

# Projet ANR-12-MONU-0003 DIONISOS.

## Premier Rapport: Analyse de la bibliographie

### 1 Introduction.

Le but de ce rapport est de mettre en évidence des résultats déjà publiés qui permettent de faire progresser les travaux menés dans le projet DIONISOS. Les points qui sont abordés concernent:

- L'étude du support des distributions limites des valeurs propres de modèles de matrices aléatoires apparaissant dans la tâche 2-1. La connaissance de ce support est crucial pour pouvoir concevoir des techniques d'estimation consistantes en grande dimension.
- L'étude du comportement des valeurs propres de matrices utilisées pour estimer les tenseurs cumulants d'ordre 4 d'un signal défini comme un mélange linéaire bruité de signaux bande étroite. Ce problème apparaît dans la tâche 2-2 qui concerne la séparation aveugle de sources bande étroite en grande dimension.
- L'étude du comportement des valeurs propres de la matrice de covariance spatio-temporelle empirique d'un signal multivariable défini comme la sortie bruitée d'un système linéaire 1 entrée / multi-sorties. Ce problème apparaît dans la tâche 3 qui concerne des problèmes de détection et d'estimation mettant en jeu des signaux large bande.
- L'étude d'estimateurs consistants de grandes matrices de covariance sur lesquelles des a priori tels que une certaine forme de parcimonie ou la structure (Toeplitz) sont disponibles. Les résultats correspondants sont utiles dans la tâche 2-1, et devraient pouvoir être adaptés dans le cadre de la tâche 3 dans laquelle nous nous poserons le problème d'estimer de façon consistante de grandes matrices de covariances spatio-temporelles.
- Enfin, la problématique de l'estimation robuste est présentée de façon relativement détaillée de façon à faire apparaître les problèmes nouveaux qui se posent dans le contexte des grandes dimensions, objet, entre autres, de la tâche 2-1.

D'une façon générale, nous mettons en évidence dans chaque rubrique le type de résultat qu'il nous faudra établir dans le futur afin de résoudre tout ou partie de nos problèmes.

### 2 Analyse du support des distributions limites.

When the limit spectral measure of a large random matrix exists, it is most frequently characterized by the fact that its Stieltjes Transform is the unique solution of an implicit equation. For some very particular matrix ensembles, it is possible to solve this equation and to invert the Stieltjes Transform to obtain a closed form expression of this limit spectral measure. However, when the matrix ensemble involves non trivial correlations between the matrix elements, a non trivial variance field, or a non trivial mean field, it is generally impossible to explicitly solve the equation characterizing the Stieltjes transform. Yet, by closely analyzing this equation, it is often possible to establish the properties of the limit spectral measure in terms of the support, the existence of a density, the behaviour at the edges of the support, and so forth. Some of the main elements of this analysis can be already found in Marchenko and Pastur's seminal paper [37]. Detailed studies were made for some popular matrix ensembles in [43, 19, 44, 34].

Having determined the support of the limit spectral measure, another interesting research direction consists in analyzing the behaviour of the individual eigenvalues with respect to this support in the large dimensional regime (see the works of Bai, Silverstein and Paul [3, 4, 41, 5], Vallet, Loubaton and Mestre [44], Capitaine, Donati-Matrin and Féral [18] relying on the results of Haagerup and Thorbjörnsen [32], and Male [36] among others.

These analyses paved the way to some very interesting applications in terms of the estimation theory. Initiated by the works of Girko [27, 28, 29], these applications were introduced and developed in the field of signal processing by Mestre [39, 38].

In this document, we shall summarize these results for two matrix ensembles frequently considered in the literature: the so called “one sided covariance” ensemble analyzed in [43], and the “information plus noise” ensemble studied in [19] and [44]. We shall also say a few words on “double sided covariance” ensemble under study in this project.

## 2.1 The one sided covariance ensemble

### 2.1.1 Limit spectral measure: existence and properties

Consider the  $N \times n$  complex random matrix  $\Sigma_n = D_n^{1/2} X_n$  where  $X_n$  is a  $N \times n$  matrix with independent and identically distributed (iid) elements having the mean zero and the variance  $1/n$ , and  $D_n$  is a  $N \times N$  deterministic nonnegative matrix, the matrix  $D_n^{1/2}$  being any square root of  $D_n$ . In multivariate analysis, the columns of  $\sqrt{n}\Sigma_n$  are iid snapshots of a vector random variable with mean zero and covariance matrix  $D_n$ .

We assume that  $n \rightarrow \infty$  and that  $N/n \rightarrow c > 0$ , and we denote this asymptotic regime as “ $n \rightarrow \infty$ ”. We also assume that the spectral measure of  $D_n$  converges weakly towards a probability measure  $\nu \neq \delta_0$  as  $n \rightarrow \infty$ , where  $\delta_x$  is the Dirac measure at  $\{x\}$ . It is well known that the (random) spectral measure of  $\Sigma_n \Sigma_n^*$  converges to a deterministic probability measure  $\mu$  in the sense that for any continuous and bounded function  $\varphi : [0, \infty) \rightarrow \mathbb{R}$ ,

$$\frac{1}{N} \sum_{\ell=0}^{N-1} \varphi(\lambda_{\ell,n}) \xrightarrow[n \rightarrow \infty]{\text{a.s.}} \int \varphi(\lambda) \mu(d\lambda) \quad (2.1)$$

where  $\lambda_{0,n}, \dots, \lambda_{N-1,n}$  are the eigenvalues of  $\Sigma_n \Sigma_n^*$ . The measure  $\mu$  is characterized by its Stieltjes Transform: given a finite positive measure  $\rho$ , this is the complex function

$$f(z) = \int \frac{1}{\lambda - z} \rho(d\lambda)$$

defined and analytic on  $\mathbb{C} - \text{support}(\rho)$ . One can straightforwardly observe that  $f(\mathbb{C}_+) \subset \mathbb{C}_+$  where  $\mathbb{C}_+ = \{z : \Im(z) > 0\}$ .

The so called Perron inversion formula shows that the measure  $\rho$  is uniquely determined by its Stieltjes Transform.

In our case, we have the following result: for each  $z \in \mathbb{C}_+$ , the equation

$$\tilde{m} = \frac{1}{-z + c \int \frac{t}{1 + \tilde{m}t} \nu(dt)} \quad (2.2)$$

has a unique solution in  $\mathbb{C}_+$ . Calling  $\tilde{m}(z)$  this solution, it is the Stieltjes Transform of the probability measure  $\tilde{\mu} = c\mu + (1 - c)\delta_0$  which is in fact the limit spectral measure of  $\Sigma_n \Sigma_n^*$ . And the Stieltjes Transform  $m(z)$  of  $\mu$  can be shown to be given by the equation

$$m(z) = \int \frac{-1}{z(1 + m(z)t)} \nu(dt). \quad (2.3)$$

The oldest version of these results dates back to the work of Marchenko and Pastur [37].

In [43], Silverstein and Choi study the properties of  $\mu$ . By analyzing Equation (2.2), which has no closed form solution for most  $\nu$ , they showed that  $\tilde{\mu}$  (or equivalently  $\mu$ ) has a continuous density away from zero, and that this density is analytical wherever it is positive.

The support of  $\tilde{\mu}$  might consist of several connected sets. This happens when  $\nu$  consists in *e.g.* several Diracs far enough

from each other. Silverstein and Choi characterized this support by applying an idea which dates back to Marchenko and Pastur’s paper: on any open interval of  $\mathbb{R} - \text{support}(\tilde{\mu})$ , the function  $\tilde{m}(z)$  exists and it is continuous, real and increasing. These properties are generically due to the fact that  $\tilde{m}(z)$  is the Stieltjes Transform of  $\tilde{\mu}$ . To determine  $\text{support}(\tilde{\mu})$ , the idea is to find the intervals on which the inverse of  $\tilde{m}(z)$  is increasing and to take out these intervals. On  $\mathbb{C}_+$ , it is clear from (2.2) that  $\tilde{m}(z)$  has a unique inverse given by

$$z(\tilde{m}) = -\frac{1}{\tilde{m}} + c \int \frac{t}{1 + \tilde{m}t} \nu(dt).$$

Now, on the set  $\mathcal{D} = \{\tilde{m} \in \mathbb{R} : \tilde{m} \neq 0, -\tilde{m}^{-1} \notin \text{support}(\nu)\}$ , simply define

$$x(\tilde{m}) = -\frac{1}{\tilde{m}} + c \int \frac{t}{1 + \tilde{m}t} \nu(dt).$$

Silverstein and Choi showed that in order to determine  $\text{support}(\tilde{\mu})$ , one only needs to study the behavior of  $x(\tilde{m})$  on  $\mathcal{D}$ . By locating the open intervals of  $\mathcal{D}$  where  $x(\tilde{m})$  is increasing and by taking out these intervals, we are left with  $\text{support}(\tilde{\mu})$ .

Finally, Silverstein and Choi showed that for most measures  $\nu$ , the density of  $\tilde{\mu}$  behaves like  $\sqrt{|x - x_0|}$  near a boundary  $x_0$  of  $\text{support}(\tilde{\mu})$ . Interestingly, it appeared later that there is a connection between this “squared root” behaviour of the density and the Tracy-Widom fluctuations of the eigenvalues of  $\Sigma_n \Sigma_n^*$  near the edges of the connected components of  $\text{support}(\pi)$ . This connection appears implicitly in El Karoui’s work [22].

### 2.1.2 Eigenvalue separation results

In [3], Bai and Silverstein showed the following result: Denoting by  $(x_{i,j})_{i,j=1}^{N;n}$  the elements of  $X_n$ , assume that  $\mathbb{E}|x_{11}|^4 < \infty$  and that  $\sup_n \|D_n\| < \infty$  where  $\|\cdot\|$  is the spectral norm. Denote by  $\nu_n$  the spectral measure of  $D_n$ . Recall that  $\nu_n$  converges weakly to  $\nu$ . Let  $\tilde{\mu}_n$  be the “finite  $n$  approximation” of  $\tilde{\mu}$  whose Stieltjes Transform is provided by Equation (2.2) in which  $\nu$  is replaced with  $\nu_n$ . Let  $[a, b]$  with  $a > 0$  be an interval outside the supports of  $\tilde{\mu}_n$  and  $\tilde{\mu}$  for all large  $n$ . Then with probability one,  $\Sigma_n \Sigma_n^*$  has no eigenvalue in  $[a, b]$  for all large  $n$ .

In [4], a finer “exact separation” result is shown. A close look at the results of [43] summarized above shows that there is a one to one correspondence between the interval  $[a, b]$  described above and an interval  $[a', b']$  lying outside  $\text{support}(\nu_n)$  and  $\text{support}(\nu)$  for all large  $n$ . In the typical cases, this correspondence is given by  $a' = -1/\tilde{m}(a)$  and  $b' = -1/\tilde{m}(b)$ . In a word, it is shown in [4] that with probability one, the number of eigenvalues of  $D_n$  and the number of eigenvalues of  $\Sigma_n \Sigma_n^*$  that lie at the same side of  $[a', b']$  and  $[a, b]$  respectively are the same for all large  $n$ .

### 2.1.3 A quick presentation of the estimation algorithms

We still assume that  $\sup_n \|D_n\| < \infty$  and that  $\mathbb{E}|x_{11}|^4 < \infty$ . Note that the assumption  $\sup_n \|D_n\| < \infty$  implies that  $\nu$  is compactly supported. In this case, one can show that  $\mu$  is also compactly supported. In [39, 38], Mestre poses the following problem: given a real analytical function  $\varphi$ , we want to consistently estimate the integral  $I = \int \varphi(\lambda) \nu(d\lambda)$  from the sole knowledge of  $\Sigma_n \Sigma_n^*$ . To solve this problem, Mestre develops an idea that can be traced back to Girko’s works [27, 28, 29]: Since  $\varphi$  is analytical, Cauchy’s integral theorem shows that

$$I = \frac{-1}{2i\pi} \int_{\mathcal{C}} \left( \oint_{\mathcal{C}} \frac{\varphi(z)}{\lambda - z} dz \right) \nu(d\lambda) = \frac{-1}{2i\pi} \oint_{\mathcal{C}} \varphi(z) m_\nu(z) dz$$

where  $\mathcal{C}$  is a positively oriented contour enclosing the support of  $\nu$ , and where  $m_\nu(z)$  is the Stieltjes Transform of  $\nu$ . Now, by doing the variable change  $z = -1/\tilde{m}(u)$ , one can show by using the results of [43] again that this variable change maps the contour  $\mathcal{C}$  into a positively oriented contour  $\mathcal{C}'$  enclosing the support of  $\mu$ . With this variable change,

we get

$$\begin{aligned}
I &= \frac{-1}{2i\pi} \oint_{\mathcal{C}'} \varphi\left(-\frac{1}{\tilde{m}(u)}\right) m_\nu\left(-\frac{1}{\tilde{m}(u)}\right) \frac{\tilde{m}'(u)}{\tilde{m}(u)^2} du \\
&= \frac{-1}{2i\pi} \oint_{\mathcal{C}'} \varphi\left(-\frac{1}{\tilde{m}(u)}\right) \left(\int \frac{1}{1+\tilde{m}(u)t} \nu(dt)\right) \frac{\tilde{m}'(u)}{\tilde{m}(u)} du \\
&= \frac{1}{2i\pi} \oint_{\mathcal{C}'} \varphi\left(-\frac{1}{\tilde{m}(u)}\right) \frac{um(u)\tilde{m}'(u)}{\tilde{m}(u)} dz
\end{aligned} \tag{2.4}$$

by Equation (2.3). Clearly, this expression depends only on the limit spectral measure  $\mu$ . Now, if we denote by  $\rho_n = N^{-1} \sum_{\ell=0}^{N-1} \delta_{\lambda_{\ell,n}}$  the spectral measure of  $\Sigma_n \Sigma_n^*$ , then its Stieltjes Transform

$$\hat{m}_n(z) = \int \frac{1}{\lambda - z} \rho_n(d\lambda) = \frac{1}{N} \text{trace}[(\Sigma_n \Sigma_n^* - zI)^{-1}]$$

converges almost surely to  $m(z)$  for any  $z$  outside support( $\mu$ ). Mestre shows that a consistent estimate of  $I$  is obtained by replacing  $m(z)$  in the last integral in (2.4) with  $\hat{m}_n(z)$ . Solving this new integral provides an easily computable expression depending only on the eigenvalues  $\lambda_{\ell,n}$  of  $\Sigma_n \Sigma_n^*$ . In the course of the proof, the results of [3] are of importance, since they ensure that no eigenvalue escapes the interior of the contour  $\mathcal{C}'$  for large  $n$ .

