Random matrices and application to detection and estimation in array processing

Julia Vinogradova

ParisTech
INSTITUT DES SCIENCES ET TECHNOLOGIES
PARIS INSTITUTE OF TECHNOLOGY

TELECOM ParisTech

2014-ENST-0073

EDITE - ED 130

Doctorat ParisTech

T H È S E

pour obtenir le grade de docteur délivré par

TELECOM ParisTech

Spécialité « Electronique et Communications »

présentée et soutenue publiquement par

Julia VINOGRADOVA

le 27 novembre 2014

Matrices aléatoires et application à la détection et estimation en traitement d'antennes

Directeur de thèse : **Walid HACHEM**
Co-encadrement de la thèse : **Romain COUILLET**

THÈSE

Jury

M. Pascal LARZABAL, Professeur à l'ENS de Cachan	Rapporteur
M. Philippe LOUBATON, Professeur à l'Université Marne-la-Vallée	Rapporteur
M. Frédéric PASCAL, Professeur à Supélec	Examinateur
M. Romain COUILLET, Maître de conférences à Supélec	Examinateur, co-directeur de thèse
M. Walid HACHEM, Directeur de Recherche au CNRS à Télécom ParisTech	Directeur de thèse
M. Jean-Yves TOURNERET, Professeur à INP-ENSEEIHT Toulouse	Président

Télécom ParisTech
Grande école de l'Institut Mines-Télécom - membre fondateur de ParisTech
46 rue Barrault 75013 Paris - (+33) 1 45 81 77 77 - www.telecom-paristech.fr

Contents

Remerciements

Tout d'abord je tiens à remercier Pascal Larzabal et Philippe Loubaton pour avoir accepté d'être les rapporteurs de cette thèse. Je remercie également Jean-Yves Tourneret d'avoir accepté d'être le président du jury et Frédéric Pascal d'en être l'examinateur.

Je souhaite exprimer ma profonde gratitude à mes directeurs de thèse Walid et Romain pour m'avoir initiée à la recherche d'une très grande rigueur et pour toutes les connaissances que j'ai pu acquérir au cours de cette thèse. Je suis également très reconnaissante pour leur patience, leur grande disponibilité et tous les conseils qu'ils m'ont apportés.

Je voudrais remercier Chantal, Yvonne et, en particulier, Zouina pour leur précieuse aide administrative mais aussi morale.

Merci à tous les collègues d'avoir contribué à créer un environnement de travail chaleureux et amical. Je remercie à part mes chers collègues de bureau, Asma, Elie et Mohamed, pour leur bonne humeur malgré tout.

Merci à tous les amis et tous ceux qui étaient là pour moi tout au long de cette thèse mais aussi à tous les autres moments de ma vie.

Je remercie mes parents Svetlana et Vassili pour leur amour infini et tous les sacrifices incommensurables qu'ils ont faits pour nous.

Abstract

Consider a sensor network with N sensors observing T successive snapshots of K source signals. The aim is to derive parameter estimators considering two main difficulties arising in modern sensor networks. Usually scenarios with large dimensional systems and fast dynamics where T is limited and is generally of the same order of magnitude as N are considered. Therefore, it is natural to assume the asymptotic regime denoted by $T \to \infty$, where T converges to infinity while $N/T \to c > 0$. Therefore, the classical parameter estimation methods fail. In this regime, large dimensional random matrix theory tools allow to construct (N, T)-consistent estimators for the system parameters. The second difficulty comes from the fact that usually the received signals are embedded in a temporally (or spatially) correlated noise, *i.e.*, there is a dependency between the noise data across successive observations (or across the sensors). Such scenarios are usually met for instance in radar systems. The aim of this thesis is to develop consistent parameter estimators under this setting.

The studies in this thesis follow two different axes. According to the first axis, we do not make any assumption on the statistics of the noise samples. We propose a detection algorithm of the number of sources and estimation methods for their powers and the directions-of-arrival which are based on the sample covariance matrix of the signal-plus-noise model. Within the second axis, we assume that the noise is a stationary process whose covariance matrix has a Toeplitz structure. We revisit the known approaches for estimation of such matrices based on a Toeplitzified version of the sample covariance matrix. The main contribution of this work consists in establishing concentrations inequalities on the spectral norm of the noise covariance matrix, whether or not the signal is present. The well-known "whitening" procedure leads back to the white noise case.

Acronyms

AIC	Akaike's Information Criterion
AR	AutoRegressive
ARMA	AutoRegressive Moving Average
a.s.	almost surely
c.d.f.	cumulative distribution function
CDR	Correct Detection Rate
CLT	Central Limit Theorem
dB	deciBel
FAR	False Alarm Rate
GLRT	Generalized Likelihood-Ratio Test
GUE	Gaussian Unitary Ensemble
LRT	Likelihood-Ratio Test
MDL	Minimum Description Length
MIMO	Multiple-Input Multiple-Output
ML	Maximum Likelihood
MSE	Mean Square Error
MUSIC	Multiple SIgnal Classification
NMSE	Normalized Mean Square Error
QPSK	Quadrature Phase-Shift Keying
SNR	Signal-to-Noise Ratio
TW	Tracy–Widom

Notations

I_N	$N \times N$ identity matrix
$\mathcal{T}\left(x_{-(T-1)}, \ldots, x_{T-1}\right)$	Toeplitz matrix formed from the coefficients $x_{-(T-1)}, \ldots, x_{T-1}$
$\mathrm{diag}(x_0, \ldots, x_{T-1})$	Diagonal matrix with entries $x_0, \ldots, x_{T-1}$
X^{T}	Transpose of X
X^{H}	Hermitian transpose of X
$\det(X)$	Determinant of X
$\mathrm{Tr}(X)$	Trace of X
$\mathrm{rank}(X)$	Rank of X
$\|X\|$	Spectral norm of X
$\|\cdot\|_{\mathrm{fro}}$	Frobenius norm
$X \odot Y$	Hadamard product of X and Y
$\mathbb{C}$, $\mathbb{R}$, $\mathbb{Z}$, $\mathbb{N}$	Set of complex, real, rational, and natural numbers
$\Re(z)$	Real part of z
$\Im(z)$	Imaginary part of z
$\|f\|_\infty$	sup of the function f
x^+	Right-limit of the real x
x^-	Left-limit of the real x
$(x)^+$	For $x \in \mathbb{R}, \max(x, 0)$
$\lfloor \cdot \rfloor$	Floor function
$\mathbb{1}_A$	Indicator function on the set A
$\boldsymbol{\delta}_{k\ell}$	Kronecker delta function ($= 1$ if $k = \ell$ and 0 otherwise)
$\mathbf{d}(x, y)$	Distance from x to y
$\mathcal{O}(\cdot)$	Landau's big-$\mathcal{O}$
$o(\cdot)$	Landau's small-o
$\#\{A\}$	Cardinality of the set A
$\mathbb{E}\{X\}$	Statistical expectation of X

$\mathbb{P}\{X\}$	Probability of the event X
$\text{supp}(\mu)$	Support of the measure μ
$\mathcal{N}(a, \sigma^2)$	Real Gaussian distribution with mean a and variance σ^2
$\mathcal{CN}(a, \sigma^2)$	Complex circular Gaussian distribution with mean a and variance σ^2
$\xrightarrow{\text{a.s.}}$	Almost sure convergence
$\xrightarrow{\mathcal{P}}$	Convergence in probability
$\xrightarrow{\mathcal{L}}$	Convergence in law
w.p. 1	With probability one

Résumé en français

1 Introduction

1.1 Modèle du système

Nous considérons un réseau linéaire constitué de N capteurs observant des signaux issus de K sources pendant une fenêtre d'observation de taille T. Le signal reçu $y_t \in \mathbb{C}^{N\times 1}$ à l'instant $t = 0, \ldots, T-1$ est donné par

$$y_t = Hs_t + w_t$$

où $s_t \in \mathbb{C}^{K\times 1}$ est le vecteur de symboles aléatoires transmis de matrice de covariance Γ, $H \in \mathbb{C}^{N\times K}$ est la matrice de canal déterministe et $w_t \in \mathbb{C}^{N\times 1}$ est le vecteur de bruit additif à éléments i.i.d. complexes gaussiens de variance σ^2. Nous allons supposer dans la suite que tous les paramètres du système, à l'exception du nombre de capteurs N et la taille de la fenêtre d'observation T, sont inconnus au récepteur. Un problème fondamental en traitement d'antennes consiste à développer des algorithmes d'inférence sur la partie signal Hs_t à partir de la matrice de covariance empirique uniquement. Le but est en particulier de détecter le nombre de sources émettrices et d'estimer certains paramètres, par exemple, les puissances des sources et les directions d'arrivée. Les méthodes proposées dans la littérature sont basées sur la structure de la matrice de covariance du signal reçu :

$$\Sigma = H\Gamma H^{\mathsf{H}} + \sigma^2 I_N$$

où $H\Gamma H^{\mathsf{H}}$ correspond à la partie signal et $\sigma^2 I_N$ est la matrice de covariance du bruit. En pratique, nous n'avons pas d'accès à la matrice Σ et elle est estimée par la matrice de covariance empirique

$$\widehat{\Sigma} = \frac{1}{T}\sum_{t=0}^{T-1} y_t y_t^{\mathsf{H}}.$$

Dans le cas du régime classique pour lequel N est fixe quand $T \to \infty$, par la loi des grands nombres, la matrice de covariance empirique est un estimateur consistant de la matrice de covariance de la population. D'où, beaucoup de méthodes de détection existantes sont basées sur l'utilisation de la matrice de covariance empirique, en particulier, sur son spectre. Ces approches exploitent le fait que quand $T \to \infty$, les plus petites valeurs propres sont proches les unes des autres et convergent vers σ^2. Parmi les méthodes de détection classiques, nous avons le célèbre critère d'information d'Akaike (AIC) et la longueur de description minimale (MDL) (voir l'approche de Wax et Kailath [67]).

1.2 Régime de grandes dimensions

Les dimensions des vecteurs d'observations dans les systèmes de communications modernes deviennent de plus en plus grandes. Souvent les changements de la dynamique du système étant très rapides, la taille de la fenêtre d'observation est de même ordre de grandeur que les dimensions du système. Dans ce cas il est pertinent de supposer que N et T convergent vers l'infini à la même vitesse. Dans ce contexte, N et T étant grands tels que $N/T \to c > 0$ quand $T \to \infty$, la matrice de covariance empirique n'est pas un estimateur consistant de la vraie matrice de covariance. Par conséquent, les méthodes de détection et d'estimation classiques basées sur la matrice de covariance empirique ne permettent pas d'obtenir des estimateurs consistants. Dans ce régime, les outils de grandes matrices aléatoires permettent de construire des estimateurs des paramètres (N,T)-consistants. En concaténant tous les vecteurs reçus dans la matrice de dimensions $N \times T$, le modèle de transmission s'écrit sous forme matricielle :

$$Y_T = A_T + W_T \tag{1}$$

où $A_T = H_T S_T^{\mathsf{H}}$ est la matrice correspondante à la partie signal avec $S_T = [s_0, \ldots, s_{T-1}]^{\mathsf{H}} \in \mathbb{C}^{T \times K}$ et $W_T = [w_0, \ldots, w_{T-1}] \in \mathbb{C}^{N \times T}$ est la matrice du bruit. La matrice A_T est supposée de petit rang K quand $T \to \infty$. La matrice de rang plein Y_T peut être vue comme une version perturbée de la matrice de bruit W_T, la perturbation A_T étant de petit rang. En théorie de grandes matrices aléatoires le modèle de transmission (1) correspond au modèle de spikes [35], [9]. Les détecteurs et les estimateurs proposés sont basés sur l'étude du comportement limite des plus grandes valeurs propres de la matrice de covariance empirique $\widehat{\Sigma}_T = \frac{1}{T} Y_T Y_T^{\mathsf{H}}$.

1.3 Bruit blanc

Les approches basées sur les tests d'hypothèse testent l'hypothèse nulle H_0 (le signal est absent) contre l'hypothèse H_1 (le signal est présent). Sous l'hypothèse H_0, la matrice de covariance empirique s'écrit $\widehat{\Sigma}'_T = \frac{1}{T} W_T W_T^{\mathsf{H}}$. Quand N et T convergent vers l'infini, W_T étant à entrées i.i.d., la distribution des valeurs propres de $\widehat{\Sigma}'_T$ converge vers la célèbre loi de Marčenko–Pastur [44]. La matrice de covariance empirique forme ce qu'on appelle le « paquet » correspondant aux valeurs propres de bruit qui sont toutes asymptotiquement situées dans l'intervalle $[a, b]$, le support de la distribution de Marčenko–Pastur. Sous l'hypothèse H_1, le spectre de la matrice de covariance empirique est composé du « paquet » de bruit et de quelques valeurs propres éventuelles isolées situées à droite du « paquet » qui sont dues à la présence du signal. La condition de présence des valeurs propres isolées est liée à la puissance du signal correspondante et au paramètre c. Beaucoup de méthodes de détection qui se basent sur la théorie des spikes ont été étudiées au cours de ces dernières années. L'une des premières contributions appartient à Nadakuditi et Edelman [46] qui ont amélioré les algorithmes basés sur les critères AIC et MDL dans le contexte des grandes dimensions. Pour le cas d'une seule source un test statistique a été proposé par Bianchi *et al.* [14] qui est basé sur le ratio de la plus grande valeur propre de $\widehat{\Sigma}_T$ sur sa trace normalisée. Dans la littérature ce test se réfère au test du rapport de vraisemblance généralisé (GLRT). Pour le cas de sources multiples, une approche basée sur des tests d'hypothèses multiples a été proposée par Kritchman et Nadler [40]. Toutefois, l'hypothèse de bruit blanc ne peut pas s'appliquer dans de nombreuses situations pratiques.

1.4 Bruit corrélé

Les observations successives du vecteur de bruit peuvent ne pas être indépendantes d'une observation à l'autre ce qui correspond au bruit corrélé temporellement. De façon alternative, du au fait de la proximité des capteurs, le vecteur de bruit pourrait présenter des corrélations spatiales. Dans le scénario de corrélations temporelles, la matrice de bruit est de la forme $V_T = W_T R_T^{1/2}$ où W_T est une matrice aléatoire $N \times T$ à éléments i.i.d. et R_T est une matrice de covariance inconnue qui capte les corrélations temporelles des échantillons de bruit reçues par une antenne. Quand R_T est connue, la matrice de covariance empirique peut être blanchie par l'inverse de R_T et nous retrouvons la situation classique de signal altéré par un bruit blanc. Dans le cas de détection d'une seule source, le test GLRT [14] est

utilisé. Quand R_T est inconnue, on suppose en général l'existence d'une séquence indépendante d'échantillons de bruit pur de taille T'. A partir de cette séquence une matrice de covariance empirique de bruit est construite qui est supposée de représenter la vraie matrice de covariance. La matrice de covariance sous le test est « blanchie » par la matrice empirique de bruit pur pour donner lieu à ce qu'on appelle « F-matrix » [58] perturbée par un signal hypothétique. Cette approche a été étudiée par Nadakuditi et Silverstein [47]. L'hypothèse d'existence d'une séquence de bruit seul est très forte et nous proposons de l'éviter. Dans cette thèse nous considérons une séquence contenant potentiellement le signal.

1.5 Contributions de la thèse

Nos études ont été menées suivant deux axes de recherche différents :

- **Axe 1** : Nous proposons un algorithme de détection du nombre de sources basé sur les espacements entre les valeurs propres successives de la matrice de covariance empirique. Cet algorithme ne fait aucune hypothèse sur les statistiques des échantillons de bruit. En outre, nous proposons une méthode d'estimation de puissances de sources et un algorithme d'estimation de directions d'arrivée basé sur l'adaptation de l'approche MUSIC [57] au contexte des grandes matrices aléatoires.

- **Axe 2** : Nous supposons que R_T est structurée, étant la matrice de covariance d'un processus de bruit stationnaire. Dans ce contexte, nous réexaminons les approches d'estimation de R_T connues basées sur une version « Toeplitzifiée » de la matrice de covariance empirique. Après avoir estimé R_T, le « blanchiment » nous fait revenir au test GLRT. L'idée est de dire que même si le signal est présent, à ce stade il est considéré comme une nuisance, l'estimée de la matrice de covariance reste consistante. Cela est du au petit rang du signal. L'originalité de notre travail consiste en établissement d'inégalités de concentration sur l'erreur en norme spectrale de la matrice de covariance de bruit, que le signal soit présent ou non.

1.6 Résumé du contenu du manuscrit

Ce manuscrit est composé de deux parties principales. La première partie introduit le contexte, présente des résultats importants en théorie de grandes matrices aléatoires utilisés par la suite et parcourt l'état de l'art dans les chapitres 1, 2 et 3, respectivement. Les contributions sont données dans les

chapitres 3 et 4 et dont le contenu est le suivant :

Chapitre 4 suppose que la structure de la matrice de covariance de bruit n'est pas connue. Un détecteur de sources et des estimateurs de puissances et d'angles d'arrivée sont proposés d'abord pour un modèle général puis pour un exemple spécifique de traitement d'antennes en bande étroite.

