A New Approach for Mutual Information Analysis of Large Dimensional Multi-Antenna Channels

W. Hachem(*), O. Khorunzhiy, Ph. Loubaton, J. Najim and L. Pastur

Abstract

This paper adresses the behaviour of the mutual information of correlated MIMO Rayleigh channels when the numbers of transmit and receive antennas converge to $+\infty$ at the same rate. Using a new and simple approach based on Poincaré-Nash inequality and on an integration by parts formula, it is rigorously established that the mutual information when properly centered and rescaled converges to a Gaussian random variable whose mean and variance are evaluated. These results confirm previous evaluations based on the powerful but non rigorous replica method. It is believed that the tools that are used in this paper are simple, robust, and of interest for the communications engineering community.

Index Terms

Central Limit Theorem, Correlated MIMO Channels, Large Random Matrix Theory, Mutual Information, Poincaré-Nash Inequality.

This work was partially supported by the “Fonds National de la Science” via the ACI program “Nouvelles Interfaces des Mathématiques”, project MALCOM number 205 .

W. Hachem and J. Najim are with CNRS and ENST (UMR 5141), Paris, France. {hachem,najim}@enst.fr ,

O. Khorunzhiy is with Equipe ”Probabilités-Statistiques” Université de Versailles - Saint-Quentin, France khorunjy@math.uvsq.fr ,

P. Loubaton is with IGM LabInfo, UMR 8049, Institut Gaspard Monge, Université de Marne La Vallée, France. loubaton@univ-mlv.fr ,

L. Pastur is with Kharkiv National University - Institute for Low Temperature Physics - Kharkiv, Ukraine lpastur@flint.ilt.kharkov.ua .

(*) Corresponding author.

I. INTRODUCTION

Consider the following channel model with additive white noise: $\mathbf{y} = \mathbf{G}\mathbf{x} + \mathbf{z}$, where $\mathbf{y}$ is the received signal, $\mathbf{G}$ is the channel matrix, $\mathbf{x}$ is the vector of transmitted symbols, and $\mathbf{z}$ is a complex white Gaussian noise. It is widely known that high spectral efficiencies are attained when multiple antennas are used at both the transmitter and the receiver of a wireless communication system. Indeed, due to the mobility and to the presence of a large number of reflected and scattered signal paths, the elements of the $N \times n$ Multiple Input Multiple Output (MIMO) channel matrix with N antennas at the receiver's site and n antennas at the transmitter's are often modeled as random variables. Assuming a random model for this matrix, Telatar [1] and Foschini [2] realized in the mid-nineties that Shannon's mutual information of such channels increases at the rate of $\min(N, n)$ for a fixed transmission power [1]. The authors of [1] and [2] assumed that the elements of the channel matrix $\mathbf{G}$ are centered, independent and identically distributed (i.i.d.) elements. In this context, a well known result in Random Matrix Theory (RMT) [3] says that the eigenvalue distribution of the Gram matrix $\mathbf{G}\mathbf{G}^*$ where $\mathbf{G}^*$ is the Hermitian adjoint of $\mathbf{G}$ converges to a deterministic probability distribution as n goes to infinity and N/n converges to a constant $c > 0$. Denote by $I(\rho) = \log\det\left(\frac{\rho}{n}\mathbf{G}\mathbf{G}^* + \mathbf{I}_N\right)$ the mutual information of channel $\mathbf{G}$ for a Signal to Noise Ratio at a receiver antenna equal to ρ/n. One consequence of [3] is that the mutual information per transmit antenna $I(\rho)/n$, being an integral of a $\log$ function with respect to the empirical eigenvalue distribution of $\mathbf{G}\mathbf{G}^*$, converges to a constant. This fact already observed in [1] sustains the assertion of the linear increase of mutual information with the number of antennas. In addition, this convergence proves to be sufficiently fast. As a matter of fact, the asymptotic results predicted by the RMT remain relevant for systems with a moderate number of antennas.

The next step was to apply this theory to channel models that include a correlation between paths (or entries of $\mathbf{G}$). One of the main purposes of this generalization is to better understand the impact of these correlations on Shannon's mutual information. Let us cite in this context the contributions [4], [5], [6], [7] and [8], all devoted to the study of the mutual information in the case where the elements of channel's matrix are centered and correlated random variables. In [9], a deterministic equivalent is computed under broad conditions for the mutual information based on Rice channels modeled by non-centered matrices with independent but not identically distributed random variables. The link between matrices with correlated entries and matrices with independent entries and a variance profile is studied in [10]. In a slightly different context, concentration inequalities for the capacity of a CDMA system are established in [11].

One of the most popular correlated channel models used for these mutual information evaluations is the so-called Kronecker model $\mathbf{G} = \mathbf{\Psi}\mathbf{W}\widetilde{\mathbf{\Psi}}$ where $\mathbf{W}$ is a $N \times n$ matrix with Gaussian centered i.i.d. entries, and $\mathbf{\Psi}$ and $\widetilde{\mathbf{\Psi}}$ are $N \times N$ and $n \times n$ matrices that capture the path correlations at the receiver and at the transmitter sides respectively [12], [13]. This model has been studied by Chuah et. al. in [5]. With some assumptions on matrices $\mathbf{\Psi}$ and $\widetilde{\mathbf{\Psi}}$, these authors showed that $I(\rho)/n$ converges to a deterministic quantity defined as the fixed point of an integral equation. Later on, Tulino et. al. [8] obtained the limit of $I(\rho)/n$ for a correlation model more general than the Kronecker model. Both these works rely on a result of Girko describing the eigenvalue distribution of the Gram matrix associated with a matrix with independent but non necessarily identically distributed entries, a close model as we shall see in a moment. In [7], Moustakas et. al. studied the mutual information for the Kronecker model by using the so-called replica method. They found an approximation $V(\rho)$ of $\mathbb{E}\left[I(\rho)\right]$ accurate to the order $1/n$ in the large n regime. Using this same method, they also showed that the variance of $I(\rho) - V(\rho)$ is of order one and were able to derive this variance for large n.

Although the replica technique is powerful and has a wide range of applications, the rigorous justification of some of its parts remains to be done. In this paper, we propose a new method to study the convergence of $\mathbb{E}I(\rho)$ and the fluctuations of $I(\rho)$. Beside recovering the results in [7] and especially the strikingly simple form of the variance, we establish the Central Limit Theorem (CLT) for $I(\rho) - V(\rho)$ (for a related CLT in a non-Gaussian context, see [14]). The practical interest of such a result is of importance since the CLT leads to an evaluation of the outage probability, i.e. the probability that $I(\rho)$ lies beneath a given threshold, by means of the Gaussian approximation. Many other works have been devoted to CLT for random matrices. Close to our present article are [15], [16], [17].

In this article, we also would like to advocate the method used to establish both the approximation of $I(\rho)$ in the large n regime and the CLT. Due to the Gaussian character of the entries of Matrix $\mathbf{G}$, two simple ingredients are available. The first one is an Integration by parts formula (17) that provides an expression for the expectation of certain functionals of Gaussian vectors. This formula has been widely used in RMT [18]–[20]. The second ingredient is Poincaré-Nash inequality (18) that bounds the variance of functionals of Gaussian vectors. Although well known [21], [22], its application to RMT is fairly recent [23], [20] (see also [24] and [25] where general concentration inequalities are derived for functions of random matrices). This inequality enables us to control the decrease rate of the approximation errors such as the order $1/n$ error $\mathbb{E}\left[I(\rho)\right] - V(\rho)$ (note that the gaussian structure enters in two places: First the reduction to matrices with independent entries et varying variance and then integration by part and

Poincaré-Nash bounds for the variance of relevant spectral characteristics[1]). We believe that these tools of rigorous and explicit analysis might be of great interest for the communications engineering community (see for instance the estimates obtained in [26] in the context of Ricean MIMO channels).

The paper is organized as follows. In Section II, we introduce the main notations; we also state the two main results of the article. In Section III, we recall general matrix results and the two aforementioned Gaussian tools. Section IV is devoted to the proof of the first order result, that is the approximation of $\mathbb{E}[I(\rho)]$. The CLT, also refered to as the second order result, is established in Section V. Proof details are in an appendix.

II. Notations and Statement of the main results

A. From a Kronecker model to a separable variance model.

Consider a MIMO system represented by a $N \times n$ matrix $\mathbf{G}$ where n is the number of antennas at the transmitter and N is the number of antennas at the receiver and where $N(n)$ is a sequence of integers such that

$$0 < \quad \ell^- = \liminf_{n\to\infty} \frac{N(n)}{n} \quad \leq \quad \ell^+ = \limsup_{n\to\infty} \frac{N(n)}{n} \quad < \infty \ , \tag{1}$$

a condition we shall refer to by writing $n, N \to \infty$. Assuming the transmitted signal is a Gaussian signal with a covariance matrix equal to $\frac{1}{n}\mathbf{I}_n$ (and thus, a total power equal to one), Shannon's mutual information of this channel is $I_n(\rho) = \log\det\left(\frac{\rho}{n}\mathbf{G}_n\mathbf{G}_n^* + \mathbf{I}_N\right)$, where $\rho > 0$ is the inverse of the additive white Gaussian noise variance at each receive antenna. The general problem we address in this paper concerns the behaviour of the mutual information for large values of N and n in the case where the channel matrix $\mathbf{G}_n$, assumed to be random, is described by the Kronecker model $\mathbf{G}_n = \mathbf{\Psi}_n\mathbf{W}_n\widetilde{\mathbf{\Psi}}_n$. In this model, $\mathbf{\Psi}_n$ and $\widetilde{\mathbf{\Psi}}_n$ are respectively $N \times N$ and $n \times n$ deterministic matrices and $\mathbf{W}_n$ is random with independent entries distributed acccording to the complex circular Gaussian law with mean zero and variance one $\mathcal{CN}(0,1)$.

It is well known that this model can be replaced by a simpler Kronecker model involving a matrix with Gaussian independent (but not necessarily identically distributed) entries. Indeed, let $\mathbf{\Psi}_n = \mathbf{U}_n\mathbf{D}_n^{\frac{1}{2}}\mathbf{V}_n^*$ (resp. $\widetilde{\mathbf{\Psi}}_n = \widetilde{\mathbf{U}}_n\widetilde{\mathbf{D}}_n^{\frac{1}{2}}\widetilde{\mathbf{V}}_n^*$) be a Singular Value Decomposition (SVD) of $\mathbf{\Psi}_n$ (resp. $\widetilde{\mathbf{\Psi}}_n$), where $\mathbf{D}_n$ (resp. $\widetilde{\mathbf{D}}_n$) is the diagonal matrix of eigenvalues of $\mathbf{\Psi}_n\mathbf{\Psi}_n^*$ (resp. $\widetilde{\mathbf{\Psi}}_n\widetilde{\mathbf{\Psi}}_n^*$), then $I_n(\rho)$ writes:

$$I_n(\rho) = \log\det\left(\frac{\rho}{n}\mathbf{Y}_n\mathbf{Y}_n^* + \mathbf{I}_N\right),$$

[1]It is interesting to note that once the first reduction has been made, others techniques are available without using the gaussian assumption - see for instance [20].

where $\mathbf{Y}_n = \mathbf{D}_n^{\frac{1}{2}} \mathbf{X}_n \widetilde{\mathbf{D}}_n^{\frac{1}{2}}$ is a $N \times n$ matrix, $\mathbf{D}_n$ and $\widetilde{\mathbf{D}}_n$ are respectively $N \times N$ and $n \times n$ diagonal matrices, i.e.

$$\mathbf{D}_n = \operatorname{diag}\left(d_i^{(n)},\ 1 \leq i \leq N\right) \quad \text{and} \quad \widetilde{\mathbf{D}}_n = \operatorname{diag}\left(\tilde{d}_j^{(n)},\ 1 \leq j \leq n\right),$$

and $\mathbf{X}_n = \mathbf{V}_n^* \mathbf{W}_n \widetilde{\mathbf{U}}_n$ has i.i.d. entries with distribution $\mathcal{CN}(0,1)$ since $\mathbf{V}_n$ and $\widetilde{\mathbf{U}}_n$ are deterministic unitary matrices. Since every individual entry of $\mathbf{Y}_n$ has the form $Y_{ij}^{(n)} = \sqrt{d_i^{(n)} \tilde{d}_j^{(n)}} X_{ij}$, we call $\mathbf{Y}_n$ a random matrix with a separable variance profile.

B. Assumptions and Notations.

The centered random variable $X - \mathbb{E}[X]$ will be denoted by $\overset{\circ}{X}$. Element (i,j) of a matrix $\mathbf{A}$ will be either denoted $[\mathbf{A}]_{ij}$ or A_{ij}. Element i of vector $\mathbf{a}$ will be denoted a_i or $[\mathbf{a}]_i$. Column j of matrix $\mathbf{A}$ will be denoted $\mathbf{a}_j$. The transpose, the Hermitian adjoint (conjugate transpose) of $\mathbf{A}$, and the matrix obtained by conjugating its elements are denoted respectively $\mathbf{A}^T$, $\mathbf{A}^*$, and $\overline{\mathbf{A}}$. The spectral norm of a matrix $\mathbf{A}$ will be denoted $\|\mathbf{A}\|$. If $\mathbf{A}$ is square, $\operatorname{Tr} \mathbf{A}$ refers to its trace. Let $\mathbf{i} = \sqrt{-1}$, then the operators $\partial/\partial z$ and $\partial/\partial \overline{z}$ where $z = x + \mathbf{i}y$ is a complex number are defined by $\frac{\partial}{\partial z} = \frac{1}{2}\left(\frac{\partial}{\partial x} - \mathbf{i}\frac{\partial}{\partial y}\right)$ and $\frac{\partial}{\partial \overline{z}} = \frac{1}{2}\left(\frac{\partial}{\partial x} + \mathbf{i}\frac{\partial}{\partial y}\right)$ where $\frac{\partial}{\partial x}$ and $\frac{\partial}{\partial y}$ are the standard partial derivatives with respect to x and y.

Throughout the paper, notation K will denote a generic constant whose main feature is *not* to depend on n. In particular, the value of K might change from a line to another as long as it never depends upon n. Constant K might depend on $t \in \mathbb{R}^+$ and whenever needed, this dependence will be made more explicit.

As usual notation $\alpha_n = \mathcal{O}(\beta_n)$ is a flexible shortcut for $|\alpha_n| \leq K\beta_n$ and $\alpha_n = o(\beta_n)$, for $\alpha_n = \varepsilon_n \beta_n$ with $\varepsilon_n \to 0$ as $n, N \to \infty$ (in the sense of (1)).

In order to study a deterministic approximation of $I_n(\rho)$ and its fluctuations, the following mild assumptions are required over the two triangular arrays $\left(d_i^{(n)},\ 1 \leq i \leq N,\ n \geq 1\right)$ and $\left(\tilde{d}_j^{(n)}, 1 \leq j \leq n,\ n \geq 1\right)$.

(A1) The real numbers $d_i^{(n)}$ and $\tilde{d}_j^{(n)}$ are nonnegative and the sequences $\left(d_i^{(n)}\right)$ and $\left(\tilde{d}_j^{(n)}\right)$ are uniformly bounded, i.e. there exist constants $d_{\max}$ and $\tilde{d}_{\max}$ such that

$$\sup_n \|\mathbf{D}_n\| < d_{\max} \quad \text{and} \quad \sup_n \|\widetilde{\mathbf{D}}_n\| < \tilde{d}_{\max}.$$

where $\|\mathbf{D}_n\|$ and $\|\widetilde{\mathbf{D}}_n\|$ are the spectral norms of $\mathbf{D}_n$ and $\widetilde{\mathbf{D}}_n$.

(A2) The normalized traces of $\mathbf{D}_n$ and $\widetilde{\mathbf{D}}_n$ satisfy

$$\inf_n \frac{1}{n} \operatorname{Tr}\left(\mathbf{D}_n\right) > 0 \quad \text{and} \quad \inf_n \frac{1}{n} \operatorname{Tr}\left(\widetilde{\mathbf{D}}_n\right) > 0.$$

In the sequel, we shall frequently omit the subscript n and the superscript (n).

The resolvent associated with $\frac{1}{n}\mathbf{Y}_n\mathbf{Y}_n^*$ is the $N \times N$ matrix $\mathbf{H}_n(t) = \left(\frac{t}{n}\mathbf{Y}_n\mathbf{Y}_n^* + \mathbf{I}_N\right)^{-1}$. Of prime importance is the random variable $\beta(t) = \frac{1}{n}\mathrm{Tr}\,\mathbf{D}_n\mathbf{H}_n(t)$ and its expectation $\alpha(t) = \frac{1}{n}\mathrm{Tr}\,\mathbf{D}_n\,\mathbb{E}\mathbf{H}_n(t)$. We furthermore introduce the $n \times n$ deterministic matrix defined by

$$\begin{aligned}\widetilde{\mathbf{R}}_n(t) &= \left(\mathbf{I} + t\alpha(t)\widetilde{\mathbf{D}}_n\right)^{-1}, \\ &= \mathrm{diag}\left(\tilde{r}_j(t),\ 1 \le j \le n\right) \quad \text{where} \quad \tilde{r}_j(t) = \frac{1}{1 + t\alpha(t)\tilde{d}_j},\end{aligned}$$

and the related quantity $\tilde{\alpha}(t) = \frac{1}{n}\mathrm{Tr}\,\widetilde{\mathbf{D}}_n\widetilde{\mathbf{R}}_n(t)$. In a symmetric fashion, the $N \times N$ matrix $\mathbf{R}_n(t)$ is defined by

$$\begin{aligned}\mathbf{R}_n(t) &= \left(\mathbf{I} + t\tilde{\alpha}(t)\mathbf{D}_n\right)^{-1}, \\ &= \mathrm{diag}\left(r_i(t),\ 1 \le i \le N\right) \quad \text{where} \quad r_i(t) = \frac{1}{1 + t\tilde{\alpha}(t)d_i}.\end{aligned}$$

We finally introduce the solutions of a deterministic 2×2 system.

Proposition 1: For every n, the system of equations in $(\delta, \tilde{\delta})$

$$\begin{cases} \delta &= \frac{1}{n}\mathrm{Tr}\,\mathbf{D}_n(\mathbf{I} + t\tilde{\delta}\mathbf{D}_n)^{-1} \\ \tilde{\delta} &= \frac{1}{n}\mathrm{Tr}\,\widetilde{\mathbf{D}}_n(\mathbf{I} + t\delta\widetilde{\mathbf{D}}_n)^{-1} \end{cases} \tag{2}$$

admits a unique solution $\left(\delta_n(t), \tilde{\delta}_n(t)\right)$ satisfying $\delta_n(t) > 0, \tilde{\delta}_n(t) > 0$. Moreover, there exist nonnegative measures μ_n and $\tilde{\mu}_n$ over $\mathbb{R}^+$ such that

$$\delta_n(t) = \int_{\mathbb{R}^+} \frac{\mu_n(d\lambda)}{1 + t\lambda} \qquad \text{and} \qquad \tilde{\delta}_n(t) = \int_{\mathbb{R}^+} \frac{\tilde{\mu}_n(d\lambda)}{1 + t\lambda}\ , \tag{3}$$

where $\mu_n(\mathbb{R}^+) = \frac{1}{n}\mathrm{Tr}\,\mathbf{D}_n$ and $\tilde{\mu}_n(\mathbb{R}^+) = \frac{1}{n}\mathrm{Tr}\,\widetilde{\mathbf{D}}_n$.

The proof is postponed to Appendix A.

