A G-Estimator of the MIMO Channel Ergodic Capacity

Pascal Vallet, Philippe Loubaton
IGM LabInfo, UMR-CNRS 8049
Université Paris-Est
5, Boulevard Descartes, Champs-sur-Marne
77454 Marne-la-Vallée Cedex 2, France
Email: {vallet,loubaton}@univ-mlv.fr

***Abstract*—In this paper, we deal with the estimation of the ergodic capacity of large MIMO systems, using training sequences whose lengths are of the same order of magnitude than the number of antennas. In this context, the traditional estimator becomes inconsistent. Following the ideas developed by Girko in the context of the so-called theory of G-estimation, we propose a new estimator. We analyze its asymptotic behaviour and show, using numerical simulations, that it tends to improve significantly the performance of the standard estimate for a realistic number of antennas.**

I. Introduction

One of the goal of channel sounding methods is to evaluate the influence of the environment on relevant figure of merits in order to predict the potential performance of digital communication systems. In this paper, we address the estimation of the ergodic capacity of mobile MIMO systems in the context of channel sounding. Due to the mobility, the size of the training sequence from which the channel can be estimated is of course limited because the MIMO channel matrix has to be constant on the duration of the training sequence. If the size L of the training sequence is of the same order of magnitude than the number of antennas of the system, the traditional estimate of the capacity, built on the training based estimate of the channel matrix, provides poor performance because the influence of the background noise is not negligible. We can mention the recent work of [1] in which an estimator of the capacity is derived using Free Probability Theory; however it is not consistent in the case considered here.

In order to design new improved estimates, we use the theory of large random matrices, and assume that the number of receive and transmit antennas M and N as well as the size of the training sequence L converge to $+\infty$ at the same rate. In this asymptotic regime, we propose a new estimator of the channel ergodic capacity inspired by the so-called G-estimation theory developed by Girko.

The paper is organized as follows. In section II, we present the signal model, the precise context of the estimation and the addressed problem. In section III, we provide some useful results in Large Random Matrix Theory (LRMT). In section IV, we propose a new consistent "G-estimator" of the MIMO channel ergodic capacity. In section V, we provide some numerical simulations which illustrate the previous results.

II. Model and problem's statement

We consider a MIMO channel sounder with N transmit antennas and M receive antennas, and we suppose that the channel is a MIMO block fading channel. In order to estimate the ergodic capacity of the channel, the transmitter sends periodically a training sequence of size L, denoted $\mathbf{S} = [\mathbf{s}_1, \dots, \mathbf{s}_L]$, where $\mathbf{s}_1, \dots, \mathbf{s}_L$ are N-dimensional vectors for which $\mathbf{S}\mathbf{S}^\dagger = L\mathbf{I}_N$. The channel matrix is assumed to be constant on each "slot", i.e during the transmission of each training sequence. We denote by T the number of avalaible slots, and we assume that the statistical properties of the channel are time invariant during the transmission of these T slots. If $\mathbf{H}_{(t)}$ (for $t = 1, \dots, T$) represents the channel matrix on slot t, normalized in such a way that $\mathbb{E}[\frac{1}{M}\mathrm{Tr}\ \mathbf{H}_{(t)}\mathbf{H}_{(t)}^\dagger] = 1$, we address the estimation problem of

$$C(\sigma^2) = \frac{1}{N}\frac{1}{T}\sum_{t=1}^{T}\log\det\left(\mathbf{I}_M + \frac{\mathbf{H}_{(t)}\mathbf{H}_{(t)}}{\sigma^2}\right) \qquad (1)$$

which clearly represents an approximation of the ergodic capacity (per transmit antenna). For each $t = 1, \dots, T$, the receive $M \times L$ channel matrix $\mathbf{Y}_{(t)}$ corresponding to the transmission of the sequence $\mathbf{S}$ can be written as

$$\mathbf{Y}_{(t)} = \mathbf{H}_{(t)}\mathbf{S} + \mathbf{V}_{(t)} \qquad (2)$$

where $\mathbf{V}_{(t)}$ is the $M \times L$ noise matrix, assumed to be iid Gaussian, with entries having zero mean and variance σ_v^2. The training-based estimate $\mathbf{Z}_{(t)}$ of $\mathbf{H}_{(t)}$ is defined by

$$\mathbf{Z}_{(t)} = \frac{1}{L}\mathbf{Y}_{(t)}\mathbf{S}^\dagger = \mathbf{H}_{(t)} + \frac{\rho}{\sqrt{N}}\mathbf{W}_{(t)} \qquad (3)$$

with $\rho = \sqrt{\frac{\sigma_v^2 N}{L}}$ and $\mathbf{W}$ a $M \times N$ Gaussian random matrix whose elements are iid, unit variance and zero mean. In order to keep ρ constant, we suppose in this paper that the length of the learning sequence L is linearly increasing with N.