Chapitre 5 suppose que R_T est la matrice de covariance d'un processus stationnaire gaussien ayant une structure Toeplitz. Des inégalités de concentration sur la norme spectrale sont obtenues en absence du signal et sous sa présence. Finalement, un test de détection sur la matrice de covariance « blanchie » est proposé.

1.7 Publications

Les publications suivantes sont associées à cette thèse :

Articles de revue

- J. Vinogradova, R. Couillet, W. Hachem, "Estimation of Toeplitz covariance matrices in large dimensional regime with application to source detection," accepté dans IEEE Transactions on Signal Processing, novembre 2014.

- J. Vinogradova, R. Couillet, W. Hachem, "Statistical inference in large antenna arrays under unknown noise pattern," IEEE Transactions on Signal Processing, 61 (22), 2013, pages 5633–5645.

Articles de congrès

- J. Vinogradova, R. Couillet, W. Hachem, "Estimation of large Toeplitz covariance matrices and application to source detection," EUSIPCO'2014, Lisbonne, Portugal.

- J. Vinogradova, R. Couillet, W. Hachem, « Nouvelle méthode de détection de sources, d'estimation de puissances et de localisation dans un système de communication sans fil avec des statistiques de bruit inconnues », GRETSI'2013, Brest, France.

- J. Vinogradova, R. Couillet, W. Hachem, "A new method for source detection, power estimation, and localization in large sensor networks under noise with unknown statistics," ICASSP'2013, Vancouver, Canada.

2 Détection/estimation d'un signal de petit rang en présence d'un bruit corrélé

2.1 Modèle général

Considérons le modèle de transmission général :

$$Y_T = A_T + V_T \tag{2}$$

où A_T est une matrice aléatoire de rang fixe K quand $T \to \infty$ représentant le nombre de signaux, $V_T = W_T R_T^{1/2}$ est la matrice de bruit corrélé temporellement avec $W_T \in \mathbb{C}^{N\times T}$ à entrées i.i.d. standard gaussiennes et R_T est une matrice semi-définie positive et hermitienne à norme spectrale bornée. Dans toute la suite nous notons le régime asymptotique par $T \to \infty$, $N/T \to c > 0$. Rappelons que la matrice de covariance empirique est donnée par

$$\widehat{\Sigma}_T = \frac{1}{T} Y_T Y_T^{\mathsf{H}}. \tag{3}$$

Le modèle (4.1) correspond également au modèle de spikes. La matrice A_T est vue comme une perturbation de rang fini de V_T. Comme dans le cas du bruit blanc, le spectre de $\widehat{\Sigma}_T$ est composé d'un « paquet » de valeurs propres correspondant au bruit et, éventuellement, de quelques valeurs propres isolées à droite du support de la mesure spectrale limite de $\frac{1}{T} V_T V_T^{\mathsf{H}}$ [9]. Les estimateurs présentés ici sont basés sur l'étude du comportement asymptotique de ces valeurs propres isolées. Nous réalisons une inférence statistique sur la partie information A_T et fournissons des estimateurs du nombre de signaux, des puissances de ces signaux et de quelques formes bilinéaires faisant intervenir la matrice de projection orthogonale sur l'espace signal (voir chapitre 4, section 4.3). Sous plus de contraintes, ces résultats sont utilisés dans le cas pratique de traitement d'antennes en bande étroite pour effectuer des algorithmes de détection et d'estimation.

2.2 Traitement d'antennes en bande étroite

2.2.1 Modèle

Nous considérons des signaux issus de K sources reçus par un réseau de N capteurs pendant T observations successives. Le signal reçu $y_t \in \mathbb{C}^{N\times 1}$ à

l'instant t est donné par

$$y_t = \sum_{k=0}^{K-1} \sqrt{p_k} h_T(\theta_k) s_{k,t} + v_t$$

où p_k est la puissance de la source k avec $p_0 \geq \ldots \geq p_{K-1}$, $\theta_k \in [-\pi/2, \pi/2]$ est son angle d'arrivée, $h_T(\theta_k) = \frac{1}{\sqrt{N}} \left[1, e^{-2\imath\pi d \sin\theta_k}, \ldots, e^{-2\imath\pi d(N-1)\sin\theta_k}\right]^{\mathsf{T}} \in \mathbb{C}^{N\times 1}$ est le vecteur directionnel avec $d > 0$. Le signal transmis par la source k à l'instant t est représenté par $s_{k,t}$ et le bruit par le vecteur v_t. La relation entre entrées-sorties du système en concaténant T réalisations du signal successives s'écrit

$$Y_T = H_T P^{1/2} S_T^{\mathsf{H}} + V_T \tag{4}$$

où $Y_T = [y_0, \ldots, y_{T-1}]$, $H_T = [h_T(\theta_0), \ldots, h_T(\theta_{K-1})]$, $P = \mathrm{diag}(p_0, \ldots, p_{K-1})$, $S_T = T^{-1/2}[s^*_{t,k}]_{t,k=0}^{T-1,K-1}$ avec $s_{t,k}$ aléatoires i.i.d. de moyenne nulle, variance unité et le moment d'ordre huit fini et $V_T = [v_0, \ldots, v_{T-1}]$. Nous supposons que le bruit est corrélé temporellement, *i.e.*, les colonnes de V_T ne sont pas indépendantes. On suppose ici que le bruit est un processus stationnaire causal ARMA (autoregressive moving average), mais cette hypothèse n'est pas nécessaire pour la validité des résultats. Chaque colonne de V_T est la réponse d'un filtre dont la fonction de transfert est donnée par $\mathbf{p}(z) = \sum_{l=0}^{\infty} \psi_l z^{-l}$ et dont l'entrée est le bruit blanc. On écrit alors $V_T = W_T R_T^{1/2}$ où $[W_T]_{i,j} \sim \mathcal{CN}(0, 1/T)$ et $R_T \in \mathbb{C}^{T\times T}$ est une matrice de Toeplitz semi-définie positive de mesure spectrale ν_T. Dans le cadre du bruit ARMA, la matrice R_T a la forme suivante :

$$R_T = \begin{bmatrix} r_0 & r_1 & \ldots & r_{T-1} \\ r_{-1} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & r_1 \\ r_{1-T} & \ldots & r_{-1} & r_0 \end{bmatrix}$$

avec $r_k \triangleq \sum_{l\geq 0} \psi_{l+k}\psi_l^*$ et $k \in \mathbb{Z}$. D'après [30, Lemma 6], ν_T converge vers ν dont le support est un intervalle compact et toutes les valeurs propres de R_T sont asymptotiquement contenues dans le support de ν. Il est aussi possible de caractériser la mesure spectrale de la matrice $\frac{1}{T} V_T V_T^{\mathsf{H}}$ qui converge vers la mesure limite notée μ dont la transformée de Stieltjes est la solution d'une équation à point fixe [59]. Le but est d'estimer le nombre de sources émettrices, leurs puissances et leurs angles d'arrivée. Les fluctuations des estimateurs de puissance sont également étudiées.

2.2.2. Préliminaires

Donnons d'abord une caractérisation du comportement limite du spectre correspondant à la partie bruit. Dans le cadre du modèle (4) et le régime asymptotique $T \to \infty$, $N/T \to c > 0$ et K fixe, la transformée de Stieltjes de μ, la mesure spectrale limite de $\frac{1}{T} V_T V_T^{\mathsf{H}}$, est donnée par la solution de l'équation

$$m(z) = \left(-z + \int_0^1 \frac{|\mathbf{p}(e^{2\pi i u})|^2}{1 + cm(z)|\mathbf{p}(e^{2\pi i u})|^2} du \right)^{-1} \tag{5}$$

avec, pour $z \in \mathbb{C}^+$, $m(z) \in \mathbb{C}^+$. Ce résultat est un corollaire de [59] et [30, Lemme 6].

La borne supérieure du support de μ est caractérisée par la proposition suivante :

Proposition 1. *Soit μ la mesure spectrale dont la transformée de Stieltjes est la solution de l'équation (5) et dont le support est l'intervalle $[a, b]$. Alors,*

$$b = -\frac{1}{m_b} + \int_0^1 \frac{|\mathbf{p}(e^{2\pi i u})|^2}{1 + cm_b|\mathbf{p}(e^{2\pi i u})|^2} du$$

où m_b est la solution unique dans $(-(c\max_u\{|\mathbf{p}(e^{2\pi i u})|^2\})^{-1}, 0)$ de l'équation en variable m

$$\int_0^1 \left(\frac{m|\mathbf{p}(e^{2\pi i u})|^2}{1 + cm|\mathbf{p}(e^{2\pi i u})|^2} \right)^2 du = \frac{1}{c}.$$

La fonction $m(z)$, $z \in \mathbb{C}^+$, est prolongeable par continuité sur (b, ∞) et $\lim_{x \to b^+} m(x) = m_b$. Le comportement des K plus grandes valeurs propres de la matrice $\widehat{\Sigma}_T$ est décrit par la proposition suivante :

Proposition 2. *Soit m la transformée de Stieltjes limite de la mesure μ de support $[a, b]$. Soient m_b et b définis comme dans la Proposition 1 et la fonction $g(x) = m(x)(xcm(x) + c - 1)$ est décroissante de $m_b(cbm_b + c - 1)$ jusqu'à zéro sur (b, ∞). Soit $k \in \mathbb{N}$ le plus grand entier pour lequel*

$$p_k > p_{lim}$$

où

$$p_{lim} \triangleq 1/m_b(cbm_b + c - 1). \tag{6}$$

Soient $\hat{\lambda}_{0,T} \geq \cdots \geq \hat{\lambda}_{N-1,T}$ les valeurs propres de $\widehat{\Sigma}_T$. Si $p_0 < p_{\lim}$, alors $\hat{\lambda}_{0,T} \xrightarrow[T\to\infty]{a.s.} b$. Sinon, pour $i = 0, \ldots, k$, soit ρ_i la solution unique de l'équation $p_i g(x) = 1$ sur (b, ∞). Alors,

$$\hat{\lambda}_{0,T} \xrightarrow[T\to\infty]{a.s.} \rho_0, \ldots, \hat{\lambda}_{i,T} \xrightarrow[T\to\infty]{a.s.} \rho_i \text{ et } \hat{\lambda}_{k+1,T} \xrightarrow[T\to\infty]{a.s.} b.$$

D'après cette proposition, si la puissance d'une source est assez grande, la valeur propre correspondante est située à droite du « paquet ». Plus précisément, si $k+1$ sources ont leurs puissances plus grandes que $p_{\lim}$ alors les $k+1$ valeurs propres correspondantes seront situées à l'extérieur du support de μ. Chacune de ces valeurs propres $\hat{\lambda}_{i,T}$ converge vers ρ_i qui est une fonction de la puissance p_i. Ainsi, la position d'une valeur propre isolée peut être associée à la puissance de la source correspondante.

2.2.3 Résultats

Détection du nombre de sources

Un algorithme de détection du nombre de sources émettrices qui satisfont la condition de détectabilité est donné ci-après. Cette méthode est basée sur l'étude des espacements entre les plus grandes valeurs propres de la matrice de covariance empirique.

Proposition 3 (Estimateur du nombre de sources)**.** *Soit L la borne supérieure du nombre de sources. Soit $k = 0, \ldots, K-1$ le plus grand entier tel que*

$$p_k > p_{\lim}$$

où $p_{\lim}$ est défini par (6). Soient $\hat{\lambda}_{0,T} \geq \ldots \geq \hat{\lambda}_{N-1,T}$ les valeurs propres de $\widehat{\Sigma}_T$. Pour $L \geq K$ et $\varepsilon > 0$, on définit (avec $\hat{\lambda}_{-1} = \infty$)

$$\hat{k}_T = \arg \max_{m \in \{0,\ldots,L-1\}} \frac{\hat{\lambda}_{m-1,T}}{\hat{\lambda}_{m,T}} > 1 + \varepsilon.$$

Alors $\hat{k}_T = k$ avec probabilité 1 pour tout T grand et ε suffisamment petit.

D'après ce résultat, nous avons un estimateur consistant du nombre de sources émettrices si la puissance $p_{K-1} > p_{\lim}$, c'est-à-dire, lorsque $k+1 = K$. Toutefois cette méthode présente un problème de choix du seuil de détection. Il est connu que dans le cas du bruit blanc, en absence de signal, la plus grande valeur propre de Σ_T centrée réduite suit la loi de Tracy–Widom [34]. D'où, le seuil de détection, qui dépend de la probabilité de fausse alarme, peut être fixé théoriquement en utilisant la distribution de Tracy–Widom.

Estimation de puissances

Le résultat suivant permet d'estimer les puissances p_i pour $i \geq k$ en remplaçant ρ_k et $g(x)$ par leurs estimées basées sur $\hat{\lambda}_{i,T}$ et $c_T \triangleq N/T$.

Proposition 4 (Estimateur des puissances). *Soit $\hat{k}_T$ défini par la proposition 3 avec pour tout T grand et ε suffisamment petit. Soit*

$$\hat{g}_T(x) = \hat{m}_T(x)(xc_T\hat{m}_T(x) + c_T - 1)$$

où $\hat{m}_T(x)$ est donné par

$$\hat{m}_T(x) = \frac{1}{N - \hat{k}_T - 1} \sum_{n=\hat{k}_T+1}^{N-1} \frac{1}{\hat{\lambda}_{n,T} - x}.$$

Pour $i = 0, \ldots, \hat{k}_T$, soit

$$\hat{p}_{i,T} = \frac{1}{\hat{g}_T(\hat{\lambda}_{i,T})}.$$

Alors,

$$\hat{p}_{i,T} - p_i \xrightarrow[T\to\infty]{a.s.} 0$$

et

$$\sqrt{T}\left(\hat{p}_i - p_i\right) \xrightarrow[T\to\infty]{\mathcal{L}} \mathcal{N}\left(0, \sigma_i^2\right)$$

où σ_i^2 a une expression connue en fonction de $m(x)$.

D'après ce théorème, l'erreur de l'estimateur de puissance suit la loi centrale limite et la variance est d'ordre $1/T$.

Localisation

La méthode de localisation proposée est basée sur l'approche MUSIC [57]. Soit $k = 0, \ldots, K-1$ le plus grand entier tel que $p_k > p_{\text{lim}}$ et soit $\Pi_{k,T}$ le projecteur orthogonal sur l'espace des colonnes de $H_{k,T} = [h_T(\theta_0), \ldots, h_T(\theta_k)]$. Les angles $\theta_0, \ldots, \theta_k$ sont les solutions de l'équation $a_T(\theta)\left(I_N - \Pi_{k,T}\right)a_T(\theta)^{\mathsf{H}} = 0$. On définit par $\gamma_T(\theta) = a_T(\theta)^{\mathsf{H}}\Pi_{k,T}a_T(\theta)$ la fonction de localisation, $\theta_0, \ldots, \theta_k$ étant les arguments des maximums locaux de $\gamma_T(\theta)$. Soient $\hat{u}_{0,T}, \ldots,$
$\hat{u}_{\hat{k}_T,T}$ les vecteurs propres de $\widehat{\Sigma}_T$ associés respectivement à $\hat{\lambda}_{0,T}, \ldots, \hat{\lambda}_{\hat{k}_T,T}$. Pour l'algorithme de MUSIC classique, l'estimateur de la fonction de localisation est donné par :

$$\hat{\gamma}_T(\theta) = \sum_{k=0}^{\hat{k}_T} h_T(\theta)^{\mathsf{H}}\hat{u}_{k,T}\hat{u}_{k,T}^{\mathsf{H}}h_T(\theta).$$

Dans le contexte des grandes dimensions, l'estimateur de la fonction de localisation est donné par la proposition suivante :

Proposition 5. *Soient* $\hat{u}_{0,T}, \ldots, \hat{u}_{\hat{k}_T,T}$ *les vecteurs propres de* $\widehat{\Sigma}_T$ *associés respectivement à* $\hat{\lambda}_{0,T}, \ldots, \hat{\lambda}_{\hat{k}_T,T}$. *Pour* $\theta \in [-\pi/2, \pi/2]$, *soit*

$$\hat{\gamma}_T(\theta) = \sum_{k=0}^{\hat{k}_T} \zeta_T(\hat{\lambda}_{k,T}) h_T(\theta)^{\mathsf{H}} \hat{u}_{k,T} \hat{u}_{k,T}^{\mathsf{H}} h_T(\theta)$$

où $\zeta_T(x) = \frac{\left(x\hat{m}_T(x)\left(c\hat{m}_T(x) - (1-c)\frac{1}{x}\right)\right)'}{x\hat{m}_T(x)^2\left(c\hat{m}_T(x) - (1-c)\frac{1}{x}\right)}$.
Alors,

$$\gamma_T(\theta) - \hat{\gamma}_T(\theta) \xrightarrow[T\to\infty]{a.s.} 0.$$

Dans la suite, nous avons effectué une série de simulations afin d'observer les performances des estimateurs proposés.