With δ and $\tilde{\delta}$ properly defined, we introduce the following $N \times N$ and $n \times n$ diagonal matrices:

$$\mathbf{T}_n = (\mathbf{I}_N + t\tilde{\delta}\mathbf{D}_n)^{-1} \quad \text{and} \quad \widetilde{\mathbf{T}}_n = (\mathbf{I}_n + t\delta\widetilde{\mathbf{D}}_n)^{-1}.$$

Notice in particular that $\delta = \frac{1}{n}\mathrm{Tr}\ \mathbf{D}_n\mathbf{T}_n$ and $\tilde{\delta} = \frac{1}{n}\mathrm{Tr}\ \widetilde{\mathbf{D}}_n\widetilde{\mathbf{T}}_n$ by (2). We finally introduce the following quantities which are required to express the fluctuations of $I_n(\rho)$:

$$\begin{cases} \gamma_n(t) = \frac{1}{n}\mathrm{Tr}\,\mathbf{D}_n^2\mathbf{T}_n^2(t) \\ \tilde{\gamma}_n(t) = \frac{1}{n}\mathrm{Tr}\,\widetilde{\mathbf{D}}_n^2\widetilde{\mathbf{T}}_n^2(t) \end{cases}. \tag{4}$$

Proposition 2: Assume that Assumptions **(A1)** and **(A2)** hold and denote by

$$\sigma_n^2(t) = -\log\left(1 - t^2\gamma_n(t)\tilde{\gamma}_n(t)\right), \quad t > 0 \tag{5}$$

where $\gamma_n(t)$ and $\tilde{\gamma}_n(t)$ are given by (4). Then $\sigma_n^2(t)$ is well-defined, i.e. $1 - t^2\gamma_n(t)\tilde{\gamma}_n(t) > 0$ for $t > 0$. Moreover there exist nonnegative real numbers m_t and M_t such that

$$0 < m_t^2 \leq \inf_n \sigma_n^2(t) \leq \sup_n \sigma_n^2(t) \leq M_t^2 < \infty \quad \text{for} \quad t > 0 \ . \tag{6}$$

Moreover, $\sigma_n^2(t)$ is upper-bounded uniformly in n and t for $t \in [0, \rho]$, i.e. $\sup_{t \leq \rho} M_t^2 < \infty$.

Proof of Proposition 2 is postponed to Appendix B.

Summary of the main notations.

In order to improve the readability of the paper, we gather all the notations in Table II-B. As expressed there, there are three kinds of quantities:

1) Random quantities,
2) Deterministic quantities depending on the law of $\mathbf{Y}_n\mathbf{Y}_n^*$ via the expectation $\mathbb{E}$ with respect to the entries of $\mathbf{Y}_n$,
3) Deterministic quantities which only depend on the matrices $\mathbf{D}_n$ and $\widetilde{\mathbf{D}}_n$, sometimes via δ and $\tilde{\delta}$ (as defined in Proposition 1) which are easily computable.

The main goal of the forthcoming computations will be to approximate elements of the first and second kind by elements of the third kind.

C. Statement of the main results.

We now state the main results. Theorem 1 describes the first order approximation of the mutual information $I_n(\rho)$ while Theorem 2 describes its fluctuations when centered with respect to its first order approximation.

Theorem 1: Let $\mathbf{X}_n$ be a $N \times n$ matrix whose elements X_{ij} are independent complex Gaussian variables such that

$$\mathbb{E}(X_{ij}) = \mathbb{E}(X_{ij}^2) = 0, \quad \mathbb{E}(|X_{ij}|^2) = 1, \quad 1 \leq i \leq N, \ 1 \leq j \leq n,$$

and $\mathbf{Y}_n = \mathbf{D}_n^{\frac{1}{2}}\mathbf{X}_n\widetilde{\mathbf{D}}_n^{\frac{1}{2}}$ where the diagonal matrices $\mathbf{D}_n$ and $\widetilde{\mathbf{D}}_n$ satisfy Assumptions **(A1)** and **(A2)**. Let $I_n(\rho) = \log\det\left(\frac{\rho}{n}\mathbf{Y}_n\mathbf{Y}_n^* + \mathbf{I}_N\right)$. Then, we have

$$\mathbb{E}[I_n(\rho)] = V_n(\rho) + \mathcal{O}\left(\frac{1}{n}\right) \tag{7}$$

as $n, N \to \infty$ (in the sense of (1)), where

$$V_n(\rho) = \log\det\left(\mathbf{I}_n + \rho\delta_n(\rho)\widetilde{\mathbf{D}}_n\right) + \log\det\left(\mathbf{I}_N + \rho\tilde{\delta}_n(\rho)\mathbf{D}_n\right) - n\rho\delta_n(\rho)\tilde{\delta}_n(\rho) \ .$$

Random quantities	Deterministic quantities	
	depending on the law of $\mathbf{YY}^*$ via $\mathbb{E}$	only depending on the variance structure via $\mathbf{D}$ and $\widetilde{\mathbf{D}}$
$\mathbf{H} = \left(\frac{t}{n}\mathbf{YY}^* + \mathbf{I}\right)^{-1}$		
$\beta = \frac{1}{n}\mathrm{Tr}\,\mathbf{DH}$	$\alpha = \frac{1}{n}\mathrm{Tr}\,\mathbf{D}(\mathbb{E}\mathbf{H})$	$\delta = \frac{1}{n}\mathrm{Tr}\,\mathbf{D}(\mathbf{I} + t\tilde{\delta}\mathbf{D})^{-1} = \frac{1}{n}\mathrm{Tr}\,\mathbf{DT}$
	$\tilde{r}_j = (1 + t\alpha\tilde{d}_j)^{-1}$	
	$\widetilde{\mathbf{R}} = (\mathbf{I} + t\alpha\widetilde{\mathbf{D}})^{-1}$	$\widetilde{\mathbf{T}} = (\mathbf{I} + t\delta\widetilde{\mathbf{D}})^{-1}$
	$\tilde{\alpha} = \frac{1}{n}\mathrm{Tr}\,\widetilde{\mathbf{D}}\widetilde{\mathbf{R}} = \frac{1}{n}\mathrm{Tr}\,\widetilde{\mathbf{D}}(\mathbf{I} + t\alpha\widetilde{\mathbf{D}})^{-1}$	$\tilde{\delta} = \frac{1}{n}\mathrm{Tr}\,\widetilde{\mathbf{D}}(\mathbf{I} + t\delta\widetilde{\mathbf{D}})^{-1} = \frac{1}{n}\mathrm{Tr}\,\widetilde{\mathbf{D}}\widetilde{\mathbf{T}}$
	$r_i = (1 + t\tilde{\alpha}d_i)^{-1}$	
	$\mathbf{R} = (\mathbf{I} + t\tilde{\alpha}\mathbf{D})^{-1}$	$\mathbf{T} = (\mathbf{I} + t\tilde{\delta}\mathbf{D})^{-1}$
		$\gamma = \frac{1}{n}\mathrm{Tr}\,\mathbf{T}^2\mathbf{D}^2, \quad \tilde{\gamma} = \frac{1}{n}\mathrm{Tr}\,\widetilde{\mathbf{T}}^2\widetilde{\mathbf{D}}^2$
		$\sigma^2(t) = -\log(1 - t^2\gamma(t)\tilde{\gamma}(t))$

TABLE I

SUMMARY OF THE MAIN NOTATIONS

and where $(\delta_n(t), \tilde{\delta}_n(t))$ is the unique positive solution of the system

$$\left\{ \begin{array}{rcl} \delta & = & \frac{1}{n}\mathrm{Tr}\,\mathbf{D}_n(\mathbf{I}_N + t\tilde{\delta}\mathbf{D}_n)^{-1} \\ \tilde{\delta} & = & \frac{1}{n}\mathrm{Tr}\,\widetilde{\mathbf{D}}_n(\mathbf{I}_n + t\delta\widetilde{\mathbf{D}}_n)^{-1} \end{array} \right. .$$

Theorem 2: Assume that the setting of Theorem 1 holds and let $\sigma_n^2(\rho) = -\log\left(1 - \rho^2\gamma_n(\rho)\tilde{\gamma}_n(\rho)\right)$. Then the following convergence holds true:

$$\frac{I_n(\rho) - V_n(\rho)}{\sigma_n(\rho)} \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0,1) \ ,$$

where $\xrightarrow{\mathcal{L}}$ stands for the convergence in distribution.

III. Mathematical Tools and Some Useful Results

In this section, we present the tools we will use extensively all along the paper. In Section III-A, we recall well-known matrix results; in Section III-B, we present two fundamental properties of Gaussian models: The Integration by parts formula and Poincaré-Nash inequality for Gaussian vectors. Section III-C is devoted to a cornerstone approximation result which roughly states that $\mathbf{R}$ and $\widetilde{\mathbf{R}}$ can be replaced by $\mathbf{T}$ and $\widetilde{\mathbf{T}}$ up to some well-quantified error. In Section III-D, various variance estimates and approximation rules are stated.

A. General results

1) Some matrix inequalities: Let $\mathbf{A}$ and $\mathbf{B}$ be two $N \times N$ matrices with complex elements. Then

$$|\mathrm{Tr}\ (\mathbf{AB})| \leq \sqrt{\mathrm{Tr}\ (\mathbf{AA}^*)}\sqrt{\mathrm{Tr}\ (\mathbf{BB}^*)}\ . \tag{8}$$

Assuming $\mathbf{A}$ is Hermitian nonnegative, we have

$$|\mathrm{Tr}\ (\mathbf{AB})| \leq \|\mathbf{B}\|\ \mathrm{Tr}\ (\mathbf{A})\ , \tag{9}$$

where $\|.\|$ is the spectral norm (see [27]).

2) The Resolvent: The Resolvent matrix $\mathbf{H}_n(t)$ of matrix $\mathbf{Y}_n\mathbf{Y}_n^*$ is defined as $\mathbf{H}_n(t) = \left(\frac{t}{n}\mathbf{Y}_n\mathbf{Y}_n^* + \mathbf{I}_N\right)^{-1}$. It is of constant use in this paper and we give here some of its properties. The following identity, also known as the *Resolvent identity*:

$$\mathbf{H}_n(t) = \mathbf{I}_N - \frac{t}{n}\mathbf{H}_n(t)\mathbf{Y}\mathbf{Y}^* \tag{10}$$

follows from the mere definition of $\mathbf{H}_n$. Furthermore, the spectral norm of the resolvent is readily bounded by one:

$$\|\mathbf{H}_n(t)\| \leq 1 \quad \text{for} \quad t \geq 0\ . \tag{11}$$

3) Bounded character of the mean of some empirical moments: Let $(\mathbf{B}_n)_{n\in\mathbb{N}} = \mathrm{diag}\left(\left[b_1^{(n)}, \ldots, b_n^{(n)}\right]\right)$, $n \in \mathbb{N}$, be a sequence of deterministic $n \times n$ diagonal matrices. Assume **(A1)**, and furthermore, that $\sup_n \|\mathbf{B}_n\| < \infty$. Then for every integer k, we have

$$\frac{1}{n}\mathbb{E}\left[\mathrm{Tr}\ \left(\frac{1}{n}\mathbf{Y}_n\mathbf{B}_n\mathbf{Y}_n^*\right)^k\right] < K_k\ . \tag{12}$$

Let us sketch a proof. Expanding the left hand side of (12) yields:

$$\frac{1}{n^{k+1}} \sum_{\substack{i_1,i_2,\ldots,i_k=1:N \\ j_1,\ldots,j_k=1:n}} b_{j_1}b_{j_2}\cdots b_{j_k}\mathbb{E}\left[Y_{i_1j_1}\overline{Y_{i_2j_1}}Y_{i_2j_2}\overline{Y_{i_3j_2}}\cdots Y_{i_kj_k}\overline{Y_{i_1j_k}}\right]\ .$$

A close look at the argument of the $\mathbb{E}$ operator implies that due to the independence of the Y_{ij}, we only have $k+1$ degrees of freedom in the choice of the indices i_p and j_q. As all moments of the Gaussian law exist and moreover $\|\mathbf{B}_n\|$, $\|\mathbf{D}_n\|$, and $\|\widetilde{\mathbf{D}}_n\|$ are bounded, this sum is of order 1 as $n \to \infty$.

4) Differentiation formulas: Let $\mathbf{A}$ be a $N \times N$ complex matrix and let $\mathbf{Q}(\mathbf{A}) = (\mathbf{I}_N + \mathbf{A})^{-1}$. Let $\boldsymbol{\delta}\mathbf{A}$ be a perturbation of $\mathbf{A}$. Then

$$\mathbf{Q}(\mathbf{A} + \boldsymbol{\delta}\mathbf{A}) = \mathbf{Q}(\mathbf{A}) - \mathbf{Q}(\mathbf{A})\ \boldsymbol{\delta}\mathbf{A}\ \mathbf{Q}(\mathbf{A}) + o\left(\|\boldsymbol{\delta}\mathbf{A}\|\right), \tag{13}$$

where $o\left(\|\boldsymbol{\delta}\mathbf{A}\|\right)$ is negligible with respect to $\|\boldsymbol{\delta}\mathbf{A}\|$ in a neighborhood of 0. Writing $\mathbf{H}(t) = [H_{pq}(t)]_{p,q=1}^{N,N}$, we need the expression of the partial derivative $\partial H_{pq}/\partial Y_{ij}$. Using (13), we have:

$$\begin{aligned}\frac{\partial H_{pq}}{\partial Y_{ij}} &= -\frac{t}{n}\left[\mathbf{H}\frac{\partial \mathbf{Y}\mathbf{Y}^*}{\partial Y_{ij}}\mathbf{H}\right]_{pq} = -\frac{t}{n}\left[\mathbf{H}\left[\delta(k-i)\overline{Y_{\ell j}}\right]_{k,\ell=1}^{N}\mathbf{H}\right]_{pq} \\ &= -\frac{t}{n}H_{pi}\left[\mathbf{Y}^*\mathbf{H}\right]_{jq} = -\frac{t}{n}H_{pi}[\mathbf{y}_j^*\mathbf{H}]_q\ ,\end{aligned} \tag{14}$$

where δ is the Kronecker function. Similarly, we can establish

$$\frac{\partial H_{pq}}{\partial \overline{Y_{ij}}} = -\frac{t}{n}\left[\mathbf{H}\mathbf{Y}\right]_{pj} H_{iq} = -\frac{t}{n}[\mathbf{H}\mathbf{y}_j]_p H_{iq}\ . \tag{15}$$

The differential of $g(\mathbf{A}) = \log\det(\mathbf{A})$ is given by $g(\mathbf{A}+\boldsymbol{\delta}\mathbf{A}) = g(\mathbf{A}) + \mathrm{Tr}\ \left(\mathbf{A}^{-1}\ \boldsymbol{\delta}\mathbf{A}\right) + o\left(\|\boldsymbol{\delta}\mathbf{A}\|\right)$. We use this equation to derive the expression of $\partial I(t)/\partial\overline{Y_{ij}}$ also needed below:

$$\frac{\partial I}{\partial \overline{Y_{ij}}} = \frac{t}{n}\mathrm{Tr}\ \left(\mathbf{H}\frac{\partial \mathbf{Y}\mathbf{Y}^*}{\partial \overline{Y_{ij}}}\right) = \frac{t}{n}\mathrm{Tr}\ \left(\mathbf{H}\left[\delta(\ell-j)Y_{kj}\right]_{k,\ell=1}^{N}\right) = \frac{t}{n}\left[\mathbf{H}\mathbf{Y}\right]_{ij} = \frac{t}{n}\left[\mathbf{H}\mathbf{y}_j\right]_i\ . \tag{16}$$

B. Gaussian tools

1) An Integration by parts formula for Gaussian functionals: Let $\boldsymbol{\xi} = [\xi_1, \ldots, \xi_M]^T$ be a complex Gaussian random vector whose law is determined by $\mathbb{E}[\boldsymbol{\xi}] = \mathbf{0}$, $\mathbb{E}[\boldsymbol{\xi}\boldsymbol{\xi}^T] = \mathbf{0}$, and $\mathbb{E}[\boldsymbol{\xi}\boldsymbol{\xi}^*] = \boldsymbol{\Xi}$. Let $\Gamma = \Gamma(\xi_1, \cdots, \xi_M, \overline{\xi_1}, \cdots, \overline{\xi_M})$ be a $\mathcal{C}^1$ complex function polynomially bounded together with its derivatives, then:

$$\mathbb{E}\left[\xi_p \Gamma(\boldsymbol{\xi})\right] = \sum_{m=1}^{M}\left[\boldsymbol{\Xi}\right]_{pm}\mathbb{E}\left[\frac{\partial\Gamma(\boldsymbol{\xi})}{\partial\overline{\xi_m}}\right]\ . \tag{17}$$

This formula relies on an integration by parts and thus is referred to as the Integration by parts formula for Gaussian vectors. It is widely used in Mathematical Physics [28] and has been used in Random Matrix Theory in [18], [19].

2) Poincaré-Nash inequality: Let $\boldsymbol{\xi}$ and Γ be as previously and let $\nabla_z\Gamma = [\partial\Gamma/\partial z_1, \ldots, \partial\Gamma/\partial z_M]^T$ and $\nabla_{\overline{z}}\Gamma = [\partial\Gamma/\partial\overline{z_1}, \ldots, \partial\Gamma/\partial\overline{z_M}]^T$. Then the following inequality holds true:

$$\mathrm{var}\left(\Gamma(\boldsymbol{\xi})\right) \leq \mathbb{E}\left[\nabla_z\Gamma(\boldsymbol{\xi})^T\ \boldsymbol{\Xi}\ \overline{\nabla_z\Gamma(\boldsymbol{\xi})}\right] + \mathbb{E}\left[(\nabla_{\overline{z}}\Gamma(\boldsymbol{\xi}))^*\ \boldsymbol{\Xi}\ \nabla_{\overline{z}}\Gamma(\boldsymbol{\xi})\right]\ . \tag{18}$$

A proof of this inequality is available in [20] in the real case; see also [23].

