

Projet ANR-12-MONU-0003 DIONISOS.

Rapport final sur le WP3 Problèmes liés aux sources large bande.

1 Introduction.

Ce rapport synthétise l'ensemble des résultats obtenus au cours du WP3 intitulé "Problèmes liés aux sources large bande", et correspond aux livrables D3.1-3, D3.2-2 et D3.3-2. Ce WP a été divisé en trois parties. Le WP3-1 a été consacré à l'étude des valeurs propres et vecteurs propres de matrices de covariance spatio-temporelles constituées à partir d'un bruit blanc temporellement et spatialement, et de la somme d'un signal défini comme la sortie d'un filtre 1 entrée / multi-sorties attaqué par un signal déterministe non observable auquel vient s'ajouter un bruit blanc temporellement et spatialement. Le WP3-2 s'est intéressé aux applications des résultats du WP3-1 à la localisation de sources bande étroite basés sur la technique dite du "spatial smoothing", et à la détection d'une source large bande dans un bruit blanc temporellement et spatialement. Le WP3-3 a été consacré aux algorithmes d'estimation de filtres prédicteurs spatio-temporels dans le contexte des grandes dimensions.

Les résultats obtenus dans le cadre du WP3 ont donné lieu aux publications suivantes:

Articles de revue

- P. Loubaton, "On the almost sure location of the singular values of certain Gaussian block-Hankel large random matrices", *J. of Theoretical Probability*, vol. 29, no. 4, pp. 1339-1443, December 2016, also available on Arxiv (<http://arxiv.org/abs/1405.2006>).
- G.T. Pham, P. Loubaton, P. Vallet, "Performance analysis of spatial smoothing schemes in the context of large arrays", *IEEE Trans. on Signal Processing*, vol. 64, no. 1, pp. 160-172, January 2016, also available on Arxiv (arXiv:1503.08196).

Articles de conférence.

- S. Bensaid, D. Slock, "Wiener filtering in the DFT transformed domain", *Proc. EUSIPCO 2014*, Lisbon, Portugal.
- G.T. Pham, P. Loubaton, P. Vallet, "Performance analysis of spatial smoothing schemes in the context of large arrays", *Acoustics, Speech and Signal Processing (ICASSP), 2015 IEEE International Conference on*, Year: 2015, Pages: 2824 - 2828.
- G.T. Pham; P. Loubaton, "Applications of large empirical spatio-temporal covariance matrix in multipath channels detection", *Signal Processing Conference (EUSIPCO)*, Nice, September 2015.
- G.T. Pham; P. Loubaton, "Performances des filtres de Wiener spatio-temporels entraînés: le cas des grandes dimensions", *Proc. Colloque GretsI*, Lyon, September 2015.
- G.T. Pham; P. Loubaton, "Optimization of the loading factor of regularized estimated spatial-temporal Wiener filters in large system case", *Proc. of Statistical Signal Processing Workshop (SSP)*, Palma de Majorque, June 2016.

2 Positionnement des travaux.

Rappelons que le WP3 est consacré à des problèmes de détection / estimation multi-capteurs mettant en jeu des signaux large bande. En présence d'une source large bande, le signal de dimension M $(\mathbf{y}_n)_{n \in \mathbb{Z}}$ recueilli sur le réseau de capteurs à M éléments peut s'écrire sous la forme

$$\mathbf{y}_n = \sum_{p=0}^{P-1} \mathbf{h}_p s_{n-p} + \mathbf{v}_n = \mathbf{x}_n + \mathbf{v}_n \quad (2.1)$$

où $(s_n)_{n \in \mathbb{Z}}$ représente un signal scalaire déterministe non observable, $(\mathbf{h}_p)_{p=0, \dots, P-1}$ sont des vecteurs déterministes inconnus de dimension M et où $(\mathbf{v}_n)_{n \in \mathbb{Z}}$ est un bruit gaussien complexe, blanc temporellement et spatialement de variance σ^2 . Le signal \mathbf{x}_n représente le signal utile sur lequel il est souhaitable d'obtenir des informations à partir des observations $(\mathbf{y}_n)_{n=1, \dots, N}$. On peut par exemple s'intéresser aux problèmes suivants:

- Détecter la présence ou l'absence de (\mathbf{x}_n) dans l'observation, ce qui est équivalent à un problème de détection.
- Estimer les vecteurs $(\mathbf{h}_p)_{p=0, \dots, P-1}$.
- Estimer la séquence (s_n) .

(voir par exemple [20], [11], [21], [1] pour plus de précisions). Un bon nombre d'algorithmes de détection / estimation mettant en jeu le signal \mathbf{y} utilisent les valeurs propres et les vecteurs propres de la matrice $\frac{\mathbf{Y}_L \mathbf{Y}_L^*}{N}$ où \mathbf{Y}_L est la matrice Hankel par blocs $ML \times N$ définie par

$$\mathbf{Y}_L = \begin{pmatrix} \mathbf{Y}_{1,L} \\ \vdots \\ \mathbf{Y}_{M,L} \end{pmatrix}$$

où pour chaque $m = 1, \dots, M$, la matrice $\mathbf{Y}_{m,L}$ est la matrice de Hankel de dimensions $L \times N$ donnée par

$$\mathbf{Y}_{m,L} = \begin{pmatrix} \mathbf{y}_{m,1} & \mathbf{y}_{m,2} & \cdots & \cdots & \mathbf{y}_{m,N} \\ \mathbf{y}_{m,2} & \mathbf{y}_{m,3} & \ddots & \ddots & \mathbf{y}_{m,N+1} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \mathbf{y}_{m,L} & \mathbf{y}_{m,L+1} & \cdots & \cdots & \mathbf{y}_{m,N+L-1} \end{pmatrix}$$

$\frac{\mathbf{Y}_L \mathbf{Y}_L^*}{N}$ peut être interprétée comme une matrice de covariance spatio-temporelle empirique car elle peut aussi s'écrire sous la forme

$$\frac{\mathbf{Y}_L \mathbf{Y}_L^*}{N} = \frac{1}{N} \sum_{n=1}^N \mathbf{y}_n^{(L)} \mathbf{y}_n^{(L)*}$$

où $\mathbf{y}_n^{(L)}$ est le vecteur de dimension ML défini par

$$\mathbf{y}_n^{(L)} = (\mathbf{y}_{1,n}, \dots, \mathbf{y}_{1,n+L-1}, \dots, \mathbf{y}_{M,n}, \dots, \mathbf{y}_{M,n+L-1})^T$$

La matrice \mathbf{Y}_L est la somme de la matrice déterministe \mathbf{X}_L et de la matrice aléatoire \mathbf{V}_L , toutes deux définies comme \mathbf{Y}_L . De plus, du fait de la structure du signal $(\mathbf{x}_n)_{n \in \mathbb{Z}}$, la matrice \mathbf{X}_L peut se mettre sous la forme

$$\mathbf{X}_L = \mathbf{H}^{(L)} \mathbf{S}_L$$

où \mathbf{S}_L est la matrice de Hankel de dimensions $(P+L-1) \times N$ définie par $(\mathbf{S}_L)_{i,j} = s_{i+j-P}$ et où $\mathbf{H}^{(L)}$ est une matrice $ML \times (P+L-1)$ définie par $\mathbf{H}^{(L)} = (\mathbf{H}_1^{(L)T}, \dots, \mathbf{H}_M^{(L)T})^T$ où chaque bloc $\mathbf{H}_m^{(L)}$ est une matrice de Toeplitz $L \times (P+L-1)$ traduisant la convolution entre la séquence s et le filtre de fonction de transfert $h_m(z) = \sum_{p=0}^{P-1} \mathbf{h}_{m,p} z^{-p}$. Dans la suite, nous notons \mathbf{H} la matrice $\mathbf{H}^{(1)}$ qui est tout simplement la matrice $M \times P$ $\mathbf{H} = (\mathbf{h}_0, \dots, \mathbf{h}_{P-1})$. On peut donc constater

que \mathbf{Y}_L est la somme d'une matrice déterministe dont le rang est au plus égal à $P + L - 1$ et d'une matrice aléatoire bloc-Hankel constituée à partir d'un bruit gaussien, blanc temporellement et spatialement.

Les algorithmes classiques de détection/ estimation sont bien compris quand ML est très petit par rapport à N , ce que l'on modélise par le régime asymptotique ML fixe et $N \rightarrow +\infty$. La raison principale est que, dans ce contexte, la loi des grands nombres aboutit à

$$\left\| \frac{\mathbf{Y}_L \mathbf{Y}_L^*}{N} - \left(\frac{\mathbf{X}_L \mathbf{X}_L^*}{N} + \sigma^2 \mathbf{I}_{ML} \right) \right\| \rightarrow 0$$

Cependant, il arrive souvent que ML et N soient du même ordre de grandeur. Il est donc intéressant de s'intéresser aux régimes asymptotiques dans lesquels

$$M \rightarrow +\infty, N \rightarrow +\infty, \frac{ML}{N} \rightarrow c \tag{2.2}$$

avec $c > 0$ et d'étudier le comportement de $\frac{\mathbf{Y}_L \mathbf{Y}_L^*}{N}$ dans ce contexte. Il convient de noter que la condition (2.2) couvre différents types de régimes: L fixe et M et N du même ordre de grandeur, tandis que L peut tendre vers l'infini à toute vitesse strictement plus petite que N . Il convient de noter que nous n'envisageons par le cas où M est fixe, ou de façon équivalente, le cas où L et N tendent vers l'infini au même rythme. Dans ce contexte, les techniques à utiliser sont complètement différentes et nettement moins matures sur le plan mathématique.

3 Localisation des valeurs propres de $\frac{\mathbf{Y}_L \mathbf{Y}_L^*}{N}$.

Afin d'étudier les algorithmes de détection et d'estimation relatifs aux modèles large bande dans le régime (2.2), nous étudions les éléments propres de la matrice $\frac{\mathbf{Y}_L \mathbf{Y}_L^*}{N}$. Pour cela, notre démarche a été de considérer dans un premier temps le cas où le signal utile est absent, et d'utiliser une démarche proche de celle qui est utilisée dans [5] dans le cas où la matrice $\mathbf{X}_L = \mathbf{H}^{(L)} \mathbf{S}_L$ est de rang fixe, c'est-à-dire lorsque P et L restent fixes lorsque M et N croissent.