One classical way to estimate (1) is to replace $\mathbf{H}_{(t)}\mathbf{H}_{(t)}^\dagger$ by its standard unbiased estimate $\mathbf{Z}_{(t)}\mathbf{Z}_{(t)}^\dagger - \rho^2\mathbf{I}_M$. However, some eigenvalues of $\mathbf{Z}_{(t)}\mathbf{Z}_{(t)}^\dagger$ may be less than ρ^2. If we denote by $\hat{\lambda}_1^{(t)}, \dots, \hat{\lambda}_M^{(t)}$ the eigenvalues of $\mathbf{Z}_{(t)}\mathbf{Z}_{(t)}^\dagger$ (in increasing order), the standard estimator of (1) is given by

$$\hat{C}_{\mathrm{trad}}(\sigma^2) = \frac{1}{N}\frac{1}{T}\sum_{t=1}^{T}\sum_{m=1}^{M}\log\left(1+\frac{\left(\hat{\lambda}_m^{(t)}-\rho^2\right)^+}{\sigma^2}\right) \quad (4)$$

where $(.)^+$ is the positive part. This estimator is relevant in the case where $L \to +\infty$ and $N/L \to 0$. However, this asymptotic regime is not realistic in the context of mobile systems. We therefore assume that $N \to +\infty$, $L \to +\infty$ in such a way that $\rho = \sqrt{\frac{\sigma_v^2 N}{L}}$ remains constant. We also assume that $M < N$, $M \to +\infty$, $N \to +\infty$ and that $c_N = \frac{M}{N}$ converges to a constant $c \in]0,1[$. In this context, (4) is not consistent because $\|\mathbf{Z}_{(t)}\mathbf{Z}_{(t)}^\dagger - (\rho^2\mathbf{I}_M + \mathbf{H}_{(t)}\mathbf{H}_{(t)}^\dagger)\|$ does not converge anymore to 0.

Using LRMT tools, we are able to find a consistent estimator of (1) in the latter case. Note that our results can be extended to the case $c > 1$ by considering matrix $\mathbf{Z}^\dagger\mathbf{Z}$ instead of $\mathbf{Z}\mathbf{Z}^\dagger$.

III. Review of some useful LRMT results

In literature, the model (3) is commonly referred as the "Information plus Noise model" and we now summarize some published results concerning the statistical spectral properties of the random matrix $\mathbf{Z}\mathbf{Z}^\dagger$ (dependency in slot t is dropped only for this section). First, we denote respectively by $\lambda_1,\dots,\lambda_M$ and $\hat{\lambda}_1,\dots,\hat{\lambda}_M$ the eigenvalues (in increasing order) of $\mathbf{H}\mathbf{H}^\dagger$ and $\mathbf{Z}\mathbf{Z}^\dagger$.

For an easier reading, we do not mention explicitely the dependency in M, N, L of all quantities involving matrix $\mathbf{H}$. Moreover, when we will say "$N \to +\infty$", it will stand for "$M, N, L \to \infty$ while $c_N \to c \in]0,1[$ and ρ remains constant". The notation "$\overline{X} \approx X$" will also stand for "$\overline{X} - X \to 0$ with probability one as $N \to +\infty$" with $\overline{X}$ deterministic and X random.

Consider the empirical spectral measure $\hat{\mu}$ of the random matrix $\mathbf{Z}\mathbf{Z}^\dagger$ given by

$$\hat{\mu} = \frac{1}{M}\sum_{m=1}^{M}\delta_{\hat{\lambda}_m} \quad (5)$$

where $\delta_{\hat{\lambda}_m}$ is the Dirac measure at the eigenvalue $\hat{\lambda}_m$. It is useful to characterize $\hat{\mu}$ in terms of its Stieltjes transform $\Upsilon_{\hat{\mu}}$ defined $\forall z \in \mathbb{C} - \{\hat{\lambda}_1,\dots,\hat{\lambda}_M\}$ by

$$\Upsilon_{\hat{\mu}}(z) = \int_{\mathbb{R}^+}\frac{\hat{\mu}(\mathrm{d}\lambda)}{\lambda - z} = \frac{1}{M}\mathrm{Tr}\left(\mathbf{Z}\mathbf{Z}^\dagger - z\mathbf{I}_M\right)^{-1} \quad (6)$$

Clearly, $\hat{\mu}$ is random and the following fundamental theorem (see [2]) gives us details about its asymptotic behaviour.