When $\boldsymbol{\xi}$ is the vector of the stacked columns of matrix $\mathbf{Y}$, i.e. $\boldsymbol{\xi} = [Y_{11}, \ldots, Y_{Nn}]^T$, formula (17) becomes:

$$\mathbb{E}\left[Y_{ij}\Gamma(\mathbf{Y})\right] = d_i\tilde{d}_j\mathbb{E}\left[\frac{\partial\Gamma(\mathbf{Y})}{\partial\overline{Y_{ij}}}\right] \ , \tag{19}$$

while inequality (18) writes:

$$\mathrm{var}\left(\Gamma(\mathbf{Y})\right) \leq \sum_{i=1}^{N}\sum_{j=1}^{n} d_i\tilde{d}_j\mathbb{E}\left[\left|\frac{\partial\Gamma(\mathbf{Y})}{\partial Y_{i,j}}\right|^2 + \left|\frac{\partial\Gamma(\mathbf{Y})}{\partial\overline{Y_{i,j}}}\right|^2\right] \ . \tag{20}$$

Poincaré-Nash inequality turns out to be extremely useful to deal with variances of various quantities of interest related with random matrices. In order to give right away the flavour of such results, we state and prove the following:

Proposition 3: Assume that the setting of Theorem 1 holds and let $\mathbf{A}_n$ be a $N \times N$ real diagonal matrix which spectral norm is uniformly bounded in n. Then

$$\mathrm{var}\left(\frac{1}{n}\mathrm{Tr}\,\mathbf{A}_n\mathbf{H}_n\right) = \mathcal{O}\left(n^{-2}\right) \ .$$

Proof: We apply inequality (20) to the function $\Gamma(\mathbf{Y}) = \frac{1}{n}\mathrm{Tr}\,\mathbf{AH}$. Using (14), we have

$$\frac{\partial\Gamma}{\partial Y_{i,j}} = \frac{1}{n}\sum_{p=1}^{N} a_p\frac{\partial H_{pp}}{\partial Y_{i,j}} = -\frac{t}{n^2}[\mathbf{y}_j^*\mathbf{HAH}]_i \ .$$

Therefore, denoting by A the upper bound $A = \sup_n \|\mathbf{A}_n\|$ and noticing that $|\partial\Gamma/\partial Y_{i,j}| = \left|\partial\Gamma/\partial\overline{Y_{i,j}}\right|$, we have:

$$\begin{aligned}
\mathrm{var}\,\Gamma(\mathbf{Y}) &\leq \frac{2t^2}{n^4}\sum_{i=1}^{N}\sum_{j=1}^{n} d_i\tilde{d}_j\mathbb{E}\left|\left[\boldsymbol{y}_j^*\mathbf{HAH}\right]_i\right|^2 \\
&= \frac{2t^2}{n^4}\sum_{j=1}^{n}\tilde{d}_j\mathbb{E}\left(\mathbf{y}_j^*\mathbf{HAHDHAH}\mathbf{y}_j\right) \\
&= \frac{2t^2}{n^3}\mathbb{E}\,\mathrm{Tr}\left(\mathbf{HAHDHAH}\frac{\mathbf{Y}\widetilde{\mathbf{D}}\mathbf{Y}^*}{n}\right) \\
&\overset{(a)}{\leq} \frac{2t^2}{n^3}\mathbb{E}\left\{\|\mathbf{H}\|^4\|\mathbf{A}\|^2\|\mathbf{D}\|\ \mathrm{Tr}\left(\frac{\mathbf{Y}\widetilde{\mathbf{D}}\mathbf{Y}^*}{n}\right)\right\} \overset{(b)}{\leq} \frac{2A^2d_{\max}t^2}{n^3}\mathbb{E}\,\mathrm{Tr}\left(\frac{\mathbf{Y}\widetilde{\mathbf{D}}\mathbf{Y}^*}{n}\right) \overset{(c)}{\leq} \frac{K}{n^2} \ ,
\end{aligned}$$

where inequality (a) follows from (9), (b) follows from (11) and from the bounded character of $\|\mathbf{A}_n\|$ and $\|\mathbf{D}_n\|$, and (c) follows from (12). ■

C. Approximation rules

The following theorem is crucial in order to prove Theorems 1 and 2. Roughly speaking it allows to replace matrices $\mathbf{R}_n$ and $\widetilde{\mathbf{R}}_n$ by $\mathbf{T}_n$ and $\widetilde{\mathbf{T}}_n$ up to a well-quantified small error.

Theorem 3: Let $(\mathbf{A}_n)$ and $(\mathbf{B}_n)$ be two sequences of respectively $N \times N$ and $n \times n$ diagonal deterministic matrices whose spectral norm are uniformly bounded in n, then the following hold true:

$$\frac{1}{n}\mathrm{Tr}\,\mathbf{A}_n\mathbf{R}_n = \frac{1}{n}\mathrm{Tr}\,\mathbf{A}_n\mathbf{T}_n + \mathcal{O}\left(\frac{1}{n^2}\right), \tag{21}$$

$$\frac{1}{n}\mathrm{Tr}\,\mathbf{B}_n\widetilde{\mathbf{R}}_n = \frac{1}{n}\mathrm{Tr}\,\mathbf{B}_n\widetilde{\mathbf{T}}_n + \mathcal{O}\left(\frac{1}{n^2}\right). \tag{22}$$

Proof of Theorem 3 is postponed to Appendix C.

D. More variance estimates and more approximations rules

We collect here a few results which proofs rely on the Integration by parts formula (19), on Poincaré-Nash inequality, and on Theorem 3. The proofs of these results, although systematic, are somewhat lengthy and are therefore postponed to the Appendix. These results will be used extensively in Section V.

Proposition 4: In the setting of Theorem 1, let $(\mathbf{A})$ and $(\mathbf{B})$ be two sequences of respectively $N \times N$ and $n \times n$ diagonal deterministic matrices whose spectral norm are uniformly bounded in n. Consider the following functions:

$$\Phi(\mathbf{Y}) = \frac{1}{n}\mathrm{Tr}\left(\mathbf{AH}\frac{\mathbf{YBY}^*}{n}\right), \qquad \Psi(\mathbf{Y}) = \frac{1}{n}\mathrm{Tr}\left(\mathbf{AHDH}\frac{\mathbf{YBY}^*}{n}\right).$$

Then,

1) The following inequalities hold true:

$$\mathrm{var}\left(\Phi(\mathbf{Y})\right) = \mathcal{O}(n^{-2}), \qquad \mathrm{var}\left(\Psi(\mathbf{Y})\right) = \mathcal{O}(n^{-2}).$$

2) The following approximations hold true:

$$\mathbb{E}\left[\Phi(\mathbf{Y})\right] = \frac{1}{n}\mathrm{Tr}\left(\widetilde{\mathbf{D}}\widetilde{\mathbf{T}}\mathbf{B}\right)\frac{1}{n}\mathrm{Tr}\left(\mathbf{ADT}\right) + \mathcal{O}\left(n^{-2}\right), \tag{23}$$

$$\mathbb{E}\left[\Psi\left(\mathbf{Y}\right)\right] = \frac{1}{1-t^2\gamma\tilde{\gamma}}\left(\frac{1}{n^2}\mathrm{Tr}\left(\widetilde{\mathbf{D}}\widetilde{\mathbf{T}}\mathbf{B}\right)\mathrm{Tr}\left(\mathbf{AD}^2\mathbf{T}^2\right) - \frac{t\gamma}{n^2}\mathrm{Tr}\left(\widetilde{\mathbf{D}}^2\widetilde{\mathbf{T}}^2\mathbf{B}\right)\mathrm{Tr}\left(\mathbf{ADT}\right)\right) + \mathcal{O}\left(\frac{1}{n^2}\right) \tag{24}$$

The variance inequalities are proved in Appendix D; the approximation rules, in Appendix E.

IV. First Order Moment Approximation: Proof of Theorem 1

This section is devoted to the proof of the following approximation:

$$\mathbb{E}[I_n(\rho)] = V_n(\rho) + \mathcal{O}\left(n^{-1}\right), \tag{25}$$

where

$$V_n(\rho) = \log\det\left(\mathbf{I}_n + \rho\delta_n(\rho)\widetilde{\mathbf{D}}_n\right) + \log\det\left(\mathbf{I}_N + \rho\tilde{\delta}_n(\rho)\mathbf{D}_n\right) - n\rho\delta_n(\rho)\tilde{\delta}_n(\rho)\ . \tag{26}$$

This result already appears in [7] and is proved under greater generality in [9]. The proof presented here is new and relies on gaussian tools.

Outline of the proof

The proof is divided into three steps. We first make some preliminary remarks. Notice that the mutual information can be expressed as $I(\rho) = \int_0^\rho \mathrm{Tr}\,\left(n^{-1}\mathbf{H}(t)\mathbf{Y}\mathbf{Y}^*\right)dt$. In particular,

$$\mathbb{E}\left[I(\rho)\right] = \int_0^\rho \mathrm{Tr}\,\left(\mathbb{E}\left[\mathbf{H}(t)\frac{\mathbf{Y}\mathbf{Y}^*}{n}\right]\right)\,dt\ . \tag{27}$$

In order to study the asymptotic behaviour of $\mathbb{E}\left[I(\rho)\right]$, it is thus enough to study $\mathrm{Tr}\,\left(\mathbf{H}(t)\frac{\mathbf{Y}\mathbf{Y}^*}{n}\right)$ for $n \to +\infty$ up to an integration. The Resolvent identity (10) yields

$$\mathrm{Tr}\,\mathbb{E}\left(\mathbf{H}(t)\frac{\mathbf{Y}\mathbf{Y}^*}{n}\right) = \mathrm{Tr}\,\mathbb{E}\left(\frac{\mathbf{I}-\mathbf{H}(t)}{t}\right)\ .$$

We are therefore led to the study of $\mathbb{E}\left[\mathrm{Tr}\,(\mathbf{H}(t))\right]$. We now describe the three steps of the proof.

A. In the first part of the proof, we expand $\mathbb{E}\mathbf{H}(t)$ with the help of the Integration by parts formula (19). This derivations will bring to the fore the deterministic diagonal matrix $\mathbf{R}$, and Poincaré-Nash inequality will then allow us to obtain the following approximation:

$$\mathbb{E}\mathrm{Tr}\,\mathbf{A}\mathbf{H} = \mathrm{Tr}\,\mathbf{A}\mathbf{R} + \mathcal{O}\left(n^{-1}\right)\ ,$$

for every diagonal matrix $\mathbf{A}$ bounded in the spectral norm. Here are the main steps, gathered in an informal way. Differentiating the term $\mathbb{E}\left(\left[\mathbf{H}\mathbf{y}_j\right]_p\overline{Y_{pj}}\right)$, we obtain:

$$\mathbb{E}\left(\left[\mathbf{H}\mathbf{y}_j\right]_p\overline{Y_{p,j}}\right) = d_p\tilde{d}_j\mathbb{E}\left[H_{pp}\right] - t\tilde{d}_j\mathbb{E}\left(\frac{1}{n}\mathrm{Tr}\,(\mathbf{D}\mathbf{H})\ \left[\mathbf{H}\mathbf{y}_j\right]_p\overline{Y_{pj}}\right)\ ,$$

from which we will extract $\mathbb{E}[H_{pp}]$ later on. At this point, Poincaré-Nash inequality yields some decorrelation up to $\mathcal{O}\left(n^{-1}\right)$ and we obtain:

$$\mathbb{E}\left[\frac{1}{n}\mathrm{Tr}\,(\mathbf{D}\mathbf{H})\,(\mathbf{H}\mathbf{y}_j)_p\overline{Y_{pj}}\right] \simeq \mathbb{E}\left[\frac{1}{n}\mathrm{Tr}\,(\mathbf{D}\mathbf{H})\right]\mathbb{E}\left[\left[\mathbf{H}\mathbf{y}_j\right]_p\overline{Y_{pj}}\right] = \alpha\mathbb{E}\left[\left[\mathbf{H}\mathbf{y}_j\right]_p\overline{Y_{pj}}\right]\ .$$

This approximation allows us to isolate $\mathbb{E}\left(\left[\mathbf{H}\mathbf{y}_j\right]_p\overline{Y_{pj}}\right)$:

$$(1+t\tilde{d}_j\alpha)\mathbb{E}\left(\left[\mathbf{H}\mathbf{y}_j\right]_p\overline{Y_{p,j}}\right) \simeq d_p\tilde{d}_j\mathbb{E}\left[H_{pp}\right] \quad\Leftrightarrow\quad \mathbb{E}\left(\left[\mathbf{H}\mathbf{y}_j\right]_p\overline{Y_{p,j}}\right) \simeq d_p\tilde{d}_j\tilde{r}_j\mathbb{E}\left[H_{pp}\right]\ .$$

Now summing over j and using the Resolvent identity $\mathbb{E}H_{pp} = 1 - \frac{t}{n}\sum_{j=1}^{n}\mathbb{E}\left[\mathbf{H}\mathbf{y}_j\right]_p\overline{Y_{pj}}$ in the previous equation yields:

$$\frac{1-\mathbb{E}H_{pp}}{t} \simeq \tilde{\alpha} d_p \mathbb{E}H_{pp}, \quad \text{that is} \quad \mathbb{E}H_{pp} \simeq r_p \ .$$

All the technical details are provided in Section IV-A.

B. The second step follows from the approximation rule (21) stated in Section III-C, which immediatly yields

$$\mathbb{E}\mathrm{Tr}\,\mathbf{A}\mathbf{H} = \mathrm{Tr}\,\mathbf{A}\mathbf{T} + \mathcal{O}\left(n^{-1}\right) \ .$$

This in turn will imply that

$$\mathbb{E}\mathrm{Tr}\left(\mathbf{H}(t)\frac{\mathbf{Y}\mathbf{Y}^*}{n}\right) = \mathrm{Tr}\left(\frac{\mathbf{I}-\mathbb{E}\mathbf{H}}{t}\right) = \mathrm{Tr}\left(\frac{\mathbf{I}-\mathbf{T}}{t}\right) + \varepsilon_n(t) \overset{(a)}{=} n\delta(t)\tilde{\delta}(t) + \varepsilon_n(t).$$

where (a) follows from the fact that $\mathbf{I}-\mathbf{T} = t\tilde{\delta}\mathbf{D}(\mathbf{I}+t\tilde{\delta}\mathbf{D})^{-1}$.

C. In the third step, we integrate the previous equality:

$$\int_0^\rho \mathbb{E}\mathrm{Tr}\left(\mathbf{H}(t)\frac{\mathbf{Y}\mathbf{Y}^*}{n}\right)dt = n\int_0^\rho \delta(t)\tilde{\delta}(t)dt + \int_0^\rho \varepsilon_n(t)dt.$$

We identify $n\int_0^\rho \delta(t)\tilde{\delta}(t)dt$ with $V_n(\rho)$ as given by (26), and check that $\int_0^\rho \varepsilon_n(t)dt = \mathcal{O}(n^{-1})$.

A. Development of $\mathbb{E}\left(\mathrm{Tr}\,\mathbf{A}\mathbf{H}(t)\right)$ *and Approximation by* $\mathrm{Tr}\,\mathbf{A}\mathbf{R}(t)$

In order to study $\mathbb{E}\left(\mathrm{Tr}\,\mathbf{A}\mathbf{H}(t)\right)$, we first consider the diagonal entries $H_{pp}(t)$ of $\mathbf{H}(t)$. For each index j, we have

$$\mathbb{E}\left(\left[\mathbf{H}\mathbf{y}_j\right]_p\overline{Y_{p,j}}\right) = \sum_{i=1}^{N}\mathbb{E}\left(H_{pi}Y_{ij}\overline{Y_{pj}}\right) \ .$$

We now apply the Integration by parts formula (19) to the summand of the right hand side for function Γ defined as $\Gamma(\mathbf{Y}) = H_{pi}\overline{Y_{pj}}$. This yields:

$$\mathbb{E}\left(H_{pi}Y_{ij}\overline{Y_{pj}}\right) = d_i\tilde{d}_j\mathbb{E}\left[H_{ii}\right]\delta(i-p) - d_i\tilde{d}_j\frac{t}{n}\mathbb{E}\left(\left[\mathbf{H}\mathbf{y}_j\right]_p H_{ii}\overline{Y_{pj}}\right) \ . \tag{28}$$

Therefore,

$$\mathbb{E}\left(\left[\mathbf{H}\mathbf{y}_j\right]_p\overline{Y_{p,j}}\right) = d_p\tilde{d}_j\mathbb{E}\left[H_{pp}\right] - t\tilde{d}_j\mathbb{E}\left(\frac{1}{n}\mathrm{Tr}\left(\mathbf{D}\mathbf{H}\right)\left[\mathbf{H}\mathbf{y}_j\right]_p\overline{Y_{pj}}\right) \ , \tag{29}$$

from which we extract $\mathbb{E}[H_{pp}]$ later on. Recall at this point that $\mathrm{var}\left(n^{-1}\mathrm{Tr}\,\mathbf{DH}(t)\right) = \mathcal{O}\left(n^{-2}\right)$ by Proposition 3. Recall also the following notations: $\beta = n^{-1}\mathrm{Tr}\left(\mathbf{DH}\right)$, $\alpha = \mathbb{E}\left[\beta\right]$, and $\overset{\circ}{\beta} = \beta - \alpha$. Plugging the relation $\beta = \alpha + \overset{\circ}{\beta}$ into (29), we get

$$\mathbb{E}\left[\left[\mathbf{H}\mathbf{y}_j\right]_p \overline{Y_{p,j}}\right] = d_p\tilde{d}_j\mathbb{E}[H_{pp}] - t\tilde{d}_j\alpha\mathbb{E}\left[\left[\mathbf{H}\mathbf{y}_j\right]_p \overline{Y_{p,j}}\right] - t\tilde{d}_j\mathbb{E}\left[\overset{\circ}{\beta}\left[\mathbf{H}\mathbf{y}_j\right]_p\overline{Y_{pj}}\right] \ . \tag{30}$$

Solving this equation w.r.t. $\mathbb{E}\left[\left[\mathbf{H}\mathbf{y}_j\right]_p\overline{Y_{p,j}}\right]$ provides:

$$\mathbb{E}\left[\left[\mathbf{H}\mathbf{y}_j\right]_p\overline{Y_{p,j}}\right] = d_p\tilde{d}_j\tilde{r}_j\mathbb{E}[H_{pp}] - t\tilde{d}_j\tilde{r}_j\mathbb{E}\left[\overset{\circ}{\beta}\left[\mathbf{H}\mathbf{y}_j\right]_p\overline{Y_{pj}}\right] \quad \text{where} \quad \tilde{r}_j(t) = \frac{1}{1+t\alpha(t)\tilde{d}_j} \quad \text{for } 1 \leq j \leq n \ . \tag{31}$$

Summing (31) over j yields:

$$\mathbb{E}\left[\mathbf{H}\frac{\mathbf{Y}\mathbf{Y}^*}{n}\right]_{pp} = \tilde{\alpha}d_p\mathbb{E}[H_{pp}] - t\mathbb{E}\overset{\circ}{\beta}\left[\mathbf{H}\frac{\mathbf{Y}\tilde{\mathbf{D}}\tilde{\mathbf{R}}\mathbf{Y}^*}{n}\right]_{pp} \ , \tag{32}$$

where $\widetilde{\mathbf{R}}$ is the diagonal matrix $\mathrm{diag}\left(\tilde{r}_j(t)\right) = \left(\mathbf{I} + \alpha t\widetilde{\mathbf{D}}\right)^{-1}$ and $\tilde{\alpha} = \frac{1}{n}\mathrm{Tr}\,\tilde{\mathbf{D}}\tilde{\mathbf{R}}$. In order to obtain an expression for $\mathbb{E}[H_{pp}]$, we plug the identity (32) into the Resolvent identity:

$$\mathbb{E}[H_{pp}] = 1 - t\mathbb{E}\left[\left[\mathbf{H}\frac{\mathbf{Y}\mathbf{Y}^*}{n}\right]_{pp}\right]$$

and obtain:

$$\mathbb{E}\left[H_{pp}\right] = r_p + t^2r_p\mathbb{E}\left[\overset{\circ}{\beta}\left[\mathbf{H}\frac{\mathbf{Y}\tilde{\mathbf{D}}\tilde{\mathbf{R}}\mathbf{Y}^*}{n}\right]_{pp}\right] \tag{33}$$

with $r_p(t) = \left(1 + t\tilde{\alpha}d_p\right)^{-1}$. Let $\mathbf{A}$ be a $N \times N$ diagonal matrix with bounded spectral norm. Multiplying (33) by $\mathbf{A}$'s components and summing over p yields:

$$\mathbb{E}\mathrm{Tr}\left(\mathbf{AH}\right) = \mathrm{Tr}\left(\mathbf{AR}\right) + nt^2\mathbb{E}\left[\overset{\circ}{\beta}\,\Phi(\mathbf{Y})\right] \ ,$$

where $\Phi(\mathbf{Y}) = \frac{1}{n}\mathrm{Tr}\left(\mathbf{ARH}\frac{\mathbf{Y}\tilde{\mathbf{D}}\tilde{\mathbf{R}}\mathbf{Y}^*}{n}\right)$. As $\overset{\circ}{\beta}$ is zero-mean, $\mathbb{E}[\overset{\circ}{\beta}\,\Phi] = \mathbb{E}[\overset{\circ}{\beta}\,\overset{\circ}{\Phi}]$. In particular, Cauchy-Schwarz inequality yields:

$$|\mathbb{E}\overset{\circ}{\beta}\,\overset{\circ}{\Phi}| \leq \sqrt{\mathrm{var}(\beta)}\sqrt{\mathrm{var}(\Phi)}.$$

Recall that $\mathrm{var}(\beta) = \mathcal{O}\left(n^{-2}\right)$ by Prop. 3. Since $\|\mathbf{R}\|$ and $\|\widetilde{\mathbf{D}}\widetilde{\mathbf{R}}\|$ are both bounded by Assumption **(A1)** and by the definitions of $\mathbf{R}$ and $\widetilde{\mathbf{R}}$, one can directly apply the result of Proposition 4 to Φ in order to get $\mathrm{var}(\Phi) = \mathcal{O}\left(n^{-2}\right)$.

