On the capacity achieving covariance matrix for frequency selective MIMO channels using the asymptotic approach

Florian Dupuy, Philippe Loubaton

On the Capacity Achieving Covariance Matrix for Frequency Selective MIMO Channels Using the Asymptotic Approach

Florian Dupuy and Philippe Loubaton, *Fellow, IEEE*

***Abstract*—In this contribution, an algorithm for evaluating the capacity-achieving input covariance matrices for frequency selective Rayleigh MIMO channels is proposed. In contrast with the flat fading Rayleigh case, no closed-form expressions for the eigenvectors of the optimum input covariance matrix are available. Classically, both the eigenvectors and eigenvalues are computed numerically and the corresponding optimization algorithms remain computationally very demanding. In this paper, it is proposed to optimize (w.r.t. the input covariance matrix) a large system approximation of the average mutual information derived by Moustakas and Simon. The validity of this asymptotic approximation is clarified thanks to Gaussian large random matrices methods. It is shown that the approximation is a strictly concave function of the input covariance matrix and that the average mutual information evaluated at the argmax of the approximation is equal to the capacity of the channel up to a $\mathcal{O}(1/t)$ term, where t is the number of transmit antennas. An algorithm based on an iterative waterfilling scheme is proposed to maximize the average mutual information approximation, and its convergence studied. Numerical simulation results show that, even for a moderate number of transmit and receive antennas, the new approach provides the same results as direct maximization approaches of the average mutual information.**

***Index Terms*—Ergodic capacity, frequency selective MIMO channels, large random matrices**

I. Introduction

When the channel state information is available at both the receiver and the transmitter of a MIMO system, the problem of designing the transmitter in order to maximize the (Gaussian) mutual information of the system has been addressed successfully in a number of papers. This problem is, however, more difficult when the transmitter has the knowledge of the statistical properties of the channel, the channel state information being still available at the receiver side, a more realistic assumption in the context of mobile systems. In this case, the mutual information is replaced by the average mutual information (EMI), which, of course, is more complicated to optimize.

The optimization problem of the EMI has been addressed extensively in the case of certain flat fading Rayleigh channels. In the context of the so-called Kronecker model, it has been shown by various authors (see, e.g., [1] 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. Similar results have been obtained for flat fading uncorrelated Rician channels ([2]).

In this paper, we consider this EMI maximization problem in the case of popular frequency selective MIMO channels (see, e.g., [3], [4]) with independent paths. In this context, the eigenvectors of the optimum transmit covariance matrix have no closed-form expressions, so that both the eigenvalues and the eigenvectors of the matrix have to be evaluated numerically. For this, it is possible to adapt the approach of [5] developed in the context of correlated Rician channels. However, the corresponding algorithms are computationally very demanding as they heavily rely on intensive Monte-Carlo simulations. We therefore propose to optimize the approximation of the EMI, derived by Moustakas and Simon ([4]), in principle valid when the number of transmit and receive antennas converge to infinity at the same rate, but accurate for realistic numbers of antennas. This will turn out to be a simpler problem. We mention that, while [4] contains some results related to the structure of the argument of the maximum of the EMI approximation, [4] does not propose any optimization algorithm.

We first review the results of [4] related to the large system approximation of the EMI. The analysis of [4] is based on the so-called replica method, an ingenious trick whose mathematical relevance has not yet been established mathematically. Using a generalization of the rigorous analysis of [6], we verify the validity of the approximation of [4] and provide the convergence speed under certain technical assumptions. Besides, the expression of the approximation depends on the solutions of a non linear system. The existence and the uniqueness of the solutions are not addressed in [4]. As our optimization algorithm needs to solve this system, we clarify this crucial point. We show in particular that the system admits a unique solution that can be evaluated numerically using the fixed point algorithm. Next, we study the properties of the EMI approximation, and briefly justify that it is a strictly concave function of the input covariance matrix. We show that the mutual information corresponding to the argmax of the EMI approximation is equal to the channel capacity up to a $\mathcal{O}\left(\frac{1}{t}\right)$ term, where t is the number of transmit antennas.

Manuscript received April 16, 2010; revised April 12, 2011. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Aris Moustakas. The work of F. Dupuy is partially supported by the French Government through a CIFRE grant between Thales Communication and Université Paris Est.

F. Dupuy is with Thales Communications, 92704 Colombes, France and also with Université Paris Est, 77454 Champs-sur-Marne, France.

P. Loubaton is with Université Paris Est, IGM LabInfo, UMR-CNRS 8049, 77454 Champs-sur-Marne, France.

Therefore it is relevant to optimize the EMI approximation to evaluate the capacity achieving covariance matrix. We finally present our maximization algorithm of the EMI approximation. It is based on an iterative waterfilling algorithm which, in some sense, can be seen as a generalization of [7] devoted to the Rayleigh context and of [8], [9] devoted to the correlated Rician case: Each iteration will be devoted to solve the above mentioned system of nonlinear equations as well as a standard waterfilling problem. It is proved that the algorithm converges towards the optimum input covariance matrix as long as it converges[1].

The paper is organized as follows. Section II is devoted to the presentation of the channel model, the underlying assumptions, the problem statement. In section III, we rigorously derive the large system approximation of the EMI with Gaussian methods and establish some properties of the asymptotic approximation as a function of the covariance matrix of the input signal. The maximization problem of the EMI approximation is then studied in section IV. Numerical results are provided in section V.

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 $\|\mathbf{M}\|$, $\rho(\mathbf{M})$ and $|\mathbf{M}|$ respectively stand for the spectral norm, the spectral radius and the determinant 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 $\mathbb{E}(\cdot)$. We denote by $\delta_{i,j}$ the Kronecker delta, i.e. $\delta_{i,j}=1$ if $i=j$ and 0 otherwise.

All along this paper, r and t stand for the number of receive and transmit antennas. Certain quantities will be studied in the asymptotic regime $t\to\infty$, $r\to\infty$ in such a way that $t/r\to c\in(0,\infty)$. In order to simplify the notations, $t\to\infty$ should be understood from now on as $t\to\infty$, $r\to\infty$ and $t/r\to c\in(0,\infty)$. 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 corrupted by a multipath propagation channel. The discrete-time propagation channel between the transmitter and the receiver is characterized by the input-output equation

$$\mathbf{y}(n)=\sum_{l=1}^{L}\mathbf{H}^{(l)}\mathbf{s}(n-l+1)+\mathbf{n}(n)=[\mathbf{H}(z)]\mathbf{s}(n)+\mathbf{n}(n), \quad (1)$$

where $\mathbf{s}(n)=[s_1(n),\ldots,s_t(n)]^T$ and $\mathbf{y}(n)=[y_1(n),\ldots,y_r(n)]^T$ represent the transmit and the receive vector at time n respectively. $\mathbf{n}(n)$ is an additive Gaussian noise such that $\mathbb{E}(\mathbf{n}(n)\mathbf{n}(n)^H)=\sigma^2\mathbf{I}$. $\mathbf{H}(z)$ denotes the transfer function of the discrete-time equivalent channel defined by

$$\mathbf{H}(z)=\sum_{l=1}^{L}\mathbf{H}^{(l)}\,z^{-(l-1)}. \quad (2)$$

Each coefficient $\mathbf{H}^{(l)}$ is assumed to be a Gaussian random matrix given by

$$\mathbf{H}^{(l)}=\frac{1}{\sqrt{t}}(\mathbf{C}^{(l)})^{1/2}\mathbf{W}_l(\tilde{\mathbf{C}}^{(l)})^{1/2}, \quad (3)$$

where $\mathbf{W}_l$ is a $r\times t$ random matrix whose entries are independent and identically distributed complex circular Gaussian random variables, with zero mean and unit variance. The matrices $\mathbf{C}^{(l)}$ and $\tilde{\mathbf{C}}^{(l)}$ are positive definite, and respectively account for the receive and transmit antenna correlation. This correlation structure is called a separable or Kronecker correlation model. We also assume that for each $k\neq l$, matrices $\mathbf{H}^{(k)}$ and $\mathbf{H}^{(l)}$ are independent. Note that our assumptions imply that $\mathbf{H}^{(l)}\neq 0$ for $l=1,\ldots,L$. However, it can be checked easily that the results stated in this paper remain valid if some coefficients $(\mathbf{H}^{(l)})_{l=1,\ldots,L}$ are zero.

In this article the channel matrices are assumed perfectly known at the receiver side. However, only the statistics of the $(\mathbf{H}^{(l)})_{l=1,\ldots,L}$, i.e. matrices $(\tilde{\mathbf{C}}^{(l)},\mathbf{C}^{(l)})_{l=1,\ldots,L}$, are available at the transmitter side.

C. Ergodic Capacity of the Channel.

Let $\mathbf{Q}(e^{2i\pi\nu})$ be the $t\times t$ spectral density matrix of the transmit signal $\mathbf{s}(n)$, which is assumed to verify the transmit power condition

$$\frac{1}{t}\int_0^1\mathrm{Tr}(\mathbf{Q}(e^{2i\pi\nu}))d\nu=1. \quad (4)$$

Then, the (Gaussian) ergodic mutual information $I(\mathbf{Q}(.))$ between the transmitter and the receiver is defined as

$$I(\mathbf{Q}(.))=\mathbb{E}_{\mathcal{W}}\left[\int_0^1\log\left|\mathbf{I}_r+\frac{1}{\sigma^2}\mathbf{H}(\cdot)\mathbf{Q}(\cdot)\mathbf{H}(\cdot)^H\right|\,d\nu\right], \quad (5)$$

where $\mathbb{E}_{\mathcal{W}}[.]=\mathbb{E}_{(\mathbf{W}_l)_{l=1,\ldots,L}}[.]$. The ergodic capacity of the MIMO channel is equal to the maximum of $I(\mathbf{Q}(.))$ over the set of all spectral density matrices satisfying the constraint (4). The hypotheses formulated on the statistics of the channel allow however to limit the optimization to the set of positive matrices which are independent of the frequency ν. This is because the probability distribution of matrix $\mathbf{H}(e^{2i\pi\nu})$ is clearly independent of the frequency ν. More precisely, the mutual information $\mathbf{I}(\mathbf{Q}(.))$ is also given by

$$I(\mathbf{Q}(.))=\mathbb{E}_{\mathbf{H}}\left[\int_0^1\log\left|\mathbf{I}_r+\frac{1}{\sigma^2}\mathbf{H}\mathbf{Q}(e^{2i\pi\nu})\mathbf{H}^H\right|\,d\nu\right],$$

where $\mathbf{H}=\sum_{l=1}^{L}\mathbf{H}^{(l)}=\mathbf{H}(1)$. Using the concavity of the logarithm, we obtain that

$$I(\mathbf{Q}(.))\leq\mathbb{E}_{\mathbf{H}}\left[\log\left|\mathbf{I}_r+\frac{1}{\sigma^2}\mathbf{H}\left(\int_0^1\mathbf{Q}(e^{2i\pi\nu})d\nu\right)\mathbf{H}^H\right|\right].$$

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

We denote by $\mathcal{C}$ the cone of non negative hermitian matrices, and by $\mathcal{C}_1$ the subset of all matrices $\mathbf{Q}$ of $\mathcal{C}$ satisfying $\frac{1}{t}\mathrm{Tr}(\mathbf{Q}) = 1$. If $\mathbf{Q}$ is an element of $\mathcal{C}_1$, the mutual information $I(\mathbf{Q})$ reduces to

$$I(\mathbf{Q}) = \mathbb{E}_{\mathbf{H}}\left[\log\left|\mathbf{I}_r + \frac{1}{\sigma^2}\mathbf{H}\mathbf{Q}\mathbf{H}^H\right|\right]. \tag{6}$$

$\mathbf{Q} \mapsto I(\mathbf{Q})$ is strictly concave on the convex set $\mathcal{C}_1$ and reaches its maximum at a unique element $\mathbf{Q}_* \in \mathcal{C}_1$. It is clear that if $\mathbf{Q}(e^{2i\pi\nu})$ is any spectral density matrix satisfying (4), then the matrix $\int_0^1 \mathbf{Q}(e^{2i\pi\nu})d\nu$ is an element of $\mathcal{C}_1$. Therefore,

$$\mathbb{E}_{\mathbf{H}}\Big[\log\Big|\mathbf{I}_r + \frac{1}{\sigma^2}\mathbf{H}\left(\int_0^1 \mathbf{Q}(e^{2i\pi\nu})d\nu\right)\mathbf{H}^H\Big|\Big] \leq \mathbb{E}_{\mathbf{H}}\Big[\log\Big|\mathbf{I}_r + \frac{1}{\sigma^2}\mathbf{H}\mathbf{Q}_*\mathbf{H}^H\Big|\Big].$$

In other words,

$$I(\mathbf{Q}(.)) \leq I(\mathbf{Q}_*)$$

for each spectral density matrix verifying (4). This shows that the maximum of function I over the set of all spectral densities satisfying (4) is reached on the set $\mathcal{C}_1$. The ergodic capacity $\mathcal{C}_E$ of the channel is thus equal to

$$\mathcal{C}_E = \max_{\mathbf{Q}\in\mathcal{C}_1} I(\mathbf{Q}). \tag{7}$$

We note that property (7) also holds if the time delays of the channel are non integer multiples of the symbol period, provided that the receiving filter coincides with the ideal low-pass filter on the $[-\frac{1}{2T}, \frac{1}{2T}]$ frequency interval, where T denotes the symbol period. If this is the case, the transfer function $\mathbf{H}(e^{2i\pi\nu})$ is equal to $\mathbf{H}(e^{2i\pi\nu}) = \sum_{l=1}^{L}\mathbf{H}^{(l)}e^{-2i\pi\nu\tau_l}$, where τ_l is the delay associated to path l for $l = 1, \ldots, L$. The probability distribution of $\mathbf{H}(e^{2i\pi\nu})$ does not depend on ν and this leads immediately to (7).

If the matrices $(\mathbf{C}^{(l)})_{l=1,\ldots,L}$ all coincide with a matrix $\mathbf{C}$, matrix $\mathbf{H}$ follows a Kronecker model with transmit and receive covariance matrices $\frac{1}{L}\sum_{l=1}^{L}\tilde{\mathbf{C}}^{(l)}$ and $\mathbf{C}$ respectively [10]. In this case, the eigenvectors of the optimum matrix $\mathbf{Q}_*$ coincide with the eigenvectors of $\frac{1}{L}\sum_{l=1}^{L}\tilde{\mathbf{C}}^{(l)}$. The situation is similar if the transmit covariance matrices $(\tilde{\mathbf{C}}^{(l)})_{l=1,\ldots,L}$ coincide. In the most general case, the eigenvectors of $\mathbf{Q}_*$ have however no closed-form expression. The evaluation of $\mathbf{Q}_*$ and of the channel capacity $\mathcal{C}_E$ is thus a more difficult problem. A possible solution consists in adapting the Vu-Paulraj approach ([5]) to the present context. However, the algorithm presented in [5] is very demanding since the needed evaluations of $I(\mathbf{Q})$ gradient and Hessian require intensive Monte-Carlo simulations.

D. The Large System Approximation of $I(\mathbf{Q})$

When t and r converge to ∞ while $t/r \to c$, $c \in (0, \infty)$, [4] showed that $I(\mathbf{Q})$ can be approximated by $\overline{I}(\mathbf{Q})$ defined by (8) at the bottom of the page, where $(\delta_1(\mathbf{Q}), \ldots, \delta_L(\mathbf{Q}))^T = \boldsymbol{\delta}(\mathbf{Q})$ and $(\tilde{\delta}_1(\mathbf{Q}), \ldots, \tilde{\delta}_L(\mathbf{Q}))^T = \tilde{\boldsymbol{\delta}}(\mathbf{Q})$ are the positive solutions of the system of $2L$ equations:

$$\begin{cases} \kappa_l = f_l(\tilde{\boldsymbol{\kappa}}) \\ \tilde{\kappa}_l = \tilde{f}_l(\boldsymbol{\kappa}, \mathbf{Q}) \end{cases} \quad \text{for } l = 1, \ldots, L, \tag{9}$$

with $\boldsymbol{\kappa} = (\kappa_1, \ldots, \kappa_L)^T$ and $\tilde{\boldsymbol{\kappa}} = (\tilde{\kappa}_1, \ldots, \tilde{\kappa}_L)^T$, and with

$$\begin{cases} f_l(\tilde{\boldsymbol{\kappa}}) = \frac{1}{t}\mathrm{Tr}\left[\mathbf{C}^{(l)}\mathbf{T}(\tilde{\boldsymbol{\kappa}})\right], \\ \tilde{f}_l(\boldsymbol{\kappa}, \mathbf{Q}) = \frac{1}{t}\mathrm{Tr}\left[\mathbf{Q}^{1/2}\tilde{\mathbf{C}}^{(l)}\mathbf{Q}^{1/2}\tilde{\mathbf{T}}(\boldsymbol{\kappa}, \mathbf{Q})\right]. \end{cases} \tag{10}$$

The $r \times r$ matrix $\mathbf{T}(\tilde{\boldsymbol{\kappa}})$ and the $t \times t$ matrix $\tilde{\mathbf{T}}(\boldsymbol{\kappa}, \mathbf{Q})$ are respectively defined by:

$$\mathbf{T}(\tilde{\boldsymbol{\kappa}}) = \left[\sigma^2\left(\mathbf{I}_r + \sum_{j=1}^{L}\tilde{\kappa}_j\mathbf{C}^{(j)}\right)\right]^{-1}, \tag{11}$$

$$\tilde{\mathbf{T}}(\boldsymbol{\kappa}, \mathbf{Q}) = \left[\sigma^2\left(\mathbf{I}_t + \sum_{j=1}^{L}\kappa_j\mathbf{Q}^{1/2}\tilde{\mathbf{C}}^{(j)}\mathbf{Q}^{1/2}\right)\right]^{-1}. \tag{12}$$

III. Deriving the Large System Approximation

A. The Canonical Equations

In [4], the existence and the uniqueness of positive solutions to (9) is assumed without justification. Moreover no algorithm is given for the calculation of the δ_l and $\tilde{\delta}_l$, $l = 1, \ldots, L$. We therefore clarify below these important points. We consider the case $\mathbf{Q} = \mathbf{I}$ in order to simplify the notations. To address the general case it is sufficient to change matrices $(\tilde{\mathbf{C}}^{(l)})_{l=1,\ldots,L}$ into $(\mathbf{Q}^{1/2}\tilde{\mathbf{C}}^{(l)}\mathbf{Q}^{1/2})_{l=1,\ldots,L}$ in what follows.

Theorem 1: The system of equations (9) admits unique positive solutions $(\delta_l)_{l=1,\ldots,L}$ and $(\tilde{\delta}_l)_{l=1,\ldots,L}$, which are the limits of the following fixed point algorithm:

- Initialization: $\delta_l^{(0)} > 0$, $\tilde{\delta}_l^{(0)} > 0$, $l = 1, \ldots, L$.
- Evaluation of the $\delta_l^{(n+1)}$ and $\tilde{\delta}_l^{(n+1)}$ from $\boldsymbol{\delta}^{(n)} = (\delta_1^{(n)}, \ldots, \delta_L^{(n)})^T$ and $\tilde{\boldsymbol{\delta}}^{(n)} = (\tilde{\delta}_1^{(n)}, \ldots, \tilde{\delta}_L^{(n)})^T$:

$$\begin{cases} \delta_l^{(n+1)} = f_l(\tilde{\boldsymbol{\delta}}^{(n)}), \\ \tilde{\delta}_l^{(n+1)} = \tilde{f}_l(\boldsymbol{\delta}^{(n)}, \mathbf{I}). \end{cases} \tag{13}$$

Proof: We prove the existence and uniqueness of positive solutions.