Theorem 1. *Assume that* $\sup_{M,N}\|\mathbf{H}\| < \infty$.
Then as $N \to +\infty$, with probability one, $\hat{\mu}$ converges in distribution toward a deterministic measure μ called the asymptotic eigenvalue distribution of $\mathbf{Z}\mathbf{Z}^\dagger$.
Measure μ is characterized by the function $\delta(x) = \rho c \Upsilon_\mu(x)$ where Υ_μ is the Stieltjes transform of μ. $\forall x \in \mathbb{R}_*^-$, $\delta(x)$ is the solution to the equation

$$\delta(x) = \frac{\rho}{N}\mathrm{Tr}\ \mathbf{T}(x)$$
$$\mathbf{T}(x) = \left[\left(-x(1+\rho\delta(x)) + \rho^2(1-c)\right)\mathbf{I}_M + \frac{\mathbf{H}\mathbf{H}^\dagger}{1+\rho\delta(x)}\right]^{-1}$$

Moreover, if we define $\mathcal{S} = \mathrm{supp}(\mu)$, i.e the support of the asymptotic eigenvalue distribution of $\mathbf{Z}\mathbf{Z}^\dagger$, then $\delta(x)$ is analytic on $\mathbb{R} - \mathcal{S}$.

Thanks to (6) and Theorem 1, we have $\forall x \in \mathbb{R}_*^-$, $\Upsilon_\mu(x) \approx \Upsilon_{\hat{\mu}}(x)$ and therefore

$$\delta(x) \approx \hat{\delta}(x) \quad (7)$$

with $\hat{\delta}(x) = \frac{\rho}{N}\mathrm{Tr}\ \left(\mathbf{Z}\mathbf{Z}^\dagger - x\mathbf{I}_M\right)^{-1}$ Thus, $\delta(x)$ is consistently estimated by $\hat{\delta}(x)$ on $\mathbb{R}_*^-$.

The next theorem (see [3]) characterizes the infimum of $\mathcal{S}$, the support of μ.

Theorem 2. *We consider the case where $\mathbf{H}\mathbf{H}^\dagger$ is full rank and $N \to +\infty$. Let $\phi(w)$ be the function defined on $\mathbb{R} - \{\lambda_1,\dots,\lambda_M\}$ by*

$$\phi(w) = w(1-\rho^2 f(w))^2 + \rho^2(1-c)(1-\rho^2 f(w))$$

with $f(w) = \frac{1}{N}\mathrm{Tr}[\mathbf{H}\mathbf{H}^\dagger - w\mathbf{I}_M]^{-1}$. Let w^- be the smallest local extremum of ϕ satisfying $1-\rho^2 f(w^-) > 0$ and $\phi(w^-) > 0$. Then we have

$$x^- = \phi(w^-) = \inf\ \mathcal{S}$$

Moreover, ϕ is a strictly increasing bijective function from $]-\infty, w^-[$ to $]-\infty, x^-[$. Let $\psi(x)$ be its inverse defined on $]-\infty, x^-[$. Then, we have

$$\psi(x) = x\left(1+\rho\delta(x)\right)^2 - \rho^2(1-c)\left(1+\rho\delta(x)\right)$$

and for $w < w^-$, the equation $\psi(x) = w$ has a unique solution in $]-\infty, x^-[$ equal to $\phi(w)$.

Figure 1 and 2 represent two possible behaviours of ϕ when $c < 1$. We also represent the point w_0, solution of $\phi(w) = 0$ in the interval $]-\infty, w^-[$, which will be useful for Proposition 1 in the next section. Note that in the two cases, 0 does not belong to $\mathcal{S}$, because $c < 1$. We now state a fundamental result

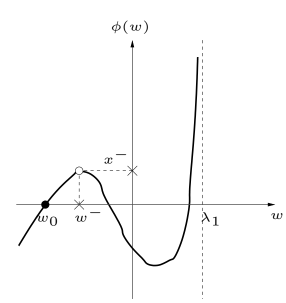

Fig. 1. Allure of ϕ on $]-\infty, \lambda_1[$ when $c < 1$ and $f(0) > \frac{1}{\rho^2}$

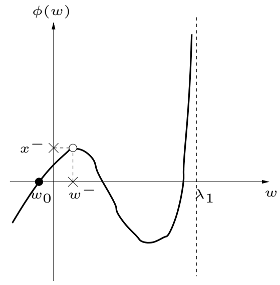

Fig. 2. Allure of ϕ on $]-\infty,\lambda_1[$ when $c<1$ and $f(0)<\frac{1}{\rho^2}$

on the eigenvalues of $\mathbf{ZZ}^\dagger$. It can be found in [4] (Theorem 23.1 p.267).

