .$$

Due to the block-diagonal nature of matrices $\check{\mathbf{A}}$, $\check{\mathbf{Q}}$, $\boldsymbol{\Delta}$ and $\tilde{\boldsymbol{\Delta}}$, the system 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 (36) 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.

3) 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

$$\forall m \geq 1, \quad \forall \lambda \in [0,1], \qquad \phi_m''(\lambda) \leq \kappa < 0 \ . \tag{45}$$

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

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

Proof of Lemma 1 is postponed to Appendix III.

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

$$\limsup_{m\to\infty} \phi_m''(\lambda) \le \kappa < 0 \ . \tag{46}$$

(46) is proved in the Appendix III.

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 $\overline{\mathbf{Q}}_*$, i.e. :

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

As we shall see in Section V, matrix $\overline{\mathbf{Q}}_*$ can be obtained by a rather simple algorithm. Provided that $\sup_t \|\overline{\mathbf{Q}}_*\|$ is bounded, Eq. (36) in Theorem 3 yields $I(\overline{\mathbf{Q}}_*) - \bar{I}(\overline{\mathbf{Q}}_*) \to 0$ as $t \to \infty$. It remains to check that $I(\mathbf{Q}_*) - I(\overline{\mathbf{Q}}_*)$ goes asymptotically to zero to be able to approximate the capacity. This is the purpose of the next proposition.

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

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

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

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

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

$$\underbrace{\left(I(\mathbf{Q}_*) - I(\overline{\mathbf{Q}}_*)\right)}_{\ge 0} + \underbrace{\left(\bar{I}(\overline{\mathbf{Q}}_*) - \bar{I}(\mathbf{Q}_*)\right)}_{\ge 0} = \underbrace{\left(I(\mathbf{Q}_*) - \bar{I}(\mathbf{Q}_*)\right)}_{\substack{= O(t^{-1}) \\ \text{by (ii) and Th. 3 Eq. (36)}}} + \underbrace{\left(\bar{I}(\overline{\mathbf{Q}}_*) - I(\overline{\mathbf{Q}}_*)\right)}_{\substack{= O(t^{-1}) \\ \text{by (i) and Th. 3 Eq. (36)}}} \tag{47}$$

where the two terms of the lefthand side are nonnegative due to the fact that $\mathbf{Q}_*$ and $\overline{\mathbf{Q}}_*$ are the maximizers of I and $\bar{I}$ respectively. As a direct consequence of (47), we have $I(\mathbf{Q}_*) - I(\overline{\mathbf{Q}}_*) = O(t^{-1})$ and the proof is completed. ■

V. Optimization of the input covariance matrix

In the previous section, we have proved that matrix $\overline{\mathbf{Q}}_*$ 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 $\overline{\mathbf{Q}}_*$'s properties. We first introduce a few notations. Let $V(\kappa, \tilde{\kappa}, \mathbf{Q})$ be the function defined by :

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

or equivalently by

$$V(\kappa, \tilde{\kappa}, \mathbf{Q}) = \log\det\left(\mathbf{I}_r + \frac{\tilde{\kappa}}{K+1}\mathbf{C} + \frac{K}{\sigma^2(K+1)}\mathbf{A}\mathbf{Q}^{\frac{1}{2}}\left(\mathbf{I}_t + \frac{\kappa}{K+1}\mathbf{Q}^{\frac{1}{2}}\tilde{\mathbf{C}}\mathbf{Q}^{\frac{1}{2}}\right)^{-1}\mathbf{Q}^{\frac{1}{2}}\mathbf{A}^H\right) + \log\det\left(\mathbf{I}_t + \frac{\kappa}{K+1}\mathbf{Q}^{1/2}\tilde{\mathbf{C}}\mathbf{Q}^{1/2}\right) - \frac{t\sigma^2\kappa\tilde{\kappa}}{K+1} . \quad (49)$$

Note that if $(\delta(\mathbf{Q}), \tilde{\delta}(\mathbf{Q}))$ is the solution of system (33), then :

$$\bar{I}(\mathbf{Q}) = V(\delta(\mathbf{Q}), \tilde{\delta}(\mathbf{Q}), \mathbf{Q}) .$$

Denote by $(\delta_*, \tilde{\delta}_*)$ the solution $(\delta(\overline{\mathbf{Q}}_*), \tilde{\delta}(\overline{\mathbf{Q}}_*))$ of (33) associated with $\overline{\mathbf{Q}}_*$. The aim of the section is to prove that $\overline{\mathbf{Q}}_*$ is the solution of the following standard waterfilling problem :

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

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

$$\mathbf{G}(\kappa, \tilde{\kappa}) = \frac{\kappa}{K+1}\tilde{\mathbf{C}} + \frac{K}{\sigma^2(K+1)}\mathbf{A}^H\left(\mathbf{I}_r + \frac{\tilde{\kappa}}{K+1}\mathbf{C}\right)^{-1}\mathbf{A} . \quad (50)$$

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

$$V(\kappa, \tilde{\kappa}, \mathbf{Q}) = \log\det\left(\mathbf{I} + \mathbf{Q}\mathbf{G}(\kappa, \tilde{\kappa})\right) + \log\det\left(\mathbf{I}_r + \frac{\tilde{\kappa}}{K+1}\mathbf{C}\right) - \frac{t\sigma^2\kappa\tilde{\kappa}}{K+1} , \quad (51)$$

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 (x taking its values in some finite-dimensional space) and by $\langle \nabla F(x), y \rangle$ the value of this derivative at point y. Sometimes, a function is not differentiable but still admits *directional derivatives* : The directional derivative of a function F at x in direction y is

$$F'(x; y) = \lim_{t \downarrow 0} \frac{F(x + ty) - F(x)}{t}$$

when the limit exists. Of course, if F is differentiable at x, then $F'(x; y) = \langle \nabla F(x), y \rangle$. The following proposition captures the main features needed in the sequel.

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

(i) The directional derivative $F'(\mathbf{Q}; \mathbf{P} - \mathbf{Q})$ exists in $(-\infty, \infty]$ for all $\mathbf{Q}, \mathbf{P}$ in $\mathcal{C}_1$.

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

$$\forall \mathbf{Q} \in \mathcal{C}_1, \quad F'(\overline{\mathbf{Q}}_*; \mathbf{Q} - \overline{\mathbf{Q}}_*) \leq 0 \ . \tag{52}$$

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

$$\forall \mathbf{Q} \in \mathcal{C}_1, \quad F'(\overline{\mathbf{Q}}_*; \mathbf{Q} - \overline{\mathbf{Q}}_*) \leq 0. \tag{53}$$

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

If F is differentiable then both conditions (52) and (53) write :

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

Although this is standard material (see for instance [4, Chapter 2]), we provide some elements of proof for the reader's convenience.

Proof: Let us first prove item (i). As $\mathbf{Q} + t(\mathbf{P} - \mathbf{Q}) = (1-t)\mathbf{Q} + t\mathbf{P} \in \mathcal{C}_1$, $\Delta(t) \stackrel{\triangle}{=} t^{-1}\left(F(\mathbf{Q} + t(\mathbf{P} - \mathbf{Q})) - F(\mathbf{Q})\right)$ is well-defined. Let $0 \leq s \leq t \leq 1$ and consider

$$\begin{aligned}
\Delta(t) - \Delta(s) &= \frac{1}{s}\left\{ \frac{s}{t} F\left((1-t)\mathbf{Q} + t\mathbf{P}\right) + \frac{t-s}{t} F(\mathbf{Q}) - F\left((1-s)\mathbf{Q} + s\mathbf{P}\right) \right\} , \\
&\stackrel{(a)}{\leq} \frac{1}{s}\left\{ F\left(s\frac{(1-t)\mathbf{Q} + t\mathbf{P}}{t} + \frac{t-s}{t}\mathbf{Q} \right) - F\left((1-s)\mathbf{Q} + s\mathbf{P}\right) \right\} , \\
&= \frac{1}{s}\left\{ F\left((1-s)\mathbf{Q} + s\mathbf{P}\right) - F\left((1-s)\mathbf{Q} + s\mathbf{P}\right) \right\} \quad = \quad 0 \, ,
\end{aligned}$$

where (a) follows from the concavity of F. This shows that $\Delta(t)$ increases as $t \downarrow 0$, and in particular always admits a limit in $(-\infty, \infty]$.

Item (ii) readily follows from the fact that $F((1-t)\overline{\mathbf{Q}}_* + t\mathbf{P}) \leq F(\overline{\mathbf{Q}}_*)$ due to the mere definition of $\overline{\mathbf{Q}}_*$. This implies that $\Delta(t) \leq 0$ which in turn yields (52).

We now prove (iii). The concavity of F yields :

$$\Delta(t) = \frac{F(\overline{\mathbf{Q}}_* + t(\mathbf{P} - \overline{\mathbf{Q}}_*)) - F(\overline{\mathbf{Q}}_*)}{t} \geq F(\mathbf{P}) - F(\overline{\mathbf{Q}}_*).$$

As $\lim_{t \downarrow 0} \Delta(t) \leq 0$ by (53), one gets : $\forall \mathbf{P} \in \mathcal{C}_1$, $F(\mathbf{P}) - F(\overline{\mathbf{Q}}_*) \leq 0$. Otherwise stated, F attains its maximum at $\overline{\mathbf{Q}}_*$ and Proposition 4 is proved. ■

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

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

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

(ii) Recall that $\overline{\mathbf{Q}}_*$ is the argmax of $\bar{I}$ over $\mathcal{C}_1$, i.e. $\forall \mathbf{Q} \in \mathcal{C}_1,\ \bar{I}(\mathbf{Q}) \leq \bar{I}(\overline{\mathbf{Q}}_*)$. Let $\mathbf{Q} \in \mathcal{C}_1$. The following property :

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

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

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

Proof: (i) is established in the Appendix. Let us establish (ii). Recall that $\bar{I}(\mathbf{Q})$ is strictly concave by Theorem 4 (and therefore its maximum is attained at 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 $\overline{\mathbf{Q}}_*$. Item (ii) follows then from Proposition 4.

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

$$\langle \nabla \bar{I}(\overline{\mathbf{Q}}_*), \mathbf{Q} - \overline{\mathbf{Q}}_* \rangle = \langle \nabla_{\mathbf{Q}} V \left(\delta_*, \tilde{\delta}_*, \overline{\mathbf{Q}}_* \right), \mathbf{Q} - \overline{\mathbf{Q}}_* \rangle \ , \tag{54}$$

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 (54) holds true. Then item (ii) implies that $\langle \nabla_{\overline{\mathbf{Q}}_*} V \left(\delta_*, \tilde{\delta}_*, \overline{\mathbf{Q}}_* \right), \mathbf{Q} - \overline{\mathbf{Q}}_* \rangle \leq 0$ for every $\mathbf{Q} \in \mathcal{C}_1$. As $\mathbf{Q} \mapsto V(\delta_*, \tilde{\delta}_*, \mathbf{Q})$ is strictly concave on $\mathcal{C}_1$, $\overline{\mathbf{Q}}_*$ is the argmax of $V(\delta_*, \tilde{\delta}_*, \cdot)$ by Proposition 4 and we are done.

It remains to prove (54). Consider $\mathbf{Q}$ and $\mathbf{P}$ in $\mathcal{C}_1$, and use the identity

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

We now compute the partial derivatives of V and obtain :

$$\left\{ \begin{array}{rcl} \frac{\partial V}{\partial \kappa} & = & -\frac{t\sigma^2}{K+1}\left(\tilde{\kappa} - \tilde{f}(\kappa, \tilde{\kappa}, \mathbf{Q})\right) \\ \frac{\partial V}{\partial \tilde{\kappa}} & = & -\frac{t\sigma^2}{K+1}\left(\kappa - f(\kappa, \tilde{\kappa}, \mathbf{Q})\right) \end{array} \right. , \tag{55}$$

where f and $\tilde{f}$ are defined by (34) and (35). The first relation follows from (48) and the second relation from (49). As $(\delta(\mathbf{Q}), \tilde{\delta}(\mathbf{Q}))$ is the solution of system (33), equations (55) imply that :

$$\frac{\partial V}{\partial \kappa}(\delta(\mathbf{Q}), \tilde{\delta}(\mathbf{Q}), \mathbf{Q}) = \frac{\partial V}{\partial \tilde{\kappa}}(\delta(\mathbf{Q}), \tilde{\delta}(\mathbf{Q}), \mathbf{Q}) = 0 \ . \tag{56}$$

Letting $\mathbf{P} = \overline{\mathbf{Q}}_*$ and taking into account (56) yields :

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

and (iii) is established. ■

Remark 9: The quantities δ_* and $\tilde{\delta}_*$ depend on matrix $\overline{\mathbf{Q}}_*$. Therefore, Proposition 5 does not provide by itself any optimization algorithm. However, it gives valuable insights on the structure of $\overline{\mathbf{Q}}_*$. Consider first the case $\mathbf{C} = \mathbf{I}$ and $\tilde{\mathbf{C}} = \mathbf{I}$. Then, $\mathbf{G}(\delta_*, \tilde{\delta}_*)$ is a linear combination of $\mathbf{I}$ and matrix $\mathbf{A}^H\mathbf{A}$. The eigenvectors of $\overline{\mathbf{Q}}_*$ thus coincide with the right singular vectors of matrix $\mathbf{A}$, a result consistent with the work [19] devoted to the maximization of the EMI $I(\mathbf{Q})$. If $\mathbf{C} = \mathbf{I}$ and $\tilde{\mathbf{C}} \neq \mathbf{I}$, $\mathbf{G}(\delta_*, \tilde{\delta}_*)$ can be interpreted as a linear combination of matrices $\tilde{\mathbf{C}}$ and $\mathbf{A}^H\mathbf{A}$. Therefore, if the transmit antennas are correlated, the eigenvectors of the optimum matrix $\overline{\mathbf{Q}}_*$ coincide with the eigenvectors of some weighted sum of $\tilde{\mathbf{C}}$ 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 $\overline{\mathbf{Q}}_*$ is however less intuitive because matrix $\mathbf{A}^H\mathbf{A}$ has to be replaced with $\mathbf{A}^H(\mathbf{I} + \tilde{\delta}_*\mathbf{C})^{-1}\mathbf{A}$.

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 w.r.t. $\mathbf{Q}$ of function $V(\kappa, \tilde{\kappa}, \mathbf{Q})$ defined by (51) 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 equations (33) 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 $\delta, \tilde{\delta}$ separately by the following iterative scheme :

- Initialization : $\mathbf{Q}_0 = \mathbf{I}$, $(\delta_1, \tilde{\delta}_1)$ are defined as the unique solutions of system (33) in which $\mathbf{Q} = \mathbf{Q}_0 = \mathbf{I}$. Then, define $\mathbf{Q}_1$ are the maximum of function $\mathbf{Q} \to V(\delta_1, \tilde{\delta}_1, \mathbf{Q})$ on $\mathcal{C}_1$, which is obtained through a standard waterfilling procedure.

- Iteration k : assume $\mathbf{Q}_{k-1}$, $(\delta_{k-1}, \tilde{\delta}_{k-1})$ available. Then, $(\delta_k, \tilde{\delta}_k)$ is defined as the unique solution of (33) in which $\mathbf{Q} = \mathbf{Q}_{k-1}$. Then, define $\mathbf{Q}_k$ are the maximum of function $\mathbf{Q} \rightarrow V(\delta_k, \tilde{\delta}_k, \mathbf{Q})$ on $\mathcal{C}_1$.

One can notice that this algorithm is the generalization of the procedure used by [44] 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 $\overline{\mathbf{Q}}_*$ of $\bar{I}$.

Proposition 6: Assume that the two sequences $(\delta_k)_{k \geq 0}$ and $(\tilde{\delta}_k)_{k \geq 0}$ verify

$$\lim_{k \rightarrow +\infty} \delta_k - \delta_{k-1} \rightarrow 0, \lim_{k \rightarrow +\infty} \tilde{\delta}_k - \tilde{\delta}_{k-1} \rightarrow 0 \tag{57}$$

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

Remark 10: 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 $\mathbf{P}_* = \overline{\mathbf{Q}}_*$. In fact, functions $\mathbf{Q} \mapsto \delta(\mathbf{Q})$ and $\mathbf{Q} \mapsto \tilde{\delta}(\mathbf{Q})$ are continuous by Proposition 5. As $\delta_k = \delta(\mathbf{Q}_{k-1})$ and $\tilde{\delta}_k = \tilde{\delta}(\mathbf{Q}_{k-1})$, the convergence of $(\mathbf{Q}_k)$ thus implies the convergence of (δ_k) and $(\tilde{\delta}_k)$, and (57) is fulfilled. Proposition 6 immediately yields $\mathbf{P}_* = \overline{\mathbf{Q}}_*$. Although we have not been able to prove the convergence of the algorithm, the above result is encouraging, and tends to indicate the algorithm is reliable. In particular, all the numerical experiments we have conducted indicates that the algorithm converges towards a certain matrix which must coincide by Proposition 6 with $\overline{\mathbf{Q}}_*$.

VI. NUMERICAL EXPERIMENTS.

A. When is the number of antennas large enough to reach the asymptotic regime ?

All our analysis is based on the approximation of the ergodic mutual information. This approximation consists in assuming the channel matrix to be large. Here we provide typical simulation results showing that the asymptotic regime is reached for relatively small number of antennas. For the simulations provided here we assume :

- $\mathbf{Q} = \mathbf{I}_t$.
- The chosen line-of-sight (LOS) component $\mathbf{A}$ is based on equation (4). The angle of arrivals are chosen randomly according to a uniform distribution.
- Antenna correlation is assumed to decrease exponentially with the inter-antenna distance i.e. $\tilde{\mathbf{C}}_{ij} \sim \rho_T^{|i-j|}$, $\mathbf{C}_{ij} \sim \rho_R^{|i-j|}$ with $0 \leq \rho_T \leq 1$ and $0 \leq \rho_R \leq 1$.
- K is equal to 1.

Figure 1 represents the EMI $I(\mathbf{Q})$ evaluated by Monte Carlo simulations and its approximation $\bar{I}(\mathbf{Q})$ as well as their relative difference (in percentage). Here, the correlation coefficients are equal to $(\rho_T, \rho_R) = (0.8, 0.3)$ and three different pairs of numbers of antenna are considered : $(t, r) \in \{(2, 2), (4, 4), (8, 8)\}$. Figure 1 shows that the approximation is reliable even for $r = t = 2$ in a wide range of SNR.

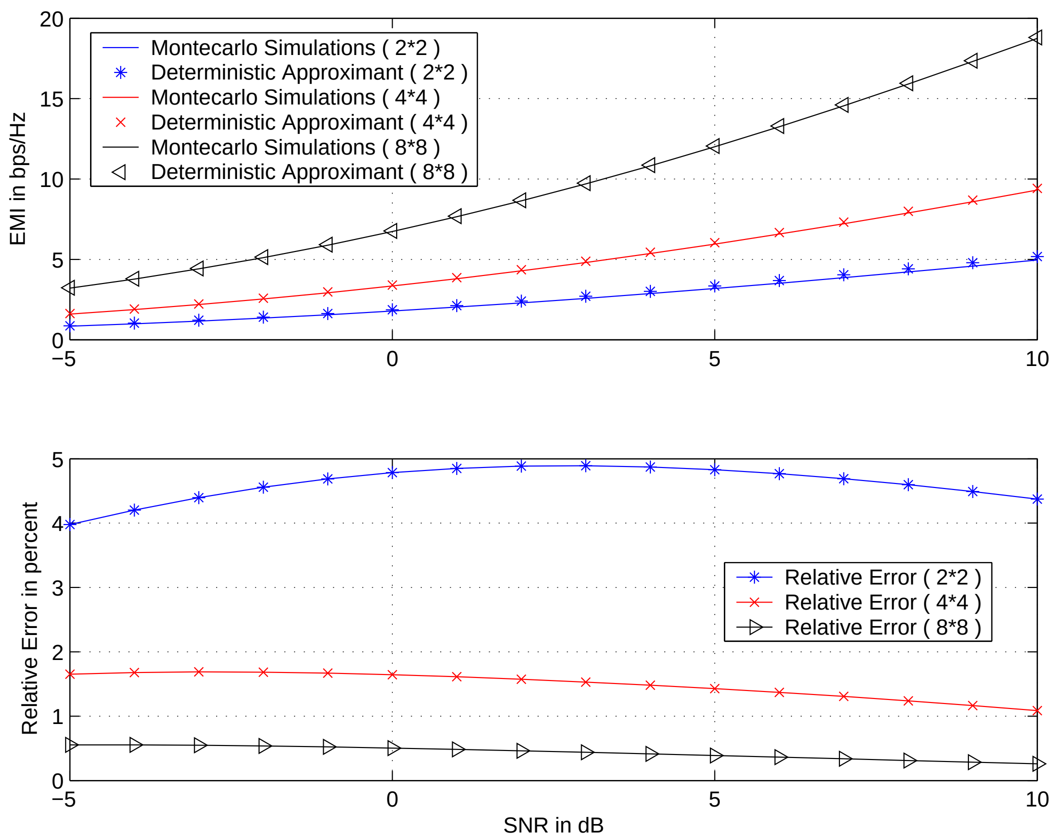

Fig. 1. The large system approximation is accurate for correlated Rician MIMO channels. The relative difference between the EMI approximation and that obtained by Monte-Carlo simulations is less than 5 % for a 2×2 system and less than 1 % for a 8×8 system.

