Contribution à l'analyse asymptotique des grandes matrices aléatoires et applications aux communications numériques

Mémoire présenté pour l'obtention de l'habilitation à diriger des recherches

Jamal Najim

June 2, 2009

Contents

List of Figures

List of publications

Papers published or accepted in refered journals

1. J. Najim. A Cramér type theorem for weighted random variables. *Electron. J. Probab.*, 7:no. 4, 32 pp. (electronic), 2002.

2. C. Léonard and J. Najim. An extention of Sanov's theorem. Application to the Gibbs conditionning principle. to be published in *Bernoulli*, 2002.

3. J. Najim. Large deviations for independent random variables—application to Erdős-Renyi's functional law of large numbers. *ESAIM Probab. Stat.*, 9:116–142 (electronic), 2005.

4. R. Douc, A. Guillin, and J. Najim. Moderate deviations for particle filtering. *Ann. Appl. Probab.*, 15(1B):587–614, 2005.

5. W. Hachem, P. Loubaton, and J. Najim. The empirical eigenvalue distribution of a gram matrix: From independence to stationarity. *Markov Process, Related Fields.*, 11(4):629–648, 2005.

6. W. Hachem, P. Loubaton, and J. Najim. The empirical distribution of the eigenvalues of a Gram matrix with a given variance profile. *Ann. Inst. H. Poincaré Probab. Statist.*, 42(6):649–670, 2006.

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

8. M. Maïda, J. Najim, and S. Péché. Large deviations for weighted empirical mean with outliers. *Stochastic Process. Appl.*, 117(10):1373–1403, 2007.

9. W. Hachem, P. Loubaton, and J. Najim. A CLT for information-theoretic statistics of gram random matrices with a given variance profile. *Ann. Appl. Probab.*, 18(6):2071–2130, 2008.

10. W. Hachem, O. Khorunzhiy, P. Loubaton, J. Najim, and L. Pastur. A new approach for mutual information analysis of large dimensional multi-antenna chennels. *IEEE Trans. Inform. Theory*, 54(9):3987–4004, 2008.

11. A. Kammoun, M. Kharouf, W. Hachem, and J. Najim. BER and Outage probability approximations for LMMSE detectors on correlated MIMO channels. *accepted for publication in IEEE Inf. Th.*, sept. 2008.

12. A. Kammoun, M. Kharouf, W. Hachem, and J. Najim. A Central Limit Theorem for the SINR at the LMMSE estimator output for large dimensional signals. *accepted for publication in IEEE Inf. Th.*, june 2008.

Proceedings published or accepted

1. W. Hachem, P. Loubaton, and J. Najim. The Mutual Information of a MIMO channel: A survey. In *in ISCCSP*, Marrakech, 2006.

2. W. Hachem, P. Loubaton, and J. Najim. On the asymptotic analysis of mutual information of MIMO rician correlated channels. In *in ISCCSP*, Marrakech, 2006.

3. W. Hachem, P. Loubaton, and J. Najim. Sur les fluctuations de l'information mutuelle des canaux MIMO de grande dimension. In *in GRETSI*, Troyes, France, 2007.

4. J. Dumont, W. Hachem, S. Lasaulce, P. Loubaton, and J. Najim. Quelques proprits d'un approximant de l'information mutuelle des canaux MIMO de rice bicorrls. In *in GRETSI*, Troyes, France, 2007.

5. J. Dumont, W. Hachem, S. Lasaulce, P. Loubaton, and J. Najim. High snr approximations of the capacity of mimo correlated rician channels: a large system approach. In *Information Theory, 2007. ISIT 2007. IEEE International Symposium on*, pages 536–540, June 2007.

6. L.S. Cardoso, M. Debbah, P. Bianchi, and J. Najim. Cooperative spectrum sensing using random matrix theory. In *Wireless Pervasive Computing, 2008. ISWPC 2008. 3rd International Symposium on*, pages 334–338, May 2008.

7. W. Hachem, P. Loubaton, and J. Najim. On the fluctuations of the mutual information of large dimensional mimo channels. In *Information Theory Workshop, 2008. ITW '08. IEEE*, pages 436–439, May 2008.

8. A. Kammoun, M. Kharouf, W. Hachem, and J. Najim. Fluctuations of the snr at the wiener filter output for large dimensional signals. In *Signal Processing Advances in Wireless Communications, 2008. SPAWC 2008. IEEE 9th Workshop on*, pages 590–594, July 2008.

9. A. Kammoun, M. Kharouf, W. Hachem, and J. Najim. Outage probability approximation for the wiener filter sinr in mimo systems. In *Signal Processing Advances in Wireless Communications, 2008. SPAWC 2008. IEEE 9th Workshop on*, pages 171–175, July 2008.

10. L. Cardoso, P. Bianchi, M. Debbah, and J. Najim. Ecoute coopérative de spectre pour la radio cognitive. *in GRETSI*, 2009.

11. P. Bianchi, M. Debbah, M. Maïda, and J. Najim. Power analysis of some hypothesis tests for collaborative sensing. In *IEEE Workshop on Stat. Signal Processing*, Cardiff, UK, August 2009.

Submitted papers

1. J. Dumont, W. Hachem, S. Lasaulce, P. Loubaton, and J. Najim. On the capacity achieving covariance matrix for rician mimo channels: an asymptotic approach. *submitted to IEEE Inf. Th.*, 2007. `http://front.math.ucdavis.edu/0710.4051`.

2. P. Bianchi, M. Debbah, and J. Najim. Asymptotic Independence in the Spectrum of the Gaussian Unitary Ensemble. *submitted to ALEA*, march 2009. `http://front.math.ucdavis.edu/0811.0979.`

Submitted contributions to conferences

1. G. Alfano, P. Bianchi, M. Debbah, and J. Najim. Asymptotics of eigenbased collaborative sensing. *submitted to IEEE Inf. Theory Workshop*, oct. 2009.

Chapter 1

Introduction

The mid-nineties have witnessed a convergence between wireless communication and large random matrices. Although both fields were separately active, the pionneering works of Telatar, Foschini, Verdù, Shamai, Tse and others related the two subjects in an extremely convincing way. From that time, the application of large random matrices to wireless communication has grown and continues to grow. It continuously extends from its original field of application, that is broadly multiple input multiple output (MIMO) channels and code division multiple access (CDMA), to reach recent and fashionable subjects such as compression and collaborative sensing. The first result that has been successfully applied by Telatar to MIMO channels is 1967 Marčenko and Pastur celebrated result. Since, Girko's equations and Bai and Silverstein numerous results and many other existing results were quickly exploited. Nevertheless, there has been, at an early stage of this story, a need to develop specific mathematical results that were not available in the huge body of knowledge related to random matrices.

This document is devoted to the presentation of my scientific contribution in the field of random matrices and their applications to wireless communication. It roughly covers a period that extends from 2002, where I began to learn about random matrices with the help of my colleagues Walid Hachem and Philippe Loubaton, to nowadays. During this period, I also had a marginal activity on the subject of large deviations, which I have been working on extensively during my Ph-D. These efforts resulted in a series of three papers [86, 29, 77] which will not be discussed any further here. As we shall see however, my interest in large deviations has been revived recently, in the context of collaborative sensing this time.

The rest of the document is organized in four chapters whose contents are briefly described below.

Chapter 2 is entitled "a random matrix approach to wireless communication". In the first part of the chapter, I present various wireless communication models (MIMO channels, CDMA, collaborative sensing networks) that have largely motivated the introduction of random matrices within the field. The presentation is intended to a mathematical audience not necessarily familiar with the field and focuses on the key-role played by the spectrum of various Gram matrices with random entries in the description of performance indicators such as the ergodic capacity or the signal to interference plus noise ratio. The second part of the chapter gives a brief historical account on Large Random matrices. This account is by no means exhaustive[1] and intends to describe the

[1]In particular, free probability techniques are largely ignored, both from a theoretical point of view and from a practical one - see [55] for a mathematical account, [114] for applications to digital communication.

practical and theoretical background where our results take place. The third part of the chapter is devoted to the description of three simple but nonetheless far-reaching concepts. The first one is the Stieltjes transform, the second one is referred to as gaussian tools and comprises an integration by part identity and Poincaré-Nash inequality; the last one is a martingale decomposition of functionals of random matrices. For each concept, I develop simple examples, in a quick but non-rigorous way, intended to illustrate their range of application. For instance, I quickly derive Marčenko-Pastur distribution with the help of the Stieltjes transform, then with the help of Gaussian tools; and finally, I outline the proof of the Central Limit Theorem for a functional of the spectrum of a given random matrix with the help of martingale techniques. These techniques have been widely used in my work [31, 52, 53, 54, 64, 65].

Chapters 3, 4 and 5 are a guideline to my research. They all begin by a summary which should ease the reading, and are devoted to the presentation of results that have either been published in referred journals, or that are parts of submitted articles. The proofs are therefore available at some point and omitted here.

A notable exception is the last part of the last chapter (section 5.3) which is more an announcement as the proofs, being part of an ongoing work, are not yet available. I decided however to include it as it gathers my past and present interests: Large deviations, random matrices and wireless communication.

Chapter 3 is devoted to the presentation of first-order results. Consider the spectrum of a Gram matrix $M = XX^*$ and a linear functional of its eigenvalues $\frac{1}{N}\sum f(\lambda_i)$, then we describe, as precisely as possible, its limiting behaviour when both dimensions of the matrix X go to infinity at the same rate, whenever matrix X fulfills specific assumptions motivated by wireless communication such as the presence of a variance profile (the variance of each entry depends on its location in matrix X) and a mean profile (the mean of each entry depends as well on its location).

Chapter 4 is devoted to the presentation of second-order results. Considering as previously a linear functional $\frac{1}{N}\sum f(\lambda_i)$ where function f is either related to the mutual information in the case of MIMO channels (cf. section 2.1.1) or to the signal plus interference plus noise in the case of CDMA models (cf. section 2.1.2), we describe its fluctuations. There are two series of results, one when the underlying entries are gaussian (cf. section 4.1) and one where they are not (cf. sections 4.2 and 4.3). The techniques differ, so do the expressions of the variance related to the fluctuations.

In chapter 5, we focus on three specific contributions to wireless communication: The optimization of the ergodic capacity, where we build upon results of chapter 3 to optimize the mutual information of a MIMO channel. An algorithm is presented, to compute both the ergodic capacity and the maximizing covariance matrix, and its theoretical study is developed. The second application is devoted to the approximation of the signal to interference plus noise and the Bit Error Rate. Approximating this quantities by a gaussian random variable as justified by the CLT is too rough for low dimension. A way to circumvent this issue is to match the distribution with a generalized gamma distribution, the third moment of which is computed by gaussian techniques. Finally, the last application is related to collaborative sensing. In the context where a secondary wireless network scans the bandwidth of a primary wireless network to check whether it can use the ressource or not, we propose a statistical test and study its type I error and type II error with the help of random matrices techniques, and especially with the study of the large deviations of the largest eigenvalue of a Gram matrix.

Acknowlegment

Walid Hachem and Philippe Loubaton have been my closest collaborators during this period. It is extremely pleasant to learn mathematics with them and I want to express them my deepest thanks. I also have been fortunate to work with Prof. Leonid Pastur who generously shared with us his scientific expertise. More mystical, although enjoyable as well, was my encounter with Prof. Vyasheslav Girko in the very building of Université de Marne La Vallée. My sincere thanks go also to all my other collaborators and to the whole statistics team at Ecole Nationale Supérieure des Télécommunications (ENST). Eric Moulines certainly deserves special thanks as he strongly supported me from the beginning, and also advised me to work with Philippe Loubaton when he learned that I was interested in random matrices.

Chapter 2

A random matrix approach to wireless communication problems

2.1 Wireless communication models

2.1.1 Multiple-Input Multiple-Output channels

During the mid-nineties, an important paradigm has emerged in wireless communication: Instead of considering a point-to-point wireless communication channel to transmit some information, people discovered that using several antennas both at the emitting and receiving side would result in a dramatic increase of the capacity of the channel, that is the amount of information per time unit one can transmit. These channels whose study has resulted in a huge amount of articles in the electrical engineering literature are now coined as **MIMO** channels, for **M**ultiple-**I**nput **M**ultiple-**O**utput channels.

The MIMO channel. In a MIMO channel, the received signal writes:

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n},$$

where $\mathbf{x}$ is a vector of dimension t representing the transmitted vector, $\mathbf{H}$ is a $r \times t$ matrix representing the channel and $\mathbf{n}$ is a vector of dimension r representing the noise. Let us be more specific:

- There are t transmitting antennas and the j^{th} component of $\mathbf{x}$ represents the signal transmitted by the j^{th} antenna.
- Similarly, there are r receiving antennas and the i^{th} component of $\mathbf{y}$ represents the signal received by the i^{th} antenna.
- The i^{th} component of $\mathbf{n}$ represents the noise received by the i^{th} antenna.
- Finally, the entry H_{ij} of matrix $\mathbf{H}$ represents the complex path gain between transmitter j and receiver i.

In the sequel, we shall assume that each entry H_{ij} is random. This is based on the fact that between transmitting antenna j and receiving antenna i, there are a large number of independent reflected and scattered paths with random amplitudes. The gain H_{ij} appears as a complex sum of random variables and a Central Limit Theorem argument yields the fact that H_{ij} can be modelled as a complex gaussian random variable.

Channel matrix H. It is worth giving a closer look at $\mathbf{H}$. Matrix $\mathbf{H}$ will be assumed to be random and the physical characteristics of the channel will translate into properties related to the statistical distribution of $\mathbf{H}$ as we shall see. The realization of $\mathbf{H}$ will be known to the receiver while the transmitter will only have access to the distribution of $\mathbf{H}$.

In the seminal work of Telatar [111], the entries of $\mathbf{H}$ are complex gaussian random variables (r.v.), independent and identically distributed (i.i.d.), with zero-mean, independent real and imaginary part, both with variance $1/2$, denoted in the sequel by $\mathcal{C}N(0,1)$ (i.e. circular normal random variables). Quoting Telatar: "Equivalently, each entry of $\mathbf{H}$ has uniform phase and Rayleigh magnitude. This choice models a *Rayleigh fading environment* with enough separation *within* the receiving antennas and the transmitting antennas such that the fades for each transmitting-receiving antenna pair are independent."

It may happen that the assumption of sufficient separation between the antennas does not hold. In fact simulation-based studies [24] report that the observed performances are lower that the predicted ones. Thus, some correlation between the gains must be taken into account. Such an environment will be referred to as a *correlated fading environment*. In a correlated Rayleigh fading environment (the entries still being jointly gaussian but not independent any more), a popular and tractable correlation structure is the so-called Kronecker model where:

$$\mathbb{E}\left(H_{pj}H_{qk}^*\right) = T_{jk}R_{pq} \ .$$

This model accounts for a correlation structure between the transmitting antennas described by matrix $\mathbf{T}$ (for example $\mathbb{E}\left(H_{pj}H_{pk}^*\right) = T_{jk}$), a correlation structure between the receiving antennas described by matrix $\mathbf{R}$, and independent correlations at the emitting and receiving sides. Thus, matrix $\mathbf{H}$ admit the representation: $\mathbf{R}^{\frac{1}{2}}\mathbf{W}\mathbf{T}^{\frac{1}{2}}$ where $\mathbf{W}$ has i.i.d. $\mathcal{C}N(0,1)$ entries. Experimental validation of such a model can be found in [68] for instance.

Another improvement toward a realistic model is the *Rician fading channel* where the channel matrix $\mathbf{H}$ is the sum of a deterministic matrix $\mathbf{A}$ and a random one with some given correlation structure. While, as previously, the random part of the channel matrix accounts for many independent paths, the deterministic one accounts for a large and known path (often referred to as a Line of Sight component). Mixing the rician component with a Kroneker-type correlation yields the channel: $\mathbf{H} = \mathbf{A} + \mathbf{R}^{\frac{1}{2}}\mathbf{W}\mathbf{T}^{\frac{1}{2}}$ (where $\mathbf{W}$ has i.i.d. $\mathcal{C}N(0,1)$ entries), often referred to as a *separately correlated Rician fading channel*.

Measurement of the channel performances. Two scenarios are considered. In the *fast fading channel* scenario, to each transmitted vector $\mathbf{x}$ with covariance matrix $\mathbf{Q}$ corresponds a new realisation of the channel $\mathbf{H}$. The mutual information between $\mathbf{x}$ and $\mathbf{y}$ is given by

$$\mathbb{E}\log\det\left(\mathbf{I}_r + \frac{\rho}{r}\mathbf{H}\mathbf{Q}\mathbf{H}^*\right),$$

where ρ stands for the variance of the noise $\mathbf{n}$ (whose covariance matrix is $\rho\mathbf{I}_r$). The distribution of the channel being known to the transmitter, one can optimize the covariance matrix $\mathbf{Q}$, and get

the *ergodic capacity*:

$$\sup_{\mathbf{Q}\geq 0, \frac{1}{t}\mathrm{Trace}(\mathbf{Q})\leq 1} \mathbb{E}\log\det\left(\mathbf{I}_r + \frac{\rho}{r}\mathbf{H}\mathbf{Q}\mathbf{H}^*\right),$$

where the trace constraint $\frac{1}{t}\mathrm{Trace}(\mathbf{Q}) \leq 1$ corresponds to the total transmit power constraint. From an information-theoretic point of view, the ergodic capacity represents the theoretical maximum rate at which one can reliably transmit the data $\mathbf{x}$ in a fast fading channel scenario. Therefore, being able to compute the maximizing covariance matrix $\mathbf{Q}$ is a challenging problem of important practical interest.

In the *slow fading channel* scenario, the realisation of the channel $\mathbf{H}$ persists while many vectors $\mathbf{x}$ are transmitted. It is not legitimate any more to average over the realisations of the channel $\mathbf{H}$ (i.e. to take the expectation $\mathbb{E}$ with respect to $\mathbf{H}$) and the mutual information writes

$$\log\det\left(\mathbf{I}_r + \frac{\rho}{r}\mathbf{H}\mathbf{Q}\mathbf{H}^*\right).$$

It can be poor if the current realisation of $\mathbf{H}$ is bad. The communication capability of a slow fading channel is now described by the **outage probability** curve

$$\mathbb{P}_{\mathrm{out}}(R) = \inf_{\mathbf{Q}\geq 0, \frac{1}{t}\mathrm{Trace}(\mathbf{Q})\leq 1} \mathbb{P}\left(\log\det\left(\mathbf{I}_r + \frac{\rho}{r}\mathbf{H}\mathbf{Q}\mathbf{H}^*\right) < R\right) .$$

and a trade-off must be made between the desired transmitting rate R and the probability $\mathbb{P}_{\mathrm{out}}(R)$ that this rate cannot be attained.

2.1.2 Code Division Multiple Access

In this section, we shall look at specific techniques for communication over wireless channel. We look at many mobile users interested in communicating with a common wireline infrastructure for a wideband system. Unlike traditional methods such as Time-Division multiple access (TDMA) or Frequency-Division multiple access (FDMA), spread-spectrum techniques are broadband in the sens that the entire transmission bandwidth is shared between all users at all time. One form of spread-spectrum is **C**ode **D**ivision **M**ultiple **A**ccess, or **CDMA**. A simple channel model for (direct-sequence) CDMA is:

$$\mathbf{y} = \sum_{k=1}^{K} x_k \mathbf{s}_k + \mathbf{w} \quad \in \mathbb{R}^N , \tag{2.1}$$

where K is the number of users, x_k the transmitted data symbol of user k, $\mathbf{w} \sim \mathcal{N}(0, \sigma^2\mathbf{I}_N)$ is the additive gaussian noise in the channel, and $\mathbf{y}$ is the received vector. Vector $\mathbf{s}_K \in \mathbb{R}^N$ is the *signature sequence* associated with the k^{th} transmitted data symbol. These sequences are known at the receiving side. Assume moreover that $\mathbb{E}x_k = 0$ and $\mathbb{E}(x_k)^2 = p_k$, the received power of user k.

To catch the idea behind CDMA, consider a noiseless system $\mathbf{y} = \sum_{k=1}^{K} x_k\mathbf{s}_k$ with the sequences $(\mathbf{s}_k)$ being orthonormal. In this case, one can perfectly extract the transmitted data symbol x_k by observing the received vector $\mathbf{y}$. Indeed, the orthogonality of the sequences directly yields $x_k = \mathbf{s}_k^t\mathbf{y}$.

In the environment described by (2.1), a natural class of receivers is the class of linear receivers. Focusing on user 1, the estimate $\hat{x}_1$ of the transmitted symbol x_1 is of the form $\hat{x}_1 = \mathbf{c}_1^t\mathbf{y}$,

where $\mathbf{c}_1 \in \mathbb{R}^N$ is the linear receiver for user 1. For example, choosing $\mathbf{c}_1 = \mathbf{s}_1$, which allows perfect reconstruction in the absence of noise, yields the well-known *matched-filter* receiver. There are other linear receivers and one way to classify them is to evaluate their performance.

A commonly used ratio to do so is the *signal-to-interference plus noise ratio* (SINR), which (in the case of user 1) writes:

$$S_1 = \frac{(\mathbf{c}_1^t \mathbf{s}_1)^2 p_1}{(\mathbf{c}_1^t \mathbf{c}_1)^2 \sigma^2 + \sum_{i \neq 1} (\mathbf{c}_1^t \mathbf{s}_i)^2 p_i} \ .$$

The SINR is the ratio of the variance of user 1 to of noise plus interference from other users, measured at the output of the linear receiver. The optimal receiver which maximizes the output SINR is the *minimum mean-square error* (MMSE) receiver which minimizes $\mathbb{E}(x_1 - \hat{x}_1)^2$. This least squares problem has a well-known solution given by:

$$\mathbf{c}_1 \ = \ (\mathbf{S}\mathbf{D}\mathbf{S}^t + \sigma^2 \mathbf{I})^{-1} \mathbf{s}_1 \quad \text{and} \quad S_1 \ = \ p_1 \mathbf{s}_1^t (\mathbf{S}\mathbf{D}\mathbf{S}^t + \sigma^2 \mathbf{I})^{-1} \mathbf{s}_1 \tag{2.2}$$

where $\mathbf{S} = [\mathbf{s}_2, \cdots, \mathbf{s}_K]$ and $\mathbf{D} = \mathrm{diag}(p_2, \cdots, p_K)$.

2.1.3 Cooperative spectrum sensing

It is a common understanding that current mobile communication systems do not make full use of the available spectrum, either due to sparse user access or to the system's inherent defficiencies. In its simplest form *spectrum sensing* means looking for a signal in the presence of noise for a given frequency band. Although this problem has been extensively studied before, it has regained attentio now as part as the cognitive radio research effort. By *cognitive radio*, we mean techniques where systems such as mobile agents are able to opportunistically exploit their environment in order to take advantage of available spectrum leftovers.

From a cognitive radio perspective, the problem of spectrum sensing has very stringent requirements among which:

- No priori knowledge of the signal structure,
- Very fast detection of the signal (real-time context),
- Ability to detect reliably in heavily faded environments.

Consider the scenario depicted in Figure 2.1.3 in which primary users (in white) communicate to their dedicated (primary) base station. Secondary base stations $BS_1, \cdots, BS_K$ are cooperatively sensing the channel in order to identify whether the spectrum is available and to exploit the medium in this case. By cooperatively, we mean that the K base stations in the secondary system share information between them. This can be performed by transmission over a wired high speed backbone for instance.

Each secondary base station $k = 1 : K$ receives a sampled signal of dimension N: $(y_k(\ell), \ell = 1 : N)$. Stacking all these signals yields the matrix:

$$\mathbf{Y} = (y_k(\ell); k = 1 : K, \ell = 1 : N) \ ,$$

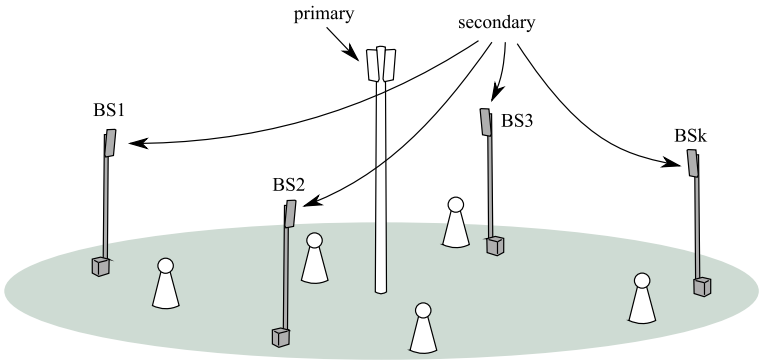

Figure 2.1: Considered scenario for cooperative spectrum sensing

of dimensions $K \times N$. If no signal is emitted from a primary base station, then:

$$y_k(\ell) = \sigma w_k(\ell) \stackrel{\triangle}{=} \hat{y}_k(\ell) \ ,$$

where $\sigma w_k(\ell) \sim \mathcal{C}N(0, \sigma^2)$ is a white noise with unknown variance σ^2. If some signal is emitted, then:

$$y_k(\ell) = h_k s(\ell) + \sigma w_k(\ell) \stackrel{\triangle}{=} \check{y}_k(\ell) \ ,$$

where h_k is the propagation channel associated to secondary base station number k and $s(\ell) \sim \mathcal{C}N(0, 1)$.

In order to decide whether a signal is emitted or not, one needs to perform a statistical test:

$$\mathbf{Y} = \hat{\mathbf{Y}} \quad \text{versus} \quad \mathbf{Y} = \check{\mathbf{Y}} \ .$$

This issue is addressed in Section 5.3.

2.2 Large random matrices come into play

2.2.1 How large are these random matrices?

In the various scenarios of wireless communications described previously, random matrices have naturally popped up. This report is based on an asymptotic study of the random matrices under investigation. By asymptotic study, we mean a study of the spectrum of the considered matrices whenever the two dimensions of the matrices under investigation are of the same order, and grow at the same pace. The first assumption is very legitimate:

- In a MIMO channel, the number of emitting antennas is of the same magnitude as the number of receiving antennas;
- In CDMA, the number of users K is comparable to the gain N of the system;
- In cooperative spectrum sensing, the number of secondary base stations is comparable to the dimension of the signal received by each station;

However, the second one may sound questionable, and indeed is.

- For a MIMO channel, quoting Tulino and Verdú [114, pp. 8-9]: "In multi antenna systems, arrays of 8 to 16 antennas would be at the forefront of what is envisioned to be feasible in the foreseeable future".
- In CDMA, channels with K symbols and gain N between 32 to 64 would be fairly typical.
- In collaborative compress sensing, 10 secondary base station, each receiving 50 samples, is a good order of magnitude.

These figures are low, especially for MIMO channels. A partial answer lies again in [114, p.9]: "Surprisingly, even quite smaller system sizes are large enough for the asymptotic limit to be an excellent approximation. Furthermore, not only do the averages of $[\cdots]$ converge to their limits surprisingly fast, but the randomness in those functionals due to the random outcome of $\mathbf{H}$ disappears extremely quickly. Naturally, such robustness has welcome consequences for operational significance of the resulting formulas."

As the entries are complex gaussian, one may expect to have direct access to the distribution of $\log\det\left(\mathbf{I}+\frac{\rho}{r}\mathbf{H}\mathbf{H}^*\right)$. It turns out that despite the small size of the system or the gaussianity of the entries, a large random matrix approach is supported by several arguments:

- The asymptotic approximations are often relatively simple closed-form formulas where the role of many parameters of the system clearly appears.
- The convergence is fast; moreover, the gaussianity of the entries fastens the convergence up to a factor proportional to the dimension of the matrix.
- Simulations match theoretical predictions for unreasonably small systems.
- Direct computations based on the exact expression of the eigenvalues of $\mathbf{H}\mathbf{H}^*$ in the gaussian case are rapidely out of reach for complex scenarios (non-separable covariance structure, Rician fading channel, etc.).

A brief survey

Law of Large Numbers. The study of the whole spectrum of large random matrices is relevant, as noticed before, whenever one wants to compute the ergodic capacity or the outage probability of a MIMO system, or the signal to interference plus noise ratio in a CDMA system.

For square matrices, this study can be traced back to the seminal work of Wigner [122, 123]. In the article *"On the distribution of the roots of certain symmetric matrices"*, Wigner derived the limiting distribution of the spectrum of a symmetric $n \times n$ matrix $\mathbf{X}_n$ with independent and identically distributed (i.i.d.) entries (up to the symmetry) and proved that the limiting distribution was the semi-circle probability distribution. Wigner proceeded in counting the limiting moments of the spectrum of the matrix, which turned out to converge toward the Catalan numbers, well-known to be the moments of the semi-circle law.

A few years later, Marčenko and Pastur [78] derived the (eponymic) limiting distribution of the spectrum of a Gram random matrix $\mathbf{X}_n\mathbf{X}_n^*$ where $\mathbf{X}_n$ is a $N \times n$ matrix whose entries are i.i.d. as the dimensions go to infinity at the same pace (i.e. $\frac{N}{n} \to c \in (0,\infty)$). They did not follow the same path as Wigner but instead studied the resolvent of the matrix and proved that the Stieltjes transform (that is the normalized trace of the resolvent) of the limiting distribution was the solution of a simple quadratic equation. Combining a systematic use of the Stieltjes transform

and computation techniques based on the gaussianity of the entries, Pastur and his collaborators developped further the technique in a series of papers [70, 92, 69, 16, 15, 89, 90, 91].

In the mid-nineties, people studying wireless network began to take advantage of theses results. In 1995, Telatar (paper published in '99 [111]) took advantage of Marčenko and Pastur's result to evaluate the ergodic capacity of a MIMO channel. In [36], Foschini also mentionned the linear growth of the capacity and based his computations on Marčenko and Pastur's result, wrongly attributed to Grenander and Silverstein [45]. In the context of CDMA, Tse and Hanly [113], and Shamai and Verdù [119] studied the performances of multiuser linear receivers for random spreading sequences in large dimension. Tse pursued the study of the fluctuations with Zeitouni in [112].

In his books [39] and [41], Girko systematically studied many models of random matrices beyond the i.i.d. case. These references soon became precious for people looking for more realistic channels, that is matrices $\mathbf{H}$ whose entries would not be i.i.d. but rather correlated or non-centered - see Müller [85], Chuah et al. [25], Mestre et al. [79], Moustakas et al. [82], Tulino et al. [116]. Girko's books are full of ready-to-use equations and beautiful results related to non-centered non-i.i.d. models but the mathematical proofs are hard to follow and extremely difficult to exploit or extend.

Silverstein, together with Bai and other collaborators, played a pivotal role in the study of large gram random matrices. Their favourite model, XTX^*, X being random with i.i.d. entries and T being deterministic enables one to consider some correlation between the entries. Unsurprisingly, many of their papers [103, 8, 30] are of constant use among electrical engineering (especially wireless communication) specialists.

In Sections 2.3.1 and 2.3.2, we quickly revisit the historical example of Marčenko and Pastur with the help of the Stieltjes transform to give a flavor of the technique.

Fluctuations. Central limit theorems have been widely studied for various models of random matrices and for various classes of linear statistics of the eigenvalues in the physics, engineering and mathematical literature.

In the mathematical literature, CLTs for Wigner matrices can be traced back to Girko [37] (see also [40]). Results for this class of matrices have also been obtained by Khorunzhy *et al.* [69], Johansson [61], Sinai and Sochnikov [105], Soshnikov [107], Cabanal-Duvillard [18]. For band matrices, let us mention the paper by Khorunzhy *et al.* [69], Boutet de Monvel and Khorunzhy [15], Guionnet [46], Anderson and Zeitouni [4]. The case of Gram matrices has been studied in Jonsson [62] and Bai and Silverstein [7]. Fluctuations for Wigner and Wishart matrices have also been studied by Chatterjee [20], and Mingo and Speicher in [80] with the help of free probability tools. A recent preprint by Pastur and Lytova [88] is of particular interest, where the authors develop tools to extend CLTs obtained with gaussian entries to models with non-gaussian entries. The major interest follows from the fact that one can rely on the body of methods (and results) developed for models with gaussian entries. The extension process yields however non-trivial computations. For a more detailed overview, the reader is referred to the introduction in [4]. In the physics literature, so-called replica methods as well as saddle-point methods have long been a popular tool to compute the moments of the limiting distributions related to the fluctuations of the statistics of the eigenvalues.

Previous results and methods have recently been exploited in the engineering literature, with the growing interest in random matrix models for wireless communications (see the seminal paper by Telatar [111] and the subsequent papers of Tse and co-workers [112], [113] - see also the monograph by Tulino and Verdu [114] and the references therein). One main interest lies in the

study of the convergence and the fluctuations of the mutual information $\frac{1}{N}\log\det\left(Y_nY_n^* + \rho I_N\right)$ for various models of matrices Y_n. General convergence results have been established by the authors in [53, 51, 52] while fluctuation results based on Bai and Silverstein [7] have been developed in Debbah and Müller [28] and Tulino and Verdu [115]. Other fluctuation results either based on the replica method or on saddle-point analysis have been developed by Moustakas, Sengupta and coauthors [82, 99], Taricco [109]. In a different fashion and extensively based on the Gaussianity of the entries, a CLT has been proved in Hachem *et al.* [50].

In section 2.3.3, we sketch the proof of the CLT for $\log\det(I + XX^*)$ in the case where X has i.i.d. entries. We rely on a technique based on a martingale decomposition of the functional of interest.

2.3 Random Matrices Techniques

2.3.1 Stieltjes transforms

Consider a hermitian matrix $\mathbf{A}_n$, of dimension $n \times n$ and eigenvalues $(\lambda_i, i = 1 : n)$. The resolvent $\mathbf{R}_n(z)$ of $\mathbf{A}_n$ is the complex-indexed matrix $\mathbf{R}_n(z) = (\mathbf{A}_n - z\mathbf{I}_n)^{-1}$ defined for any $z \in \mathbb{C}$ different from the eigenvalues od $\mathbf{A}_n$. A function-theoretic study of $z \mapsto \mathbf{R}_n(z)$ has proven to be an efficient way to study the spectrum of $\mathbf{A}_n$ (see for instance [67]).

A natural extension of the resolvent whenever the dimension of $\mathbf{A}_n$ grows to infinity is the normalized trace of the resolvent:

$$f_n(z) = \frac{1}{n}\text{Trace } \mathbf{R}_n(z) \ .$$

As we shall see, this function is the **Stieltjes transform** of the empirical probability distribution $L_n = \frac{1}{n}\sum_{i=1}^n \delta_{\lambda_i}$.

Let μ be a probability measure over $\mathbb{R}$. Its Stieltjes transform f is defined by:

$$f(z) = \int_{\mathbb{R}} \frac{\mu(d\lambda)}{\lambda - z}, \quad z \in \mathbb{C}^+ \stackrel{\triangle}{=} \{z \in \mathbb{C},\ \text{Im}\,(z) > 0\}.$$

It is now clear that:

$$f_n(z) = \frac{1}{n}\text{Trace } \mathbf{R}_n(z) = \frac{1}{n}\sum_{i=1}^n \frac{1}{z - \lambda_i} = \int \frac{L_n(d\lambda)}{\lambda - z} \ .$$

As one should expect, the Stieltjes transform carries a lot of information on the underlying probability measure: One can get back the underlying probability measure with the help of its Stieltjes transform (2.3) and one can characterize the convergence of probability measures toward the convergence of their Stieltjes transforms (2.4).

Proposition 2.1. *The following properties hold true:*

1. *Let f be the Stieltjes transform of μ, then*
   - *the function f is analytic over $\mathbb{C}^+$,*
   - *if $z \in \mathbb{C}^+$ then $f(z) \in \mathbb{C}^+$,*
   - *the function f satisfies:* $\lim_{y\to+\infty} -iy\, f(iy) = 1$.

2. *Conversely, let f be a function analytic over $\mathbb{C}^+$ such that $f(z) \in \mathbb{C}^+$ if $z \in \mathbb{C}^+$. If $\lim_{y\to+\infty} -iy\, f(iy) = 1$, then f is the Stieltjes transform of a probability measure μ and the following inversion formula holds:*

$$\mu([a,b]) = \lim_{\eta\to 0^+} \frac{1}{\pi} \int_a^b \operatorname{Im} f(\xi + \mathbf{i}\eta)\, d\xi, \tag{2.3}$$

whenever a and b are continuity points of μ.

3. *If, for $x_0 \in \mathbb{R}$, $\operatorname{Im} f(x_0) \stackrel{\triangle}{=} \lim_{z\in\mathbb{C}^+\to x_0} \operatorname{Im} f(z)$, then $x \mapsto \mu(-\infty, x]$ is differentiable at x_0 with value $\frac{1}{\pi}\operatorname{Im} f(x_0)$.*

4. *Let $\mathbb{P}_n$ and $\mathbb{P}$ be probability measures over $\mathbb{R}$ and denote by f_n and f their Stieltjes transforms. Then*

$$\Big(\forall z \in \mathbb{C}^+,\ f_n(z) \xrightarrow[n\to\infty]{} f(z)\Big) \quad \Longleftrightarrow \quad \mathbb{P}_n \xrightarrow[n\to\infty]{\mathcal{D}} \mathbb{P}. \tag{2.4}$$

where $\mathcal{D}$ stands for convergence in distribution.

In their seminal paper [78], Marčenko and Pastur have first used Stieltjes transforms techniques to characterize the (eponymic) limiting probability measure of the spectrum of $\mathbf{X}_n\mathbf{X}_n^*$ where $\mathbf{X}_n$ is a rectangular matrix with i.i.d. entries whose dimensions grow at the same pace.