Theorem 3. *If M,N are large enough, then almost surely, the eigenvalues of $\mathbf{ZZ}^\dagger$ belong to $\mathcal{S}$.*

We now give a theorem which is a direct consequence of Theorem 1 and which can be found in [5] and [6].

Theorem 4. *Consider the function $\overline{C}$ defined on $\mathbb{R}_*^-$ by*

$$\begin{aligned}&\overline{C}(x)=\\&\frac{1}{N}\log\det\left(\mathbf{I}_M-\frac{\mathbf{HH}^\dagger}{\psi(x)}\right)\\&\quad+\log\left(1+\rho\delta(x)\right)\\&\quad+c\log\left(1+\rho\delta(x)-\frac{\rho^2(1-c)}{x}\right)\\&\quad+x\delta(x)\left(\delta(x)-\frac{\rho(1-c)}{x}\right)\end{aligned}$$

Then,

$$\overline{C}(x)\approx\frac{1}{N}\log\det\left(\mathbf{I}_M-\frac{1}{x}\mathbf{ZZ}^\dagger\right)\tag{8}$$

Moreover, we have

$$\mathbb{E}\left[\frac{1}{N}\log\det\left(\mathbf{I}_M-\frac{1}{x}\mathbf{ZZ}^\dagger\right)\right]=\overline{C}(x)+\mathcal{O}\left(\frac{1}{N^2}\right)\tag{9}$$

$$\mathrm{Var}\left[\frac{1}{N}\log\det\left(\mathbf{I}_M-\frac{1}{x}\mathbf{ZZ}^\dagger\right)\right]=\mathcal{O}(\frac{1}{N^2})\tag{10}$$

IV. Derivation of the G-estimator

We present a new consistent estimator of (1), based on Theorem 4. In what follows, we add the notation (t) for all the quantities of the previous sections which depend on the matrix $\mathbf{H}_{(t)}$ (i.e on a certain slot time t). In particular, we have $x_{(t)}^-=\inf\mathcal{S}_{(t)}$.

In order to have a better understanding of the techniques involved in the derivation of the G-estimator, we derive here an estimator in a simple case.

Proposition 1. *As $N\to+\infty$, $\forall\ \sigma^2>-w_{0,(t)}$ we have*

$$\begin{aligned}&\frac{1}{N}\log\det\left(\mathbf{I}_M+\frac{\mathbf{H}_{(t)}\mathbf{H}_{(t)}^\dagger}{\sigma^2}\right)\approx\\&\frac{1}{N}\log\det\left(\mathbf{I}_M-\frac{\mathbf{Z}_{(t)}\mathbf{Z}_{(t)}^\dagger}{\hat{x}_*^{(t)}}\right)-\log(1+\rho\hat{\delta}_{(t)}(\hat{x}_*^{(t)}))\\&-c\log\left(1+\rho\hat{\delta}_{(t)}(\hat{x}_*^{(t)})-\frac{\rho^2(1-c)}{\hat{x}_*^{(t)}}\right)\\&-\hat{x}_*^{(t)}\hat{\delta}_{(t)}(\hat{x}_*^{(t)})\left(\hat{\delta}_{(t)}(\hat{x}_*^{(t)})-\frac{\rho(1-c)}{\hat{x}_*^{(t)}}\right)\end{aligned}$$

with $\hat{x}_*^{(t)}$ the solution to the equation $\hat{\psi}_{(t)}(x)=-\sigma^2$ and $\hat{\psi}_{(t)}$ defined on $\mathbb{R}_*^-$ by

$$\hat{\psi}_{(t)}(x)=x\left(1+\rho\hat{\delta}_{(t)}(x)\right)^2-\rho^2(1-c)\left(1+\rho\hat{\delta}_{(t)}(x)\right)\tag{11}$$