In some applications,  $\nu$  is written as  $\nu = \nu^1 + \dots + \nu^L$  where  $L$  is fixed and where the measures  $\nu^\ell$  have disjoint supports. The observer wants now to estimate objects as  $I = \int \varphi d\nu^\ell$  for some fixed  $\ell$ . The approach described above can be adapted to that case without difficulty, provided the supports of the components  $\nu^k$  are distant enough from each other. In this situation, the separation results of [4] are needed.

## 2.2 The information plus Noise ensemble

### 2.2.1 Limit spectral measure

Here we consider the matrix ensemble  $\Sigma_n = X_n + A_n$  where  $X_n$  is a  $N \times n$  matrix with iid elements having the mean zero and the variance  $1/n$ , and  $A_n$  is a deterministic matrix. As above, we assume that  $n \rightarrow \infty$  and that  $N/n \rightarrow c > 0$ . We also assume that  $\sup_n \|A_n\| < \infty$ , and that the spectral measure of  $A_n A_n^*$  converges weakly to a probability measure  $\nu$  as  $n \rightarrow \infty$ . In the fields of signal processing and communication theory, the matrix  $A_n$  often represents an ‘‘information signal’’ received on *e.g.* an array of  $N$  antennas, and  $X_n$  represents a corrupting ‘‘noise’’.

We are interested in the asymptotic behavior of the spectral measure of  $\Sigma_n \Sigma_n^*$ . In [20], Dozier and Silverstein showed the following facts. For each  $z \in \mathbb{C}_+$ , the equation

$$m = \int \frac{1}{-z(1+cm) + (1-c) + \frac{t}{1+cm}} \nu(dt)$$

has a unique solution  $m \in \mathbb{C}_+$ . The function  $m(z)$  so defined is the Stieltjes Transform of a probability measure  $\mu$ . Furthermore, as  $n \rightarrow \infty$ , the spectral measure of  $\Sigma_n \Sigma_n^*$  converges to  $\mu$  in the sense of (2.1).

The properties of  $\mu$  were studied in [19]. Dozier and Silverstein showed there that  $\mu$  has a density away from zero and that this density is analytical wherever it is positive. For most measures  $\nu$ , they also established the  $\sqrt{|x-x_0|}$  behavior of the density near the boundary points  $x_0$  of the support of  $\mu$ . They also proposed a procedure for determining the support of  $\mu$ . A more convenient formulation of the support determination problem was provided by Vallet, Loubaton and Mestre in [44]. In a word, by denoting as  $m_\nu(z)$  is the Stieltjes Transform of  $\nu$ , the fixed point equation defining  $m(z)$  can be rewritten as

$$\frac{m(z)}{1+cm(z)} = m_\nu(w(z))$$

where  $w(z) = z(1+cm(z))^2 - (1-c)(1+cm(z))$ . On  $\mathbb{R} - \text{support}(\mu)$ , the real function  $w(x)$  can be shown to be strictly increasing, and it is a solution of the equation  $\phi(w(x)) = x$ , where

$$\phi(w) = w(1 - cm_\nu(w))^2 + (1-c)(1 - cm_\nu(w)).$$

In [44], a simple procedure for determining support( $\mu$ ) by studying the function  $\phi(w)$  on  $\mathbb{R} - \text{support}(\nu)$  is proposed.

### 2.2.2 Eigenvalue separation results and estimation algorithms

Establishing the absence of eigenvalues of  $\Sigma_n \Sigma_n^*$  outside the support of the limit spectral measure for the information plus noise ensemble is not an easy problem. In [5], Bai and Silverstein establish a “no eigenvalue” result, but under a rather stringent assumption (Condition 1.10) on  $A_n$ . In [35, 34, 44] the “no eigenvalues” and the “exact separation” results were established in the case where the elements of  $X_n$  are Gaussian. In these references, no assumption on  $A_n$  except for the boundedness of its spectral norm was needed.

Given a real analytical function  $\varphi$ , we now want to estimate the integral  $I = \int \varphi(\lambda) \nu(d\lambda)$  from the sole knowledge of  $\Sigma_n \Sigma_n^*$ . Similarly to Section 2.1.3, we write

$$I = \frac{-1}{2i\pi} \oint_{\mathcal{C}} \varphi(z) m_{\nu}(z) dz$$

where the positively oriented contour  $\mathcal{C}$  encloses  $\text{support}(\nu)$ . By doing the change of variables  $z = w(u)$ , the contour  $\mathcal{C}$  is mapped to a contour  $\mathcal{C}'$  that encloses the support of  $\mu$ , and we get

$$I = \frac{-1}{2i\pi} \oint_{\mathcal{C}'} \varphi(w(u)) \frac{m(u)}{1 + cm(u)} w'(u) du.$$

This integral can be consistently estimated by replacing  $m(z)$  with  $\hat{m}_n(z) = N^{-1} \text{trace}[(\Sigma_n \Sigma_n - zI)^{-1}]$ . The “no eigenvalues” and the “exact separation” results are needed, the latter being needed in the case we want to estimate the integral of  $\varphi$  with respect to a component of  $\nu$ .

### 2.3 The double sided correlation ensemble

Here we have  $\Sigma_n = D_n^{1/2} X_n \tilde{D}_n^{1/2}$  where  $X_n$  is a  $N \times n$  random matrix with iid elements having the mean zero and the variance  $1/n$ ,  $D_n$  is a  $N \times N$  deterministic nonnegative matrix, and  $\tilde{D}_n$  is a  $n \times n$  deterministic nonnegative diagonal matrix. In the asymptotic regime as  $n \rightarrow \infty$  and  $N/n \rightarrow c > 0$ , we assume that the spectral measures of  $D_n$  and  $\tilde{D}_n$  converge weakly towards two probability measures. The asymptotic behavior of the spectral measure of  $\Sigma_n \Sigma_n^*$  has been studied in many works, including [13, 48, 33].

A “no eigenvalues” result was shown in [41] without a study of the support of the limit eigenvalue distribution. Work on this subject (support, density, behaviour near boundary points, etc.) is in progress in the framework of the present project.

## 3 Matrices mettant en jeu des estimateurs des tenseurs de cumulants d’ordre 4.

Cette section synthétise des résultats existants qui présentent un intérêt dans le cadre de la tâche 2-2 ”Analyse des performances de certains algorithmes de séparation aveugle de sources”.

### 3.1 Rappel de la problématique.

Dans le contexte de la séparation aveugle de sources bande étroite, on observe une série temporelle multivariable de dimension  $M$ , notée  $(\mathbf{y}_n)$ , et qui est supposée être un mélange linéaire de  $K$  ( $K < M$ ) signaux aléatoires stationnaires non gaussiens mutuellement indépendants perturbé par un bruit blanc additif gaussien. L’objet de la séparation aveugle de sources est d’estimer la matrice de mélange et les différentes composantes du mélange. Plus formellement,  $\mathbf{y}_n$  est défini à chaque instant  $n$  par

$$\mathbf{y}_n = \sum_{k=1}^K \mathbf{a}_k s_{k,n} + \mathbf{v}_n = \mathbf{A} \mathbf{s}_n + \mathbf{v}_n$$

où  $(s_k)_{k=1,\dots,K}$  sont des suites aléatoires stationnaires non gaussiennes non observables, mutuellement indépendentes, de moyenne nulle et de variance 1, et où les vecteurs de dimension  $M$   $(\mathbf{a}_k)_{k=1,\dots,K}$  sont inconnus et supposés déterministes.  $\mathbf{v}$  représente un bruit additif gaussien complexe.

$\mathbf{y}_n$  est observé entre les instants  $n = 1$  et  $n = N$ , et le problème est d'identifier la matrice  $\mathbf{A}$  de façon à pouvoir, dans un second temps, estimer les signaux non observables  $(s_{k,n})_{n=1,\dots,N}$ . Le problème de la séparation aveugle de sources est apparu dans les années 80, et de très nombreux travaux lui ont été consacrés. Un bon nombre d'algorithmes permettant d'estimer  $\mathbf{A}$  ont été développés, mais en supposant que le nombre de capteurs  $M$  est très inférieur à  $N$  le nombre d'observations. Le but de la tâche 2-2 est de reconsidérer le problème de la séparation de sources dans le contexte où  $M$  est grand et où  $N$  n'est pas très nettement plus grand que  $N$ . Dans ce contexte, les estimateurs classiques des statistiques de l'observation sur lesquels sont basés les algorithmes existants n'ont pas le même comportement que si  $M$  est fixe et que  $N \rightarrow +\infty$ .

Nous nous proposons de commencer à aborder le problème si le nombre de sources  $K$  est petit devant le nombre de capteurs  $M$  et le nombre d'observations  $N$ . L'approche la plus simple consiste sans doute à essayer de se ramener au cas d'un signal observé de dimension  $K$  en projetant  $\mathbf{y}_n$  sur l'espace  $\text{Im}(\mathbf{A})$  engendré par les colonnes de  $\mathbf{A}$ . Si on appelle  $\mathbf{U}$  la matrice  $M \times K$  constituée à partir d'une base orthonormée de  $\text{Im}(\mathbf{A})$ , on calcule ainsi le signal de dimension  $K$   $\mathbf{z}_n = \mathbf{U}^* \mathbf{y}_n$  sur lequel on peut appliquer des algorithmes classiques sans crainte des problèmes de dimension. Pour pouvoir mettre oeuvre cette approche, il faut être capable d'estimer l'espace  $\text{Im}(\mathbf{A})$ . Dans le cas où le bruit est blanc temporellement et spatialement, i.e. si  $\mathbb{E}(\mathbf{v}_m \mathbf{v}_n^*) = \delta(m-n) \sigma^2 \mathbf{I}_M$ , l'espace engendré par les colonnes de  $\mathbf{A}$  coïncide avec l'espace propre associé aux  $K$  plus grandes valeurs propres de la matrice de covariance  $\mathbb{E}(\mathbf{y}_n \mathbf{y}_n^H) = \mathbf{A} \mathbf{A}^* + \sigma^2 \mathbf{I}$ . Dans le cas où  $M$  est fixe et où  $N \rightarrow +\infty$ , la matrice des vecteurs propres  $\hat{\mathbf{U}}$  associés aux  $K$  plus grandes valeurs propres de la matrice de covariance empirique  $\frac{1}{N} \sum_{n=1}^N \mathbf{y}_n \mathbf{y}_n^*$  est un estimateur consistant d'une matrice  $\mathbf{U}$  formée à partir d'une base orthonormée de  $\text{Im}(\mathbf{A})$ . Dans le cas où  $M$  et  $N$  sont du même ordre de grandeur, la situation est plus complexe car  $\hat{\mathbf{U}}$  perd sa consistance. Cependant, la projection sur l'espace colonne de  $\hat{\mathbf{U}}$  reste pertinente si les  $K$  valeurs propres non nulles de la matrice  $\mathbf{A} \mathbf{A}^*$  sont supérieures au seuil  $\sigma^2 \sqrt{\frac{M}{N}}$ , mais cause en quelque sorte une dégradation du rapport signal sur bruit qui est d'autant plus importante qu'une ou plusieurs valeurs propres se rapprochent de  $\sigma^2 \sqrt{\frac{M}{N}}$  (voir [9] par exemple).