To illustrate the strength of this transformation, we propose:

A quick derivation of Marčenko and Pastur's distribution. Consider a $N \times n$ matrix $\mathbf{X}_n$ with entries $X_{ij}^{(n)} = \frac{\sigma}{\sqrt{n}} Y_{ij}$, the Y_{ij}'s being i.i.d. Denote by $\mathbf{R}_n(z) = (r_{ij}(z), 1 \le i,j \le N)$ the resolvent of matrix $\mathbf{X}_n\mathbf{X}_n^T$ and by $\tilde{\mathbf{R}}_n(z)$ the co-resolvent , i.e. the resolvent of $\mathbf{X}_n^T\mathbf{X}_n$. Let $f_n(z) = \frac{1}{N}\operatorname{Trace}\mathbf{R}_n(z) = \frac{1}{N}\sum_{i=1}^N r_{ii}(z)$ and $\tilde{f}_n(z) = \frac{1}{n}\operatorname{Trace}\tilde{\mathbf{R}}_n(z)$. Assume that $c_n = \frac{N}{n} \to c \in (0,\infty)$, then:

$$\begin{aligned}
r_{ii}(z) &\stackrel{(a)}{=} -\frac{1}{z\left(1 + \xi_i \left(\mathbf{X}_n^{(i)T}\mathbf{X}_n^{(i)} - z\mathbf{I}_n\right)^{-1} \xi_i^T\right)} \\
&\stackrel{(b)}{\approx} -\frac{1}{z\left(1 + \frac{\sigma^2}{n}\operatorname{Trace}\left(\mathbf{X}_n^{(i)T}\mathbf{X}_n^{(i)} - z\mathbf{I}_n\right)^{-1}\right)} \\
&\stackrel{(c)}{\approx} -\frac{1}{z\left(1 + \frac{\sigma^2}{n}\operatorname{Trace}\left(\mathbf{X}_n^T\mathbf{X}_n - z\mathbf{I}_n\right)^{-1}\right)} \\
&= -\frac{1}{z\left(1 + \sigma^2\tilde{f}_n(z)\right)}
\end{aligned}$$

where (a) is a standard matrix identity (cf. [57]) [denote by $\mathbf{X}_n^{(i)}$ matrix $\mathbf{X}_n$ where the ith line ξ_i has been removed], (b) follows from standard results related to quadratic forms: $x\mathbf{A}x \sim_\infty \frac{1}{n}\operatorname{Trace}\mathbf{A}$ when x and $\mathbf{A}$ are independent (see for instance [6, Lemma 2.7]), and (c) follows from a rank-one perturbation argument. Summing over i and dividing by N yields:

$$f_n(z) \approx -\frac{1}{z\left(1 + \sigma^2\tilde{f}_n(z)\right)}$$

Similarly, we obtain $\tilde{f}_n(z) \approx -\frac{1}{z(1+c_n\sigma^2 f_n(z))}$. This yields the asymptotic equation:

$$f(z) = \left(-z + \frac{\sigma^2}{1 + c\sigma^2 f(z)} \right)^{-1} . \tag{2.5}$$

This quadratic equation admits two solutions. Pick up the one that is a Stieltjes transform:

$$f(z) = \frac{(1-c) - |1-c|}{2zc} + \frac{-z + |1-c|\sigma^2 + \sqrt{(z - \sigma^2(1+\sqrt{c})^2)\,(z - \sigma^2(1-\sqrt{c})^2)}}{2zc\sigma^2} ,$$

where $\sqrt{\ }$ is the branch of the complex logarithm such that[1] $z \in \mathbb{C}^+ \Rightarrow f(z) \in \mathbb{C}^+$ and let $\mathbb{P}_{\check{M}P}$ be the probability associated to f (Marčenko-Pastur distribution). Denote by $\lambda^+ = \sigma^2(1+\sqrt{c})^2$ and $\lambda^- = \sigma^2(1-\sqrt{c})^2$. If $c \leq 1$, then using Prop. 2.1-(3) yields:

$$\mathbb{P}_{\check{M}P}(d\lambda) = \frac{\sqrt{(\lambda^+ - \lambda)_+ (\lambda - \lambda^-)_+}}{2c\pi\lambda\sigma^2}\, d\lambda ,$$

where $x_+ = \sup(x, 0)$. If $c > 1$, then:

$$f(z) = \left(1 - \frac{1}{c}\right) f_1(z) + \frac{1}{c} f_2(z) \quad \text{with} \quad f_1(z) = -\frac{1}{z} .$$

Note that f_1 is the Stieltjes transform of the dirac measure at zero and that f_2 can be handled with Prop. 2.1-(3). Thus

$$\mathbb{P}_{\check{M}P}(d\lambda) = \left(1 - \frac{1}{c}\right) \delta_0(d\lambda) + \frac{1}{c} \frac{\sqrt{(\lambda^+ - \lambda)_+ (\lambda - \lambda^-)_+}}{2\pi\lambda\sigma^2}\, d\lambda .$$

2.3.2 Gaussian calculus

If one deals with random matrices with gaussian entries, then two powerful tools are available. The first one is an integration by parts formula, which enables to get identities between random quantities of interest; the second one is Poincaré-Nash inequality which yields sharp estimates form the variance of random quantities and enables one to decorrelate these quantities.

In the sequel, consider a $N \times n$ matrice $\mathbf{Y}_n = [\mathbf{y}_1, \cdots, \mathbf{y}_n]$, the $\mathbf{y}_j$'s being its columns with each entry Y_{ij} being i.i.d. $\mathcal{C}N(0, \sigma^2)$. Denote by $\mathbf{R}_n(z) = (-zI + n^{-1}\mathbf{Y}_n\mathbf{Y}_n^*)^{-1}$. The following differentiation formulas will be of interest:

$$\frac{\partial r_{pq}}{\partial Y_{ij}} = -\frac{1}{n} r_{pi} [\mathbf{Y}^*\mathbf{R}]_{jq} = -\frac{1}{n} r_{pi} [\mathbf{y}_j^*\mathbf{R}]_q \quad \text{and} \quad \frac{\partial r_{pq}}{\partial \overline{Y_{ij}}} = -\frac{1}{n} [\mathbf{R}\mathbf{Y}]_{pj} r_{iq} = -\frac{1}{n} [\mathbf{R}\mathbf{y}_j]_p r_{iq} . \tag{2.6}$$

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:

$$\mathbb{E}\left[\xi_p \Gamma(\boldsymbol{\xi})\right] = \sum_{m=1}^{M} [\boldsymbol{\Xi}]_{pm}\, \mathbb{E}\left[\frac{\partial \Gamma(\boldsymbol{\xi})}{\partial \overline{\xi_m}}\right] . \tag{2.7}$$

[1] Note that one must consider a different branch according to the fact that $z \in \mathbb{C}^+$ is such that $\Re(z)$ is below or above the average between $\sigma^2(1-\sqrt{c})^2$ and $\sigma^2(1+\sqrt{c})^2$.

This formula relies on an integration by parts and thus is referred to as the Integration by parts formula for Gaussian vectors.

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:

$$\text{var}\left(\Gamma(\boldsymbol{\xi})\right) \leq \mathbb{E}\left[\nabla_z\Gamma(\boldsymbol{\xi})^T \;\boldsymbol{\Xi}\; \overline{\nabla_z\Gamma(\boldsymbol{\xi})}\right] + \mathbb{E}\left[(\nabla_{\overline{z}}\Gamma(\boldsymbol{\xi}))^* \;\boldsymbol{\Xi}\; \nabla_{\overline{z}}\Gamma(\boldsymbol{\xi})\right] \ . \tag{2.8}$$

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 (2.7) becomes:

$$\mathbb{E}\left[Y_{ij}\Gamma(\mathbf{Y})\right] = \sigma^2\mathbb{E}\left[\frac{\partial\Gamma(\mathbf{Y})}{\partial\overline{Y_{ij}}}\right] \ , \tag{2.9}$$

while inequality (2.8) writes:

$$\text{var}\left(\Gamma(\mathbf{Y})\right) \leq \sum_{i=1}^{N}\sum_{j=1}^{n}\sigma^2\mathbb{E}\left[\left|\frac{\partial\Gamma(\mathbf{Y})}{\partial Y_{i,j}}\right|^2 + \left|\frac{\partial\Gamma(\mathbf{Y})}{\partial\overline{Y_{i,j}}}\right|^2\right] \ . \tag{2.10}$$

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

Proposition 2.2. *Let $\mathbf{A}_n$ be a $N \times N$ real diagonal matrix whose spectral norm is uniformly bounded in n. Then*

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

Proof. We apply inequality (2.10) to the function $\Gamma(\mathbf{Y}) = \frac{1}{n}\text{Tr}\mathbf{AR}$. Using (2.6), we have

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

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:

$$\begin{aligned}
\text{var}\,\Gamma(\mathbf{Y}) &\leq \frac{2}{n^4}\sum_{i=1}^{N}\sum_{j=1}^{n}\sigma^2\mathbb{E}\left|\left[\boldsymbol{y}_j^*\mathbf{HAH}\right]_i\right|^2 \\
&= \frac{2}{n^4}\sum_{j=1}^{n}\sigma^2\mathbb{E}\left(\mathbf{y}_j^*\mathbf{RAR^2AR}\mathbf{y}_j\right) \\
&= \frac{2\sigma^2}{n^3}\mathbb{E}\,\text{Tr}\left(\mathbf{RAR^2AR}\frac{\mathbf{YY}^*}{n}\right) \\
&\overset{(a)}{\leq} \frac{2\sigma^2}{n^3}\mathbb{E}\left\{\|\mathbf{R}\|^4\|\mathbf{A}\|^2\;\text{Tr}\left(\frac{\mathbf{YY}^*}{n}\right)\right\} \overset{(b)}{\leq} \frac{2A^2\sigma^2}{n^3}\mathbb{E}\,\text{Tr}\left(\frac{\mathbf{YY}^*}{n}\right) \leq \frac{K}{n^2} \ ,
\end{aligned}$$

where inequality (a) follows from the standard inequality $\text{Tr}(AB) \leq \|B\|\text{Tr}A$, (b) follows from the fact that the spectral norm of $\mathbf{R}$ is bounded and from the boundedness of $\|\mathbf{A}_n\|$. □

Another derivation of Marčenko-Pastur distribution. Let us introduce the following notations:

$$f_n(z) = \frac{1}{N}\mathrm{Tr}\mathbf{R}_n(z), \quad g_n(z) = \mathbb{E}f_n(z), \quad \overset{\circ}{f_n} = f_n - g_n \quad \text{and} \quad c_n = \frac{N}{n} \ .$$

Recall that $\mathbf{R}_n = (-zI + n^{-1}\mathbf{Y}_n\mathbf{Y}_n^*)^{-1}$, then the resolvent identity yields:

$$\mathbf{R}_n\frac{\mathbf{Y}_n\mathbf{Y}_n^*}{n} = I - y\mathbf{R}_n \ .$$

In particular,

$$\left[\mathbf{R}_n\frac{\mathbf{Y}_n\mathbf{Y}_n^*}{n}\right]_{pp} = 1 + z\, r_{pp} \ . \tag{2.11}$$

Now $\mathbb{E}\left[\mathbf{R}_n\frac{\mathbf{Y}_n\mathbf{Y}_n^*}{n}\right]_{pp} = \sum_{i=1}^N \mathbb{E}[r_{pi}Y_{ij}\overline{Y_{pj}}]$. Using the integration by part formula, we get:

$$\mathbb{E}[r_{pi}Y_{ij}\overline{Y_{pj}}] = \sigma^2\mathbb{E}r_{ii}\delta_{ip} - \frac{\sigma^2}{n}\mathbb{E}[(\mathbf{RY})_{pj}r_{ii}\overline{Y_{pj}}]$$

where δ_{ip} is zero except if $i = p$. Now sum over i:

$$\begin{aligned}\mathbb{E}[(\mathbf{RY})_{pj}\overline{Y_{pj}}] &= \sigma^2\mathbb{E}r_{pp} - \sigma^2 c_n\mathbb{E}f_n(\mathbf{RY})_{pj}\overline{Y_{pj}} \\ &= \sigma^2\mathbb{E}r_{pp} - \sigma^2 c_n g_n\mathbb{E}(\mathbf{RY})_{pj}\overline{Y_{pj}} - \sigma^2 c_n\mathbb{E}\overset{\circ}{f_n}(\mathbf{RY})_{pj}\overline{Y_{pj}} \ .\end{aligned}$$

Using Schwarz inequality plus Poincaré-Nash inequality yields $\mathbb{E}\overset{\circ}{f_n}(\mathbf{RY})_{pj}\overline{Y_{pj}} = \mathcal{O}(n^{-2})$. Thus

$$\begin{aligned}\mathbb{E}[(\mathbf{RY})_{pj}\overline{Y_{pj}}](1 + \sigma^2 c_n g_n) &= \sigma^2\mathbb{E}r_{pp} + \mathcal{O}(n^{-2}) \\ \Rightarrow \quad \mathbb{E}\left(\mathbf{R}\frac{\mathbf{YY}^*}{n}\right)_{pp}(1 + \sigma^2 c_n g_n) &= \sigma^2\mathbb{E}r_{pp} + \mathcal{O}(n^{-2}) \ .\end{aligned}$$

Using (2.11) yields

$$(1 + z\mathbb{E}r_{pp})(1 + \sigma^2 c_n g_n) = \sigma^2\mathbb{E}r_{pp} + \mathcal{O}(n^{-2}) \ ,$$

which after summation over p yields the equation:

$$g_n = \left(-z + \frac{\sigma^2}{1 + \sigma^2 c_n g_n}\right)^{-1} + \mathcal{O}(n^{-2}) \ .$$

Asymptotically, we obtain:

$$f = \left(-z + \frac{\sigma^2}{1 + \sigma^2 cf}\right)^{-1}$$

which can be solved as previously.

2.3.3 Martingale techniques: Fluctuations of the mutual information

When considering functionals of random matrices, it is often interesting to introduce the filtration associated to the columns of the random matrix and to decompose a functional as the sum of martingale increments.

In order to illustrate the technique, we apply it to the mutual information of a Gram matrix with i.i.d. entries:

$$\mathcal{I}_n(\rho) = \frac{1}{N} \log\det\left(\rho I_N + X_n X_n^*\right) \ .$$

We will first decompose $\mathcal{I}_n - \mathbb{E}\mathcal{I}_n$ as a sum of martingale increments, and then study its fluctuations.

Denote by $\tilde{R}(z) = (-zI_n + X_n^* X_n)^{-1}$ the co-resolvent

$$\tilde{r}_{jj} = \frac{\det(X_n^{jT} X_n^j - zI_{n-1})}{\det(X_n^T X_n - zI_n)} \ , \tag{2.12}$$

where X_n^j is the $N \times (n-1)$ matrix for which column j has been removed.

Consider $\chi_n = \log\det(\rho I_N + X_n X_n^*) - \mathbb{E}\log\det(\rho I_N + X_n X_n^*)$, and introduce the filtration $\mathcal{F}_j = \sigma(x_k;\ j \le k \le n)$, x_k being column k of X_n and $\mathcal{F}_{n+1} = \{\emptyset, \Omega\}$. Then χ_n writes as the sum of martingale differences:

$$\chi_n = \sum_{j=1}^n \mathbb{E}_j \log\det(\rho I_N + X_n X_n^*) - \mathbb{E}_{j+1} \log\det(\rho I_N + X_n X_n^*).$$

As $\mathbb{E}_j \log\det(\rho I_N + X_n^j X_n^{j*}) = \mathbb{E}_{j+1} \log\det(\rho I_N + X_n^j X_n^{j*})$, we obtain, by substracting on both sides:

$$\begin{aligned}
\chi_n &= -\sum_{j=1}^n \mathbb{E}_j \log \frac{\det(\rho I_n + X_n^j X_n^{j*})}{\det(\rho I_n + X_n X_n^*)} + \mathbb{E}_{j+1} \log \frac{\det(\rho I_n + X_n^j X_n^{j*})}{\det(\rho I_n + X_n X_n^*)} \\
&= -\sum_{j=1}^n \mathbb{E}_j \log \frac{\det(\rho I_{n-1} + X_n^{j*} X_n^j)}{\det(\rho I_n + X_n^* X_n)} + \mathbb{E}_{j+1} \log \frac{\det(\rho I_{n-1} + X_n^{j*} X_n^j)}{\det(\rho I_n + X_n^* X_n)} \\
&= -\sum_{j=1}^n \mathbb{E}_j \log \tilde{R}_{jj}(-\rho) + \sum_{j=1}^n \mathbb{E}_{j+1} \log \tilde{R}_{jj}(-\rho) \\
&= \sum_{j=1}^n \mathbb{E}_j \log(1 + x_j^* R_j x_j) - \sum_{j=1}^n \mathbb{E}_{j+1} \log(1 + x_j^* R_j x_j)
\end{aligned}$$

Use now the fact that $\mathbb{E}_j \log\left(1 + \frac{\sigma^2}{n}\mathrm{Trace}\, R_j\right) = \mathbb{E}_{j+1} \log\left(1 + \frac{\sigma^2}{n}\mathrm{Trace}\, R_j\right)$, we obtain:

$$\begin{aligned}
\chi_n &= \sum_{j=1}^n \mathbb{E}_j \log\left(\frac{1 + x_j^* R_j x_j}{1 + \frac{\sigma^2}{n}\mathrm{Trace}\, R_j}\right) - \sum_{j=1}^n \mathbb{E}_{j+1} \log\left(\frac{1 + x_j^* R_j x_j}{1 + \frac{\sigma^2}{n}\mathrm{Trace}\, R_j}\right) \\
&= \sum_{j=1}^n (\mathbb{E}_j - \mathbb{E}_{j+1}) \log\left(1 + \frac{x_j^* R_j x_j - \frac{\sigma^2}{n}\mathrm{Trace}\, R_j}{1 + \frac{\sigma^2}{n}\mathrm{Trace}\, R_j}\right) \\
&\triangleq \sum_{j=1}^n (\mathbb{E}_j - \mathbb{E}_{j+1}) \log(1 + A_j) \quad \text{with} \quad A_j = \frac{x_j^* R_j x_j - \frac{\sigma^2}{n}\mathrm{Trace}\, R_j}{1 + \frac{\sigma^2}{n}\mathrm{Trace}\, R_j} \\
&\triangleq \sum_{j=1}^n \gamma_j.
\end{aligned}$$

This decomposition can be used to study the fluctuations of $\mathcal{I}_n - \mathbb{E}\mathcal{I}_n$ with the help of specific Central Limit Theorems for martingales such that:

Theorem 2.3 (CLT for martingales, Th. 35.12 in [14]). *Let $\gamma_n^{(n)}, \gamma_{n-1}^{(n)}, \ldots, \gamma_1^{(n)}$ be a martingale difference sequence with respect to the increasing filtration $\mathcal{F}_n^{(n)}, \ldots, \mathcal{F}_1^{(n)}$. Assume that there exists a sequence of real positive numbers Θ_n^2 such that*

$$\frac{1}{\Theta_n^2} \sum_{j=1}^{n} \mathbb{E}_{j+1} {\gamma_j^{(n)}}^2 \xrightarrow[n\to\infty]{\mathcal{P}} 1 \ . \tag{2.13}$$

Assume further that the Lindeberg condition holds:

$$\forall \epsilon > 0, \ \frac{1}{\Theta_n^2} \sum_{j=1}^{n} \mathbb{E}\left({\gamma_j^{(n)}}^2 \mathbf{1}_{\left|\gamma_j^{(n)}\right| \geq \epsilon \Theta_n} \right) \xrightarrow[n\to\infty]{} 0 \ .$$

Then $\Theta_n^{-1} \sum_{j=1}^{n} \gamma_j^{(n)}$ converges in distribution to $\mathcal{N}(0,1)$.

Let us sketch the proof of the following:

Theorem 2.4. *Consider*

$$\mathcal{I}_n = \frac{1}{N} \log\det(\rho I_N + X_n X_n^*).$$

Then, the following Central Limit Theorem holds true:

$$N(\mathcal{I}_N - \mathbb{E}\mathcal{I}_N) \xrightarrow[N\to\infty]{\mathcal{P}} \mathcal{N}(0, \Theta^2)$$

where the variance Θ^2 is given by:

$$\Theta^2 = \log\left(1 - \frac{c\sigma^4 \mathbf{f}(-\varsigma^2)}{(1 + c\sigma^2 \mathbf{f}(-\varsigma^2))^2}\right) + \kappa \frac{c\sigma^4 \mathbf{f}(-\varsigma^2)}{(1 + c\sigma^2 \mathbf{f}(-\varsigma^2))^2} \ ,$$

with $\mathbf{f}(z)$ being the Stieltjes transform of Marčenko-Pastur distribution and $\kappa = \mathbb{E}|Y_{11}|^4 - 2$.

Now the Lindeberg (or even the stronger Lyapounov) condition can be proved to hold quite easily. It remains to study $\sum_{j=1}^{n} \mathbb{E}_{j+1}\gamma_j^2$. We can prove that:

$$\begin{aligned}
\sum_{j=1}^{n} \mathbb{E}_{j+1}\gamma_j^2 &= \sum_{j=1}^{n} \mathbb{E}_{j+1} \left[(\mathbb{E}_j - \mathbb{E}_{j+1}) \log(1 + A_j) \right]^2 \\
&\approx \sum_{j=1}^{n} \mathbb{E}_{j+1} \left(\mathbb{E}_j A_j - \mathbb{E}_{j+1} A_j \right)^2 \\
&= \sum_{j=1}^{n} \left\{ \mathbb{E}_{j+1} (\mathbb{E}_j A_j)^2 - \mathbb{E}_{j+1} A_j^2 \right\} \\
&\approx \sum_{j=1}^{n} \mathbb{E}_{j+1} (\mathbb{E}_j A_j)^2 \ .
\end{aligned}$$

Now,

$$\begin{aligned}
\mathbb{E}_{j+1} (\mathbb{E}_j A_j)^2 &= \mathbb{E}_{j+1} \left(\frac{y_j^* R^{(j)} y_j - \frac{\sigma^2}{n} \mathrm{Trace}\, R^{(j)}}{1 + \frac{\sigma^2}{n} \mathrm{Trace}\, R^{(j)}} \right)^2 \\
&\approx \frac{1}{\left(1 + \frac{\sigma^2}{n} \mathrm{Trace}\, \mathbb{E}Q\right)^2} \mathbb{E}_{j+1} \left(y_j^* R^{(j)} y_j - \frac{\sigma^2}{n} \mathrm{Trace}\, R^{(j)} \right)^2 .
\end{aligned}$$

The square can be developped:

$$\mathbb{E}_{j+1}\left(y_j^* R^{(j)} y_j - \frac{\sigma^2}{n}\mathrm{Trace}\, R^{(j)}\right)^2 = \frac{1}{n^2}\left(\sigma^4 \mathrm{Trace}\,(\mathbb{E}_{j+1}R^{(j)})(\mathbb{E}_{j+1}R^{(j)}) + \kappa\sum_{i=1}^{N}\sigma^4\mathbb{E}_{j+1}[R^{(j)}]_{ii}^2\right).$$

Introduce $\xi_j = \frac{\sigma^4}{n}\mathrm{Trace}\,\mathbb{E}_{j+1}(Q)Q$. We can prove that:

$$\xi_j \approx \frac{\sigma^4}{n}\mathrm{Trace}\,\mathbb{E}Q^2 + \frac{n-j-1}{n}\xi_j\frac{\frac{\sigma^4}{n}\mathrm{Trace}\,\mathbb{E}Q^2}{(1+\frac{\sigma^2}{n}\mathrm{Trace}\,D\mathbb{E}Q)^2}\ ,$$

and therefore extract ξ_j

$$\xi_j = \frac{\frac{\sigma^4}{n}\mathrm{Trace}\,\mathbb{E}Q^2}{1-\frac{n-j-1}{n}\frac{\frac{\sigma^4}{n}\mathrm{Trace}\,\mathbb{E}Q^2}{(1+\frac{\sigma^2}{n}\mathrm{Trace}\,\mathbb{E}Q)^2}}$$

Letting $\frac{j}{n}\to x$ and $\xi_j \to \xi(x)$ yields the following asymptotic equation:

$$\xi(x) = \frac{c\sigma^4\mathbf{f}^2(-\varsigma^2)}{1-(1-x)\frac{c\sigma^4\mathbf{f}(-\varsigma^2)}{(1+c\sigma^2\mathbf{f}(-\varsigma^2))^2}}.$$

Finally,

$$\begin{aligned}\sum_{j=1}^{n}\mathbb{E}_{j+1}(\mathbb{E}_j A_j)^2 \quad &\approx \quad \sum_{j=1}^{n}\frac{1}{n}\frac{\xi_j}{(1+\frac{\sigma^2}{n}\mathrm{Trace}\,\mathbb{E}Q)^2} + \frac{\kappa}{n}\sum_{i=1}^{N}\sigma^4\mathbb{E}_{j+1}[R^{(j)}]_{ii}^2 \\ &\xrightarrow[n\to\infty]{} \quad \frac{1}{(1+c\sigma^2\mathbf{f})^2}\int_0^1\xi(x)dx + \kappa\frac{c\sigma^4\mathbf{f}^2}{(1+\sigma^2 c\mathbf{f})^2}.\end{aligned}$$

It remains to integrate (note that ξ has the form u'/u) to get the desired formula.

Chapter 3

First order results: Laws of Large Numbers

Summary

This chapter is mainly devoted to the description of the results of paper [53]. Partial results of paper [31] together with the results of [51] are also presented.

[31] J. Dumont, W. Hachem, S. Lasaulce, P. Loubaton, and J. Najim. On the capacity achieving covariance matrix for Rician MIMO channels: an asymptotic approach. *to be published in IEEE Inf. Th.*, submitted in 2007.

[51] W. Hachem, P. Loubaton, and J. Najim. The empirical eigenvalue distribution of a Gram matrix: From independence to stationarity. *Markov Process, Related Fields.*, 11(4):629–648, 2005.

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

Section 3.1. Let $\Sigma_n = Y_n + A_n$ be a $N \times n$ random matrix, A_n being deterministic, with columns and rows uniformly bounded in the Euclidean norm. Matrix $Y_n = (Y_{ij}^n)$ is random, its entries are given by $Y_{ij}^n = \frac{\sigma_{ij}(n)}{\sqrt{n}} X_{ij}^n$, the X_{ij}^n being independent and identically distributed, centered with unit variance and satisfying some mild moment assumption. We will refer to Y_n as a matrix with a given variance profile. The main result of [53] presented in Section 3.1 is a description of the behaviour of the spectrum of $\Sigma_n \Sigma_n^*$ under the asymptotic regime $n \to \infty$, $\frac{N}{n} \to c \in (0, \infty)$. We assert that there exists a deterministic $N \times N$ matrix-valued function $T_n(z)$ analytic in $\mathbb{C} - \mathbb{R}^+$ such that, almost surely,

$$\lim_{n \to +\infty, \frac{N}{n} \to c} \left(\frac{1}{N} \text{Trace}(\Sigma_n \Sigma_n^T - z I_N)^{-1} - \frac{1}{N} \text{Trace}\, T_n(z) \right) = 0.$$

Otherwise stated, there exists a deterministic equivalent to the empirical Stieltjes transform of the distribution of the eigenvalues of $\Sigma_n \Sigma_n^T$. For each n, the entries of matrix $T_n(z)$ are defined as the unique solutions of a certain system of nonlinear functional equations. It is also proved that $\frac{1}{N}\text{Trace}\, T_n(z)$ is the Stieltjes transform of a probability measure $\pi_n(d\lambda)$, and that for every bounded continuous function f, the following convergence holds almost surely $\frac{1}{N} \sum_{k=1}^N f(\lambda_k) - \int_0^\infty f(\lambda) \pi_n(d\lambda) \xrightarrow[n \to \infty]{} 0$, where the $(\lambda_k)_{1 \leq k \leq N}$ are the eigenvalues of $\Sigma_n \Sigma_n^T$.

Girko's work [39], although difficult to exploit from a mathematical point of view, has been a major source of inspiration. In particular, the equations defining $T_n(z)$ already appear in [53]. Our contribution is to provide clear proofs, to lower the assumptions related to matrix A_n (in order to apply the results to standard wireless channel models), and to prove that $\frac{1}{N}\text{Trace } T_n(z)$ is the Stieltjes transform of a probability measure $\pi_n(d\lambda)$.

Section 3.2. Results related to the approximation of the mutual information $C_n(\sigma^2) = \frac{1}{N}\mathbb{E}\log\det\left(I_N + \frac{\Sigma_n\Sigma_n^T}{\sigma^2}\right)$, (where σ^2 is a known parameter) are exposed here. Except the work of Taricco [110] (based on the replica method), there has been little study devoted to the mutual information of a Rician channel (that is involving non-centered matrices). We first provide a deterministic equivalent to the mutual information. Notice that $C_n(\sigma^2)$ writes:

$$C_n(\sigma^2) = \int_{\sigma^2}^{+\infty}\left(\frac{1}{\omega} - \frac{1}{N}\mathbb{E}\ \text{Tr}\left(\Sigma_n\Sigma_n^T + \omega I_N\right)^{-1}\right)d\omega\ .$$

In view of the previous results, one can expect $C_n(\sigma^2)$ to be close to $\overline{C}_n(\sigma^2) = \int_{\sigma^2}^{+\infty}\left(\frac{1}{\omega} - \frac{1}{N}\text{Tr}\, T_n(-\omega)\right)d\omega$. This is indeed true: $C_n(\sigma^2) - \overline{C}_n(\sigma^2) \xrightarrow[n\to\infty,\frac{N}{n}\to c]{} 0$. Moreover, $\overline{C}_n(\sigma^2)$ admits a completely explicit representation:

$$\overline{C}_n(\sigma^2) = \frac{1}{N}\log\det\left[\frac{\Psi(-\sigma^2)^{-1}}{\sigma^2} + A\tilde{\Psi}(-\sigma^2)A^T\right] + \frac{1}{N}\log\det\frac{\tilde{\Psi}(-\sigma^2)^{-1}}{\sigma^2} - \frac{\sigma^2}{nN}\sum_{i,j}\sigma_{ij}^2 T_{ii}(-\sigma^2)\tilde{T}_{jj}(-\sigma^2),$$

where $\Psi, \tilde{\Psi}, T, \tilde{T}$ satisfy the system of equations described in Section 3.1. This result is established in [53] and substantially extends those in [114], proved in the case of a centered matrix Σ_n with i.i.d. entries, and in [82] proved at a physical level of rigor (with the help of the replica method) for a centered matrix Σ_n with a separable variance profile i.e. $\sigma_{ij}^2 = d_i\tilde{d}_j$.

We then specialize the results to the Gaussian non-centered separable case (cf. [31]), that is in the case where the entries of matrix Y_n write: $Y_{ij}^n = \frac{d_i\tilde{d}_j}{\sqrt{n}}X_{ij}$ where the X_{ij} are i.i.d. complex gaussian variables. In this case, we are able to get a speed of convergence $C_n(\sigma^2) - \overline{C}_n(\sigma^2) = \mathcal{O}\left(N^{-2}\right)$. This rate is quite fast and partly explains the quality of approximations for small dimensions.

Section 3.3. In this section, we study the spectrum of a Gram matrix $\Sigma_n\Sigma_n^*$, where $\Sigma_n = Z_n + A_n$, A_n is deterministic with uniform assumptions related to its columns and rows, and Z_n is a Gaussian correlated random matrix such that $\text{cov}(Z_{j_1j_2}^n, Z_{j_1'j_2'}^n) = n^{-1}C(j_1 - j_1', j_2 - j_2')$. The main result states that the spectrum of $\Sigma_n\Sigma_n^*$ behaves like the spectrum of a companion matrix whose entries are independent gaussian random variables but not identically distributed. This fact has more or less been taken for granted in the electrical engineering community and is now explicitely proved. Close to our work but relying on different methods, Anderson and Zeitouni subsequently studied non-gaussian models [3] with finite-range correlation.

Here is a precise description of the main statement: Let $Z_n = (Z_{j_1j_2}^n)$ be a $N \times n$ random matrix with entries

$$Z_{j_1j_2}^n = \frac{1}{\sqrt{n}}\sum_{(k_1,k_2)\in\mathbb{Z}^2} h(k_1,k_2)U(j_1 - k_1, j_2 - k_2),$$

where $(U(j_1,j_2),\ (j_1,j_2)\in\mathbb{Z}^2)$ is a sequence of independent complex Gaussian random variables $\mathcal{C}N(0,1)$ and $(h(k_1,k_2),\ (k_1,k_2)\in\mathbb{Z}^2)$ is deterministic and satisfies $\sum_{(k_1,k_2)\in\mathbb{Z}^2}|h(k_1,k_2)| < \infty$. Then $\text{cov}(Z_{j_1j_2}^n, Z_{j_1'j_2'}^n) = n^{-1}C(j_1 - j_1', j_2 - j_2')$ where:

$$C(j_1,j_2) = \sum_{(k_1,k_2)\in\mathbb{Z}^2} h(k_1,k_2)h^*(k_1 - j_1, k_2 - j_2)\ .$$

Introduce the $p \times p$ Fourier unitary matrix $F_p = (F^p_{j_1,j_2})_{0\le j_1,j_2<p}$ defined by $F^p_{j_1,j_2} = \frac{1}{\sqrt{p}} \exp 2\mathbf{i}\pi \left(\frac{j_1 j_2}{p}\right)$, and the function:

$$\Phi(t_1,t_2) = \sum_{(\ell_1,\ell_2)\in\mathbb{Z}^2} h(\ell_1,\ell_2) e^{2\pi\mathbf{i}(\ell_1 t_1 - \ell_2 t_2)} \ .$$

Then the empirical distributions of the eigenvalues satisfy

$$F^{(Z_n+A_n)(Z_n+A_n)^*} - F^{(Y_n+F_N^* A_n F_n)(Y_n+F_N^* A_n F_n)^*} \to 0,$$

where $Y^n_{\ell_1\ell_2} = \frac{1}{\sqrt{n}} \Phi\left(\frac{\ell_1}{N}, \frac{\ell_2}{n}\right) X^n_{\ell_1\ell_2}$ where the $X^n_{\ell_1\ell_2}$ are independent $\mathcal{C}N(0,1)$ random variables. Now the later distribution has been studied previously. We are therefore able to fully describe the behaviour of the spectrum of $(Z_n + A_n)(Z_n + A_n)^*$.

3.1 Spectrum of a Gram matrix with independent entries

The model. Consider an $N \times n$ random matrix Y_n where the entries are given by

$$Y^n_{ij} = \frac{\sigma_{ij}(n)}{\sqrt{n}} X^n_{ij} \ , \tag{3.1}$$

where $(\sigma_{ij}(n),\ 1 \le i \le N, 1 \le j \le n)$ is a bounded sequence of real numbers (i.e. $\sup_n \max_{(i,j)} |\sigma_{ij}(n)| = \sigma_{\max} < +\infty$) called a variance profile and the real random variables X^n_{ij} are centered, independent and identically distributed (i.i.d.) with finite $4+\varepsilon$ moment. Consider a real deterministic $N \times n$ matrix $A_n = (A^n_{ij})$ whose columns $(\mathbf{a}^n_k)_{1\le k\le n}$ and rows $(\tilde{\mathbf{a}}^n_\ell)_{1\le \ell\le N}$ satisfy

$$\sup_{n\ge 1} \max_{k,\ell} (\|\mathbf{a}^n_k\|, \|\tilde{\mathbf{a}}^n_\ell\|) < +\infty \ , \tag{3.2}$$

where $\|\cdot\|$ stands for the Euclidean norm. Denote by $\Sigma_n = Y_n + A_n$. This model has two interesting features: The random variables are independent but not i.i.d. since the variance may vary and A_n, the centering perturbation of Y_n, can have a very general form.

About the literature. If Z_n is a zero-mean $N \times n$ random matrix, the asymptotics of the spectrum of the $N \times N$ Gram random matrices $Z_n Z_n^T$ have been widely studied: Marčenko and Pastur [78], Yin [124], Silverstein et al. [30, 102, 103] for i.i.d. entries, Girko [38], Khorunzhy et al. [69], Boutet de Monvel et al. [16] (see also Shlyakhtenko [100]) for non i.i.d. entries. In the centered case, it turns out that very often the empirical distribution of the eigenvalues converges towards a limiting distribution.

The case where matrix Z_n has non zero mean has been comparatively less studied. Among the related works, we first mention [30] which studies the eigenvalue asymptotics of the matrix $(R_n + Y_n)(R_n + Y_n)^T$ in the case where the matrices Y_n and R_n are independent random matrices, Y_n has i.i.d. entries and the empirical distribution of $R_n R_n^T$ converges to a non-random distribution. It is shown there that the eigenvalue distribution converges almost surely towards a deterministic distribution whose Stieltjes transform is uniquely defined by a certain functional equation. The case $(Y_n + \Delta_n)(Y_n + \Delta_n)^T$ where Y_n is given by (3.1) and Δ_n is deterministic pseudo-diagonal has been studied in [52] where it is shown that under suitable assumptions the eigenvalue distribution converges. In the general case $\Sigma_n = Y_n + A_n$, the convergence of the empirical distribution of the eigenvalues of $\Sigma_n \Sigma_n^T$ may fail to happen even if the variance profile exists in some limit and the spectral distribution of $A_n A_n^T$ converges.

Girko proposed in [39, Chapter 7] to study a deterministic equivalent of some functionals of the eigenvalues of $\Sigma_n \Sigma_n^T$ in the case where the following condition holds for A_n:

$$\sup_n \max_i \sum_{j=1}^{n} |A_{i,j}^n| < +\infty \quad \text{and} \quad \sup_N \max_j \sum_{i=1}^{N} |A_{i,j}^n| < +\infty \ . \tag{3.3}$$

Girko showed that the entries of the resolvent $(\Sigma_n \Sigma_n^T - zI)^{-1}$ have the same asymptotic behaviour as the entries of a certain deterministic holomorphic $N \times N$ matrix-valued function $T_n(z)$ characterized by a nonlinear system of $(n+N)$ coupled functional equations. Condition (3.3) is however rather restrictive. In particular, it does not hold in the context of wireless MIMO system in which (3.2) is actually much more relevant. We thus extend some of the results of Girko [39] to the case where A_n satisfies (3.2). For this, we do not follow the approach of Girko [39] based on the use of Cramer's rule but rather take the approach of Dozier and Silverstein [30] as a starting point. This approach not only allows to extend the result of [39] to deterministic matrices satisfying (3.2), but also provides a simpler proof. It also enables us to prove that the deterministic equivalent $\frac{1}{N}\mathrm{Tr}\, T_n(z)$ of the Stieltjes transform $\frac{1}{N}\mathrm{Tr}(\Sigma_n \Sigma_n^T - zI)^{-1}$ is itself the Stieltjes transform of a probability measure, which is a result of interest from a practical point of view.