3.1 Localisation des valeurs propres de la matrice de covariance spatio-temporelle empirique en absence de signal.

Dans ce cadre, la matrice \mathbf{Y}_L est réduite à \mathbf{V}_L , de sorte qu'il convient d'étudier les valeurs propres de $\frac{\mathbf{V}_L \mathbf{V}_L^*}{N}$. Afin de permettre au lecteur de mieux connecter ce rapport avec l'article [9], nous utilisons les notations de [9]. Dans [9], la matrice étudiée, notée \mathbf{W}_L , est définie par

$$\mathbf{W}_L = \frac{1}{\sqrt{N}} \mathbf{V}_L \tag{3.3}$$

Dans [9], nous démontrons les résultats suivants:

- La distribution empirique des valeurs propres de $\mathbf{W}_L \mathbf{W}_L^*$ a le même comportement que si les éléments de \mathbf{W}_L étaient i.i.d., c'est-à-dire qu'elle converge presque sûrement vers la distribution de Marcenko-Pastur $\mu_{\sigma^2, c}$ de paramètres (σ^2, c) .
- Par ailleurs, si $L = \mathcal{O}(N^\alpha)$ avec $\alpha < 2/3$, alors, pour tout $\epsilon > 0$ aussi petit que l'on veut, presque sûrement, pour tout N assez grand, les $\min(ML, N)$ valeurs propres non nulles de $\mathbf{W}_L \mathbf{W}_L^*$ sont localisées dans l'intervalle $[\sigma^2(1 - \sqrt{c})^2 - \epsilon, \sigma^2(1 + \sqrt{c})^2 + \epsilon]$

Rappelons que la distribution de Marcenko-Pastur $\mu_{\sigma^2, c}$ ([10]) est la mesure de probabilité donnée par

$$d\mu_{\sigma^2, c}(x) = \delta_0[1 - c^{-1}]_+ + \frac{\sqrt{(x - x^-)(x^+ - x)}}{2\sigma^2 c \pi x} \mathbb{1}_{[x^-, x^+]}(x) dx$$

où $x^- = \sigma^2(1 - \sqrt{c})^2$ and $x^+ = \sigma^2(1 + \sqrt{c})^2$, et que de façon informelle, la convergence de la distribution empirique des valeurs de $\mathbf{W}_L \mathbf{W}_L^*$ vers $\mu_{\sigma^2, c}$ signifie que l'histogramme des valeurs propres de toute réalisation de $\mathbf{W}_L \mathbf{W}_L^*$ tend à se rapprocher du graphe de la densité de probabilité de $\mu_{\sigma^2, c}$. Ce résultat est valable dès que M tend vers l'infini,

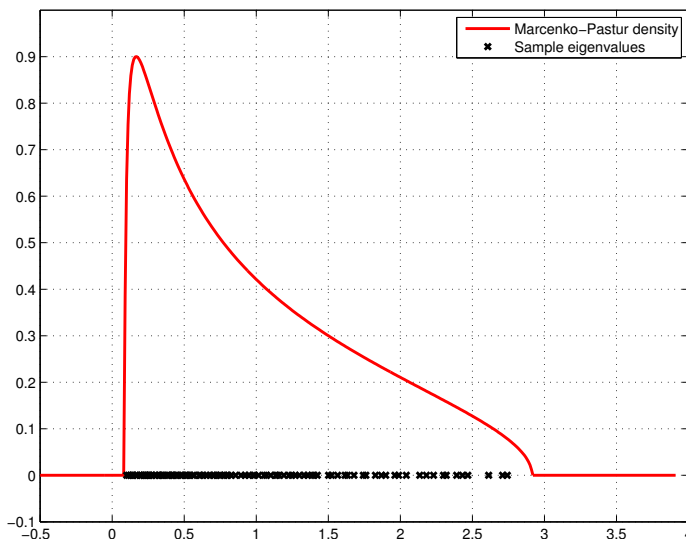


Figure 1: $M = 20, L = 5, L/M^2 = 0.0125, N = 200, c = 1/2, \sigma^2 = 1$

même si cette convergence est très lente par rapport à celle de N . Cette convergence n'est pas très étonnante au vu de publications récentes consacrées aux valeurs propres de certaines grandes matrices aléatoires structurées, par exemple des matrices hermitiennes dont les blocs sont des matrices de Toeplitz ([3]) indépendantes entre elles et construites à partir de variables aléatoires i.i.d.. La localisation des valeurs propres non nulles de $\mathbf{W}_L \mathbf{W}_L^*$ au voisinage du support de la densité de probabilité de $\mu_{\sigma^2, c}$ est également un résultat attendu, mais il semble qu'il ne soit pas valable si L tend vers l'infini trop vite par rapport à M . La condition $\alpha < 2/3$, qui correspond à $\frac{L}{M^2} \rightarrow 0$ apparaît en effet au cours de diverses étapes indépendantes de la preuve de [9], et il semble qu'elle ne puisse pas se généraliser à des valeurs de α plus grandes que $2/3$. En tout état de cause, les valeurs de M, N, L susceptibles d'être modélisées par la condition $L = \mathcal{O}(N^\alpha)$ avec $\alpha < 2/3$ nous semblent suffisamment diverses pour couvrir les applications usuelles dans le cas où le nombre de capteurs disponibles est grand. En pratique, pour de telles valeurs de M, N, L , le résultat de localisation que nous avons établi signifie que la probabilité que l'une des valeurs propres de $\mathbf{W}_L \mathbf{W}_L^*$ s'échappe du support de la distribution de Marcenko-Pastur est faible. Nous préférons ne pas évoquer les aspects techniques de la preuve des résultats annoncés, et nous contentons de mentionner que notre approche est basée sur les techniques développées dans [6] et [17] dans le cas $L = 1$.

Nous donnons quelques exemples numériques permettant d'illustrer la localisation des valeurs propres. La figure 1 représente les valeurs propres d'une réalisation de $\mathbf{W}_N \mathbf{W}_N^*$ dans le cas où $M = 20, L = 5, N = 200, c = 1/2, \sigma^2 = 1$. Dans ce scénario, le rapport L/M^2 est égal à $L/M^2 = 0.0125$, de sorte que les paramètres modélisent bien une situation dans laquelle $L/M^2 \simeq 0$. Avec ces paramètres, il est difficile de trouver une réalisation de $\mathbf{W}_L \mathbf{W}_L^*$ dans laquelle l'une des valeurs propres sort du support de la distribution de Marcenko-Pastur.

Dans la figure 2, les paramètres valent $M = 20, L = 60, N = 2400, c = 1/2, \sigma^2 = 1$, ce qui correspond à $L/M^2 = 0.15$ qui n'est pas précisément négligeable. Dans la réalisation choisie, toutes les valeurs propres restent dans le support, mais contrairement au contexte de la figure 3.1, il est difficile de trouver une telle réalisation. Ainsi, la figure 3, correspondant à un autre tirage de $\mathbf{W}_L \mathbf{W}_L^*$, fait apparaître des valeurs propres s'échappant du support de la distribution de Marcenko-Pastur.

Il convient de mentionner que si M ne tend pas l'infini, les résultats que nous avons établi dans [9] ne sont plus valables. Dans ce contexte, la distribution des valeurs propres de $\mathbf{W}_L \mathbf{W}_L^*$ ne converge pas vers la distribution de Marcenko-Pastur, mais vers une distribution à support non compact plus ou moins caractérisée par ses moments, voir

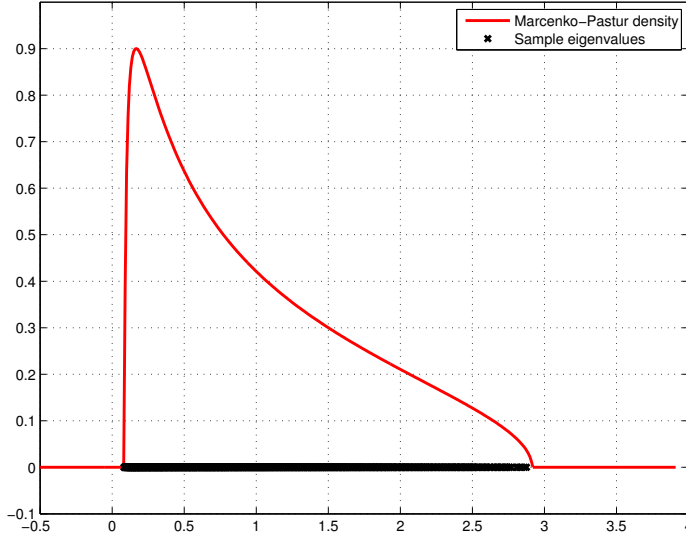


Figure 2: $M = 20, L/M^2 = 0.15, L = 60, N = 2400, c = 1/2, \sigma^2 = 1$

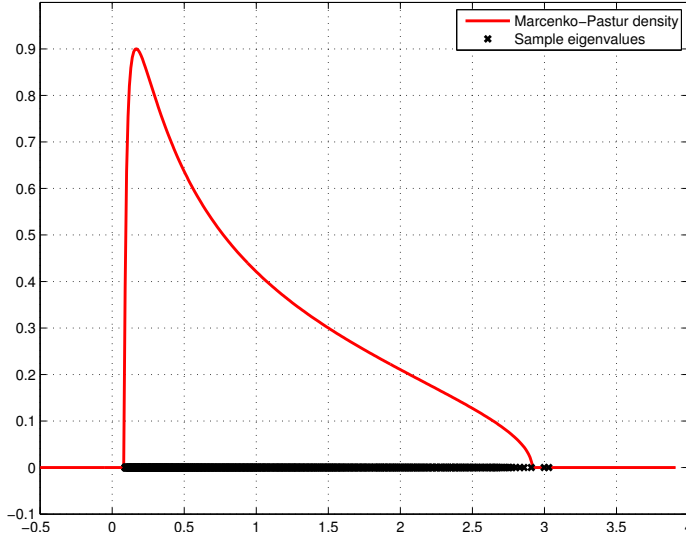


Figure 3: $M = 20, L/M^2 = 0.15, L = 60, N = 2400, c = 1/2, \sigma^2 = 1$

par exemple [2] qui a considéré le cas $M = 1$. Le comportement de la plus grande valeur propre de $\mathbf{W}_L \mathbf{W}_L^*$ n'est pas étudié dans [2], mais, au vu des résultats de [18] (qui concernent un modèle différent), il semble probable qu'elle tende vers ∞ à un rythme qui devrait être $\log N$. En pratique, la rupture entre le régime $M \rightarrow +\infty$ et celui dans lequel M reste fixe alors que L et N tendent vers l'infini au même rythme est progressive, et se matérialise par des histogrammes de valeurs propres de $\mathbf{W}_L \mathbf{W}_L^*$ s'ajustant de moins en moins bien à la densité de probabilité de la distribution de Marcenko-Pastur.