Dans le cas où le bruit n'a pas une structure aussi simple, il faut, pour pouvoir s'y ramener, estimer ses statistiques du second ordre. Bien que cette opération soit théoriquement possible, elle n'est pas si simple à réaliser en pratique. Cette constatation a été à l'origine d'approches permettant d'estimer l'espace  $\text{Im}(\mathbf{A})$ , voire même la matrice  $\mathbf{A}$  elle-même, à partir de statistiques insensibles à la présence de bruit additif gaussien, en l'occurrence les cumulants coinjoints d'ordre supérieurs à 2 de l'observation (voir par exemple [2]). C'est naturellement avec les cumulants d'ordre 4 qu'il est le plus simple de travailler. Si on désigne par  $(\mathbf{y}_{i,n})_{i=1,\dots,M}$  les composantes de  $\mathbf{y}_n$ , alors, l'indépendance mutuelle des sources et la gaussiannité du bruit additif implique immédiatement que

$$\text{cum}(\mathbf{y}_{i_1}, \mathbf{y}_{i_2}^*, \mathbf{y}_{i_3}, \mathbf{y}_{i_4}^*) = \sum_{k=1}^K c_4(s_k) \mathbf{a}_{i_1,k} \mathbf{a}_{i_2,k}^* \mathbf{a}_{i_3,k} \mathbf{a}_{i_4,k}^* \quad (3.5)$$

où  $c_4(s_k) = \text{cum}(s_k, s_k^*, s_k, s_k^*)$  (nous avons omis de préciser la dépendance temporelle). Si l'on range les divers cumulants dans une matrice  $M^2 \times M^2$   $\mathbf{C}$ , (3.5) apparaît équivalent à

$$\mathbf{C} = \sum_{k=1}^K c_4(s_k) (\mathbf{a}_k \otimes \mathbf{a}_k) (\mathbf{a}_k \otimes \mathbf{a}_k)^* \quad (3.6)$$

où  $\otimes$  désigne le produit de Kronecker. Par conséquent, la matrice  $\mathbf{C}$  est de rang  $K$ . Sous certaines hypothèses, (par exemple  $c_4(s_k) < 0$  pour  $k = 1, \dots, K$ ), il est possible d'extraire grâce à un algorithme relativement simple les vecteurs  $(\mathbf{a}_k)_{k=1,\dots,K}$  à partir des  $K$  valeurs propres non nulles et vecteurs propres associés de  $\mathbf{C}$  ([2]). En pratique, la matrice

$\mathbf{C}$  est estimée par la matrice  $\hat{\mathbf{C}}$  définie par

$$\hat{\mathbf{C}} = \frac{1}{N} \sum_{n=1}^N (\mathbf{y}_n \otimes \mathbf{y}_n) (\mathbf{y}_n \otimes \mathbf{y}_n)^* - \frac{1}{N^2} \sum_{(m,n)=1,\dots,N} (\mathbf{y}_m \otimes \mathbf{y}_n) (\mathbf{y}_m \otimes \mathbf{y}_n + \mathbf{y}_n \otimes \mathbf{y}_m)^* \quad (3.7)$$

qui remplace la matrice  $\mathbf{C}$  dans l'algorithme précédent. Cette approche fournit des estimateurs consistants des vecteurs  $(\mathbf{a}_k)_{k=1,\dots,K}$  dans le cas où  $M$  est fixe et que  $N \rightarrow +\infty$ . L'un des objectifs de la tâche 2-2 est d'étudier le comportement de  $\hat{\mathbf{C}}$  dans le cas où  $M$  et  $N$  tendent vers l'infini simultanément. Il convient de préciser que nous supposons dans la suite que lorsque  $M$  augmente, les normes des vecteurs  $(\mathbf{a}_k)_{k=1,\dots,K}$  restent bornées. Cette hypothèse implique implicitement que lorsque  $M \rightarrow +\infty$ , le rapport signal sur bruit **par capteurs** tend vers 0, mais que, au moins si le bruit est blanc temporellement et spatialement, pour chaque source  $k$ , le rapport signal sur bruit en sortie du filtre adapté spatial est un terme  $\mathcal{O}(1)$ . Ce régime nous semble le plus adapté au contexte des grands réseaux de capteurs car, dans le cas d'un nombre de sources  $K$  modéré, l'un des intérêts de ce genre de dispositif est de permettre de traiter avec succès des rapports signaux à bruits très faibles.

Pour comprendre le comportement asymptotique de  $\hat{\mathbf{C}}$ , nous commençons par étudier ses valeurs propres, et la façon dont elles se différencient de celles de  $\mathbf{C}$ . Pour aborder ce problème, nous étudions dans un premier temps les valeurs propres des 2 termes définissant  $\hat{\mathbf{C}}$ .

Soit  $\hat{\mathbf{D}}$  la matrice définie par

$$\hat{\mathbf{D}} = \frac{1}{N^2} \sum_{(m,n)=1,\dots,N} (\mathbf{y}_m \otimes \mathbf{y}_n) (\mathbf{y}_m \otimes \mathbf{y}_n + \mathbf{y}_n \otimes \mathbf{y}_m)^*$$

Nous avons établi que  $\text{Rank}(\hat{\mathbf{D}}) = \frac{M(M+1)}{2}$ , et que si  $(\hat{\lambda}_{k,N})_{k=1,\dots,M}$  représentent les valeurs propres de la matrice de covariance empirique  $\frac{1}{N} \sum_{n=1}^N \mathbf{y}_n \mathbf{y}_n^*$ , alors les valeurs propres non nulles de  $\hat{\mathbf{D}}$  coïncident avec les  $2(\hat{\lambda}_{k,N} \hat{\lambda}_{j,N})_{1 \leq k < j \leq M}$ . Ainsi que nous allons le voir, le comportement des valeurs propres  $(\hat{\lambda}_{k,N})_{k=1,\dots,M}$  est bien connu dès que le bruit  $\mathbf{v}$  est blanc temporellement dans les régimes asymptotiques définis plus loin.

Le terme qui semble le plus délicat à étudier est le terme  $\hat{\mathbf{E}}$  défini par

$$\hat{\mathbf{E}} = \frac{1}{N} \sum_{n=1}^N (\mathbf{y}_n \otimes \mathbf{y}_n) (\mathbf{y}_n \otimes \mathbf{y}_n)^*$$

La littérature existante n'est guère prolixe en la matière, mais des contributions récentes se sont intéressées à des problèmes connexes.

### 3.2 Le régime asymptotique $M \rightarrow +\infty, N \rightarrow +\infty$ de telle sorte que $\frac{M}{N} \rightarrow c, c > 0$ .

Nous supposons pour simplifier la présentation des résultats que le bruit  $\mathbf{v}$  est blanc temporellement et spatialement de variance  $\sigma^2$ . Dans ce cadre, le comportement des valeurs propres  $(\hat{\lambda}_{k,N})_{k=1,\dots,M}$  de la matrice de covariance empirique  $\frac{1}{N} \sum_{n=1}^N \mathbf{y}_n \mathbf{y}_n^*$  est bien connu. Si  $K = 0$ , les valeurs propres se distribuent selon la distribution de Marcenko-Pastur, et la plus grande  $\hat{\lambda}_{1,N}$  et la plus petite  $\hat{\lambda}_{M,N}$  convergent vers  $\sigma^2(1 + \sqrt{c})^2$  et  $\sigma^2(1 - \sqrt{c})^2$  respectivement (voir par exemple le livre [6]). Le comportement des valeurs propres a été établi plus récemment si  $K \neq 0$  (voir par exemple [9]). Les  $M - K$  plus petites valeurs propres restent dans l'intervalle  $[\sigma^2(1 - \sqrt{c})^2, \sigma^2(1 + \sqrt{c})^2]$ , et pour tout  $1 \leq k \leq K$ , la  $k$  ième valeur propre  $\hat{\lambda}_{k,N}$  tend vers  $\sigma^2(1 + \sqrt{c})^2$  si la  $k$  ième valeur propre de  $\mathbf{A}\mathbf{A}^*$  est plus petite que le seuil  $\sigma^2\sqrt{c}$ , et elle converge vers une quantité plus grande que  $\sigma^2(1 + \sqrt{c})^2$ , que l'on peut parfaitement caractériser, dans le cas contraire. Tout cela implique que le comportement des valeurs propres de  $\hat{\mathbf{D}}$  est bien compris.

Concernant les valeurs propres de la matrice  $\hat{\mathbf{E}}$ , les premières contributions que nous avons identifiées dans la littérature sont dues à El-Karoui ([24], [25]). Dans le régime asymptotique considéré, la matrice  $M^2 \times M^2$   $\hat{\mathbf{E}}$  est au maximum de rang  $N$  qui est très petit par rapport à sa dimension. Les travaux de El-Karoui ne concernent pas directement le comportement de la matrice  $\hat{\mathbf{E}}$ , mais la matrice  $N \times N$   $\tilde{\mathbf{E}}$  définie par

$$\left(\tilde{\mathbf{E}}\right)_{m,n} = \frac{(\mathbf{y}_m^* \mathbf{y}_n)^2}{N^2}$$

Il est facile de vérifier que

$$\lambda_j(\tilde{\mathbf{E}}) = \frac{1}{N} \lambda_j(\hat{\mathbf{E}})$$

pour  $j = 1, \dots, N$ , où d'une façon générale,  $\lambda_j(\mathbf{Q})$  désigne la  $j$ -ième plus grande valeur propre de la matrice hermitienne  $\mathbf{Q}$ . Ainsi, l'étude de la matrice  $\tilde{\mathbf{E}}$  permet d'avoir des informations sur les  $N$  valeurs propres non nulles de  $\hat{\mathbf{E}}$ .

### 3.2.1 En absence de source.

Dans le cas où aucune source n'est présente, c'est-à-dire que  $K = 0$ , ou de façon équivalente que  $\mathbf{y}_n = \mathbf{v}_n$ , il est établi dans [24] que

$$\text{Tr} \left[ \left( \tilde{\mathbf{E}} - c^2 \sigma^4 \mathbf{I}_N \right) \left( \tilde{\mathbf{E}} - c^2 \sigma^4 \mathbf{I}_N \right)^* \right]^2 \rightarrow 0$$

ce qui implique que

$$\|\tilde{\mathbf{E}}_N - c^2 \sigma^4 \mathbf{I}_N\| \rightarrow 0$$

et que pour tout  $j = 1, N$ ,

$$\lambda_j(\hat{\mathbf{E}}) = c^2 \sigma^4 N + o(N) \tag{3.8}$$

En d'autres termes, les  $N$  plus grandes valeurs propres de  $\hat{\mathbf{E}}$  tendent vers l'infini linéairement en  $N$ . En tenant compte des valeurs propres de la matrice  $\hat{\mathbf{D}}$ , les inégalités de Weyl impliquent que:

$$\begin{aligned} \lambda_k(\hat{\mathbf{C}}) &= \mathcal{O}(N), \quad 1 \leq k \leq N \\ -2 \hat{\lambda}_{1,N}^2 &\leq \lambda_k(\hat{\mathbf{C}}) \leq -2 \hat{\lambda}_{M,N}^2, \quad N < k \leq \frac{M(M+1)}{2} \end{aligned}$$

Dans le cas considéré, nous avons rappelé que  $\hat{\lambda}_{1,N} \rightarrow \sigma^2(1 + \sqrt{c})^2$  et que  $\hat{\lambda}_{M,N} \rightarrow \sigma^2(1 - \sqrt{c})^2$ ; la seconde inégalité implique alors que les  $\frac{M(M+1)}{2} - N$  plus petites valeurs propres de  $\hat{\mathbf{C}}$  sont presque sûrement localisées dans un voisinage de  $[-4\sigma^4(1 + \sqrt{c})^4, -4\sigma^4(1 - \sqrt{c})^4]$  pour  $N$  assez grand. Tout ceci montre que les valeurs propres de la matrice  $\hat{\mathbf{C}}$  se comportent d'une façon radicalement différentes de celles de la "vraie" matrice  $\mathbf{C}$  qui, dans le contexte  $K = 0$ , se réduit à la matrice identiquement nulle du fait de la gaussiannité du bruit.

### 3.3 Le régime $M \rightarrow +\infty, N \rightarrow +\infty$ de telle sorte que $\frac{M^2}{N} \rightarrow d$ .

Ce régime asymptotique présente également un intérêt puisqu'il modélise des situations dans lesquelles  $M$  est grand, mais où  $N$  peut être sensiblement plus grand que  $M$ . Dans ce contexte, la matrice de covariance empirique devrait avoir le même comportement en norme spectrale que la "vraie" matrice de covariance  $\mathbb{E}(\mathbf{y}_n \mathbf{y}_n^*) = \mathbf{A} \mathbf{A}^* + \sigma^2$ . Ceci permet de régler le problème de la caractérisation des valeurs propres de la matrice  $\hat{\mathbf{D}}$ .

L'étude des valeurs propres de  $\hat{\mathbf{E}}$  ne semble pas avoir été effectuée en tant que telle. Toutefois, certains résultats connexes ont été publiés récemment. Ainsi, les résultats de [31] montrent que si  $(\mathbf{v}_{1,n})_{n \in \mathbb{Z}}$  et  $(\mathbf{v}_{2,n})_{n \in \mathbb{Z}}$  sont 2 bruits gaussiens blancs temporellement et spatialement de même variance  $\sigma^2$ , alors la distribution des valeurs propres de la matrice

$$\frac{1}{N} \sum_{n=1}^N (\mathbf{v}_{1,n} \otimes \mathbf{v}_{2,n}) (\mathbf{v}_{1,n} \otimes \mathbf{v}_{2,n})^H$$



converge vers la distribution de Marcenko-Pastur de paramètre  $\sigma^4$  et  $d$ , c'est à-dire la distribution que l'on observerait si l'on remplaçait les vecteurs  $(\mathbf{v}_{1,n} \otimes \mathbf{v}_{2,n})_{n=1,\dots,N}$  par des vecteurs  $(\mathbf{w}_n)_{n=1,\dots,N}$  gaussiens complexes i.i.d. de dimension  $M^2$  et de matrice de covariance  $\sigma^4 \mathbf{I}_{M^2}$ . De plus, il est montré dans [31] par des techniques combinatoires que la plus grande (resp. plus petite) valeur propre de la matrice précédente converge vers  $\sigma^4(1 + \sqrt{d})^2$  (resp.  $\sigma^4(1 - \sqrt{d})^2$ ).

