

Débit théorique maximal de la communication MIMO

Jérôme Casse Thibaut Horel

Sujet proposé par Bartek Blaszczyzyn

21 juin 2009

Résumé

Cet exposé vise à calculer le débit théorique maximal d'une communication MIMO (Multiple Input Multiple Output). Plus précisément, on calcule la capacité (au sens de la théorie de l'information) d'un canal gaussien multi-antennes. Après quelques résultats préliminaires sur les vecteurs gaussiens réels et complexes et sur la théorie de l'information, on calcule effectivement la capacité du canal. La dernière partie vise à effectuer (de façon moins formelle), une étude asymptotique du résultat en lien avec la théorie de la distribution des valeurs propres des matrices aléatoires.

Table des matières

1	Introduction	3
2	Vecteurs gaussiens réels et complexes	4
2.1	Vecteurs gaussiens réels	4
2.2	Vecteurs gaussiens complexes	6
3	Les notions de théorie de l'information	8
3.1	Entropie et information mutuelle	9
3.2	Concavité de $\log \det$	13
3.3	Canal	14
4	Calcul de la capacité maximal théorique du canal MIMO	16
4.1	Propriété de la distribution de Rayleigh indépendante sur les matrices	16
4.2	Maximisation de $I(\mathbf{x}; \mathbf{y} \mathbf{H} = H)$ avec la contrainte $\mathbb{E}[\mathbf{x}\mathbf{x}^*] = Q$. .	18
4.3	Maximisation de $\Psi(Q)$ avec $\text{tr}(Q) \leq P$	19
5	Calcul asymptotique de la capacité du canal	22
5.1	Quand r est fixé et $t \rightarrow \infty$	22
5.2	Capacité MIMO et mesure spectrale	22
5.3	Quand $r, t \rightarrow \infty$ et $r/t \rightarrow \beta$	23
5.4	Étude numérique	25

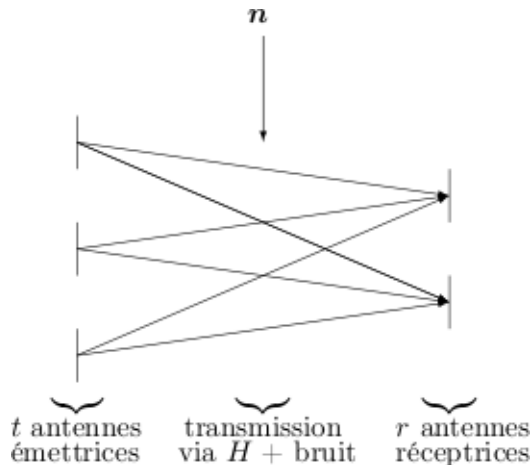
1 Introduction

La technologie MIMO consiste à augmenter le débit des transmissions sans-fil à puissance émise donnée en multipliant le nombre d'antennes tant au niveau de l'émetteur que du récepteur. On notera par la suite t le nombre d'antennes émettrices et r le nombre d'antennes réceptrices.

On considère que les antennes transmettent et reçoivent des nombres complexes (typiquement l'amplitude et la polarisation de l'onde émise). Plus précisément toutes les données sont supposées être des variables aléatoires complexes (on ne connaît pas a priori le message transmis). Le canal modélise la façon dont le signal émis est modifié avant d'être reçu. Le modèle du canal gaussien s'appuie sur les hypothèses suivantes :

- La sortie du canal (c'est à dire le signal reçu) est une fonction linéaire de l'entrée à laquelle on ajoute un bruit (par exemple le bruit d'agitation thermique) que l'on modélise par un vecteur gaussien complexe.
- La fonction linéaire ci-dessus modélise ce que reçoit chaque antenne : c'est la somme des signaux émis modifiés en amplitude suite aux réflexions des ondes dans l'environnement. Du fait des variations de l'environnement, cette fonction varie également au cours du temps et est supposée connue du récepteur.

On cherche donc quelle est la quantité maximale d'information que le canal est capable de transmettre par unité de temps. On définit pour cela la notion d'information mutuelle : elle représente la connaissance du signal émis apportée par l'observation du signal reçu. Il reste alors à rechercher le maximum de l'information mutuelle sur toutes les distributions possibles de l'entrée. Cette maximisation est effectuée sous une contrainte portant sur la puissance moyenne du signal émis.



2 Vecteurs gaussiens réels et complexes

Nous rappelons ici les résultats principaux sur les vecteurs gaussiens réels pour ensuite généraliser au cas complexe qui nous sera utile par la suite.

Si A est une matrice réelle ou complexe, on notera A^* la transconjugué de A . La matrice de covariance d'un vecteur aléatoire X sera notée K_X . Dans le cas d'un vecteur centré, on a $K_X = E[XX^*]$.

2.1 Vecteurs gaussiens réels

Définition 2.1. Soit $X = (X_1, \dots, X_d)$ une variable aléatoire (va) à valeurs dans \mathbb{R}^d . On dit que X est un *vecteur gaussien centré* si et seulement si pour toute forme linéaire $\varphi : \mathbb{R}^d \rightarrow \mathbb{R}$, $\varphi(X)$ est une variable aléatoire *réelle* gaussienne centrée.

Remarques. On sait que si X et X' sont deux variables aléatoires gaussiennes centrées et indépendantes, alors la somme $X + X'$ est une variable aléatoire gaussienne centrée (on peut le voir en regardant les fonctions caractéristiques). Par conséquent toute combinaison linéaire de vecteurs gaussiens indépendants est encore un vecteur gaussien.

Si $H \in \mathcal{L}(\mathbb{R}^d, \mathbb{R}^r)$, alors HX est également un vecteur gaussien centré. En effet, si φ est une forme linéaire, $\varphi \circ H$ l'est également.

Proposition 2.2. Soit $X = (X_1, \dots, X_d)$ un vecteur gaussien centré de matrice de covariance K alors la fonction caractéristique de X est donnée par :

$$\Phi_X(\xi) = E \left[e^{i\xi \cdot X} \right] = \exp \left(-\frac{1}{2} \xi^* K \xi \right), \quad \xi \in \mathbb{R}^d$$

Réciproquement, si $X = (X_1, \dots, X_d)$ est une variable aléatoire à valeurs dans \mathbb{R}^d , de même fonction caractéristique que ci-dessus, alors X est un vecteur gaussien centré de matrice de covariance K .

On notera $X \sim \mathcal{N}(0, K)$ pour dire que X est distribué selon la loi d'un vecteur gaussien centré de matrice de covariance K .

Démonstration. On a $\Phi_X(\xi) = \Phi_{\xi \cdot X}(1)$. Or par définition $Y = \xi \cdot X$ est une variable aléatoire gaussienne centrée. Et :

$$\begin{aligned} \text{Var}(Y) &= E \left[(\xi \cdot X)^2 \right] = E \left[(\xi^* X)^2 \right] = E \left[(\xi^* X) (X^* \xi) \right] \\ &= \xi^* E[XX^*] \xi = \xi^* K \xi \end{aligned}$$

Donc :

$$\Phi_Y(t) = \exp \left(-\frac{t^2}{2} \xi^* K \xi \right)$$

En prenant $t = 1$ dans l'égalité précédente, on obtient la première partie de la proposition.

Pour la réciproque, soit $X = (X_1, \dots, X_d)$ une variable aléatoire de fonction caractéristique :

$$\Phi_X(\xi) = \exp\left(-\frac{1}{2}\xi^* K \xi\right), \quad \xi \in \mathbb{R}^d$$

et soit A un vecteur (déterministe) de \mathbb{R}^d . Il suffit de montrer que la fonction caractéristique de $A \cdot X = A^* X$ est celle d'une variable aléatoire réelle gaussienne centrée. Soit $\xi \in \mathbb{R}$:

$$\begin{aligned} \Phi_{A^* X}(\xi) &= \mathbb{E}[\exp(i\xi A^* X)] = \mathbb{E}[\exp(i(\xi A)^* X)] \\ &= \exp\left(-\frac{1}{2}(\xi A)^* K (\xi A)\right) \\ &= \exp\left(-\frac{\xi^2}{2} A^* K A\right) \end{aligned}$$

Donc $A^* X$ est une variable aléatoire gaussienne centrée de variance $A^* K A$. Ceci étant vrai pour tout vecteur A , on en déduit que X est un vecteur gaussien centré de matrice de covariance K . \square

Remarques. – Si X_1, \dots, X_n sont n variables aléatoires gaussiennes indépendantes, alors (X_1, \dots, X_d) est un vecteur gaussien centré de matrice de covariance $K_X = \text{diag}(\text{Var}(X_1), \dots, \text{Var}(X_d))$.

– Réciproquement, si X est un vecteur gaussien centré de matrice de covariance $K_X = \text{diag}(\sigma_1^2, \dots, \sigma_d^2)$, la fonction caractéristique de X vérifie :

$$\Phi_X(\xi) = \exp\left(-\frac{1}{2} \sum_{k=1}^d \sigma_k^2 \xi_k^2\right) = \prod_{k=1}^d \Phi_{X_k}(\xi_k)$$

Donc les composantes de X sont indépendantes. On vient de montrer que *si X est un vecteur gaussien centré, ses composantes sont indépendantes si et seulement si elles sont non corrélées.*

– Ce dernier exemple est l'exemple canonique de vecteur gaussien. En effet, si X est un vecteur gaussien de matrice de covariance K_X , on peut diagonaliser K_X et écrire $K_X = P D P^*$ où P est une matrice orthogonale réelle. Le vecteur $P X$ est alors un vecteur gaussien de matrice de covariance D .