B. Comparison with the Vu-Paulraj method.

In this paragraph, we compare our algorithm with the method presented in [42] based on the maximization of $I(\mathbf{Q})$. We 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. 3, 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 [42]. 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

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

Fig. 2. Average time per iteration in seconds

its first and second derivatives is equal to 30.000, and the maximum number of iterations of the algorithm in [42] is fixed to 10. The dashed line corresponds to the results provided by our algorithm : Each point represents $I(\overline{\mathbf{Q}}_*)$ at the corresponding SNR, where $\overline{\mathbf{Q}}_*$ is the argmax of $\bar{I}$; the average mutual information at point $\overline{\mathbf{Q}}_*$ is evaluted by Monte-Carlo simulation (30.000 trials are used). The number of iterations is also limited to 10. Figure 3 shows that our asymptotic approach provides the same results than the Vu-Paulraj's algorithm. However, our algorithm is computationally much more efficient as the above table shows. The table gives the average executation time (in sec.) of one iteration for both algorithms for $r = t = 2, r = t = 4, r = t = 8$.

In fig. 4, we again compare Vu-Paulraj's algorithm and our proposal. Matrix $\mathbf{A}$ is generated according to (4), the angles being chosen at random. 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 converge at the 10th iteration.

VII. Conclusions

In this paper, an explicit approximation for the ergodic mutual information for Rician MIMO channels with transmit and receive antenna correlation is provided. This approximation is based on the asymptotic Random Matrix Theory. The accuracy of the approximation has been studied both analytically and numerically. It has been shown to be very accurate even for small MIMO systems : The relative error is less than 5% for a 2×2 MIMO channel and less $1\ \%$ for an 8×8 MIMO channel.

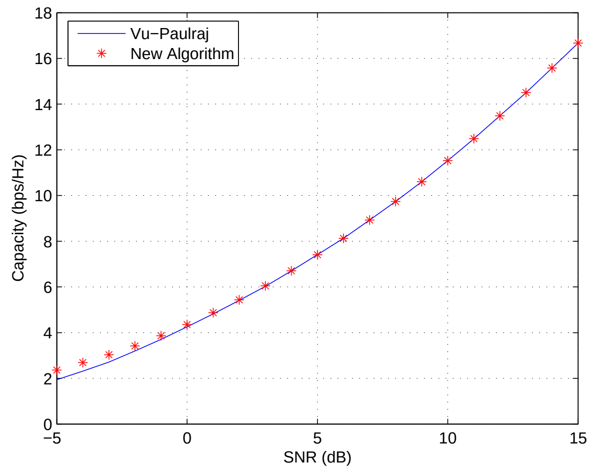

Fig. 3. Comparison with the Vu-Paulraj algorithm I

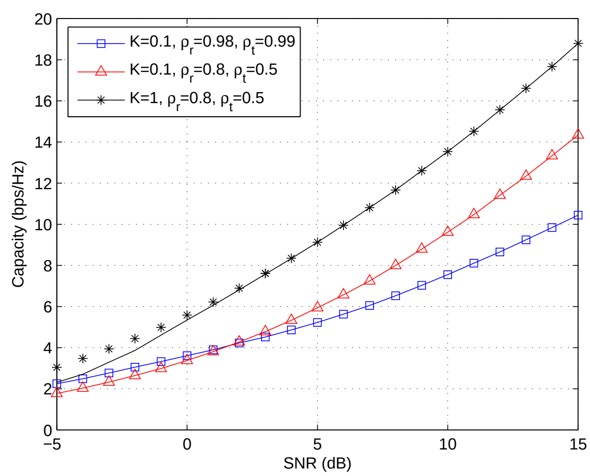

Fig. 4. Comparison with the Vu-Paulraj algorithm II

The derived expression for the EMI has been exploited to derive an efficient optimization algorithm providing the optimum covariance matrix.

Appendix I

Proof of the existence and uniqueness of the system (11).

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

$$\begin{aligned} g(\kappa,\tilde{\kappa}) &= \frac{1}{\kappa}\frac{1}{t}\mathrm{Tr}\left[\mathbf{D}\left(\sigma^2(\mathbf{I}_r+\mathbf{D}\tilde{\kappa})+\mathbf{B}(\mathbf{I}_t+\tilde{\mathbf{D}}\kappa)^{-1}\mathbf{B}^H\right)^{-1}\right] \\ \tilde{g}(\kappa,\tilde{\kappa}) &= \frac{1}{\tilde{\kappa}}\frac{1}{t}\mathrm{Tr}\left[\tilde{\mathbf{D}}\left(\sigma^2(\mathbf{I}_t+\tilde{\mathbf{D}}\kappa)+\mathbf{B}^H(\mathbf{I}_r+\mathbf{D}\tilde{\kappa})^{-1}\mathbf{B}\right)^{-1}\right] \end{aligned} . \quad (58)$$

For each $\tilde{\kappa} > 0$ fixed, function $\kappa \to g(\kappa, \tilde{\kappa})$ is clearly strictly decreasing, converges toward $+\infty$ 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,

– (ii) Function $\tilde{\kappa} \to \tilde{\kappa} h(\tilde{\kappa})$ is strictly increasing.

In fact, consider $\tilde{\kappa}_2 > \tilde{\kappa}_1$. 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

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

and use the identity

$$\mathbf{U}_1^{-1} - \mathbf{U}_2^{-1} = \mathbf{U}_1^{-1} \left(\mathbf{U}_2 - \mathbf{U}_1 \right) \mathbf{U}_2^{-1} \ . \tag{59}$$

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

$$g(h(\tilde{\kappa}_1), \tilde{\kappa}_1) - g(h(\tilde{\kappa}_2), \tilde{\kappa}_2) = u(h(\tilde{\kappa}_2) - h(\tilde{\kappa}_1)) + v(\tilde{\kappa}_2 h(\tilde{\kappa}_2) - \tilde{\kappa}_1 h(\tilde{\kappa}_1)) \ ,$$

where u and v are the strictly positive terms defined by

$$u = \frac{1}{t}\mathrm{Tr}\mathbf{D}\mathbf{U}_1^{-1} \left(\sigma^2 \mathbf{I} + \mathbf{B}(\mathbf{I} + h(\tilde{\kappa}_2)\tilde{\mathbf{D}})^{-1}(\mathbf{I} + h(\tilde{\kappa}_1)\tilde{\mathbf{D}})^{-1}\mathbf{B}^H \right) \mathbf{U}_2^{-1}$$

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. This function converges to $+\infty$ if $\tilde{\kappa} \to 0$ and to 0 if $\tilde{\kappa} \to +\infty$. Therefore, the equation

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

has a unique strictly positive solution $\tilde{\beta}$. If $\beta = h(\tilde{\beta})$, it is clear that $g(\beta, \tilde{\beta}) = 1$ and $\tilde{g}(\beta, \tilde{\beta}) = 1$. Therefore, we have shown that $(\beta, \tilde{\beta})$ is the unique solution of (11) satisfying $\beta > 0$ and $\tilde{\beta} > 0$.

APPENDIX II
PROOF OF THEOREM 2

This section is organized as follows. We first recall in subsection II-A some useful mathematical tools. In subsection II-B, we establish (26). In II-C, we prove (27) and (28). Technical details that are needed to establish (27) and (28) are also given in subsections II-D and II-E.

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 $\frac{1}{2}\left(\frac{\partial}{\partial x} - \mathbf{i}\frac{\partial}{\partial y}\right)$ and $\frac{1}{2}\left(\frac{\partial}{\partial x} + \mathbf{i}\frac{\partial}{\partial y}\right)$. 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: 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 :

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

where $\nabla_z \Phi = [\partial\Phi/\partial z_1, \ldots, \partial\Phi/\partial z_M]^T$ and $\nabla_{\bar{z}}\Phi = [\partial\Phi/\partial \bar{z}_1, \ldots, \partial\Phi/\partial \bar{z}_M]^T$.

This inequality is well known (see e.g. [7] and [21]).

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. $\mathcal{C}N(0,1)$ entries and consider the stacked $rt \times 1$ vector $\mathbf{x} = [Y_{11}, \ldots, Y_{rt}]^T$. In this case, Poincaré-Nash inequality writes :

$$\mathrm{Var}\left(\Phi(\mathbf{Y})\right) \le \frac{1}{t}\sum_{i=1}^{r}\sum_{j=1}^{t} d_i \tilde{d}_j \mathbb{E}\left[\left|\frac{\partial \Phi(\mathbf{Y})}{\partial Y_{i,j}}\right|^2 + \left|\frac{\partial \Phi(\mathbf{Y})}{\partial \overline{Y}_{i,j}}\right|^2\right] \ . \tag{60}$$

2) The differentiation formula for functions of Gaussian random vectors: With $\mathbf{x}$ and Φ given as above, we have the following

$$\mathbb{E}\left[x_p \Phi(\mathbf{x})\right] = \sum_{m=1}^{M} [\mathbf{\Xi}]_{pm}\, \mathbb{E}\left[\frac{\partial \Phi(\mathbf{x})}{\partial \bar{x}_m}\right] \ . \tag{61}$$

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 ([14]) and has been used in Random Matrix Theory in [25] and [32].

If $\mathbf{x}$ coincides with the $rt \times 1$ vector $\mathbf{x} = [Y_{11}, \ldots, Y_{rt}]^T$, relation (61) becomes

$$\mathbb{E}\left[Y_{pq}\Phi(\mathbf{Y})\right] = \frac{d_p \tilde{d}_q}{t}\mathbb{E}\left[\frac{\partial \Phi(\mathbf{Y})}{\partial \overline{Y}_{pq}}\right] \ . \tag{62}$$

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

$$\mathbb{E}\left[\overline{Y}_{pq}\Phi(\mathbf{Y})\right] = \frac{d_p\tilde{d}_q}{t}\mathbb{E}\left[\frac{\partial\Phi(\mathbf{Y})}{\partial Y_{pq}}\right] \ . \tag{63}$$

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

$$\begin{cases} \frac{\partial S_{pq}}{\partial Y_{ij}} &= -S_{p,i}\left(\boldsymbol{\xi}_j^H\mathbf{S}\right)_q \\ \frac{\partial S_{pq}}{\partial \overline{Y}_{ij}} &= -S_{i,q}\left(\mathbf{S}\boldsymbol{\xi}\right)_p \end{cases} . \tag{64}$$

B. Proof of (26)

We just prove that the variance of $\frac{1}{t}\mathrm{Tr}(\mathbf{MS})$ is a $O(\frac{1}{t^2})$ 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 (60) 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 (64), coincides with

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

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,

$$\sum_{i=1}^{r}\mathbb{E}\left|\frac{\partial\Phi(\mathbf{Y})}{\partial Y_{ij}}\right|^2$$

is a $O(\frac{1}{t^2})$ 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

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

The conclusion follows from Poincaré-Nash inequality (60).

C. Proof of (27) and (28).

As we shall see, proofs of (27) and (28) are demanding. We first introduce the following notations : Define scalar parameters $\eta(\sigma^2), \alpha(\sigma^2), \tilde{\alpha}(\sigma^2)$ as

$$\begin{array}{rcl} \eta(\sigma^2) & = & \frac{1}{t}\mathrm{Tr}\left(\mathbf{D}\mathbf{S}(\sigma^2)\right) \\ \alpha(\sigma^2) & = & \mathbb{E}\left[\frac{1}{t}\mathrm{Tr}\left(\mathbf{D}\mathbf{S}(\sigma^2)\right)\right] \\ \tilde{\alpha}(\sigma^2) & = & \frac{1}{t}\mathrm{Tr}\left[\tilde{\mathbf{D}}\left(\sigma^2(\mathbf{I}+\alpha\tilde{\mathbf{D}})\right)^{-1}\left(\mathbf{I}-\mathbf{B}^H\,\mathbb{E}(\mathbf{S}(\sigma^2))\,\mathbf{B}(\mathbf{I}+\alpha\tilde{\mathbf{D}})^{-1}\right)\right] \end{array} \tag{65}$$

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

$$\begin{array}{rcl} \mathbf{R}(\sigma^2) & = & \left[\sigma^2\left(\mathbf{I}+\tilde{\alpha}\mathbf{D}\right)+\mathbf{B}\left(\mathbf{I}+\alpha\tilde{\mathbf{D}}\right)^{-1}\mathbf{B}^H\right]^{-1} \\ \tilde{\mathbf{R}}(\sigma^2) & = & \left[\sigma^2\left(\mathbf{I}+\tilde{\alpha}\mathbf{D}\right)+\mathbf{B}\left(\mathbf{I}+\alpha\tilde{\mathbf{D}}\right)^{-1}\mathbf{B}^H\right]^{-1} \end{array} \ . \tag{66}$$

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) first, studying $\frac{1}{r}\mathrm{Tr}\mathbf{M}(\mathbb{E}(\mathbf{S})-\mathbf{R})$ and 2) studying $\frac{1}{r}\mathrm{Tr}\mathbf{M}(\mathbf{R}-\mathbf{T})$.

1) First step: 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 $r\to\infty$, we have :

$$\frac{1}{t}\mathrm{Tr}\left[\mathbf{M}\left(\mathbb{E}(\mathbf{S})-\mathbf{R}\right)\right] = O\left(\frac{1}{t^2}\right) \ . \tag{67}$$

We just sketch the proof of this proposition, and provide the detailed proof in subsection II-D. In order to use the Integration by parts formula (62), remark that

$$\sigma^2\mathbf{S}(\sigma^2)+\mathbf{S}(\sigma^2)\mathbf{\Sigma}\mathbf{\Sigma}^H = \mathbf{I} \ . \tag{68}$$

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

$$\sigma^2\mathbb{E}(S_{pq})+\mathbb{E}\left[(\mathbf{S}\mathbf{\Sigma}\mathbf{\Sigma}^H)_{pq}\right] = \delta(p-q) \ . \tag{69}$$

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

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

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

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

Using (62) with function $\Phi(\mathbf{Y}) = S_{pi}\overline{\Sigma_{qj}}$ and (63) with $\Phi(\mathbf{Y}) = S_{pi}$, and summing over index i yields :

$$\mathbb{E}\left[(\mathbf{S}\boldsymbol{\xi}_j)_p\overline{\Sigma_{q,j}}\right] = \frac{d_q\tilde{d}_j}{t}\mathbb{E}(S_{pq}) - \tilde{d}_j\mathbb{E}\left[\eta(\mathbf{S}\boldsymbol{\xi}_j)_p\overline{\Sigma_{q,j}}\right] - \frac{d_q\tilde{d}_j}{t}\mathbb{E}\left[S_{pq}\boldsymbol{\xi}_j^H\mathbf{S}\mathbf{b}_j\right] + \mathbb{E}\left[(\mathbf{S}\mathbf{b}_j)_p\right]\overline{B_{qj}} \ . \tag{70}$$

Eq. (26) for $\mathbf{M} = \mathbf{D}$ implies that $\mathrm{Var}(\eta) = O(\frac{1}{t^2})$, or equivalently that $\mathbb{E}(\overset{\circ}{\eta}{}^2) = O(\frac{1}{t^2})$. This, in some sense (details are given in subsection II-D), allows to approximate $\mathbb{E}\left[\eta(\mathbf{S}\boldsymbol{\xi}_j)_p\overline{\Sigma_{q,j}}\right]$ by $\mathbb{E}(\eta)\mathbb{E}\left[(\mathbf{S}\boldsymbol{\xi}_j)_p\overline{\Sigma_{q,j}}\right] = \alpha\mathbb{E}\left[(\mathbf{S}\boldsymbol{\xi}_j)_p\overline{\Sigma_{q,j}}\right]$. Therefore, Eq. (70) can be written as

$$\mathbb{E}\left[(\mathbf{S}\boldsymbol{\xi}_j)_p\overline{\Sigma_{q,j}}\right] \simeq \frac{d_q\tilde{d}_j}{t}\mathbb{E}(S_{pq}) - \alpha\tilde{d}_j\mathbb{E}\left[(\mathbf{S}\boldsymbol{\xi}_j)_p\overline{\Sigma_{q,j}}\right] - \frac{d_q\tilde{d}_j}{t}\mathbb{E}\left[S_{pq}\boldsymbol{\xi}_j^H\mathbf{S}\mathbf{b}_j\right] + \mathbb{E}\left[(\mathbf{S}\mathbf{b}_j)_p\right]\overline{B_{qj}} \ .$$

Solving w.r.t. $\mathbb{E}\left[(\mathbf{S}\boldsymbol{\xi}_j)_p\overline{\Sigma_{q,j}}\right]$, we obtain

$$\mathbb{E}\left[(\mathbf{S}\boldsymbol{\xi}_j)_p\overline{\Sigma_{q,j}}\right] \simeq \frac{1}{t}\frac{d_q\tilde{d}_j}{1+\alpha\tilde{d}_j}\mathbb{E}(S_{pq}) + \frac{1}{1+\alpha\tilde{d}_j}\mathbb{E}\left[(\mathbf{S}\mathbf{b}_j)_p\right]\overline{B_{qj}} - \frac{1}{t}\frac{d_q\tilde{d}_j}{1+\alpha\tilde{d}_j}\mathbb{E}\left[S_{pq}\boldsymbol{\xi}_j^H\mathbf{S}\mathbf{b}_j\right] \ .$$

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

$$\begin{aligned}\mathbb{E}\left[(\mathbf{S}\boldsymbol{\Sigma}\boldsymbol{\Sigma}^H)_{pq}\right] \quad &\simeq \quad d_q\,\frac{1}{t}\mathrm{Tr}\left[\tilde{\mathbf{D}}(\mathbf{I}+\alpha\tilde{\mathbf{D}})^{-1}\right]\mathbb{E}(S_{pq}) \\ &+ \mathbb{E}\left[\left(\mathbf{S}\mathbf{B}(\mathbf{I}+\alpha\tilde{\mathbf{D}})^{-1}\mathbf{B}^H\right)_{pq}\right] - d_q\,\mathbb{E}\left[S_{pq}\frac{1}{t}\mathrm{Tr}\left(\mathbf{S}\mathbf{B}\tilde{\mathbf{D}}(\mathbf{I}+\alpha\tilde{\mathbf{D}})^{-1}\mathbf{B}^H\right)\right] \\ &- d_q\,\mathbb{E}\left[S_{pq}\frac{1}{t}\mathrm{Tr}\left(\mathbf{S}\mathbf{B}\tilde{\mathbf{D}}(\mathbf{I}+\alpha\tilde{\mathbf{D}})^{-1}\mathbf{Y}^H\right)\right] \ .\end{aligned} \tag{71}$$