1) Existence: Using analytic continuation technique, we show in Appendix A that the fixed point algorithm introduced converges to positive coefficients δ_l and $\tilde{\delta}_l$, $l = 1, \ldots, L$. As functions $\tilde{\boldsymbol{\kappa}} \mapsto f_l(\tilde{\boldsymbol{\kappa}})$ and $\boldsymbol{\kappa} \mapsto \tilde{f}_l(\boldsymbol{\kappa}, \mathbf{I})$ are clearly continuous, the limit of $(\boldsymbol{\delta}^{(n)}, \tilde{\boldsymbol{\delta}}^{(n)})$ when $n \to \infty$ satisfies (9). Hence, the convergence of the algorithm yields the existence of a positive solution to (9).

$$\overline{I}(\mathbf{Q}) = \log\left|\mathbf{I}_r + \sum_{l=1}^{L}\tilde{\delta}_l(\mathbf{Q})\mathbf{C}^{(l)}\right| + \log\left|\mathbf{I}_t + \mathbf{Q}\left(\sum_{l=1}^{L}\delta_l(\mathbf{Q})\tilde{\mathbf{C}}^{(l)}\right)\right| - \sigma^2 t\left(\sum_{l=1}^{L}\delta_l(\mathbf{Q})\tilde{\delta}_l(\mathbf{Q})\right) \tag{8}$$

2) Uniqueness: : Let $(\boldsymbol{\delta}, \tilde{\boldsymbol{\delta}})$ and $(\boldsymbol{\delta}', \tilde{\boldsymbol{\delta}}')$ be two solutions of the canonical equation (9) with $\mathbf{Q} = \mathbf{I}$. We denote $(\mathbf{T}, \tilde{\mathbf{T}})$ and $(\mathbf{T}', \tilde{\mathbf{T}}')$ the associated matrices defined by (11) and (12), where $(\boldsymbol{\kappa}, \tilde{\boldsymbol{\kappa}})$ respectively coincide with $(\boldsymbol{\delta}, \tilde{\boldsymbol{\delta}})$ and $(\boldsymbol{\delta}', \tilde{\boldsymbol{\delta}}')$. Introducing $\mathbf{e} = \boldsymbol{\delta} - \boldsymbol{\delta}' = (e_1, \ldots, e_L)^T$ we have:

$$\begin{aligned} e_l &= \frac{1}{t}\mathrm{Tr}\left[\mathbf{C}^{(l)}\mathbf{T}(\mathbf{T}'^{-1} - \mathbf{T}^{-1})\mathbf{T}'\right] \\ &= \frac{\sigma^2}{t}\sum_{k=1}^{L}(\tilde{\delta}'_k - \tilde{\delta}_k)\mathrm{Tr}\left(\mathbf{C}^{(l)}\mathbf{T}\mathbf{C}^{(k)}\mathbf{T}'\right). \end{aligned} \tag{14}$$

Similarly, with $\tilde{\boldsymbol{e}} = \tilde{\boldsymbol{\delta}} - \tilde{\boldsymbol{\delta}}' = (\tilde{e}_1, \ldots, \tilde{e}_L)^T$,

$$\tilde{e}_k = \frac{\sigma^2}{t}\sum_{l=1}^{L}(\delta'_l - \delta_l)\mathrm{Tr}\left(\tilde{\mathbf{C}}^{(k)}\tilde{\mathbf{T}}\tilde{\mathbf{C}}^{(l)}\tilde{\mathbf{T}}'\right). \tag{15}$$

And (14) and (15) can be written together as

$$\begin{bmatrix} \mathbf{I} & \sigma^2\mathbf{A}(\mathbf{T}, \mathbf{T}') \\ \sigma^2\tilde{\mathbf{A}}(\tilde{\mathbf{T}}, \tilde{\mathbf{T}}') & \mathbf{I} \end{bmatrix}\begin{bmatrix} \mathbf{e} \\ \tilde{\mathbf{e}} \end{bmatrix} = \mathbf{0}, \tag{16}$$

where $L \times L$ matrices $\mathbf{A}(\mathbf{T}, \mathbf{T}')$ and $\tilde{\mathbf{A}}(\tilde{\mathbf{T}}, \tilde{\mathbf{T}}')$ are defined by $\mathbf{A}_{kl}(\mathbf{T}, \mathbf{T}') = \frac{1}{t}\mathrm{Tr}\left(\mathbf{C}^{(k)}\mathbf{T}\mathbf{C}^{(l)}\mathbf{T}'\right)$ and $\tilde{\mathbf{A}}_{kl}(\tilde{\mathbf{T}}, \tilde{\mathbf{T}}') = \frac{1}{t}\mathrm{Tr}(\tilde{\mathbf{C}}^{(k)}\tilde{\mathbf{T}}\tilde{\mathbf{C}}^{(l)}\tilde{\mathbf{T}}')$. We will now prove that $\rho(\mathbf{M}) < 1$, with $\mathbf{M} = \sigma^4\tilde{\mathbf{A}}(\tilde{\mathbf{T}}, \tilde{\mathbf{T}}')\mathbf{A}(\mathbf{T}, \mathbf{T}')$. This will imply that the matrix governing the linear system (16) is invertible, and thus that $\mathbf{e} = \tilde{\mathbf{e}} = \mathbf{0}$, i.e. the uniqueness.

$$\begin{aligned} |\mathbf{M}_{kl}| &= \left|\frac{\sigma^4}{t^2}\sum_{j=1}^{L}\mathrm{Tr}(\tilde{\mathbf{C}}^{(k)}\tilde{\mathbf{T}}\tilde{\mathbf{C}}^{(j)}\tilde{\mathbf{T}}')\mathrm{Tr}(\mathbf{C}^{(j)}\mathbf{T}\mathbf{C}^{(l)}\mathbf{T}')\right| \\ &\leq \frac{\sigma^4}{t^2}\sum_{j=1}^{L}\left|\mathrm{Tr}(\tilde{\mathbf{C}}^{(k)}\tilde{\mathbf{T}}\tilde{\mathbf{C}}^{(j)}\tilde{\mathbf{T}}')\right|\left|\mathrm{Tr}(\mathbf{C}^{(j)}\mathbf{T}\mathbf{C}^{(l)}\mathbf{T}')\right|. \end{aligned} \tag{17}$$

Using Cauchy-Schwarz inequality $|\mathrm{Tr}(\mathbf{A}\mathbf{B})|^2 \leq \mathrm{Tr}(\mathbf{A}\mathbf{A}^H)\cdot\mathrm{Tr}(\mathbf{B}\mathbf{B}^H)$, we have:

$$\begin{cases} \frac{1}{t}\left|\mathrm{Tr}(\tilde{\mathbf{C}}^{(k)}\tilde{\mathbf{T}}\tilde{\mathbf{C}}^{(j)}\tilde{\mathbf{T}}')\right| \leq \sqrt{\tilde{\mathbf{A}}_{kj}(\tilde{\mathbf{T}}, \tilde{\mathbf{T}})\tilde{\mathbf{A}}_{kj}(\tilde{\mathbf{T}}', \tilde{\mathbf{T}}')}, \\ \frac{1}{t}\left|\mathrm{Tr}(\mathbf{C}^{(j)}\mathbf{T}\mathbf{C}^{(l)}\mathbf{T}')\right| \leq \sqrt{\mathbf{A}_{jl}(\mathbf{T}, \mathbf{T})\mathbf{A}_{jl}(\mathbf{T}', \mathbf{T}')}. \end{cases}$$

Using these two inequalities in (17) gives

$$|\mathbf{M}_{kl}| \leq \sigma^4\sum_{j=1}^{L}\sqrt{\tilde{\mathbf{A}}_{kj}(\tilde{\mathbf{T}})\tilde{\mathbf{A}}_{kj}(\tilde{\mathbf{T}}')\mathbf{A}_{jl}(\mathbf{T})\mathbf{A}_{jl}(\mathbf{T}')},$$

where matrices $\mathbf{A}(\mathbf{T})$ and $\tilde{\mathbf{A}}(\tilde{\mathbf{T}})$ are defined by

$$\begin{cases} \mathbf{A}_{kl}(\mathbf{T}) = \frac{1}{t}\mathrm{Tr}(\mathbf{C}^{(k)}\mathbf{T}\mathbf{C}^{(l)}\mathbf{T}) = \mathbf{A}_{kl}(\mathbf{T}, \mathbf{T}), \\ \tilde{\mathbf{A}}_{kl}(\tilde{\mathbf{T}}) = \frac{1}{t}\mathrm{Tr}(\tilde{\mathbf{C}}^{(k)}\tilde{\mathbf{T}}\tilde{\mathbf{C}}^{(l)}\tilde{\mathbf{T}}) = \tilde{\mathbf{A}}_{kl}(\tilde{\mathbf{T}}, \tilde{\mathbf{T}}). \end{cases} \tag{18}$$

Using Cauchy-Schwarz inequality then yields:

$$|\mathbf{M}_{kl}| \leq \mathbf{P}_{kl},$$

where $\mathbf{P}$ is the $L \times L$ matrix whose entries are defined by $\mathbf{P}_{kl} = \sqrt{\left(\sigma^4\tilde{\mathbf{A}}(\tilde{\mathbf{T}})\mathbf{A}(\mathbf{T})\right)_{kl}}\sqrt{\left(\sigma^4\tilde{\mathbf{A}}(\tilde{\mathbf{T}}')\mathbf{A}(\mathbf{T}')\right)_{kl}}$. Theorem 8.1.18 of [11] then yields $\rho(\mathbf{M}) \leq \rho(\mathbf{P})$. Besides, Lemma 5.7.9 of [12] used on the definition of $\mathbf{P}$ gives:

$$\rho(\mathbf{P}) \leq \sqrt{\rho\left(\sigma^4\tilde{\mathbf{A}}(\tilde{\mathbf{T}})\mathbf{A}(\mathbf{T})\right)}\sqrt{\rho\left(\sigma^4\tilde{\mathbf{A}}(\tilde{\mathbf{T}}')\mathbf{A}(\mathbf{T}')\right)}. \tag{19}$$

Lemma 1 (ii) in Appendix C implies that $\rho(\sigma^4\tilde{\mathbf{A}}(\tilde{\mathbf{T}})\mathbf{A}(\mathbf{T})) < 1$ and $\rho(\sigma^4\tilde{\mathbf{A}}(\tilde{\mathbf{T}}')\mathbf{A}(\mathbf{T}')) < 1$, so that (19) finally implies:

$$\rho(\mathbf{M}) \leq \rho(\mathbf{P}) < 1.$$

This completes the proof of Theorem 1. □

B. Deriving the Approximation of $I(\mathbf{Q} = \mathbf{I}_t)$ With Gaussian Methods

We consider in this section the case $\mathbf{Q} = \mathbf{I}_t$. We note $I = I(\mathbf{I}_t)$, $\overline{I} = \overline{I}(\mathbf{I}_t)$. We have proved in the previous section the consistency of $\overline{I}(\mathbf{Q})$ definition. To establish the approximation of $I(\mathbf{Q})$, [4] used the replica method, a useful and simple trick whose mathematical relevance is not yet proved in the present context. Moreover, no assumptions were specified for the convergence of $I(\mathbf{Q})$ towards $\overline{I}(\mathbf{Q})$. However, using large random matrix techniques similar to those of [6] and [8], it is possible to prove rigorously the following theorem, in which the (mild) suitable technical assumptions are clarified.

Theorem 2: Assume that $\sup_t \|\mathbf{C}^{(j)}\| < +\infty$, $\sup_t \|\tilde{\mathbf{C}}^{(j)}\| < +\infty$, $\inf_t\left(\frac{1}{t}\mathrm{Tr}\,\mathbf{C}^{(j)}\right) > 0$ and $\inf_t\left(\frac{1}{t}\mathrm{Tr}\,\tilde{\mathbf{C}}^{(j)}\right) > 0$, for any $j \in \{1, \ldots, L\}$. Then,

$$I = \overline{I} + \mathcal{O}\left(\frac{1}{t}\right).$$

Sketch of proof: The proof is done in three steps:

1) In a first step we derive a large system approximation of $\mathbb{E}_{\mathbf{H}}[\mathrm{Tr}\,\mathbf{S}]$, where $\mathbf{S} = (\mathbf{H}\mathbf{H}^H + \sigma^2\mathbf{I}_r)^{-1}$ is the resolvent of $\mathbf{H}\mathbf{H}^H$ at point $-\sigma^2$. Nonetheless the approximation is expressed with the terms $\alpha_l = \frac{1}{t}\mathbb{E}_{\mathbf{H}}[\mathrm{Tr}\left(\mathbf{C}^{(l)}\mathbf{S}\right)]$, $l = 1, \ldots, L$, which still depend on the entries of $\mathbb{E}_{\mathbf{H}}[\mathbf{S}]$.
2) A second step refines the previous approximation to obtain an approximation which this time only depends on the variance structure of the channels, i.e. matrices $(\mathbf{C}^{(l)})_{l\in\{1,\ldots,L\}}$ and $(\tilde{\mathbf{C}}^{(l)})_{l\in\{1,\ldots,L\}}$.
3) The previous approximation is used to get the asymptotic behavior of mutual information by a proper integration.

Proof: We now sketch the three steps stated above. We provide the missing details in the Appendix.

1) A First Large System Approximation of $\mathbb{E}_{\mathbf{H}}[\mathrm{Tr}\,\mathbf{S}]$: We introduce vectors $\boldsymbol{\alpha} = [\alpha_1, \ldots, \alpha_L]^T$ and $\tilde{\boldsymbol{\alpha}} = [\tilde{\alpha}_1, \ldots, \tilde{\alpha}_L]^T$ defined by

$$\begin{cases} \alpha_l = \frac{1}{t}\mathrm{Tr}\left[\mathbf{C}^{(l)}\mathbb{E}_{\mathbf{H}}[\mathbf{S}]\right] \\ \tilde{\alpha}_l = \frac{1}{t}\mathrm{Tr}\left[\tilde{\mathbf{C}}^{(l)}\tilde{\mathbf{R}}\right] \end{cases} \text{for } l = 1, \ldots, L, \tag{20}$$

where $t \times t$ matrix $\tilde{\mathbf{R}}$ is defined by $\tilde{\mathbf{R}}(\boldsymbol{\alpha}) = \left[\sigma^2\left(\mathbf{I}_t + \sum_{j=1}^{L}\alpha_j\tilde{\mathbf{C}}^{(j)}\right)\right]^{-1}$. Using large random matrix techniques similar to those of [6] and [8], the following proposition is proved in Appendix B.

Proposition 1: Assume that, for every $j \in \{1, \ldots, L\}$, $\sup_t \|\mathbf{C}^{(j)}\| < +\infty$, $\sup_t \|\tilde{\mathbf{C}}^{(j)}\| < +\infty$. Then $\mathbb{E}_{\mathbf{H}}[\mathbf{S}]$ can be written as

$$\mathbb{E}_{\mathbf{H}}[\mathbf{S}] = \mathbf{R} + \boldsymbol{\Upsilon}, \tag{21}$$

where matrix $\boldsymbol{\Upsilon}$ is such that $\frac{1}{t}\mathrm{Tr}(\boldsymbol{\Upsilon}\mathbf{A}) = \mathcal{O}\left(\frac{1}{t^2}\right)$ for any uniformly bounded matrix $\mathbf{A}$ and where matrix $\mathbf{R}$ is defined by $\mathbf{R}(\tilde{\boldsymbol{\alpha}}) = \left[\sigma^2\left(\mathbf{I}_r + \sum_{j=1}^{L}\tilde{\alpha}_j\mathbf{C}^{(j)}\right)\right]^{-1}$.

One can check that the entries of matrix $\boldsymbol{\Upsilon}$ are $\mathcal{O}\left(\frac{1}{t^{3/2}}\right)$; nevertheless this result is not needed here. It follows from Proposition 1 that, for any $r \times r$ matrix $\mathbf{A}$ uniformly bounded in r,

$$\frac{1}{t}\mathbb{E}_{\mathbf{H}}[\mathrm{Tr}(\mathbf{SA})] = \frac{1}{t}\mathrm{Tr}(\mathbf{RA}) + \mathcal{O}\left(\frac{1}{t^2}\right). \tag{22}$$

Taking $\mathbf{A} = \mathbf{I}$ gives a first approximation of $\mathbb{E}_{\mathbf{H}}[\mathrm{Tr}\,\mathbf{S}]$:

$$\mathbb{E}_{\mathbf{H}}[\mathrm{Tr}\,\mathbf{S}] = \mathrm{Tr}\,\mathbf{R} + \mathcal{O}\left(\frac{1}{t}\right). \tag{23}$$

Nonetheless matrix $\mathbf{R}$ depends on $\mathbb{E}_{\mathbf{H}}[\mathbf{S}]$ through vector $\boldsymbol{\alpha}$.

2) A refined large system approximation of $\mathbb{E}_{\mathbf{H}}[\mathrm{Tr}\ \mathbf{S}]$*:* We first recall from Section III-A that $\mathbf{T}$ is the matrix defined by (11) associated to the solutions $(\boldsymbol{\delta}, \tilde{\boldsymbol{\delta}})$ of the canonical equation (9) with $\mathbf{Q} = \mathbf{I}_t$: $\mathbf{T} = \left(\sigma^2\left(\mathbf{I}_r + \sum_{l=1}^L \tilde{\delta}_l \mathbf{C}^{(l)}\right)\right)^{-1}$. We introduce the following proposition which will lead to the desired approximation of $\mathbb{E}_{\mathbf{H}}[\mathrm{Tr}\ \mathbf{S}]$:

Proposition 2: Assume that, for every $j \in \{1, \ldots, L\}$, $\sup_t \|\mathbf{C}^{(j)}\| < +\infty$, $\sup_t \|\tilde{\mathbf{C}}^{(j)}\| < +\infty$, $\inf_t \left(\frac{1}{t}\mathrm{Tr}\,\mathbf{C}^{(j)}\right) > 0$ and $\inf_t \left(\frac{1}{t}\mathrm{Tr}\,\tilde{\mathbf{C}}^{(j)}\right) > 0$. Let $\mathbf{A}$ be a $r \times r$ matrix uniformly bounded in r, then

$$\frac{1}{t}\mathrm{Tr}(\mathbf{RA}) = \frac{1}{t}\mathrm{Tr}(\mathbf{TA}) + \mathcal{O}\left(\frac{1}{t^2}\right). \tag{24}$$

The proof is given in Appendix C. It relies on the similarity of the systems of equations verified by the $(\alpha_l, \tilde{\alpha}_l)$ and the $(\delta_l, \tilde{\delta}_l)$. Actually, taking $\mathbf{A} = \mathbf{C}^{(l)}$ in (22) yields $\alpha_l = \frac{1}{t}\mathrm{Tr}(\mathbf{C}^{(l)}\mathbf{R}) + \mathcal{O}\left(\frac{1}{t^2}\right)$ and therefore

$$\begin{cases} \alpha_l = \frac{1}{t}\mathrm{Tr}\left[\mathbf{C}^{(l)}\left[\sigma^2(\mathbf{I} + \sum_{j=1}^L \tilde{\alpha}_j \mathbf{C}^{(j)})\right]^{-1}\right] + \mathcal{O}\left(\frac{1}{t^2}\right) \\ \tilde{\alpha}_l = \frac{1}{t}\mathrm{Tr}\left[\tilde{\mathbf{C}}^{(l)}\left[\sigma^2(\mathbf{I} + \sum_{j=1}^L \alpha_j \tilde{\mathbf{C}}^{(j)})\right]^{-1}\right] \end{cases}$$

for $l = 1, \ldots, L$. Taking $\mathbf{A} = \mathbf{I}_r$ in (24) together with (23) leads to

$$\mathbb{E}_{\mathbf{H}}[\mathrm{Tr}\,\mathbf{S}] = \mathrm{Tr}\,\mathbf{T} + \mathcal{O}\left(\frac{1}{t}\right) \tag{25}$$

3) The resulting large system approximation of I*:* The ergodic mutual information I can be written in terms of the resolvent $\mathbf{S}$:

$$I = \mathbb{E}_{\mathbf{H}}\left[\log\left|\mathbf{I}_r + \frac{\mathbf{H}\mathbf{H}^H}{\sigma^2}\right|\right] = \mathbb{E}_{\mathbf{H}}\left[\log\left|\sigma^2\mathbf{S}(\sigma^2)\right|^{-1}\right].$$

As the differential of $g(\mathbf{A}) = \log|\mathbf{A}|$ is given by $g(\mathbf{A}+\delta\mathbf{A}) = g(\mathbf{A}) + \mathrm{Tr}[\mathbf{A}^{-1}\delta\mathbf{A}] + o(\|\delta\mathbf{A}\|)$, we obtain:

$$\begin{aligned} \frac{dI}{d\sigma^2} &= -\mathbb{E}_{\mathbf{H}}\left[\frac{\mathrm{Tr}[\mathbf{S}(\sigma^2)\mathbf{H}\mathbf{H}^H]}{\sigma^2}\right] \\ &= -\mathbb{E}_{\mathbf{H}}\left[\frac{\mathrm{Tr}[\mathbf{I}_r - \sigma^2\mathbf{S}(\sigma^2)]}{\sigma^2}\right], \end{aligned}$$

where the last equality follows from the so-called resolvent identity

$$\sigma^2\mathbf{S}(\sigma^2) = \mathbf{I}_r - \mathbf{S}(\sigma^2)\mathbf{H}\mathbf{H}^H. \tag{26}$$

The resolvent identity is inferred easily from the definition of $\mathbf{S}(\sigma^2)$. As $I(\sigma^2 = +\infty) = 0$, we now have the following expression of mutual information:

$$I(\sigma^2) = \int_{\sigma^2}^{+\infty}\left(\frac{r}{\rho} - \mathbb{E}_{\mathbf{H}}\left[\mathrm{Tr}\ \mathbf{S}(\rho)\right]\right)d\rho.$$

This equality clearly justifies the search of a large system equivalent of $\mathbb{E}_{\mathbf{H}}[\mathrm{Tr}\,\mathbf{S}]$ done in the previous sections. The term under the integral sign can be written as

$$\frac{r}{\sigma^2} - \mathbb{E}_{\mathbf{H}}[\mathrm{Tr}\,\mathbf{S}] = t\sum_{l=1}^L \tilde{\delta}_l\delta_l + \mathbb{E}_{\mathbf{H}}\left[\mathrm{Tr}\left(\mathbf{T} - \mathbf{S}\right)\right],$$

as $\frac{r}{\sigma^2} - \mathrm{Tr}\,\mathbf{T} = \mathrm{Tr}\left[((\sigma^2\mathbf{T})^{-1} - \mathbf{I}_r)\mathbf{T}\right] = \mathrm{Tr}\left[(\sum_l \tilde{\delta}_l\mathbf{C}^{(l)})\mathbf{T}\right] = t\sum_l \tilde{\delta}_l\delta_l$. We need to integrate $\varepsilon(t, \sigma^2) = \mathbb{E}_{\mathbf{H}}\left[\mathrm{Tr}\left(\mathbf{T} - \mathbf{S}\right)\right]$ with respect to σ^2 on $(\rho > 0, +\infty)$. We therefore introduce the following proposition:

Proposition 3: $\varepsilon(t, \sigma^2) = \mathbb{E}_{\mathbf{H}}\left[\mathrm{Tr}\left(\mathbf{T} - \mathbf{S}\right)\right]$ is integrable with respect to σ^2 on $(\rho > 0, +\infty)$ and

$$\int_{\rho}^{+\infty}\varepsilon(t, \sigma^2)d\sigma^2 = \mathcal{O}\left(\frac{1}{t}\right).$$

Proof: We prove in Appendix D that there exists t_0 such that, for $t > t_0$, $|\varepsilon(t, \sigma^2)| \leq \frac{1}{\sigma^8 t}P\left(\frac{1}{\sigma^2}\right)$, where P is a polynomial whose coefficients are real positive and do not depend on σ^2 nor on t. Therefore $\int_{\rho}^{+\infty}\varepsilon(t, \sigma^2)d\sigma^2 = \mathcal{O}\left(\frac{1}{t}\right)$. □

We now prove that the term $t\sum_l \tilde{\delta}_l\delta_l$ corresponds to the derivative of $\overline{I}(\sigma^2)$ with respect to σ^2. To this end, we consider the function $\mathcal{V}_0(\sigma^2, \boldsymbol{\kappa}, \tilde{\boldsymbol{\kappa}})$ defined by

$$\mathcal{V}_0(\sigma^2, \boldsymbol{\kappa}, \tilde{\boldsymbol{\kappa}}) = \log|\mathbf{I} + \mathbf{C}(\tilde{\boldsymbol{\kappa}})| + \log|\mathbf{I} + \tilde{\mathbf{C}}(\boldsymbol{\kappa})| - \sigma^2 t\sum_{l=1}^L \kappa_l\tilde{\kappa}_l,$$

where $\tilde{\mathbf{C}}(\boldsymbol{\kappa}) = \sum_{l=1}^L \kappa_l\tilde{\mathbf{C}}^{(l)}$ and $\mathbf{C}(\tilde{\boldsymbol{\kappa}}) = \sum_{l=1}^L \tilde{\kappa}_l\mathbf{C}^{(l)}$. Note that $\mathcal{V}_0(\sigma^2, \boldsymbol{\delta}, \tilde{\boldsymbol{\delta}}) = \overline{I}(\sigma^2)$. The derivative of $\overline{I}(\sigma^2)$ can then be expressed in terms of the partial derivatives of $\mathcal{V}_0$.

