
Estimation robuste et matrices aléatoires

Romain Couillet

Centrale Supélec
91192 Gif sur Yvette, France
romain.couillet@centralesupelec.fr

RÉSUMÉ. Cet article propose une revue de littérature des récentes avancées en estimation robuste de matrices de dispersion vue sous l'angle de la théorie des grandes matrices aléatoires. Une exposition des résultats théoriques sera effectuée avant de proposer des applications aux domaines des statistiques et du traitement du signal. Les résultats théoriques montrent qu'alors que les estimateurs robustes de matrices de dispersion prennent une forme implicite et ainsi difficile à analyser, dans le régime des grandes matrices aléatoires où la grande taille des échantillons est commensurable à leur nombre, ces estimateurs tendent à se comporter de manière similaire à des modèles matriciels aléatoires simples. Cette observation se traduit par l'apparition de nombreuses propriétés statistiques liées à ces estimateurs. En termes applicatifs, les estimateurs robustes de matrices de dispersion ont une longue histoire en tant qu'éléments structurants permettant d'appréhender des données à queues lourdes ou contenant des observations aberrantes. Cette gestion des impulsivités des données est mieux comprise ici sous l'angle des grandes matrices aléatoires et permet d'introduire des méthodes d'estimation robustes en traitement d'antennes, optimisation de portefeuilles, etc.

ABSTRACT. This article provides a technical survey of the recent advances between the fields of robust estimation of scatter and of large dimensional random matrix theory. An exposition of the theoretical results will be made which we shall apply to various contexts in the area of statistics and signal processing at large. The theoretical results essentially show that, while robust estimators of scatter are implicitly defined and thus difficult objects to manipulate, in the large dimensional random matrix regime where both the population size and the number of samples are simultaneously large, these implicit robust estimators tend to behave similar to much simpler random matrix models, amenable to analysis. This induces that many statistical properties of these estimators could be unearthed which we shall discuss. In terms of applications, these robust estimators of scatter are long-standing structural elements to handle both outliers and heavy-tailed behavior in the observed data. These impulsiveness harnessing effects will be precisely documented and shall be instrumental to develop improved robust statistics methods for detection and estimation in antenna arrays, portfolio optimization, etc.

MOTS-CLÉS : matrices aléatoires, estimation robuste, traitement d'antennes, statistiques.

KEYWORDS: random matrix theory, robust estimation, array processing, statistics.

DOI:10.3166/TS.33.273-320 © 2016 Lavoisier

Extended Abstract

This article provides an overview of the recent works in the area of *robust estimation* and related signal processing applications, in the *large dimensional data regime*. Precisely, the article focuses on the so-called robust estimators of scatter initially developed in the seventies and eighties (Huber, 1964 ; Maronna, 1976). These estimators are designed to compensate for the lack of robustness of classical sample covariance matrices against spurious outliers such as impulsive noise. While showing superior performance in these scenarios, these estimators are however much more complex to study as they assume *implicit* expressions. For instance, letting $x_1, \dots, x_n \in \mathbb{C}^N$ be (potentially heavy-tailed or outlier-corrupted) observed data, the generic expression for the robust estimator of scatter \hat{C}_N reads

$$\hat{C}_N = \frac{1}{n} \sum_{i=1}^n u \left(\frac{1}{N} x_i^* \hat{C}_N^{-1} x_i \right) x_i x_i^*$$

for some nonnegative u function or, whenever $n < N$, reads

$$\hat{C}_N \triangleq \hat{C}_N(\rho) = \frac{1}{n} \sum_{i=1}^n u \left(\frac{1}{N} x_i^* \hat{C}_N(\rho)^{-1} x_i \right) x_i x_i^* + \rho I_N$$

for some $\rho > 0$. That is, the estimator \hat{C}_N is implicitly defined as the (generally made unique) solution to a fixed point equation. The statistical works to understand the behaviour of such estimators have only been restricted so far to the case where N is fixed and $n \rightarrow \infty$, which is of little interest in today's Big Data applications.

Moving to the large dimensional regime with observed data x_i following a pre-described heavy-tailed distribution allows for simplified theoretical analysis thanks to an advantageous concentration of measure phenomenon. In particular, we show in the article that there exists *non observable* but *simple* matrices \hat{S}_N which satisfy

$$\left\| \hat{C}_N - \hat{S}_N \right\| \xrightarrow{\text{p.s.}} 0$$

in operator norm, for a large family of such robust estimators (Couillet, Pascal, Silverstein, 2014 ; Couillet *et al.*, 2015 ; Couillet, McKay, 2014). These matrices \hat{S}_N are claimed simple in that they take on explicit forms and are easily studied using standard methods of random matrix theory. As such, even though not directly observable, these matrices play the important role of a proxy to the understanding and improvement of the performance of robust estimators of scatter.

These theoretical results are notably applied to a wide range of scientific domains, from array processing (Couillet, 2015 ; Couillet, Kammoun, Pascal, 2014 ; Kammoun *et al.*, 2015) down to statistical finance (Yang *et al.*, 2015), where observations are particularly impulsive by nature. In these applications, it is notably possible to appropriately tune the hyperparameters defining the robust estimators (the function u or the regularization parameter ρ) so to optimize a sought-for objective.

1. Introduction

1.1. Statistiques robustes

La récente impulsion du big data remet au goût du jour le domaine de l'estimation robuste, né il y a plus de cinquante ans maintenant sous l'impulsion d'auteurs tels que Tukey et Huber (Huber, 1964 ; Donoho, 2000). Déjà à l'époque, le développement rapide de l'informatique a permis à ces auteurs de comprendre que l'avenir se trouvait dans la multiplication des données et dans le développement de méthodes automatisées de traitement de ces données. Outre le développement parallèle du domaine de l'apprentissage automatisé permettant effectivement de classer et catégoriser les données, un des points importants levés par Tukey et Huber est la capacité de traiter des données à la chaîne en dépit de l'intrusion possible d'erreurs de lectures, d'une absence partielle de données ou encore en présence de bruits impulsifs dans les données. Ces erreurs ne pouvant plus être écartées manuellement dès lors que la taille des données à traiter est importante, il convenait alors de vérifier que les méthodes statistiques standard, basées essentiellement sur des modèles sans erreurs de lecture, étaient fiables dans de tels scénarios. Ce n'est malheureusement pas le cas comme en témoigne tout estimateur de moyenne ou variance empirique (de type $\frac{1}{n} \sum_{i=1}^n x_i$ pour des observations i.i.d. x_i dont on chercherait à estimer la moyenne $E[x_1]$) qui, dès qu'une observation est arbitrairement erronée (disons que x_j prend une valeur arbitrairement grande), crée un biais arbitrairement grand de l'estimateur, résultant en une erreur d'estimation potentiellement non bornée.

Cet état de fait a lancé dans les années soixante et soixante-dix une importante vague de travaux sur les statistiques appelées alors robustes. Ces dernières consistent premièrement en la mise en place de métriques de robustesse d'estimateurs (ainsi que de détecteurs, régresseurs, etc.) contre des observations impulsives ou erronées. Parmi ces critères, on trouve par exemple la fonction d'influence qui mesure l'effet d'un nombre arbitrairement petit mais d'amplitude arbitrairement grande de données erronées (par exemple, l'effet d'une seule erreur de lecture) sur un estimateur donné. Un autre critère est celui du point de rupture, correspondant à la proportion maximale de données erronées qu'un estimateur est capable de tolérer sans devenir arbitrairement biaisé. Une fois les critères de robustesse définis, il s'agit alors de générer de nouveaux estimateurs qui ont la particularité de rester bien sûr fiables sous l'hypothèse de données propres, mais qui ont maintenant l'avantage de résister aux effets indésirables de données erronées. C'est ainsi qu'une large classe d'estimateurs (et derrière eux, tout un domaine de recherche, celui des statistiques robustes) est née.

Les estimateurs qui nous intéressent ici sont les estimateurs de matrices de dispersion (dits aussi de matrices de covariance), qui consistent en l'extension multivariée d'estimateurs d'échelle (aussi dits de variance). Les premiers travaux en ce sens datent de l'article (Huber, 1964) et ont alors pour la plupart suivi l'intuition de ce dernier. L'idée originale d'Huber, qui considérait plutôt des paramètres de localisation (type moyenne) que d'échelle (type variance), a été de définir comme estimateur robuste le M-estimateur étant maximalelement robuste à des observations arbitrairement erro-

nées en ce sens que cet estimateur résout un problème d'optimisation de forme minimax (l'estimateur minimise une certaine fonction de coût paramétrée par un pire cas). Par la suite, sous l'impulsion de (Maronna, 1976), il a été remarqué que l'estimateur minimax d'Huber fait partie d'une classe bien plus large d'estimateurs comprenant également les estimateurs de maximum de vraisemblance pour les paramètres (de localisation et de dispersion) de lois elliptiques. C'est ainsi que la connection entre estimateurs robustes vus comme résilients à des données erronées et estimateurs robustes vus comme adaptés à des modèles non gaussiens (particulièrement à queues lourdes) a été réalisée. En effet, la classe des distributions elliptique a la particularité d'être une extension paramétrable très naturelle des données gaussiennes. Il est spécifiquement possible de paramétrer de telles lois de manière à réduire le nombre de moments finis de la distribution, introduisant ainsi naturellement des queues lourdes. En tant qu'extension simple du modèle gaussien, ces lois elliptiques ont en particulier la propriété agréable d'être facilement étudiables. Il reste néanmoins remarquable que les deux problèmes que sont l'estimation des paramètres d'une distribution à queue lourde et l'estimation robuste vue comme sécurisée face à des données erronées donnent toutes deux lieu aux mêmes objets d'intérêt.

1.2. Estimateur robuste de matrice de dispersion

Cet objet d'intérêt, et qui sera au centre des discussions dans cet article, est défini comme suit : étant données des observations $x_1, \dots, x_n \in \mathbb{C}^N$ (ou \mathbb{R}^N), telles que $n > N$ et telles que tout sous-ensemble de N vecteurs d'entre eux demeure linéairement indépendant (ce qui est vrai avec probabilité un pour des données elliptiques indépendantes par exemple), on appelle \hat{C}_N la matrice définie comme l'unique solution de l'équation:

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

où $u(t)$ est une fonction continue de \mathbb{R}^+ dans \mathbb{R}^+ , décroissante et telle que $t \mapsto \phi(t) \triangleq tu(t)$ est une fonction croissante et bornée. On assurera en particulier également que $\phi_\infty \triangleq \lim_{t \rightarrow \infty} \phi(t) > 1$. Dans le cas de Huber, $u(t)$ est une fonction constante sur un intervalle puis décroissante en $1/t$ au delà d'une valeur seuil. Dans le cas de Maronna, $u(t)$ prend une forme très générale et s'adapte, pour ce qui est des lois elliptiques, aux paramètres de la loi.

Notons premièrement que \hat{C}_N n'a pas de raison d'être définie, étant solution ici d'une équation matricielle à point fixe. Par ailleurs, quand bien même définie, il n'est pas garanti que l'équation à point fixe n'ait qu'une seule solution. Un travail important, qui a malheureusement mobilisé la plupart des travaux dans le domaine, commence donc par l'établissement de l'existence de \hat{C}_N ainsi que par la démonstration qu'une seule solution à l'équation à point fixe existe. Si ceci est admis (ce sera tout le temps le cas dans notre article), il devient alors intéressant d'analyser de près le comportement de \hat{C}_N . Remarquons en particulier que \hat{C}_N n'est ni plus ni moins qu'une matrice de

covariance empirique “pondérée”. Les pondérations sont précisément données par la fonction $u(t)$ qui a pour effet de donner de l’importance aux données pour lesquelles $\frac{1}{N}x_i^* \hat{C}_N^{-1} x_i$ est faible. Ceci est d’autant plus visible dans l’exemple d’Huber qui alloue une valeur constante à tous les $\frac{1}{N}x_i^* \hat{C}_N^{-1} x_i$ de faible amplitude et réduit l’impact des autres. Les données pour lesquelles $\frac{1}{N}x_i^* \hat{C}_N^{-1} x_i$ est fort (celles pour lesquelles les données x_i seront atténuées) correspondent intuitivement (dans le cas où \hat{C}_N est supposé plus ou moins proche de $\mathbb{E}[x_i x_i^*]$) aux données fortement peu alignées à $\mathbb{E}[x_i x_i^*]$ ou ayant des normes importantes, et donc essentiellement à des données erronées. On obtient ainsi bien l’effet escompté. Ainsi, l’estimateur d’Huber, qui stabilise les “bonnes” données et amortit les “mauvaises” données, est essentiellement un estimateur empirique de matrice de covariance ayant la propriété additionnelle de pouvoir atténuer toute donnée absurde. Notons finalement que l’effet de l’équation à point fixe est très remarquable ici. En effet, pour prendre conscience qu’une donnée erronée est marginale, elle doit être comparée à toutes les autres données. Pour ce faire, elle doit donc être projetée sur l’espace engendré par l’ensemble des vecteurs observés. C’est un peu ce qu’intuitivement réalise \hat{C}_N .

1.3. Vers les grandes dimensions

Un certain nombre de résultats sont connus dans la littérature des statistiques robustes concernant \hat{C}_N , et plus particulièrement des résultats asymptotiques lorsque le nombre d’observations n tend vers l’infini. Cependant, il n’en demeure pas moins que les difficultés liées à la définition implicite de \hat{C}_N sont longtemps demeurées problématiques, de sorte que la théorie des statistiques robustes est tombée un moment en désuétude.

Pourtant, la résurgence des modèles de traitement d’antennes tout d’abord (pour les applications en radar) puis le domaine du BigData par la suite ont relancé récemment l’intérêt pour ces méthodes d’estimation robuste sous un nouvel angle. En effet, il est aujourd’hui très classique de supposer dans ces domaines applicatifs que la dimension N des échantillons observés soit commensurable, sinon plus large, que le nombre n de tels échantillons. Les problèmes de bruits impulsifs, propres au radar, et la possibilité d’observer des données erronées en BigData sont quant à eux toujours valables sous cette nouvelle hypothèse et rendent ainsi attractifs le développement d’estimateurs robustes avec $N \simeq n$. Une approche possible, qui s’éloigne du spectre des statistiques robustes à proprement parler, est la vague d’intérêt récente inculquée par les méthodes dites parcimonieuses qui, sous l’hypothèse d’une réduction possible de dimensionalité du problème, permettent de développer des méthodes d’estimation naturellement robustes aux données absurdes. Cependant, ce cadre s’éloigne de notre intérêt présent pour lequel une telle hypothèse de parcimonie n’est pas nécessairement désirable.

Notons dans un premier temps que l’hypothèse $N > n$ conduit à un problème de définition même de l’estimateur \hat{C}_N en (1) étant donné que le membre de droite est singulier. Cette hypothèse amène donc naturellement à considérer de nouveaux

estimateurs, en particulier des estimateurs régularisés de type (Pascal *et al.*, 2014 ; Chen *et al.*, 2011)

$$\hat{C}_N(\rho) \triangleq (1 - \rho) \frac{1}{n} \sum_{i=1}^n u \left(\frac{1}{N} x_i^* \hat{C}_N^{-1}(\rho) x_i \right) x_i x_i^* + \rho I_N$$

où $\rho \in (0, 1]$. Ces estimateurs, qu'on peut démontrer être à nouveau définis de manière univoque, ont l'avantage d'exister pour toute hypothèse sur N et n . Leur analyse ouvre bien évidemment la question d'un choix optimal du paramètre ρ , que nous discuterons dans cet article.

Outre l'introduction de nouveaux estimateurs, la réelle nouveauté de l'hypothèse $N \simeq n$ avec N, n potentiellement larges réside dans la possibilité d'exploiter des résultats de moyennage de variables aléatoires à la fois dans la direction du temps mais aussi dans la direction de l'espace, nous amenant ainsi dans le domaine plus puissant des matrices aléatoires. Grâce aux matrices aléatoires, un phénomène nouveau va apparaître qui n'est aucunement valable dans le cas où N est supposé fixe. Nous allons en effet pouvoir montrer, plus spécifiquement dans l'hypothèse où les x_i sont des vecteurs aléatoires elliptiques (mais aussi lorsque des observations erronées et considérées arbitraires sont présentes), que la matrice \hat{C}_N a la propriété de se comporter asymptotiquement (i.e., lorsque $N, n \rightarrow \infty$ et $N/n \rightarrow c \in (0, \infty)$) de la même manière qu'une matrice aléatoire \hat{S}_N de définition explicite (et non plus solution d'une équation implicite) dont le modèle est très facile à analyser grâce aux outils de la théorie des grandes matrices aléatoires. Précisément, sous des hypothèses appropriées mais en général légères, on montrera que

$$\left\| \hat{C}_N - \hat{S}_N \right\| \rightarrow 0 \quad (2)$$

presque sûrement, où la norme considérée est la norme spectrale, i.e., la valeur propre de valeur absolue dominante. Ce résultat, que nous détaillerons précisément dans l'article, permet de transférer de nombreuses propriétés de \hat{C}_N vers \hat{S}_N et ainsi de permettre leur étude de manière bien plus précise que dans le cas classique où $n \gg N$ où un tel résultat n'est pas valable. Dans le cas régularisé où $\hat{C}_N = \hat{C}_N(\rho)$, on trouvera même que

$$\sup_{\rho \in \mathcal{I}} \left\| \hat{C}_N(\rho) - \hat{S}_N(\rho) \right\| \rightarrow 0$$

pour une certaine matrice $\hat{S}_N(\rho)$ paramétrisée par ρ et où l'ensemble \mathcal{I} est un intervalle dont nous préciserons plus tard la valeur exacte. Cette convergence uniforme en ρ permet en particulier de sélectionner un ρ optimum pour toute métrique donnée de $\hat{C}_N(\rho)$ en transférant cette métrique sur $\hat{S}_N(\rho)$. Une grande partie de cet article sera dédiée à l'introduction précise de ces résultats ainsi qu'à quelques éléments intuitifs de démonstration de ces derniers. Des applications diverses de ces résultats seront alors présentées.

Les convergences de type (2) ne sont par contre pas toujours suffisantes lorsqu'il s'agit d'utiliser \hat{C}_N dans l'expression d'un estimateur consistant d'un paramètre du

système et que les performances de cet estimateur doivent être étudiées. En effet, pour analyser les performances d'estimateurs consistants, il s'agit souvent d'obtenir les fluctuations de l'estimateur, et en particulier des théorèmes de la limite centrale. Dans ces cas particuliers, le comportement de $\sqrt{N}(\hat{C}_N - \mathbb{E}[\hat{C}_N])$ et non plus de \hat{C}_N est demandé. Malheureusement, il ne semble pas correct que (2) reste valable avec un pré-facteur \sqrt{N} , de sorte que le transfert des théorèmes de la limite centrale sont bien moins évidents. Malgré tout, pour des cas précis, nous allons montrer que cela reste tout de même possible. Une application précise en détection pour les systèmes radar sous bruits impulsifs sera en particulier présentée.

Le reste de l'article s'organise comme suit. Dans la section 2, nous passons en revue les résultats concernant le comportement asymptotique de l'estimateur robuste de matrice de dispersion de Maronna \hat{C}_N dans le régime des grandes matrices aléatoires, sous l'hypothèse où les x_i sont indépendants, identiquement distribués de loi elliptique. Des applications en traitement d'antennes sont discutées avec en particulier l'introduction d'un nouvel estimateur de sous-espace de type MUSIC pour l'estimation d'angles d'arrivées sous bruit impulsif. Dans un deuxième temps, en section 3, l'estimateur de Maronna est étudié dans le cadre où les x_i sont gaussiens indépendants avec haute probabilité ou déterministes inconnus avec faible probabilité. Cela permet en particulier de modéliser l'occurrence peu fréquente de données absurdes parmi les données observées. Pour permettre d'étendre l'étude des estimateurs de Maronna au cas où $N > n$, nous évoquons alors en section 4 les estimateurs régularisés $\hat{C}_N(\rho)$ discutés plus haut ; pour des raisons de simplicité d'analyse, nous prendrons par ailleurs dans ce chapitre l'hypothèse que $u(x) = 1/x$. Enfin, en section 5, nous évoquons les résultats de type théorèmes de la limite centrale pour des formes bilinéaires de noyau \hat{C}_N , i.e., de la forme $a^* \hat{C}_N^k b$ pour des vecteurs $a, b \in \mathbb{C}^N$ de norme unité et $k \in \mathbb{Z}$. L'article se termine par des remarques de conclusion en section 6.

