The mutual information of a MIMO non-centered time and frequency selective channel: an ergodic operator approach

Walid Hachem

CNRS LTCI; Telecom ParisTech

with Aris Moustakas and Leonid Pastur

---

[a] Random matrix theory and statistical physics workshop 2014, IHP

Context

Frequency and time selective MIMO transmission with T transmitting and N receiving antennae. Received signal sequence:

$$Y(k) = \sum_{\ell=-L}^{L} H(k,\ell)S(\ell) + V(k)$$

- $(S(k))_{k\in\mathbb{Z}}$: independent $\mathcal{CN}(0, I_T)$ input process,
- $(V(k))_{k\in\mathbb{Z}}$: independent $\mathcal{CN}(0, I_N)$ noise process,
- $(\mathbf{H}(k) = [H(k, k-L), \ldots, H(k, k+L)])_{k\in\mathbb{Z}}$: Gaussian $\mathbb{C}^{N\times(2L+1)T}$-valued **ergodic, generally non-centered** process representing the MIMO channel.

The three processes are independent.

Shannon's mutual information

Assuming the channel known at receiver,

$$I(S;(Y,H)) = \limsup_n \frac{1}{2n+1} \mathbb{E} \log\det(H^n H^{n*} + I_{(2n+1)N})$$

where

$$H^n = \begin{bmatrix} H(-n,-n-L) & \cdots & H(-n,-n+L) & & 0 \\ & \ddots & & \ddots & \\ 0 & & H(n,n-L) & \cdots & H(n,n+L) \end{bmatrix}.$$

Purpose: Behavior of $I(S;(Y,H))$ w.r.t. parameters of the channel statistical model.

Channel representation as an ergodic operator

Assume $N = T = 1$ for simplicity. On the Hilbert space $l^2(\mathbb{Z})$, let $H(\omega)$ be the random unbounded operator represented by the doubly infinite matrix

$$H(\omega) = \begin{bmatrix} \ddots & \ddots & \ddots & \ddots & & & 0 \\ \ddots & H(-1,-2) & H(-1,-1) & H(-1,1) & \ddots & & \\ & \ddots & H(0,-1) & H(0,0) & H(0,1) & \ddots & \\ & & \ddots & H(1,0) & H(1,1) & H(1,2) & \ddots \\ 0 & & & \ddots & \ddots & \ddots & \ddots \end{bmatrix}$$

$$= \left[H(k,\ell), k,\ell \in \mathbb{Z}, |k-\ell| \leq L \right].$$

Channel representation as an ergodic operator

Write

$$\mathbf{H}(\omega) = \Big(\mathbf{H}(\omega, k) = [H(k, k-L), \ldots, H(k, k+L)]\Big)_{k\in\mathbb{Z}}.$$

- By assumption, the shift $B : \Omega \to \Omega$ characterized by the equation $\mathbf{H}(B\omega, k) = \mathbf{H}(\omega, k+1)$ is **ergodic**.
- Operator $H(\omega)$ clearly satisfies the equation

$$H(B\omega) = UH(\omega)U^{-1}$$

  where U is the (unitary) shift operator $Ua = \sum_k \alpha_{k+1} e_k$ for $a = \sum_k \alpha_k e_k$, and e_k is the k^{th} canonical basis vector.
- Such an operator is **ergodic**, see Pastur and Figotin's book.

Ergodicity of HH^*

- The self-adjoint operator HH^* exists and is also **ergodic**, since

$$[HH^*](B\omega) = U[HH^*](\omega)U^{-1}.$$

- For any $z \in \mathbb{C}_+ = \{z \in \mathbb{C} : \Im z > 0\}$, let

$$Q(z) = (HH^* - z)^{-1} = [Q(z)(k, \ell)]_{k,\ell \in \mathbb{Z}}$$

  be the **resolvent** of HH^*. The resolution of identity of HH^*, hence $Q(z)$, are also ergodic. It results that

$$\frac{1}{2n+1} \sum_{i=-n}^{n} Q(z)(i,i) \xrightarrow[n\to\infty]{\text{a.s.}} \mathbb{E}Q(z)(0,0)$$

- More generally, $\mathbb{E}Q(z)(k,\ell)$ depends on $k - \ell$ only.
  $\Rightarrow$ $\mathbb{E}Q(z)$ is a bounded **Laurent operator**.
  We redenote $\mathbb{E}Q(z)(k,\ell)$ as $\mathbb{E}Q(z)(k-\ell)$.