2.2.4 Résultats numériques

Pour ces simulations les signaux $s_{t,k}$ sont modulés QPSK. La puissance du signal p_k définit le rapport du signal au bruit (RSB). Le bruit est supposé autorégressif d'ordre 1 et de paramètre a avec $[R_T]_{k,l} = a^{|k-l|}$. Tous les autres paramètres sont précisés dans la légende.

Les probabilités de fausse alarme (PFA) et les probabilités de détection correcte (PDC) pour une seule source sont évaluées dans la figure 1 pour des valeurs de ε différentes et pour les taux c_T croissants. Nous observons l'impact d'un choix inapproprié de ε qui, s'il est trop petit, génère une grande probabilité de fausse alarme lorsque les valeurs propres du bruit tendent à se disperser (*i.e.*, pour c_T grand). Au contraire, si ε est trop grand, il ne permet pas de détecter correctement la source dont la puissance est trop proche du seuil de détectabilité (*i.e.*, pour c_T grand).

Les erreurs quadratiques moyennes normalisées (EQMN) $\mathbb{E}[(\hat{p}_0 - p_0)^2 / p_0^{-2}]$ de l'estimateur de puissance donné par la proposition 4 sont tracées dans la figure 2 et comparées à la variance théorique donnée par la loi centrale limite de cet estimateur. Ces courbes de l'estimateur proposé sont comparées à celles de l'estimateur obtenu après le blanchiment du modèle par la vraie matrice du bruit. Nous observons que lorsqu'on est proche du seul de détectabilité la variance théorique diverge. Toutefois, à dimensions finies, les erreurs de l'estimateur restent bornées à petit RSB. Ceci est expliqué par le fait qu'« à l'horizon fini », le comportement des valeurs propres n'est pas aussi brutal. Le gap entre les courbes de l'estimateur proposé et celui avec le blanchiment est du à la corrélation du bruit. Plus le paramètre de corrélation est grand et plus le gap est grand. Les erreurs quadratiques moyennes

$\mathbb{E}[(\hat{\gamma}(\theta_0) - \gamma(\theta_0))^2]$ de la fonction de localisation en $\theta_0 = 10°$ sont données par la figure 3 et sont comparées avec les performances de l'estimateur oracle et à l'estimateur MUSIC traditionnel avec la fonction de localisation $\hat{\gamma}_{\text{trad},T}(\theta) \triangleq \sum_{k=0}^{\hat{k}_T} h(\theta)^{\mathsf{H}} \hat{u}_{k,T} \hat{u}_{k,T}^{\mathsf{H}} h(\theta)$. Nous notons que l'estimateur proposé donne bien de meilleurs performances que l'estimateur traditionnel. Le gap entre la version blanchie est toujours du aux corrélations.

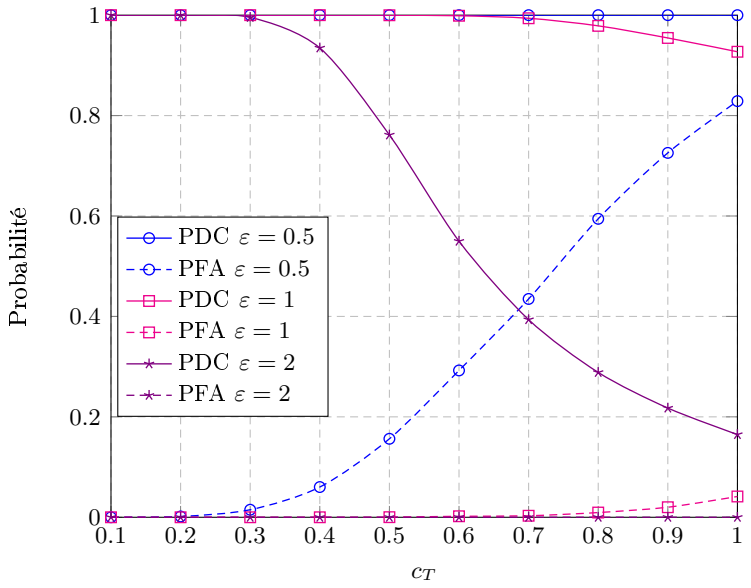

Figure 1: Probabilité de détection correcte et probabilité de fausse alarme (en pointillé) en fonction de c_T avec $K = 1$, $N = 20$, RSB= 10 dB, $L = 5$ et $a = 0.6$.

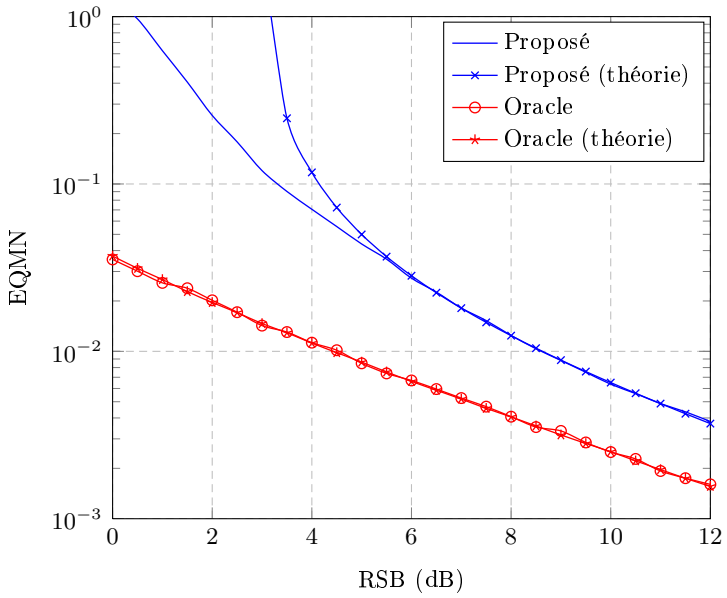

Figure 2: Erreurs quadratiques moyennes normalisées en fonction du rapport du signal au bruit avec $K = 1$, $N = 20$, $c_T = 0.5$ et $a = 0.6$.

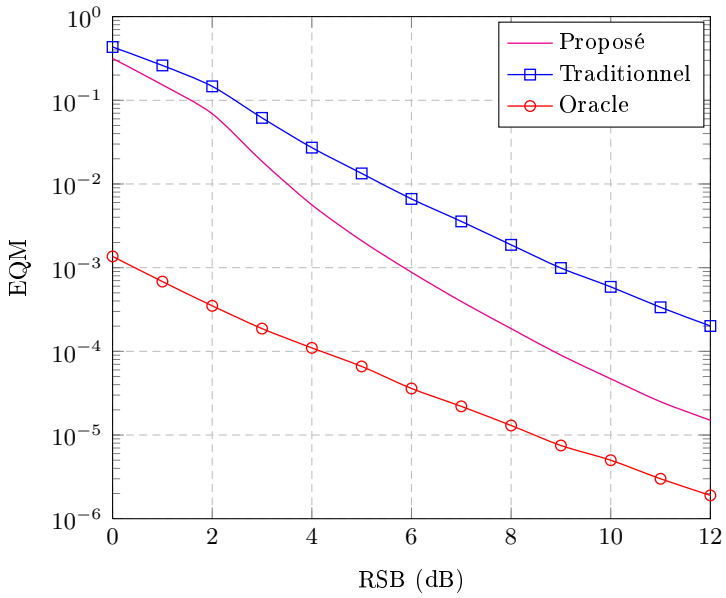

Figure 3: Erreurs quadratiques moyennes de la fonction de localisation en fonction du rapport du signal au bruit avec $K = 1$, $N = 20$, $c_T = 0.2$ et $a = 0.6$.

3 Estimation de matrices de covariance de Toeplitz et application à la détection de source

3.1 Modèle et résultats

Soit $(v_t)_{t\in\mathbb{Z}}$ un processus complexe symétrique gaussien stationnaire avec une moyenne nulle et une fonction de covariance $(r_k)_{k\in\mathbb{Z}}$ avec $r_k = \mathbb{E}[v_{t+k}v_t^*]$ et $r_k \to 0$ quand $k \to \infty$. Nous observons N copies indépendantes de $(v_t)_{t\in\mathbb{Z}}$ sous une fenêtre de taille $t \in \{0, \dots, T-1\}$, et nous concaténons les observations dans la matrice $V_T = [v_{n,t}]_{n,t=0}^{N-1,T-1}$. Cette matrice peut être écrite sous la forme $V_T = W_T R_T^{1/2}$, où $W_T \in \mathbb{C}^{N\times T}$ a des entrées indépendantes standard circulaires symétriques complexes gaussiennes et $R_T^{1/2}$ est la matrice carrée de la matrice de Toeplitz $T \times T$ semi-définie positive et hermitienne :

$$R_T \triangleq [r_{i-j}]_{0\le i,j\le T-1} = \begin{bmatrix} r_0 & r_1 & \dots & r_{T-1} \\ r_{-1} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & r_1 \\ r_{1-T} & \dots & r_{-1} & r_0 \end{bmatrix}.$$

Récemment ce problème d'estimation a attiré une nouvelle attention dans le contexte des grandes dimensions. En général, les méthodes d'estimation de R_T se basent sur les estimateurs des coefficients r_k biaisé et non biaisé classiques $\hat{r}^b_{k,T}$ et $\hat{r}^u_{k,T}$, définis respectivement par :

$$\hat{r}^b_{k,T} = \frac{1}{NT} \sum_{n=0}^{N-1} \sum_{t=0}^{T-1} v_{n,t+k} v^*_{n,t} \mathbb{1}_{0 \leq t+k \leq T-1}$$

$$\hat{r}^u_{k,T} = \frac{1}{N(T-|k|)} \sum_{n=0}^{N-1} \sum_{t=0}^{T-1} v_{n,t+k} v^*_{n,t} \mathbb{1}_{0 \leq t+k \leq T-1}$$

où $\mathbb{1}_A$ est la fonction indicatrice de l'ensemble A. En fonction du taux de convergence de N et T les estimées $\widehat{R}^b_T = [\hat{r}^b_{i-j,T}]_{0 \leq i,j \leq T-1}$ et $\widehat{R}^u_T = [\hat{r}^u_{i-j,T}]_{0 \leq i,j \leq T-1}$ peuvent ne pas être consistantes. Les approches d'estimations développées pendant la dernière décennie proposent toutes de construire des versions « fenêtrées » de la matrice estimée $\widehat{R}_T$ en réduisant ou en mettant à zéro les entrées qui sont suffisamment loin de la diagonale principale [71, 15, 72, 19, 18]. Ces méthodes donnent lieu à un estimateur consistant $\widehat{R}_{\gamma,T} = [[\widehat{R}_T]_{i,j} \mathbb{1}_{|i-j| \leq \gamma}]$ pour une fonction $\gamma(T)$ bien choisie satisfaisant $\gamma(T) \to \infty$ et $\gamma(T)/T \to 0$. Toutefois, elles présentent les limitations suivantes :

(i) Elles supposent la connaissance *a priori* du taux de décroissance des r_k (en restreignant ces taux à des classes spécifiques) ;

(i) Les résultats sont asymptotiques par nature et ne donnent pas de règles explicites pour choisir le paramètre $\gamma(T)$ pour des valeurs finies de N et de T ;

(i) Les opérations de « fenêtrage » ne garantissent pas la positivité de l'estimateur de covariance résultant.

Nous proposons de considérer des estimateurs de R_T sans fenêtrage. La seule hypothèse sur les r_k est leur sommabilité $\sum_{k=-\infty}^{\infty} |r_k| < \infty$.

On définit la fonction de covariance pour $\lambda \in [0, 2\pi)$

$$\boldsymbol{\Upsilon}(\lambda) \triangleq \sum_{k=-\infty}^{\infty} r_k e^{-\imath k \lambda}. \tag{7}$$

Nous supposons que les coefficients r_k sont absolument sommables et que $r_0 \neq 0$. D'où, $\boldsymbol{\Upsilon}(\lambda)$ est continue sur l'intervalle $[0, 2\pi]$. Comme $\|R_T\| \leq$

$\|\boldsymbol{\Upsilon}\|_\infty$ (voir [30, Lemma 4.1]), la sommabilité absolue des r_k implique que $\sup_T \|R_T\| < \infty$. Nous rappelons le régime asymptotique noté « $T \to \infty$ » pour lequel $N/T \to c > 0$ quand $T \to \infty$. Les résultats principaux sont donnés par les théorèmes 1 et 2 sous forme d'inégalités de concentration sur $\|\widehat{R}_T^b - R_T\|$ et $\|\widehat{R}_T^u - R_T\|$ et sont présentés par :

Theorem 1 (Estimateur biaisé)**.** *Soit l'estimée biaisée de* R_T

$$\widehat{R}_T^b = \mathcal{T}(\hat{r}^b_{-(T-1),T}, \ldots, \hat{r}^b_{(T-1),T})$$

où $\mathcal{T}(\hat{r}^b_{-(T-1),T}, \ldots, \hat{r}^b_{(T-1),T})$ *est la matrice Toeplitz formée des coefficients* $\hat{r}^b_{-(T-1),T}, \ldots, \hat{r}^b_{(T-1),T}$*. Alors, pour tout* $x > 0$*, on a :*

$$\mathbb{P}\left[\left\|\widehat{R}_T^b - R_T\right\| > x\right] \leq \exp\left(-cT\left(\frac{x}{\|\boldsymbol{\Upsilon}\|_\infty} - \log\Big(1 + \frac{x}{\|\boldsymbol{\Upsilon}\|_\infty}\Big) + o(1)\right)\right).$$

Theorem 2 (Estimateur non biaisé)**.** *Soit l'estimée non biaisée de* R_T

$$\widehat{R}_T^u = \mathcal{T}(\hat{r}^u_{-(T-1),T}, \ldots, \hat{r}^u_{(T-1),T})$$

où $\mathcal{T}(\hat{r}^u_{-(T-1),T}, \ldots, \hat{r}^u_{(T-1),T})$ *est la matrice de Toeplitz formée des coefficients* $\hat{r}^u_{-(T-1),T}, \ldots, \hat{r}^u_{(T-1),T}$*. Alors, pour tout* $x > 0$*, on a :*

$$\mathbb{P}\left[\left\|\widehat{R}_T^u - R_T\right\| > x\right] \leq \exp\left(-\frac{cT}{\log T}\frac{x^2}{4\|\boldsymbol{\Upsilon}\|_\infty^2}(1 + o(1))\right).$$

D'après ces résultats, l'erreur en norme spectrale est bornée par une fonction qui décroit exponentiellement quand T converge vers infini. La conséquence directe de ces théorèmes est que $\|\widehat{R}_T^b - R_T\| \to 0$ et $\|\widehat{R}_T^u - R_T\| \to 0$ presque surement $T \to \infty$. Un taux de décroissance plus petit $T/\log(T)$ pour l'exposant de l'estimateur non biaisé peut être interprété par une imprécision plus accrue des estimateurs de r_k pour les valeurs de k proches de $T-1$. La figure 4 trace les courbes de $\mathbb{P}[\|\widehat{R}_T - R_T\| > x]$ (les courbes notées biaisé et non biaisé), avec $\widehat{R}_T \in \{\widehat{R}_T^b, \widehat{R}_T^u\}$, $T = 2N$, $x = 2$. Ces courbes sont comparées aux bornes exponentielles théoriques données par les théorèmes 1 et 2 (les courbes notées biaisé théorie et non biaisé théorie). Nous observons que les taux donnés par les théorèmes sont asymptotiquement proches des taux optimaux.

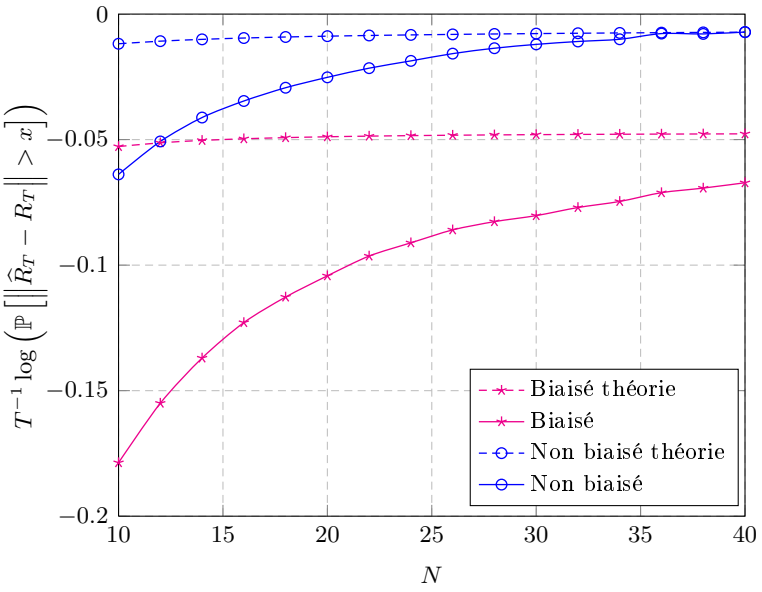

Figure 4: Probabilité d'erreur en norme spectrale pour $x = 2$, $c_T = 0.5$, $[R_T]_{k,l} = a^{|k-l|}$ avec $a = 0.6$.