3.2 Influence de la présence d'une source dans le cas où P et L restent fixes.

Nous supposons à présent la présence d'une source large bande, et considérons le cas où P et L restent fixes quand M et N augmentent. Dans ce contexte, M et N convergent vers $+\infty$ au même rythme, et $\frac{ML}{N} \rightarrow c$ signifie que le rapport

$\frac{M}{N} \rightarrow d$ avec $c = dL$. Nous posons $\mathbf{\Sigma}_L = \frac{\mathbf{Y}_L}{\sqrt{N}}$ et $\mathbf{B}_L = \frac{\mathbf{X}_L}{\sqrt{N}}$, de sorte que

$$\mathbf{\Sigma}_L = \mathbf{B}_L + \mathbf{W}_L$$

où nous rappelons que \mathbf{W}_L est défini par (3.3). Nous désignons par $(\hat{\lambda}_{k,N}^{(L)})_{k=1,\dots,ML}$ les valeurs propres de $\mathbf{\Sigma}_L \mathbf{\Sigma}_L^*$ et par $\lambda_{1,N}^{(L)} \geq \lambda_{2,N}^{(L)} \dots \geq \lambda_{P+L-1,N}^{(L)}$ les valeurs propres non nulles de $\mathbf{B}_L \mathbf{B}_L^*$. Les résultats de [9], ainsi que quelques éléments techniques supplémentaires présentés dans [12], impliquent les valeurs propres de $\mathbf{\Sigma}_L \mathbf{\Sigma}_L^*$ se comportent comme si les éléments de la matrice \mathbf{W}_L étaient i.i.d. Plus précisément:

Théorème 1 *Supposons que:*

Hypothèse 1 *Les $P+L-1$ valeurs propres non nulles $(\lambda_{k,N}^{(L)})_{k=1,\dots,P+L-1}$ de la matrice $\mathbf{B}_L \mathbf{B}_L^*$ convergent vers $\lambda_1^{(L)} \geq \lambda_2^{(L)} \geq \dots \geq \lambda_{P+L-1}^{(L)}$ quand $N \rightarrow +\infty$.*

Désignons par K_L , $0 \leq K_L \leq P+L-1$, le plus grand entier pour lequel $\lambda_{K_L}^{(L)} > \sigma^2 \sqrt{dL}$. Alors, pour $k = 1, \dots, K_L$, il apparaît que

$$\hat{\lambda}_{k,N}^{(L)} \xrightarrow[N \rightarrow \infty]{a.s.} \rho_k^{(L)} = \phi_L(\lambda_k^{(L)}) = \frac{(\lambda_k^{(L)} + \sigma^2)(\lambda_k^{(L)} + \sigma^2 dL)}{\lambda_k^{(L)}} > \sigma^2(1 + \sqrt{dL})^2.$$

et que pour $k = K_L + 1, \dots, P+L-1$, alors $\hat{\lambda}_{k,N}^{(L)} \rightarrow \sigma^2(1 + \sqrt{dL})^2$ a.s.

4 Application à la détection d'une source large bande.

Dans cette section, nous résumons les travaux présentés dans [13] dont le but est d'étudier le comportement d'algorithmes de détection d'une source large bande. Plus précisément, nous reprenons le modèle (2.1), et souhaitons tester l'hypothèse H_0 définie par $\mathbf{y}_n = \mathbf{v}_n$ pour $n = 1, \dots, N$ contre l'hypothèse H_1 que \mathbf{y}_n contient une partie déterministe définie comme la sortie d'un filtre 1 entrée / M sorties inconnu, mais d'ordre maximum P connu, excité par une séquence scalaire non observable. Nous supposons que M et N sont grands, et que par contre, P , qui en pratique, peut représenter un nombre de trajets multiples, est nettement plus petit que M et N . Un tel contexte est motivé par le cadre des systèmes de communication de type "massive MIMO".

Le classique test du rapport de vraisemblance généralisé (GLRT) ne peut pas être implémenté dans ce cadre car sous l'hypothèse H_1 , il est impossible de caractériser de façon analytique les estimateurs au sens du maximum de vraisemblance des coefficients $(\mathbf{h}_p)_{p=0,\dots,P-1}$ et de la séquence $(s_n)_{n=1,\dots,N}$. Dans ces conditions, on ne peut que se contenter de proposer des statistiques de test plus empiriques. La façon la plus simple de procéder consiste à remarquer que le signal $\sum_{p=0}^{P-1} \mathbf{h}_p s_{n-p}$ peut être vu comme la somme des P signaux à bande étroite $(\mathbf{h}_p s_{n-p})_{n=1,\dots,N}$ pour $p = 0, \dots, P-1$, et utiliser le GLRT destiné à détecter P signaux bande étroite. Bien entendu, en faisant cela, on n'utilise pas la structure particulière des P signaux (les séquences scalaires correspondant aux différents signaux sont des versions retardées les unes des autres). Cette démarche conduit à comparer à un seuil la statistique η_N définie par

$$\eta_N = \sum_{i=1}^P \lambda_i \left(\frac{\mathbf{Y}\mathbf{Y}^*}{N} \right) \quad (4.4)$$

où $\mathbf{Y} = (\mathbf{y}_1, \dots, \mathbf{y}_N)$ et où $(\lambda_i \left(\frac{\mathbf{Y}\mathbf{Y}^*}{N} \right))_{i=1,\dots,M}$ représentent les valeurs propres de $\frac{\mathbf{Y}\mathbf{Y}^*}{N}$ ordonnées dans le sens décroissant. Il convient de noter que \mathbf{Y} correspond à la matrice \mathbf{Y}_1 correspondant au choix $L = 1$. Afin d'utiliser la structure convolutive du signal $\sum_{p=0}^{P-1} \mathbf{h}_p s_{n-p}$, il est tout à fait naturel de remplacer la matrice de covariance spatiale $\frac{\mathbf{Y}\mathbf{Y}^*}{N}$ par une matrice de covariance spatio-temporelle $\frac{\mathbf{Y}_L \mathbf{Y}_L^*}{N}$ pour une valeur bien choisie de L . Puisque en présence de signal, \mathbf{Y}_L est la somme d'une matrice déterministe de rang $P+L+1$ avec la matrice de bruit \mathbf{V}_L , il est raisonnable d'étudier le test consistant à comparer à un seuil la statistique $\eta_N^{(L)}$ définie par

$$\eta_N^{(L)} = \sum_{i=1}^{P+L-1} \lambda_i \left(\frac{\mathbf{Y}_L \mathbf{Y}_L^*}{N} \right) \quad (4.5)$$

Afin de bien comprendre les performances de ce type de test, nous avons étudié le comportement asymptotique au premier ordre de $\eta_N^{(L)}$ dans le régime asymptotique où M et N tendent vers $+\infty$ de telle sorte que $\frac{M}{N} \rightarrow d$. Pour cela, nous avons évidemment utilisé le Théorème 1, et réutilisons à partir de maintenant les notations qui y sont utilisées, i.e. $\Sigma_L = \frac{\mathbf{Y}_L}{\sqrt{N}}$ et $\mathbf{B}_L = \frac{\mathbf{X}_L}{\sqrt{N}}$. Puisque ce théorème suppose que les valeurs propres non nulles de $\frac{\mathbf{B}_L \mathbf{B}_L^*}{N}$ doivent converger vers des limites $\lambda_1^{(L)} \geq \lambda_2^{(L)} \geq \dots \geq \lambda_{P+L-1}^{(L)}$, nous formulons des hypothèses raisonnables sur les coefficients $(\mathbf{h}_0, \dots, \mathbf{h}_{P-1})$ et sur la séquence s_n permettant d'assurer un tel comportement. Nous rappelons que \mathbf{H} est la matrice $M \times P$ définie par $\mathbf{H} = (\mathbf{h}_0, \dots, \mathbf{h}_{P-1})$.

Hypothèse 2 • (i) Quand $N \rightarrow +\infty$, la matrice $\mathbf{H}^* \mathbf{H}$ converge vers une matrice $P \times P$ Δ

- (ii) Si $i, j \geq 1$, $\frac{1}{N} \sum_{n=1}^N s_{n+i-P} s_{n+j-P}^*$ converge vers une limite, qui nécessairement, ne dépend que de $i - j$, et qui est notée R_{i-j} .

L'hypothèse (i) est vérifiée dès que la matrice coïncide avec la réalisation d'une matrice aléatoire $M \times P$ Gaussienne raisonnable, normalisée de telle sorte que les éléments de $\mathbb{E}(\mathbf{H}^* \mathbf{H})$ restent finis. Le cas où Δ est une matrice diagonale correspond au cas où les P trajets sont décorrélés. L'hypothèse (ii) est vérifiée si la séquence s est une réalisation d'une suite aléatoire stationnaire, ce qui est tout à fait naturel. Dans la suite, nous notons $\mathbf{R}^{(L)}$ la matrice de Toeplitz $(P + L - 1) \times (P + L - 1)$ définie par $\mathbf{R}_{i,j}^{(L)} = R_{i-j}$.

Puisque les éléments de $\mathbf{H}^{(L)*} \mathbf{H}^{(L)}$ dépendent de ceux de $\mathbf{H}^* \mathbf{H}$, l'hypothèse (i) implique que la matrice $\mathbf{H}^{(L)*} \mathbf{H}^{(L)}$ converge vers une matrice $\Delta^{(L)}$ dont les éléments sont fonctions de ceux de Δ . Comme les valeurs propres non nulles de $\mathbf{B}_L \mathbf{B}_L^*$ coïncident avec les valeurs propres de $\mathbf{H}^{(L)*} \mathbf{H}^{(L)} \frac{\mathbf{S}^{(L)} \mathbf{S}^{(L)*}}{N}$, il est clair que l'hypothèse 2 implique que les conditions 1 sont vérifiées, et que $\lambda_k^{(L)} = \lambda_k \left(\Delta^{(L)} \mathbf{R}^{(L)} \right)$.