Quelques notations : dans la suite de cet article, les matrices seront dénotées en majuscules, les vecteurs en minuscule. Pour une matrice C hermitienne, nous écrivons $C \succeq 0$ pour indiquer qu'elle est définie positive. Dans toute la suite, la norme $\|\cdot\|$ est la norme spectrale pour les matrices hermitiennes et la norme euclidienne pour les vecteurs. La convergence presque sûre sera dénotée " $\xrightarrow{\text{p.s.}}$ ". La mesure de Dirac en x est dénotée par δ_x . On définit $\mathbb{C}_+ = \{z \in \mathbb{C}, \Im[z] > 0\}$.

2. Comportement asymptotique des estimateurs robustes de type Maronna

2.1. Notations et hypothèses

Notre première étude concerne l'estimateur de matrice de dispersion de Maronna, introduit brièvement en (1), où nous rappelons que $u(x)$ est une fonction décroissante telle que $\phi(x) = xu(x)$ est croissant et de limite finie $\phi_\infty = \lim_{x \rightarrow \infty} \phi(x) > 1$. Nous considérons tout d'abord que les dimensions N des échantillons x_1, \dots, x_n et leur nombre n sont telles que, lorsque $n \rightarrow \infty$, $c_N \triangleq N/n$ vérifie l'hypothèse suivante.

SUPPOSITION 1. — Pour chaque n , $c_N < 1$ et

$$c_- < \liminf_n c_N \leq \limsup_n c_N < c_+$$

où $0 < c_- < c_+ < 1$. □

Nous travaillerons particulièrement ici sous l’hypothèse d’observations $x_1, \dots, x_n \in \mathbb{C}^N$ indépendantes et de distribution similaire (en fait plus générale dans un certain sens) à une loi elliptique. Plus précisément, nous effectuerons l’hypothèse suivante.

SUPPOSITION 2. — Les vecteurs $x_i = \sqrt{\tau_i} C_N^{\frac{1}{2}} w_i$, $i \in \{1, \dots, n\}$, satisfont les propriétés suivantes :

1. la distribution empirique des amplitudes aléatoires $\tilde{\nu}_n = \frac{1}{n} \sum_{i=1}^n \delta_{\tau_i}$ vérifie $\int \tau \tilde{\nu}_n(d\tau) \xrightarrow{\text{p.s.}} 1$
2. il existe $\varepsilon < 1 - \phi_\infty^{-1} < 1 - c_+$ et $m > 0$ tel que, pour tout n large presque sûrement, $\tilde{\nu}_n([0, m]) < \varepsilon$
3. $C_N \succ 0$ et $\limsup_N \|C_N\| < \infty$
4. $w_1, \dots, w_n \in \mathbb{C}^N$ sont indépendants, identiquement distribués, complexe unitairement invariant, de moyenne nulle telle que, pour tout i , $\|w_i\|^2 = N$, et sont indépendants des τ_1, \dots, τ_n . □

Ce qu’il est important de noter ici est que le vecteur x_i peut être essentiellement vu comme un vecteur unitairement invariant (donc sans direction de pointage particulière) de norme contrôlée par le paramètre τ_i sur lequel les contraintes sont légères. Les τ_i peuvent être arbitrairement définis mais s’ils sont indépendants et de même distribution, alors les observations x_i sont indépendantes et de loi elliptique centrée, paramétrée par la distribution des τ_i . En choisissant les τ_i de sorte que $2N\tau_i$ suit une loi du chi-carré de $2N$ degrés de libertés, alors les x_i sont gaussiens et sont à queue légère (comme en témoigne le fait que $\tau_i \xrightarrow{\text{p.s.}} 1$ dans ce cas). Si au contraire, on choisit pour les τ_i des lois à support (asymptotique) non compact, la nature des x_i devient alors impulsive. Par exemple, si $1/\tau_i$ est de loi chi-carré de degré de liberté fixe, indépendant de N , alors x_i suit une loi de Student centrée de paramètre adapté au degré de liberté de la loi du chi-carré. Il est ainsi possible de paramétrer les observations pour les rendre plus ou moins impulsives.

Les différentes hypothèses de supposition 2 non discutées jusqu’alors son essentiellement techniques et sans grande portée pratique.

Nous considérons maintenant la définition précise de l’estimateur de matrice de dispersion de Maronna. Nous appelons \hat{C}_N la matrice définie, lorsqu’elle existe, comme la solution unique de l’équation en $Z \in \mathbb{C}^{N \times N}$

$$Z = \frac{1}{n} \sum_{i=1}^n u \left(\frac{1}{N} x_i^* Z^{-1} x_i \right) x_i x_i^* \tag{3}$$

où $x_1, \dots, x_n \in \mathbb{C}^N$ sont vus ici comme une collection quelconque de vecteurs et où u satisfait les propriétés suivantes :

- (i) $u : [0, \infty) \rightarrow (0, \infty)$ est positive, continue et décroissante
- (ii) $\phi : x \mapsto xu(x)$ est croissante et bornée, avec $\lim_{x \rightarrow \infty} \phi(x) \triangleq \phi_\infty > 1$
- (iii) $\phi_\infty < c_+^{-1}$.

La dernière propriété est une hypothèse technique, traditionnellement non rencontrée dans les travaux originels de Maronna. Elle nous permettra ici d'opérer un changement de variable important pour la suite. Notons que cette hypothèse est assez légère en ce sens que pour un estimateur robuste, il est souvent demandé que ϕ_∞ soit très proche de 1 (renforçant ainsi son caractère robuste). Comme $c_N > 1$, cette hypothèse est assez peu restrictive.

Une dernière hypothèse technique, sur laquelle nous allons nous pencher quelques instants, s'avère importante (et vraisemblablement nécessaire) pour l'obtention de notre résultat central. Celle-ci peut être écrite comme suit.

SUPPOSITION 3. — Pour chaque $a > b > 0$, presque sûrement,

$$\limsup_{t \rightarrow \infty} \frac{\limsup_n \tilde{\nu}_n((t, \infty))}{\phi(at) - \phi(bt)} = 0.$$

□

Cette hypothèse nous dit en essence que le poids de la queue de la distribution empirique des τ_i doit rester faible relativement au taux de croissance de la fonction ϕ . Notons en effet que le numérateur, dans le cas de τ_i indépendants et identiquement distribués s'écrit simplement $P(\tau_1 > t)$. Si notre étude permet d'appréhender des distributions elliptiques, elle reste cependant limitée à cette contrainte. Malgré tout, pour des fonctions $u(t)$ classiques, telles que $u(t) = (1+t)/(t+\alpha)$, $\alpha > 0$, utilisée classiquement dans la littérature, et pour τ_i identiquement distribués et indépendants, il est facile de voir que, pour satisfaire la supposition 3, il est suffisant que $E[\tau_1^{(1+\varepsilon)}] < \infty$ pour un $\varepsilon > 0$ quelconque. Pour des raisons propres à la théorie des matrices aléatoires, il est important que $E[\tau_1] < \infty$, sans quoi la plupart des résultats à suivre ne seraient pas corrects. Il paraît donc relativement léger d'imposer, dans le cas particulier que nous venons d'évoquer, que $E[\tau_1^{(1+\varepsilon)}] < \infty$ pour ε arbitrairement petit.

Equipés de ces notations, nous allons maintenant décrire de manière intuitive et non rigoureuse les quelques éléments de la théorie des matrices aléatoires permettant d'obtenir notre résultat central, à savoir la convergence de type (2) que nous introduirons précisément plus tard à l'occasion du théorème au cœur de cette section.

2.2. *Éléments de calcul intuitifs*

2.2.1. *Préliminaire sur l'existence de \hat{C}_N*

Nous introduisons tout d'abord un résultat concernant l'existence et l'unicité de \hat{C}_N , ainsi qu'une méthode itérative permettant d'évaluer \hat{C}_N en pratique. Ce résultat

est essentiellement la conséquence de l'article (Kent, Tyler, 1991) et n'est pas l'objet précis de la présente étude.

THÉORÈME 4 (Existence et unicité). — *Supposons que les hypothèses 1 et 2 soient en place, avec $\limsup_N \|C_N\|$ non nécessairement bornée. Alors, pour tout N, n , presque sûrement, (3) admet une unique solution \hat{C}_N définie par*

$$\hat{C}_N = \lim_{\substack{t \rightarrow \infty \\ t \in \mathbb{N}}} Z^{(t)}$$

où $Z^{(0)} \succ 0$ est arbitraire et, pour $t \in \mathbb{N}$,

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

Etant donné le résultat du théorème, nous pouvons dorénavant librement utiliser la notation \hat{C}_N , en se restreignant au pire à un espace de probabilité un.

2.2.2. Une intuition de matrices aléatoires

Notre première étape consiste en une réécriture de \hat{C}_N sous une forme plus pratique pour notre analyse. Remarquons tout d'abord qu'il est possible de se restreindre à l'étude de \hat{C}_N avec $C_N = I_N$. En effet, en définissant par exemple $\tilde{C}_N = C_N^{-\frac{1}{2}} \hat{C}_N C_N^{\frac{1}{2}}$, il est facile de voir que \tilde{C}_N est solution de (3) mais pour les vecteurs $C_N^{-\frac{1}{2}} x_i$ au lieu de x_i . Il nous suffit donc de comprendre le comportement de \hat{C}_N pour $C_N = I_N$ en premier lieu, avant de pré- et post-multiplier \hat{C}_N par $C_N^{\frac{1}{2}}$ pour obtenir le cas général. A partir de maintenant et pour le reste de cette section, nous supposons donc que $C_N = I_N$.

Notre intuition principale, mais pourtant non rigoureuse, consiste à remarquer que, quand bien même \hat{C}_N impose une structure de dépendance forte entre les données x_1, \dots, x_n , la matrice

$$\hat{C}_{(j)} \triangleq \frac{1}{n} \sum_{i \neq j} u \left(\frac{1}{N} x_i^* \hat{C}_N^{-1} x_i \right) x_i x_i^*$$

a elle une dépendance apparemment "assez faible" en x_j . En effet, le vecteur x_j n'apparaît pas dans la somme des produits vectoriels $x_i x_i^*$ mais apparaît à travers \hat{C}_N dans les formes quadratiques $\frac{1}{N} x_i^* \hat{C}_N^{-1} x_i$. Cependant, cette apparition à travers \hat{C}_N reste elle-même d'effet a priori très restreint étant donné que x_j n'est qu'un vecteur parmi n (grandissant) contribuant à \hat{C}_N . Il paraît donc naturel à première vue de quasiment supposer $\hat{C}_{(j)}$ comme indépendant de x_j , même si cela n'est pas vrai.

De cette intuition naît alors la conjecture fondamentale suivante : il existe un paramètre scalaire déterministe γ_N tel que

$$\max_{1 \leq j \leq n} \left| \frac{1}{N} w_j^* \hat{C}_{(j)}^{-1} w_j - \gamma_N \right| \xrightarrow{\text{p.s.}} 0. \tag{4}$$

Cette intuition provient de trois résultats classiques en matrices aléatoires, que nous rappelons présentement (voir par exemple (Bai, Silverstein, 2009, Annexe B) pour plus de détails). Le premier résultat est le lemme de trace suivant.

LEMME 5 (Lemme de la trace). — *Soit $x \in \mathbb{C}^N$ un vecteur à entrées indépendantes et identiquement distribuées, de moyenne nulle, de variance un, et de moment d'ordre $8 + \varepsilon$ fini. Soit $A_N \in \mathbb{C}^{N \times N}$ une matrice aléatoire ou déterministe, indépendante de w et telle que $\limsup_N \|A_N\| < \infty$. Alors,*

$$\frac{1}{N} w^* A_N w - \frac{1}{N} \operatorname{tr} A_N \xrightarrow{\text{p.s.}} 0.$$

Ce résultat fondateur des matrices aléatoires (utilisé en particulier dans l'article (Silverstein, Bai, 1995)) se généralise aisément à la convergence uniforme pour N^β réalisations de vecteurs w si tant est que l'hypothèse sur le moment d'ordre $8 + \varepsilon$ fini s'étend à des moments plus élevés. En particulier, pour w gaussien (distribution ayant des moments de tous ordres), le résultat s'étend très facilement à N vecteurs w . Dans notre cas d'intérêt, deux problèmes se posent. Le premier concerne le fait que w_i n'est pas gaussien mais unitairement invariant. Ce problème est très vite contourné par le fait qu'alors on peut écrire $w_i = \sqrt{N} \tilde{w}_i / \|\tilde{w}_i\|$ pour \tilde{w}_i un vecteur gaussien à entrées de moyenne nulle et disons de variance un. En utilisant le fait que $\|\tilde{w}_i\|/\sqrt{N} \xrightarrow{\text{p.s.}} 1$ (convergence que l'on peut aisément rendre uniforme pour $i = 1, \dots, n$), on se ramène alors au cas gaussien. Le second problème est par contre plus délicat et concerne précisément le fait que w_i et $\hat{C}_{(j)}$ ne sont effectivement pas indépendants. Le résultat (4) ne peut donc pas être validé et restera sous la forme de conjecture. En l'état, nous conjecturerons que

$$\max_{1 \leq j \leq n} \left| \frac{1}{N} w_j^* \hat{C}_{(j)}^{-1} w_j - \frac{1}{N} \operatorname{tr} \hat{C}_{(j)}^{-1} \right| \xrightarrow{\text{p.s.}} 0.$$

Le second lemme standard dont nous avons maintenant besoin concerne l'approximation de l'inverse d'une matrice par l'inverse de la même matrice perturbée par une matrice de rang un. Précisément, nous avons le résultat suivant :

LEMME 6 (Perturbation de rang-1). — *Soit A_N une matrice inversible telle que $\limsup \|A_N^{-1}\| < \infty$ et soit $w \in \mathbb{C}^N$ un vecteur quelconque (déterministe ou aléatoire). Alors,*

$$\frac{1}{N} \operatorname{tr} A_N^{-1} - \frac{1}{N} \operatorname{tr} (A_N + ww^*)^{-1} \rightarrow 0.$$

Ce lemme, pour peu que $\|\hat{C}_{(j)}^{-1}\|$ reste borné uniformément sur i et sur N (chose que nous conjecturons également), nous permet alors d'assurer que

$$\max_{1 \leq j \leq n} \left| \frac{1}{N} w_j^* \hat{C}_{(j)}^{-1} w_j - \frac{1}{N} \operatorname{tr} \hat{C}_{(j)}^{-1} \right| \xrightarrow{\text{p.s.}} 0$$

et donc que toutes les formes quadratiques $\frac{1}{N} w_j^* \hat{C}_{(j)}^{-1} w_j$ tendent à s'approcher d'une même valeur. La prochaine étape dans la pratique usuelle de la théorie des matrices aléatoires consisterait en temps normal à démontrer que $\frac{1}{N} \text{tr} \hat{C}_N^{-1}$ s'approche d'une quantité déterministe que l'on cherchera à identifier. Cependant, du fait de la nature implicite de \hat{C}_N , les méthodes usuelles ne permettent pas de s'approcher d'un tel résultat. Pourtant, étant donnée la proximité dans son écriture de \hat{C}_N à la matrice de covariance empirique $\sum_i x_i x_i^*$ et étant donnée notre intuition selon laquelle les formes quadratiques $\frac{1}{N} w_j^* \hat{C}_{(j)}^{-1} w_j$ se concentrent, il vient naturellement à l'esprit que $\frac{1}{N} \text{tr} \hat{C}_N^{-1}$ devrait devenir déterministe dans la limite des grands N, n . Nous conjecturons donc à nouveau que cela est le cas et dénotons cet équivalent déterministe γ_N . Cela donne finalement le résultat escompté (4), mais au prix d'un grand nombre d'hypothèses potentiellement fallacieuses.

Dans l'article (Couillet *et al.*, 2015) dont le résultat que nous cherchons à démontrer est issu, le cœur technique de la preuve rigoureuse du résultat consiste précisément à démontrer (4) de manière rigoureuse, en utilisant des outils extrêmement différents. Ces outils ne présentent cependant pas d'intérêt pédagogique et ne seront donc pas discutés plus longuement. Toutefois, il est important de mentionner qu'une difficulté majeure de la preuve rigoureuse consiste en un contrôle très fin des τ_i . Cette difficulté amène à introduire la supposition 3 discutée précédemment.

2.2.3. Quelques relations algébriques

Admettons donc que (4) est valable. Ce résultat porte en lui l'essence même de l'apport des matrices aléatoires au domaine de l'estimation robuste. En effet, par ce résultat, les formes quadratiques $\frac{1}{N} x_j^* \hat{C}_{(j)}^{-1} x_j = \tau_j \frac{1}{N} w_j^* \hat{C}_{(j)}^{-1} w_j$ deviennent bien plus simples d'analyse. Cependant, dans la définition de \hat{C}_N , les formes quadratiques qui entrent en jeu sont les formes $\frac{1}{N} x_j^* \hat{C}_N^{-1} x_j$ et non pas les formes $\frac{1}{N} x_j^* \hat{C}_{(j)}^{-1} x_j$ qui ne sont pas nécessairement aisées à relier entre elles. Notre seconde étape consiste donc à exhiber un lien entre ces deux objets.

Cette étape est essentiellement algébrique et ne présente pas de difficulté majeure. Ecrivons pour cela \hat{C}_N comme une fonction de $\frac{1}{N} w_i^* \hat{C}_{(i)}^{-1} w_i$. Pour ce faire, utilisons l'identité $(A + \tau z z^*)^{-1} z = A^{-1} z / (1 + \tau z^* A^{-1} z)$ valable pour toute matrice A inversible, vecteur z et scalaire $\tau > 0$ que nous appliquons à notre problème. Nous obtenons

$$\frac{1}{N} w_i^* \hat{C}_N^{-1} w_i = \frac{\frac{1}{N} w_i^* \hat{C}_{(i)}^{-1} w_i}{1 + \tau_i u \left(\tau_i \frac{1}{N} w_i^* \hat{C}_N^{-1} w_i \right) \frac{1}{n} w_i^* \hat{C}_{(i)}^{-1} w_i}$$

de sorte que

$$\frac{1}{N} w_i^* \hat{C}_{(i)}^{-1} w_i \left(1 - c_N \tau_i u \left(\tau_i \frac{1}{N} w_i^* \hat{C}_N^{-1} w_i \right) \frac{1}{N} w_i^* \hat{C}_N^{-1} w_i \right) = \frac{1}{N} w_i^* \hat{C}_N^{-1} w_i$$

qui, par la définition de ϕ , n'est autre que

$$\frac{1}{N} w_i^* \hat{C}_{(i)}^{-1} w_i \left(1 - c_N \phi \left(\tau_i \frac{1}{N} w_i^* \hat{C}_N^{-1} w_i \right) \right) = \frac{1}{N} w_i^* \hat{C}_N^{-1} w_i.$$

D'après la Supposition 1 et le fait que $\phi_\infty < c_+^{-1}$, en prenant n assez large pour assurer que $\phi(x) \leq \phi_\infty < 1/c_N$, cette équation peut se réécrire

$$\frac{1}{N} w_i^* \hat{C}_{(i)}^{-1} w_i = \frac{\frac{1}{N} w_i^* \hat{C}_N^{-1} w_i}{1 - c_N \phi \left(\tau_i \frac{1}{N} w_i^* \hat{C}_N^{-1} w_i \right)}. \tag{5}$$

C'est à ce moment qu'intervient l'hypothèse importante selon laquelle ϕ est une fonction strictement croissante. En effet, de par cette hypothèse, la fonction $g : [0, \infty) \rightarrow [0, \infty)$, $x \mapsto x/(1 - c_N \phi(x))$ est croissante, positive et projette $[0, \infty)$ sur $[0, \infty)$. Ainsi, g est une fonction inversible dont nous écrivons l'inverse g^{-1} . En particulier, grâce à (5), nous avons

$$\tau_i \frac{1}{N} w_i^* Z^{-1} w_i = g^{-1} \left(\tau_i \frac{1}{N} w_i^* \hat{C}_{(i)}^{-1} w_i \right).$$

Définissons maintenant la fonction $v : [0, \infty) \rightarrow [0, \infty)$, $x \mapsto u \circ g^{-1}$. Comme g est strictement croissante et positive, et u est décroissante, v est décroissante et positive. Par ailleurs, $\psi : x \mapsto xv(x)$ est telle que

$$\psi(x) = xu(g^{-1}(x)) = g(g^{-1}(x))u(g^{-1}(x)) = \frac{\phi(g^{-1}(x))}{1 - c_N \phi(g^{-1}(x))}$$

elle même croissante, positive et de limite $\psi_\infty \triangleq \phi_\infty/(1 - c_N \phi_\infty)$ lorsque $x \rightarrow \infty$. Ainsi, v et ψ conservent les propriétés de u et ϕ , respectivement.