3.2 Résultats pour le modèle perturbé

Considérons maintenant le modèle signal-plus-bruit :

$$Y_T = [y_{n,t}]_{\substack{0\leq n\leq N-1\\ 0\leq t\leq T-1}} = P_T + V_T \tag{8}$$

où la matrice $V_T = W_T R_T^{1/2} \in \mathbb{C}^{N\times T}$ est la matrice de bruit définie comme précédemment et $P_T \triangleq \boldsymbol{h}_T \boldsymbol{s}_T^{\mathsf{H}} \Gamma_T^{1/2}$ avec $\boldsymbol{h}_T \in \mathbb{C}^N$ un vecteur déterministe tel que $\sup_T \|\boldsymbol{h}_T\| < \infty$, $\boldsymbol{s}_T = (s_0, \ldots, s_{T-1})^{\mathsf{T}} \in \mathbb{C}^T$ un vecteur aléatoire indépendant de W_T avec la distribution $\mathcal{CN}(0, I_T)$ et $\Gamma_T = [\gamma_{ij}]_{i,j=0}^{T-1}$ une matrice hermitienne semi-définie positive telle que $\sup_T \|\Gamma_T\| < \infty$. Nous avons ici un modèle de bruit gaussien blanc spatialement et corrélé temporellement perturbé par un signal de rang un. Le signal est aussi corrélé temporellement. Notre but est toujours d'estimer la matrice de covariance de bruit R_T. Pour cela, les observations du bruit pur $v_{n,t}$ sont à présent remplacées par les échantillons $y_{n,t}$. Il se trouve que les estimées obtenues sont toujours consistantes en norme spectrale. Intuitivement, la matrice P_T ne casse pas la consistance de ces estimées car elle peut être vue comme une perturbation de rang un du terme de bruit V_T dans laquelle le sous espace

vectoriel engendré par $(\Gamma_T^{1/2})^{\mathsf{H}}\boldsymbol{s}_T$ est « délocalisé » assez pour ne pas trop perturber les estimateurs de R_T. Nous avons le résultat suivant :

Theorem 3 (Estimateurs pour le modèle « signal-plus-bruit »). *On considère les estimées*

$$\hat{r}_{k,T}^{bp} = \frac{1}{NT}\sum_{n=0}^{N-1}\sum_{t=0}^{T-1} y_{n,t+k}y_{n,t}^{*}\mathbb{1}_{0\leq t+k\leq T-1}$$

et

$$\hat{r}_{k,T}^{up} = \frac{1}{N(T-|k|)}\sum_{n=0}^{N-1}\sum_{t=0}^{T-1} y_{n,t+k}y_{n,t}^{\mathsf{H}}\mathbb{1}_{0\leq t+k\leq T-1}.$$

Soit $\widehat{R}_T^{bp} = \mathcal{T}(\hat{r}_{-(T-1),T}^{bp},\ldots,\hat{r}_{(T-1),T}^{bp})$ *et* $\widehat{R}_T^{up} = \mathcal{T}(\hat{r}_{-(T-1),T}^{up},\ldots,\hat{r}_{(T-1),T}^{up})$. *Alors pour tout* $x>0$, *on a :*

$$\mathbb{P}\left[\left\|\widehat{R}_T^{bp}-R_T\right\|>x\right]\leq \exp\left(-cT\left(\frac{x}{\|\boldsymbol{\Upsilon}\|_\infty}-\log\left(1+\frac{x}{\|\boldsymbol{\Upsilon}\|_\infty}\right)+o(1)\right)\right)$$

et

$$\mathbb{P}\left[\left\|\widehat{R}_T^{up}-R_T\right\|>x\right]\leq \exp\Big(-\frac{T}{\log T}\frac{cx^2}{4\,\|\boldsymbol{\Upsilon}\|_\infty^2}(1+o(1))\Big).$$

Notons que nous avons les mêmes taux de convergence que précédemment.

3.3 Application à la détection

Après avoir estimé la matrice de covariance du bruit, l'estimée est utilisée pour « blanchir » le modèle :

$$Y_T\widehat{R}_T^{-1/2} = \begin{cases} W_T R_T^{1/2}\widehat{R}_T^{-1/2}, & H_0 \\ \boldsymbol{h}_T\boldsymbol{s}_T^{\mathsf{H}}\widehat{R}_T^{-1/2} + W_T R_T^{1/2}\widehat{R}_T^{-1/2}, & H_1. \end{cases} \tag{9}$$

Comme la quantité $\|R_T\widehat{R}_T^{-1}-I_T\|\to 0$ presque surement (par le théorème 3 quand $\inf_{\lambda\in[0,2\pi)}\boldsymbol{\Upsilon}(\lambda)>0$), pour T grand, la décision sur l'hypothèse (9) peut être résolue par le GLRT [14] en approximant $W_T R_T^{1/2}\widehat{R}_T^{-1/2}$ par le bruit blanc. Nous avons le résultat suivant :

Theorem 4 (Détection de source). *Considérons* $\widehat{R}_T$ *correspondant soit à* $\widehat{R}_T^{bp}$ *soit à* $\widehat{R}_T^{up}$ *strictement défini dans le théorème 3 pour* Y_T *du modèle* (8).

Supposons que $\inf_{\lambda\in[0,2\pi)} \boldsymbol{\Upsilon}(\lambda) > 0$ *et que* $\liminf_T \ \|\boldsymbol{h}_T\|^2 \operatorname{Tr}\left(R_T^{-1}\right)/T \geq \sqrt{c}$. *Le test est défini par :*

$$\alpha = \frac{N\left\|Y_T \widehat{R}_T^{-1} Y_T^{\mathsf{H}}\right\|}{\operatorname{Tr}\left(Y_T \widehat{R}_T^{-1} Y_T^{\mathsf{H}}\right)} \underset{H_1}{\overset{H_0}{\lessgtr}} \gamma \tag{10}$$

où $\gamma \in \mathbb{R}^+$ *satisfait* $\gamma > (1+\sqrt{c})^2$. *Alors, quand* $T \to \infty$,

$$\mathbb{P}\left[\alpha \geq \gamma\right] \to \begin{cases} 0, & H_0 \\ 1, & H_1. \end{cases}$$

Rappelons que le seuil de détection $(1+\sqrt{c})^2$ correspond à la limite presque sure de la plus grande valeur propre de $\frac{1}{T} W_T W_T^{\mathsf{H}}$ qui est en fait le bord droit du support de la loi de Marčenko–Pastur.

3.4 Résultats numériques

Nous effectuons des simulations pour montrer les performances du test (5.24). Le vecteur de canal est un vecteur directionnel $h_T = \sqrt{p/T}[1, \ldots, e^{2i\pi\theta(T-1)}]$ avec $\theta = 10°$ l'angle d'arrivée et p le paramètre de puissance. La matrice R_T modélise un bruit autorégressif d'ordre 1 et de paramètre a tel que $[R_T]_{k,l} = a^{|k-l|}$.

La figure 5 trace les erreurs de détection $1 - \mathbb{P}[\alpha \geq \gamma | H_1]$ du test (5.24) pour la PFA égale à $\mathbb{P}[\alpha \geq \gamma | H_0] = 0.05$ pour $\widehat{R}_T = \widehat{R}_T^{up}$ (non biaisé) ou $\widehat{R}_T = \widehat{R}_T^{bp}$ (biaisé) et les compare aux erreurs de l'estimateur dit oracle pour lequel R_T est supposée parfaitement connue (oracle), *i.e.*, on pose $\widehat{R}_T = R_T$ dans (5.24). On compare les résultats également au GLRT qui suppose faussement que le bruit est blanc (blanc), *i.e.*, en supposant $\widehat{R}_T = I_T$ dans l'expression (5.24). La puissance de la source est fixée à $p = 1$, pour laquelle le ratio du signal au bruit (RSB) est égal à 0 dB, N varie de 10 à 50 et $T = N/c_T$ pour $c_T = 0.5$. Comme plus haut, le nombre de capteurs est fixé à $N = 20$, $T = N/c_T = 40$ et le RSB (d'où p) varie de -10 dB à 4 dB.

Les puissances des différents tests sont présentées dans la figure 6 et comparées aux méthodes de détection qui estiment R_T à partir d'une séquence de bruit pur appelé biaisé BP (bruit pur) et non biaisé BP. Les résultats de la méthode proposée sont proches de ceux de biaisé/non biaisé BP, le dernier présentant le désavantage de supposer qu'une séquence de bruit pur est disponible au récepteur qui est une hypothèse très forte.

Les deux figures suggèrent que les deux méthodes proposés biaisé et l'oracle ont des performances très proches, alors que le non biaisé présente des

performances moins bonnes. Le gap entre le biaisé et le non biaisé confirme bien les prévisions théoriques.

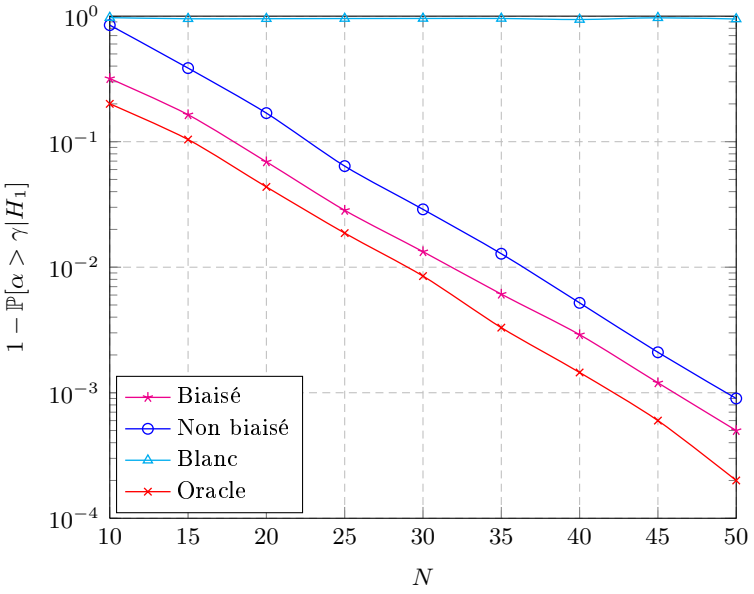

Figure 5: Erreurs de détection en fonction de N avec PFA= 0.05, $p = 1$, RSB= 0 dB, $c = 0.5$ et $a = 0.6$.

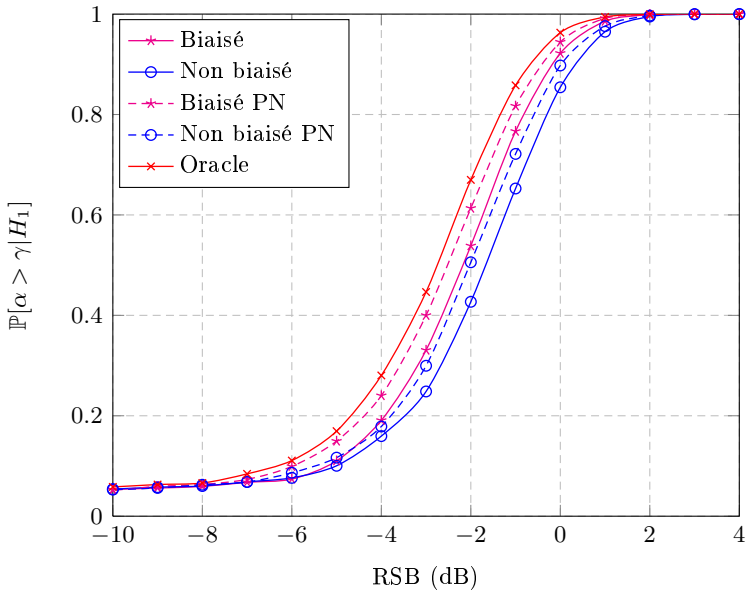

Figure 6: Puissances de test de détection en fonction du rapport du signal au bruit (dB) avec PFA= 0.05, $N = 20$, $c = 0.5$ et $a = 0.6$.

4 Conclusion

Dans cette thèse nous avons traité deux difficultés souvent rencontrées dans les réseaux de capteurs modernes : (i) le nombre de capteurs et la taille de la fenêtre d'observation sont grands tous les deux et de même ordre de grandeur et (ii) les capteurs sont situés dans un environnement de bruit corrélé avec une matrice de covariance inconnue. Compte tenu l'hypothèse sur la matrice de covariance de bruit deux approches ont été élaborées. La première a consisté à estimer les paramètres du système sans hypothèse quelconque sur les statistiques des échantillons de bruit. La seconde approche a supposé que le bruit est un processus stationnaire et a proposé des estimateurs consistants pour la matrice de covariance de bruit conduisant à la procédure de « blanchiment ». La première approche ne peut être appliquée dans les scénarios où le bruit et le signal sont simultanément temporellement ou spatialement corrélés. En faisant une hypothèse sur la structure de la matrice de covariance de bruit nous permet d'éviter cette restriction. En parallèle avec la structure de cette thèse deux directions de recherche pourraient être envisagées.

4.1 Matrices non structurées

Concernant le comportement des valeurs propres isolées de grandes matrices aléatoires et les algorithmes de détection et d'estimation qui leur sont associés, des modèles de matrices plus sophistiqués que ceux considérés dans cette thèse pourraient être exploités. Ce qu'on appelle le modèle bi-corrélé pourrait être étudié, pour lequel la matrice de bruit présente des corrélations spatiales et temporelles en même temps. La matrice de bruit pour ce modèle s'écrit $V_T = \widetilde{R}_T^{1/2} W_T R_T^{1/2}$ où W_T a des entrées i.i.d. et $\widetilde{R}_T$ et R_T sont des matrices de covariance. De point de vue d'applications, ce modèle présente un intérêt particulier dans les systèmes de radar où les corrélations temporelles et spatiales surviennent simultanément.

L'analyse en composantes parcimonieuses est devenu récemment un sujet de grand intérêt puisqu'il trouve des applications dans de nombreux domaines d'analyse multivariée et de traitement de signal. Une perspective intéressante consisterait à exploiter le lien entre les modèles de spikes et les méthodes d'acquisition comprimée dans le cas où les vecteurs propres associés aux spikes ont une structure creuse. Dans ce cadre, la détection en composantes parcimonieuses a été par exemple explorée dans [12].

4.2 Matrices structurées

Dans cette thèse, la matrice de covariance d'un processus stationnaire a été estimée sans avoir recours à un fenêtrage. L'utilisation d'un fenêtrage optimal pourrait donner lieu à un estimateur plus efficace. Toutefois, la connaissance *a priori* du taux de décroissance des coefficients de Toeplitz est requise afin de construire un fenêtrage optimal. Une des directions à prendre pour résoudre ce problème serait la méthode itérative.

Le contexte d'application des résultats de la thèse concerne principalement les radars où les bruits à queue lourde sont souvent rencontrés. Dans le cadre du travail de [23], nous supposons que la matrice de covariance de bruit est de Toeplitz et nous nous proposons de l'estimer. L'estimation se fait en deux étapes. Tout d'abord nous procédons à l'estimation robuste de la matrice de covariance en nous basant sur l'algorithme appelé M-estimation de Maronna. Puis, une régularisation est appliquée en toeplizifiant la matrice estimée. L'objectif est de démontrer la consistance de la matrice estimée que le signal soit présent ou non. Comme exemple d'application, on devrait proposer un test de détection dans le contexte des grandes dimensions avec le bruit corrélé spatialement.

Chapter 1

Introduction

1.1 Model and problem statement

1.1.1 System model

Consider a linear array composed of up to K possibly emitting sources and N sensors embedded in an additive noise. Consider an observation window of size T. The received signal vector $y_t \in \mathbb{C}^{N\times 1}$ at time interval $t = 0, \ldots, T-1$ is given by

$$y_t = Hs_t + w_t$$

where $s_t \in \mathbb{C}^{K\times 1}$ is the vector of random transmitted signal symbols of covariance matrix Γ, $H \in \mathbb{C}^{N\times K}$ is a deterministic channel matrix, and $w_t \in \mathbb{C}^{N\times 1}$ is the noise vector with i.i.d complex Gaussian entries of variance σ^2. It will be assumed in the following that all the parameters of the system are unknown at the receiver side, excepted the number of the sensors N and the size of the observation window T. A fundamental problem in array processing consists in developing algorithms to infer on the signal part Hs_t using the sample covariance matrix only. In particular, the purpose is to detect the number of the emitting signals and to estimate some of their parameters, as for instance, their powers and the directions-of-arrival. The methods proposed in the literature relay on the structure of the received covariance matrix given by

$$\Sigma = H\Gamma H^{\mathsf{H}} + \sigma^2 I_N$$

where $H\Gamma H^{\mathsf{H}}$ corresponds to the signal part and $\sigma^2 I_N$ is the noise covariance matrix. In practice we do not access to Σ which is estimated by the sample covariance matrix

$$\widehat{\Sigma} = \frac{1}{T}\sum_{t=0}^{T-1} y_t y_t^{\mathsf{H}}.$$

It is known that for the classical regime where N is fixed and $T \to \infty$ from the law of large numbers the sample covariance matrix is a consistent estimator of the population covariance matrix. Therefore, many existing detection methods are based on the use of the sample covariance matrix, more particularly on its spectrum. They exploit specifically the fact that as $T \to \infty$, the smallest eigenvalues are close to each other and converge to σ^2. Among classical detection approaches, we find the famous Akaike Information Criterion (AIC) and the Minimum Description Length (MDL) (see the method of Wax and Kailath in [67]).