Using similar calculations, it is possible to establish that

$$\begin{aligned}\mathbb{E}\left[S_{pq}\frac{1}{t}\mathrm{Tr}\left(\mathbf{S}\mathbf{B}\tilde{\mathbf{D}}(\mathbf{I}+\alpha\tilde{\mathbf{D}})^{-1}\mathbf{Y}^H\right)\right] \quad &\simeq \quad -\mathbb{E}(S_{pq})\mathbb{E}\left[\frac{1}{t}\mathrm{Tr}\left(\mathbf{S}\mathbf{B}\tilde{\mathbf{D}}(\mathbf{I}+\alpha\tilde{\mathbf{D}})^{-1}\mathbf{Y}^H\right)\right] \\ \mathbb{E}\left[S_{pq}\frac{1}{t}\mathrm{Tr}\left(\mathbf{S}\mathbf{B}\tilde{\mathbf{D}}(\mathbf{I}+\alpha\tilde{\mathbf{D}})^{-1}\mathbf{B}^H\right)\right] \quad &\simeq \quad \mathbb{E}(S_{pq})\mathbb{E}\left[\frac{1}{t}\mathrm{Tr}\left(\mathbf{S}\mathbf{B}\tilde{\mathbf{D}}(\mathbf{I}+\alpha\tilde{\mathbf{D}})^{-1}\mathbf{B}^H\right)\right]\end{aligned}$$

and that

$$\mathbb{E}\left[\frac{1}{t}\mathrm{Tr}\left(\mathbf{S}\mathbf{B}\tilde{\mathbf{D}}(\mathbf{I}+\alpha\tilde{\mathbf{D}})^{-1}\mathbf{Y}^H\right)\right] \simeq -\alpha\,\frac{1}{t}\mathrm{Tr}\left(\mathbb{E}(\mathbf{S})\mathbf{B}\tilde{\mathbf{D}}^2(\mathbf{I}+\alpha\tilde{\mathbf{D}})^{-2}\mathbf{B}^H\right) \ . \tag{72}$$

Therefore, $\mathbb{E}\left[(\mathbf{S}\boldsymbol{\Sigma}\boldsymbol{\Sigma}^H)_{pq}\right]$ can be approximated by

$$\begin{aligned}&d_q\,\frac{1}{t}\mathrm{Tr}\left[\tilde{\mathbf{D}}(\mathbf{I}+\alpha\tilde{\mathbf{D}})^{-1}\right]\mathbb{E}(S_{pq}) + \mathbb{E}\left[\left(\mathbf{S}\mathbf{B}(\mathbf{I}+\alpha\tilde{\mathbf{D}})^{-1}\mathbf{B}^H\right)_{pq}\right] \\ &\qquad - d_q\,\mathbb{E}(S_{pq})\mathbb{E}\left[\frac{1}{t}\mathrm{Tr}\left(\mathbf{S}\mathbf{B}\tilde{\mathbf{D}}(\mathbf{I}+\alpha\tilde{\mathbf{D}})^{-1}\mathbf{B}^H\right)\right] \\ &\qquad + \alpha d_q\,\mathbb{E}(S_{pq})\mathbb{E}\left[\frac{1}{t}\mathrm{Tr}\left(\mathbf{S}\mathbf{B}\tilde{\mathbf{D}}^2(\mathbf{I}+\alpha\tilde{\mathbf{D}})^{-2}\mathbf{B}^H\right)\right] \ .\end{aligned} \tag{73}$$

Plugging the above approximate expression of $\mathbb{E}\left[(\mathbf{S}\boldsymbol{\Sigma}\boldsymbol{\Sigma}^H)_{pq}\right]$ into (69), and solving with respect to $\mathbb{E}(S_{pq})$, we obtain after some algebra that

$$\left(\mathbb{E}\left[\mathbf{S}\left(\sigma^2(\mathbf{I}+\tilde{\alpha}\mathbf{D})+\mathbf{B}(\mathbf{I}+\alpha\tilde{\mathbf{D}})^{-1}\mathbf{B}^H\right)\right]\right)_{pq} \simeq \delta(p-q)$$

or equivalently that

$$\mathbb{E}(\mathbf{S}_{pq}) \simeq \mathbf{R}_{pq} \ .$$

In order to prove Proposition 7, it is of course necessary to evaluate the convergence speed toward 0 of the error terms $\mathbb{E}(\mathbf{S}_{pq}) - \mathbf{R}_{pq}$. Fortunately, Poincaré-Nash inequality allows to study these terms rather easily, and to prove (67). More details are given in subsection II-D.

2) Second step: The second step consists in showing the Proposition :

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

$$\frac{1}{t}\mathrm{Tr}\left[\mathbf{M}\left(\mathbf{R}-\mathbf{T}\right)\right] = O\left(\frac{1}{t^2}\right) \ . \tag{74}$$

We first observe that $\mathbf{R}-\mathbf{T} = \mathbf{R}\left(\mathbf{T}^{-1}-\mathbf{R}^{-1}\right)\mathbf{T}$. Using the expressions of $\mathbf{R}^{-1}$ and $\mathbf{T}^{-1}$, multiplying by $\mathbf{M}$, and taking the trace gives

$$\begin{aligned}\frac{1}{t}\mathrm{Tr}\left[\mathbf{M}\left(\mathbf{R}-\mathbf{T}\right)\right] \ &= \ (\tilde{\beta}-\tilde{\alpha})\,\sigma^2\frac{1}{t}\mathrm{Tr}(\mathbf{MRDT}) + \\ &\quad (\alpha-\beta)\,\frac{1}{t}\mathrm{Tr}\left[\mathbf{MRB}(\mathbf{I}+\beta\tilde{\mathbf{D}})^{-1}\tilde{\mathbf{D}}(\mathbf{I}+\beta\tilde{\mathbf{D}})^{-1}\mathbf{B}^H\mathbf{T}\right] \ . \end{aligned} \tag{75}$$

As the terms $\frac{\sigma^2}{t}\mathrm{Tr}(\mathbf{MRDT})$ and $\frac{1}{t}\mathrm{Tr}\left[\mathbf{MRB}(\mathbf{I}+\beta\tilde{\mathbf{D}})^{-1}\tilde{\mathbf{D}}(\mathbf{I}+\beta\tilde{\mathbf{D}})^{-1}\mathbf{B}^H\mathbf{T}\right]$ are uniformly bounded, it is sufficient to establish that $(\alpha-\beta)$ and $(\tilde{\alpha}-\tilde{\beta})$ are themselves $O(\frac{1}{t^2})$ terms. For this, we first prove the following lemma.

Lemma 2: α and $\tilde{\alpha}$ can be written as

$$\alpha \ = \ \frac{1}{t}\mathrm{Tr}\left(\mathbf{DR}\right)+\epsilon, \qquad \tilde{\alpha} \ = \ \frac{1}{t}\mathrm{Tr}\left(\tilde{\mathbf{D}}\tilde{\mathbf{R}}\right)+\tilde{\epsilon} \ , \tag{76}$$

where ϵ and $\tilde{\epsilon}$ are $O(\frac{1}{t^2})$ terms.

Proof: The first relation of (76) follows immediately from Proposition 7 when matrix $\mathbf{M}$ is equal to $\mathbf{D}$. To establish the second relation, we again use Proposition 7 for a relevant matrix $\mathbf{M}$, and obtain that

$$\tilde{\alpha}(\sigma^2) \ = \ \frac{1}{t}\mathrm{Tr}\left[\tilde{\mathbf{D}}\left(\sigma^2(\mathbf{I}+\alpha\tilde{\mathbf{D}})\right)^{-1}\left(\mathbf{I}-\mathbf{B}^H\,\mathbf{R}\,\mathbf{B}(\mathbf{I}+\alpha\tilde{\mathbf{D}})^{-1}\right)\right]+\tilde{\epsilon}$$

where $\tilde{\epsilon} = O(\frac{1}{t^2})$. We claim that

$$\frac{1}{t}\mathrm{Tr}\left[\tilde{\mathbf{D}}\left(\sigma^2(\mathbf{I}+\alpha\tilde{\mathbf{D}})\right)^{-1}\left(\mathbf{I}-\mathbf{B}^H\,\mathbf{R}\,\mathbf{B}(\mathbf{I}+\alpha\tilde{\mathbf{D}})^{-1}\right)\right] = \frac{1}{t}\mathrm{Tr}\left(\tilde{\mathbf{D}}\tilde{\mathbf{R}}\right). \tag{77}$$

In fact, using the definition of $\mathbf{R}$, we get that

$$\left(\mathbf{B}^H \mathbf{R} \mathbf{B}(\mathbf{I}+\alpha\tilde{\mathbf{D}})^{-1}\right) = \mathbf{B}^H \left[\mathbf{I} + \left(\sigma^2(\mathbf{I}+\tilde{\alpha}\mathbf{D})\right)^{-1} \mathbf{B}(\mathbf{I}+\alpha\tilde{\mathbf{D}})^{-1}\mathbf{B}^H\right]^{-1} \left(\sigma^2(\mathbf{I}+\tilde{\alpha}\mathbf{D})\right)^{-1} \mathbf{B}(\mathbf{I}+\alpha\tilde{\mathbf{D}})^{-1}.$$

In order to simplify the notations, we put

$$\mathbf{G} = \left(\sigma^2(\mathbf{I}+\tilde{\alpha}\mathbf{D})\right)^{-1} \mathbf{B}(\mathbf{I}+\alpha\tilde{\mathbf{D}})^{-1}.$$

Using the identities

$$\begin{array}{rcl} \mathbf{B}^H\left(\mathbf{I}+\mathbf{G}\mathbf{B}^H\right)^{-1}\mathbf{G} & = & \left(\mathbf{I}+\mathbf{B}^H\mathbf{G}\right)^{-1}\mathbf{B}^H\mathbf{G} \\ \mathbf{I} - \left(\mathbf{I}+\mathbf{B}^H\mathbf{G}\right)^{-1}\mathbf{B}^H\mathbf{G} & = & \left(\mathbf{I}+\mathbf{B}^H\mathbf{G}\right)^{-1} \end{array},$$

we get that

$$\mathbf{I} - \left(\mathbf{B}^H \mathbf{R} \mathbf{B}(\mathbf{I}+\alpha\tilde{\mathbf{D}})^{-1}\right) = \left(\mathbf{I}+\mathbf{B}^H\mathbf{G}\right)^{-1}.$$

Hence,

$$\left(\sigma^2(\mathbf{I}+\alpha\tilde{\mathbf{D}})\right)^{-1} \left(\mathbf{I} - \mathbf{B}^H \mathbf{R} \mathbf{B}(\mathbf{I}+\alpha\tilde{\mathbf{D}})^{-1}\right) = \tilde{\mathbf{R}},$$

which eventually yields (77). This establishes the second equation of (76). ∎

We now establish that $\alpha-\beta$ and $\tilde{\alpha}-\tilde{\beta}$ are both $O(\frac{1}{t^2})$ terms.

First express $(\alpha-\beta) = \frac{1}{t}\mathrm{Tr}\mathbf{D}(\mathbf{R}-\mathbf{T}) + \epsilon$. Using (75) for $\mathbf{M}=\mathbf{D}$ yields

$$(\alpha-\beta)\left(1 - \frac{1}{t}\mathrm{Tr}\left[\mathbf{D}\mathbf{R}\mathbf{B}(\mathbf{I}+\beta\tilde{\mathbf{D}})^{-1}\tilde{\mathbf{D}}(\mathbf{I}+\beta\tilde{\mathbf{D}})^{-1}\mathbf{B}^H\mathbf{T}\right]\right) + (\tilde{\alpha}-\tilde{\beta})\,\sigma^2\frac{1}{t}\mathrm{Tr}(\mathbf{D}\mathbf{R}\mathbf{D}\mathbf{T}) = \epsilon. \tag{78}$$

Similarly, $(\tilde{\alpha}-\tilde{\beta}) = \frac{1}{t}\mathrm{Tr}\left[\tilde{\mathbf{D}}(\tilde{\mathbf{R}}-\tilde{\mathbf{T}})\right] + \tilde{\epsilon}$. Expressing $(\tilde{\mathbf{R}}-\tilde{\mathbf{T}})$ as $(\tilde{\mathbf{R}}-\tilde{\mathbf{T}}) = \tilde{\mathbf{R}}\left(\tilde{\mathbf{T}}^{-1}-\tilde{\mathbf{R}}^{-1}\right)\tilde{\mathbf{T}}$ and replacing $\tilde{\mathbf{T}}^{-1}$ and $\tilde{\mathbf{R}}^{-1}$ by their expressions, we obtain after straighforward computations :

$$(\alpha-\beta)\,\sigma^2\frac{1}{t}\mathrm{Tr}(\tilde{\mathbf{D}}\tilde{\mathbf{R}}\tilde{\mathbf{D}}\tilde{\mathbf{T}}) + (\tilde{\alpha}-\tilde{\beta})\left(1 - \frac{1}{t}\mathrm{Tr}\left[\tilde{\mathbf{D}}\tilde{\mathbf{R}}\mathbf{B}^H(\mathbf{I}+\tilde{\beta}\mathbf{D})^{-1}\mathbf{D}(\mathbf{I}+\tilde{\beta}\mathbf{D})^{-1}\mathbf{D}\tilde{\mathbf{T}}\right]\right) = \tilde{\epsilon}. \tag{79}$$

Equations (78) and (79) can be interpreted as a linear systems w.r.t. $(\alpha-\beta)$ and $(\tilde{\alpha}-\tilde{\beta})$. More precisely, if we define $(u_0, v_0, \tilde{u}_0, \tilde{v}_0)$ by

$$\begin{array}{rcl} u_0 & = & 1 - \frac{1}{t}\mathrm{Tr}(\mathbf{D}\mathbf{R}\mathbf{B}(\mathbf{I}+\beta\tilde{\mathbf{D}})^{-1}\tilde{\mathbf{D}}(\mathbf{I}+\beta\tilde{\mathbf{D}})^{-1}\mathbf{B}^H\mathbf{T}) \\ \tilde{v}_0 & = & 1 - \frac{1}{t}\mathrm{Tr}(\tilde{\mathbf{D}}\tilde{\mathbf{R}}\mathbf{B}^H(\mathbf{I}+\tilde{\beta}\mathbf{D})^{-1}\mathbf{D}(\mathbf{I}+\tilde{\beta}\mathbf{D})^{-1}\mathbf{B}\tilde{\mathbf{T}}) \\ v_0 & = & \sigma^2\frac{1}{t}\mathrm{Tr}(\mathbf{D}\mathbf{R}\mathbf{D}\mathbf{T}) \\ \tilde{u}_0 & = & \sigma^2\frac{1}{t}\mathrm{Tr}(\tilde{\mathbf{D}}\tilde{\mathbf{R}}\tilde{\mathbf{D}}\tilde{\mathbf{T}}) \end{array} \tag{80}$$

then, (78) and (79) can be written as

$$\begin{pmatrix} u_0 & v_0 \\ \tilde{u}_0 & \tilde{v}_0 \end{pmatrix} \begin{pmatrix} \alpha - \beta \\ \tilde{\alpha} - \tilde{\beta} \end{pmatrix} = \begin{pmatrix} \epsilon \\ \tilde{\epsilon} \end{pmatrix} . \tag{81}$$

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

$$\alpha - \beta \ = \ \frac{\tilde{v}_0\epsilon - v_0\tilde{\epsilon}}{u_0\tilde{v}_0 - \tilde{u}_0 v_0}, \qquad \tilde{\alpha} - \tilde{\beta} \ = \ \frac{u_0\tilde{\epsilon} - \tilde{u}_0\epsilon}{u_0\tilde{v}_0 - \tilde{u}_0 v_0} \ , \tag{82}$$

$u_0, v_0, \tilde{u}_0, \tilde{v}_0$ being uniformly bounded. As ϵ and $\tilde{\epsilon}$ are $O(\frac{1}{t^2})$ terms, $(\alpha - \beta)$ and $(\tilde{\alpha} - \tilde{\beta})$ will themselves be $O(\frac{1}{t^2})$ terms as long as the inverse $(u_0\tilde{v}_0 - \tilde{u}_0 v_0)^{-1}$ of the determinant is uniformly bounded. In order to state the corresponding result, we define $(u, v, \tilde{u}, \tilde{v})$ by

$$\begin{array}{rcl} u & = & 1 - \frac{1}{t}\mathrm{Tr}(\mathbf{DTB}(\mathbf{I} + \beta\tilde{\mathbf{D}})^{-1}\tilde{\mathbf{D}}(\mathbf{I} + \beta\tilde{\mathbf{D}})^{-1}\mathbf{B}^H\mathbf{T}) \\ \tilde{v} & = & 1 - \frac{1}{t}\mathrm{Tr}(\tilde{\mathbf{D}}\tilde{\mathbf{T}}\mathbf{B}^H(\mathbf{I} + \tilde{\beta}\mathbf{D})^{-1}\mathbf{D}(\mathbf{I} + \tilde{\beta}\mathbf{D})^{-1}\mathbf{B}\tilde{\mathbf{T}}) \\ v & = & \sigma^2\frac{1}{t}\mathrm{Tr}(\mathbf{DTDT}) \\ \tilde{u} & = & \sigma^2\frac{1}{t}\mathrm{Tr}(\tilde{\mathbf{D}}\tilde{\mathbf{T}}\tilde{\mathbf{D}}\tilde{\mathbf{T}}) \end{array} . \tag{83}$$

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 matrices $\mathbf{R}$ and $\tilde{\mathbf{R}}$ are replaced in the definition of $(u, v, \tilde{u}, \tilde{v})$ by matrices $\mathbf{T}$ and $\tilde{\mathbf{T}}$ respectively. The following result, proved in subsection II-E, suggests that the study of $u\tilde{v} - \tilde{u}v$ provides useful informations on $u_0\tilde{v}_0 - \tilde{u}_0 v_0$.

Lemma 3: $(u_0, v_0, \tilde{u}_0, \tilde{v}_0)$ can be written as

$$\begin{array}{rcl} u_0 & = & u + \epsilon_u \\ \tilde{v}_0 & = & \tilde{v} + \tilde{\epsilon}_v \\ v_0 & = & v + \epsilon_v \\ \tilde{u}_0 & = & \tilde{u} + \tilde{\epsilon}_u \end{array} \tag{84}$$

where $\epsilon_u, \tilde{\epsilon}_v, \tilde{\epsilon}_u, \epsilon_v$ converge to 0 when $t \to +\infty$.

The behaviour of $u\tilde{v} - \tilde{u}v$ is provided in the following Lemma, whose proof is given in subsection II-E.

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 $\sup_t \frac{1}{u\tilde{v} - \tilde{u}v} < +\infty$.

Lemmas 3 and 4 immediately imply that it exists t_0 such that $0 < u_0\tilde{v}_0 - \tilde{u}_0 v_0 \leq 1$ for each $t \geq t_0$ and

$$\sup_{t \geq t_0} \frac{1}{u_0\tilde{v}_0 - \tilde{u}_0 v_0} < +\infty \ . \tag{85}$$

This eventually shows $\alpha - \beta$ and $\tilde{\alpha} - \tilde{\beta}$ are of the same order of magnitude than ϵ and $\tilde{\epsilon}$, i.e. are $O(\frac{1}{t^2})$ terms which in turn establishes Proposition 8.

Eq. (28) eventually follows from the integral representation (27)

$$\overline{J}(\sigma^2) - J(\sigma^2) = \int_{\sigma^2}^{+\infty} \mathrm{Tr}\left(\mathrm{E}(\mathbf{S}(\omega)) - \mathbf{T}(\omega)\right) \, d\,\omega. \tag{86}$$

as well as a dominated convergence argument that is omitted.

D. Details of the proof of Step 1

We provide the detailed proof of Proposition 7. We first state a useful Lemma.

Lemma 5: 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 $\mathbf{Y}$.

$$\begin{aligned} \mathbf{\Phi}(\mathbf{Y}) &= \frac{1}{t}\mathrm{Tr}\left[\mathbf{S}\mathbf{P}\mathbf{\Sigma}^H\right] \ , & (87) \\ \mathbf{\Psi}(\mathbf{Y}) &= \frac{1}{t}\mathrm{Tr}\left[\mathbf{S}\mathbf{\Sigma}\mathbf{P}_1\mathbf{\Sigma}^H\mathbf{P}_2\right] \ , & (88) \\ \mathbf{\Psi}'(\mathbf{Y}) &= \frac{1}{t}\mathrm{Tr}\left[\mathbf{S}\mathbf{\Sigma}\mathbf{P}_1\mathbf{Y}^H\mathbf{P}_2\right] \ . & (89) \end{aligned}$$

Then, the following estimates hold true :

$$\begin{aligned} \mathrm{Var}(\mathbf{\Phi}) &= O\left(\frac{1}{t^2}\right) \ , & (90) \\ \mathrm{Var}(\mathbf{\Psi}) &= O\left(\frac{1}{t^2}\right) \ , & (91) \\ \mathrm{Var}(\mathbf{\Psi}') &= O\left(\frac{1}{t^2}\right) \ . & (92) \end{aligned}$$

The proof, based on the Poincaré-Nash inequality (60), is omitted.