Finalement, grâce à ces nouvelles notations, \hat{C}_N peut se définir de manière équivalente comme

$$\hat{C}_N = \frac{1}{n} \sum_{i=1}^n \tau_i v \left(\tau_i \frac{1}{N} w_i^* \hat{C}_{(i)}^{-1} w_i \right) w_i w_i^*.$$

Définissons $d_j \triangleq \frac{1}{N} w_j^* \hat{C}_{(j)}^{-1} w_j$. En soustrayant l'entrée j de la somme et en prenant la forme quadratique de l'inverse du résultat, il est alors assez facile de voir que l'existence et l'unicité de \hat{C}_N est strictement équivalente à l'existence et l'unicité du vecteur $(d_1, \dots, d_n) \in (\mathbb{R}^+)^n$ comme solution des n équations :

$$d_j = \frac{1}{N} w_j^* \left(\frac{1}{n} \sum_{i \neq j} \tau_i v(\tau_i d_i) w_i w_i^* \right)^{-1} w_j, \quad 1 \leq j \leq n. \tag{6}$$

2.2.4. Finalisation de la preuve, identification de γ_N

En revenant maintenant légèrement sur nos pas, rappelons que nous avons conjecturé dans (4) que

$$\max_{1 \leq j \leq n} |d_j - \gamma_N| \xrightarrow{\text{p.s.}} 0 \tag{7}$$

pour un certain γ_N qu’il nous reste à identifier. Grâce à la réécriture (6), il vient en particulier que

$$\begin{aligned} & \left\| \hat{C}_N - \frac{1}{n} \sum_{i=1}^n \tau_i v(\tau_i \gamma_N) w_i w_i^* \right\| \\ &= \left\| \frac{1}{n} \sum_{i=1}^n \tau_i (v(\tau_i d_i) - v(\tau_i \gamma_N)) w_i w_i^* \right\| \\ &\leq \max_{1 \leq i \leq n} |\tau_i v(\tau_i d_i) - \tau_i v(\tau_i \gamma_N)| \left\| \frac{1}{n} \sum_{i=1}^n w_i w_i^* \right\| \\ &= \max_{1 \leq i \leq n} |\psi(\tau_i d_i) d_i^{-1} - \psi(\tau_i \gamma_N) \gamma_N^{-1}| \left\| \frac{1}{n} \sum_{i=1}^n w_i w_i^* \right\|. \end{aligned}$$

En utilisant le fait que ψ est bornée et la conjecture (7), il vient facilement que le premier terme tend vers zéro. Quant à la matrice du second terme, elle peut être écrite sous la forme $\frac{1}{n} \tilde{W} T \tilde{W}^*$ où $\tilde{W} = [\tilde{w}_1, \dots, \tilde{w}_n]$ avec, comme précédemment, $w_i = \sqrt{N} \tilde{w}_i / \|\tilde{w}_i\|$ pour \tilde{w}_i gaussien à entrées de moyenne nulle et de variance un, et avec $T = \text{diag}(\sqrt{N} / \|\tilde{w}_i\|)$. Comme $T \xrightarrow{\text{p.s.}} I_N$ par des arguments standard de loi des grands nombres (uniformément sur les éléments de T), on a alors que $\left\| \frac{1}{n} \tilde{W} T \tilde{W}^* - \frac{1}{n} \tilde{W} \tilde{W}^* \right\| \xrightarrow{\text{p.s.}} 0$ et il suffit donc de contrôler la norme de $\frac{1}{n} \tilde{W} \tilde{W}^*$. Mais ce contrôle est très connu, et provient de l’article (Bai, Silverstein, 1998), selon lequel $\left\| \frac{1}{n} \tilde{W} \tilde{W}^* \right\| - (1 + \sqrt{c_N})^2 \xrightarrow{\text{p.s.}} 0$, à savoir que la valeur propre la plus large de $\frac{1}{n} \tilde{W} \tilde{W}^*$ converge vers le bord droit du spectre de la loi de Marčenko–Pastur. Ainsi, $\left\| \frac{1}{n} \sum_{i=1}^n w_i w_i^* \right\|$ est borné pour tout N presque sûrement, et on obtient finalement que

$$\left\| \hat{C}_N - \frac{1}{n} \sum_{i=1}^n \tau_i v(\tau_i \gamma_N) w_i w_i^* \right\| \xrightarrow{\text{p.s.}} 0. \tag{8}$$

Nous obtenons donc que \hat{C}_N se comporte asymptotiquement comme une matrice plus simple, constituée par la somme de produits vectoriels $w_i w_i^*$ (qu’on peut supposer comme essentiellement gaussiens) pondérés par un scalaire indépendant des w_i et qu’on peut donc considérer comme déterministe. Il reste cependant toujours à identifier γ_N .

Pour cela, à nouveau nous allons devoir utiliser des résultats standard de la théorie des matrices aléatoires. Rappelons, dans le déroulement de la preuve intuitive de (4) que nous avons défini γ_N comme un équivalent déterministe de la quantité $\frac{1}{N} \text{tr} \hat{C}_N^{-1}$. Cette dernière quantité n'est autre que la transformée de Stieltjes de la distribution des valeurs propres de \hat{C}_N évaluée en zéro. En effet, rappelons ici que pour une mesure de probabilité μ , la transformée de Stieltjes est la quantité $m(z) = \int (t - z)^{-1} \mu(dt)$ définie pour tout $z \in \mathbb{C} \setminus \text{supp}(\mu)$, avec $\text{supp}(\mu)$ le support de la loi μ . En prenant pour loi μ la distribution empirique des valeurs propres $\lambda_1, \dots, \lambda_N$ de \hat{C}_N , i.e., $\mu = \frac{1}{N} \sum_{i=1}^N \delta_{\lambda_i}$, on trouve alors que $m(z) = \frac{1}{N} \text{tr} \text{diag}((\lambda_i - z)^{-1}) = \frac{1}{N} \text{tr}(\hat{C}_N - zI_N)^{-1}$. En particulier $m(0) = \frac{1}{N} \text{tr} \hat{C}_N^{-1}$. Par le résultat (8) et par des arguments de convergence dominée, il vient que

$$m(0) - \frac{1}{N} \text{tr} \left(\frac{1}{n} \sum_{i=1}^n \tau_i v(\tau_i \gamma_N) w_i w_i^* \right)^{-1} \xrightarrow{\text{p.s.}} 0$$

Le deuxième terme du membre de gauche est bien connu et il a été démontré successivement dans (Silverstein, Bai, 1995 ; Couillet *et al.*, 2011) et bien d'autres articles que

$$\frac{1}{N} \text{tr} \left(\frac{1}{n} \sum_{i=1}^n \tau_i v(\tau_i \gamma_N) w_i w_i^* \right)^{-1} - \tilde{m}(0) \xrightarrow{\text{p.s.}} 0$$

où $\tilde{m}(0)$ est l'unique solution positive de

$$\tilde{m}(0) = \left(\frac{1}{n} \sum_{i=1}^n \frac{\tau_i v(\tau_i \gamma_N)}{1 + c_N \tau_i v(\tau_i \gamma_N) \tilde{m}(0)} \right)^{-1}.$$

Mais comme $\tilde{m}(0)$ est asymptotiquement équivalent à $m(0)$ lui-même asymptotiquement équivalent à γ_N , il vient que l'on peut simplement définir $\gamma_N = \tilde{m}(0)$ qui devient ainsi défini comme l'unique solution positive de l'équation

$$\gamma_N = \left(\frac{1}{n} \sum_{i=1}^n \frac{\tau_i v(\tau_i \gamma_N)}{1 + c_N \tau_i v(\tau_i \gamma_N) \gamma_N} \right)^{-1}$$

ou autrement dit, l'unique solution de l'équation

$$1 = \frac{1}{n} \sum_{i=1}^n \frac{\psi(\tau_i \gamma_N)}{1 + c_N \psi(\tau_i \gamma_N)}. \quad (9)$$

Cette identité conclut notre preuve intuitive et nous sommes dès lors en position de présenter notre résultat principal et ses conséquences.

2.3. Résultats principaux et interprétations

Notre résultat principal peut être présenté comme suit.

THÉORÈME 7 (Comportement limite). — *Sous les hypothèses Suppositions 1–3, soit \hat{C}_N comme défini dans Théorème 4 (et pris arbitraire lorsque non défini). Alors*

$$\left\| \hat{C}_N - \hat{S}_N \right\| \xrightarrow{\text{p.s.}} 0$$

où

$$\hat{S}_N \triangleq \frac{1}{n} \sum_{i=1}^n v(\tau_i \gamma_N) x_i x_i^*$$

avec γ_N l'unique solution strictement positive de l'équation en γ

$$1 = \frac{1}{n} \sum_{i=1}^n \frac{\psi(\tau_i \gamma)}{1 + c_N \psi(\tau_i \gamma)}$$

et avec $v : x \mapsto (u \circ g^{-1})(x)$, $\psi : x \mapsto xv(x)$, et $g : \mathbb{R}_+ \rightarrow \mathbb{R}_+$, $x \mapsto x/(1 - c_N \phi(x))$.

Le résultat stipule donc que \hat{C}_N , dont la définition est donnée comme solution de l'équation implicite (3) en Z , devient asymptotiquement bien approximée par \hat{S}_N dont la définition est complètement explicite, si ce n'est pour le paramètre γ_N dont la définition demeure implicite ; cependant, remarquons que γ_N dépend uniquement de τ_1, \dots, τ_n et est indépendant des vecteurs w_1, \dots, w_n . Conditionnellement aux τ_i , \hat{S}_N est un modèle matriciel aléatoire bien connu et étudié (essentiellement) déjà dans l'article de Marčenko–Pastur en 1967 (Marčenko, Pastur, 1967) pour $C_N = I_N$, puis plus tard dans (Silverstein, Bai, 1995 ; Silverstein, Choi, 1995), et par la suite dans (Hachem *et al.*, 2007 ; Couillet, Hachem, 2014).

Du fait de la convergence $\|\hat{C}_N - \hat{S}_N\| \xrightarrow{\text{p.s.}} 0$, un certain nombre des propriétés de \hat{S}_N déterminées dans ces articles se transfère immédiatement à \hat{C}_N . En particulier, notons que théorème 7 implique les propriétés suivantes :

– les valeurs propres ordonnées $\lambda_1(\hat{C}_N) \geq \dots \geq \lambda_N(\hat{C}_N)$ et $\lambda_1(\hat{S}_N) \geq \dots \geq \lambda_N(\hat{S}_N)$ vérifient

$$\max_{1 \leq i \leq N} \left| \lambda_i(\hat{C}_N) - \lambda_i(\hat{S}_N) \right| \xrightarrow{\text{p.s.}} 0$$

et donc nous avons ici une caractérisation très précise des valeurs propres individuelles de \hat{C}_N ;

– si $\mu_N = \frac{1}{N} \sum_{i=1}^N \delta_{\lambda_i(\hat{C}_N)}$ et $\bar{\mu}_N = \frac{1}{N} \sum_{i=1}^N \delta_{\lambda_i(\hat{S}_N)}$ dénotent les distributions empiriques des valeurs propres de \hat{C}_N et \hat{S}_N , respectivement, alors, $\mu_N - \bar{\mu}_N \xrightarrow{\text{p.s.}} 0$ où la convergence est comprise ici comme la convergence faible en mesure. En particulier, si $\bar{\mu}_N$ converge vers une mesure limite déterministe (ou possède une mesure de probabilité déterministe équivalente), alors μ_N possède la même mesure limite.

Le cas elliptique où les τ_i sont indépendants et identiquement distribués est en particulier intéressant et donne lieu au corollaire suivant.

COROLLAIRE 8 (Cas elliptique). — *Sous les Suppositions 1–3 avec τ_i i.i.d. de loi $\tilde{\nu}$ et $c_N \rightarrow c \in (0, 1)$,*

$$\left\| \hat{C}_N - \frac{1}{n} \sum_{i=1}^n v(\tau_i \gamma^\infty) x_i x_i^* \right\| \xrightarrow{\text{p.s.}} 0$$

où γ^∞ est l'unique solution positive de l'équation en γ

$$1 = \int \frac{\psi_c(t\gamma)}{1 + c\psi_c(t\gamma)} \tilde{\nu}(dt)$$

avec $\psi_c = \lim_{c_N \rightarrow c} \psi$. Par ailleurs, si $\frac{1}{n} \sum_{i=1}^n \delta_{\lambda_i(C_N)} \rightarrow \nu$ faiblement, alors

$$\frac{1}{n} \sum_{i=1}^n \delta_{\lambda_i(\hat{C}_N)} \xrightarrow{\text{p.s.}} \mu$$

faiblement où μ est une mesure de probabilité à densité de support \mathcal{S} borné, dont la transformée de Stieltjes $m(z) = \int (t - z)^{-1} \mu(dt)$ est donnée pour $z \in \mathbb{C}_+$ par

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

avec $\tilde{\delta}(z)$ la solution unique de \mathbb{C}_+ des équations en $\tilde{\delta}$

$$\tilde{\delta} = -\frac{1}{z} \int \frac{\psi_c(t\gamma^\infty)}{\gamma^\infty + \psi_c(t\gamma^\infty)\tilde{\delta}} \tilde{\nu}(dt)$$

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

Enfin, pour tout ensemble fermé $\mathcal{A} \subset \mathbb{R}$ tel que $\mathcal{A} \cap \mathcal{S} = \emptyset$,

$$\left| \left\{ \lambda_i(\hat{C}_N) \right\}_{i=1}^N \cap \mathcal{A} \right| \xrightarrow{\text{p.s.}} 0.$$

Ici, nous trouvons donc que le spectre des valeurs propres de \hat{C}_N converge (presque sûrement) vers une mesure à densité connue à travers sa transformée de Stieltjes. De plus, en utilisant des résultats existants pour le modèle équivalent qui nous intéresse ici (Bai, Silverstein, 1998), il s'avère qu'asymptotiquement aucune valeur propre de \hat{C}_N ne se trouve en dehors du support limite de cette mesure. La Figure 1 présente un exemple graphique où les τ_i sont de loi Gamma. Notons de manière fondamentale ici qu'en écrivant

$$\frac{1}{n} \sum_{i=1}^n v(\tau_i \gamma^\infty) x_i x_i^* = \frac{1}{\gamma^\infty} \frac{1}{n} \sum_{i=1}^n \psi(\tau_i \gamma^\infty) C_N^{\frac{1}{2}} w_i w_i^* C_N^{\frac{1}{2}}$$

si $\limsup_N \|C_N\| < \infty$ (à savoir, les valeurs propres de C_N restent bornées), alors, comme ψ est borné, d'après le corollaire on trouve que $\limsup_N \|\hat{C}_N\| < \infty$. Ce résultat implique donc qu'alors que la matrice de covariance empirique $\frac{1}{n} \sum_{i=1}^n x_i x_i^*$ a un support asymptotique souvent non borné, les valeurs propres de \hat{C}_N restent quant à elles bornées. Nous allons fondamentalement utiliser cette propriété dans la section suivante.

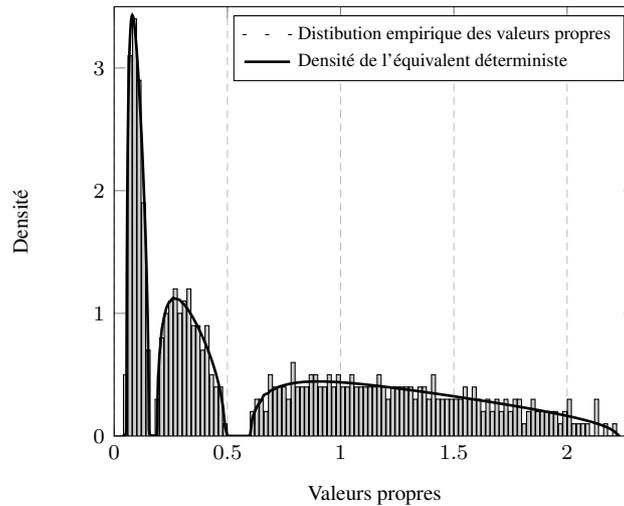


Figure 1. Histogramme des valeurs propres de \hat{C}_N pour $n = 2500$, $N = 500$, $C_N = \text{diag}(I_{125}, 3I_{125}, 10I_{250})$, τ_1 de loi $\Gamma(.5, 2)$. Comparaison avec la loi limite des valeurs propres de la matrice \hat{S}_N

2.4. Application au traitement d'antennes

Un domaine d'application classique de l'estimation robuste de matrices de dispersion est celui du traitement d'antennes et en particulier le radar, où les bruits subis par le réseau d'antenne sont typiquement impulsifs. Dans ce cadre, les méthodes à base de matrices de covariance empiriques sont en général très sensibles au bruit et donc inadaptées. On leur préfère alors des méthodes à base d'estimateurs robustes (Pascal *et al.*, 2014).

Une application spécifique consiste en l'estimation d'angles d'arrivée de sources multiples, au nombre de L , corrompues par un bruit impulsif. Précisément, supposons le modèle suivant

$$x_i = \sum_{l=1}^L \sqrt{p_l} a_l s_{li} + \sqrt{\tau_i} w_i \quad (10)$$

dans lequel, comparativement à la section précédente, on suppose que $C_N = I_N$ (pour des raisons de simplicité) et surtout la présence d'une partie signal $\sum_{l=1}^L \sqrt{p_l} a_l s_{li}$. Cette partie signal est définie comme suit :

1. L est fixe (indépendant de N, n)
2. $p_1 \geq \dots \geq p_L > 0$ sont les puissances positives des sources
3. $a_1, \dots, a_L \in \mathbb{C}^N$ sont les vecteurs directionnels déterministes tels que

$$A^* A \xrightarrow{\text{p.s.}} \text{diag}(p_1, \dots, p_L)$$

quand $N \rightarrow \infty$, où $A \triangleq [\sqrt{p_1} a_1, \dots, \sqrt{p_L} a_L] \in \mathbb{C}^{N \times L}$. En particulier, pour des antennes linéaires uniformément distribuées, il est classique de supposer que $[a_j]_k = \exp(-2\pi i k \sin(\theta_j))$ pour un certain angle d'incidence $\theta_j \in [0, 2\pi)$.

4. $s_{1,1}, \dots, s_{L,n} \in \mathbb{C}$ sont les signaux indépendants de moyenne nulle, de variance un et de moments de tout ordre fini.