Proposition 2.3. *Soit X un vecteur gaussien centré de matrice de covariance K_X . Si K_X est non dégénérée, alors X a une densité par rapport à la mesure de Lebesgue sur \mathbb{R}^d égale à :*

$$p_X(x) = \frac{1}{(\sqrt{2\pi})^d \sqrt{\det K_X}} \exp\left(-\frac{1}{2} x^* K_X^{-1} x\right)$$

Démonstration. On commence par le cas où $K_X = I_d$ où I_d est la matrice identité sur \mathbb{R}^d . Dans ce cas les composantes de X sont d variables aléatoires gaussiennes

centrées de variance 1, donc la loi de X admet une densité par rapport à la mesure de Lebesgue sur \mathbb{R}^d donnée par :

$$p_g(x) = \prod_{i=1}^d p_{X_i}(x_i) = \frac{1}{(\sqrt{2\pi})^d} \exp\left(-\frac{1}{2}x^*x\right)$$

Dans le cas général, on peut écrire K_X est symétrique définie positive donc admet une racine carrée symétrique définie positive, on peut donc écrire $K_X = AA^*$ avec A une matrice diagonale inversible. Si on pose Y le vecteur gaussien dont toutes les composantes sont des variables aléatoires gaussiennes centrées indépendantes et de variance 1, alors AY a la même loi que X . On va donc étudier la loi de AY . Si f est une fonction mesurable positive définie sur \mathbb{R}^d , on a :

$$E[f(AY)] = \int_{\mathbb{R}^d} f(Ay)p_g(y)dy$$

En faisant le changement de variable $u = Ay$ dans l'intégrale précédente on obtient :

$$E[f(AY)] = \int_{\mathbb{R}^d} f(y)p_Y(A^{-1}u) \frac{1}{|\det A|} p_g(A^{-1}u)du$$

Mais on remarque que $|\det A| = \sqrt{\det K_X}$ et que $p_g(A^{-1}u) = p_X(u)$, d'où le résultat. \square

2.2 Vecteurs gaussiens complexes

Définition 2.4. Si $z \in \mathbb{C}^n$ et $A \in \mathcal{M}_{n \times m}(\mathbb{C})$, on définit respectivement $\widehat{z} \in \mathbb{R}^{2n}$ et $\widehat{A} \in \mathcal{M}_{2n \times 2m}(\mathbb{R})$ par :

$$\widehat{z} = \begin{pmatrix} \Re(z) \\ \Im(z) \end{pmatrix} \quad \text{et} \quad \widehat{A} = \begin{pmatrix} \Re(A) & -\Im(A) \\ \Im(A) & \Re(A) \end{pmatrix}$$

Proposition 2.5. 1. $z \mapsto \widehat{z}$ est un isomorphisme de groupes de \mathbb{C}^n dans \mathbb{R}^{2n} .

2. $\Re(x^*y) = \widehat{x}^*\widehat{y}$.

3. $A \mapsto \widehat{A}$ est un morphisme injectif d'algèbres de $\mathcal{M}_{n \times m}(\mathbb{C})$ dans $\mathcal{M}_{2n \times 2m}(\mathbb{R})$.

4. $\widehat{A^*} = \widehat{A}^*$.

5. $\det \widehat{A} = \det AA^* = |\det A|^2$

6. $\widehat{Ax} = \widehat{A}\widehat{x}$.

Démonstration. Tout est évident sauf le point 5. Pour le montrer, on remarque que :

$$\det(\widehat{A}) = \det \left[\begin{pmatrix} I_d & iI_d \\ 0 & I_d \end{pmatrix} \widehat{A} \begin{pmatrix} I_d & -iI_d \\ 0 & I_d \end{pmatrix} \right] = \det \begin{pmatrix} A & 0 \\ \Im(A) & \Re(A) \end{pmatrix} = |\det A|^2 \quad \square$$

Définition 2.6. Soit X une variable aléatoire à valeurs dans \mathbb{C}^n , on dit que X est un *vecteur gaussien* (complexe) si et seulement si \widehat{X} est un vecteur gaussien réel.

On sait qu'un vecteur gaussien réel est entièrement déterminé par sa matrice de covariance et son espérance. Ceci n'est pas vrai pour un vecteur gaussien complexe. En effet la connaissance de K_X ne donne pas celle de $K_{\widehat{X}}$. Ceci justifie la définition suivante :

Définition 2.7. On dit que X est un *vecteur gaussien à symétrie circulaire* si et seulement si $K_{\widehat{X}}$ la matrice de covariance de \widehat{X} est de la forme :

$$K_{\widehat{X}} = \frac{1}{2}\widehat{Q}$$

pour une certaine matrice $Q \in \mathcal{M}_n(\mathbb{C})$ hermitienne positive. Dans ce cas, on a $K_X = Q = E[(X - E[X])(X - E[X])^*]$.

Un vecteur gaussien complexe à symétrie circulaire est donc totalement déterminé par son espérance et sa matrice de covariance. On notera $X \sim \mathcal{N}(E[X], K)$ quand le vecteur aléatoire X suivra la loi d'un vecteur gaussien complexe à symétrie circulaire de moyenne $E[X]$ et de matrice de covariance K .

Proposition 2.8. 1. Si $X \in \mathbb{C}^n$ est un vecteur gaussien à symétrie circulaire et $A \in \mathcal{M}_{m \times n}$, alors $Y = AX$ également.

2. Si X et Y sont deux vecteurs gaussiens complexes indépendants à symétrie circulaire, alors $X + Y$ également.

Démonstration. 1. Par propriétés de morphisme de $z \mapsto \widehat{z}$, $\widehat{Y} = \widehat{A}\widehat{X}$ donc Y est un vecteur gaussien complexe. Soit $Q = E[XX^*]$, alors :

$$E[\widehat{y}\widehat{y}^*] = \widehat{A}E[\widehat{x}\widehat{x}^*]\widehat{A}^* = \frac{1}{2}\widehat{A}\widehat{Q}\widehat{A}^* = \frac{1}{2}\widehat{K}$$

2. De même, $X + Y$ est un vecteur gaussien complexe, on vérifie la symétrie circulaire. Soit $A = E[XX^*]$ et $B = E[YY^*]$, alors :

$$E[\widehat{X+Y}\widehat{X+Y}^*] = E[\widehat{X}\widehat{X}^*] + E[\widehat{Y}\widehat{Y}^*] = \frac{1}{2}(\widehat{A} + \widehat{B}) = \frac{1}{2}(\widehat{A+B}) \quad \square$$

Proposition 2.9. Soit $X \in \mathbb{C}^d$ un vecteur gaussien à symétrie circulaire. Si $K_X \in \mathcal{M}_d(\mathbb{C})$ est hermitienne définie positive, alors $K_{\widehat{X}} \in \mathcal{M}_{2d}(\mathbb{R})$ est symétrique définie positive. Dans ce cas la loi de X admet une densité par rapport à la mesure de Lebesgue sur \mathbb{C}^d donnée par :

$$p(x) = \frac{1}{\det(\pi K_X)} \exp(-x^* K_X^{-1} x) = \frac{1}{\sqrt{\det(\pi \widehat{K}_X)}} \exp(-\widehat{x}^* \widehat{K}_X^{-1} \widehat{x})$$

Démonstration. Soit \hat{x} un vecteur dans \mathbb{R}^{2n} . On écrit :

$$\begin{aligned}\hat{x}^* K_{\hat{X}} \hat{x} &= \frac{1}{2} \hat{x}^* \widehat{K_X} \hat{x} \\ &= \frac{1}{2} \hat{x}^* \widehat{K_X} x \\ &= \frac{1}{2} \Re(x^* K_X x) \\ &= \frac{1}{2} x^* K_X x \geq 0\end{aligned}$$

Où on a utilisé les propriétés du morphisme $z \mapsto \hat{z}$, la dernière égalité provenant de la positivité de K_X . Enfin on a égalité si et seulement si on a égalité dans la dernière inégalité, i.e. (par positivité de K_X) si et seulement si $x = 0$ et donc si et seulement si $\hat{x} = 0$.

Pour obtenir la densité de la loi de X . Soit f une fonction mesurable positive définie sur \mathbb{C}^d . On définit \hat{f} sur \mathbb{R}^{2d} par :

$$\hat{f}(a_1, \dots, a_d, b_1, \dots, b_d) = f(a_1 + ib_1, \dots, a_d + ib_d)$$

De sorte que si $\hat{x} \in \mathbb{C}^d$, on a $f(x) = \hat{f}(\hat{x})$. On peut donc écrire :

$$\mathbb{E}[f(X)] = \mathbb{E}[\hat{f}(\hat{X})] = \int_{\mathbb{R}^{2d}} \hat{f}(\hat{x}) \mathbb{P}_{\hat{X}}(d\hat{x})$$

On s'est donc ramené au cas réel, on sait que $\mathbb{P}_{\hat{X}}$ a une densité par rapport à la mesure de Lebesgue. Donc on peut écrire :

$$\begin{aligned}\mathbb{E}[f(X)] &= \int_{\mathbb{R}^{2d}} \hat{f}(\hat{x}) \frac{1}{(\sqrt{2\pi})^{2d} \sqrt{\det K_{\hat{X}}}} \exp\left(-\frac{1}{2} \hat{x}^* K_{\hat{X}}^{-1} \hat{x}\right) d\hat{x} \\ &= \int_{\mathbb{R}^{2d}} \hat{f}(\hat{x}) \frac{1}{\pi^d \sqrt{\det \widehat{K_X}}} \exp\left(-\hat{x}^* \widehat{K_X}^{-1} \hat{x}\right) d\hat{x} \\ &= \int_{\mathbb{R}^{2d}} \hat{f}(\hat{x}) \frac{1}{\sqrt{\det(\pi \widehat{K_X})}} \exp\left(-\hat{x}^* \widehat{K_X}^{-1} \hat{x}\right) d\hat{x} \\ &= \int_{\mathbb{C}^d} f(x) \frac{1}{\sqrt{\det(\pi \widehat{K_X})}} \exp\left(-x^* K_X^{-1} x\right) dx \\ &= \int_{\mathbb{C}^d} f(x) \frac{1}{\det(\pi K_X)} \exp\left(-x^* K_X^{-1} x\right) dx\end{aligned}$$

D'où le résultat. □

3 Les notions de théorie de l'information

Nous allons, dans cette partie, définir les outils de théorie de l'information dont nous aurons besoin pour étudier la capacité du canal qui nous intéresse.