We now complete proof of Step 1. We take Eq. (70) as a starting point, and write η as $\eta = \mathbb{E}(\eta) + \overset{\circ}{\eta} = \alpha + \overset{\circ}{\eta}$. Therefore,

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

Plugging this relation into (70), and solving w.r.t. $\mathbb{E}\left[(\mathbf{S}\boldsymbol{\xi}_j)_p\overline{\Sigma_q, j}\right]$ yields

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

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

$$\begin{aligned} \mathbb{E}\left[(\mathbf{S}\mathbf{\Sigma}\mathbf{\Sigma}^H)_{pq}\right] &= d_q\,\frac{1}{t}\mathrm{Tr}\left[\tilde{\mathbf{D}}(\mathbf{I}+\alpha\tilde{\mathbf{D}})^{-1}\right]\mathbb{E}(S_{pq}) \\ &\quad + \mathbb{E}\left[\left(\mathbf{S}\mathbf{B}(\mathbf{I}+\alpha\tilde{\mathbf{D}})^{-1}\mathbf{B}^H\right)_{pq}\right] - d_q\,\mathbb{E}\left[S_{pq}\frac{1}{t}\mathrm{Tr}\left(\mathbf{S}\mathbf{B}\tilde{\mathbf{D}}(\mathbf{I}+\alpha\tilde{\mathbf{D}})^{-1}\mathbf{B}^H\right)\right] \\ &\quad - d_q\,\mathbb{E}\left[S_{pq}\frac{1}{t}\mathrm{Tr}\left(\mathbf{S}\mathbf{B}\tilde{\mathbf{D}}(\mathbf{I}+\alpha\tilde{\mathbf{D}})^{-1}\mathbf{Y}^H\right)\right] - \mathbb{E}\left[\overset{\circ}{\eta}\left(\mathbf{S}\mathbf{\Sigma}\tilde{\mathbf{D}}(\mathbf{I}+\alpha\tilde{\mathbf{D}})^{-1}\mathbf{\Sigma}^H\right)_{p,q}\right] \ . \end{aligned} \tag{93}$$

The resolvent identity (68) thus implies that

$$\begin{aligned}\delta(p-q) \quad &= \quad \sigma^2\mathbb{E}(S_{pq}) + \frac{d_q}{t}\mathrm{Tr}\left[\tilde{\mathbf{D}}(\mathbf{I}+\alpha\tilde{\mathbf{D}})^{-1}\right]\mathbb{E}(S_{pq}) \\ &+ \mathbb{E}\left[\left(\mathbf{SB}(\mathbf{I}+\alpha\tilde{\mathbf{D}})^{-1}\mathbf{B}^H\right)_{pq}\right] - d_q\,\mathbb{E}\left[S_{pq}\frac{1}{t}\mathrm{Tr}\left(\mathbf{SB}\tilde{\mathbf{D}}(\mathbf{I}+\alpha\tilde{\mathbf{D}})^{-1}\mathbf{B}^H\right)\right] \\ &- d_q\,\mathbb{E}\left[S_{pq}\frac{1}{t}\mathrm{Tr}\left(\mathbf{SB}\tilde{\mathbf{D}}(\mathbf{I}+\alpha\tilde{\mathbf{D}})^{-1}\mathbf{Y}^H\right)\right] - \mathbb{E}\left[\overset{\circ}{\eta}\left(\mathbf{S\Sigma}\tilde{\mathbf{D}}(\mathbf{I}+\alpha\tilde{\mathbf{D}})^{-1}\mathbf{\Sigma}^H\right)_{p,q}\right]\ . \end{aligned} \tag{94}$$

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

$$\rho_1 \;=\; \frac{1}{t}\mathrm{Tr}\left(\mathbf{SB}\tilde{\mathbf{D}}(\mathbf{I}+\alpha\tilde{\mathbf{D}})^{-1}\mathbf{B}^H\right) \quad \text{and} \quad \rho_2 \;=\; \frac{1}{t}\mathrm{Tr}\left(\mathbf{SB}\tilde{\mathbf{D}}(\mathbf{I}+\alpha\tilde{\mathbf{D}})^{-1}\mathbf{Y}^H\right)\ .$$

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

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

Thus, (94) can thus be written as

$$\begin{aligned}\delta(p-q) \quad &= \quad \sigma^2\mathbb{E}(S_{pq}) + d_q\,\frac{1}{t}\mathrm{Tr}\left[\tilde{\mathbf{D}}(\mathbf{I}+\alpha\tilde{\mathbf{D}})^{-1}\right]\mathbb{E}(S_{pq}) \\ &+ \left(\mathbb{E}(\mathbf{S})\mathbf{B}(\mathbf{I}+\alpha\tilde{\mathbf{D}})^{-1}\mathbf{B}^H\right)_{pq} - d_q\,\mathbb{E}(S_{pq})\frac{1}{t}\mathrm{Tr}\left(\mathbb{E}(\mathbf{S})\mathbf{B}\tilde{\mathbf{D}}(\mathbf{I}+\alpha\tilde{\mathbf{D}})^{-1}\mathbf{B}^H\right) \\ &- d_q\,\mathbb{E}(S_{pq})\mathbb{E}\left[\frac{1}{t}\mathrm{Tr}\left(\mathbf{SB}\tilde{\mathbf{D}}(\mathbf{I}+\alpha\tilde{\mathbf{D}})^{-1}\mathbf{Y}^H\right)\right] - d_q\,\mathbb{E}\left(\overset{\circ}{S}_{pq}\ \overset{\circ}{\rho}_1\right) - d_q\,\mathbb{E}\left(\overset{\circ}{S}_{pq}\ \overset{\circ}{\rho}_2\right) \\ &- \mathbb{E}\left[\overset{\circ}{\eta}\left(\mathbf{S\Sigma}\tilde{\mathbf{D}}(\mathbf{I}+\alpha\tilde{\mathbf{D}})^{-1}\mathbf{\Sigma}^H\right)_{p,q}\right]\ . \end{aligned} \tag{95}$$

We now establish the following lemma.

Lemma 6:

$$\begin{aligned}\mathbb{E}\rho_2 \quad &= \quad \mathbb{E}\left[\frac{1}{t}\mathrm{Tr}\left(\mathbf{SB}\tilde{\mathbf{D}}(\mathbf{I}+\alpha\tilde{\mathbf{D}})^{-1}\mathbf{Y}^H\right)\right] \\ &= \quad -\alpha\,\frac{1}{t}\mathrm{Tr}\left(\mathbb{E}(\mathbf{S})\mathbf{B}\tilde{\mathbf{D}}^2(\mathbf{I}+\alpha\tilde{\mathbf{D}})^{-2}\mathbf{B}^H\right) - \mathbb{E}\left(\overset{\circ}{\eta}\ \overset{\circ}{\rho}_3\right)\ , \end{aligned} \tag{96}$$

where ρ_3 is defined by

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

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

$$\begin{aligned}\mathbb{E}(\rho_2) \quad &= \quad \tfrac{1}{t}\textstyle\sum_{j=1}^t \frac{\tilde{d}_j}{1+\alpha\tilde{d}_j}\ \mathbb{E}(\mathbf{y}_j^H\mathbf{S}\mathbf{b}_j) \\ &= \quad \tfrac{1}{t}\textstyle\sum_{j=1}^t \frac{\tilde{d}_j}{1+\alpha\tilde{d}_j}\ \sum_{i=1}^r \mathbb{E}\left((\mathbf{S}\mathbf{b}_j)_i\overline{Y_{ij}}\right) \end{aligned} \tag{97}$$

and evaluate $\mathbb{E}\left((\mathbf{S}\mathbf{b}_j)_i\overline{Y_{ij}}\right)$ using formula (63) for $\Phi(\mathbf{Y}) = (\mathbf{S}\mathbf{b}_j)_i$. This gives

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

By (64),

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

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

$$\begin{aligned}\mathbb{E}\left(\mathbf{y}_j^H\mathbf{S}\mathbf{b}_j\right) &= -\alpha\tilde{d}_j\mathbb{E}\left(\mathbf{b}_j^H\mathbf{S}\mathbf{b}_j\right) - \alpha\tilde{d}_j\mathbb{E}\left(\mathbf{y}_j^H\mathbf{S}\mathbf{b}_j\right) \\ &\quad -\tilde{d}_j\mathbb{E}\left(\overset{\circ}{\eta}\,\mathbf{b}_j^H\mathbf{S}\mathbf{b}_j\right) - \tilde{d}_j\mathbb{E}\left(\overset{\circ}{\eta}\,\mathbf{y}_j^H\mathbf{S}\mathbf{b}_j\right)\end{aligned} \tag{98}$$

Solving this equation w.r.t. $\mathbb{E}\left(\mathbf{y}_j^H\mathbf{S}\mathbf{b}_j\right)$ yields

$$\mathbb{E}\left(\mathbf{y}_j^H\mathbf{S}\mathbf{b}_j\right) = -\frac{\alpha\tilde{d}_j}{1+\alpha\tilde{d}_j}\mathbb{E}\left(\mathbf{b}_j^H\mathbf{S}\mathbf{b}_j\right) - \frac{\tilde{d}_j}{1+\alpha\tilde{d}_j}\mathbb{E}\left(\overset{\circ}{\eta}\,\mathbf{b}_j^H\mathbf{S}\mathbf{b}_j\right) - \frac{\tilde{d}_j}{1+\alpha\tilde{d}_j}\mathbb{E}\left(\overset{\circ}{\eta}\,\mathbf{y}_j^H\mathbf{S}\mathbf{b}_j\right) \tag{99}$$

or equivalently

$$\mathbb{E}\left(\mathbf{y}_j^H\mathbf{S}\mathbf{b}_j\right) = -\frac{\alpha\tilde{d}_j}{1+\alpha\tilde{d}_j}\mathbb{E}\left(\mathbf{b}_j^H\mathbf{S}\mathbf{b}_j\right) - \frac{\tilde{d}_j}{1+\alpha\tilde{d}_j}\mathbb{E}\left(\overset{\circ}{\eta}\,\boldsymbol{\xi}_j^H\mathbf{S}\mathbf{b}_j\right) \;. \tag{100}$$

Eq. (96) immediately follows from (97), (100), and the relation $\mathbb{E}(\overset{\circ}{\eta}\,\rho_3) = \mathbb{E}(\overset{\circ}{\eta}\,\overset{\circ}{\rho}_3)$. ∎

Plugging (96) into (95) yields

$$\begin{aligned}&\delta(p-q) + \Delta_{pq} \\ &= \mathbb{E}(S_{pq})\left[\sigma^2 + d_q\left(\frac{1}{t}\mathrm{Tr}\tilde{\mathbf{D}}(\mathbf{I}+\alpha\tilde{\mathbf{D}})^{-1} - \mathbb{E}(\rho_1) + \alpha\frac{1}{t}\mathrm{Tr}\mathbb{E}(\mathbf{S})\mathbf{B}\tilde{\mathbf{D}}^2(\mathbf{I}+\alpha\tilde{\mathbf{D}})^{-2}\mathbf{B}^H\right)\right] \\ &\qquad + \left[\mathbb{E}(\mathbf{S})\mathbf{B}(\mathbf{I}+\alpha\tilde{\mathbf{D}})^{-1}\mathbf{B}^H\right]_{pq}\end{aligned} \tag{101}$$

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

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

for each p, q or equivalently by

$$\boldsymbol{\Delta} = \mathbb{E}\left[\overset{\circ}{\eta}\left(\mathbf{S}\boldsymbol{\Sigma}\tilde{\mathbf{D}}(\mathbf{I}+\alpha\tilde{\mathbf{D}})^{-1}\boldsymbol{\Sigma}^H\right)\right] + \mathbb{E}\left((\overset{\circ}{\rho}_1 + \overset{\circ}{\rho}_2)\,\overset{\circ}{\mathbf{S}}\right)\,\mathbf{D} - \mathbb{E}\left(\overset{\circ}{\eta}\,\overset{\circ}{\rho}_3\right)\,\mathbb{E}(\mathbf{S})\,\mathbf{D} \;.$$

Using the relation

$$\alpha\tilde{\mathbf{D}}(\mathbf{I}+\alpha\tilde{\mathbf{D}})^{-1} = \mathbf{I} - (\mathbf{I}+\alpha\tilde{\mathbf{D}})^{-1} \;,$$

we obtain that

$$\begin{aligned}&\alpha\frac{1}{t}\mathrm{Tr}\left(\mathbb{E}(\mathbf{S})\mathbf{B}\tilde{\mathbf{D}}^2(\mathbf{I}+\alpha\tilde{\mathbf{D}})^{-2}\mathbf{B}^H\right) \\ &= \frac{1}{t}\mathrm{Tr}\left(\mathbb{E}(\mathbf{S})\mathbf{B}\tilde{\mathbf{D}}(\mathbf{I}+\alpha\tilde{\mathbf{D}})^{-1}\mathbf{B}^H\right) - \frac{1}{t}\mathrm{Tr}\left(\mathbb{E}(\mathbf{S})\mathbf{B}\tilde{\mathbf{D}}(\mathbf{I}+\alpha\tilde{\mathbf{D}})^{-2}\mathbf{B}^H\right) \\ &= \mathbb{E}(\rho_1) - \frac{1}{t}\mathrm{Tr}\left(\mathbb{E}(\mathbf{S})\mathbf{B}\tilde{\mathbf{D}}(\mathbf{I}+\alpha\tilde{\mathbf{D}})^{-2}\mathbf{B}^H\right) \;.\end{aligned} \tag{102}$$

Therefore, the term

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

is equal to

$$\frac{1}{t}\mathrm{Tr}\tilde{\mathbf{D}}(\mathbf{I}+\alpha\tilde{\mathbf{D}})^{-1}-\frac{1}{t}\mathrm{Tr}\left(\mathbb{E}(\mathbf{S})\mathbf{B}\tilde{\mathbf{D}}(\mathbf{I}+\alpha\tilde{\mathbf{D}})^{-2}\mathbf{B}^H\right) = \frac{1}{t}\mathrm{Tr}\left[\tilde{\mathbf{D}}(\mathbf{I}+\alpha\tilde{\mathbf{D}})^{-1}\left(\mathbf{I}-\mathbf{B}^H\mathbb{E}(\mathbf{S})\mathbf{B}(\mathbf{I}+\alpha\tilde{\mathbf{D}})^{-1}\right)\right]$$

which, in turn, coincides with $\sigma^2\,\tilde{\alpha}$ (see Eq. (65)). Eq. (101) is thus equivalent to

$$\left(\mathbb{E}(\mathbf{S})\left[\sigma^2(\mathbf{I}+\tilde{\alpha}\mathbf{D})+\mathbf{B}(\mathbf{I}+\alpha\tilde{\mathbf{D}})^{-1}\mathbf{B}^H\right]\right)_{pq}=\delta(p-q)+\Delta_{pq} \tag{103}$$

or, in matrix form,

$$\mathbb{E}(\mathbf{S})\,\mathbf{R}^{-1}=\mathbf{I}+\mathbf{\Delta} \tag{104}$$

i.e.

$$\mathbb{E}(\mathbf{S})=\mathbf{R}+\mathbf{\Delta}\,\mathbf{R}\ . \tag{105}$$

In order to complete the proof of Proposition 7, it remains to check that if $\mathbf{M}$ is a deterministic, uniformly bounded matrix for the spectral norm as $n\to\infty$, then

$$\frac{1}{t}\mathrm{Tr}\mathbf{\Delta}\mathbf{R}\mathbf{M}=O\left(\frac{1}{t^2}\right)\ .$$

For this, we write $\frac{1}{t}\,\mathrm{Tr}\mathbf{\Delta}\mathbf{R}\mathbf{M}$ as $\frac{1}{t}\,\mathrm{Tr}\mathbf{\Delta}\mathbf{R}\mathbf{M}=T_1+T_2-T_3$ where

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

We denote by ρ_4 the term

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

and remark that $T_1=\mathbb{E}(\overset{\circ}{\eta}\ \overset{\circ}{\rho}_4)$. Eq. (26) implies that $\mathbb{E}(\overset{\circ}{\eta}^2)$ and $\mathbb{E}\left[\frac{1}{t}\,\mathrm{Tr}\left(\overset{\circ}{\mathbf{S}}\mathbf{D}\mathbf{R}\mathbf{M})\right)\right]^2$ are $O(\frac{1}{t^2})$ terms. Moreover, Lemma 5 immediately shows that for each $i=1,2,3$, $\mathbb{E}(\overset{\circ}{\rho}_i^{\,2})$ is a $O(\frac{1}{t^2})$ term. The Cauchy-Schwarz inequality eventually provides $\frac{1}{t}\mathrm{Tr}\mathbf{\Delta}\mathbf{R}\mathbf{M}=O(\frac{1}{t^2})$, which completes the proof of Proposition 7.

E. Details of the proof of Step 2

1) Proof of Lemma 3: In order to establish Lemma 3, we first prove that :

Lemma 7:

$$\alpha-\beta\ =\ o(1)\quad\text{and}\quad\tilde{\alpha}-\tilde{\beta}\ =\ o(1)\ . \tag{106}$$

Proof: We first prove that if σ^2 is large enough, then (106) holds. For this, we take (82) as a starting point, and study the behaviour of coefficients $u_0,\tilde{u}_0,v_0,\tilde{v}_0$ for large enough values of σ^2. As matrices $\mathbf{R}$ and $\mathbf{T}$ are less than $\frac{1}{\sigma^2}\mathbf{I}_r$ and matrices $\tilde{\mathbf{R}}$ and $\tilde{\mathbf{T}}$ are less than $\frac{1}{\sigma^2}\mathbf{I}_t$, it is clear that :

$$\begin{array}{rclcl} u_0 & \geq & 1-\frac{1}{\sigma^4}\frac{1}{t}\mathrm{Tr}\left[\mathbf{D}\mathbf{B}(\mathbf{I}+\beta\tilde{\mathbf{D}})^{-1}\tilde{\mathbf{D}}(\mathbf{I}+\beta\tilde{\mathbf{D}})^{-1}\mathbf{B}^H\right] & \geq & 1-\frac{1}{\sigma^4}\,\frac{r}{t}\,d_{\max}\tilde{d}_{\max}b_{\max}^2 \ , \\ \tilde{v}_0 & \geq & 1-\frac{1}{\sigma^4}\frac{1}{t}\mathrm{Tr}\left[\tilde{\mathbf{D}}\mathbf{B}^H(\mathbf{I}+\tilde{\beta}\mathbf{D})^{-1}\mathbf{D}(\mathbf{I}+\tilde{\beta}\mathbf{D})^{-1}\mathbf{B}\right] & \geq & 1-\frac{1}{\sigma^4}\,d_{\max}\tilde{d}_{\max}b_{\max}^2 \ , \\ \tilde{u}_0 & \leq & \frac{\tilde{d}_{\max}^2}{\sigma^2} \ , & & \\ v_0 & \leq & \frac{r}{t}\,\frac{d_{\max}^2}{\sigma^2} \ . & & \end{array} \tag{107}$$

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, $u_0\tilde{v}_0-\tilde{u}_0 v_0>\frac{3}{16}$ for $t\geq t_0$ and $\sigma^2\geq\sigma_0^2$. Eq. (82) thus implies that if $\sigma^2\geq\sigma_0^2$, then $\alpha-\beta$ and $\tilde{\alpha}-\tilde{\beta}$ are of the same order of magnitude as $\epsilon=O(\frac{1}{t^2})$, 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. [5]), a tool frequently used in the context of large random matrices. It is based on the observation that, considered as functions of parameter σ^2, $\alpha(\sigma^2)-\beta(\sigma^2)$ and $\tilde{\alpha}(\sigma^2)-\tilde{\beta}(\sigma^2)$ 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 $\sup_t\sup_{z\in K}\ |\alpha(z)-\beta(z)|<+\infty$ and $\sup_t\sup_{z\in K}\ |\tilde{\alpha}(z)-\tilde{\beta}(z)|<+\infty$. Using Montel's theorem, it can thus be shown that if $\alpha(\sigma^2)-\beta(\sigma^2)$ and $\tilde{\alpha}(\sigma^2)-\tilde{\beta}(\sigma^2)$ converge toward zero for each $\sigma^2>\sigma_0^2$, then for each $z\in\mathbb{C}-\mathbb{R}^-$, $\alpha(z)-\beta(z)$ and $\tilde{\alpha}(z)-\tilde{\beta}(z)$ converge as well towards 0. This in particular implies that $\alpha(\sigma^2)-\beta(\sigma^2)$ and $\tilde{\alpha}(\sigma^2)-\tilde{\beta}(\sigma^2)$ converge towards 0 for each $\sigma^2>0$. This proves Lemma 7. For more details, the reader may e.g. refer to [17]. ■

We note that Montel's theorem does not guarantee that $\alpha-\beta$ and $\tilde{\alpha}-\tilde{\beta}$ are still $O(\frac{1}{t^2})$ terms for $\sigma^2<\sigma_0^2$. It is therefore necessary to prove Lemmas 3 and 4 to obtain this result from Eq. (82).