Les méthodes d'estimation d'angle traditionnelles (de type MUSIC en particulier (Schmidt, 1986)) consistent à évaluer la matrice de covariance empirique $\frac{1}{n} \sum_{i=1}^n x_i x_i^*$, à en extraire ses plus larges valeurs propres et à utiliser les vecteurs propres associés pour évaluer les paramètres angulaires. Ces méthodes reposent fondamentalement sur le fait qu'asymptotiquement les L plus grandes valeurs propres de $\frac{1}{n} \sum_{i=1}^n x_i x_i^*$ s'isolent des $N - L$ autres valeurs propres. Dans le cadre des matrices aléatoires, on parle de modèles "spike", voir par exemple (Baik, Silverstein, 2006 ; Benaych-Georges, Nadakuditi, 2012 ; Couillet, Hachem, 2013). Dans ce cadre, les résultats principaux peuvent être résumés comme suit :

– d'après (Baik, Silverstein, 2006), dans l'hypothèse où $\max_{1 \leq i \leq n} |\tau_i - 1| \xrightarrow{\text{p.s.}} 0$, si $p_k > \sqrt{c}$ mais $p_{k+1} \leq \sqrt{c}$ (où $c = \lim_n N/n$), alors la distribution empirique des $N - k$ plus petites valeurs propres de $\frac{1}{n} \sum_{i=1}^n x_i x_i^*$ convergent faiblement vers la loi de Marčenko–Pastur et ces valeurs propres ne s'échappent pas de son support $[(1 - \sqrt{c})^2, (1 + \sqrt{c})^2]$, tandis que les k plus grandes valeurs propres convergent vers k valeurs supérieures à $(1 + \sqrt{c})^2$.

– d'après (Couillet, Hachem, 2013), si les τ_i sont i.i.d. de support borné, alors il existe $p_- > 0$ tel que, si $p_k > p_-$ mais $p_{k+1} \leq p_-$, alors les $N - k$ plus petites valeurs propres de $\frac{1}{n} \sum_{i=1}^n x_i x_i^*$ s'amassent dans le support limite de sa distribution empirique, tandis que les k plus grandes s'isolent au-delà du support.

– d'après (Silverstein, Bai, 1995 ; Couillet, Hachem, 2014), si les τ_i sont i.i.d. de support non compact, alors la distribution limite des valeurs propres de $\frac{1}{n} \sum_{i=1}^n x_i x_i^*$ est à support non compact et donc, en particulier, aucune valeur propre ne s'isole.

Pour des modèles de bruits elliptiques, il est naturel de supposer la dernière hypothèse selon laquelle les τ_i sont de support non borné, et il est donc essentiel de recourir à des estimateurs plus performants qui évitent ce problème élémentaire. D'après les conclusions de la section précédente, établis dans le cas d'un bruit elliptique seul, le support limite de la distribution des valeurs propres de \hat{C}_N est borné. Il semble alors naturel, du fait des conclusions du théorème 7, d'imaginer que l'extension au

cas signal-plus-bruit de ce théorème induise les comportements décrits précédemment pour la matrice $\frac{1}{n} \sum_{i=1}^n x_i x_i^*$ lorsque les τ_i sont bornés.

2.4.1. Modèles robustes spike

Cette intuition est confirmée par le résultat suivant, établi dans (Couillet, 2015).

THÉORÈME 9 (Equivalence du modèle signal-plus-bruit). — *Sous les hypothèses du modèle signal-plus-bruit avec τ_i i.i.d. et $c_N \rightarrow c \in (0, 1)$, nous avons*

$$\|\hat{C}_N - \hat{S}_N\| \xrightarrow{\text{p.s.}} 0$$

où

$$\hat{S}_N \triangleq \frac{1}{n} \sum_{i=1}^n v_c(\tau_i \gamma) x_i x_i^*$$

(x_i modélisé différemment ici!) et γ défini comme auparavant comme l'unique solution positive de

$$1 = \int \frac{\psi_c(t\gamma)}{1 + c\psi_c(t\gamma)} \tilde{\nu}(dt).$$

Le nouveau modèle pour \hat{S}_N du théorème précédent s'analyse encore à l'aide de résultats de la théorie des matrices aléatoires pour les modèles "spike". En particulier, il apparaît qu'il existe un seuil p_- tel que, si $p_k > p_-$ tandis que $p_{k+1} \leq p_-$, alors les k plus grandes valeurs propres de \hat{C}_N se situent asymptotiquement au-delà d'une autre valeur seuil. La présence (ou l'absence) de ces valeurs extrêmes permet alors de déterminer un critère de détection et d'énumération de sources. Les vecteurs propres associés à ces valeurs propres permettent alors d'estimer les vecteurs a_i ou, plus spécifiquement ici, d'estimer les angles d'indices des $k \leq L$ sources. Plus précisément, nous avons le résultat suivant, établi dans l'article (Couillet, 2015).

THÉORÈME 10 (Estimation robuste du cas signal-plus-bruit). — *Définissons u_k le vecteur propre associé à la k -ième plus large valeur propre de AA^* (en cas de multiplicité, n'importe quelle base du sous espace propre convient) et $\hat{u}_1, \dots, \hat{u}_N$ les vecteurs propres de \hat{C}_N associés aux valeurs propres $\lambda_1 \geq \dots \geq \lambda_N$ ($\lambda_i = \lambda_i(\hat{C}_N)$). Définissons par ailleurs, pour $x > 0$ suffisamment large, $\delta(x)$ l'unique solution positive de l'équation*

$$\delta(x) = c \left(-x + \int \frac{tv_c(t\gamma)}{1 + \delta(x)tv_c(t\gamma)} \tilde{\nu}(dt) \right)^{-1}.$$

Alors, nous avons les résultats suivants:

1. Valeurs propres extrêmes. Il existe $p_- > 0$ tel que pour chaque j tel que $p_j > p_-$,

$$\lambda_j \xrightarrow{\text{p.s.}} \Lambda_j > S^+ \triangleq \frac{\phi_\infty(1 + \sqrt{c})^2}{\gamma(1 - c\phi_\infty)}$$

tandis que, si $p_k \leq p_-$, $\limsup_n \lambda_k \leq S^+$ a.s., où Λ_j est la solution positive unique de

$$-c \left(\delta(\Lambda_j) \int \frac{v_c(\tau\gamma)}{1 + \delta(\Lambda_j)\tau v_c(\tau\gamma)} \tilde{\nu}(d\tau) \right)^{-1} = p_j.$$

2. Estimation de puissance. Si $p_j > p_-$,

$$-c \left(\delta(\hat{\lambda}_j) \int \frac{v_c(\tau\gamma)}{1 + \delta(\hat{\lambda}_j)\tau v_c(\tau\gamma)} \tilde{\nu}(d\tau) \right)^{-1} \xrightarrow{\text{p.s.}} p_j.$$

3. Estimation de forme bilinéaire. Pour tout $a, b \in \mathbb{C}^N$ tels que $\|a\| = \|b\| = 1$, si $p_j > p_-$, alors

$$\sum_{k, p_k = p_j} a^* u_k u_k^* b - \sum_{k, p_k = p_j} w_k a^* \hat{u}_k \hat{u}_k^* b \xrightarrow{\text{p.s.}} 0$$

où

$$w_k = \frac{\int \frac{v_c(t\gamma)}{(1 + \delta(\lambda_k)t v_c(t\gamma))^2} \tilde{\nu}(dt)}{\int \frac{v_c(t\gamma)}{1 + \delta(\lambda_k)t v_c(t\gamma)} \tilde{\nu}(dt) \left(1 - \frac{1}{c} \int \frac{\delta(\lambda_k)^2 t^2 v_c(t\gamma)^2}{(1 + \delta(\lambda_k)t v_c(t\gamma))^2} \tilde{\nu}(dt) \right)}.$$

Ce résultat prévoit donc l'existence de valeurs propres isolées, au nombre de $k \leq L$, qui permettent d'évaluer les puissances des k sources d'amplitude maximale. Par ailleurs, pour ces k sources, en prenant $a = b = u_j$ dans le résultat 3), et en supposant que $p_j \neq p_i$ pour tout $i \neq j$, le résultat permet d'évaluer la valeur de l'alignement $|u_j^* \hat{u}_j|^2$ entre le vecteur propre de population u_j et celui du vecteur propre correspondant \hat{u}_j de \hat{C}_N ; cet alignement peut être approximé par $1/w_j$.

Au contraire des résultats classiques liés aux modèles dits "spiked", il est important de noter ici que le seuil S^+ défini plus haut n'est pas optimal dans le sens suivant : comme il est permis aux τ_i de prendre des grandes valeurs, pour tout n large, il est possible que les τ_i eux-mêmes engendrent des valeurs propres isolées dans le spectre de \hat{S}_N . Par contre, du fait de la bornitude du spectre de \hat{S}_N dans le cas d'absence de perturbation dans le modèle ($L = 0$), on peut assurer l'absence asymptotique de valeurs propres isolées au-delà d'un seuil S^+ si $L = 0$. Ainsi, toute valeur propre découverte au delà de ce seuil témoigne de l'action d'une perturbation du modèle et non pas d'un τ_i .

2.4.2. Méthode MUSIC robuste

En définitive, dans des situations de bruits elliptiques avec τ_i de loi non bornée, le spectre des valeurs propres de la matrice $\frac{1}{n} \sum_{i=1}^n x_i x_i^*$ a un support asymptotiquement

non borné et donc ne permet pas la détection et l'estimation de sources à l'aide de méthodes sous-espaces. A contrario, la matrice \hat{C}_N a la propriété de réduire le spectre de ses valeurs propres à un support borné dont peuvent s'échapper un certain nombre de valeurs propres ; ceci permet en particulier d'exhiber de nouvelles méthodes sous-espaces de détection et estimation de sources. En particulier rappelons que, dans le cadre présent d'estimation d'angles d'incidence, l'objectif est d'évaluer les angles $\theta_1, \dots, \theta_L$ paramétrant les vecteurs de sources a_1, \dots, a_L , $a_i = a(\theta_i)$, respectivement. La méthode de l'algorithme MUSIC (Schmidt, 1986) consiste précisément à estimer les angles θ_i en utilisant la propriété selon laquelle la fonction

$$\eta(\theta) \triangleq 1 - \sum_{l=1}^L a(\theta)^* u_l u_l^* a(\theta)$$

s'annule uniquement pour $\theta \in \{\theta_1, \dots, \theta_L\}$. Grâce au résultat du théorème 10, il vient alors le corollaire suivant.

COROLLAIRE 11 (Estimateur MUSIC robuste). — *Définissons la fonction*

$$\hat{\eta}(\theta) \triangleq 1 - \sum_{k=1}^L w_k a(\theta)^* \hat{u}_k \hat{u}_k^* a(\theta)$$

et dénotons $\hat{\theta}_1, \dots, \hat{\theta}_L$ ses L plus profonds minima. Alors, si $p_k > p_-$ et $p_{k+1} \leq p_-$, pour $j \leq k$,

$$\hat{\theta}_j - \theta_j \xrightarrow{\text{P.S.}} 0.$$

Ce corollaire permet donc d'estimer de manière consistante les angles correspondant aux sources de plus fortes puissances dans une situation de bruit impulsif que des méthodes à base de matrices de covariance empiriques ne peuvent pas appréhender.

Nous illustrons les résultats obtenus dans cette section par quelques histogrammes de valeurs propres ainsi que par des résultats de simulations pour l'estimateur MUSIC robuste. La première courbe que nous présentons est la figure 2 qui présente l'histogramme des valeurs propres de \hat{C}_N dans un modèle signal-plus-bruit dont les sources de signal sont suffisamment fortes pour générer des valeurs propres à l'extérieur du support limite. Ce comportement est en effet confirmé sur l'histogramme. Cette courbe doit être comparée à la figure 3 qui, dans les mêmes conditions sur les parties signal et bruit du modèle, présente l'histogramme des valeurs propres de la matrice de covariance empirique $\frac{1}{n} \sum_{i=1}^n x_i x_i^*$. Comme argumenté plus haut, nous observons ici qu'aucune valeur ne s'isole du spectre des valeurs propres qui est bien plus large que celui de \hat{C}_N et ne permet ainsi pas aux méthodes de détection et estimation de type "spike" d'être valables.

Les deux figures suivantes présentent les performances de l'estimateur MUSIC robuste du Corollaire 11 qui, rappelons le, prend en compte à la fois l'effet du régime des grandes matrices aléatoires $N, n \rightarrow \infty$ avec $N/n \rightarrow c \in (0, 1)$ et la maîtrise des

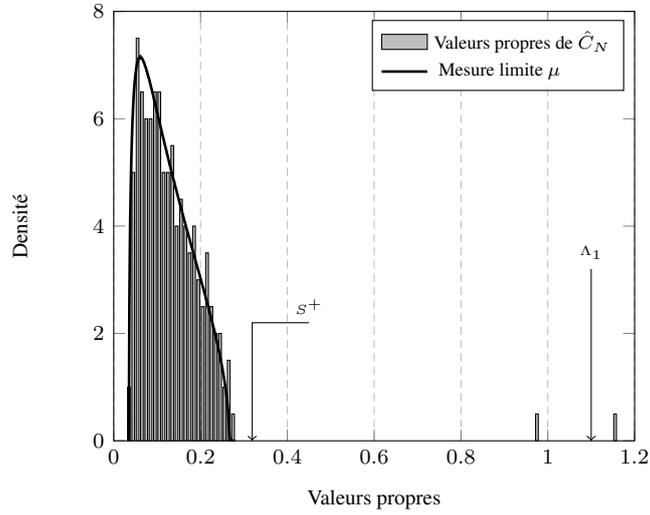


Figure 2. Histogramme des valeurs propres de \hat{C}_N et mesure spectrale limite, pour $u(x) = (1 + \alpha)/(\alpha + x)$ avec $\alpha = 0.2$, $L = 2$, $p_1 = p_2 = 1$, $N = 200$, $n = 1000$, impulsions de loi Student

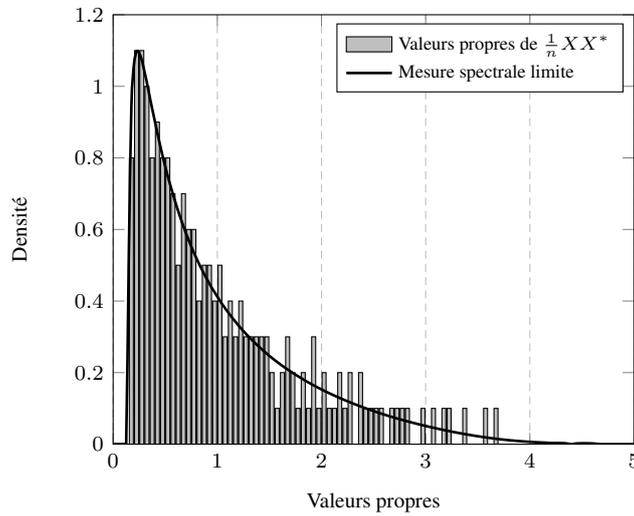


Figure 3. Histogramme des valeurs propres de $\frac{1}{n}XX^*$ et mesure spectrale limite, $L = 2$, $p_1 = p_2 = 1$, $N = 200$, $n = 1000$, impulsions de loi Student

conditions de bruits impulsifs grâce à l'estimateur \hat{C}_N . Cet estimateur est comparé à l'estimateur non robuste qui ne prend en compte que l'effet des matrices aléatoires et suppose (à tort) que le support des valeurs propres de $\frac{1}{n} \sum_{i=1}^n x_i x_i^*$ reste borné; cet estimateur est obtenu par une adaptation simple de l'estimateur originel proposé dans (Mestre, 2008). Enfin, nous comparons également l'estimateur MUSIC robuste à l'estimateur MUSIC originel (Schmidt, 1986) qui ne prend lui ni en compte l'effet du régime des grandes matrices aléatoires (et suppose donc à tort que $N \ll n$) ni celui du bruit impulsif. Les deux figures présentent les performances d'estimation angulaire d'une source à 10° d'incidence dans une situation de deux sources d'angles proches (en général difficile à discriminer et donc pertinente pour la comparaison).

La figure 4 se place dans un contexte où le bruit impulsif est elliptique de loi de Student. Il apparaît que les méthodes MUSIC robuste et non robuste sont toutes deux consistantes dans le régime des hauts rapports signaux-à-bruit. Par contre, l'estimateur robuste présente un gain de plus de 10 dB par rapport à l'estimateur non robuste. Il peut paraître surprenant à première vue que l'estimateur non robuste soit consistant quand bien même il a été établi que le support des valeurs propres de la matrice de covariance empirique ne laisse pas échapper de valeurs propres isolées. Ce raisonnement, correct en théorie asymptotiquement, ne l'est par contre pas à horizon fini où, si les sources sont suffisamment puissantes, des valeurs propres isolées peuvent s'échapper. Néanmoins, la distance entre ces valeurs propres isolées et le bord supérieur du spectre des valeurs propres "du bruit" est bien inférieur à celle induite dans le cas robuste; d'où le gain de performance. Notons par ailleurs, et ceci confirme notre affirmation précédente, que l'estimateur non robuste présente un effet d'avalanche autour d'un rapport signal-à-bruit de 10 dB, tandis que cet effet d'avalanche est bien inférieur pour le cas robuste. Cet effet, dans les deux cas, témoigne de l'impossibilité asymptotique de récupérer une quelconque information sur les angles d'incidence dès lors que les valeurs propres extrêmes de l'estimateur de covariance ou dispersion ne s'échappent plus du support des valeurs propres du bruit seul. Enfin, remarquons que l'estimateur robuste originel ne fonctionne ici pas du tout. La raison est ici liée au fait que le rapport N/n est trop grand pour pouvoir discriminer deux angles proches sous l'hypothèse erronée où $N \ll n$.

La figure 5 se place quant à elle dans le contexte où le bruit impulsif est de nature "impulsion isolée" en ce sens que τ_i est pris égal à 1 pour tous les indices i , sauf pour $i = n$ où $\tau_n = 100$. Dans ce deuxième cadre, la matrice de covariance empirique $\frac{1}{n} \sum_{i=1}^n x_i x_i^*$ a un spectre limite correspondant au spectre de Marčenko–Pastur ainsi qu'à un maximum de trois valeurs propres isolées générées pour deux d'entre elles par les sources et pour la troisième par l'impulsion à $i = n$. Cette dernière génère ainsi une fausse alarme de détection de source et donne lieu à une estimation arbitraire d'une première source. Dans le cadre où seulement deux sources sont estimées, cette fausse alarme reste ainsi présente tant que la valeur propre correspondant au bruit impulsif reste plus grande que les deux valeurs propres des sources. C'est ce qui est observé dans la figure où une avalanche apparaît à un rapport signal-à-bruit de 20 dB pour l'estimateur non robuste. A contrario, l'estimateur robuste ne souffre pas de cette limitation.

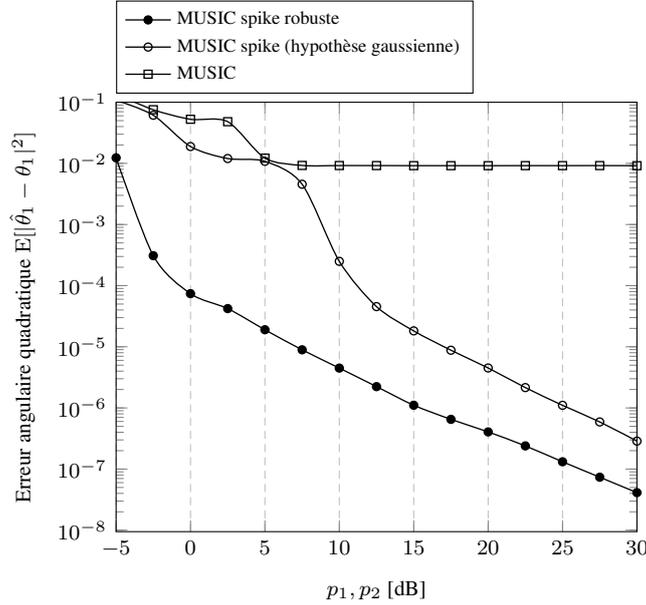


Figure 4. Erreur quadratique angulaire pour l'estimation de $\theta_1 = 10^\circ$, $N = 20$, $n = 100$, deux sources à 10° et 12° , impulsions de type Student de paramètre $\beta = 10$, $u(x) = (1 + \alpha)/(\alpha + x)$ avec $\alpha = 0.2$, $p_1 = p_2$

Ces quelques figures concluent cette première section sur les estimateurs robustes de la classe “Maronna” pour des distributions de bruit elliptique (ou assimilable à des bruits elliptiques). Cependant, nous n’avons pas encore abordé l’objectif originel de Huber de rejet de données erronées, à savoir des observations déterministes génériques $a_i \in \mathbb{C}^N$, non nécessairement modélisées comme $\sqrt{\tau_i}w_i$ comme ici. La prochaine section est dédiée à l’analyse des performances de l’estimateur de Maronna sous cette nouvelle hypothèse.