Dans le contexte qui est le nôtre, nous avons besoin d'un résultat concernant le cas où  $\mathbf{v}_1 = \mathbf{v}_2 = \mathbf{v}$  pour analyser les valeurs propres de  $\hat{\mathbf{D}}$  dans le cas  $K = 0$  avec un bruit blanc temporellement et spatialement. Il sera ensuite nécessaire d'évaluer la situation dans le cas où  $K > 0$ . En tout état de cause, dans le cas où  $N$  et  $M^2$  sont du même ordre de grandeur, il est fort probable les valeurs propres de  $\hat{\mathbf{D}}$  restent confinées dans un intervalle compact à la différence du cas où  $\frac{M}{N} \rightarrow c$ . Pour autant, les valeurs propres de  $\hat{\mathbf{C}}$  ne se comporteront sans doute pas du tout comme celles de la vraie matrice  $\mathbf{C}$ . Pour qu'il en soit ainsi, il faudrait vraisemblablement que  $N$  soit encore plus grand que  $M^2$ . Cela tendrait à indiquer que pour bien estimer le tenseur cumulant d'ordre 4, il faudrait au minimum que  $\frac{M^2}{N} \rightarrow 0$ , ce qui, en soi, serait un résultat très intéressant.

## 4 Etude du comportement asymptotique de matrices Hankel par bloc.

L'étude de grandes matrices aléatoires Hankel par bloc est motivée par la tâche 3 consacrée à des problèmes de détection / estimation mettant en jeu des signaux large bande. Lorsque qu'un réseau de capteurs à  $M$  éléments recueille des données générées par une source large bande, le signal de dimension  $M$   $(\mathbf{y}_n)_{n \in \mathbb{Z}}$  reçu peut s'écrire sous la forme

$$\mathbf{y}_n = \sum_{p=0}^{P-1} \mathbf{a}_p s_{n-p} + \mathbf{v}_n = \mathbf{x}_n + \mathbf{v}_n \quad (4.9)$$

où  $(s_n)_{n \in \mathbb{Z}}$  représente un signal scalaire déterministe non observable,  $(\mathbf{a}_p)_{p=0,\dots,P-1}$  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  $\sigma^2$ . 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  $(\mathbf{a}_p)_{p=0,\dots,P-1}$ .
- Estimer la séquence  $(s_n)$ .

(voir par exemple [45], [40], [46], [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  $\mathbf{Y}_L$  is the block-Hankel  $ML \times N$  defined by

$$\mathbf{Y}_L = \begin{pmatrix} \mathbf{y}_1 & \mathbf{y}_2 & \cdots & \cdots & \mathbf{y}_{N-L+1} \\ \mathbf{y}_2 & \mathbf{y}_3 & \ddots & \ddots & \mathbf{y}_{N-L+2} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \mathbf{y}_L & \mathbf{y}_{L+1} & \cdots & \cdots & \mathbf{y}_N \end{pmatrix}$$

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  $\mathbf{Y}_L$ . En principe, l'entier  $L$  est choisi plus grand que  $P$  car, dans ce cas, la matrice  $ML \times ML$   $\frac{\mathbf{X}_L \mathbf{X}_L^*}{N}$  est de rang déficient quand  $ML < N$  dès que  $M \geq 2$ , ce qui s'avère utile dans le cadre des problèmes de détection / estimation listés plus haut.

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

$$ML \rightarrow +\infty, N \rightarrow +\infty, \frac{ML}{N} \rightarrow c \quad (4.10)$$

et d'étudier le comportement de  $\frac{\mathbf{Y}_L \mathbf{Y}_L^*}{N}$  dans ce contexte. Il convient de noter que la condition (4.10) couvre différents types de régimes:  $L$  fixe et  $M$  et  $N$  du même ordre de grandeur,  $L$  et  $M$  tendent tous les deux vers  $+\infty$ , et  $M$  fixe et  $L$  et  $N$  du même ordre de grandeur.

Dans le cadre du projet DIONISOS, nous commencerons à nous intéresser au comportement des valeurs propres de  $\frac{\mathbf{Y}_L \mathbf{Y}_L^*}{N}$  dans le cas où la source large bande n'est pas présente, c'est-à-dire que  $\mathbf{y}$  se réduit à un bruit blanc temporellement et spatialement. La deuxième étape consistera à considérer le cas où le signal utile est présent, à voir dans quelle mesure cela affecte les valeurs propres, et, le cas échéant si on peut préciser le comportement des vecteurs propres associés aux plus grandes valeurs propres.

L'étude du spectre de  $\frac{\mathbf{Y}_L \mathbf{Y}_L^*}{N}$  en absence de source dans le régime (4.10) peut-être déduite de travaux récents dans certains cas. A notre connaissance, rien n'a été fait dans le cas où le signal utile  $\mathbf{x}$  est présent.

**Si  $L$  est fixe.** Divers travaux concernant les matrices aléatoires par bloc peuvent être utilisés, par exemple [28] (chapitre 16), [14]. Ils montrent que la distribution des valeurs propres coïncide avec la distribution de Marcenko-Pastur de paramètres  $\sigma^2$  et  $c$ , c'est-à-dire celle que l'on observerait la matrice  $\mathbf{Y}_L$  était à éléments i.i.d. Le travail récent [36] implique également que la plus grande valeur propre et la plus petite valeur propre de  $\frac{\mathbf{Y}_L \mathbf{Y}_L^*}{N}$  convergent respectivement vers  $\sigma^2(1 + \sqrt{c})^2$  et  $\sigma^2(1 - \sqrt{c})^2$ .

**Si  $M$  est fixe.** Il s'agit d'un contexte plus difficile que le précédent. [7] a considéré le cas  $M = 1$  and  $\frac{L}{N} \rightarrow c$ , mais quand les  $y_n$  sont forcés à 0 pour  $N < n < N + L$ . Dans ce cadre, la matrice  $\frac{\mathbf{Y}_L \mathbf{Y}_L^*}{N}$  n'est rien d'autre que la matrice d'autocovariance empirique du signal  $y$ . En utilisant la méthode des moments, [7] établit que la distribution des valeurs propres converge vers une distribution déterministe à support non compact. Le comportement de la plus grande valeur propre de  $\frac{\mathbf{Y}_L \mathbf{Y}_L^*}{N}$  n'est pas étudiée dans [7], mais, au vu des résultats de [42] (qui concernent un modèle différent), il semble probable qu'elle tende vers  $\infty$  à un rythme qui devrait être  $\log N$ .

Il reste donc un travail important à effectuer pour avoir une bonne compréhension des valeurs propres de  $\frac{\mathbf{Y}_L \mathbf{Y}_L^*}{N}$ :

- En absence de source:
  - Lorsque  $L$  et  $M$  tendent vers  $+\infty$ , étude de la distribution de ses valeurs propres, et évaluation du comportement de la plus grande valeur propre.
  - Lorsque  $M$  est fixe, caractérisation du comportement de la plus grande valeur propre.
- En présence d'une source, il conviendra d'étudier les conditions dans lesquelles les plus grandes valeurs propres de  $\frac{\mathbf{Y}_L \mathbf{Y}_L^*}{N}$  se comportent différemment du cas où  $K = 0$ .

## 5 Estimation convergente en norme de grandes matrices de covariance (Bickel-Levina et autres).

La problématique de l'estimation de grandes matrices de covariance a connu une certaine accélération à partir des années 2005, et nous nous proposons d'en résumer ici les résultats qui peuvent être d'une certaine utilité dans le cadre du projet DIONISOS.

## 5.1 Estimation d'une matrice de covariance dont les termes éloignés de la diagonale tendent vers 0 suffisamment vite.

Dans le cadre de ce paragraphe, on observe des vecteurs aléatoires Gaussiens  $(\mathbf{y}_n)_{n=1,\dots,N}$  de dimension  $M$ , indépendants identiquement distribués, de moyenne nulle, et de matrice de covariance  $\mathbf{R}$ . Il s'agit ici d'estimer  $\mathbf{R}$  par une matrice  $\hat{\mathbf{R}}$  vérifiant

$$\|\hat{\mathbf{R}} - \mathbf{R}\| \rightarrow 0 \quad (5.11)$$

dans le cas où  $M$  et  $N$  tendent tous les deux vers  $+\infty$  de telle sorte que  $\frac{M}{N} \rightarrow c$ . En absence d'hypothèse a priori sur  $\mathbf{R}$ , cela est impossible car la matrice de covariance empirique, qui coïncide avec l'estimateur du maximum de vraisemblance de  $\mathbf{R}$ , ne vérifie pas (5.11). Nous allons essentiellement évoquer le cas où les éléments  $\mathbf{R}_{i,j}$  tendent vers 0 à un rythme suffisant lorsque  $|i-j| \rightarrow +\infty$ . Ainsi, [11] suppose que  $\mathbf{R}$  appartient à la classe de matrices de covariance  $M \times M$   $\mathcal{U}(\alpha)$  définie comme l'ensemble des matrices définies positives  $\mathbf{Q}$  vérifiant les conditions suivantes:

$$\begin{aligned} \mathbf{Q} > 0, \sup_1 \lambda_1(\mathbf{Q}) < \kappa_1, \inf_M \lambda_1(\mathbf{Q}) > \kappa_2 \\ \sup_j \sum_{|i-j|>k} |\mathbf{Q}_{i,j}| \leq \mu k^{-\alpha} \end{aligned}$$

où  $\kappa_1, \kappa_2, \mu$  sont des constantes.

Dans la suite, nous appelons  $\hat{\mathbf{R}}_e$  la matrice de covariance empirique définie par

$$\hat{\mathbf{R}}_e = \frac{1}{N} \sum_{n=1}^N \mathbf{y}_n \mathbf{y}_n^*$$

[11] propose de considérer l'estimateur  $\mathbf{B}_k(\hat{\mathbf{R}}_e)$  obtenu en mettant à zéro les éléments  $(\hat{\mathbf{R}}_e)_{i,j}$  pour lesquels  $|i-j| > k$ . Plus formellement,  $\mathbf{B}_k(\hat{\mathbf{R}}_e)$  est défini par

$$\left(\mathbf{B}_k(\hat{\mathbf{R}}_e)\right)_{i,j} = (\hat{\mathbf{R}}_e)_{i,j} \mathbb{1}_{|i-j| \leq k} \quad (5.12)$$

L'un des résultats principaux de [11] est que si la matrice  $\mathbf{R}$  appartient à l'ensemble  $\mathcal{U}(\alpha)$ , alors l'estimateur  $\mathbf{B}_{k(N)}(\hat{\mathbf{R}}_e)$ , avec

$$k(N) = \mathcal{O}\left(\frac{N}{\log M}\right)^{\frac{1}{2(\alpha+1)}} \quad (5.13)$$

vérifie

$$\|\mathbf{R} - \mathbf{B}_{k(N)}(\hat{\mathbf{R}}_e)\| = \mathcal{O}\left(\frac{\log M}{N}\right)^{\frac{\alpha}{2(\alpha+1)}} \quad (5.14)$$

On peut ainsi constater que si la matrice  $\mathbf{R}$  est suffisamment parcimonieuse, alors il est possible d'estimer  $\mathbf{R}$  de façon consistante dès que  $\frac{\log M}{N} \rightarrow 0$ , condition qui est évidemment vérifiée si  $M$  et  $N$  sont du même ordre de grandeur, mais aussi si  $N$  est de l'ordre de  $M^p$  avec  $p$  un entier quelconque.

Ainsi que cela est mentionné dans [11], l'estimateur  $\mathbf{B}_k(\hat{\mathbf{R}}_e)$  a l'inconvénient de ne pas être toujours une matrice positive. Pour résoudre ce problème, [11] propose de multiplier les éléments de  $\hat{\mathbf{R}}_e$  par ceux d'une matrice positive bien choisie. Considérons par exemple la matrice  $\mathbf{F}_\beta$  définie par

$$\mathbf{F}_\beta(i, j) = \left(1 - \frac{|i-j|}{\beta}\right)_+$$

Si  $\bullet$  représente le produit terme à terme des matrices,  $\mathbf{F}_\beta \bullet \hat{\mathbf{R}}_e$  est une matrice positive, et si on choisit  $\beta = \beta_N = \mathcal{O}\left(\frac{N}{\log M}\right)^{\frac{1}{2\alpha+1}}$ , alors,

$$\|\mathbf{R} - \mathbf{F}_{\beta_N} \bullet \hat{\mathbf{R}}_e\| = \mathcal{O}\left(\frac{\log M}{N}\right)^{\frac{\alpha}{2\alpha+1}}$$

On retrouve le même type de résultat que dans le cadre de l'estimateur  $\mathbf{B}_{k_N}(\hat{\mathbf{R}}_e)$ , mais la positivité de l'estimateur est préservée, ce qui est un avantage.