L'utilisation du Théorème 1 permet alors de montrer que $\eta_N^{(L)}$ converge presque sûrement vers la quantité $\eta^{(L)}$ définie par

$$\eta^{(L)} = \sum_{k=1}^{K_L} \rho_k^{(L)} + (P + L - 1 - K_L) \sigma^2 (1 + \sqrt{dL})^2 \quad (4.6)$$

L'analyse de cette expression permet de mettre en évidence quelques conclusions intéressantes, notamment concernant le choix du paramètre L . Si $K_L = 0$, c'est-à-dire quand toutes les valeurs propres limites de $\mathbf{B}_L \mathbf{B}_L^*$ sont inférieures au seuil de détectabilité $\sigma^2 \sqrt{dL}$, le test consistant à comparer $\eta_N^{(L)}$ à un seuil est non consistant, en ce sens que $\eta_N^{(L)}$ tend à avoir le même comportement asymptotique sous l'hypothèse H_0 et sous l'hypothèse H_1 . Dans le cas contraire, c'est-à-dire si

$$\frac{\lambda_1^{(L)}}{\sigma^2 \sqrt{dL}} > 1 \quad (4.7)$$

le test est consistant. Une façon de choisir L de façon optimale pourrait donc consister à trouver la valeur de L pour laquelle $\frac{\lambda_1^{(L)}}{\sigma^2 \sqrt{dL}}$ est maximum. La réponse n'est pas triviale puisque le dénominateur augmente avec L , et qu'il en est de même pour le numérateur car $\lambda_1^{(L)}$ tend à augmenter avec L car la matrice $\mathbf{B}_L \mathbf{B}_L^*$ est une sous-matrice de $\mathbf{B}_{L+1} \mathbf{B}_{L+1}^*$. Lorsque la matrice Δ est égale à $\delta \mathbf{I}$, c'est-à-dire que les P trajets sont indépendants et de même puissance δ , et que la suite s est asymptotiquement décorrélée, i.e. $\mathbf{R}^L = \mathbf{I}$ pour tout L , alors

$$\frac{\lambda_1^{(L)}}{\sigma^2 \sqrt{dL}} = \sqrt{L} \frac{\delta}{\sigma^2 \sqrt{d}}$$

si $L \leq P$, tandis que

$$\frac{\lambda_1^{(L)}}{\sigma^2 \sqrt{dL}} = \frac{P}{\sqrt{L}} \frac{\delta}{\sigma^2 \sqrt{d}}$$

si $L \geq P$. Par conséquent, lorsque les P trajets sont indépendants et de même puissance, la meilleure valeur de L est $L = P$, et permet de gagner, par rapport au cas où $L = 1$, un facteur \sqrt{P} sur le rapport signal minimal $\frac{\delta}{\sigma^2}$ qu'il faut avoir pour assurer la consistance du test. Lorsque les trajets sont indépendants et de puissances différentes, la valeur

optimale apparaît toujours inférieure ou égale à P . Par conséquent, il est tout à fait inutile de choisir des valeurs de L supérieures au nombre de trajets P .

Nous terminons en présentant les courbes ROC de ces tests. Pour construire ces courbes, nous générons un grand nombre de réalisations de $\eta_N^{(L)}$ sous l'hypothèse H_0 et sous l'hypothèse H_1 , et estimons empiriquement les points des courbes ROC. Dans la première expérience, $M = 80$, $N = 160$, $P = 5$, la suite s est une réalisation d'une suite de Bernoulli i.i.d, et la matrice \mathbf{H} est une réalisation d'une matrice Gaussienne à éléments i.i.d. (trajets indépendants de même puissance). Le rapport signal sur bruit est par ailleurs égal à 4 dB. Dans ce contexte, $L = 5$ est la valeur optimale. Nous représentons figure 4 les courbes ROC correspondant au choix de $L = 5$ (spatio-temporal) et $L = 1$ (spatial), ainsi que celles qui sont associées aux statistiques de test consistant à ne considérer que la plus grande valeur propre de la matrice $\frac{\mathbf{Y}_L \mathbf{Y}_L^*}{N}$ pour $L = 5$ (lmax s-t) et $L = 1$ (lmax spa). Les résultats confirment que choisir $L = 5$ permet d'augmenter les performances de façon considérable, et que les tests prenant en compte les $P + L - 1$ plus grandes valeurs propres de $\frac{\mathbf{Y}_L \mathbf{Y}_L^*}{N}$ sont préférables à ceux qui se contentent de ne considérer que la plus grande valeur propre.

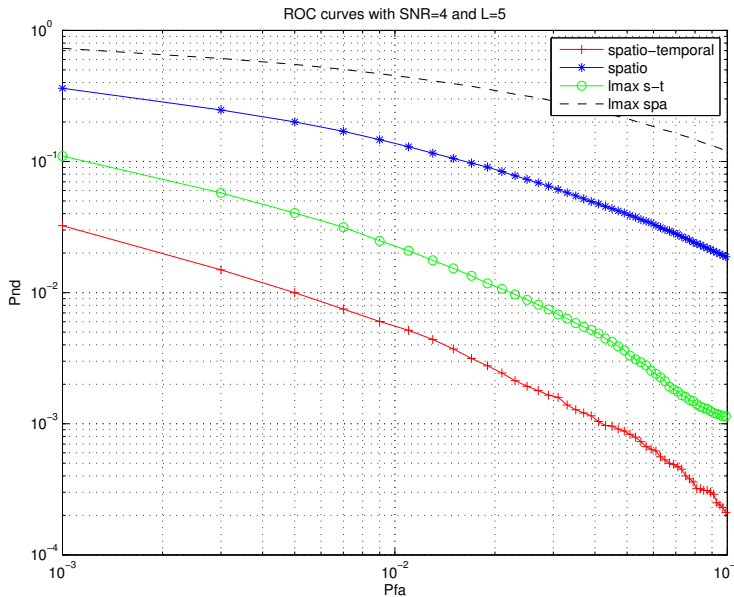


Figure 4: ROC curves of different statistics, $\Delta = \delta \mathbf{I}$

Comme le modèle de trajets indépendants n'est pas très réaliste dans le domaine des communications numériques, nous considérons un modèle de canal prenant en compte la présence d'un filtre de mise en forme à l'émetteur. P reste toujours égal à 5, mais le nombre de trajets indépendants est cette fois égal à 2. Les autres paramètres prennent les mêmes valeurs que dans le cadre de l'expérience précédente. Les conclusions que l'on peut tirer de la figure 5 diffèrent assez profondément. Les courbes spatio temporel et spatial coïncident, ce qui signifie que la valeur optimale de L n'est ni 1 ni 5. Par ailleurs, le test consistant à comparer la plus grande valeur propre de $\frac{\mathbf{Y}_L \mathbf{Y}_L^*}{N}$ pour $L = 5$ fournit de meilleures performances que le test prenant en compte ses $P + L - 1$ plus grandes valeurs propres, tandis que lorsque $L = 1$, la hiérarchie s'inverse. A notre sens, dans le contexte des communications numériques, la meilleure valeur de L devrait être égale au nombre de trajets physiques, et le meilleur test ne devrait conserver qu'un nombre réduit de valeurs propres.

5 Application à l'évaluation de filtres de Wiener spatio-temporels régularisés.

Nous nous plaçons toujours dans le cadre du modèle (2.1), avec M et N grands, qui modélise nombre de situations d'intérêt pratique dans lesquelles la suite $(s_n)_{n \in \mathbb{Z}}$, supposée i.i.d., représente une information à extraire à partir des obser-

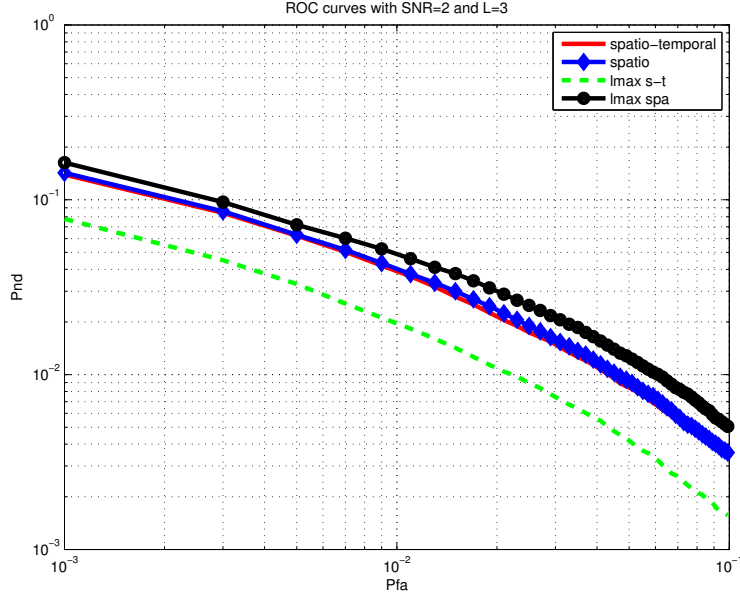


Figure 5: ROC curves of different statistics, Δ not diagonal

variations $(\mathbf{y}_n)_{n \in \mathbb{Z}}$. La façon la plus simple de procéder consiste à estimer s_n par une combinaison linéaire des composantes du vecteur $\mathbf{y}_n^{(L)} = (\mathbf{y}_{1,n}, \dots, \mathbf{y}_{1,n+L-1}, \dots, \mathbf{y}_{M,n}, \dots, \mathbf{y}_{M,n+L-1})^T$, c'est-à-dire par $(\mathbf{g}_w^{(L)})^* \mathbf{y}_n^{(L)}$ où $\mathbf{g}_w^{(L)}$ est le filtre de Wiener spatio-temporel de profondeur L défini par le fait que $\mathbb{E}|s_n - (\mathbf{g}_w^{(L)})^* \mathbf{y}_n^{(L)}|^2$ est minimum. L'expression de $\mathbf{g}_w^{(L)}$ dépendant des coefficients $(\mathbf{h}_p)_{p=0, \dots, P-1}$ qui sont inconnus, il est fréquent que dans une phase d'apprentissage préliminaire, les $(s_n)_{n=1, \dots, N}$ soient connus, et que grâce à cette connaissance, on estime $\mathbf{g}_w^{(L)}$ par le vecteur $\hat{\mathbf{g}}^{(L)}$ solution du problème des moindres carrés $\sum_{n=1}^N |\hat{\mathbf{g}}^{(L)*} \mathbf{y}_n^{(L)} - s_n|^2$, c'est-à-dire

$$\hat{\mathbf{g}}^{(L)} = \left(\frac{1}{N} \sum_{n=1}^N \mathbf{y}_n^{(L)} \mathbf{y}_n^{(L)*} \right)^{-1} \left(\frac{1}{N} \sum_{n=1}^N \mathbf{y}_n^{(L)} s_n^* \right) = \left(\frac{\mathbf{Y}_L \mathbf{Y}_L^*}{N} \right)^{-1} \left(\frac{1}{N} \sum_{n=1}^N \mathbf{y}_n^{(L)} s_n^* \right) \quad (5.8)$$

Cet estimateur n'a pas de sens si $ML > N$ car la matrice $\frac{\mathbf{Y}_L \mathbf{Y}_L^*}{N}$ n'est pas inversible dans ce contexte. D'autre part, même quand $ML < N$, il est bien connu que remplacer $\left(\frac{\mathbf{Y}_L \mathbf{Y}_L^*}{N} \right)^{-1}$ par l'inverse d'une version régularisée $\left(\frac{\mathbf{Y}_L \mathbf{Y}_L^*}{N} + \lambda \mathbf{I} \right)^{-1}$ a des effets souvent bénéfiques lorsque ML n'est pas très petit devant N , situation que nous considérons ici. Dans ce cas, on est donc conduit à estimer $\mathbf{g}_w^{(L)}$ par le vecteur $\hat{\mathbf{g}}_\lambda^{(L)}$ donné par

$$\hat{\mathbf{g}}_\lambda^{(L)} = \left(\frac{1}{N} \sum_{n=1}^N \mathbf{y}_n^{(L)} \mathbf{y}_n^{(L)*} + \lambda \mathbf{I}_{ML} \right)^{-1} \left(\frac{1}{N} \sum_{n=1}^N \mathbf{y}_n^{(L)} s_n^* \right) \quad (5.9)$$