The Integrated Density of States

- The operator HH^* has an **Integrated Density of States (IDS)**: there exists a deterministic probability measure μ such that

$$\frac{1}{2n+1} \operatorname{tr} g(H^n H^{n*}) \xrightarrow[n\to\infty]{\text{a.s.}} \int g(\lambda)\mu(d\lambda)$$

  for all continuous and bounded functions g.
- Taking $g(\lambda) = (\lambda - z)^{-1}$, we also have

$$\int_0^\infty \frac{1}{\lambda - z}\mu(d\lambda) = \mathbb{E}Q(z)(0).$$

  This is the **Stieltjes Transform (ST)** of μ.

IDS and mutual information

Theorem 1: The sequence $(2n+1)^{-1} \log\det(H^n H^{n*} + I_{2n+1})$ converges a.s. and in expectation as $n \to \infty$, and the limit is the mutual information

$$I(S;(Y,H)) = \int_0^\infty \log(1+\lambda)\,\mu(d\lambda) < \infty.$$

Mutual information approximation

- Theorem 1 characterizes the mutual information. However, **the integral is not very informative** in general. Some sort of **asymptotic regime** is needed.
- All shown properties of HH^* still hold true when N **or** T **is** > 1. In particular, IDS exists, and writing

$$Q(z) = (HH^* - z)^{-1} = [Q(z)(k,\ell)]_{k,\ell \in \mathbb{Z}}$$

  where the blocks $Q(z)(k,\ell)$ are $N \times N$ matrices, $\mathbb{E}Q(z) = [\mathbb{E}Q(z)(k-\ell)]_{k,\ell\in\mathbb{Z}}$ is **block-Laurent**, and the ST of μ is now $m_\mu(z) = \dfrac{\operatorname{tr} \mathbb{E}Q(z)(0)}{N}$.
- Theorem 1 becomes

$$\frac{1}{N} I(S;(Y,H)) = \int_0^\infty \log(1+\lambda)\,\mu(d\lambda) < \infty.$$

Mutual information approximation

- In order to make the integral w.r.t. μ more informative given the parameters of the channel statistical model, we make

$$T \to \infty, \quad 0 < \liminf N/T \leq \sup N/T < \infty.$$

and we study the asymptotic behavior of $m_\mu(z)$ in this regime, along the lines of Khorunzhyi-Pastur'93.

- Specifically, we find a sequence $m_{\boldsymbol{\pi}_T}(z)$ of ST of probability measures $\boldsymbol{\pi}_T$ such that

$$m_{\boldsymbol{\pi}_T}(z) - \frac{\operatorname{tr} \mathbb{E} Q(z)(0)}{N} \xrightarrow[T \to \infty]{} 0$$

for $\Im z > 0$. We then deduce an expression for $\int \log(1+\lambda)\boldsymbol{\pi}_T(d\lambda)$.

- **Other asymptotic regimes are possible**.

Preliminaries

Ergodic operators: a quick overview

Model, assumptions and the results

Main steps of the proof

The channel model

Recall that $Y(k) = \sum_\ell H(k,\ell)S(\ell) + V(k)$.

We assume $H(k,\ell) = A(k-\ell) + X(k,\ell) \in \mathbb{C}^{N\times T}$ where

- $(A(-L),\ldots,A(L))$ is deterministic (**frequency selective specular part**),
- $X(k,\ell) = \dfrac{1}{\sqrt{T}}\phi(k-\ell)W(k,\ell)$, where $\phi : \{-L,\ldots,L\} \to \mathbb{R}$ is a function whose square is the **multipath variance profile**,
- $(W(k,\ell) = [W_{n,t}(k,\ell), n = 0:(N-1), t = 0:(T-1)])_{k,\ell\in\mathbb{Z}}$ is a complex Gaussian proper centered random field such that

$$\mathbb{E}[W_{n_1,t_1}(k_1,\ell_1)\bar{W}_{n_2,t_2}(k_2,\ell_2)] = \mathbb{1}_{n_1=n_2}\mathbb{1}_{t_1=t_2}\mathbb{1}_{k_1-\ell_1=k_2-\ell_2}\gamma(k_1-k_2)$$

  where $\gamma(k)$ is a summable covariance function modeling the **Doppler effect**.