Définition 3.1. La mesure de Lebesgue sur \mathbb{C}^d est la mesure de Lebesgue sur \mathbb{R}^{2d} où \mathbb{C}^d est vu comme un \mathbb{R} -espace vectoriel. Si X et Y sont deux variables aléatoires respectivement dans \mathbb{C}^d et $\mathbb{C}^{d'}$, on notera :

- $f(x, y)$, la densité de la variable aléatoire (X, Y) par rapport à la mesure de Lebesgue sur $\mathbb{C}^d \times \mathbb{C}^{d'} \simeq \mathbb{C}^{d+d'}$.
- $f(x)$, la densité de la variable aléatoire X par rapport à la mesure de Lebesgue sur \mathbb{C}^d .
- $f(x|y)$, la densité de la variable X sachant Y , le rapport $\frac{f(x,y)}{f(y)}$.

3.1 Entropie et information mutuelle

Définition 3.2 (Entropie). Soit X une variable aléatoire à valeur dans \mathbb{C}^d de fonction de densité f par rapport à la mesure de Lebesgue sur \mathbb{C}^d . On définit l'entropie $h(X)$ de la variable aléatoire X par :

$$h(X) = - \int_{\mathbb{C}^d} f(x) \log(f(x)) dx = \mathbb{E}[-\log(f(X))]$$

quand cette intégrale existe.

Cette valeur peut être interprétée comme la quantité d'information donnée par l'observation de X , ou encore comme l'incertitude de X .

Exemple 1 (Entropie d'un vecteur gaussien). On calcule l'entropie d'un vecteur gaussien $\mathbf{X} = (X_1, \dots, X_d)$ à symétrie circulaire de moyenne μ et de matrice de covariance K hermitienne définie positive. On utilise la formule de la densité de la proposition 2.9.

$$\begin{aligned} h(\mathbf{X}) &= -\mathbb{E} \left[(\mathbf{x}) [-(\mathbf{x} - \mu)^* K^{-1} (\mathbf{x} - \mu) - \log((\pi)^d |K|)] \right] \\ &= \mathbb{E} \left[\sum_{i,j} (X_i - \mu_i) (K^{-1})_{ij} (X_j - \mu_j) \right] + \log(\pi^d |K|) \\ &= \sum_{i,j} \mathbb{E} [(X_i - \mu_i) (X_j - \mu_j)] (K^{-1})_{ij} + \log(\pi^d |K|) \\ &= \sum_j \sum_i K_{ji} (K^{-1})_{ij} + \log(\pi^d |K|) \\ &= \sum_j (K K^{-1})_{jj} + \log(\pi^d |K|) \\ &= \sum_j I_{jj} + \log(\pi^d |K|) \\ &= d + \log \pi^d + \log |K| \\ &= \log(e^d) + \log \pi^d + \log |K| \\ &= \log |\pi e K| \end{aligned}$$

On a même mieux, en effet le lemme suivant nous dit que le vecteur gaussien à symétrie circulaire est le vecteur qui a la plus grande entropie à covariance fixée.

Théorème 3.3. *Soit X une variable aléatoire avec densité, de moyenne nulle et de matrice de covariance K hermitienne définie positive. Alors $h(X) \leq \log \det(\pi e K)$. Avec égalité si $X \sim \mathcal{N}(0, K)$.*

Démonstration. On note g , la fonction de densité de X . On va la comparer à $\phi(x) = \frac{1}{\det(\pi K)} e^{-x^* K^{-1} x}$. Pour cela, on va étudier la quantité $\int g \log g / \phi$. On va d'abord montrer qu'elle est positive, puis ensuite on va la calculer.

– Positivité (utilise Jensen).

$$\begin{aligned} - \int g(x) \log \frac{g(x)}{\phi(x)} dx &= \int g(x) \log \frac{\phi(x)}{g(x)} dx \\ &\leq \log \int \left(g(x) \frac{\phi(x)}{g(x)} \right) dx \\ &\leq \log \int \phi(x) dx \\ &= 0 \end{aligned}$$

– Calcul de $\int g(x) \log \frac{g(x)}{\phi(x)} dx$.

$$\begin{aligned} \int g(x) \log \frac{g(x)}{\phi(x)} dx &= \int g(x) \log g(x) dx - \int g(x) \log \phi(x) dx \\ &= -h(X) - \int g(x) \left(\log \frac{1}{\det(\pi K)} - x^* K^{-1} x \right) dx \\ &= -h(X) - \left(\log \frac{1}{\det(\pi K)} \right) \int g(x) dx + \int g(x) x^* K^{-1} x dx \\ &= -h(X) - \left(\log \frac{1}{\det(\pi K)} \right) + \int g(x) \sum_{i,j} x_i^* (K^{-1})_{ij} x_j dx \\ &= -h(X) - \left(\log \frac{1}{\det(\pi K)} \right) + \sum_{i,j} (K^{-1})_{ij} \int g(x) x_i^* x_j dx \\ &= -h(X) - \left(\log \frac{1}{\det(\pi K)} \right) + \sum_{i,j} (K^{-1})_{ij} K_{ji} dx \\ &= -h(X) - \left(\log \frac{1}{\det(\pi K)} \right) + \sum_i I_{ii} \\ &= -h(X) + \log \det(\pi K) + d \\ &= -h(X) + \log \det(\pi e K) \end{aligned}$$

Ainsi

$$h(X) \leq \log \det(\pi e K)$$

On a bien l'égalité pour une distribution gaussienne (voir exemple 1) □

Définition 3.4 (Entropie conditionnelle). Soit X et Y deux variables aléatoires à valeurs dans \mathbb{C}^d et $\mathbb{C}^{d'}$. Soit f la fonction de densité de la loi jointe. On définit l'entropie $h(X|Y)$ de la variable aléatoire X conditionnellement à la variable aléatoire Y par :

$$h(X|Y) = - \int_{\mathbb{C}^d \times \mathbb{C}^{d'}} f(x, y) \log(f(x|y)) dx = \mathbb{E}[-\log(f((X|Y)))]$$

Lemme 3.5. Soit X une variable aléatoire et Z une variable aléatoire. Alors

$$h(X + Z|X) = h(Z)$$

Démonstration.

$$\begin{aligned} h(X + Z|X) &= - \int f(x + z, x) \log f(x + z|x) dx dy \\ &= - \int f(x + z|x) f(x) \log f(z|x) dx dy \\ &= - \int f(z|x) f(x) \log f(z|x) dx dy \\ &= - \int f(z, x) \log f(z|x) dx dy \\ &= h(Z|X) \end{aligned}$$

□

Lemme 3.6. Soit $(X_1, \dots, X_d) \in E_1 \times \dots \times E_d$ vecteurs aléatoires. Alors

$$h(X_1, \dots, X_d) = \sum_{i=1}^d h(X_i|X_1, \dots, X_{i-1})$$

Démonstration.

$$\begin{aligned} h(X_1, \dots, X_d) &= - \int_{E_1} \dots \int_{E_d} f(x_1, \dots, x_d) \log(f(x_1, \dots, x_d)) \\ &= - \int_{E_1} \dots \int_{E_d} f(x_1, \dots, x_d) \log\left(\prod_{i=1}^d f(x_i|x_1, \dots, x_{i-1})\right) \\ &\quad (\text{car } f(x, y) = f(x)f(y|x)) \\ &= - \sum_{i=1}^d \int_{E_1} \dots \int_{E_d} f(x_1, \dots, x_d) \log(f(x_i|x_1, \dots, x_{i-1})) \end{aligned}$$

(on peut inverser somme et intégrale car tout est négatif)

$$\begin{aligned}
h(X_1, \dots, X_d) &= - \sum_{i=1}^d \int_{E_1} \dots \int_{E_i} \log(f(x_i|x_1, \dots, x_{i-1})) \int_{E_{i+1}} \dots \int_{E_d} f(x_1, \dots, x_d) \\
&= \sum_{i=1}^d - \int_{E_1} \dots \int_{E_i} f(x_1, \dots, x_i) \log(f(x_i|x_1, \dots, x_{i-1})) \\
&= \sum_{i=1}^d h(X_i|X_1, \dots, X_{i-1})
\end{aligned}$$

□

Lemme 3.7. Soit X et Y deux variables aléatoires dans \mathbb{C}^d et dans $\mathbb{C}^{d'}$. Alors

$$h(X|Y) \leq h(X)$$

Le cas d'égalité est obtenu quand X et Y sont indépendantes.

Démonstration. Si les deux intégrales précédentes existent, on a, alors

$$I(X; Y) = h(X) - h(X|Y)$$

Il reste à prouver que $I(X; Y) \geq 0$ pour tous X et Y .

$$\begin{aligned}
-I(X, Y) &= \int_{\mathbb{C}^d} \int_{\mathbb{C}^{d'}} f(x, y) \log \frac{f(x)f(y)}{f(x, y)} dx dy \\
&\leq \log \int_{\mathbb{C}^d} \int_{\mathbb{C}^{d'}} f(x, y) \frac{f(x)f(y)}{f(x, y)} dx dy && \text{(par Jensen)} \\
&= \log \int_{\mathbb{C}^d} \int_{\mathbb{C}^{d'}} f(x)f(y) dx dy \\
&= \log 1 \\
&= 0
\end{aligned}$$

Ainsi, $I(X; Y) \geq 0$ et par suite $h(X) \geq h(X|Y)$.