[17] a mis en évidence des estimateurs convergeant plus vite que ceux de [11]. De surcroît, il est montré que le taux de convergence de [17] est optimal au sens minimax. Comme dans [11], [17] proposent de multiplier terme à terme  $\hat{\mathbf{R}}_e$  par une matrice (non nécessairement positive) bien choisie. Ils considèrent la matrice  $\mathbf{W}_\gamma$  définie par

$$\begin{aligned} \mathbf{W}_\gamma(i, j) &= 1, \quad |i - j| \leq \gamma/2 \\ \mathbf{W}_\gamma(i, j) &= 2 \left( 1 - \frac{|i - j|}{\gamma} \right), \quad \gamma/2 \leq |i - j| \leq \gamma \\ \mathbf{W}_\gamma(i, j) &= 0, \quad |i - j| \geq \gamma \end{aligned}$$

et supposent que  $M \geq N^{1/(2\alpha+1)}$ . Ils considèrent le choix de  $\gamma = \gamma_N = N^{1/(2\alpha+1)}$ , et établissent que

$$\|\mathbf{R} - \mathbf{W}_{\gamma_N} \bullet \hat{\mathbf{R}}_e\|^2 = \mathcal{O} \left( N^{-\frac{2\alpha}{2\alpha+1}} + \frac{\log M}{N} \right)$$

Si l'on considère le cas où  $M$  et  $N$  sont du même ordre de grandeur,  $\frac{\log M}{N}$  est négligeable devant  $N^{-\frac{\alpha}{2\alpha+1}}$ , et l'on obtient que

$$\|\mathbf{R} - \mathbf{W}_{\gamma_N} \bullet \hat{\mathbf{R}}_e\| = \mathcal{O} \left( N^{-\frac{\alpha}{2\alpha+1}} \right)$$

alors que les 2 estimateurs de [11] produisent une erreur de l'ordre de

$$\|\mathbf{R} - \mathbf{B}_{k(N)}(\hat{\mathbf{R}}_e)\| = \|\mathbf{R} - \mathbf{F}_{\beta_N} \bullet \hat{\mathbf{R}}_e\| = \mathcal{O} \left( \frac{\log N}{N} \right)^{\frac{\alpha}{2(\alpha+1)}}$$

qui, bien entendu, est supérieur à  $\mathcal{O} \left( N^{-\frac{\alpha}{2\alpha+1}} \right)$ .

Il convient de mentionner que [12] a également considéré une classe de matrice de covariance parcimonieuse qui diffère de  $\mathcal{U}(\alpha)$ , et a analysé les performances d'un estimateur obtenu en seuillant les éléments de  $\hat{\mathbf{R}}_e$ . [23] s'est enfin intéressé à une classe de matrices parcimonieuses plus générale que celle de [12], et établit la consistance en norme d'estimateurs obtenus en seuillant les éléments de  $\hat{\mathbf{R}}_e$ .

## 5.2 Estimation de grandes matrices de covariance de Toeplitz.

On suppose dans ce paragraphe que la matrice de covariance  $\mathbf{R}$  de  $\mathbb{E}(\mathbf{y}_n \mathbf{y}_n^H)$  est une matrice de Toeplitz, et l'on souhaite estimer  $\mathbf{R}$  de façon consistante au sens de la norme spectrale dans le cas où  $M \rightarrow +\infty$ . Il convient de noter qu'il s'agit d'un contexte dans lequel le cas où  $N = 1$  est pertinent: dans ce cadre, l'unique vecteur  $\mathbf{y}$  disponible coïncide avec  $(y_1, \dots, y_M)^T$  où  $(y_m)_{m \in \mathbb{Z}}$  est une série temporelle stationnaire centrée.

Cette problématique a fait l'objet de quelques travaux dans le passé dans le cas  $N = 1$  (voir par exemple [47] et les références qui y sont données), mais le travail le plus achevé est [15] qui considère le cas le plus général, et dont les résultats sont les plus puissants. [15] considère l'ensemble  $\mathcal{V}(\alpha)$  de matrices de covariance  $\mathbf{Q} = (q(i - j))_{(i,j)=1,\dots,M}$  de Toeplitz défini par

$$\mathcal{V}(\alpha) = \left\{ \mathbf{Q} = (q(i - j))_{(i,j)=1,\dots,M} : |q(m)| \leq \frac{C}{(m+1)^{\alpha+1}}, \mathbf{Q} > 0 \right\}$$

On désigne par  $\hat{\mathbf{R}}_t$  la matrice obtenue en Toeplitzifiant de façon classique la matrice de covariance empirique  $\hat{\mathbf{R}}_e$ , c'est-à-dire:

$$\left( \hat{\mathbf{R}}_t \right)_{i,j} = \frac{1}{M - |i - j|} \sum_{m_1 - m_2 = i - j} \left( \hat{\mathbf{R}}_e \right)_{m_1, m_2}$$

L'estimateur  $\hat{\mathbf{R}}_t$  n'est évidemment pas adapté au contexte des grandes dimensions, et [15] propose, à la suite de [47], de l'améliorer en s'inspirant des estimateurs de [11] et [17]. [15] étudie en particulier le comportement de  $\mathbf{B}_k(\hat{\mathbf{R}}_t)$  et de  $\mathbf{W}_\beta \bullet \hat{\mathbf{R}}_t$  pour des choix pertinents de  $k$  et de  $\beta$ . Ainsi, il est prouvé que si

$$\left( \frac{MN}{\log MN} \right)^{\frac{1}{2\alpha+1}} \leq \frac{M}{2}$$

alors, si  $k = k(N) = \beta = \beta(N) = \left( \frac{MN}{\log MN} \right)^{\frac{1}{2\alpha+1}}$ ,

$$\begin{aligned} \|\mathbf{R} - \mathbf{B}_{k(N)}(\hat{\mathbf{R}}_t)\| &= \mathcal{O} \left( \left( \frac{\log MN}{MN} \right)^{\frac{2\alpha}{2\alpha+1}} \right) \\ \|\mathbf{R} - \mathbf{W}_{\beta(N)} \bullet \hat{\mathbf{R}}_t\| &= \mathcal{O} \left( \left( \frac{\log MN}{MN} \right)^{\frac{2\alpha}{2\alpha+1}} \right) \end{aligned}$$

### 5.3 Applications possibles dans le cadre du projet DIONISOS.

Les problèmes d'estimation de grandes matrices de covariance, ayant éventuellement une structure de type Toeplitz, apparaissent dans le cadre des tâches 2-1 et 3.

#### 5.4 Dans le cadre de la tâche 2-1

Dans le domaine de la localisation de source bande étroite, la matrice de covariance des observations a une structure Toeplitz si l'antenne de réception est linéaire à capteurs équidistants. Dès lors, il apparaît raisonnable d'utiliser cette information a priori pour remplacer la classique matrice de covariance empirique par une matrice de Toeplitz bien choisie. Dans le cas où  $M$  est fixe et où  $N \rightarrow +\infty$ , l'estimateur  $\hat{\mathbf{R}}_t$  est consistant, et les gains possibles de la Toeplitzification sont bien connus. Dans le cas où  $M$  et  $N$  tendent vers l'infini au même rythme, la consistance de  $\hat{\mathbf{R}}_t$  n'est pas évidente. Les travaux de [15] ne peuvent être utilisés dans le contexte du traitement d'antenne car les matrices de covariance que l'on y rencontre n'appartiennent pas à un ensemble du type  $\mathcal{V}(\alpha)$ . Le problème de l'estimation consistante de  $\mathbf{R}$  et des bénéfices potentiels que cela peut procurer par aux méthodes sous-espaces adaptées aux grandes dimensions ([39], [44]) doivent donc être étudiés.

Dans la tâche 2-1, nous nous sommes également donnés comme objectif de traiter du problème de la détection et de la localisation de sources bande étroite dans le cas où le bruit est blanc spatialement, mais possède une corrélation temporelle inconnue. Dans le cas où les bruits sur chacun des capteurs sont des signaux stationnaires de même spectre, on doit pouvoir utiliser les techniques de [15] afin de pouvoir estimer la matrice de covariance du bruit de façon consistante, et se ramener au cas bien connu où le bruit est blanc temporellement et spatialement.

#### 5.5 Dans le cadre de la tâche 3.

Dans le cadre de la tâche 3, il est pertinent de se demander si il est possible d'estimer de façon consistante la matrice de covariance spatio-temporelle, c'est-à-dire  $\mathbb{E}(\mathbf{y}_{n,L} \mathbf{y}_{n,L}^*)$ , où  $\mathbf{y}_{n,L} = (\mathbf{y}_n^T, \mathbf{y}_{n+1}^T, \dots, \mathbf{y}_{n+L-1}^T)^T$ , et d'utiliser un éventuel estimateur consistant en lieu et place de la matrice de covariance empirique spatio-temporelle  $\frac{\mathbf{Y}_L \mathbf{Y}_L^H}{N}$ .  $\mathbb{E}(\mathbf{y}_{n,L} \mathbf{y}_{n,L}^*)$  est une matrice bloc-Toeplitz, de sorte qu'il conviendrait de combiner les techniques de [11] et [17] avec celles de [15]. Cela supposerait cependant que l'on dispose d'a priori sur la parcimonie de  $\mathbb{E}(\mathbf{y}_{n,L} \mathbf{y}_{n,L}^*)$  dans le domaine temporel, ce qui est raisonnable, mais aussi dans le domaine spatial, ce qui est plus ou moins équivalent à supposer que les vecteurs  $(\mathbf{a}_p)_{p=0, \dots, P-1}$  sont parcimonieux en un sens à définir. Les implications pratiques de ce type d'hypothèse seraient évidemment à analyser soigneusement.

## 6 Estimation robuste.

### 6.1 Introduction

L'origine de la théorie de l'estimation robuste est souvent associée à 1964 et l'article de Huber [49] dans lequel il considère le problème d'estimation d'un paramètre (dit de *localisation*) d'une distribution de probabilité  $F$  lorsque les échantillons scalaires observés issus de  $F$  sont remplacés avec faible probabilité par des échantillons issus d'une autre distribution inconnue  $G$ . Dans son expression moderne plus générale, l'un des objectifs de l'estimation robuste est ainsi d'identifier des paramètres associés à une variable aléatoire multivariée  $X$  (moyenne ou covariance par exemple) à partir d'observations parfois fallacieuses de la distribution de  $X$ . L'un des intérêts premiers des estimateurs robustes est d'éviter l'introduction d'un biais d'estimation arbitrairement large en cas de pollution des observations par un petit nombre de données fallacieuses d'amplitude arbitraire. Ainsi la moyenne empirique  $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$  d'une variable aléatoire scalaire  $X$  de moyenne  $E[X]$  observée à travers  $X_1, \dots, X_n$  est-elle un estimateur de  $E[X]$  extrêmement peu robuste à l'introduction d'une observation fallacieuse  $X_{n+1}$  d'amplitude arbitraire. A contrario, l'estimateur de  $E[X]$  donné par l'échantillon médian  $\text{MED}(X_1, \dots, X_n)$  est un estimateur robuste (cependant potentiellement biaisé ou moins précis) à l'introduction d'une observation de valeur arbitraire.

L'autre aspect de la théorie de l'estimation robuste est celui de l'estimation de paramètres de variables aléatoires multivariées dont les lois sont à queues lourdes<sup>1</sup> ou tout du moins sont suffisamment distinctes de la distribution gaussienne multivariée. Dans ce cadre, les estimateurs robustes ont pour buts d'avoir des performances précises à horizon  $n$  fini, au contraire des estimateurs standards tels que la moyenne et la covariance empirique. En effet, bien que beaucoup de résultats d'universalité de ces estimateurs standards (basés sur la loi des grands nombres) assurent leur convergence presque sûre vers les paramètres estimés, la précision de tels estimateurs et en particulier leur vitesse de convergence présentent de grandes fluctuations selon la distribution sous-jacente des variables observées. En particulier, les moyenne et covariance empiriques d'une variable aléatoire gaussienne multivariée sont précisément les estimateurs maximisant la vraisemblance  $P(\cdot | X_1, \dots, X_n)$  des moyenne  $E[X]$  et covariance  $E[(X - E[X])(X - E[X])^*]$  de  $X$ , respectivement. Du fait de cette propriété, du vaste usage de la distribution gaussienne dans le domaine du traitement du signal et de la simplicité d'analyse de ces deux estimateurs, les moyenne et covariance empiriques sont quasiment les seuls estimateurs employés à cet effet. Pourtant, force est de constater que ces estimateurs présentent d'importantes faiblesses lorsqu'il s'agit d'estimer les moyenne et covariance de variables aléatoires à queues plus lourdes. Dans ce cadre, il est plus efficace (l'efficacité étant bien sur un critère à définir) d'utiliser des estimateurs plus appropriés. C'est dans cet esprit que Maronna [50] introduisit une famille plus large d'estimateurs robustes que celui proposé par Huber et dont une sous-classe correspond aux estimateurs maximisant la vraisemblance des moyenne et covariance (ou matrice de diffusion lorsque la covariance n'est pas définie) de variables aléatoires elliptiques. Les distributions elliptiques forment une famille intéressante de distributions qui contiennent la distribution gaussienne comme cas particulier, dont les queues peuvent être paramétrées aisément et qui sont relativement exploitables mathématiquement. En outre, les estimateurs de Maronna se prêtent facilement à l'étude mathématique de leur comportement asymptotique. Par ailleurs, ils montrent la propriété étonnante de contenir comme cas particulier l'estimateur de Huber initialement obtenu pour des raisons très différentes.