The channel model

$$H = A + X$$

$$= \underbrace{\begin{bmatrix} \ddots & A(1) & A(0) & A(-1) & \ddots & & 0 \\ & \ddots & A(1) & A(0) & A(-1) & \ddots & \\ 0 & & \ddots & A(1) & A(0) & A(-1) & \ddots \end{bmatrix}}_{\text{Constant block-Laurent banded operator}}$$

$$+ \underbrace{\begin{bmatrix} \ddots & X(-1,-2) & X(-1,-1) & X(-1,1) & \ddots & & 0 \\ & \ddots & X(0,-1) & X(0,0) & X(0,1) & \ddots & \\ 0 & & \ddots & X(1,0) & X(1,1) & X(1,2) & \ddots \end{bmatrix}}$$

- Elements of each $X(k,\ell)$ are iid,
- Mutually independent block diagonals,
- Variance of an element depends on the b.-diagonal,
- Time correlations on every b.-diagonal.

Since A is block-Laurent and γ is summable, H is ergodic.

Assumptions

We add the index T when necessary to stress dependency on T.
We assume

1. $\sigma_T^2 = \sum_\ell \phi_T(\ell)^2$ satisfies $\sup_T \sigma_T^2 < \infty$.
2. $\sup_T \sum_\ell |\gamma_T(\ell)| < \infty$.
3. $\sup_T \sum_\ell \|A_T(\ell)\| < \infty$ where $\| \cdot \|$ is the spectral norm.

Comments:

- Received power due to random part of the channel is σ_T^2.
- Assumption 2 means that the **coherence time** of the channel does not grow with T. The channel will become harder and harder to estimate !
- Assumption 3 can be lightened.

Results : characterization of $\boldsymbol{\pi}_T$

Fourier transforms :

$$\gamma_T(f) = \sum_k \exp(2\imath\pi kf)\gamma_T(k), \quad \mathbf{A}_T(f) = \sum_k \exp(2\imath\pi kf)A_T(k)$$

Write $(\mathbf{A}_T\mathbf{A}_T^*)(f) = \mathbf{A}_T(f)\mathbf{A}_T^*(f)$ and $(\mathbf{A}_T^*\mathbf{A}_T)(f) = \mathbf{A}_T^*(f)\mathbf{A}_T(f)$.
Consider the system of equations

$$\boldsymbol{\varphi}_T(f,z) = \frac{\operatorname{tr}\mathbf{S}_T(f,z)}{T} \quad \text{and} \quad \tilde{\boldsymbol{\varphi}}_T(f,z) = \frac{\operatorname{tr}\widetilde{\mathbf{S}}_T(f,z)}{T}$$

where $\mathbf{S}_T(f,z)$ and $\widetilde{\mathbf{S}}_T(f,z)$ are the $N \times N$ and $T \times T$ matrices

$$\begin{aligned}\mathbf{S}_T(f,z) &= \Big[-z\big(1+\sigma_T^2\gamma_T(f)\star\tilde{\boldsymbol{\varphi}}_T(f,z)\big) \\ &\qquad +\big(1+\sigma_T^2\gamma_T(-f)\star\boldsymbol{\varphi}_T(f,z)\big)^{-1}(\mathbf{A}_T\mathbf{A}_T^*)(f)\Big]^{-1}, \\ \widetilde{\mathbf{S}}_T(f,z) &= \Big[-z\big(1+\sigma_T^2\gamma_T(-f)\star\boldsymbol{\varphi}_T(f,z)\big) \\ &\qquad +\big(1+\sigma_T^2\gamma_T(f)\star\tilde{\boldsymbol{\varphi}}_T(f,z)\big)^{-1}(\mathbf{A}_T^*\mathbf{A}_T)(f)\Big]^{-1},\end{aligned}$$

