The Mutual Information of a MIMO Channel: A Survey

Walid Hachem, Philippe Loubaton and Jamal Najim

Abstract— **In this paper, we survey recent mathematical results devoted to the study of the mutual information of MIMO channels in the case where transmit and receive antennas converge to ∞ at the same rate.**

We express the different results in a unified framework and the emphasis is put on non-asymptotic deterministic approximations of the mutual information, asymptotic limits (when existing) and Rician correlated channels.

I. Introduction

It is well-known that the mutual information of a MIMO channel is given by

$$C(\varsigma^2) = \mathbb{E} \log\det \left(I + \frac{H_n H_n^*}{\varsigma^2} \right)$$

where ς^2 is the variance of an additive corrupting noise and the $N \times n$ matrix $\sqrt{n} H_n = (H_{ij}^n)$ represents the complex gain between transmit and receive antennas. In his seminal paper [10], Telatar has proved that in the case where the entries of the matrix are i.i.d. centered Gaussian random variables with variance $\frac{\sigma^2}{n}$, the mutual information properly normalized, i.e. $C_n(\varsigma^2) = \frac{C(\varsigma^2)}{N}$ converges toward a deterministic quantity involving Marčenko-Pastur probability distribution in the case where $\frac{N}{n} \to c > 0$. Telatar relied on Marčenko-Pastur's theorem from the theory of Large Random Matrices. Of importance is the fact that the mutual information of the channel grows proportionally to the number of emitting antennas (or receiving ones since their ratio is assumed to be constant).

The question soon arised to extend these results to more realistic models, especially to those models where the entries of the matrix are no longer independent and have a covariance function of the form:

$$\mathrm{cov}(H_{ij}^n, H_{i'j'}^n) = \frac{a(i-i')b(j-j')}{n}$$

where f and g are two given functions. Such results, based on an extensive use of the Stieltjes transform $\mathbf{f}$ of a probability measure μ:

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

have been developed by Chuah et al. in [2], relying on results by Girko [4].

Relying on replica methods, Moustakas et al. [9] have been able to compute an equivalent of the mean and the variance of $\frac{1}{N} \log\det \left(I + \frac{H_n H_n^*}{\varsigma^2} \right)$ (the variance being of order $\frac{1}{N^2}$).

We shall survey all this line of results and present recent results [8] where an equivalent of the mutual information is computed in the case where the covariance of Y_n is of a general form

$$\mathrm{cov}(H_{ij}^n, H_{i'j'}^n) = \frac{\kappa(i-i', j-j')}{n}$$

and in the case where H_n is no longer centered, i.e. $H_n = Z_n + B_n$ where B_n is deterministic and $\mathbb{E}(Z_n) = 0$. Such a case is known as the Rician channel.

In the sequel, we deal with the following model of non-centered random matrices with a variane profile:

$$\Sigma_n = Y_n + A_n$$

where Σ_n, Y_n and A_n are $N \times n$ random matrices. Matrix Y_n has a variance profile, i.e. the entries of $Y_n = (Y_{ij}^n)$ have the form $Y_{ij}^n = \frac{\sigma_{ij}(n)}{\sqrt{n}} X_{ij}^n$, the X_{ij}^n being independent and identically distributed $(0,1)$ complex circular gaussian (denoted $\mathcal{CN}(0,1)$) random variables. Matrix A_n is assumed to be deterministic. Otherwise stated, $\Sigma_n = Y_n + A_n$ where $\mathbb{E} Y_n = 0$ and $\mathbb{E}\Sigma_n = A_n$.

In Section II, we survey mutual informations results in the case where Σ_n is centered, that is in the case where $A_n = 0$. Non-asymptotic formulas are given for a general variance profile $\sigma_{ij}^{(n)}$ and asymptotic formulas are provided in the case where the variance profile is the sampling of a continuous function, i.e. $\sigma_{ij}^{(n)} = \sigma(i/N, j/n)$.

In Section III, the general case is adressed. Non-asymptotic formulas for the mutual information are provided while no asymptotic formulas are given. In fact,

As will be shown in Section V (based on [6]), the case of a Gaussian matrix Z_n with correlated entries is very close to the case of a matrix Y_n with a variance profile. The intuitive equivalence $Y_n \approx F_N Z_n F_n^*$ where $F_p = (F_{j_1,j_2}^p)_{0 \le j_1, j_2 < p}$ is the $p \times p$ Fourier matrix:

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

is fully explained.