1.1.2 Large dimensional regime

Modern communication systems usually deal with large dimensional vectors of observations. Usually the system dynamics change very fast, the size of the observation window being of the same order of magnitude as that the system dimension. In this case it is relevant to assume that N and T converge to infinity at the same speed. In this setting as N and T are large, such that $N/T \to c > 0$ as $T \to \infty$, the sample covariance matrix is not a consistent estimator of the true covariance matrix. Therefore, the classical detection and estimation methods fail. In this regime, large dimensional random matrix theory tools allow to construct (N, T)-consistent estimators for the system parameters. Stacking all the received signal vectors into the $N \times T$ matrix, the transmission model can be written in the matrix form

$$Y_T = A_T + W_T \tag{1.1}$$

where $A_T = H_T S_T^{\mathsf{H}}$ is the signal matrix with $S_T = [s_0, \ldots, s_{T-1}]^{\mathsf{H}} \in \mathbb{C}^{T\times K}$, and $W_T = [w_0, \ldots, w_{T-1}] \in \mathbb{C}^{N\times T}$ is the noise matrix. The signal matrix A_T is assumed to be of small rank K as $T \to \infty$. The full rank matrix Y_T can be viewed as a perturbed version of the noise matrix W_T, the additive perturbation A_T having a small rank. In large dimensional random matrix theory the model (1.1) belongs to the class of spiked models [35], [9]. The proposed detectors/estimators are based on the study of the limiting behavior of the largest eigenvalues of the sample covariance matrix $\widehat{\Sigma}_T = \frac{1}{T} Y_T Y_T^{\mathsf{H}}$.

1.1.3 White noise setting

The hypothesis testing approaches test the null hypothesis H_0 (the signal is absent) against the hypothesis H_1 (the signal is present). Under H_0, the sample covariance matrix is written as $\widehat{\Sigma}'_T = \frac{1}{T} W_T W_T^{\mathsf{H}}$. As N and T converge to infinity, W_T being of i.i.d. entries, the distribution of eigenvalues of $\widehat{\Sigma}'_T$ converges to the celebrated Marčenko–Pastur law [44]. The sample covariance matrix eigenvalues form actually a so-called "bulk" corresponding to the noise eigenvalues which are all asymptotically located in the interval $[a, b]$, the support of the Marčenko–Pastur distribution. Under H_1, the spectrum of the sample covariance is composed of the main "bulk" and may present some isolated eigenvalues called outliers due to the signal presence. The condition of apparition of these outliers is related to the power of the signal and the parameter c.

Based on the spiked models theory there exist many detection methods studied during these last years. One of the first contributions belongs to Nadakuditi and Edelman in [46] who improved the AIC and MDL-based algorithms in the large dimensional context. For the single source case, a suboptimal statistical test was derived by Bianchi *et al.* in [14] and is based on the ratio of the largest eigenvalue of $\widehat{\Sigma}_T$ to its trace. In the literature, it is referred to as the Generalized Likelihood-Ratio Test (GLRT). For a multiple source case, a multiple hypothesis testing approach was provided by Kritchman and Nadler in [40]. However, the hypothesis of white noise environment does not hold in many practical situations.

1.1.4 Correlated noise

The successive observations of the noise samples may not be independent from one observation to another corresponding to the temporally correlated noise. Alternatively, due to the closeness of the sensors, the noise vector may present spatial correlations. In the case of a temporally correlated scenario, the noise matrix is of the form $V_T = W_T R_T^{1/2}$ where W_T is an $N \times T$ random matrix with i.i.d. entries and R_T is an unknown covariance matrix capturing the temporal correlations of the noise samples received by one antenna.

When R_T is known, the sample covariance matrix can be whitened by the inverse of R_T, and we fall back to the classical situation of a signal corrupted by a white noise. In this case, when a single source is possibly present and the noise is Gaussian, the GLRT procedure of [14] is applied in order to perform source detection.

When R_T is unknown, one generally assumes the existence of an inde-

pendent sequence of T' pure-noise samples. From this independent sequence an empirical noise covariance matrix is constructed which is supposed to represent the true noise covariance matrix. The empirical noise covariance matrix is then used to "whiten" the observations from which one wants to perform detection. The covariance matrix under testing which is "whitened" by the pure-noise empirical covariance matrix gives rise to the so-called F-matrix [58], perturbed by an hypothetical signal. This approach was studied by Nadakuditi and Silverstein in [47].

The existence of a pure noise sequence is a strong hypothesis that we propose to avoid. In this thesis we consider a sequence which potentially contains the signal.

1.2 Contributions of this thesis

Our studies lead two different axes:

- **Axis 1**: we propose a detection algorithm of the number of sources based on the spacings between the successive eigenvalues of the empirical covariance matrix. This algorithm does not make any assumption on the statistics of the noise samples. We proposed in addition a source power estimation method and a direction-of-arrival estimation algorithm based on the adaptation of the well-known MUSIC approach to the context of random matrices.

- **Axis 2**: we assume that R_T is structured as being the covariance matrix of a stationary noise process. Within this context, we revisit the known approaches for estimation of such matrices based on a Toeplitz-ified version of the sample covariance matrix. After having estimated this matrix, "whitening" is performed which leads back to the GLRT. The idea is that even if the signal is present, considered as a nuisance at this stage, the estimator of the covariance matrix remains consistent. This is due to the small rank of the signal. The originality of our work consists in establishing concentrations inequalities on the spectral norm of the noise covariance matrix, whether or not the signal is present.

1.3 Outline

The remainder of this thesis is composed of four chapters.

Chapter 2 introduces first the basic tools of random matrix theory and presents some important results on asymptotic spectrum analysis of large dimensional random matrices which will be necessary in the following chapters. It provides also the important theoretical results on the advanced spiked models which are used in the remaining chapters.

Chapter 3 overviews the existing signal detection and localization methods. Detection/localization algorithms are presented for the white and correlated noise models.

Chapter 4 assumes that the noise covariance matrix structure is not known. A source detector, power and direction-of-arrival estimators are provided for a generic model first, then for a specific narrow processing example. Second order statistics of some estimators are also studied.

Chapter 5 assumes that R_T is the covariance matrix of a stationary Gaussian process having a Toeplitz structure. Concentration inequalities on the spectral norm are derived in a signal free case and under its presence. A detection test on the "whitened" sample covariance matrix is also proposed.

1.4 Publications

The following publications are associated with this thesis:

1.4.1 Journal papers

- "Estimation of Toeplitz covariance matrices in large dimensional regime with application to source detection," J. Vinogradova, R. Couillet, W. Hachem, submitted to IEEE Transactions on Signal Processing, March 2014, revised on June 2014.

- "Statistical inference in large antenna arrays under unknown noise pattern," J. Vinogradova, R. Couillet, W. Hachem, IEEE Transactions on Signal Processing, 61 (22), 2013, pages 5633-5645.

1.4.2 Conference papers

- "Estimation of large Toeplitz covariance matrices and application to source detection," J. Vinogradova, R. Couillet, W. Hachem, EUSIPCO'2014, Lisbonne, Portugal.

- "Nouvelle méthode de détection de sources, d'estimation de puissances et de localisation dans un système de communication sans fil avec des statistiques de bruit inconnues," J. Vinogradova, R. Couillet, W. Hachem, GRETSI'2013, Brest, France.
- "A new method for source detection, power estimation, and localization in large sensor networks under noise with unknown statistics," J. Vinogradova, R. Couillet, W. Hachem, ICASSP'2013, Vancouver, Canada.

Chapter 2

Some results of random matrix theory

This chapter provides a theoretical background useful in the remaining chapters. After introducing the necessary tools from the random matrix theory, we overview the classical results in asymptotic spectral behavior of large random matrices, in particular for some covariance matrix models. They include the famous Marčenko–Pastur law as well as a characterization of the limiting spectral measure by using Stieltjes transform for some advanced models. Further, the limiting behavior of the extreme eigenvalues is provided. The models where the sample covariance matrix is perturbed by a small rank matrix are related to the so-called spiked models. Under these models, the study of the limiting behavior of the sample covariance matrix isolated eigenvalues is of main importance in this thesis.

2.1 Basic tools

We consider sequences of random matrices $X_1, X_2, \ldots$ with $X_T \in \mathbb{C}^{T\times T}$. The limiting spectral behavior of Hermitian random matrices will be studied. We give first the definition of the spectral measure:

Definition 1. *The spectral measure μ_T of the Hermitian matrix $X_T \in \mathbb{C}^{T\times T}$ is defined by*

$$\mu_T = \frac{1}{T}\sum_{t=0}^{T-1} \boldsymbol{\delta}_{\lambda_t}$$

where $\lambda_0, \ldots, \lambda_{T-1}$ are the eigenvalues of X_T.

Definition 2. *The sequence of random measures μ_T is said to converge weakly to a deterministic probability measure μ in the almost sure sense if for every bounded and continuous real function f, we have*

$$\int f(t)\mu_T(dt) \xrightarrow[T\to\infty]{a.s.} \int f(t)\mu(dt)$$

where a.s. stands for the almost sure convergence.

We introduce now an important tool called Stieltjes transform and give some of its useful properties.

Definition 3. *Let μ be a probability measure defined on $\mathbb{R}$. The Stieltjes transform $m(z)$ of μ, for $z \in \mathbb{C} - \mathrm{supp}(\mu)$ where $\mathrm{supp}(\mu)$ is the support of μ, is defined by*

$$m(z) \triangleq \int_{\mathbb{R}} \frac{1}{t-z}\mu(dt).$$

Any measure μ is uniquely defined by its Stieltjes transform from the following property:

Property 1. *For any continuous real function φ with compact support in $\mathbb{R}$,*

$$\int \varphi(t)\mu(dt) = \frac{1}{\pi}\lim_{y\to 0}\int \varphi(x)\Im(m(x+iy))dx.$$

Equivalently, for any a and b continuity points of μ, we have

$$\mu([a,b]) = \frac{1}{\pi}\lim_{y\to 0^+}\int_a^b \Im[m(x+iy)]dx.$$

The Stieltjes transform presents the following properties:

Property 2. *Let $m(z)$ be the Stieltjes transform of a probability measure defined on $\mathbb{R}$. Then*

- *$m(z)$ is analytic on $\mathbb{C} - \mathrm{supp}(\mu)$*
- *$z \in \mathbb{C}^+$ implies that $m(z) \in \mathbb{C}^+$*
- *$\sup_{y>0} |ym(iy)| = 1$.*

The Stieltjes transform of the spectral measure μ_T of the Hermitian matrix $X_T \in \mathbb{C}^{T\times T}$ is

$$m(z) = \int_{\mathbb{R}} \frac{1}{t-z}\mu_T(dt) = \frac{1}{T}\mathrm{tr}(X_T - zI_T)^{-1}.$$

2.2 Asymptotic spectrum analysis

2.2.1 Basic results on asymptotic spectrum

The first result on limiting spectral measure of large dimensional random matrices was given by Wigner [69], [68]. He considered a $T \times T$ symmetric matrix W_T whose diagonal entries are equal to 0 and whose upper-triangle entries $[W_T]_{i,j}$ are independent and take the values ± 1 with equal probability. In [69] it was shown that, as $T \to \infty$, the spectral measure of the eigenvalues of $T^{-1/2}W_T$ converges to the semi-circle law whose density f is defined as

$$f(x) = \frac{1}{2\pi}\sqrt{(4-x^2)^+}.$$

Later, in [68] it was shown that the same result holds if the elements of W_T are drawn from a zero-mean (real or complex) Gaussian distribution. A generalization of this result was provided by Bai and Silverstein [5] who considered a $T \times T$ Hermitian matrix W_T with independent zero-mean entries $[W_T]_{i,j}$ of unit variance and finite moment of order $2+\epsilon$, for $\epsilon > 0$. We are interested into the study of the matrix models of the type $\frac{1}{T}W_T R_T W_T^{\mathsf{H}}$ where W_T has i.i.d. entries and R_T is a deterministic covariance matrix. In the particular case where $R_T = I_T$, the asymptotic spectral measure is given by the Marčenko–Pastur law [44], provided by the following theorem:

Theorem 5 ([44], [59]). *Consider a matrix $W_T \in \mathbb{C}^{N\times T}$ with zero mean i.i.d. entries of unit variance. As T, $N \to \infty$ with $N/T \to c > 0$, the empirical measure of $\frac{1}{T}W_T W_T^{\mathsf{H}}$ converges weakly and almost surely to a nonrandom measure μ_c with density f_c given by*

$$f_c(x) = (1-c^{-1})^+\delta(x) + \frac{1}{2\pi c x}\sqrt{(x-a)^+(b-x)^+}$$

where $a = (1-\sqrt{c})^2$, $b = (1+\sqrt{c})^2$, and $\delta(x) = 1_{\{0\}}(x)$.

The histogram of the eigenvalues of $\frac{1}{T}W_T W_T^{\mathsf{H}}$ and their limit law are depicted in Figure 2.1.

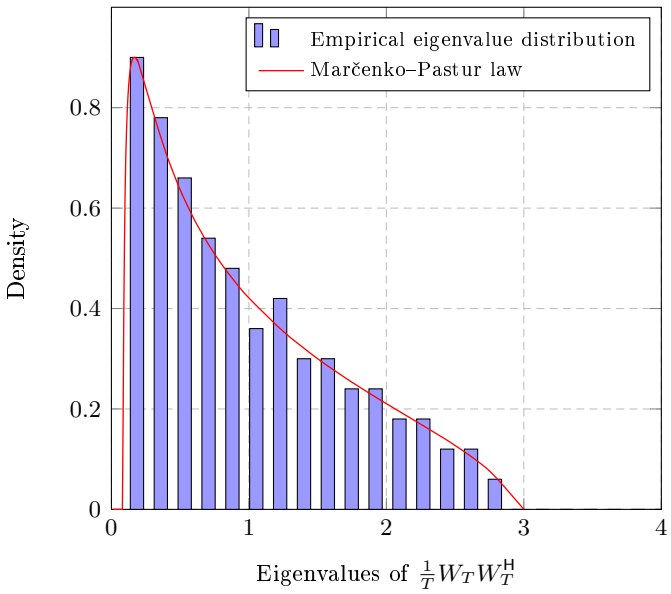

Figure 2.1: Histogram of the empirical eigenvalues of $\frac{1}{T}W_T W_T^{\mathsf{H}}$ and the Marčenko–Pastur law for $N = 100$, $T = 200$, $c = 0.5$.

As the population covariance matrix associated with the random column vectors of W_T is the identity matrix, its spectrum is a Dirac mass at one. Observe that, when N is fixed while T converges to infinity, by a simple application of the law of large numbers we can show that the spectral measure of $\frac{1}{T}W_T W_T^{\mathsf{H}}$ converges to a Dirac mass at one. In our regime where N and T converge to infinity at the same rate, Theorem 2.2.1 shows that the support of μ is an interval around one which is small when c is small as one can expect intuitively (see Figure 2.2).

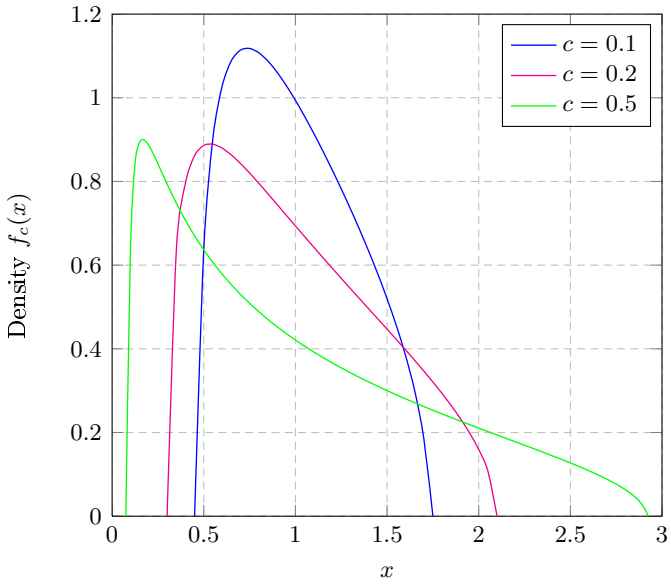

Figure 2.2: Marčenko–Pastur law for different limiting ratios c.