In order to complete the proof of Lemma 3, we observe that, by (75) and (106), $\frac{1}{t}\mathrm{Tr}\left[\mathbf{M}(\mathbf{R}-\mathbf{T})\right]$ converges towards 0 for each uniformly bounded matrix $\mathbf{M}$. It can be shown similarly that $\frac{1}{t}\mathrm{Tr}\left[\tilde{\mathbf{M}}(\tilde{\mathbf{R}}-\tilde{\mathbf{T}})\right]$ converges towards 0 for each uniformly bounded matrix $\tilde{\mathbf{M}}$. Using these properties for relevant matrices $\mathbf{M}$ and $\tilde{\mathbf{M}}$ immediately yields Lemma 3.

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

$$\tilde{\mathbf{T}}\mathbf{B}^H(\mathbf{I}+\tilde{\beta}\mathbf{D})^{-1}=(\mathbf{I}+\beta\tilde{\mathbf{D}})^{-1}\mathbf{B}^H\mathbf{T} \ . \tag{108}$$

The equality $u=\tilde{v}$ immediately follows from (108).

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 Eq. (16) 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 $d\mu_b(\lambda) = \frac{1}{t}\mathrm{Tr}(\mathbf{D}d\boldsymbol{\mu}(\lambda))$ and $d\tilde{\mu}_b(\lambda) = \frac{1}{t}\mathrm{Tr}(\tilde{\mathbf{D}}d\tilde{\boldsymbol{\mu}}(\lambda))$. Remark that $\mu_b(\mathbb{R}^+) = \frac{1}{t}\mathrm{Tr}(\mathbf{D})$ and that $\tilde{\mu}_b(\mathbb{R}^+) = \frac{1}{t}\mathrm{Tr}(\tilde{\mathbf{D}})$. 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 w.r.t. σ^2. Then, $(\sigma^2\beta)' > 0$ and $\tilde{\beta}' < 0$ for each σ^2. We now differentiate relations (15) w.r.t. σ^2. After some algebra, we obtain :

$$\begin{array}{rcl} u\,(\sigma^2\beta)' + \sigma^2 v\,\tilde{\beta}' &=& \frac{1}{t}\mathrm{Tr}(\mathbf{DTB}(\mathbf{I}+\beta\tilde{\mathbf{D}})^{-1}(\mathbf{I}+\beta\tilde{\mathbf{D}})^{-1}\mathbf{B}^H\mathbf{T}) \\ \frac{\tilde{u}}{\sigma^2}\,(\sigma^2\beta)' + \tilde{v}\tilde{\beta}' &=& -\frac{1}{t}\mathrm{Tr}\tilde{\mathbf{T}}\tilde{\mathbf{D}}\tilde{\mathbf{T}} \end{array} . \tag{109}$$

As $\tilde{\beta}' < 0$, the first equation of (109) implies that $u\,(\sigma^2\beta)' > 0$. As $(\sigma^2\beta)' > 0$, this yields $u > 0$. As $u < 1$ clearly holds, the first part of (ii) is proved.

We now prove that $\inf_t u > 0$. The first equation of (109) yields :

$$u > -\sigma^2 v\tilde{\beta}' \frac{1}{(\sigma^2\beta)'} . \tag{110}$$

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 representation (16),

$$-\tilde{\beta}' = \int_{\mathbb{R}^+} \frac{d\tilde{\mu}_b(\lambda)}{(\lambda+\sigma^2)^2} \quad \text{and} \quad (\sigma^2\beta(\sigma^2))' = \int_{\mathbb{R}^+} \frac{\lambda d\mu_b(\lambda)}{(\lambda+\sigma^2)^2} .$$

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 :

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

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

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

As mentioned above, $(\frac{1}{t}\mathrm{Tr}\tilde{\mathbf{D}})^{-1}$ 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 $(t_k)_{k\geq 0}$ 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 $(\tilde{\mu}_b^{(t_k)})_{k\geq 0}$ is tight. One can therefore extract from $(\tilde{\mu}_b^{(t_k)})_{k\geq 0}$ a subsequence $(\tilde{\mu}_b^{(t_l^{'})})_{l\geq 0}$ 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 $(\tilde{\mu}_b^{(t_l^{'})})_{l\geq 0}$ 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 $\mathbf{M}$,

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

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, $\inf_t \frac{1}{t}\mathrm{Tr}\mathbf{DTDT} \geq \inf_t \beta^2$. 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 (109) to express $(\sigma^2\beta)^{'}$ in terms of $\tilde{\beta}^{'}$, and plug this relation into the second equation of (109). This gives :

$$\left(\tilde{v} - \frac{1}{u}\tilde{u}v\right)\ \tilde{\beta}^{'} = -\frac{1}{t}\mathrm{Tr}\tilde{\mathbf{T}}\tilde{\mathbf{D}}\tilde{\mathbf{T}} - \frac{\tilde{u}}{\sigma^2 u}\frac{1}{t}\mathrm{Tr}(\mathbf{DTB}(\mathbf{I}+\beta\tilde{\mathbf{D}})^{-1}(\mathbf{I}+\beta\tilde{\mathbf{D}})^{-1}\mathbf{B}^H\mathbf{T}) \ . \qquad (111)$$

The righthand side of (111) is negative as well as $\tilde{\beta}^{'}$. Therefore, $\tilde{v} - \frac{1}{u}\tilde{u}v > 0$. As u 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 remark that by (111),

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

$|\tilde{\beta}'|$ clearly satisfies $|\tilde{\beta}'| \le \frac{1}{\sigma^4}\frac{1}{t}\mathrm{Tr}\tilde{\mathbf{D}}$ and is thus upper bounded by $\frac{\tilde{d}_{\max}}{\sigma^4}$. (ii) implies that $\sup_t \frac{1}{u} < +\infty$. 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 remark that $\frac{1}{t}\sum_{i=1}^{t}\tilde{\kappa}_i = 1$. x can be written as

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

By Jensen's inequality

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

Moreover,

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

Finally,

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

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

STRICT CONCAVITY OF $\bar{I}(\mathbf{Q})$: REMAINING PROOFS

A. Proof of Lemma 1

Remark that ϕ_m is strictly concave due to (45). Remark also that $\overline{\phi}$ is concave as a pointwise limit of the ϕ_m's. Now in order to prove the strict concavity of $\overline{\phi}$, assume that there exists a subinterval, say $(a,b) \subset [0,1]$ with $a < b$ where $\overline{\phi}$ fails to be strictly concave :

$$\forall \lambda \in [0,1], \quad \overline{\phi}(\lambda a + (1-\lambda)b) = \lambda\overline{\phi}(a) + (1-\lambda)\overline{\phi}(b) \ .$$

Otherwise stated,

$$\forall x \in (a,b), \quad \overline{\phi}(x) = \frac{\overline{\phi}(b) - \overline{\phi}(a)}{b-a}(x-a) + \overline{\phi}(a).$$

Let $x \in (a,b)$ and $h > 0$ be small enough so that $x-h$ and $x+h$ belong to (a,b) ; recall the following inequality, valid for differentiable concave functions :

$$\frac{\phi_m(x) - \phi_m(x-h)}{h} \ge \phi_m'(x) \ge \frac{\phi_m(x+h) - \phi_m(x)}{h} \ .$$

Letting $m \to \infty$, we obtain :

$$\frac{\overline{\phi}(x) - \overline{\phi}(x-h)}{h} \;\geq\; \limsup_{m\to\infty} \phi'_m(x) \;\geq\; \liminf_{m\to\infty} \phi'_m(x) \;\geq\; \frac{\overline{\phi}(x+h) - \overline{\phi}(x)}{h} \;.$$

In particular, for all $x \in (a,b)$, $\lim_{m\to\infty} \phi'_m(x) = \frac{\overline{\phi}(b)-\overline{\phi}(a)}{b-a}$. Now let $[x, x+h] \in (a,b)$. Fatou's lemma together with (45) yield :

$$\begin{aligned} 0 \;<\; \kappa h \;\leq\; & \int_x^{x+h} \liminf_{m\to\infty} \phi''_m(u)\, du \\ & \leq\; \liminf_{m\to\infty} \int_x^{x+h} \phi''_m(u)\, du \;=\; \lim_{m\to\infty} \left(\phi'_m(x+h) - \phi'_m(x) \right) \;=\; 0 \;. \end{aligned}$$

This yields a contradiction, therefore $\overline{\phi}$ must be strictly convex on $[0,1]$.

B. Proof of (46).

We define $\check{\mathbf{M}}$ as the $tm \times tm$ matrix given by

$$\check{\mathbf{M}} = \check{\mathbf{H}}^H \left(\mathbf{I} + \frac{\check{\mathbf{H}}\check{\mathbf{Q}}\check{\mathbf{H}}^H}{\sigma^2} \right)^{-1} \frac{\check{\mathbf{H}}}{\sigma^2} \;.$$

We have :

$$\phi''_m(\lambda) = -\frac{1}{m} \mathbb{E}\, \mathrm{Tr} \left[\check{\mathbf{M}}(\check{\mathbf{Q}}_1 - \check{\mathbf{Q}}_2)\check{\mathbf{M}}(\check{\mathbf{Q}}_1 - \check{\mathbf{Q}}_2) \right]$$

or equivalently

$$\phi''_m(\lambda) = -\frac{1}{m} \mathbb{E}\, \mathrm{Tr} \left[\left(\mathbf{I} + \frac{\check{\mathbf{H}}\check{\mathbf{Q}}\check{\mathbf{H}}^H}{\sigma^2} \right)^{-1} \frac{\check{\mathbf{H}}}{\sigma^2} (\check{\mathbf{Q}}_1 - \check{\mathbf{Q}}_2)\check{\mathbf{M}}(\check{\mathbf{Q}}_1 - \check{\mathbf{Q}}_2)\check{\mathbf{H}}^H \right]$$

Recall that $\mathrm{Tr}(\mathbf{AB}) \geq \lambda_{\min}(\mathbf{A})\mathrm{Tr}(\mathbf{B})$ for $\mathbf{A}$, $\mathbf{B}$ Hermitian and nonnegative matrices. In particular :

$$\begin{aligned} \mathrm{Tr} & \left[\left(\mathbf{I} + \frac{\check{\mathbf{H}}\check{\mathbf{Q}}\check{\mathbf{H}}^H}{\sigma^2} \right)^{-1} \frac{\check{\mathbf{H}}}{\sigma^2} (\check{\mathbf{Q}}_1 - \check{\mathbf{Q}}_2)\check{\mathbf{M}}(\check{\mathbf{Q}}_1 - \check{\mathbf{Q}}_2)\check{\mathbf{H}}^H \right] \\ & \geq \lambda_{\min} \left(\mathbf{I} + \frac{\check{\mathbf{H}}\check{\mathbf{Q}}\check{\mathbf{H}}^H}{\sigma^2} \right)^{-1} \mathrm{Tr} \left[\frac{\check{\mathbf{H}}}{\sigma^2} (\check{\mathbf{Q}}_1 - \check{\mathbf{Q}}_2)\check{\mathbf{M}}(\check{\mathbf{Q}}_1 - \check{\mathbf{Q}}_2)\check{\mathbf{H}}^H \right] \;. \end{aligned}$$

Similarly, we obtain that

$$\begin{aligned} \mathrm{Tr} & \left[\frac{\check{\mathbf{H}}}{\sigma^2} (\check{\mathbf{Q}}_1 - \check{\mathbf{Q}}_2)\check{\mathbf{M}}(\check{\mathbf{Q}}_1 - \check{\mathbf{Q}}_2)\check{\mathbf{H}}^H \right] \\ & \geq \lambda_{\min} \left(\mathbf{I} + \frac{\check{\mathbf{H}}\check{\mathbf{Q}}\check{\mathbf{H}}^H}{\sigma^2} \right)^{-1} \mathrm{Tr} \left[\frac{\check{\mathbf{H}}}{\sigma^2} (\check{\mathbf{Q}}_1 - \check{\mathbf{Q}}_2) \frac{\check{\mathbf{H}}^H\check{\mathbf{H}}}{\sigma^2} (\check{\mathbf{Q}}_1 - \check{\mathbf{Q}}_2)\check{\mathbf{H}}^H \right] \;. \end{aligned}$$

This eventually implies that

$$\begin{aligned} \mathrm{Tr} & \left[\left(\mathbf{I} + \frac{\check{\mathbf{H}}\check{\mathbf{Q}}\check{\mathbf{H}}^H}{\sigma^2} \right)^{-1} \frac{\check{\mathbf{H}}}{\sigma^2} (\check{\mathbf{Q}}_1 - \check{\mathbf{Q}}_2)\check{\mathbf{M}}(\check{\mathbf{Q}}_1 - \check{\mathbf{Q}}_2)\check{\mathbf{H}}^H \right] \geq \\ & \lambda_{\min}^2 \left(\mathbf{I} + \frac{\check{\mathbf{H}}\check{\mathbf{Q}}\check{\mathbf{H}}^H}{\sigma^2} \right)^{-1} \mathrm{Tr} \left[\frac{\check{\mathbf{H}}^H\check{\mathbf{H}}}{\sigma^2} (\check{\mathbf{Q}}_1 - \check{\mathbf{Q}}_2) \frac{\check{\mathbf{H}}^H\check{\mathbf{H}}}{\sigma^2} (\check{\mathbf{Q}}_1 - \check{\mathbf{Q}}_2) \right] \;. \end{aligned}$$

As

$$\lambda_{\min}^2\left(\mathbf{I}+\frac{\check{\mathbf{H}}\check{\mathbf{Q}}\check{\mathbf{H}}^H}{\sigma^2}\right)^{-1} \geq \frac{1}{\lambda_{\max}^2\left(\mathbf{I}+\frac{\check{\mathbf{H}}\check{\mathbf{Q}}\check{\mathbf{H}}^H}{\sigma^2}\right)} \geq \frac{1}{\left(1+\sigma^{-2}\|\check{\mathbf{Q}}\|\,\|\check{\mathbf{H}}^H\check{\mathbf{H}}\|\right)^2}\ ,$$

we have :

$$\phi_m''(\lambda) \leq -\frac{1}{m}\mathbb{E}\left[\left(\frac{1}{\left(1+\sigma^{-2}\|\check{\mathbf{Q}}\|\,\|\check{\mathbf{H}}^H\check{\mathbf{H}}\|\right)^2}\right)\times \mathrm{Tr}\left(\frac{\check{\mathbf{H}}^H\check{\mathbf{H}}}{\sigma^2}(\check{\mathbf{Q}}_1-\check{\mathbf{Q}}_2)\frac{\check{\mathbf{H}}^H\check{\mathbf{H}}}{\sigma^2}(\check{\mathbf{Q}}_1-\check{\mathbf{Q}}_2)\right)\right]\ .$$

Let us introduce the following notations :

$$\alpha_m=\frac{1}{\left(1+\sigma^{-2}\|\check{\mathbf{Q}}\|\,\|\check{\mathbf{H}}^H\check{\mathbf{H}}\|\right)^2},\qquad \beta_m=\frac{1}{m}\mathrm{Tr}\left[\frac{\check{\mathbf{H}}^H\check{\mathbf{H}}}{\sigma^2}(\check{\mathbf{Q}}_1-\check{\mathbf{Q}}_2)\frac{\check{\mathbf{H}}^H\check{\mathbf{H}}}{\sigma^2}(\check{\mathbf{Q}}_1-\check{\mathbf{Q}}_2)\right]\ .$$

The following properties whose proofs are postponed to Appendix III-C hold true :

Proposition 9: (i) $\lim_{m\to\infty}\mathrm{var}(\beta_m)=0$,

(ii) For all $m\geq 1$, $\mathbb{E}(\beta_m)=\mathbb{E}(\beta_1)=\mathbb{E}\mathrm{Tr}\left[\frac{\mathbf{H}^H\mathbf{H}}{\sigma^2}(\mathbf{Q}_1-\mathbf{Q}_2)\frac{\mathbf{H}^H\mathbf{H}}{\sigma^2}(\mathbf{Q}_1-\mathbf{Q}_2)\right]>0$,

(iii) There exists $\delta>0$ such that for all $\lambda\in[0,1]$, $\liminf_{m\to\infty}\mathbb{E}(\alpha_m)\geq\delta>0$.

We are now in position to establish (46). By Proposition 9-(i), we have

$$|\mathbb{E}(\alpha_m\beta_m)-\mathbb{E}(\alpha_m)\mathbb{E}(\beta_m)|\leq\sqrt{\mathrm{var}(\beta_m)}\sqrt{\mathbb{E}(\alpha_m^2)}\leq\sqrt{\mathrm{var}(\beta_m)}\xrightarrow[m\to\infty]{}0\ .$$

By Proposition 9-(ii),(iii), we have :

$$\liminf_{m\to\infty}\mathbb{E}(\alpha_m\beta_m)=\liminf_{m\to\infty}\mathbb{E}(\alpha_m)\mathbb{E}(\beta_m)=\mathbb{E}(\beta_1)\liminf_{m\to\infty}\mathbb{E}(\alpha_m)\geq\delta\mathbb{E}(\beta_1)>0\ .$$

The bound (46) is now established for $\kappa=-\delta\mathrm{E}(\beta_1)$. Applying Lemma 1 to $\phi_m(\lambda)$, we conclude that $\lambda\mapsto\bar{\phi}(\lambda)$ is strictly concave for every $\mathbf{Q}_1,\mathbf{Q}_2$ in $\mathcal{C}_1$ ($\mathbf{Q}_1\neq\mathbf{Q}_2$), and so is $\mathbf{Q}\mapsto\bar{I}(\mathbf{Q})$ by Proposition 2.