Notations and assumptions. Let $N = N(n)$ be a sequence of integers such that $\lim_{n\to\infty} \frac{N}{n} = c \in (0, \infty)$. We denote by $\mathbf{i}$ the complex number $\sqrt{-1}$ and by $\mathrm{Im}\,(z)$ the imaginary part of $z \in \mathbb{C}$. Consider an $N \times n$ random matrix Y_n where the entries are given by

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

where X_{ij}^n and $(\sigma_{ij}(n))$ are defined below.

Assumption A-1. *The random variables* $(X_{ij}^n \ ; \ 1 \le i \le N, 1 \le j \le n \, , \, n \ge 1)$ *are real, independent and identically distributed. They are centered with* $\mathbb{E}(X_{ij}^n)^2 = 1$. *Moreover there exists* $\varepsilon > 0$ *such that:*

$$\mathbb{E}|X_{ij}^n|^{4+\varepsilon} < \infty \ ,$$

where $\mathbb{E}$ *denotes expectation.*

Assumption A-2. *There exists a non-negative finite real number* $\sigma_{\max}$ *such that the family of real numbers* $(\sigma_{ij}(n), \ 1 \le i \le N, \ 1 \le j \le n, \ n \ge 1)$ *satisfies:*

$$\sup_{n \ge 1} \max_{\substack{1 \le i \le N \\ 1 \le j \le n}} |\sigma_{ij}(n)| \le \sigma_{\max}.$$

Denote by $\mathrm{var}(Z)$ the variance of the random variable Z. Since $\mathrm{var}(Y_{ij}^n) = \sigma_{ij}^2(n)/n$, the family $(\sigma_{ij}(n))$ will be called a variance profile.

Let $A_n = (A_{ij}^n)$ be an $N \times n$ real deterministic matrix. We introduce the $N \times n$ matrix

$$\Sigma_n = Y_n + A_n.$$

For every matrix M, we will denote by M^T (resp. M^*) its transpose (resp. its Hermitian adjoint), by $\mathrm{Tr}\,(M)$ (resp. $\det(M)$) its trace (resp. its determinant if M is square) and by $F^{M\,M^T}$, the empirical distribution function of the eigenvalues of $M\,M^T$. The matrix I_n will always refer to the $n \times n$ identity.

Let A be an $n \times n$ matrix with complex entries. If A is Hermitian and positive semi-definite (that is $z^*Az \geq 0$ for every $z \in \mathbb{C}^n$), we write $A \geq 0$. If A and B are Hermitian matrices, $A \geq B$ means $A - B \geq 0$.

We define the matrix $\mathrm{Im}\,(A)$ by

$$\mathrm{Im}\,(A) = \frac{1}{2\mathbf{i}}(A - A^*).$$

Let $\mathcal{B}$ be the Borel σ-field on $\mathbb{R}$. An $n \times n$ matrix-valued measure M on $\mathcal{B}$ is a matrix-valued function on $\mathcal{B}$, the entries of which are complex measures. A positive matrix-valued measure M is such that for every $\Delta \in \mathcal{B}$, $M(\Delta) \geq 0$ (i.e. $M(\Delta)$ is a Hermitian and positive semi-definite matrix). In this case, the diagonal entries are non-negative measures and for every $z \in \mathbb{C}^n$, z^*Mz is a scalar non-negative measure.

Denote by $\|\cdot\|$ the Euclidean norm for vectors. In the case of matrices, the norm $\|\cdot\|_{\mathrm{sp}}$ will refer to the spectral norm and if Z is a complex-valued random variable, denote by $\|Z\|_p = (\mathbb{E}|Z|^p)^{\frac{1}{p}}$ for $p > 0$. If $\mathcal{X}$ is a topological space, we endow it with its Borel σ-algebra $\mathcal{B}(\mathcal{X})$ and we denote by $\mathcal{P}(\mathcal{X})$ the set of probability measures over $\mathcal{X}$.

We denote by $\mathbb{C}^- = \{z \in \mathbb{C},\ \mathrm{Im}\,(z) < 0\}$ and by $\mathbb{C}^+ = \{z \in \mathbb{C},\ \mathrm{Im}\,(z) > 0\}$.

Assumption A-3. *Denote by $\mathbf{a}_k^n$ the k^{th} column of A_n, by $\tilde{\mathbf{a}}_\ell^n$ its ℓ^{th} row and by*

$$\mathbf{a}_{\max,n} = \max\left(\|\mathbf{a}_k^n\|,\ \|\tilde{\mathbf{a}}_\ell^n\|;\ 1 \leq k \leq n,\ 1 \leq \ell \leq N\right).$$

We assume that

$$\mathbf{a}_{\max} = \sup_{n \geq 1}\ \mathbf{a}_{\max,n} < \infty.$$

If $(\alpha_1, \cdots, \alpha_k)$ is a finite sequence of real numbers, we denote by $\mathrm{diag}(\alpha_1, \cdots, \alpha_k)$ the $k \times k$ diagonal matrix whose diagonal elements are the α_i's. Let

$$D_j = \mathrm{diag}(\sigma_{ij}^2(n);\ 1 \leq i \leq N) \quad \text{and} \quad \tilde{D}_i = \mathrm{diag}(\sigma_{ij}^2(n);\ 1 \leq j \leq n). \tag{3.4}$$

We will denote by $D_j^{1/2} = \mathrm{diag}(\sigma_{ij}, i \leq N)$ and $\tilde{D}_i^{1/2} = \mathrm{diag}(\sigma_{ij}, j \leq n)$. Let μ be a probability measure over $\mathbb{R}$. Its Stieltjes transform f is defined by:

$$f(z) = \int_{\mathbb{R}} \frac{\mu(d\lambda)}{\lambda - z}, \quad z \in \mathbb{C}^+.$$

Recall that if f be a function analytic over $\mathbb{C}^+$ such that $f(z) \in \mathbb{C}^+$ if $z \in \mathbb{C}^+$. If $\lim_{y \to +\infty} -iy\, f(iy) = 1$, then f is the Stieltjes transform of a probability measure μ and the following inversion formula holds:

$$\mu([a,b]) = \lim_{\eta \to 0^+} \frac{1}{\pi} \int_a^b \mathrm{Im}\, f(\xi + \mathbf{i}\eta)\, d\xi,$$

whenever a and b are continuity points of μ.

In the sequel we shall denote by $\mathcal{S}(\mathbb{R}^+)$ the set of Stieltjes transforms of probability measures over $\mathbb{R}^+$.

The deterministic equivalent: Existence and asymptotic behaviour. Let us introduce the following resolvents:

$$\begin{aligned} Q_n(z) &= (\Sigma_n\Sigma_n^T - zI_N)^{-1} &= (q_{ij}(z))_{1\le i,j\le N}\,, \quad z\in\mathbb{C}-\mathbb{R}^+,\\ \tilde{Q}_n(z) &= (\Sigma_n^T\Sigma_n - zI_n)^{-1} &= (\tilde{q}_{ij}(z))_{1\le i,j\le n}\,, \quad z\in\mathbb{C}-\mathbb{R}^+. \end{aligned}$$

The function $f_n(z) = \frac{1}{N}\mathrm{Tr}\, Q_n(z)$ is the Stieltjes transform of the empirical distribution of the eigenvalues of $\Sigma_n\Sigma_n^T$. The aim of this paper is to get some insight on f_n by the means of a deterministic equivalent. We will often drop subscripts n. In Theorem 3.1, we prove the existence of matrix-valued deterministic functions $T(z)$ and $\tilde{T}(z)$ which satisfy a system of $N+n$ functional equations. In Theorem 3.2, we prove that asymptotically,

$$\frac{1}{N}\mathrm{Tr}\, Q(z) \sim \frac{1}{N}\mathrm{Tr}\, T(z) \quad \text{and} \quad \frac{1}{n}\mathrm{Tr}\, \tilde{Q}(z) \sim \frac{1}{n}\mathrm{Tr}\, \tilde{T}(z)$$

in a sense to be defined.

Theorem 3.1. *Let A_n be an $N\times n$ deterministic matrix. The deterministic system of $N+n$ equations:*

$$\psi_i(z) = \frac{-1}{z\left(1+\frac{1}{n}\mathrm{Tr}\,\tilde{D}_i\tilde{T}(z)\right)} \quad \text{for } 1\le i\le N, \tag{3.5}$$

$$\tilde{\psi}_j(z) = \frac{-1}{z\left(1+\frac{1}{n}\mathrm{Tr}\, D_jT(z)\right)} \quad \text{for } 1\le j\le n, \tag{3.6}$$

where

$$\begin{aligned} &\Psi(z) = \mathrm{diag}(\psi_i(z), 1\le i\le N) \quad , \quad \tilde{\Psi}(z) = \mathrm{diag}(\tilde{\psi}_j(z), 1\le j\le n)\\ &T(z) = \left(\Psi^{-1}(z) - zA\tilde{\Psi}(z)A^T\right)^{-1}, \quad \tilde{T}(z) = \left(\tilde{\Psi}^{-1}(z) - zA^T\Psi(z)A\right)^{-1}. \end{aligned} \tag{3.7}$$

admits a unique solution $(\psi_1,\ldots,\psi_N,\tilde{\psi}_1,\ldots,\tilde{\psi}_n)$ in $\mathcal{S}(\mathbb{R}^+)^{N+n}$.

Moreover, there exist a positive $N\times N$ matrix-valued measure $\mu=(\mu_{ij})$ and a positive $n\times n$ matrix-valued measure $\tilde{\mu}=(\tilde{\mu}_{ij})$ such that

$$\mu(\mathbb{R}^+) = I_N,\ \tilde{\mu}(\mathbb{R}^+) = I_n \quad \text{and} \quad T(z) = \int_{\mathbb{R}^+}\frac{\mu(d\lambda)}{\lambda - z}, \quad \tilde{T}(z) = \int_{\mathbb{R}^+}\frac{\tilde{\mu}(d\lambda)}{\lambda - z}$$

for $z\in\mathbb{C}-\mathbb{R}^+$. In particular, $\frac{1}{N}\mathrm{Tr}\ T(z)$ and $\frac{1}{n}\mathrm{Tr}\ \tilde{T}(z)$ are Stieltjes transforms of probability measures.

Proof of Theorem 3.1 relies on an iteration scheme and on properties of matrix-valued Stieltjes transforms.

Theorem 3.2. *Assume that Assumptions (A-1), (A-2) and (A-3) hold, then the following limits hold true almost everywhere:*

$$\lim_{n\to\infty,\frac{N}{n}\to c}\left(\frac{1}{N}\mathrm{Tr}\, Q(z) - \frac{1}{N}\mathrm{Tr}\, T(z)\right) = 0, \quad \forall z\in\mathbb{C}-\mathbb{R}^+,$$

$$\lim_{n\to\infty,\frac{N}{n}\to c}\left(\frac{1}{n}\mathrm{Tr}\, \tilde{Q}(z) - \frac{1}{n}\mathrm{Tr}\, \tilde{T}(z)\right) = 0, \quad \forall z\in\mathbb{C}-\mathbb{R}^+.$$

Remark 3.3 (limiting behaviour)**.** *There are two well-known cases where the empirical distribution of the eigenvalues of $\Sigma_n \Sigma_n^T$ converges towards a limit expressed in terms of its Stieltjes transform: The case where the variance profile $\sigma_{ij} = \sigma$ is a constant [30] and the case where the centering matrix A (which can be rectangular) has elements equal to zero outside the diagonal [52] (with $F^{A_n A_n^T}$ converging to a probability measure in both cases). Interestingly, one can obtain discretized versions of the limiting equations in [30] and [52] by combining (3.5)-(3.7).*

Corollary 3.4. *Assume that Assumptions (A-1), (A-2) and (A-3) hold. Denote by $\mathbb{P}_n$ and π_n the probability measures whose Stieltjes transform are respectively $\frac{1}{N}\mathrm{Tr}\, Q_n$ and $\frac{1}{N}\mathrm{Tr}\, T_n$. Then the following limit holds true almost everywhere:*

$$\int_0^\infty f(\lambda)\mathbb{P}_n(d\lambda) - \int_0^\infty f(\lambda)\pi_n(d\lambda) \xrightarrow[n\to\infty]{} 0,$$

where $f : \mathbb{R}^+ \to \mathbb{R}$ is a continuous and bounded function. The same results hold for the probability measures related to $\frac{1}{n}\mathrm{Tr}\, \tilde{Q}_n$ and $\frac{1}{n}\mathrm{Tr}\, \tilde{T}_n$.

Remark 3.5 (Concentration and Martingale arguments)**.** *If one is interested in proving that the empirical measure is close to its expectation, one can rely on concentration arguments (see e.g. [47]) at least when the entries are Gaussian, bounded or satisfy the Poincaré inequality. One can also rely on martingale arguments, regardless of the nature of the entries (as long as they are independent and satisfy some mild moment assumptions - see [39, Chapter 16], and also [31]). The purpose here is to provide a "computable" deterministic equivalent which is not expressed in terms of expectations. In fact, although expectations can be computed by Monte-Carlo methods, these methods quickly imply a huge amount of computations when the size of the matrix models increases.*

Outline of the proof of Theorem 3.2. The proof relies on the introduction of intermediate quantities which are the stochastic counterparts of Ψ, $\tilde{\Psi}$, T and $\tilde{T}$. Denote by

$$b_i(z) = \frac{-1}{z\left(1 + \frac{1}{n}\mathrm{Tr} D_i \tilde{Q}(z)\right)} \text{ for } 1 \le i \le N, \quad B(z) = \mathrm{diag}(b_i(z), 1 \le i \le N)\ , \tag{3.8}$$

$$\tilde{b}_j(z) = \frac{-1}{z\left(1 + \frac{1}{n}\mathrm{Tr} D_j Q(z)\right)} \text{ for } 1 \le j \le n, \quad \tilde{B}(z) = \mathrm{diag}(\tilde{b}_j(z), 1 \le j \le n)\ , \tag{3.9}$$

and by

$$R(z) = (B^{-1}(z) - zA\tilde{B}(z)A^T)^{-1} \quad \text{and} \quad \tilde{R}(z) = (\tilde{B}^{-1}(z) - zA^T B(z) A)^{-1}. \tag{3.10}$$

The introduction of the quantities $R, \tilde{R}, T$ and $\tilde{T}$ can be traced back to the work of Girko. We first prove that $\frac{1}{N}\mathrm{Tr}\, Q(z) \sim \frac{1}{N}\mathrm{Tr}\, R(z)$ and $\frac{1}{N}\mathrm{Tr}\, \tilde{Q}(z) \sim \frac{1}{N}\mathrm{Tr}\, \tilde{R}(z)$. These computations, quite involved, are along the same lines as the computations by Dozier and Silverstein in [30]. We then prove that $\frac{1}{N}\mathrm{Tr}\, R(z) \sim \frac{1}{N}\mathrm{Tr}\, T(z)$ and $\frac{1}{N}\mathrm{Tr}\, \tilde{R}(z) \sim \frac{1}{N}\mathrm{Tr}\, \tilde{T}(z)$.

3.2 A closed-form formula for the "log-det" functional

The mutual information is the maximum number of bits per second per Hertz per antenna that can be transmitted reliably on a MIMO channel with channel matrix H_n. It is equal to

$$C_n(\sigma^2) = \frac{1}{N}\mathbb{E}\log\det\left(I_N + \frac{H_n H_n^*}{\sigma^2}\right), \tag{3.11}$$

where σ^2 represents the variance of an additive noise corrupting the received signals. The mutual information $C_n(\sigma^2)$ is related to $\frac{1}{N}\mathrm{Tr}(H_nH_n^* + \sigma^2 I_N)^{-1}$ by the formula

$$\frac{\partial C_n}{\partial \sigma^2} = \frac{1}{N}\mathbb{E}\ \mathrm{Tr}(H_nH_n^* + \sigma^2 I_N)^{-1} - \frac{1}{\sigma^2}$$

or equivalently by

$$C_n(\sigma^2) = \int_{\sigma^2}^{+\infty} \left(\frac{1}{\omega} - \frac{1}{N}\mathbb{E}\ \mathrm{Tr}\left(H_nH_n^* + \omega I_N\right)^{-1} \right) d\omega\ ,$$

which follows from (3.11) by Fubini's theorem. In certain cases, channel matrix H_n is unitarily equivalent to $\Sigma_n = Y_n + A_n$ which has complex entries. Without loss of generality, we shall work with matrix Σ_n with real entries in order to remain consistent with the general exposition. Showing that $\frac{1}{N}\mathrm{Tr}\left(\Sigma_n\Sigma_n^T + \omega I_N\right)^{-1} \simeq \frac{1}{N}\mathrm{Tr}\, T_n(-\omega)$ for the deterministic matrix-valued function $T_n(z)$ defined in Theorem 3.1 allows one to approximate $C_n(\sigma^2)$ by $\overline{C}_n(\sigma^2) = \int_{\sigma^2}^{+\infty} \left(\frac{1}{\omega} - \frac{1}{N}\mathrm{Tr}\, T_n(-\omega)\right) d\omega$. This approximant can be written more explicitly.

Theorem 3.6. *Assume that Assumptions (A-1), (A-2) and (A-3) hold and denote by*

$$\overline{C}_n(\sigma^2) = \int_{\sigma^2}^{+\infty} \left(\frac{1}{\omega} - \frac{1}{N}\mathrm{Tr}\, T_n(-\omega) \right) d\omega\ , \tag{3.12}$$

where T is given by Theorem 3.1. Then the following limit holds true:

$$C_n(\sigma^2) - \overline{C}_n(\sigma^2) \xrightarrow[n\to+\infty,\, \frac{N}{n}\to c]{} 0,$$

where $\sigma^2 \in \mathbb{R}^+$. Moreover,

$$\begin{aligned} \overline{C}_n(\sigma^2) = \frac{1}{N}\log\det\left[\frac{\Psi(-\sigma^2)^{-1}}{\sigma^2} + A\tilde{\Psi}(-\sigma^2)A^T\right] \\ + \frac{1}{N}\log\det\frac{\tilde{\Psi}(-\sigma^2)^{-1}}{\sigma^2} - \frac{\sigma^2}{nN}\sum_{i,j}\sigma_{ij}^2 T_{ii}(-\sigma^2)\tilde{T}_{jj}(-\sigma^2). \end{aligned} \tag{3.13}$$

In certain cases, the study of the behaviour of $\overline{C}_n(\sigma^2)$ is simpler than the behaviour of $C_n(\sigma^2)$, and allows one to get some insight on the behaviour of the mutual information of certain MIMO wireless channels (see e.g. [32] for preliminary results).

Remark 3.7. *Eq. (3.13) has already been established in the zero mean case ($A_n = 0$): the centered case with no variance profile has been studied by Verdú and Shamai [119], the centered case with a variance profile by Sengupta and Mitra [99] (with a separable variance profile) and Tulino and Verdú ([114], Theorem 2.44).*

Sharper results for the Gaussian separable case. Let $H_n = Y_n + A_n$ and assume the following:

Assumption A-4. *1. Matrix Y_n writes:*

$$Y_n = \frac{1}{\sqrt{n}} D_n^{1/2} X_n \tilde{D}_n^{1/2}$$

where D_n (resp. $\tilde{D}_n$) is a $N \times N$ (resp. $n \times n$) diagonal matrix whose spectral norm is uniformly bounded, X_n is a $N \times n$ matrix with i.i.d. $\mathcal{C}N(0,1)$ entries.

2. Matrix A_n is $N \times n$ and has a uniformly bounded spectral norm.

As previously, denote by:

$$C_n(\sigma^2) = \frac{1}{N}\mathbb{E}\log\det\left(I_N + \frac{H_n H_n^*}{\sigma^2}\right),$$

Lemma 3.8. *The following system of equations admits, for every $\sigma^2 > 0$, a unique pair of solutions $(\delta, \tilde{\delta})$:*

$$\begin{cases} \delta = \dfrac{1}{n}\mathrm{Tr}\left[D\left(\sigma^2(I_N + D\tilde{\delta}) + A(I_n + \tilde{D}\delta)^{-1}A^*\right)^{-1}\right] \\ \tilde{\delta} = \dfrac{1}{n}Tr\left[\tilde{D}\left(\sigma^2(I_n + \tilde{D}\delta) + A^*(I_N + D\tilde{\delta})^{-1}A\right)^{-1}\right] \end{cases}.$$

Remark 3.9. *Interestingly, the set of $N + n$ equations in Theorem 3.1 reduces to 2 equations in the case where the variance profile is separable.*

The following theorem holds true:

Theorem 3.10. *Assume that (A-4) holds true, then:*

$$C_n(\sigma^2) - \overline{C}_n(\sigma^2) = \mathcal{O}\left(\frac{1}{t^2}\right),$$

where

$$\overline{C}_n(\sigma^2) = \log\det\left[I_n + \delta\tilde{D} + \sigma^{-2}A^*(I_N + \tilde{\delta}D)^{-1}A\right] + \log\det\left[I_N + \tilde{\delta}D\right] - \sigma^2\delta\tilde{\delta},$$

and δ and $\tilde{\delta}$ are given by the previous lemma.

3.3 From independence to stationarity

The model Let $Z_n = (Z^n_{j_1 j_2}, 0 \le j_1 < N,\ 0 \le j_2 < n)$ be a $N \times n$ random matrix with entries

$$Z^n_{j_1 j_2} = \frac{1}{\sqrt{n}} \sum_{(k_1,k_2)\in\mathbb{Z}^2} h(k_1, k_2)U(j_1 - k_1, j_2 - k_2),$$

where $(U(j_1, j_2),\ (j_1, j_2) \in \mathbb{Z}^2)$ is a sequence of independent complex Gaussian random variables (r.v.) such that $\mathbb{E}U(j_1, j_2) = 0$, $\mathbb{E}U(j_1, j_2)^2 = 0$ and $\mathbb{E}\left|U(j_1, j_2)\right|^2 = 1$, and $(h(k_1, k_2),\ (k_1, k_2) \in \mathbb{Z}^2)$ is a deterministic complex sequence satisfying

$$\sum_{(k_1,k_2)\in\mathbb{Z}^2} |h(k_1, k_2)| < \infty\ .$$

The bidimensional process $Z^n_{j_1 j_2}$ is a stationary gaussian field. Indeed, $\mathrm{cov}(Z^n_{j_1 j_2}, Z^n_{j'_1 j'_2}) = n^{-1}C(j_1 - j'_1, j_2 - j'_2)$ where

$$C(j_1, j_2) = \sum_{(k_1,k_2)\in\mathbb{Z}^2} h(k_1, k_2)h^*(k_1 - j_1, k_2 - j_2) \tag{3.14}$$

(we denote by a^* the complex conjugate of $a \in \mathbb{C}$ - we also denote by A^* the hermitian adjoint of matrix A).

Aim and motivation The purpose here is to study the asymptotic behaviour of the empirical distribution of the eigenvalues of various Gram matrices based on Z_n in the large limit $n \to \infty$, $\frac{N}{n} \to c \in (0, \infty)$. More precisely, we shall study the convergence of the spectral distribution of $(Z_n + A_n)(Z_n + A_n)^*$ where A_n is a deterministic matrix. The contribution in this section is to provide a new method to study Gram matrices based on Gaussian fields. The main idea is to approximate the matrix Z_n by a matrix $\tilde{Z}_n$ unitarily congruent to a matrix with independent but not identically distributed entries. This method will allow us to revisit the centered case $Z_n Z_n^*$, already studied by Boutet de Monvel et al. in [16] and to state asymptotic results for the spectral distribution of the non-centered case $(Z_n + A_n)(Z_n + A_n)^*$.

The motivations for such a work are twofold. First of all, we believe that this line of proof is new. Let us briefly describe the three main elements of it. The first one is a periodization scheme popular in signal processing and described as follows:

$$\tilde{Z}_n = (\tilde{Z}^n_{j_1 j_2}) \quad \text{where} \quad \tilde{Z}^n_{j_1 j_2} = \frac{1}{\sqrt{n}} \sum_{(k_1,k_2)\in\mathbb{Z}^2} h(k_1, k_2) U\left((j_1 - k_1) \bmod N, (j_2 - k_2) \bmod n\right),$$

where mod denotes modulo.

The second element is an inequality due to Bai [5] involving the Lévy distance $\mathcal{L}$ between distribution functions:

$$\mathcal{L}^4(F^{AA^*}, F^{BB^*}) \leq \frac{2}{N^2} \mathrm{Tr}(A-B)(A-B)^* \mathrm{Tr}(AA^* + BB^*), \tag{3.15}$$

where F^{AA^*} denotes the empirical distribution function of the eigenvalues of the matrix AA^* and $\mathrm{Tr}(X)$ denotes the trace of matrix X. With the help of this inequality, we shall prove that $Z_n Z_n^*$ and $\tilde{Z}_n \tilde{Z}_n^*$ have the same limiting spectral distribution.

The third element comes from the advantage of considering $\tilde{Z}_n$. In fact, $\tilde{Z}_n$ is congruent (via Fourier unitary transforms) to a random matrix with independent but not identically distributed entries. Therefore, we can (and will) rely on results established in [52] for Gram matrices with independent but not identically distributed entries.

The second motivation comes from the field of wireless communications. In a communication system employing antenna arrays at the transmitter and at the receiver sides, random matrices extracted from Gaussian fields are often good models for representing the radio communication channel. In this course, the stationary model as considered above is often a realistic channel model. The computations of popular receiver performance indexes such as Signal to Interference plus Noise Ratio or Shannon channel capacity heavily rely on the knowledge of the limiting spectral distribution of matrices of the type $Z_n Z_n^*$ (see [26],[74] and also the tutorial [114] for further references).

About the literature Various Gram matrices based on Gaussian fields have already been studied in the literature. The study of the general case $(Z_n + A_n)(Z_n + A_n)^*$ has been undertaken by Girko in [39, 41]. His approach is based on more general results valid in the case of a Gram matrix with asymptotically independent entries. In this context, Girko shows that the normalized trace of its resolvent has the same asymptotic behavior as the normalized trace of a deterministic matrix verifying a certain non-linear "canonical equation". Since no assumptions are done on the structure of A_n, there might not be any limiting spectral distribution. In the case where Z_n is a stationary field and A_n is Toeplitz, the equations have a simpler form, and depend on the spectral measure of Z_n and on the Fourier transform of the entries of A_n. Note that the Gaussianity is not necessary in this approach.

Boutet de Monvel et al. [16] have also studied Gram matrices based on stationary Gaussian

fields in the case where the matrix has the form $V_n + Z_n Z_n^*$, V_n being a deterministic Toeplitz matrix. Their line of proof is based on a direct study of the resolvent, taking advantage of the gaussianity of the entries.

Following the idea behing [51], Anderson and Zeitouni [3] extended to a non-gaussian setup (and in the case of Hermition matrices) some of the results presented here. The price to pay is a more stringent assumption related to the covariance function.

Disclaimer. We study in detail the case where the entries of matrix Z_n are complex and where A_n is not zero. In the real case, the general framework of the proof works as well if one considers the real counterpart of the Fourier unitary transforms, however the computations are more involved. We provide some details in Section 3.3. In the case where A_n is equal to zero, one would get a limiting probability distribution result (for the details, see [51]).

Assumptions and notations Let $N = N(n)$ be a sequence of integers such that

$$\lim_{n\to\infty} \frac{N(n)}{n} = c \in (0, \infty).$$

We denote by $\mathbf{i}$ the complex number $\sqrt{-1}$, by $\mathbf{1}_A(x)$ the indicator function over set A and by $\delta_{x_0}(x)$ the Dirac measure at point x_0. A sum will be equivalently written as $\sum_{k=1}^n$ or $\sum_{k=1:n}$. We denote by $\mathcal{CN}(0,1)$ the distribution of the Gaussian complex random variable U satisfying $\mathbb{E}U = 0$, $\mathbb{E}U^2 = 0$, and $\mathbb{E}\,|U|^2 = 1$ (equivalently, $U = A + \mathbf{i}B$ where A and B are real independent Gaussian r.v.'s with mean 0 and standard deviation $\frac{1}{\sqrt{2}}$ each).

Assumption A-5. *The entries* $(Z_{j_1 j_2}^n\ ,\ 0 \le j_1 < N,\, 0 \le j_2 < n\,,\, n \ge 1)$ *of the* $N \times n$ *matrix* Z_n *are random variables defined as:*

$$Z_{j_1 j_2}^n = \frac{1}{\sqrt{n}} \sum_{(k_1,k_2)\in\mathbb{Z}^2} h(k_1,k_2) U(j_1 - k_1, j_2 - k_2),$$

where $(h(k_1,k_2),\ (k_1,k_2) \in \mathbb{Z}^2)$ *is a deterministic complex sequence satisfying*

$$h_{\max} \stackrel{\triangle}{=} \sum_{(k_1,k_2)\in\mathbb{Z}^2} |h(k_1,k_2)| < \infty$$

and $(U(j_1,j_2),\ (j_1,j_2) \in \mathbb{Z}^2)$ *is a sequence of independent random variables with distribution* $\mathcal{CN}(0,1)$.

Remark 3.11. *Assumption (A-5) is a bit more restrictive than the related assumption [16], which only relies on the summability of the covariance function of the stationary process.*

For every matrix A, we denote by $F^{A\,A^*}$ the empirical distribution function of the eigenvalues of $A\,A^*$. Since we will study at the same time the limiting spectrum of the matrices $Z_n Z_n^*$ (resp. $(Z_n + A_n)(Z_n + A_n)^*$) and $Z_n^* Z_n$ (resp. $(Z_n + A_n)^*(Z_n + A_n)$), we can assume without loss of generality that $c \le 1$. We also assume for simplicity that $N \le n$.

When dealing with vectors, the norm $\|\cdot\|$ will denote the Euclidean norm. In the case of matrices, the norm $\|\cdot\|$ will refer to the spectral norm. Denote by $\mathbb{C}^+$ the set $\mathbb{C}^+ = \{z \in \mathbb{C},\ \mathrm{Im}\,(z) > 0\}$ and by $C(\mathcal{X})$ the set of bounded continuous functions over a given topological space $\mathcal{X}$ endowed with the supremum norm $\|\cdot\|_\infty$.

The asymptotic behaviour in the non centered case Following [53], it is possible to rely on Theorem 3.2 to provide a deterministic equivalent for a matrix $(Z_n + A_n)(Z_n + A_n)^*$

We first introduce the following complex-valued function $\Phi : [0,1] \times [0,1] \to \mathbb{C}$ defined by:

$$\Phi(t_1, t_2) = \sum_{(\ell_1,\ell_2)\in\mathbb{Z}^2} h(\ell_1, \ell_2) e^{2\pi \mathbf{i}(\ell_1 t_1 - \ell_2 t_2)} \tag{3.16}$$

We also introduce the $p \times p$ Fourier matrix $F_p = (F^p_{j_1,j_2})_{0\le j_1,j_2<p}$ defined by:

$$F^p_{j_1,j_2} = \frac{1}{\sqrt{p}} \exp 2\mathbf{i}\pi \left(\frac{j_1 j_2}{p}\right). \tag{3.17}$$

Note that matrix F_p is a unitary matrix.

Theorem 3.12. *Let $\Sigma_n = Z_n + B_n$ where B_n is a complex $N \times n$ matrix which satisfies (A-3). Denote by $Q(z) = (\Sigma_n\Sigma_n^* - zI_N)^{-1}$ and $\tilde{Q}_n(z) = (\Sigma_n^*\Sigma_n - zI_n)^{-1}$. Then*

$$\lim_{n\to\infty, \frac{N}{n}\to c} \left(\frac{1}{N}\mathrm{Tr}\, Q(z) - \frac{1}{N}\mathrm{Tr}\, T(z)\right) = 0 \quad \textit{and} \lim_{n\to\infty, \frac{N}{n}\to c} \left(\frac{1}{n}\mathrm{Tr}\, \tilde{Q}(z) - \frac{1}{n}\mathrm{Tr}\, \tilde{T}(z)\right) = 0$$

for every $z \in \mathbb{C} - \mathbb{R}^+$ where T and $\tilde{T}$ in Theorem 3.1 must be replaced by the following:

$$T(z) = \left(\Psi^{-1}(z) - zA\tilde{\Psi}(z)A^*\right)^{-1}, \quad \tilde{T}(z) = \left(\tilde{\Psi}^{-1}(z) - zA^*\Psi(z)A\right)^{-1},$$

and where the variance profile $\sigma_{ij}(n)$ must be replaced by $\left|\Phi\left(\frac{i}{N}, \frac{j}{n}\right)\right|$, and matrix A by $F_N B F_n^*$. The convergence holds in probability.

Sketch of proof. The proof is based on two key lemmas. The first one, Lemma 3.13, is about deconvolution; the second one, Lemma 3.14, fully relies on Bai's inequality (3.15).

We introduce the $N \times n$ matrix $\tilde{Z}_n$ whose entries are defined by

$$\tilde{Z}^n_{j_1 j_2} = \frac{1}{\sqrt{n}} \sum_{(k_1,k_2)\in\mathbb{Z}^2} h(k_1, k_2) U(j_1 - k_1 \bmod N, j_2 - k_2 \bmod n).$$

For simplicity, we shall write $\tilde{U}^n(j_1, j_2)$ instead of $U(j_1 \bmod N, j_2 \bmod n)$. Recall that $\mathcal{L}$ stands for the Lévy distance between distribution functions. The main interest in dealing with matrix $\tilde{Z}_n$ lies in the following two lemmas.

Lemma 3.13. *Consider the $N \times n$ matrix $Y_n = F_N \tilde{Z}_n F_n^*$. Then the entries $Y^n_{\ell_1\ell_2}$ of Y_n can be written*

$$Y^n_{\ell_1\ell_2} = \frac{1}{\sqrt{n}} \Phi\left(\frac{\ell_1}{N}, \frac{\ell_2}{n}\right) X^n_{\ell_1\ell_2}$$

where Φ is defined in (3.16) and the complex random variables $\{X^n_{\ell_1\ell_2}, 0 \le \ell_1 < N, 0 \le \ell_2 < n\}$ are independent with distribution $\mathcal{CN}(0,1)$.

Lemma 3.14. *Let B_n be a $N \times n$ deterministic matrix such that the sequence $\frac{1}{n}\mathrm{Tr} B_n B_n^*$ is bounded over n. Then*

$$\mathcal{L}\left(F^{(Z_n+B_n)(Z_n+B_n)^*}, F^{(\tilde{Z}_n+B_n)(\tilde{Z}_n+B_n)^*}\right) \xrightarrow[n\to\infty]{P} 0,$$

where $\xrightarrow{P}$ denotes convergence in probability.

□

Remarks on the real case In the case where the entries of matrix Z_n are given by

$$Z^n_{j_1 j_2} = \frac{1}{\sqrt{n}} \sum_{(k_1,k_2)\in\mathbb{Z}^2} h(k_1,k_2)U(j_1-k_1, j_2-k_2),$$

where $(h(k_1,k_2),\ (k_1,k_2)\in\mathbb{Z}^2)$ is a deterministic real and summable sequence and where $U(j_1,j_2)$ are real standard independent gaussian r.v.'s, the conclusion of Lemma 3.13 is no longer valid. In fact the entries of $Y_n = F_N \tilde{Z}_n F_n^*$ are far from being independent since straightforward computation yields:

$$Y^n_{\ell_1,\ell_2} = Y^{n^*}_{N-\ell_1,n-\ell_2} \quad \text{for} \quad 0<\ell_1<N \text{ and } 0<\ell_2<n.$$

We introduce the $p\times p$ orthogonal matrix $Q_p = (Q^P_{j_1 j_2})_{0\le j_1,j_2<p}$ defined by:

$$Q^p_{0,j_2} = \frac{1}{\sqrt{p}}, \quad 0\le j_2<p.$$

In the case where p is even, the entries $Q^p(j_1,j_2)$ $(j_1\ge 1)$ are defined by

$$\begin{cases} Q^p_{2j_1-1,j_2} = \sqrt{\frac{2}{p}}\cos\left(\frac{2\pi j_1 j_2}{p}\right) & \text{if} \quad 1\le j_1\le \frac{p}{2}-1, 0\le j_2<p; \\ Q^p_{2j_1,j_2} = \sqrt{\frac{2}{p}}\sin\left(\frac{2\pi j_1 j_2}{p}\right) & \text{if} \quad 1\le j_1\le \frac{p}{2}-1, 0\le j_2<p; \\ Q^p_{p-1,j_2} = \frac{(-1)^{j_2}}{\sqrt{p}} & \text{if} \quad 0\le j_2<p. \end{cases}$$

In the case where p is odd, they are defined by

$$\begin{cases} Q^p_{2j_1-1,j_2} = \sqrt{\frac{2}{p}}\cos\left(\frac{2\pi j_1 j_2}{p}\right) & \text{if} \quad 1\le j_1\le \frac{p-1}{2}, 0\le j_2<p; \\ Q^p_{2j_1,j_2} = \sqrt{\frac{2}{p}}\sin\left(\frac{2\pi j_1 j_2}{p}\right) & \text{if} \quad 1\le j_1\le \frac{p-1}{2}, 0\le j_2<p. \end{cases}$$

In the sequel, $\lfloor x\rfloor$ stands for the integer part of x. The following result is the counterpart of Lemma 3.13 in the real case.