We have therefore proved the following:

Proposition 5: In the setting of Theorem 1, let $\mathbf{A}_n$ be a uniformly bounded diagonal $N \times N$ matrix. Then for every $t \in \mathbb{R}^+$,

$$\mathbb{E}(\mathrm{Tr}\, \mathbf{A}_n \mathbf{H}_n(t)) = \mathrm{Tr}\, \mathbf{A}_n \mathbf{R}_n(t) + \mathcal{O}\left(n^{-1}\right) \ . \tag{34}$$

B. The Deterministic Approximation $\mathbf{T}_n(t)$.

Proposition 5 provides a deterministic equivalent to $\mathbb{E}\left(\mathrm{Tr}\, \mathbf{A}_n \mathbf{H}_n\right)$ since the matrix $\mathbf{R}_n$ is deterministic; however its elements still depend on $\tilde{\alpha} = n^{-1}\mathrm{Tr}\left(\tilde{\mathbf{D}}_n \tilde{\mathbf{R}}_n\right)$, which itself depends on $\alpha = \mathbb{E}\left(n^{-1}\mathrm{Tr}\, \mathbf{D}_n \mathbf{H}_n\right)$, an unknown parameter. The next step is therefore to apply Theorem 3 to approximate the matrix $\mathbf{R}_n$ by $\mathbf{T}_n$, which only depends on $\mathbf{D}_n$ and $\widetilde{\mathbf{D}}_n$ and on δ and $\tilde{\delta}$, the solutions of (2). Theorem 3 together with Equation (34) imply that:

$$\mathbb{E}(\mathrm{Tr}\, \mathbf{A}_n \mathbf{H}_n) = \mathrm{Tr}\left(\mathbf{A}_n \mathbf{T}_n\right) + \mathcal{O}\left(n^{-1}\right) \ . \tag{35}$$

Since $\mathbf{T}_n$ only depends on δ and $\tilde{\delta}$, (35) provides a deterministic equivalent of $\mathbb{E}(\mathrm{Tr}\, \mathbf{A}_n \mathbf{H}_n)$ in terms of δ and $\tilde{\delta}$. Note that taking $\mathbf{A} = \mathbf{D}$ yields in particular $\alpha = \delta + \mathcal{O}(n^{-2})$ while a direct application of Theorem 3 for $\widetilde{\mathbf{A}}_n = \widetilde{\mathbf{D}}_n$ yields $\tilde{\alpha} = \tilde{\delta} + \mathcal{O}(n^{-2})$.

We are now in a position to describe the behaviour of $\mathbb{E}\,\mathrm{Tr}\left(\mathbf{H}_n(t)\frac{\mathbf{Y}_n\mathbf{Y}_n^*}{n}\right)$ by using the Resolvent identity. From (10) and (35), taking $\mathbf{A}_n = \mathbf{I}_n$, we immediately obtain:

$$\mathbb{E}\,\mathrm{Tr}\left(\mathbf{H}_n(t)\frac{\mathbf{Y}_n\mathbf{Y}_n^*}{n}\right) = \frac{1}{t}\mathrm{Tr}\left(\mathbf{I}_n - \mathbf{T}_n(t)\right) + \mathcal{O}\left(n^{-1}\right) \ .$$

As $\mathbf{I}_n - \mathbf{T}_n(t) = (\mathbf{T}_n(t)^{-1} - \mathbf{I}_n)\mathbf{T}_n(t) = t\tilde{\delta}(t)\mathbf{D}_n\mathbf{T}_n(t)$, we eventually get that

$$\mathbb{E}\left[\mathrm{Tr}\left(\mathbf{H}_n(t)\frac{\mathbf{Y}_n\mathbf{Y}_n^*}{n}\right)\right] = n\delta(t)\tilde{\delta}(t) + \varepsilon_n(t), \tag{36}$$

where the error $\varepsilon_n(t)$ is a $\mathcal{O}(n^{-1})$ term.

C. Recovering the Deterministic Approximation $V(\rho)$ *of* $\mathbb{E}[I(\rho)]$.

As mentionned previously, $\varepsilon_n(t)$ is a $\mathcal{O}(n^{-1})$ term, i.e. $|\varepsilon_n(t)| \leq K_t\, n^{-1}$. One can easily keep track of K_t in the derivations that lead to (36) and prove that K_t is bounded on the compact interval $[0, \rho]$. In particular, $|\varepsilon_n(t)| < Kn^{-1}$ on the compact interval $[0, \rho]$ for some $K > 0$. The proof of this fact is omitted.

As $n\varepsilon_n(t)$ is uniformly bounded on $[0, \rho]$, we have $\left|\int_0^\rho \varepsilon_n(t)dt\right| = \mathcal{O}(n^{-1})$. Therefore,

$$\mathbb{E}[I(\rho)] = \int_0^\rho n\delta(t)\tilde{\delta}(t) + \mathcal{O}\left(n^{-1}\right) \ .$$

Consider now

$$V(\rho) = W\left(\rho, \delta(\rho), \tilde{\delta}(\rho)\right),$$

where function $W(\rho, \delta, \tilde{\delta})$ is defined by

$$W\left(\rho, \delta, \tilde{\delta}\right) = \log\det\left(\mathbf{I}_n + \rho\delta\tilde{\mathbf{D}}_n\right) + \log\det\left(\mathbf{I}_N + \rho\tilde{\delta}\mathbf{D}_n\right) - n\rho\delta\tilde{\delta}\ .$$

One can easily check that:

$$\frac{\partial W}{\partial \delta} = \rho\left(\mathrm{Tr}\,\left(\widetilde{\mathbf{D}}_n(\mathbf{I}_n + \rho\delta\widetilde{\mathbf{D}}_n)^{-1}\right) - n\tilde{\delta}\right) \qquad \text{and} \qquad \frac{\partial W}{\partial \tilde{\delta}} = \rho\left(\mathrm{Tr}\,\left(\mathbf{D}_n(\mathbf{I}_N + \rho\tilde{\delta}\mathbf{D}_n)^{-1}\right) - n\delta\right)\ .$$

As the pair $(\delta(\rho), \tilde{\delta}(\rho))$ satisfies (2), the above partial derivatives evaluated at point $(\rho, \delta(\rho), \tilde{\delta}(\rho))$ are zero. Therefore,

$$\frac{dV}{d\rho} = \left(\frac{\partial W}{\partial \rho}\right)_{(\rho,\delta(\rho),\tilde{\delta}(\rho))} = n\delta(\rho)\tilde{\delta}(\rho) \tag{37}$$

which in turn implies (7). Theorem 1 is proved.

Remark 1 (On the deterministic approximation $\mathbf{T}_n$): The deterministic approximation $\mathbf{T}_n$ can be used to approximate functionals of the eigenvalues of $\mathbf{Y}_n\mathbf{Y}_n^*$ other that the mutual information $\log\det(\rho n^{-1}\mathbf{Y}_n\mathbf{Y}_n^* + \mathbf{I}_N)$ (see for instance [9]). This relies on a specific representation of $\mathbf{T}_n$: The spectral theorem for Hermitian matrices yields the integral representation:

$$\frac{1}{n}\mathrm{Tr}\,\mathbf{H}_n(z) = \int_0^\infty \frac{N_n(d\lambda)}{1+\lambda z}, \quad z \in \mathbb{C}\setminus\mathbb{R}_-\ ,$$

where N_n represents the empirical distribution of the eigenvalues of $\mathbf{Y}_n\mathbf{Y}_n^*$. It can be shown that $n^{-1}\mathrm{Tr}\,\mathbf{T}_n$ admits a similar representation:

$$\frac{1}{n}\mathrm{Tr}\,\mathbf{T}_n(z) = \int_0^\infty \frac{\pi_n(d\lambda)}{1+\lambda z}, \quad z \in \mathbb{C}\setminus\mathbb{R}_-\ ,$$

where π_n is a probability measure. Finally, one can prove that $\int_0^\infty f(\lambda)N_n(d\lambda) - \int_0^\infty f(\lambda)\pi_n(d\lambda)$ converges to zero almost surely for every continuous bounded function (see [9] for details).

V. Second order Analysis: Proof of Theorem 2

This section is devoted to the proof of the Central Limit Theorem:

$$\sigma_n^{-1}(\rho)\left(I_n(\rho) - V_n(\rho)\right) \xrightarrow[n,N\to\infty]{\mathcal{L}} \mathcal{N}(0,1)\ . \tag{38}$$

Denote by $\psi_n(u,\rho) = \mathbb{E}\left[e^{\mathbf{i}u(I_n(\rho)-V_n(\rho))}\right]$ the characteristic function of $I_n(\rho) - V_n(\rho)$. We first reduce the problem in the following way:

Proposition 6: Assume that for every $u \in \mathbb{R}$,

$$h_n(u) = \psi_n(u,\rho) - e^{-u^2\sigma_n^2(\rho)/2} \xrightarrow[n,N\to\infty]{} 0 \ , \tag{39}$$

then (38) holds true.

Proof: Let Z be a real $\mathcal{N}(0,1)$ random variable. If (39) holds true, then for every bounded continuous function f,

$$\mathbb{E}f(I_n(\rho) - V_n(\rho)) - \mathbb{E}f(\sigma_n(\rho)Z) \xrightarrow[n\to\infty]{} 0. \tag{40}$$

Assume not, then is there exists some bounded continuous function f^* and an extracted subsequence $(I_{(n)} - V_{(n)}, \sigma_{(n)})$ such that

$$\mathbb{E}f^*(I_{(n)} - V_{(n)}) - \mathbb{E}f^*(\sigma_{(n)}Z) \xrightarrow[n\to\infty]{} a \neq 0 \ . \tag{41}$$

Due to Proposition 2, the family of probability distributions of $\sigma_{(n)}Z$ is tight and one can extract a new converging subsequence, say $\sigma_{(m)}Z$ such that $\sigma_{(m)}Z \xrightarrow{\mathcal{L}} \sigma^*Z$. But in this case, $I_{(m)} - V_{(m)} \xrightarrow{\mathcal{L}} \sigma^*Z$ by (39), which in turn implies that

$$\mathbb{E}f(I_{(m)} - V_{(m)}) - \mathbb{E}f(\sigma_{(m)}Z) \quad = \quad \big(\mathbb{E}f(I_{(m)} - V_{(m)}) - \mathbb{E}f(\sigma^*Z)\big) + \big(\mathbb{E}f(\sigma^*Z) - \mathbb{E}f(\sigma_{(m)}Z)\big)$$

goes to zero; and this contradicts (41). Convergence (40) is proved.

Recall now the following distance between probability distributions:

$$\beta(\mathbb{P},\tilde{\mathbb{P}}) = \sup\left\{\int f \ d\,\mathbb{P} - \int f \ d\,\tilde{\mathbb{P}},\ \|f\|_{BL} \leq 1\right\},$$

where $\|f\|_{BL} = \|f\|_{Lip} + \|f\|_\infty$, $\|\cdot\|_{Lip}$ being the Lipschitz norm and $\|\cdot\|_\infty$, the supremum norm. By using the tightness of $(\sigma_n(\rho)Z)$ and mimicking [29, Theorem 11.3.3], one can prove that

$$\beta\left(\mathcal{L}(I_n - V_n)\,,\,\mathcal{L}(\sigma_n Z)\,\right) \xrightarrow[n\to\infty]{} 0 \ , \tag{42}$$

where $\mathcal{L}(X)$ stands for the probability distribution of X. Let us now prove that

$$\beta\left(\mathcal{L}\left(\frac{I_n - V_n}{\sigma_n}\right)\,,\,\mathcal{L}(Z)\right) \xrightarrow[n\to\infty]{} 0 \ .$$

This will prove (38) by [29, Theorem 11.3.3]. Let f be such that $\|f\|_{BL} \leq 1$.

$$\mathbb{E}f\left(\frac{I_n - V_n}{\sigma_n}\right) - \mathbb{E}f(Z) = \frac{2M_t}{m_t}\left(\mathbb{E}g_n(I_n - V_n) - \mathbb{E}g_n(\sigma_n Z)\right)$$

where M_t, m_t are defined in Proposition 2 and $g_n(x) = \frac{m_t}{2M_t} f\left(\frac{x}{\sigma_n}\right)$. One can readily check that $\|g_n\|_{BL} \leq 1$, thus:

$$\mathbb{E}f\left(\frac{I_n - V_n}{\sigma_n}\right) - \mathbb{E}f(Z) \leq \frac{2M_t}{m_t}\beta\left(\mathcal{L}(I_n - V_n)\,,\,\mathcal{L}(\sigma_n Z)\,\right) \ .$$

It remains to take the supremum over the f's such that $\|f\|_{BL} \leq 1$ to get:

$$\beta\left(\mathcal{L}\left(\frac{I_n - V_n}{\sigma_n}\right),\, \mathcal{L}(Z)\right) \leq \frac{2M_t}{m_t}\beta\left(\mathcal{L}(I_n - V_n),\, \mathcal{L}(\sigma_n Z)\right) \xrightarrow[n\to\infty]{(a)} 0\ ,$$

where (a) follows from (42). Proposition 6 is proved. ■

Outline of the proof of (39).

The proof of the convergence of $h_n(u)$ towards zero is divided into two steps.

A. We first differentiate $\psi_n(u,t)$ with respect to t in order to obtain a differential equation of the form:

$$\frac{\partial \psi_n(u,t)}{\partial t} = -\frac{u^2}{2}\eta_n(t)\psi_n(u,t) + \varepsilon_n(u,t)\ . \tag{43}$$

In order to obtain the differential equation (43), we first develop $\partial\psi/\partial t$ with the help of the Integration by parts formula (19). We then use Poincaré-Nash inequality to prove that relevant variances are of order $\mathcal{O}(n^{-2})$. This will enable us to decorrelate various expectations, i.e. to express them as products of expectations up to negligible terms. We shall then use the approximation rules stated in Proposition 4 in Section III-D to deal with the obtained expectations.

B. The second step is devoted to identify the variance, that is to prove the identity

$$\int_0^\rho \eta_n(t)\, dt = \sigma_n^2(\rho),$$

where σ_n^2 is given by (5), i.e. $\sigma_n^2(\rho) = -\log(1 - \rho^2\gamma_n(\rho)\tilde{\gamma}_n(\rho))$.

C. The third step is devoted to the integration of (43). Instead of directly integrating (43), we introduce $K_n(u,\rho) = \psi_n(u,\rho)e^{\frac{u^2}{2}\sigma_n^2(\rho)}$ which satisfies the following differential equation:

$$\frac{\partial K_n(u,t)}{\partial t} = \varepsilon_n(u,t)e^{\frac{u^2}{2}\sigma_n^2(t)}\ . \tag{44}$$

Taking into account the obvious facts that $\psi_n(u,0) = 1$, $\sigma_n^2(0) = 0$ and therefore that $K_n(u,0) = 1$, we shall obtain that

$$K_n(u,\rho) = 1 + \int_0^\rho \varepsilon_n(u,t)e^{\frac{u^2}{2}\sigma_n^2(t)}\, dt\ ,$$

and prove that $\int_0^\rho \varepsilon_n(u,t)e^{\frac{u^2}{2}\sigma_n^2(t)}\, dt = \mathcal{O}(n^{-1})$. This will yield in turn that:

$$\psi_n(u,\rho) \quad = \quad \left(1 + \mathcal{O}(n^{-1})\right)e^{-\frac{u^2}{2}\sigma_n^2(\rho)} \quad \overset{(a)}{=} \quad e^{-\frac{u^2}{2}\sigma_n^2(\rho)} + \mathcal{O}(n^{-1})\ .$$

where (a) follows from Proposition 2.

The theorem will then be proved.

A. The differential equation $\partial_t \psi_n = -\frac{u^2}{2}\eta_n \psi_n + \varepsilon_n$

Recall that $\psi_n(u,\rho) = \varphi_n(u,\rho)e^{-\mathbf{i}uV_n(\rho)}$ where $\varphi(u,t) = \mathbb{E}\left(e^{\mathbf{i}uI(t)}\right)$. As $V_n'(t) = n\delta(t)\tilde{\delta}(t)$ by (37), we obtain:

$$\frac{\partial \psi(u,t)}{\partial t} = e^{-\mathbf{i}uV(t)}\frac{\partial \varphi(u,t)}{\partial t} - \mathbf{i}un\delta(t)\tilde{\delta}(t)\psi(u,t) \ . \tag{45}$$

Since $I'(t) = n^{-1}\mathrm{Tr}\,\mathbf{H}(t)\mathbf{Y}\mathbf{Y}^*$ by (27), we have:

$$\frac{\partial \varphi(u,t)}{\partial t} = \mathbf{i}u\ \mathbb{E}\left[\mathrm{Tr}\ \left(\mathbf{H}(t)\frac{\mathbf{Y}\mathbf{Y}^*}{n}\right)e^{\mathbf{i}uI(t)}\right] = \frac{\mathbf{i}u}{n}\sum_{p,i=1}^{N}\sum_{j=1}^{n}\mathbb{E}\left[Y_{ij}H_{pi}\overline{Y_{pj}}e^{\mathbf{i}uI}\right] \ . \tag{46}$$

Applying the Integration by parts formula (19) to $\mathbb{E}\left[Y_{ij}H_{pi}\overline{Y_{pj}}e^{\mathbf{i}uI}\right]$ (which can be written $\mathbb{E}\left(Y_{ij}\Gamma(\mathbf{Y})\right)$ for $\Gamma(\mathbf{Y}) = H_{pi}\overline{Y_{pj}}e^{\mathbf{i}uI}$) and using the differentiation formulas (15) and (16) yields:

$$\begin{aligned}
\mathbb{E}\left[Y_{ij}H_{pi}\overline{Y_{pj}}e^{\mathbf{i}uI}\right] &= d_i\tilde{d}_j\mathbb{E}\left[\frac{\partial}{\partial\overline{Y_{ij}}}\left(H_{pi}\overline{Y_{pj}}e^{\mathbf{i}uI}\right)\right], \\
&= -\frac{t}{n}d_i\tilde{d}_j\mathbb{E}\left[\left[\mathbf{H}\mathbf{y}_j\right]_p H_{ii}\overline{Y_{pj}}e^{\mathbf{i}uI}\right] + d_i\tilde{d}_j\delta(i-p)\mathbb{E}\left[H_{pi}e^{\mathbf{i}uI}\right] \\
&\quad + \frac{\mathbf{i}ut}{n}d_i\tilde{d}_j\mathbb{E}\left[H_{pi}\overline{Y_{pj}}\left[\mathbf{H}\mathbf{y}_j\right]_i e^{\mathbf{i}uI}\right] \ .
\end{aligned} \tag{47}$$