An important case, both practically and theoretically (computations are heavily simplified), is when the variance profile is separable, i.e.

$$\sigma_{ij}^n = \alpha_i^n \beta_i^n$$

This work was partially supported by the "Fonds National de la Science" via the ACI program "Nouvelles Interfaces des Mathématiques", project MALCOM number 205, and by the European network of excellence NEWCOM

W. Hachem is with Ecole Supérieure d'Electricité (Supélec), Plateau de Moulon, 3 rue Joliot-Curie 91192 Gif Sur Yvette Cedex, France. walid.hachem@supelec.fr

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

J. Najim is with CNRS and Télécom Paris 46, rue Barrault, 75013 Paris, France. najim@tsi.enst.fr

(we shall soon drop superscript n). Particular attention will be devoted to the case of Rician channel, i.e. $A_n \neq 0$.

II. The centered case

In this section, we take $A_n = 0$ that is $\Sigma_n = Y_n$. As a major consequence of this assumption, the normalized mutual information $\frac{1}{N}\mathbb{E}\log\det\left(I_N + \frac{\Sigma\Sigma^*}{\varsigma^2}\right)$ converges toward a deterministic limit in the case where the variance profile is the sampling of a continuous function (see Assumption (A-1) and Theorem 2.3).

We introduce the following notations:

$$\begin{array}{ll} D_j = \mathrm{diag}\left(\sigma_{ij}^2,\ 1 \le i \le N\right), & T = \mathrm{diag}(T_i,\ 1 \le i \le N), \\ \tilde{D}_i = \mathrm{diag}\left(\sigma_{ij}^2,\ 1 \le j \le n\right), & \tilde{T} = \mathrm{diag}(\tilde{T}_j,\ 1 \le j \le n) \end{array}$$

where both T and $\tilde{T}$ are defined by the following system of $N+n$ equations.

Theorem 2.1 (see [8]): Consider the following system of $N+n$ equations:

$$\begin{aligned} T_i(z) &= \frac{-1}{z(1+\frac{1}{n}\mathrm{Tr}\tilde{D}_i\tilde{T}(z))}, \quad 1 \le i \le N, \\ \tilde{T}_j(z) &= \frac{-1}{z(1+\frac{1}{n}\mathrm{Tr}D_jT(z))}, \quad 1 \le j \le n \end{aligned}$$

then this system admits a unique solution $(T, \tilde{T})$ among the class of diagonal matrices such that $T_i(z)$ and $\tilde{T}_j(z)$ are Stieltjes transforms of probability measures.

Theorem 2.2 (see [8]): Denote by $\bar{C}_n(\varsigma^2)$ the quantity

$$\bar{C}_n(\varsigma^2) = -\frac{1}{N}\sum_{i=1}^{N}\log\varsigma^2 T_i(-\varsigma^2) - \frac{1}{N}\sum_{j=1}^{n}\log\varsigma^2\tilde{T}_j(-\varsigma^2) - \frac{\varsigma^2}{Nn}\sum_{\substack{i=1:N \\ j=1:n}}\sigma_{ij}^2 T_i(-\varsigma^2)\tilde{T}_j(-\varsigma^2)$$

Then the following holds true:

$$\frac{1}{N}\mathbb{E}\log\det\left(I_N + \frac{\Sigma\Sigma^*}{\varsigma^2}\right) - \bar{C}_n(\varsigma^2) \xrightarrow[n\to\infty]{} 0.$$

Of interest is the case where the convergence of $\bar{C}_n(\varsigma^2)$ occurs. This is the aim of next assumption and next theorem.

Assumption A-1: The variance profile is the sampling of a continuous function:

$$\sigma_{ij}^{(n)} = \sigma\left(\frac{i}{N}, \frac{j}{n}\right) \quad (2)$$

where $\sigma(x, y)$ is continuous.

Theorem 2.3: Assume now that (A-1) holds and consider the following functional equation:

$$k(u,z) = \frac{1}{-z + \int_0^1 \frac{\sigma^2(u,t)}{1+c\int_0^1\sigma^2(x,t)k(x,z)dx}dt}.$$

This equation admits a unique solution in the class of functions k such that

1) $z \mapsto k(u,z)$ is the Stieltjes transform of a probability measure,
2) $[0,1] \ni u \mapsto k(u,z)$ is continuous.