Lemma 3.15. *Consider the $N\times n$ matrix $W_n = Q_N \tilde{Z}_n Q_n^{\mathrm{T}}$ where A^{T} is the transpose of matrix A. Then the entries $W^n_{\ell_1\ell_2}$ of W_n can be written as*

$$W^n_{\ell_1\ell_2} = \frac{1}{\sqrt{n}}\left|\Phi\left(\frac{1}{N}\left\lfloor\frac{\ell_1+1}{2}\right\rfloor, \frac{1}{n}\left\lfloor\frac{\ell_2+1}{2}\right\rfloor\right)\right| X^n_{\ell_1\ell_2}$$

where Φ is defined in (3.16) and the real random variables $\{X^n_{\ell_1\ell_2}, 0\le\ell_1<N, 0\le\ell_2<n\}$ are independent standard gaussian r.v.'s.

The proof is computationally more involved but similar in spirit to that of Theorem 3.12.

As a consequence of this lemma, Theorem 3.12 remains true.

Chapter 4

Second order analysis: Central Limit Theorems

Summary

This chapter is devoted to the description of the results of papers [50], [54] and [64]

[50] W. Hachem, O. Khorunzhiy, P. Loubaton, J. Najim, and L. Pastur. A new approach for mutual information analysis of large dimensional multi-antenna chennels. *IEEE Trans. Inform. Theory*, 54(9):3987–4004, 2008.

[54] W. Hachem, P. Loubaton, and J. Najim. A CLT for information-theoretic statistics of gram random matrices with a given variance profile. *Ann. Appl. Probab.*, 18(6):2071–2130, 2008.

[64] A. Kammoun, M. Kharouf, W. Hachem, and J. Najim. A Central Limit Theorem for the SINR at the LMMSE estimator output for large dimensional signals. *accepted for publication in IEEE Inf. Th.*, june 2008.

Section 4.1. In this section, we state a Central Limit Theorem for $I_n(\rho) = \log\det\left(\mathbf{I}_N + \rho n^{-1}\mathbf{Y}_n\mathbf{Y}_n^*\right)$, where $\mathbf{Y}_n = \mathbf{D}_n^{1/2}\mathbf{X}_n\widetilde{\mathbf{D}}_n^{1/2}$ 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, and $\mathbf{X}_n$ has i.i.d. entries with distribution $\mathcal{CN}(0,1)$. Notice that the order of the fluctuations is N. This section follows [50] where the CLT is proved.

Consider the system of equations in $(\delta, \tilde{\delta})$ and the deterministic matrices $\mathbf{T}_n$ and $\tilde{\mathbf{T}}_n$:

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

The results stated in Chapter 3 (see e.g. Section 3.2) yield that under the asymptotic regime: $n \to \infty$ and $0 < \liminf \frac{N}{n} \leq \limsup \frac{N}{n} < \infty$, we have $\mathbb{E}[I_n(\rho)] = V_n(\rho) + \mathcal{O}\left(n^{-1}\right)$ where:

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

Now, denote by $\gamma_n = \frac{1}{n}\mathrm{Tr}\mathbf{D}_n^2\mathbf{T}_n^2$ and $\tilde{\gamma}_n = \frac{1}{n}\mathrm{Tr}\widetilde{\mathbf{D}}_n^2\widetilde{\mathbf{T}}_n^2$, and by $\sigma_n^2(\rho) = -\log(1-\rho^2\gamma_n\tilde{\gamma}_n)$. Then the following convergence (in distribution) holds true:

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

The Central Limit Theorem is stated for Gaussian random variables and its proof relies on Gaussian tools such as Poincaré-Nash inequality and the integration by part formula. Interestingly, this result confirms previous computations of asymptotic first and second moments obtained at a physical level of rigor by the replica method , and fully establishes the convergence towards a Gaussian random variable. Although the results in [50] could be inferred from those in [54], we prefer to present them separately because they rely on Gaussian tools of interest on their own (cf. [31], [65]).

Section 4.2. This section is devoted to the study of the fluctuations of $\mathcal{I}_n(\rho) = \frac{1}{N}\log\det\left(Y_nY_n^* + \rho I_N\right)$, where $Y_n = (Y_{ij}^n)$ is a $N \times n$ random matrix whose entries are given by $Y_{ij}^n = \frac{\sigma_{ij}(n)}{\sqrt{n}}X_{ij}^n$, the X_{ij}'s being complex, centered and i.i.d. with unit variance. The main differences with the result in Section 4.1 lie in the fact that the $X_{ij}'s$ are no longer gaussian, and that the variance profile is no longer separable.

A Central Limit Theorem (CLT) is stated for $\mathcal{I}_n(\rho)$ whenever $n \to \infty$ and $0 < \liminf \frac{N}{n} \leq \limsup \frac{N}{n} < \infty$. In [53] (see Section 3.2), it has been proved that there exists a sequence of deterministic probability measures (π_n) such that the mathematical expectation $\mathbb{E}\mathcal{I}_n(\rho)$ satisfies:

$$\mathbb{E}\mathcal{I}_n(\rho) - \int \log(\lambda+\rho)\pi_n(\,d\lambda) \xrightarrow[n\to\infty]{} 0\ .$$

Moreover, $\int \log(\lambda+\rho)\pi_n(\,d\lambda)$ has a closed form formula and is easier to compute than $\mathbb{E}\mathcal{I}_n$ (whose evaluation would rely on massive Monte-Carlo simulations). Accordingly, we study the fluctuations of

$$\frac{1}{N}\log\det(Y_nY_n^* + \rho I_N) - \int \log(\rho+t)\pi_n(\,dt)\ ,$$

and prove that this quantity properly rescaled converges in distribution toward a Gaussian random variable.

In order to prove the CLT, we study separately the quantity $N(\mathcal{I}_n(\rho) - \mathbb{E}\mathcal{I}_n(\rho))$ from which the fluctuations arise and the quantity $N(\mathbb{E}\mathcal{I}_n(\rho) - \int \log(\lambda+\rho)\pi_n(\,d\lambda))$ which yields a bias.

As we shall see, the variance Θ_n^2 of $N(\mathcal{I}_n(\rho) - \mathbb{E}\mathcal{I}_n(\rho))$ takes a remarkably simple closed-form expression. In fact, there exists a $n\times n$ deterministic matrix A_n whose entries depend on the variance profile (σ_{ij}) and such that the variance takes the form: $\Theta_n^2 = -\log\det(I_n - A_n) + \kappa\mathrm{Tr}A_n$, where $\kappa = \mathbb{E}|X_{11}|^4 - 2$ in the fourth cumulant of the complex variable X_{11}. The CLT expresses as:

$$\frac{N}{\Theta_n}\left(\mathcal{I}_n - \mathbb{E}\mathcal{I}_n\right) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0,1).$$

Finally, there exists a deterministic quantity $\mathcal{B}_n(\rho)$ such that:

$$N\left(\mathbb{E}\mathcal{I}_n(\rho) - \int \log(\lambda+\rho)\pi_n(\,d\lambda)\right) - \mathcal{B}_n(\rho) \xrightarrow[n\to\infty]{} 0\ .$$

A special attention is devoted to the case where the variance profile is the sample of a continuous function, i.e. $\sigma_{ij} = \sigma(i/N, j/n)$ where $(x,y) \mapsto \sigma(x,y)$ is a bounded continuous function.

There are many overlaps between the results presented here and the articles from Bai and Silverstein [7] and from Anderson and Zeitouni [4]. Among the main differences: We do not need random variables with gaussian-like second and fourth moments (which is the case in [7]); we only need mild moments assumptions whereas all the moments are required in [4]; and our method of proof works both for the real and complex case while the combinatorics methods of [4] heavily rely on the fact that the variables are real. On the other hand, we only deal with one functional (the logdet) for simplicity while larger classes of functions are considered in [7, 4]; A precise comparison is made in Section 4.2.

Section 4.3. This section is devoted to the study of the fluctuations of the quadratic form $\beta_K = \mathbf{y}^* \left(\mathbf{Y}\mathbf{Y}^* + \rho \mathbf{I}_N\right)^{-1} \mathbf{y}$, where $\mathbf{y}$ is a $N \times 1$ vector, $\mathbf{Y}$ is a $N \times K$ matrix, both extracted from matrix $\mathbf{\Sigma} = [\mathbf{y}\ \mathbf{Y}]$:

$$\mathbf{\Sigma} = \left(\Sigma_{nk}\right)_{n=1,k=0}^{N,K} = \left(\frac{\sigma_{nk}}{\sqrt{K}} W_{nk}\right)_{n=1,k=0}^{N,K}.$$

The complex random variables W_{nk} are i.i.d. with $\mathbb{E}W_{nk} = 0$, $\mathbb{E}W_{nk}^2 = 0$ and $\mathbb{E}|W_{nk}|^2 = 1$ and where $(\sigma_{nk}^2;\ 1 \leq n \leq N;\ 0 \leq k \leq K)$ is an array of real numbers.

The study of such a quadratic form arises from performance analysis of large dimensional signals. If one considers the communication model $\mathbf{r} = \mathbf{\Sigma}\mathbf{s} + \mathbf{n}$ (where $\mathbf{s}$, $\mathbf{n}$, $\mathbf{r}$ and $\mathbf{\Sigma}$ are respectively the transmitted signal, white noise, received signal and channel - see Section 2.1.2 for more details), then β is the Signal to Interference plus noise (SINR) associated to the Linear Minimum Mean Square Estimator (LMMSE) which is known to be the best (that is the one with lowest SINR) estimator in the class of linear estimators.

Matrix $\mathbf{\Sigma}$ is centered with a given variance profile and therefore the study of its asymptotic behaviour falls into the framework developed in Section 3.2. We denote by $\overline{\beta}_K$ the deterministic equivalent of β_K. We can now state the Central Limit Theorem:

$$\frac{\sqrt{K}}{\Theta_K}\left(\beta_K - \overline{\beta}_K\right) \xrightarrow[K\to\infty]{} \mathcal{N}(0,1)$$

where Θ_K^2 depends on the variance profile and the fourth moment of the entries.

Although simpler than the study of the mutual information (as exposed in Section 4.2), the study of the quadratic form β_K bears interesting features: The order of the fluctuations $\sqrt{K}$ is due to the random vector $\mathbf{y}$ (in fact, one could prove that the fluctuations of $\frac{1}{K}\mathrm{Trace}\left(\mathbf{Y}\mathbf{Y}^* + \rho\mathbf{I}_N\right)^{-1}$ are of order K). Matrix $\mathbf{A}$ appearing in the variance of the mutual information (see Section 4.2) also appears in the expression of the variance for the quadratic form. As the random variables are not gaussian, the variance depends upon $\mathbb{E}|W_{01}|^4$, which is expected. The results presented here generalize and simplify Tse and Zeitouni's result [112] and Pan et al. results [87].

4.1 A CLT for the mutual information (I)

This section is devoted to the study of the fluctuations of the mutual information in the case where the entries are gaussian and the variance profile is separable.

Introduction 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, consider the classical transmission model $\mathbf{y} = \mathbf{G}\mathbf{x} + \mathbf{z}$, where $\mathbf{y}$ is the received signal, $\mathbf{x}$ is the vector of transmitted symbols, $\mathbf{z}$ is a complex white Gaussian noise, and $\mathbf{G}$ is 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. In this section, the emphasis is put on 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 [85], [25], [79], [82] and [116], 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 [53], 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 [51].

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 [27], [68]. This model has been studied by Chuah et al. in [25]. 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. [116] 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 not necessarily identically distributed entries, a close model as we shall see in a moment.

In [82], 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 $\mathbb{E}[I(\rho)]$ 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. We propose here a new method to study the convergence of $\mathbb{E}I(\rho)$ and the fluctuations of $I(\rho)$. Beside recovering the results in [82] 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 [54]). 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 [4], [7], [15].

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 nature of the entries of Matrix $\mathbf{G}$, two simple ingredients are available. The first one is an Integration by parts formula that provides an expression for the expectation of certain functionals of Gaussian vectors. This formula has been widely used in RMT [70, 92, 93]. The second ingredient is Poincaré-Nash inequality that bounds the variance of functionals of Gaussian vectors. Although well-known [23, 58], its application to RMT is fairly recent [21], [93] (see also [20] and [47] 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 $\mathbb{E}[I(\rho)] - V(\rho)$ (note that the Gaussian structure enters in two places: First the reduction to matrices with independent entries and varying variance and then integration by part and Poincaré-Nash bounds for the variance of relevant spectral characteristics). 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 [31] in the context of Ricean MIMO channels).

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

$$0 < \quad \ell^- = \liminf_{n\to\infty} \frac{N(n)}{n} \quad \leq \quad \ell^+ = \limsup_{n\to\infty} \frac{N(n)}{n} \quad < \infty \,, \tag{4.1}$$

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 here 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 with 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:

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

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.

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

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.

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, 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.

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 $\left(d_i^{(n)},\ 1 \le i \le N,\ n \ge 1\right)$ and $\left(\tilde{d}_j^{(n)}, 1 \le j \le n,\ n \ge 1\right)$.

(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

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

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

Proposition 4.1. *For every n, the system of equations in $(\delta, \tilde{\delta})$*

$$\left\{ \begin{array}{rcl} \delta & = & \frac{1}{n}\mathrm{Tr}\mathbf{D}_n(\mathbf{I} + t\tilde{\delta}\mathbf{D}_n)^{-1} \\ \tilde{\delta} & = & \frac{1}{n}\mathrm{Tr}\widetilde{\mathbf{D}}_n(\mathbf{I} + t\delta\widetilde{\mathbf{D}}_n)^{-1} \end{array} \right. \tag{4.2}$$

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

$$\delta_n(t) = \int_{\mathbb{R}^+} \frac{\mu_n(d\lambda)}{1 + t\lambda} \qquad \text{and} \qquad \tilde{\delta}_n(t) = \int_{\mathbb{R}^+} \frac{\tilde{\mu}_n(d\lambda)}{1 + t\lambda}\ , \tag{4.3}$$

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

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 (4.2). We finally introduce the following quantities which are required to express the fluctuations of $I_n(\rho)$:

$$\left\{ \begin{array}{l} \gamma_n(t) = \frac{1}{n}\mathrm{Tr}\mathbf{D}_n^2\mathbf{T}_n^2(t) \\ \tilde{\gamma}_n(t) = \frac{1}{n}\mathrm{Tr}\widetilde{\mathbf{D}}_n^2\widetilde{\mathbf{T}}_n^2(t) \end{array} \right. . \tag{4.4}$$

Proposition 4.2. *Assume that Assumptions* **(A1)** *and* **(A2)** *hold and denote by*

$$\sigma_n^2(t) = -\log\left(1 - t^2\gamma_n(t)\tilde{\gamma}_n(t)\right), \quad t > 0 \tag{4.5}$$

where $\gamma_n(t)$ and $\tilde{\gamma}_n(t)$ are given by (4.4). 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

$$0 < m_t^2 \leq \inf_n \sigma_n^2(t) \leq \sup_n \sigma_n^2(t) \leq M_t^2 < \infty \quad \text{for} \quad t > 0\ . \tag{4.6}$$

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

Statement of the main results. We now state the main results. Theorem 4.3 describes the first order approximation of the mutual information $I_n(\rho)$ while Theorem 4.4 describes its fluctuations when centered with respect to its first order approximation. Theorem 4.3 is a consequence of results in [53] and [31].

Theorem 4.3. *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*

$$\mathbb{E}[I_n(\rho)] = V_n(\rho) + \mathcal{O}\left(\frac{1}{n}\right) \tag{4.7}$$

as $n, N \to \infty$ *(in the sense of* (4.1)*), where*

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

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 4.4. *Assume that the setting of Theorem 4.3 holds and let*

$$\sigma_n^2(\rho) = -\log\left(1 - \rho^2\gamma_n(\rho)\tilde{\gamma}_n(\rho)\right) \ ,$$

where $\gamma_n(\rho)$ *and* $\tilde{\gamma}_n(\rho)$ *are defined in* (4.4). *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.*

4.2 A CLT for the mutual information (II)

This section is devoted to the study of the fluctuations of the mutual information in the case where the entries are non-gaussian and the variance profile is non-separable.

Introduction. Consider a $N \times n$ random matrix $Y_n = (Y_{ij}^n)$ whose entries are given by

$$Y_{ij}^n = \frac{\sigma_{ij}(n)}{\sqrt{n}} X_{ij}^n \ , \tag{4.8}$$

where $(\sigma_{ij}(n),\ 1 \le i \le N,\ 1 \le j \le n)$ is a uniformly bounded sequence of real numbers, and the random variables X_{ij}^n are complex, centered, independent and identically distributed (i.i.d.) with unit variance and finite 8^{th} moment. Consider the following linear statistics of the eigenvalues:

$$\mathcal{I}_n(\rho) = \frac{1}{N}\log\det\left(Y_nY_n^* + \rho I_N\right) = \frac{1}{N}\sum_{i=1}^{N}\log(\lambda_i + \rho)$$

where I_N is the $N \times N$ identity matrix, $\rho > 0$ is a given parameter and the λ_i's are the eigenvalues of matrix $Y_nY_n^*$. This functional known as the mutual information for multiple antenna radio channels is very popular in wireless communication. Understanding its fluctuations and in particular being able to approximate its standard deviation is of major interest for various purposes such as for instance the computation of the so-called outage probability.

The purpose of this section is to state a Central Limit Theorem (CLT) for $\mathcal{I}_n(\rho)$ whenever $n \to \infty$ and $\frac{N}{n} \to c \in (0, \infty)$.

In [53], it has been proved that there exists a sequence of deterministic probability measures (π_n) such that the mathematical expectation $\mathbb{E}\mathcal{I}_n(\rho)$ satisfies:

$$\mathbb{E}\mathcal{I}_n(\rho) - \int \log(\lambda + \rho)\pi_n(\, d\lambda) \xrightarrow[n\to\infty]{} 0 \ .$$

Moreover, $\int \log(\lambda+\rho)\pi_n(\, d\lambda)$ has a closed form formula (see Section 3.2) and is easier to compute than $\mathbb{E}\mathcal{I}_n$ (whose evaluation would rely on massive Monte-Carlo simulations). Accordingly, we study in this article the fluctuations of

$$\frac{1}{N}\log\det(Y_nY_n^* + \rho I_N) - \int \log(\rho + t)\pi_n(\, dt) \ ,$$

and prove that this quantity properly rescaled converges in distribution toward a Gaussian random variable. Although phrased differently, such a centering procedure relying on a deterministic equivalent is used in [4] and [7].

In order to prove the CLT, we study separately the quantity $N(\mathcal{I}_n(\rho) - \mathbb{E}\mathcal{I}_n(\rho))$ from which the fluctuations arise and the quantity $N(\mathbb{E}\mathcal{I}_n(\rho) - \int \log(\lambda+\rho)\pi_n(\, d\lambda))$ which yields a bias.

We will see that the variance Θ_n^2 of $N(\mathcal{I}_n(\rho) - \mathbb{E}\mathcal{I}_n(\rho))$ takes a remarkably simple closed-form expression. In fact, there exists a $n \times n$ deterministic matrix A_n (described in Theorem 4.9) whose entries depend on the variance profile (σ_{ij}) such that the variance takes the form:

$$\Theta_n^2 = -\log\det(I_n - A_n) + \kappa \mathrm{Tr} A_n,$$

where $\kappa = \mathbb{E}|X_{11}|^4 - 2$ in the fourth cumulant of the complex variable X_{11} and the CLT expresses as:

$$\frac{N}{\Theta_n}\left(\mathcal{I}_n - \mathbb{E}\mathcal{I}_n\right) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0,1).$$

The presence in the variance of a term directly depending on the cumulant of the variable X_{11} ($\kappa = \mathbb{E}X_{11}^4 - 3\mathbb{E}X_{11}^2$ if X_{11} is real; $\kappa = \mathbb{E}|X_{11}|^4 - 2\mathbb{E}|X_{11}|^2$ if X_{11} is complex) can be traced back to the article by Khorunzhy *et al.* [69, Formula (I.15)] and also appears in the recent paper by Anderson and Zeitouni [4]. In the case where $\kappa = 0$ (which happens if X_{ij} is a complex Gaussian random variable for instance), the variance has the log-form $\Theta_n^2 = \log\det(I_n - A_n)$. This has already been noticed for different models in the engineering literature by Moustakas *et al.* [82], Taricco [109]. See also Hachem *et al.* in [49].

Finally, there exists a deterministic quantity $\mathcal{B}_n(\rho)$ (described in Theorem 4.11) such that:

$$N\left(\mathbb{E}\mathcal{I}_n(\rho) - \int \log(\lambda+\rho)\pi_n(\, d\lambda)\right) - \mathcal{B}_n(\rho) \xrightarrow[n\to\infty]{} 0 \ .$$

If $\kappa = 0$, then $\mathcal{B}_n(\rho) = 0$ and there is no bias in the CLT.

About the literature - comparison with existing work. Central limit theorems have been widely studied for various models of random matrices and for various classes of linear statistics of the eigenvalues in the physics, engineering and mathematical literature.

In the mathematical literature, CLTs for Wigner matrices can be traced back to Girko [37] (see also [40]). Results for this class of matrices have also been obtained by Khorunzhy *et al.*

[69], Johansson [61], Sinai and Sochnikov [105], Soshnikov [107], Cabanal-Duvillard [18]. For band matrices, let us mention the paper by Khorunzhy *et al.* [69], Boutet de Monvel and Khorunzhy [15], Guionnet [46], Anderson and Zeitouni [4]. The case of Gram matrices has been studied in Jonsson [62] and Bai and Silverstein [7]. Fluctuations for Wigner and Wishart matrices have also been studied by Mingo and Speicher in [80] with the help of free probability tools. For a more detailed overview, the reader is referred to the introduction in [4]. In the physics literature, so-called replica methods as well as saddle-point methods have long been a popular tool to compute the moments of the limiting distributions related to the fluctuations of the statistics of the eigenvalues.

Previous results and methods have recently been exploited in the engineering literature, with the growing interest in random matrix models for wireless communications (see the seminal paper by Telatar [111] and the subsequent papers of Tse and co-workers [112], [113] - see also the monograph by Tulino and Verdu [114] and the references therein). One main interest lies in the study of the convergence and the fluctuations of the mutual information $\frac{1}{N}\log\det\left(Y_nY_n^* + \rho I_N\right)$ for various models of matrices Y_n. General convergence results have been established by the authors in [53, 51, 52] while fluctuation results based on Bai and Silverstein [7] have been developed in Debbah and Müller [28] and Tulino and Verdu [115]. Other fluctuation results either based on the replica method or on saddle-point analysis have been developed by Moustakas, Sengupta and coauthors [82, 99], Taricco [109]. In a different fashion and extensively based on the Gaussianity of the entries, a CLT has been proved in Hachem *et al.* [49].

There are many overlaps between this work and other works in the literature, in particular with the paper by Bai and Silverstein [7] and the paper by Anderson and Zeitouni [4] (although this last paper is primarily devoted to band matrix models, *i.e.* symmetric matrices with a symmetric variance profile). The computation of the variance and the obtention of a closed-form formula significantly extend the results obtained in [49].

Here, we deal with complex variables which are more relevant for wireless communication applications. The case of real random variables would have led to very similar computation, the cumulant $\kappa = \mathbb{E}|X|^4 - 2$ being replaced by $\tilde{\kappa} = \mathbb{E}X^4 - 3$. In [4], Anderson and Zeitouni deal with band matrices with real variables. Due to the complex nature of the variables herein, the standard trick of considering the symmetric matrix $\left(\begin{smallmatrix} 0 & X \\ X^* & 0 \end{smallmatrix}\right)$ to study the spectral distribution of XX^* does not help and one cannot rely on the CLT in [4]. Moreover, we substantially relax the moment assumptions concerning the entries with respect to [4] where the existence of moments of all order is required [1]. In this paper, we shall only assume the finiteness of the 8^{th} moment. Bai and Silverstein [7] consider the model $T_n^{\frac{1}{2}}X_nX_n^*T_n^{\frac{1}{2}}$ where the entries of X_n are i.i.d. and have Gaussian fourth moment.

This assumption can be skipped in our framework, where a good understanding of the behaviour of the diagonal individual entries of the resolvent $(-zI_n + Y_nY_n^*)^{-1}$ enables us to deal with non-Gaussian entries.

On the other hand, it must be noticed that we establish the CLT for the single functional $\log\det(Y_nY_n^* + \rho I_N)$ and do not provide results for a general class of functionals as in [4] and [7]. We do believe however that all the computations performed in this article are a good starting point to address this issue.

The approach developed in this article is conceptually simple. The quantity $\mathcal{I}_n(\rho) - \mathbb{E}\mathcal{I}_n(\rho)$ is decomposed into a sum of martingale differences; we then systematically approximate random

[1] However, provided one is willing to make strong moment and distribution assumptions and consider real rather than complex random variables, one can in principle get a CLT for $\mathcal{I}$ from [4] although the closed-form formula for the variance obtained here would still require a specific effort.

quantities such as quadratic forms $\mathbf{x}^T A\mathbf{x}$ where $\mathbf{x}$ is some random vector and A is some deterministic matrix, by their deterministic counterparts $\frac{1}{n}\mathrm{Trace}\, A$ (in the case where the entries of $\mathbf{x}$ are i.i.d. with variance $\frac{1}{n}$) as the size of the vectors and the matrices goes to infinity. A careful study of the deterministic quantities that arise, mainly based on (deterministic) matrix analysis is carried out and yields the closed-form formula for the variance. The martingale method which is used to establish the fluctuations of $\mathcal{I}_n(\rho)$ can be traced back to Girko's REFORM (REsolvent, FORmula and Martingale) method (see [37, 40]) and is close to the one developed in [7].

Notations and assumptions. Let $N = N(n)$ be a sequence of integers such that

$$\lim_{n\to\infty} \frac{N(n)}{n} = c \in (0,\infty) \ .$$

In the sequel, we shall consider a $N \times n$ random matrix Y_n with individual entries:

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

where X_{ij}^n are complex centered i.i.d random variables with unit variance and $(\sigma_{ij}(n);\ 1 \le i \le N, 1 \le j \le n)$ is a triangular array of real numbers. Denote by $\mathrm{var}(Z)$ the variance of the random variable Z. Since $\mathrm{var}(Y_{ij}^n) = \sigma_{ij}^2(n)/n$, the family $(\sigma_{ij}(n))$ will be referred to as a variance profile.

Assumption A-6. *The random variables $(X_{ij}^n\ ;\ 1 \le i \le N, 1 \le j \le n\,, n \ge 1)$ are complex, independent and identically distributed. They satisfy*

$$\mathbb{E}X_{ij}^n = \mathbb{E}(X_{ij}^n)^2 = 0, \quad \mathbb{E}|X_{ij}^n|^2 = 1 \quad \text{and} \quad \mathbb{E}|X_{ij}^n|^8 < \infty \ .$$

Assumption A-7. *There exists a finite positive real number $\sigma_{\max}$ such that the family of real numbers $(\sigma_{ij}(n),\ 1 \le i \le N,\ 1 \le j \le n,\ n \ge 1)$ satisfies:*

$$\sup_{n\ge 1} \max_{\substack{1\le i\le N\\ 1\le j\le n}} |\sigma_{ij}(n)| \le \sigma_{\max} \ .$$

Assumption A-8. *There exists a real number $\sigma_{\min}^2 > 0$ such that*

$$\liminf_{n\ge 1} \min_{1\le j\le n} \frac{1}{n}\sum_{i=1}^{N} \sigma_{ij}^2(n) \ge \sigma_{\min}^2 \ .$$

Sometimes we shall assume that the variance profile is obtained by sampling a function on the unit square of $\mathbb{R}^2$. This helps to get limiting expressions and limiting behaviours (cf. Theorem 4.7):

Assumption A-9. *There exists a continuous function $\sigma^2 : [0,1]\times[0,1] \to (0,\infty)$ such that $\sigma_{ij}^2(n) = \sigma^2(i/N, j/n)$.*

Remark 4.5. *(Remarks related to the assumptions)*

1. *One may readily relax the assumption $\frac{N}{n} \to c \in (0,\infty)$ and assume instead:*

$$0 < \liminf_n \frac{N}{n} \le \limsup_n \frac{N}{n} < \infty \ ,$$

as done in [49]. We stick to the initial assumption to remain coherent with the companion paper [53].

2. *Using truncation arguments à la Bai and Silverstein [6, 102, 103], one may lower the moment assumption related to the X_{ij}'s in* **A-6**.

3. *Obviously, assumption* **A-8** *holds if σ_{ij}^2 is uniformly lower bounded by some nonnegative quantity.*

4. *Obviously, assumption* **A-9** *implies both* **A-7** *and* **A-8**. *When* **A-9** *holds, we shall say that there exists a limiting variance profile.*

5. *If necessary, assumption* **A-8** *can be slightly improved by stating:*

$$\max\left(\liminf_{n\geq 1}\min_{1\leq j\leq n}\frac{1}{n}\sum_{i=1}^{N}\sigma_{ij}^2(n)\ ,\ \liminf_{n\geq 1}\min_{1\leq i\leq N}\frac{1}{n}\sum_{j=1}^{n}\sigma_{ij}^2(n)\right)>0\ .$$

In the case where the first liminf is zero, one may notice that $\log\det(Y_nY_n^*+\rho I_N)=\log\det(Y_n^*Y_n+\rho I_n)+(n-N)\log\rho$ *and consider $Y_n^*Y_n$ instead.*

The indicator function of the set $\mathcal{A}$ will be denoted by $\mathbf{1}_{\mathcal{A}}(x)$, its cardinality by $\#\mathcal{A}$. As usual, $\mathbb{R}^+=\{x\in\mathbb{R}\ :\ x\geq 0\}$ and $\mathbb{C}^+=\{z\in\mathbb{C}\ :\ \mathrm{Im}\,(z)>0\}$.

We denote by $\xrightarrow{\mathcal{P}}$ the convergence in probability of random variables and by $\xrightarrow{\mathcal{D}}$ the convergence in distribution of probability measures.

Denote by $\mathrm{diag}(a_i;\ 1\leq i\leq k)$ the $k\times k$ diagonal matrix whose diagonal entries are the a_i's. Element (i,j) of matrix M will be either denoted m_{ij} or $[M]_{ij}$ depending on the notational context. Denote by M^T the matrix transpose of M, by M^* its Hermitian adjoint, by $\mathrm{Tr}(M)$ its trace and $\det(M)$ its determinant (if M is square), and by $F^{M\,M^*}$, the empirical distribution function of the eigenvalues of $M\,M^*$, *i.e.*

$$F^{M\,M^*}(x)=\frac{1}{N}\#\{i:\lambda_i\leq x\}\ ,$$

where $M\,M^*$ has dimensions $N\times N$ and the λ_i's are the eigenvalues of $M\,M^*$.

When dealing with vectors, $\|\cdot\|$ will refer to the Euclidean norm, and $\|\cdot\|_\infty$, to the max (or ℓ_∞) norm. In the case of matrices, $\|\cdot\|$ will refer to the spectral norm and $|\!|\!|\cdot|\!|\!|_\infty$ to the maximum row sum norm (referred to as the max-row norm), *i.e.*, $|\!|\!|M|\!|\!|_\infty=\max_{1\leq i\leq N}\sum_{j=1}^{N}|[M]_{ij}|$ when M is a $N\times N$ matrix. We shall denote by $r(M)$ the spectral radius of matrix M.

When no confusion can occur, we shall often drop subscripts and superscripts n for readability.

Let ν be a bounded non-negative measure over $\mathbb{R}$. Its Stieltjes transform f is defined as:

$$f(z)=\int_{\mathbb{R}}\frac{\nu(d\lambda)}{\lambda-z},\quad z\in\mathbb{C}\setminus\mathrm{supp}(\nu)\ ,$$

where $\mathrm{supp}(\nu)$ is the support of the measure ν. We shall denote by $\mathcal{S}(\mathbb{R}^+)$ the set of Stieltjes transforms of probability measures with support in $\mathbb{R}^+$.

First Order Results.

Theorem 4.6 ([53], [39])**.** *Consider the family of random matrices $(Y_nY_n^*)$ and assume that* **A-6** *and* **A-7** *hold. Then, the following hold true:*

1. *The system of N functional equations:*

$$t_i(z) = \frac{1}{-z + \frac{1}{n}\sum_{j=1}^{n} \frac{\sigma_{ij}^2(n)}{1+\frac{1}{n}\sum_{\ell=1}^{N}\sigma_{\ell j}^2(n)t_\ell(z)}} \tag{4.9}$$

admits a unique solution $(t_1(z), \cdots, t_N(z))$ in $\mathcal{S}(\mathbb{R}^+)^N$. In particular, $m_n(z) = \frac{1}{N}\sum_{i=1}^{N} t_i(z)$ belongs to $\mathcal{S}(\mathbb{R}^+)$ and there exists a probability measure π_n on $\mathbb{R}^+$ such that:

$$m_n(z) = \int_0^\infty \frac{\pi_n(d\lambda)}{\lambda - z} \ .$$

2. *For every continuous and bounded function g on $\mathbb{R}^+$,*

$$\int_{\mathbb{R}^+} g(\lambda)\, dF^{Y_n Y_n^*}(\lambda) - \int_{\mathbb{R}^+} g(\lambda)\, \pi_n(d\lambda) \xrightarrow[n\to\infty]{} 0 \quad a.e.$$

3. *The function $V_n(\rho) = \int_{\mathbb{R}^+} \log(\lambda+\rho)\pi_n(d\lambda)$ is finite for every $\rho > 0$ and*

$$\mathbb{E}\mathcal{I}_n(\rho) - V_n(\rho) \xrightarrow[n\to\infty]{} 0 \qquad \text{where} \quad \mathcal{I}_n(\rho) = \frac{1}{N}\log\det\left(Y_n Y_n^* + \rho I_N\right) \ .$$

Moreover, $V_n(\rho)$ admits the following closed form formula:

$$V_n(\rho) = -\frac{1}{N}\sum_{i=1}^{N}\log t_i(-\rho) + \frac{1}{N}\sum_{j=1}^{n}\log\left(1 + \frac{1}{n}\sum_{\ell=1}^{N}\sigma_{\ell j}^2(n)t_\ell(-\rho)\right) - \frac{1}{Nn}\sum_{i=1:N, j=1:n}\frac{\sigma_{ij}^2(n)t_i(-\rho)}{1+\frac{1}{n}\sum_{\ell=1}^{N}\sigma_{\ell j}^2(n)t_\ell(-\rho)} \ .$$

where the t_i's are defined above.

In the case where there exists a limiting variance profile, the results can be expressed in the following manner:

Theorem 4.7 ([16], [38], [52])**.** *Consider the family of random matrices $(Y_n Y_n^*)$ and assume that* **A-6** *and* **A-9** *hold. Then:*

1. *The functional equation*

$$\tau(u,z) = \left(-z + \int_0^1 \frac{\sigma^2(u,v)}{1 + c\int_0^1 \sigma^2(x,v)\tau(x,z)\,dx}dv\right)^{-1} \tag{4.10}$$

admits a unique solution among the class of functions $\Phi : [0,1]\times\mathbb{C}\setminus\mathbb{R} \to \mathbb{C}$ such that $u \mapsto \Phi(u,z)$ is continuous over $[0,1]$ and $z \mapsto \Phi(u,z)$ belongs to $\mathcal{S}(\mathbb{R}^+)$.

2. *The function $f(z) = \int_0^1 \tau(u,z)\,du$ where $\tau(u,z)$ is defined above is the Stieltjes transform of a probability measure $\mathbb{P}$. Moreover, we have*

$$F^{Y_n Y_n^*} \xrightarrow[n\to\infty]{\mathcal{D}} \mathbb{P} \ a.s.$$

Remark 4.8. *If one is interested in the Stieltjes function related to the limit of $F^{Y_n^* Y_n}$, then one must introduce the following function $\tilde{\tau}$, which is the counterpart of τ:*

$$\tilde{\tau}(v,z) = \left(-z + c \int_0^1 \frac{\sigma^2(t,v)}{1 + \int_0^1 \sigma^2(t,s)\tilde{\tau}(s,z)\, ds} dt \right)^{-1} .$$

Functions τ and $\tilde{\tau}$ are related via the following equations:

$$\left\{ \begin{array}{l} \tau(u,z) = -\left[z \left(1 + \int_0^1 \sigma^2(u,v)\tilde{\tau}(v,z)\, dv \right) \right]^{-1} \\ \tilde{\tau}(v,z) = -\left[z \left(1 + c \int_0^1 \sigma^2(t,v)\tau(t,z)\, dt \right) \right]^{-1} \end{array} \right. . \tag{4.11}$$

The Central Limit Theorem for $\mathcal{I}_n(\rho)$. When given a variance profile, one can consider the t_i's defined in Theorem 4.6-(1). Recall that

$$T(z) = \mathrm{diag}(t_i(z), 1 \leq i \leq N) \quad \text{and} \quad D_j = \mathrm{diag}(\sigma_{ij}^2, 1 \leq i \leq N) .$$

We shall first define in Theorem 4.9 a non-negative real number that will play the role of the variance in the CLT. We then state the CLT in Theorem 4.19. Theorem 4.11 deals with the bias term $N(\mathbb{E}\mathcal{I} - V)$.

Theorem 4.9 (Definition of the variance)**.** *Consider a variance profile (σ_{ij}) which fulfills assumptions* **A-7** *and* **A-8** *and the related t_i's defined in Theorem 4.6-(1). Let $\rho > 0$.*

1. *Let $A_n = (a_{\ell,m})$ be the matrix defined by:*

$$a_{\ell,m} = \frac{1}{n} \frac{\frac{1}{n}\mathrm{Tr} D_\ell D_m T(-\rho)^2}{\left(1 + \frac{1}{n}\mathrm{Tr} D_\ell T(-\rho)\right)^2} , \ 1 \leq \ell, m \leq n ,$$

   then the quantity $\mathcal{V}_n = -\log\det(I_n - A_n)$ is well-defined.