3. Propriétés de rejet de données erronées

Dans cette section, nous étudions le comportement de l’estimateur de matrice de dispersion de Maronna \hat{C}_N dans un nouveau contexte où une partie des données observées est déterministe et inconnue. L’idée est ici de déterminer les quantités fondamentales régissant l’effet de rejet de données absurdes par les estimateurs robustes.

Le modèle que nous considérons ici est celui d’une observation matricielle $X \in \mathbb{C}^{N \times n}$ formées des n observations de dimension N en colonnes comme suit

$$X = [x_1, \dots, x_{(1-\varepsilon_n)n}, a_1, \dots, a_{\varepsilon_n n}]$$

où $\varepsilon_n \in \{k/n \mid k = 0, \dots, n\}$ est la proportion de données erronées (ou absurdes) observées, où $x_1, \dots, x_{(1-\varepsilon_n)n} \in \mathbb{C}^N$ sont des vecteurs de “données correctes” modéli-

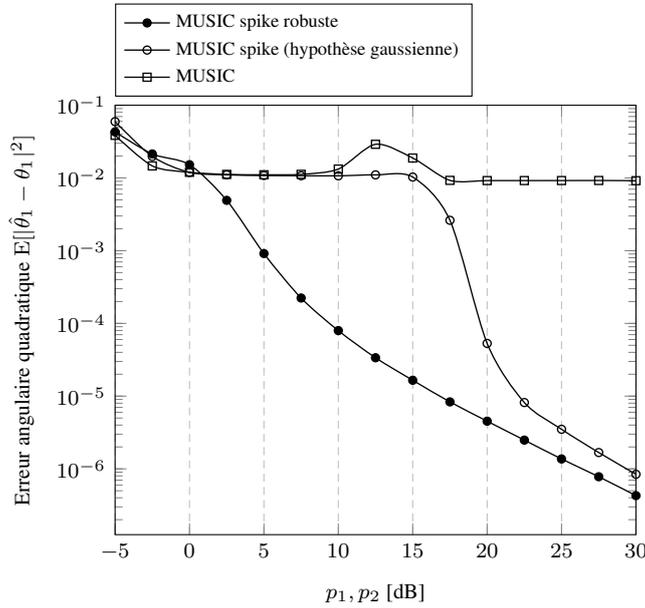


Figure 5. Erreur quadratique angulaire pour l'estimation de $\theta_1 = 10^\circ$, $N = 20$, $n = 100$, deux sources à 10° et 12° , scénario impulsif instantané: $\tau_i = 1$, $i < n$, $\tau_n = 100$, $u(x) = (1 + \alpha)/(\alpha + x)$ avec $\alpha = 0.2$, $p_1 = p_2$

sées par $x_i = C_N^{\frac{1}{2}} w_i$, $C_N \in \mathbb{C}^{N \times N}$ déterministe et définie positive et $w_1, \dots, w_{(1-\varepsilon_n)n}$ i.i.d. gaussiens de moyenne nulle et de covariance I_N tandis que $a_1, \dots, a_{\varepsilon_n n} \in \mathbb{C}^N$ sont déterministes et telles que

$$0 < \min_i \liminf_n \frac{\|a_i\|}{\sqrt{N}} \leq \max_i \limsup_n \frac{\|a_i\|}{\sqrt{N}} < \infty.$$

Nous faisons également l'hypothèse ici que $\varepsilon_n \rightarrow \varepsilon \in [0, 1)$ et que $N/n = c_N \rightarrow c \in (0, 1 - \varepsilon)$.

Notons que nous supposons ici que les données absurdes apparaissent toutes à la fin de la séquence d'observation. Cette hypothèse n'est cependant pas réductrice en ce sens que les estimateurs de matrices de covariance ou de dispersion ne prennent pas en compte l'ordre d'apparition des données. Idéalement, dans un contexte oracle où les indices des données erronées nous seraient donnés, un estimateur optimal (en tout cas dans le sens du maximum de vraisemblance) de la matrice de covariance C_N serait défini comme

$$\frac{1}{(1 - \varepsilon_n)n} \sum_{i=1}^{(1-\varepsilon_n)n} x_i x_i^*.$$

Dans cette section, nous allons donc évaluer selon plusieurs critères la proximité entre l'estimateur robuste \hat{C}_N et cet estimateur optimal. De manière très surprenante, nous

allons en particulier observer que la puissance de rejet de l'estimateur \hat{C}_N est intimement lié au conditionnement de la matrice C_N (en particulier à ces valeurs propres extrêmes).

Notons enfin que le modèle des x_i est un simple modèle gaussien et non elliptique. Ceci permet d'une part une analyse simplifiée et, d'autre part, de mettre plus facilement en lumière la différence de traitement des données correctes et erronées.

Sans autre hypothèse, nous sommes en position de présenter notre résultat. Rappelons tout d'abord que \hat{C}_N est définie ici comme l'unique solution (presque sûrement) de l'équation en Z

$$Z = \frac{1}{n} \sum_{i=1}^{(1-\varepsilon_n)n} u \left(\frac{1}{N} x_i^* Z^{-1} x_i \right) x_i x_i^* + \frac{1}{n} \sum_{i=1}^{\varepsilon_n n} u \left(\frac{1}{N} a_i^* Z^{-1} a_i \right) a_i a_i^*. \quad (11)$$

Nous avons alors le résultat central suivant, établi dans (Morales-Jimenez *et al.*, 2015).

THÉORÈME 12. — *Sous les hypothèses ci-dessous, avec \hat{C}_N la solution presque sûrement unique de (11), lorsque $N, n \rightarrow \infty$,*

$$\left\| \hat{C}_N - \hat{S}_N \right\| \xrightarrow{\text{p.s.}} 0$$

où

$$\hat{S}_N \triangleq \frac{1}{n} \sum_{i=1}^{(1-\varepsilon_n)n} v(\gamma_n) x_i x_i^* + \frac{1}{n} \sum_{i=1}^{\varepsilon_n n} v(\alpha_{i,n}) a_i a_i^*$$

avec γ_n et $\alpha_{1,n}, \dots, \alpha_{\varepsilon_n n, n}$ les solutions uniques du système de $\varepsilon_n n + 1$ équations ($i = 1, \dots, \varepsilon_n n$)

$$\begin{aligned} \gamma_n &= \frac{1}{N} \operatorname{tr} C_N \left(\frac{(1-\varepsilon)v_c(\gamma_n)}{1+cv_c(\gamma_n)\gamma_n} C_N + \frac{1}{n} \sum_{i=1}^{\varepsilon_n n} v(\alpha_{i,n}) a_i a_i^* \right)^{-1} \\ \alpha_{i,n} &= \frac{1}{N} a_i^* \left(\frac{(1-\varepsilon)v_c(\gamma_n)}{1+cv_c(\gamma_n)\gamma_n} C_N + \frac{1}{n} \sum_{j \neq i}^{\varepsilon_n n} v(\alpha_{j,n}) a_j a_j^* \right)^{-1} a_i \end{aligned}$$

et, comme auparavant, $v_c(x) = u(g^{-1}(x))$, $g(x) = x/(1 - c\phi(x))$.

Ainsi, \hat{C}_N est ici asymptotiquement équivalent à la somme d'une matrice correspondant à la somme des $x_i x_i^*$, pondérée par le terme constant $v(\gamma_n)$, et d'une matrice rassemblant les sommes pondérées par $v(\alpha_{i,n})$ des données erronées $a_i a_i^*$. Comparativement à la matrice de covariance empirique standard, disons pondérée par $v(\gamma_n)$, la différence majeure règne donc dans la pondération inégale des $a_i a_i^*$. Il est évidemment intéressant d'analyser les constantes $v(\alpha_{i,n})/v(\gamma_n)$ pour mettre en lumière les

effets d'atténuation des données erronées. Le système implicite qui régit les γ_n et $\alpha_{i,n}$ est cependant assez compliqué et peu interprétable de prime abord. Nous allons donc nous contenter de son analyse dans des cas simplifiés.

Un cas d'analyse simplifiée consiste en le cas où $\varepsilon_n = 1/n$ correspondant à une seule donnée absurde. Dans ce cas, remarquons que $\gamma_n \rightarrow \gamma$ où γ est la solution unique de l'équation

$$\gamma = \frac{1 + cv_c(\gamma)\gamma}{v_c(\gamma)}$$

qui se résout explicitement, suite à quelques manipulations algébriques, en

$$\gamma = \frac{\phi^{-1}(1)}{1 - c}.$$

Quant à $\alpha_{1,n}$ (le seul des termes $\alpha_{i,n}$ ici), cette quantité se définit alors simplement, et également de manière explicite, comme

$$\alpha_{1,n} = \left(\frac{\phi^{-1}(1)}{1 - c} + o(1) \right) \frac{1}{N} a_1^* C_N^{-1} a_1.$$

Ainsi, si $\liminf_N \frac{1}{N} a_1^* C_N^{-1} a_1 > 1$, nous avons $v(\alpha_{1,n})/v(\gamma_n) \leq 1$ pour tout n large et donc l'impact de la donnée absurde a_1 sera amorti. Au contraire, sous l'hypothèse où $\limsup_N \frac{1}{N} a_1^* C_N^{-1} a_1 < 1$, son impact sera renforcé. Deux paramètres jouent naturellement dans la quantité $\frac{1}{N} a_1^* C_N^{-1} a_1$: la norme de a_1 et, de manière plus intéressante, l'alignement entre le vecteur normé porté par a_1 et les sous-espaces propres de C_N correspondant à ces valeurs propres extrêmes. En particulier, à norme égale, $\frac{1}{N} a_1^* C_N^{-1} a_1$ est d'autant plus grand (et donc $v(\alpha_{1,n})$ est d'autant plus petit et a_1 d'autant plus atténué) que a_1 ne s'aligne pas aux espace-propres dominants de C_N (ou de manière équivalente s'aligne aux espaces propres associés aux plus petites valeurs propres de C_N).

Le constat précédent est très intéressant en ce sens que toute donnée "arbitraire" qui s'aligne bien aux espaces propres dominants de C_N est alors renforcée et donc considérée comme "utile" tandis qu'une donnée arbitraire qui s'aligne mal sur les espaces propres des plus grandes valeurs propres de C_N sera rejeté. Un exemple pertinent est celui où a_1 est issu d'un bruit blanc, et non pas d'un bruit corrélé par C_N , auquel cas, par le lemme de la trace évoqué brièvement dans la Section 2, $\frac{1}{N} a_1^* C_N^{-1} a_1 \simeq \frac{1}{N} \text{tr} C_N^{-1}$. Ainsi, pour une trace $\frac{1}{N} \text{tr} C_N$ donnée (et donc pour une énergie totale du bruit donnée), la quantité $\frac{1}{N} \text{tr} C_N^{-1}$ va naturellement donner la propension de l'estimateur robuste à rejeter des données de bruit blanc. L'estimateur \hat{C}_N est ainsi d'autant plus puissant dans sa capacité de rejet de données absurdes que C_N est mal conditionné. Par contre, si $C_N = I_N$, les données absurdes corrélées ne seront pas rejetées. En effet, dans ce cas, $\frac{1}{N} a_1^* C_N^{-1} a_1$ se réduit à la norme $\frac{1}{N} \|a_1\|^2$ de a_1 et n'est pas lié au modèle statistique de a_1 .

Etant donnée cette étude, un cas pertinent consiste en l'hypothèse que les données x_i suivent une loi gaussienne centrée de covariance C_N et que les données a_i

sont également aléatoires, gaussiennes, centrées, mais de covariance différente, disons $D_N \in \mathbb{C}^{N \times N}$. Ce cas particulier du théorème 12 donne lieu au corollaire suivant.

COROLLAIRE 13. — *Sous les hypothèses précédentes, supposons désormais que $a_1, \dots, a_{\varepsilon_n n}$ sont aléatoires, modélisées par $a_i = D_N^{\frac{1}{2}} \tilde{w}_i$, où $D_N \in \mathbb{C}^{N \times N}$ est hermitien, définie positive et telle que $\limsup_N \|D_N\| < \infty$ et $\hat{w}_1, \dots, \hat{w}_{\varepsilon_n n}$ sont indépendants d'entrées i.i.d. de moyenne nulle et variance unité. Alors, lorsque $N, n \rightarrow \infty$,*

$$\left\| \hat{C}_N - \hat{S}_N^{\text{rnd}} \right\| \xrightarrow{\text{p.s.}} 0$$

où

$$\hat{S}_N^{\text{rnd}} \triangleq \frac{1}{n} \sum_{i=1}^{(1-\varepsilon_n)n} v(\gamma_n) x_i x_i^* + \frac{1}{n} \sum_{i=1}^{\varepsilon_n n} v(\alpha_n) a_i a_i^*$$

avec γ_n et α_n les solutions uniques du système de deux équations

$$\begin{aligned} \gamma_n &= \frac{1}{N} \operatorname{tr} C_N \left(\frac{(1-\varepsilon)v_c(\gamma_n)}{1+cv_c(\gamma_n)\gamma_n} C_N + \frac{\varepsilon v_c(\alpha_n)}{1+cv_c(\alpha_n)\alpha_n} D_N \right)^{-1} \\ \alpha_n &= \frac{1}{N} \operatorname{tr} D_N \left(\frac{(1-\varepsilon)v_c(\gamma_n)}{1+cv_c(\gamma_n)\gamma_n} C_N + \frac{\varepsilon v_c(\alpha_n)}{1+cv_c(\alpha_n)\alpha_n} D_N \right)^{-1}. \end{aligned}$$

De la même manière que précédemment, il est intéressant de considérer le cas où $\varepsilon = 0$, auquel cas $\gamma_n = \gamma$ tel que $v_c(\gamma) = 1/\phi^{-1}(1)$. Quant à α_n , il est dans ce cas donné par

$$\alpha_n = \frac{\phi^{-1}(1)}{1-c} \frac{1}{N} \operatorname{tr} D_N C_N^{-1}.$$

Désormais, le paramètre d'intérêt qui permet de quantifier la puissance d'amortissement de l'estimateur robuste est la quantité $\frac{1}{N} \operatorname{tr} D_N C_N^{-1}$. Il apparaît en particulier clairement que les rôles de C_N et D_N ne sont ici pas symétriques. Si C_N est proche de I_N , le pouvoir de rejet des données erronées de covariance D_N sera faible et surtout indépendant de D_N de trace donnée. A contrario, pour C_N mal conditionné, le pouvoir de rejet sera plus important. Pour illustrer les performances de l'estimateur robuste, nous comparons en figure 6 la loi limite des valeurs propres de l'estimateur robuste à celles de la matrice de covariance empirique $\frac{1}{n} X X^*$ et de la matrice de covariance empirique des bonnes données seules. Nous observons un net gain de performance de l'estimateur robuste comparativement à $\frac{1}{n} X X^*$. En particulier, nous observons des gains importants en termes d'approximation des queues de la distribution limite.

Cette simulation conclut notre étude des effets sur les données erronées des estimateurs robustes et conclue par là même l'étude de l'estimateur de Maronna. Cependant, comme mentionné dans l'introduction, l'estimateur de Maronna est fortement limité dans son existence par l'hypothèse $n > N$. Nous nous penchons maintenant sur une version généralisée de l'estimateur robuste qui permet de prendre en compte l'effet d'un trop faible nombre d'observations $n < N$.

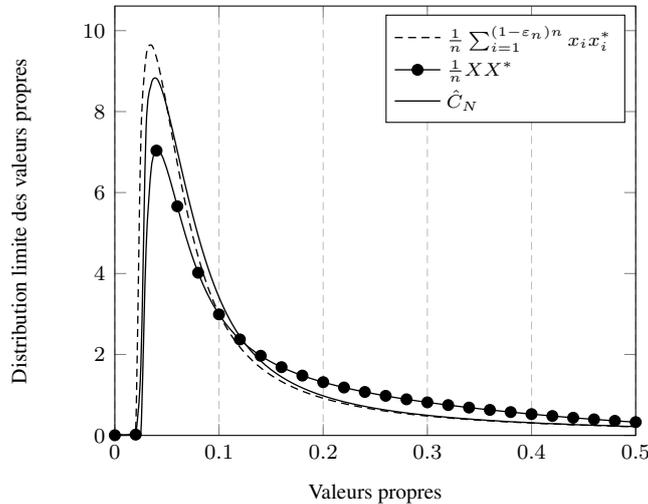


Figure 6. Distribution limite des valeurs propres, pour $[C_N]_{ij} = .9^{|i-j|}$, $D_N = I_N$, $\varepsilon = .05$

4. Comportement asymptotique des estimateurs régularisés

4.1. Résultats théoriques

Dans cette nouvelle section, nous considérons une autre forme d’estimateurs robustes, plus récents, à la croisée des chemins entre les estimateurs originels de Huber et Maronna (en fait plus précisément basés sur l’estimateur de Tyler que nous évoquerons dans un instant) et l’estimateur dit de “shrinkage” (Ledoit, Wolf, 2004). L’estimateur de Ledoit et Wolf consiste en une régularisation élémentaire de la matrice de covariance empirique $\frac{1}{n} \sum_{i=1}^n x_i x_i^*$ obtenu pour des données observées x_1, \dots, x_n , de la forme

$$(1 - \rho) \frac{1}{n} \sum_{i=1}^n x_i x_i^* + \rho F$$

où $\rho \in [0, 1]$ et $F \in \mathbb{C}^{N \times N}$ est une matrice définie positive, souvent prise égale à l’identité, i.e., $F = I_N$. L’intérêt de cette régularisation est de permettre premièrement l’existence d’une inverse pour l’estimateur de covariance quand bien même $N > n$, et dans un second temps de donner un degré de liberté supplémentaire à l’estimateur de covariance qui permet une performance accrue sous certaines métriques d’erreur d’estimation.

L’estimateur (en fait, les estimateurs) considéré dans cette partie consiste en une régularisation à la Ledoit–Wolf d’un estimateur robuste de matrice de dispersion $\hat{C}_N(0)$,

introduit par Tyler comme la solution, unique à constante multiplicative près de l'équation en $Z \in \mathbb{C}^{N \times N}$

$$Z = \frac{1}{n} \sum_{i=1}^n \frac{x_i x_i^*}{\frac{1}{N} x_i^* Z^{-1} x_i}.$$

L'estimateur de Tyler n'est évidemment défini que pour $N \leq n$ et n'est autre que l'estimateur de Maronna pris pour $u(x) = 1/x$. Cependant, cet estimateur ne satisfait pas les conditions imposées par Maronna ; en particulier, $u(x)$ n'est pas défini en $x = 0$ et $xu(x) = 1$ n'est pas croissante. Cette particularité donne lieu à des différences notoires par rapport à l'estimateur de Maronna, comme en particulier le fait que $\hat{C}_N(0)$ n'est défini qu'à constante près (et donc perd en particulier l'information sur la trace de $C_N = E[x_i x_i^*]$). Mais ici, comme nous nous penchons sur une régularisation de cet estimateur, ce problème d'unicité disparaîtra naturellement. Les deux estimateurs que nous allons étudier ont été récemment proposés dans la littérature, parallèlement dans (Abramovich, Besson, 2013 ; Pascal *et al.*, 2014) pour le premier, que nous dénotons $\hat{C}_N(\rho)$ et (Chen *et al.*, 2011) pour le second, que nous appelons $\check{C}_N(\rho)$.