Le cas d'égalité, X et Y sont indépendantes, alors

$$\begin{aligned}
h(X|Y) &= - \int_{\mathbb{C}^d} \int_{\mathbb{C}^{d'}} f(x, y) \log f(x|y) dx dy \\
&= - \int_{\mathbb{C}^d} \int_{\mathbb{C}^{d'}} f(x, y) \log f(x) dx dy \\
&= - \int_{\mathbb{C}^d} f(x) \log f(x) dx \\
&= h(X)
\end{aligned}$$

□

Définition 3.8 (Information mutuelle). Soit X et Y deux variables aléatoires à valeur dans \mathbb{C}^d et $\mathbb{C}^{d'}$. Soit f la fonction de densité de la loi jointe. On définit l'information mutuelle $I(X; Y)$ entre X et Y par :

$$I(X; Y) = - \int_{\mathbb{C}^d} f(x, y) \log \frac{f(x, y)}{f(x)f(y)} dx = \mathbb{E}[-\log(f((X, Y)))] = h(Y) - h(Y|X)$$

Cette valeur peut être interprétée comme la quantité d'information donnée sur X par la connaissance de Y .

3.2 Concavité de log det

Proposition 3.9 (Concavité de log det). Soit K_0 et K_1 deux matrices hermitiennes définies positives de dim $n \times n$. Soit $\lambda \in [0, 1]$. On a, alors

$$\log \det(\lambda K_0 + (1 - \lambda)K_1) \geq \lambda \log \det K_0 + (1 - \lambda) \log \det K_1$$

Démonstration. Soit θ , une variable aléatoire suivant une loi de Bernoulli de paramètre λ (i.e. $P(\theta = 0) = 1 - \lambda$ et $P(\theta = 1) = \lambda$). Soient \mathbf{X}_i , le vecteur gaussien de loi $\mathcal{N}(0, K_i)$ pour $i \in 0, 1$ et dont la fonction de densité sera noté f_i . On définit le vecteur aléatoire $\mathbf{Z} = \theta \mathbf{X}_0 + (1 - \theta) \mathbf{X}_1$. On a, par le lemme 3.7, $h(\mathbf{Z}) \geq h(\mathbf{Z}|\theta)$. Il ne reste plus qu'à calculer ces deux valeurs et prendre le log.

– Calcul de $h(\mathbf{Z}|\theta)$.

$$\begin{aligned} h(\mathbf{Z}|\theta) &= - \sum_{i=0}^1 \int_{\mathbb{C}^n} f(\mathbf{x}, i) \log f(\mathbf{x}|i) d\mathbf{x} \\ &= - \sum_{i=0}^1 \int_{\mathbb{C}^n} f(\mathbf{x}|i) P(\theta = i) \log f(\mathbf{x}|i) d\mathbf{x} \\ &= - \sum_{i=0}^1 P(\theta = i) \int_{\mathbb{C}^n} f_i(\mathbf{x}) \log f_i(\mathbf{x}) d\mathbf{x} \\ &= P(\theta = 0) \int_{\mathbb{C}^n} -f_0(\mathbf{x}) \log f_0(\mathbf{x}) d\mathbf{x} \\ &\quad + P(\theta = 1) \int_{\mathbb{C}^n} -f_1(\mathbf{x}) \log f_1(\mathbf{x}) d\mathbf{x} \\ &= \lambda h(\mathbf{X}_0) + (1 - \lambda) h(\mathbf{X}_1) \\ &= \lambda \log \det(\pi e K_0) + (1 - \lambda) \log \det(\pi e K_1) \\ &= \lambda \log((\pi e)^d \det K_0) + (1 - \lambda) \log((\pi e)^d \det K_1) \\ &= \log(\pi e)^d + \lambda \log \det K_0 + (1 - \lambda) \log \det K_1 \end{aligned}$$

– Majoration de $h(\mathbf{Z})$.

Pour majorer $h(\mathbf{Z})$, on va utiliser le théorème 3.3. Ainsi, on calcule la covariance de \mathbf{Z} .

$$\begin{aligned}
\mathbb{E}[\mathbf{Z}\mathbf{Z}^*] &= \mathbb{E}[(\theta\mathbf{X}_0 + (1-\theta)\mathbf{X}_1)(\theta\mathbf{X}_0 + (1-\theta)\mathbf{X}_1)^*] \\
&= \mathbb{E}[\theta^2\mathbf{X}_0\mathbf{X}_0^* + (1-\theta)^2\mathbf{X}_1\mathbf{X}_1^* \\
&\quad + \theta(1-\theta)\mathbf{X}_0\mathbf{X}_1^* + \theta(1-\theta)\mathbf{X}_1\mathbf{X}_0^*] \\
&= \mathbb{E}[\theta^2\mathbf{X}_0\mathbf{X}_0^*] + \mathbb{E}[(1-\theta)^2\mathbf{X}_1\mathbf{X}_1^*] \\
&\quad + \mathbb{E}[\theta(1-\theta)\mathbf{X}_0\mathbf{X}_1^*] + \mathbb{E}[\theta(1-\theta)\mathbf{X}_1\mathbf{X}_0^*] \\
&= \mathbb{E}[\theta^2]\mathbb{E}[\mathbf{X}_0\mathbf{X}_0^*] + \mathbb{E}[(1-\theta)^2]\mathbb{E}[\mathbf{X}_1\mathbf{X}_1^*] \\
&\quad + \mathbb{E}[\theta(1-\theta)]\mathbb{E}[\mathbf{X}_0]\mathbb{E}[\mathbf{X}_1^*] + \mathbb{E}[\theta(1-\theta)]\mathbb{E}[\mathbf{X}_1]\mathbb{E}[\mathbf{X}_0^*] \\
&= \lambda K_0 + (1-\lambda)K_1 + 0 \\
&= \lambda K_0 + (1-\lambda)K_1
\end{aligned}$$

Ainsi, en appliquant le théorème 3.3, on obtient

$$\begin{aligned}
h(\mathbf{Z}) &\leq \log \det(\pi e(\lambda K_0 + (1-\lambda)K_1)) \\
&= \log(\pi e)^d + \log \det(\lambda K_0 + (1-\lambda)K_1)
\end{aligned}$$

Ainsi

$$\log(\pi e)^d + \lambda \log |K_0| + (1-\lambda) \log |K_1| \leq \log(\pi e)^d + \log |\lambda K_0 + (1-\lambda)K_1|$$

Ainsi, on obtient

$$\lambda \log \det K_0 + (1-\lambda) \log \det K_1 \leq \log \det(\lambda K_0 + (1-\lambda)K_1)$$

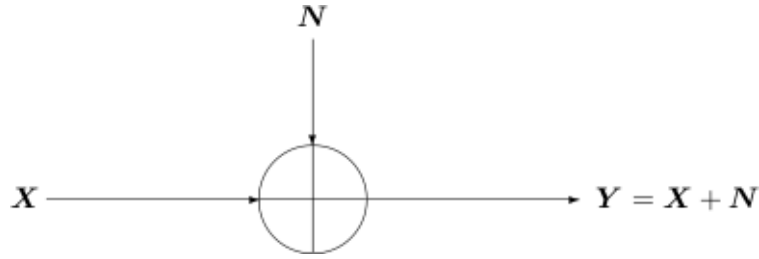
□

3.3 Canal

Définition 3.10 (Canal). Soient X et Y deux variables aléatoires. X est appelée entrée et Y , sortie. Le canal est la fonction qui donne la probabilité que l'on obtienne la réalisation y sachant la réalisation x , c'est à dire la fonction $f(Y|X)$.

Moins formellement, le canal est un moyen de transmettre des informations en entrée (X) et d'obtenir la sortie (Y), cette sortie pouvant subir des modifications du fait de la transmission.

Exemple 2 (Canal gaussien). Soit N une variable aléatoire complexe gaussienne à symétrie circulaire centrée réduite ($N \sim \mathcal{N}(0, 1)$), ceci représente le bruit. La sortie du canal gaussien associé à l'entrée X , variable aléatoire dans \mathbb{C} , est la variable aléatoire complexe Y qui vaut $Y = X + N$. Ce canal se représente ainsi



Nous avons pris en exemple, le canal gaussien complexe qui est un cas particulier du canal qui nous intéresse qui est le canal MIMO et que nous allons définir maintenant.

Définition 3.11 (Canal MIMO). Un canal MIMO est la donnée d'une matrice aléatoire \mathbf{H} et d'un vecteur aléatoire $\mathbf{n} \sim \mathcal{N}(0, I)$. La sortie obtenue (\mathbf{y}, \mathbf{H}) vaut pour l'entrée \mathbf{x}

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

Définition 3.12 (Capacité d'un canal). On définit la capacité C d'un canal de fonction de transition f sous la contrainte c ainsi :

$$C = \max_{(X;Y):f(Y|X)\wedge c} I(X;Y)$$

Exemple 3 (Capacité d'un canal gaussien). La contrainte prise est une contrainte de puissance, on a que $E[|X|^2] \leq P$ où P est un réel positif. On va chercher à calculer la capacité du canal gaussien sous cette contrainte. On a

$$\begin{aligned} I(X;Y) &= h(Y) - h(Y|X) \\ &= h(Y) - h(X + N|X) \\ &= h(Y) - h(N|X) && \text{(par le lemme 3.5)} \\ &= h(Y) - h(N) && \text{(par le lemme 3.7)} \\ &= h(Y) - \log(\pi e) \end{aligned}$$

Nous avons donc plus qu'à chercher un Y qui maximise $h(Y)$ et qui permet d'avoir un entrée réalisant la contrainte. On a

$$\begin{aligned} E[|Y|^2] &= E[(X + N)(X + N)^*] \\ &= E[|X|^2 + |N|^2 + X\bar{N} + N\bar{X}] \\ &= E[|X|^2] + E[|N|^2] + E[X]E[\bar{N}] + E[N]E[\bar{X}] \\ &\leq P + 1 + 0 + 0 \\ &\leq 1 + P \end{aligned}$$

Ainsi, par le théorème 3.3, on a que

$$h(Y) \leq \log(\pi e(1 + P))$$

Et donc

$$\begin{aligned} I(X; Y) &\leq \log(\pi e(P + 1)) - \log(\pi e) \\ &= \log(\pi e) + \log(P + 1) - \log(\pi e) \\ &= \log(1 + P) \end{aligned}$$

On remarque que cette valeur est atteinte pour $X \sim \mathcal{N}(0, P)$ qui vérifie bien sûr la contrainte $\mathbb{E}[|X|^2] \leq P$ car alors Y est somme de deux variables gaussiennes indépendantes de moyenne nulle donc une gaussienne de moyenne nulle et de variance la somme des deux, c'est à dire $P + 1$ ici.