Les estimateurs robustes étant bien souvent des estimateurs de maximum de vraisemblance (dans le cas ci-dessus des estimateurs de Maronna pour des observations elliptiques) ou satisfaisant l'optimisation d'une autre métrique, ils appartiennent naturellement à la large classe des  $M$ - ou  $Z$ -estimateurs ( $M$ -estimateurs pour ceux réalisant le maximum d'une métrique et  $Z$ -estimateurs  $\hat{\theta}$  pour ceux satisfaisant une équation de type  $f_n(\hat{\theta}) = 0$ ), les deux sous-classes étant fortement reliées par le fait que des métriques différentiables de  $M$ -estimateurs induisent des  $Z$ -estimateurs. C'est ainsi que nombre d'estimateurs robustes traditionnels  $\hat{\theta}$  sont définis comme solution d'une équation implicite de type

---

<sup>1</sup>Dans la suite, la terminologie "queue lourde" sera utilisée pour caractériser une distribution dont les moments successifs sont larges ou infini; nous ne suivons pas la terminologie standard consistant à utiliser "queue lourde" pour des distributions ayant un moment d'ordre deux infini.

$f_n(\hat{\theta}) = 0$ . L'analyse d'un estimateur robuste débute alors bien souvent par une preuve de son existence en tant que solution de l'équation implicite, puis par une preuve d'unicité de la solution de cette équation implicite (au moins parmi une classe de solutions donnée). En particulier, la paramétrisation fine des estimateurs de Maronna que nous discuterons plus bas ainsi que quelques hypothèses additionnelles sur les échantillons observés (hypothèses qui ne sont pas un obstacle majeur pour Maronna mais qui en seront un pour notre étude) assurent cette unicité [50, Théorème 1]. Cette analyse est souvent constructive en ce sens que la preuve d'existence consiste bien souvent en l'élaboration d'une méthode itérative permettant d'obtenir la solution de l'équation implicite. Dès lors que l'estimateur robuste est identifié comme bien défini, l'étape naturelle pour l'analyse des performances de l'estimateur, est de décrire ses propriétés statistiques. Cependant, comme dans la plupart des études d'estimation en traitement du signal sous des modèles plus compliqués que linéaires gaussiens, les performances des estimateurs robustes se prêtent plus facilement à une analyse asymptotique, c'est-à-dire lorsque le nombre  $n$  d'échantillons est supposé grandissant, qu'à une analyse exacte à dimension finie. C'est pourquoi les premiers indices de performance sont la convergence asymptotique, pour démontrer par exemple que l'estimateur robuste est consistant (souvent au même titre que l'estimateur empirique standard) ainsi que les statistiques limites de second ordre, souvent des théorèmes de la limite centrale assurant une performance accrue de l'estimateur robuste par rapport aux estimateurs classiques [49, 50, 51]. Ces indices de performance résultent généralement de résultats génériques ou améliorés de la théorie asymptotique des  $M$ - et  $Z$ -estimateurs, voir par exemple [53]. Ils trouvent alors un usage fondamental dans l'analyse des performances d'algorithmes et méthodes de traitement du signal basés sur ces estimateurs robustes (la delta-méthode et le lemme de Slutsky [53] permettent souvent de déplacer les résultats asymptotiques sur les estimateurs robustes vers les méthodes en question). Un dernier indicateur de performances, souvent plus accessible mathématiquement, est la comparaison des performances des estimateurs traditionnels et des estimateurs robustes à des observations gaussiennes. Cette analyse permet de mesurer la détérioration de la précision induite par l'estimateur robuste sur données gaussiennes. Conjointement à l'analyse des performances dans un environnement non gaussien, cette dernière permet de mesurer le compromis robustesse-précision de l'estimateur et est particulièrement utile en pratique lorsque l'environnement observé est non stationnaire avec des périodes impulsives et des périodes non impulsives.

Pour ce qui concerne l'analyse des méthodes robustes à des observations à queue lourde, ces analyses asymptotiques constituent l'essentiel du traitement nécessaire en traitement du signal. La robustesse à des observations fallacieuses appelle quant à elle des critères supplémentaires qui évaluent au sens littéral du terme la robustesse à des données d'amplitude arbitrairement grande. Généralement, la robustesse est mesurée à l'aide de deux critères: (i) la fonction d'influence qui mesure la sensibilité de l'estimateur à l'introduction d'une quantité très faible d'observations fallacieuses d'amplitude donnée, et (ii) le point de bascule (breakdown point) qui mesure la quantité maximale de données fallacieuses d'amplitude arbitraire pouvant être introduites tout en maintenant l'estimateur borné, voir [54, Chapitre 3].

## 6.2 Origine de la problématique

L'intérêt du projet dans le cadre de l'ANR DIONISOS porte plus spécifiquement sur les estimateurs robustes à des observations à queue lourde. En effet, le traitement d'antennes (plus particulièrement à bande étroite) au cœur du projet trouve de nombreuses applications pour lesquelles les signaux observés sont entachés de bruits impulsifs. C'est le cas notamment dans certaines applications de type radar civil ou militaire lorsque les échantillons temporels sont récupérés successivement de différentes positions physiques, donnant ainsi lieu à un bruit de fond présentant des amplitudes très variées (présence d'un objet tel qu'un arbre dans l'axe du radar), ou lorsque le bruit de fond évolue physiquement à une position fixe donnée (effet régulier des vagues en observation maritime, mouvement de fond, etc.). Dans ces cas pratiques, l'hypothèse classique du bruit gaussien est souvent mise à mal de sorte que les simulations numériques n'adhèrent que très peu aux observations physiques. Plusieurs modèles physiques de bruits ont ainsi été proposés qui permettent à la fois une bonne proximité aux phénomènes physiques observés, une paramétrisation aisée, ainsi qu'une possibilité d'analyse mathématique accessible, voir par exemple [55, 56, 57]. Dans le domaine du traitement du signal radar, la distinction est d'ailleurs faite entre le bruit blanc thermique toujours présent et le bruit additionnel non gaussien, souvent appelé "clutter".

Le lien entre la théorie de l'estimation robuste et le domaine du traitement du signal radar ont ainsi été clairement mis en relation bien qu'un assez faible nombre de travaux n'ait exploré finement cette voie [58, 59]. Dans ces travaux, la matrice de covariance de la composante bruit (ou du clutter) est estimée à l'aide d'estimateurs robustes, ces estimateurs étant alors utilisés au sein de détecteur et estimateurs classiques en traitement d'antennes. Le modèle de détection en réception d'antennes utilisé est de la forme

$$y = \begin{cases} w & , \mathcal{H}_0 \\ \alpha p + w & , \mathcal{H}_1 \end{cases}$$

où  $y, p, w \in \mathbb{C}^N$ ,  $p$  un signal souvent connu présent exclusivement sous l'hypothèse  $\mathcal{H}_1$ ,  $\alpha > 0$  un paramètre scalaire d'amplitude et  $w$  le bruit (ou bruit et clutter) additif de moyenne nulle et de covariance  $C_N$ . Se reposant sur des observations de bruit pur  $w_1, \dots, w_n$ , le détecteur de source utilisé en [58] consiste en le test d'hypothèse suivant

$$\frac{|y^* \hat{C}_N^{-1} p|^2}{y^* \hat{C}_N^{-1} y \cdot p^* \hat{C}_N^{-1} p} \underset{\mathcal{H}_1}{\overset{\mathcal{H}_0}{\leq}} \gamma$$

pour un pallier de test donné  $\gamma > 0$ , où  $\hat{C}_N$  est un estimateur de  $C_N$ . Il est démontré, à la fois théoriquement et par le biais de simulations, que l'utilisation d'estimateurs robustes pour  $C_N$  améliore sensiblement la performance du test de détection de source lorsque le bruit  $w$  est fortement impulsif. Dans un contexte de bruit gaussien, un étonnant lien entre les performances obtenues à l'aide de l'estimateur robuste et de la covariance empirique est exhibé qui montre que, dans ce contexte, l'estimateur robuste a sensiblement les mêmes performances (asymptotiques) que la covariance empirique obtenue à partir d'une proportion non triviale de ses observations. Ainsi, l'estimateur robuste apporte une résistance aux observations impulsives au prix d'une nécessité accrue en nombre d'échantillons. Mentionnons par ailleurs que l'estimateur robuste utilisé ici n'est défini (avec probabilité un) que si  $n > N$ .

Evidemment, au même titre que les détecteurs et estimateurs classiques en traitement du signal, l'analyse asymptotique des performances des estimateurs robustes n'a jusqu'alors portée que sur le régime où  $N$  est fixe et  $n \rightarrow \infty$ . Pour tenir compte des performances à horizon fini ( $N$  et  $n$  petits) ou pour des systèmes antennaires de grande dimension ( $N$  et  $n$  grands), il est donc nécessaire dans un premier temps d'étudier les performances de ces estimateurs dans le régime où  $N, n \rightarrow \infty$  et  $N/n$  est non trivial. Pour ce faire, conformément à la discussion de la Section 6.1, il est important d'étudier avant toute chose le comportement asymptotique des estimateurs robustes. C'est l'objectif même de ce projet. Cette étude devrait permettre de mieux comprendre le comportement des détecteurs et estimateurs obtenus par le biais d'outils robustes dans le régime des grandes matrices aléatoires. L'obtention de tels résultats, qui à n'en pas douter vont témoigner du caractère souvent biaisé des estimateurs classiques (au même titre que l'algorithme MUSIC [39]), permettra alors dans un deuxième temps de produire de nouveaux détecteurs et estimateurs améliorés tenant compte à la fois de l'aspect impulsif des observations et du rapport non trivial  $N/n$  (tel que l'algorithme G-MUSIC [38]) même si ce deuxième objectif, très ambitieux, est quelque peu en marge du projet. C'est donc bien un pont encore non exploré entre les deux domaines que ce projet tente de dresser. Notons tout de même que [60] a dressé une première étude d'un estimateur robuste de covariance dans le cas où  $N, n \rightarrow \infty$  mais où  $N/n \rightarrow 0$ . L'étude porte spécifiquement sur la localisation des valeurs propres extrêmes de l'estimateur robuste dans ce contexte.

## 6.3 Les estimateurs de Maronna et de Tyler

### 6.3.1 L'estimateur de Maronna

La famille des estimateurs robustes étant très large, nous n'avons jusqu'alors pas explicité l'expression prise par des estimateurs spécifiques. Notre intérêt, centré sur l'estimation des matrices de covariance, ne se porte donc pas sur des estimateurs robustes de moyenne (ou de localisation en général) mais sur des estimateurs robustes de covariance pour des variables aléatoires centrées. Nous supposons donc que nous avons accès à  $n$  observations vectorielles  $x_1, \dots, x_n$  d'une variable aléatoire  $x \in \mathbb{C}^N$  de moyenne nulle et de covariance (ou dispersion)  $C_N$ . Ces observations seront supposées indépendantes.



Notre étude consiste en l'analyse mathématique du comportement d'une classe d'estimateurs  $\hat{C}_N$  de  $C_N$  définis (lorsque c'est possible) comme une solution de l'équation en  $Z$

$$Z = \frac{1}{n} \sum_{i=1}^n u \left( \frac{1}{N} x_i^* Z^{-1} x_i \right) x_i x_i^* \quad (6.15)$$

où  $u : \mathbb{R}^+ \rightarrow \mathbb{R}^+$  est une fonction satisfaisant un certain nombre de critères. Cette expression générale a été exhibée pour la première fois par Maronna [50] sous les conditions suivantes pour  $u$ :

- $u$  est décroissante sur  $[0, \infty)$  (en particulier  $u(0)$  existe)
- $\phi : x \mapsto xu(x)$  est croissante et bornée supérieurement par  $\phi_\infty > 1$
- $\phi$  est strictement croissante sur l'intervalle maximal où  $\phi(x) \leq 1$ .

Avant toute chose, remarquons que  $n \geq N$  est nécessaire à la bonne définition d'un tel objet. Comme rappelé plus haut, cet estimateur coïncide avec l'estimateur du maximum de vraisemblance pour un choix approprié de la fonction  $u$  lorsque  $x$  suit une distribution elliptique centrée. Maronna généralise même l'estimateur au cas non centré qui ne nous préoccupe pas ici. Il est à noter que l'estimateur robuste initialement développé par Huber [49] fait partie de la famille des estimateurs de Maronna. Ce dernier correspond au cas simple où  $\phi(x) = x$  pour  $x \in [0, K]$  pour un certain  $K > 1$  et  $\phi(x) = K$  sinon. Cet estimateur a un effet de seuillage qui conserve à l'identique les données "mutuellement semblables" et plafonne les données fallacieuses. D'ailleurs, si  $K$  est proche de 1, un nombre important de données sont saturées, renforçant ainsi le caractère robuste de l'estimateur, tandis qu'une valeur plus large de  $K$  rapproche l'estimateur de Huber d'un simple estimateur de covariance empirique (de manière exacte lorsque  $K = \infty$ ).