L'estimateur $\hat{C}_N(\rho)$ est défini, pour $\rho \in (\max\{0, 1 - 1/c_N\}, 1]$, comme l'unique solution de l'équation en Z

$$Z = (1 - \rho) \frac{1}{n} \sum_{i=1}^n \frac{x_i x_i^*}{\frac{1}{N} x_i^* Z^{-1} x_i} + \rho I_N.$$

Cet estimateur est très naturel en tant que généralisation immédiate à la Ledoit–Wolf de l'estimateur de Tyler. Il a en particulier la spécificité de satisfaire $\frac{1}{N} \text{tr} \hat{C}_N^{-1} = 1$, ce qui peut être observé en multipliant l'équation précédente à droite et à gauche par $Z^{-\frac{1}{2}}$ et en prenant la trace.

Quant à l'estimateur $\check{C}_N(\rho)$, il a été proposé dans l'optique d'assurer qu'il soit de trace constante indépendamment de ρ , c'est-à-dire que $\frac{1}{N} \text{tr} \check{C}_N(\rho) = 1$. Pour ce faire, l'estimateur est défini, pour tout $\rho \in (0, 1]$, comme la solution unique de l'équation en Z

$$Z = \frac{\tilde{Z}}{\frac{1}{N} \text{tr} \tilde{Z}}$$

$$\tilde{Z} = (1 - \rho) \frac{1}{n} \sum_{i=1}^n \frac{x_i x_i^*}{\frac{1}{N} x_i^* Z^{-1} x_i} + \rho I_N.$$

Notez que les domaines de définition de $\hat{C}_N(\rho)$ et $\check{C}_N(\rho)$ diffèrent et qu'il n'est pas évident que les deux estimateurs robustes se comportent de manière similaire ou distincte. En fait, de l'aveu même des auteurs de (Pascal *et al.*, 2014) et (Chen *et al.*, 2011), il est a priori vraisemblable que les estimateurs aient des comportements très différents. Dans cette section nous allons démontrer que, pour ce qui est du cas où les x_i sont des variables elliptiques centrées, il n'en est en fait rien, dans ce sens où

nous verrons qu'asymptotiquement les deux estimateurs sont strictement équivalents, à changement de variable près du paramètre ρ .

Maintenant que les notions de base d'estimation robuste deviennent familières au lecteur, nous pouvons entrer rapidement dans le vif de cette section et présenter nos résultats centraux. Rappelons donc que nous travaillons ici sous l'hypothèse où $x_i = \sqrt{\tau_i} C_N^{\frac{1}{2}} w_i$ avec les hypothèses habituelles. Notons maintenant que, du fait de la forme des estimateurs $\hat{C}_N(\rho)$ et $\check{C}_N(\rho)$, le paramètre τ_i ne revêt plus d'importance et est naturellement annihilé. Il suffit de supposer que x_i est un vecteur gaussien centré de matrice de covariance C_N .

Nos résultats centraux, qui se démontrent de la même manière que pour l'estimateur de Maronna (à une difficulté technique près concernant l'estimateur $\check{C}_N(\rho)$), sont récapitulés dans les deux théorèmes suivants.

THÉORÈME 14 (Estimateur $\hat{C}_N(\rho)$). — *Pour $\varepsilon \in (0, \min\{1, c^{-1}\})$, définissons $\hat{\mathcal{R}}_\varepsilon = [\varepsilon + \max\{0, 1 - c^{-1}\}, 1]$. Pour chaque $\rho \in (\max\{0, 1 - c^{-1}\}, 1]$, dénotons $\hat{C}_N(\rho)$ la matrice satisfaisant*

$$\hat{C}_N(\rho) = (1 - \rho) \frac{1}{n} \sum_{i=1}^n \frac{x_i x_i^*}{\frac{1}{N} x_i^* \hat{C}_N(\rho)^{-1} x_i} + \rho I_N.$$

Alors, lorsque $N, n \rightarrow \infty$ avec $N/n \rightarrow c$,

$$\sup_{\rho \in \hat{\mathcal{R}}_\varepsilon} \left\| \hat{C}_N(\rho) - \hat{S}_N(\rho) \right\| \xrightarrow{\text{p.s.}} 0$$

où

$$\hat{S}_N(\rho) = \frac{1}{\hat{\gamma}(\rho)} \frac{1 - \rho}{1 - (1 - \rho)c} \frac{1}{n} \sum_{i=1}^n z_i z_i^* + \rho I_N$$

et $\hat{\gamma}(\rho)$ est l'unique solution positive de l'équation en $\hat{\gamma}$

$$1 = \int \frac{t}{\hat{\gamma}\rho + (1 - \rho)t} \nu(dt).$$

Par ailleurs, la fonction $\rho \mapsto \hat{\gamma}(\rho)$ définie ainsi est continue sur $(0, 1]$.

THÉORÈME 15 (Estimateur $\check{C}_N(\rho)$). — *Pour $\varepsilon \in (0, 1)$, dénotons $\check{\mathcal{R}}_\varepsilon = [\varepsilon, 1]$. Pour chaque $\rho \in (0, 1]$, soit $\check{C}_N(\rho)$ l'unique matrice satisfaisant*

$$\check{C}_N(\rho) = \frac{\check{B}_N(\rho)}{\frac{1}{N} \text{tr} \check{B}_N(\rho)}$$

et

$$\check{B}_N(\rho) = (1 - \rho) \frac{1}{n} \sum_{i=1}^n \frac{x_i x_i^*}{\frac{1}{N} x_i^* \check{C}_N(\rho)^{-1} x_i} + \rho I_N.$$

Alors, pour $N, n \rightarrow \infty$ avec $N/n \rightarrow c \in (0, 1)$,

$$\sup_{\rho \in \tilde{\mathcal{R}}_\varepsilon} \|\check{C}_N(\rho) - \check{S}_N(\rho)\| \xrightarrow{\text{p.s.}} 0$$

où

$$\check{S}_N(\rho) = \frac{1 - \rho}{1 - \rho + T_\rho} \frac{1}{n} \sum_{i=1}^n z_i z_i^* + \frac{T_\rho}{1 - \rho + T_\rho} I_N$$

dans laquelle $T_\rho = \rho \check{\gamma}(\rho) F(\check{\gamma}(\rho); \rho)$ avec, pour chaque $x > 0$,

$$F(x; \rho) = \frac{1}{2} (\rho - c(1 - \rho)) + \sqrt{\frac{1}{4} (\rho - c(1 - \rho))^2 + (1 - \rho) \frac{1}{x}}$$

et $\check{\gamma}(\rho)$ l'unique solution positive en $\check{\gamma}$ de

$$1 = \int \frac{t}{\check{\gamma}\rho + \frac{1-\rho}{(1-\rho)c+F(\check{\gamma};\rho)}t} \nu(dt).$$

De plus, la fonction $\rho \mapsto \check{\gamma}(\rho)$ ainsi définie est continue sur $(0, 1]$.

Les résultats sont donc, de manière assez prédictible, très proches du théorème 7 de la section précédente. L'expression du résultat pour $\hat{C}_N(\rho)$ est plus simple d'écriture (et en fait plus directe à établir). Cependant, quand bien même les expressions de $\hat{S}_N(\rho)$ et $\check{S}_N(\rho)$ sont très différentes, il est possible d'établir l'important résultat suivant selon lequel les deux matrices sont en fait essentiellement égales.

LEMME 16 (Equivalence de modèles). — Pour chaque $\rho \in (0, 1]$, il existe une unique paire $\hat{\rho} \in (\max\{0, 1 - c^{-1}\}, 1]$ et $\check{\rho} \in (0, 1]$ telle que

$$\frac{\hat{S}_N(\hat{\rho})}{\lim_N \frac{1}{N} \text{tr} \hat{S}_N(\hat{\rho})} = \check{S}_N(\check{\rho}) = (1 - \rho) \frac{1}{n} \sum_{i=1}^n C_N^{\frac{1}{2}} w_i w_i^* C_N^{\frac{1}{2}} + \rho I_N.$$

De plus, les fonctions $(0, 1] \rightarrow (\max\{0, 1 - c^{-1}\}, 1]$, $\rho \mapsto \hat{\rho}$ et $(0, 1] \rightarrow (0, 1]$, $\rho \mapsto \check{\rho}$ ainsi définies sont continues, croissantes et injectives.

Ainsi, à normalisation près, les matrices $\hat{S}_N(\hat{\rho})$ et $\check{S}_N(\check{\rho})$ sont les mêmes et en fait identiques à l'estimateur de Ledoit–Wolf pour les données (non observables) $C_N^{\frac{1}{2}} w_i$ et pour ρ . Ceci indique donc qu'asymptotiquement, sous un modèle de données elliptiques tout du moins, les estimateurs $\hat{C}_N(\rho)$ et $\check{C}_N(\rho)$ sont essentiellement identiques, quand bien même à horizon fini ils soient potentiellement très différents. Les figures 7 et 8 présentent les histogrammes respectifs des estimateurs $\hat{C}_N(\rho)$ et $\check{C}_N(\rho)$ pour une même valeur de ρ , comparativement à la loi limite des valeurs propres de leurs approximations $\hat{S}_N(\rho)$ et $\check{S}_N(\rho)$.

Grâce aux résultats précédents, il devient plus simple d'appréhender les estimateurs robustes $\hat{C}_N(\rho)$ et $\check{C}_N(\rho)$ que dans les publications dédiées (Abramovich, Beson, 2013 ; Pascal *et al.*, 2014 ; Chen *et al.*, 2011). En particulier, une question naturelle qui se pose suite à cette étude est celle de la possibilité d'estimer en ligne la

meilleure valeur de ρ à sélectionner pour optimiser tel ou tel estimateur basé sur C_N . Du fait du lemme 16, il est strictement équivalent de travailler avec l'un ou l'autre des estimateurs, à changement de variable près.

Une métrique naturelle à minimiser lors de l'évaluation par un estimateur robuste de la matrice de covariance C_N consiste en la norme de Frobenius de la différence. A savoir, (pour ce qui est de $\check{C}_N(\rho)$)

$$\|\check{C}_N(\rho) - C_N\|_{\text{Fro}}^2 \triangleq \text{tr}(\check{C}_N(\rho) - C_N)^2.$$

Nous nous posons donc ici les questions suivantes :

- est-il possible de déterminer, au moins asymptotiquement, une expression explicite du choix de ρ qui minimise la métrique ci-dessus ?
- si oui, cette valeur sera vraisemblablement fonction de C_N , mais est-il alors possible de l'estimer empiriquement, de manière à établir un algorithme permettant de choisir en ligne le meilleur estimateur robuste pour C_N ?

Nous apportons ci-après une réponse affirmative à chacune de ces deux questions. Les résultats successifs sont les suivants.

PROPOSITION 17 (Shrinkage optimal). — *Pour chaque $\rho \in (0, 1]$, nous définissons*

$$\hat{D}_N(\rho) = \frac{1}{N} \text{tr} \left(\left(\frac{\hat{C}_N(\rho)}{\frac{1}{N} \text{tr} \hat{C}_N(\rho)} - C_N \right)^2 \right)$$

$$\check{D}_N(\rho) = \frac{1}{N} \text{tr} \left((\check{C}_N(\rho) - C_N)^2 \right).$$

En faisant l'hypothèse que $\frac{1}{N} \text{tr} C_N = 1$ (sans restreindre la généralité), nous écrivons également $D^* = c \frac{M_2 - 1}{c + M_2 - 1}$, $\rho^* = \frac{c}{c + M_2 - 1}$ où $M_k \triangleq \lim_N \frac{1}{N} \text{tr} C_N^k$, et $\hat{\rho}^* \in (\max\{0, 1 - c^{-1}\}, 1]$, $\check{\rho}^* \in (0, 1]$ les solutions uniques de

$$\frac{\hat{\rho}^*}{\frac{1}{\hat{\gamma}(\hat{\rho}^*)} \frac{1 - \hat{\rho}^*}{1 - (1 - \hat{\rho}^*)c} + \hat{\rho}^*} = \frac{T_{\hat{\rho}^*}}{1 - \hat{\rho}^* + T_{\hat{\rho}^*}} = \rho^*.$$

Alors, en prenant $\varepsilon < \min(\hat{\rho}^* - \max\{0, 1 - c^{-1}\}, \check{\rho}^*)$, sous les hypothèses des Théorèmes 14 et 15,

$$\inf_{\rho \in \hat{\mathcal{R}}_\varepsilon} \hat{D}_N(\rho) \xrightarrow{\text{p.s.}} D^*, \quad \inf_{\rho \in \check{\mathcal{R}}_\varepsilon} \check{D}_N(\rho) \xrightarrow{\text{p.s.}} D^*$$

et

$$\hat{D}_N(\hat{\rho}^*) \xrightarrow{\text{p.s.}} D^*, \quad \check{D}_N(\check{\rho}^*) \xrightarrow{\text{p.s.}} D^*.$$

PROPOSITION 18 (Estimateur de shrinkage optimal). — *Sous les hypothèses de la Proposition 17, dénotons $\hat{\rho}_N \in (\max\{0, 1 - c^{-1}\}, 1]$ et $\check{\rho}_N \in (0, 1]$ les solutions (non nécessairement uniques) de*

$$\frac{\hat{\rho}_N}{\frac{1}{N} \operatorname{tr} \hat{C}_N(\hat{\rho}_N)} = \frac{c_N}{\frac{1}{N} \operatorname{tr} \left[\left(\frac{1}{n} \sum_{i=1}^n \frac{x_i x_i^*}{\frac{1}{N} \|x_i\|^2} \right)^2 \right] - 1}$$

$$\frac{\check{\rho}_N \frac{1}{n} \sum_{i=1}^n \frac{x_i^* \check{C}_N(\check{\rho}_N)^{-1} x_i}{\|x_i\|^2}}{1 - \check{\rho}_N + \check{\rho}_N \frac{1}{n} \sum_{i=1}^n \frac{x_i^* \check{C}_N(\check{\rho}_N)^{-1} x_i}{\|x_i\|^2}} = \frac{c_N}{\frac{1}{N} \operatorname{tr} \left[\left(\frac{1}{n} \sum_{i=1}^n \frac{x_i x_i^*}{\frac{1}{N} \|x_i\|^2} \right)^2 \right] - 1}$$

définies arbitrairement si ces solutions n'existent pas. Alors, nous avons $\hat{\rho}_N \xrightarrow{\text{p.s.}} \hat{\rho}^*$ et $\check{\rho}_N \xrightarrow{\text{p.s.}} \check{\rho}^*$, de telle manière que $\hat{D}_N(\hat{\rho}_N) \xrightarrow{\text{p.s.}} D^*$ et $\check{D}_N(\check{\rho}_N) \xrightarrow{\text{p.s.}} D^*$.

Les propositions ci-dessus donnent ainsi des réponses aux deux questions que nous nous posions précédemment en proposant en particulier un choix optimal pour ρ pour chacun des estimateurs robustes étudiés ici. La découverte d'un tel paramètre ρ avait déjà été tentée dans l'article (Chen *et al.*, 2011) mais dans un cadre fini-dimensionnel. Malheureusement, devant la difficulté liée à l'expression implicite de $\check{C}_N(\rho)$, seule une approche inexacte a été proposée qui n'est intrinsèquement valable que pour des petites valeurs du rapport N/n . Nous utiliserons cet estimateur, dit "clairvoyant", pour nos comparaisons. Nos résultats de simulations sont proposés dans les figures 9 et 10 qui évaluent l'erreur de Frobenius (normalisée) entre $\hat{C}_N(\rho)$ ou $\check{C}_N(\rho)$ et C_N . Nous observons en particulier la très bonne qualité de l'estimateur empirique $\check{\rho}_N$ comparativement à l'optimum obtenu numériquement, ainsi que le gain important en performance comparativement à la méthode clairvoyante approximative proposée dans (Chen *et al.*, 2011).

4.2. Application à l'optimisation de portefeuille

Dans la section précédente, notre intérêt s'est porté sur l'optimisation du paramètre ρ pour la minimisation de la norme de Frobenius de l'erreur d'estimation de C_N par \hat{C}_N . Cependant, ce choix n'était pas lié à une application spécifique. Historiquement, l'estimateur de Ledoit–Wolf, proposé en (Ledoit, Wolf, 2004) et dont le paramètre ρ était également choisi de manière à minimiser la norme de Frobenius de l'erreur, a été appliqué à des problèmes d'optimisation en statistiques financières (Ledoit, Wolf, 2003) et en particulier à l'optimisation de portefeuilles d'actions.

Une approche différente de celle considérée par Ledoit et Wolf consiste plutôt à considérer ρ comme un paramètre libre que l'on choisit, non pas de sorte à minimiser une métrique arbitraire, mais de sorte à minimiser une métrique dédiée au problème spécifique d'intérêt. Cette brève section est dédiée à la question d'optimisation de portefeuille financier au sens de la minimisation de la variance (associée au risque encouru) des gains apportés par le portefeuille choisi.