4 Calcul de la capacité maximal théorique du canal MIMO

On va, dans cette partie, calculer la capacité du canal MIMO avec la matrice \mathbf{H} distribuée selon la distribution de Rayleigh indépendante sur la matrice.

Définition 4.1 (distribution de Rayleigh indépendante sur les matrices). On dit qu'une matrice \mathbf{H} est distribuée suivant une distribution de Rayleigh indépendante de paramètre σ^2 si les coefficients de \mathbf{H} sont indépendants, de moyenne nulle avec une partie réelle et imaginaire indépendantes de variance $\frac{1}{2}\sigma^2$, ce qui revient à dire que les coefficients sont distribués suivant la distribution de Rayleigh de paramètre σ^2 .

Pour la suite, on prend \mathbf{H} est distribuée selon une distribution de Rayleigh indépendante de paramètre 1. De plus, \mathbf{H} est supposé indépendante de l'entrée \mathbf{x} et du bruit \mathbf{n} . En plus de cela, on suppose que la réalisation H de \mathbf{H} est inconnue de l'émetteur. L'entrée du canal est, donc, juste l'entrée \mathbf{x} et la sortie, le couple $(\mathbf{y}, \mathbf{H}) = (\mathbf{H}\mathbf{x} + \mathbf{n}, \mathbf{H})$.

Le but de cette partie va être de démontrer le théorème suivant

Théorème 4.2. *La capacité du canal MIMO, sous les hypothèses précédentes et la contrainte $\mathbb{E}[\mathbf{x}^*\mathbf{x}] \leq P$, est*

$$C_{MIMO} = \mathbb{E} \left[\log \det \left(I_r + \frac{P}{t} \mathbf{H}\mathbf{H}^* \right) \right]$$

Remarque. La contrainte $\mathbb{E}[\mathbf{x}^*\mathbf{x}] \leq P$ est équivalente à $\mathbb{E}[\mathbf{x}\mathbf{x}^*] = Q$ où $\text{tr}(Q) \leq P$. Dans les deux cas, cela signifie que la moyenne de la norme de \mathbf{x} est inférieure à P .

4.1 Propriété de la distribution de Rayleigh indépendante sur les matrices

Proposition 4.3. *Soit \mathbf{H} distribuée selon une distribution de Rayleigh indépendante de paramètres σ^2 . Alors, pour tout $U \in \mathbb{C}^{r \times r}$ et $V \in \mathbb{C}^{t \times t}$ unitaires (i.e. $U = U^*$), la distribution de $U\mathbf{H}V$ est la même que celle de \mathbf{H} .*

Démonstration. On va d'abord montrer que cela est vrai pour $\mathbf{G} = U\mathbf{H}$. Pour ce faire, on va montrer que les colonnes de \mathbf{G} sont indépendantes et, ensuite, qu'elles ont même distribution que celles de \mathbf{H} . On va noter \mathbf{g}_i (respectivement \mathbf{h}_i) la i^{e} colonne de \mathbf{G} (respectivement \mathbf{H}). Montrons l'indépendance des colonnes de \mathbf{G} . On a, évidemment, par le produit par bloc, que $\mathbf{g}_i = U\mathbf{h}_i$. Soit i, j avec $i \neq j$ on a

$$\begin{aligned} \mathbb{E}[\mathbf{g}_i \mathbf{g}_j^*] &= \mathbb{E}[U\mathbf{h}_i \mathbf{h}_j^* U^*] \\ &= U\mathbb{E}[\mathbf{h}_i \mathbf{h}_j^*] U^* \\ &= U\mathbb{E}[\mathbf{h}_i] \mathbb{E}[\mathbf{h}_j^*] U^* \\ &\text{(car les colonnes de } \mathbf{H} \text{ sont indépendantes)} \\ &= \mathbb{E}[U\mathbf{h}_i] \mathbb{E}[\mathbf{h}_j^* U^*] \\ &= \mathbb{E}[\mathbf{g}_i] \mathbb{E}[\mathbf{g}_j^*] \end{aligned}$$

Donc les colonnes de \mathbf{G} sont indépendantes.

Il reste à montrer qu'elles ont mêmes lois. Tout d'abord, \mathbf{g}_i est un vecteur gaussien de moyenne nulle. En effet, une combinaison linéaire des coordonnées de \mathbf{g}_i est une combinaison linéaire des coordonnées de \mathbf{h}_i (car les coordonnées de \mathbf{g}_i sont des combinaisons linéaires des coordonnées de \mathbf{h}_i). Comme \mathbf{h}_i est un vecteur gaussien, toute combinaison linéaire de ses coordonnées suit une loi gaussienne. En particulier, toute combinaison linéaire des coordonnées de \mathbf{g}_i suit une loi gaussienne. Et donc, \mathbf{g}_i est un vecteur gaussien. La moyenne nulle est due au fait que chacune de ses coordonnées est combinaison linéaire de gaussiennes de moyenne nulle. Il suffit donc, maintenant de montrer que \mathbf{g}_i et \mathbf{h}_i ont même variance pour montrer qu'ils ont même loi. On calcule, d'abord, la covariance de \mathbf{h}_i .

$$\begin{aligned} \mathbb{E}[\mathbf{h}_i \mathbf{h}_i^*] &= \mathbb{E}[(h_{ij} \bar{h}_{ik})_{j,k}] \\ &= (\mathbb{E}[h_{ij} \bar{h}_{ik}])_{j,k} \\ &= (\mathbb{E}[h_{ij} \bar{h}_{ik}])_{j,k} \\ &= \left(\begin{array}{cc} \sigma^2 & \text{si } j = k \\ 0 & \text{si } j \neq k \end{array} \right)_{j,k} \\ &= \sigma^2 I \end{aligned}$$

Il ne reste plus qu'à calculer la covariance de \mathbf{g}_i .

$$\begin{aligned} \mathbb{E}[\mathbf{g}_i \mathbf{g}_i^*] &= \mathbb{E}[U\mathbf{h}_i \mathbf{h}_i^* U^*] \\ &= U\mathbb{E}[\mathbf{h}_i \mathbf{h}_i^*] U^* \\ &= U(\sigma^2 I) U^* \\ &= \sigma^2 I \\ &= \mathbb{E}[\mathbf{h}_i \mathbf{h}_i^*] \end{aligned}$$

Nous avons donc montré que les colonnes de \mathbf{G} sont indépendantes et distribuées comme celle de \mathbf{H} . Donc \mathbf{G} a même distribution que \mathbf{H} . \square

Muni de cette proposition, on peut commencer le calcul de la capacité du canal par le calcul de l'information mutuelle entre l'entrée \mathbf{x} et la sortie $(\mathbf{y}, \mathbf{H}) = (\mathbf{H}\mathbf{x} + \mathbf{n}, \mathbf{H})$.

$$\begin{aligned} I(\mathbf{x}; (\mathbf{y}, \mathbf{H})) &= I(\mathbf{x}|\mathbf{H}) + I(\mathbf{x}; \mathbf{y}|\mathbf{H}) \\ &= 0 + I(\mathbf{x}; \mathbf{y}|\mathbf{H}) \\ &= \mathbb{E}[I(\mathbf{x}; \mathbf{y}|\mathbf{H} = H)] \end{aligned}$$

Il faut, donc, en particulier chercher à maximiser $I(\mathbf{x}; \mathbf{y}|\mathbf{H} = H)$ sous la contrainte $\mathbb{E}[\mathbf{x}\mathbf{x}^*] = Q$ où $\text{tr}(Q) \leq P$.

4.2 Maximisation de $I(\mathbf{x}; \mathbf{y}|\mathbf{H} = H)$ avec la contrainte $\mathbb{E}[\mathbf{x}\mathbf{x}^*] = Q$

On va effectuer tous les calculs, dans cette partie, en fixant \mathbf{H} à H et on va faire l'abus de notation dans cette partie $I(\mathbf{x}; \mathbf{y}|\mathbf{H} = H) = I(\mathbf{x}; \mathbf{y})$.