We now sum over index i and obtain:

$$\begin{aligned}
\mathbb{E}\left[\left[\mathbf{H}\mathbf{y}_j\right]_p\overline{Y_{pj}}e^{\mathbf{i}uI}\right] &= -t\tilde{d}_j\mathbb{E}\left[\beta\ \left[\mathbf{H}\mathbf{y}_j\right]_p\overline{Y_{pj}}e^{\mathbf{i}uI}\right] + d_p\tilde{d}_j\mathbb{E}\left[H_{pp}e^{\mathbf{i}uI}\right] \\
&\quad + \frac{\mathbf{i}ut}{n}\tilde{d}_j\mathbb{E}\left[\left[\mathbf{H}\mathbf{D}\mathbf{H}\mathbf{y}_j\right]_p\overline{Y_{pj}}e^{\mathbf{i}uI}\right] \ ,
\end{aligned}$$

where $\beta = n^{-1}\mathrm{Tr}\,\mathbf{D}\mathbf{H}$. Writing $\beta = \overset{\circ}{\beta} + \alpha$ yields:

$$\begin{aligned}
(1+t\alpha\tilde{d}_j)\mathbb{E}\left[\left[\mathbf{H}\mathbf{y}_j\right]_p\overline{Y_{pj}}e^{\mathbf{i}uI}\right] &= -t\tilde{d}_j\mathbb{E}\left[\overset{\circ}{\beta}\ \left[\mathbf{H}\mathbf{y}_j\right]_p\overline{Y_{pj}}e^{\mathbf{i}uI}\right] + d_p\tilde{d}_j\mathbb{E}\left[H_{pp}e^{\mathbf{i}uI}\right] \\
&\quad + \frac{\mathbf{i}ut}{n}\tilde{d}_j\mathbb{E}\left[\left[\mathbf{H}\mathbf{D}\mathbf{H}\mathbf{y}_j\right]_p\overline{Y_{pj}}e^{\mathbf{i}uI}\right] \ .
\end{aligned} \tag{48}$$

We now take into account that $\tilde{r}_j(t) = (1+t\alpha\tilde{d}_j)^{-1}$ and sum over j:

$$\begin{aligned}
\mathbb{E}\left[\left[\mathbf{H}\mathbf{Y}\mathbf{Y}^*\right]_{pp}e^{\mathbf{i}uI}\right] &= -t\mathbb{E}\left[\overset{\circ}{\beta}\ \left[\mathbf{H}\mathbf{Y}\widetilde{\mathbf{D}}\widetilde{\mathbf{R}}\mathbf{Y}^*\right]_{pp}e^{\mathbf{i}uI}\right] + n\tilde{\alpha}d_p\mathbb{E}\left[H_{pp}e^{\mathbf{i}uI}\right] \\
&\quad + \frac{\mathbf{i}ut}{n}\mathbb{E}\left[\left[\mathbf{H}\mathbf{D}\mathbf{H}\mathbf{Y}\widetilde{\mathbf{D}}\widetilde{\mathbf{R}}\mathbf{Y}^*\right]_{pp}e^{\mathbf{i}uI}\right] \ .
\end{aligned} \tag{49}$$

By the Resolvent identity (10), $\mathbb{E}\left[H_{pp}e^{\mathbf{i}uI}\right] = \mathbb{E}\left[e^{\mathbf{i}uI}\right] - \frac{t}{n}\mathbb{E}\left[\left[\mathbf{H}\mathbf{Y}\mathbf{Y}^*\right]_{pp}e^{\mathbf{i}uI}\right]$. Replace now in (49),

recall that $r_p(t) = (1 + t\tilde{\alpha}(t)d_p)^{-1}$ and sum over p to obtain:

$$\begin{aligned}\mathbb{E}\left[\mathrm{Tr}\,\left(\mathbf{H}\frac{\mathbf{Y}\mathbf{Y}^*}{n}\right)e^{\mathbf{i}uI}\right] &= \mathrm{Tr}\,(\mathbf{D}\mathbf{R})\,\tilde{\alpha}\mathbb{E}\left[e^{\mathbf{i}uI}\right] \\ &\quad + \mathbf{i}ut\mathbb{E}\left[\frac{1}{n}\mathrm{Tr}\,\left(\mathbf{R}\mathbf{H}\mathbf{D}\mathbf{H}\frac{\mathbf{Y}\widetilde{\mathbf{D}}\widetilde{\mathbf{R}}\mathbf{Y}^*}{n}\right)e^{\mathbf{i}uI}\right] \\ &\quad - t\,\mathbb{E}\left[\overset{\circ}{\beta}\,\mathrm{Tr}\,\left(\mathbf{R}\mathbf{H}\frac{\mathbf{Y}\widetilde{\mathbf{D}}\widetilde{\mathbf{R}}\mathbf{Y}^*}{n}\right)e^{\mathbf{i}uI}\right] \\ &\triangleq \chi_1 + \chi_2 + \chi_3\ . \end{aligned} \tag{50}$$

Thanks to Theorem 3,

$$\chi_1 = \mathrm{Tr}\,(\mathbf{D}\mathbf{R})\,\tilde{\alpha}\mathbb{E}\left[e^{\mathbf{i}uI}\right] = \mathrm{Tr}\,(\mathbf{D}\mathbf{T})\,\tilde{\alpha}\mathbb{E}\left[e^{\mathbf{i}uI}\right] + \mathcal{O}(n^{-1}) = n\delta\tilde{\delta}\mathbb{E}\left[e^{\mathbf{i}uI}\right] + \mathcal{O}(n^{-1}). \tag{51}$$

In order to deal with χ_2, we apply the results of Proposition 4 related to $\Psi(\mathbf{Y})$ in the particular case where $\mathbf{A} = \mathbf{R}$ and $\mathbf{B} = \widetilde{\mathbf{D}}\widetilde{\mathbf{R}}$. In this case, χ_2 writes $\chi_2 = \mathbf{i}ut\mathbb{E}\left(\Psi(\mathbf{Y})e^{\mathbf{i}uI}\right)$, and Cauchy-Schwarz inequality yields:

$$\left|\mathbb{E}\left(\Psi e^{\mathbf{i}uI}\right) - \mathbb{E}\left(e^{\mathbf{i}uI}\right)\mathbb{E}\left(\Psi\right)\right| = \left|\mathbb{E}[e^{\mathbf{i}uI}\ \overset{\circ}{\Psi}]\right| \leq \sqrt{\mathbb{E}\left[\left|\overset{\circ}{\Psi}(\mathbf{Y})\right|^2\right]} = \mathcal{O}(n^{-1})\ .$$

Therefore,

$$\mathbb{E}\left(\Psi e^{\mathbf{i}uI}\right) = \mathbb{E}\left(e^{\mathbf{i}uI}\right)\mathbb{E}\left(\Psi\right) + \mathcal{O}(n^{-1})\ .$$

We now use the approximation for $\mathbb{E}\Psi(\mathbf{Y})$ given in Proposition 4. By Theorem 3, we can replace $\tilde{\mathbf{R}}$ (resp. $\widetilde{\mathbf{R}}$ by $\tilde{\mathbf{T}}$ (resp. $\widetilde{\mathbf{T}}$) in the obtained expression. We therefore obtain:

$$\begin{aligned}\mathbb{E}\left(\Psi(\mathbf{Y})e^{\mathbf{i}uI}\right) &= \mathbb{E}\Psi(\mathbf{Y})\mathbb{E}\left[e^{\mathbf{i}uI}\right] + \mathcal{O}\left(n^{-1}\right) \\ &= \frac{1}{1-t^2\gamma\tilde{\gamma}}\left(\tilde{\gamma}\frac{1}{n}\mathrm{Tr}\,\left(\mathbf{D}^2\mathbf{T}^3\right) - t\gamma\frac{1}{n}\mathrm{Tr}\,\left(\widetilde{\mathbf{D}}^3\widetilde{\mathbf{T}}^3\right)\frac{1}{n}\mathrm{Tr}\,\left(\mathbf{D}\mathbf{T}^2\right)\right)\mathbb{E}\left[e^{\mathbf{i}uI}\right] + \mathcal{O}\left(n^{-1}\right)\end{aligned} \tag{52}$$

The term χ_3 can be handled similarly: We apply the results of Proposition 4 related to $\Phi(\mathbf{Y})$ in the particular case where $\mathbf{A} = \mathbf{R}$ and $\mathbf{B} = \widetilde{\mathbf{D}}\widetilde{\mathbf{R}}$. In this case, χ_3 writes $\chi_2 = -tn\mathbb{E}\left(\overset{\circ}{\beta}\Phi(\mathbf{Y})e^{\mathbf{i}uI}\right)$, and Cauchy-Schwarz inequality yields:

$$\left|\mathbb{E}\left(\overset{\circ}{\beta}\Phi e^{\mathbf{i}uI}\right) - \mathbb{E}\left(\overset{\circ}{\beta}e^{\mathbf{i}uI}\right)\mathbb{E}\left(\Phi\right)\right| = \left|\mathbb{E}[\overset{\circ}{\beta}\,e^{\mathbf{i}uI}\ \overset{\circ}{\Phi}]\right| \leq \sqrt{\mathbb{E}\left[\overset{\circ}{\beta}^2\right]}\sqrt{\mathbb{E}\left[\overset{\circ}{\Phi}^2\right]} = \mathcal{O}(n^{-2})\ .$$

We therefore obtain

$$\begin{aligned}\mathbb{E}\left[\overset{\circ}{\beta}\,\mathrm{Tr}\,\left(\mathbf{R}\mathbf{H}\frac{\mathbf{Y}\widetilde{\mathbf{D}}\widetilde{\mathbf{R}}\mathbf{Y}^*}{n}\right)e^{\mathbf{i}uI}\right] &= \mathbb{E}\left[\overset{\circ}{\beta}\,e^{\mathbf{i}uI}\right]\ \mathrm{Tr}\,\left(\widetilde{\mathbf{D}}^2\widetilde{\mathbf{T}}\widetilde{\mathbf{R}}\right)\frac{1}{n}\mathrm{Tr}\,\left(\mathbf{D}\mathbf{T}\mathbf{R}\right) + \mathcal{O}\left(n^{-1}\right) \\ &\overset{(a)}{=} \mathbb{E}\left[\overset{\circ}{\beta}\,e^{\mathbf{i}uI}\right]\ \tilde{\gamma}\ \mathrm{Tr}\,\left(\mathbf{D}\mathbf{T}^2\right) + \mathcal{O}\left(n^{-1}\right)\ ,\end{aligned} \tag{53}$$

where (a) follows from Theorem 3. It remains to deal with the term $\mathbb{E}\left[\overset{\circ}{\beta}\, e^{\mathbf{i}uI}\right]$. To this end, we shall rely on (49) and develop the term $\mathbb{E}\left[H_{pp}e^{\mathbf{i}uI}\right]$. The Resolvent identity yields:

$$\mathbb{E}\left[\left[\mathbf{HYY}^*\right]_{pp} e^{\mathbf{i}uI}\right] = \frac{n}{t}\mathbb{E}\left[e^{\mathbf{i}uI}\right] - \frac{n}{t}\mathbb{E}\left[H_{pp}e^{\mathbf{i}uI}\right] \ .$$

Plugging this equality into (49) and using $r_p = (1+t\tilde{\alpha}d_p)^{-1}$, we obtain after some computations

$$\begin{aligned}
\mathbb{E}\left[\overset{\circ}{\beta}\, e^{\mathbf{i}uI}\right] &= t^2\mathbb{E}\left[\overset{\circ}{\beta}\, e^{\mathbf{i}uI}\, \frac{1}{n}\mathrm{Tr}\left(\mathbf{RDH}\frac{\mathbf{Y}\widetilde{\mathbf{D}}\widetilde{\mathbf{R}}\mathbf{Y}^*}{n}\right)\right] \\
&\quad - \frac{\mathbf{i}ut^2}{n}\mathbb{E}\left[\frac{1}{n}\mathrm{Tr}\left(\mathbf{RDHDH}\frac{\mathbf{Y}\widetilde{\mathbf{D}}\widetilde{\mathbf{R}}\mathbf{Y}^*}{n}\right)e^{\mathbf{i}uI}\right] + \frac{1}{n}\mathrm{Tr}\left(\mathbf{D}\left(\mathbf{R}-\mathbb{E}\left[\mathbf{H}\right]\right)\right)\mathbb{E}\left[e^{\mathbf{i}uI}\right] \\
&\overset{(a)}{=} t^2\gamma\tilde{\gamma}\mathbb{E}\left[\overset{\circ}{\beta}\, e^{\mathbf{i}uI}\right] - \frac{1}{n}\frac{\mathbf{i}ut^2}{(1-t^2\gamma\tilde{\gamma})}\left(\tilde{\gamma}\frac{1}{n}\mathrm{Tr}\left(\mathbf{D}^3\mathbf{T}^3\right) - t\gamma^2\frac{1}{n}\mathrm{Tr}\left(\widetilde{\mathbf{D}}^3\widetilde{\mathbf{T}}^3\right)\right)\varphi + \mathcal{O}(n^{-2}) \qquad (54)
\end{aligned}$$

where (a) follows from Theorem 3, Proposition 4 and Proposition 5. We therefore obtain:

$$\mathbb{E}\left[\overset{\circ}{\beta}\, e^{\mathbf{i}uI}\right] = -\frac{1}{n}\frac{\mathbf{i}ut^2}{(1-t^2\gamma\tilde{\gamma})^2}\left(\tilde{\gamma}\frac{1}{n}\mathrm{Tr}\left(\mathbf{D}^3\mathbf{T}^3\right) - t\gamma^2\frac{1}{n}\mathrm{Tr}\left(\widetilde{\mathbf{D}}^3\widetilde{\mathbf{T}}^3\right)\right)\varphi + \mathcal{O}\left(\frac{1}{n^2}\right) \ .$$

Plugging (54) into (53), and the result together with (51) and (52) into (50), and getting back to (46) and (45), we obtain:

$$\frac{\partial\psi_n(u,t)}{\partial t} = -u^2\eta_n(t)\psi_n(u,t) + \mathcal{O}(n^{-1}) \ ,$$

where

$$\eta_n(t) = \frac{1}{1-t^2\gamma\tilde{\gamma}}\left(-\frac{t^2\gamma\frac{1}{n}\mathrm{Tr}\left(\widetilde{\mathbf{D}}^3\widetilde{\mathbf{T}}^3\right)\frac{1}{n}\mathrm{Tr}\left(\mathbf{DT}^2\right)}{1-t^2\gamma\tilde{\gamma}} + t\tilde{\gamma}\frac{1}{n}\mathrm{Tr}\left(\mathbf{D}^2\mathbf{T}^3\right) + \frac{t^3\tilde{\gamma}^2\frac{1}{n}\mathrm{Tr}\left(\mathbf{D}^3\mathbf{T}^3\right)\frac{1}{n}\mathrm{Tr}\left(\mathbf{DT}^2\right)}{1-t^2\gamma\tilde{\gamma}}\right) \ . \qquad (55)$$

Equation (43) is established, and the first step of the proof is completed.

B. Identification of the variance

In order to finish the proof, it remains to prove that:

$$\eta_n(t) = \frac{1}{2}\frac{d\sigma_n^2(t)}{dt} \qquad \text{where} \quad \sigma_n^2(t) = -\log\left(1-t\gamma_n(t)\tilde{\gamma}_n(t)\right) \ . \qquad (56)$$

To this end, we first begin by computing the derivatives of $\gamma_n(t)$ and $\tilde{\gamma}_n(t)$. We shall prove that

$$\frac{d\tilde{\gamma}}{dt} = -2\frac{\frac{1}{n}\mathrm{Tr}\left(\widetilde{\mathbf{D}}^3\widetilde{\mathbf{T}}^3\right)\frac{1}{n}\mathrm{Tr}\left(\mathbf{DT}^2\right)}{1-t^2\gamma\tilde{\gamma}} \qquad \text{and} \qquad \frac{d\gamma}{dt} = -2\frac{\frac{1}{n}\mathrm{Tr}\left(\mathbf{D}^3\mathbf{T}^3\right)\frac{1}{n}\mathrm{Tr}\left(\widetilde{\mathbf{D}}\widetilde{\mathbf{T}}^2\right)}{1-t^2\gamma\tilde{\gamma}} \ . \qquad (57)$$

We only derive $\frac{d\tilde{\gamma}}{dt}$, the computations being similar in the other case. We first expand the expression of $\tilde{\gamma}$, and obtain:

$$\frac{d\tilde{\gamma}}{dt} = \frac{1}{n}\sum_{j=1}^n \tilde{d}_j^2\frac{d}{dt}\widetilde{T}_{jj}^2 = \frac{1}{n}\sum_{j=1}^n \tilde{d}_j^2\frac{d}{dt}\left(\frac{1}{1+t\delta(t)\tilde{d}_j}\right)^2 = -2\frac{d}{dt}\left(t\delta(t)\right)\frac{1}{n}\mathrm{Tr}\left(\widetilde{\mathbf{D}}^3\widetilde{\mathbf{T}}^3\right) \ . \qquad (58)$$

Let us now compute $\delta'(t)$:

$$\delta'(t) = \frac{1}{n}\sum_{i=1}^{N} d_i \left(\frac{1}{1+t\tilde{\delta}(t)d_i}\right)' = -\gamma\tilde{\delta}(t) - \gamma t\tilde{\delta}'(t) \ . \tag{59}$$

A similar computation yields $\tilde{\delta}'(t) = -\tilde{\gamma}\delta(t) - \tilde{\gamma}t\delta'(t)$. Combining both equations yields:

$$\delta' = \frac{t\gamma\tilde{\gamma}\delta - \gamma\tilde{\delta}}{1-t^2\gamma\tilde{\gamma}} \ .$$

We now plug this into (58) and obtain:

$$\frac{d\tilde{\gamma}}{dt} = -2\frac{\frac{1}{n}\mathrm{Tr}\left(\widetilde{\mathbf{D}}^3\widetilde{\mathbf{T}}^3\right)\left(\delta - t\gamma\tilde{\delta}\right)}{1-t^2\gamma\tilde{\gamma}} \ . \tag{60}$$

Recall now that the mere definition of $\mathbf{T}$, $\widetilde{\mathbf{T}}$, δ and $\tilde{\delta}$ yields

$$\begin{cases} t\delta\widetilde{\mathbf{D}}\widetilde{\mathbf{T}} = \mathbf{I} - \widetilde{\mathbf{T}} \\ t\tilde{\delta}\mathbf{D}\mathbf{T} = \mathbf{I} - \mathbf{T} \end{cases} . \tag{61}$$

Using (61), we obtain:

$$n^{-1}\mathrm{Tr}\left(\mathbf{D}\mathbf{T}^2\right) = n^{-1}\mathrm{Tr}\left(\mathbf{D}\mathbf{T}\left(\mathbf{I} - t\tilde{\delta}\mathbf{D}\mathbf{T}\right)\right) = \delta - t\tilde{\delta}\gamma \ , \tag{62}$$

$$n^{-1}\mathrm{Tr}\left(\widetilde{\mathbf{D}}\widetilde{\mathbf{T}}^2\right) = n^{-1}\mathrm{Tr}\left(\widetilde{\mathbf{D}}\widetilde{\mathbf{T}}\left(\mathbf{I} - t\delta\widetilde{\mathbf{D}}\widetilde{\mathbf{T}}\right)\right) = \tilde{\delta} - t\delta\tilde{\gamma} \ . \tag{63}$$

It remains to plug (62) in (60) to conclude the proof of (57).