Proof: Let $x_*^{(t)}$ be the solution to the equation

$$\psi_{(t)}(x)=-\sigma^2\tag{12}$$

As we have $-\sigma^2<w_{0,(t)}<w_{(t)}^-$ and $\phi_{(t)}(w_{0,(t)})=0$, we deduce from Theorem 2 that it exists a unique solution $x_*^{(t)}$ to the equation (12) such that $x_*^{(t)}<0$. Therefore we can use Theorem 4 (because it is valid on $\mathbb{R}_*^-$) to obtain

$$\begin{aligned}&\frac{1}{N}\log\det\left(\mathbf{I}_M+\frac{\mathbf{H}_{(t)}\mathbf{H}_{(t)}^\dagger}{\sigma^2}\right)=\\&\overline{C}_{(t)}(x_*^{(t)})-x_*^{(t)}\delta_{(t)}(x_*^{(t)})\left(\delta_{(t)}(x_*^{(t)})-\frac{\rho(1-c)}{x_*^{(t)}}\right)\\&-\log\left(1+\rho\delta_{(t)}(x_*^{(t)})\right)-c\log\left(1+\rho\delta_{(t)}(x_*^{(t)})-\frac{\rho^2(1-c)}{x_*^{(t)}}\right)\end{aligned}\tag{13}$$

From (7), we can approximate $\delta_{(t)}(x)$ by $\hat{\delta}_{(t)}(x)$, and therefore function $\psi_{(t)}(x)\approx\hat{\psi}_{(t)}(x)$ for $x\in\mathbb{R}_*^-$. If $\hat{x}_*^{(t)}$ is the random solution to $\hat{\psi}_{(t)}(x)=-\sigma^2$, we clearly have $x_*^{(t)}\approx\hat{x}_*^{(t)}$. Thus we deduce from (7) and (8)

$$\begin{aligned}\delta_{(t)}(x_*^{(t)})&\approx\hat{\delta}_{(t)}(\hat{x}_*^{(t)})\\\psi_{(t)}(x_*^{(t)})&\approx\hat{\psi}_{(t)}(\hat{x}_*^{(t)})\\\overline{C}_{(t)}(x_*^{(t)})&\approx\frac{1}{N}\log\det\left(\mathbf{I}_M-\frac{1}{\hat{x}_*^{(t)}}\mathbf{Z}_{(t)}\mathbf{Z}_{(t)}^\dagger\right)\end{aligned}$$

By replacing the previous approximations in the righthand side of (13) and in equation (12) we get the final result. ■

We now extend the domain of convergence for (7) and (8).

Proposition 2. *As $N\to+\infty$, we have $\forall x\in]-\infty,x_{(t)}^-[$,*

$$\delta_{(t)}(x)\approx\hat{\delta}_{(t)}(x)\tag{14}$$

with $\hat{\delta}_{(t)}(x)=\frac{\rho}{N}\mathrm{Tr}\ \left(\mathbf{Z}_{(t)}\mathbf{Z}_{(t)}^\dagger-x\mathbf{I}_M\right)^{-1}$.

Proof: The proof, which relies on Theorem 2, 3 and Montel's theorem, is omitted here. ■

Proposition 3. *As* $N \to +\infty$, $\forall\ \sigma^2 > -\min_{t=1,\ldots,T}\left\{w_{(t)}^-\right\}$, $C(\sigma^2)$ *given by* (1) *is consistently estimated by*

$$\hat{C}_{\text{new}}\left(\sigma^2\right) = \frac{1}{T}\sum_{t=1}^{T}\Bigg[\frac{1}{N}\log\det\left(\mathbf{Z}_{(t)}\mathbf{Z}_{(t)}^\dagger - \hat{x}_*^{(t)}\mathbf{I}_M\right) + (c-1)\log\left(1+\rho\hat{\delta}_{(t)}(\hat{x}_*^{(t)})\right) - \hat{x}_*^{(t)}\left(\hat{\delta}_{(t)}(\hat{x}_*^{(t)}) - \frac{\rho(1-c)}{\hat{x}_*^{(t)}}\right)\Bigg] - c\log\left(\sigma^2\right)$$

with $\hat{x}_*^{(t)}$ *the solution to the equation*

$$\hat{\psi}_{(t)}(x) = -\sigma^2 \tag{15}$$

Proof: We consider a particular slot t. Thanks to Theorem 3, we know that for M, N large enough and $x < x_{(t)}^-$, functions $\hat{\delta}_{(t)}(x)$ and $\frac{1}{N}\log\det\left(\mathbf{Z}_{(t)}\mathbf{Z}_{(t)}^\dagger - x\mathbf{I}_M\right)$ are defined and integrable on $]-\infty, x[$, with probability one. We have

$$\frac{\mathrm{d}}{\mathrm{d}x}\left[\frac{1}{N}\log\det\left(\mathbf{Z}_{(t)}\mathbf{Z}_{(t)}^\dagger - x\mathbf{I}_M\right)\right] = -\frac{1}{\rho}\hat{\delta}_{(t)}(x)$$