On va commencer par calculer l'information mutuelle.

$$\begin{aligned} I(\mathbf{x}; \mathbf{y}) &= h(\mathbf{y}) - h(\mathbf{y}|\mathbf{x}) \\ &= h(\mathbf{y}) - h(H\mathbf{x} + \mathbf{n}|\mathbf{x}) \\ &= h(\mathbf{y}) - h(\mathbf{n}|\mathbf{x}) \\ &= h(\mathbf{y}) - h(\mathbf{n}) \end{aligned}$$

Ainsi pour maximiser $I(\mathbf{x}, \mathbf{y})$, il suffit de maximiser $h(\mathbf{y})$. On va, dans un premier temps, nous restreindre aux entrées \mathbf{x} de moyenne nulle.

$$\begin{aligned} \mathbb{E}[(\mathbf{x} - \mathbb{E}[\mathbf{x}])^*(\mathbf{x} - \mathbb{E}[\mathbf{x}])] &= \mathbb{E}[(\mathbf{x}^* - \mathbb{E}[\mathbf{x}^*])(\mathbf{x} - \mathbb{E}[\mathbf{x}])] \\ &= \mathbb{E}[\mathbf{x}^*\mathbf{x}] - \mathbb{E}[\mathbf{x}^*\mathbb{E}[\mathbf{x}]] - \mathbb{E}[\mathbb{E}[\mathbf{x}^*]\mathbf{x}] + \mathbb{E}[\mathbb{E}[\mathbf{x}^*]\mathbb{E}[\mathbf{x}]] \\ &= \mathbb{E}[\mathbf{x}^*\mathbf{x}] - \mathbb{E}[\mathbf{x}^*]\mathbb{E}[\mathbf{x}] - \mathbb{E}[\mathbf{x}^*]\mathbb{E}[\mathbf{x}] + \mathbb{E}[\mathbf{x}^*]\mathbb{E}[\mathbf{x}] \\ &= \mathbb{E}[\mathbf{x}^*\mathbf{x}] - \mathbb{E}[\mathbf{x}^*]\mathbb{E}[\mathbf{x}] \\ &\leq \mathbb{E}[\mathbf{x}^*\mathbf{x}] \\ &= P \end{aligned}$$

(car $\mathbb{E}[\mathbf{x}^*]\mathbb{E}[\mathbf{x}]$ est la norme de $\mathbb{E}[\mathbf{x}]$ qui est positif).

Ce calcul montre que si \mathbf{x} vérifie la contrainte de puissance $\mathbf{x} - \mathbb{E}[\mathbf{x}]$ aussi. On peut donc restreindre notre étude aux entrées de moyenne nulle. Maintenant, prenons \mathbf{x} de moyenne nulle et de covariance $\mathbb{E}[\mathbf{x}\mathbf{x}^*] = Q$. On a alors, pour la

covariance de la sortie \mathbf{y}

$$\begin{aligned}
\mathbb{E}[\mathbf{y}\mathbf{y}^*] &= \mathbb{E}[(H\mathbf{x} + \mathbf{n})(H\mathbf{x} + \mathbf{n})^*] \\
&= \mathbb{E}[(H\mathbf{x} + \mathbf{n})(\mathbf{x}^*H^* + \mathbf{n}^*)] \\
&= \mathbb{E}[H\mathbf{x}\mathbf{x}^*H^*] + \mathbb{E}[H\mathbf{x}\mathbf{n}^*] + \mathbb{E}[\mathbf{n}\mathbf{x}^*H^*] + \mathbb{E}[\mathbf{n}\mathbf{n}^*] \\
&= H\mathbb{E}[\mathbf{x}\mathbf{x}^*]H^* + H\mathbb{E}[\mathbf{x}]\mathbb{E}[\mathbf{n}^*] + \mathbb{E}[\mathbf{n}]\mathbb{E}[\mathbf{x}^*]H^* + \mathbb{E}[\mathbf{n}\mathbf{n}^*] \\
&= HQH^* + 0 + 0 + I \\
&= I + HQH^*
\end{aligned}$$

Pour maximiser $I(\mathbf{x}; \mathbf{y})$, il suffit de choisir, par le théorème 3.3, \mathbf{y} qui suit la loi d'un vecteur gaussien de moyenne nulle et de variance $I + HQH^*$. Dans ce cas-là, comme $\mathbf{x} = H^{-1}(\mathbf{y} - \mathbf{n})$, on peut appliquer la proposition 2.8, pour voir qu'alors \mathbf{x} suit la loi d'un vecteur gaussien. Donc, la capacité est atteinte pour le vecteur gaussien \mathbf{x} de moyenne nulle et de covariance Q et elle vaut

$$\begin{aligned}
I(\mathbf{x}; \mathbf{y}) &= h(\mathbf{y}) - h(\mathbf{n}) \\
&= \log \det(\pi e(I + HQH^*)) - \log \det(\pi eI) \\
&= \log((\pi e)^r \det(I + HQH^*)) - \log(\pi e)^r \\
&= \log(\pi e)^r + \log \det(I + HQH^*) - \log(\pi e)^r \\
&= \log \det(I + HQH^*)
\end{aligned}$$

Ainsi, pour maximiser $I(\mathbf{x}; \mathbf{y} | \mathbf{H} = H)$ sous la contrainte $\mathbb{E}[\mathbf{x}\mathbf{x}^*] = Q$, il faut choisir \mathbf{x} vecteur aléatoire de covariance Q . On atteint alors le maximum pour l'information mutuelle qui vaut $\Psi(Q, H) = \log \det(I_r + \mathbf{H}Q\mathbf{H}^*)$. Ainsi, il ne reste plus qu'à maximiser

$$\Psi(Q) = \mathbb{E}[\Psi(Q, \mathbf{H})] = \mathbb{E}[\log \det(I_r + \mathbf{H}Q\mathbf{H}^*)]$$

sur les matrices Q définies positives avec $\text{tr}(Q) \leq P$.

4.3 Maximisation de $\Psi(Q)$ avec $\text{tr}(Q) \leq P$

Cette partie est la dernière et va nous amener vers la valeur de C_{MIMO} . La démonstration sera articulée autour de lemmes qui en formeront les points clés.

Comme Q une matrice définie positive, il existe U unitaire (ie $UU^* = I$) et D matrice diagonale à coefficients diagonaux strictement positifs tel que $Q = UDU^*$.

Ainsi, on peut écrire

$$\begin{aligned}
\Psi(Q) &= \mathbb{E}[\log \det(I_r + \mathbf{H}U D U^* \mathbf{H}^*)] \\
&= \mathbb{E}[\log \det(I_r + (\mathbf{H}U) D (\mathbf{H}U)^*)]
\end{aligned}$$

Ainsi, par la propriété 4.3, nous avons $\Psi(Q) = \Psi(D)$ et nous pouvons donc restreindre notre étude sur les matrices Q diagonales définies et positives. Maintenant, nous allons chercher le maximum de $\Psi(Q)$. Pour cela nous allons introduire les matrices de permutation et on va également montrer que $\Psi(Q)$ est concave.

Définition 4.4 (Matrice de permutation). Une matrice Π est une matrice de permutation si et seulement si elle possède un et un seul 1 par ligne, elle possède un et un seul 1 par colonne et tous les autres coefficients sont nuls.

Remarque. La matrice $\Pi Q \Pi^*$ est la matrice qui correspond à la matrice Q dans laquelle on aurait permuter les éléments de la base sous-jacente. En effet, Π peut être vu comme la matrice de passage de la base (e_1, \dots, e_t) à la base (e'_1, \dots, e'_t) avec $e'_j = e_i$ où $(\Pi)_{ij} = 1$. En particulier, si Q est diagonale, $\Pi Q \Pi^*$ est diagonale avec les mêmes éléments sur la diagonale.

Toujours par la propriété 4.3, on n'a que $\psi(Q) = \psi(\Pi Q \Pi^*)$.

Lemme 4.5. *L'image des matrices définies positives de $\mathbb{C}^{t \times t}$ par l'application $Q \rightarrow I_r + H Q H^*$ est contenu dans l'ensemble des matrices définies positives de $\mathbb{C}^{r \times r}$.*

Démonstration. Soit $X \in \mathbb{C}^r - \{0\}$. On a

$$\begin{aligned} X^*(I_r + H Q H^*)X &= X^* I_r X + X^* H Q H^* X \\ &> 0 + (X^* H) Q (X^* H)^* && \text{(car } X^* X = \|X\|^2 > 0) \\ &> 0 && \text{(car } Q \text{ est positive)} \end{aligned}$$

Ainsi, $I_r + H Q H^*$ est définie positive. □

Lemme 4.6. Ψ est concave sur les matrices définies positives $\mathbb{C}^{t \times t}$.

Démonstration. Soit $\lambda \in [0, 1]$, Q_0 et Q_1 deux matrices définies positives de $\mathbb{C}^{t \times t}$. On a

$$\begin{aligned} \log |I_r + H(\lambda Q_0 + (1 - \lambda) Q_1) H^*| &= \log |\lambda(I_r + H Q_0 H^*) + (1 - \lambda)(I_r + H Q_1 H^*)| \\ &\geq \lambda \log |I_r + H Q_0 H^*| \\ &\quad + (1 - \lambda) \log |I_r + H Q_1 H^*| \end{aligned}$$

par la proposition 3.9. Ainsi, on en déduit que $\psi : Q \rightarrow \mathbb{E}[\log \det(I_r + H Q H^*)]$ est concave. □

On prend maintenant la matrice $\tilde{Q} = \frac{1}{t!} \sum_{\Pi} \Pi Q \Pi^*$. On va montrer que ce sont les matrices qui vérifie $Q = \tilde{Q}$ qui nous intéressent. Pour cela, on va montrer quelques propriétés des matrices \tilde{Q} .

Lemme 4.7. *Soit Q une matrice définie positive et diagonale, alors on a*

$$\psi(\tilde{Q}) \geq \psi(Q)$$

et

$$\text{tr}(\tilde{Q}) = \text{tr}(Q)$$

C'est à dire que \tilde{Q} est meilleure que Q pour maximiser l'information mutuelle tout en respectant la même contrainte de puissance.