Results : characterization of $\boldsymbol{\pi}_T$

$$\gamma_T(f) \star \tilde{\varphi}_T(f,z) = \int_0^1 \gamma_T(f-u)\tilde{\varphi}_T(u,z)\,du, \text{ and}$$

$$\gamma_T(-f) \star \varphi_T(f,z) = \int_0^1 \gamma_T(u-f)\varphi_T(u,z)\,du,$$

Theorem 2: For any $z \in \mathbb{C}_+$, this system admits a unique solution $(\varphi_T(\cdot,z), \tilde{\varphi}_T(\cdot,z))$ such that

$$\varphi_T(\cdot,z), \tilde{\varphi}_T(\cdot,z) : [0,1] \to \mathbb{C}$$

are both measurable and Lebesgue-integrable on $[0,1]$ and such that $\Im\varphi(f,z)$, $\Im\tilde{\varphi}(f,z)$, $\Im(z\varphi(f,z))$ and $\Im(z\tilde{\varphi}(f,z))$ are nonnegative for any $f \in [0,1]$.

The complex function $N^{-1}\int_0^1 \operatorname{tr}\mathbf{S}(f,z)\,df$ is the ST of a probability measure $\boldsymbol{\pi}_T$ carried by $[0,\infty)$.

Large-T approximation of the IDS

Theorem 3: For any $z \in \mathbb{C}_+$,

$$\int \frac{1}{\lambda - z} \mu_T(d\lambda) \; - \; \int \frac{1}{\lambda - z} \boldsymbol{\pi}_T(d\lambda) \xrightarrow[T\to\infty]{} 0.$$

Moreover, the sequences μ_T and $\boldsymbol{\pi}_T$ are tight, and

$$\int g(\lambda)\, \mu_T(d\lambda) \; - \; \int g(\lambda)\, \boldsymbol{\pi}_T(d\lambda) \xrightarrow[T\to\infty]{} 0$$

for any continuous and bounded real function g.

Mutual information approximation

Theorem 4: It holds that

$$N^{-1} I_T(S;(Y,H)) - \mathcal{I}_T \xrightarrow[T\to\infty]{} 0$$

where $\mathcal{I}_T = \int \log(1+\lambda)\, \boldsymbol{\pi}_T(d\lambda)$. This integral is given by

$$\mathcal{I}_T = \frac{1}{N} \int_0^1 \log\det\Big(1 + \sigma_T^2 \gamma_T(f) \star \tilde{\boldsymbol{\varphi}}_T(f,-1) + \frac{(\mathbf{A}_T \mathbf{A}_T^*)(f)}{1 + \sigma_T^2 \gamma_T(-f) \star \boldsymbol{\varphi}_T(f,-1)}\Big)\, df$$
$$+ \frac{T}{N} \int_0^1 \log(1 + \sigma_T^2 \gamma_T(-f) \star \boldsymbol{\varphi}_T(f,-1))\, df$$
$$- \frac{T}{N} \int_0^1 \int_0^1 \sigma_T^2 \gamma_T(f-v) \tilde{\boldsymbol{\varphi}}_T(v,-1) \boldsymbol{\varphi}_T(f,-1)\, dv\, df.$$

Some particular cases

It turns out that

- The form of the variance profile $\phi_T(k)^2$ has no influence on $\mathcal{I}_T$.
- If the channel is centered ($A_T = 0$), then $\boldsymbol{\pi}_T$ is a Marchenko-Pastur distribution.
- In the limit of small coherence times ($\gamma(f) \to 1$), the law $\boldsymbol{\pi}_T$ is the one obtained with a so-called "Information plus Noise" model.