C. Proof of Proposition 9

Proof: [Proof of (i)] In order to prove that $\lim_m\mathrm{var}(\beta_m)=0$, we shall rely on Poincaré-Nash inequality. We shall use the following decomposition[3] :

$$\frac{\mathbf{C}^{\frac12}}{\sqrt{K+1}}=\mathbf{U}\mathbf{D}^{\frac12}\mathbf{U}^H;\qquad \tilde{\mathbf{C}}^{\frac12}=\tilde{\mathbf{U}}\tilde{\mathbf{D}}^{\frac12}\tilde{\mathbf{U}}^H.$$

In particular, $\mathbf{H}$ writes

$$\begin{aligned}\mathbf{U}^H\mathbf{H}\tilde{\mathbf{U}} &= \sqrt{\frac{K}{K+1}}\mathbf{U}^H\mathbf{A}\tilde{\mathbf{U}}+\mathbf{D}^{\frac12}\frac{\mathbf{U}^H\mathbf{W}\tilde{\mathbf{U}}}{\sqrt{t}}\tilde{\mathbf{D}}^{\frac12}\\ &\triangleq \mathbf{B}+\mathbf{D}^{\frac12}\frac{\mathbf{X}}{\sqrt{t}}\tilde{\mathbf{D}}^{\frac12}\ \triangleq\ \mathbf{B}+\mathbf{Y}\\ &\triangleq \mathbf{\Sigma}\ ,\end{aligned}$$

[3]Note that the notations introduced hereafter slightly differ from those introduced in Section III-B but this should not disturb the reader.

where $\mathbf{X}$ is a $r \times t$ matrix with i.i.d. $\mathcal{C}N(0,1)$ entries. Consider now the following matrices :

$$\check{\mathbf{B}} = \mathbf{I}_m \otimes \mathbf{B}, \quad \mathbf{\Gamma} = \mathbf{I}_m \otimes \mathbf{D}, \quad \tilde{\mathbf{\Gamma}} = \mathbf{I}_m \otimes \tilde{\mathbf{D}}, \quad \mathbf{V} = \mathbf{I}_m \otimes \mathbf{U}, \quad \tilde{\mathbf{V}} = \mathbf{I}_m \otimes \tilde{\mathbf{U}}.$$

Similarly, $\check{\mathbf{H}}$ writes :

$$\mathbf{V}^H \check{\mathbf{H}} \tilde{\mathbf{V}} = \check{\mathbf{B}} + \mathbf{\Gamma}^{\frac{1}{2}} \frac{\check{\mathbf{X}}}{\sqrt{mt}} \tilde{\mathbf{\Gamma}}^{\frac{1}{2}} \quad \stackrel{\triangle}{=} \quad \check{\mathbf{B}} + \check{\mathbf{Y}} \quad \stackrel{\triangle}{=} \quad \check{\mathbf{\Sigma}} \ ,$$

where $\check{\mathbf{X}}$ is a $mr \times mt$ matrix with i.i.d. $\mathcal{C}N(0,1)$ entries. Denote by $\mathbf{\Theta} = \tilde{\mathbf{U}}^H(\mathbf{Q}_1 - \mathbf{Q}_2)\tilde{\mathbf{U}}$ and by $\check{\mathbf{\Theta}} = \tilde{\mathbf{V}}^H(\check{\mathbf{Q}}_1 - \check{\mathbf{Q}}_2)\tilde{\mathbf{V}}(= \mathbf{I}_m \otimes \mathbf{\Theta})$. The quantity β_m writes then : $\beta_m = \frac{1}{\sigma^4 m} \mathrm{Tr}\, \check{\mathbf{\Theta}} \check{\mathbf{\Sigma}}^H \check{\mathbf{\Sigma}} \check{\mathbf{\Theta}} \check{\mathbf{\Sigma}}^H \check{\mathbf{\Sigma}}$. Considering β_m as a function of the entries of $\check{\mathbf{X}} = (\check{X}_{ij})$, i.e. $\beta_m = \phi(\check{\mathbf{X}})$, standard computations yield

$$\frac{\partial \phi(\check{\mathbf{X}})}{\partial \check{X}_{ij}} = \frac{2}{m} \left(\check{\mathbf{\Theta}} \check{\mathbf{\Sigma}}^H \check{\mathbf{\Sigma}} \check{\mathbf{\Theta}} \check{\mathbf{\Sigma}}^H \right)_{ji} \ .$$

Poincaré-Nash inequality yields then

$$\begin{aligned}
\mathrm{var}\,(\beta_m) &\leq \frac{1}{mt} \sum_{i,j} \mathbf{\Gamma}_i \tilde{\mathbf{\Gamma}}_j \mathbb{E} \left| \frac{\partial \phi(\check{\mathbf{X}})}{\partial \check{X}_{ij}} \right|^2 \\
&= \frac{1}{mt} \sum_{i,j} \mathbf{\Gamma}_i \tilde{\mathbf{\Gamma}}_j \frac{4}{m^2 t^2} \mathbb{E} \left| \left(\check{\mathbf{\Theta}} \check{\mathbf{\Sigma}}^H \check{\mathbf{\Sigma}} \check{\mathbf{\Theta}} \check{\mathbf{\Sigma}}^H \right)_{ji} \right|^2 \\
&\leq \frac{4 d_{\max} \tilde{d}_{\max}}{m^3 t^3} \mathbb{E}\mathrm{Tr}\, \left(\check{\mathbf{\Theta}} \check{\mathbf{\Sigma}}^H \check{\mathbf{\Sigma}} \check{\mathbf{\Theta}} \check{\mathbf{\Sigma}}^H \check{\mathbf{\Sigma}} \check{\mathbf{\Theta}}^H \check{\mathbf{\Sigma}}^H \check{\mathbf{\Sigma}} \check{\mathbf{\Theta}}^H \right) \\
&\leq \frac{4 d_{\max} \tilde{d}_{\max}}{m^2 t^2} \|\check{\mathbf{\Theta}}^H \check{\mathbf{\Theta}}\| \mathbb{E} \left(\frac{1}{mt} \mathrm{Tr} \check{\mathbf{\Sigma}}^H \check{\mathbf{\Sigma}} \check{\mathbf{\Theta}} \check{\mathbf{\Sigma}}^H \check{\mathbf{\Sigma}} \check{\mathbf{\Theta}}^H \check{\mathbf{\Sigma}}^H \check{\mathbf{\Sigma}} \right) \ .
\end{aligned}$$

Moreover, Schwarz inequality yields

$$\frac{1}{mt} \mathrm{Tr} \check{\mathbf{\Sigma}}^H \check{\mathbf{\Sigma}} \check{\mathbf{\Theta}}^H (\check{\mathbf{\Sigma}}^H \check{\mathbf{\Sigma}})^2 \check{\mathbf{\Theta}} \leq \left[\frac{1}{mt} \mathrm{Tr} (\check{\mathbf{\Sigma}}^H \check{\mathbf{\Sigma}})^2 \right]^{1/2} \left[\frac{1}{mt} \mathrm{Tr} \left(\check{\mathbf{\Theta}}^H (\check{\mathbf{\Sigma}}^H \check{\mathbf{\Sigma}})^2 \check{\mathbf{\Theta}} \check{\mathbf{\Theta}}^H (\check{\mathbf{\Sigma}}^H \check{\mathbf{\Sigma}})^2 \check{\mathbf{\Theta}} \right) \right]^{1/2}$$

so that

$$\frac{1}{mt} \mathrm{Tr} \check{\mathbf{\Sigma}}^H \check{\mathbf{\Sigma}} \check{\mathbf{\Theta}}^H \check{\mathbf{\Sigma}}^H \check{\mathbf{\Sigma}} \check{\mathbf{\Theta}} \leq \|\check{\mathbf{\Theta}}^H \check{\mathbf{\Theta}}\| \left[\frac{1}{mt} \mathrm{Tr} (\check{\mathbf{\Sigma}}^H \check{\mathbf{\Sigma}})^2 \right]^{1/2} \left[\frac{1}{mt} \mathrm{Tr} (\check{\mathbf{\Sigma}}^H \check{\mathbf{\Sigma}})^4 \right]^{1/2} \ .$$

Schwarz inequality yields then

$$\mathbb{E} \left(\frac{1}{mt} \mathrm{Tr} \check{\mathbf{\Sigma}}^H \check{\mathbf{\Sigma}} \check{\mathbf{\Theta}}^H \check{\mathbf{\Sigma}}^H \check{\mathbf{\Sigma}} \check{\mathbf{\Theta}} \right) \leq \|\check{\mathbf{\Theta}}^H \check{\mathbf{\Theta}}\| \left[\mathbb{E} \left(\frac{1}{mt} \mathrm{Tr} (\check{\mathbf{\Sigma}}^H \check{\mathbf{\Sigma}})^2 \right) \right]^{1/2} \left[\mathbb{E} \left(\frac{1}{mt} \mathrm{Tr} (\check{\mathbf{\Sigma}}^H \check{\mathbf{\Sigma}})^4 \right) \right]^{1/2} \ .$$

It is tedious, but straightforward, to check that

$$\sup_m \mathbb{E} \left(\frac{1}{mt} \mathrm{Tr} (\check{\mathbf{\Sigma}}^H \check{\mathbf{\Sigma}})^2 \right) < +\infty$$

and

$$\sup_m \mathbb{E} \left(\frac{1}{mt} \mathrm{Tr} (\check{\mathbf{\Sigma}}^H \check{\mathbf{\Sigma}})^4 \right) < +\infty$$

which, in turn, imply that $\mathrm{var}(\beta_m) = O(\frac{1}{m^2})$. ∎

Proof: [Proof of (ii)] Write $\mathbb{E}\,\beta_m$ as

$$\begin{aligned}
\mathbb{E}\,\beta_m &= \frac{1}{\sigma^4 m}\mathbb{E}\,\mathrm{Tr}\,\check{\mathbf{\Sigma}}^H\check{\mathbf{\Sigma}}\check{\mathbf{\Theta}}\check{\mathbf{\Sigma}}^H\check{\mathbf{\Sigma}}\check{\mathbf{\Theta}} \\
&= \frac{1}{\sigma^4 m}\mathbb{E}\,\mathrm{Tr}\,\left(\check{\mathbf{B}}^H\check{\mathbf{B}}+\check{\mathbf{B}}^H\check{\mathbf{Y}}+\check{\mathbf{Y}}^H\check{\mathbf{B}}+\check{\mathbf{Y}}^H\check{\mathbf{Y}}\right)\check{\mathbf{\Theta}}\left(\check{\mathbf{B}}^H\check{\mathbf{B}}+\check{\mathbf{B}}^H\check{\mathbf{Y}}+\check{\mathbf{Y}}^H\check{\mathbf{B}}+\check{\mathbf{Y}}^H\check{\mathbf{Y}}\right)\check{\mathbf{\Theta}} \\
&\overset{(a)}{=} \frac{1}{\sigma^4 m}\,\mathrm{Tr}\,\check{\mathbf{B}}^H\check{\mathbf{B}}\check{\mathbf{\Theta}}\check{\mathbf{B}}^H\check{\mathbf{B}}\check{\mathbf{\Theta}} \quad + \quad \frac{1}{\sigma^4 m}\,\mathbb{E}\,\mathrm{Tr}\,\check{\mathbf{B}}^H\check{\mathbf{B}}\check{\mathbf{\Theta}}\check{\mathbf{Y}}^H\check{\mathbf{Y}}\check{\mathbf{\Theta}} \\
&\quad +\frac{1}{\sigma^4 m}\,\mathbb{E}\,\mathrm{Tr}\,\check{\mathbf{B}}^H\check{\mathbf{Y}}\check{\mathbf{\Theta}}\check{\mathbf{Y}}^H\check{\mathbf{B}}\check{\mathbf{\Theta}} \quad + \quad \frac{1}{\sigma^4 m}\,\mathbb{E}\,\mathrm{Tr}\,\check{\mathbf{Y}}^H\check{\mathbf{B}}\check{\mathbf{\Theta}}\check{\mathbf{B}}^H\check{\mathbf{Y}}\check{\mathbf{\Theta}} \\
&\quad +\frac{1}{\sigma^4 m}\,\mathbb{E}\,\mathrm{Tr}\,\check{\mathbf{Y}}^H\check{\mathbf{Y}}\check{\mathbf{\Theta}}\check{\mathbf{B}}^H\check{\mathbf{B}}\check{\mathbf{\Theta}} \quad + \quad \frac{1}{\sigma^4 m}\,\mathbb{E}\,\mathrm{Tr}\,\check{\mathbf{Y}}^H\check{\mathbf{Y}}\check{\mathbf{\Theta}}\check{\mathbf{Y}}^H\check{\mathbf{Y}}\check{\mathbf{\Theta}}\ ,
\end{aligned}$$

where (a) follows from the fact that the terms where $\check{\mathbf{Y}}$ appears one or three times are readily zero, and so are the terms like $\mathbb{E}\,\mathrm{Tr}\,\check{\mathbf{B}}^H\check{\mathbf{Y}}\check{\mathbf{\Theta}}\check{\mathbf{B}}^H\check{\mathbf{Y}}\check{\mathbf{\Theta}}$. Therefore, it remains to compute the following four terms :

$$\begin{aligned}
T_1 &\triangleq \frac{1}{m}\mathrm{Tr}\,\check{\mathbf{B}}^H\check{\mathbf{B}}\check{\mathbf{\Theta}}\check{\mathbf{B}}^H\check{\mathbf{B}}\check{\mathbf{\Theta}}\ , \\
T_2 &\triangleq \frac{1}{m}\mathbb{E}\,\mathrm{Tr}\,\check{\mathbf{B}}^H\check{\mathbf{B}}\check{\mathbf{\Theta}}\check{\mathbf{Y}}^H\check{\mathbf{Y}}\check{\mathbf{\Theta}}\ , \\
T_3 &\triangleq \frac{1}{m}\mathbb{E}\,\mathrm{Tr}\,\check{\mathbf{B}}^H\check{\mathbf{Y}}\check{\mathbf{\Theta}}\check{\mathbf{Y}}^H\check{\mathbf{B}}\check{\mathbf{\Theta}}\ , \\
T_4 &\triangleq \frac{1}{m}\mathbb{E}\,\mathrm{Tr}\,\check{\mathbf{Y}}^H\check{\mathbf{Y}}\check{\mathbf{\Theta}}\check{\mathbf{Y}}^H\check{\mathbf{Y}}\check{\mathbf{\Theta}}\ .
\end{aligned}$$

Due to the block nature of the matrices involved, $T_1 = \mathrm{Tr}\,\mathbf{B}^H\mathbf{B}\mathbf{\Theta}\mathbf{B}^H\mathbf{B}\mathbf{\Theta}$; in particular, T_1 does not depend on m. Let us now compute T_2. We have $T_2 = m^{-1}\mathrm{Tr}\,\check{\mathbf{B}}^H\check{\mathbf{B}}\check{\mathbf{\Theta}}\mathbb{E}\left(\check{\mathbf{Y}}^H\check{\mathbf{Y}}\right)\check{\mathbf{\Theta}}$ and $\mathbb{E}\left(\check{\mathbf{Y}}^H\check{\mathbf{Y}}\right) = (mt)^{-1}\tilde{\mathbf{\Gamma}}^{\frac12}\mathbb{E}\left(\check{\mathbf{X}}\mathbf{\Gamma}\check{\mathbf{X}}\right)\tilde{\mathbf{\Gamma}}^{\frac12} = (mt)^{-1}\mathrm{Tr}(\mathbf{\Gamma})\tilde{\mathbf{\Gamma}}$. Therefore, T_2 writes :

$$T_2 = \frac{1}{m}\mathrm{Tr}\,(\mathbf{\Gamma})\,\frac{1}{mt}\mathrm{Tr}\,\left(\check{\mathbf{B}}^H\check{\mathbf{B}}\check{\mathbf{\Theta}}\tilde{\mathbf{\Gamma}}\check{\mathbf{\Theta}}\right) = \mathrm{Tr}\,(\mathbf{D})\,\frac{1}{t}\mathrm{Tr}\,\left(\mathbf{B}^H\mathbf{B}\mathbf{\Theta}\tilde{\mathbf{D}}\mathbf{\Theta}\right),$$

and this quantity does not depend on m. We now turn to the term T_3. We have $T_3 = m^{-1}\mathrm{Tr}\,\check{\mathbf{B}}^H\mathbb{E}\,\left(\check{\mathbf{Y}}\check{\mathbf{\Theta}}\check{\mathbf{Y}}^H\right)\check{\mathbf{B}}\check{\mathbf{\Theta}}$. The same computations as before yield $\mathbb{E}\,\left(\check{\mathbf{Y}}\check{\mathbf{\Theta}}\check{\mathbf{Y}}^H\right) = (mt)^{-1}\mathrm{Tr}\,\left(\tilde{\mathbf{\Gamma}}^{\frac12}\check{\mathbf{\Theta}}\tilde{\mathbf{\Gamma}}^{\frac12}\right)\mathbf{\Gamma}$. Therefore T_3 writes :

$$T_3 = \frac{1}{m}\mathrm{Tr}\,\left(\tilde{\mathbf{\Gamma}}^{\frac12}\check{\mathbf{\Theta}}\tilde{\mathbf{\Gamma}}^{\frac12}\right)\frac{1}{mt}\mathrm{Tr}\,\left(\check{\mathbf{B}}^H\mathbf{\Gamma}\check{\mathbf{B}}\check{\mathbf{\Theta}}\right) = \mathrm{Tr}\,\left(\tilde{\mathbf{D}}^{\frac12}\mathbf{\Theta}\tilde{\mathbf{D}}^{\frac12}\right)\frac{1}{t}\mathrm{Tr}\,\left(\mathbf{B}^H\mathbf{D}\mathbf{B}\mathbf{\Theta}\right)\ ,$$

which does not depend on m. It remains to compute $T_4 = \frac{1}{m}\mathrm{Tr}\left[\mathbb{E}\left(\check{\mathbf{Y}}^H\check{\mathbf{Y}}\check{\mathbf{\Theta}}\check{\mathbf{Y}}^H\check{\mathbf{Y}}\right)\check{\mathbf{\Theta}}\right]$.

$$\mathbb{E}\left(\check{\mathbf{Y}}^H\check{\mathbf{Y}}\check{\mathbf{\Theta}}\check{\mathbf{Y}}^H\check{\mathbf{Y}}\right) = \frac{1}{(mt)^2}\tilde{\mathbf{\Gamma}}^{\frac12}\mathbb{E}\left(\check{\mathbf{X}}\mathbf{\Gamma}\check{\mathbf{X}}\tilde{\mathbf{\Gamma}}^{\frac12}\check{\mathbf{\Theta}}\tilde{\mathbf{\Gamma}}^{\frac12}\check{\mathbf{X}}\mathbf{\Gamma}\check{\mathbf{X}}\right)\tilde{\mathbf{\Gamma}}^{\frac12}\ .$$

Computing the individual terms of matrix $\mathbb{E}\left(\check{\mathbf{X}}\mathbf{\Gamma}\check{\mathbf{X}}\tilde{\mathbf{\Gamma}}^{\frac12}\check{\mathbf{\Theta}}\tilde{\mathbf{\Gamma}}^{\frac12}\check{\mathbf{X}}\mathbf{\Gamma}\check{\mathbf{X}}\right)$ yields (denote by $\mathbf{G} = \tilde{\mathbf{\Gamma}}^{\frac12}\check{\mathbf{\Theta}}\tilde{\mathbf{\Gamma}}^{\frac12}$ for the sake of simplicity) :

$$\begin{aligned}
\left[\mathbb{E}\left(\check{\mathbf{X}}\mathbf{\Gamma}\check{\mathbf{X}}\mathbf{G}\check{\mathbf{X}}\mathbf{\Gamma}\check{\mathbf{X}}\right)\right]_{k\ell} &= \sum_{i_1,j_1,j_2,i_2}\mathbb{E}\left(\overline{\check{\mathbf{X}}}_{i_1,k}\check{\mathbf{X}}_{i_1,j_1}\overline{\check{\mathbf{X}}}_{i_2,j_2}\check{\mathbf{X}}_{i_2,\ell}\right)\mathbf{\Gamma}_{i_1,i_1}\mathbf{G}_{j_1,j_2}\mathbf{\Gamma}_{i_2,i_2} \\
&= \left(\mathrm{Tr}\,\mathbf{\Gamma}\right)^2\mathbf{G}_{k\ell} + \mathrm{Tr}\,\left(\mathbf{\Gamma}^2\right)\,\mathrm{Tr}\,\mathbf{G}\ \delta_{k\ell}\ ,
\end{aligned}$$

where $\delta_{k\ell}$ stands for the Kronecker symbol (i.e. $\delta_{k\ell} = 1$ if $k = \ell$, and 0 otherwise). This yields

$$\mathbb{E}\left(\check{\mathbf{Y}}^H \check{\mathbf{Y}} \check{\boldsymbol{\Theta}} \check{\mathbf{Y}}^H \check{\mathbf{Y}}\right) = \frac{1}{(mt)^2} \left(\mathrm{Tr}\, \boldsymbol{\Gamma}\right)^2 \tilde{\boldsymbol{\Gamma}} \check{\boldsymbol{\Theta}} \tilde{\boldsymbol{\Gamma}} + \frac{1}{(mt)^2} \mathrm{Tr}\left(\boldsymbol{\Gamma}^2\right) \mathrm{Tr}\left(\check{\boldsymbol{\Theta}} \tilde{\boldsymbol{\Gamma}}\right) \tilde{\boldsymbol{\Gamma}}$$

and

$$\begin{aligned} T_4 &= \frac{1}{t^2} \left(\frac{\mathrm{Tr}\boldsymbol{\Gamma}}{m}\right)^2 \frac{1}{m} \mathrm{Tr}\left(\tilde{\boldsymbol{\Gamma}} \check{\boldsymbol{\Theta}} \tilde{\boldsymbol{\Gamma}} \check{\boldsymbol{\Theta}}\right) + \frac{1}{t^2} \frac{1}{m} \mathrm{Tr}\left(\boldsymbol{\Gamma}^2\right) \frac{1}{m^2} \left(\mathrm{Tr} \check{\boldsymbol{\Theta}} \tilde{\boldsymbol{\Gamma}}\right)^2 \\ &= \frac{1}{t^2} \left(\mathrm{Tr}\mathbf{D}\right)^2 \mathrm{Tr}\left(\tilde{\mathbf{D}} \boldsymbol{\Theta} \tilde{\mathbf{D}} \boldsymbol{\Theta}\right) + \frac{1}{t^2} \mathrm{Tr}\left(\mathbf{D}^2\right) \left(\mathrm{Tr} \boldsymbol{\Theta} \tilde{\mathbf{D}}\right)^2 , \end{aligned}$$

which does not depend on m. This shows that $\mathbb{E}\beta_m$ does not depend on m, and thus coincides with $\mathbb{E}\beta_1$. In order to complete the proof of (ii), it remains to verify that $\mathbb{E}\beta_1 > 0$, or equivalenlty that $\mathbb{E}\beta_1$ is not equal to 0. If $\mathbb{E}\beta_1$ was indeed equal to 0, then, matrix

$$\left(\mathbf{H}^H \mathbf{H}\right)^{1/2} (\mathbf{Q}_1 - \mathbf{Q}_2) \left(\mathbf{H}^H \mathbf{H}\right)^{1/2}$$

or equivalently matrix

$$\mathbf{H}^H \mathbf{H} (\mathbf{Q}_1 - \mathbf{Q}_2)$$

would be equal to zero almost everywhere. As $\mathbf{Q}_1 \neq \mathbf{Q}_2$, it would exist a deterministic non zero vector $\mathbf{x}$ such that $\mathbf{x}^H \mathbf{H}^H \mathbf{H} \mathbf{x} = 0$ almost everywhere, i.e. $\mathbf{H}\mathbf{x} = 0$, or equivalently

$$\mathbf{W} \tilde{\mathbf{C}}^{1/2} \mathbf{x} = -\sqrt{Kt}\, \mathbf{C}^{-1/2} \mathbf{A} \mathbf{x} . \tag{112}$$