$$\begin{aligned} \frac{d\overline{I}}{d\sigma^2} =& \frac{\partial\mathcal{V}_0}{\partial\sigma^2}(\sigma^2, \boldsymbol{\delta}, \tilde{\boldsymbol{\delta}}) + \sum_{l=1}^L \frac{\partial\mathcal{V}_0}{\partial\kappa_l}(\sigma^2, \boldsymbol{\delta}, \tilde{\boldsymbol{\delta}}) \cdot \frac{d\delta_l}{d\sigma^2} \\ &+ \sum_{l=1}^L \frac{\partial\mathcal{V}_0}{\partial\tilde{\kappa}_l}(\sigma^2, \boldsymbol{\delta}, \tilde{\boldsymbol{\delta}}) \cdot \frac{d\tilde{\delta}_l}{d\sigma^2}. \end{aligned}$$

It is straightforward to check that

$$\begin{cases} \frac{\partial\mathcal{V}_0}{\partial\kappa_l}(\sigma^2, \boldsymbol{\kappa}, \tilde{\boldsymbol{\kappa}}) = -\sigma^2 t\big(\tilde{f}_l(\boldsymbol{\kappa}, \mathbf{I}_t) - \tilde{\kappa}_l\big), \\ \frac{\partial\mathcal{V}_0}{\partial\tilde{\kappa}_l}(\sigma^2, \boldsymbol{\kappa}, \tilde{\boldsymbol{\kappa}}) = -\sigma^2 t\big(f_l(\tilde{\boldsymbol{\kappa}}) - \kappa_l\big). \end{cases} \tag{27}$$

Both partial derivatives are equal to zero at point $(\sigma^2, \boldsymbol{\delta}, \tilde{\boldsymbol{\delta}})$, as $(\boldsymbol{\delta}, \tilde{\boldsymbol{\delta}})$ verifies by definition (9) with $\mathbf{Q} = \mathbf{I}_t$. Therefore,

$$\frac{d\overline{I}}{d\sigma^2} = \frac{\partial\mathcal{V}_0}{\partial\sigma^2}(\sigma^2, \boldsymbol{\delta}, \tilde{\boldsymbol{\delta}}) = -t\sum_{l=1}^L \delta_l\tilde{\delta}_l,$$

which, together with Proposition 3, leads to $I = \overline{I} + \mathcal{O}\left(\frac{1}{t}\right)$. □

C. The Approximation $\overline{I}(\mathbf{Q})$

We now consider the dependency in $\mathbf{Q}$ of the approximation $\bar{I}(\mathbf{Q})$. We previously considered the case $\mathbf{Q} = \mathbf{I}_t$; to address the general case it is sufficient to change matrices

$(\tilde{\mathbf{C}}^{(l)})_{l=1,\ldots,L}$ into $(\mathbf{Q}^{1/2}\tilde{\mathbf{C}}^{(l)}\mathbf{Q}^{1/2})_{l=1,\ldots,L}$ in III-A and III-B. Hence the following Corollary of Theorem 2:

Corollary 1: Assume that, for every $j \in \{1,\ldots,L\}$, $\sup_t \|\mathbf{C}^{(j)}\| < +\infty$, $\sup_t \|\tilde{\mathbf{C}}^{(j)}\| < +\infty$, $\inf_t \left(\frac{1}{t}\mathrm{Tr}\,\mathbf{C}^{(j)}\right) > 0$ and $\inf_t \lambda_{\min}(\tilde{\mathbf{C}}^{(j)}) > 0$. Then, for $\mathbf{Q}$ such as $\sup_t \|\mathbf{Q}\| < +\infty$,

$$I(\mathbf{Q}) = \overline{I}(\mathbf{Q}) + \mathcal{O}\left(\frac{1}{t}\right).$$

Note that the technical assumptions on matrices $(\tilde{\mathbf{C}}^{(l)})_{l=1,\ldots,L}$ are slightly stronger than in Theorem 2 in order to ensure that $\inf_t \left(\frac{1}{t}\mathrm{Tr}\left[\mathbf{Q}\tilde{\mathbf{C}}^{(j)}\right]\right) > 0$.

We can now state an important result about the concavity of the function $\mathbf{Q} \mapsto \overline{I}(\mathbf{Q})$, a result which will be highly needed for its optimization in section IV.

Theorem 3: $\mathbf{Q} \mapsto \overline{I}(\mathbf{Q})$ is a strictly concave function over the compact set $\mathcal{C}_1$.

Proof: We here only prove the concavity of $\overline{I}(\mathbf{Q})$. The proof of the strict concavity is quite tedious, but essentially the same as in [8] section IV (see also the extended version [9]). It is therefore omitted.

Denote by $\otimes$ the Kronecker product of matrices. Let us introduce the following matrices:

$$\mathbf{\Delta}^{(l)} = \mathbf{I}_m \otimes \mathbf{C}^{(l)},\ \tilde{\mathbf{\Delta}}^{(l)} = \mathbf{I}_m \otimes \tilde{\mathbf{C}}^{(l)},\ \check{\mathbf{Q}} = \mathbf{I}_m \otimes \mathbf{Q}.$$

We now denote $\check{\mathbf{H}}(z) = \sum_{l=1}^{L} \check{\mathbf{H}}^{(l)} z^{-(l-1)}$ with $\check{\mathbf{H}}^{(l)} = \frac{1}{\sqrt{mt}}(\mathbf{\Delta}^{(l)})^{1/2}\check{\mathbf{W}}_l(\tilde{\mathbf{\Delta}}^{(l)})^{1/2}$. where $\check{\mathbf{W}}$ is a $rm \times tm$ matrix whose entries are independent and identically distributed complex circular Gaussian random variables with variance 1. Introducing $I_m(\check{\mathbf{Q}})$ the ergodic mutual information associated with channel $\check{\mathbf{H}}(z)$:

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

where $\check{\mathbf{H}} = \check{\mathbf{H}}(1) = \sum_l \check{\mathbf{H}}^{(l)}$. Using the results of [4] and Theorem 2, it is clear that $I_m(\check{\mathbf{Q}})$ admits an asymptotic approximation $\bar{I}_m(\check{\mathbf{Q}})$. Due to the block-diagonal nature of matrices $\mathbf{\Delta}^{(l)}, \tilde{\mathbf{\Delta}}^{(l)}$ and $\check{\mathbf{Q}}$, it is straightforward to show that $\delta_l(\mathbf{Q}) = \delta_l(\check{\mathbf{Q}}), \tilde{\delta}_l(\mathbf{Q}) = \tilde{\delta}_l(\check{\mathbf{Q}})$ and that, as a consequence,

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

and thus

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

As $\check{\mathbf{Q}} \mapsto I_m(\check{\mathbf{Q}})$ is concave, we can conclude that $\bar{I}(\mathbf{Q})$ is concave as a pointwise limit of concave functions. □

As $\overline{I}(\mathbf{Q})$ is strictly concave on $\mathcal{C}_1$ by Theorem 3, it admits a unique argmax that we denote $\overline{\mathbf{Q}}_*$. We recall that $I(\mathbf{Q})$ is strictly concave on $\mathcal{C}_1$ and that we denoted $\mathbf{Q}_*$ its argmax. In order to clarify the relation between $\overline{\mathbf{Q}}_*$ and $\mathbf{Q}_*$ we introduce the following proposition which establishes that the maximization of $\overline{I}(\mathbf{Q})$ is equivalent to the maximization of $I(\mathbf{Q})$ over $\mathcal{C}_1$, up to a $\mathcal{O}\left(\frac{1}{t}\right)$ term.

Proposition 4: Assume that, for every $j \in \{1,\ldots,L\}$, $\sup_t \|\mathbf{C}^{(j)}\| < +\infty$, $\sup_t \|\tilde{\mathbf{C}}^{(j)}\| < +\infty$, $\inf_t \lambda_{\min}(\mathbf{C}^{(j)}) > 0$ and $\inf_t \lambda_{\min}(\tilde{\mathbf{C}}^{(j)}) > 0$. Then

$$I(\overline{\mathbf{Q}}_*) = I(\mathbf{Q}_*) + \mathcal{O}\left(\frac{1}{t}\right).$$

Proof: The proof is very similar to the one of [8, Proposition 3]. Assuming that $\sup_t \|\overline{\mathbf{Q}}_*\| < +\infty$ and $\sup_t \|\mathbf{Q}_*\| < +\infty$ we can apply Theorem 1 on $\overline{\mathbf{Q}}_*$ and $\mathbf{Q}_*$, hence

$$\begin{aligned}&\Big(I(\mathbf{Q}_*) - I(\overline{\mathbf{Q}}_*)\Big) + \Big(\overline{I}(\overline{\mathbf{Q}}_*) - \overline{I}(\mathbf{Q}_*)\Big)\\ &\quad= \Big(I(\mathbf{Q}_*) - \overline{I}(\mathbf{Q}_*)\Big) + \Big(\overline{I}(\overline{\mathbf{Q}}_*) - I(\overline{\mathbf{Q}}_*)\Big) = \mathcal{O}\left(\frac{1}{t}\right).\end{aligned}$$

Besides $I(\mathbf{Q}_*) - I(\overline{\mathbf{Q}}_*) \geq 0$ and $\overline{I}(\overline{\mathbf{Q}}_*) - \overline{I}(\mathbf{Q}_*) \geq 0$, as $\mathbf{Q}_*$ and $\overline{\mathbf{Q}}_*$ respectively maximize $I(\mathbf{Q})$ and $\overline{I}(\mathbf{Q})$. Therefore $I(\mathbf{Q}_*) - I(\overline{\mathbf{Q}}_*) = \mathcal{O}\left(\frac{1}{t}\right)$.

One can prove $\sup_t \|\overline{\mathbf{Q}}_*\| < +\infty$ using the same arguments as in [8, Appendix III]. It essentially lies in the fact that $\overline{\mathbf{Q}}_*$ is the solution of a waterfilling algorithm, which will be shown independently from this result in next section (see Proposition 7).

Concerning $\sup_t \|\mathbf{Q}_*\| < +\infty$, the proof is identical to [8, Appendix III], one just needs to replace $\frac{\sqrt{K}}{\sqrt{K+1}}\mathbf{A}$ by $\frac{1}{\sqrt{t}}\sum_{l=2}^{L}(\mathbf{C}^{(l)})^{1/2}\mathbf{W}_l(\tilde{\mathbf{C}}^{(l)})^{1/2}$ and $\frac{1}{\sqrt{K+1}}\frac{1}{\sqrt{t}}\mathbf{C}_R^{1/2}\mathbf{W}\mathbf{C}_T^{1/2}$ by $\frac{1}{\sqrt{t}}(\mathbf{C}^{(1)})^{1/2}\mathbf{W}_1(\tilde{\mathbf{C}}^{(1)})^{1/2}$ in the definition of $\mathbf{H}$. Then S_j, defined in [8, (134)], can be written as (28) at the bottom of the page, where $\mathbf{R}_j$ has the same definition as in [8], $\mathbf{z}_{l,j}$ is the j^{th} column of matrix $\mathbf{W}_l(\tilde{\mathbf{C}}^{(l)})^{1/2}$ and $\mathbf{z}_j = \mathbf{z}_{1,j} = \mathbf{u}_j + \mathbf{u}_j^{\perp}$ with $\mathbf{u}_j$ the conditional expectation $\mathbf{u}_j = \mathbb{E}\big[\mathbf{z}_{1,j}\big|(\mathbf{z}_{1,k})_{1\leq k\leq t, k\neq j}\big]$. As the vector $\mathbf{u}_j^{\perp}$ is independent from $\mathbf{R}_j$ and from $\mathbf{z}_{l,k}$, $k = 1,\ldots,t$, $l = 2,\ldots,L$, we can easily prove that the first term of the right-hand side of (28) is a $\mathcal{O}\left(\frac{1}{t}\right)$. The second term of the right-hand side of (28) is moreover close from $\rho_j = \frac{1}{t}\big[(\tilde{\mathbf{C}}^{(1)})^{-1}\big]_{jj}^{-1}\mathrm{Tr}(\mathbf{R}_j\mathbf{C}^{(1)})$. In fact it is possible to prove that there exists a constant C_1 such that $\mathbb{E}\big[(S_j - \rho_j)^2\big] < \frac{C_1}{t}$ (see [8] for more details).

The rest of the proof of [8, Proposition 3 (ii)] can then follow.

□

IV. Maximization Algorithm

Proposition 4 shows that it is relevant to maximize $\overline{I}(\mathbf{Q})$ over $\mathcal{C}_1$. In this section we propose a maximization algorithm for the large system approximation $\overline{I}(\mathbf{Q})$. We first introduce some classical concepts and results needed for the optimization of $\mathbf{Q} \mapsto \overline{I}(\mathbf{Q})$.

Definition 1: Let ϕ be a function defined on the convex set $\mathcal{C}_1$. Let $\mathbf{P}, \mathbf{Q} \in \mathcal{C}_1$. Then ϕ is said to be differentiable in the

$$S_j = 2\mathrm{Re}\left\{\frac{1}{t}\mathbf{u}_j^{\perp H}\big(\mathbf{C}^{(1)}\big)^{1/2}\mathbf{R}_j\left(\sum_{l=2}^{L}\big(\mathbf{C}^{(l)}\big)^{1/2}\mathbf{z}_{l,j} + \big(\mathbf{C}^{(1)}\big)^{1/2}\mathbf{u}_j\right)\right\} + \frac{1}{t}\mathbf{u}_j^{\perp H}\big(\mathbf{C}^{(1)}\big)^{1/2}\mathbf{R}_j\big(\mathbf{C}^{(1)}\big)^{1/2}\mathbf{u}_j^{\perp}, \qquad (28)$$

Gâteaux sense (or Gâteaux differentiable) at point $\mathbf{Q}$ in the direction $\mathbf{P}-\mathbf{Q}$ if the following limit exists:

$$\lim_{\lambda\to 0^+}\frac{\phi(\mathbf{Q}+\lambda(\mathbf{P}-\mathbf{Q}))-\phi(\mathbf{Q})}{\lambda}.$$

In this case, this limit is noted $\langle\phi'(\mathbf{Q}),\mathbf{P}-\mathbf{Q}\rangle$.

Note that $\phi(\mathbf{Q}+\lambda(\mathbf{P}-\mathbf{Q}))$ makes sense for $\lambda\in[0,1]$, as $\mathbf{Q}+\lambda(\mathbf{P}-\mathbf{Q})=(1-\lambda)\mathbf{Q}+\lambda\mathbf{P}$ naturally belongs to $\mathcal{C}_1$. We now establish the following result:

Proposition 5: For each $\mathbf{P},\mathbf{Q}\in\mathcal{C}_1$, functions $\mathbf{Q}\mapsto\delta_l(\mathbf{Q})$, $\mathbf{Q}\mapsto\tilde{\delta}_l(\mathbf{Q})$, $l=1,\ldots,L$, as well as function $\mathbf{Q}\mapsto\overline{I}(\mathbf{Q})$ are Gâteaux differentiable at $\mathbf{Q}$ in the direction $\mathbf{P}-\mathbf{Q}$.

Proof: See Appendix E. □

In order to characterize matrix $\overline{\mathbf{Q}}_*$ we recall the following result:

Proposition 6: Let $\phi:\mathcal{C}_1\to\mathbb{R}$ be a strictly concave function. Then,

(i) ϕ is Gâteaux differentiable at $\mathbf{Q}$ in the direction $\mathbf{P}-\mathbf{Q}$ for each $\mathbf{P},\mathbf{Q}\in\mathcal{C}_1$,

(ii) $\mathbf{Q}_{opt}$ is the unique argmax of ϕ on $\mathcal{C}_1$ if and only if it verifies:

$$\forall\mathbf{Q}\in\mathcal{C}_1,\ \langle\phi'(\mathbf{Q}_{opt}),\mathbf{Q}-\mathbf{Q}_{opt}\rangle\leq 0. \tag{29}$$

This proposition is standard (see for example [13, Chapter 2]).

In order to introduce our maximization algorithm, we consider the function $\mathcal{V}(\mathbf{Q},\boldsymbol{\kappa},\tilde{\boldsymbol{\kappa}})$ defined by:

$$\begin{aligned}\mathcal{V}(\mathbf{Q},\boldsymbol{\kappa},\tilde{\boldsymbol{\kappa}})=&\log|\mathbf{I}_r+\mathbf{C}(\tilde{\boldsymbol{\kappa}})|+\log|\mathbf{I}_t+\mathbf{Q}\tilde{\mathbf{C}}(\boldsymbol{\kappa})|\\&-\sigma^2t\sum_{l=1}^{L}\kappa_l\tilde{\kappa}_l.\end{aligned} \tag{30}$$

We recall that $\tilde{\mathbf{C}}(\boldsymbol{\kappa})=\sum_l\kappa_l\tilde{\mathbf{C}}^{(l)}$ and $\mathbf{C}(\tilde{\boldsymbol{\kappa}})=\sum_l\tilde{\kappa}_l\mathbf{C}^{(l)}$. Note that we have $\mathcal{V}(\mathbf{Q},\boldsymbol{\delta}(\mathbf{Q}),\tilde{\boldsymbol{\delta}}(\mathbf{Q}))=\overline{I}(\mathbf{Q})$. We then have the following result:

Proposition 7: Denote by $\boldsymbol{\delta}_*$ and $\tilde{\boldsymbol{\delta}}_*$ the quantities $\boldsymbol{\delta}(\overline{\mathbf{Q}}_*)$ and $\tilde{\boldsymbol{\delta}}(\overline{\mathbf{Q}}_*)$. Matrix $\overline{\mathbf{Q}}_*$ is the solution of the standard waterfilling problem: maximize over $\mathbf{Q}\in\mathcal{C}_1$ the function $\log|\mathbf{I}_t+\mathbf{Q}\tilde{\mathbf{C}}(\boldsymbol{\delta}_*)|$.

Proof: We first remark that maximizing function $\mathbf{Q}\mapsto\log|\mathbf{I}+\mathbf{Q}\tilde{\mathbf{C}}(\boldsymbol{\delta}_*)|$ is equivalent to maximizing function $\mathbf{Q}\mapsto\mathcal{V}(\mathbf{Q},\boldsymbol{\delta}_*,\tilde{\boldsymbol{\delta}}_*)$ by (30). The proof then relies on the observation hereafter proven that, for each $\mathbf{P}\in\mathcal{C}_1$,

$$\langle\overline{I}'(\overline{\mathbf{Q}}_*),\mathbf{P}-\overline{\mathbf{Q}}_*\rangle=\langle\mathcal{V}'(\overline{\mathbf{Q}}_*,\boldsymbol{\delta}_*,\tilde{\boldsymbol{\delta}}_*),\mathbf{P}-\overline{\mathbf{Q}}_*\rangle, \tag{31}$$

where $\langle\mathcal{V}'(\overline{\mathbf{Q}}_*,\boldsymbol{\delta}_*,\tilde{\boldsymbol{\delta}}_*),\mathbf{P}-\overline{\mathbf{Q}}_*\rangle$ is the Gâteaux differential of function $\mathbf{Q}\mapsto\mathcal{V}(\mathbf{Q},\boldsymbol{\delta}_*,\tilde{\boldsymbol{\delta}}_*)$ at point $\overline{\mathbf{Q}}_*$ in direction $\mathbf{P}-\overline{\mathbf{Q}}_*$. Assuming (31) is verified, (29) yields that $\langle\mathcal{V}'(\overline{\mathbf{Q}}_*,\boldsymbol{\delta}_*,\tilde{\boldsymbol{\delta}}_*),\mathbf{P}-\overline{\mathbf{Q}}_*\rangle\leq 0$ for each matrix $\mathbf{P}\in\mathcal{C}_1$. And as the function $\mathbf{Q}\mapsto\mathcal{V}(\mathbf{Q},\boldsymbol{\delta}_*,\tilde{\boldsymbol{\delta}}_*)$ is strictly concave on $\mathcal{C}_1$, its unique argmax on $\mathcal{C}_1$ coincides with $\overline{\mathbf{Q}}_*$.

It now remains to prove (31). Consider $\mathbf{P},\mathbf{Q}\in\mathcal{C}_1$. Then,

$$\begin{aligned}\langle\overline{I}'(\mathbf{Q}),\mathbf{P}-\mathbf{Q}\rangle=&\langle\mathcal{V}'(\mathbf{Q},\boldsymbol{\delta}(\mathbf{Q}),\tilde{\boldsymbol{\delta}}(\mathbf{Q})),\mathbf{P}-\mathbf{Q}\rangle\\&+\sum_{l=1}^{L}\frac{\partial\mathcal{V}}{\partial\kappa_l}(\mathbf{Q},\boldsymbol{\delta}(\mathbf{Q}),\tilde{\boldsymbol{\delta}}(\mathbf{Q}))\langle\delta_l'(\mathbf{Q}),\mathbf{P}-\mathbf{Q}\rangle\\&+\sum_{l=1}^{L}\frac{\partial\mathcal{V}}{\partial\tilde{\kappa}_l}(\mathbf{Q},\boldsymbol{\delta}(\mathbf{Q}),\tilde{\boldsymbol{\delta}}(\mathbf{Q}))\langle\tilde{\delta}_l'(\mathbf{Q}),\mathbf{P}-\mathbf{Q}\rangle.\end{aligned} \tag{32}$$

Similarly to (27), partial derivatives $\frac{\partial\mathcal{V}}{\partial\kappa_l}(\mathbf{Q},\boldsymbol{\kappa},\tilde{\boldsymbol{\kappa}})=-\sigma^2t\big(\tilde{f}_l(\boldsymbol{\kappa},\mathbf{Q})-\tilde{\kappa}_l\big)$ and $\frac{\partial\mathcal{V}}{\partial\tilde{\kappa}_l}(\mathbf{Q},\boldsymbol{\kappa},\tilde{\boldsymbol{\kappa}})=-\sigma^2t\big(f_l(\tilde{\boldsymbol{\kappa}})-\kappa_l\big)$ are equal to zero at point $(\mathbf{Q},\boldsymbol{\delta}(\mathbf{Q}),\tilde{\boldsymbol{\delta}}(\mathbf{Q}))$, as $(\boldsymbol{\delta}(\mathbf{Q}),\tilde{\boldsymbol{\delta}}(\mathbf{Q}))$ verifies (9) by definition. Therefore, letting $\mathbf{Q}=\overline{\mathbf{Q}}_*$ in (32) yields:

$$\langle\overline{I}'(\overline{\mathbf{Q}}_*),\mathbf{P}-\overline{\mathbf{Q}}_*\rangle=\langle\mathcal{V}'(\overline{\mathbf{Q}}_*,\boldsymbol{\delta}(\overline{\mathbf{Q}}_*),\tilde{\boldsymbol{\delta}}(\overline{\mathbf{Q}}_*)),\mathbf{P}-\overline{\mathbf{Q}}_*\rangle.$$

□

Proposition 7 shows that the optimum matrix is solution of a waterfilling problem associated to the covariance matrix $\tilde{\mathbf{C}}(\boldsymbol{\delta}_*)$. This result cannot be used to evaluate $\overline{\mathbf{Q}}_*$, because the matrix $\tilde{\mathbf{C}}(\boldsymbol{\delta}_*)$ itself depends of $\overline{\mathbf{Q}}_*$. However, it provides some insight on the structure of the optimum matrix: the eigenvectors of $\overline{\mathbf{Q}}_*$ coincide with the eigenvectors of a linear combination of matrices $\tilde{\mathbf{C}}^{(l)}$, the $\delta_l(\mathbf{Q}_*)$ being the coefficients of this linear combination. This is in line with the result of [4, Appendix VI].