We are now in position to prove (56). The main idea in the following computations is to express (55) as a symmetric quantity with respect to δ and $\mathbf{T}$ on the one hand and $\tilde{\delta}$ and $\tilde{\mathbf{T}}$ on the other hand. To this end, we split $\eta_n(t)$ in (55) as $\eta_n(t) = \frac{1}{1-t^2\gamma\tilde{\gamma}}\left(\eta^{(1)} + \eta^{(2)} + \eta^{(3)}\right)$. We first work on $\eta^{(3)}$:

$$\begin{aligned} \eta^{(3)} &\overset{(a)}{=} \frac{t^3\delta\tilde{\gamma}^2\frac{1}{n}\mathrm{Tr}\left(\mathbf{D}^3\mathbf{T}^3\right)}{1-t^2\gamma\tilde{\gamma}} - \frac{t^4\tilde{\delta}\tilde{\gamma}^2\gamma\frac{1}{n}\mathrm{Tr}\left(\mathbf{D}^3\mathbf{T}^3\right)}{1-t^2\gamma\tilde{\gamma}} \ , \\ &\overset{(b)}{=} \frac{-t^2\tilde{\gamma}\frac{1}{n}\mathrm{Tr}\left(\mathbf{D}^3\mathbf{T}^3\right)\frac{1}{n}\mathrm{Tr}\left(\widetilde{\mathbf{D}}\widetilde{\mathbf{T}}^2\right)}{1-t^2\gamma\tilde{\gamma}} + t^2\tilde{\gamma}\tilde{\delta}\frac{1}{n}\mathrm{Tr}\left(\mathbf{D}^3\mathbf{T}^3\right) \ . \end{aligned}$$

where (a) follows from (62), and (b) from (63). We now look at $\eta^{(2)}$:

$$\eta^{(2)} + t^2\tilde{\gamma}\tilde{\delta}\frac{1}{n}\mathrm{Tr}\left(\mathbf{D}^3\mathbf{T}^3\right) = t\tilde{\gamma}\left(\frac{1}{n}\mathrm{Tr}\left(\mathbf{D}^2\mathbf{T}^3 + \frac{1}{n}\mathrm{Tr}\left(\mathbf{D}^2\mathbf{T}^2\left(t\tilde{\delta}\mathbf{D}\mathbf{T}\right)\right)\right)\right) = t\gamma\tilde{\gamma}$$

where the last equality follows (61) again. We therefore have

$$\begin{aligned} \eta_n(t) &= \frac{1}{1-t^2\gamma\tilde{\gamma}}\left(-\frac{t^2\gamma\frac{1}{n}\mathrm{Tr}\left(\widetilde{\mathbf{D}}^3\widetilde{\mathbf{T}}^3\right)\frac{1}{n}\mathrm{Tr}\left(\mathbf{D}\mathbf{T}^2\right)}{1-t^2\gamma\tilde{\gamma}} - \frac{t^2\tilde{\gamma}\frac{1}{n}\mathrm{Tr}\left(\mathbf{D}^3\mathbf{T}^3\right)\frac{1}{n}\mathrm{Tr}\left(\widetilde{\mathbf{D}}\widetilde{\mathbf{T}}^2\right)}{1-t^2\gamma\tilde{\gamma}} + t\gamma\tilde{\gamma}\right) \ , \\ &\overset{(a)}{=} \frac{1}{2}\frac{t^2\gamma\tilde{\gamma}' + t^2\gamma'\tilde{\gamma} + 2t\gamma\tilde{\gamma}}{1-t^2\gamma\tilde{\gamma}} \ , \\ &= -\frac{1}{2}\frac{d}{dt}\log\left(1-t^2\gamma\tilde{\gamma}\right) \ , \end{aligned}$$

where (a) follows from (57). This concludes the identification of the variance.

C. Integration of the differential equation (43)

Let us introduce $K_n(u,\rho) = \psi_n(u,\rho)e^{\frac{u^2}{2}\sigma_n^2(\rho)}$. Due to (43), $K_n(u,\rho)$ readily satisfies the following differential equation:

$$\frac{\partial K_n(u,t)}{\partial t} = \varepsilon_n(u,t)e^{\frac{u^2}{2}\sigma_n^2(t)} \ . \tag{64}$$

As in Section IV-C, one can easily prove that $|\varepsilon_n(t)| \leq \frac{K}{n}$ for every $t \in [0,\rho]$. As $K_n(u,0) = 1$, we get

$$K_n(u,\rho) = 1 + \int_0^\rho \varepsilon_n(u,t)e^{\frac{u^2}{2}\sigma_n^2(t)}\,dt \ .$$

Due to Proposition 5, $\sigma_n^2(t)$ is bounded from above uniformly in n and $t \in [0,\rho]$. This fact, together with $|\varepsilon_n(t)| \leq \frac{K}{n}$ implies that:

$$K_n(u,\rho) = 1 + \mathcal{O}\left(\frac{1}{n}\right) \ .$$

This in turn yields

$$\begin{aligned}\Psi_n(u,\rho) &= \left(1+\mathcal{O}\left(n^{-1}\right)\right)e^{-\frac{u^2}{2}\sigma_n^2(\rho)} \\ &= e^{-\frac{u^2}{2}\sigma_n^2(\rho)} + \mathcal{O}(n^{-1}) \ ,\end{aligned}$$

where the last equality follows from the fact that $\sigma_n^2(\rho)$ is uniformly bounded by n by Proposition 2.

APPENDIX

A. Proof of Proposition 1

Let us first establish the existence and uniqueness of the solution of (2). To this end, we plug the expression of $\tilde{\delta}$ in (2). The system of two equations reduces to the single equation $\delta = f(t,\delta)$ where $f(t,\delta)$ is defined by

$$f(t,\delta) = \frac{1}{n}\mathrm{Tr}\left(\mathbf{D}_n\left(\mathbf{I}_N + \frac{t}{n}\mathrm{Tr}\left(\widetilde{\mathbf{D}}_n\left(\mathbf{I}_n + t\delta\widetilde{\mathbf{D}}_n\right)^{-1}\right)\mathbf{D}_n\right)^{-1}\right) \tag{65}$$

which is itself equivalent to $g(\delta,t) = 1$ where

$$g(t,\delta) = \frac{f(t,\delta)}{\delta} = \frac{1}{n}\mathrm{Tr}\left(\mathbf{D}_n\left(\delta\mathbf{I}_N + \frac{t}{n}\mathrm{Tr}\left(\delta\widetilde{\mathbf{D}}_n\left(\mathbf{I}_n + t\delta\widetilde{\mathbf{D}}_n\right)^{-1}\right)\mathbf{D}_n\right)^{-1}\right) \ .$$

The function $\delta \mapsto g(t,\delta)$ is continuous, decreasing and satisfies $g(t,0) = +\infty$ and $g(t,+\infty) = 0$. Therefore, the equation $g(t,\delta) = 1$ has a unique solution $\delta(t) > 0$.

The integral representation (3) of δ and $\tilde{\delta}$ is related to the Stieltjes representation of a class of analytic functions. We refer to [9, Section 3.2] where a more general result is proven and skip the details. Be aware however that δ in this paper and δ in [9] slightly differ (but are related by $\delta_{\text{IEEE}}(z) = z^{-1}\delta_{\text{AAP}}(-z^{-1})$), so do the $\tilde{\delta}$'s.

B. Proof of Proposition 2

In order to prove Proposition 2, it is sufficient to first prove that $1 - t^2\gamma\tilde{\gamma}$ is bounded away from zero and then to prove that the same quantity is strictly lower than 1, uniformly in n. We shall proceed into four steps.

1) A priori estimates for δ, $\tilde{\delta}$, γ and $\tilde{\gamma}$: The mere definition of δ and $\tilde{\delta}$ yields:

$$\delta = \frac{1}{n}\sum_{i=1}^{N} \frac{d_i}{1 + td_i\tilde{\delta}} \leq \frac{Nd_{\max}}{n} \quad \text{and} \quad \tilde{\delta} = \frac{1}{n}\sum_{j=1}^{n} \frac{\tilde{d}_j}{1 + t\tilde{d}_j\delta} \leq \tilde{d}_{\max} \ . \tag{66}$$

Using these upper estimates, one gets the following lower estimates:

$$\delta \geq \frac{\frac{1}{n}\mathrm{Tr}\ \mathbf{D}_n}{1 + td_{\max}\tilde{d}_{\max}} \quad \text{and} \quad \tilde{\delta} \geq \frac{\frac{1}{n}\mathrm{Tr}\ \widetilde{\mathbf{D}}_n}{1 + \frac{N}{n}td_{\max}\tilde{d}_{\max}} \ . \tag{67}$$

One can notice that due to Assumption **(A2)**, these lower bounds are uniformly bounded in n away from zero. Finally a straightforward application of Jensen's inequality yields:

$$\delta^2 = \left(\frac{1}{n}\sum_{i=1}^{N} d_i T_{ii}\right)^2 \leq \frac{N\gamma}{n} \quad \text{i.e.} \quad \frac{n}{N}\delta^2 \leq \gamma \ , \quad \text{Similarly} \quad \tilde{\delta}^2 \leq \tilde{\gamma} \ . \tag{68}$$

2) An estimate over $\frac{d\tilde{\delta}}{dt}$: The following equalities are straightforward (see for instance (59)):

$$\delta'(t) = -\gamma\tilde{\delta}(t) - \gamma t\tilde{\delta}'(t) \quad \text{and} \quad \tilde{\delta}'(t) = -\tilde{\gamma}\delta(t) - \tilde{\gamma}t\delta'(t) \ . \tag{69}$$

In particular, $|\tilde{\delta}'(0)| = \tilde{\gamma}(0)\delta(0) \leq Nn^{-1}\tilde{d}_{\max}^2 d_{\max}$ which is eventually bounded. Recall that $\tilde{\delta}$ admits the following representation:

$$\tilde{\delta}(t) = \int_0^\infty \frac{\tilde{\mu}(d\lambda)}{1 + t\lambda},$$

where $\tilde{\mu}$ is a nonnegative mesure satisfying $\tilde{\mu}(\mathbb{R}^+) = \frac{1}{n}\mathrm{Tr}\ \widetilde{\mathbf{D}}_n$. In particular, one obtains:

$$0 < -\tilde{\delta}'(t) \ = \ \int_0^\infty \frac{\lambda\tilde{\mu}(d\lambda)}{(1 + t\lambda)^2} \ \leq \ -\tilde{\delta}'(0) \ \leq \ Nn^{-1}\tilde{d}_{\max}^2 d_{\max} \ . \tag{70}$$

3) The quantity $1 - t^2\gamma\tilde{\gamma}$ is bounded away from zero, uniformly in n and for $t \in [0, \rho]$: Eliminating δ' between the two equations in (69) yields:

$$\frac{d\tilde{\delta}}{dt}(1 - t^2\gamma\tilde{\gamma}) \ = \ \tilde{\gamma}(t\tilde{\delta}\gamma - \delta) \ = \ \frac{\tilde{\gamma}}{n}\mathrm{Tr}\,\mathbf{D}_n\mathbf{T}_n\left(t\tilde{\delta}\mathbf{D}_n\mathbf{T}_n - \mathbf{I}_N\right) \ = \ -\frac{\tilde{\gamma}}{n}\mathrm{Tr}\,\mathbf{D}_n\mathbf{T}_n^2 \ ,$$

where the last equality follows from the identity $\mathbf{T}_n = (\mathbf{I}_N + t\tilde{\delta}\mathbf{D}_n)^{-1}$ which yields $(t\tilde{\delta}\mathbf{D}_n\mathbf{T}_n - \mathbf{I}_N) = -\mathbf{T}_n$. Otherwise stated:

$$1 - t^2\gamma\tilde{\gamma} = \frac{\tilde{\gamma}\mathrm{Tr}\,\mathbf{D}_n\mathbf{T}_n^2}{n(-\tilde{\delta}'(t))}.$$

This immediatly implies that $1 - t^2\gamma\tilde{\gamma}$ is positive. In order to check that it is bounded away from zero uniformly in n, notice first that $n^{-1}\mathrm{Tr}\,\mathbf{D}_n\mathbf{T}_n^2 \geq d_{\max}^{-1}\gamma$. Collecting now the previous estimates (68) and (70), we obtain:

$$1 - t^2\gamma\tilde{\gamma} \geq \frac{n^2}{N^2}\frac{\delta^2\tilde{\delta}^2}{d_{\max}^2\tilde{d}_{\max}^2}.$$

Using (67) and Assumption **(A1)**, we obtain that $1 - t^2\gamma\tilde{\gamma}$ is bounded away from zero, uniformly in n and for $t \in [0, \rho]$.

4) The quantity $1 - t^2\gamma\tilde{\gamma}$ is strictly bounded above from 1, uniformly in n: The inequalities (68) together with (67) yield:

$$\sup_n \left(1 - t^2\gamma\tilde{\gamma}\right) \;\leq\; \sup_n \left(1 - t^2\frac{n}{N}\delta^2\tilde{\delta}^2\right) \;<\; 1\ .$$

This completes the proof of Proposition 2.

C. Proof of Theorem 3

We first give a sketch of the proof to emphasize the main ideas over the technical aspects of the proof.

1) We first prove that the asymptotic behaviour of $n^{-1}\mathrm{Tr}\,(\mathbf{A}_n(\mathbf{R}_n - \mathbf{T}_n))$ is directly related to the behaviour of $\alpha(t) - \delta(t)$. Similarly, $n^{-1}\mathrm{Tr}\,\widetilde{\mathbf{A}}_n\left(\widetilde{\mathbf{R}}_n - \widetilde{\mathbf{T}}_n\right)$ is related to $\tilde{\alpha}(t) - \tilde{\delta}(t)$.
2) We extend the definition of α from $t \in \mathbb{R}^+$ to $z \in \mathbb{C} \setminus \mathbb{R}_-$ and establish an integral representation:

$$\alpha(t) = \int_{\mathbb{R}^+} \frac{\nu(d\lambda)}{1 + \lambda t}\ .$$

   As a consequence of the integral representations for δ, $\tilde{\delta}$ and α, we prove that δ, $\tilde{\delta}$ and α are bounded analytic functions on every compact subset of $\mathbb{C} \setminus \mathbb{R}_-$.
3) As a consequence of this detour in the complex plane, we prove the following weaker result. For every uniformly bounded diagonal matrix $\mathbf{A}_n$, the following holds true:

$$\begin{cases} n^{-1}\mathrm{Tr}\,(\mathbf{A}_n\mathbf{R}_n) &= n^{-1}\mathrm{Tr}\,(\mathbf{A}_n\mathbf{T}_n) + o(1) \\ n^{-1}\mathrm{Tr}\,(\tilde{\mathbf{A}}_n\tilde{\mathbf{R}}_n) &= n^{-1}\mathrm{Tr}\,(\tilde{\mathbf{A}}_n\tilde{\mathbf{T}}_n) + o(1) \end{cases}.$$

4) We then refine the previous result in order the get the sharper rate of convergence $\mathcal{O}(n^{-2})$ instead of $o(1)$.

The theorem will then be proved.

1) The asymptotic behaviour of $n^{-1}\mathrm{Tr}\,(\mathbf{A}_n(\mathbf{R}_n-\mathbf{T}_n))$ *and its relation with* $\alpha(t)-\delta(t)$*:* The standard matrix identity

$$\mathbf{R}_n-\mathbf{T}_n=\mathbf{R}_n(\mathbf{T}_n^{-1}-\mathbf{R}_n^{-1})\mathbf{T}_n$$

immediatly yields

$$\begin{aligned} n^{-1}\mathrm{Tr}\,(\mathbf{A}_n(\mathbf{R}_n-\mathbf{T}_n)) &= t(\tilde{\delta}(t)-\tilde{\alpha}(t))\frac{1}{n}\mathrm{Tr}\,(\mathbf{A}_n\mathbf{R}_n\mathbf{D}_n\mathbf{T}_n) \quad \text{and} \\ n^{-1}\mathrm{Tr}\,\left(\widetilde{\mathbf{D}}_n(\widetilde{\mathbf{T}}_n-\widetilde{\mathbf{R}}_n)\right) &= \tilde{\delta}(t)-\tilde{\alpha}(t)=t(\alpha(t)-\delta(t))\frac{1}{n}\mathrm{Tr}\,\left(\widetilde{\mathbf{D}}_n\widetilde{\mathbf{R}}_n\widetilde{\mathbf{D}}_n\widetilde{\mathbf{T}}_n\right)\ . \end{aligned}$$

Therefore,

$$n^{-1}\mathrm{Tr}\,(\mathbf{A}_n(\mathbf{R}_n-\mathbf{T}_n))=t^2(\alpha(t)-\delta(t))\frac{1}{n}\mathrm{Tr}\,\left(\widetilde{\mathbf{D}}_n\widetilde{\mathbf{R}}_n\widetilde{\mathbf{D}}_n\widetilde{\mathbf{T}}_n\right)\frac{1}{n}\mathrm{Tr}\,(\mathbf{A}_n\mathbf{R}_n\mathbf{D}_n\mathbf{T}_n)\ . \tag{71}$$

2) An integral representation for α*, and bounds over* α*,* δ *and* $\tilde{\delta}$*:* Recall that $\alpha(t)=\mathbb{E}[n^{-1}\mathrm{Tr}\,(\mathbf{D}_n(\mathbf{I}_N+tn^{-1}\mathbf{Y}_n\mathbf{Y}_n^*)^{-1})]$. This function readily extends from $t\in\mathbb{R}^+$ to $z\in\mathbb{C}\setminus\mathbb{R}^-$. Moreover, the following representation holds true:

$$\alpha(z)=\int_0^{+\infty}\frac{\nu(d\lambda)}{1+\lambda z}\ , \tag{72}$$

where ν is a uniquely defined positive measure on $\mathbb{R}^+$ such that $\nu(\mathbb{R}^+)=\frac{1}{n}\mathrm{Tr}\,\mathbf{D}_n$. To prove this, we introduce the eigenvalue/eigenvector decomposition of matrix $n^{-1}\mathbf{Y}_n\mathbf{Y}_n^*=\sum_{i=1}^N\lambda_i\mathbf{u}_i\mathbf{u}_i^*$ where $(\lambda_i,\ 1\le i\le N)$ and $(\mathbf{u}_i,\ 1\le i\le N)$ represent its eigenvalues and eigenvectors respectively. The random variable $\beta(z)=\frac{1}{n}\mathrm{Tr}\,\mathbf{D}_n(\mathbf{I}+z\frac{\mathbf{Y}_n\mathbf{Y}_n^*}{n})^{-1}$ can be written as

$$\beta(z)=\frac{1}{n}\sum_{i=1}^N\frac{\mathbf{u}_i^*\mathbf{D}_n\mathbf{u}_i}{\lambda_i-z}=\int_0^{+\infty}\frac{\omega(d\lambda)}{1+\lambda z}\ ,$$

where ω is the nonnegative random measure defined by

$$\omega=\frac{1}{n}\sum_{i=1}^N\mathbf{u}_i^*\mathbf{D}_n\mathbf{u}_i\delta(\lambda-\lambda_i)\ .$$

Consider now the measure ν defined by $\nu=\mathbb{E}[\omega]$, that is $\nu(B)=\mathbb{E}[\omega(B)]$ for every Borel set $B\subset\mathbb{R}^+$. It is clear that $\alpha(z)=\mathbb{E}[\beta(z)]$ is given by (72), and that $\nu(\mathbb{R}^+)=\mathbb{E}[\omega(\mathbb{R}^+)]$ is given by

$$\nu(\mathbb{R}^+)=\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^N\mathbf{u}_i^*\mathbf{D}_n\mathbf{u}_i\right]=\mathbb{E}\left[\frac{1}{n}\mathrm{Tr}\,\mathbf{D}_n(\sum_i\mathbf{u}_i\mathbf{u}_i^*)\right]\ .$$

As $\sum_i\mathbf{u}_i\mathbf{u}_i^*=\mathbf{I}_N$, $\nu(\mathbb{R}^+)=\frac{1}{n}\mathrm{Tr}\,\mathbf{D}_n$ as expected and representation (72) implies that $\alpha(z)$ is analytic over $\mathbb{C}\setminus\mathbb{R}^-$.