L'estimateur $\hat{\mathbf{g}}_\lambda^{(L)}$ peut alors être utilisé pour estimer s_n par $(\hat{\mathbf{g}}_\lambda^{(L)})^* \mathbf{y}_n^{(L)}$ pour $n > N$. L'utilisation de cette approche pose évidemment le problème du choix du paramètre de régularisation λ . Pour cela, l'idéal serait de pouvoir choisir λ de telle façon que les performances de l'estimateur $\hat{\mathbf{g}}_\lambda^{(L)*} \mathbf{y}_n^{(L)}$ de s_n soit optimales. Dans le contexte des communications numériques, il est usuel de mesurer la qualité de l'estimateur par le rapport signal à interférence plus bruit qui est égal à

$$\text{SINR}(\hat{g}_\lambda^{(L)}) = \frac{|\hat{g}_\lambda^{(L)*} \mathbf{h}_P^{(L)}|^2}{\hat{g}_\lambda^{(L)*} \mathbf{H}_1^{(L)} \mathbf{H}_1^{(L)*} \hat{g}_\lambda^{(L)} + \sigma^2 \|\hat{g}_\lambda^{(L)}\|^2} \quad (5.10)$$

où $\mathbf{h}_P^{(L)}$ est la colonne P de $\mathbf{H}^{(L)}$ et $\mathbf{H}_1^{(L)}$ la matrice $\mathbf{H}^{(L)}$ privée de sa P -ième colonne. Il serait donc idéal de choisir λ de telle sorte que $\text{SINR}(\hat{g}_\lambda^{(L)})$ soit optimal. Cependant, les vecteurs $(\mathbf{h}_p)_{p=0, \dots, P-1}$ étant inconnus, la fonction $\lambda \rightarrow \text{SINR}(\hat{g}_\lambda^{(L)})$

l'est tout autant. Il serait évidemment possible de remplacer les vecteurs $(\mathbf{h}_p)_{p=0,\dots,P-1}$ par leurs estimateurs au sens des moindres carrés évalués grâce à la connaissance des $(s_n)_{n=1,\dots,N}$, mais dans le cas où M et N sont du même ordre de grandeur, ces estimateurs sont de très mauvaise qualité. Il s'avère que dans le régime asymptotique dans lequel M et N convergent vers $+\infty$ de telle sorte que $\frac{M}{N} \rightarrow d$, et où P et L restent fixes, $\text{SINR}(\hat{g}_\lambda^{(L)})$ qui est une variable aléatoire dépendant du bruit additif perturbant les $(\mathbf{y}_n)_{n=1,\dots,N}$, converge vers une quantité déterministe, notée $\phi(\lambda)$, que l'on peut dans un premier temps calculer analytiquement en fonction de σ^2 et des $(\mathbf{h}_p)_{p=0,\dots,P-1}$, et dans un deuxième temps, estimer de façon consistante à partir des $(\mathbf{y}_n)_{n=1,\dots,N}$. On dispose ainsi d'une technique permettant d'estimer pour chaque λ le SINR $\text{SINR}(\hat{g}_\lambda^{(L)})$, et de sélectionner la valeur de λ qui maximise l'estimateur.

Ces résultats sont présentés dans les articles [14] et [15]. Sans donner plus de détails techniques, nous mentionnons que le point de départ de notre travail repose sur le fait que si on appelle $\mathbf{Q}(z)$ la résolvante de $\frac{\mathbf{Y}_L \mathbf{Y}_L^*}{N}$ définie par

$$\mathbf{Q}(z) = \left(\frac{\mathbf{Y}_L \mathbf{Y}_L^*}{N} - z\mathbf{I} \right)^{-1}$$

alors le vecteur $\hat{\mathbf{g}}_\lambda^{(L)}$ se met sous la forme

$$\hat{\mathbf{g}}_\lambda^{(L)} = \mathbf{Q}(-\lambda) \frac{\mathbf{Y}_L \mathbf{s}^*}{N}$$

où $\mathbf{s} = (s_1, \dots, s_N)$. Tout repose alors sur le fait que en un certains sens, la résolvante a un comportement déterministe qui s'apparente à celui qu'elle aurait si la matrice \mathbf{Y}_L était la somme du terme déterministe $\mathbf{H}^{(L)} \mathbf{S}_L$ et d'une matrice aléatoire à éléments i.i.d., ce qui bien sur n'est pas le cas car \mathbf{V}_L est une matrice bloc Hankel. Ce résultat est donc dans la lignée du Théorème 1. Nous préférons ne pas donner plus de détails, et renvoyons le lecteur intéressé à [14] et [15].

Les preuves utilisées dans [14] et [15] reposent fondamentalement sur l'hypothèse que P et L restent fixes lorsque M et N tendent vers $+\infty$. Nous sommes en train de terminer un travail visant à établir que les résultats obtenus sont toujours valides si $L \rightarrow +\infty$, mais de telle façon que $\frac{L}{M} \rightarrow 0$. Nous pensons même que modulo l'introduction de termes correctifs supplémentaires, il devrait même être possible de se passer de cette hypothèse.

Pour terminer, nous donnons quelques résultats numériques permettant de se faire une idée de l'utilité des techniques de [14] et [15]. Nous allons tout d'abord évaluer dans quelle mesure l'approximation de $\text{SINR}(\hat{g}_\lambda^{(L)})$ par $\phi(\lambda)$ est raisonnable. Afin de générer $\mathbf{H} = (\mathbf{h}_0, \dots, \mathbf{h}_{P-1})$, nous considérons la matrice $(\mathbf{a}(\theta_1), \dots, \mathbf{a}(\theta_P))$, avec le vecteur $\mathbf{a}(\theta) = \frac{1}{\sqrt{M}}(1, \dots, e^{i(M-1)\theta})^T$ où les angles $(\theta_p)_{p=0,\dots,P-1}$ sont tirés aléatoirement sur $[0, 2\pi]$. La séquence $(s_n)_{n=1,\dots,N}$ est une réalisation d'une séquence i.i.d prenant des valeurs dans ± 1 avec la probabilité 1/2. Enfin, nous prenons $N = 200$, $M = 40$ et $P = 5$. Dans la figure 6, SNR est égal à 8 dB. Nous évaluons $\text{SINR}(\hat{g}_\lambda^{(L)})$ par simulations de Monte-Carlo, et représentons le graphe de la fonction $\lambda \rightarrow \phi(\lambda)$ ainsi que 2 courbes représentant les bornes inférieure et supérieure d'un intervalle de confiance à 95 pour cent de $\lambda \rightarrow \text{SINR}(\hat{g}_\lambda^{(L)})$. Ainsi que nous pouvons le constater, les 3 courbes sont relativement proches l'une de l'autre.

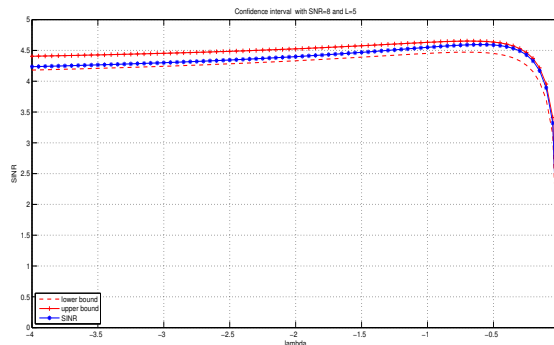


Figure 6: Intervalle de confiance sur $\phi(\lambda)$ versus λ

Ensuite, dans la figure 7, nous montrons à quel point il peut être avantageux de bien choisir le paramètre de régularisation λ en représentant le graphe de $\lambda \rightarrow \phi(\lambda)$ pour diverses valeurs de L .

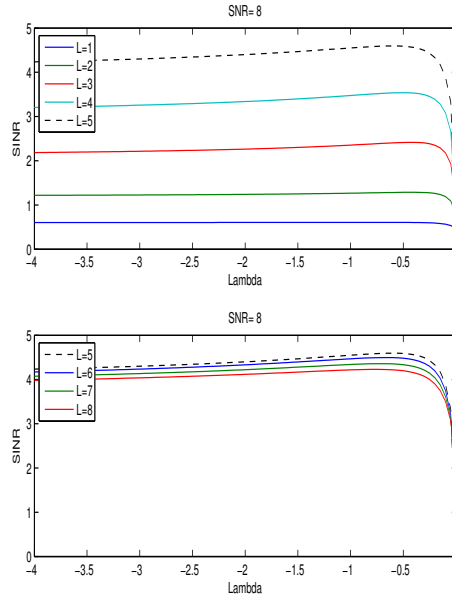


Figure 7: $\phi(\lambda)$ en fonction de L et λ

Nous nous intéressons à la qualité de la technique d'estimation du paramètre de régularisation λ . Pour cela, nous avons déterminé la valeur λ_{opt} de λ maximisant la fonction $\lambda \rightarrow \phi(\lambda)$, puis la comparons à son estimateur $\hat{\lambda}_{opt}$ maximisant l'estimateur $\hat{\phi}(\lambda)$ en évaluant la racine carrée de l'erreur quadratique moyenne sur $\phi(\hat{\lambda}_{opt}) - \phi(\lambda_{opt})$ par simulations de Monte-Carlo. Notons qu'évaluer l'erreur sur le SINR optimal est la façon la plus pertinente de qualifier les performances de ce genre de technique. Les résultats sont consignés dans le tableau 1 permettant de comparer pour divers choix de L l'approche proposée à des techniques empiriques de détermination de λ .

RMMSE	L=1	L=2	L=3	L=4	L=5	L=6	L=7	L=8	L=9	L=10
Proposed	0,0072	0,0116	0,0103	0,0118	0,0151	0,0178	0,0182	0,0189	0,0218	0,0208
NaiveEstimate	0,0038	0,0191	0,1201	0,1794	0,18	0,1833	0,1706	0,166	0,1664	0,1637
Ledoit-Wolf	0,9798	0,9678	0,9659	0,9686	0,9693	0,9665	0,9607	0,9539	0,9465	0,9403
M1	0,8596	0,79	0,7293	0,6731	0,5648	0,587	0,556	0,5217	0,5143	0,4744
M2	0,2326	0,4143	0,5486	0,6528	0,7	0,6882	0,6458	0,6038	0,585	0,5437
M3	0,6662	0,8386	0,8986	0,9296	0,9424	0,9436	0,9394	0,9337	0,9283	0,9237

Table 1: Erreur quadratique moyenne sur le SINR optimal

6 Application à l'analyse des performances des techniques de localisation de sources bande étroite utilisant la technique du lissage spatial.