We now introduce our iterative algorithm for optimizing $\overline{I}(\mathbf{Q})$:

- Initialization: $\mathbf{Q}_0=\mathbf{I}$.
- Evaluation of $\mathbf{Q}_k$ from $\mathbf{Q}_{k-1}$: $(\boldsymbol{\delta}^{(k)},\tilde{\boldsymbol{\delta}}^{(k)})$ is defined as the unique solution of (9) in which $\mathbf{Q}=\mathbf{Q}_{k-1}$. Then $\mathbf{Q}_k$ is defined as the maximum of function $\mathbf{Q}\mapsto\log\big|\mathbf{I}_t+\mathbf{Q}\tilde{\mathbf{C}}(\boldsymbol{\delta}^{(k)})\big|$ on $\mathcal{C}_1$.

We now establish a result which implies that, if the algorithm converges, then it converges towards the optimal covariance matrix $\overline{\mathbf{Q}}_*$.

Proposition 8: Assume that

$$\lim_{k\to\infty}\boldsymbol{\delta}^{(k)}-\boldsymbol{\delta}^{(k-1)}=\lim_{k\to\infty}\tilde{\boldsymbol{\delta}}^{(k)}-\tilde{\boldsymbol{\delta}}^{(k-1)}=0. \tag{33}$$

Then, the algorithm converges towards matrix $\overline{\mathbf{Q}}_*$.

Proof: The sequence $(\mathbf{Q}_k)$ belongs to the set $\mathcal{C}_1$. As $\mathcal{C}_1$ is compact, we just have to verify that every convergent subsequence $(\mathbf{Q}_{\psi(k)})_{k\in\mathbb{N}}$ extracted from $(\mathbf{Q}_k)_{k\in\mathbb{N}}$ converges towards $\overline{\mathbf{Q}}_*$. For this, we denote by $\overline{\mathbf{Q}}_{\psi,*}$ the limit of the above subsequence, and prove that this matrix verifies property (29) with $\phi=\overline{I}$. Vectors $\boldsymbol{\delta}^{\psi(k)+1}$ and $\tilde{\boldsymbol{\delta}}^{\psi(k)+1}$ are defined as the solutions of (9) with $\mathbf{Q}=\mathbf{Q}_{\psi(k)}$. Hence, due to the continuity of functions $\mathbf{Q}\mapsto\delta_l(\mathbf{Q})$ and $\mathbf{Q}\mapsto\tilde{\delta}_l(\mathbf{Q})$, sequences $(\boldsymbol{\delta}^{\psi(k)+1})_{k\in\mathbb{N}}$ and $(\tilde{\boldsymbol{\delta}}^{\psi(k)+1})_{k\in\mathbb{N}}$ converge towards $\boldsymbol{\delta}^{\psi,*}=\boldsymbol{\delta}(\overline{\mathbf{Q}}_{\psi,*})$ and $\tilde{\boldsymbol{\delta}}^{\psi,*}=\tilde{\boldsymbol{\delta}}(\overline{\mathbf{Q}}_{\psi,*})$ respectively. Moreover, $(\boldsymbol{\delta}^{\psi,*},\tilde{\boldsymbol{\delta}}^{\psi,*})$ is solution of system (9) in which matrix $\mathbf{Q}$ coincides with $\overline{\mathbf{Q}}_{\psi,*}$. Therefore,

$$\frac{\partial\mathcal{V}}{\partial\kappa_l}\left(\overline{\mathbf{Q}}_{\psi,*},\boldsymbol{\delta}^{\psi,*},\tilde{\boldsymbol{\delta}}^{\psi,*}\right)=\frac{\partial\mathcal{V}}{\partial\tilde{\kappa}_l}\left(\overline{\mathbf{Q}}_{\psi,*},\boldsymbol{\delta}^{\psi,*},\tilde{\boldsymbol{\delta}}^{\psi,*}\right)=0.$$

As in the proof of Proposition 7, this leads to

$$\langle\overline{I}'(\overline{\mathbf{Q}}_{\psi,*}),\mathbf{P}-\overline{\mathbf{Q}}_{\psi,*}\rangle=\langle\mathcal{V}'(\overline{\mathbf{Q}}_{\psi,*},\boldsymbol{\delta}_{\psi,*},\tilde{\boldsymbol{\delta}}_{\psi,*}),\mathbf{P}-\overline{\mathbf{Q}}_{\psi,*}\rangle \tag{34}$$

for every $\mathbf{P}\in\mathcal{C}_1$. It remains to show that the right-hand side of (34) is negative to complete the proof. For this, we use that $\mathbf{Q}_{\psi(k)}$ is the argmax over $\mathcal{C}_1$ of function $\mathbf{Q}\mapsto\mathcal{V}\big(\mathbf{Q},\boldsymbol{\delta}^{\psi(k)},\tilde{\boldsymbol{\delta}}^{\psi(k)}\big)$. Therefore,

$$\langle\mathcal{V}'(\mathbf{Q}_{\psi(k)},\boldsymbol{\delta}_{\psi(k)},\tilde{\boldsymbol{\delta}}_{\psi(k)}),\mathbf{P}-\mathbf{Q}_{\psi(k)}\rangle\leq 0\ \ \forall\ \mathbf{P}\in\mathcal{C}_1. \tag{35}$$

By condition (33), sequences $(\boldsymbol{\delta}_{\psi(k)})$ and $(\tilde{\boldsymbol{\delta}}_{\psi(k)})$ also converge towards $\boldsymbol{\delta}^{\psi,*}$ and $\tilde{\boldsymbol{\delta}}^{\psi,*}$ respectively. Taking the limit of (35) when $k \to \infty$ eventually shows that $\langle \mathcal{V}'(\overline{\mathbf{Q}}_{\psi,*}, \boldsymbol{\delta}_{\psi,*}, \tilde{\boldsymbol{\delta}}_{\psi,*}), \mathbf{P} - \overline{\mathbf{Q}}_{\psi,*} \rangle \leq 0$ as required. □

To conclude, if the algorithm is convergent, that is, if the sequence of $(\mathbf{Q}_k)_{k\in\mathbb{N}}$ converges towards a certain matrix, then the $\delta_l^{(k)} = \delta_l(\mathbf{Q}_{k-1})$ and the $\tilde{\delta}_l^{(k)} = \tilde{\delta}_l(\mathbf{Q}_{k-1})$ converge as well when $k \to \infty$. Condition (33) is then verified, hence, if the algorithm is convergent, it converges towards $\overline{\mathbf{Q}}_*$. Although the convergence of the algorithm has not been proved, this result is encouraging and suggests that the algorithm is reliable. In particular, in all the conducted simulations the algorithm was converging. In any case, condition (33) can be easily checked. If it is not satisfied, it is possible to modify the initial point $\mathbf{Q}_0$ as many times as needed to ensure the convergence.

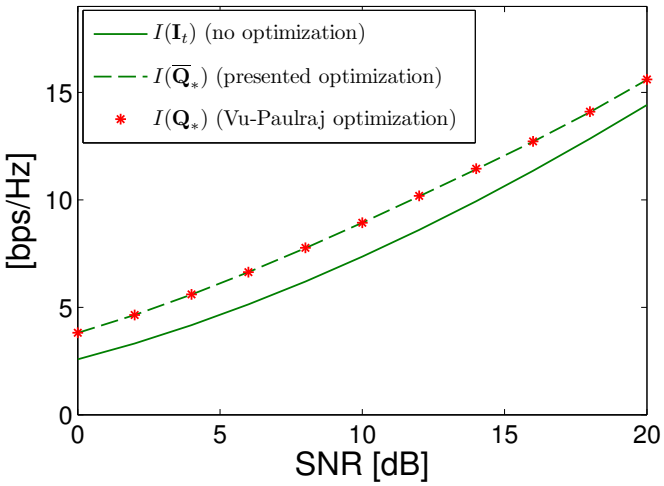
V. Numerical Results

We provide here some simulations results to evaluate the performance of the proposed approach. We use the propagation model introduced in [3], in which each path corresponds to a scatterer cluster characterized by a mean angle of departure, a mean angle of arrival and an angle spread for each of these two angles.

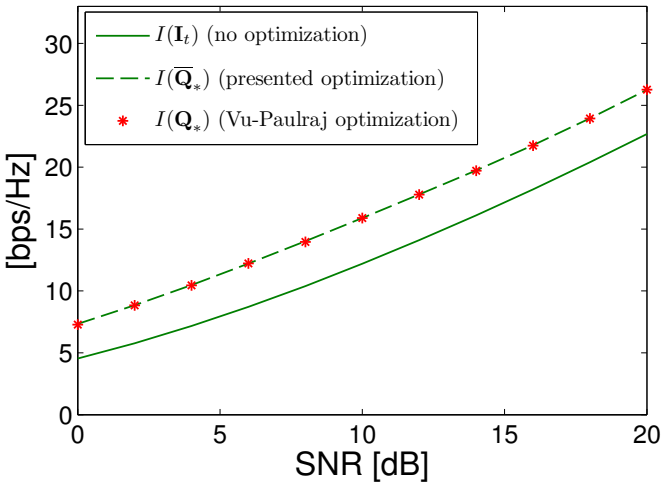
In the featured simulations for Fig. 1(a) (respectively Fig. 1(b)), we consider a frequency selective MIMO system with $r = t = 4$ (respectively $r = t = 8$), a carrier frequency of 2GHz, a number of paths $L = 5$. The paths share the same power, and their mean departure angles and angles spreads are given in Table I in radians. In both Fig. 1(a) and 1(b), we have represented the EMI $I(\mathbf{I}_t)$ (i.e. without optimization), and the optimized EMI $I(\overline{\mathbf{Q}}_*)$ (i.e. with an input covariance matrix maximizing the approximation $\overline{I}$). The EMI are evaluated by Monte-Carlo simulations, with $2 \cdot 10^4$ channel realizations. The EMI optimized with Vu-Paulraj algorithm [5] is also represented for comparison.

Vu-Paulraj's algorithm is composed of two nested iterative loops. The inner loop evaluates $\mathbf{Q}_*^{(n)} = \operatorname{argmax}\{I(\mathbf{Q}) + k_{\text{barrier}} \log|\mathbf{Q}|\}$ thanks to the Newton algorithm with the constraint $\frac{1}{t}\operatorname{Tr}\mathbf{Q} = 1$, for a given value of k_{barrier} and a given starting point $\mathbf{Q}_0^{(n)}$. Maximizing $I(\mathbf{Q}) + k_{\text{barrier}} \log|\mathbf{Q}|$ instead of $I(\mathbf{Q})$ ensures that $\mathbf{Q}$ remains positive semi-definite through the steps of the Newton algorithm; this is the so-called barrier interior-point method. The outer loop then decreases k_{barrier} by a certain constant factor μ and gives the inner loop the next starting point $\mathbf{Q}_0^{(n+1)} = \mathbf{Q}_*^{(n)}$. The algorithm stops when the desired precision is obtained, or, as the Newton algorithm requires heavy Monte-Carlo simulations for the evaluation of the gradient and of the Hessian of $I(\mathbf{Q})$, when the number of iterations of the outer loop reaches a given number $N_{\max}$. As in [5] we took $N_{\max} = 10$, $\mu = 100$, $2 \cdot 10^4$ trials for the Monte-Carlo simulations, and we started with $k_{\text{barrier}} = \frac{1}{100}$.

Both Fig. 1(a) and 1(b) show that maximizing $\overline{I}(\mathbf{Q})$ over the input covariance leads to significant improvement for $I(\mathbf{Q})$. Our approach provides the same results as Vu-Paulraj's algorithm. Moreover our algorithm is computationally much more efficient: in Vu-Paulraj's algorithm, the evaluation of the gradient and of the Hessian of $I(\mathbf{Q})$ needs heavy Monte-Carlo simulations. Table II gives for both algorithms the average execution time in seconds to obtain the input covariance matrix, on a 3.16GHz Intel Xeon CPU with 8GB of RAM, for a number of paths $L = 3$, $L = 4$ and $L = 5$, given $r = t = 4$.



(a) $r = t = 4$



(b) $r = t = 8$

Fig. 1. Comparison with Vu-Paulraj algorithm

TABLE I
PATHS ANGULAR PARAMETERS *(in radians)*

	$l=1$	$l=2$	$l=3$	$l=4$	$l=5$
mean departure angle	6.15	3.52	4.04	2.58	2.66
departure angle spread	0.06	0.09	0.05	0.05	0.03
mean arrival angle	4.85	3.48	1.71	5.31	0.06
arrival angle spread	0.06	0.08	0.05	0.02	0.11

VI. Conclusion

In this paper we have addressed the evaluation of the capacity achieving covariance matrices of frequency selective MIMO channels. We have first clarified the definition of

TABLE II
AVERAGE EXECUTION TIME *(in seconds)*

	$L=3$	$L=4$	$L=5$
Vu-Paulraj	681	884	1077
New algorithm	$7.0\cdot10^{-3}$	$7.4\cdot10^{-3}$	$8.3\cdot10^{-3}$

the large system approximation of the EMI and rigorously proved its expression and convergence speed with Gaussian methods. We have then proposed to optimize the EMI through this approximation, and have introduced an attractive iterative algorithm based on an iterative waterfilling scheme. Numerical results have shown that our approach provides the same results as a direct approach, but in a more efficient way in terms of computation time.

APPENDIX A
PROOF OF THE EXISTENCE OF A SOLUTION

To study (9), it is quite useful to interpret functions f_l and $\tilde{f}_l$ as functions of the parameter $-\sigma^2 \in \mathbb{R}^-$, to extend their domain of validity from $\mathbb{R}^-$ to $\mathbb{C}-\mathbb{R}^+$, and to use powerful results concerning certain class of analytic functions. We therefore define the functions $g(\tilde{\boldsymbol{\psi}})(z)$ and $g(\boldsymbol{\psi})(z)$, with $\boldsymbol{\psi}(z)=[\psi_1(z),...,\psi_L(z)]^T$, $\tilde{\boldsymbol{\psi}}(z)=[\tilde{\psi}_1(z),...,\tilde{\psi}_L(z)]^T$, as

$$g(\tilde{\boldsymbol{\psi}})(z)=\begin{bmatrix} g_1(\tilde{\boldsymbol{\psi}})(z) \\ ... \\ g_L(\tilde{\boldsymbol{\psi}})(z)\end{bmatrix},\ \tilde{g}(\boldsymbol{\psi})(z)=\begin{bmatrix} \tilde{g}_1(\boldsymbol{\psi})(z) \\ ... \\ \tilde{g}_L(\boldsymbol{\psi})(z)\end{bmatrix},$$

where functions $g_l(\tilde{\boldsymbol{\psi}})$ and $\tilde{g}_l(\boldsymbol{\psi})$ are defined by $g_l(\tilde{\boldsymbol{\psi}})(z)=\frac{1}{t}\mathrm{Tr}\big[\mathbf{C}^{(l)}\mathbf{T}^{\tilde{\boldsymbol{\psi}}}(z)\big]$, $\tilde{g}_l(\boldsymbol{\psi})(z)=\frac{1}{t}\mathrm{Tr}\big[\tilde{\mathbf{C}}^{(l)}\tilde{\mathbf{T}}^{\boldsymbol{\psi}}(z)\big]$. Matrices $\mathbf{T}^{\tilde{\boldsymbol{\psi}}}(z)$ and $\tilde{\mathbf{T}}^{\boldsymbol{\psi}}(z)$ are defined by

$$\mathbf{T}^{\tilde{\boldsymbol{\psi}}}(z)=\left[-z\left(\mathbf{I}_r+\sum_{j=1}^{L}\tilde{\psi}_j(z)\mathbf{C}^{(j)}\right)\right]^{-1}, \quad (36)$$

$$\tilde{\mathbf{T}}^{\boldsymbol{\psi}}(z)=\left[-z\left(\mathbf{I}_t+\sum_{j=1}^{L}\psi_j(z)\tilde{\mathbf{C}}^{(j)}\right)\right]^{-1}. \quad (37)$$

In order to explain the following results, we now have to introduce the concept of Stieltjès transforms.

Definition 2: Let μ be a finite[2] positive measure carried by $\mathbb{R}^+$. The Stieltjès transform of μ is the function $s(z)$ defined for $z\in\mathbb{C}-\mathbb{R}^+$ by

$$s(z)=\int_{\mathbb{R}^+}\frac{d\mu(\lambda)}{\lambda-z}. \quad (38)$$

In the following, the class of all Stieltjès transforms of finite positive measures carried by $\mathbb{R}^+$ is denoted $\mathcal{S}(\mathbb{R}^+)$. We now state some of the properties of the elements of $\mathcal{S}(\mathbb{R}^+)$.

Proposition 9: Let $s(z)\in\mathcal{S}(\mathbb{R}^+)$, and μ its associated measure. Then we have the following results:

(i) $s(z)$ is analytic on $\mathbb{C}-\mathbb{R}^+$,
(ii) $\mathrm{Im}(s(z))>0$ if $\mathrm{Im}(z)>0$, and $\mathrm{Im}(s(z))<0$ if $\mathrm{Im}(z)<0$,
(iii) $\mathrm{Im}(zs(z))>0$ if $\mathrm{Im}(z)>0$, and $\mathrm{Im}(zs(z))<0$ if $\mathrm{Im}(z)<0$,
(iv) $s(-\sigma^2)>0$ for $\sigma^2>0$,
(v) $|s(z)|\le\frac{\mu(\mathbb{R}^+)}{d(z,\mathbb{R}^+)}$ for $z\in\mathbb{C}-\mathbb{R}^+$,
(vi) $\mu(\mathbb{R}^+)=\lim_{y\to\infty}-iy\,s(iy)$.

[2]finite means that $\mu(\mathbb{R}^+)<\infty$

Proof: All the stated properties are standard material, see e.g. Appendix of [14]. □

Conversely, a useful tool to prove that a certain function belongs to $\mathcal{S}(\mathbb{R}^+)$ is the following proposition:

Proposition 10: Let s be a function holomorphic on $\mathbb{C}-\mathbb{R}^+$ which verifies the three following properties:

(i) $\mathrm{Im}(s(z))>0$ if $\mathrm{Im}(z)>0$,
(ii) $\mathrm{Im}(zs(z))>0$ if $\mathrm{Im}(z)>0$,
(iii) $\sup_{y>0}|iy\,s(iy)|<\infty$.

Then $s\in\mathcal{S}(\mathbb{R}^+)$, and if μ represents the corresponding positive measure, then $\mu(\mathbb{R}^+)=\lim_{y\to\infty}(-iy\,s(iy))$.

Proof: see Appendix of [14]. □

Now that we have recalled the notion of Stieltjès transforms and its associated basic properties we can introduce the following proposition:

Proposition 11: Let $(\psi_l,\tilde{\psi}_l)_{l=1,...,L}\in\mathcal{S}(\mathbb{R}^+)$. We define functions $\varphi_l(z)$ and $\tilde{\varphi}_l(z)$, $l=1,\dots,L$, as

$$\begin{cases}\varphi_l(z)=\frac{1}{t}\mathrm{Tr}\big[\mathbf{C}^{(l)}\mathbf{T}^{\tilde{\boldsymbol{\psi}}}(z)\big],\\ \tilde{\varphi}_l(z)=\frac{1}{t}\mathrm{Tr}\big[\tilde{\mathbf{C}}^{(l)}\tilde{\mathbf{T}}^{\boldsymbol{\psi}}(z)\big].\end{cases}$$

Then we have the following results:

(i) $\mathbf{T}^{\tilde{\boldsymbol{\psi}}}$, $\tilde{\mathbf{T}}^{\boldsymbol{\psi}}$ are holomorphic on $\mathbb{C}-\mathbb{R}^+$,
(ii) $\|\mathbf{T}^{\tilde{\boldsymbol{\psi}}}(z)\|\le\frac{1}{d(z,\mathbb{R}^+)}$, $\|\tilde{\mathbf{T}}^{\boldsymbol{\psi}}(z)\|\le\frac{1}{d(z,\mathbb{R}^+)}$ on $\mathbb{C}-\mathbb{R}^+$,
(iii) $\varphi_l\in\mathcal{S}(\mathbb{R}^+)$ with the corresponding mass μ_l verifying $\mu_l(\mathbb{R}^+)=\frac{1}{t}\mathrm{Tr}\,\mathbf{C}^{(l)}$, $\tilde{\varphi}_l\in\mathcal{S}(\mathbb{R}^+)$ with the corresponding mass $\tilde{\mu}_l$ verifying $\tilde{\mu}_l(\mathbb{R}^+)=\frac{1}{t}\mathrm{Tr}\,\tilde{\mathbf{C}}^{(l)}$.

Proof: For item (i) we only have to check that $z\big(\mathbf{I}_r+\sum_{j=1}^{L}\tilde{\psi}_j(z)\mathbf{C}^{(j)}\big)$ is invertible for every $z\in\mathbb{C}-\mathbb{R}^+$ to prove that $\mathbf{T}^{\tilde{\boldsymbol{\psi}}}$ is holomorphic on $\mathbb{C}-\mathbb{R}^+$. The key point is to notice that, for any vector $\mathbf{v}$, for z such that $\mathrm{Im}(z)>0$,

$$\mathrm{Im}\Big\{\mathbf{v}^H z\Big(\mathbf{I}_r+\sum_{j=1}^{L}\tilde{\psi}_j(z)\mathbf{C}^{(j)}\Big)\mathbf{v}\Big\}$$
$$=\mathrm{Im}\{z\}\mathbf{v}^H\mathbf{v}+\sum_{j=1}^{L}\mathrm{Im}\left\{z\tilde{\psi}_j(z)\right\}\mathbf{v}^H\mathbf{C}^{(j)}\mathbf{v}>0.$$

A similar inequality holds for $\mathrm{Im}(z)<0$, and the case $z\in\mathbb{R}^-$ is straightforward.