Thus, for $x \in]-\infty, x_{(t)}^-[$ and $x_0 > 0$ such that $-x_0 < x$, we can write

$$\frac{1}{N}\log\det\left(\mathbf{Z}_{(t)}\mathbf{Z}_{(t)}^\dagger - x\mathbf{I}_M\right) = \frac{1}{N}\log\det\left(\mathbf{Z}_{(t)}\mathbf{Z}_{(t)}^\dagger + x_0\mathbf{I}_M\right) - \frac{1}{\rho}\int_{-x_0}^{x}\hat{\delta}_{(t)}(u)\mathrm{d}u \tag{16}$$

From Proposition 2, we have $\hat{\delta}_{(t)}(u) \approx \delta_{(t)}(u)$ on $]-\infty, x_{(t)}^-[$. $\hat{\delta}_{(t)}$ is integrable on $]-\infty, x[$, and by using the dominated convergence theorem, we get

$$\int_{-x_0}^{x}\hat{\delta}_{(t)}(u)\mathrm{d}u \approx \int_{-x_0}^{x}\delta_{(t)}(u)\mathrm{d}u \tag{17}$$

From Theorem 4, as $x_0 > 0$, we have

$$c\log x_0 + \overline{C}_{(t)}(-x_0) \approx \frac{1}{N}\log\det\left(\mathbf{Z}_{(t)}\mathbf{Z}_{(t)}^\dagger + x_0\mathbf{I}_M\right) \tag{18}$$

Straightforward calculations lead to the following derivative

$$-\frac{1}{\rho}\delta_{(t)}(x) = -\frac{1}{N}\mathrm{Tr}\ \mathbf{T}_{(t)}(x) = \frac{\mathrm{d}}{\mathrm{d}x}\Bigg[\frac{1}{N}\log\det\left[\mathbf{T}_{(t)}^{-1}(x)\right] + \log\left(1+\rho\delta_{(t)}(x)\right) + x\delta_{(t)}(x)\left(\delta_{(t)}(x) - \frac{\rho(1-c)}{x}\right)\Bigg]$$

Therefore, we deduce that

$$-\frac{1}{\rho}\int_{-x_0}^{x}\delta_{(t)}(u)\mathrm{d}u = \frac{1}{N}\log\det\left[\mathbf{T}_{(t)}^{-1}(x)\right] + \log\left(1+\rho\delta_{(t)}(x)\right) + x\delta_{(t)}(x)\left(\delta_{(t)}(x) - \frac{\rho(1-c)}{x}\right) - \overline{C}_{(t)}(-x_0) - c\log(x_0) \tag{19}$$

It follows from (16), (18) and (19) that

$$\frac{1}{N}\log\det\left(\mathbf{Z}_{(t)}\mathbf{Z}_{(t)}^\dagger - x\mathbf{I}_M\right) \approx \frac{1}{N}\log\det\left[\mathbf{T}_{(t)}^{-1}(x)\right] + \log\left(1+\rho\delta_{(t)}(x)\right) + x\delta_{(t)}(x)\left(\delta_{(t)}(x) - \frac{\rho(1-c)}{x}\right) \tag{20}$$

Moreover, we notice that

$$\frac{1}{N}\log\det\left[\mathbf{T}_{(t)}^{-1}(x)\right] = c\log\left(\psi(x)\right) - c\log\left(1+\rho\delta_{(t)}(x)\right) + \frac{1}{N}\log\det\left[-\frac{\mathbf{H}_{(t)}\mathbf{H}_{(t)}^\dagger}{\psi(x)} + \mathbf{I}_M\right] \tag{21}$$

As $-\sigma^2 < w_{(t)}^-$, we know from Theorem 2 that the equation $\psi(x) = -\sigma^2$ has a unique solution $x_*^{(t)}$ on $]-\infty, x_{(t)}^-[$. Then, from (20) and (21), we deduce that

$$\frac{1}{N}\log\det\left[\frac{\mathbf{H}_{(t)}\mathbf{H}_{(t)}^\dagger}{\sigma^2} + \mathbf{I}_M\right] \approx \frac{1}{N}\log\det\left(\mathbf{Z}_{(t)}\mathbf{Z}_{(t)}^\dagger - x_*^{(t)}\mathbf{I}_M\right) - x_*^{(t)}\left(\delta_{(t)}(x_*^{(t)}) - \frac{\rho(1-c)}{x_*^{(t)}}\right) + (c-1)\log\left(1+\rho\delta_{(t)}(x_*^{(t)})\right) - c\log(\sigma^2) \tag{22}$$