2.2.2 Further results on asymptotic spectrum

Many applications use the matrix models for which either the columns or the rows of the matrix W_T are not independent. Specifically throughout the manuscript we will consider the matrix under study is written as a product of two matrices $V_T = W_T R_T^{1/2}$ where W_T has i.i.d. entries of mean zero and unit variance and R_T is a covariance matrix. Note that for this case, V_T has dependent columns but its rows are independent. In practical models, this model corresponds to the temporally correlated noise. We would like to characterize the limiting spectral behavior of the sample covariance matrix $\frac{1}{T} V_T V_T^{\mathsf{H}}$. The following result was derived by Marčenko and Pastur in [44] and generalized by Silverstein and Bai in [59]:

Theorem 6 ([44], [59]). *Let $\widehat{\Sigma}_T = \frac{1}{T} W_T R_T W_T^{\mathsf{H}}$, where $W_T \in C^{N \times T}$ has i.i.d. entries of mean zero and unit variance and R_T is a deterministic Hermitian nonnegative matrix whose spectral measure ν_T converges weakly to ν as $T \to \infty$. Assume that as $T \to \infty$, $N/T \to c > 0$. Then, the spectral measure μ_T of $\widehat{\Sigma}_T$ converges weakly and almost surely to μ whose Stieltjes transform $m(z)$, $z \in \mathbb{C}^+$, is given by the unique solution in $\mathbb{C}^+$ of*

the equation

$$m(z) = \left(z - \int \frac{t}{1 + cm(z)t} \nu(dt) \right)^{-1} \tag{2.1}$$

where $\mathbb{C}^+ = \{z \in \mathbb{C} : \Im z > 0\}$.

These results concern the limiting behavior of the spectrum. However, from this we cannot say anything about the asymptotic behavior of the sample covariance matrix extreme eigenvalues (*i.e.*, the smallest and the largest eigenvalues). The following theorem provided by Bai and Silverstein in [4] gives the conditions under which there is no eigenvalue that can be found away from the limiting support. The following statement is slightly more restrictive than in its original form:

Theorem 7 ([4]). *Let* $\widehat{\Sigma}_T = \frac{1}{T} W_T R_T W_T^{\mathsf{H}} \in \mathbb{C}^{N \times N}$, *where* $W_T \in \mathbb{C}^{N \times T}$ *has i.i.d. entries with mean zero, unit variance, and finite fourth order moment. Let* $R_T \in \mathbb{C}^{T \times T}$ *be a deterministic Hermitian nonnegative matrix with uniformly bounded spectral norm* $\|R_T\|$ *whose spectral measure* ν_T *converges weakly to* ν *as* $T \to \infty$. *Let* $\sigma_{0,T}^2, \ldots, \sigma_{T-1,T}^2$ *be the eigenvalues of* R_T. *Assume that* $\max_{t \in \{0,\ldots,T-1\}} \mathbf{d}\left(\sigma_{t,T}^2, \operatorname{supp}(\nu)\right) \to 0$ *as* $T \to \infty$, *where* $\mathbf{d}(\sigma_{t,T}^2, \operatorname{supp}(\nu)\cdot)$ *is the distance from* $\sigma_{t,T}^2$ *to* $\operatorname{supp}(\nu)$. *Let* $\lambda_{0,T}, \ldots, \lambda_{N-1,T}$ *be the eigenvalues of* $\widehat{\Sigma}_T$ *with spectral measure denoted by* μ_T *converging weakly and almost surely to* μ. *Assume that as* $T \to \infty$, $N/T \to c > 0$. *For any interval* $[x_1, x_2] \subset \mathbb{R} - \operatorname{supp}(\mu)$,

$$\sharp\{i : \lambda_{i,T} \in [x_1, x_2]\} = 0 \text{ a.s. for all large } T.$$

2.2.3 Some background on the limiting support

Consider the model $\widehat{\Sigma}_T = \frac{1}{T} W_T R_T W_T^{\mathsf{H}}$ from Theorem 7. We are looking for a characterization of the limiting support of μ. A procedure for its determination from the knowledge of c and ν is provided by Silverstein and Choi in [60] and is presented hereafter.

As Equation (2.1) has a unique solution in $\mathbb{C}^+$, it admits an inverse expressed by

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

for $m \in \mathbb{C}^+$.

We present the following theorems [60]:

Theorem 8 ([60]). *Let $\widehat{\Sigma}_T = \frac{1}{T} W_T R_T W_T^{\mathsf{H}} \in \mathbb{C}^{N\times N}$ be defined as above with the limit measure μ. Then, for $x \in \mathbb{R}^*$*

$$\lim_{\substack{z\to x \\ z\in\mathbb{C}^+}} m(z) \triangleq m_0(x)$$

exists. The function m_0 is continuous on $\mathbb{R} - \{0\}$. Hence, F, the limit eigenvalue distribution function of $\widehat{\Sigma}_T$, has a continuous derivative f on $\mathbb{R} - \{0\}$ given by $f(x) = \frac{1}{\pi}\Im(m_0(x))$. The density f is analytic for every $x \in \mathbb{R}^*$ for which $f(x) > 0$.

By studying the function m_0 it is possible to characterize the complete support of μ. The following theorem provides its precise description.

Theorem 9 ([60]). *Let $E = \{m \text{ s.t. } m \neq 0, -1/m \in \mathrm{supp}(\nu)^c\}$, with $\mathrm{supp}(\nu)^c$ the complementary of $\mathrm{supp}(\nu)$, and let $x(m)$ be the function defined on E by*

$$x(m) = -\frac{1}{m} + \int \frac{t}{1+cmt}\nu(dt). \tag{2.2}$$

Then for $m \in E$, if $x'(m) > 0$, then $x(m) \in \mathrm{supp}(\mu)^c$. Conversely, for any $x \in \mathrm{supp}(\mu)^c$ there exists $m \in E$ such that $x = x(m)$ and $x'(m) > 0$.

The support of the limiting spectral measure of $\widehat{\Sigma}_T$ is determined by using these rules, in particular, the fact that the function $x(m)$ is increasing outside the support of μ. For practical application, we are rather interested in determining the right-edge b. The next proposition specifies this point

Proposition 6. *Assume that ν is compactly supported and let $b_\nu = \max(\mathrm{supp}(\nu))$. Then $b < \infty$ and it coincides with the infimum of $x(m)$ on the interval $(-1/b_\nu, 0)$. In particular if the infimum is achieved then $b = x(m_b)$ where m_b is the unique solution in $(-1/b_\nu, 0)$ of the equation $x'(m) = 0$.*

2.2.4 Fluctuations of the largest eigenvalue

Many detection and estimation methods in array processing rely on the largest eigenvalues of a random matrix. When we need to study the performance of such estimators, a study of fluctuations of these eigenvalues is required. The first results on the fluctuations of the largest eigenvalue of a Wigner Hermitian matrix with i.i.d. Gaussian entries above the diagonal were introduced by Tracy and Widom [62]. The fluctuations of the largest eigenvalue of a Wishart matrix were provided later by Johansson in [34]. We

recall that the $N \times N$ matrix $X_T X_T^{\mathsf{H}}$ is a central complex Wishart matrix with N degrees of freedom and covariance matrix R_T if the rows of the $N \times T$ matrix X_T are zero-mean independent real or complex Gaussian vectors with covariance matrix R_T. We have the following theorem:

Theorem 10 ([34]). *Let $W_T \in \mathbb{C}^{N \times T}$ have i.i.d. complex Gaussian entries of mean zero and unit variance. Denote by $\lambda_{0,T}$ the largest eigenvalue of $\frac{1}{T} W_T W_T^{\mathsf{H}}$. Assume that $c_T = N/T < 1$. Define:*

$$b_T = (1 + \sqrt{c_T})^2$$
$$\sigma_T = (1 + \sqrt{c_T})^{4/3} \sqrt{c_T}.$$

Then, as $T \to \infty$, for any real x in a compact set

$$\mathbb{P}\left(N^{\frac{2}{3}} \frac{\lambda_{0,T} - b_T}{\sigma_T} \geq x \right) \to F_{TW}(x)$$

with the Tracy–Widom law distribution function F_{TW} defined as

$$F_{TW}(x) = exp\left(- \int_t^\infty (x - t)^2 q^2(x) dx \right)$$

where q is the Painlevé II function that solves the differential equation

$$q''(x) = x q(x) + 2q^3(x)$$
$$q(x) \underset{x \to \infty}{\sim} Ai(x)$$

with $Ai(x)$ the Airy function defined by, for any $x \in \mathbb{R}$

$$Ai(x) = \frac{1}{2\pi} \int e^{ixt + \frac{it^3}{3}} dt.$$

In Figure 2.3 we see that the histogram of the properly centered and scaled largest eigenvalue of $\frac{1}{T} W_T W_T^{\mathsf{H}}$ converges to the Tracy–Widom law.

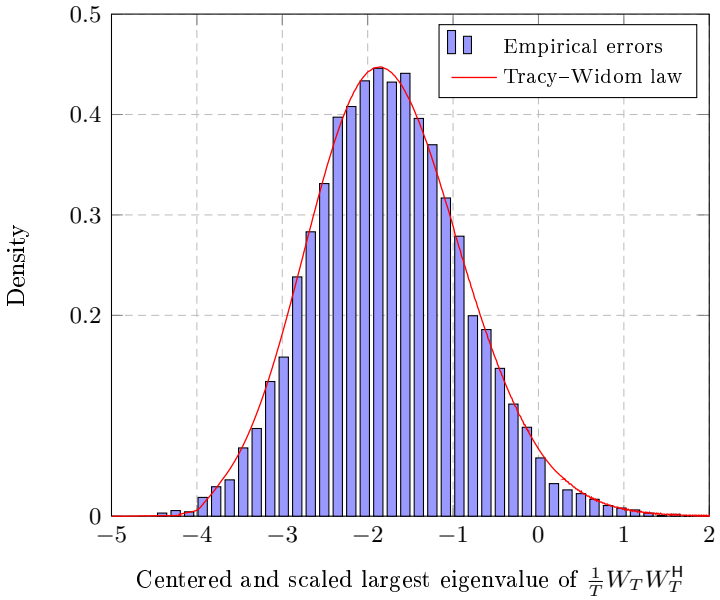

Figure 2.3: Histogram of $N^{\frac{2}{3}}\frac{\lambda_{0,T}-b_T}{\sigma_T}$ and the Tracy–Widom law for $N = 100$, $T = 300$.

A more general result, when the covariance matrix is not identity was proposed by El Karoui in [26]. Let V_T be an $N \times T$ matrix, and let its rows be i.i.d. complex normal vectors with mean zero and covariance R_T. The fluctuation behavior of the centered and scaled largest eigenvalue of a sample covariance matrix is given by the following theorem:

Theorem 11 ([26]). *Let $V_T = W_T R_T^{1/2}$ be a matrix product where $W_T \in \mathbb{C}^{N\times T}$ has i.i.d. complex Gaussian entries with zero mean and unit variance, and $R_T \in \mathbb{C}^{T\times T}$ is a deterministic nonnegative matrix with spectral measure ν_T converging to ν. Let $\sigma_{0,T}^2 \geq \ldots \geq \sigma_{T-1,T}^2$ be the eigenvalues of R_T. We assume that $c_T = N/T < 1$ is uniformly bounded. Let m_b be the unique solution in $(-1/\sigma_{0,T}^2, 0)$ to the equation in m*

$$\int \left(\frac{mt}{1+mt}\right)^2 \nu_T(dt) = \frac{1}{c_T}.$$

Assume that $\limsup\ \sigma_{0,T}^2 < \infty$, $\liminf\ \sigma_{T-1,T}^2 > 0$, *and* $\limsup\ \sigma_{0,T}^2 m_b < 1$.

We call

$$b_T = -\frac{1}{m_b} + \int \frac{t}{1+m_b t}\nu_T(dt)$$
$$\sigma_T = \frac{1}{m_b}\left(1 + c_T \int \left(\frac{t}{1+m_b t}\right)^3 \nu_T(dt)\right)^{1/3}.$$

Let $\lambda_{0,T}$ be the largest eigenvalue of $\frac{1}{T}V_T V_T^{\mathsf{H}}$. Then, as $T \to \infty$, for any real x in a compact set

$$\mathbb{P}\left(N^{2/3}\frac{\hat{\lambda}_{0,T} - b_T}{\sigma_T} \geq x\right) \to F_{TW}(x).$$

We will not provide a comprehensive interpretation of the crucial condition $\limsup\ \sigma_{0,T}^2 m_b < 1$. Let us simply say that it will be satisfied if ν "decreases" sufficiently fast at the left of b and if $\max_{t\in\{0,\ldots,T-1\}} \mathbf{d}\left(\sigma_{t,T}^2, \mathrm{supp}(\nu)\right) \to 0$ as $T \to \infty$. Actually, the condition $\limsup \sigma_{0,T}^2 m_b < 1$ will be satisfied in most cases of practical interest.

From this result we see that the error follows the Tracy–Widom law but the scaling and centering factors do not depend only on the limiting ratio c but on the limit spectral measure ν of the covariance matrix.

2.3 Spiked models

2.3.1 Background on spiked models

Spiked models refer to the models where the full rank random matrix $W_T \in \mathbb{C}^{N\times T}$ with i.i.d. entries of zero mean and unit variance is perturbed by a small rank matrix. The matrices of the type $\widehat{\Sigma}_T = T^{-1}W_T(I_T + P_T)W_T^{\mathsf{H}}$ with $P_T \in \mathbb{C}^{T\times T}$ of small rank K or $\widehat{\Sigma}_T = (T^{-1/2}W_T + P_T)(T^{-1/2}W_T + P_T)^{\mathsf{H}}$ with $P_T \in \mathbb{C}^{N\times T}$ of small rank K, correspond respectively to the multiplicative and additive spiked models. As $T \to \infty$ such that $N/T \to c > 0$, the spectrum of $\widehat{\Sigma}_T$ still converges to the Marčenko–Pastur law. However, under some conditions that will be specified later, some isolated eigenvalues called outliers may be found outside the Marčenko–Pastur law support. This is observed in Figure 2.4 where $\widehat{\Sigma}_T = (T^{-1/2}W_T + P_T)(T^{-1/2}W_T + P_T)^{\mathsf{H}}$ with W_T with i.i.d. complex Gaussian zero mean entries of variance one, and P_T is of rank 2. Two eigenvalues of $\widehat{\Sigma}_T$ are found outside the interval $[(1-\sqrt{c})^2, (1+\sqrt{c})^2]$. A fundamental task is to analyze the limiting behavior of these isolated eigenvalues, as N and T go to infinity. In many applications

we are interested in finding their limiting positions and in characterizing their fluctuations.

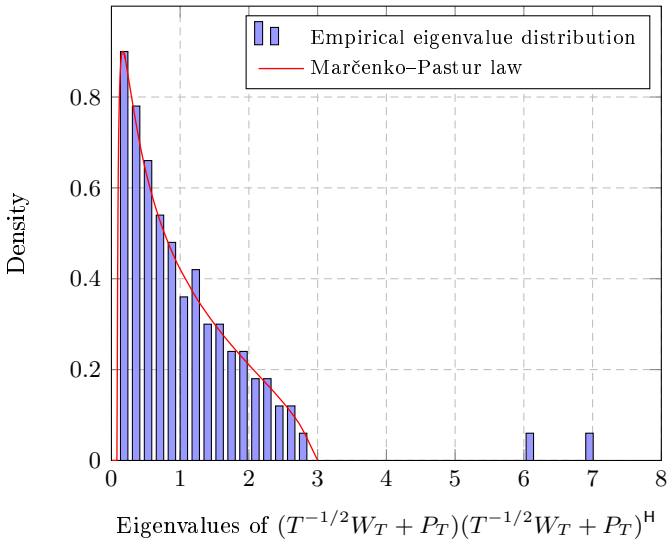

Figure 2.4: Histogram of the empirical eigenvalues of $\widehat{\Sigma}_T$ and the Marčenko–Pastur law for $N = 100$, $T = 200$, P_T of rank 2.

One of the first contributions on spiked belongs to the work of Johnstone [35]. He considered the population covariance matrix whose all the eigenvalues equal to one, except a fixed and relatively small number among them. Noticing that few eigenvalues of the sample covariance matrix can be separated from the bulk formed from the rest of eigenvalues, he raised the question under which condition on the population eigenvalues, isolated sample covariance matrix eigenvalues can be observed. A further contribution on this model was provided by Baik and Silverstein in [9] who gave the limiting position of the spikes for all N and T large. This is presented in the following theorem [9]

Theorem 12 ([9]). *Let $Y_T = (B_T + I_N)^{1/2} W_T$ where $W_T \in \mathbb{C}^{N \times T}$ has i.i.d. entries of mean zero, unit variance, and finite fourth order moment and $B_T \in \mathbb{C}^{N \times N}$ is a Hermitian nonnegative matrix of fixed rank K as $T \to \infty$. Let $\omega_0^2 \geq \ldots \geq \omega_{K-1}^2$ and $\hat{\lambda}_{0,T} \geq \ldots \geq \hat{\lambda}_{N-1,T}$ be the ordered eigenvalues of B_T and $\frac{1}{T} Y_T Y_T^{\mathsf{H}}$, respectively. Then*

- *If* $\omega_k^2 > \sqrt{c}$

$$\hat{\lambda}_{k,T} \xrightarrow{a.s.} \left(1+\omega_k^2\right)\left(1+\frac{c}{\omega_k^2}\right)$$

- *Otherwise*

$$\hat{\lambda}_{k,T} \xrightarrow{a.s.} (1+\sqrt{c})^2.$$

This theorem typically says under which condition an isolated eigenvalue $\hat{\lambda}_{k,T}$ of the sample covariance matrix can be observed outside the limiting support of the Marčenko–Pastur law. We note that this condition depends on the corresponding $(k+1)$th largest eigenvalues of B_T and on the limiting ratio c.