Item (iii) can easily be proved thanks to Proposition 10.

As for item (ii), the proof is essentially the same as the proof of Proposition 5.1 item 3 in [15], and is therefore omitted. □

We consider the following iterative scheme:

$$\begin{cases}\boldsymbol{\psi}^{(n+1)}(z)=g(\tilde{\boldsymbol{\psi}}^{(n)})(z),\\ \tilde{\boldsymbol{\psi}}^{(n+1)}(z)=\tilde{g}(\boldsymbol{\psi}^{(n)})(z),\end{cases} \quad (39)$$

with a starting point $(\boldsymbol{\psi}^{(0)}(z),\tilde{\boldsymbol{\psi}}^{(0)}(z))$ in $(\mathcal{S}(\mathbb{R}^+))^{2L}$. Item (iii) of Proposition 11 then ensures that, for each $n\ge1$,

$\boldsymbol{\psi}^{(n)}(z)$ and $\tilde{\boldsymbol{\psi}}^{(n)}(z)$ belong to $(\mathcal{S}(\mathbb{R}^+))^L$. Moreover,

$$\begin{aligned}\left|(\psi_l^{(n+1)} - \psi_l^{(n)})(z)\right| &= \left|g_l(\boldsymbol{\psi}^{(n)})(z) - g_l(\boldsymbol{\psi}^{(n-1)})(z)\right| \\ &= \frac{1}{t}\left|\mathrm{Tr}\Big[\mathbf{C}^{(l)}(\mathbf{T}^{(n)}(z) - \mathbf{T}^{(n-1)}(z))\Big]\right|, \end{aligned} \quad (40)$$

where matrices $\mathbf{T}^{(n)}(z)$ and $\tilde{\mathbf{T}}^{(n)}(z)$ are defined by $\mathbf{T}^{(n)}(z) = \mathbf{T}^{\tilde{\boldsymbol{\psi}}^{(n)}}(z)$, $\tilde{\mathbf{T}}^{(n)}(z) = \tilde{\mathbf{T}}^{\boldsymbol{\psi}^{(n)}}(z)$. Note that in the following we may not always mention the dependency in z of $\mathbf{T}^{(n)}$, $\tilde{\mathbf{T}}^{(n)}$, $\psi_j^{(n)}$ and $\tilde{\psi}_j^{(n)}$ for reading ease. Using the equality $\mathbf{A} - \mathbf{B} = \mathbf{A}\left(\mathbf{B}^{-1} - \mathbf{A}^{-1}\right)\mathbf{B}$, we then obtain:

$$\begin{aligned}&\mathbf{T}^{(n)} - \mathbf{T}^{(n-1)} \\ &= \mathbf{T}^{(n)}\Big(- z\sum_{j=1}^{L}\left(\tilde{\psi}_j^{(n-1)} - \tilde{\psi}_j^{(n)}\right)\mathbf{C}^{(j)}\Big)\mathbf{T}^{(n-1)}. \end{aligned} \quad (41)$$

Using (41) in (40) then yields:

$$\begin{aligned}&\left|\psi_l^{(n+1)} - \psi_l^{(n)}\right| \\ &= \frac{|z|}{t}\left|\sum_{j=1}^{L}\left(\tilde{\psi}_j^{(n-1)} - \tilde{\psi}_j^{(n)}\right)\mathrm{Tr}\left[\mathbf{C}^{(l)}\mathbf{T}^{(n)}\mathbf{C}^{(j)}\mathbf{T}^{(n-1)}\right]\right| \\ &\leq \frac{|z|}{t}\sum_{j=1}^{L}\left|\tilde{\psi}_j^{(n-1)} - \tilde{\psi}_j^{(n)}\right|\left|\mathrm{Tr}\left[\mathbf{C}^{(l)}\mathbf{T}^{(n)}\mathbf{C}^{(j)}\mathbf{T}^{(n-1)}\right]\right|. \end{aligned}$$

The trace in the above expression can be bounded with the help of $C_{\max} = \max_j\{\|\mathbf{C}^{(j)}\|, \|\tilde{\mathbf{C}}^{(j)}\|\}$:

$$\begin{aligned}&\left|\psi_l^{(n+1)} - \psi_l^{(n)}\right| \\ &\leq |z|\frac{r}{t}\sum_{j=1}^{L}\left|\tilde{\psi}_j^{(n)} - \tilde{\psi}_j^{(n-1)}\right|\|\mathbf{C}^{(l)}\|\|\mathbf{T}^{(n)}\|\|\mathbf{C}^{(j)}\|\|\mathbf{T}^{(n-1)}\| \\ &\leq |z|C_{\max}^2\frac{r}{t}\|\mathbf{T}^{(n)}\|\|\mathbf{T}^{(n-1)}\|\sum_{j=1}^{L}\left|\tilde{\psi}_j^{(n)} - \tilde{\psi}_j^{(n-1)}\right|. \end{aligned}$$

For $z \in \mathbb{C} - \mathbb{R}^+$, $\mathbf{T}^{(n)}(z)$ and $\mathbf{T}^{(n-1)}(z)$ have a spectral norm less than $\frac{1}{d(z,\mathbb{R}^+)}$ by item (ii) of Proposition 11. Therefore,

$$\begin{aligned}&\left|(\psi_l^{(n+1)} - \psi_l^{(n)})(z)\right| \\ &\quad\leq \frac{rC_{\max}^2}{t}\frac{|z|}{(d(z,\mathbb{R}^+))^2}\sum_{j=1}^{L}\left|\left(\tilde{\psi}_j^{(n)} - \tilde{\psi}_j^{(n-1)}\right)(z)\right|. \end{aligned} \quad (42)$$

A similar computation leads to

$$\begin{aligned}&\left|(\tilde{\psi}_j^{(n+1)} - \tilde{\psi}_j^{(n)})(z)\right| \\ &\quad\leq C_{\max}^2\frac{|z|}{(d(z,\mathbb{R}^+))^2}\sum_{l=1}^{L}\left|\left(\psi_l^{(n)} - \psi_l^{(n-1)}\right)(z)\right|. \end{aligned} \quad (43)$$

We now introduce the following maximum:

$$M^{(n)}(z) = \max_j\left\{\left|(\psi_j^{(n+1)} - \psi_j^{(n)})(z)\right|, \left|(\tilde{\psi}_j^{(n+1)} - \tilde{\psi}_j^{(n)})(z)\right|\right\}$$

Equations (42) and (43) can then be combined into:

$$M^{(n)}(z) \leq \varepsilon(z)M^{(n-1)}(z),$$

where $\varepsilon(z) = \frac{\varepsilon_1|z|}{(d(z,\mathbb{R}^+))^2}$, with $\varepsilon_1 = LC_{\max}^2\max\left\{\frac{r}{t}, 1\right\}$. We now define the following domain: $U = \left\{z \in \mathbb{C}, d(z,\mathbb{R}^+) \geq \frac{2\varepsilon_1}{K}, \frac{|z|}{d(z,\mathbb{R}^+)} \leq 2\right\}$, with $0 \leq K < 1$. On this domain U we have $M^{(n)}(z) \leq KM^{(n-1)}(z)$. Hence, for $z \in U$, $\psi_l^{(n)}(z)$ and $\tilde{\psi}_j^{(n)}(z)$ are Cauchy sequences and, as such, converge. We denote by $\psi_l(z)$ and $\tilde{\psi}_j(z)$ their respective limit.

One wants to extend this convergence result on $\mathbb{C} - \mathbb{R}^+$. We first notice that, as $\psi_l^{(n)}$ is a Stieltjès transform whose associated measure has mass $\frac{1}{t}\mathrm{Tr}\,\mathbf{C}^{(l)}$ by Proposition 11 item (iii), item (v) of Proposition 9 implies

$$\psi_l^{(n)}(z) \leq \frac{\frac{1}{t}\mathrm{Tr}\mathbf{C}^{(l)}}{d(z,\mathbb{R}^+)}.$$

The $\psi_l^{(n)}$ are thus bounded on any compact set included in $\mathbb{C} - \mathbb{R}^+$, uniformly in n. By Montel's theorem, $\left(\psi_l^{(n)}\right)_{n\in\mathbb{N}}$ is a normal family. Therefore one can extract a subsequence converging uniformly on compact sets of $\mathbb{C}-\mathbb{R}^+$, whose limit is thus analytic over $\mathbb{C} - \mathbb{R}^+$. This limit coincides with ψ_l on domain U. The limit of any converging subsequence of $\left(\psi_l^{(n)}\right)$ thus coincides with ψ_l on U. Therefore, these limits all coincide on $\mathbb{C}-\mathbb{R}^+$ with a function analytic on $\mathbb{C}-\mathbb{R}^+$, that we still denote ψ_l. The converging subsequences of $\left(\psi_l^{(n)}\right)$ have thus the same limit. We have therefore showed the convergence of the whole sequence $\left(\psi_l^{(n)}\right)_{n\geq 0}$ on $\mathbb{C} - \mathbb{R}^+$ towards an analytic function ψ_l. Moreover, as one can check that ψ_l verifies Proposition 10, we have $\psi_l(z) \in \mathcal{S}(\mathbb{R}^+)$. The same arguments hold for the $\tilde{\psi}_l(z)$.

We have proved the convergence of iterative sequence (39). Taking $z = -\sigma^2$ then yields the convergence of the fixed point algorithm (13). Note that the starting point $(\boldsymbol{\delta}^{(0)}, \tilde{\boldsymbol{\delta}}^{(0)})$ only needs to verify $\delta_l^{(0)} > 0$, $\tilde{\delta}_l^{(0)} > 0$ $(l = 1, \ldots, L)$, as any positive real number can be interpreted as the value at point $z = -\sigma^2$ of some element $s(z) \in \mathcal{S}(\mathbb{R}^+)$. Moreover, the limits $\psi_l(z)$, $\tilde{\psi}_l(z)$ $(l = 1, \ldots, L)$ of the iterative sequence (39) are positive for any $z = -\sigma^2$ by item (iv) of Proposition 9, as they all are Stieltjès transforms. Therefore, the limits δ_l, $\tilde{\delta}_l$ $(l = 1, \ldots, L)$ are positive.

Appendix B
A first large system approximation of $\mathbb{E}_{\mathbf{H}}[\mathrm{Tr}\,\mathbf{S}]$ – Proof of Proposition 1

In this section, if x is a random variable we denote by $\mathring{x}$ the zero mean random variable $\mathring{x} = x - \mathbb{E}(x)$.

We will prove Proposition 1 by deriving the matrix $\boldsymbol{\Upsilon}$ defined by (21), before proving that it satisfies $\frac{1}{t}\mathrm{Tr}\left(\boldsymbol{\Upsilon}\mathbf{A}\right) = \mathcal{O}\left(\frac{1}{t^2}\right)$ for any uniformly bounded matrix $\mathbf{A}$. To that end, as the entries of matrices $\mathbf{H}^{(l)}$ are Gaussian, we can use the classical Gaussian methods: we introduce here two Gaussian tools, an Integration by Parts formula and the Nash-Poincaré inequality, both widely used in Random Matrix Theory (see e.g. [16]).

We first present an Integration by Parts formula which provides the expectation of some functionals of Gaussian vectors (see e.g. [17]).

Theorem 4: Let $\boldsymbol{\xi} = [\xi_1, \ldots, \xi_M]^T$ a complex Gaussian random vector such that $\mathbb{E}[\boldsymbol{\xi}] = \mathbf{0}$, $\mathbb{E}[\boldsymbol{\xi}\boldsymbol{\xi}^T] = \mathbf{0}$ and $\mathbb{E}[\boldsymbol{\xi}\boldsymbol{\xi}^H] =$

$\boldsymbol{\Omega}$. If $\Gamma = \Gamma(\boldsymbol{\xi}, \boldsymbol{\xi}^*)$ is a $\mathcal{C}^1$ complex function polynomially bounded together with its derivatives, then

$$\mathbb{E}[\xi_p \Gamma(\boldsymbol{\xi})] = \sum_{m=1}^{M} \boldsymbol{\Omega}_{pm} \mathbb{E}\left[\frac{\partial \Gamma(\boldsymbol{\xi})}{\partial \xi_m^*}\right]. \tag{44}$$

In the present context we consider $\boldsymbol{\xi}$ being the vector of the stacked columns of matrices $\mathbf{H}^{(l)}$, where the channels $\mathbf{H}^{(l)}$ are independent and follow the Kronecker model, i.e. $\mathbb{E}_{\mathbf{H}}\big[\mathbf{H}_{ij}^{(k)} \mathbf{H}_{mn}^{(l)*}\big] = \delta_{k,l} \frac{1}{t} \mathbf{C}_{im}^{(l)} \tilde{\mathbf{C}}_{jn}^{(l)}$. Then (44) becomes

$$\mathbb{E}_{\mathbf{H}}\big[\mathbf{H}_{ij}^{(l)} \Gamma\big((\mathbf{H}^{(l)})_{l=1,\ldots,L}\big)\big] = \frac{1}{t} \sum_{m=1}^{r} \sum_{n=1}^{t} \mathbf{C}_{im}^{(l)} \tilde{\mathbf{C}}_{jn}^{(l)} \mathbb{E}_{\mathbf{H}}\left[\frac{\partial \Gamma}{\partial \mathbf{H}_{mn}^{(l)*}}\right]. \tag{45}$$

The second useful tool is the Poincaré Nash inequality which bounds the variance of certain functionals of Gaussian vectors (see e.g. [16], [6]).

Theorem 5: Let $\boldsymbol{\xi} = [\xi_1, \ldots, \xi_M]^T$ a complex Gaussian random vector such that $\mathbb{E}[\boldsymbol{\xi}] = \mathbf{0}$, $\mathbb{E}[\boldsymbol{\xi}\boldsymbol{\xi}^T] = \mathbf{0}$ and $\mathbb{E}[\boldsymbol{\xi}\boldsymbol{\xi}^H] = \boldsymbol{\Omega}$. If $\Gamma = \Gamma(\boldsymbol{\xi}, \boldsymbol{\xi}^*)$ is a $\mathcal{C}^1$ complex function polynomially bounded together with its derivatives, then, noting $\nabla_{\boldsymbol{\xi}} \Gamma = [\frac{\partial \Gamma}{\partial \xi_1}, \ldots, \frac{\partial \Gamma}{\partial \xi_M}]^T$ and $\nabla_{\boldsymbol{\xi}^*} \Gamma = [\frac{\partial \Gamma}{\partial \xi_1^*}, \ldots, \frac{\partial \Gamma}{\partial \xi_M^*}]^T$,

$$\begin{aligned}\mathrm{var}(\Gamma(\boldsymbol{\xi})) \leq & \mathbb{E}\left[\nabla_{\boldsymbol{\xi}} \Gamma(\boldsymbol{\xi})^T \ \boldsymbol{\Omega} \ \overline{\nabla_{\boldsymbol{\xi}} \Gamma(\boldsymbol{\xi})}\right] \\ & + \mathbb{E}\left[\nabla_{\boldsymbol{\xi}^*} \Gamma(\boldsymbol{\xi})^H \ \boldsymbol{\Omega} \ \nabla_{\boldsymbol{\xi}^*} \Gamma(\boldsymbol{\xi})\right].\end{aligned} \tag{46}$$

In the following we will use the Nash-Poincaré inequality with $\boldsymbol{\xi}$ being the vector of the stacked columns of independent matrices $\mathbf{H}^{(l)}$, where the channels $\mathbf{H}^{(l)}$ follow the Kronecker model. Then (46) can be written under the form (47) at the bottom of the page.

Using these two Gaussian tools we now prove Proposition 1. In order to derive the matrix $\boldsymbol{\Upsilon}$ defined by $\mathbb{E}_{\mathbf{H}}[\mathbf{S}] = \mathbf{R} + \boldsymbol{\Upsilon}$ we study the entries of $\mathbb{E}_{\mathbf{H}}[\mathbf{S}]$. Using the resolvent identity (26) we have $\sigma^2 \mathbb{E}_{\mathbf{H}}[\mathbf{S}_{pq}] = (\mathbf{I} - \mathbb{E}_{\mathbf{H}}[\mathbf{S}\mathbf{H}\mathbf{H}^H])_{pq}$. We evaluate $\mathbb{E}_{\mathbf{H}}[(\mathbf{S}\mathbf{H}\mathbf{H}^H)_{pq}]$ by first studying $\mathbb{E}_{\mathbf{H}}\big[\mathbf{S}_{pi} \mathbf{H}_{ij}^{(l)} \mathbf{H}_{qk}^{(l')*}\big]$. Calculation begins with an integration by parts on $\mathbf{H}_{ij}^{(l)}$ (45):

$$\begin{aligned}&\mathbb{E}_{\mathbf{H}}\left[\mathbf{S}_{pi} \mathbf{H}_{ij}^{(l)} \mathbf{H}_{qk}^{(l')*}\right] \\ &= \frac{1}{t} \sum_{m,n} \mathbf{C}_{im}^{(l)} \tilde{\mathbf{C}}_{jn}^{(l)} \mathbb{E}_{\mathbf{H}}\left[\frac{\partial (\mathbf{S}_{pi} \mathbf{H}_{qk}^{(l')*})}{\partial \mathbf{H}_{mn}^{(l)*}}\right] \\ &= \frac{1}{t} \sum_{m,n} \mathbf{C}_{im}^{(l)} \tilde{\mathbf{C}}_{jn}^{(l)} \mathbb{E}_{\mathbf{H}}\left[\mathbf{S}_{pi} \delta_{l,l'} \delta_{q,m} \delta_{k,n} + \mathbf{H}_{qk}^{(l')*} \frac{\partial \mathbf{S}_{pi}}{\partial \mathbf{H}_{mn}^{(l)*}}\right].\end{aligned}$$

As $\frac{\partial \mathbf{S}_{pi}}{\partial \mathbf{H}_{mn}^{(l)*}} = -\Big(\mathbf{S} \frac{\partial \mathbf{S}^{-1}}{\partial \mathbf{H}_{mn}^{(l)*}} \mathbf{S}\Big)_{pi} = -(\mathbf{S}\mathbf{H})_{pn} \mathbf{S}_{mi}$, we obtain

$$\begin{aligned}\mathbb{E}_{\mathbf{H}}\Big[\mathbf{S}_{pi} \mathbf{H}_{ij}^{(l)} \mathbf{H}_{qk}^{(l')*}\Big] = & \frac{1}{t} \mathbf{C}_{iq}^{(l)} \tilde{\mathbf{C}}_{jk}^{(l)} \mathbb{E}_{\mathbf{H}}[\mathbf{S}_{pi}] \delta_{l,l'} \\ & - \frac{1}{t} \sum_{n} \tilde{\mathbf{C}}_{jn}^{(l)} \mathbb{E}_{\mathbf{H}}\Big[\mathbf{H}_{qk}^{(l')*} (\mathbf{S}\mathbf{H})_{pn} (\mathbf{C}^{(l)} \mathbf{S})_{ii}\Big].\end{aligned}$$

Summing over i, l and l' then leads to:

$$\begin{aligned}\mathbb{E}_{\mathbf{H}}\left[(\mathbf{S}\mathbf{H})_{pj} \mathbf{H}_{qk}^*\right] = & \sum_{l} \frac{1}{t} \mathbb{E}_{\mathbf{H}}[(\mathbf{S}\mathbf{C}^{(l)})_{pq}] \tilde{\mathbf{C}}_{jk}^{(l)} \\ & - \sum_{n,l} \tilde{\mathbf{C}}_{jn}^{(l)} \mathbb{E}_{\mathbf{H}}\Big[\mathbf{H}_{qk}^* (\mathbf{S}\mathbf{H})_{pn} \frac{1}{t} \mathrm{Tr}(\mathbf{S}\mathbf{C}^{(l)})\Big].\end{aligned}$$

To separate the terms under the last expectation, we denote $\eta_l = \frac{1}{t} \mathrm{Tr}(\mathbf{S}\mathbf{C}^{(l)}) = \alpha_l + \mathring{\eta}_l$, where $\alpha_l = \mathbb{E}_{\mathbf{H}}[\eta_l]$. We can then write $\mathbb{E}_{\mathbf{H}}\big[\mathbf{H}_{qk}^* (\mathbf{S}\mathbf{H})_{pn} \eta_l)\big] = \alpha_l \mathbb{E}_{\mathbf{H}}\big[\mathbf{H}_{qk}^* (\mathbf{S}\mathbf{H})_{pn}\big] + \mathbb{E}_{\mathbf{H}}\big[\mathbf{H}_{qk}^* (\mathbf{S}\mathbf{H})_{pn} \mathring{\eta}_l\big]$, hence

$$\begin{aligned}\mathbb{E}_{\mathbf{H}}\left[(\mathbf{S}\mathbf{H})_{pj} \mathbf{H}_{qk}^*\right] = & \sum_{l} \frac{1}{t} \mathbb{E}_{\mathbf{H}}[(\mathbf{S}\mathbf{C}^{(l)})_{pq}] \tilde{\mathbf{C}}_{jk}^{(l)} \\ & - \sum_{n,l} \alpha_l \tilde{\mathbf{C}}_{jn}^{(l)} \mathbb{E}_{\mathbf{H}}\left[(\mathbf{S}\mathbf{H})_{pn} \mathbf{H}_{qk}^*\right] - \boldsymbol{\Xi}_{jk}^{(p,q)},\end{aligned} \tag{48}$$

where $\boldsymbol{\Xi}_{jk}^{(p,q)} = \sum_l \mathbb{E}_{\mathbf{H}}\big[\mathring{\eta}_l \mathbf{H}_{qk}^* (\mathbf{S}\mathbf{H}\tilde{\mathbf{C}}^{(l)T})_{pj}\big]$. We here notice the presence of $\mathbb{E}_{\mathbf{H}}\big[(\mathbf{S}\mathbf{H})_{p\_} \mathbf{H}_{qk}^*\big]$ on both sides of equation (48). Hence, let us denote $\Delta_{jk}^{(p,q)} = \mathbb{E}_{\mathbf{H}}\big[(\mathbf{S}\mathbf{H})_{pj} \mathbf{H}_{qk}^*\big]$. Then (48) becomes

$$\begin{aligned}\Delta_{jk}^{(p,q)} = & \sum_{l} \frac{1}{t} \mathbb{E}_{\mathbf{H}}[(\mathbf{S}\mathbf{C}^{(l)})_{pq}] \tilde{\mathbf{C}}_{jk}^{(l)} \\ & - \Big(\sum_{l} \alpha_l \tilde{\mathbf{C}}^{(l)} \Delta^{(p,q)}\Big)_{jk} - \boldsymbol{\Xi}_{jk}^{(p,q)}.\end{aligned}$$