As matrix $\tilde{\mathbf{C}}^{1/2}$ is positive definite, vector $\tilde{\mathbf{C}}^{1/2}\mathbf{x}$ is non zero. Relation (112) leads to a contradiction because the joint distribution of the entries of $\mathbf{W}$ is absolutely continuous. This shows that $\mathbb{E}\beta_1 > 0$. The proof of (ii) is complete. ■

Proof: [Proof of (iii)] In order to control $\alpha_m = \frac{1}{\left(1+\sigma^{-2} \|\tilde{\mathbf{Q}}\| \, \|\tilde{\mathbf{H}}^H \tilde{\mathbf{H}}\|\right)^2}$, first notice that $\|\tilde{\mathbf{Q}}\| = \|\mathbf{Q}\|$. Now $\|\tilde{\mathbf{H}}^H \tilde{\mathbf{H}}\| = \|\tilde{\mathbf{H}}\|^2$ and

$$\|\tilde{\mathbf{H}}\| \leq \sqrt{\frac{K}{K+1}} \|\check{\mathbf{A}}\| + \frac{1}{\sqrt{K+1}} \|\boldsymbol{\Delta}^{\frac{1}{2}}\| \, \|\tilde{\boldsymbol{\Delta}}^{\frac{1}{2}}\| \left\| \frac{\check{\mathbf{W}}}{\sqrt{mt}} \right\| .$$

Now $\|\check{\mathbf{A}}\| = \|\mathbf{A}\|$, $\|\boldsymbol{\Delta}^{\frac{1}{2}}\| = \|\mathbf{C}^{\frac{1}{2}}\|$ and $\|\tilde{\boldsymbol{\Delta}}^{\frac{1}{2}}\| = \|\tilde{\mathbf{C}}^{\frac{1}{2}}\|$. The behaviour of the spectral norm of $(mt)^{-\frac{1}{2}} \check{\mathbf{W}}$ is well-known (see for instance [36], [1]) : $\left\|(mt)^{-\frac{1}{2}} \check{\mathbf{W}}\right\| \to_{m\to\infty} 1 + \sqrt{1/c}$ almost surely. Therefore, Fatou's lemma yields the desired result : $\liminf_m \mathbb{E}\alpha_m \geq \delta > 0$, and (iii) is proved. ■

Appendix IV

Proof of Proposition 5, item (i).

By (50) and (51), $(\kappa, \tilde{\kappa}, \mathbf{Q}) \mapsto V(\kappa, \tilde{\kappa}, \mathbf{Q})$ is differentiable from $\mathbb{R}^+ \times \mathbb{R}^+ \times \mathcal{C}_1$ to $\mathbb{R}$. In order to prove that $\bar{I}(\mathbf{Q}) = V(\delta(\mathbf{Q}), \tilde{\delta}(\mathbf{Q}), \mathbf{Q})$ is differentiable, it is sufficient to prove the differentiability of $\delta, \tilde{\delta} : \mathcal{C}_1 \to \mathbb{R}$. Recall that δ and $\tilde{\delta}$ are solution of system (33) associated with matrix $\mathbf{Q}$. In order to apply the implicit function theorem, which will immediatly yield the differentiablity of δ and $\tilde{\delta}$ with respect to $\mathbf{Q}$, we must check that :

1) The function

$$(\delta, \tilde{\delta}, \mathbf{Q}) \mapsto \Upsilon(\delta, \tilde{\delta}, \mathbf{Q}) = \begin{pmatrix} \delta - f(\delta, \tilde{\delta}, \mathbf{Q}) \\ \tilde{\delta} - \tilde{f}(\delta, \tilde{\delta}, \mathbf{Q}) \end{pmatrix}$$

is differentiable.

2) The partial jacobian

$$D_{(\delta,\tilde{\delta})}\Upsilon(\delta, \tilde{\delta}, \mathbf{Q}) = \begin{pmatrix} 1 - \frac{\partial f}{\partial \delta}(\delta, \tilde{\delta}, \mathbf{Q}) & -\frac{\partial f}{\partial \tilde{\delta}}(\delta, \tilde{\delta}, \mathbf{Q}) \\ -\frac{\partial \tilde{f}}{\partial \delta}(\delta, \tilde{\delta}, \mathbf{Q}) & 1 - \frac{\partial \tilde{f}}{\partial \tilde{\delta}}(\delta, \tilde{\delta}, \mathbf{Q}) \end{pmatrix}$$

is invertible for every $\mathbf{Q} \in \mathcal{C}_1$.

In order to check the differentiability of Υ, recall the following matrix equality

$$(\mathbf{I} + \mathbf{U}\mathbf{V})^{-1}\mathbf{U} = \mathbf{U}(\mathbf{I} + \mathbf{V}\mathbf{U})^{-1} \tag{113}$$

which follows from elementary matrix manipulations (cf. [20, Section 0.7.4]). Applying this equality to $\mathbf{U} = \mathbf{Q}^{\frac{1}{2}}$ and $V = \delta\tilde{\mathbf{C}}\mathbf{Q}^{\frac{1}{2}}$, we obtain :

$$\mathbf{A}\mathbf{Q}^{\frac{1}{2}}\left(\mathbf{I} + \delta\mathbf{Q}^{\frac{1}{2}}\tilde{\mathbf{C}}\mathbf{Q}^{\frac{1}{2}}\right)^{-1}\mathbf{Q}^{\frac{1}{2}}\mathbf{A}^H = \mathbf{A}\mathbf{Q}\left(\mathbf{I} + \delta\tilde{\mathbf{C}}\mathbf{Q}\right)^{-1}\mathbf{A}^H$$

which yields

$$f(\delta, \tilde{\delta}, \mathbf{Q}) = \frac{1}{t}\mathrm{Tr}\left\{\mathbf{C}\left[\sigma^2\left(\mathbf{I}_r + \frac{\tilde{\delta}}{K+1}\mathbf{C}\right) + \frac{K}{K+1}\mathbf{A}\mathbf{Q}\left(\mathbf{I}_t + \frac{\delta}{K+1}\tilde{\mathbf{C}}\mathbf{Q}\right)^{-1}\mathbf{A}^H\right]^{-1}\right\}.$$

Clearly, f is differentiable with respect to the three variables $\delta, \tilde{\delta}$ and $\mathbf{Q}$. Similar computations yield

$$\tilde{f}(\delta, \tilde{\delta}, \mathbf{Q}) = \frac{1}{t}\,\mathrm{Tr}\left\{\mathbf{Q}\tilde{\mathbf{C}}\left[\sigma^2\left(\mathbf{I}_t + \frac{\delta}{K+1}\tilde{\mathbf{C}}\mathbf{Q}\right) + \frac{K}{K+1}\mathbf{A}^H\left(\mathbf{I}_r + \frac{\tilde{\delta}}{K+1}\mathbf{C}\right)^{-1}\mathbf{A}\mathbf{Q}\right]^{-1}\right\},$$

and the same conclusion holds for $\tilde{f}$. Therefore, $(\delta, \tilde{\delta}, \mathbf{Q}) \mapsto \Upsilon(\delta, \tilde{\delta}, \mathbf{Q})$ is differentiable and 1) is proved.

In order to study the jacobian $D_{(\delta,\tilde{\delta})}\Upsilon$, let us compute first $\frac{\partial f}{\partial \delta}$.

$$\begin{aligned}
\frac{\partial f}{\partial \delta}(\delta, \tilde{\delta}, \mathbf{Q}) &= \frac{1}{t}\mathrm{Tr}\ \mathbf{C}\mathbf{T}_K\mathbf{A}\mathbf{Q}\left(\mathbf{I} + \frac{\delta}{K+1}\tilde{\mathbf{C}}\mathbf{Q}\right)^{-1}\frac{\tilde{\mathbf{C}}\mathbf{Q}}{K+1}\left(\mathbf{I} + \frac{\delta}{K+1}\tilde{\mathbf{C}}\mathbf{Q}\right)^{-1}\mathbf{A}^H\mathbf{T}_K\frac{K}{K+1}, \\
&= \frac{1}{t}\mathrm{Tr}\ \mathbf{C}\mathbf{T}_K\mathbf{A}\mathbf{Q}^{\frac{1}{2}}\left(\mathbf{I} + \frac{\delta}{K+1}\mathbf{Q}^{\frac{1}{2}}\tilde{\mathbf{C}}\mathbf{Q}^{\frac{1}{2}}\right)^{-1}\frac{\mathbf{Q}^{\frac{1}{2}}\tilde{\mathbf{C}}\mathbf{Q}^{\frac{1}{2}}}{K+1} \\
&\quad \times\frac{\mathbf{Q}^{\frac{1}{2}}\tilde{\mathbf{C}}\mathbf{Q}^{\frac{1}{2}}}{K+1}\left(\mathbf{I} + \frac{\delta}{K+1}\mathbf{Q}^{\frac{1}{2}}\tilde{\mathbf{C}}\mathbf{Q}^{\frac{1}{2}}\right)^{-1}\mathbf{Q}^{\frac{1}{2}}\mathbf{A}^H\mathbf{T}_K\frac{K}{K+1}, \\
&\overset{(a)}{=} \frac{1}{t}\mathrm{Tr}\ (\mathbf{D}\mathbf{T}\mathbf{B}(\mathbf{I} + \beta\tilde{\mathbf{D}})^{-1}\tilde{\mathbf{D}}(\mathbf{I} + \beta\tilde{\mathbf{D}})^{-1}\mathbf{B}^H\mathbf{T})
\end{aligned}$$

where (a) follows from the virtual channel equivalences (31), (32) together with (39) and (41). Finally, we end up with the following :

$$1-\frac{\partial f}{\partial \delta}(\delta,\tilde{\delta},\mathbf{Q}) = 1-\frac{1}{t}\mathrm{Tr}(\mathbf{DTB}(\mathbf{I}+\beta\tilde{\mathbf{D}})^{-1}\tilde{\mathbf{D}}(\mathbf{I}+\beta\tilde{\mathbf{D}})^{-1}\mathbf{B}^H\mathbf{T}) \ .$$

Similar computations yield

$$\begin{aligned}
1-\frac{\partial \tilde{f}}{\partial \tilde{\delta}}(\delta,\tilde{\delta},\mathbf{Q}) &= 1-\frac{1}{t}\mathrm{Tr}(\tilde{\mathbf{D}}\tilde{\mathbf{T}}\mathbf{B}^H(\mathbf{I}+\tilde{\beta}\mathbf{D})^{-1}\mathbf{D}(\mathbf{I}+\tilde{\beta}\mathbf{D})^{-1}\mathbf{B}\tilde{\mathbf{T}}) \ , \\
-\frac{\partial f}{\partial \tilde{\delta}}(\delta,\tilde{\delta},\mathbf{Q}) &= \frac{\sigma^2}{t}\mathrm{Tr}\ (\mathbf{DTDT}) \ , \\
-\frac{\partial \tilde{f}}{\partial \delta}(\delta,\tilde{\delta},\mathbf{Q}) &= \frac{\sigma^2}{t}\mathrm{Tr}\ (\tilde{\mathbf{D}}\tilde{\mathbf{T}}\tilde{\mathbf{D}}\tilde{\mathbf{T}}) \ .
\end{aligned}$$

The invertibility of the jacobian $D_{(\delta,\tilde{\delta})}\Upsilon$ follows then from Lemma 4 in Appendix II-C and 2) is proved. In particular, we can assert that $\mathcal{C}_1 \ni \mathbf{Q} \mapsto \delta(\mathbf{Q})$ and $\mathcal{C}_1 \ni \mathbf{Q} \mapsto \tilde{\delta}(\mathbf{Q})$ are differentiable due to the Implicit function theorem. Item (i) is proved.

Appendix V
Proof of Proposition 6

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

$$< \nabla \bar{I}(\mathbf{Q}_*^{\psi}), \mathbf{Q}-\mathbf{Q}_*^{\psi} > \leq 0 \tag{114}$$

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

In order to prove (114), consider the iteration $\psi(k)$ of the algorithm. The matrix $\mathbf{Q}_{\psi(k)}$ maximizes the function $\mathbf{Q} \mapsto V(\delta_{\psi(k)},\tilde{\delta}_{\psi(k)},\mathbf{Q})$. As this function is strictly concave andd differentiable, Proposition 4 implies that

$$< \nabla_{\mathbf{Q}} V(\delta_{\psi(k)},\tilde{\delta}_{\psi(k)},\mathbf{Q}_{\psi(k)}), \mathbf{Q}-\mathbf{Q}_{\psi(k)} > \leq 0 \tag{115}$$

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 $(\delta_{\psi(k)+1},\tilde{\delta}_{\psi(k)+1})$ of the system (33) associated with matrix $\mathbf{Q}_{\psi(k)}$.

Due to the continuity of $\delta(\mathbf{Q})$ and $\tilde{\delta}(\mathbf{Q})$, the convergence of the subsequence $\mathbf{Q}_{\psi(k)}$ implies the convergence of the subsequences $(\delta_{\psi(k)+1},\tilde{\delta}_{\psi(k)+1})$ towards a limit $(\delta_*^{\psi},\tilde{\delta}_*^{\psi})$. The pair

$(\delta_*^\psi, \tilde{\delta}_*^\psi)$ is the solution of system (33) associated with $\mathbf{Q}_*^\psi$ i.e. $\delta_*^\psi = \delta(\mathbf{Q}_*^\psi)$ and $\tilde{\delta}_*^\psi = \tilde{\delta}(\mathbf{Q}_*^\psi)$; in particular :

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

(see for instance (56)). Using the same computation as in the proof of Proposition 5, we obtain

$$\langle \nabla \bar{I}(\mathbf{Q}_*^\psi), \mathbf{Q} - \mathbf{Q}_*^\psi \rangle = \langle \nabla V \left(\delta_*^\psi, \tilde{\delta}_*^\psi, \mathbf{Q}_*^\psi \right), \mathbf{Q} - \mathbf{Q}_*^\psi \rangle \tag{116}$$

for every $\mathbf{Q} \in \mathcal{C}_1$. Now condition (57) implies that the subsequence $(\delta_{\psi(k)}, \tilde{\delta}_{\psi(k)})$ also converges toward $(\delta_*^\psi, \tilde{\delta}_*^\psi)$. As a consequence,

$$\lim_{k \to +\infty} \langle \nabla V(\delta_{\psi(k)}, \tilde{\delta}_{\psi(k)}, \mathbf{Q}_{\psi(k)}), \mathbf{Q} - \mathbf{Q}_{\psi(k)} \rangle = \langle \nabla V(\delta_*^\psi, \tilde{\delta}_*^\psi, \mathbf{Q}_*^\psi), \mathbf{Q} - \mathbf{Q}_*^\psi \rangle \ .$$

Inequality (115) thus implies that $\langle \nabla V(\delta_*^\psi, \tilde{\delta}_*^\psi, \mathbf{Q}_*^\psi), \mathbf{Q} - \mathbf{Q}_*^\psi \rangle \leq 0$ and relation (116) allows us to conclude the proof.

APPENDIX VI
END OF PROOF OF PROPOSITION 3

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

$$A = \max\left(\sup_t \|\mathbf{A}\|, \sup_t \|\tilde{\mathbf{C}}\|, \sup_t \|\mathbf{C}\| \right) < \infty \quad \text{and} \quad a = \min\left(\inf_t \lambda_{\min}(\tilde{\mathbf{C}}), \inf_t \lambda_{\min}(\mathbf{C}) \right) > 0 \ .$$

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

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

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

$$t = \sum_j \lambda_j(\overline{\mathbf{Q}}_*) = \gamma t - \sum_j \frac{1}{\lambda_j(\mathbf{G})} \ ,$$

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

$$\|\overline{\mathbf{Q}}_*\| \leq 1 + \frac{1}{\lambda_{\min}(\mathbf{G})} \ . \tag{117}$$

It remains to prove

$$\forall\ \mathbf{Q} \in \mathcal{C}_1, \quad \inf_t \lambda_{\min}\left(\mathbf{G}(\delta(\mathbf{Q}), \tilde{\delta}(\mathbf{Q})) \right) > 0 \tag{118}$$

and we are done. To this end, we first show that $\inf_t \delta(\mathbf{Q}) > 0$ for all $\mathbf{Q} \in \mathcal{C}_1$. From Equations (40) and (42), we have :

$$\begin{aligned}
\delta(\mathbf{Q}) &= \frac{1}{t}\mathrm{tr}\mathbf{C}\mathbf{T}_K(\sigma^2) \\
&\geq \lambda_{\min}(\mathbf{C})\frac{1}{t}\mathrm{tr}\mathbf{T}_K(\sigma^2) \\
&\overset{(a)}{\geq} \lambda_{\min}(\mathbf{C})\left[\frac{1}{t}\mathrm{tr}\left(\sigma^2\mathbf{I}_r + \frac{\sigma^2}{K+1}\tilde{\delta}\tilde{\mathbf{C}}\right.\right. \\
&\quad \left.\left. + \frac{K}{K+1}\mathbf{A}\mathbf{Q}^{1/2}\left(\mathbf{I}_t + \frac{\delta}{K+1}\mathbf{Q}^{1/2}\tilde{\mathbf{C}}\mathbf{Q}^{1/2}\right)^{-1}\mathbf{Q}^{1/2}\mathbf{A}^H\right)\right]^{-1} \\
&\overset{(b)}{\geq} \lambda_{\min}(\mathbf{C})\left(\frac{1}{t}\mathrm{tr}\left(\sigma^2\mathbf{I}_r + \frac{\sigma^2}{K+1}\tilde{\delta}\mathbf{C} + \frac{K}{K+1}\mathbf{A}\mathbf{Q}\mathbf{A}^H\right)\right)^{-1} \qquad (119)
\end{aligned}$$

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 $\tilde{\delta}$. From (41) and (13), we have $\|\tilde{\mathbf{T}}_K(\sigma^2)\| \leq 1/\sigma^2$. Using (42) we then have

$$\tilde{\delta} \leq \|\tilde{\mathbf{T}}_K\|\frac{1}{t}\mathrm{tr}\tilde{\mathbf{C}}\mathbf{Q} \leq \|\tilde{\mathbf{T}}_K\|\|\tilde{\mathbf{C}}\|\frac{1}{t}\mathrm{tr}\mathbf{Q} \leq \frac{A}{\sigma^2}$$