Parmi les résultats démontrés par Maronna et qui nous intéressent au premier chef, nous trouvons les suivants (écrits sous une forme moins générale que Maronna mais témoignant suffisamment des limitations discutées par la suite):

1. Sous les conditions définies pour  $u$  et sous les hypothèses additionnelles que  $n > N$ , que toute sous-famille de  $N$  vecteurs de  $\{x_1, \dots, x_n\}$  est libre et que  $\phi_\infty > n/(n - N)$ , l'équation (6.15) a une solution unique que l'on dénotera  $\hat{C}_N$  et qui est obtenue à l'aide de l'algorithme standard du point fixe en itérant à partir de  $Z = I_N$
2. Sous les mêmes hypothèses que ci-dessus, dans le régime  $N$  fixe et  $n \rightarrow \infty$ ,  $\hat{C}_N \rightarrow V_N$  presque sûrement, où  $V_N$  est l'unique solution de l'équation

$$V_N = E \left[ \frac{1}{n} \sum_{i=1}^n u \left( \frac{1}{N} x_i^* V_N^{-1} x_i \right) x_i x_i^* \right].$$

Le résultat 1. est exprimé sous forme statistique, en ce sens qu'il est vrai pour tout  $\{x_1, \dots, x_n\}$  indépendamment de toute considération probabiliste, ce qui le rend extrêmement intéressant en ce sens qu'il permet à  $\hat{C}_N$  d'être toujours défini (du moins avec probabilité un) si  $x$  est une variable à densité. Cependant, ce résultat dépend fortement de l'hypothèse  $\phi_\infty > n/(n - N)$  qui, dans le cadre où  $n \gg N$  n'a que peu d'incidence mais qui, lorsque  $N/n \rightarrow c > 0$  induit une contrainte sur la fonction  $u$ . Cette contrainte, comme discuté dans [50] et comme vu plus haut dans le cas particulier de l'estimateur de Huber, a pour effet de limiter la robustesse de  $\hat{C}_N$ . Intuitivement, plus  $\phi_\infty$  est large, plus faibles sont les possibilités d'estomper l'effet d'observations impulsives, réduisant ainsi le pouvoir de robustesse. Parmi les objectifs de notre étude, il sera ainsi question de généraliser le résultat 1. de Maronna pour tolérer des valeurs plus faibles de  $\phi_\infty$ . De premières avancées dans ce sens suggèrent qu'il est possible de retrouver le résultat de Maronna sans cette contrainte mais au prix d'une expression probabiliste, à savoir que l'existence et l'unicité de  $\hat{C}_N$  ne peuvent être généralement garanties que pour une distribution donnée de  $x$  et seront souvent exprimés sous forme probabiliste.

Quant au résultat 2., il est notamment utilisé en pratique lorsque  $x$  est elliptique centré et  $u$  choisi de sorte que  $V_N = C_N$  (ce qui est toujours possible). Dans ce cas,  $\hat{C}_N$  est un estimateur consistant pour  $C_N$ , au même titre que  $\frac{1}{n} \sum_{i=1}^n x_i x_i^*$  lorsque  $E[xx^*] = C_N$ . Cependant, d'une part les vitesses de convergence obtenues à l'aide de  $\hat{C}_N$  sont

accrues et d'autre part  $\hat{C}_N$  est toujours un estimateur consistant de la matrice de dispersion lorsque la covariance n'est pas définie, rendant ainsi  $\hat{C}_N$  bien plus attractive. Dans le cadre de notre étude du régime  $N, n \rightarrow \infty, N/n \rightarrow c$ , il est assez naturellement attendu que la différence  $\|\hat{C}_N - C_N\|$  ne tendent pas vers zéro (où la norme matricielle choisie ici sera la norme spectrale). L'objectif central de ce projet porte précisément sur la caractérisation (notamment spectrale) de  $\hat{C}_N$  dans le régime des matrices aléatoires pour une certaine classe de distribution de  $x$ .

Il faut bien voir que le projet semble ambitieux tant la structure matricielle de  $\hat{C}_N$  comme solution implicite de (6.15) est fondamentalement différente des structures classiques des matrices aléatoires étudiées à ce jour [?, 61]. En effet, les objets matriciels standards en théorie des matrices aléatoires à grandes dimensions se basent fondamentalement sur l'indépendance, ou toute structure simple de dépendance (souvent linéaire), entre les entrées des matrices étudiées. Les matrices de covariance empirique étudiées dans le cadre de DIONISOS s'expriment en particulier sous la forme  $XX^*$  où  $X$  est une matrice dont les entrées ont une structure simple de dépendance. Ici, si l'on écrit  $\hat{C}_N = XDX^*$  avec  $X = [x_1, \dots, x_n]$  et  $D = \text{diag}(\{u(\frac{1}{N}x_i^*V_N^{-1}x_i)\})$ , il est clair que les éléments de  $D$  ont une dépendance forte avec les éléments de  $X$ , ce qui fait de  $\hat{C}_N$  un objet encore largement inexploré dans le domaine des matrices aléatoires et contre lequel aucun outil connu à ce jour ne peut s'attaquer simplement. Néanmoins, nous allons voir en Section 6.4 que certaines intuitions du domaine des matrices aléatoires restent valides et permettent tout du moins d'inférer le comportement asymptotique de  $\hat{C}_N$ , étape essentielle afin d'obtenir une preuve de ces résultats.

### 6.3.2 L'estimateur de Tyler

Dans le domaine de l'estimation robuste de matrices de covariance (ou de dispersion), l'estimateur de Tyler [51, 52] est souvent préféré à l'estimateur de Maronna et ce pour plusieurs raisons. D'une part, l'estimateur de Tyler s'exprime de manière unique, sans paramétrisation possible, qui évite la nécessité parfois d'un choix arbitraire pour la fonction  $u$ . Plus important, les fluctuations limites de l'estimateur sont indépendants de la distribution elliptique de  $x$  sous-jacente. Par ailleurs, en cas d'absence totale de connaissance sur les paramètres de la loi elliptique pour  $x$ , l'estimateur de Tyler est en fait l'estimateur le plus robuste en ce sens que l'estimateur a une variance plus faible que la pire des variances des estimateurs de Maronna mal-adaptés [51, Remarque 3.1]. Enfin, l'expression de l'estimateur de Tyler est simple, mathématiquement tractable dans l'horizon  $n \rightarrow \infty$  et donne des résultats très simples d'analyse.

Cet estimateur de Tyler se présente en fait sous la forme d'une solution de (6.15) mais pour  $u(x) = 1/x$ , fonction qui ne satisfait pas les conditions de Maronna et pour laquelle, en particulier,  $\phi(x) = xu(x) = 1$ . La matrice  $\hat{C}_N$  est ainsi solution de l'équation en  $Z$

$$Z = \frac{1}{n} \sum_{i=1}^n \frac{x_i x_i^*}{\frac{1}{N} x_i^* \hat{Z}^{-1} x_i}. \quad (6.16)$$

A nouveau,  $n \geq N$  est une condition nécessaire à l'existence de  $\hat{C}_N$ . Cependant, et de manière fondamentale, si (6.16) a une solution  $\hat{C}_N$ , toute la droite  $\mathbb{R}\hat{C}_N$  est aussi clairement une solution. La question d'unicité se pose dans ce cas à un paramètre scalaire près. En particulier, un usage classique de l'estimateur de Tyler est de considérer des distributions elliptiques pour  $x$  qui rend le problème équivalent à supposer  $x$  gaussien (voir Section 6.4 pour plus de détails), simplifiant ainsi grandement l'analyse.

Tyler, le premier à étudier précisément cet objet, démontra [51] que  $\hat{C}_N$  est défini de manière unique (à scalaire près) dès lors que  $n > N$ , que toute sous-famille de taille  $N$  de  $\{x_1, \dots, x_n\}$  est libre et que  $n > N(N-1)$ . Les techniques de preuve de Tyler suivent sensiblement celles de Huber et Maronna, toutes basées sur des arguments d'algèbre linéaire. Plus tard, Pascal ses et co-auteurs [59] généralisèrent grandement ce résultat en supprimant l'hypothèse contraignante  $n > N(N-1)$ . L'approche élégante utilisée est très différente et exploite le fait que  $\hat{C}_N$  est l'argument d'un problème de maximum de vraisemblance, utilisant ainsi pleinement le caractère  $M$ -estimateur de  $\hat{C}_N$ . Ce résultat rend l'estimateur de Tyler très stable et de fait plus intéressant que l'estimateur de Maronna dont l'existence n'est assurée que dès lors qu'une hypothèse liant  $u, N$  et  $n$  est satisfaite. De la même manière que Maronna, Tyler assure la convergence  $\hat{C}_N \rightarrow V_N$  avec  $V_N = E[\frac{1}{n} \sum_{i=1}^n \frac{x_i x_i^*}{\frac{1}{N} x_i^* V_N^{-1} x_i}]$  dans le régime  $N$  fixe et  $n \rightarrow \infty$ . Dans le cas particulier où  $x$  a une distribution elliptique,  $V_N = \alpha C_N$  pour un certain paramètre  $\alpha > 0$ .

Dans le contexte de l'analyse de  $\hat{C}_N$  lorsque  $N, n \rightarrow \infty, N/n \rightarrow c$ , la question d'existence et unicité ne se pose pas puisque déjà résolue entièrement dans [59]. Quant à la question du comportement asymptotique de  $\hat{C}_N$ , à nouveau, des intuitions naturelles de la théorie des matrices aléatoires suggèrent que, pour  $x$  elliptique (en fait, même pour  $x$  suivant des lois un peu plus générales satisfaisant de bonnes propriétés de concentrations de mesure),  $\hat{C}_N$  devrait avoir un comportement asymptotique assez simple. De nombreuses simulations suggèrent par ailleurs que ces intuitions sont vraisemblablement correctes. Néanmoins, le fait que  $u(x) = 1/x$  ne satisfasse pas les conditions de Maronna pose des difficultés importantes qui rendent ce problème véritablement plus inaccessible que le problème équivalent posé par l'estimateur de Maronna.

## 6.4 Pistes d'étude

Le projet portant sur l'estimation robuste se divise en plusieurs pistes dont les niveaux de difficultés et les probabilités de succès sont variés. Ces pistes se divisent ainsi:

1. **Estimateur de Maronna pour des vecteurs i.i.d.** Ce projet préliminaire consiste en l'analyse de  $\hat{C}_N$  du type Maronna lorsque  $x \in \mathbb{C}^N$  est à entrées indépendantes et identiquement distribuées de distribution quelconque. L'intérêt théorique de cette étude est d'obtenir un premier résultat sur  $\hat{C}_N$  dans un cas d'application simple et pour lequel l'intuition sur le résultat attendu est assez élémentaire. Les applications pratiques en traitement du signal se limitent au cas de réseaux d'antennes distribués étant soumis aux mêmes conditions de bruits impulsifs, ainsi qu'à des réseaux d'antennes proches subissant indépendamment des bruits impulsifs (par exemple, saturations, problèmes RF, etc.).
2. **Estimateur de Maronna pour des vecteurs elliptiques.** Ce second projet étend la piste 1. à un contexte plus standard où  $x$  suit une distribution elliptique. La difficulté mathématique est accrue en ce sens que l'intuition sur le résultat finale est moins claire d'une part et que le caractère "queue lourde" de  $x$  représente une difficulté supplémentaire non négligeable. D'un point de vue pratique, l'analyse de cet estimateur a un intérêt fondamental dans le contexte de réseaux d'antennes colocalisées, telles que des lignes radar, et qui subissent conjointement des bruits impulsifs.
3. **Estimateur de Tyler pour des vecteurs elliptiques.** Etant donné que l'étude de l'estimateur de Tyler pour des distributions elliptiques est équivalent à supposer les distributions gaussiennes, la difficulté de cette étude se trouve être bien moindre que celle des pistes 1. et 2. Cependant, les difficultés intrinsèques à l'estimateur de Tyler, évoquées précédemment, soulèvent énormément de problèmes qui, à ce jour, n'ont pas trouvé de directions raisonnables d'exploration. Ainsi, cette piste 3. se trouve être le défi le plus ambitieux du projet.
4. **Performance des algorithmes basés sur les estimateurs robustes.** Dès lors qu'une des pistes 2. ou 3. aura apporté des résultats concluant, les performances des méthodes connues d'estimation robuste à des algorithmes de traitement d'antennes à bande étroite pourront être évaluées dans le régime des matrices aléatoires. Parmi ces méthodes, les techniques de détections rappelées plus haut ainsi que les méthodes sous-espaces basées sur l'algorithme MUSIC seront étudiées.
5. **Algorithmes améliorés reposant sur les estimateurs robustes.** De manière bien plus hypothétique, si la piste 4. apporte des résultats concluant, la finalité du projet consisterait à exhiber de nouveaux algorithmes permettant d'améliorer les performances des méthodes standards étudiées en piste 4.

La piste 1. repose sur l'hypothèse où  $x = [x^{(1)}, \dots, x^{(N)}]^T$  où les  $x^{(i)}$  sont des variables i.i.d. centrées, de variance unité et de moments d'ordres supérieurs larges ou infinis. L'intuition provenant des objets standards de théorie des matrices aléatoires suggère que les formes quadratiques  $\frac{1}{N} x_i^* \hat{C}_N^{-1} x_i, i = 1, \dots, n$ , vont avoir un comportement de convergence conjointe vers une même valeur limite déterministe, disons  $\ell$ . Si cet état de fait est avéré, on serait alors en droit de penser que  $\hat{C}_N$  elle-même a alors un comportement asymptotique proche de celui de la matrice  $\hat{S}_n = \frac{1}{n} \sum_{i=1}^n u(\ell) x_i x_i^*$

qui est un objet bien connu en matrice aléatoire et dont la loi limite des valeurs propres suit la loi de Marčenko–Pastur [37]. La généralisation au cas où  $x = C_N^{\frac{1}{2}}y$  avec  $y$  a entrées i.i.d. centrées est alors aisée.