Démonstration. On montre d'abord $\psi(\tilde{Q}) \geq \psi(Q)$:

$$\begin{aligned} \psi(\tilde{Q}) &= \psi\left(\frac{1}{t!} \sum_{\Pi} \Pi Q \Pi^*\right) \\ &\geq \frac{1}{t!} \sum_{\Pi} \psi(\Pi Q \Pi^*) && \text{(par le lemme 4.6)} \\ &= \frac{1}{t!} \sum_{\Pi} \psi(Q) && \text{(par la propriété 4.3)} \\ &= \frac{t!}{t!} \psi(Q) \\ &= \psi(Q) \end{aligned}$$

(car $\text{card}\{\text{permutation de } \{1, \dots, t\}\} = t!$)

Puis, ensuite $\text{tr}(\tilde{Q}) = \text{tr}(Q)$:

$$\begin{aligned} \text{tr}(\tilde{Q}) &= \text{tr}\left(\frac{1}{t!} \sum_{\Pi} \Pi Q \Pi^*\right) \\ &= \frac{1}{t!} \sum_{\Pi} \text{tr}(\Pi Q \Pi^*) \\ &= \frac{1}{t!} \sum_{\Pi} \text{tr}(Q) && (\text{tr}(Q) = \sum_{i=1}^t q_{ii} \text{ est invariant par permutation}) \\ &= \frac{t!}{t!} \text{tr}(Q) \\ &= \text{tr}(Q) \end{aligned}$$

□

Ainsi, par ce lemme, on a donc que pour maximiser l'information mutuelle, il faut chercher une matrice Q diagonale définie et positive telle que $Q = \tilde{Q}$. En choisissant Q sous la forme αI , on a évidemment que $\tilde{Q} = Q$. Ainsi, nous concluons que le maximum pour $\psi(Q)$ est atteint pour une matrice de la forme αI . Le maximum est évidemment atteint pour α le plus grand possible. La condition $\text{tr}(Q) = P$ nous donne $\text{tr}(\alpha I) = \alpha t$. On prend donc $\alpha = \frac{P}{t}$.

Ainsi, nous avons que la capacité du canal que l'on étudie est atteinte pour une distribution de \mathbf{x} gaussienne et de covariance $(P/t)I_t$ et vaut donc

$$C_{\text{MIMO}} = \mathbb{E} \left[\log \det \left(I_r + \frac{P}{t} \mathbf{H} \mathbf{H}^* \right) \right]$$

5 Calcul asymptotique de la capacité du canal

On présente ici quelques ouvertures possibles du résultat précédent. On ne justifiera donc pas les arguments de manière aussi formelle que précédemment.

5.1 Quand r est fixé et $t \rightarrow \infty$

Pour calculer la capacité dans ce cas là, on va étudier les coefficients de $\frac{1}{t} \mathbf{H} \mathbf{H}^*$.

$$\begin{aligned} \frac{1}{t} (\mathbf{H} \mathbf{H}^*)_{ij} &= \frac{1}{t} \sum_{k=1}^t h_{ik} \bar{h}_{jk} \\ &\stackrel{\text{ps}}{\rightarrow} \mathbb{E} [h_{ik} \bar{h}_{jk}] \\ &= \begin{cases} \frac{1}{2} + \frac{1}{2} = 1 & \text{si } i = j \\ 0 & \text{si } i \neq j \end{cases} \end{aligned}$$

(par la loi des grands nombres)

On peut donc espérer que

$$\begin{aligned} C_{\text{MIMO}} &\rightarrow \mathbb{E} [\log \det (I_r + P I_r)] \\ &= \log \det ((1 + P) I_r) \\ &= \log (1 + P)^r \\ &= r \log (1 + P) \end{aligned}$$

Remarque. On peut interpréter le résultat précédent comme le fait que le canal MIMO avec une infinité d'antennes d'entrées et r antennes de sortie est équivalent à r canaux gaussiens indépendants.

5.2 Capacité MIMO et mesure spectrale

On va expliciter l'expression de la capacité MIMO obtenue dans la partie précédente en termes de mesure spectrale que l'on définit maintenant.

Définition 5.1. Soit $W \in \mathcal{M}_r(\mathbb{C})$ une matrice hermitienne. On note $\lambda_1 \geq \dots \geq \lambda_r \geq 0$ les valeurs propres de W . On définit la *mesure spectrale* associée à W . C'est la mesure de probabilité sur \mathbb{R}^+ définie par :

$$\mu_W(B) = \frac{1}{r} \sum_{k=1}^r \delta_{\lambda_k}(B), \quad \forall B \in \mathcal{B}(\mathbb{R}^+)$$

On repart de l'expression :

$$C_{\text{MIMO}} = \mathbb{E} \left[\log \det \left(I_r + \frac{P}{t} \mathbf{H} \mathbf{H}^* \right) \right]$$

La matrice $\mathbf{H} \mathbf{H}^*$ est hermitienne positive, donc on peut écrire $\mathbf{H} \mathbf{H}^* = U \Delta U^*$ avec $\Delta = \text{diag}(\lambda_1, \dots, \lambda_r)$ avec $\lambda_i \geq 0$ pour tout i . D'où :

$$\begin{aligned} \log \det \left(I_r + \frac{P}{t} \mathbf{H} \mathbf{H}^* \right) &= \log \det \left(I_r + \frac{P}{t} U \Delta U^* \right) \\ &= \log \det \left(I_r + \frac{P}{t} \Delta \right) = \log \prod_{k=1}^r \left(1 + \frac{P}{t} \lambda_i \right) \\ &= \sum_{k=1}^r \log \left(1 + \frac{P}{t} \lambda_i \right) \end{aligned}$$

Remarque. À ce stade du calcul, on peut remarquer que le canal MIMO est équivalent à r canaux gaussiens indépendants à une dimension, leurs capacités étant modifiées par les valeurs propres de la matrice $\mathbf{H} \mathbf{H}^*$.

On voit que pour calculer explicitement la capacité MIMO, il faut connaître la loi jointe des valeurs propres de $\mathbf{H} \mathbf{H}^*$. Les matrices qui suivent une distribution de Rayleigh indépendante sont un cas de particulier de matrices de Wishat pour lesquelles cette loi est connue et s'exprime à l'aide des polynômes de Laguerre.

Pour l'étude asymptotique que nous proposons, il est préférable de ré-exprimer le résultat précédent avec la mesure spectrale de $\mathbf{H} \mathbf{H}^*$. Plus précisément on a :

$$\begin{aligned} \sum_{k=1}^r \log \left(1 + \frac{P}{t} \lambda_i \right) &= r \cdot \frac{1}{r} \sum_{k=1}^r \log \left(1 + \frac{P}{t} \lambda_i \right) \\ &= r \int_0^{+\infty} \log \left(1 + \frac{P}{t} x \right) \mu_{\mathbf{H} \mathbf{H}^*}(\mathrm{d}x) \end{aligned}$$

On obtient finalement l'expression suivante pour C_{MIMO} :

$$C_{\text{MIMO}} = \mathbb{E} \left[r \int_0^{+\infty} \log \left(1 + \frac{P}{t} x \right) \mu_{\mathbf{H} \mathbf{H}^*}(\mathrm{d}x) \right]$$

5.3 Quand $r, t \rightarrow \infty$ et $r/t \rightarrow \beta$

Dans cette section, nous allons chercher la limite de C_{MIMO} dans le cas où le nombre d'antennes d'entrée et de sortie devient infini, mais où le rapport t/r , lui, tend vers une valeur finie strictement positive β . Pour cela, nous avons besoin de définir la convergence d'une mesure de probabilité et d'admettre un théorème sur la convergence des valeurs propres des matrices aléatoires.

Définition 5.2 (Convergence d'une mesure de probabilité). Soit $\mu_{\mathbf{W}}$ une suite de mesure de probabilité aléatoire et soit μ une mesure de probabilité. On dit que $\mu_{\mathbf{W}} \rightarrow \mu$ si

$$\forall f \text{ continue bornée } \mathbb{E} \left[\int f(x) \mu_{\mathbf{W}}(dx) \right] \rightarrow \int f(x) \mu(dx)$$

On va admettre le théorème suivant

Théorème 5.3 (Marcenko et Pastur (1967)). *On suppose*

- \mathbf{A} est une matrice de dimension $r \times t$ dont tous les coefficients a_{ij} sont des variables aléatoires indépendantes et de même loi. La loi étant de moyenne nulle, de variance 1 et tel que $\mathbb{E}[|a_{ij}|^4] < \infty$.
 - $r/t \rightarrow \beta > 0$ quand $r, t \rightarrow \infty$.
 - $\mathbf{W}_r = \frac{1}{t} \mathbf{A} \mathbf{A}^*$
- Alors

$$\mu_{\mathbf{W}_r} \xrightarrow{r \rightarrow \infty} \mu_{M-P}$$

où

$$\mu_{M-P}(dx) = \left(1 - \frac{1}{\beta}\right)^+ \delta_0(x) + \frac{\sqrt{(x-a)^+(b-x)^+}}{2\pi\beta x} dx$$

avec $a = (1 - \sqrt{\beta})^2$ et $b = (1 + \sqrt{\beta})^2$.

On continue, alors, le calcul de C_{MIMO} .

$$\begin{aligned} C_{\text{MIMO}} &= r \mathbb{E} \left[\int_0^{+\infty} \log\left(1 + \frac{P}{t} x\right) \mu_{\mathbf{H}\mathbf{H}^*}(dx) \right] \\ &= r \mathbb{E} \left[\int_0^{+\infty} \log\left(1 + P \frac{t}{t} x\right) \mu_{\frac{1}{t} \mathbf{H}\mathbf{H}^*}(dx) \right] \\ &\quad (\text{car } \lambda \text{ est valeur propre de } \mathbf{H}\mathbf{H}^* \text{ ssi } \frac{1}{t} \lambda \text{ est valeur propre de } \frac{1}{t} \mathbf{H}\mathbf{H}^*) \\ &= r \mathbb{E} \left[\int_0^{+\infty} \log(1 + Px) \mu_{\frac{1}{t} \mathbf{H}\mathbf{H}^*}(dx) \right] \end{aligned}$$

On a presque l'application du théorème, pour l'avoir il faudrait que $x \rightarrow \log(1 + Px)$ soit bornée.