Recalling that $\tilde{\mathbf{R}} = \big(\sigma^2 \big(\mathbf{I}_t + \sum_l \alpha_l \tilde{\mathbf{C}}^{(l)}\big)\big)^{-1}$, this leads to

$$\Delta^{(p,q)} = \sigma^2 \sum_{l} \frac{1}{t} \mathbb{E}_{\mathbf{H}}[(\mathbf{S}\mathbf{C}^{(l)})_{pq}] \tilde{\mathbf{R}} \tilde{\mathbf{C}}^{(l)} - \sigma^2 \tilde{\mathbf{R}} \boldsymbol{\Xi}^{(p,q)}.$$

We now come back to the calculation of $\mathbb{E}_{\mathbf{H}}[\mathbf{S}_{pq}] = \frac{1}{\sigma^2}(\mathbf{I}_r - \mathbb{E}_{\mathbf{H}}[\mathbf{S}\mathbf{H}\mathbf{H}^H])_{pq}$ by noticing that $\mathbb{E}_{\mathbf{H}}[(\mathbf{S}\mathbf{H}\mathbf{H}^H)_{pq}] = \sum_j \mathbb{E}_{\mathbf{H}}\left[(\mathbf{S}\mathbf{H})_{pj} \mathbf{H}_{qj}^*\right] = \mathrm{Tr}(\Delta^{(p,q)})$. Therefore

$$\mathbb{E}_{\mathbf{H}}[\mathbf{S}_{pq}] = \frac{\delta_{p,q}}{\sigma^2} - \sum_{l} \tilde{\alpha}_l \mathbb{E}_{\mathbf{H}}[(\mathbf{S}\mathbf{C}^{(l)})_{pq}] + \mathrm{Tr}\left(\tilde{\mathbf{R}} \boldsymbol{\Xi}^{(p,q)}\right),$$

recalling from (20) that $\tilde{\alpha}_l = \frac{1}{t} \mathrm{Tr}\big(\tilde{\mathbf{R}} \tilde{\mathbf{C}}^{(l)}\big)$. Coming back to the definition of matrix $\boldsymbol{\Xi}^{(p,q)}$, we notice that $\mathrm{Tr}\big(\tilde{\mathbf{R}} \boldsymbol{\Xi}^{(p,q)}\big) = \sum_l \mathbb{E}_{\mathbf{H}}\big[\mathring{\eta}_l (\mathbf{S}\mathbf{H}\tilde{\mathbf{C}}^{(l)T} \tilde{\mathbf{R}}^T \mathbf{H}^H)_{pq}\big]$. Hence matrix $\mathbb{E}_{\mathbf{H}}[\mathbf{S}]$ can be written as

$$\begin{aligned}\mathbb{E}_{\mathbf{H}}[\mathbf{S}] = & \frac{1}{\sigma^2} \mathbf{I}_r - \mathbb{E}_{\mathbf{H}}[\mathbf{S}] \sum_{l} \tilde{\alpha}_l \mathbf{C}^{(l)} \\ & + \sum_{l} \mathbb{E}_{\mathbf{H}}\left[\mathring{\eta}_l \mathbf{S}\mathbf{H}\tilde{\mathbf{C}}^{(l)T} \tilde{\mathbf{R}}^T \mathbf{H}^H\right].\end{aligned}$$

And finally, $\mathbb{E}_{\mathbf{H}}[\mathbf{S}] = \mathbf{R} + \boldsymbol{\Upsilon}$, where we recall that $\mathbf{R} = \big(\sigma^2 \big(\mathbf{I}_r + \sum_l \tilde{\alpha}_l \mathbf{C}^{(l)}\big)\big)^{-1}$ and where matrix $\boldsymbol{\Upsilon}$ is defined as

$$\boldsymbol{\Upsilon} = \sigma^2 \sum_{l} \mathbb{E}_{\mathbf{H}}\left[\mathring{\eta}_l \mathbf{S}\mathbf{H}\tilde{\mathbf{C}}^{(l)T} \tilde{\mathbf{R}}^T \mathbf{H}^H\right] \mathbf{R}. \tag{49}$$

$$\mathrm{var}\left(\Gamma\big((\mathbf{H}^{(l)})_{l=1,\ldots,L}\big)\right) \leq \frac{1}{t} \sum_{i,m=1}^{r} \sum_{j,n=1}^{t} \sum_{l=1}^{L} \mathbf{C}_{im}^{(l)} \tilde{\mathbf{C}}_{jn}^{(l)} \mathbb{E}_{\mathbf{H}}\left[\frac{\partial \Gamma}{\partial \mathbf{H}_{ij}^{(l)}} \left(\frac{\partial \Gamma}{\partial \mathbf{H}_{mn}^{(l)}}\right)^* + \left(\frac{\partial \Gamma}{\partial \mathbf{H}_{ij}^{(l)*}}\right)^* \frac{\partial \Gamma}{\partial \mathbf{H}_{mn}^{(l)*}}\right] \tag{47}$$

To end Proposition 1 proof, we now need to prove that $\frac{1}{t}\mathrm{Tr}\left(\mathbf{\Upsilon A}\right) = \mathcal{O}\left(\frac{1}{t^2}\right)$ for any uniformly bounded matrix $\mathbf{A}$. Let $\mathbf{A}$ be a $r \times r$ matrix uniformly bounded in r. Using (49),

$$\begin{aligned}\frac{1}{t}\mathrm{Tr}\left(\mathbf{\Upsilon A}\right) &= \frac{\sigma^2}{t}\sum_l \mathbb{E}_{\mathbf{H}}\left[\mathring{\eta}_l \mathrm{Tr}\left(\mathbf{S}\mathbf{H}\tilde{\mathbf{C}}^{(l)T}\tilde{\mathbf{R}}^T\mathbf{H}^H\mathbf{R}\mathbf{A}\right)\right] \\ &= \frac{\sigma^2}{t}\sum_l \mathbb{E}_{\mathbf{H}}\left[\mathring{\eta}_l \overset{\circ}{\overbrace{\mathrm{Tr}(\mathbf{S}\mathbf{H}\tilde{\mathbf{C}}^{(l)T}\tilde{\mathbf{R}}^T\mathbf{H}^H\mathbf{R}\mathbf{A})}}\right].\end{aligned}$$

We can now bound $\frac{1}{t}\mathrm{Tr}\left(\mathbf{\Upsilon A}\right)$ thanks to Cauchy-Schwartz inequality.

$$\begin{aligned}&\left|\frac{1}{t}\mathrm{Tr}\left(\mathbf{\Upsilon A}\right)\right| \\ &\leq \frac{\sigma^2}{t}\sum_l \sqrt{\mathbb{E}_{\mathbf{H}}\left[\left|\mathring{\eta}_l\right|^2\right]\mathbb{E}_{\mathbf{H}}\left[\left|\overset{\circ}{\overbrace{\mathrm{Tr}(\mathbf{S}\mathbf{H}\tilde{\mathbf{C}}^{(l)T}\tilde{\mathbf{R}}^T\mathbf{H}^H\mathbf{R}\mathbf{A})}}\right|^2\right]} \\ &= \frac{\sigma^2}{t}\sum_l \sqrt{\mathrm{var}\left(\eta_l\right)\mathrm{var}\left(\mathrm{Tr}(\mathbf{S}\mathbf{H}\tilde{\mathbf{C}}^{(l)T}\tilde{\mathbf{R}}^T\mathbf{H}^H\mathbf{R}\mathbf{A})\right)}, \quad (50)\end{aligned}$$

as $\mathbb{E}_{\mathbf{H}}\left[\left|\mathring{x}\right|^2\right] = \mathrm{var}\left(x\right)$ for any random variable x. We now prove that $\mathrm{var}(\eta_l) = \mathcal{O}\left(\frac{1}{t^2}\right)$. The Nash-Poincaré inequality (47) states that

$$\begin{aligned}\mathrm{var}(\eta_l) \leq \frac{1}{t}\sum_{i,j,m,n,k}\mathbf{C}^{(k)}_{im}\tilde{\mathbf{C}}^{(k)}_{jn}\mathbb{E}\Bigg[&\frac{\partial\eta_l}{\partial\mathbf{H}^{(k)}_{ij}}\left(\frac{\partial\eta_l}{\partial\mathbf{H}^{(k)}_{mn}}\right)^* \\ &+\left(\frac{\partial\eta_l}{\partial\mathbf{H}^{(k)*}_{ij}}\right)^*\frac{\partial\eta_l}{\partial\mathbf{H}^{(k)*}_{mn}}\Bigg]. \quad (51)\end{aligned}$$

As $\partial\mathbf{S}_{pq}/\partial\mathbf{H}^{(k)}_{ij} = -(\mathbf{S}(\partial\mathbf{S}^{-1}/\partial\mathbf{H}^{(k)}_{ij})\mathbf{S})_{pq} = -\mathbf{S}_{pi}(\mathbf{H}^H\mathbf{S})_{jq}$ the partial derivative $\partial\eta_l/\partial\mathbf{H}^{(k)}_{ij}$ can be written as

$$\begin{aligned}\frac{\partial\eta_l}{\partial\mathbf{H}^{(k)}_{ij}} &= \frac{1}{t}\mathrm{Tr}\left(\frac{\partial\mathbf{S}}{\partial\mathbf{H}^{(k)}_{ij}}\mathbf{C}^{(l)}\right) = \frac{1}{t}\sum_{p,q}\frac{\partial\mathbf{S}_{pq}}{\partial\mathbf{H}^{(k)}_{ij}}\mathbf{C}^{(l)}_{qp} \\ &= -\frac{1}{t}(\mathbf{H}^H\mathbf{S}\mathbf{C}^{(l)}\mathbf{S})_{ji}.\end{aligned}$$

Similarly we obtain $\partial\eta_l/\partial\mathbf{H}^{(k)*}_{ij} = -\frac{1}{t}(\mathbf{S}\mathbf{C}^{(l)}\mathbf{S}\mathbf{H})_{ij}$. Therefore (51) leads to (52) at the bottom of the page. Both traces of (52) can be upper bounded thanks to inequality $|\mathrm{Tr}(\mathbf{B}_1\mathbf{B}_2)| \leq \|\mathbf{B}_1\|\mathrm{Tr}\,\mathbf{B}_2$, where $\mathbf{B}_2$ is non-negative hermitian,

$$\begin{aligned}\mathrm{var}(\eta_l) &\leq \frac{2}{t^3}\|\mathbf{C}^{(l)}\|^2\sum_k\|\mathbf{C}^{(k)}\|\ \mathbb{E}\left[\|\mathbf{S}\|^4\mathrm{Tr}\left(\mathbf{H}\tilde{\mathbf{C}}^{(k)T}\mathbf{H}^H\right)\right] \\ &\leq \frac{2}{t^3}\|\mathbf{C}^{(l)}\|^2\sum_k\|\mathbf{C}^{(k)}\|\|\tilde{\mathbf{C}}^{(k)}\|\ \mathbb{E}\left[\|\mathbf{S}\|^4\mathrm{Tr}\left(\mathbf{H}\mathbf{H}^H\right)\right] \\ &\leq \frac{1}{t^2}\frac{2LC^4_{\mathrm{sup}}}{\sigma^8}\mathbb{E}\left[\frac{1}{t}\mathrm{Tr}\left(\mathbf{H}\mathbf{H}^H\right)\right], \quad (53)\end{aligned}$$

where the second inequality follows from $\|\mathbf{S}\| \leq \frac{1}{\sigma^2}$ and from the definition of C_{sup}:

$$C_{\mathrm{sup}} = \sup_t C_{\max} = \sup_t\ \left\{\max_k\left\{\|\mathbf{C}^{(k)}\|, \|\tilde{\mathbf{C}}^{(k)}\|\right\}\right\}. \quad (54)$$

The hypotheses of Proposition 1 ensure that $C_{\mathrm{sup}} < +\infty$. We now prove that $\mathbb{E}\left[\frac{1}{t}\mathrm{Tr}\left(\mathbf{H}\mathbf{H}^H\right)\right] = \mathcal{O}\left(1\right)$. Using the fact that the channels $\mathbf{H}^{(l)}$ are independent and follow the Kronecker model, that is $\mathbb{E}_{\mathbf{H}}\left[\mathbf{H}^{(k)}_{ij}\mathbf{H}^{(l)*}_{mn}\right] = \delta_{k,l}\frac{1}{t}\mathbf{C}^{(l)}_{im}\tilde{\mathbf{C}}^{(l)}_{jn}$,

$$\begin{aligned}&\mathbb{E}_{\mathbf{H}}\left[\frac{1}{t}\mathrm{Tr}\left(\mathbf{H}\mathbf{H}^H\right)\right] \\ &= \frac{1}{t}\sum_{i,j,k,l}\mathbb{E}_{\mathbf{H}}\left[\mathbf{H}^{(k)}_{ij}\mathbf{H}^{(l)*}_{ij}\right] = \frac{1}{t^2}\sum_{i,j,l}\mathbf{C}^{(l)}_{ii}\tilde{\mathbf{C}}^{(l)}_{jj} \\ &= \frac{1}{t^2}\sum_l \mathrm{Tr}\,\mathbf{C}^{(l)}\mathrm{Tr}\,\tilde{\mathbf{C}}^{(l)} \leq \frac{r}{t}LC^2_{\mathrm{sup}}.\end{aligned}$$

Therefore we proved that $\mathbb{E}_{\mathbf{H}}\left[\frac{1}{t}\mathrm{Tr}\left(\mathbf{H}\mathbf{H}^H\right)\right] = \mathcal{O}\left(1\right)$. Coming back to (53) gives $\mathrm{var}(\eta_l) \leq \frac{1}{t^2}\left(\frac{r}{t}\frac{2C^6_{\mathrm{sup}}L^2}{\sigma^8}\right)$, hence $\mathrm{var}(\eta_l) = \mathcal{O}\left(\frac{1}{t^2}\right)$.

We evaluate similarly the behavior of the second term of the right-hand side of (50) and we obtain $\mathrm{var}\big(\mathrm{Tr}(\mathbf{S}\mathbf{H}\tilde{\mathbf{C}}^{(l)T}\tilde{\mathbf{R}}^T\mathbf{H}^H\mathbf{R}\mathbf{A})\big) \leq \frac{k}{\sigma^{12}}\left(1+\frac{1}{\sigma^2}\right)\|\mathbf{A}\|^2 = \mathcal{O}\left(1\right)$, where k does not depend on σ^2 nor on t. Hence we eventually have:

$$\frac{1}{t}\mathrm{Tr}(\mathbf{\Upsilon A}) = \mathcal{O}\left(\frac{1}{t^2}\right),$$

which completes the proof of Proposition 1.

Remark 1: Note that, as $\mathrm{var}(\eta_l) \leq \frac{1}{\sigma^8t^2}\left(2\frac{r}{t}C^6_{\mathrm{sup}}L^2\right)$ and $\mathrm{var}\big(\mathrm{Tr}(\mathbf{S}\mathbf{H}\tilde{\mathbf{C}}^{(l)T}\tilde{\mathbf{R}}^T\mathbf{H}^H\mathbf{R}\mathbf{A})\big) \leq \frac{1}{\sigma^{12}}\left(k\|\mathbf{A}\|^2\left(1+\frac{1}{\sigma^2}\right)\right)$, (50) leads to $\frac{1}{t}\mathrm{Tr}(\mathbf{\Upsilon A}) \leq \frac{1}{\sigma^8t^2}P\left(\frac{1}{\sigma^2}\right)$, where P is a polynomial with real positive coefficients which do not depend on σ^2 nor on t.

Appendix C
A refined large system approximation of $\mathbb{E}_{\mathbf{H}}[\mathrm{Tr}\,\mathbf{S}]$ – Proof of Proposition 2

We prove in this section that $\frac{1}{t}\mathrm{Tr}(\mathbf{R}\mathbf{A}) = \frac{1}{t}\mathrm{Tr}(\mathbf{T}\mathbf{A}) + \mathcal{O}\left(\frac{1}{t^2}\right)$ for any $r \times r$ matrix $\mathbf{A}$ uniformly bounded in r. We first note that the difference $\frac{1}{t}\mathrm{Tr}\left(\mathbf{R}\mathbf{A}\right) - \frac{1}{t}\mathrm{Tr}\left(\mathbf{T}\mathbf{A}\right)$ can be written as

$$\begin{aligned}\frac{1}{t}\mathrm{Tr}\left((\mathbf{R}-\mathbf{T})\mathbf{A}\right) &= \frac{1}{t}\mathrm{Tr}\left(\mathbf{R}\left(\mathbf{T}^{-1}-\mathbf{R}^{-1}\right)\mathbf{T}\mathbf{A}\right) \\ &= -\frac{\sigma^2}{t}\sum_l(\tilde{\alpha}_l-\tilde{\delta}_l)\mathrm{Tr}(\mathbf{R}\mathbf{C}^{(l)}\mathbf{T}\mathbf{A}). \quad (55)\end{aligned}$$

As $\|\mathbf{T}\| \leq \frac{1}{\sigma^2}$ and $\|\mathbf{R}\| \leq \frac{1}{\sigma^2}$, equation (55) yields

$$\frac{1}{t}\left|\mathrm{Tr}\left((\mathbf{R}-\mathbf{T})\mathbf{A}\right)\right| \leq \frac{r}{t}\frac{C_{\mathrm{sup}}\|\mathbf{A}\|}{\sigma^2}\sum_l\left|\tilde{\alpha}_l-\tilde{\delta}_l\right|, \quad (56)$$

where $C_{\mathrm{sup}} < +\infty$ is defined by (54). We derive similarly the difference $\frac{1}{t}\mathrm{Tr}(\tilde{\mathbf{R}}\tilde{\mathbf{A}}) - \frac{1}{t}\mathrm{Tr}(\tilde{\mathbf{T}}\tilde{\mathbf{A}})$ for any $t \times t$ matrix $\tilde{\mathbf{A}}$ uniformly bounded in t.

$$\frac{1}{t}\left|\mathrm{Tr}\left(\left(\tilde{\mathbf{R}}-\tilde{\mathbf{T}}\right)\tilde{\mathbf{A}}\right)\right| \leq \frac{C_{\mathrm{sup}}\|\tilde{\mathbf{A}}\|}{\sigma^2}\sum_l|\alpha_l-\delta_l| \quad (57)$$

$$\mathrm{var}(\eta_l) \leq \frac{1}{t^3}\sum_k\mathbb{E}\Bigg[\mathrm{Tr}\left((\mathbf{H}^H\mathbf{S}\mathbf{C}^{(l)}\mathbf{S})\mathbf{C}^{(k)}(\mathbf{H}^H\mathbf{S}\mathbf{C}^{(l)}\mathbf{S})^H\tilde{\mathbf{C}}^{(k)T}\right) + \mathrm{Tr}\left(\tilde{\mathbf{C}}^{(k)T}(\mathbf{S}\mathbf{C}^{(l)}\mathbf{S}\mathbf{H})^H\mathbf{C}^{(k)}(\mathbf{S}\mathbf{C}^{(l)}\mathbf{S}\mathbf{H})\right)\Bigg]. \quad (52)$$

Taking $\mathbf{A} = \mathbf{C}^{(k)}$ in (56), $\tilde{\mathbf{A}} = \tilde{\mathbf{C}}^{(k)}$ in (57) and using Proposition 1 gives

$$\left|\alpha_k - \delta_k\right| \leq \frac{r}{t}\frac{C_{\sup}^2}{\sigma^2}\sum_l \left|\tilde{\alpha}_l - \tilde{\delta}_l\right| + \mathcal{O}\left(\frac{1}{t^2}\right), \tag{58}$$

$$\left|\tilde{\alpha}_k - \tilde{\delta}_k\right| \leq \frac{C_{\sup}^2}{\sigma^2}\sum_l \left|\alpha_l - \delta_l\right|, \tag{59}$$

which leads to

$$\left(1 - \frac{r}{t}\frac{C_{\sup}^4 L^2}{\sigma^4}\right)\sum_k \left|\alpha_k - \delta_k\right| \leq \mathcal{O}\left(\frac{1}{t^2}\right).$$

Therefore it is clear that there exists σ_0^2 such that $\left|\alpha_k - \delta_k\right| = \mathcal{O}\left(\frac{1}{t^2}\right)$ for $\sigma^2 > \sigma_0^2$ for any $k \in \{1, \ldots, L\}$. In particular, $\left|\alpha_k - \delta_k\right| \xrightarrow{t\to\infty} 0$ for $\sigma^2 > \sigma_0^2$. We now extend this result to any $\sigma^2 > 0$. To this end, similarly to Appendix A, it is useful to consider α_l and δ_l as functions of the parameter $(-\sigma^2) \in \mathbb{R}^-$ and to extend their domain of validity from $\mathbb{R}^-$ to $\mathbb{C} - \mathbb{R}^+$ in order to use the results about Stieltjès transforms. The function $\delta_l(z)$ then corresponds to the function $\psi_l(z)$ of Appendix A and therefore belongs to $\mathcal{S}(\mathbb{R}^+)$ with an associated measure of mass $\frac{1}{t}\mathrm{Tr}\,\mathbf{C}^{(l)}$, for $l = 1, \ldots, L$. It is easy to check that function $\alpha_l(z)$ also belongs to $\mathcal{S}(\mathbb{R}^+)$ with an associated measure of mass $\frac{1}{t}\mathrm{Tr}\,\mathbf{C}^{(l)}$ for any $l \in \{1, \ldots, L\}$. Hence, by Proposition 9 (v), we can upper bound the Stieltjès transforms $\alpha_l(z)$ and $\delta_l(z)$ on $\mathbb{C} - \mathbb{R}^+$, yielding:

$$|\alpha_l(z) - \delta_l(z)| \leq 2\frac{\frac{1}{t}\mathrm{Tr}\,\mathbf{C}^{(l)}}{d(z, \mathbb{R}^+)} \leq 2\frac{\frac{r}{t}C_{\sup}}{d(z, \mathbb{R}^+)}.$$