Dans un premier temps, nous avons choisi d'appliquer le résultat précédent dans un contexte de sources bande étroite mais dans lequel les matrices de covariances empiriques ont des structures qui sont proches de celles qui viennent d'être évoquées. Les résultats qui vont être présentés ont fait l'objet de l'article [12]. Nous supposons que K sources bande étroite situées à l'infini se propagent, et que les signaux qu'elles produisent sont reçus par un réseau linéaire à capteurs

équidistants de M antennes, avec $K < M$. Dans ce contexte, le signal de dimension M $(\mathbf{y}_n)_{n \geq 1}$ reçu par le réseau de capteurs peut se mettre sous la forme

$$\mathbf{y}_n = \mathbf{A} \mathbf{s}_n + \mathbf{v}_n,$$

où $\mathbf{A} = [\mathbf{a}_M(\theta_1), \dots, \mathbf{a}_M(\theta_K)]$ est la matrice $M \times K$ collectant les K vecteurs directionnels de dimension M $\mathbf{a}_M(\theta_1), \dots, \mathbf{a}_M(\theta_K)$, où $\theta_1, \dots, \theta_K$ sont les angles d'arrivée des sources et où $\mathbf{a}_M(\theta) = \frac{1}{\sqrt{M}}[1, \dots, e^{i(M-1)\theta}]^T$. Les composantes du vecteur $\mathbf{s}_n \in \mathbb{C}^K$ coïncident avec les signaux sources reçus à l'instant n , supposés déterministes et non observables. $(\mathbf{v}_n)_{n \geq 1}$ est un bruit blanc gaussien complexe dont la matrice de covariance est $\mathbb{E}[\mathbf{v}_n \mathbf{v}_n^*] = \sigma^2 \mathbf{I}$. Nous supposons que les vecteurs \mathbf{y}_n sont observés entre les instants 1 et N , et nous nous intéressons spécifiquement à des contextes dans lesquels le nombre d'observations N est nettement plus petit que le nombre de capteurs M . Les classiques méthodes de type sous-espace ne fonctionnent plus, et il est d'usage d'utiliser une approche, qualifiée de lissage spatial, consistant à augmenter artificiellement le nombre d'observations (voir par exemple [19], [16]), et qui, à l'origine, avait été introduite pour améliorer les méthodes sous-espaces dans le case de signaux sources totalement corrélés. Si $L < M$, le lissage spatial consiste à former L sous réseaux d'antennes de $M - L + 1$ capteurs se recouvrant. A chaque instant n , l'observation disponible sur le réseau l est le vecteur $\mathbf{y}_n^{(l)} = (\mathbf{y}_{l,n}, \dots, \mathbf{y}_{M-L+1,n})^T$, de sorte que l'on fait apparaître NL observations de dimension $M - L + 1$, c'est-à-dire les vecteurs $(\mathbf{y}_n^{(l)})_{l=1, \dots, L, n=1, \dots, N}$. Ces observations sont collectées dans la matrice Hankel par bloc de dimensions $(M - L + 1) \times NL$ $\mathbf{Y}_N^{(L)}$ définie par

$$\mathbf{Y}_N^{(L)} = (\mathbf{y}_1^{(1)}, \dots, \mathbf{y}_1^{(L)}, \dots, \mathbf{y}_N^{(1)}, \dots, \mathbf{y}_N^{(L)})$$

Afin de bien comprendre la façon dont les signaux sources s'organisent au sein de $\mathbf{Y}_N^{(L)}$, nous définissons la matrice de Hankel de rang 1 $\mathcal{A}^{(L)}(\theta)$ donnée par

$$\mathcal{A}^{(L)}(\theta) = \sqrt{L(M-L+1)/M} \mathbf{a}_{M-L+1}(\theta) (\mathbf{a}_L(\theta))^T \quad (6.11)$$

et la matrice $(M - L + 1) \times KL$ de rang K $\mathbf{A}^{(L)}$ s'écrivant

$$\mathbf{A}^{(L)} = \left(\mathcal{A}^{(L)}(\theta_1), \mathcal{A}^{(L)}(\theta_2), \dots, \mathcal{A}^{(L)}(\theta_K) \right) \quad (6.12)$$

$\mathbf{Y}_N^{(L)}$ peut alors se mettre sous la forme

$$\mathbf{Y}_N^{(L)} = \mathbf{A}^{(L)} (\mathbf{S}_N \otimes \mathbf{I}_L) + \mathbf{V}_N^{(L)} \quad (6.13)$$

où \mathbf{S}_N est la matrice $K \times N$ $\mathbf{S}_N = (\mathbf{s}_1, \dots, \mathbf{s}_N)$ supposée de rang K , et où \otimes représente le produit de Kronecker. Il est facile de vérifier que $\mathbf{A}^{(L)} (\mathbf{S}_N \otimes \mathbf{I}_L)$ est de rang K , et que son espace image coïncide avec l'espace de dimension K engendré par les vecteurs directionnels de dimension $M - L + 1$ $\mathbf{a}_{M-L+1}(\theta_1), \dots, \mathbf{a}_{M-L+1}(\theta_K)$. Quand NL est suffisamment grand et que $M - L + 1$ est suffisamment petit, la loi des grands nombres permet de se convaincre que la matrice de covariance empirique $\mathbf{Y}_N^{(L)} \mathbf{Y}_N^{(L)*} / NL$ se comporte comme $\mathbf{A}^{(L)} (\mathbf{S} \mathbf{S}^* \otimes \mathbf{I}_L / NL) \mathbf{A}^{(L)*} + \sigma^2 \mathbf{I}_{M-L+1}$. On peut alors utiliser l'approche sous-espace pour estimer les $(\theta_k)_{k=1, \dots, K}$. Plus précisément, désignons par $\mathbf{\Pi}^{(L)}$ la matrice de projection orthogonale sur ce que l'on appelle l'espace bruit, c'est-à-dire le complément orthogonal de $\text{sp}\{\mathbf{a}_{M-L+1}(\theta_1), \dots, \mathbf{a}_{M-L+1}(\theta_K)\}$, qui coïncide aussi avec l'espace propre associé à la plus petite valeur propre (égale à σ^2) de $\mathbf{A}^{(L)} (\mathbf{S} \mathbf{S}^* \otimes \mathbf{I}_L / NL) \mathbf{A}^{(L)*} + \sigma^2 \mathbf{I}_{M-L+1}$. Soit $\eta(\theta)$ le pseudo-spectre spatial défini par

$$\eta(\theta) = \mathbf{a}_{M-L+1}(\theta)^* \mathbf{\Pi}^{(L)} \mathbf{a}_{M-L+1}(\theta)$$

Alors, les angles $(\theta_k)_{k=1, \dots, K}$ sont les solutions uniques de l'équation $\eta(\theta) = 0$. En pratique, la méthode sous-espace traditionnelle consiste à remplacer $\mathbf{\Pi}^{(L)}$ par la matrice de projection orthogonale $\hat{\mathbf{\Pi}}_N^{(L)}$ sur l'espace propre associé aux $M - L + 1 - K$ plus petites valeurs propres de $\mathbf{Y}_N^{(L)} \mathbf{Y}_N^{(L)*} / NL$, et à estimer les angles comme les K minimas les plus significatifs de la fonction $\theta \rightarrow \hat{\eta}_N^{(t)}(\theta)$ définie par

$$\hat{\eta}_N^{(t)}(\theta) = \mathbf{a}_{M-L+1}(\theta)^* \hat{\mathbf{\Pi}}_N^{(L)} \mathbf{a}_{M-L+1}(\theta)$$

Quand $NL \rightarrow +\infty$ et que $M - L + 1$ reste fixe, la loi des grands nombres implique que $\|\hat{\mathbf{\Pi}}_N^{(L)} - \mathbf{\Pi}^{(L)}\| \rightarrow 0$ et que

$$\sup_{\theta \in [-\pi, \pi]} \left| \hat{\eta}_N^{(t)}(\theta) - \eta(\theta) \right| \xrightarrow[N \rightarrow \infty]{a.s.} 0. \quad (6.14)$$

Cela implique alors que les estimateurs $(\hat{\theta}_{k,N}^{(t)})_{k=1,\dots,K}$ définis par

$$\hat{\theta}_{k,N}^{(t)} = \underset{\theta \in \mathcal{I}_k}{\text{argument}} \min |\hat{\eta}_N^{(t)}(\theta)|,$$

où les $(\mathcal{I}_k)_{k=1,\dots,K}$ sont des intervalles disjoints contenant les $(\theta_k)_{k=1,\dots,K}$, sont consistants, c'est-à-dire que

$$\hat{\theta}_{k,N}^{(t)} \xrightarrow[N \rightarrow \infty]{a.s.} \theta_k. \quad (6.15)$$

Quand le nombre d'antennes M est grand, le régime asymptotique dans lequel $NL \rightarrow +\infty$ et $M - L + 1$ fixe n'est pas très intéressant en pratique car il implique évidemment que l'antenne virtuelle associée à l'approche de lissage spatial est composée d'un nombre de capteurs nettement plus petit que M , ce qui limite considérablement la résolution de la méthode sous-espace correspondante. Nous étudions donc la méthode sous-espace dans le cas où les dimensions $M - L + 1$ et NL de la matrice $\mathbf{Y}_N^{(L)}$ sont du même ordre de grandeur et que $L \ll M$ afin que le schéma de lissage spatial ne conduise pas à une réduction significative du nombre de capteurs. Afin de formaliser cela, nous considérons le régime asymptotique suivant. Nous supposons que N et L dépendent de M et que

$$M \rightarrow +\infty, N = \mathcal{O}(M^\beta), 1/3 < \beta \leq 1, \frac{M - L + 1}{NL} \rightarrow c \quad (6.16)$$

où $0 < c < \infty$. Dans la suite, le symbole $N \rightarrow +\infty$ sera employé pour représenter le régime (6.16). Dans la suite, nous allons montrer que les résultats de [9] permettent de mettre en évidence des estimateurs de type sous-espace consistants des angles d'arrivée. Puisque $\frac{M-L+1}{NL} \rightarrow c$, il est clair que $\frac{L}{M} \rightarrow 0$ et que $L = \mathcal{O}(M^\alpha)$ où $\alpha = 1 - \beta$ vérifie $0 \leq \alpha < 2/3$. Par conséquent, le régime (6.16) est similaire à (2.2) si ce n'est qu'il convient d'échanger (M, N) avec $(N, M - L + 1)$ pour passer de (2.2) à (6.16). Par ailleurs, si \mathbf{W}_L est la matrice définie par (3.3), on peut constater que l'échange $(M, N) \rightarrow (N, M - L + 1)$ fait que les propriétés de la matrice $\sqrt{\frac{N}{ML}} \mathbf{W}_L^*$ sont identiques à celles de $\frac{1}{\sqrt{NL}} \mathbf{V}_N^{(L)}$. Dès lors, les résultats de [9] peuvent être utilisés pour analyser les propriétés des valeurs propres de $\frac{1}{NL} \mathbf{V}_N^{(L)} \mathbf{V}_N^{(L)*}$, qui, en gros se comportent comme si les éléments de $\mathbf{V}_N^{(L)}$ étaient des variables aléatoires i.i.d. En particulier, la distribution empirique des valeurs propres de $\frac{1}{NL} \mathbf{V}_N^{(L)} \mathbf{V}_N^{(L)*}$ converge vers la distribution de Marcenko-Pastur de paramètres (σ^2, c) , et les valeurs propres restent localisées au voisinage de $[\sigma^2(1 - \sqrt{c})^2, \sigma^2(1 + \sqrt{c})^2]$.