We denote $k_\varsigma(u) = k(u, -\varsigma^2)$. The following convergence holds true:

$$\bar{C}_n(\varsigma^2) \xrightarrow[n\to\infty]{} C^*(\varsigma^2)$$

where $C^*(\varsigma^2)$ is given by the following formula

$$\begin{aligned} C^*(\varsigma^2) = &-\int_0^1 \log k_\varsigma(u)\,du \\ &- \frac{1}{c}\int_0^1 \log\left(\frac{1}{\varsigma^2(1+c\int_0^1\sigma^2(x,u)k_\varsigma(x)dx)}\right)du \\ &- \int_{[0,1]^2}\frac{\sigma^2(x,y)k_\varsigma(x)}{1+c\int_0^1\sigma^2(u,y)k_\varsigma(u)du}dx\,dy \end{aligned}$$

Mathematical details are provided in [7] and [8].

III. The general case

In the general case, that is when $A_n \neq 0$ one cannot expect the convergence of the empirical distribution of the eigenvalues of $\Sigma_n\Sigma_n^*$ in the case where $A_n \neq 0$. Only very specific cases can be studied ([3], [7]) in a fully asymptotic perspective. However, one can still compute a deterministic approximation as in Theorem 2.2.

Assumption A-2: We assume that the $N \times n$ matrix $A_n = (A_{ij}^n)$ whose columns $(\mathbf{a}_k^n)_{1\le k\le n}$ and rows $(\tilde{\mathbf{a}}_\ell^n)_{1\le\ell\le N}$ satisfies

$$\sup_{n\ge1}\max_{k,\ell}(\|\mathbf{a}_k^n\|, \|\tilde{\mathbf{a}}_\ell^n\|) < +\infty \quad (3)$$

where $\|\cdot\|$ stands for the Euclidean norm.

Theorem 3.1 (see [8], see also [5]): Assume that (A-2) holds and let A_n be a $N \times n$ deterministic matrix. The deterministic system of $N+n$ equations:

$$\psi_i(z) = \frac{-1}{z\left(1+\frac{1}{n}\mathrm{Tr}\,\tilde{D}_i\tilde{T}(z)\right)} \quad \text{for } 1 \le i \le N, \quad (4)$$

$$\tilde{\psi}_j(z) = \frac{-1}{z\left(1+\frac{1}{n}\mathrm{Tr}\,D_jT(z)\right)} \quad \text{for } 1 \le j \le n, \quad (5)$$

where

$$\begin{aligned} \Psi(z) &= \mathrm{diag}(\psi_i(z), 1 \le i \le N), &(6)\\ \tilde{\Psi}(z) &= \mathrm{diag}(\tilde{\psi}_j(z), 1 \le j \le n), &(7)\\ T(z) &= \left(\Psi^{-1}(z) - zA\tilde{\Psi}(z)A^*\right)^{-1}, &(8)\\ \tilde{T}(z) &= \left(\tilde{\Psi}^{-1}(z) - zA^*\Psi(z)A\right)^{-1}. &(9) \end{aligned}$$

admits a unique solution $(\psi_1, \ldots, \psi_N, \tilde{\psi}_1, \ldots, \tilde{\psi}_n)$ in the class of the functions which are Stieltjes transforms.

In the sequel, we denote by $\Psi_\varsigma = \Psi(-\varsigma^2)$ and by $\tilde{\Psi}_\varsigma = \tilde{\Psi}(-\varsigma^2)$.

Theorem 3.2: Denote by $\bar{C}_n(\varsigma^2)$ the quantity

$$\begin{aligned} \bar{C}_n(\varsigma^2) = &\frac{1}{N}\sum_{i=1}^{N}\log\det\left[\frac{\Psi_\varsigma^{-1}}{\varsigma^2} + A\tilde{\Psi}_\varsigma A^*\right] \\ &+ \frac{1}{N}\log\det\frac{\tilde{\Psi}_\varsigma^{-1}}{\varsigma^2} \\ &- \frac{\varsigma^2}{Nn}\sum_{\substack{i=1:N \\ j=1:n}}\sigma_{ij}^2 T_i(-\varsigma^2)\tilde{T}_j(-\varsigma^2) \quad (10) \end{aligned}$$

Then the following holds true:

$$\frac{1}{N}\mathbb{E}\log\det\left(I_N+\frac{\Sigma\Sigma^*}{\varsigma^2}\right)-\bar{C}_n(\varsigma^2)\xrightarrow[n\to\infty]{}0.$$

Mathematical details are provided in [8].