2. *Denote by $\mathcal{W}_n = \mathrm{Tr} A_n$ and let κ be a real number*[2] *satisfying $\kappa \geq -1$. The sequence $(\mathcal{V}_n + \kappa\mathcal{W}_n)$ satisfies*

$$0 \quad < \quad \liminf_n \left(\mathcal{V}_n + \kappa\mathcal{W}_n\right) \quad \leq \quad \limsup_n \left(\mathcal{V}_n + \kappa\mathcal{W}_n\right) \quad < \quad \infty$$

   as $n \to \infty$ and $N/n \to c > 0$. We shall denote by:

$$\Theta_n^2 \stackrel{\triangle}{=} -\log\det(I - A_n) + \kappa \mathrm{Tr} A_n .$$

In the sequel and for obvious reasons, we shall refer to matrix A_n as the **variance matrix**. In order to study the CLT for $N(\mathcal{I}_n(\rho) - V_n(\rho))$, we decompose it into a random term from which the fluctuations arise:

$$N\left(\mathcal{I}_n(\rho) - \mathbb{E}\mathcal{I}_n(\rho)\right) = \log\det(Y_n Y_n^* + \rho I_N) - \mathbb{E}\log\det(Y_n Y_n^* + \rho I_N) ,$$

and into a deterministic one which yields a bias in the CLT:

$$N\left(\mathbb{E}\mathcal{I}_n(\rho) - V_n(\rho)\right) = \mathbb{E}\log\det(Y_n Y_n^* + \rho I_N) - N \int \log(\lambda + \rho)\pi_n(\, d\lambda) .$$

We can now state the CLT.

[2] In the sequel, κ is defined as $\kappa = \mathbb{E}|X_{11}|^4 - 2$.

Theorem 4.10 (The CLT)**.** *Consider the family of random matrices $(Y_nY_n^*)$ and assume that* **A-6**, **A-7** *and* **A-8** *hold true. Let $\rho > 0$, let $\kappa = \mathbb{E}|X_{11}|^4 - 2$, and let Θ_n^2 be given by Theorem 4.9. Then*

$$\Theta_n^{-1}\Big(\log\det(Y_nY_n^* + \rho I_N) - \mathbb{E}\log\det(Y_nY_n^* + \rho I_N)\Big) \quad \xrightarrow[n\to\infty,\ \frac{N}{n}\to c]{\mathcal{D}} \quad \mathcal{N}(0,1)\ .$$

The asymptotic bias is described in the following theorem:

Theorem 4.11 (The bias)**.** *Assume that the setting of Theorem 4.19 holds true. Then*

1. *For every $\omega \in [\rho, +\infty)$, the system of n linear equations with unknown parameters $(\boldsymbol{w}_{\ell,n}(\omega); 1 \le \ell \le n)$:*

$$\boldsymbol{w}_{\ell,n}(\omega) = \frac{1}{n}\sum_{m=1}^{n} \frac{\frac{1}{n}\mathrm{Tr} D_\ell D_m T(-\omega)^2}{(1+\frac{1}{n}\mathrm{Tr} D_\ell T(-\omega))^2}\boldsymbol{w}_{m,n}(\omega) + \boldsymbol{p}_{\ell,n}(\omega), \tag{4.12}$$

   with

$$\boldsymbol{p}_{\ell,n}(\omega) = \kappa\ \omega^2 \tilde{t}_\ell(-\omega)^2 \times \left(\frac{\omega}{n}\sum_{i=1}^{N}\left(\frac{\sigma_{i\ell}^2 t_i(-\omega)^3}{n}\mathrm{Tr}\tilde{D}_i^2\tilde{T}(-\omega)^2\right) - \frac{\tilde{t}_\ell(-\omega)}{n}\mathrm{Tr} D_\ell^2 T(-\omega)^2\right) \tag{4.13}$$

   admits a unique solution for n large enough. In particular if $\kappa = 0$, then $\boldsymbol{p}_{\ell,n} = 0$ and $\boldsymbol{w}_{\ell,n} = 0$.

2. *Let*

$$\beta_n(\omega) = \frac{1}{n}\sum_{\ell=1}^{n}\boldsymbol{w}_{\ell,n}(\omega)\ . \tag{4.14}$$

   Then $\mathcal{B}_n(\rho) \stackrel{\triangle}{=} \int_\rho^\infty \beta_n(\omega)\,d\omega$ is well-defined, moreover,

$$\limsup_n \int_\rho^\infty |\beta_n(\omega)|\,d\omega \ <\ \infty. \tag{4.15}$$

   Furthermore,

$$N\left(\mathbb{E}\mathcal{I}_n(\rho) - V_n(\rho)\right) - \mathcal{B}_n(\rho) \quad \xrightarrow[n\to\infty,\ \frac{N}{n}\to c]{} \quad 0\ . \tag{4.16}$$

Remark 4.12 (The Gaussian case)**.** *In the case where the entries X_{ij} are complex Gaussian (*i.e. *with independent normal real and imaginary parts, each of them centered with variance 2^{-1}) then $\kappa = 0$ and the CLT writes:*

$$N[-\log\det(I - A_n)]^{-\frac{1}{2}}\left(\mathcal{I}_n(\rho) - V_n(\rho)\right) \xrightarrow[n\to\infty,\ \frac{N}{n}\to c]{\mathcal{D}} \quad \mathcal{N}(0,1)\ .$$

The CLT for a limiting variance profile. In this section, we shall assume that **A-9** holds, *i.e.* $\sigma_{ij}^2(n) = \sigma^2(i/N, j/n)$ for some continuous nonnegative function $\sigma^2(x,y)$. Recall the definitions (4.10) of function τ and of the t_i's (defined in Theorem 4.6-(1)). In the sequel, we take $\rho > 0$, $z = -\rho$ and denote $\tau(t) \stackrel{\triangle}{=} \tau(t, -\rho)$.

Let $K : [0,1]^2 \to \mathbb{R}$ be some non-negative continuous function we shall refer to as a kernel. Consider the associated operator (similarly denoted with a slight abuse of notations):

$$\begin{aligned} K : C[0,1] &\to C[0,1] \\ f &\mapsto Kf(x) = \int_{[0,1]} K(x,y) f(y)\, dy \ . \end{aligned}$$

Then one can define (see for instance [106, Theorem 5.3.1]) the Fredholm determinant $\det(1+\lambda K)$, where $1 : f \mapsto f$ is the identity operator, as

$$\det(1-\lambda K) = \sum_{k=0}^{\infty} \frac{(-1)^k \lambda^k}{k!} \int_{[0,1]^k} K \begin{pmatrix} x_1 & \cdots & x_k \\ x_1 & \cdots & x_k \end{pmatrix} \otimes_{i=1}^k d\, x_i \tag{4.17}$$

where

$$K \begin{pmatrix} x_1 & \cdots & x_k \\ y_1 & \cdots & y_k \end{pmatrix} = \det(K(x_i, y_j),\ 1 \le i,j \le k) \ ,$$

for every $\lambda \in \mathbb{C}$. One can define the trace of the iterated kernel as:

$$\mathrm{Tr} K^k = \int_{[0,1]^k} K(x_1,x_2) \cdots K(x_{k-1}, x_k) K(x_k, x_1) dx_1 \cdots dx_k$$

In the sequel, we shall focus on the following kernel:

$$K_\infty(x,y) = \frac{c \int_{[0,1]} \sigma^2(u,x)\sigma^2(u,y)\tau^2(u)\, du}{\left(1 + c \int_{[0,1]} \sigma^2(u,x)\tau(u)\, du\right)^2} \ . \tag{4.18}$$

Theorem 4.13 (The variance)**.** *Assume that assumptions* **A-6** *and* **A-9** *hold. Let $\rho > 0$ and recall the definition of matrix A_n:*

$$a_{\ell,m} = \frac{1}{n} \frac{\frac{1}{n} \sum_{i=1}^N \sigma^2\left(\frac{i}{N}, \frac{\ell}{n}\right) \sigma^2\left(\frac{i}{N}, \frac{m}{n}\right) t_i^2}{\left(1 + \frac{1}{n} \sum_{i=1}^N \sigma^2\left(\frac{i}{N}, \frac{\ell}{n}\right) t_i\right)^2} \ , \ 1 \le \ell, m \le n \ .$$

Then:

1. $\mathrm{Tr} A_n \xrightarrow[n\to\infty]{} \mathrm{Tr} K_\infty$.

2. $\det(I_n - A_n) \xrightarrow[n\to\infty]{} \det(1 - K_\infty)$ *and* $\det(1-K_\infty) \neq 0$.

3. *Let* $\kappa = \mathbb{E}|X_{11}|^4 - 2$, *then*

$$0 < -\log\det(1-K_\infty) + \kappa \mathrm{Tr} K_\infty < \infty \ .$$

Corollary 4.14 (Fluctuations)**.** *Assume that* **A-6** *and* **A-9** *hold. Denote by*

$$\Theta_\infty^2 = -\log\det(1-K_\infty) + \kappa \mathrm{Tr} K_\infty \ ,$$

then

$$\begin{aligned}&\frac{N}{\Theta_\infty}\left(\mathcal{I}_n(\rho) - \mathbb{E}\mathcal{I}_n(\rho)\right)\\ &\quad = \quad \Theta_\infty^{-1}\left(\log\det\left(Y_nY_n^* + \rho I_N\right) - \mathbb{E}\log\det\left(Y_nY_n^* + \rho I_N\right)\right) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0,1) \ .\end{aligned}$$

Recall the definition of $\tilde{\tau}$ (cf. Remark 4.8).

We now state a consequence of Corollary 4.14 in the case where the variance profile is separable. Recall the definitions of τ and $\tilde{\tau}$ given in (4.11).

Corollary 4.15 (Separable variance profile)**.** *Assume that* **A-6** *and* **A-9** *hold. Assume moreover that* $\rho > 0$ *and that* σ^2 *is separable,* i.e. *that*

$$\sigma^2(x,y) = d(x)\tilde{d}(y) \ ,$$

where both $d : [0,1] \to (0,\infty)$ *and* $\tilde{d} : [0,1] \to (0,\infty)$ *are continuous functions. Denote by*

$$\gamma = c\int_0^1 d^2(t)\tau^2(t)\,dt \qquad \textit{and} \qquad \tilde{\gamma} = \int_0^1 \tilde{d}^2(t)\tilde{\tau}^2(t)\,dt \ .$$

Then

$$\Theta_\infty^2 = -\log\left(1-\rho^2\gamma\tilde{\gamma}\right) + \kappa\rho^2\gamma\tilde{\gamma} \ . \tag{4.19}$$

Remark 4.16. *In the case where the random variables* X_{ij} *are standard complex circular Gaussian (i.e.* $X_{ij} = U_{ij} + \mathbf{i}V_{ij}$ *with* U_{ij} *and* V_{ij} *independent real centered Gaussian random variables with variance* 2^{-1}*) and where the variance profile is separable, then*

$$N(\mathcal{I}_n(\rho) - V_n(\rho)) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}\left(0, -\log\left(1-\rho^2\gamma\tilde{\gamma}\right)\right) \ .$$

This result is in accordance with those in [82] and in [49].

4.3 A CLT for the Signal to Interference plus Noise ratio

Introduction Large Random Matrix Theory (LRMT) is a powerful mathematical tool used to study the performance of multi-user and multi-access communication systems such as Multiple Input Multiple Output (MIMO) digital wireless systems, antenna arrays for source detection and localization, spread spectrum communication systems as Code Division Multiple Access (CDMA) and Multi-Carrier CDMA (MC-CDMA) systems. In most of these communication systems, the N dimensional received random vector $\mathbf{r} \in \mathbb{C}^N$ is described by the model

$$\mathbf{r} = \mathbf{\Sigma}\mathbf{s} + \mathbf{n} \tag{4.20}$$

where $\mathbf{s} = [s_0, s_1, \ldots, s_K]^{\mathrm{T}}$ is the unknown random vector of transmitted symbols with size $K+1$ satisfying $\mathbb{E}\mathbf{s}\mathbf{s}^* = \mathbf{I}_{K+1}$, the noise $\mathbf{n}$ is an independent Additive White Gaussian Noise (AWGN) with covariance matrix $\mathbb{E}\mathbf{n}\mathbf{n}^* = \rho\mathbf{I}_N$ whose variance $\rho > 0$ is known, and matrix $\mathbf{\Sigma}$ represents the

known "channel" in the wide sense whose structure depends on the particular system under study. One typical problem addressed by LRMT concerns the estimation performance by the receiver of a given transmitted symbol, say s_0.

In this paper we focus on one of the most popular estimators, namely the linear Wiener estimator, also called LMMSE for Linear Minimum Mean Squared Error estimator: the LMMSE estimate $\hat{s}_0 = \mathbf{g}^* \mathbf{r}$ of signal s_0 is the one for which the $N \times 1$ vector $\mathbf{g}$ minimizes $\mathbb{E}|\hat{s}_0 - s_0|^2$. If we partition the channel matrix as $\mathbf{\Sigma} = [\mathbf{y}\ \mathbf{Y}]$ where $\mathbf{y}$ is the first column of $\mathbf{\Sigma}$ and where matrix $\mathbf{Y}$ has dimensions $N \times K$, then it is well known that vector $\mathbf{g}$ is given by $\mathbf{g} = \left(\mathbf{\Sigma}\mathbf{\Sigma}^* + \rho \mathbf{I}_N\right)^{-1} \mathbf{y}$. Usually, the performance of this estimator is evaluated in terms of the Signal to Interference plus Noise Ratio (SINR) at its output. Writing the received vector $\mathbf{r}$ as $\mathbf{r} = s_0 \mathbf{y} + \mathbf{r}_{\text{in}}$ where $s_0 \mathbf{y}$ is the relevant term and $\mathbf{r}_{\text{in}}$ represents the so-called interference plus noise term, the SINR is given by $\beta_K = |\mathbf{g}^* \mathbf{y}|^2 / \mathbb{E}|\mathbf{g}^* \mathbf{r}_{\text{in}}|^2$. Plugging the expression of $\mathbf{g}$ given above into this expression, one can prove that the SINR β_K is given by the well-known expression:

$$\beta_K = \mathbf{y}^* \left(\mathbf{Y}\mathbf{Y}^* + \rho \mathbf{I}_N\right)^{-1} \mathbf{y} \ . \tag{4.21}$$

In general, this expression does not provide a clear insight on the impact of the channel model parameters (such as the load factor KN^{-1}, the power distribution of the transmission data streams, or the correlation structure of the channel paths in the context of multi-antenna transmissions) on the performance of the LMMSE estimator.

An alternative approach, justified by the fluctuating nature of the channel paths in the context of MIMO communications and by the pseudo-random nature of the spreading sequences in spread spectrum applications consists to model matrix $\mathbf{\Sigma}$ as a random matrix (in this case, β_K becomes a random SINR). The simplest random matrix model for $\mathbf{\Sigma}$, corresponding to the most canonical MIMO or CDMA transmission channels, corresponds to independent and identically distributed (i.i.d.) entries with mean zero and variance N^{-1}. In that case, LRMT shows that when $K \to \infty$ and the load factor KN^{-1} converges to a limiting load factor $\alpha > 0$, the SINR β_K converges almost surely (a.s.) to an explicit deterministic quantity $\overline{\beta}(\alpha, \rho)$ which simply depends on the limiting load factor α and on the noise variance ρ. As a result, the impact of these two parameters on the LMMSE performance can be easily evaluated [113, 119].

The LMMSE SINR large dimensional behavior for more sophisticated random matrix models has also been thoroughly studied (cf. [113, 35, 13, 96, 22, 72, 116, 94]) and it has been proved that there exists a deterministic sequence $(\overline{\beta}_K)$, generally defined as the solution of an implicit equation, such that $\beta_K - \overline{\beta}_K \to 0$ almost surely as $K \to \infty$ and $\frac{K}{N}$ remains bounded away from zero and from infinity.

Beyond the convergence $\beta_K - \overline{\beta}_K \to 0$, a natural question arises concerning the accuracy of $\overline{\beta}_K$ for finite values of K. A first answer to this question consists in evaluating the Mean Squared Error (MSE) of the SINR $\mathbb{E}|\beta_K - \overline{\beta}_K|^2$ for large K. A further problem is the computation of outage probability, that is the probability for $\beta_K - \overline{\beta}_K$ to be below a certain level. Both problems can be addressed by establishing a Central Limit Theorem (CLT) for $\beta_K - \overline{\beta}_K$. In this paper, we establish such a CLT (Theorem 4.19 below) for a large class of random matrices $\mathbf{\Sigma}$. We prove that there exists a sequence $\Theta_K^2 = \mathcal{O}(1)$ such that $\frac{\sqrt{K}}{\Theta_K}(\beta_K - \overline{\beta}_K)$ converges in distribution to the standard normal law $\mathcal{N}(0,1)$ in the asymptotic regime. One can therefore infer that the MSE asymptotically behaves like $\frac{\Theta_K^2}{K}$ and that the outage probability can be simply approximated by a Gaussian tail function.

The class of random matrices $\mathbf{\Sigma}$ we consider in this paper is described by the following

statistical model: Assume that

$$\mathbf{\Sigma} = \left(\Sigma_{nk}\right)_{n=1,k=0}^{N,K} = \left(\frac{\sigma_{nk}}{\sqrt{K}} W_{nk}\right)_{n=1,k=0}^{N,K} \tag{4.22}$$

where the complex random variables W_{nk} are i.i.d. with $\mathbb{E}W_{nk} = 0$, $\mathbb{E}W_{nk}^2 = 0$ and $\mathbb{E}|W_{nk}|^2 = 1$ and where $(\sigma_{nk}^2;\ 1 \leq n \leq N;\ 0 \leq k \leq K)$ is an array of real numbers. Due to the fact that $\mathbb{E}|\Sigma_{nk}|^2 = \frac{\sigma_{nk}^2}{K}$, the array (σ_{nk}^2) is referred to as a variance profile. An important particular case is when σ_{nk}^2 is *separable*, that is, writes:

$$\sigma_{nk}^2 = d_n \tilde{d}_k \ , \tag{4.23}$$

where $(d_1, \ldots, d_N)$ and $(\tilde{d}_0, \ldots, \tilde{d}_K)$ are two vectors of real positive numbers.

The asymptotic approximation $\overline{\beta}_K$ (first order result) is connected with the asymptotic eigenvalue distribution of Gram matrices $\mathbf{YY}^*$ where elements of $\mathbf{Y}$ are described by the model (4.22), and can be found in the mathematical LRMT literature in the work of Girko [38] (see also [103] and [100]). Applications in the field of wireless communications can be found in e.g. [22] in the separable case and in [116] in the general variance profile case.

Concerning the CLT for $\beta_K - \overline{\beta}_K$ (second order result), only some particular cases of the general model (4.22) have been considered in the literature among which the i.i.d. case ($\sigma_{nk}^2 = 1$) is studied in [112] (and based on a result of [101] pertaining to the asymptotic behavior of the eigenvectors of $\mathbf{YY}^*$). The more general CDMA model has been considered in [87], using a result of [44]. The model used in this paper includes the models of [112] and [87] as particular cases.

Fluctuations of other performance indexes such as Shannon's mutual information $\mathbb{E}\log\det\left(\frac{\Sigma\Sigma^*}{\rho} + \mathbf{I}_N\right)$ have also been studied at length. Let us cite [50] where the CLT is established in the separable case and [54] for a CLT in the general variance profile case. Similar results concerning the mutual information are found in [82] and in [83].

The statements about these deterministic approximations are valid within the following asymptotic regime:

$$K \to \infty, \quad \liminf \frac{K}{N} > 0 \quad \text{and} \quad \limsup \frac{K}{N} < \infty \ . \tag{4.24}$$

Note that $\frac{K}{N}$ is not required to converge. In the remainder of the paper, the notation "$K \to \infty$" will refer to (4.24).

We note that in the particular case where $\frac{K}{N} \to \alpha > 0$ and the variance profile is obtained by a regular sampling of a continuous function f i.e. $\sigma_{nk}^2 = f\left(\frac{n}{N}, \frac{k}{K+1}\right)$, it is possible to prove that $\overline{\beta}_K$ and Θ_K^2 converge towards limits that can be characterized by integral equations.

The approach used here is simple and powerful. It is based on the approximation of β_K by the sum of a martingale difference sequence and on the use of the CLT for martingales [14]. We note that apart from the LRMT context, such a technique has been used recently in [10] to establish a CLT on general quadratic forms of the type $\mathbf{z}^*\mathbf{Az}$ where $\mathbf{A}$ is a deterministic matrix and $\mathbf{z}$ is a random vector with i.i.d. elements.

Notations. Given a complex $N \times N$ matrix $\mathbf{X} = [x_{ij}]_{i,j=1}^N$, denote by $\|\mathbf{X}\|$ its spectral norm, and by $\|\!|\mathbf{X}|\!\|_\infty$ its maximum row sum norm, i.e., $\|\!|\mathbf{X}|\!\|_\infty = \max_{1\leq i\leq N} \sum_{j=1}^N |x_{ij}|$. Denote by $\|\cdot\|$ the Euclidean norm of a vector and by $\|\cdot\|_\infty$ its max (or ℓ_∞) norm.

First Order Results: The SINR Deterministic Approximation In the sequel, we shall often show explicitly the dependence on K in the notations. Consider the quadratic form (4.21):

$$\beta_K = \mathbf{y}^* \left(\mathbf{Y}\mathbf{Y}^* + \rho \mathbf{I}_N\right)^{-1} \mathbf{y} \ ,$$

where the sequence of matrices $\mathbf{\Sigma}(K) = [\mathbf{y}(K)\ \mathbf{Y}(K)]$ is given by

$$\mathbf{\Sigma}(K) \ = \ (\Sigma_{nk}(K))_{n=1,k=0}^{N,K} \ = \ \left(\frac{\sigma_{nk}(K)}{\sqrt{K}} W_{nk}\right)_{n=1,k=0}^{N,K} \ .$$

Let us state the main assumptions:

Assumption A-10. *The complex random variables* $(W_{nk};\ n \geq 1,\ k \geq 0)$ *are i.i.d. with* $\mathbb{E}W_{10} = 0$, $\mathbb{E}W_{10}^2 = 0$, $\mathbb{E}|W_{10}|^2 = 1$ *and* $\mathbb{E}|W_{10}|^8 < \infty$.

Assumption A-11. *There exists* $\sigma_{\max} < \infty$ *such that* $\sup_{K \geq 1} \ \max_{\substack{1 \leq n \leq N \\ 0 \leq k \leq K}} |\sigma_{nk}(K)| \leq \sigma_{\max}$.

Let $(a_m; 1 \leq m \leq M)$ be complex numbers, then $\mathrm{diag}(a_m; 1 \leq m \leq M)$ refers to the $M \times M$ diagonal matrix whose diagonal elements are the a_m's. If $\mathbf{A} = (a_{ij})$ is a square matrix, then $\mathrm{diag}(\mathbf{A})$ refers to the matrix $\mathrm{diag}(a_{ii})$. Consider the following diagonal matrices based on the variance profile along the columns and the rows of $\mathbf{\Sigma}$:

$$\begin{array}{rcll} \mathbf{D}_k(K) & = & \mathrm{diag}(\sigma_{1k}^2(K), \cdots, \sigma_{Nk}^2(K)), & 0 \leq k \leq K \\ \widetilde{\mathbf{D}}_n(K) & = & \mathrm{diag}(\sigma_{n1}^2(K), \cdots, \sigma_{nK}^2(K)), & 1 \leq n \leq N. \end{array} \tag{4.25}$$

Assumption A-12. *The variance profile satisfies*

$$\liminf_{K \geq 1} \min_{0 \leq k \leq K} \frac{1}{K} \mathrm{Tr} \mathbf{D}_k(K) > 0 \ .$$

Since $\mathbb{E}|W_{10}|^2 = 1$, one has $\mathbb{E}|W_{10}|^4 \geq 1$. The following is needed:

Assumption A-13. *At least one of the following conditions is satisfied:*

$$\mathbb{E}|W_{10}|^4 > 1 \quad or \quad \liminf_K \frac{1}{K^2} \mathrm{Tr} \left(\mathbf{D}_0(K) \sum_{k=1}^K \mathbf{D}_k(K) \right) > 0 \ .$$

Remark 4.17. *If needed, one can attenuate the assumption on the eighth moment in A-10. For instance, one can adapt without difficulty the proofs in this paper to the case where* $\mathbb{E}|W_{10}|^{4+\epsilon} < \infty$ *for* $\varepsilon > 0$. *We assumed* $\mathbb{E}|W_{10}|^8 < \infty$ *because at some places we rely on results of [54] which are stated with the assumption on the eighth moment. Assumption A-12 is technical. It has already appeared in [53]. Assumption A-13 is necessary to get a non-vanishing variance* Θ_K^2 *in Theorem 4.19.*

The following definitions will be of help in the sequel. A complex function $t(z)$ belongs to class $\mathcal{S}$ if $t(z)$ is analytical in the upper half plane $\mathbb{C}_+ = \{z \in \mathbb{C}\ ;\ \mathrm{im}(z) > 0\}$, if $t(z) \in \mathbb{C}_+$ for all $z \in \mathbb{C}_+$ and if $\mathrm{im}(z)|t(z)|$ is bounded over the upper half plane $\mathbb{C}_+$.

Denote by $\mathbf{Q}_K(z)$ and $\widetilde{\mathbf{Q}}_K(z)$ the resolvents of $\mathbf{Y}(K)\mathbf{Y}(K)^*$ and $\mathbf{Y}(K)^*\mathbf{Y}(K)$ respectively, that is the $N \times N$ and $K \times K$ matrices defined by:

$$\mathbf{Q}_K(z) = \left(\mathbf{Y}(K)\mathbf{Y}(K)^* - z\mathbf{I}_N\right)^{-1} \qquad \text{and} \qquad \widetilde{\mathbf{Q}}_K(z) = \left(\mathbf{Y}(K)^*\mathbf{Y}(K) - z\mathbf{I}_K\right)^{-1} \ .$$

It is known [38, 53] that there exists a deterministic diagonal $N \times N$ matrix function $\mathbf{T}(z)$ that approximates the resolvent $\mathbf{Q}(z)$. As we shall see, matrix $\mathbf{T}(z)$ also plays a fundamental role in the second order result (Theorem 4.19). In the following theorem, we recall the definition and some of the main properties of $\mathbf{T}(z)$.

Theorem 4.18. *The following hold true:*

1. *[53, Theorem 2.4] Let $(\sigma_{nk}^2(K);\ 1 \le n \le N;\ 1 \le k \le K)$ be a sequence of arrays of real numbers and consider the matrices $\mathbf{D}_k(K)$ and $\widetilde{\mathbf{D}}_n(K)$ defined in (4.25). The system of $N+K$ functional equations*

$$\left\{ \begin{array}{rcl} t_{n,K}(z) & = & \dfrac{-1}{z\left(1+\frac{1}{K}\mathrm{Tr}(\widetilde{\mathbf{D}}_n(K)\widetilde{\mathbf{T}}_K(z))\right)}, \quad 1 \le n \le N \\ \tilde{t}_{k,K}(z) & = & \dfrac{-1}{z\left(1+\frac{1}{K}\mathrm{Tr}(\mathbf{D}_k(K)\mathbf{T}_K(z))\right)}, \quad 1 \le k \le K \end{array} \right. \tag{4.26}$$

   where

$$\mathbf{T}_K(z) = \mathrm{diag}(t_{1,K}(z), \ldots, t_{N,K}(z)), \quad \widetilde{\mathbf{T}}_K(z) = \mathrm{diag}(\tilde{t}_{1,K}(z), \ldots, \tilde{t}_{K,K}(z))$$

   admits a unique solution $(\mathbf{T}, \widetilde{\mathbf{T}})$ among the diagonal matrices for which the $t_{n,K}$'s and the $\tilde{t}_{k,K}$'s belong to class $\mathcal{S}$. Moreover, functions $t_{n,K}(z)$ and $\tilde{t}_{k,K}(z)$ admit an analytical continuation over $\mathbb{C} - \mathbb{R}_+$ which is real and positive for $z \in (-\infty, 0)$.

2. *Let $\overline{\beta}_K = \dfrac{1}{K}\mathrm{Tr}(\mathbf{D}_0(K)\mathbf{T}_K(-\rho))$ where $\mathbf{T}_K$ is given by Theorem 4.18–(1). Assume that A-10 and A-11. Then*

$$\beta_K - \overline{\beta}_K \xrightarrow[K\to\infty]{} 0 \quad a.s.$$

Second order results: The Central Limit Theorem The following theorem is the main result of this section.

Theorem 4.19. *1. Assume that A-11, A-12 and A-13 hold true. Let $\mathbf{A}_K$ and $\mathbf{\Delta}_K$ be the $K \times K$ matrices*

$$\begin{array}{rcl} \mathbf{A}_K & = & \left[\dfrac{1}{K} \dfrac{\frac{1}{K}\mathrm{Tr}\mathbf{D}_\ell \mathbf{D}_m \mathbf{T}(-\rho)^2}{\left(1+\frac{1}{K}\mathrm{Tr}\mathbf{D}_\ell \mathbf{T}(-\rho)\right)^2} \right]_{\ell,m=1}^{K} \quad \text{and} \\ \mathbf{\Delta}_K & = & \mathrm{diag}\left(\left(1+\dfrac{1}{K}\mathrm{Tr}\mathbf{D}_\ell \mathbf{T}(-\rho)\right)^2 ; 1 \le \ell \le K \right), \end{array} \tag{4.27}$$

   where $\mathbf{T}$ is defined in Theorem 4.18–(1). Let $\mathbf{g}_K$ be the $K \times 1$ vector

$$\mathbf{g}_K = \left[\frac{1}{K}\mathrm{Tr}\mathbf{D}_0\mathbf{D}_1\mathbf{T}(-\rho)^2, \cdots, \frac{1}{K}\mathrm{Tr}\mathbf{D}_0\mathbf{D}_K\mathbf{T}(-\rho)^2 \right]^{\mathrm{T}}.$$

   Then the sequence of real numbers

$$\Theta_K^2 = \frac{1}{K}\mathbf{g}^{\mathrm{T}}(\mathbf{I}_K - \mathbf{A})^{-1}\mathbf{\Delta}^{-1}\mathbf{g} + (\mathbb{E}|W_{10}|^4 - 1)\frac{1}{K}\mathrm{Tr}\mathbf{D}_0^2\mathbf{T}(-\rho)^2 \tag{4.28}$$

   is well defined and furthermore

$$0 < \liminf_K \Theta_K^2 \le \limsup_K \Theta_K^2 < \infty\ .$$

2. Assume in addition A-10. Then the sequence $\beta_K = \mathbf{y}^*(\mathbf{Y}\mathbf{Y}^* + \rho\mathbf{I})^{-1}\mathbf{y}$ *satisfies*

$$\frac{\sqrt{K}}{\Theta_K}\left(\beta_K - \overline{\beta}_K\right) \xrightarrow[K\to\infty]{} \mathcal{N}(0,1)$$

in distribution where $\overline{\beta}_K = \frac{1}{K}\mathrm{Tr}\,\mathbf{D}_0\mathbf{T}_K$ *is defined in the statement of Theorem 4.18-*(2).

Remark 4.20. (Comparison with other performance indexes) *It is interesting to compare the "Mean Squared Error" (MSE) related to the SINR* β_K*:* $MSE(\beta_K) = \mathbb{E}(\beta_K - \overline{\beta}_K)^2$*, with the MSE related to Shannon's mutual information per transmit dimension* $I = \frac{1}{K}\log\det(\rho\mathbf{\Sigma}\mathbf{\Sigma}^* + \mathbf{I})$ *(studied in [82, 54] for instance):*

$$MSE(\beta_K) \quad \propto \quad \mathcal{O}\left(\frac{1}{K}\right) \qquad \textit{while} \qquad MSE(I) \quad \propto \quad \mathcal{O}\left(\frac{1}{K^2}\right) \ .$$

Remark 4.21. (On the achievability of the minimum of the variance) *Recall that the variance writes*

$$\Theta_K^2 = \frac{1}{K}\mathbf{g}^{\mathrm{T}}(\mathbf{I}_K - \mathbf{A})^{-1}\mathbf{\Delta}^{-1}\mathbf{g} + (\mathbb{E}|W_{10}|^4 - 1)\frac{1}{K}\mathrm{Tr}\,\mathbf{D}_0^2\mathbf{T}^2 \ .$$

As $\mathbb{E}|W_{10}|^2 = 1$*, one clearly has* $\mathbb{E}|W_{10}|^4 - 1 \geq 0$ *with equality if and only if* $|W_{10}| = 1$ *with probability one. Moreover, we can prove that* $\liminf_K \frac{1}{K}\mathbf{D}_0(K)\mathbf{T}_K^2 > 0$*. Therefore* $(\mathbb{E}|W_{10}|^4 - 1)\frac{1}{K}\mathrm{Tr}\,\mathbf{D}_0^2\mathbf{T}^2$ *is nonnegative, and is zero if and only if* $|W_{10}| = 1$ *with probability one. As a consequence,* Θ_K^2 *is minimum with respect to the distribution of the* W_{nk} *if and only if these random variables have their values on the unit circle. In the context of CDMA and MC-CDMA, this is the case when the signature matrix elements are elements of a PSK constellation. In multi-antenna systems, the* W_{nk}*'s are frequently considered as Gaussian which induces a penalty on the SINR asymptotic MSE with respect to the unit norm case.*

The deterministic approximation in the separable case In the separable case $\sigma_{nk}^2(K) = d_n(K)\tilde{d}_k(K)$, matrices $\mathbf{D}_k(K)$ and $\widetilde{\mathbf{D}}_n(K)$ are written as $\mathbf{D}_k(K) = \tilde{d}_k(K)\mathbf{D}(K)$ and $\widetilde{\mathbf{D}}_n(K) = d_n(K)\widetilde{\mathbf{D}}(K)$ where $\mathbf{D}(K)$ and $\widetilde{\mathbf{D}}(K)$ are the diagonal matrices

$$\mathbf{D}(K) = \mathrm{diag}(d_1(K), \ldots, d_N(K)), \quad \widetilde{\mathbf{D}}(K) = \mathrm{diag}(\tilde{d}_1(K), \ldots, \tilde{d}_K(K)) \ . \tag{4.29}$$

and one can check that the system of $N + K$ equations leading to $\mathbf{T}_K$ and $\widetilde{\mathbf{T}}_K$ simplifies into a system of two equations, and Theorem 4.18 takes the following form:

Proposition 4.22. *[53, Sec. 3.2]*

1. Assume $\sigma_{nk}^2(K) = d_n(K)\tilde{d}_k(K)$*. Given* $\rho > 0$*, the system of two equations*

$$\left\{\begin{array}{rcl} \delta_K(\rho) & = & \frac{1}{K}\mathrm{Tr}\left(\mathbf{D}\left(\rho(\mathbf{I}_N + \tilde{\delta}_K(\rho)\mathbf{D})\right)^{-1}\right) \\ \tilde{\delta}_K(\rho) & = & \frac{1}{K}\mathrm{Tr}\left(\widetilde{\mathbf{D}}\left(\rho(\mathbf{I}_K + \delta_K(\rho)\widetilde{\mathbf{D}})\right)^{-1}\right) \end{array}\right. \tag{4.30}$$

where $\mathbf{D}$ *and* $\widetilde{\mathbf{D}}$ *are given by* (4.29) *admits a unique solution* $(\delta_K(\rho), \tilde{\delta}_K(\rho))$*. Moreover, in this case matrices* $\mathbf{T}(-\rho)$ *and* $\widetilde{\mathbf{T}}(-\rho)$ *provided by Theorem 4.18–(1) coincide with*

$$\mathbf{T}(-\rho) = \frac{1}{\rho}(\mathbf{I} + \tilde{\delta}(\rho)\mathbf{D})^{-1} \quad \textit{and} \quad \widetilde{\mathbf{T}}(-\rho) = \frac{1}{\rho}(\mathbf{I} + \delta(\rho)\widetilde{\mathbf{D}})^{-1} \ . \tag{4.31}$$

2. *Notice that* $\mathbf{D}_0 = \tilde{d}_0\mathbf{D}$ *and that* $\delta(\rho)$ *given by the system* (4.30) *coincides with* $\frac{1}{K}\mathrm{Tr}(\mathbf{DT})$. *Assume that A-10 and A-11 hold true. Then*

$$\frac{\beta_K}{\tilde{d}_0} - \delta_K(\rho) \xrightarrow[K\to\infty]{} 0 \quad a.s.$$

where $\delta_K(\rho)$ *is given by Proposition 4.22–(1).*

3. *Assume that A-11 is satisfied and that* $\sigma_{nk}^2 = d_n\tilde{d}_k$. *Assume moreover that*

$$\min\left(\liminf_K \frac{1}{K}\mathrm{Tr}(\mathbf{D}(K)), \liminf_K \frac{1}{K}\mathrm{Tr}(\widetilde{\mathbf{D}}(K))\right) > 0 \tag{4.32}$$

where $\mathbf{D}$ *and* $\widetilde{\mathbf{D}}$ *are given by* (4.29). *Let* $\gamma = \frac{1}{K}\mathrm{Tr}\mathbf{D}^2\mathbf{T}^2$ *and* $\tilde{\gamma} = \frac{1}{K}\mathrm{Tr}\widetilde{\mathbf{D}}^2\widetilde{\mathbf{T}}^2$. *Then the sequence*

$$\Omega_K^2 = \gamma\left(\frac{\rho^2\gamma\tilde{\gamma}}{1-\rho^2\gamma\tilde{\gamma}} + \left(\mathbb{E}|W_{10}|^4 - 1\right)\right) \tag{4.33}$$

satisfies $0 < \liminf_K \Omega_K^2 \leq \limsup_K \Omega_K^2 < \infty$. *If, in addition, A-10 holds true, then:*

$$\frac{\sqrt{K}}{\Omega_K}\left(\frac{\beta_K}{\tilde{d}_0} - \delta_K\right) \xrightarrow[K\to\infty]{} \mathcal{N}(0,1)$$

in distribution.