Ainsi, on peut écrire que la capacité par antenne réceptrice du canal MIMO est alors

$$\begin{aligned} \frac{1}{r} C_{\text{MIMO}} &= \mathbb{E} \left[\int_0^{+\infty} \log(1 + Px) \mu_{\frac{1}{t} \mathbf{H}\mathbf{H}^*}(dx) \right] \\ &\xrightarrow{r \rightarrow \infty} \int_0^{+\infty} \log(1 + Px) \mu_{M-P}(dx) \\ &= \int_a^b \log(1 + Px) \mu_{M-P}(dx) \end{aligned}$$

(car $\log(1 + Px) = 0$ en 0 et que $\mu_{M-P}(x) = 0$ en dehors de $[a, b]$ et de 0)

On peut également écrire la capacité par antenne émettrice du canal MIMO

$$\begin{aligned} \frac{1}{t} C_{\text{MIMO}} &= \frac{r}{t} \mathbb{E} \left[\int_0^{+\infty} \log(1 + Px) \mu_{\frac{1}{t} \mathbf{H}\mathbf{H}^*}(dx) \right] \\ &\xrightarrow{r \rightarrow \infty} \beta \int_0^{+\infty} \log(1 + Px) \mu_{M-P}(dx) \\ &= \beta \int_a^b \log(1 + Px) \mu_{M-P}(dx) \end{aligned}$$

On peut aussi regarder la capacité par antenne (émettrice et réceptrice) du canal MIMO

$$\begin{aligned} \frac{1}{t+r} C_{\text{MIMO}} &= \frac{r}{r+t} \mathbb{E} \left[\int_0^{+\infty} \log(1 + Px) \mu_{\frac{1}{t} \mathbf{H}\mathbf{H}^*}(dx) \right] \\ &= \frac{1}{1 + \frac{t}{r}} \int_0^{+\infty} \log(1 + Px) \mu_{M-P}(dx) \\ &\xrightarrow{r \rightarrow \infty} \frac{1}{1 + \frac{1}{\beta}} \int_0^{+\infty} \log(1 + Px) \mu_{M-P}(dx) \\ &= \frac{1}{1 + \frac{1}{\beta}} \int_a^b \log(1 + Px) \mu_{M-P}(dx) \end{aligned}$$

5.4 Étude numérique

Cette section va présenter brièvement les résultats que l'on obtient à partir de la partie précédente. En particulier, elle contiendra le graphe des trois fonctions qui concluaient la partie précédente pour quelques valeurs de P . En ordonnée, se trouve la capacité par antenne du canal et en abscisse, $B = \beta$, la proportion d'antennes réceptrices par rapport aux antennes émettrices. En rouge, c'est le graphe de la capacité par antenne émettrice du canal MIMO; en vert, c'est la capacité par antenne réceptrice et en jaune, c'est la capacité par antenne.

Tout d'abord, il faut prendre $P \geq 1$, c'est à dire dans notre modèle une puissance supérieure à celle du bruit pour obtenir une capacité significative.

La deuxième application correspond à une puissance de l'ordre du bruit ($P=1$).

Enfin, les troisième et quatrième applications numériques sont prises pour une puissance plus importante que le bruit.

Ces trois derniers graphes, nous montre qu'il y a apparemment un bon compromis pour la capacité d'un tel canal qui est atteint pour $\beta = 1$, c'est à dire quand on a autant d'antennes émettrices que réceptrices.

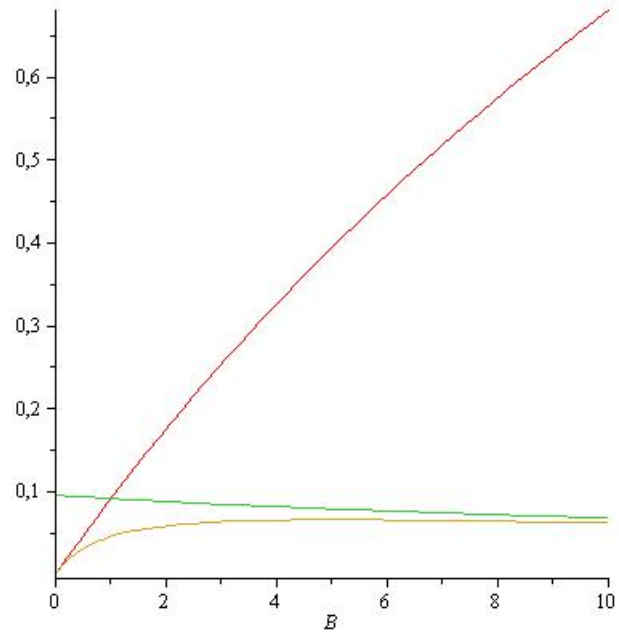


FIG. 1 – La puissance des antennes est de $P=0.1$ ce qui est faible par rapport au bruit

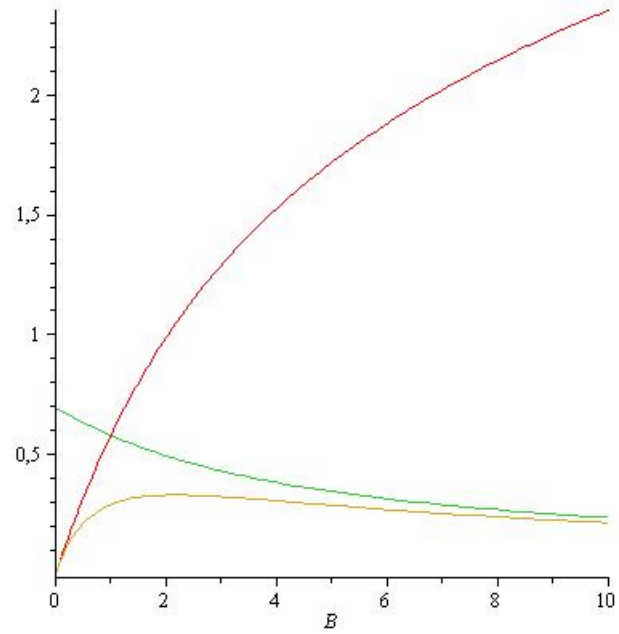


FIG. 2 – La puissance des antennes est de $P=1$ ce qui est de l'ordre de la puissance du bruit

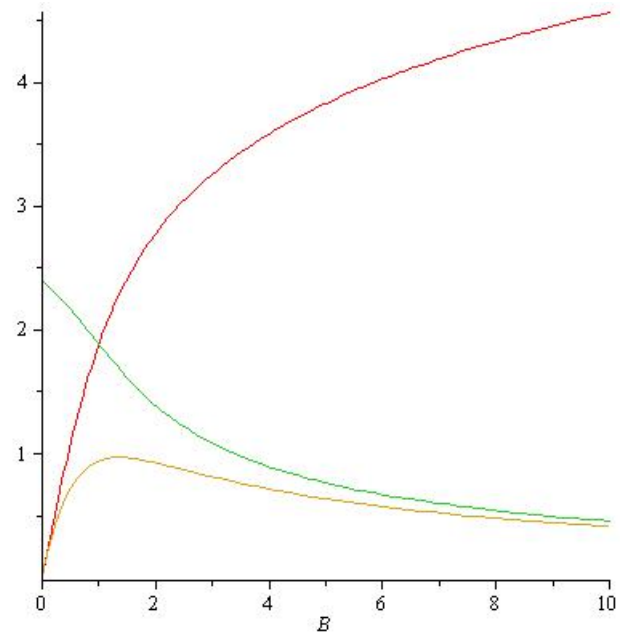


FIG. 3 – La puissance des antennes est de $P=10$

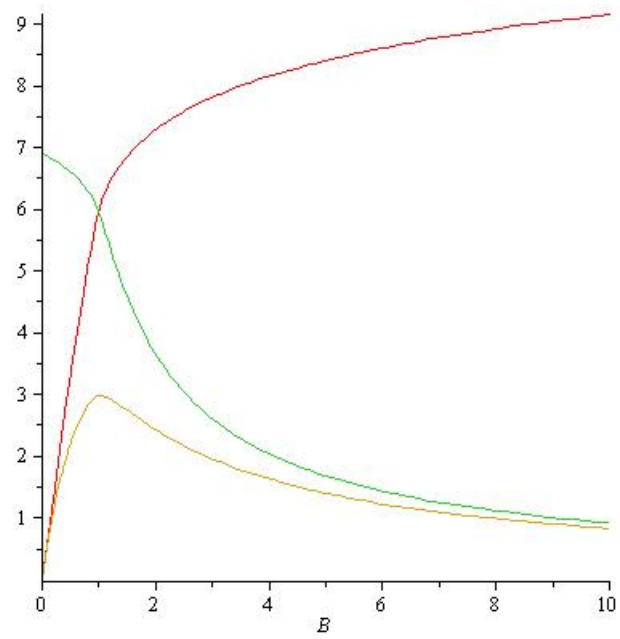


FIG. 4 – La puissance des antennes est de $P=1000$

Références

- [1] Thomas M. COVER et Joy A. THOMAS : *Elements of Information theory*. Wiley, 2^{de} édition, July 2006.
- [2] A. MARČENKO et L. A. PASTUR : Distribution of eigenvalues for some sets of random matrices. *Math. USSR-Sb*, 1(4):457–483, 1967.
- [3] Emre TELATAR : Capacity of multi-antenna gaussian channels. *European Transactions on Telecommunications*, 10(6):585–596, 1999.
- [4] David TSE et Pramod VISWANATH : *Fundamentals of Wireless Communication*. Cambridge University Press, 2005.