Il convient maintenant d'étudier le comportement des K plus grandes valeurs propres et vecteurs propres associés de la matrice de covariance empirique $\frac{1}{NL} \mathbf{Y}_N^{(L)} \mathbf{Y}_N^{(L)*}$ où d'après (6.13), $\mathbf{Y}_N^{(L)}$ peut être vue comme une perturbation déterministe de rang K de la matrice $\mathbf{V}_N^{(L)}$. Si $\mathbf{V}_N^{(L)}$ était à éléments i.i.d., les résultats de [5] permettraient de caractériser le comportement des K plus grandes valeurs propres et vecteurs propres de $\frac{1}{NL} \mathbf{Y}_N^{(L)} \mathbf{Y}_N^{(L)*}$. On peut alors établir que les résultats de [9], quelque peu enrichis, impliquent que les K plus grandes valeurs propres et vecteurs propres de $\frac{1}{NL} \mathbf{Y}_N^{(L)} \mathbf{Y}_N^{(L)*}$ se comportent également comme si $\mathbf{V}_N^{(L)}$ était à éléments i.i.d. En particulier, on peut établir le résultat suivant.

Théorème 2 *Supposons que les K valeurs propres non nulles $(\lambda_{k,N})_{k=1,\dots,K}$ de $\frac{1}{NL} \mathbf{A}^{(L)} (\mathbf{S}_N \mathbf{S}_N^* \otimes \mathbf{I}_L) \mathbf{A}^{(L)*}$ convergent vers des quantités $\lambda_1 > \lambda_2 > \dots > \lambda_K$. Désignons par ailleurs par $(\hat{\lambda}_{k,N})_{k=1,\dots,M-L+1}$ et $(\hat{\mathbf{u}}_{k,N})_{k=1,\dots,M-L+1}$ les valeurs propres (ordonnées par ordre décroissant) et vecteurs propres associés de $\frac{1}{NL} \mathbf{Y}_N^{(L)} \mathbf{Y}_N^{(L)*}$. Soit s , $0 \leq s \leq K$, le plus grand entier pour lequel $\lambda_s > \sigma^2 \sqrt{c}$. Alors, les s plus grandes valeurs propres $(\hat{\lambda}_{k,N})_{k=1,\dots,s}$ vérifient*

$$\hat{\lambda}_{k,N} \xrightarrow[N \rightarrow \infty]{a.s.} \rho_k = \phi(\lambda_k) = \frac{(\lambda_k + \sigma^2)(\lambda_k + \sigma^2 c)}{\lambda_k} > x^+ = \sigma^2(1 + \sqrt{c})^2$$

tandis que les $(\hat{\lambda}_{k,N})_{k=s+1,\dots,K}$ convergent presque sûrement vers x^+ . De plus, pour tous vecteurs déterministes de norme 1 (\mathbf{a}_N) et (\mathbf{b}_N) , pour $k = 1, \dots, s$,

$$\mathbf{a}_N^* (\hat{\mathbf{u}}_{k,N} \hat{\mathbf{u}}_{k,N}^* - h(\rho_k) \mathbf{u}_{k,N} \mathbf{u}_{k,N}^*) \mathbf{b}_N \rightarrow 0 \text{ a.s.} \quad (6.17)$$

où la fonction $h(z)$ apparaissant dans (6.17) dépend de la transformée de Stieltjes de la distribution de Marcenko-Pastur $\mu_{\sigma^2,c}$ et a une expression analytique connue.

Le Théorème 2 permet de montrer que la méthode sous-espace développée dans le cas $L = 1$ dans [7] continue à fonctionner dans le contexte du régime asymptotique (6.16). Plus précisément:

Théorème 3 *Si la condition de séparation*

$$\lambda_K > \sigma^2 \sqrt{c} \quad (6.18)$$

est vérifiée, l'estimateur $\hat{\eta}_N(\theta)$ défini par

$$\hat{\eta}_N(\theta) = (\mathbf{a}_{M-L+1}(\theta))^* \left(\mathbf{I} - \sum_{k=1}^K \frac{1}{h(\hat{\lambda}_{k,N})} \hat{\mathbf{u}}_{k,N} \hat{\mathbf{u}}_{k,N}^* \right) \mathbf{a}_{M-L+1}(\theta). \quad (6.19)$$

vérifie

$$\sup_{\theta \in [-\pi, \pi]} |\hat{\eta}_N(\theta) - \eta_N(\theta)| \xrightarrow[N \rightarrow \infty]{a.s.} 0, \quad (6.20)$$

Enfin, les estimateurs des angles $(\hat{\theta}_{k,N})_{k=1,\dots,K}$ obtenus en minimisant $\hat{\eta}_N(\theta)$ sont consistants, et ils vérifient

$$M(\hat{\theta}_{k,N} - \theta_k) \rightarrow 0 \text{ a.s.} \quad (6.21)$$

Ce théorème montre que si la condition (6.18) est vérifiée, alors on peut estimer les angles d'arrivées par méthode sous-espace avec une vitesse de convergence meilleure que $\frac{1}{M}$ dans le cas où le nombre de capteurs M est grand et où le nombre d'observations N est très inférieur à M tout en étant suffisamment grand. En pratique, pour M, N, L finis, la condition (6.18) signifie que $\lambda_{K,N}$ doit suffisamment éloigné de $\sigma^2 \sqrt{c_N}$ où c_N est défini par $c_N = \frac{M-L+1}{NL}$, et cela doit être interprété comme un critère permettant de bien distinguer les valeurs propres dues au signal utile des valeurs propres dues au bruit. Il est évidemment intéressant de rendre la condition (6.18) plus explicite. Pour cela, nous supposons que $\frac{\mathbf{S}_N \mathbf{S}_N^*}{N}$ converge une matrice diagonale \mathbf{D} . Les éléments diagonaux $(d_k)_{k=1,\dots,K}$ de \mathbf{D} représentent les puissances des différents signaux source. Dans ce cas, si l'on désigne par \mathbf{A}_{M-L+1} la matrice $\mathbf{A}_{M-L+1} = (\mathbf{a}_{M-L+1}(\theta_1), \dots, \mathbf{a}_{M-L+1}(\theta_K))$, il est facile de voir que la matrice $\frac{1}{NL} \mathbf{A}^{(L)} (\mathbf{S}_N \mathbf{S}_N^* \otimes \mathbf{I}_L) \mathbf{A}^{(L)*}$ se comporte comme $\mathbf{A}_{M-L+1} \mathbf{D} \mathbf{A}_{M-L+1}^*$. Si l'on suppose que les angles θ_k restent fixes quand M, N, L augmentent, la matrice $\mathbf{A}_{M-L+1}^* \mathbf{A}_{M-L+1} \rightarrow \mathbf{I}_K$, et les valeurs propres de $\frac{1}{NL} \mathbf{A}^{(L)} (\mathbf{S}_N \mathbf{S}_N^* \otimes \mathbf{I}_L) \mathbf{A}^{(L)*}$ convergent vers les $(d_k)_{k=1,\dots,K}$. Dans ces conditions, la condition de séparation (6.18) signifie que les puissances de toutes les sources dépassent le seuil de "détection" $\sigma^2 \sqrt{c}$. Pour des valeurs finies de M, N, L , ceci veut dire que pour tout k ,

$$\frac{1}{N} \sum_{n=1}^N |s_{k,n}|^2 > \sigma^2 \sqrt{c_N} \quad (6.22)$$

où $(s_{k,n})_{n=1,\dots,N}$ est le signal transmis par la source k . Dans le cas où L reste fini, ou de façon équivalente, que M et N sont du même ordre de grandeur, il est intéressant d'étudier la façon dont le seuil de détection $\sigma^2 \sqrt{c_N}$ évolue quand L varie. c_N est égal à $\frac{M-L+1}{NL}$ et se comporte donc comme $\frac{1}{L} \frac{M}{N}$ quand M et N sont suffisamment grands par rapport à L . Dès lors, utiliser une technique de lissage spatial avec L réseaux d'antennes permet d'abaisser le seuil de détection du facteur $\frac{1}{\sqrt{L}}$. Cette analyse n'est cependant justifiée que si L reste très inférieur à M et N . Si L augmente trop, la diminution de $M - L + 1$ va impliquer que la matrice $\mathbf{A}_{M-L+1}^* \mathbf{A}_{M-L+1}$ s'éloigne de l'identité, que la plus petite valeur propre de $\frac{1}{NL} \mathbf{A}^{(L)} (\mathbf{S}_N \mathbf{S}_N^* \otimes \mathbf{I}_L) \mathbf{A}^{(L)*}$ va diminuer et que la condition (6.22) n'est plus une bonne approximation de la condition (6.18). En d'autres termes, on peut s'attendre à ce que les performances des estimateurs commencent à

s'améliorer quand L augmente, mais qu'à partir d'une certaine valeur de L , elles se dégradent.

Afin d'illustrer ces résultats théoriques, nous présentons des résultats de simulations numériques. Nous considérons le cas de $K = 2$ sources dont les angles sont séparés de $\frac{\pi}{2M}$, et supposons que $M = 160$ et $N = 20$. La matrice $2 \times N$ des signaux sources est la réalisation d'une matrice aléatoire dont les éléments sont i.i.d. et de loi gaussienne complexe centrée de variance 1. Les signaux sources sont normalisés de telle façon que leur puissance moyenne soit égale à 1. Ainsi, le rapport signal sur bruit est égal à $\text{SNR} = 1/\sigma^2$. Le tableau 2 met en évidence la valeur minimale du SNR pour laquelle la condition de séparation, dans sa version à M, N, L finie, est vérifiée, i.e.

$$(\sigma^2)^{-1} = \frac{1}{\lambda_{K,N}} \sqrt{(M - L + 1)/NL}$$

Ainsi que cela a été expliqué, quand L augmente, $\sqrt{(M - L + 1)/NL}$ décroît, mais $\lambda_{K,N}$ tend à diminuer car $M - L + 1$ diminue. Cela explique pourquoi le SNR minimal décroît, puis croît quand L augmente.

L	2	4	8	16	32	64	96	128
SNR	33.46	30.30	27.46	25.31	24.70	28.25	36.11	51.52

Table 2: Valeur minimale du SNR assurant la condition de séparation.