Remark 4.23. *Let us provide a more explicit expression of* δ_K. *By combining the two equations in System* (4.30), *it turns out that* $\delta = \delta_K(\rho)$ *is the unique solution of the implicit equation:*

$$\delta = \frac{1}{K}\sum_{n=0}^{N-1} \frac{d_n}{\rho + \frac{1}{K}d_n\sum_{k=1}^{K}\frac{p_k}{1+p_k\delta}}\ . \tag{4.34}$$

Chapter 5

Specific contributions to wireless communication

Summary

This chapter is devoted to the description of three specific applications of random matrix theory to wireless communication. Results of papers [31] (optimization of the ergodic capacity), [65] (Bit Error Rate and Outage probability approximation) and [11] (Collaborative compress sensing) are presented.

[31] J. Dumont, W. Hachem, S. Lasaulce, P. Loubaton, and J. Najim. On the capacity achieving covariance matrix for rician mimo channels: an asymptotic approach. *to be published in IEEE Inf. Th.*, submitted in 2007.

[64] A. Kammoun, M. Kharouf, W. Hachem, and J. Najim. BER and Outage probability approximations for LMMSE detectors on correlated MIMO channels. *accepted for publication in IEEE Inf. Th.*, submitted in sept. 2008.

[11] P. Bianchi, M. Debbah, M. Maïda, and J. Najim. Power analysis of some hypothesis tests for collaborative sensing. In *IEEE Workshop on Stat. Signal Processing*, Cardiff, UK, August 2009.

Section 5.1. Consider the capacity C_E associated to a MIMO channel $\mathbf{H}$:

$$C_E = \sup_{\mathbf{Q}\geq 0,\ \frac{1}{t}\mathrm{Trace}\,\mathbf{Q}\leq 1} I(\mathbf{Q}) \qquad \text{with} \qquad I(\mathbf{Q}) = \mathbb{E}_{\boldsymbol{H}}\left[\log\det\left(\mathbf{I}_r + \frac{1}{\sigma^2}\mathbf{H}\mathbf{Q}\mathbf{H}^H\right)\right] ,$$

In this section, we propose an algorithm to approximate both C_E and the Capacity Achieving Covariance Matrix $\mathbf{Q}^*$, that is the maximizer $\mathbf{Q}^*$ of I over the set of constraints $\mathcal{C}_1 = \left\{\mathbf{Q}\geq 0,\ \frac{1}{t}\mathrm{Trace}\,\mathbf{Q}\leq 1\right\}$ in the case of a Rician channel:

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

where matrix $\mathbf{A}$ is deterministic with $\frac{1}{r}\mathrm{Tr}(\mathbf{A}\mathbf{A}^H) = 1$ and accounts for a line of sight component, matrix $\mathbf{V}$ is a random matrix and constant $K \geq 0$ is the so-called Rician factor which expresses the relative strength

of the direct and scattered components of the received signal. Matrix $\mathbf{V}$ is given by $\mathbf{V} = \frac{1}{\sqrt{t}}\mathbf{C}^{\frac{1}{2}}\mathbf{W}\tilde{\mathbf{C}}^{\frac{1}{2}}$, where $\mathbf{W} = (W_{ij})$ is a $r \times t$ matrix whose entries are independent and identically distributed (i.i.d.) $\mathcal{C}N(0,1)$.

Being able to compute the Capacity Achieving Covariance Matrix $\mathbf{Q}^*$ is fundamental because it enables to precode the transmitted vector $\mathbf{x}$ by $\mathbf{Q}^{*\frac{1}{2}}\mathbf{x}$ in order to achieve the capacity.

Contrary to the cases where $\mathbf{A} = 0$ (Rayleigh channel) or $\mathbf{C}, \tilde{\mathbf{C}} = \mathbf{I}$ (Rician uncorrelated channel), the problem does not admit a closed-form analytical solution and one often relies on massive Monte-Carlo solutions to optimize C_E. We propose a different approach based on Large Random Matrix theory in the asymptotic regime $r, t \to \infty$, $0 < \liminf \frac{r}{t} \leq \limsup \frac{r}{t} < \infty$. The main results of the section are summarized below:

1. Construction of a deterministic approximation $\overline{I}(\mathbf{Q})$ of $I(\mathbf{Q})$ easily computable following the lines exposed in Section 3: $I(\mathbf{Q}) - \overline{I}(\mathbf{Q}) = \mathcal{O}(t^{-1})$.
2. Maximization of $\overline{I}(\mathbf{Q})$ over the set of constraints $\mathcal{C}_1$: $\overline{I}(\overline{\mathbf{Q}}^*) = \sup_{\mathbf{Q}\in\mathcal{C}_1} \overline{I}(\mathbf{Q})$.
3. Proof that $I(\overline{\mathbf{Q}}^*) - I(\mathbf{Q}^*) \to 0$, where $\mathbf{Q}^*$ is the Capacity Achieving Covariance Matrix. This fact insures that the precoding can be performed by $\overline{\mathbf{Q}}^*$.
4. Description of an algorithm to efficiently compute $\overline{I}(\overline{\mathbf{Q}}^*) = \sup_{\mathbf{Q}\in\mathcal{C}_1} \overline{I}(\mathbf{Q})$.

The presented results substantially improve previous algorithms devoted to the computation of the CACM, and based on a direct optimization of I over $\mathcal{C}_1$ by computationally-intensive Monte-Carlo methods (cf. Vu and Paulraj [120]). Simulations are provided and a very good fit can be observed with the true capacity even for very small dimensions.

Section 5.2. Consider a N dimensional received signal $\mathbf{r} = \mathbf{\Sigma s} + \mathbf{n}$, where $\mathbf{s} = [s_0, \cdots, s_K]^{\mathrm{T}}$ is the transmitted complex vector signal with size $K+1$, $\mathbf{\Sigma} = [\mathbf{y}\ \mathbf{S}]$ is the $N \times (K+1)$ channel matrix and $\mathbf{n}$ is the white noise. The signal to interference plus noise writes:

$$\beta_K = \mathbf{y}^* \left(\mathbf{S}\mathbf{S}^* + \rho I_N\right)^{-1} \mathbf{y} \ .$$

The purpose of the section is to compute the three asymptotic moments of the SINR in the case of a correlated channel. Such a computation is motivated by applications: The first asymptotic moment gives a deterministic approximation of the SINR; the first two moments allow to approximate the distribution of the SINR by a gaussian one, using a CLT argument.

It turns out however that for small K, N, the gaussian approximation is no longer accurate. Li, Paul, Narasimhan and Cioffi [73] proposed to approximate the SINR distribution by so-called generalized Gamma distribution, which is parametrized by its three first moments and allows to take into account the positivity of the SINR. Having a good approximation of the distribution of the SINR is of prime importance since it allows one to compute the so-called Bit Error Rate and the outage probability of the channel.

Consider a correlated channel modeled by a matrix $\mathbf{D}^{1/2}\mathbf{Z}\tilde{\mathbf{D}}^{1/2}$, where $\mathbf{Z}$ has i.i.d. $\mathcal{C}N(0,1)$ entries, and $\mathbf{D}$ and $\tilde{\mathbf{D}}$ are deterministic diagonal matrices. The SINR writes:

$$\beta_K \quad = \quad \frac{p_0}{\rho K}\mathbf{z}^*\mathbf{D}^{\frac{1}{2}}\left(\frac{1}{K\rho}\mathbf{D}^{\frac{1}{2}}\mathbf{Z}\tilde{\mathbf{D}}\mathbf{Z}^*\mathbf{D}^{\frac{1}{2}} + \mathbf{I}\right)^{-1}\mathbf{D}^{\frac{1}{2}}\mathbf{z}$$

where $\mathbf{z}$ is a $N \times 1$ vector with complex independent standard Gaussian entries independent from $\mathbf{Z}$. Introduce the following well-known 2×2 system and the deterministic matrices $\mathbf{T}$ and $\tilde{\mathbf{T}}$,

$$\left\{\begin{array}{rcl} \delta & = & \frac{1}{K}\mathrm{Tr}\mathbf{D}_K\left(\mathbf{I} + t\tilde{\delta}\mathbf{D_K}\right)^{-1} \\ \tilde{\delta} & = & \frac{1}{K}\mathrm{Tr}\widetilde{\mathbf{D}}_K\left(\mathbf{I} + t\delta\widetilde{\mathbf{D}}_K\right)^{-1} \end{array}\right. \quad \text{and} \quad \left\{\begin{array}{l} \mathbf{T} = \left(\mathbf{I} + t\tilde{\delta}\mathbf{D}\right)^{-1} \\ \widetilde{\mathbf{T}} = \left(\mathbf{I} + t\delta\widetilde{\mathbf{D}}\right)^{-1} \end{array}\right. ,$$

and let $\delta = \frac{1}{K}\mathrm{Tr}\mathbf{DT}$, $\tilde{\delta} = \frac{1}{K}\mathrm{Tr}\widetilde{\mathbf{D}}\widetilde{\mathbf{T}}$, $\gamma = \frac{1}{K}\mathrm{Tr}\mathbf{D}^2\mathbf{T}^2$ and $\tilde{\gamma} = \frac{1}{K}\mathrm{Tr}\widetilde{\mathbf{D}}^2\widetilde{\mathbf{T}}^2$. The following approximations of the first, second and third moment of the SINR hold true (by $K \to \infty$, we mean $K \to \infty$ and $0 < \liminf \frac{K}{N} \leq \limsup \frac{K}{N} < \infty$):

1. First asymptotic moment:
$$\mathbb{E}\left(\frac{\beta_K}{p_0}\right) - \frac{\delta}{\rho} \xrightarrow[K\to\infty]{} 0,$$
2. Second asymptotic moment:
$$\Omega_K^2 = \frac{\gamma}{\rho^2}\left(\frac{\gamma\tilde{\gamma}}{\rho^2 - \gamma\tilde{\gamma}} + 1\right) \quad \text{and} \quad K\mathbb{E}\left(\frac{\beta_K}{p_0} - \mathbb{E}\left(\frac{\beta_K}{p_0}\right)\right)^2 - \Omega_K^2 \xrightarrow[K\to\infty]{} 0,$$
3. Third asymptotic moment:
$$\nu_K = \frac{2\rho^3}{K\left(\rho^2 - \gamma\tilde{\gamma}\right)^3}\left[\mathrm{Tr}\mathbf{D}^3\mathbf{T}^3 - \frac{\gamma^3}{\rho^3}\mathrm{Tr}\widetilde{\mathbf{D}}^3\widetilde{\mathbf{T}}^3\right] = \mathcal{O}(1) \quad \text{and} \quad K^2\mathbb{E}\left(\frac{\beta_K}{p_0} - \mathbb{E}\left(\frac{\beta_K}{p_0}\right)\right)^3 - \nu_K \xrightarrow[K\to\infty]{} 0.$$

This result extends Li et al.'s results [73] to the correlated case and is extensively based on gaussian tools -the integration by part formula and Poincaré-Nash inequality.

Section 5.3. In this contribution, we provide a theoretical study of a new technique for collaborative sensing proposed in [12] based on the analysis of the normalized (by the trace) maximum eigenvalue of the sample covariance matrix. The analytical study is based on Large Deviation Principles and asymptotic random matrix theory where we compute the error exponent of the probability of error of the test. The performance of the test is compared with the eigen-based ratio test (ratio of the extremal eigenvalues) and a proof, through the Bahadur efficiency, of the superiority of the new technique is provided. To be more specific, consider a $K \times K$ matrix $\hat{R} = \frac{1}{N}\mathbf{Y}\mathbf{Y}^H$ where matrix $\mathbf{Y}$ is $K \times N$. Under assumption (H_0), matrix $\mathbf{Y}$ has i.i.d. $\mathcal{C}N(0,1)$ entries and under assumption (H_1), the matrix follows a spiked model, i.e. the largest eigenvalue converges outside the bulk. In this section, we study the test based on the satistics $\frac{\lambda_1}{\frac{1}{K}\mathrm{Tr}\ \hat{R}}$ where λ_1 is the largest eigenvalue of $\hat{R}$. The Tracy-Widom fluctuations of λ_1 enables us to compute the threshold of the test. The study of the type II error relies on the study of the large deviations of λ_1. These large deviations enable us to compute the error exponent of the test.

5.1 Optimization of the ergodic capacity

Introduction. Since the seminal work of Telatar [111], the advantage of considering multiple antennas at the transmitter and the receiver in terms of capacity, for Gaussian and fast Rayleigh fading single-user channels, is well understood. In that paper the figure of merit chosen for characterizing the performance of a coherent[1] communication over a fading Multiple Input Multiple Output (MIMO) channel is the Ergodic Mutual Information (EMI). This choice will be justified later. Assuming the knowledge of the channel statistics at the transmitter, one important issue is then to maximize the EMI with respect to the channel input distribution. Without loss of optimality, the search for the optimal input distribution can be restricted to circularly Gaussian inputs. The problem then amounts to finding the optimum covariance matrix.

This optimization problem has been addressed extensively in the case of certain Rayleigh channels. In the context of the so-called Kronecker model, it has been shown by various authors (see

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

e.g. [43] for a review) that the eigenvectors of the optimal input covariance matrix must coincide with the eigenvectors of the transmit correlation matrix. It is therefore sufficient to evaluate the eigenvalues of the optimal matrix, a problem which can be solved by using standard optimization algorithms. Note that [1] extended this result to more general (non Kronecker) Rayleigh channels.

Rician channels have been comparatively less studied from this point of view. Let us mention the work [56] devoted to the case of uncorrelated Rician channels, where the authors proved that the eigenvectors of the optimal input covariance matrix are the right-singular vectors of the line of sight component of the channel. As in the Rayleigh case, the eigenvalues can then be evaluated by standard routines. The case of correlated Rician channels is undoubtedly more complicated because the eigenvectors of the optimum matrix have no closed form expressions. Moreover, the exact expression of the EMI being complicated (see e.g. [66]), both the eigenvalues and the eigenvectors have to be evaluated numerically. In [120], a barrier interior-point method is proposed and implemented to directly evaluate the EMI as an expectation. The corresponding algorithms are however not very attractive because they rely on computationally-intensive Monte-Carlo simulations.

In this section, we address the optimization of the input covariance of Rician channels with a two-sided (Kronecker) correlation. As the exact expression of the EMI is very complicated, we propose to evaluate an approximation of the EMI, valid when the number of transmit and receive antennas converge to $+\infty$ at the same rate, and then to optimize this asymptotic approximation. This will turn out to be a simpler problem.

The asymptotic approximation of the mutual information has been obtained by various authors in the case of MIMO Rayleigh channels, and has shown to be reliable even for a quite moderate number of antennas. The general case of a Rician correlated channel has recently been established in [53] using large random matrix theory and completes a number of previous works among which [26], [114] and [82] (Rayleigh channels), [2] and [84] (Rician uncorrelated channels), [59] (Rician receive correlated channel) and, later on, [110] (Rician correlated channels). Notice that the latest work (together with [82] and [84]) relies on the powerful but non-rigorous replica method. It also gives an expression for the variance of the mutual information.

In this paper, we rely on the results of [53] in which a closed form asymptotic approximation for the mutual information is provided, and present new results concerning its accuracy. We then address the optimization of the large system approximation w.r.t. the input covariance matrix and propose a simple iterative maximization algorithm which, in some sense, can be seen as a generalization to the Rician case of [121] devoted to the Rayleigh context: Each iteration will be intended to solve a system of two non linear equations as well as a standard waterfilling problem. In contrast with [121], we give some convergence results and prove that the algorithm converges towards the optimum input covariance matrix if it converges at all. Finally, simulation results confirm the relevance of our approach.

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

All along this paper, r and t stand for the number of transmit and receive antennas. Certain

quantities will be studied in the asymptotic regime $t \to \infty$, $r \to \infty$ in such a way that

$$0 < \liminf \frac{t}{r} \leq \limsup \frac{t}{r} < \infty \ .$$

In order to simplify the notations, $t \to +\infty$ should be understood from now on as $t \to \infty$, $r \to \infty$ and $0 < \liminf tr^{-1} \leq \limsup tr^{-1} < \infty$. Let $\mathbf{M}_t$ a matrix whose size depends on t. Then, $\mathbf{M}_t$ is said to be uniformly bounded if $\sup_t \|\mathbf{M}_t\| < +\infty$.

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

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

$$\mathbf{y}(\tau) = \mathbf{H}(\tau)\mathbf{x}(\tau) + \mathbf{z}(\tau) \tag{5.1}$$

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

$$\mathbf{H} = \sqrt{\frac{K}{K+1}}\mathbf{A} + \frac{1}{\sqrt{K+1}}\mathbf{V} \ , \tag{5.2}$$

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

$$\mathbf{V} = \frac{1}{\sqrt{t}}\mathbf{C}^{\frac{1}{2}}\mathbf{W}\tilde{\mathbf{C}}^{\frac{1}{2}} \ , \tag{5.3}$$

where $\mathbf{W} = (W_{ij})$ is a $r \times t$ matrix whose entries are independent and identically distributed (i.i.d.) complex circular Gaussian random variables $\mathcal{C}N(0,1)$. The matrices $\tilde{\mathbf{C}} > 0$ and $\mathbf{C} > 0$ account for the transmit and receive antenna correlation effects respectively and satisfy $\frac{1}{t}\mathrm{Tr}(\tilde{\mathbf{C}}) = 1$ and $\frac{1}{r}\mathrm{Tr}(\mathbf{C}) = 1$. This correlation structure is often referred to as a separable or Kronecker correlation model.

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

$$I(\mathbf{Q}) = \mathbb{E}_{\boldsymbol{H}}\left[\log\det\left(\mathbf{I}_r + \frac{1}{\sigma^2}\mathbf{H}\mathbf{Q}\mathbf{H}^H\right)\right]. \tag{5.4}$$

Maximizing the EMI with respect to the input covariance matrix $\mathbf{Q} = \mathbb{E}(\mathbf{x}\mathbf{x}^H)$ leads to the channel Shannon capacity for *fast* fading MIMO channels i.e. when the channel vary from symbol to symbol. This capacity is achieved by averaging over channel variations over time. For slow fading MIMO channels, i.e. when the channel matrix remains constant over a certain block duration much smaller than the channel coherence time, such an averaging is not possible and one has to communicate

at rates smaller than the ergodic capacity. The maximum EMI is therefore a rate upper bound for *slow* fading MIMO channels and only a fraction of it can be achieved[2]. A more suited performance metric to study slow-fading channels is the outage capacity whose computation would require the knowledge of the variance of the mutual information. This is beyond the aim of this paper where we limit ourselves to the calculation of an asymptotic approximation of the mean of the mutual information. The computations performed in this article can be seen as a first step toward the evaluation of the variance of the EMI and its outage probability.

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

$$C_E = \sup_{\mathbf{Q}\in\mathcal{C}_1} I(\mathbf{Q}). \tag{5.5}$$

The optimal input covariance matrix thus coincides with the argument of the above maximization problem. Note that $I : \mathbf{Q} \mapsto I(\mathbf{Q})$ is a strictly concave function on the convex set $\mathcal{C}_1$, which guarantees the existence of a unique maximum [75]. When $\tilde{\mathbf{C}} = \mathbf{I}_t$, $\mathbf{C} = \mathbf{I}_r$, [56] shows that the eigenvectors of the optimal input covariance matrix coincide with the right-singular vectors of $\mathbf{A}$. By adapting the proof of [56], one can easily check that this result also holds when $\tilde{\mathbf{C}} = \mathbf{I}_t$ and $\mathbf{C}$ and $\mathbf{A}\mathbf{A}^H$ share a common eigenvector basis. Apart from these two simple cases, it seems difficult to find a closed-form expression for the eigenvectors of the optimal covariance matrix. Therefore the evaluation of C_E requires the use of numerical techniques (see e.g. [120]) which are very demanding since they rely on computationally-intensive Monte-Carlo simulations. This problem can be circumvented as the EMI $I(\mathbf{Q})$ can be approximated by a simple expression denoted by $\bar{I}(\mathbf{Q})$ as $t \to \infty$ which in turn will be optimized with respect to $\mathbf{Q}$ (see below).

Remark 5.1. *Finding the optimum covariance matrix is useful in practice, in particular if the channel input is assumed to be Gaussian. In fact, there exist many practical space-time encoders that produce near-Gaussian outputs (these outputs are used as inputs for the linear precoder $\mathbf{Q}^{1/2}$). See for instance [98].*

Summary of the main results. The main contributions of this section can be summarized as follows:

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

   $$\bar{I}(\mathbf{Q}) = \log\det\left[\mathbf{I}_t + \mathbf{G}(\delta(\mathbf{Q},\tilde{\delta}(\mathbf{Q}))\mathbf{Q}\right] + i(\delta(\mathbf{Q}),\tilde{\delta}(\mathbf{Q})) \tag{5.6}$$

   where $\delta(\mathbf{Q})$ and $\tilde{\delta}(\mathbf{Q})$ are two positive terms defined as the solutions of a system of 2 equations (see Eq. (5.7)). The functions $\mathbf{G}$ and i depend on $(\delta(\mathbf{Q}),\tilde{\delta}(\mathbf{Q}))$, K, $\mathbf{A}$, $\mathbf{C}$, $\tilde{\mathbf{C}}$, and on the noise variance σ^2. They are given in closed form.

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

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

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

[2]This fraction is called the multiplexing gain in [126] where the authors introduced the famous diversity multiplexing trade-off.

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

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

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

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

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

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

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

Approximation of the capacity $I(\mathbf{Q}*)$.

Theorem 5.2. *For $\mathbf{Q} \in \mathcal{C}_1$, consider the system of equations*

$$\left\{ \begin{array}{rcl} \delta & = & f(\delta, \tilde{\delta}, \mathbf{Q}) \\ \tilde{\delta} & = & \tilde{f}(\delta, \tilde{\delta}, \mathbf{Q}) \end{array} \right. , \tag{5.7}$$

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

$$f(\delta, \tilde{\delta}, \mathbf{Q}) = \frac{1}{t}\mathrm{Tr}\Big\{\mathbf{C}\Big[\sigma^2\left(\mathbf{I}_r + \frac{\tilde{\delta}}{K+1}\mathbf{C}\right) + \frac{K}{K+1}\mathbf{A}\mathbf{Q}^{\frac{1}{2}}\left(\mathbf{I}_t + \frac{\delta}{K+1}\mathbf{Q}^{\frac{1}{2}}\tilde{\mathbf{C}}\mathbf{Q}^{\frac{1}{2}}\right)^{-1}\mathbf{Q}^{\frac{1}{2}}\mathbf{A}^H\Big]^{-1}\Big\} , \tag{5.8}$$

$$\tilde{f}(\delta, \tilde{\delta}, \mathbf{Q}) = \frac{1}{t}\mathrm{Tr}\Big\{\mathbf{Q}^{\frac{1}{2}}\tilde{\mathbf{C}}\mathbf{Q}^{\frac{1}{2}}\Big[\sigma^2\left(\mathbf{I}_t + \frac{\delta}{K+1}\mathbf{Q}^{\frac{1}{2}}\tilde{\mathbf{C}}\mathbf{Q}^{\frac{1}{2}}\right) + \frac{K}{K+1}\mathbf{Q}^{\frac{1}{2}}\mathbf{A}^H\left(\mathbf{I}_r + \frac{\tilde{\delta}}{K+1}\mathbf{C}\right)^{-1}\mathbf{A}\mathbf{Q}^{\frac{1}{2}}\Big]^{-1}\Big\} . \tag{5.9}$$

Then the system of equations (5.7) has a unique strictly positive solution $(\delta(\mathbf{Q}), \tilde{\delta}(\mathbf{Q}))$.
Furthermore, assume that $\sup_t \|\mathbf{Q}\| < +\infty$, $\sup_t \|\mathbf{A}\| < +\infty$, $\sup_t \|\mathbf{C}\| < +\infty$, and $\sup_t \|\tilde{\mathbf{C}}\| <$

$+\infty$. *Assume also that* $\inf_t \lambda_{\min}(\tilde{\mathbf{C}}) > 0$ *where* $\lambda_{\min}(\tilde{\mathbf{C}})$ *represents the smallest eigenvalue of* $\tilde{\mathbf{C}}$. *Then, as* $t \to +\infty$,

$$I(\mathbf{Q}) = \bar{I}(\mathbf{Q}) + O\left(\frac{1}{t}\right) \tag{5.10}$$

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

$$\begin{aligned}\bar{I}(\mathbf{Q}) = \log\det\left(\mathbf{I}_t + \frac{\delta(\mathbf{Q})}{K+1}\mathbf{Q}^{\frac{1}{2}}\tilde{\mathbf{C}}\mathbf{Q}^{\frac{1}{2}} + \frac{1}{\sigma^2}\frac{K}{K+1}\mathbf{Q}^{\frac{1}{2}}\mathbf{A}^H\left(\mathbf{I}_r + \frac{\tilde{\delta}(\mathbf{Q})}{K+1}\mathbf{C}\right)^{-1}\mathbf{A}\mathbf{Q}^{\frac{1}{2}}\right) \\ + \log\det\left(\mathbf{I}_r + \frac{\tilde{\delta}(\mathbf{Q})}{K+1}\mathbf{C}\right) - \frac{t\sigma^2}{K+1}\delta(\mathbf{Q})\,\tilde{\delta}(\mathbf{Q})\ ,\end{aligned} \tag{5.11}$$

or equivalently by

$$\begin{aligned}\bar{I}(\mathbf{Q}) = \log\det\left(\mathbf{I}_r + \frac{\tilde{\delta}(\mathbf{Q})}{K+1}\mathbf{C} + \frac{1}{\sigma^2}\frac{K}{K+1}\mathbf{A}\mathbf{Q}^{\frac{1}{2}}\left(\mathbf{I}_t + \frac{\delta(\mathbf{Q})}{K+1}\mathbf{Q}^{\frac{1}{2}}\tilde{\mathbf{C}}\mathbf{Q}^{\frac{1}{2}}\right)^{-1}\mathbf{Q}^{\frac{1}{2}}\mathbf{A}^H\right) \\ + \log\det\left(\mathbf{I}_t + \frac{\delta(\mathbf{Q})}{K+1}\mathbf{Q}^{1/2}\tilde{\mathbf{C}}\mathbf{Q}^{1/2}\right) - \frac{t\sigma^2}{K+1}\delta(\mathbf{Q})\,\tilde{\delta}(\mathbf{Q}).\end{aligned} \tag{5.12}$$

Theorem 5.3. *The function* $\mathbf{Q} \mapsto \bar{I}(\mathbf{Q})$ *is strictly concave on* $\mathcal{C}_1$.

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

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

As we shall see, matrix $\overline{\mathbf{Q}}_*$ can be obtained by a rather simple algorithm. Provided that $\sup_t \|\overline{\mathbf{Q}}_*\|$ is bounded, Eq. (5.10) in Theorem 5.2 yields $I(\overline{\mathbf{Q}}_*) - \bar{I}(\overline{\mathbf{Q}}_*) \to 0$ as $t \to \infty$. It remains to check that $I(\mathbf{Q}_*) - I(\overline{\mathbf{Q}}_*)$ goes asymptotically to zero to be able to approximate the capacity. This is the purpose of the next proposition.

Proposition 5.4. *Assume that* $\sup_t \|\mathbf{A}\| < \infty$, $\sup_t \|\tilde{\mathbf{C}}\| < \infty$, $\sup_t \|\mathbf{C}\| < \infty$, $\inf_t \lambda_{\min}(\tilde{\mathbf{C}}) > 0$, *and* $\inf_t \lambda_{\min}(\mathbf{C}) > 0$. *Let* $\overline{\mathbf{Q}}_*$ *and* $\mathbf{Q}_*$ *be the maximizers over* $\mathcal{C}_1$ *of* $\bar{I}$ *and* I *respectively. Then the following facts hold true:*

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

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

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

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

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

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

(5.13)
$$V(\kappa, \tilde{\kappa}, \mathbf{Q}) = \log\det\left(\mathbf{I}_t + \frac{\kappa}{K+1}\,\mathbf{Q}^{\frac{1}{2}}\tilde{\mathbf{C}}\mathbf{Q}^{\frac{1}{2}} + \frac{K}{\sigma^2(K+1)}\,\mathbf{Q}^{\frac{1}{2}}\mathbf{A}^H\left(\mathbf{I}_r + \frac{\tilde{\kappa}}{K+1}\,\mathbf{C}\right)^{-1}\mathbf{A}\mathbf{Q}^{\frac{1}{2}}\right) + \log\det\left(\mathbf{I}_r + \frac{\tilde{\kappa}}{K+1}\mathbf{C}\right) - \frac{t\sigma^2\kappa\tilde{\kappa}}{K+1}\,.$$

or equivalently by

(5.14)
$$V(\kappa, \tilde{\kappa}, \mathbf{Q}) = \log\det\left(\mathbf{I}_r + \frac{\tilde{\kappa}}{K+1}\mathbf{C} + \frac{K}{\sigma^2(K+1)}\mathbf{A}\mathbf{Q}^{\frac{1}{2}}\left(\mathbf{I}_t + \frac{\kappa}{K+1}\mathbf{Q}^{\frac{1}{2}}\tilde{\mathbf{C}}\mathbf{Q}^{\frac{1}{2}}\right)^{-1}\mathbf{Q}^{\frac{1}{2}}\mathbf{A}^H\right) + \log\det\left(\mathbf{I}_t + \frac{\kappa}{K+1}\mathbf{Q}^{1/2}\tilde{\mathbf{C}}\mathbf{Q}^{1/2}\right) - \frac{t\sigma^2\kappa\tilde{\kappa}}{K+1}\,.$$

Note that if $(\delta(\mathbf{Q}), \tilde{\delta}(\mathbf{Q}))$ is the solution of system (5.7), then:

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

Denote by $(\delta_*, \tilde{\delta}_*)$ the solution $(\delta(\overline{\mathbf{Q}}_*), \tilde{\delta}(\overline{\mathbf{Q}}_*))$ of (5.7) associated with $\overline{\mathbf{Q}}_*$. The aim of the section is to prove that $\overline{\mathbf{Q}}_*$ is the solution of the following standard waterfilling problem:

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

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

$$\mathbf{G}(\kappa, \tilde{\kappa}) = \frac{\kappa}{K+1}\tilde{\mathbf{C}} + \frac{K}{\sigma^2(K+1)}\mathbf{A}^H\left(\mathbf{I}_r + \frac{\tilde{\kappa}}{K+1}\,\mathbf{C}\right)^{-1}\mathbf{A}\,. \tag{5.15}$$

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

$$V(\kappa, \tilde{\kappa}, \mathbf{Q}) = \log\det\left(\mathbf{I} + \mathbf{Q}\mathbf{G}(\kappa, \tilde{\kappa})\right) + \log\det\left(\mathbf{I}_r + \frac{\tilde{\kappa}}{K+1}\mathbf{C}\right) - \frac{t\sigma^2\kappa\tilde{\kappa}}{K+1}\,, \tag{5.16}$$

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

Proposition 5.5. *Consider the functions $\delta(\mathbf{Q}), \tilde{\delta}(\mathbf{Q})$ and $\bar{I}(\mathbf{Q})$ from $\mathcal{C}_1$ to $\mathbb{R}$. The following properties hold true:*

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

Remark 5.6. *The quantities δ_* and $\tilde{\delta}_*$ depend on matrix $\overline{\mathbf{Q}}_*$. Therefore, Proposition 5.5 does not provide by itself any optimization algorithm. However, it gives valuable insights on the structure of $\overline{\mathbf{Q}}_*$. Consider first the case $\mathbf{C} = \mathbf{I}$ and $\tilde{\mathbf{C}} = \mathbf{I}$. Then, $\mathbf{G}(\delta_*, \tilde{\delta}_*)$ is a linear combination of $\mathbf{I}$ and*

matrix $\mathbf{A}^H\mathbf{A}$*. The eigenvectors of* $\overline{\mathbf{Q}}_*$ *thus coincide with the right singular vectors of matrix* $\mathbf{A}$*, a result consistent with the work [56] devoted to the maximization of the EMI* $I(\mathbf{Q})$*. If* $\mathbf{C} = \mathbf{I}$ *and* $\tilde{\mathbf{C}} \neq \mathbf{I}$*,* $\mathbf{G}(\delta_*, \tilde{\delta}_*)$ *can be interpreted as a linear combination of matrices* $\tilde{\mathbf{C}}$ *and* $\mathbf{A}^H\mathbf{A}$*. Therefore, if the transmit antennas are correlated, the eigenvectors of the optimum matrix* $\overline{\mathbf{Q}}_*$ *coincide with the eigenvectors of some weighted sum of* $\tilde{\mathbf{C}}$ *and* $\mathbf{A}^H\mathbf{A}$*. This result provides a simple explanation of the impact of correlated transmit antennas on the structure of the optimal input covariance matrix. The impact of correlated receive antennas on* $\overline{\mathbf{Q}}_*$ *is however less intuitive because matrix* $\mathbf{A}^H\mathbf{A}$ *has to be replaced with* $\mathbf{A}^H(\mathbf{I} + \tilde{\delta}_*\mathbf{C})^{-1}\mathbf{A}$*.*

The optimization algorithm.

We are now in position to introduce our maximization algorithm of $\bar{I}$. It is mainly motivated by the simple observation that for each fixed $(\kappa, \tilde{\kappa})$, the maximization w.r.t. $\mathbf{Q}$ of function $V(\kappa, \tilde{\kappa}, \mathbf{Q})$ defined by (5.16) can be achieved by a standard waterfilling procedure, which, of course, does not need the use of numerical techniques. On the other hand, for $\mathbf{Q}$ fixed, the equations (5.7) have unique solutions that, in practice, can be obtained using a standard fixed-point algorithm. Our algorithm thus consists in adapting parameters $\mathbf{Q}$ and $\delta, \tilde{\delta}$ separately by the following iterative scheme:

- Initialization: $\mathbf{Q}_0 = \mathbf{I}$, $(\delta_1, \tilde{\delta}_1)$ are defined as the unique solutions of system (5.7) in which $\mathbf{Q} = \mathbf{Q}_0 = \mathbf{I}$. Then, define $\mathbf{Q}_1$ are the maximum of function $\mathbf{Q} \to V(\delta_1, \tilde{\delta}_1, \mathbf{Q})$ on $\mathcal{C}_1$, which is obtained through a standard waterfilling procedure.
- Iteration k: assume $\mathbf{Q}_{k-1}$, $(\delta_{k-1}, \tilde{\delta}_{k-1})$ available. Then, $(\delta_k, \tilde{\delta}_k)$ is defined as the unique solution of (5.7) in which $\mathbf{Q} = \mathbf{Q}_{k-1}$. Then, define $\mathbf{Q}_k$ are the maximum of function $\mathbf{Q} \to V(\delta_k, \tilde{\delta}_k, \mathbf{Q})$ on $\mathcal{C}_1$.

One can notice that this algorithm is the generalization of the procedure used by [121] for optimizing the input covariance matrix for correlated Rayleigh MIMO channels.

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

Proposition 5.7. *Assume that the two sequences* $(\delta_k)_{k\geq 0}$ *and* $(\tilde{\delta}_k)_{k\geq 0}$ *verify*

$$\lim_{k\to+\infty} \delta_k - \delta_{k-1} \to 0, \lim_{k\to+\infty} \tilde{\delta}_k - \tilde{\delta}_{k-1} \to 0 \tag{5.17}$$

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

Remark 5.8. *If the algorithm is convergent, i.e. if sequence* $(\mathbf{Q}_k)_{k\geq 0}$ *converges towards a matrix* $\mathbf{P}_*$*, Proposition 5.7 implies that* $\mathbf{P}_* = \overline{\mathbf{Q}}_*$*. In fact, functions* $\mathbf{Q} \mapsto \delta(\mathbf{Q})$ *and* $\mathbf{Q} \mapsto \tilde{\delta}(\mathbf{Q})$ *are continuous by Proposition 5.5. As* $\delta_k = \delta(\mathbf{Q}_{k-1})$ *and* $\tilde{\delta}_k = \tilde{\delta}(\mathbf{Q}_{k-1})$*, the convergence of* $(\mathbf{Q}_k)$ *thus implies the convergence of* (δ_k) *and* $(\tilde{\delta}_k)$*, and (5.17) is fulfilled. Proposition 5.7 immediately yields* $\mathbf{P}_* = \overline{\mathbf{Q}}_*$*. Although we have not been able to prove the convergence of the algorithm, the above result is encouraging, and tends to indicate the algorithm is reliable. In particular, all the numerical experiments we have conducted indicates that the algorithm converges towards a certain matrix which must coincide by Proposition 5.7 with* $\overline{\mathbf{Q}}_*$*.*

Numerical experiments. *How large do the numbers of antennas need to be to reach the asymptotic regime?* All our analysis is based on the approximation of the ergodic mutual information.

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

- $\mathbf{Q} = \mathbf{I}_t$.
- The chosen line-of-sight (LOS) component $\mathbf{A}$ is based on the following typical structure:

$$\mathbf{A} = \frac{1}{\sqrt{t}} \left[\mathbf{a}(\theta_1), \ldots, \mathbf{a}(\theta_t)\right] \mathbf{\Lambda} \ , \tag{5.18}$$

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

  The angle of arrivals are chosen randomly according to a uniform distribution.

- Antenna correlation is assumed to decrease exponentially with the inter-antenna distance i.e. $\tilde{\mathbf{C}}_{ij} \sim \rho_T^{|i-j|}$, $\mathbf{C}_{ij} \sim \rho_R^{|i-j|}$ with $0 \leq \rho_T \leq 1$ and $0 \leq \rho_R \leq 1$.
- K is equal to 1.