Now denote by $\hat{x}_*^{(t)}$ the solution to the equation $\hat{\psi}_{(t)}(x) = -\sigma^2$. In the same way than in the proof of Proposition 1, and because $-\sigma^2 < w_{(t)}^-$, we have

$$x_*^{(t)} \approx \hat{x}_*^{(t)}$$
$$\delta_{(t)}(x_*^{(t)}) \approx \hat{\delta}_{(t)}(\hat{x}_*^{(t)})$$
$$\psi_{(t)}(x_*^{(t)}) \approx \hat{\psi}_{(t)}(\hat{x}_*^{(t)})$$

Finally, by replacing the previous approximation in (22), we get the final form of the G-estimator. ■

We now study the performance of the previous estimator in terms of bias and MSE.

Property 1. $\forall\ \sigma^2 > -\min_{t=1,\ldots,T}\left\{w_{0,(t)}\right\}$ *the quantity* $\hat{C}_{new}\left(\sigma^2\right)$ *defined in* (15) *satisfies*

$$\mathbb{E}\left[\hat{C}_{new}\left(\sigma^2\right)\right] = C\left(\sigma^2\right) + \mathcal{O}\left(\frac{1}{N^2}\right) \tag{23}$$

$$\mathbb{E}\left[\left(\hat{C}_{new}\left(\sigma^2\right) - C\left(\sigma^2\right)\right)^2\right] = \mathcal{O}\left(\frac{1}{N^2}\right) \tag{24}$$

Proof: The proof relies on formulae (9) and (10) and analysis of higher moments of $\hat{\delta}_{(t)}$. It is omitted here due to lack of space. ■

Remark: Notice that the previous results of bias and MSE are valid $\forall\ \sigma^2 > -\min_{t=1,\ldots,T}\{w_{0,(t)}\}$. They probably remain valid for $\sigma^2 > -\min_{t=1,\ldots,T}\{w_{(t)}^-\}$, but it has not been proven yet.

V. Numerical simulations

In this section, we compare the performance of the traditional and new estimators in terms of bias and MSE.

The channel matrices $\mathbf{H}_{(t)}$ follow the well-known "Kronecker model", i.e

$$\mathbf{H}_{(t)} = \frac{1}{\sqrt{N}} \mathbf{C}_{(t)}^{\frac{1}{2}} \mathbf{X} \widetilde{\mathbf{C}}_{(t)}^{\frac{1}{2}}$$

with $\mathbf{C}_{(t)}, \widetilde{\mathbf{C}}_{(t)}$ real positive definite matrices (respectively of size $M \times M$ and $N \times N$) and $\mathbf{X}$ a random matrix such that $\mathbf{X}_{ij} \sim \mathcal{CN}(0,1)$. For the simulations, $\mathbf{C}_{(t)}, \widetilde{\mathbf{C}}_{(t)}$ are defined by

$$\mathbf{C}_{(t)} = \left[\eta_{1,(t)}^{|i-j|}\right]_{1\leq i,j\leq M} \text{ and } \widetilde{\mathbf{C}}_{(t)} = \left[\eta_{2,(t)}^{|i-j|}\right]_{1\leq i,j\leq N}$$

$\eta_{1,(t)}, \eta_{2,(t)}$ being constants chosen between $]-1,1[$.

In order to keep the quantity ρ constant in (3), we set $L = 2N$. Computing the G-estimator implies to solve equation (15) with numerical techniques. The solution is obtained by using the Newton-Raphson method. Moreover we set $T = 30$, $\sigma_v^2 = 0.25$ and $\sigma^2 = 1$.

In Figure 3 we compared the bias of (4) and (15) when N, M increase ($c = 0.5$). We clearly notice that the bias of (4) is constant with N while the bias of (15) is decreasing. Unfortunately, the rate of convergence of the bias in Figure 3 does not fit with $\frac{1}{N^2}$ for large values of N. This is because the corresponding values of the bias are of the same order of magnitude than the accuracy of the numerical technique which is used to solve the equation $\hat{\psi}(x) = -\sigma^2$.