(recall that $\frac{1}{t}\mathrm{tr}\,\mathbf{Q} = 1$). Getting back to (119), we easily obtain

$$\frac{1}{t}\mathrm{tr}\left(\sigma^2\mathbf{I}_r + \frac{\sigma^2}{K+1}\tilde{\delta}\mathbf{C} + \frac{K}{K+1}\mathbf{A}\mathbf{Q}\mathbf{A}^H\right) \leq \frac{r}{t}\left(\sigma^2 + \frac{A}{K+1}\right) + \frac{A^2K}{K+1} \leq C_0 \quad \forall(t,r),\ \frac{t}{r} \to c$$

where C_0 is a certain constant term. Hence we have $\delta(\mathbf{Q}) \geq aC_0^{-1}$. By inspecting the expression (50) of $\mathbf{G}(\delta, \tilde{\delta})$, we then obtain

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

and (118) is proven. It remains to plug this estimate into (117) 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

$$\frac{\partial I(\mathbf{Q})}{\partial q_j} = \frac{\partial}{\partial q_j}\mathbb{E}\left[\mathcal{I}(\mathbf{Q})\right] = \eta - \beta_j \qquad (120)$$

where $\mathcal{I}(\mathbf{Q}) = \log\det\left(\mathbf{I}_r + \frac{1}{\sigma^2}\mathbf{H}\mathbf{Q}\mathbf{H}^H\right)$ 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 j^{th} column of $\mathbf{H}$. By consequence, $\mathbb{E}\left[\partial \mathcal{I}(\mathbf{Q})/\partial q_j\right] \leq \frac{1}{\sigma^2}\mathbb{E}\left[\|\mathbf{h}_j\|^2\right]$. As $\mathbf{h}_j$ is a Gaussian vector, the righthand side of this inequality is defined and therefore, by the Dominated Convergence Theorem, we can exchange $\partial/\partial q_j$ with $\mathbb{E}$ in Equation (120) and write

$$\frac{\partial I(\mathbf{Q})}{\partial q_j} = \frac{1}{\sigma^2}\mathbb{E}\left[\mathbf{h}_j^H\left(\mathbf{I}_r + \frac{1}{\sigma^2}\mathbf{H}\mathbf{Q}\mathbf{H}^H\right)^{-1}\mathbf{h}_j\right] \tag{121}$$

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 $\mathbf{R}_j = \left(\mathbf{I}_r + \frac{1}{\sigma^2}\mathbf{H}_j\mathbf{Q}_j\mathbf{H}_j^H\right)^{-1}$, we have by the Matrix Inversion Lemma ([20, §0.7.4])

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

By plugging this expression into the righthand side of Equation (121), the Lagrange-KKT conditions become

$$\mathbb{E}\left[\frac{X_j}{\sigma^2 + q_jX_j}\right] = \eta - \beta_j \tag{122}$$

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 $\mathbb{E}\left[\frac{X_j}{\sigma^2+q_jX_j}\right] < \eta$, therefore $\beta_j > 0$ (122), 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

$$\max_{j=1,\ldots,t}\mathbb{P}\left(X_j \leq C\right) \xrightarrow[t\to\infty]{} 0 \ . \tag{123}$$

Indeed, let us admit (123) temporarily. We have

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

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

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

by (123), 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 (122) we therefore have for t large enough $\eta-\beta_j > \frac{C/2}{\sigma^2+q_jC/2}$ for every $j=1,\dots,t$. 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

$$\sup_t \|\mathbf{Q}_*^{\mathrm{d}}\| < \frac{2\sigma^2}{C} + 1 \tag{124}$$

which is the desired result. To prove (123), we make use of MMSE estimation theory. Recall that $\mathbf{H} = \sqrt{\frac{K}{K+1}}\mathbf{A} + \frac{1}{\sqrt{K+1}}\frac{1}{\sqrt{t}}\mathbf{C}^{1/2}\mathbf{W}\tilde{\mathbf{C}}^{1/2}$. Denoting by $\mathbf{a}_j$ and $\mathbf{z}_j$ the j^{th} columns of the matrices $\mathbf{A}$ and $\mathbf{W}\tilde{\mathbf{C}}^{1/2}$ respectively, we have

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

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

$$\begin{aligned} S_j &= 2\Re\left(\frac{1}{\sqrt{K+1}}\frac{1}{\sqrt{t}}\mathbf{u}_j^{\perp H}\mathbf{C}^{1/2}\mathbf{R}_j\left(\sqrt{\frac{K}{K+1}}\mathbf{a}_j + \frac{1}{\sqrt{K+1}}\frac{1}{\sqrt{t}}\mathbf{C}^{1/2}\mathbf{u}_j\right)\right) \\ &\quad + \frac{1}{t(K+1)}\mathbf{u}_j^{\perp H}\mathbf{C}^{1/2}\mathbf{R}_j\mathbf{C}^{1/2}\mathbf{u}_j^{\perp} . \end{aligned} \tag{125}$$

Then

$$\begin{aligned} X_j &= S_j + \left(\sqrt{\frac{K}{K+1}}\mathbf{a}_j^H + \frac{1}{\sqrt{K+1}}\frac{1}{\sqrt{t}}\mathbf{u}_j^H\mathbf{C}^{1/2}\right)\mathbf{R}_j\left(\sqrt{\frac{K}{K+1}}\mathbf{a}_j + \frac{1}{\sqrt{K+1}}\frac{1}{\sqrt{t}}\mathbf{C}^{1/2}\mathbf{u}_j\right) \\ &\geq S_j . \end{aligned} \tag{126}$$

Let us study the asymptotic behaviour of S_j. First, we note that due to the fact that the joint distribution of the elements of $\mathbf{W}\tilde{\mathbf{C}}^{1/2}$ is the Gaussian distribution, $\mathbf{u}_j^{\perp}$ and $\mathbf{v}_j = [\mathbf{z}_1^T,\dots,\mathbf{z}_{j-1}^T,\mathbf{z}_{j+1}^T,\dots,\mathbf{z}_t^T]^T$ 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 ([35]), we have $\mathbf{R}_{\mathbf{u}} = \mathbb{E}[\mathbf{z}_j\mathbf{z}_j^H] - \mathbb{E}[\mathbf{z}_j\mathbf{v}_j^H]\left(\mathbb{E}[\mathbf{v}_j\mathbf{v}_j^H]\right)^{-1}\mathbb{E}[\mathbf{v}_j\mathbf{z}_j^H]$. To obtain $\mathbf{R}_{\mathbf{u}}$, we note that the covariance matrix of the vector $\mathbf{z} = [\mathbf{z}_1^T,\dots,\mathbf{z}_t^T]^T$ is $\mathbb{E}[\mathbf{z}\mathbf{z}^H] = \tilde{\mathbf{C}}^T \otimes \mathbf{I}_r$ (just check that $\mathbb{E}\left[[\mathbf{W}\tilde{\mathbf{C}}^{1/2}]_{ij}\overline{[\mathbf{W}\tilde{\mathbf{C}}^{1/2}]_{kl}}\right] = \delta(i-k)[\tilde{\mathbf{C}}]_{lj}$). Let us denote by $\tilde{c}_j$, $\tilde{\mathbf{c}}_j$ and $\tilde{\mathbf{C}}_j$ the scalar $\tilde{c}_j = [\tilde{\mathbf{C}}]_{jj}$, the j^{th} vector column of $\tilde{\mathbf{C}}$ without element $\tilde{c}_j$, and the $(t-1)\times(t-1)$ matrix that remains after extracting row and column j from $\tilde{\mathbf{C}}$ respectively. With these notations we have $\mathbf{R}_{\mathbf{u}} = \left(\tilde{c}_j - \tilde{\mathbf{c}}_j^H\tilde{\mathbf{C}}_j^{-1}\tilde{\mathbf{c}}_j\right)\mathbf{I}_r$. Recalling that $\mathbf{u}_j^{\perp}$ and $(\mathbf{R}_j,\mathbf{u}_j)$ are independent, one may see that the first term of the righthand side of (125) is negligible while the second is close to $\rho_j = \frac{1}{t}\frac{\tilde{c}_j - \tilde{\mathbf{c}}_j^H \tilde{\mathbf{C}}_j^{-1}\tilde{\mathbf{c}}_j}{K+1}\mathrm{tr}(\mathbf{R}_j\mathbf{C})$. More rigorously, using this independence in addition to $A = \max(\|\mathbf{A}\|, \|\mathbf{C}\|, \|\tilde{\mathbf{C}}\|) < \infty$ 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

$$\mathbb{E}\left[(S_j - \rho_j)^2\right] \leq \frac{C_1}{t} \ . \tag{127}$$

In order to prove (123), we will prove that the ρ_j are bounded away from zero in some sense. First, we have

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

(for (a) see [20, §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,

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

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

$$\liminf_t \min_{j=1,\ldots,t} \rho_j \geq \frac{a^2}{K+1}\left(c^{-1} + \frac{A^2}{\sigma^2}\left(2 + c^{-1/2}\right)^2\right)^{-1} = C_2 \ . \tag{128}$$

Choose the constant C in the lefthand side of (123) as $C = C_2/4$. From (126) we have

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

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 (127) and to (128).

We have proven (123) and hence that $\mathbf{Q}_*^{\mathrm{d}} = \arg\max_{\mathbf{Q}\in\mathcal{C}_1^{\mathrm{d}}} I(\mathbf{Q})$ satisfies $\sup_t \|\mathbf{Q}_*^{\mathrm{d}}\| < \infty$.

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

$$\max_{\mathbf{Q}\in\mathcal{C}_1} I(\mathbf{Q}) = \max_{\mathbf{U}\in\mathcal{U}_t} \max_{\mathbf{\Lambda}\in\mathcal{C}_1^{\mathrm{d}}} \mathbb{E}\left[\log\det\left(\mathbf{I}_r + \frac{1}{\sigma^2}\mathbf{H}\mathbf{U}\mathbf{\Lambda}\mathbf{U}^H\mathbf{H}^H\right)\right] \tag{129}$$

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 (129) 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 $\mathbf{H}' = \mathbf{H}\mathbf{U} = \sqrt{\frac{K}{K+1}}\mathbf{A}' + \frac{1}{\sqrt{K+1}}\frac{1}{\sqrt{t}}\mathbf{C}^{1/2}\mathbf{W}'\tilde{\mathbf{C}}'^{1/2}$. Here, matrix $\tilde{\mathbf{C}}'$ is defined by $\tilde{\mathbf{C}}' = \mathbf{U}^H\tilde{\mathbf{C}}\mathbf{U}$, $\mathbf{A}' = \mathbf{A}\mathbf{U}$, $\mathbf{W}' = \mathbf{W}\mathbf{\Theta}$ where $\mathbf{\Theta}$ is the unitary matrix $\mathbf{\Theta} = \tilde{\mathbf{C}}^{1/2}\mathbf{U}\tilde{\mathbf{C}}'^{-1/2}$. As $\mathbf{U}\in\mathcal{U}_t$, we clearly have $\|\mathbf{A}'\| = \|\mathbf{A}\|$, $\|\tilde{\mathbf{C}}'\| = \|\tilde{\mathbf{C}}\|$, and $\|\tilde{\mathbf{C}}'^{-1}\| = \|\tilde{\mathbf{C}}^{-1}\|$. By consequence, the bounds a and A, and hence the constant C in the left hand member of (123) (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 $\mathbb{E}\left[\log\det\left(\mathbf{I}_r + \frac{1}{\sigma^2}\mathbf{H}\mathbf{U}\mathbf{\Lambda}\mathbf{U}^H\mathbf{H}^H\right)\right]$ satisfies $\|\mathbf{\Lambda}_*(\mathbf{U})\| < 2\sigma^2/C + 1$ (see (124)) 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, J.W. Silverstein, "No eigenvalues outside the support of the limiting spectral distribution of large dimensionnal sample covariance matrices", Ann. of Prob., 26(1) :316-345, 1998.

[2] Z. Bai, J. Silverstein, ”CLT for linear statistics of large-dimensional sample covariance matrices“, Ann. Probab., 32(1A) :553-605, 2004.

[3] E. Biglieri, G. Taricco, A. Tulino, ”How far is infinite ? Using asymptotic analyses in multiple-antennas systems“, Proc. of ISSTA-02, vol. 1, pp. 1-6, 2002.

[4] Borwein, J. and Lewis, A., “Convex analysis and nonlinear optimization”, Springer-Verlag, 2000.

[5] H. Cartan, "Théorie Elementaire des Fonctions Analytiques d’une ou Plusieurs Variables Complexes", Hermann, 1978.

[6] S. Chatterjee, A. Bose, "A new method for bounding rates of convergence of empirical spectral distributions", J. of Theoretical Probability, vol. 17, no. 4, October 2004, pp. 1003-1019.

[7] L.H. Chen, "An inequality for the multivariate normal distribution", J. Multivariate Anal., vol. 12, no. 2, pp. 306-315, 1982.

[8] L. Cottatellucci, M. Debbah, ”The Effect of Line of Sight Components on the Asymptotic Capacity of MIMO Systems“, in Proc. ISIT 04, Chicago, June-July, 2004.

[9] C.N. Chuah, D.N.C. Tse, J.M. Kahn, R.A. Valenzuela, ”Capacity Scaling in MIMO Wireless Systems under Correlated Fading”, IEEE Trans. Inf. Theo.,vol.48,no 3,pp 637-650,March 2002

[10] J.Dumont, Ph.Loubaton, S.Lasaulce, M.Debbah, ”On the Asymptotic Performance of MIMO Correlated Rician Channels“, Proc ICASSP05, vol. 5, pp. 813-816, March 2005.

[11] J. Dumont, W. Hachem, P. Loubaton, J. Najim, ”On the asymptotic analysis of mutual information of MIMO Rician correlated channels“, Proc. Int. Symp. on Com., Control and Sig. Proc., Marrakech, March 2006.

[12] J. Dumont, P. Loubaton, S. Lasaulce, "On the capacity achieving covariance matrix for Rician MIMO channels : an asymptotic approach", *In Proc. Globecom 2006*, San Francisco, December 2006.

[13] V.L. Girko, ”Theory of Stochastic Canonical Equations, Volume I“, Chap. 7, Kluwer Academic Publishers, Dordrecht, The Netherlands, 2001.

[14] J. Glimm, A. Jaffe, *Quantum Physics : A functional point of view*, Springer Verlag, New York, second edition, 1987.

[15] A. Goldsmith, S.A. Jafar, N. Jindal, S. Vishwanath, ”Capacity Limits of MIMO Channels“, IEEE J. Sel. Areas in Comm.,vol.21,no 5,June 2003

[16] W. Hachem and O. Khorunzhiy and P. Loubaton and J. Najim and L. Pastur, “A new approach for capacity analysis of large dimensional multi-antenna channels” *submitted to IEEE trans. Inform. Theory, 2006.* available at http ://arxiv.org/abs/cs.IT/0612076

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

[18] J. Hansen and H. Bölcskei, ”A Geometrical Investigation of the Rank-1 Rician MIMO Channel at High SNR“, *IEEE Int. Symp. on Inf. Theo.*, Chicago, June/July 2004.

[19] D. Hoesli, Y.H. Kim, A. Lapidoth, ”Monotonicity results for coherent MIMO Rician channels“, IEEE Trans. on Information Theory, vol. 51, no. 12, pp. 4334-4339, December 2005.

[20] Horn, R. A. and Johnson, C. R., “Matrix Analysis”, Cambridge University Press, 1994.

[21] C. Houdré, V. Pérez-Abreu, D. Surgailis, "Interpolation, correlation identities, and inequalities for infinitely divisible variables", J. Fourier Anal. Appl., vol. 4, no. 6, pp. 651-668, 1998.

[22] M. Kang, M.S. Alouini, ”Capacity of MIMO Rician channels“, IEEE Trans. on Wireless Communications, vol. 5, no. 1, pp. 112-122, January 2006.

[23] M. Kang, M.S. Alouini, ”Capacity of correlated MIMO Rayleigh channels“, IEEE Trans. on Wireless Communications, vol. 5, no. 1, pp.143-155, January 2006.

[24] M.G. Krein, A.A. Nudelman, "The Markov Moment Problem and Extremal Problems", American Mathematical Society, Providence, Rhode Island, 1997.

[25] A.M. Khorunzhy, L.A. Pastur, "Limits of infinite interaction radius, dimensionality and the number of components for random operators with off-diagonal randomness", Comm. Math. Phys., vol. 153, no. 3, pp. 605-646, 1993.

[26] G. Lebrun, M. Faulkner, M. Shafi, P.J. Smith, "MIMO Ricean channel capacity : an asymptotic analysis", IEEE Trans. on Wireless Communications, vol. 5, no. 6, June 2006, pp. 1343-1350.

[27] A. Lozano, A.M. Tulino, S. Verdú, ”Multiple-Antenna Capacity in the Low-Power Regime“, Trans. on Inf. Theo., vol.49,no 10,pp 2527-2544, Oct. 03

[28] D.G. Luenberger, ”Optimization by Vector Space Methods“, John Wiley and Sons, Inc, New-York, 1969.

[29] V.A. Marcenko, L.A. Pastur, "Distribution of eigenvalues in a certain set of random matrices", Mat. Sb., 72(114) :507-526, 1967.

[30] A.L. Moustakas, S.H. Simon, A.M. Sengupta, ”MIMO Capacity Through Correlated Channels in the Presence of Correlated Interference and Noise : A (Not so) Large N Analysis“, Trans. on Inf. Theo., vol.49,no 10,pp 2545-2561, Oct. 03.

[31] A.L. Moustakas, S.H. Simon, ”Random matrix theory of multi-antenna communications : the Rician case“, J. Phys. A : Math. Gen. 38 (2005) 10859-10872.

[32] L.A. Pastur, A.M. Khorunzhy, V.Yu Vasil'chuck, "On the asymptotic property of the spectrum of the sum of one-dimensional independent random operators", Dopov. Nats. Akad. Nauk Ukra#ıni, no. 2, pp. 27-30, 1995.

[33] L.A. Pastur, "A simple approach for the study of global regime of large random matrices", Ukrainian Mathematical Journal, vol. 57, no. 6, pp. 936-966, June 2005.

[34] G. Rekaya, J. C. Belfiore and E. Viterbo, "Algrebric 3×3 and 4×4 6×6 space-time codes with non-vanishing determinants", *In Proc. of ISITA*, Parma, Italy, Oct. 2004.

[35] L.L. Scharf, "Statistical Signal Processing", Addison-Wesley, 1991.

[36] J. Silverstein, "The smallest eigenvalue of a large-dimensional Wishart matrix", Ann. Probab., vol. 13, no. 4, pp. 1364–1368, 1985.

[37] G. Taricco, "On the Capacity of Separately-Correlated MIMO Rician Fading Channels" *In Proc. Globecom 2006*, San Francisco, December 2006.

[38] G. Tarrico, E. Riegler, "On the ergodic capacity of the asymptotic separately-correlated Rician fading MIMO channel with interference", in Proc. ISIT 2007, Nice, France, June 24-June 29 2007, pp. 531-535.

[39] E. Telatar, "Capacity of Multi-antenna Gaussian Channels", Europ. Trans. Telecom., vol.10, pp 585-595, Nov. 99

[40] A.M. Tulino, A. Lozano, S. Verdú, "Capacity-achieving Input Covariance for Correlated Multi-Antenna Channels", 41^{th} Annual Allerton Conf. on Comm., Control and Computing, Monticello, Oct. 03

[41] A.M. Tulino, S. Verdu, "Random Matrix Theory and Wireless Communications", in Foundations and Trends in Communications and Information Theory, vol. 1, pp. 1-182, Now Publishers, June 2004.

[42] M. Vu, A. Paulraj, "Capacity optimization for Rician correlated MIMO wireless channels", in Proc. Asilomar Conference, pp. 133-138, Asilomar, November 2005.

[43] C-K. Wen, K-K. Wong, P. Ting, C-L. I, "Optimal power-saving input covariance for MIMO wireless systems exploiting only channel spatial correlations", Proc. of IEEE ICC 2005, Vol.3, May 2005, pp 2011-2015

[44] C-K. Wen, P. Ting, J-T. Chen "Asymptotic analysis of MIMO wireless systems with spatial correlation at the receiver", IEEE Trans. on Communications, Vol. 54, No. 2, pp. 349-363, February 2006.