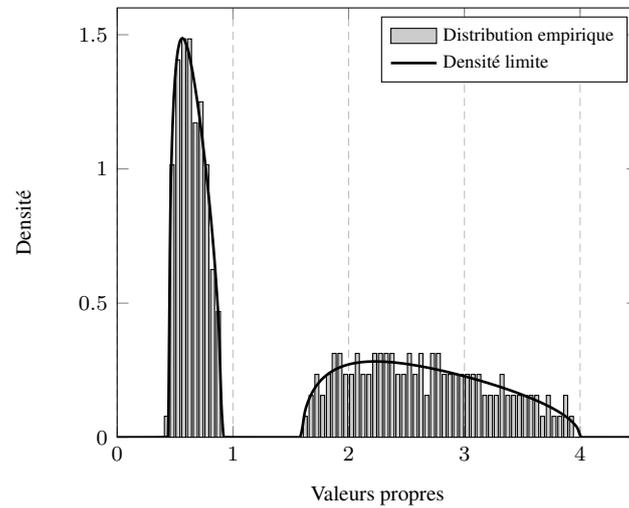


Figure 7. Histogramme des valeurs propres de $\hat{C}_N(\rho)$ pour $n = 2048$, $N = 256$, $C_N = \frac{1}{3} \text{diag}(I_{128}, 5I_{128})$, $\rho = 0.2$, et comparaison au spectre limite

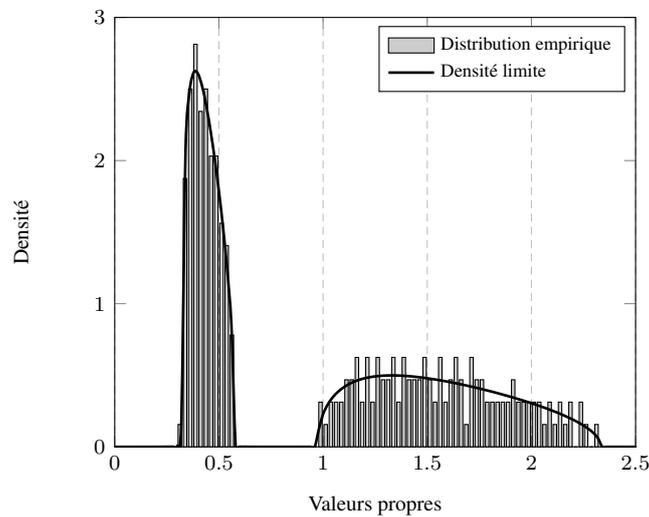


Figure 8. Histogramme des valeurs propres de \check{C}_N pour $n = 2048$, $N = 256$, $C_N = \frac{1}{3} \text{diag}(I_{128}, 5I_{128})$, $\rho = 0.2$, comparativement au spectre limite

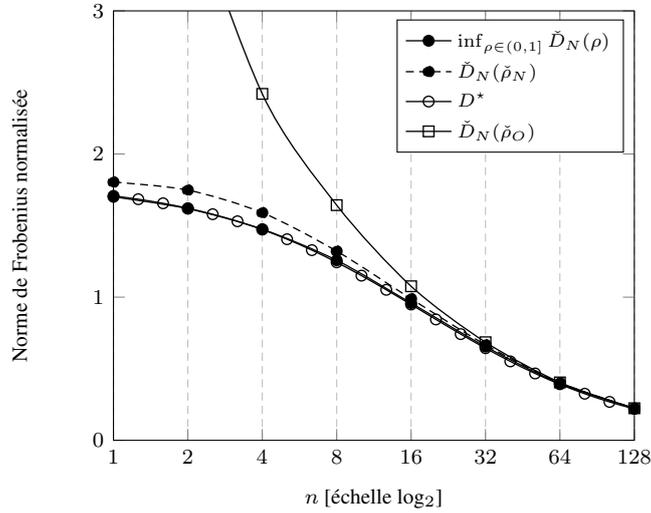


Figure 9. Performance du shrinkage optimal moyenné sur 10 000 simulations de Monte Carlo, pour $N = 32$, $[C_N]_{ij} = r^{|i-j|}$ avec $r = 0.7$; $\check{\rho}_N$ donné en Proposition 18; $\check{\rho}_O$ est l'estimateur clairvoyant proposé dans (Chen et al., 2011) (Equation (12)); D^* est pris pour $c = N/n$

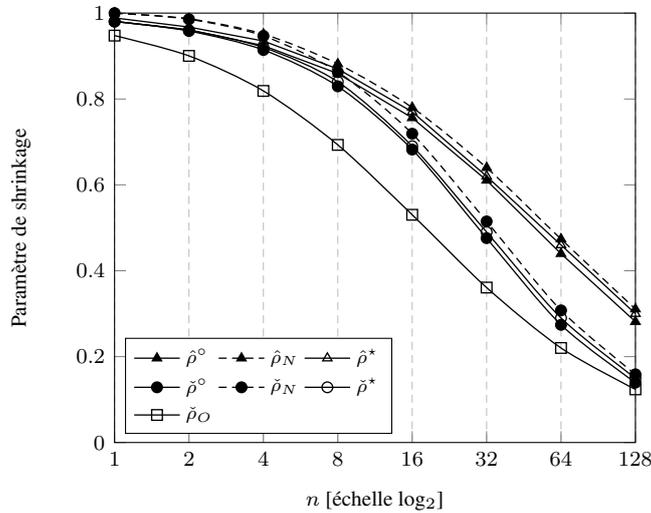


Figure 10. Paramètre de shrinkage ρ moyenné sur 10 000 simulations de Monte Carlo, pour $N = 32$, $[C_N]_{ij} = r^{|i-j|}$ avec $r = 0.7$; $\hat{\rho}_N$ et $\check{\rho}_N$ donné par la Proposition 18; $\check{\rho}_O$ est l'estimateur clairvoyant proposé dans (Chen et al., 2011) (Equation (12)); ρ^* , $\hat{\rho}^*$, et $\check{\rho}^*$ pris pour $c = N/n$;

$$\hat{\rho}^o = \operatorname{argmin}_{\rho \in (\max\{0, 1 - c_N^{-1}\}, 1]} \hat{D}_N(\rho) \text{ et } \check{\rho}^o = \operatorname{argmin}_{\rho \in (0, 1]} \check{D}_N(\rho)$$

Dans le cadre de ce problème, la question est de déterminer un portefeuille $h \in \mathbb{C}^N$ de N actions financières, où h est tel que $\sum_i h_i = 1$ (distribution complète des ressources) et où on admet que h_i puisse être négatif. Dans ce cadre, on cherche à minimiser le risque encouru qui est donné par

$$\sigma^2(h) = h^* C_N h$$

où C_N est la variance des observations successives du marché (supposées indépendantes) $\tilde{x}_1, \dots, \tilde{x}_n$. Comme la moyenne des \tilde{x}_i n'est pas nulle a priori, on prendra ici $x_i = \tilde{x}_i - \frac{1}{n} \sum_{j=1}^n \tilde{x}_j$ de manière à recentrer (empiriquement) les données. Cette particularité induit des complications techniques que nous n'évoquerons pas ici étant donné qu'elles n'impactent finalement que relativement peu le résultat final.

Si C_N était connu, le choix optimal pour h serait de prendre

$$h_N = \frac{C_N^{-1} 1_N}{1_N^* C_N^{-1} 1_N}$$

avec $1_N \in \mathbb{C}^N$ le vecteur d'entrées toutes égales à un, qui induit alors

$$\sigma^2(h_N) = \frac{1}{1_N^* C_N^{-1} 1_N}. \quad (12)$$

Comme C_N n'est cependant pas connu parfaitement, nous proposons ici de considérer le portefeuille

$$\hat{h}_N(\rho) = \frac{\hat{C}_N^{-1}(\rho) 1_N}{1_N^* \hat{C}_N^{-1}(\rho) 1_N}$$

dont le risque associé peut être calculé comme étant égal à

$$\sigma^2(\hat{h}_N(\rho)) = \frac{1_N^* \hat{C}_N^{-1}(\rho) C_N \hat{C}_N^{-1}(\rho) 1_N}{(1_N^* \hat{C}_N^{-1}(\rho) 1_N)^2}. \quad (13)$$

L'objectif est de déterminer ρ de telle manière à minimiser le risque (13), au moins asymptotiquement, puis de déterminer un estimateur consistant pour ce ρ optimal. Ces résultats sont donnés ci-après, issus de l'article (Yang *et al.*, 2014).

THÉORÈME 19. — *Avec les mêmes notations et hypothèses que dans la section précédente, nous avons, lorsque $N, n \rightarrow \infty$ et $c_N \rightarrow c \in (0, \infty)$,*

$$\sup_{\rho \in \hat{\mathcal{R}}_\varepsilon} \left| \sigma^2(\hat{h}_N(\rho)) - \bar{\sigma}^2(\hat{h}_N(\rho)) \right| \xrightarrow{\text{p.s.}} 0$$

où

$$\bar{\sigma}^2(\hat{h}_N(\rho)) \triangleq \frac{1_N^* \left(\frac{k(\rho)}{(\gamma + \alpha k(\rho))} C_N + \rho I_N \right)^{-1} C_N \left(\frac{k(\rho)}{(\gamma + \alpha k(\rho))} C_N + \rho I_N \right)^{-1} 1_N}{\left(1 - \frac{\beta k(\rho)^2}{(\gamma + \alpha k(\rho))^2} \right) \left(1_N^* \left(\frac{k(\rho)}{(\gamma + \alpha k(\rho))} C_N + \rho I_N \right)^{-1} 1_N \right)^2}$$

$k(\rho) = (1 - \rho)/(1 - c(1 - \rho))$ et α l'unique solution de

$$\alpha = \frac{1}{n} \operatorname{tr} \left[C_N \left(\frac{k(\rho)}{(\gamma + \alpha k(\rho))} C_N + \rho I_N \right)^{-1} \right].$$

Ce résultat donne une expression déterministe limite du risque associé à $h_N(\rho)$ et va nous permettre de choisir, de manière déterministe, le ρ minimisant ce risque. Il nous reste alors à estimer $\sigma^2(\hat{h}_N(\rho))$. Ce résultat est donné ci-après.

THÉORÈME 20. — Lorsque $N, n \rightarrow \infty$ et $c_N \rightarrow c \in (0, \infty)$,

$$\sup_{\rho \in \mathcal{R}_\varepsilon} \left| \hat{\sigma}_{\text{sc}}^2(\hat{h}_N(\rho)) - \frac{1}{\kappa} \sigma^2(\hat{h}_N(\rho)) \right| \xrightarrow{\text{p.s.}} 0 \quad (14)$$

pour une certaine constante κ , où nous avons défini

$$\hat{\sigma}_{\text{sc}}^2(\hat{h}_N(\rho)) = \frac{(\hat{\gamma}_{\text{sc}} + \hat{\alpha}_{\text{sc}} \hat{k}(\rho))^2 \mathbf{1}_N^* \hat{C}_N^{-1}(\rho) (\hat{C}_N(\rho) - \rho I_N) \hat{C}_N^{-1}(\rho) \mathbf{1}_N}{\hat{k}(\rho) \hat{\gamma}_{\text{sc}} (\mathbf{1}_N^* \hat{C}_N^{-1}(\rho) \mathbf{1}_N)^2}$$

ainsi que

$$\begin{aligned} \hat{k}(\rho) &= \frac{1 - \rho}{1 - (1 - \rho)c_N} \\ \hat{\gamma}_{\text{sc}} &= \frac{1}{1 - (1 - \rho)c_N} \frac{1}{|\mathcal{B}|} \sum_{i \in \mathcal{B}} \frac{\tilde{x}_i^* \hat{C}_N^{-1}(\rho) \tilde{x}_i}{\|\tilde{x}_i\|^2} \\ \hat{\alpha}_{\text{sc}} &= \frac{\hat{\gamma}_{\text{sc}} \frac{1}{N} \operatorname{tr} [I_N - \rho \hat{C}_N^{-1}(\rho)]}{\hat{k}(\rho) \left(\frac{n}{N} - \frac{1}{N} \operatorname{tr} [I_N - \rho \hat{C}_N^{-1}(\rho)] \right)} \end{aligned}$$

et $\mathcal{B} \triangleq \{t : \|\tilde{x}_t\|^2 > \xi\}$ avec $\xi > 0$ suffisamment petit.

Les convergences uniformes des théorèmes précédents permettent en particulier d'assurer que

$$|\sigma^2(\hat{h}_N(\hat{\rho}^*)) - \sigma^2(\hat{h}_N(\rho^*))| \xrightarrow{\text{p.s.}} 0$$

où $\hat{\rho}^*$ est l'argument du minimum de $\hat{\sigma}_{\text{sc}}^2(\hat{h}_N(\rho))$ et ρ^* l'argument du minimum de $\sigma^2(h_N(\rho))$.

Il suffit donc d'exhiber (empiriquement) le minimum de la fonction $\hat{\sigma}_{\text{sc}}^2(\hat{h}_N(\rho))$ pour identifier un $\hat{h}_N(\hat{\rho}^*)$ dont les performances sont asymptotiquement optimales.

Une courbe de performances établies sur des données financières réelles est présentée en figure 11. On remarque un gain de performance important pour toute valeur de la fenêtre temporelle de données obtenu par la méthode robuste dédiée au problème.

Les méthodes non robustes sont quant à elles en retrait. Remarquons qu’au-delà d’une certaine valeur de n les performances diminuent pour tous les estimateurs. Ceci est dû à la perte de stationnarité des données au cours du temps qui demandent, pour ce qui est des données réelles, de maîtriser le nombre d’échantillons passés utilisé pour effectuer l’estimation de la variance C_N .

Cet exemple clôt notre section sur les estimateurs de type “robuste shrinkage” ainsi que l’ensemble des sections portant sur les estimations de premier ordre des estimateurs robustes. Alors que les performances étudiées dans ces sections ont souvent été caractérisées par des quantités qui, contrairement au cas $n \gg N$, ne tendent pas vers zéro (comme ici dans le cadre du portefeuille de risque minimal), il est possible que les estimateurs en question soient consistants dans le régime où $N, n \rightarrow \infty$ (comme par exemple pour l’estimateur d’angle MUSIC robuste). Dans ce cas les figures de mérite, au-delà de la simple consistance permettant en particulier de sélectionner des paramètres libres (ρ dans le cas des estimateurs shrinkage ou la fonction $u(t)$ dans le cas des estimateurs de Maronna), sont liées à des résultats de second ordre, du type théorèmes de la limite centrale. Dans la prochaine section, nous évoquons ces questions de manière assez succincte, avec un exemple parlant en détection de signal de source connue.

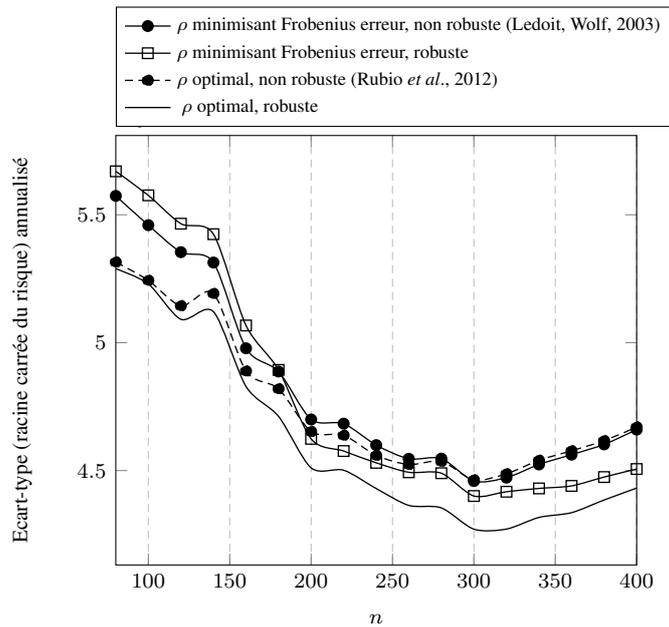


Figure 11. Risque généré par le portefeuille, moyenné sur une année pour différents estimateurs. Données issues du marché HSI de 2011 à 2013

5. Résultats de second ordre

Dans cette section, nous travaillerons, essentiellement pour des raisons de simplicité, avec l'estimateur robuste $\hat{C}_N(\rho)$ étudié dans la section 4. Il a en particulier été établi qu'uniformément sur ρ , cet estimateur vérifie $\|\hat{C}_N(\rho) - \hat{S}_N(\rho)\| \xrightarrow{\text{p.s.}} 0$ pour un certain $\hat{S}_N(\rho)$ facilement analysable grâce aux outils de matrices aléatoires.

Dans beaucoup de problèmes liés à l'évaluation des performances asymptotiques de tel ou tel estimateur, il est demandé que quantifier des fonctionnelles du type $N^\alpha(f(\hat{C}_N(\rho)) - E[f(\hat{C}_N(\rho))])$ ou $N^\alpha(a^*F(\hat{C}_N(\rho))b - E[a^*F(\hat{C}_N(\rho))b])$ où $f : \mathbb{C}^{N \times N} \rightarrow \mathbb{R}$ et $F : \mathbb{C}^{N \times N} \rightarrow \mathbb{C}^{N \times N}$, et α est souvent pris égal à $1/2$ ou 1 . Cette quantification donne lieu en général à des théorèmes de la limite centrale ou à des convergences vers d'autres lois (par exemple pour ce qui est de résultats en théorie des valeurs extrêmes). Nous nous concentrerons ici sur l'étude des fluctuations de $\sqrt{N}a^*\hat{C}_N^{-1}(\rho)b$. En effet, de par l'équivalence entre \hat{C}_N et \hat{S}_N , les formes quadratiques $a^*\hat{C}_N^{-1}(\rho)b$ doivent être sensiblement équivalentes aux formes quadratiques $a^*\hat{S}_N^{-1}(\rho)b$ et donc doivent fluctuer (selon par exemple (Kammoun *et al.*, 2009) qui établit un théorème de la limite centrale pour $a^*\hat{S}_N^{-1}(\rho)b$) au rythme $1/\sqrt{N}$.

Cependant, contrairement à nos résultats précédents selon lesquels tout résultat concernant \hat{S}_N peut se prolonger à \hat{C}_N , il n'est pas clair que tout résultat concernant $\sqrt{N}\hat{S}_N$ puisse se prolonger à $\sqrt{N}\hat{C}_N$. En particulier, s'il est très naturel que les formes bilinéaires $a^*\hat{C}_N^{-1}(\rho)b$ et $a^*\hat{S}_N^{-1}(\rho)b$ fluctuent au même rythme, il n'est pas clair que les variances asymptotiques soient les mêmes et que les fluctuations de $a^*\hat{C}_N^{-1}(\rho)b$ soient toujours gaussiennes. Pour assurer ce résultat simplement, il faudrait que $\sqrt{N}\|\hat{C}_N(\rho) - \hat{S}_N(\rho)\| \xrightarrow{\text{p.s.}} 0$. Malheureusement, ce résultat ne semble pas vrai. La preuve de la convergence $\|\hat{C}_N(\rho) - \hat{S}_N(\rho)\| \xrightarrow{\text{p.s.}} 0$ permet en effet (plutôt facilement) d'obtenir le résultat plus grossier suivant

$$\sup_{\rho \in \mathcal{R}_\kappa} N^{\frac{1}{2}-\varepsilon} \left\| \hat{C}_N(\rho) - \hat{S}_N(\rho) \right\| \xrightarrow{\text{p.s.}} 0 \quad (15)$$

pour tout $\varepsilon > 0$. C'est-à-dire que la convergence peut être améliorée d'un facteur $N^{\frac{1}{2}-\varepsilon}$, mais certainement pas d'un facteur $N^{\frac{1}{2}}$. Ceci pose un problème fondamental qu'il s'agit donc de résoudre au cas par cas pour chaque théorème de la limite centrale souhaité.

Dans le cadre de notre étude, quand bien même $\sqrt{N}\|\hat{C}_N(\rho) - \hat{S}_N(\rho)\| \xrightarrow{\text{p.s.}} 0$ n'est certainement pas vrai, le résultat moins fort suivant peut en fait être prouvé.

THÉORÈME 21 (Fluctuation des formes bilinéaires). — Soit $a, b \in \mathbb{C}^N$ deux vecteurs tels que $\|a\| = \|b\| = 1$. Alors, si $N, n \rightarrow \infty$ avec $c_N \rightarrow c \in (0, \infty)$, pour tout $\varepsilon > 0$,

$$\sup_{\rho \in \mathcal{R}_\kappa} N^{1-\varepsilon} \left| a^*\hat{C}_N^{-1}(\rho)b - a^*\hat{S}_N^{-1}(\rho)b \right| \xrightarrow{\text{p.s.}} 0.$$

Ainsi, en se penchant plus précisément sur la forme bilinéaire $a^* \hat{C}_N^{-1}(\rho)b$ et non plus sur la matrice \hat{C}_N , nous avons gagné ici un facteur $N^{\frac{1}{2}}$ dans le résultat de convergence, qui nous permet en particulier d'assurer que $\sqrt{N}(a^* \hat{C}_N^{-1}(\rho)b - a^* \hat{S}_N^{-1}(\rho)b) \xrightarrow{\text{p.s.}} 0$ uniformément sur ρ . Ce résultat est maintenant suffisant pour assurer le transfert des théorèmes de la limite centrale pour $a^* \hat{S}_N^{-1}(\rho)b$ au même théorème de la limite centrale pour $a^* \hat{C}_N^{-1}(\rho)b$.

En guise d'application, considérons le problème de détection suivant

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

où x est un bruit centré elliptique comme auparavant, $\alpha > 0$ un scalaire inconnu et $p \in \mathbb{C}^N$ un vecteur source connu. La question est d'identifier, à partir de la donnée y et des données x_1, \dots, x_n issues de l'hypothèse \mathcal{H}_0 (nous sommes donc dans un cas d'apprentissage supervisé), si nous nous trouvons sous hypothèse \mathcal{H}_0 ou \mathcal{H}_1 .