A further study of spiked models was provided by Baik *et al.* in [10] where the fluctuations of the K largest sample covariance eigenvalues were studied. This is provided in the following theorem:

Theorem 13 ([10]). *Let* $Y_T = (B_T + I_N)^{1/2} W_T$ *where* $W_T \in \mathbb{C}^{N\times T}$ *has i.i.d. complex Gaussian entries of mean zero and unit variance. Let* B_T *be a matrix of fixed rank* K *as* $T \to \infty$. *Denote by* $\hat{\lambda}_{0,T} \geq \ldots \geq \hat{\lambda}_{N-1,T}$ *and* $\omega_0^2 \geq \ldots \geq \omega_{K-1}^2$ *the ordered eigenvalues of* $\frac{1}{T} Y_T Y_T^{\mathsf{H}}$ *and* B_T, *respectively. As* $T \to \infty$, *assume* $N/T \to c \in (0,1]$. *The following holds for any real* x:

(i) If $\omega_0^2 < \sqrt{c}$ *then*

$$\mathbb{P}\left(N^{2/3}\frac{\hat{\lambda}_{0,T}-b}{\sigma} \geq x\right) \to F_{TW}(x)$$

(ii) When for some $0 \leq k \leq K-1$,

$$\omega_0^2 = \ldots = \omega_k^2 > \sqrt{c}$$

then

$$\mathbb{P}\left(\sqrt{N}\frac{\hat{\lambda}_{0,T}-\rho_0}{\sigma'} \geq x\right) \to G_k(x)$$

where

$$\rho_0 = \left(1+\omega_0^2\right)\left(1+\frac{c}{\omega_0^2}\right)$$
$$\sigma' = (1+\omega_0^2)\sqrt{1-\frac{c}{\omega_0^4}}$$

and G_k is the distribution function of the largest eigenvalue of the $(k+1)\times(k+1)$ GUE[1].

From this theorem if ω_0^2, the largest eigenvalue of B_T, is below the limit given by $\sqrt{c}$, the largest eigenvalue of $\frac{1}{T}Y_T Y_T^{\mathsf{H}}$, follows the Tracy–Widom law. If $\omega_0^2 > \sqrt{c}$, then the centered and scaled $\hat{\lambda}_{0,T}$ fluctuates as a Gaussian variable with a slower convergence rate than the Tracy–Widom variable.

A further contribution on the second order largest eigenvalue analysis when W_T is a real Gaussian matrix were provided by Paul in [53]. A non-Gaussian case of the entries of W_T was considered in [8] and then extended in [7] to the generalized spiked population model for which the limiting spectral measure of $\frac{1}{T}W_T W_T^{\mathsf{H}}$ is not restricted to the Marčenko–Pastur law. A rank one deformation of a Wigner matrix was studied by [28] and small rank perturbations of Hermitian Wigner matrix by [25], [54]. The work of [28] was extended to the non Wigner case when the sample distribution has some conditions on its moments in [29].

2.3.2 Advanced spiked models

We have presented the important results for spiked models when the full-rank matrix has i.i.d. entries. In practical applications, the models for which the rows or the columns of the perturbed matrix are not independent is of main importance. The results of this thesis are based on the model:

$$Y_T = P_T + V_T \tag{2.3}$$

where $Y_T \in \mathbb{C}^{N\times T}$, $P_T \in \mathbb{C}^{N\times T}$ is a random signal matrix of small rank K such that $\sup_T \|P_T\|/T$ is bounded, $V_T = W_T R_T^{1/2}$ is the noise matrix with $W_T \in \mathbb{C}^{N\times T}$ with i.i.d. standard complex Gaussian entries and $R_T \in \mathbb{C}^{T\times T}$ is a covariance matrix.

The limiting position of the outliers and their fluctuations in this setting depend on the interaction between the small rank matrix P_T and the covariance matrix R_T. In order for the outliers to converge, we shall make an additional technical assumption which will be met in practice.

Assumption 1. *Let $P_T = U_T(B_T)^{\mathsf{H}}$ be the Gram-Schmidt factorization of P_T where $U_T \in \mathbb{C}^{N\times K}$ is an isometry matrix and $B_T \in \mathbb{C}^{T\times K}$ is an upper triangular matrix whose first nonzero coefficient of each row is positive.*

[1] The matrix M is a GUE matrix if $M_{ii} \sim \mathcal{N}(0,1)$ and $M_{ij} \sim \mathcal{CN}(0,1)$ for $i<j$, these random variables being independent.

Define the $K \times K$ Hermitian nonnegative matrix-valued measure $\Lambda_T(dt)$ as

$$\Lambda_T(dt) = (B_T)^{\mathsf{H}} \begin{bmatrix} \delta_{\sigma^2_{0,T}}(dt) & & \\ & \ddots & \\ & & \delta_{\sigma^2_{T-1,T}}(dt) \end{bmatrix} B_T. \tag{2.4}$$

The accumulation points of Λ_T, denoted by Λ_*, are of the form $\nu(dt) \times U\Omega^2 U$ where

$$\Omega^2 = \begin{bmatrix} \omega_0^2 I_{j_0} & & \\ & \ddots & \\ & & \omega_t^2 I_{j_t} \end{bmatrix}, \ \omega_0^2 > \ldots > \omega_t^2, \ j_0 + \ldots + j_t = K$$

and where U is a unitary matrix.

The following result was proved by Chapon *et al.* in [21] and gives the limiting behavior of the spikes:

Theorem 14 ([21])**.** *Consider the model (2.3). Assume that $N/T \to c > 0$ as $T \to \infty$. Let $\sigma^2_{0,T}, \ldots, \sigma^2_{T-1,T}$ be the eigenvalues of R_T with spectral measure ν_T converging to ν with a compact support such that $\nu(\{0\}) = 0$. Assume that $\max_{t \in \{1,\ldots,T-1\}} \mathbf{d}\left(\sigma^2_{t,T}, \mathrm{supp}(\nu)\right) \to 0$ as $T \to \infty$. Let Assumption 1 hold. Let $[a, b]$ be the compact support of μ. Let $\hat{\lambda}_{0,T} \geq \ldots \geq \hat{\lambda}_{N-1,T}$ be the ordered eigenvalues of $\frac{1}{T} Y_T Y_T^{\mathsf{H}}$. Define the function*

$$g(x) \triangleq xm(x)\left(cm(x) - \frac{1-c}{x}\right)$$

which is positive and decreases from $g(b^+)$ to zero on (b, ∞). If $\omega_0^2 g(b^+) \leq 1$, then $\hat{\lambda}_{0,T} \xrightarrow{\text{a.s.}} b$. Otherwise, let $q \in \{0, \ldots, t\}$ be the largest index for which $\omega_q^2 g(b^+) > 1$. For $k = 0, \ldots, q$, let ρ_k be the unique solution x in (b, ∞) of $\omega_k^2 g(x) = 1$. Then, for $k = 0, \ldots, q$,

$$\hat{\lambda}_{j_0+\cdots+j_{k-1}+1,T}, \ldots, \hat{\lambda}_{j_0+\cdots+j_k,T} \xrightarrow[T\to\infty]{a.s.} \rho_k$$
$$\hat{\lambda}_{j_0+\cdots+j_q+1,T} \xrightarrow[T\to\infty]{a.s.} b.$$

This theorem shows in particular that the number of isolated eigenvalues of $\frac{1}{T} Y_T Y_T^{\mathsf{H}}$ is upper bounded by the rank K of P_T and it reaches this rank if ω_{K-1}^2 is large enough.

The following assumption is required for the second order result:

Assumption 2. *The following hold true:*

$$\sup_T \sqrt{T}|c_T - c| < \infty,$$
$$\limsup_T \sqrt{T}\left|\int \frac{1}{t-x}\nu_T(dt) - \int \frac{1}{t-x}\nu(dt)\right| < \infty, \text{ for all } x \in \mathbb{R} - \text{supp}(\nu).$$

Moreover, there exists a sequence of factorizations of P_T such that the measures Λ_T defined by (2.4) converge to $\nu(dt) \times \Omega$ and such that

$$\limsup_T \sqrt{T}\left\|\int \frac{1}{t-x}\Lambda_T(dt) - \int \frac{1}{t-x}\nu_T(dt) \times \Omega\right\| < \infty, \text{ for all } x \in \mathbb{R} - \text{supp}(\nu).$$

Define the following matrix-valued function, for $z \in \mathbb{C} - \text{supp}(\mu)$,

$$H_T(z) = \int \frac{m_T(z)}{1 + c_T m_T(z) t}\Lambda_T(dt).$$

Denote by $H_{0,T}(z), \ldots, H_{q,T}(z)$ the first upper left diagonal blocks of $H_T(z)$ where $H_{i,T}(z) \in \mathbb{C}^{j_i \times j_i}$. The following theorem provides the fluctuations of the outliers.

Theorem 15 ([21])**.** *Keeping the conditions of Theorem 22, let*

$$M_k = \sqrt{T}\left(\begin{bmatrix} \hat{\lambda}_{j_0+\ldots+j_{k-1}+1,T} \\ \vdots \\ \hat{\lambda}_{j_0+\ldots+j_k,T} \end{bmatrix} - \rho_k \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}\right)$$

where the eigenvalues $\hat{\lambda}_{k,T}$ of $\frac{1}{T}Y_T Y_T^{\mathsf{H}}$ are arranged in decreasing order. Let $G_0, \ldots, G_q$ be independent GUE matrices such that G_k is a $j_k \times j_k + 1$ matrix. Then, for any bounded continuous $f : \mathbb{R}^{j_0+\ldots+j_q} \to \mathbb{R}$,

$$\mathbb{E}[f(M_0, \ldots, M_q)] - \mathbb{E}[f(\chi_0, \ldots, \chi_q)] \xrightarrow[T\to\infty]{a.s.} 0$$

where $\chi_k \in \mathbb{R}^{j_k}$ is the random vector of the decreasingly ordered eigenvalues of the matrix

$$\frac{1}{\omega_k^2 g(\rho_k)'}(\alpha_k G_k + \sqrt{T}(H_{k,T}(\rho_k) + I_{j_k}))$$

with

$$\alpha_k^2 = \frac{m^2(\rho_k)}{\Delta(\rho_k)}\left[\int \frac{t^2 + 2\omega_k^2 t}{(1 + cm(\rho_k)t)^2}\nu(dt) + c\left(\int \frac{\omega_k^2 m(\rho_k)t}{(1 + cm(\rho_k)t)^2}\nu(dt)\right)^2\right]$$

and

$$\Delta(\rho_k) = 1 - c \int \left(\frac{m(\rho_k)t}{1 + cm(\rho_k)t} \right)^2 \nu(dt)$$

is positive for every $k = 0, \ldots, q$.

From these results, the errors $\hat{\lambda}_{k,T} - \rho_k$ are normally distributed and have a variance of order $1/T$.

Chapter 3

Detection techniques of a small rank signal

3.1 Problem statement and motivation

Consider a wireless sensor network composed from K possibly emitting sources and N sensors observing the sources during T successive snapshots. The sensors are supposed to be embedded in an additive noise. The received signal vector $y_t \in \mathbb{C}^{N\times 1}$ at the sample t is given by:

$$y_t = Hs_t + v_t \tag{3.1}$$

where $s_t \in \mathbb{C}^{K\times 1}$ is the signal vector, $H \in \mathbb{C}^{N\times K}$ is the channel matrix, and $v_t \in \mathbb{C}^{N\times 1}$ is the noise vector of covariance matrix R. A key question in array processing is estimation of parameters from the signal part Hs_t such a detection of the number of the emitting signals or their direction-of-arrival estimation from the T observations.

We propose in this thesis to deal with two main difficulties usually arising in modern sensor networks. The first difficulty is the large dimensions of the systems which in our setting means that the dimension of the received signal vector is assumed to be large. Due to the fast changing propagation environment occurring in the network, the scenarios such that $T >> N$ are no longer met. The size of the observation window T is usually of the same order of magnitude that N. Therefore, the regime that is going to be assumed is $N/T \to c > 0$ as $N, T \to \infty$. The second difficulty comes from the noise model. It is mostly assumed that the noise is an additive white Gaussian noise with some known or unknown variance. However, in reality it does not correspond to a lot of practical situations where spatial

or temporal correlation can occur. Such a noise model finds its applications in different fields as wireless communications and signal processing. In civil radar processing the received signal is usually superposed to echoes coming from a scattering environment. The scattering is due to earth's (woods, mountains) or sea surfaces, buildings in urban area, weather scatters as, *e.g.*, rain clouds, *etc.*. Interference in such dense scattering environments may contain multi-path and the induced noise is correlated in time. Stacking all the independent observations of the noise vectors into the matrix $V = [v_0, \ldots, v_{T-1}]$, we can write $V = WR^{1/2}$ where $W \in \mathbb{C}^{N\times T}$ has i.i.d. entries and $R \in \mathbb{C}^{T\times T}$ is a deterministic matrix.

When the sensors are too close to each other the noise is often correlated along the sensor arrays, and then the elements of the vector v_t are not independent. The resulting noise is spatially correlated, and we can write $V = \widetilde{R}^{1/2}W$ where $\widetilde{R} \in \mathbb{C}^{N\times N}$ is a deterministic matrix. The aim is to estimate the number of the emitting sources and the angle-of-arrival from the received signal estimations $y_0, \ldots, y_{T-1}$ only, under temporally or spatially correlated noise.

This chapter is organized as follows. In Section 3.2 the state-of-art of the existing detection techniques is provided. Section 3.3 overviews the MUSIC-based localization algorithms.

3.2 Detection techniques

In this section we provide an overview of the existing signal detection methods in different noise environments. Firstly, the white noise model is considered. The classical techniques for estimation of the number of unknown signals are provided followed by the approaches applied in the large dimensional setting. Secondly, signal detection in the presence of an uncorrelated noise model is studied assuming that a pure noise sequence is available at the receiver. Finally, the model of our interest is presented rising the problem of the so-called block detection, *i.e.* without assumption that a pure noise sequence is available. Further, the MUSIC-based direction-of-arrival estimation methods are discussed.

3.2.1 White noise environment

Recall the transmission model (3.1) and assume that the vector v_t has i.i.d. Gaussian entries with zero mean and variance σ^2. Depending on the available information on the noise and signal characteristics at the receiver side, there exist different detection approaches. When the noise variance σ^2 is

assumed to be known, for a single transmitter-receiver case with unknown deterministic signal, a classical technique is the energy detection method proposed by Urkowitz in [63] consisting in evaluation of the total received power. It is based on the fact that knowing the noise variance, although the signal is assumed to be unknown, the decision statistic at the output of the energy detector follows a non central chi-squared law. Using an approximation based on a modified central chi-square distribution, a detection threshold is fixed as a function of a target false alarm rate. The decision test consists then in comparing the total received power to this threshold. In the same setting an extension to the random signal case was considered in [38] and [24] making this applicable to the fading channel scenarios. However, the main limitation of such methods is that the knowledge of the noise variance is a quite unrealistic assumption. Further, information theoretic criteria-based methods are presented.