Let dist$(w, \mathbb{R}^+)$ stand for the distance from element $w \in \mathbb{C}$ to $\mathbb{R}^+$. Then the following holds true for every $z \in \mathbb{C} \setminus \mathbb{R}^-$:

$$|\alpha(z)| \leq \frac{1}{n}\mathrm{Tr}\,(\mathbf{D}_n)\frac{1}{|z|}\frac{1}{\mathrm{dist}(-\frac{1}{z},\mathbb{R}^+)} \leq \frac{N}{n}d_{\max}\frac{1}{|z|}\frac{1}{\mathrm{dist}(-\frac{1}{z},\mathbb{R}^+)} \ . \tag{73}$$

Similarly, (3) yields that

$$|\delta(z)| \leq \frac{Nd_{\max}}{n|z|}\frac{1}{\mathrm{dist}(-\frac{1}{z},\mathbb{R}^+)} \ . \tag{74}$$

A similar result holds for $\tilde{\delta}(z)$. These upper bounds imply in particular that $\alpha(z)$, $\delta(z)$ and $\tilde{\delta}(z)$ are uniformly bounded on each compact subset of $\mathbb{C} \setminus \mathbb{R}_-$.

3) A weaker result as a consequence of Montel's theorem: We first establish that for every diagonal matrix $\mathbf{A}_n$ uniformly bounded,

$$\left\{ \begin{array}{ll} n^{-1}\mathrm{Tr}\,(\mathbf{A}_n\mathbf{R}_n) & = n^{-1}\mathrm{Tr}\,(\mathbf{A}_n\mathbf{T}_n) + o(1) \\ n^{-1}\mathrm{Tr}\,(\tilde{\mathbf{A}}_n\tilde{\mathbf{R}}_n) & = n^{-1}\mathrm{Tr}\,(\tilde{\mathbf{A}}_n\tilde{\mathbf{T}}_n) + o(1) \end{array} \right. . \tag{75}$$

We take (71) as a starting point. Matrices $\mathbf{R}_n, \widetilde{\mathbf{R}}_n, \mathbf{T}_n$, and $\widetilde{\mathbf{T}}_n$ have their spectral norms bounded by one for $t \in \mathbb{R}^+$ and matrices $\mathbf{A}_n, \mathbf{D}_n$, and $\widetilde{\mathbf{D}}_n$ are also uniformly bounded by assumption. Therefore, the terms $n^{-1}\mathrm{Tr}\left(\widetilde{\mathbf{A}}_n\widetilde{\mathbf{R}}_n\widetilde{\mathbf{D}}_n\widetilde{\mathbf{T}}_n\right)$ and $n^{-1}\mathrm{Tr}\,(\mathbf{A}_n\mathbf{R}_n\mathbf{D}_n\mathbf{T}_n)$ are also bounded. In order to prove (75), it is sufficient to prove that $\alpha(t) - \delta(t) = o(1)$. To this end, we make use of Proposition 5 and write $\alpha(t) - \delta(t)$ as

$$\alpha(t) - \delta(t) = \frac{1}{n}\mathrm{Tr}\,(\mathbf{D}_n(\mathbf{R}_n - \mathbf{T}_n)) + \varepsilon_n(t) \ ,$$

where $\varepsilon_n(t) = \mathcal{O}(n^{-2})$. Using relation (71) for $\mathbf{A}_n = \mathbf{D}_n$, we immediately get that:

$$\alpha(t) - \delta(t) = (\alpha(t) - \delta(t))t^2\frac{1}{n}\mathrm{Tr}\left(\widetilde{\mathbf{D}}_n\widetilde{\mathbf{R}}_n\widetilde{\mathbf{D}}_n\widetilde{\mathbf{T}}_n\right)\frac{1}{n}\mathrm{Tr}\,(\mathbf{D}_n\mathbf{R}_n\mathbf{D}_n\mathbf{T}_n) + \varepsilon_n(t) \ . \tag{76}$$

As $\sup_n\left(\|\mathbf{R}_n\|, \|\widetilde{\mathbf{R}}_n\|, \|\mathbf{T}_n\|, \|\widetilde{\mathbf{T}}_n\|\right) \leq 1$, we have:

$$\frac{1}{n}\mathrm{Tr}\left(\widetilde{\mathbf{D}}_n\widetilde{\mathbf{R}}_n\widetilde{\mathbf{D}}_n\widetilde{\mathbf{T}}_n\right)\frac{1}{n}\mathrm{Tr}\,(\mathbf{D}_n\mathbf{R}_n\mathbf{D}_n\mathbf{T}_n) \leq \frac{N}{n}d_{\max}^2\tilde{d}_{\max}^2 \leq 2cd_{\max}^2\tilde{d}_{\max}^2$$

as soon as $\frac{N}{n} \leq 2\ell^+$, where ℓ^+ is defined in (1). Therefore, if $t < t_0 := (2d_{\max}\tilde{d}_{\max}\sqrt{\ell^+})^{-1}$, then

$$t^2\frac{1}{n}\mathrm{Tr}\left(\widetilde{\mathbf{D}}_n\widetilde{\mathbf{R}}_n\widetilde{\mathbf{D}}_n\widetilde{\mathbf{T}}_n\right)\frac{1}{n}\mathrm{Tr}\,(\mathbf{D}_n\mathbf{R}_n\mathbf{D}_n\mathbf{T}_n) < \frac{1}{2}$$

for n large enough. Eq. (76) thus implies that

$$|\alpha_n(t) - \delta_n(t)| < 2|\varepsilon_n(t)|, \quad i.e. \quad \alpha(t) - \delta(t) = \mathcal{O}(n^{-2}) \quad \text{for} \quad t < t_0 \ . \tag{77}$$

This in particular implies that $\alpha_n(t) - \delta_n(t) = o(1)$ for $t < t_0$; however, it remains to establish this convergence for $t \geq t_0$. To this end, observe that $\alpha_n(z) - \delta_n(z)$ is analytic in $\mathbb{C} \setminus \mathbb{R}_-$ and bounded on each compact subset of $\mathbb{C} \setminus \mathbb{R}_-$. Montel's theorem asserts that the sequence of functions $\alpha_n(z) - \delta_n(z)$ is compact and therefore that there exists a converging subsequence which converges towards an analytic function. Since this limiting function is zero on $[0, t_0[$ by (77), it must be zero everywhere due to the analycity. Therefore from every subsequence, one can extract a subsequence that converges toward zero. Necessarily, $\alpha_n(z) - \delta_n(z)$ converges to zero for every $z \in \mathbb{C} \setminus \mathbb{R}^-$ and in particular for $t \geq 0$. This establishes (75).

Even if the convergence rate of $\alpha_n(t) - \delta_n(t)$ is $\mathcal{O}(n^{-2})$ for $t < t_0$, Montel's theorem does not imply that the convergence rate of $\alpha_n(z) - \delta_n(z)$ remains $\mathcal{O}(n^{-2})$ elsewhere. Therefore, there remains some work to be done in order to prove that $\alpha_n(t) - \delta_n(t) = \mathcal{O}(n^{-2})$ for each $t > 0$.

4) End of the proof: We take (76) as a starting point. Equations (75) imply that for each $t \geq 0$,

$$\left\{ \begin{array}{ll} n^{-1}\mathrm{Tr}\ \left(\mathbf{D}_n \mathbf{R}_n(t) \mathbf{D}_n \mathbf{T}_n(t)\right) - \gamma(t) & = o(1) \\ n^{-1}\mathrm{Tr}\ \left(\widetilde{\mathbf{D}}_n \widetilde{\mathbf{R}}_n(t) \widetilde{\mathbf{D}}_n \widetilde{\mathbf{T}}_n(t)\right) - \tilde{\gamma}(t) & = o(1) \end{array} \right. . \tag{78}$$

where $\gamma_n = n^{-1}\mathrm{Tr}\, \mathbf{D}_n^2 \mathbf{T}_n^2$ and $\tilde{\gamma}_n = n^{-1}\mathrm{Tr}\, \widetilde{\mathbf{D}}_n^2 \widetilde{\mathbf{T}}_n^2$. Thanks to Proposition 6, (78) implies that

$$\inf_n \left(1 - t^2 \frac{1}{n}\mathrm{Tr}\ \left(\mathbf{D}_n \mathbf{R}_n(t) \mathbf{D}_n \mathbf{T}_n(t)\right) \frac{1}{n}\mathrm{Tr}\ \left(\widetilde{\mathbf{D}}_n \widetilde{\mathbf{R}}_n(t) \widetilde{\mathbf{D}}_n \widetilde{\mathbf{T}}_n(t)\right) \right) > 0 \ .$$

Equation (76) thus clearly implies that $\alpha(t) - \delta(t)$ is of the same order of magnitude as $\varepsilon_n(t)$, i.e. that $\alpha(t) - \delta(t) = \mathcal{O}(n^{-2})$. Theorem 3 is proved.

D. Proof of Proposition 4-(1) *- Variance controls*

Consider first $\Phi(\mathbf{Y}) = \frac{1}{n}\mathrm{Tr}\ \left(\mathbf{A}\mathbf{H}\frac{\mathbf{Y}\mathbf{B}\mathbf{Y}^*}{n}\right)$. We use Poincaré-Nash inequality (20) to control the variance of Φ. It writes

$$\mathbb{E}\left[\mathring{\Phi}(\mathbf{Y})^2\right] \leq \sum_{i=1}^{N}\sum_{j=1}^{n} d_i \tilde{d}_j \mathbb{E}\left[\left|\frac{\partial \Phi}{\partial Y_{i,j}}\right|^2\right] + \sum_{i=1}^{N}\sum_{j=1}^{n} d_i \tilde{d}_j \mathbb{E}\left[\left|\frac{\partial \Phi}{\partial \overline{Y_{i,j}}}\right|^2\right] \ . \tag{79}$$

We have $\Phi(\mathbf{Y}) = (1/n^2)\sum_{p,r=1}^{N}\sum_{q=1}^{n} a_p b_q H_{pr} Y_{rq}\overline{Y_{pq}}$. From the differentiation formula (14) we have

$$\frac{\partial}{\partial Y_{ij}}\left(H_{pr}Y_{rq}\overline{Y_{pq}}\right) = -\frac{t}{n}H_{pi}[\mathbf{y}_j^*\mathbf{H}]_r Y_{rq}\overline{Y_{pq}} + H_{pr}\overline{Y_{pq}}\delta(r-i)\delta(q-j) \ .$$

Therefore, after a straightforward computation we obtain $\partial\Phi/\partial Y_{ij} = \phi_{ij}^{(1)} + \phi_{ij}^{(2)}$ with

$$\phi_{ij}^{(1)} = -\frac{t}{n^3}\left[\mathbf{y}_j^*\mathbf{H}\mathbf{Y}\mathbf{B}\mathbf{Y}^*\mathbf{A}\mathbf{H}\right]_i \quad \text{and} \quad \phi_{ij}^{(2)} = \frac{1}{n^2}b_j\left[\mathbf{y}_j^*\mathbf{A}\mathbf{H}\right]_i \ .$$

The first term of the right hand side of inequality (79) can be treated as follows:

$$\begin{aligned}\sum_{i=1}^{N}\sum_{j=1}^{n} d_i\tilde{d}_j\mathbb{E}\left[\left|\frac{\partial\Phi}{\partial Y_{i,j}}\right|^2\right] &\leq 2\sum_{i=1}^{N}\sum_{j=1}^{n} d_i\tilde{d}_j\left(\mathbb{E}\left[\left|\phi_{ij}^{(1)}\right|^2\right]+\mathbb{E}\left[\left|\phi_{ij}^{(2)}\right|^2\right]\right)\\ &= \frac{2t^2}{n^6}\mathbb{E}\left[\mathrm{Tr}\,\left(\mathbf{HYBY}^*\mathbf{AHDHAYBY}^*\mathbf{HY}\widetilde{\mathbf{D}}\mathbf{Y}^*\right)\right]\\ &\quad+\frac{2}{n^4}\mathbb{E}\left[\mathrm{Tr}\,\left(\mathbf{AHDHAYB}^2\widetilde{\mathbf{D}}\mathbf{Y}^*\right)\right]\ . \end{aligned}\tag{80}$$

Let $A=\sup\|\mathbf{A}\|$. Using inequalities (8), (9), (11) and Cauchy-Schwarz inequality, we have

$$\begin{aligned}&\frac{2t^2}{n^6}\mathbb{E}\left[\mathrm{Tr}\,\left(\mathbf{H\ YBY}^*\ \mathbf{AHDHA\ YBY}^*\ \mathbf{H\ Y}\widetilde{\mathbf{D}}\mathbf{Y}^*\right)\right]\\ &\qquad\leq \frac{2t^2}{n^6}\mathbb{E}\left[\sqrt{\mathrm{Tr}\,\left((\mathbf{HYBY}^*\mathbf{AHDHAYBY}^*\mathbf{H})^2\right)}\sqrt{\mathrm{Tr}\,\left(\left(\mathbf{Y}\widetilde{\mathbf{D}}\mathbf{Y}^*\right)^2\right)}\right]\\ &\qquad\leq \frac{2t^2}{n^6}\mathbb{E}\left[\|\mathbf{H}\|^4\|\mathbf{A}\|^2\|\mathbf{D}\|\sqrt{\mathrm{Tr}\,\left((\mathbf{YBY}^*)^4\right)}\sqrt{\mathrm{Tr}\,\left(\left(\mathbf{Y}\widetilde{\mathbf{D}}\mathbf{Y}^*\right)^2\right)}\right]\\ &\qquad\leq \frac{2d_{\max}A^2t^2}{n^2}\sqrt{\frac{1}{n}\mathbb{E}\left[\mathrm{Tr}\,\left(\left(\frac{\mathbf{YBY}^*}{n}\right)^4\right)\right]}\sqrt{\frac{1}{n}\mathbb{E}\left[\mathrm{Tr}\,\left(\left(\frac{\mathbf{Y}\widetilde{\mathbf{D}}\mathbf{Y}^*}{n}\right)^2\right)\right]}\\ &\qquad< \frac{K}{n^2}\ , \end{aligned}\tag{81}$$

where the last inequality is due to (12). Turning to the second term of the right hand side of (80), we have

$$\frac{2}{n^4}\mathbb{E}\left[\mathrm{Tr}\,\left(\mathbf{AHDHAYB}^2\widetilde{\mathbf{D}}\mathbf{Y}^*\right)\right]\leq\frac{2A^2d_{\max}}{n^2}\mathbb{E}\left[\frac{1}{n}\mathrm{Tr}\,\left(\frac{1}{n}\mathbf{YB}^2\widetilde{\mathbf{D}}\mathbf{Y}^*\right)\right]<\frac{K}{n^2}\ .\tag{82}$$

The second term of the right hand side of Inequality (79) is treated similarly. This proves that $\mathrm{var}(\Phi)=\mathcal{O}(n^{-2})$.

Consider now $\Psi(\mathbf{Y})=\frac{1}{n}\mathrm{Tr}\,\left(\mathbf{AHDH}\frac{\mathbf{YBY}^*}{n}\right)$. The proof being quite similar to the previous one, we just give its main steps. By (20) we have $\mathbb{E}[\mathring{\Psi}(\mathbf{Y})^2]\leq\sum_{i=1}^{N}\sum_{j=1}^{n}d_i\tilde{d}_j\left(\mathbb{E}[|\partial\Psi/\partial Y_{i,j}|^2]+\mathbb{E}[|\partial\Psi/\partial\overline{Y_{i,j}}|^2]\right)$. A computation similar to above yields $\partial\Psi/\partial Y_{ij}=\psi_{ij}^{(1)}+\psi_{ij}^{(2)}+\psi_{ij}^{(3)}$ where

$$\begin{aligned}\psi_{ij}^{(1)} &= -\frac{t}{n^3}\left[\mathbf{y}_j^*\mathbf{HDHYBY}^*\mathbf{AH}\right]_i,\\ \psi_{ij}^{(2)} &= -\frac{t}{n^3}\left[\mathbf{y}_j^*\mathbf{HYBY}^*\mathbf{AHDH}\right]_i,\\ \psi_{ij}^{(3)} &= \frac{1}{n^2}b_j\left[\mathbf{y}_j^*\mathbf{AHDH}\right]_i\ .\end{aligned}$$

We have

$$\begin{aligned}\sum_{i=1}^{N}\sum_{j=1}^{n} d_i\tilde{d}_j\mathbb{E}\left[\left|\frac{\partial\Psi}{\partial Y_{i,j}}\right|^2\right] &\leq 3\sum_{i=1}^{N}\sum_{j=1}^{n} d_i\tilde{d}_j\left(\mathbb{E}\left[\left|\psi_{ij}^{(1)}\right|^2\right]+\mathbb{E}\left[\left|\psi_{ij}^{(2)}\right|^2\right]+\mathbb{E}\left[\left|\psi_{ij}^{(3)}\right|^2\right]\right)\\ &= \frac{3t^2}{n^6}\mathbb{E}\left[\mathrm{Tr}\left(\mathbf{HDH\ YBY}^*\ \mathbf{AHDHA\ YBY}^*\ \mathbf{HDH\ Y}\widetilde{\mathbf{D}}\mathbf{Y}^*\right)\right]\\ &\quad+\frac{3t^2}{n^6}\mathbb{E}\left[\mathrm{Tr}\left(\mathbf{H\ YBY}^*\ \mathbf{A}\left(\mathbf{HD}\right)^3\mathbf{HA\ YBY}^*\ \mathbf{H\ Y}\widetilde{\mathbf{D}}\mathbf{Y}^*\right)\right]\\ &\quad+\frac{3}{n^4}\mathbb{E}\left[\mathrm{Tr}\left(\mathbf{A}\left(\mathbf{HD}\right)^3\mathbf{HA\ YB}^2\widetilde{\mathbf{D}}\mathbf{Y}^*\right)\right]\ .\end{aligned}$$

The first two terms of the right hand side can be bounded by a series of inequalities similar to inequalities (81). The third term can be bounded as in (82). This ends the proofs of the variance controls in Proposition 4.