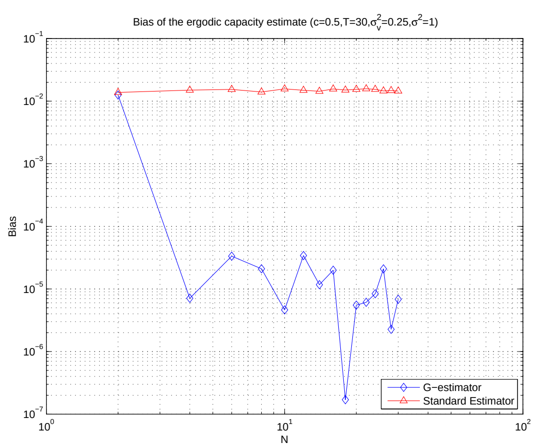

Fig. 3. Bias of the traditional and new estimator versus N

Figure 4 represents the evolution of the MSE with respect to N ($c = 0.5$). We notice that the MSE of (4) is constant while the MSE of (15) is clearly decreasing in $\frac{1}{N^2}$, as predicted in Property 1.

Proposition 3 states that we are not entirely free to choose the value of σ^2 at which we want to evaluate the capacity. Indeed, we are restricted to the condition $\sigma^2 > -\min_{t=1,\ldots,T}\left\{w_{(t)}^-\right\}$ (or to $\sigma^2 > -\min_{t=1,\ldots,T}\left\{w_{0,(t)}^-\right\}$ for the validity of bias and MSE). It can be interesting to see how restrictive is this condition. In Figure 5, we compute for several values of σ_v^2 the minimum of w^- and w_0 over a large number of channel matrices $\mathbf{H}$ ($c = 0.5$ and $L = 2N$) and it proves that the condition on σ^2 is not as restrictive as it appears.

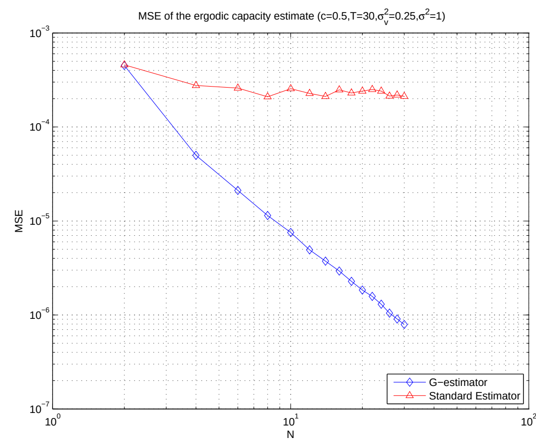

Fig. 4. MSE of the traditional and new estimator versus N

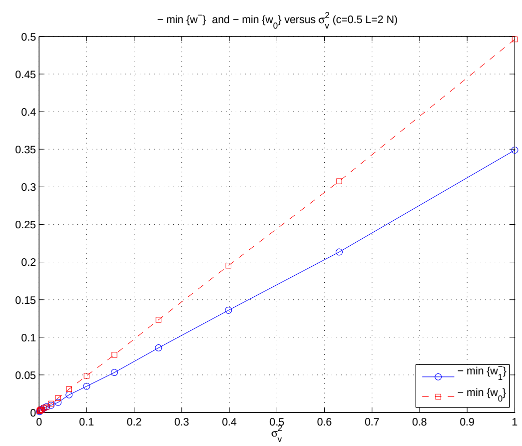

Fig. 5. $-\min\{w^-\}$ and $-\min\{w_0\}$ versus σ_v^2

References

[1] O.Ryan and M.Debbah, "Channel capacity estimation using free probability theory," *IEEE Transactions on Signal Processing*, vol. 56, no. 11, pp. 5654–5667, November 2008.

[2] R. Dozier and J. Silverstein, "On the empirical distribution of eigenvalues of large dimension information-plus-noise type matrices," *Journal of Multivariate Analysis 98*, pp. 678–694, 2007.

[3] P.Vallet, P.Loubaton, and X.Mestre, "Improved subspace doa estimation with large arrays: The deterministic signal case," in *IEEE International Conference on Acoustics, Speech, and Signal Processing*, Taipei (Taiwan), 19-24 Apr. 2009, 4p.

[4] V.L.Girko, *An Introduction to Statistical Analysis of Random Arrays*. VSP Science, December 1998.

[5] 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 Transactions on Information Theory*, available at http://arxiv.org/abs/0710.4051.

[6] G.Taricco, "Asymptotic mutual information statistics of separately correlated rician fading mimo channels," *IEEE Transactions on Information Theory*, vol. 54, no. 8, pp. 3490 – 3504, August 2008.