Figure 5.1 represents the EMI $I(\mathbf{I})$ evaluated by Monte Carlo simulations and its approximation $\bar{I}(\mathbf{I})$ as well as their relative difference (in percentage). Here, the correlation coefficients are equal to $(\rho_T, \rho_R) = (0.8, 0.3)$ and three different pairs of numbers of antenna are considered: $(t, r) \in \{(2, 2), (4, 4), (8, 8)\}$. Figure 5.1 shows that the approximation is reliable even for $r = t = 2$ in a wide range of SNR.

Comparison with the Vu-Paulraj method. In this paragraph, we compare our algorithm with the method presented in [120] based on the maximization of $I(\mathbf{Q})$. We recall that Vu-Paulraj's algorithm is based on a Newton method and a barrier interior point method. Moreover, the average mutual informations and their first and second derivatives are evaluated by Monte-Carlo simulations. In fig. 5.3, we have evaluated $C_E = \max_{\mathbf{Q} \in \mathcal{C}_1} I(\mathbf{Q})$ versus the SNR for $r = t = 4$. Matrix $\mathbf{H}$ coincides with the example considered in [120]. The solid line corresponds to the results provided by the Vu-Paulraj's algorithm; the number of trials used to evaluate the mutual informations and its first and second derivatives is equal to 30.000, and the maximum number of iterations of the algorithm in [120] is fixed to 10. The dashed line corresponds to the results provided by our algorithm: each point represent $I(\overline{\mathbf{Q}}_*)$ at the corresponding SNR, where $\overline{\mathbf{Q}}_*$ is the "optimal" matrix provided by our approach; the average mutual information at point $\overline{\mathbf{Q}}_*$ is evaluted by Monte-Carlo simulation (30.000 trials are used). The number of iterations is also limited to 10. Figure 5.3 shows that our asymptotic approach provides the same results than the Vu-Paulraj's algorithm. However, our algorithm is computationally much more efficient as the above table shows. The table gives the average executation time (in sec.) of one iteration for both algorithms for $r = t = 2, r = t = 4, r = t = 8$.

In fig. 5.4, we again compare Vu-Paulraj's algorithm and our proposal. Matrix $\mathbf{A}$ is generated according to (5.18), the angles being chosen at random. The transmit and receive antennas correlations are exponential with parameter $0 < \rho_T < 1$ and $0 < \rho_R < 1$ respectively. In the experiments, $r = t = 4$, while various values of ρ_T, ρ_R and of the Rice factor K have been considered. As in the previous experiment, the maximum number of iterations for both algorithms is 10, while the number of trials generated to evaluate the average mutual informations and their derivatives is equal to 30.000. Our approach again provides the same results than Vu-Paulraj's algorithm, except for low SNRs for $K = 1, \rho_T = 0.5, \rho_R = 0.8$ where our method gives better results: at these points, the Vu-Paulraj's algorithm seems not to have converge at the 10th iteration.

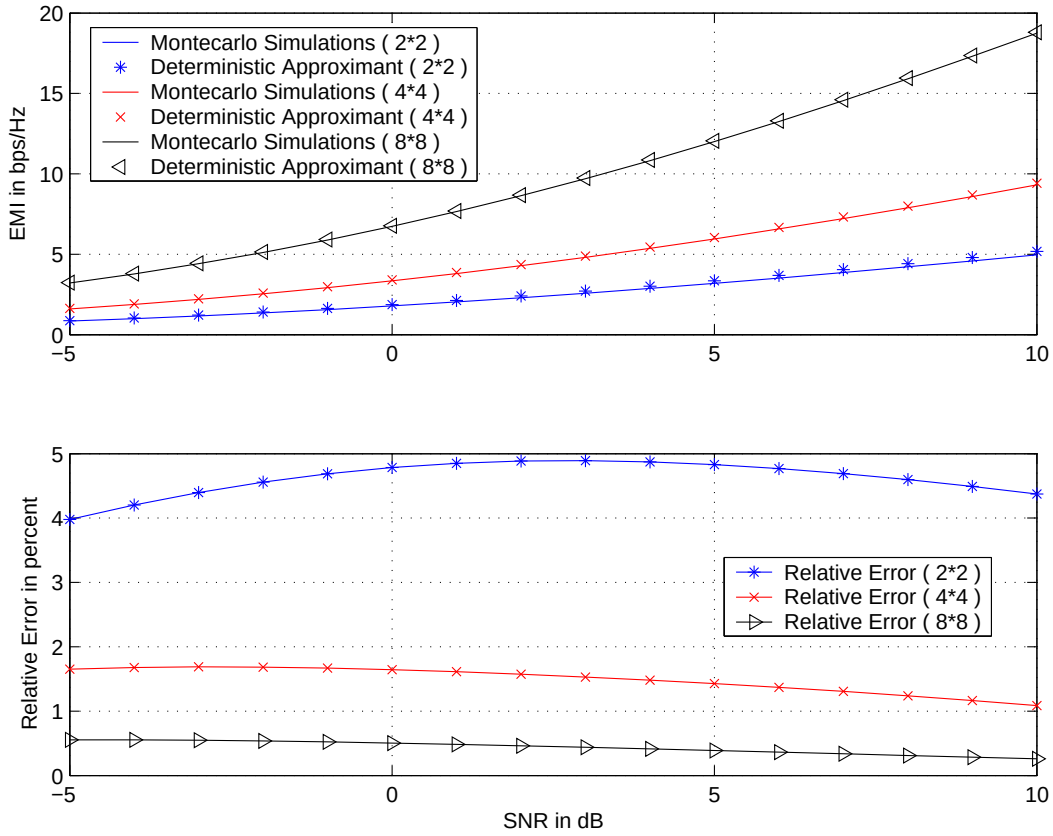

Figure 5.1: The large system approximation is accurate for correlated Rician MIMO channels. The relative difference between the EMI approximation and that obtained by Monte-Carlo simulations is less than 5 % for a 2×2 system and less than 1 % for a 8×8 system.

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

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

5.2 Bit Error Rate and Outage probability approximation

Introduction. In this section, we focus on the study of the Signal to interference plus noise ratio (SINR) for the linear Wiener receiver, also called LMMSE for Linear Minimum Mean Squared Error receiver. In this context, an *outage* event occurs when the SINR at the LMMSE output lies beneath a given threshold. One purpose of this section is to approximate the associated outage probability

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

Figure 5.2: Average time per iteration in seconds

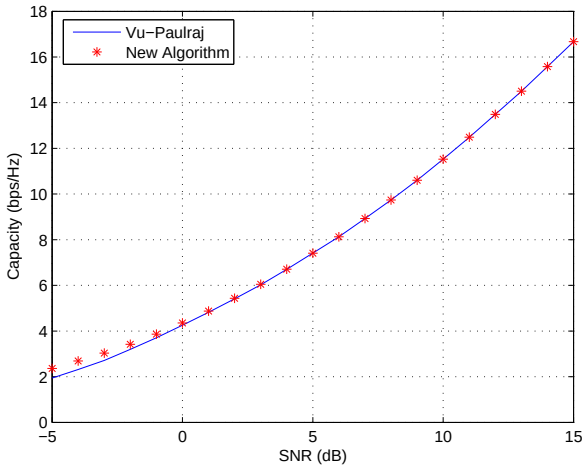

Figure 5.3: Comparison with the Vu-Paulraj algorithm II

for an important class of MIMO channel models. Another performance index associated with the SINR is the Bit Error Rate (BER) which will be also studied herein.

Outage probability approximations has been provided in recent works for various channels, under very specific technical conditions (in the case where the moment generating function [71] or the probability density function [42] have closed form expressions; when a first order expansion of the probability density function can be derived [60]; in the more general case where the moment generating function can be approximated by using Padé approximations [108]; etc.). All these results deal with specific situations where the statistics of the SINR could be derived for finite system dimensions.

Alternatively, by making use of large random matrix theory, one can study the behavior of the SINR in the asymptotic regime where the channel matrix dimensions grow to infinity. For fairly general channel statistical models, it is then possible to prove the convergence of the SINR to deterministic values and even establish its asymptotic normality (see for instance [64]). However, this Gaussian approximation is not accurate when the channel dimensions are small. This is confirmed in [48] where it is shown that the asymptotic BER based on the sole Gaussian approximation is significantly smaller than the empirical estimate. A more precise approximation of the BER or the outage probability is expected if one chooses to approximate the SINR probability distribution with a distribution 1) which is supported by $\mathbb{R}_+$ (indeed, a Gaussian random variable takes negative values which is not realistic), 2) which is adjusted to the first three moments of the SINR instead of the first two moments needed by the Gaussian approximation.

In this line of thought, Li, Paul, Narasimhan and Cioffi [73] proposed to use alternative parameterized distributions (Gamma and generalized Gamma distributions) whose parameters are

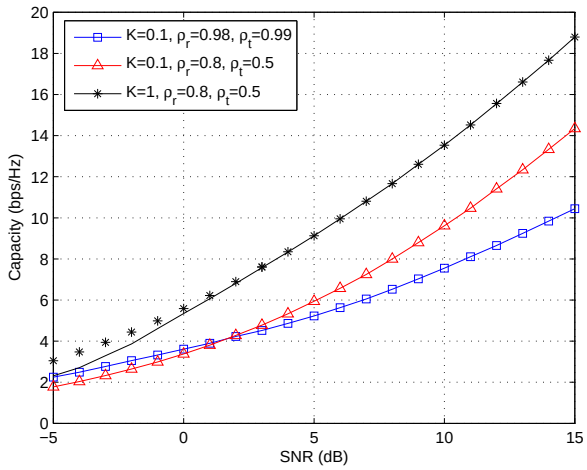

Figure 5.4: Comparison with the Vu-Paulraj algorithm III

set to coincide with the asymptotic moments of the output SINR. This approach was derived for (transmit) correlated channels and asymptotic moments were provided for the special case of uncorrelated or equicorrelated channels. For the general correlated channel case, only limiting upper bounds for the first three asymptotic moments were provided. Based on Random Matrix Theory and especially on the Gaussian mathematical tools elaborated in [50] and further used in [31], we derive closed-form expressions for the first three moments, generalizing the work of [73] to a general (receive) correlated channel. Using the generalized Gamma approximation, we provide closed-form expressions for the BER and numerical approximations for the outage probability.

System Model and SINR expression. We consider an uplink transmission system, in which a base station equipped by N correlated antennas detects the symbols of a given user of interest in the presence of K interfering users. The N dimensional received signal writes $\mathbf{r} = \mathbf{\Sigma s} + \mathbf{n}$, where $\mathbf{s} = [s_0, \cdots, s_K]^{\mathrm{T}}$ is the transmitted complex vector signal with size $K+1$ satisfying $\mathbb{E}\mathbf{s}\mathbf{s}^* = \mathbf{I}_{K+1}$, and $\mathbf{\Sigma}$ is the $N \times (K+1)$ channel matrix. We assume that this matrix writes as

$$\mathbf{\Sigma} = \frac{1}{\sqrt{K}} \mathbf{\Psi}^{\frac{1}{2}} \mathbf{W} \mathbf{P}^{\frac{1}{2}},$$

where $\mathbf{\Psi}$ is a $N \times N$ Hermitian nonnegative matrix that captures the correlations at the receiver, $\mathbf{P} = \mathrm{diag}\,(p_0, \cdots, p_K)$ is the deterministic matrix of the powers allocated to the different users and $\mathbf{W} = [\mathbf{w}_0, \cdots, \mathbf{w}_K]$ ($\mathbf{w}_k$ being the kth column) is a $N \times (K+1)$ complex Gaussian matrix with centered unit variance (standard) i.i.d entries. To detect symbol s_0 and to mitigate the interference caused by users $1, \ldots, K$, the base station applies the LMMSE estimator, which minimizes the following metric $\mathbf{g} = \min_{\mathbf{h}} \mathbb{E}\left|\mathbf{h}^*\mathbf{r} - s_0\right|^2$. Let $\mathbf{y} = \sqrt{\frac{p_0}{K}} \mathbf{\Psi}^{\frac{1}{2}} \mathbf{w}_0$, then it is well known that the LMMSE estimator is given by:

$$\mathbf{g} = \left(\mathbf{\Sigma}\mathbf{\Sigma}^* + \rho \mathbf{I}_N\right)^{-1} \mathbf{y}.$$

Writing the received vector $\mathbf{r} = s_0\mathbf{y} + \mathbf{r}_{\mathrm{in}}$ where $s_0\mathbf{y}$ is the relevant term and $\mathbf{r}_{\mathrm{in}}$ represents the interference plus noise term, the SINR at the output of the LMMSE estimator is given by : $\beta_K = \left|\mathbf{g}^*\mathbf{y}\right|^2 / \mathbb{E}\left|\mathbf{g}^*\mathbf{r}_{\mathrm{in}}\right|^2$. Plugging the expression of $\mathbf{g}$ given above into this expression, one can

show that the SINR β_K is given by:

$$\beta_K = \mathbf{y}^* \left(\frac{1}{K} \mathbf{\Psi}^{\frac{1}{2}} \widetilde{\mathbf{W}} \widetilde{\mathbf{P}} \widetilde{\mathbf{W}}^* \mathbf{\Psi}^{\frac{1}{2}} + \rho \mathbf{I}_N \right)^{-1} \mathbf{y},$$

with $\widetilde{\mathbf{P}} = \mathrm{diag}\,(p_1, \cdots, p_K)$ and $\widetilde{\mathbf{W}} = [\mathbf{w}_1, \cdots, \mathbf{w}_K]$. Let $\mathbf{\Psi} = \mathbf{U}\mathbf{D}\mathbf{U}^*$ be a spectral decomposition of $\mathbf{\Psi}$. Then, β_K writes:

$$\beta_K \quad = \quad \frac{p_0}{\rho K} \mathbf{z}^* \mathbf{D}^{\frac{1}{2}} \left(\frac{1}{K\rho} \mathbf{D}^{\frac{1}{2}} \mathbf{Z} \tilde{\mathbf{D}} \mathbf{Z}^* \mathbf{D}^{\frac{1}{2}} + \mathbf{I} \right)^{-1} \mathbf{D}^{\frac{1}{2}} \mathbf{z}$$

where: $\mathbf{z} = \mathbf{U}^* \mathbf{w}_0$ (resp. $\mathbf{Z} = \mathbf{U}^* \widetilde{\mathbf{W}}$) is a $N \times 1$ vector with complex independent standard Gaussian entries (resp. $N \times K$ matrix with independent Gaussian entries).

Under appropriate assumptions, it can be proved that β_K admits a deterministic approximation as $K, N \to \infty$, the ratio being bounded below by a positive constant and above by a finite constant. Furthermore, its fluctuations can be precisely described under the same asymptotic regime (for a full and rigorous computation based on random matrix theory, see[64]). As it will appear shortly, a deterministic approximation of the third centered moment of β_K is needed and will be computed in the sequel.

A quick reminder of the generalised Gamma distribution. Recall that if a random variable X follows a generalized gamma distribution $G(\alpha, b, \xi)$, where α and b are respectively referred to as the shape and scale parameters, then:

$$\mathbb{E}X = \alpha b, \quad \mathrm{var}(X) = \alpha b^2 \quad \text{and} \quad \mathbb{E}(X - \mathbb{E}X)^3 = (\xi + 1)\alpha b^3 \ .$$

The probability density function (pdf) of the generalized Gamma distribution with parameters (α, b, ξ) does not have a closed form expression but its moment generating function (MGF) writes:

$$\mathrm{MGF}(s) = \begin{cases} \exp(\frac{\alpha}{\xi - 1}(1 - (1 - b\xi s)^{\frac{\xi - 1}{\xi}})) & \text{if } \xi > 1, \\ \exp(\frac{\alpha}{1 - \xi}((1 - b\xi s)^{\frac{\xi - 1}{\xi}} - 1)) & \text{if } \xi \leq 1. \end{cases}$$

BER approximation. Under QPSK constellations with Gray encoding and assuming that the noise at the LMMSE output is Gaussian, the BER is given by:

$$\mathrm{BER} = \mathbb{E}Q(\sqrt{\beta_K})$$

where $Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-t^2/2}\, dt$ and the expectation is taken over the distribution of the SINR β_K. Based on the asymptotic normality of the SINR, [47] and [97] proposed to use the limiting BER value given by:

$$\mathrm{BER} = \frac{1}{\sqrt{2\pi}} \int_{\sqrt{\overline{\beta}_K}}^\infty e^{-t^2/2} dt,$$

where $\overline{\beta}_K$ denotes an asymptotic deterministic approximation of the first moment of β_K. It was shown however in [73] that this expression is inaccurate since a Gaussian random variable allows negative values and has a zero third moment while the output SINR is always positive and has a non-zero third moment for finite system dimensions. To overcome these difficulties, Li *et al.* [73] approximate the BER by considering first that the SINR follows a Gamma distribution with scale α and shape b, these parameters being tuned by equating the first two moments of the Gamma

distribution with the first two asymptotic moments of the SINR. However, the third asymptotic moment was shown to be different from the third moment of the Gamma distribution which only depends on the scale α and shape b. In light of this consideration, Li *et al.* [73] refine this approximation and consider that the SINR follows a generalized Gamma distribution which is adjusted by assuming that its first three moments equate the first three asymptotic moments of the SINR. As expected, this approximation has proved to be more accurate than the Gamma approximation, and so will be the one considered in this paper. Next, we briefly review this technique, which we will rely on to provide accurate approximations for the BER and outage probability.

Let $\mathbb{E}_\infty(\beta_K)$, $\mathrm{var}_\infty(\beta_K)$ and $\mathrm{S}_\infty(\beta_K)$ denote respectively the deterministic approximations of the asymptotic central moments of β_K. Then, the parameters ξ, α and b are determined by solving:

$$\mathbb{E}_\infty(\beta_K) = \alpha b, \quad \mathrm{var}_\infty(\beta_K) = \alpha b^2 \quad \text{and} \quad \mathrm{S}_\infty(\beta_K) = (\xi+1)\alpha b^3,$$

thus giving the following values:

$$\alpha = \frac{(\mathbb{E}_\infty(\beta_K))^2}{\mathrm{var}_\infty(\beta_K)}, \quad \beta = \frac{\mathrm{var}_\infty(\beta_K)}{\mathbb{E}_\infty(\beta_K)} \quad \text{and} \quad \xi = \frac{\mathrm{S}_\infty(\beta_K)\mathbb{E}_\infty(\beta_K)}{(\mathrm{var}_\infty(\beta_K))^2} - 1.$$

Using the MGF, one can evaluate the BER by using the following relation [104], that holds for QPSK constellation:

$$\mathrm{BER} = \frac{1}{\pi}\int_0^{\frac{\pi}{2}} \mathrm{MGF}\left(-\frac{1}{2\sin^2\phi}\right) d\phi. \tag{5.19}$$

Note that similar expressions for the BER exist for other constellations and can be derived by plugging the following identity involving the function $Q(x)$ [104]:

$$Q(x) = \frac{1}{\pi}\int_0^{\frac{\pi}{2}} \exp\left(-\frac{x^2}{2\sin^2\theta}\right) d\theta$$

into the BER expression.

Outage probability approximation. Only the moment generation function (MGF) has a closed form expression. Knowing the MGF, one can compute numerically the cumulative distribution function by applying the saddle point approximation technique [17]. Denote by $K(y) = \log(\mathrm{MGF}(y))$ the cumulative generating function, by y the threshold SINR and by t_y the solution of $K'(t_y) = y$. Let w_0 and u_0 be given by: $w_0 = \mathrm{sign}(t_y)\sqrt{2\left(t_y y - K(t_y)\right)}$ and $u_0 = t_y\sqrt{K"(t_y)}$. The saddle point approximate of the outage probability is given by:

$$\mathrm{P}_{out} = \Phi(w_0) + \phi(w_0)\left(\frac{1}{w_0} - \frac{1}{u_0}\right), \tag{5.20}$$

where $\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\, dt$ and $\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ denote respectively the standard normal cumulative distribution function and probability distribution function.

So far, we have presented the technique that will be used in simulations for the evaluation of the BER and outage probability. This technique is heavily based on the computation of the three first asymptotic moments of the SINR β_K, an issue that is handled in the next section.

Assumptions. Recall from Section 5.2 the various definitions $K, N, \mathbf{D}, \widetilde{\mathbf{D}}$. In the following, we assume that both K and N go to $+\infty$, their ratio being bounded below and above as follows:

$$0 \ < \ \ell^- = \liminf \frac{K}{N} \ \leq \ \ell^+ = \limsup \frac{K}{N} \ < \ +\infty \ .$$

In the sequel, the notation $K \to \infty$ will refer to this asymptotic regime. We will frequently write $\mathbf{D}_K$ and $\widetilde{\mathbf{D}}_K$ to emphasize the dependence in K, but may drop the subscript K as well. Assume the following mild conditions:

Assumption A-14. *There exist real numbers $d_{\max} < \infty$ and $\tilde{d}_{\max} < \infty$ such that:*

$$\sup_K \|\mathbf{D}_K\| \leq d_{\max} \quad \text{and} \quad \sup_K \|\widetilde{\mathbf{D}}_K\| \leq \tilde{d}_{\max},$$

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

Assumption A-15. *The normalized traces of $\mathbf{D}_K$ and $\widetilde{\mathbf{D}}_K$ satisfy:*

$$\inf_K \frac{1}{K}\mathrm{Tr}(\mathbf{D}_K) > 0 \quad \text{and} \quad \inf_K \frac{1}{K}\mathrm{Tr}(\widetilde{\mathbf{D}}_K) > 0.$$

Asymptotic moments computation In this section, we provide closed form expressions for the first three asymptotic moments. We shall first introduce some deterministic quantities that are used for the computation of the first, second and third asymptotic moments.

Proposition 5.9. *(cf. [50]) For every integer K and any $t > 0$, the system of equations in $(\delta, \tilde{\delta})$*

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

admits a unique solution $\left(\delta_K(t), \tilde{\delta}_K(t)\right)$ satisfying $\delta_K(t) > 0$, $\tilde{\delta}_K(t) > 0$.

Let $\mathbf{T}$ and $\widetilde{\mathbf{T}}$ be the $N \times N$ and $K \times K$ diagonal matrices defined by:

$$\mathbf{T} = \left(\mathbf{I} + t\tilde{\delta_K}\mathbf{D}\right)^{-1} \quad \text{and} \quad \widetilde{\mathbf{T}} = \left(\mathbf{I} + t\delta_K \widetilde{\mathbf{D}}\right)^{-1}.$$

Note that in particular: $\delta = \frac{1}{K}\mathrm{Tr}\mathbf{D}\mathbf{T}$ and $\tilde{\delta} = \frac{1}{K}\mathrm{Tr}\widetilde{\mathbf{D}}\widetilde{\mathbf{T}}$. Define also γ and $\tilde{\gamma}$ as $\gamma = \frac{1}{K}\mathrm{Tr}\mathbf{D}^2\mathbf{T}^2$ and $\tilde{\gamma} = \frac{1}{K}\mathrm{Tr}\widetilde{\mathbf{D}}^2\widetilde{\mathbf{T}}^2$. Finally, replace t by $\frac{1}{\rho}$ and introduce the following deterministic quantities:

$$\Omega_K^2 = \frac{\gamma}{\rho^2}\left(\frac{\gamma\tilde{\gamma}}{\rho^2 - \gamma\tilde{\gamma}} + 1\right) \quad \text{and} \quad \nu_K = \frac{2\rho^3}{K\left(\rho^2 - \gamma\tilde{\gamma}\right)^3}\left[\mathrm{Tr}\mathbf{D}^3\mathbf{T}^3 - \frac{\gamma^3}{\rho^3}\mathrm{Tr}\widetilde{\mathbf{D}}^3\widetilde{\mathbf{T}}^3\right].$$

As usual, the notation $\alpha_K = \mathcal{O}(\beta_K)$ means that $\alpha_K(\beta_K)^{-1}$ is uniformly bounded as $K \to \infty$. Then, the first three asymptotic moments are given by the following theorem:

Theorem 5.10. *Assuming that the matrices $\mathbf{D}$ and $\widetilde{\mathbf{D}}$ satisfy the conditions stated in 14 and 15, then the following convergences hold true:*

1. *First asymptotic moment [63, 64]:*

$$\frac{\delta_K}{\rho} = \mathcal{O}(1) \quad \textit{and} \quad \mathbb{E}\left(\frac{\beta_K}{p_0}\right) - \frac{\delta_K}{\rho} \xrightarrow[K\to\infty]{} 0,$$

2. Second asymptotic moment [63, 64]:

$$\Omega_K = \mathcal{O}(1) \quad \textit{and} \quad K\mathbb{E}\left(\frac{\beta_K}{p_0} - \mathbb{E}\left(\frac{\beta_K}{p_0}\right)\right)^2 - \Omega_K^2 \xrightarrow[K\to\infty]{} 0,$$

3. Third asymptotic moment:

$$\nu_K = \mathcal{O}(1) \quad \textit{and} \quad K^2\mathbb{E}\left(\frac{\beta_K}{p_0} - \mathbb{E}\left(\frac{\beta_K}{p_0}\right)\right)^3 - \nu_K \xrightarrow[K\to\infty]{} 0.$$

The two first items of the theorem are proved in [64] (beware that the notations used in this article are the same as those in [50] and slightly differ from those used in [64]).

Remark 5.11. *One can note that the third asymptotic moment is of order $\mathcal{O}(K^{-2})$. This is in accordance with the asymptotic normality of the SINR, where the third moment of $\sqrt{K}(\beta_K - \mathbb{E}(\beta_K))$ will eventually vanish, as this quantity becomes closer to a Gaussian random variable. However, its value remains significant for small dimension systems.*

5.3 Collaborative sensing

In the context of cognitive networks [81], sensing is one the major of steps in order for the flexible network to adapt its parameters to the environment context. In general, the sensing procedure requires the knowledge of the noise variance as well as a high number of samples for a successful test. This is rarely compatible with the mobile constraints of the users and has pushed the community to propose alternative methods based on collaborative sensing to reduce the number of samples required [19, 125]. The techniques proposed trade the time dimension (samples) with the space dimension (antennas or base stations) and do not require the knowledge of the noise variance, which is one of the drawbacks of energy detector techniques [117]. The general idea of these new techniques compute some functionals of the eigenvalues of the sample covariance matrix which cancel out the noise variance. The work was initiated in [125, 19] using the ratio of the extremal eigenvalues and further formally justified in [12] through the derivation of Generalized Likelihood Ratio Test (GLRT). It was shown in particular that the normalized (by the trace) of the maximum eigenvalue of the sample covariance matrix represents an adequate statistic. In this contribution, we provide an analytical study of both tests using recent results of asymptotic random matrix theory based on spiked models. Using Large Deviation Principles (LDP), we compute the error exponents of the probability of error and show the superiority of the test proposed in [12] through the Bahadur efficiency [118].

The section is articulated as follows: next Section focuses on the problem formulation and contains the signal model, while in Section III the two tests are detailed. The asymptotic analytical study is detailed in Section IV with some simulations. Section V is devoted to the evaluation of the Bahadur relative efficiency of the tests. Conclusions are given in Section V.

5.3.1 Signal Model

Consider a secondary wireless network formed by K nodes, working in sensing mode. We assume that all K nodes are simultaneously sensing a given sub-band $\mathcal{B}$ of the spectrum. For each $k = 1, \ldots, K$, we denote by $y_k(n)$ the complex envelope of the signal received by the kth sensor in band

$\mathcal{B}$ after proper filtering and sampling. Denote by $\boldsymbol{y}(n) = [y_1(n), \ldots, y_K(n)]^T$ the vector obtained when stacking all K sensors' observations at time n into a column vector. The aim is to detect the presence of a primary transmitter in band $\mathcal{B}$. We respectively denote by H_0 and H_1 the hypotheses corresponding to the case where "band $\mathcal{B}$ is free" and "a primary device is already transmitting in band $\mathcal{B}$":

$$\boldsymbol{y}(n) = \begin{cases} \boldsymbol{w}(n): & H_0 \\ \boldsymbol{h}\, s(n) + \boldsymbol{w}(n): & H_1 \end{cases}, \tag{5.21}$$

where $\boldsymbol{w}(n)$ represents a complex circular temporally-white Gaussian noise vector with zero mean and covariance matrix equal to $\sigma^2 \mathbf{I}_K$. In the H_1-case, vector $\boldsymbol{h} \in \mathbb{C}^{K1}$ represents the complex-valued Single-Input Multiple-Output (SIMO) channel between the primary transmitter and the K receiving nodes. Sequence $s(n)$ denotes the unknown data process sent by the active primary device. Sequence $s(n)$ is assumed to be an independent identically distributed (i.i.d.) zero mean random sequence. We assume without restriction that $s(n)$ has unit variance. In order to be able to derive hypothesis testing procedures and to analyze their performance in terms of probability of false alarm and power, we make the usual assumption (see [9] and ref. therein) that the transmitted symbols are Gaussian distributed, say $s(n) \sim \mathcal{CN}(0, 1)$. We assume that

- the noise variance σ^2 is unknown,
- the channel matrix $\boldsymbol{h}$ is unknown.

In the sequel, we denote by N the number of samples observed by each sensor k. Consider the following KN data matrix $\mathbf{Y}$:

$$\mathbf{Y} = [\boldsymbol{y}(0), \ldots, \boldsymbol{y}(N-1)]\,. \tag{5.22}$$

In order to test hypothesis H_0 versus H_1, the aim is to construct a relevant test function $\varphi : \mathbb{C}^{KN} \to \{0, 1\}$ with the sense that one decides hypothesis H_0 (resp. H_1) whenever $\varphi(\mathbf{Y}) = 0$ (resp. $\varphi(\mathbf{Y}) = 1$). As usual, we restrict ourselves to the search for test functions such that the probability of false alarm does not exceed a predefined level α *i.e.*,

$$\mathbb{P}_{H_0}\left[\varphi(\mathbf{Y}) = 1\right] \leq \alpha\,, \tag{5.23}$$

where $\mathbb{P}_{H_i}[\mathcal{E}]$ represents the probability of a given event $\mathcal{E}$ under hypothesis H_i, $i = 0, 1$. On the otherhand, the power of the test is given by $\mathbb{P}_{H_1}\left[\varphi(\mathbf{Y}) = 1\right]$.

5.3.2 Eigen-based Hypothesis Tests

Generalized Likelihood Ratio Test. We respectively denote by $p_0(\mathbf{Y}; \sigma^2)$ and $p_1(\mathbf{Y}; \boldsymbol{h}, \sigma^2)$ the likelihood functions of the observation matrix $\boldsymbol{y}$ indexed by the unknown parameters $\boldsymbol{h}$ and σ^2 under hypotheses H_0 and H_1 respectively:

$$p_0(\mathbf{Y}; \sigma^2) = (\pi\sigma^2)^{-NK} \exp\left(-\frac{N}{\sigma^2}\mathrm{Tr}\,\hat{\mathbf{R}}\right) \tag{5.24}$$

$$p_1(\mathbf{Y}; \boldsymbol{h}, \sigma^2) = (\pi^K \det \mathbf{R})^{-N} \exp\left(-N\mathrm{Tr}\left(\hat{\mathbf{R}}\mathbf{R}^{-1}\right)\right) \tag{5.25}$$

where $\mathbf{R} = \mathbf{R}(\boldsymbol{h}, \sigma^2)$ is the true covariance matrix under H_1 defined by

$$\mathbf{R} = \boldsymbol{h}\boldsymbol{h}^H + \sigma^2 \mathbf{I}_K$$

and where $\hat{\mathbf{R}}$ is the sampled covariance matrix:

$$\hat{\mathbf{R}} = \frac{1}{N}\mathbf{Y}\mathbf{Y}^H \ .$$

In the ideal case where parameters $\boldsymbol{h}$ and σ^2 are supposed to be available, it is well known that a uniformly most powerful test is obtained through the computation of the likelihood ratio statistic $p_0(\mathbf{Y};\sigma^2)/p_1(\mathbf{Y};\boldsymbol{h},\sigma^2)$. Unfortunately, parameters $\boldsymbol{h}$ and σ^2 are unknown in our context so that a uniformly powerful test can no longer be that easily defined. In this case, a suboptimal but classical approach consists in replacing the true likelihood ratio by the following generalized likelihood ratio (GLR)

$$L_N = \frac{\sup_{\sigma^2} p_0(\mathbf{Y};\sigma^2)}{\sup_{\boldsymbol{h},\sigma^2} p_1(\mathbf{Y};\boldsymbol{h},\sigma^2)} \ . \tag{5.26}$$

In the GLRT procedure, one rejects hypothesis H_0 whenever $L_N < \xi_N$, where ξ_N is a certain threshold which is selected so that the probability of false alarm (5.23) does not exceed a given level α. A closed form expression of the above GLR L_N has been recently derived in [9, 12]. Denote by $\lambda_1 > \lambda_2 \cdots > \lambda_K \geq 0$ the ordered eigenvalues of $\hat{\mathbf{R}}$ (all distincts with probability one). As we shall see below, the GLR can be written as a function of the ratio

$$T_N^{(1)} = \frac{\lambda_1}{\frac{1}{K}\operatorname{Tr}\hat{\mathbf{R}}} \ . \tag{5.27}$$

The following result is a direct consequence of [9, 12].

Proposition 5.12. *The GLR (5.26) writes*

$$L_N = C\, T_N^{(1)} \left(1 - \frac{T_N^{(1)}}{K}\right)^{K-1}$$

where $C = \left(1 - \frac{1}{K}\right)^{K-1}$ *is a constant.*

By definition, $T_N^{(1)}$ belongs to the interval $(1, K)$ with probability one. Now it is straightforward to show that function $x \mapsto x\left(1 - \frac{x}{K}\right)^{K-1}$ is decreasing for x in the interval $(1, K)$. Therefore, we propose the following test.

Proposed hypothesis test.

$$T_N^{(1)} \underset{H_0}{\overset{H_1}{\gtrless}} \gamma_N^{(1)} \tag{5.28}$$

where $\gamma_N^{(1)}$ is a certain threshold which is such that the probability of false alarm does not exceed a given level α. Clearly, the above test is equivalent to the GLRT in the sense that both tests have identical power under the PFA constraint (5.23). Before studying the performance of the above test, we must now complete the definition of this test by providing a practical way to set the threshold $\gamma_N^{(1)}$ in (5.28).

Setting the Threshold $\gamma_N^{(1)}$. In order to maximize the power of our test while keeping the PFA constraint (5.23) satisfied, we must select the threshold $\gamma_N^{(1)}$ such that $\mathbb{P}_{H_0}\left[T_N^{(1)} \leq \gamma_N^{(1)}\right] = \alpha$. This requires the tedious computation of the distribution function of $T_N^{(1)}$ under H_0 for each N, K. Such a computation is usually impractical in cognitive radio applications, due to complexity/delay constraints, along with the fact that the number of sensors nodes K and the number of observations N are frequently varying. In order to simplify the selection of $\gamma_N^{(1)}$, we recently investigated in [12] the asymptotic case where both the number K of sensors and the number of observations are assumed to be large. In this case, simple expressions of the threshold $\gamma_N^{(1)}$ can be derived. More precisely, we studied the behaviour of $T_N^{(1)}$ under H_0 in the asymptotic regime

$$N \to \infty,\ K \to \infty,\ \ K/N \to c, \tag{5.29}$$

where $0 < c < 1$ is a constant. This asymptotic regime is relevant under cognitive radio constraints, as the secondary system must be able to decide the presence/absence of primary transmitters in a moderate amount of time: the number K of sensors and the number N of samples have therefore the same order of magnitude. In the asymptotic regime (5.29), it was proved in [12] that

$$N^{2/3}\left(\frac{T_N^{(1)} - (1+\sqrt{c})^2}{(1+\sqrt{c})\left(\frac{1}{\sqrt{c}}+1\right)^{1/3}}\right) \xrightarrow[H_0]{\mathcal{D}} X$$

where $\xrightarrow[H_0]{\mathcal{D}}$ stands for the convergence in distribution under H_0 and where X is a random variable which follows the Tracy-Widom distribution function $F_{TW}(.)$ associated with the Gaussian unitary Ensemble (see [12] for details). As a consequence, we obtain the following result.

Proposition 5.13. *The power of test (5.28) is maximum under constraint (5.23) only if the threshold $\gamma_N^{(1)}$ writes*

$$\gamma_N^{(1)} = (1+\sqrt{c})^2 + \frac{\beta_N^{(1)}}{N^{2/3}} \tag{5.30}$$

for some $\beta_N^{(1)}$ which tends to $(1+\sqrt{c})\left(\frac{1}{\sqrt{c}}+1\right)^{1/3} F_{TW}^{-1}(\alpha)$.

The above Proposition was used in [12] to derive practical guidelines to select the threshold $\gamma_N^{(1)}$ without resorting to a tedious computation of the exact distribution function $x \mapsto \mathbb{P}_{H_0}\left[T_N^{(1)} \leq x\right]$. This result will also be useful in Section 5.3.3 in order to analyze the performance of the proposed hypothesis test. Before providing such a performance analysis, we mention the existence of an other hypothesis testing approach which has been recently developed in [125, 19, 95] for cognitive radio contexts.