IV. The general case (revisited)

In this section, we assume that the variance profile $\sigma_{ij}^{(n)}$ is separable:

Assumption A-3: The variance profile $\sigma_{ij}^{(n)}$ is assumed to be separable, i.e.:

$$\sigma_{ij}^{(n)}=d_i\tilde{d}_j;\quad 1\le i\le N,\ 1\le j\le n.$$

As we shall see, Assumption (A-3) induces major simplification over the system of $N+n$ equations of Theorem 3.1 since the system is reduced to 2 equations in this case (in accordance with [9] for instance). Denote by

$$\begin{aligned} D &= \mathrm{diag}(d_i,\ 1\le i\le N)\\ \tilde{D} &= \mathrm{diag}(\tilde{d}_j,\ 1\le j\le n)\end{aligned}$$

Theorem 4.1 (see [8]): Assume that (A-3) holds and consider the following system of equations

$$\left\{\begin{array}{l}\delta(z)=\frac{1}{n}\mathrm{Tr}\left[D\left(-z(I+D\tilde{\delta})+A(I+\tilde{D}\delta)^{-1}A^T\right)^{-1}\right]\\ \\ \tilde{\delta}(z)=\frac{1}{n}\mathrm{Tr}\left[\tilde{D}\left(-z(I+\tilde{D}\delta)+A^T(I+D\tilde{\delta})^{-1}A\right)^{-1}\right]\end{array}\right.$$

Then this system admits a unique solution in the class of Stieltjes transforms of positive measures μ and $\tilde{\mu}$ such that $\mu(\mathbb{R}^+)=\frac{1}{n}\mathrm{Tr}D$ and $\tilde{\mu}(\mathbb{R}^+)=\frac{1}{n}\mathrm{Tr}\tilde{D}$.

We can now define properly the related quantities $T,\tilde{T},\Psi$ and $\tilde{\Psi}$ as:

$$\begin{aligned}\Psi(z) &= -\frac{(I+\tilde{\delta}D)^{-1}}{z},\ \tilde{\Psi}(z)=-\frac{(I+\delta\tilde{D})^{-1}}{z} \quad (11)\\ T(-z) &= \left(-z(1+\tilde{\delta}D)+A(I+\delta\tilde{D})^{-1}A^*\right)^{-1} \quad (12)\\ \tilde{T}(-z) &= \left(-z(1+\delta\tilde{D})+A(I+\tilde{\delta}D)^{-1}A^*\right)^{-1} \quad (13)\end{aligned}$$

and accordingly their evaluations at the point $z=-\varsigma^2$: Ψ_ς, $\tilde{\Psi}_\varsigma$, T_ς and $\tilde{T}_\varsigma$.

Theorem 4.2: The statement of Theorem 3.2 remains valid with $T,\tilde{T},\Psi$ and $\tilde{\Psi}$ given by (11), (12) and (13).

V. From independence to stationarity: The case of Gaussian matrices

We now turn to the relation between random matrices based on a Gaussian stationary field and matrices with independent entries and a variance profile.

Assumption A-4: Consider the $N\times n$ matrix whose entries are given by

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

where h is a deterministic complex summable sequence and $(U(j_1,j_2);(j_1,j_2)\in\ \mathbb{Z}^2)$ is a sequence of $\mathcal{C}N(0,1)$ random variables.

Such a matrix is a good model for a Gaussian stationary field since every entry $Z_{j_1j_2}^n$ is complex gaussian, centered and

$$\mathrm{cov}(Z_{j_1j_2}^n,Z_{j_1'j_2'}^n)=\frac{\kappa(j_1-j_1',j_2-j_2')}{n}$$

where

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

Consider on the other hand the $N\times n$ matrix $Y_n=(Y_{j_1,j_2}^n)$ where

$$Y_{j_1,j_2}^n=\frac{\Phi\left(\frac{j_1}{N},\frac{j_2}{n}\right)}{\sqrt{n}}X_{j_1,j_2} \quad (14)$$

where the (X_{j_1,j_2}) are i.i.d. $\mathcal{C}N(0,1)$ random variables and

$$\Phi(t_1,t_2)=\sum_{(\ell_1,\ell_2)\in\mathbb{Z}^2}h(\ell_1,\ell_2)e^{2\pi\mathbf{i}(\ell_1t_1-\ell_2t_2)} \quad (15)$$