E. Proof of Proposition 4-(2) *- Approximation rules*

Consider first $\Phi(Y)=\frac{1}{n}\mathrm{Tr}\left(\mathbf{AH}\frac{\mathbf{YBY}^*}{n}\right)$. we write $\Phi(\mathbf{Y})=n^{-2}\sum_{p,i=1}^{N}\sum_{j=1}^{n}a_pb_j\mathbb{E}\left[Y_{ij}H_{pi}\overline{Y_{pj}}\right]$ and apply the Integration by parts formula (19) to the summand. Using identity (15), we have

$$\mathbb{E}\left[Y_{ij}H_{pi}\overline{Y_{pj}}\right]=d_i\tilde{d}_j\mathbb{E}\left[\frac{\partial}{\partial\overline{Y_{ij}}}\left(H_{pi}\overline{Y_{pj}}\right)\right]=-\frac{t}{n}d_i\tilde{d}_j\mathbb{E}\left[\left[\mathbf{Hy}_j\right]_p H_{ii}\overline{Y_{pj}}\right]+d_i\tilde{d}_j\delta(i-p)\mathbb{E}\left[H_{pi}\right]\ .$$

By taking the sum over the index i, we obtain $\mathbb{E}\left[\left[\mathbf{Hy}_j\right]_p\overline{Y_{pj}}\right]=-t\tilde{d}_j\mathbb{E}\left[\beta\ \left[\mathbf{Hy}_j\right]_p\overline{Y_{pj}}\right]+d_p\tilde{d}_j\mathbb{E}\left[H_{pp}\right]$. Writing now $\beta=\overset{\circ}{\beta}+\alpha$ and then grouping together the terms with $\mathbb{E}\left[\left[\mathbf{Hy}_j\right]_p\overline{Y_{pj}}\right]$, we obtain:

$$\mathbb{E}\left[\left[\mathbf{Hy}_j\right]_p\overline{Y_{pj}}\right]=-t\tilde{d}_j\tilde{r}_j\mathbb{E}\left[\overset{\circ}{\beta}\ \left[\mathbf{Hy}_j\right]_p\overline{Y_{pj}}\right]+d_p\tilde{d}_j\tilde{r}_j\mathbb{E}\left[H_{pp}\right]\ .$$

We now sum over j and p, and obtain:

$$\mathbb{E}\left[\frac{1}{n}\mathrm{Tr}\left(\mathbf{AH}\frac{\mathbf{YBY}^*}{n}\right)\right]=\frac{1}{n}\mathrm{Tr}\left(\widetilde{\mathbf{D}}\widetilde{\mathbf{R}}\mathbf{B}\right)\frac{1}{n}\mathrm{Tr}\left(\mathbf{AD}\ \mathbb{E}\left[\mathbf{H}\right]\right)+\varepsilon\ ,$$

with

$$\varepsilon=-t\ \mathbb{E}\left[\overset{\circ}{\beta}\frac{1}{n}\mathrm{Tr}\left(\mathbf{AH}\frac{\mathbf{Y}\widetilde{\mathbf{D}}\widetilde{\mathbf{R}}\mathbf{BY}^*}{n}\right)\right]=-t\ \mathbb{E}\left[\overset{\circ}{\beta}\frac{1}{n}\overbrace{\mathrm{Tr}\left(\mathbf{AH}\frac{\mathbf{Y}\widetilde{\mathbf{D}}\widetilde{\mathbf{R}}\mathbf{BY}^*}{n}\right)}^{\circ}\right]\ .$$

Applying Cauchy-Schwarz inequality, Proposition 3 and the variance controls in Proposition 4, we get $|\varepsilon|=\mathcal{O}(n^{-2})$.

By Theorem 3, $n^{-1}\mathrm{Tr}\left(\widetilde{\mathbf{D}}\widetilde{\mathbf{R}}\mathbf{B}\right)=n^{-1}\mathrm{Tr}\left(\widetilde{\mathbf{D}}\widetilde{\mathbf{T}}\mathbf{B}\right)+\mathcal{O}(n^{-2})$. By Theorem 3 and Proposition 5, we obtain $n^{-1}\mathrm{Tr}\left(\mathbf{AD}\ \mathbb{E}\left[\mathbf{H}\right]\right)=n^{-1}\mathrm{Tr}\left(\mathbf{ADT}\right)+\mathcal{O}(n^{-2})$. This ends the proof of (23).

Consider now $\Psi(\mathbf{Y}) = \frac{1}{n}\mathrm{Tr}\left(\mathbf{AHDH}\frac{\mathbf{YBY}^*}{n}\right)$. In order to compute $\mathbb{E}\Psi(\mathbf{Y})$, we shall need the following intermediate result:

Lemma 1: In the setting of Theorem 1, let $\Upsilon(\mathbf{Y}) = \frac{1}{n}\mathrm{Tr}\ (\mathbf{DHDH})$. Then

1) The following estimate holds true:

$$\mathrm{var}\left[\Upsilon(\mathbf{Y})\right] = \mathcal{O}\left(\frac{1}{n^2}\right) \ ,$$

2) moreover,

$$\mathbb{E}\left[\Upsilon(\mathbf{Y})\right] = \frac{\gamma}{1-t^2\gamma\tilde{\gamma}} + \mathcal{O}\left(\frac{1}{n^2}\right) \ .$$

Proof: In order to prove Lemma 1-(1), we use the Resolvent identity (10) and write:

$$\mathbf{DHDH} = \mathbf{DHD} - tn^{-1}\mathbf{DHDHYY}^* \ .$$

Since $\mathrm{var}(X+Y) \leq 2\mathrm{var}(X) + 2\mathrm{var}(Y)$, we only need to deal with each term of the right handside. By Proposition 3, $\mathrm{var}(n^{-1}\mathrm{Tr}\ \mathbf{DHD}) = \mathcal{O}(n^{-2})$ and by Proposition 4-(1), $\mathrm{var}(tn^{-2}\mathrm{Tr}\ \mathbf{DHDHYY}^*) = \mathcal{O}(n^{-2})$ and the proof of Lemma 1-(1) is completed.

Let us now prove Lemma 1-(2). The Resolvent identity (10) yields:

$$\mathbb{E}\left[\left[\mathbf{HDH}\right]_{pp}\right] = d_p\mathbb{E}\left[H_{pp}\right] - t\mathbb{E}\left[\left[\mathbf{HDH}\frac{\mathbf{YY}^*}{n}\right]_{pp}\right] \ . \tag{83}$$

We then write $\mathbb{E}\left[\left[\mathbf{HDH}\frac{\mathbf{YY}^*}{n}\right]_{pp}\right] = n^{-1}\sum_{k,i=1}^{N}\sum_{j=1}^{n} d_k H_{pk}H_{ki}Y_{ij}\overline{Y_{pj}}$, and apply the differentiation formula (14) to the summand. After derivations similar to (47–48), we obtain:

$$\begin{aligned}\frac{1}{n}\mathbb{E}\left[\left[\mathbf{HDHy}_j\right]_p\overline{Y_{pj}}\right] &= -\frac{t}{n}\tilde{d}_j\tilde{r}_j\mathbb{E}\left[\left[\mathbf{Hy}_j\right]_p\overline{Y_{pj}}\frac{1}{n}\mathrm{Tr}\ (\mathbf{DHDH})\right] \\ &\quad -\frac{t}{n}\tilde{d}_j\tilde{r}_j\mathbb{E}\left[\overset{\circ}{\beta}\ \left[\mathbf{HDHy}_j\right]_p\overline{Y_{pj}}\right] \\ &\quad +\frac{1}{n}d_p\tilde{d}_j\tilde{r}_j\mathbb{E}\left[\left[\mathbf{HDH}\right]_{pp}\right] \ . \end{aligned} \tag{84}$$

Taking the sum over j and combining with (83) yields:

$$\begin{aligned}\mathbb{E}\left[\left[\mathbf{HDH}\right]_{pp}\right] &= t^2r_p\mathbb{E}\left[\left[\mathbf{H}\frac{\mathbf{Y}\widetilde{\mathbf{D}}\widetilde{\mathbf{R}}\mathbf{Y}^*}{n}\right]_{pp}\frac{1}{n}\mathrm{Tr}\ (\mathbf{DHDH})\right] \\ &\quad +t^2r_p\mathbb{E}\left[\overset{\circ}{\beta}\left[\mathbf{HDH}\frac{\mathbf{Y}\widetilde{\mathbf{D}}\widetilde{\mathbf{R}}\mathbf{Y}^*}{n}\right]_{pp}\right] \\ &\quad +r_pd_p\mathbb{E}\left[H_{pp}\right] \ . \end{aligned} \tag{85}$$

Taking now the sum over p, we obtain:

$$\mathbb{E}\left[\frac{1}{n}\mathrm{Tr}\,\left(\mathbf{DHDH}\right)\right] = \frac{1}{n}\sum_{p=1}^{N} d_p \mathbb{E}\left[\left[\mathbf{HDH}\right]_{pp}\right] = \chi_1 + \chi_2 + \chi_3 \ , \tag{86}$$

where

$$\begin{aligned}
\chi_1 &= t^2\,\mathbb{E}\left[\frac{1}{n}\mathrm{Tr}\left(\mathbf{DRH}\frac{\mathbf{Y}\widetilde{\mathbf{D}}\widetilde{\mathbf{R}}\mathbf{Y}^*}{n}\right)\frac{1}{n}\mathrm{Tr}\,\left(\mathbf{DHDH}\right)\right] \ , \\
\chi_2 &= t^2\,\mathbb{E}\left[\overset{\circ}{\beta}\frac{1}{n}\mathrm{Tr}\left(\mathbf{DRHDH}\frac{\mathbf{Y}\widetilde{\mathbf{D}}\widetilde{\mathbf{R}}\mathbf{Y}^*}{n}\right)\right] \ , \\
\chi_3 &= \frac{1}{n}\mathrm{Tr}\left(\mathbf{D}^2\mathbf{R}\mathbb{E}\left[\mathbf{H}\right]\right) \ .
\end{aligned}$$

Let us first deal with the terms χ_2 and χ_3. Cauchy-Schwarz inequality together with Proposition 3 and Proposition 4-(1) yield $\chi_2 = \mathcal{O}(n^{-2})$. Proposition 5 together with Theorem 3 yield $\chi_3 = \gamma + \mathcal{O}(n^{-2})$. We now look at χ_1. Due to Proposition 4-(1) and to Lemma 1-(1), we have:

$$\begin{aligned}
\chi_1 &= t^2\,\mathbb{E}\left[\frac{1}{n}\mathrm{Tr}\left(\mathbf{DRH}\frac{\mathbf{Y}\widetilde{\mathbf{D}}\widetilde{\mathbf{R}}\mathbf{Y}^*}{n}\right)\right]\mathbb{E}\left[\frac{1}{n}\mathrm{Tr}\,\left(\mathbf{DHDH}\right)\right] + \mathcal{O}\left(\frac{1}{n^2}\right) \ , \\
&\overset{(a)}{=} t^2\gamma\tilde{\gamma}\mathbb{E}\left[\frac{1}{n}\mathrm{Tr}\,\left(\mathbf{DHDH}\right)\right] + \mathcal{O}\left(\frac{1}{n^2}\right) \ ,
\end{aligned}$$

where (a) follows from (23) in Proposition 4. It remains to plug the values obtained for χ_1, χ_2 and χ_3 into (86) to obtain:

$$(1 - t^2\gamma\tilde{\gamma})\mathbb{E}\left[\frac{1}{n}\mathrm{Tr}\,\left(\mathbf{DHDH}\right)\right] = \gamma + \mathcal{O}(n^{-2}) \ .$$

Recalling Proposition 2, we can divide by $(1 - t^2\gamma\tilde{\gamma})$ and obtain the desired result. ■

We can now go back to the computation of $\mathbb{E}\Psi(\mathbf{Y})$. Let us give the main steps of the derivation. Expanding $\mathbb{E}\Psi(\mathbf{Y})$ yields:

$$\mathbb{E}\left[\frac{1}{n}\mathrm{Tr}\left(\mathbf{AHDH}\frac{\mathbf{YBY}^*}{n}\right)\right] = \frac{1}{n^2}\sum_{p=1}^{N}\sum_{j=1}^{n} a_p b_j \mathbb{E}\left[\left[\mathbf{HDHy}_j\right]_p \overline{Y_{pj}}\right] \ .$$

We replace the summand $n^{-1}\mathbb{E}\left[\left[\mathbf{HDHy}_j\right]_p \overline{Y_{pj}}\right]$ by the expression given by (84). We then replace the term $\mathbb{E}\left[\left[\mathbf{HDH}\right]_{pp}\right]$ in (84) by the expression given by (85). We sum over p and j and notice afterwards that the terms where $\overset{\circ}{\beta}$ is involved are of order $\mathcal{O}(n^{-2})$. We therefore end up with:

$$\begin{aligned}
\mathbb{E}\left[\frac{1}{n}\mathrm{Tr}\left(\mathbf{AHDH}\frac{\mathbf{YBY}^*}{n}\right)\right] &= -t\,\mathbb{E}\left[\frac{1}{n}\mathrm{Tr}\,\left(\mathbf{DHDH}\right)\frac{1}{n}\mathrm{Tr}\left(\mathbf{AH}\frac{\mathbf{Y}\widetilde{\mathbf{D}}\widetilde{\mathbf{R}}\mathbf{BY}^*}{n}\right)\right] \\
&\quad + \frac{t^2}{n}\mathrm{Tr}\left(\widetilde{\mathbf{D}}\widetilde{\mathbf{R}}\mathbf{B}\right)\mathbb{E}\left[\frac{1}{n}\mathrm{Tr}\,\left(\mathbf{DHDH}\right)\frac{1}{n}\mathrm{Tr}\left(\mathbf{ARDH}\frac{\mathbf{Y}\widetilde{\mathbf{D}}\widetilde{\mathbf{R}}\mathbf{Y}^*}{n}\right)\right] \\
&\quad + \frac{1}{n}\mathrm{Tr}\left(\widetilde{\mathbf{D}}\widetilde{\mathbf{R}}\mathbf{B}\right)\frac{1}{n}\mathrm{Tr}\,\left(\mathbf{AD}^2\mathbf{R}\,\mathbb{E}\mathbf{H}\right) + \mathcal{O}\left(\frac{1}{n^2}\right) \ .
\end{aligned}$$

We first decorrelate by using the variance estimates in Proposition 4-(1) and Lemma 1-(1) and obtain:

$$\begin{aligned}\mathbb{E}\left[\frac{1}{n}\mathrm{Tr}\left(\mathbf{AHDH}\frac{\mathbf{YBY}^*}{n}\right)\right] &= -t\,\mathbb{E}\left[\frac{1}{n}\mathrm{Tr}\,(\mathbf{DHDH})\right]\mathbb{E}\left[\frac{1}{n}\mathrm{Tr}\left(\mathbf{AH}\frac{\mathbf{Y}\widetilde{\mathbf{D}}\widetilde{\mathbf{R}}\mathbf{BY}^*}{n}\right)\right] \\ &\quad +t^2\frac{1}{n}\mathrm{Tr}\left(\widetilde{\mathbf{D}}\widetilde{\mathbf{R}}\mathbf{B}\right)\mathbb{E}\left[\frac{1}{n}\mathrm{Tr}\,(\mathbf{DHDH})\right]\mathbb{E}\left[\frac{1}{n}\mathrm{Tr}\left(\mathbf{ARDH}\frac{\mathbf{Y}\widetilde{\mathbf{D}}\widetilde{\mathbf{R}}\mathbf{Y}^*}{n}\right)\right] \\ &\quad +\frac{1}{n}\mathrm{Tr}\left(\widetilde{\mathbf{D}}\widetilde{\mathbf{R}}\mathbf{B}\right)\frac{1}{n}\mathrm{Tr}\,\left(\mathbf{AD}^2\mathbf{R}\,\mathbb{E}\mathbf{H}\right)+\mathcal{O}\left(\frac{1}{n^2}\right)\ .\end{aligned}$$

It remains to apply Theorem 3, Proposition 4 and Lemma 1-(2) to the terms in the right hand side of the previous equality to conclude.

REFERENCES

[1] I.E. Telatar, "Capacity of Multi-Antenna Gaussian Channels," published in *European Transactions on Telecommunications*, vol. 10, no. 6, pp. 585–595, Nov/Dec 1999, Technical Memorandum, Bell Laboratories, Lucent Technologies, October 1995.

[2] Foschini G.J., "Layered space-time architecture for wireless communication in a fading environment when using multiple antennas," *Bell Labs Technical Journal*, vol. 1, no. 2, pp. 41–59, 1996.

[3] V.A. Marchenko and L.A. Pastur, "Distribution of Eigenvalues for Some Sets of Random Matrices," *Math. USSR – Sbornik*, vol. 1, no. 4, pp. 457–483, 1967.

[4] R.R. Müller, "A Random Matrix Model of Communication Via Antenna Arrays," *IEEE Trans. on Information Theory*, vol. 48, no. 9, pp. 2495–2506, Sept. 2002.

[5] C.-N. Chuah, D.N.C. Tse, J.M. Kahn, and R.A. Valenzuela, "Capacity Scaling in MIMO Wireless Systems Under Correlated Fading," *IEEE Trans. on Information Theory*, vol. 48, no. 3, pp. 637–650, Mar. 2002.

[6] X. Mestre, J.R. Fonollosa, and A. Pagès-Zamora, "Capacity of MIMO channels: Asymptotic Evaluation Under Correlated Fading," *IEEE Journal on Selected Areas in Communications*, vol. 21, no. 5, pp. 829–838, June 2003.

[7] A.L. Moustakas, S.H. Simon, and A.M. Sengupta, "MIMO Capacity Through Correlated Channels in the Presence of Correlated Interferers and Noise: A (Not So) Large N Analysis ," *IEEE Trans. on Information Theory*, vol. 49, no. 10, pp. 2545–2561, Oct. 2003.

[8] A.M. Tulino, A. Lozano, and S. Verdú, "Impact of Antenna Correlation on the Capacity of Multiantenna Channels," *IEEE Trans. on Information Theory*, vol. 51, no. 7, pp. 2491–2509, July 2005.

[9] 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.

[10] W. Hachem, P. Loubaton, and J. Najim, "The empirical eigenvalue distribution of a gram matrix: from independence to stationarity," *Markov Process. Related Fields*, vol. 11, no. 4, pp. 629–648, 2005.

[11] S.B. Korada and Macris N., "On the concentration of the capacity for a code division multiple access system," Nice, France, june 2007, ISIT.

[12] Shiu D.-S., G.J. Foschini, M.J. Gans, and J.M. Kahn, "Fading Correlation and its Effect on the Capacity of Multielement Antenna Systems," *IEEE Trans. on Communications*, vol. 48, no. 3, pp. 502–513, Mar. 2000.

[13] J.P. Kermoal, L. Schumacher, K.I. Pedersen, P.E. Mogensen, and F. Frederiken, "A Stochastic MIMO Radio Channel Model with Experimental Validation," *IEEE Journal on Selected Areas in Communications*, vol. 20, no. 6, pp. 1211–1225, 2002.

[14] W. Hachem, P. Loubaton, and J. Najim, "A CLT for information-theoretic statistics of gram random matrices with a given variance profile," Tech. Rep., 2007, submitted, available as arXiv:0706.0166.

[15] G. W. Anderson and O. Zeitouni, "A CLT for a band matrix model," *Probab. Theory Related Fields*, vol. 134, no. 2, pp. 283–338, 2006.

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

[17] A. Boutet de Monvel and A. Khorunzhy, "Limit theorems for random matrices," *Markov Process. Related Fields*, vol. 4, no. 2, pp. 175–197, 1998.

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

[19] L. A. Pastur, A. M. Khorunzhiĭ, and V. Yu. Vasil′chuk, "On an asymptotic property of the spectrum of the sum of one-dimensional independent random operators," *Dopov. Nats. Akad. Nauk Ukraïni*, , no. 2, pp. 27–30, 1995.

[20] L.A. Pastur, "A Simple Approach to the Global Regime of Gaussian Ensembles of Random Matrices," *Ukrainian Mathematical Journal*, vol. 57, no. 6, pp. 936–966, June 2005.

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

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

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

[24] Chatterjee S., "Fluctuations of eigenvalues and second order poincar inequalities," *to be published in Probab. Theory Related Fields, available at arXiv:0705.1224v2*, 2007.

[25] A. Guionnet and O. Zeitouni, "Concentration of the spectral measure for large matrices," *Electron. Comm. Probab.*, vol. 5, pp. 119–136 (electronic), 2000.

[26] S. Lasaulce Ph. Loubaton J. Dumont, W. Hachem and J. Najim, "On the capacity achieving covariance matrix for Rician mimo channels: An asymptotic approach," Tech. Rep., 2007, submitted.

[27] R. Horn and C. Johnson, *Matrix Analysis*, Cambridge Univ. Press, 1994.

[28] J. Glimm and A. Jaffe, *Quantum physics*, Springer-Verlag, New York, second edition, 1987, A functional integral point of view.

[29] R. M. Dudley, *Real analysis and probability*, vol. 74 of *Cambridge Studies in Advanced Mathematics*, Cambridge University Press, Cambridge, 2002, Revised reprint of the 1989 original.