The $(\alpha_l(z) - \delta_l(z))_{t\in\mathbb{N}}$ are thus bounded on any compact set included in $\mathbb{C} - \mathbb{R}^+$, uniformly in t. Moreover $(\alpha_l(z) - \delta_l(z))_{t\in\mathbb{N}}$ is a family of analytic functions. Using Montel's theorem similarly to Appendix A, we obtain that $\left|\alpha_l(z) - \delta_l(z)\right| \xrightarrow{t\to\infty} 0$ on $\mathbb{C} - \mathbb{R}^+$ for any $l \in \{1, \ldots, L\}$, thus in particular

$$\left|\alpha_l - \delta_l\right| \xrightarrow{t\to\infty} 0 \tag{60}$$

for any $\sigma^2 > 0$, $l \in \{1, \ldots, L\}$. And (59) then yields

$$\left|\tilde{\alpha}_l - \tilde{\delta}_l\right| \xrightarrow{t\to\infty} 0 \tag{61}$$

for any $\sigma^2 > 0$, $l \in \{1, \ldots, L\}$. Using (61) in (56) and (60) in (57) gives

$$\frac{1}{t}\mathrm{Tr}\left(\mathbf{A}(\mathbf{R} - \mathbf{T})\right) \xrightarrow{t\to\infty} 0, \tag{62}$$

$$\frac{1}{t}\mathrm{Tr}\left(\tilde{\mathbf{A}}(\tilde{\mathbf{R}} - \tilde{\mathbf{T}})\right) \xrightarrow{t\to\infty} 0. \tag{63}$$

We now refine (62) and (63) to prove that these two traces are $\mathcal{O}\left(\frac{1}{t^2}\right)$. Taking $\mathbf{A} = \mathbf{C}^{(l)}$ in (55) leads to $\alpha_k - \delta_k = \frac{\sigma^2}{t}\sum_l(\tilde{\delta}_l - \tilde{\alpha}_l)\mathrm{Tr}\left(\mathbf{C}^{(l)}\mathbf{T}\mathbf{C}^{(k)}\mathbf{R}\right) + \frac{1}{t}\mathrm{Tr}\left(\mathbf{C}^{(k)}\boldsymbol{\Upsilon}\right)$, where $\boldsymbol{\Upsilon} = \mathbb{E}_{\mathbf{H}}[\mathbf{S}] - \mathbf{R}$, and similarly $\tilde{\delta}_k - \tilde{\alpha}_k = \frac{\sigma^2}{t}\sum_l(\alpha_l - \delta_l)\mathrm{Tr}\big(\tilde{\mathbf{C}}^{(l)}\tilde{\mathbf{T}}\tilde{\mathbf{C}}^{(k)}\tilde{\mathbf{R}}\big)$. We can rewrite these two equalities under the following matrix form:

$$\left(\mathbf{I}_{2L} - \mathbf{N}(\mathbf{R}, \mathbf{T}, \tilde{\mathbf{R}}, \tilde{\mathbf{T}})\right)\begin{bmatrix}\boldsymbol{\alpha} - \boldsymbol{\delta} \\ \tilde{\boldsymbol{\delta}} - \tilde{\boldsymbol{\alpha}}\end{bmatrix} = \begin{bmatrix}\boldsymbol{\varepsilon} \\ \mathbf{0}\end{bmatrix}, \tag{64}$$

where $\boldsymbol{\varepsilon}$ is a $L \times 1$ vector whose entries defined by $\boldsymbol{\varepsilon}_k = \frac{1}{t}\mathrm{Tr}\left(\mathbf{C}^{(k)}\boldsymbol{\Upsilon}\right)$ verify $\boldsymbol{\varepsilon}_k = \mathcal{O}\left(\frac{1}{t^2}\right)$, $k = 1, \ldots, L$, by Proposition 1, and where matrix $\mathbf{N}(\mathbf{R}, \mathbf{T}, \tilde{\mathbf{R}}, \tilde{\mathbf{T}})$ is defined by

$$\mathbf{N}(\mathbf{R}, \mathbf{T}, \tilde{\mathbf{R}}, \tilde{\mathbf{T}}) = \sigma^2\begin{bmatrix}\mathbf{0} & \mathbf{B}(\mathbf{R}, \mathbf{T}) \\ \tilde{\mathbf{B}}(\tilde{\mathbf{R}}, \tilde{\mathbf{T}}) & \mathbf{0}\end{bmatrix}, \tag{65}$$

where matrices $\mathbf{B}(\mathbf{R}, \mathbf{T})$ and $\tilde{\mathbf{B}}(\tilde{\mathbf{R}}, \tilde{\mathbf{T}})$ are $L \times L$ matrices whose entries are defined by $\mathbf{B}_{kl}(\mathbf{R}, \mathbf{T}) = \frac{1}{t}\mathrm{Tr}\left(\mathbf{C}^{(l)}\mathbf{T}\mathbf{C}^{(k)}\mathbf{R}\right)$ and $\tilde{\mathbf{B}}_{kl}(\tilde{\mathbf{R}}, \tilde{\mathbf{T}}) = \frac{1}{t}\mathrm{Tr}\big(\tilde{\mathbf{C}}^{(l)}\tilde{\mathbf{T}}\tilde{\mathbf{C}}^{(k)}\tilde{\mathbf{R}}\big)$. Besides, taking $\mathbf{A} = \mathbf{C}^{(l)}\mathbf{T}\mathbf{C}^{(k)}$ in (62) and $\tilde{\mathbf{A}} = \tilde{\mathbf{C}}^{(l)}\tilde{\mathbf{T}}\tilde{\mathbf{C}}^{(k)}$ in (63) leads to

$$\begin{cases}\mathbf{B}_{kl}(\mathbf{R}, \mathbf{T}) \xrightarrow{t\to\infty} \frac{1}{t}\mathrm{Tr}\left(\mathbf{C}^{(l)}\mathbf{T}\mathbf{C}^{(k)}\mathbf{T}\right), \\ \tilde{\mathbf{B}}_{kl}(\tilde{\mathbf{R}}, \tilde{\mathbf{T}}) \xrightarrow{t\to\infty} \frac{1}{t}\mathrm{Tr}\left(\tilde{\mathbf{C}}^{(l)}\tilde{\mathbf{T}}\tilde{\mathbf{C}}^{(k)}\tilde{\mathbf{T}}\right).\end{cases} \tag{66}$$

We now introduce the following lemma:

Lemma 1: Let $\mathbf{T}$, $\tilde{\mathbf{T}}$ be the matrices defined by (11) and (12) with $(\boldsymbol{\delta}, \tilde{\boldsymbol{\delta}})$ verifying the canonical equation (9) with $\mathbf{Q} = \mathbf{I}_t$. Let $\mathbf{A}(\mathbf{T})$ and $\tilde{\mathbf{A}}(\mathbf{T})$ be the $L \times L$ matrices whose entries are defined by $\mathbf{A}_{kl}(\mathbf{T}) = \frac{1}{t}\mathrm{Tr}\left(\mathbf{C}^{(k)}\mathbf{T}\mathbf{C}^{(l)}\mathbf{T}\right)$ and $\tilde{\mathbf{A}}_{kl}(\tilde{\mathbf{T}}) = \frac{1}{t}\mathrm{Tr}(\tilde{\mathbf{C}}^{(k)}\tilde{\mathbf{T}}\tilde{\mathbf{C}}^{(l)}\tilde{\mathbf{T}})$ and $\mathbf{M}(\mathbf{T}, \tilde{\mathbf{T}})$ the matrix defined by

$$\mathbf{M}(\mathbf{T}, \tilde{\mathbf{T}}) = \sigma^2\begin{bmatrix}0 & \mathbf{A}(\mathbf{T}) \\ \tilde{\mathbf{A}}(\tilde{\mathbf{T}}) & 0\end{bmatrix}.$$

Assume that, for every $l \in \{1, \ldots, L\}$, $\sup_t \|\mathbf{C}^{(l)}\| < +\infty$, $\sup_t \|\tilde{\mathbf{C}}^{(l)}\| < +\infty$, $\inf_t\left(\frac{1}{t}\mathrm{Tr}\,\mathbf{C}^{(l)}\right) > 0$ and $\inf_t\left(\frac{1}{t}\mathrm{Tr}\,\tilde{\mathbf{C}}^{(l)}\right) > 0$. Then there exists $k_0 > 0$ and $k_1 < \infty$ both independent of σ^2 such that

(i) $\sup_t\left[\rho\left(\mathbf{M}\right)\right] \leq 1 - \frac{k_0\sigma^4}{(\sigma^2 + k_1)^2} < 1$,

(ii) $\sup_t\left[\rho\left(\sigma^4\tilde{\mathbf{A}}(\tilde{\mathbf{T}})\mathbf{A}(\mathbf{T})\right)\right] \leq \left(1 - \frac{k_0\sigma^4}{(\sigma^2+k_1)^2}\right)^2 < 1$,

(iii) $\sup_t\left[\left\|\left|(\mathbf{I}_{2L} - \mathbf{M}(\mathbf{T}, \tilde{\mathbf{T}}))^{-1}\right|\right\|_\infty\right] \leq \frac{(\sigma^2+k_1)^2}{k_0\sigma^4}$,

where $\|\|\cdot\|\|_\infty$ is the max-row ℓ_1 norm defined by $\|\|\mathbf{P}\|\|_\infty = \max_{j\in\{1,\ldots,M\}}\sum_{k=1}^N |\mathbf{P}_{jk}|$ for a $M \times N$ matrix $\mathbf{P}$.

Proof: Using the expression of $\mathbf{T}^{-1} = \sigma^2(\mathbf{I}_r + \sum_k \tilde{\delta}_k\mathbf{C}^{(k)})$, δ_l can be written as:

$$\begin{aligned}\delta_l &= \frac{1}{t}\mathrm{Tr}(\mathbf{C}^{(l)}\mathbf{T}\mathbf{T}^{-1}\mathbf{T}) \\ &= \frac{\sigma^2}{t}\mathrm{Tr}(\mathbf{C}^{(l)}\mathbf{T}\mathbf{T}) + \frac{\sigma^2}{t}\sum_{k=1}^L \tilde{\delta}_k\mathrm{Tr}(\mathbf{C}^{(l)}\mathbf{T}\mathbf{C}^{(k)}\mathbf{T}).\end{aligned}$$

Similarly $\tilde{\delta}_l$ verifies

$$\tilde{\delta}_l = \frac{\sigma^2}{t}\mathrm{Tr}\big(\tilde{\mathbf{C}}^{(l)}\tilde{\mathbf{T}}\tilde{\mathbf{T}}\big) + \frac{\sigma^2}{t}\sum_k \delta_k\mathrm{Tr}\big(\tilde{\mathbf{C}}^{(l)}\tilde{\mathbf{T}}\tilde{\mathbf{C}}^{(k)}\tilde{\mathbf{T}}\big).$$

Thus,

$$\begin{bmatrix}\boldsymbol{\delta} \\ \tilde{\boldsymbol{\delta}}\end{bmatrix} = \sigma^2\begin{bmatrix}0 & \mathbf{A}(\mathbf{T}) \\ \tilde{\mathbf{A}}(\tilde{\mathbf{T}}) & 0\end{bmatrix}\begin{bmatrix}\boldsymbol{\delta} \\ \tilde{\boldsymbol{\delta}}\end{bmatrix} + \begin{bmatrix}\mathbf{w} \\ \tilde{\mathbf{w}}\end{bmatrix},$$

where $\mathbf{w}$ and $\tilde{\mathbf{w}}$ are $L \times 1$ vectors such that $\mathbf{w}_l = \frac{\sigma^2}{t}\mathrm{Tr}(\mathbf{C}^{(l)}\mathbf{T}\mathbf{T})$ and $\tilde{\mathbf{w}}_l = \frac{\sigma^2}{t}\mathrm{Tr}(\tilde{\mathbf{C}}^{(l)}\tilde{\mathbf{T}}\tilde{\mathbf{T}})$. This equality is of the form $\mathbf{u} = \mathbf{M}(\mathbf{T}, \tilde{\mathbf{T}})\mathbf{u} + \mathbf{v}$, with $\mathbf{u} = \left[\boldsymbol{\delta}^T, \tilde{\boldsymbol{\delta}}^T\right]^T$ and $\mathbf{v} = \left[\mathbf{w}^T, \tilde{\mathbf{w}}^T\right]^T$, the entries of $\mathbf{u}$ and $\mathbf{v}$ being positive, and the entries of $\mathbf{M}(\mathbf{T}, \tilde{\mathbf{T}})$ non-negative. A direct application

of Corollary 8.1.29 of [11] then implies $\rho(\mathbf{M}(\mathbf{T},\tilde{\mathbf{T}})) \leq 1-\frac{\min \mathbf{v}_l}{\max \mathbf{u}_l}$.

We first consider $\sup_t \{\max \mathbf{u}_l\}$. As $\mathbf{u} = \left[\boldsymbol{\delta}^T, \tilde{\boldsymbol{\delta}}^T\right]^T$ we need to upper bound δ_k and $\tilde{\delta}_k$. As $\|\mathbf{T}\| \leq \frac{1}{\sigma^2}$ and $\|\mathbf{C}^{(l)}\| \leq C_{\sup}$ we have

$$\delta_k = \frac{1}{t}\mathrm{Tr}\left(\mathbf{C}^{(k)}\mathbf{T}\right) \leq \frac{r}{\sigma^2 t} C_{\sup}. \tag{67}$$

Similarly, as $\|\tilde{\mathbf{T}}\| \leq \frac{1}{\sigma^2}$ and $\|\tilde{\mathbf{C}}^{(l)}\| \leq C_{\sup}$,

$$\tilde{\delta}_k = \frac{1}{t}\mathrm{Tr}\left(\tilde{\mathbf{C}}^{(k)}\tilde{\mathbf{T}}\right) \leq \frac{1}{\sigma^2} C_{\sup}. \tag{68}$$

As $t/r \xrightarrow{t\to\infty} c > 0$ we have that $\sup_t [r/t] < +\infty$. Therefore $\sup_t \{\max \mathbf{u}_l\} \leq \frac{\lambda_0}{\sigma^2} < +\infty$, where $\lambda_0 = C_{\sup}\max\{1, \sup_t [r/t]\}$.

We now consider $\inf_t \{\min_l \mathbf{v}_l\}$. As $\min_l \mathbf{v}_l = \min_k \left\{\frac{\sigma^2}{t}\mathrm{Tr}(\mathbf{C}^{(k)}\mathbf{T}\mathbf{T}), \frac{\sigma^2}{t}\mathrm{Tr}(\tilde{\mathbf{C}}^{(k)}\tilde{\mathbf{T}}\tilde{\mathbf{T}})\right\}$, we need to lower bound $\frac{\sigma^2}{t}\mathrm{Tr}(\mathbf{C}^{(k)}\mathbf{T}\mathbf{T})$ and $\frac{\sigma^2}{t}\mathrm{Tr}(\tilde{\mathbf{C}}^{(k)}\tilde{\mathbf{T}}\tilde{\mathbf{T}})$. We use the Cauchy-Schwarz inequality:

$$|\mathrm{Tr}(\mathbf{A}\mathbf{B})| \leq \sqrt{\mathrm{Tr}(\mathbf{A}\mathbf{A}^H)}\,\sqrt{\mathrm{Tr}(\mathbf{B}\mathbf{B}^H)}. \tag{69}$$

Taking $\mathbf{A} = \left(\mathbf{C}^{(l)}\right)^{1/2}\mathbf{T}$ and $\mathbf{B} = \left(\mathbf{C}^{(l)}\right)^{1/2}$ in (69) leads to

$$\frac{\sigma^2}{t}\mathrm{Tr}\left(\mathbf{C}^{(l)}\mathbf{T}\mathbf{T}\right) \geq \frac{\sigma^2\left(\frac{1}{t}\mathrm{Tr}\left(\mathbf{C}^{(l)}\mathbf{T}\right)\right)^2}{\frac{1}{t}\mathrm{Tr}\,\mathbf{C}^{(l)}} = \frac{\sigma^2\delta_l^2}{\frac{1}{t}\mathrm{Tr}\,\mathbf{C}^{(l)}}. \tag{70}$$

We now need to lower bound δ_l. Using again inequality (69) with $\mathbf{A} = \left(\mathbf{C}^{(l)}\right)^{1/2}\mathbf{T}^{1/2}$ and $\mathbf{B} = \mathbf{T}^{-1/2}\left(\mathbf{C}^{(l)}\right)^{1/2}$ yields

$$\delta_l = \frac{1}{t}\mathrm{Tr}\left(\mathbf{C}^{(l)}\mathbf{T}\right) \geq \frac{\left(\frac{1}{t}\mathrm{Tr}\,\mathbf{C}^{(l)}\right)^2}{\frac{1}{t}\mathrm{Tr}\left(\mathbf{C}^{(l)}\mathbf{T}^{-1}\right)}. \tag{71}$$

Thanks to (68), $\|\mathbf{T}^{-1}\| = \|\sigma^2(\mathbf{I}_r + \sum_l \tilde{\delta}_l\mathbf{C}^{(l)})\| \leq \sigma^2 + LC_{\sup}^2$. Hence (71) leads to

$$\delta_l \geq \frac{\frac{1}{t}\mathrm{Tr}\,\mathbf{C}^{(l)}}{\|\mathbf{T}^{-1}\|} \geq \frac{\frac{1}{t}\mathrm{Tr}\,\mathbf{C}^{(l)}}{\sigma^2 + LC_{\sup}^2}. \tag{72}$$

Eventually, using (72) in (70) gives

$$\frac{\sigma^2}{t}\mathrm{Tr}\left(\mathbf{C}^{(l)}\mathbf{T}\mathbf{T}\right) \geq \frac{\sigma^2\frac{1}{t}\mathrm{Tr}\,\mathbf{C}^{(l)}}{\left(\sigma^2 + LC_{\sup}^2\right)^2}. \tag{73}$$

Similarly, we prove that

$$\frac{\sigma^2}{t}\mathrm{Tr}\left(\tilde{\mathbf{C}}^{(l)}\tilde{\mathbf{T}}\tilde{\mathbf{T}}\right) \geq \frac{\sigma^2\frac{1}{t}\mathrm{Tr}\,\tilde{\mathbf{C}}^{(l)}}{\left(\sigma^2 + \frac{r}{t}LC_{\sup}^2\right)^2}.$$

Therefore $\inf_t\left\{\min_l \mathbf{v}_l\right\} \geq \frac{\sigma^2\lambda_1}{(\sigma^2+k_1)^2}$, where $\lambda_1 = \min_l\left\{\inf_t\left[\frac{1}{t}\mathrm{Tr}\,\mathbf{C}^{(l)}\right], \inf_t\left[\frac{1}{t}\mathrm{Tr}\,\tilde{\mathbf{C}}^{(l)}\right]\right\} > 0$ and $k_1 = LC_{\sup}^2\max\{1, \inf_t[r/t]\} = LC_{\sup}\lambda_0 < +\infty$. Noting $k_0 = \frac{\lambda_1}{\lambda_0} > 0$ we can now conclude about statement (i) of the lemma:

$$\sup_t \rho(\mathbf{M}(\mathbf{T},\tilde{\mathbf{T}})) \leq 1 - \frac{\inf_t(\min_l \mathbf{v}_l)}{\sup_t(\max_l \mathbf{u}_l)} \leq 1 - \frac{k_0\sigma^4}{(\sigma^2+k_1)^2}.$$

As for statement (ii) of the lemma, we note that $\left|\mathbf{M}(\mathbf{T},\tilde{\mathbf{T}}) - \lambda\mathbf{I}_{2L}\right| = \left|\sigma^4\tilde{\mathbf{A}}(\tilde{\mathbf{T}})\mathbf{A}(\mathbf{T}) - \lambda^2\mathbf{I}_L\right|$. Hence $\rho(\sigma^4\tilde{\mathbf{A}}(\tilde{\mathbf{T}})\mathbf{A}(\mathbf{T})) = (\rho(\mathbf{M}(\mathbf{T},\tilde{\mathbf{T}})))^2 \leq \left(1 - \frac{k_0\sigma^4}{(\sigma^2+k_1)^2}\right)^2 < 1$.

Concerning statement (iii), the proof is the same as in [18, Lemma 5.2]. Nonetheless we provide it here for the sake of completeness. As $\rho(\mathbf{M}(\mathbf{T},\tilde{\mathbf{T}})) < 1$, the series $\sum_{k\in\mathbb{N}}\mathbf{M}(\mathbf{T},\tilde{\mathbf{T}})^k$ converges, matrix $\mathbf{I}_{2L} - \mathbf{M}(\mathbf{T},\tilde{\mathbf{T}})$ is invertible and its inverse can be written as $\left(\mathbf{I}_{2L} - \mathbf{M}(\mathbf{T},\tilde{\mathbf{T}})\right)^{-1} = \sum_{k\in\mathbb{N}}\mathbf{M}(\mathbf{T},\tilde{\mathbf{T}})^k$. Therefore the entries of $\left(\mathbf{I}_{2L} - \mathbf{M}(\mathbf{T},\tilde{\mathbf{T}})\right)^{-1}$ are non-negative. Hence,

$$\begin{aligned}\mathbf{u}_k &= \sum_{l=1}^{2L}\left[(\mathbf{I}_{2L} - \mathbf{M}(\mathbf{T},\tilde{\mathbf{T}}))^{-1}\right]_{kl}\mathbf{v}_l \\ &\geq \min_l(\mathbf{v}_l)\sum_{l=1}^{2L}\left[(\mathbf{I}_{2L} - \mathbf{M}(\mathbf{T},\tilde{\mathbf{T}}))^{-1}\right]_{kl}.\end{aligned}$$

Therefore $\max_k\sum_{l=1}^{2L}\left[\left(\mathbf{I}_{2L} - \mathbf{M}(\mathbf{T},\tilde{\mathbf{T}})\right)^{-1}\right]_{kl} \leq \frac{\max_l(\mathbf{u}_l)}{\min_l(\mathbf{v}_l)}$ and it eventually follows that:

$$\begin{aligned}\sup_t\left[\left\|\left(\mathbf{I}_{2L} - \mathbf{M}(\mathbf{T},\tilde{\mathbf{T}})\right)^{-1}\right\|_\infty\right] &\leq \frac{\sup_t(\max_l \mathbf{u}_l)}{\inf_t(\min_l \mathbf{v}_l)} \\ &\leq \frac{(\sigma^2+k_1)^2}{k_0\sigma^4}.\end{aligned}$$

□

Remark 2: Lemma 1 (ii) is used in the proof of Theorem 1 for the uniqueness of solutions to (9), but we took care not to use any consequences of this uniqueness in the proof above; this proof only requires the existence of solutions to (9).