The similar asymptotic behavior of the spectral measure of $Z_nZ_n^*$ and $Y_nY_n^*$ are part of the folklore of the digital communication literature. We give here a formal justification to this fact, based on [6], and extend Theorem 3.2 to the case of matrices with Gaussian stationary entries. The following holds true:

Theorem 5.1 (see [6]): Let $H_n=Z_n+B_n$ where B_n satisfies (A-2) and Z_n satisfies (A-4). Then the conclusions of Theorems 3.1 and 3.2 remain valid with the following slight modifications:

$$\begin{aligned} D_j &= \mathrm{diag}\left\{|\Phi|^2\left(\frac{i}{N},\frac{j}{n}\right);\ 1\le i\le N\right\};\\ \tilde{D}_i &= \mathrm{diag}\left\{|\Phi|^2\left(\frac{i}{N},\frac{j}{n}\right);\ 1\le j\le n\right\};\\ A &= F_N^*BF_n.\end{aligned}$$

where Φ is given by (15) and F_N and F_n are Fourier matrices defined by (1). Moreover,

$$\frac{1}{N}\mathbb{E}\log\det\left(I+\frac{HH^*}{\varsigma^2}\right)-\bar{C}_n(\varsigma^2)\xrightarrow[n\to\infty]{}0,$$

where $\bar{C}_n(\varsigma^2)$ is given by (10).

Elements of proof

The proof of Theorem 5.1 relies on two main components.

1) A periodization scheme popular in signal processing. We introduce the matrix $\tilde{Z}_n=(\tilde{Z}_{j_1j_2}^n)$ where

$$\begin{aligned}\tilde{Z}_{j_1j_2}^n=\frac{1}{\sqrt{n}}&\sum_{(k_1,k_2)\in\mathbb{Z}^2}h(k_1,k_2)\\ &\times U\left((j_1-k_1)\bmod N,(j_2-k_2)\bmod n\right),\end{aligned}$$

and mod denotes modulo. The main interest of matrix $\tilde{Z}_n$ comes from the fact that it can be fully decorrelated by Fourier multiplication:

$$F_N^*\tilde{Z}_nF_n=Y_n,$$

where Y_n is defined by (14).

2) The second element is an inequality due to Bai [1] involving the Lévy distance $\mathcal{L}$ between distribution functions:

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

where F^{AA^*} denotes the empirical distribution function of the eigenvalues of the matrix AA^*. This inequality turns out to be perfectly suited to evaluate the difference between the spectrum of matrices $Z_n Z_n^*$ (resp. $(Z_n + B_n)(Z_n + B_n)^*$) and $\tilde{Z}_n \tilde{Z}_n^*$ (resp. $(\tilde{Z}_n + B_n)(\tilde{Z}_n + B_n)^*$)

Mathematical details are provided in [6].

References

[1] Z. D. Bai. Convergence rate of expected spectral distributions of large random matrices. II. Sample covariance matrices. *Ann. Probab.*, 21(2):649–672, 1993.

[2] 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.

[3] R. B. Dozier and J. W. Silverstein. On the empirical distribution of eigenvalues of large dimensional information-plus-noise type matrices. preprint 2004.

[4] V. L. Girko. *Theory of random determinants*, volume 45 of *Mathematics and its Applications (Soviet Series)*. Kluwer Academic Publishers Group, Dordrecht, 1990.

[5] V. L. Girko. *Theory of stochastic canonical equations. Vol. I*, volume 535 of *Mathematics and its Applications*. Kluwer Academic Publishers, Dordrecht, 2001.

[6] 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.

[7] W. Hachem, P. Loubaton, and J. Najim. The empirical distribution of the eigenvalues of a Gram matrix with a given variance profile. available at http://arxiv.org/abs/math.PR/0411333, 2004, to be published in Annales de l'I.H.P., Probab. & Stat.

[8] W. Hachem, P. Loubaton, and J. Najim. Deterministic equivalents for certain functionals of large random matrices. available at http://arxiv.org/abs/math.PR/0507172, 2005.

[9] 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.

[10] I.E. Telatar. Capacity of multi-antenna gaussian channel. *European Transactions on Telecommunications*, 1999.

[11] 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, 2005.

[12] 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.