An Other Existing Hypothesis Testing Approach. A different approach introduced in several papers devoted to cognitive radio contexts [125, 19, 95] consists in rejecting the H_0-hypothesis when the following statistic

$$T_N^{(2)} = \frac{\lambda_1}{\lambda_K} \tag{5.31}$$

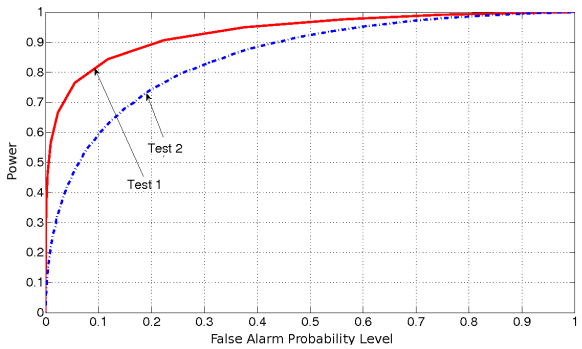

Figure 5.5: ROC curves of tests $T_N^{(1)}$ and $T_N^{(2)}$ – K=10, N=50, ρ=1.

lies above a well chosen threshold $\gamma_N^{(2)}$. The introduction of the above statistic $T_N^{(2)}$ can be motivated by the following observation. Assume that the following *limiting signal-to-noise ratio* (SNR) is well defined :

$$\rho = \lim_{K\to\infty} \frac{\|\boldsymbol{h}\|^2}{\sigma^2} \tag{5.32}$$

where $\|\boldsymbol{h}\|$ denotes the L2-norm of the $K1$ vector $\boldsymbol{h}$. Then, in the asymptotic regime (5.29),

$$T_N^{(2)} \xrightarrow[H_0]{a.s.} \frac{(1-\sqrt{c})^2}{(1+\sqrt{c})^2}$$

$$T_N^{(2)} \xrightarrow[H_1]{a.s.} \begin{cases} \frac{(1+\rho)(1+c/\rho)}{(1+\sqrt{c})^2} & \text{if } \rho > \sqrt{c} \\ \frac{(1-\sqrt{c})^2}{(1+\sqrt{c})^2} & \text{if } \rho < \sqrt{c}\,. \end{cases} \tag{5.33}$$

Provided that the SNR is large enough ($\rho > \sqrt{c}$), $T_N^{(2)}$ converges to different values depending on the true hypothesis. This motivates the fact that the value of $T_N^{(2)}$ can be used to decide which hypothesis is true.

Simulation Analysis. In the following, we compare by simulations both tests in the case where $K = 10$, $N = 50$ and $\rho = 1$. For a fixed level α, the thresholds $\gamma_N^{1,2}$ corresponding to the probability of error under H_0 are given by $P_{H_0}(T_N^{(1,2)} \geq \gamma_N^{(1,2)}) = \alpha$. The power of the test is then given by $P_{H_1}(T_N^{(1,2)} \geq \gamma_N^{(1,2)})$. Figure (5.5) provides the ROC curve for both tests. tests. It clearly shows that the test T_1 outperforms the test T_2.

In the rest of the paper, we provide a theoretical performance study of the tests based on $T_N^{(1)}$ and $T_N^{(2)}$ respectively to sustain the experimental claims of Figure 5.5. Using large deviations arguments, we rigorously prove that the test (5.28) based on $T_N^{(1)}$ outperforms the test based on $T_N^{(2)}$.

5.3.3 Error Exponents

Definition. The most natural approach to characterize the performance of the tests associated with statistics $T_N^{(1)}$ and $T_N^{(2)}$ is to evalute the power of each of these tests, or equivalently the miss probability $\mathbb{P}_{H_1}\left(T_N^{(i)} < \gamma_N^{(i)}\right)$, $i = 1, 2$. As the miss probability has no simple expression in the general case, we propose to study the asymptotic behaviour of the miss probability in the asymptotic regime (5.29) of interest. More precisely, for each test $i = 1, 2$, we prove the existence and provide the expression of the following error exponents

$$\mathcal{E}_{i,\rho} = \lim_{N\to\infty} -\frac{1}{N} \log \inf \left\{ \mathbb{P}_{H_1}\left(T_N^{(i)} < \gamma\right) \right\}, \tag{5.34}$$

where the infemum is taken w.r.t. all γ such that PFA constraint (5.23) holds for a fixed level α.

Main Result. The main conclusion is two-fold: First we give analytic closed form formulas for both limits, based on a precise Large Deviation study inspired from [76]; second, we show that the two limits are equal. Otherwise stated, the two tests have the same error exponents. Such a conclusion needs to be taken with caution as the presented results are asymptotic in nature, while the tests will be used in a finite-dimension context. A further study will be carried in the next section to discriminate between the two tests.

In order to express the error exponents of interest, we need further notations. Recall that the limiting probability distribution of the empirical distribution of the eigenvalues $F_N(x) = \frac{\#\{i,\, \lambda_i \leq x\}}{K}$ of $\hat{\mathbf{R}} = \frac{1}{N}\mathbf{Y}\mathbf{Y}^H$ is (under both assumptions H_0 or H_1) Marčhenko-Pastur distribution:

$$\mathbb{P}_{\check{\mathrm{M}}\mathrm{P}}(dy) = \mathbf{1}_{(\lambda^-,\lambda^+)}(y) \frac{\sqrt{(\lambda^+ - y)(y - \lambda^-)}}{2\pi c y}\, dy,$$

where $\lambda^+ = (1 + \sqrt{c})^2$ and $\lambda^- = (1 - \sqrt{c})^2$. We also introduce $\lambda_{\mathrm{spk}}^{\infty} = (1 + \rho)\left(1 + \frac{c}{\rho}\right)$ (recall that the largest eigenvalue λ_1 converges toward λ^+ under H_0 and toward $\lambda_{\mathrm{spk}}^{\infty}$ under H_1 - see for instance [8]). Of prime importance is the Stieltjes transform of $\mathbb{P}_{\check{\mathrm{M}}\mathrm{P}}$:

$$\mathbf{f}(x) = \int \frac{\mathbb{P}_{\check{\mathrm{M}}\mathrm{P}}(dy)}{y - x}$$

which admits the following well-known closed-form representations: If $x > \lambda^+$, then

$$\mathbf{f}(x) = \frac{(1 - x - c) + \sqrt{(1 - x - c)^2 - 4cx}}{2cx},$$

else, that is if $x \in (0, \lambda^-)$, then

$$\mathbf{f}(x) = \frac{(1 - x - c) - \sqrt{(1 - x - c)^2 - 4cx}}{2cx}.$$

Denote by

$$\begin{cases} \mathbf{F}^+(x) = \int \log(x - y)\mathbb{P}_{\check{\mathrm{M}}\mathrm{P}}(dy) & \text{for } x > \lambda^+, \\ \mathbf{F}^-(x) = \int \log(y - x)\mathbb{P}_{\check{\mathrm{M}}\mathrm{P}}(dy) & \text{for } x \in (0, \lambda^-) \end{cases}.$$

and let $\tilde{\mathbf{f}}(x) = -\frac{1}{x(1+c\mathbf{f}(x))}$.

Lemma 5.14. *The following representations hold true:*

$$\begin{aligned}\mathbf{F}^+(x) &= \log(x) + \frac{1}{c}\log(1+c\mathbf{f}(x)) + \log(1+\tilde{\mathbf{f}}(x)) \\ &\quad + x\mathbf{f}(x)\tilde{\mathbf{f}}(x) \\ \mathbf{F}^-(x) &= \log(x) + \frac{1}{c}\log(1+c\mathbf{f}(x)) + \log(-(1+\tilde{\mathbf{f}}(x))) \\ &\quad + x\mathbf{f}(x)\tilde{\mathbf{f}}(x)\ .\end{aligned}$$

For similar computations, see for example [53, Section 4]. We are now in position to introduce the functions that will help to express the error exponents. Denote by $\Delta(\cdot \mid A)$ the convex indicator function defined by:

$$\Delta(x \mid A) = \begin{cases} 0 & \text{if } x \in A, \\ \infty & \text{else .} \end{cases}$$

Define for each $\rho \geq 0$:

$$\begin{aligned}I_\rho^+(x) &= \frac{x - \lambda_{\mathrm{spk}}^\infty}{(1+\rho)} - (1-c)\log\left(\frac{x}{\lambda_{\mathrm{spk}}^\infty}\right) \\ &\quad - c\left(\mathbf{F}^+(x) - \mathbf{F}^+(\lambda_{\mathrm{spk}}^\infty)\right) + \Delta(x \mid [\lambda^+, \infty))\ , \\ I^-(y) &= y - \lambda_{\mathrm{spk}}^\infty - (1-c)\log\left(\frac{y}{\lambda_{\mathrm{spk}}^\infty}\right) \\ &\quad - c\left(\mathbf{F}^-(y) - \mathbf{F}^-(\lambda^-)\right) + \Delta(y \mid (0, \lambda^-])\ .\end{aligned}$$

As one may expect, I_ρ^+ and I^- are associated to the Large Deviation Principle (LDP) governing λ_1 and λ_K respectively. Thus, $\Gamma_\rho(t) = \inf\left\{I_\rho^+(x) + I^-(y), \quad \frac{x}{y} = t\right\}$ is associated to the LDP governing λ_1/λ_K.

Theorem 5.15. *Assume that $\rho > \sqrt{c}$. Error exponents $\mathcal{E}_{1,\rho}$ and $\mathcal{E}_{2,\rho}$ are well defined and are given by the following formulas:*

$$\mathcal{E}_{1,\,\rho} = I_\rho^+(\lambda^+) \quad \textit{and} \quad \mathcal{E}_{2,\,\rho} = \Gamma_\rho\left(\lambda^+/\lambda^-\right)\ .$$

Moreover, $I_\rho^+(\lambda^+) = \Gamma_\rho\left(\lambda^+/\lambda^-\right)$ and thus

$$\mathcal{E}_\rho \stackrel{\triangle}{=} \mathcal{E}_{1,\,\rho} = \mathcal{E}_{2,\,\rho}\ .$$

It is easy to check the equality $I_\rho^+(\lambda^+) = \Gamma_\rho\left(\lambda^+/\lambda^-\right)$. In fact, take $x = \lambda^+$ and $y = \lambda^-$, then the constraint $\frac{x}{y} = \frac{\lambda^+}{\lambda^-}$ is satisfied and

$$I_\rho^+(\lambda^+) + I^-(\lambda^-) = I_\rho^+(\lambda^+)\ .$$

Now assume that the infemum of $I_\rho^+(x) + I^-(y)$ over the constraint $\frac{x}{y} = \frac{\lambda^+}{\lambda^-}$ is attained for some different (x, y). Either $x < \lambda^+$, but in this case $I_\rho^+(x) = \infty$ due to the convex indicator function, or $x > \lambda^+$ but in this case $y > \lambda^-$, which implies that $I^-(y) = \infty$. Necessarily, the infemum is achieved for $x = \lambda^+$ and $y = \lambda^-$. This yields $I_\rho^+(\lambda^+) = \Gamma_\rho\left(\lambda^+/\lambda^-\right)$.

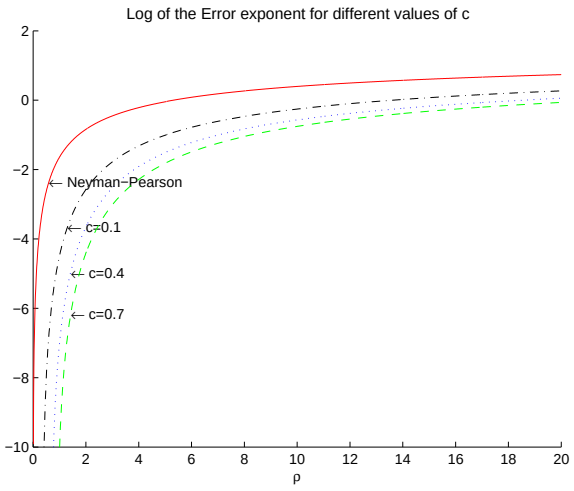

Figure 5.6: Logarithm of the error exponent $\mathcal{E}_\rho$ as a function of ρ (in dB) for different values of c.

Comments and Numerical Results. In terms of error exponents, both tests $T_N^{(1)}$ and $T_N^{(2)}$ admit the same error exponent. Figure 5.6 represents the error exponent in log-scale as a function of the SNR ρ in dB. Error exponents are compared with the error exponent associated with the Neyman-Pearson test, achieved in the ideal case where parameters $\mathbf{H}$ and σ are known. The error exponent of the Neyman-Pearson test can be derived from Stein's Lemma, and provides an upper bound on the achievable error exponents. Note that when $\rho < \sqrt{c}$, the test statistic $T_N^{(1)}$ converges to the same limit under H_0 and under H_1. A similar behaviour occurs for $T_N^{(2)}$ due to equation (5.33). Thus, both tests fail when $\rho < \sqrt{c}$. Therefore, it is not surprising that the error exponent tends to zero when ρ is close to $\sqrt{c}$.

5.3.4 Bahadur Relative Efficiency

The previous study provides the error exponents associated with the power of both tests. As both error exponents coincide, one should consider other efficiency metrics in order to discriminate both tests. A widespread approach consists in using the *Bahadur relative efficiency* [118] which coincides with the ratio $\mathbb{B}_1/\mathbb{B}_2$ of the *Bahadur slopes* $\mathbb{B}_1$ and $\mathbb{B}_2$ associated with tests $T_N^{(1)}$ and $T_N^{(2)}$ respectively. For each test $i = 1, 2$, the Bahadur slope[3] $\mathbb{B}_i$ can be interpreted as the error exponent associated with the significance level (p-value) under H_1:

$$\mathbb{B}_i = \lim_{N\to\infty} -\frac{1}{N}\log P_{H_0}\left(T_N^{(i)} \leq T_{\infty,H_1}^{(i)}\right) \tag{5.35}$$

where $T_{\infty,H_1}^{(i)}$ represents the (deterministic) limit in probability under H_1 of $T_N^{(i)}$. When $\rho > \sqrt{c}$, limits $T_{\infty,H_1}^{(1)}$ and $T_{\infty,H_1}^{(2)}$ are given by $T_{\infty,H_1}^{(1)} = \lambda_{\text{spk}}^{\infty}$ and $T_{\infty,H_1}^{(2)} = \frac{\lambda_{\text{spk}}^{\infty}}{\lambda_-}$. Whereas the evaluation of error exponents in Section 5.3.3 relied on a LDP under H_1, it is clear from (5.35) that the evaluation of the slopes relies on a LDP under H_0.

[3]Strictly speaking, Bahadur slopes are defined by $2\mathbb{B}_1$ and $2\mathbb{B}_2$. However we still refer to $\mathbb{B}_1$ and $\mathbb{B}_2$ as the slopes with slight language abuse.

Theorem 5.16. *Assume $\rho > \sqrt{c}$. Quantities $\mathbb{B}_1$ and $\mathbb{B}_2$ are given by:*

$$\mathbb{B}_1 = I_0^+(\lambda_{\text{spk}}^\infty) \tag{5.36}$$

$$\mathbb{B}_2 = \Gamma_0(\frac{\lambda_{\text{spk}}^\infty}{\lambda_-}) \tag{5.37}$$

Moreover, $\mathbb{B}_1 < \mathbb{B}_2$ for each $c \in (0,1)$ and each $\rho > \sqrt{c}$.

The proof of (5.36) and (5.37) is omitted due to the lack of space. One can show inequality $\mathbb{B}_1 < \mathbb{B}_2$ by noting that:

$$\begin{aligned} \mathbb{B}_2 &= \inf\left(I_0^+(x) + I_-(y) \,:\, \frac{x}{y} = \frac{\lambda_{\text{spk}}^\infty}{\lambda_-} \right) \\ &< I_0^+(\lambda_{\text{spk}}^\infty) + I^-(\lambda_-) = I_0^+(\lambda_{\text{spk}}^\infty)\,. \end{aligned}$$

In particular, Theorem 5.16 proves that test $T_N^{(1)}$ is more efficient than test $T_N^{(2)}$ in Bahadur's sense. Furthermore, the Bahadur efficiency provides useful information on the number of samples required by each of the two tests to achieve a target level α and a target power β (see [118]). These aspects will be developed in an extended version of this paper.

Conclusion

In this section, we have studied the performance of two eigen-based collaborative sensing test and showed the superiority of the one introduced in [12] using asymptotic random matrix tools.

Bibliography

[1] S. Verdú A.M. Tulino, A. Lozano. Capacity-achieving input covariance for correlated multi-antenna channels. In *41th Annual Allerton Conf. on Comm., control and computing*, Oct. 2003.

[2] L. Cottatellucci and M. Debbah. The effect of line of sight components on the asymptotic capacity of mimo systems. In *Proc. ISIT*, Chicago, July 2004.

[3] Greg W. Anderson and Ofer Zeitouni. A law of large numbers for finite-range dependent random matrices. *Comm. Pure Appl. Math.*, 61(8):1118–1154, 2008.

[4] G.W. Anderson and O. Zeitouni. A CLT for a band matrix model. *Probab. Theory Related Fields*, 134(2):283–338, 2006.

[5] Z. D. Bai. Convergence rate of expected spectral distributions of large random matrices. II. Sample covariance matrices. *Ann. Probab.*, 21(2):649–672, 1993.

[6] Z. D. Bai and J. W. Silverstein. No eigenvalues outside the support of the limiting spectral distribution of large-dimensional sample covariance matrices. *Ann. Probab.*, 26(1):316–345, 1998.

[7] Z. D. Bai and J. W. Silverstein. CLT for linear spectral statistics of large-dimensional sample covariance matrices. *Ann. Probab.*, 32(1A):553–605, 2004.

[8] J. Baik and J. W. Silverstein. Eigenvalues of large sample covariance matrices of spiked population models. *J. Multivariate Anal.*, 97(6):1382–1408, 2006.

[9] O. Besson, L. L. Scharf, and S. Kraut. Adaptive detection of a signal known only to lie on a line in a known subspace, when primary and secondary data are partially homogeneous. *IEEE Trans. on Sign. Process.*, 54(12):4698–4705, December 2006.

[10] R.J. Bhansali, L. Giraitis, and P.S. Kokoszka. Convergence of quadratic forms with nonvanishing diagonal. *Stat. Probab. Letters*, 77:726–734, 2007.

[11] P. Bianchi, M. Debbah, M. Maïda, and J. Najim. Power analysis of some hypothesis tests for collaborative sensing. In *IEEE Workshop on Stat. Signal Processing*, Cardiff, UK, August 2009.

[12] P. Bianchi, J. Najim, G. Alfano, and M. Debbah. Asymptotics of eigenbased collaborative sensing. In *submitted to Information Theory Workshop ITW*, 2009.

[13] E. Biglieri, G. Caire, and G. Taricco. CDMA system design through asymptotic analysis. *IEEE Trans. on Communications*, 48(11):1882–1896, November 2000.

[14] P. Billingsley. *Probability and measure*. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons Inc., New York, third edition, 1995.

[15] A. Boutet de Monvel and A. Khorunzhy. Limit theorems for random matrices. *Markov Process. Related Fields*, 4(2):175–197, 1998.

[16] A. Boutet de Monvel, A. Khorunzhy, and V. Vasilchuk. Limiting eigenvalue distribution of random matrices with correlated entries. *Markov Process. Related Fields*, 2(4):607–636, 1996.

[17] R. W. Butler and A. T. A. Wood. Saddlepoint approximation for moment generating functions of truncated random variables. *Ann. Statist.*, 32(6):2712–2730, 2004.

[18] T. Cabanal-Duvillard. Fluctuations de la loi empirique de grandes matrices aléatoires. *Ann. Inst. H. Poincaré Probab. Statist.*, 37(3):373–402, 2001.

[19] L.S. Cardoso, M. Debbah, P. Bianchi, and J. Najim. Cooperative spectrum sensing using random matrix theory. In *Wireless Pervasive Computing, 2008. ISWPC 2008. 3rd International Symposium on*, pages 334–338, May 2008.

[20] S. Chatterjee. Fluctuations of eigenvalues and second order Poincaré inequalities. *Probab. Theory Related Fields*, 143(1-2):1–40, 2009.

[21] S. Chatterjee and A. Bose. A new method for bounding rates of convergence of empirical spectral distributions. *J. Theoret. Probab.*, 17(4):1003–1019, 2004.

[22] J.-M. Chaufray, W. Hachem, and Ph. Loubaton. Asymptotic Analysis of Optimum and Sub-Optimum CDMA Downlink MMSE Receivers. *IEEE Trans. Inform. Theory*, 50(11):2620–2638, November 2004.

[23] L. H. Y. Chen. An inequality for the multivariate normal distribution. *J. Multivariate Anal.*, 12(2):306–315, 1982.

[24] C. Chuah, J.M. Kahn, and D. Tse. Capacity of multi-antenna array systems in indoor wireless environment. In *GLOBECOM*, volume 4, pages 1894–1899, Sydney, Nov. 1998. IEEE.

[25] 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. Inform. Theory*, 48(3):637–650, March 2002.

[26] 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. Inform. Theory*, 48(3):637–650, 2002.

[27] 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. Communications*, 48(3):502–513, March 2000.

[28] M. Debbah and R. Müller. On the limiting behavior of directional mimo models. In *Proceeding of the Sixth Baiona Workshop On Signal Processing in Communication*, Spain, September 2003.

[29] R. Douc, A. Guillin, and J. Najim. Moderate deviations for particle filtering. *Ann. Appl. Probab.*, 15(1B):587–614, 2005.

[30] R. B. Dozier and J. W. Silverstein. On the empirical distribution of eigenvalues of large dimensional information-plus-noise-type matrices. *J. Multivariate Anal.*, 98(4):678–694, 2007.

[31] J. Dumont, W. Hachem, S. Lasaulce, P. Loubaton, and J. Najim. On the capacity achieving covariance matrix for rician mimo channels: an asymptotic approach. *submitted to IEEE Inf. Th.*, 2007. `http://front.math.ucdavis.edu/0710.4051`.

[32] J. Dumont, P. Loubaton, S. Lasaulce, and M. Debbah. On the asymptotic performance of mimo correlated ricean channels. In *ICASSP Proceedings*, 2005.

[33] A. Edelman. Eigenvalues and condition numbers of random matrices. *SIAM J. Matrix Anal. Appl.*, 9(4):543–560, 1988.

[34] Alan Edelman. The distribution and moments of the smallest eigenvalue of a random matrix of Wishart type. *Linear Algebra Appl.*, 159:55–80, 1991.

[35] J. Evans and D.N.C Tse. Large system performance of linear multiuser receivers in multipath fading channels. *IEEE Trans. Inform. Theory*, 46(6):2059–2078, September 2000.

[36] G. Foschini. Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas. *Bell Labs Technical Journal*, pages 41–59, 1996.

[37] V. L. Girko. *Sluchainye matritsy (Russian) [Random Matrices]*. Izdat. Ob″ed. "Višča Škola" pri Kiev. Gosudarstv. Univ., Kiev, 1975.

[38] V. L. Girko. *Theory of random determinants*, volume 45 of *Mathematics and its Applications (Soviet Series)*. Kluwer Academic Publishers Group, Dordrecht, 1990.

[39] V. L. Girko. *Theory of stochastic canonical equations. Vol. I*, volume 535 of *Mathematics and its Applications*. Kluwer Academic Publishers, Dordrecht, 2001.

[40] V. L. Girko. Thirty years of the central resolvent law and three laws on the $1/n$ expansion for resolvent of random matrices. *Random Oper. Stochastic Equations*, 11(2):167–212, 2003.

[41] Vyacheslav L. Girko. *Theory of stochastic canonical equations. Vol. II*, volume 535 of *Mathematics and its Applications*. Kluwer Academic Publishers, Dordrecht, 2001.

[42] G.L. Stüber. *Principles of Mobile Communications*. Norwell, MA:Kluwer, 1996.

[43] A. Goldsmith, S.A. Jafar, N. Jindal, and S. Vishwanath. Capacity Limits of MIMO Channels. *IEEE J. Sel. Areas in Comm.*, 21(5), june 2003.

[44] F. Götze and A. Tikhomirov. Asymptotic distributions of quadratic forms and applications. *J. Theoret. Probab.*, 15:424–475, 2002.

[45] U. Grenander and J. W. Silverstein. Spectral analysis of networks with random topologies. *SIAM J. Appl. Math.*, 32(2):499–519, 1977.

[46] A. Guionnet. Large deviations upper bounds and central limit theorems for non-commutative functionals of Gaussian large random matrices. *Ann. Inst. H. Poincaré Probab. Statist.*, 38(3):341–384, 2002.

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

[48] D. Guo, S. Verdu, and L.K. Rasmussen. Asymptotic Normality of Linear Multiuser Receiver Outputs. *IEEE Trans. Info. Theory*, 48(12):3080–3095, December 2002.

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

[50] W. Hachem, O. Khorunzhiy, P. Loubaton, J. Najim, and L. Pastur. A new approach for mutual information analysis of large dimensional multi-antenna chennels. *IEEE Trans. Inform. Theory*, 54(9):3987–4004, 2008.

[51] W. Hachem, P. Loubaton, and J. Najim. The empirical eigenvalue distribution of a gram matrix: From independence to stationarity. *Markov Process, Related Fields.*, 11(4):629–648, 2005.

[52] W. Hachem, P. Loubaton, and J. Najim. The empirical distribution of the eigenvalues of a Gram matrix with a given variance profile. *Ann. Inst. H. Poincaré Probab. Statist.*, 42(6):649–670, 2006.

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

[54] W. Hachem, P. Loubaton, and J. Najim. A CLT for information-theoretic statistics of gram random matrices with a given variance profile. *Ann. Appl. Probab.*, 18(6):2071–2130, 2008.

[55] F. Hiai and D. Petz. *The semicircle law, free random variables and entropy*, volume 77 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 2000.

[56] D. Hoesli, Y.-H. Kim, and A. Lapidoth. Monotonicity results for coherent MIMO Rician channels. *IEEE Trans. Inform. Theory*, 51(12):4334–4339, 2005.

[57] R. A. Horn and C. R. Johnson. *Topics in matrix analysis*. Cambridge University Press, Cambridge, 1994.

[58] C. Houdré, V. Pérez-Abreu, and D. Surgailis. Interpolation, correlation identities, and inequalities for infinitely divisible variables. *J. Fourier Anal. Appl.*, 4(6):651–668, 1998.

[59] J.Dumont, Ph. Loubaton, S. Lasaulce, and M.Debbah. On the asymptotic performance of mimo correlated rician channels. In *Proc. ICASSP*, pages 813–816, 2005.

[60] S. Jin, M.R. McKay, X. Gao, and I.B. Collings. Asymptotic SER and Outage Probability of MIMO MRC in Correlated Fading. *IEEE Signal Processing Letters*, 14(1), 2007.

[61] K. Johansson. On fluctuations of eigenvalues of random Hermitian matrices. *Duke Math. J.*, 91(1):151–204, 1998.

[62] D. Jonsson. Some limit theorems for the eigenvalues of a sample covariance matrix. *J. Multivariate Anal.*, 12(1):1–38, 1982.

[63] A. Kammoun, M. Kharouf, W. Hachem, and J. Najim. A central limit theorem for the snr at the wiener filter output for large dimensional signals. *IEEE Workshop on Signal Processing Advances in Wireless Communications*, 2008.

[64] A. Kammoun, M. Kharouf, W. Hachem, and J. Najim. A Central Limit Theorem for the SINR at the LMMSE estimator output for large dimensional signals. *accepted for publication in IEEE Inf. Th.*, june 2008.

[65] A. Kammoun, M. Kharouf, W. Hachem, and J. Najim. BER and Outage probability approximations for LMMSE detectors on correlated MIMO channels. *accepted for publication in IEEE Inf. Th.*, sept. 2008.

[66] M. Kang and M.S. Alouini. Capacity of mimo rician channels. *IEEE Trans. on Wireless Communications*, 5(1):112–122, january 2006.

[67] T. Kato. *Perturbation theory for linear operators*. Classics in Mathematics. Springer-Verlag, Berlin, 1995. Reprint of the 1980 edition.

[68] J-P. Kermoal, L. Schumacher, K-I. Pedersen, P-E. Mogensen, and F. Frederiksen. A stochastic mimo radio channel model with experimental validation. *IEEE Journal on Selected Areas in Communications*, 20(6):1211–1226, Aug 2002.

[69] A. Khorunzhy, B. Khoruzhenko, and L. Pastur. Asymptotic properties of large random matrices with independent entries. *J. Math. Phys.*, 37(10):5033–5060, 1996.

[70] 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.*, 153(3):605–646, 1993.

[71] Y.C. Ko, M.S. Alouini, and M.K. Simon. Outage Probability of Diversity Systems over Generalized Fading Channels. *IEEE Transactions on Communications*, 48(11):1783–1787, November 2000.

[72] L. Li, A.M. Tulino, and S. Verdu. Design of reduced-rank mmse multiuser detectors using random matrix methods. *IEEE Trans. Inform. Theory*, 50(6):986–1008, june 2004.

[73] P. Li, D. Paul, R. Narasimhan, and J. Cioffi. On the Distribution of SINR for the MMSE MIMO Receiver and Performance Analysis. *IEEE Inf. Theory*, 52(1):271–286, January 2006.

[74] K. Liu, V. Raghavan, and A.M. Sayeed. Capacity scaling and spectral efficiency in wide-band correlated MIMO channels. *IEEE Trans. Inform. Theory*, 49(10):2504–2526, 2003.

[75] David G. Luenberger. *Optimization by vector space methods*. John Wiley & Sons Inc., New York, 1969.

[76] M. Maïda. Large deviations for the largest eigenvalue of rank one deformations of Gaussian ensembles. *Electron. J. Probab.*, 12:1131–1150 (electronic), 2007.

[77] M. Maïda, J. Najim, and S. Péché. Large deviations for weighted empirical mean with outliers. *Stochastic Process. Appl.*, 117(10):1373–1403, 2007.

[78] V. A. Marčenko and L. A. Pastur. Distribution of eigenvalues in certain sets of random matrices. *Mat. Sb. (N.S.)*, 72 (114):507–536, 1967.

[79] X. Mestre, J.R. Fonollosa, and A. Pagès-Zamora. Capacity of MIMO channels: Asymptotic Evaluation Under Correlated Fading. 21(5):829–838, June 2003.

[80] J. A. Mingo and R. Speicher. Second order freeness and fluctuations of random matrices. I. Gaussian and Wishart matrices and cyclic Fock spaces. *J. Funct. Anal.*, 235(1):226–270, 2006.

[81] J. Mitola. *Cognitive Radio An Integrated Agent Architecture for Software Defined Radio*. PhD thesis, Royal Institute of Technology (KTH), May 2000.

[82] 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. Inform. Theory*, 49(10):2545–2561, 2003.

[83] A.L. Moustakas and S.H. Simon. On the outage capacity of correlated multiple-path MIMO channels. *IEEE Trans. Inform. Theory*, 53(11):3887–3903, November 2007.

[84] Aris L. Moustakas and Steven H. Simon. Random matrix theory of multi-antenna communications: the Ricean channel. *J. Phys. A*, 38(49):10859–10872, 2005.

[85] R.R. Müller. A Random Matrix Model of Communication Via Antenna Arrays. *IEEE Trans. on Information Theory*, 48(9):2495–2506, September 2002.

[86] J. Najim. Large deviations for independent random variables—application to Erdős-Renyi's functional law of large numbers. *ESAIM Probab. Stat.*, 9:116–142 (electronic), 2005.

[87] G. Pan, M. Guo, and W. Zhou. Asymptotic distributions of the signal-to-interference ratios of LMMSE detection in multiuser communications. *Ann. Appl. Probab.*, 17(1):181–206, 2007.

[88] L. Pastur and A. Lytova. Central limit theorem for linear eigenvalue statistics of random matrices with independent entries. Technical report, Institute for low temperatures, Kharkiv, Ukraine, 2008. Available on ArXiv `http://front.math.ucdavis.edu/0809.4698`.

[89] L. Pastur and V. Vasilchuk. On the law of addition of random matrices. *Comm. Math. Phys.*, 214(2):249–286, 2000.

[90] L. Pastur and V. Vasilchuk. On the moments of traces of matrices of classical groups. *Comm. Math. Phys.*, 252(1-3):149–166, 2004.

[91] L. Pastur and V. Vasilchuk. On the law of addition of random matrices: covariance and the central limit theorem for traces of resolvent. In *Probability and mathematical physics*, volume 42 of *CRM Proc. Lecture Notes*, pages 399–416. Amer. Math. Soc., Providence, RI, 2007.

[92] 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*, 2:27–30, 1995.

[93] L.A. Pastur. A Simple Approach to the Global Regime of Gaussian Ensembles of Random Matrices. *Ukrainian Mathematical Journal*, 57(6):936–966, June 2005.

[94] M.J.M. Peacock, I.B. Collings, and M.L. Honig. Asymptotic spectral efficiency of multiuser multisignature CDMA in frequency-selective channels. *IEEE Trans. Inform. Theory*, 52(3):1113–1129, March 2006.

[95] F. Penna, R. Garello, and M. A. Spirito. Cooperative Spectrum Sensing based on the Limiting Eigenvalue Ratio Distribution in Wishart Matrices. *IEEE Communication Letters*, submitted, 2009.

[96] W. Phoel and M.L. Honig. Performance of coded DS-CDMA with pilot-assisted channel estimation and linear interference suppression. *IEEE Trans. Communications*, 50(5):822–832, May 2002.

[97] H. V. Poor and S. Verdu. Probability of error in MMSE multiuser detection. *IEEE Trans. Info. Theory*, 43(3):858–871, May 1997.

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

[99] A. M. Sengupta and P.P. Mitra. Capacity of multivariate channels with multiplicative noise: I. random matrix techniques and large-n expansions for full transfer matrices. available at http://arxiv.org/abs/physics/0010081, 2000.

[100] D. Shlyakhtenko. Random Gaussian band matrices and freeness with amalgamation. *Internat. Math. Res. Notices*, 20:1013–1025, 1996.

[101] J. W. Silverstein. Weak convergence of random functions defined by the eigenvectors of sample covariance matrices. *Ann. Probab.*, 18(3):1174–1194, 1990.

[102] J. W. Silverstein. Strong convergence of the empirical distribution of eigenvalues of large-dimensional random matrices. *J. Multivariate Anal.*, 55(2):331–339, 1995.

[103] J. W. Silverstein and Z. D. Bai. On the empirical distribution of eigenvalues of a class of large-dimensional random matrices. *J. Multivariate Anal.*, 54(2):175–192, 1995.

[104] M. K. Simon and M. S. alouini. *Digital Communication Over Fading Channels*. Wiley Series in Telecommunications and Signal Processing, second edition, 2005.

[105] Ya. Sinai and A. Soshnikov. Central limit theorem for traces of large random symmetric matrices with independent matrix elements. *Bol. Soc. Brasil. Mat. (N.S.)*, 29(1):1–24, 1998.

[106] F. Smithies. *Integral equations*. Cambridge Tracts in Mathematics and Mathematical Physics, no. 49. Cambridge University Press, New York, 1958.

[107] A. Soshnikov. The central limit theorem for local linear statistics in classical compact groups and related combinatorial identities. *Ann. Probab.*, 28(3):1353–1370, 2000.

[108] J.W. Stokes and J.A. Ritcey. A General Method for Evaluating Outage Probabilities Using Padé Approximations. *Globecom*, 3(8-12):1485–1490, 1998.

[109] G. Taricco. On the capacity of separately-correlated MIMO Rician fading channels. In *Proc. 49th Annual IEEE Globecom Conference, 2006*, San Francisco (CA), USA, November 2006.

[110] G. Taricco. Asymptotic Mutual Information Statistics of Seperately-Correlated Rician Fading MIMO Channels. *IEEE Transactions on Information Theory*, 2008.

[111] I.E. Telatar. Capacity of multi-antenna gaussian channel. *European Transactions on Telecommunications*, 1999.

[112] D. Tse and O. Zeitouni. Linear multiuser receivers in random environments. *IEEE Trans. Inform. Theory*, 46(1):171–188, 2000.

[113] D. N. C. Tse and S. V. Hanly. Linear multiuser receivers: effective interference, effective bandwidth and user capacity. *IEEE Trans. Inform. Theory*, 45(2):641–657, 1999.

[114] A. Tulino and S. Verdú. Random matrix theory and wireless communications. In *Foundations and Trends in Communications and Information Theory*, volume 1, pages 1–182. Now Publishers, june 2004.

[115] A. Tulino and S. Verdu. Asymptotic outage capacity of multiantenna channels. In *Proceedings of IEEE ICASSP '05*, volume 5, pages 828–835, march 2005.

[116] A.M. Tulino, A. Lozano, and S. Verdú. Impact of Antenna Correlation on the Capacity of Multiantenna Channels. *IEEE Trans. Inform. Theory*, 51(7):2491–2509, July 2005.

[117] H. Urkowitz. Energy detection of unknown deterministic signals. *Proceedings of the IEEE*, 55:523–531, 1967.

[118] A. W. van der Vaart. *Asymptotic statistics*, volume 3 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge, 1998.

[119] S. Verdu and S. Shamai. Spectral efficiency of cdma with random spreading. *IEEE Trans. Inform. Theory*, 45(2):622–640, 1999.

[120] M. Vu and A. Paulraj. Capacity optimization for rician correlated MIMO wireless channels. In *in Proc. Asilomar Conference*, pages 133–138, Asilomar, November 2005.

[121] C-K. Wen, P. Ting, and J-T. Chen. Asymptotic analysis of mimo wireless systems with spatial correlation at the receiver. *ieeeCOM*, 54(2):349–363, Feb. 2006.

[122] E. P. Wigner. Characteristic vectors of bordered matrices with infinite dimensions. *Ann. of Math. (2)*, 62:548–564, 1955.

[123] E. P. Wigner. On the distribution of the roots of certain symmetric matrices. *Ann. of Math. (2)*, 67:325–327, 1958.

[124] Y. Q. Yin. Limiting spectral distribution for a class of random matrices. *J. Multivariate Anal.*, 20(1):50–68, 1986.

[125] Y. H. Zeng and Y. C. Liang. Eigenvalue based spectrum sensing algorithms for cognitive radio. *to appear in IEEE Trans on Communications volume=arXiv:0804.2960v1.*

[126] L. Zheng and D.N. Tse. Optimal diversity-multiplexing tradeoff in multiple antenna channels. In *Proc. Allerton Conf. Comm., Control, Computing*, pages 835–844, Monticello, IL, Oct. 2001.