La figure 8 représente l'erreur quadratique moyenne de l'estimateur $\hat{\theta}_1$ (appelé G-MUSIC SS) pour $L = 2, 4, 8, 16$ versus SNR. La borne de Cramer-Rao est également représentée. Comme cela était prévisible, les performances tendent à s'améliorer quand L augmente jusqu'à $L = 16$.

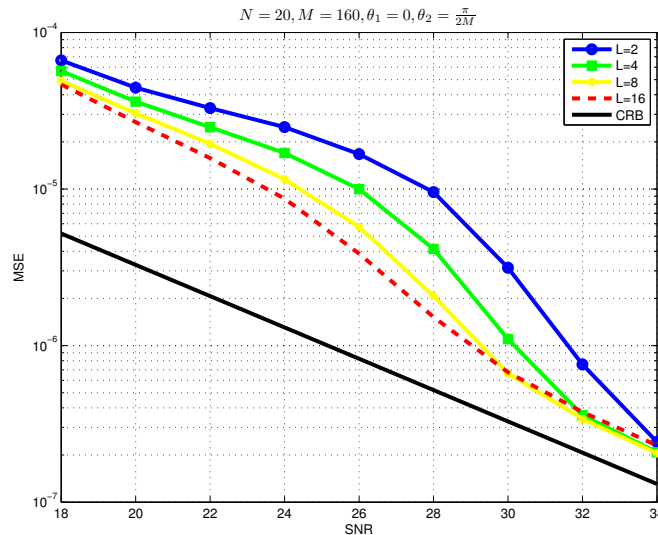


Figure 8: MSE de l'estimateur G-MUSIC SS $\hat{\theta}_1$ versus SNR

Sur la figure 9, L est égal à 16, 32, 64, 96, 128. Pour $L = 32$, l'erreur quadratique moyenne tend à se dégrader à haut SNR par rapport au cas $L = 16$, alors que la perte de performance devient sévère pour les plus grandes valeurs de L .

Sur la figure 10, L est fixé à 16. Nous comparons les performances de G-MUSIC SS avec celles de l'algorithme MUSIC standard avec lissage spatial, et représentons également les performances obtenues sans lissage spatial, c'est-à-dire quand $L = 1$. Il est clair que l'algorithme G-MUSIC SS permet une très nette amélioration des performances par rapport aux

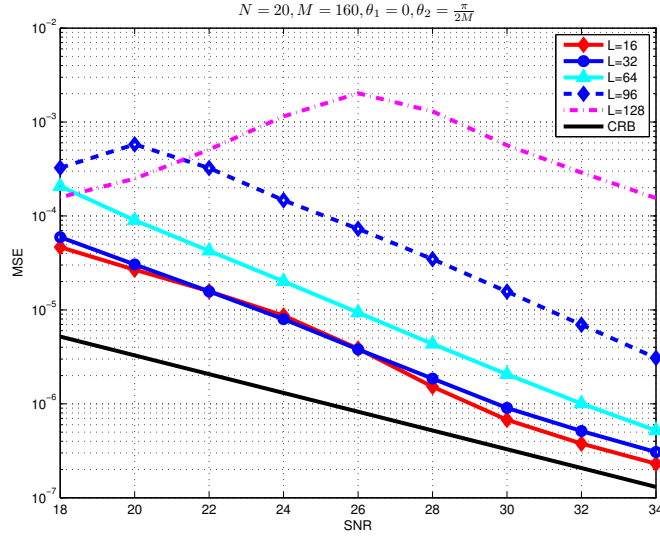


Figure 9: MSE de l'estimateur G-MUSIC SS $\hat{\theta}_1$ $\hat{\theta}_1$ versus SNR

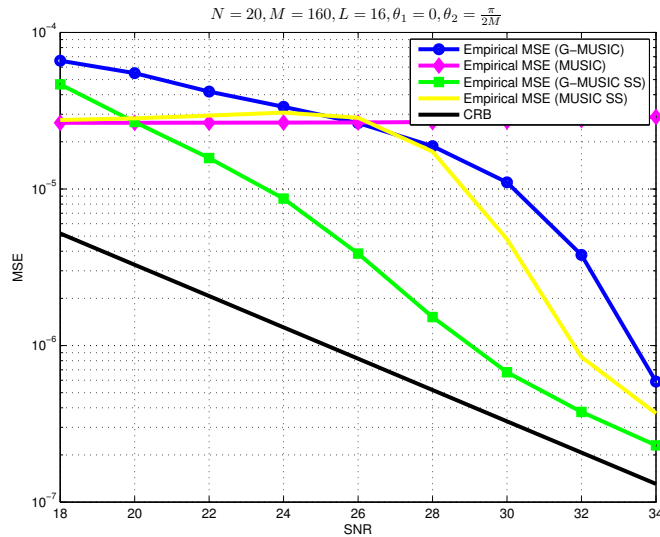


Figure 10: Empirical MSE of different estimators of θ_1 when $L=16$

autres.

Nous considérons finalement le cas $L = 128$ dans la figure 11, et comparons comme plus haut G-MUSIC SS, MUSIC SS, G-MUSIC et MUSIC pour $L = 1$. G-MUSIC SS a des performances catastrophiques car L et M sont du même ordre de grandeur, une situation dans laquelle le Théorème 3 n'est plus valide.

Références

- [1] K. Abed-Meraim, E. Moulines, Ph. Loubaton, "Prediction error method for second-order blind identification", *IEEE Trans. on Signal Processing*, vol. 45, no. 3, pp. 694-705, March 1997.
- [2] A. Basak, A. Bose, S. Sen, "Limiting spectral distribution of sample autocovariance matrices", to appear in *Bernoulli*, can be downloaded on Arxiv, arXiv:1108.3147v1.

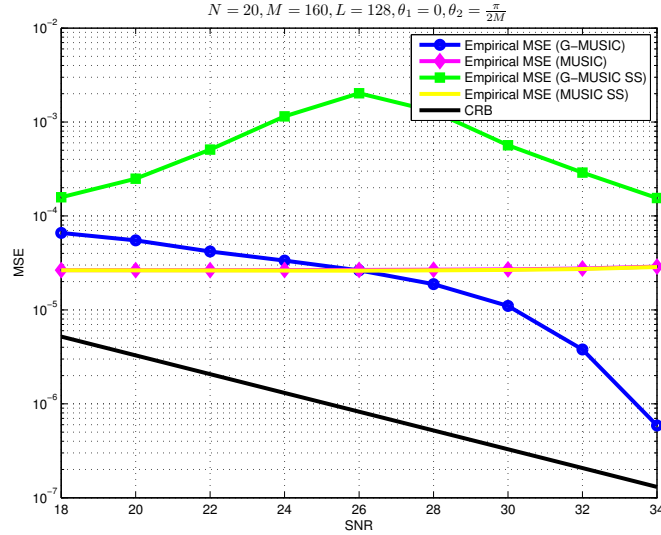


Figure 11: MSE de différents estimateur de θ_1 when $L=128$

- [3] R. Basu, A. Bose, S. Ganguly, R.S. Hazra, "Limiting spectral distribution of block matrices with Toeplitz block structure", *Statist. and Probab. Lett.*, 82 (2012), no. 7, 1430-1438.
- [4] F. Benaych-Georges, R.R. Nadakuditi, "The eigenvalues and eigenvectors of finite, low rank perturbations of large random matrices", *Adv. in Math.*, vol. 227, no. 1 (2011), 494–521.
- [5] F. Benaych-Georges, R.R. Nadakuditi, "The singular values and vectors of low rank perturbations of large rectangular random matrices", *J. Multivariate Anal.*, Vol. 111 (2012), 120–135.
- [6] U. Haagerup and S. Thorbjørnsen. A new application of random matrices: $\text{Ext}(C_{\text{red}}^*(F_2))$ is not a group. *Ann. of Math. (2)*, 162(2):711–775, 2005.
- [7] Hachem, W. and Loubaton, P. and Mestre, X. and Najim, J. and Vallet, P., "A Subspace Estimator for Fixed Rank Perturbations of Large Random Matrices", *Journal of Multivariate Analysis*, vol. 114, pp. 427-447, 2012.
- [8] S. Kay, "Modern Spectral Estimation: Theory and Application", Prentice Hall, 1988.
- [9] P. Loubaton, "On the almost sure location of the singular values of certain Gaussian block-Hankel large random matrices", arXiv:1405.2006.
- [10] V. A. Marčenko and L. A. Pastur. Distribution of eigenvalues in certain sets of random matrices. *Mat. Sb. (N.S.)*, 72 (114):507–536, 1967.
- [11] E. Moulines, P. Duhamel, J.F. Cardoso, S. Mayrargue, "Subspace methods for blind identification of multichannel FIR filters", *IEEE Trans. on Signal Processing*, vol. 43, pp. 516-525, February 1995.
- [12] G.T. Pham, P. Loubaton, P. Vallet, "Performance analysis of spatial smoothing schemes in the context of large arrays", *IEEE Trans. on Signal Processing*, vol. 64, no. 1, pp. 160-172, January 2016, also available on Arxiv (arXiv:1503.08196).
- [13] G.T. Pham; P. Loubaton, "Applications of large empirical spatio-temporal covariance matrix in multipath channels detection", Signal Processing Conference (EUSIPCO), Nice, September 2015.
- [14] G.T. Pham; P. Loubaton, "Performances des filtres de Wiener spatio-temporels entraînés: le cas des grandes dimensions", Proc. Colloque Gretsi, Lyon, September 2015.

- [15] G.T. Pham; P. Loubaton, "Optimization of the loading factor of regularized estimated spatial-temporal Wiener filters in large system case", Proc. of Statistical Signal Processing Workshop (SSP), Palma de Majorque, June 2016.
- [16] S.U. Pillai, B.H. Kwon, "Performance analysis of MUSIC type high resolution estimators for direction finding in correlated and coherent scenes", *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 37, no. 8, pp. 1176-1189, 1989.
- [17] H. Schultz, "Non commutative polynomials of independent Gaussian random matrices", *Probab. Theory Relat. Fields* 131, 261-309 (2005)
- [18] A. Sen, B. Virág, "The top eigenvalue of the random Toeplitz matrix and the sine kernel", *Ann. Probab.*, vol. 41, no. 6, pp. 4050-4079, 2013.
- [19] T.J. Shan, M. Wax, T. Kailath, "On spatial smoothing for estimation of coherent signals", *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 33, no. 8, pp. 806-811, 1985.
- [20] A.J. Van der Veen, S. Talwar, A. Paulraj, "A subspace approach to blind space-time signal processing for wireless communication systems", *IEEE Trans. on Signal Processing*, vol. 45, no. 1, January 1997.
- [21] A.J. Van der Veen, M. Vanderveen, A. Paulraj, "Joint angle and delay estimation using shift-invariant techniques", *IEEE Trans. on Signal Processing*, vol. 46, no. 2, pp. 405-418, February 1998.