Une technique classique pour atteindre cet objectif est d'utiliser un estimateur de vraisemblance généralisé (GLRT) qui, sous l'hypothèse où $C_N = E[x_i x_i^*]$ est connu, assure que le détecteur optimal est donné par le comparateur

$$T_N(\rho) \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \Gamma$$

où

$$T_N(\rho) = \frac{|y^* C_N^{-1} p|}{\sqrt{y^* C_N^{-1} y} \sqrt{p^* C_N^{-1} p}}.$$

A défaut de connaître C_N , nous remplaçons donc cet estimateur par le suivant

$$T_N(\rho) \triangleq \frac{|y^* \hat{C}_N^{-1}(\rho) p|}{\sqrt{y^* \hat{C}_N^{-1}(\rho) y} \sqrt{p^* \hat{C}_N^{-1}(\rho) p}}$$

et le problème consiste alors à déterminer le choix de ρ assurant une détection optimale, à savoir une probabilité de fausse alarme minimale sous contrainte de probabilité de détection sous \mathcal{H}_1 assurée, ou a contrario une probabilité de détection maximale (sous \mathcal{H}_1) sous une contrainte de fausses alarmes donnée.

Il est assez aisé de voir que, si $\Gamma = O(1)$, dans le régime où $N, n \rightarrow \infty$ et $c_N \rightarrow c \in (0, \infty)$, $T_N(\rho) \xrightarrow{\text{p.s.}} 0$ sous \mathcal{H}_0 et $\liminf_N T_N(\rho) > 0$ presque sûrement sous \mathcal{H}_1 et ce, indépendamment de la valeur de ρ . Le détecteur est donc assez trivial, puisque la probabilité de fausse alarme tend vers zéro et la puissance du test vers un. Il n'est donc pas très utile de vouloir sélectionner un ρ optimal. Il est par contre intéressant de se pencher plutôt sur le régime où $\alpha = a/\sqrt{N}$ (à savoir au régime de détection à bas

rapport signal à bruit) pour lequel il faut alors choisir le seuil de détection $\Gamma = \gamma/\sqrt{N}$. Dans ce cadre, le comportement asymptotique de $T_N(\rho)$ devient plus intéressant et les erreurs de première et deuxième espèces maintenant non triviales.

Ce problème revient donc essentiellement à étudier les fluctuations de la quantité $\sqrt{N}T_N(\rho)$ et plus précisément la probabilité que $\sqrt{N}T_N(\rho)$ soit supérieur à γ . Etant donnée la convergence établie dans le théorème 21, le résultat que nous cherchons à établir ici se réduit à une delta-méthode appliquée aux formes bilinéaires de noyau \hat{S}_N^{-1} . Nous avons précisément le résultat suivant.

THÉORÈME 22 (Performance asymptotique du détecteur). — *Sous l'hypothèse \mathcal{H}_0 , lorsque $N, n \rightarrow \infty$ avec $c_N \rightarrow c \in (0, \infty)$,*

$$\sup_{\rho \in \mathcal{R}_c} \left| P \left(T_N(\rho) > \frac{\gamma}{\sqrt{N}} \right) - \exp \left(-\frac{\gamma^2}{2\sigma_N^2(\rho)} \right) \right| \rightarrow 0$$

$$\text{où } \underline{\rho} = \rho \left(\rho + \gamma_N(\rho)^{-1} \frac{1-\rho}{1-(1-\rho)c} \right)^{-1},$$

$$\sigma_N^2(\rho) \triangleq \frac{1}{2} \frac{p^* C_N Q_N^2(\rho) p}{p^* Q_N(\rho) p \cdot \frac{1}{N} \text{tr} C_N Q_N(\rho) \cdot (1-c(1-\rho)^2 m(-\rho))^2 \frac{1}{N} \text{tr} C_N^2 Q_N^2(\rho)}$$

et $Q_N(\rho) \triangleq (I_N + (1-\rho)m(-\rho)C_N)^{-1}$; finalement, $m(-\rho)$ est l'unique solution de l'équation

$$m(-\rho) = \left(\rho + c \int \frac{(1-\rho)t}{1+(1-\rho)tm(-\rho)} \nu(dt) \right)^{-1}.$$

Le résultat assure donc que la probabilité de fausse alarme est asymptotiquement équivalente à une loi de Rayleigh $R_N(\rho)$ de paramètre $\sigma_N(\rho)$. Ce résultat est donné dans les figures 12–15 ci après pour $N = 20$ et $N = 100$, respectivement. Comme il est très classique en théorie des matrices aléatoires, les fluctuations de deuxième ordre ont des convergences bien plus faibles que les convergences de premier ordre. Ceci est confirmé dans les figures où le cas $N = 20$ montre des approximations asymptotiques bien moins bonnes que le cas $N = 100$, particulièrement pour des probabilités de fausse alarme faibles.

Cet exemple conclut cette section sur les fluctuations des formes bilinéaires de type $a^* \hat{C}_N^{-1}(\rho) b$ et conclut par là même les sections techniques de cet article.

6. Conclusions

Dans cet article, nous avons fait un tour d'horizon des récents résultats de la recherche en matrices aléatoires et estimation robuste. Ce domaine présente la difficulté majeure de devoir parvenir à approximer correctement les estimateurs robustes par des quantités aléatoires plus facilement maîtrisables par les outils de la théorie des

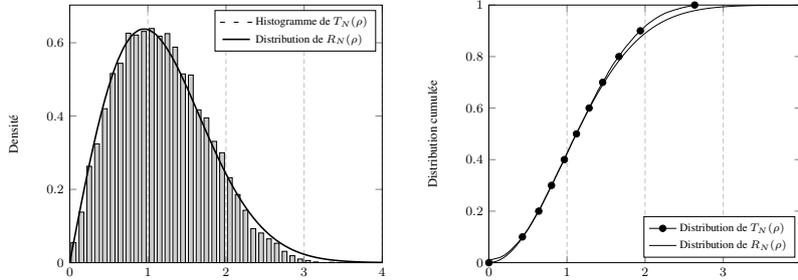


Figure 12. Histogramme et distribution cumulée de $\sqrt{N}T_N(\rho)$ comparativement à $R_N(\rho)$, $N = 20$, $p = N^{-\frac{1}{2}}[1, \dots, 1]^T$, $[C_N]_{ij} = 0.7^{|i-j|}$, $c_N = 1/2$, $\rho = 0.2$

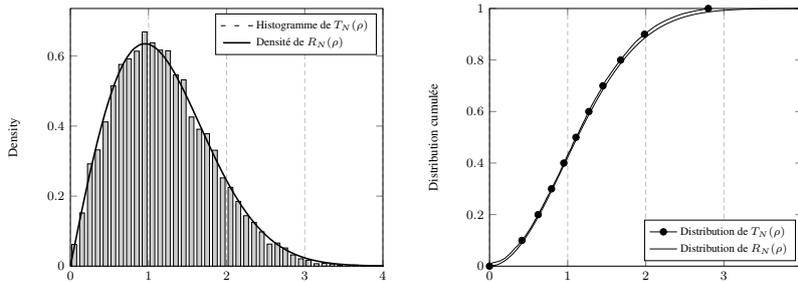


Figure 13. Histogramme et distribution cumulée de $\sqrt{N}T_N(\rho)$ comparativement à $R_N(\rho)$, $N = 100$, $p = N^{-\frac{1}{2}}[1, \dots, 1]^T$, $[C_N]_{ij} = 0.7^{|i-j|}$, $c_N = 1/2$, $\rho = 0.2$

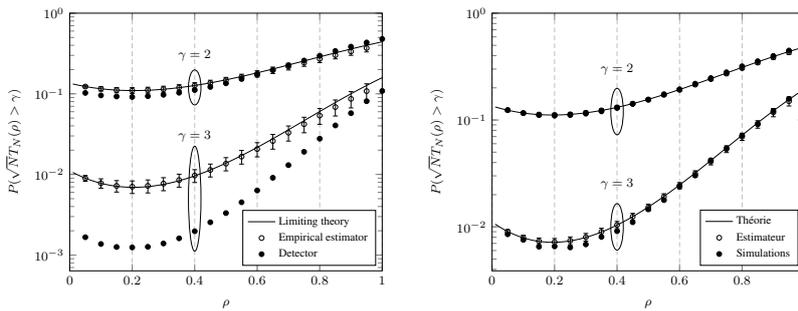


Figure 14. Probabilité de fausse alarme $P(\sqrt{N}T_N(\rho) > \gamma)$, $N = 20$ (gauche), $N = 100$ (droite), $p = N^{-\frac{1}{2}}[1, \dots, 1]^T$, $[C_N]_{ij} = 0.7^{|i-j|}$, $c_N = 1/2$

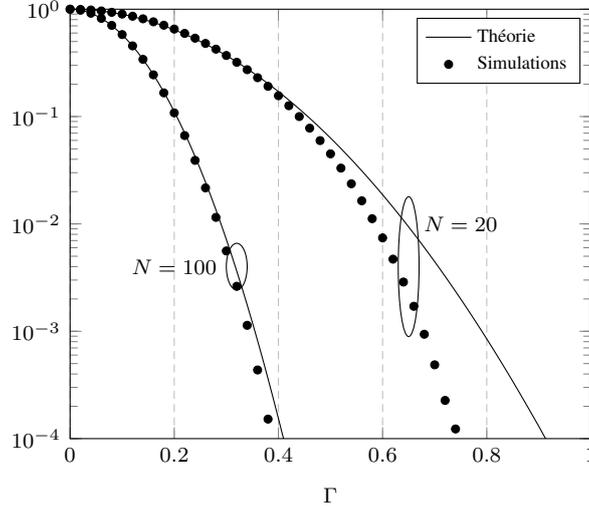


Figure 15. Probabilité de fausse alarme $P(T_N(\hat{\rho}_N^*) > \Gamma)$ pour $N = 20$ et $N = 100$, $p = N^{-\frac{1}{2}}[1, \dots, 1]^T$, $[C_N]_{ij} = 0.7^{|i-j|}$, $c_N = 1/2$

matrices aléatoires. Une fois cette difficulté dépassée, la plupart des informations pertinentes concernant ces estimateurs sont facilement évaluées via leurs approximations aléatoires.

Nous avons notamment discuté de l'étude des estimateurs robustes de matrices de dispersion C_N et plus particulièrement sous des hypothèses d'observations elliptiques. Le domaine des statistiques robustes s'étend cependant bien au-delà de ces simples cas. Concernant l'estimation robuste, non seulement les matrices de dispersions mais également les paramètres de localisation (moyenne ou médiane) sont des objets centraux qu'il convient également d'étudier dans le régime des grandes matrices aléatoires. A ce jour, il n'existe pas à la connaissance de l'auteur d'études avancées sur ce sujet. Au niveau de l'estimation robuste de matrices de dispersion, il convient également de remarquer que l'hypothèse de modèles x_i de moyenne nulle est assez restrictive et que l'écriture $x_i = \sqrt{\tau_i} C_N^{\frac{1}{2}} w_i$ avec w_i unitairement invariant est aussi très restrictive et potentiellement inadéquate dans de nombreux scénarios. Par exemple, nous avons remarqué que les performances des estimateurs de type shrinkage peuvent s'apparenter aux performances d'estimateurs bien plus simples, telles que $(1 - \rho) \frac{1}{n} \sum_{i=1}^n \frac{x_i x_i^*}{\|x_i\|^2} + \rho I_N$; ceci pose la question de la pertinence de cet estimateur s'il se comporte asymptotiquement comme un estimateur bien plus simple. Cependant, notons ici que cette équivalence n'est valable que pour des observations x_i elliptiques et nous avons en particulier remarqué en section 4 que pour des x_i arbitraires il en est tout autre et que les estimateurs robustes jouissent d'autres propriétés les rendant plus favorables que des estimateurs de type $(1 - \rho) \frac{1}{n} \sum_{i=1}^n \frac{x_i x_i^*}{\|x_i\|^2} + \rho I_N$. Un

certain nombre de questions restent ouvertes ici qui permettront de mieux comprendre l'entière des capacités de ces estimateurs robustes.

Maintenant, venons en à d'autres quantités des statistiques robustes. Outre les estimateurs, les régresseurs robustes ont aussi connu leur heure de gloire dans les années soixante-dix pour tomber également en désuétude devant leur difficulté d'analyse. Aujourd'hui, dans plusieurs articles parus en parallèle, de nouvelles considérations ont été faites sur ces régresseurs qui, dans le régime des grandes matrices aléatoires, se comportent également de manière bien plus analysable qu'à horizon fini, voir par exemple (El Karoui, 2013). Les résultats sont cependant techniquement plus compliqués à obtenir et valables, pour l'heure, seulement pour des données gaussiennes (ainsi que des résultats non rigoureux pour des données elliptiques). Les arguments théoriques utilisés pour prouver les résultats de cet article ne semblent en particulier plus appropriés pour ces nouveaux objets et un pan entier de théorie nouvelle doit être repensée ici.

Enfin, nous avons débuté cet article en mentionnant l'importance de notre étude dans le cadre récent et à fort potentiel attractif du BigData. L'estimation robuste n'est qu'un outil parmi beaucoup d'autres qui permettront demain de mieux appréhender les grandes et nombreuses données des années à venir. Dans ce futur, la théorie des matrices aléatoires aura un rôle important à jouer. C'est déjà le cas du domaine récent des méthodes d'estimation parcimonieuse pour lequel les matrices aléatoires sont au cœur des dites propriétés d'isométrie restreintes (RIP). De très récents travaux de l'auteur portent également à penser que de nouvelles méthodes dans le cadre de l'analyse parcimonieuse en composantes principales peut fortement jouir des résultats récents en matrices aléatoires.

Remerciements

Cette étude a été soutenue par le projet ANR DIONISOS (ANR-12-MONU-0003) ainsi que par le projet "Jeunes Chercheurs" du GdR ISIS-GRETSI.

Bibliographie

- Abramovich Y., Besson O. (2013). Regularized covariance matrix estimation in complex elliptically symmetric distributions using the expected likelihood approach-part 1: The over-sampled case. *IEEE Transactions on Signal Processing*, vol. 61, n° 23, p. 5807-5818.
- Bai Z. D., Silverstein J. W. (1998). No eigenvalues outside the support of the limiting spectral distribution of large dimensional sample covariance matrices. *The Annals of Probability*, vol. 26, n° 1, p. 316-345.
- Bai Z. D., Silverstein J. W. (2009). *Spectral analysis of large dimensional random matrices* (second éd.). New York, NY, USA, Springer Series in Statistics.
- Baik J., Silverstein J. W. (2006). Eigenvalues of large sample covariance matrices of spiked population models. *Journal of Multivariate Analysis*, vol. 97, n° 6, p. 1382-1408.

- Benaych-Georges F., Nadakuditi R. R. (2012). The singular values and vectors of low rank perturbations of large rectangular random matrices. *Journal of Multivariate Analysis*, vol. 111, p. 120–135.
- Chen Y., Wiesel A., Hero A. O. (2011). Robust shrinkage estimation of high-dimensional covariance matrices. *IEEE Transactions on Signal Processing*, vol. 59, n° 9, p. 4097–4107.
- Couillet R. (2015). Robust spiked random matrices and a robust G-MUSIC estimator. *J. Multivar. Anal.*, vol. 140, p. 139–151.
- Couillet R., Debbah M., Silverstein J. W. (2011, juin). A deterministic equivalent for the analysis of correlated MIMO multiple access channels. *IEEE Transactions on Information Theory*, vol. 57, n° 6, p. 3493–3514.
- Couillet R., Hachem W. (2013). Fluctuations of spiked random matrix models and failure diagnosis in sensor networks. *IEEE Transactions on Information Theory*, vol. 59, n° 1, p. 509–525.
- Couillet R., Hachem W. (2014). Analysis of the limit spectral measure of large random matrices of the separable covariance type. *Random Matrix Theory and Applications*, vol. 3, n° 4, p. 1–23.
- Couillet R., Kammoun A., Pascal F. (2014). Second order statistics of robust estimators of scatter. Application to GLRT detection for elliptical signals. *arXiv preprint arXiv:1410.0817*.
- Couillet R., McKay M. (2014). Large dimensional analysis and optimization of robust shrinkage covariance matrix estimators. *Journal of Multivariate Analysis*, vol. 131, p. 99–120.
- Couillet R., Pascal F., Silverstein J. W. (2014). Robust Estimates of Covariance Matrices in the Large Dimensional Regime. *IEEE Transactions on Information Theory*, vol. 60, n° 11, p. 7269–7278.
- Couillet R., Pascal F., Silverstein J. W. (2015). The random matrix regime of Maronna’s M-estimator with elliptically distributed samples. *Journal of Multivariate Analysis*, vol. 139, p. 56–78.
- Donoho D. L. (2000). High-dimensional data analysis: The curses and blessings of dimensionality. *AMS Math Challenges Lecture*, p. 1–32.
- El Karoui N. (2013). Asymptotic behavior of unregularized and ridge-regularized high-dimensional robust regression estimators: rigorous results. *arXiv preprint arXiv:1311.2445*.
- Hachem W., Loubaton P., Najim J. (2007). Deterministic equivalents for certain functionals of large random matrices. *Annals of Applied Probability*, vol. 17, n° 3, p. 875–930.
- Huber P. J. (1964). Robust estimation of a location parameter. *The Annals of Mathematical Statistics*, vol. 35, n° 1, p. 73–101.
- Kammoun A., Couillet R., Pascal F., Alouini M.-S. (2015). Optimal design of the adaptive normalized matched filter detector. *arXiv preprint arXiv:1501.06027*.
- Kammoun A., Kharouf M., Hachem W., Najim J. (2009). A central limit theorem for the sinr at the lmmse estimator output for large-dimensional signals. *IEEE Transactions on Information Theory*, vol. 55, n° 11, p. 5048–5063.
- Kent J. T., Tyler D. E. (1991). Redescending M-estimates of multivariate location and scatter. *The Annals of Statistics*, p. 2102–2119.

- Ledoit O., Wolf M. (2003). Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. *Journal of Empirical Finance*, vol. 10, n° 5, p. 603–621.
- Ledoit O., Wolf M. (2004). A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis*, vol. 88, n° 2, p. 365–411.
- Maronna R. A. (1976). Robust M-estimators of multivariate location and scatter. *The Annals of Statistics*, vol. 4, p. 51-67.
- Marčenko V. A., Pastur L. A. (1967). Distribution of eigenvalues for some sets of random matrices. *Math USSR-Sbornik*, vol. 1, n° 4, p. 457-483.
- Mestre X. (2008, novembre). Improved estimation of eigenvalues of covariance matrices and their associated subspaces using their sample estimates. *IEEE Transactions on Information Theory*, vol. 54, n° 11, p. 5113-5129.
- Morales-Jimenez D., Couillet R., McKay M. (2015). Large dimensional analysis of Maronna's M-estimator with outliers. In *Ieee international conference on acoustics, speech and signal processing (icassp'15)*. Brisbane, Australia.
- Pascal F., Chitour Y., Quek Y. (2014). Generalized robust shrinkage estimator and its application to stap detection problem. *Signal Processing, IEEE Transactions on*, vol. 62, n° 21, p. 5640–5651.
- Rubio F., Mestre X., Palomar D. P. (2012). Performance analysis and optimal selection of large minimum variance portfolios under estimation risk. *IEEE Journal of Selected Topics in Signal Processing*, vol. 6, n° 4, p. 337–350.
- Schmidt R. (1986). Multiple emitter location and signal parameter estimation. *IEEE Transactions on Antennas and Propagation*, vol. 34, n° 3, p. 276-280.
- Silverstein J. W., Bai Z. D. (1995). On the empirical distribution of eigenvalues of a class of large dimensional random matrices. *Journal of Multivariate Analysis*, vol. 54, n° 2, p. 175-192.
- Silverstein J. W., Choi S. (1995). Analysis of the limiting spectral distribution of large dimensional random matrices. *Journal of Multivariate Analysis*, vol. 54, n° 2, p. 295-309.
- Yang L., Couillet R., McKay M. (2014). Minimum variance portfolio optimization with robust shrinkage covariance estimation. In *Proc. ieee asilomar conference on signals, systems, and computers*. Pacific Grove, CA, USA.
- Yang L., Couillet R., McKay M. (2015). A robust statistics approach to minimum variance portfolio optimization. *IEEE Transactions on Signal Processing*, vol. 63, n° 24, p. 6684–6697.

Article soumis le 19/01/2015

Accepté le 12/10/2015