Expectations of resolvents

Let

$$\begin{aligned} Q(z) &= (HH^* - z)^{-1} = [Q(z)(k,\ell)]_{k,\ell\in\mathbb{Z}} \quad (N \times N \text{ blocks}) \\ \widetilde{Q}(z) &= (H^*H - z)^{-1} = [\widetilde{Q}(z)(k,\ell)]_{k,\ell\in\mathbb{Z}} \quad (T \times T \text{ blocks}) \end{aligned}$$

Recall that the bounded operators $\mathbb{E}Q(z)$ and $\mathbb{E}\widetilde{Q}(z)$ are block-Laurent.

$$\mathbb{E}Q(z) = [\mathbb{E}Q(z)(k-\ell)(z)]_{k,\ell\in\mathbb{Z}}, \quad \mathbb{E}\widetilde{Q}(z) = [\mathbb{E}\widetilde{Q}(z)(k-\ell)]_{k,\ell\in\mathbb{Z}}$$

We want to approximate $\operatorname{tr} \mathbb{E}Q_T(z)(0)/N$, the ST of μ_T.

Tools

The basic tools are :

- **Stein's lemma**, *i.e.*, the integration by parts formula for $\mathbb{E}[X_i \Gamma(X)]$ where $X = (X_1, \ldots, X_M)$ is a Gaussian vector and $\Gamma : \mathbb{C}^M \to \mathbb{C}$ is a C^1 function.
- **Poincaré-Nash inequality** for bounding $\mathbb{V}\mathrm{ar}\, \Gamma(X)$.

A perturbed infinite system of equations

By adapting these tools to the infinite dimensional context, we get (omitting z)

$$\mathbb{E}Q(k) = -z^{-1}I(k) - \sigma^2 \sum_r \gamma(r)\Big[\frac{\operatorname{tr}\mathbb{E}\widetilde{Q}(r)}{T}\Big]\mathbb{E}Q(k-r) + z^{-1}\mathbb{E}[AH^*Q](k) + E(k),$$

$$\mathbb{E}[AH^*Q](k) = -\sigma^2 \sum_r \gamma(-r)\Big[\frac{\operatorname{tr}\mathbb{E}Q(r)}{T}\Big]\mathbb{E}[AH^*Q](k-r) + \mathbb{E}[AA^*Q](k) + E'(k)$$

$\mathbb{E}[AH^*Q](k)$ is the k^{th} diagonal block of the block-Laurent operator $\mathbb{E}[AH^*Q]$.
Elements of "perturbation" matrices $E(k)$ and $E'(K)$ are **bounded by Constant**$(z)/T$.
Similar equations for $\mathbb{E}\widetilde{Q}$.

Operators S and $\widetilde{S}$

Identifying the function $\mathbf{S}(\cdot, z)$ with a multiplication operator on the Hilbert space $\mathcal{L}^2([0,1] \to \mathbb{C}^N)$, and letting $\mathcal{F}$ be the operator who sends $\mathbf{g} \in \mathcal{L}^2([0,1] \to \mathbb{C}^N)$, to the sequence of its Fourier coefficients in $l^2(\mathbb{Z})$, the operator

$$S_T(z) = \mathcal{F}\mathbf{S}(\cdot, z)\mathcal{F}^*$$

is bounded and block-Laurent.

Blocks of $S_T(z) = [S_T(z)(k-\ell)]_{k,\ell\in\mathbb{Z}}$ and those of $\widetilde{S}_T(z) = [\widetilde{S}_T(z)(k-\ell)]_{k,\ell\in\mathbb{Z}}$ defined similarly satisfy a system of equations **similar to $\mathbb{E}Q$, $\mathbb{E}\widetilde{Q}$, but without perturbation**.

ST of $\boldsymbol{\pi}_T$ and mutual information approximation

We show that $\frac{\operatorname{tr} \mathbb{E}Q(z)(0)}{N} - \frac{\operatorname{tr} S(z)(0)}{N} \to 0$ for large T. Since

$$m_{\boldsymbol{\pi}_T}(z) = \frac{\operatorname{tr} S(z)(0)}{N} = \int_0^1 \frac{\mathbf{S}(f,z)}{N} df$$

we get the large-T approximation of the ST m_{μ_T} of the IDS μ_T (Theorem 3).

To pass from the ST to the mutual information, we use

$$\int \log(1+\lambda)\mu_T = \int_1^\infty \left(\frac{1}{t} - m_{\mu_T}(-t) \right) dt$$

We therefore need to find an antiderivative for $(t^{-1} - m_{\boldsymbol{\pi}_T}(-t))$.
Derivation done in HLN'07 $\Rightarrow$ Theorem 4.