La piste 2. reprend la piste 1. mais avec  $x = \tau C_N^{\frac{1}{2}}y$ , où  $y$  est de norme fixe (que l'on prendra égale à  $\sqrt{N}$ ) et unitairement invariant et où  $\tau > 0$  est une variable aléatoire indépendante de  $x$ . Cette écriture caractérise complètement l'ensemble des variables aléatoires de  $\mathbb{C}^N$  a distribution elliptique centrée. Pour cette seconde étude, les formes quadratiques  $\frac{1}{N}x_i^* \hat{C}_N^{-1}x_i$ ,  $i = 1, \dots, n$ , ne convergent pas vers une limite commune. Il est cependant naturel de concevoir que chacune aura une limite qui ne dépendra que de  $\tau_i$  (où on écrit  $x_i = \tau_i \hat{C}_N^{\frac{1}{2}}y_i$ ) et d'une fonction  $f$  de l'ensemble des  $\{\tau_j\}$ , de telle manière que  $\hat{C}_N$  sera asymptotiquement proche d'une matrice de la forme  $\frac{1}{n} \sum_{i=1}^n f(\tau_i; \{\tau_j\}_{j=1}^n) x_i x_i^*$ .

L'écriture  $x = \tau C_N^{\frac{1}{2}}y$  lorsque  $x$  suit une distribution elliptique justifie pleinement les arguments précédents mentionnant que l'estimateur de Tyler pour  $x$  elliptique est équivalent au même estimateur pour  $x$  gaussien. En effet, dans l'expression de  $\hat{C}_N$ , le rapport  $x_i x_i^* / (x_i^* \hat{C}_N^{-1} x_i)$  permet de modifier arbitrairement la norme de  $x_i$  sans affecter  $\hat{C}_N$ . Comme  $y_i$  est unitaire et unitairement invariant,  $\tilde{x}_i = \kappa_i C_N^{\frac{1}{2}}y_i$  avec  $\kappa_i$  indépendant de  $y_i$  et de distribution Rayleigh rend  $\tilde{x}_i$  gaussien et ne diffère de  $x_i$  que par la norme. On se retrouve donc en ce sens dans un cas simplifié de la piste 1. où  $\tilde{x}_i$  a des entrées i.i.d. (et même gaussiennes donc ayant des moments de tout ordre). L'intuition de matrice aléatoire pour ce modèle est également assez simple tant il est attendu que  $\frac{1}{N}x_i^* \hat{C}_N^{-1}x_i$ ,  $i = 1, \dots, n$ , convergent conjointement vers une même limite  $\ell'$  (indépendamment de la constante multiplicative cette fois ci). A un scalaire près,  $\hat{C}_N$  devrait donc se comporter comme  $\frac{1}{N} \sum_{i=1}^n x_i x_i^*$ , à nouveau un objet bien connu. Cependant, comme annoncé plus haut, l'estimateur de Tyler ne se prête que difficilement à une preuve de ce résultat, tant la propriété  $\phi$  croissante de l'estimateur de Maronna apporte de facilité d'analyse.

## 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] L. Albéra, A. Ferréol, P. Chevalier, P. Comon, "ICAR, a tool for blind source separation using fourth order statistics only", *IEEE Transactions On Signal Processing*, vol. 53, n°10, part 1, pp. 3633-3643, October 2005.
- [3] Z. D. Bai and J. W. Silverstein. No eigenvalues outside the support of the limiting spectral distribution of large-dimensional sample covariance matrices. *Ann. Probab.*, 26(1):316–345, 1998.
- [4] Z. D. Bai and J. W. Silverstein. Exact separation of eigenvalues of large-dimensional sample covariance matrices. *Ann. Probab.*, 27(3):1536–1555, 1999.
- [5] Z. D. Bai and J. W. Silverstein. No eigenvalues outside the support of the limiting spectral distribution of information-plus-noise type matrices. *Random Matrices: Theory and Applications*, 1(1), 2012.
- [6] Z. Bai, J.W. Silverstein, "Spectral analysis of large dimensional random matrices", Springer Series in Statistics, 2nd ed., 2010.
- [7] A. Basak, A. Bose, S. Sen, "Limiting spectral distribution of sample autocovariance matrices", Preprint, can be downloaded on Arxiv, arXiv:1108.3147v1.
- [8] 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.
- [9] F. Benaych-Georges, R.R. Nadakuditi, "The singular values and vectors of low rank perturbations of large rectangular random matrices", submitted, arXiv:1103.2221.

- [10] P. Bianchi, M. Debbah, M. Maïda and J. Najim, "Performance of Statistical Tests for Single Source Detection using Random Matrix Theory", *IEEE Inf. Theory*, Vol. 57 (4), april 2011, 2400–2419.
- [11] P.J. Bickel, E. Levina, "Regularized estimation of large covariance matrices", *Ann. of Statistics*, 2008, vol. 36, 199-227.
- [12] P.J. Bickel, E. Levina, "Covariance regularization by thresholding", *Ann. of Statistics*, vol. 36, no. 6, pp. 2577–2604, 2008.
- [13] A. Boutet de Monvel, A. Khorunzhy, and V. Vasilchuk. Limiting eigenvalue distribution of random matrices with correlated entries. *Markov Process. Related Fields*, 2(4):607–636, 1996.
- [14] R.R. Far, T. Oraby, W. Bryc, R. Speicher, "Spectra of large block matrices", Preprint available on Arxiv, arXiv:cs/0610045.
- [15] T. Cai, Z. Ren, H.H. Zhou, "Optimal Rates of Convergence for estimating Toeplitz covariance matrices", *Probability Theory and Related Fields*, 156, 101-143.
- [16] J. B. Billingsley, "Ground clutter measurements for surface-sited radar", Massachusetts Inst. Technol., Cambridge, MA, Tech. Rep. 780, Feb. 1993.
- [17] T. Cai, C-H Zhang, H.H. Zhou, "Optimal rates of convergence for covariance matrix estimation", *Ann. of Statistics*, 2010, vol. 38, no. 4, 2118-2144.
- [18] M. Capitaine, C. Donati-Martin, and D. Féral. The largest eigenvalues of finite rank deformation of large Wigner matrices: convergence and nonuniversality of the fluctuations. *Ann. Probab.*, 37(1):1–47, 2009.
- [19] R. Brent Dozier and J. W. Silverstein. Analysis of the limiting spectral distribution of large dimensional information-plus-noise type matrices. *J. Multivariate Anal.*, 98(6):1099–1122, 2007.
- [20] R. Brent Dozier and J. W. Silverstein. On the empirical distribution of eigenvalues of large dimensional information-plus-noise-type matrices. *J. Multivariate Anal.*, 98(4):678–694, 2007.
- [21] Julien Dumont, Walid Hachem, Samson Lasaulce, Philippe Loubaton, Jamal Najim, "On the capacity achieving transmit covariance matrices for MIMO Rician channels: an asymptotic approach", *IEEE Transactions on Information Theory*, vol. 56, no. 3, pp. 1048-1069, March 2010.
- [22] Noureddine El Karoui. Tracy-Widom limit for the largest eigenvalue of a large class of complex sample covariance matrices. *Ann. Probab.*, 35(2):663–714, 2007.
- [23] N. El-Karoui, "Operator norm consistent estimation of large dimension matrices", *Annals of Statistics*, vol. 36, no. 6, pp. 2717-2756, December 2008.
- [24] N. El-Karoui, "The spectrum of kernel random matrices", *Annals of Statistics*, vol. 38, no. 1, pp. 1-50, January 2010.
- [25] N. El-Karoui, "On information plus noise kernel random matrices", *Annals of Statistics*, vol. 38, no. 5, pp. 3191-3216, May 2010.
- [26] J. Evans, D.C.N. Tse, "Large system performance of linear multiuser receivers in multipath fading channels", *IEEE Trans. on Information Theory*, pp. 2059-2078, September 2000.
- [27] V. L. Girko. *Sluchainye matritsy (Russian) [Random Matrices]*. Izdat. Obed. "Višča Škola" pri Kiev. Gosudarstv. Univ., Kiev, 1975.

- [28] V. L. Girko. *Theory of stochastic canonical equations. Vol. I*, volume 535 of *Mathematics and its Applications*. Kluwer Academic Publishers, Dordrecht, 2001.
- [29] V. L. Girko. *Theory of stochastic canonical equations. Vol. II*, volume 535 of *Mathematics and its Applications*. Kluwer Academic Publishers, Dordrecht, 2001.
- [30] V.L. Girko, "An introduction to statistical analysis of random arrays", VSP, The Netherlands, 1998.
- [31] A. Ambainis, A.W. Harrow, M.B. Hastings, "Random tensor theory: extending random matrix theory to mixtures of random products states", *Commun. Math. Phys.*, vol. 310, no. 1, pp. 25-74 (2012).
- [32] 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.
- [33] W. Hachem, O. Khorunzhy, P. Loubaton, J. Najim, and L. Pastur. A new approach for mutual information analysis of large dimensional multi-antenna channels. *IEEE Trans. Inform. Theory*, 54(9):3987–4004, 2008.
- [34] W. Hachem, P. Loubaton, X. Mestre, J. Najim, and P. Vallet. Large information plus noise random matrix models and consistent subspace estimation in large sensor networks. *Random Matrices: Theory and Applications*, 01(02):1150006, 2012.
- [35] P. Loubaton and P. Vallet. Almost sure localization of the eigenvalues in a gaussian information plus noise model—application to the spiked models. *Electronic Journal of Probability*, 16:1934–1959, 2011.
- [36] C. Male, "The norm of polynomials in large random and deterministic matrices", *Probab. Theory Related Fields*, 154 (2012), no. 3-4, 477-532.
- [37] 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.
- [38] X. Mestre. Improved estimation of eigenvalues and eigenvectors of covariance matrices using their sample estimates. *IEEE Trans. Inform. Theory*, 54(11):5113–5129, 2008.
- [39] X. Mestre. On the asymptotic behavior of the sample estimates of eigenvalues and eigenvectors of covariance matrices. *IEEE Trans. Signal Process.*, 56(11):5353–5368, 2008.
- [40] 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.
- [41] D. Paul and J. W. Silverstein. No eigenvalues outside the support of the limiting empirical spectral distribution of a separable covariance matrix. *Journal of Multivariate Analysis*, 100(1):37 – 57, 2009.
- [42] A. Sen, B. Virag, "The top eigenvalue of the random Toeplitz matrix and the sine kernel", submitted, can be downloaded on Arxiv, arxiv:1109.5494, 2011.
- [43] J. W. Silverstein and Sang-Il Choi. Analysis of the limiting spectral distribution of large-dimensional random matrices. *J. Multivariate Anal.*, 54(2):295–309, 1995.
- [44] Pascal Vallet, Philippe Loubaton, Xavier Mestre, "Improved subspace estimation for multivariate observations of high dimension: the deterministic signals case", *IEEE Transactions on Information Theory*, vol. 58, no. 2, pp. 1043-1068, February 2012.
- [45] 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.

- [46] 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.
- [47] W.B. Wu, M. Pourahmadi, "Banding sample autocovariance matrices of stationary processes", *Statistica Sinica*, 19(2009), 1755-1768.
- [48] L. Zhang. *Spectral Analysis of Large Random Matrices*. PhD thesis, National University of Singapore, 2006.
- [49] Huber, Peter J. "Robust estimation of a location parameter." *The Annals of Mathematical Statistics* 35.1 (1964): 73-101.
- [50] Maronna, Ricardo Antonio. "Robust M-estimators of multivariate location and scatter." *The annals of statistics* (1976): 51-67.
- [51] Tyler, David E. "A distribution-free  $M$ -estimator of multivariate scatter." *The Annals of Statistics* 15.1 (1987): 234-251.
- [52] Ollila, Esa, and David E. Tyler. "Distribution-free detection under complex elliptically symmetric clutter distribution." *Sensor Array and Multichannel Signal Processing Workshop (SAM), 2012 IEEE 7th. IEEE*, 2012.
- [53] Van der Vaart, Aad W. *Asymptotic statistics*. Vol. 3. Cambridge university press, 2000.
- [54] Maronna, Ricardo A., R. Douglas Martin, and Victor J. Yohai. *Robust statistics*. J. Wiley, 2006.
- [55] K.D. Ward, "Compound representation of high resolution sea clutter", *Electronics letters*, vol. 17, no. 16, pp. 561-563, August 1981.
- [56] S. Watts, "Radar Detection Prediction in Sea Clutter Using the Compound K-Distribution model", *IEE Proceeding*, Part. F, vol. 132, no. 7, pp. 613-620, December 1985.
- [57] E. Conte and M. Longo, "Characterization of radar clutter as a spherically invariant random process", *IEE Proc.-Pt.F*, vol. 134, no. 2, pp. 191-197, April 1987.
- [58] Pascal, Frédéric, et al. "Performance analysis of covariance matrix estimates in impulsive noise." *Signal Processing, IEEE Transactions on* 56.6 (2008): 2206-2217.
- [59] Pascal, Frédéric, et al. "Covariance structure maximum-likelihood estimates in compound Gaussian noise: Existence and algorithm analysis." *Signal Processing, IEEE Transactions on* 56.1 (2008): 34-48.
- [60] Dümbgen, Lutz. "On Tyler's  $M$ -functional of scatter in high dimension." *Annals of the Institute of Statistical Mathematics* 50.3 (1998): 471-491.
- [61] Pastur, L. Leonid Andreevich, and Mariâ Vladimirovna Scherbina. *Eigenvalue distribution of large random matrices*. Vol. 171. AMS Bookstore, 2011.
- [62] Kent, John T., and David E. Tyler. "Redescending  $M$ -Estimates of Multivariate Location and Scatter." *The Annals of Statistics* 19.4 (1991): 2102-2119.