Remark 3: Unfortunately assumptions $\inf_t\left(\frac{1}{t}\mathrm{Tr}\,\mathbf{C}^{(l)}\right) > 0$ and $\inf_t\left(\frac{1}{t}\mathrm{Tr}\,\tilde{\mathbf{C}}^{(l)}\right) > 0$ made in Lemma 1 cannot be restrained, as $\frac{1}{t}\mathrm{Tr}\left(\mathbf{C}^{(l)}\mathbf{T}\mathbf{T}\right) \leq \frac{1}{\sigma^4}\left(\frac{1}{t}\mathrm{Tr}\,\mathbf{C}^{(l)}\right)$ and similarly $\frac{1}{t}\mathrm{Tr}\left(\tilde{\mathbf{C}}^{(l)}\tilde{\mathbf{T}}\tilde{\mathbf{T}}\right) \leq \frac{1}{\sigma^4}\left(\frac{1}{t}\mathrm{Tr}\,\tilde{\mathbf{C}}^{(l)}\right)$.

Equation (66) shows that the entries of $\mathbf{B}(\mathbf{R},\mathbf{T})$ and $\tilde{\mathbf{B}}(\tilde{\mathbf{R}},\tilde{\mathbf{T}})$ respectively converge to the entries of $\mathbf{A}(\mathbf{T})$ and $\tilde{\mathbf{A}}(\tilde{\mathbf{T}})$. Hence there exists t_0 such that, for $t > t_0$,

- the matrix $\mathbf{I}_{2L} - \mathbf{N}(\mathbf{R},\mathbf{T},\tilde{\mathbf{R}},\tilde{\mathbf{T}})$ is invertible,
- $\sup_t\left[\left\|(\mathbf{I}_{2L} - \mathbf{N}(\mathbf{R},\mathbf{T},\tilde{\mathbf{R}},\tilde{\mathbf{T}}))^{-1}\right\|_\infty\right] \leq \frac{2(\sigma^2+k_1)^2}{k_0\sigma^4}$.

Then, for $t > t_0$, (64) yields

$$\begin{bmatrix}\boldsymbol{\alpha}-\boldsymbol{\delta}\\ \tilde{\boldsymbol{\delta}}-\tilde{\boldsymbol{\alpha}}\end{bmatrix} = \left(\mathbf{I}_{2L} - \mathbf{N}(\mathbf{R},\mathbf{T},\tilde{\mathbf{R}},\tilde{\mathbf{T}})\right)^{-1}\begin{bmatrix}\boldsymbol{\varepsilon}\\ \mathbf{0}\end{bmatrix}. \tag{74}$$

Hence $\max_l\left\{\left|\alpha_l - \delta_l\right|, \left|\tilde{\alpha}_l - \tilde{\delta}_l\right|\right\} \leq \left\|(\mathbf{I}_{2L} - \mathbf{N}(\mathbf{R},\mathbf{T},\tilde{\mathbf{R}},\tilde{\mathbf{T}}))^{-1}\right\|_\infty \max_k|\boldsymbol{\varepsilon}_k|$, and as $\boldsymbol{\varepsilon}_l = \mathrm{Tr}\left(\mathbf{C}^{(l)}\boldsymbol{\Upsilon}\right) = \mathcal{O}\left(\frac{1}{t^2}\right)$ for $l = 1,\ldots,L$, we eventually have that

$$\tilde{\alpha}_l - \tilde{\delta}_l = \mathcal{O}\left(\frac{1}{t^2}\right). \tag{75}$$

Using (75) in (56) completes the proof of Proposition 2.

Appendix D
Integrability of $\mathbb{E}_{\mathbf{H}}[\mathrm{Tr}(\mathbf{T}-\mathbf{S})]$ - Proof of Proposition 3

We first consider $\mathbb{E}_{\mathbf{H}}[\mathrm{Tr}(\mathbf{R}-\mathbf{S})]$, which is equal to $\mathrm{Tr}\,\boldsymbol{\Upsilon}$ by Proposition 1. As noted in Remark 1 of Appendix B, we have $\left|\frac{1}{t}\mathrm{Tr}(\boldsymbol{\Upsilon}\mathbf{A})\right| \leq \frac{1}{\sigma^8 t^2}P_0\left(\frac{1}{\sigma^2}\right)$, where P_0 is a polynomial with real positive coefficients which do not depend on σ^2 nor on t. Therefore

$$\left|\mathbb{E}_{\mathbf{H}}[\mathrm{Tr}(\mathbf{R}-\mathbf{S})]\right| \leq \frac{P_0\left(\frac{1}{\sigma^2}\right)}{\sigma^8 t}. \tag{76}$$

We now consider $\mathrm{Tr}\,(\mathbf{R}-\mathbf{T})$. We showed in Appendix C that there exists t_0 such that, for $t>t_0$, $\mathbf{I}_{2L}-\mathbf{N}(\mathbf{R},\mathbf{T},\tilde{\mathbf{R}},\tilde{\mathbf{T}})$ is invertible and such that $\left\|\left(\mathbf{I}_{2L}-\mathbf{N}(\mathbf{R},\mathbf{T},\tilde{\mathbf{R}},\tilde{\mathbf{T}})\right)^{-1}\right\|_\infty \leq \frac{2(\sigma^2+k_1)^2}{k_0\sigma^4}$, where k_0 and k_1 are given by Lemma 1. Equation (64) then implies

$$\begin{aligned}\left|\tilde{\alpha}_l-\tilde{\delta}_l\right| &\leq \left\|\left(\mathbf{I}_{2L}-\mathbf{N}(\mathbf{R},\mathbf{T},\tilde{\mathbf{R}},\tilde{\mathbf{T}})\right)^{-1}\right\|_\infty \max_k|\boldsymbol{\varepsilon}_k| \\ &\leq \frac{2(\sigma^2+k_1)^2}{k_0\sigma^4}\max_k|\boldsymbol{\varepsilon}_k|\,,\end{aligned}$$

where $\boldsymbol{\varepsilon}_k=\mathrm{Tr}\left(\mathbf{C}^{(k)}\boldsymbol{\Upsilon}\right)$. Besides, Remark 1 of Appendix B ensures that $|\boldsymbol{\varepsilon}_k|\leq\frac{1}{\sigma^8t^2}P_1\left(\frac{1}{\sigma^2}\right)$, where P_1 is a polynomial with real positive coefficients which do not depend on σ^2 nor on t. Hence, for $t>t_0$,

$$\left|\tilde{\alpha}_l-\tilde{\delta}_l\right|\leq\frac{P_1\left(\frac{1}{\sigma^2}\right)}{\sigma^8t^2}\frac{2(\sigma^2+k_1)^2}{k_0\sigma^4} \tag{77}$$

for any $l\in\{1,\ldots,L\}$. Using (77) in (56) with $\mathbf{A}=\mathbf{I}_r$ then gives, for $t>t_0$,

$$\left|\mathrm{Tr}\,(\mathbf{R}-\mathbf{T})\right|\leq\frac{1}{\sigma^8t}\left(\frac{k_2}{\sigma^2}\left(1+\frac{k_1}{\sigma^2}\right)^2P_1\left(\frac{1}{\sigma^2}\right)\right) \tag{78}$$

where $k_2=\frac{2LC_{\sup}}{k_0}\sup_t\{r/t\}<+\infty$.

Eventually, (76) and (78) yield $\left|\mathbb{E}_{\mathbf{H}}\left[\mathrm{Tr}\left(\mathbf{T}-\mathbf{S}\right)\right]\right|\leq\frac{1}{\sigma^8t}P\left(\frac{1}{\sigma^2}\right)$ for $t>t_0$, where the coefficients of the polynomial $P\left(\frac{1}{\sigma^2}\right)=P_0\left(\frac{1}{\sigma^2}\right)+\frac{k_2}{\sigma^2}\left(1+\frac{k_1}{\sigma^2}\right)^2P_1\left(\frac{1}{\sigma^2}\right)$ are real positive coefficients and do not depend on σ^2 nor on t. This completes the proof of Proposition 3.

Appendix E
Differentiability of $\mathbf{Q}\mapsto\boldsymbol{\delta}(\mathbf{Q})$, $\mathbf{Q}\mapsto\tilde{\boldsymbol{\delta}}(\mathbf{Q})$ and $\mathbf{Q}\mapsto\overline{I}(\mathbf{Q})$ – Proof of Proposition 5

We prove in this section that for all $\mathbf{P},\mathbf{Q}\in\mathcal{C}_1$ functions $\boldsymbol{\delta}$ and $\tilde{\boldsymbol{\delta}}$ are Gâteaux differentiable at point $\mathbf{Q}$ in the direction $\mathbf{P}-\mathbf{Q}$, where $\boldsymbol{\delta},\tilde{\boldsymbol{\delta}}$ are defined as the solutions of system (9). The proof is based on the implicit function theorem.

Let $\mathbf{P},\mathbf{Q}\in\mathcal{C}_1$. We introduce the function $\Gamma:\mathbb{R}_+^L\times\mathbb{R}_+^L\times[0,1]\to\mathbb{R}^{2L}$ defined by

$$\Gamma(\boldsymbol{\delta},\tilde{\boldsymbol{\delta}},\lambda)=\begin{bmatrix}\boldsymbol{\delta}-f(\tilde{\boldsymbol{\delta}})\\ \tilde{\boldsymbol{\delta}}-\tilde{f}(\boldsymbol{\delta},\mathbf{Q}+\lambda(\mathbf{P}-\mathbf{Q}))\end{bmatrix},$$

with $f(\tilde{\boldsymbol{\delta}})=\left[f_1(\tilde{\boldsymbol{\delta}}),\ldots,f_L(\tilde{\boldsymbol{\delta}})\right]^T$ and $\tilde{f}(\boldsymbol{\delta},\mathbf{Q})=\left[\tilde{f}_1(\boldsymbol{\delta},\mathbf{Q}),\ldots,\tilde{f}_L(\boldsymbol{\delta},\mathbf{Q})\right]^T$, where the f_l and the $\tilde{f}_l$ are defined by (10). Note that $\boldsymbol{\delta}(\mathbf{Q}+\lambda(\mathbf{P}-\mathbf{Q}))$ and $\tilde{\boldsymbol{\delta}}(\mathbf{Q}+\lambda(\mathbf{P}-\mathbf{Q}))$ are defined by $\Gamma(\boldsymbol{\delta},\tilde{\boldsymbol{\delta}},\lambda)=0$. We want to apply the implicit theorem on a neighbourhood of $\lambda=0$; this requires the differentiability of Γ on this neighbourhood, and the invertibility of the partial Jacobian $D_{(\boldsymbol{\delta},\tilde{\boldsymbol{\delta}})}(\Gamma(\boldsymbol{\delta},\tilde{\boldsymbol{\delta}},\lambda))$ at point $\lambda=0$.

We first note that $f_l:\tilde{\boldsymbol{\delta}}\mapsto\frac{1}{\sigma^2t}\mathrm{Tr}\left[\mathbf{C}^{(l)}\left(\mathbf{I}_r+\sum_k\tilde{\delta}_k\mathbf{C}^{(k)}\right)^{-1}\right]$ is clearly continuously differentiable on $\mathbb{R}_+^L$. Concerning $\tilde{f}_l$, we first need to use the matrix equality $(\mathbf{I}+\mathbf{AB})^{-1}\mathbf{B}=\mathbf{B}(\mathbf{I}+\mathbf{BA})^{-1}$, with $\mathbf{A}=\mathbf{Q}^{1/2}$ and $\mathbf{B}=\tilde{\mathbf{C}}\mathbf{Q}^{1/2}$:

$$\begin{aligned}\tilde{f}_l(\boldsymbol{\delta},\mathbf{Q})&=\frac{1}{\sigma^2t}\mathrm{Tr}\left[\mathbf{Q}^{1/2}\tilde{\mathbf{C}}^{(l)}\mathbf{Q}^{1/2}\left(\mathbf{I}_t+\mathbf{Q}^{1/2}\tilde{\mathbf{C}}(\boldsymbol{\delta})\mathbf{Q}^{1/2}\right)^{-1}\right]\\&=\frac{1}{\sigma^2t}\mathrm{Tr}\left[\tilde{\mathbf{C}}^{(l)}\mathbf{Q}(\mathbf{I}_t+\tilde{\mathbf{C}}(\boldsymbol{\delta})\mathbf{Q})^{-1}\right].\end{aligned} \tag{79}$$

Recall that $\tilde{\mathbf{C}}(\boldsymbol{\delta})=\sum_k\delta_k\tilde{\mathbf{C}}^{(k)}$. Function $(\boldsymbol{\delta},\lambda)\mapsto\tilde{f}(\boldsymbol{\delta},\mathbf{Q}+\lambda(\mathbf{P}-\mathbf{Q}))$ is therefore clearly continuously differentiable on $\mathbb{R}_+^L\times[0,1]$. Nevertheless, as we want to use the implicit theorem for $\lambda=0$, we need to enlarge the continuous differentiability on an open set including $\lambda=0$. Note that for $\lambda<0$, $\mathbf{Q}+\lambda(\mathbf{P}-\mathbf{Q})$ might have negative eigenvalues. Yet, $\det\left[\mathbf{I}_t+\tilde{\mathbf{C}}(\boldsymbol{\delta})(\mathbf{Q}+\lambda(\mathbf{P}-\mathbf{Q}))\right]>0$ for $\boldsymbol{\delta}=\boldsymbol{\delta}(\mathbf{Q})$ and $\lambda=0$. Therefore it exists a neighbourhood V of $(\boldsymbol{\delta}(\mathbf{Q}),0)$ on which $\det\left[\mathbf{I}_t+\tilde{\mathbf{C}}(\boldsymbol{\delta})(\mathbf{Q}+\lambda(\mathbf{P}-\mathbf{Q}))\right]>0$. Defining $\tilde{f}_l$ by (79), the functions $(\boldsymbol{\delta},\lambda)\mapsto\tilde{f}_l(\boldsymbol{\delta},\mathbf{Q}+\lambda(\mathbf{P}-\mathbf{Q}))$ are continuously differentiable on V. Hence, $\Gamma(\boldsymbol{\delta},\tilde{\boldsymbol{\delta}},\lambda)$ is continuously differentiable on $\mathbb{R}_+^L\times V$.

We still have to check that the partial Jacobian $D_{(\boldsymbol{\delta},\tilde{\boldsymbol{\delta}})}(\Gamma(\boldsymbol{\delta},\tilde{\boldsymbol{\delta}},\lambda))$ is invertible at the point $\lambda=0$.

$$\begin{aligned}D_{(\boldsymbol{\delta},\tilde{\boldsymbol{\delta}})}\Gamma_{(\boldsymbol{\delta},\tilde{\boldsymbol{\delta}},0)}&=\begin{bmatrix}\mathbf{I}_L-D_{\boldsymbol{\delta}}f_{(\tilde{\boldsymbol{\delta}})} & -D_{\tilde{\boldsymbol{\delta}}}f_{(\tilde{\boldsymbol{\delta}})}\\ -D_{\boldsymbol{\delta}}\tilde{f}_{(\boldsymbol{\delta},0)} & \mathbf{I}_L-D_{\tilde{\boldsymbol{\delta}}}\tilde{f}_{(\boldsymbol{\delta},0)}\end{bmatrix}\\&=\begin{bmatrix}\mathbf{I}_L & -\sigma^2\mathbf{A}(\mathbf{T})\\ -\sigma^2\tilde{\mathbf{A}}(\tilde{\mathbf{T}}) & \mathbf{I}_L\end{bmatrix}=\mathbf{M}(\mathbf{T},\tilde{\mathbf{T}}),\end{aligned}$$

where $\mathbf{A}_{kl}(\mathbf{T})=\frac{1}{t}\mathrm{Tr}(\mathbf{C}^{(k)}\mathbf{T}\mathbf{C}^{(l)}\mathbf{T})$ and $\tilde{\mathbf{A}}_{kl}(\tilde{\mathbf{T}})=\frac{1}{t}\mathrm{Tr}(\mathbf{Q}^{1/2}\tilde{\mathbf{C}}^{(k)}\mathbf{Q}^{1/2}\tilde{\mathbf{T}}\mathbf{Q}^{1/2}\tilde{\mathbf{C}}^{(l)}\mathbf{Q}^{1/2}\tilde{\mathbf{T}})$, and with $\mathbf{T}=\mathbf{T}(\tilde{\boldsymbol{\delta}}(\mathbf{Q}))$ and $\tilde{\mathbf{T}}=\tilde{\mathbf{T}}(\boldsymbol{\delta}(\mathbf{Q}))$ respectively defined by (11) and (12). Matrices $\mathbf{A}(\mathbf{T})$, $\tilde{\mathbf{A}}(\tilde{\mathbf{T}})$ and $\mathbf{M}(\mathbf{T},\tilde{\mathbf{T}})$ correspond to those defined in Lemma 1, but in which $\tilde{\mathbf{C}}^{(l)}$ is replaced by $\mathbf{Q}^{1/2}\tilde{\mathbf{C}}^{(l)}\mathbf{Q}^{1/2}$. Lemma 1 item (i) therefore gives the invertibility of $D_{(\boldsymbol{\delta},\tilde{\boldsymbol{\delta}})}\Gamma$ at point $\lambda=0$.

We now are in position to apply the implicit function theorem, which asserts that functions $\lambda\mapsto\boldsymbol{\delta}(\mathbf{Q}+\lambda(\mathbf{P}-\mathbf{Q}))$ and $\lambda\mapsto\tilde{\boldsymbol{\delta}}(\mathbf{Q}+\lambda(\mathbf{P}-\mathbf{Q}))$ are continuously differentiable on a neighbourhood of 0. Hence, $\boldsymbol{\delta}$ and $\tilde{\boldsymbol{\delta}}$ are Gâteaux differentiable at point $\mathbf{Q}$ in the direction $\mathbf{P}-\mathbf{Q}$. As $\overline{I}(\mathbf{Q})=\log\left|\mathbf{I}_r+\sum_l\tilde{\delta}_l(\mathbf{Q})\mathbf{C}^{(l)}\right|+\log\left|\mathbf{I}_t+\mathbf{Q}\left(\sum_l\delta_l(\mathbf{Q})\tilde{\mathbf{C}}^{(l)}\right)\right|-\sigma^2t\left(\sum_l\delta_l(\mathbf{Q})\tilde{\delta}_l(\mathbf{Q})\right)$ it is clear that $\mathbf{Q}\mapsto\overline{I}(\mathbf{Q})$ is as well Gâteaux differentiable at point $\mathbf{Q}$ in the direction $\mathbf{P}-\mathbf{Q}$.

References

[1] A. Goldsmith, S. Jafar, N. Jindal, and S. Vishwanath, "Capacity limits of MIMO channels," *IEEE J. Sel. Areas Commun.*, vol. 21, no. 5, pp. 684–702, Jun 2003.

[2] D. Hösli, Y. Kim, and A. Lapidoth, "Monotonicity results for coherent MIMO Rician channels," *IEEE Trans. Inform. Theory*, vol. 51, no. 12, pp. 4334–4339, Dec 2005.

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

[4] A. Moustakas and S. Simon, "On the outage capacity of correlated multiple-path MIMO channels," *IEEE Trans. Inform. Theory*, vol. 53, no. 11, p. 3887, Nov 2007.

[5] M. Vu and A. Paulraj, "Capacity optimization for Rician correlated MIMO wireless channels," in *Proc. Asilomar Conference*, 2005, pp. 133–138.

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

[7] C. Wen, P. Ting, and J. Chen, "Asymptotic analysis of MIMO wireless systems with spatial correlation at the receiver," *IEEE Trans. Commun.*, vol. 54, no. 2, pp. 349–363, Feb 2006.

[8] J. Dumont, W. Hachem, S. Lasaulce, P. Loubaton, and J. Najim, "On the capacity achieving covariance matrix for Rician MIMO channels: an asymptotic approach," *IEEE Trans. Inform. Theory*, vol. 56, no. 3, pp. 1048–1069, Mar 2010.

[9] ——, "On the capacity achieving covariance matrix for Rician MIMO channels: an asymptotic approach," 2007, extended version of [8]. [Online]. Available: http://arxiv.org/abs/0710.4051
[10] C. Artigue, P. Loubaton, and B. Mouhouche, "On the ergodic capacity of frequency selective MIMO systems equipped with MMSE receivers: An asymptotic approach," in *Proc. Globecom*, 2008.
[11] R. Horn and C. Johnson, *Matrix analysis*. Cambridge Univ Pr, 1990.
[12] ——, *Topics in matrix analysis*. Cambridge Univ Pr, 1994.
[13] J. Borwein and A. Lewis, *Convex analysis and nonlinear optimization : theory and examples*. New York: Springer Verlag, 2000.
[14] M. Kreĭn and A. Nudel'man, *The Markov moment problem and extremal problems*. Translations of Mathematical Monographs, American Mathematical Society, Providence, 1977, vol. 50.
[15] W. Hachem, P. Loubaton, and J. Najim, "Deterministic equivalents for certain functionals of large random matrices," *Ann. Appl. Probab.*, vol. 17, no. 3, pp. 875–930, 2007.
[16] L. Pastur, "A simple approach to the global regime of Gaussian ensembles of random matrices," *Ukrainian Math. J.*, vol. 57, no. 6, pp. 936–966, Jun 2005.
[17] E. Novikov, "Functionals and the random-force method in turbulence theory," *Soviet Physics-JETP*, vol. 20, pp. 1290–1294, 1965.
[18] W. Hachem, P. Loubaton, and J. Najim, "A CLT for information-theoretic statistics of Gram random matrices with a given variance profile," *Ann. Appl. Probab.*, vol. 18, no. 6, pp. 2071–2130, Dec 2008.

Florian Dupuy Florian Dupuy was born in Herbolzheim, Germany, in 1983. He received the M.Sc. degree from Supélec, France, and Helsinki University of Technology (now Aalto University), Finland, in 2008.

Since 2009 he has been working towards the Ph.D. degree through a CIFRE convention between Thales Communications and Université Paris Est under the direction of both Prof. Chevalier and Prof. Loubaton.

His research interests are MIMO capacity optimization, random matrix theory, widely linear receivers for noncircular signals and interference cancellation.

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

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

Dr. Loubaton served as associate editor for the IEEE Transactions on Signal Processing (1999-2001, 2005-2007) and for the IEEE Communications Letters (2001-2003). He is currently associate editor for Eurasip Signal Processing. He was a member of the IEEE Signal Processing for Communications technical committee (1998-2004).