Information theoretic criteria-based methods

When the noise variance is assumed unknown and the number of the signals to be detected is equal to $K \geq 1$, one of the first historically known approaches was provided by Wax and Kailath in [67] and is based on the famous Akaike Information Criterion (AIC) [1] and the Minimum Description Length (MDL) [56]. We recall the transmission model with the received signal vector $y_t \in \mathbb{C}^{N\times 1}$

$$y_t = Hs_t + v_t \tag{3.2}$$

where $H = [h(\theta_0), \ldots, h(\theta_{K-1})] \in \mathbb{C}^{N\times K}$ is the channel matrix with $h(\theta_k) \in \mathbb{C}^{N\times 1}$ a complex vector parameterized by an unknown parameter θ_k, $s_t \in \mathbb{C}^{K\times 1}$ and $v_t \in \mathbb{C}^{N\times 1}$ are respectively the signal and the noise vectors modeled as zero mean stationary ergodic Gaussian vector processes, independent of each other with respective covariance matrices $\Gamma \in C^{K\times K}$ and $R = \sigma^2 I_N$ where σ^2 is unknown. The detection is based on the received signal covariance matrix structure given by

$$\Sigma = H\Gamma H^{\mathsf{H}} + \sigma^2 I_N.$$

Assuming that the channel matrix H is of full rank K and that the signal covariance matrix Γ is nonsingular, then the matrix $H\Gamma H^{\mathsf{H}}$ is of rank K. Denote by $\lambda_0 \geq \ldots \geq \lambda_{N-1}$ the eigenvalues of Σ. The smallest $N - K$ eigenvalues of Σ are all equal to σ^2 and correspond to the noise eigenvalues:

$$\lambda_K = \lambda_{K+1} = \ldots = \lambda_{N-1} = \sigma^2.$$

Hence, the number of signals can be obtained from the multiplicity of the smallest eigenvalue of Σ. However, in practice we do not have access to the true covariance matrix Σ but only to the sample covariance matrix which is obtained from the T observations of the received signal vector and defined by

$$\widehat{\Sigma} = \frac{1}{T}\sum_{t=0}^{T-1} y_t y_t^{\mathsf{H}}.$$

Denote the decreasingly ordered eigenvalues of $\widehat{\Sigma}$ by $\hat{\lambda}_0 > \hat{\lambda}_1 > \ldots > \hat{\lambda}_{N-1}$. The detection method of [67] relies on the closeness of the $N-K$ noise eigenvalues of $\widehat{\Sigma}$. Given T observations $y_0, \ldots, y_{T-1}$, the Akaike's information criteria consists in selecting the model which best fits the data by minimizing the function

$$\mathrm{AIC} = -2\log\left(y_0, \ldots, y_{T-1}|\widehat{\Theta}\right) + 2k$$

where $\widehat{\Theta}$ is the maximum likelihood estimate of the parameter vector Θ and k is the number of the parameters in the model. After computation of the maximum likelihood function of the unknown parameters, the source number estimation is given by

$$\begin{aligned}\widehat{K}_{\mathrm{AIC}} &= \arg\ \min\ \mathrm{AIC}(k) \\ &= \arg\ \min\ -2(N-k)T\log\left(\frac{\prod_{i=k}^{N-1}\hat{\lambda}_i^{1/(N-k-1)}}{\frac{1}{N-k}\sum_{i=k}^{N-1}\hat{\lambda}_i}\right) + 2k(2N-k)\end{aligned}$$

for $k = 0, \ldots, N-1$. The MDL-based estimator proposed as well in [67] differs by the penalty term and given by

$$\begin{aligned}\widehat{K}_{\mathrm{MDL}} &= \arg\ \min\ \mathrm{MDL}(k) \\ &= \arg\ \min\ -(N-k)T\log\left(\frac{\prod_{i=k}^{N-1}\hat{\lambda}_i^{1/(N-k-1)}}{\frac{1}{N-k}\sum_{i=k}^{N-1}\hat{\lambda}_i}\right) + \frac{1}{2}k(2N-k)\log T\end{aligned}$$

for $k = 0, \ldots, N-1$. It was shown in [67] that as N is fixed and $T \to \infty$, the estimator $\widehat{K}_{\mathrm{AIC}}$ tends to overestimate the number of signals while $\widehat{K}_{\mathrm{MDL}}$ is consistent. An alternative consistent information theoretic criterion based algorithm was established in [75]. As an extension of the work of [67], more general expressions for the AIC and the MDL-based estimators are provided for which the observations $y_0, \ldots, y_{T-1}$ are not necessarily independent and the noise vector v_t is not necessarily Gaussian.

In many works, the performances of the AIC and the MDL estimators were studied (see among others [67], [75], [73], [43]). In the large dimensional setting for which $N, T \to \infty$ with $N/T \to c > 0$, both of these algorithms underestimate the real number of the signals (see for instance [46]). This is due to the fact that the eigenvalues of the sample covariance matrix corresponding to the noise part are not concentrated around σ^2 anymore but spread and this depends on the parameter c [44] (see also Chapter 2 for more details). The noise spectrum spreading increases with the increase of c. In this setting the random matrix theory tools are used in order to estimate consistently K as $T \to \infty$. One of the first studies of signal detection in the the framework of large random matrix theory was considered by Silverstein and Combettes in [61]. The detection method is based on the fact that the noise eigenvalues corresponding to the $N - K$ smallest eigenvalues of the sample covariance matrix are asymptotically close to each other. They showed that with probability one the spectrum of the sample covariance matrix splits into two distinct intervals corresponding to the noise eigenvalues and the signal eigenvalues with respective proportions close to $(N-K)/N$ and K/N with high probability. However, the condition under which the splitting phenomenon occurs have been not investigated in their work. A more accurate detection is related to the study of the extreme sample covariance matrix eigenvalue behavior and, particularly, to the results on the spiked models. We stack now all the received-vector T observations into the matrix:

$$Y_T = H_T S_T^{\mathsf{H}} + W_T \tag{3.3}$$

where $Y_T = [y_0, \ldots, y_{T-1}] \in \mathbb{C}^{N \times T}$, $S_T = [s_0, \ldots, s_{T-1}]^{\mathsf{H}} \in \mathbb{C}^{T \times K}$ is the signal matrix whose columns are independent with zero mean complex Gaussian entries and $K \times K$ covariance matrix Γ, and $W_T = [w_0, \ldots, w_{T-1}] \in \mathbb{C}^{N \times T}$ is the noise matrix with i.i.d. zero mean complex Gaussian entries with variance σ^2, H_T is an unknown $N \times K$ deterministic channel matrix. In the model (3.3), the full-rank matrix Y_T can be viewed as a perturbed version of the noise matrix W_T, the additive perturbation $H_T S_T^{\mathsf{H}}$ having a small rank K. This is referred to as the additive spiked models (see Chapter 2). Based on this model signal detection was extensively studied during the last decade and is related to the limiting spectrum behavior analysis of the sample covariance matrix $\widehat{\Sigma}_T = \frac{1}{T} Y_T Y_T^{\mathsf{H}}$. Consider the decreasingly ordered eigenvalues of $\widehat{\Sigma}_T$ denoted by $\hat{\lambda}_{0,T} \geq \ldots \geq \hat{\lambda}_{N-1,T}$. Under the hypothesis that only noise is present, the spectrum of the sample covariance matrix converges to the Marčenko–Pastur law [44] and from [4] with probability one no eigenvalues can be found outside the support of the Marčenko–Pastur

distribution given by the interval $[\sigma^2(1-\sqrt{c})^2, \sigma^2(1+\sqrt{c})^2]$. Under the hypothesis that K signals are present such that K is fixed as $T \to \infty$, the eigenvalues of $\widehat{\Sigma}_T$ converge as well to the Marčenko–Pastur law [3]. However, under some conditions on the signal power and the parameter c, up to K isolated eigenvalues can be observed on the right side of the Marčenko–Pastur law support. The condition under which isolated eigenvalues are observed is based on the K signal eigenvalues of the matrix $H_T \Gamma H_T^{\mathsf{H}}$ denoted by $\omega_0^2 \geq \ldots \geq \omega_{K-1}^2$. From the relevant results on spiked models such as [9], [53], [10], for $k = 0, \ldots, K-1$, if the signal eigenvalue satisfies $\omega_k^2 > \sigma^2\sqrt{c}$, then as $T \to \infty$ the corresponding eigenvalue of the sample covariance matrix $\hat{\lambda}_k$ converges almost surely to the limit $\rho_k = \left(1 + \frac{\omega_k^2}{\sigma^2}\right)\left(1 + \frac{c\sigma^2}{\omega_k^2}\right) > \sigma^2(1+\sqrt{c})^2$ and hence the corresponding signal can be detected. When the signal power is below the detectability level, *i.e.*, $\omega_k^2 < \sigma^2\sqrt{c}$, then the corresponding eigenvalue $\hat{\lambda}_{k,T}$ converges almost surely to the right edge of the limiting support $\hat{\lambda}_k \xrightarrow{\text{a.s.}} \sigma^2(1+\sqrt{c})^2$.

In large dimensional regime, most of the detection methods are based on the results on spiked models. In this context, based on the above discussions, an improved version of the information theoretic approach of [67] was proposed by Nadakuditi and Edelman in [46]. It is assumed that the signal and the noise matrices have i.i.d. real Gaussian entries. The signal number estimator is given by

$$\widehat{K}_{\text{NE}} = \arg\min \left\{ \frac{\beta}{4} \left[\frac{T}{N}\right]^2 \alpha_k^2 \right\} + 2(k+1) \text{ for } k = 0, \ldots, \min(N,T) - 1,$$

with $\beta = 1$ if $y_t \in \mathbb{R}^{N\times 1}$, and $\beta = 2$ if $y_t \in \mathbb{C}^{N\times 1}$, and α_k defined by

$$\alpha_k = \left[(N-k) \frac{\sum_{i=k}^{N-1} \hat{\lambda}_{i,T}^2}{\left(\sum_{i=k}^{N-1} \hat{\lambda}_{i,T}\right)^2} - \left(1 + \frac{N}{T}\right) \right] N - \left(\frac{2}{\beta} - 1\right) \frac{N}{T}.$$

Recall that $\omega_0^2 \geq \ldots \geq \omega_{K-1}^2$ are the eigenvalues of the covariance matrix of the signal part $H_T \Gamma H_T^{\mathsf{H}}$. From the discussions on the spiked model results, if ω_k^2 is above the detectability level, the corresponding eigenvalue is asymptotically located outside the Marčenko–Pastur law support and hence, the corresponding signal is detectable. If not all the signal eigenvalues satisfy the detectability condition, the underestimation of the number of signals in unavoidable. The number of effective signals is heuristically defined by

$$K_{\text{eff}} = \# \left\{ k : \ \omega_k > \sigma^2 \sqrt{\frac{N}{T}} \right\}.$$

It is conjectured that as N, $T \to \infty$, the estimator $\widehat{K}_{\text{NE}}$ is a consistent estimate of K_{eff}. Although the derivation of the estimator of [46] is based on the test statistic $\sum_{i=k}^{N-1} \hat{\lambda}_{i,T}^2 / \left(\sum_{i=k}^{N-1} \hat{\lambda}_{i,T} \right)^2$, it does not consider the setting of the threshold level. However, when the signal has a low power, the information criterion based methods do not provide a good detection performance. Indeed, it was shown in [39], [40] that the detectability level for the MDL algorithm is significantly larger (by approximately $\sqrt{2 \log T}$) than the asymptotic limit given by $\sigma^2 \sqrt{c}$. Therefore, a finer statistical analysis is needed. It consists in hypothesis testing approaches principally involving the statistical behavior of the sample covariance matrix largest eigenvalues under hypothesis where only noise is present versus the hypothesis of the signal presence.

Hypothesis testing based detection methods

Hypothesis testing based approaches consider testing an hypothesis H_0 versus its alternative H_1. For a single signal detection this is presented by the following decision:

$$\begin{aligned} H_0 &: \text{no signal} \\ H_1 &: \text{signal present.} \end{aligned}$$

Regarding the above discussions, an heuristic method for single signal detection was proposed by Cardoso *et al.* in [20] assuming that the noise variance σ^2 is unknown. Recall that $\hat{\lambda}_{0,T} \geq \ldots \geq \hat{\lambda}_{N-1,T}$ are the ordered eigenvalues of the sample covariance matrix $\widehat{\Sigma}_T$. Under H_0 they are all asymptotically located in the interval $[\sigma^2(1-\sqrt{c})^2, \sigma^2(1+\sqrt{c})^2]$ with probability one. The test statistic consists in rejecting the null hypothesis for large values of the statistics given by the ratio of the largest to the smallest eigenvalues of $\widehat{\Sigma}_T$:

$$\frac{\hat{\lambda}_{0,T}}{\hat{\lambda}_{N-1,T}} \underset{H_1}{\overset{H_0}{\lessgtr}} \frac{(1+\sqrt{c})^2}{(1-\sqrt{c})^2}.$$

The above ratio test is referred in the literature to as the condition number of the sample covariance matrix. Note that the detection threshold given by $(1+\sqrt{c})^2/(1-\sqrt{c})^2$ is independent of the unknown noise variance. However, this approach tests the finite dimension based ratio $\hat{\lambda}_{0,T}/\hat{\lambda}_{N-1,T}$ (it is obtained from $\widehat{\Sigma}_T$ with N and T large but finite) against an asymptotic detection threshold. The threshold approximation turns out to be inaccurate in many practical situations. This approach can be improved by using the

limiting law of the ratio $\hat{\lambda}_{0,T}/\hat{\lambda}_{N-1,T}$ allowing to set the detection threshold as a function of the false alarm probability. Taking into account the limiting statistics of the largest eigenvalue $\hat{\lambda}_{0,T}$ which follows the Tracy–Widom distribution and making an approximation on the minimum eigenvalue by considering its limit, a detection method with a control of a false alarm rate was proposed in [74]. With the result on the smallest eigenvalue limiting distribution derived in [27], an analytical expression of the limiting distribution of the ratio test $\hat{\lambda}_{0,T}/\hat{\lambda}_{N-1,T}$ is given in [55] and a more accurate threshold, which is a function of an admissible probability of false alarm, is calculated. This method gives better detection performances in comparison to the asymptotic or semi-asymptotic threshold based methods of [20] and [74], respectively.

In was shown by Bianchi *et al.* in [14] that the condition number based test of [55] is asymptotically outperformed by an improved method of [14] based on the ratio of the largest eigenvalue of $\widehat{\Sigma}_T$ to its normalized trace. The hypothesis H_0 is rejected for high values of the test:

$$\frac{\hat{\lambda}_{0,T}}{\frac{1}{N}\operatorname{Tr}\widehat{\Sigma}_T} \underset{H_1}{\overset{H_0}{\lessgtr}} \gamma_T$$

where γ_T is the detection threshold to be defined later. The above test refers in the literature to as the so-called Generalized Likelihood-Ratio Test (GLRT). It is based on the commonly used Neyman-Pearson detection procedure (see *e.g.* [42]) and consists in replacing the unknown parameters by their maximum likelihood estimates in the likelihood-ratio test. Hereafter we provide the main steps of the derivation of the test (*cf.* [14]). Recall the matrix based transmission model under two hypotheses:

$$Y_T = \begin{cases} W_T \text{ under } H_0, \\ \boldsymbol{h}_T \boldsymbol{s}_T^{\mathsf{H}} + W_T \text{ under } H_1 \end{cases}$$

where $\boldsymbol{h}_T \in \mathbb{C}^{N\times 1}$ is the deterministic channel vector, $\boldsymbol{s}_T \in \mathbb{C}^{T\times 1}$ is the signal vector with i.i.d. standard complex Gaussian entries, and the noise is defined as previously. It is known that if the channel vector $\boldsymbol{h}_T$ and the noise variance σ^2 are known at the receiver, then the optimal test is the Neyman-Pearson detector based on the Likelihood-Ratio Test (LRT)

$$\text{LRT} = \frac{p_1(Y_T; \boldsymbol{h}_T, \sigma^2)}{p_0(Y_T; \sigma^2)}$$

where the likelihood functions $p_0(Y_T; \sigma^2)$ and $p_1(Y_T; \boldsymbol{h}_T, \sigma^2)$ under H_0 and

H_1 hypotheses respectively are defined as

$$p_0(Y_T;\sigma^2) = \left(\pi\sigma^2\right)^{-NT} \exp\left(-\frac{T}{\sigma^2}\operatorname{Tr}\widehat{\Sigma}_T\right)$$
$$p_1(Y_T;\boldsymbol{h}_T,\sigma^2) = \left(\pi^N \det\left(\boldsymbol{h}_T\boldsymbol{h}_T^{\mathsf{H}} + \sigma^2 I_N\right)\right)^{-T} \times \exp\left(-T\operatorname{Tr}\left(\widehat{\Sigma}_T\left(\boldsymbol{h}_T\boldsymbol{h}_T^{\mathsf{H}} + \sigma^2 I_N\right)^{-1}\right)\right).$$

When the channel and the noise variance are unknown at the receiver as it is the case in our model setting, it is necessary to integrate over all possible channel realizations and the noise variances. A suboptimal GLRT is then obtained and is given by the ratio:

$$\text{GLRT} = \frac{\sup_{\boldsymbol{h}_T,\sigma^2} p_1(Y_T;\boldsymbol{h}_T,\sigma^2)}{\sup_{\sigma^2} p_0(Y_T;\sigma^2)}. \tag{3.4}$$

A closed-form expression for (3.4) is given by

$$\text{GLRT} = \frac{\left(1-\frac{1}{T}\right)^{(1-T)N}}{(\mathcal{T}_T)^N\left(1-\frac{\mathcal{T}_T}{T}\right)^{(T-1)N}}$$

where

$$\mathcal{T}_T = \frac{\hat{\lambda}_{0,T}}{\frac{1}{N}\operatorname{Tr}\widehat{\Sigma}_T}.$$

This leads to an equivalent detection test:

$$\mathcal{T}_T = \frac{\hat{\lambda}_{0,T}}{\frac{1}{N}\operatorname{Tr}\widehat{\Sigma}_T} \underset{H_1}{\overset{H_0}{\lessgtr}} \gamma_T \tag{3.5}$$

where γ_T is the detection threshold whose setting is provided hereafter. There exist two types of detection errors. The error of the first kind is an incorrect rejection of the hypothesis H_0 when it is true. The probability of this error is called the probability of false alarm and is given by $\mathbb{P}_0(\mathcal{T}_T > \gamma_T)$ for a given detection threshold γ_T. The error of the second type is an incorrect rejection of the hypothesis H_1 when it is true and called the probability of missed detection and given by $\mathbb{P}_1(\mathcal{T}_T < \gamma_T)$. Based on that the detection threshold γ_T is defined such that the power of the test $\mathbb{P}_1(\mathcal{T}_T > \gamma_T)$, *i.e.*, the probability to detect the signal under hypothesis H_1, is maximized while keeping the probability of false alarm $\mathbb{P}_0(\mathcal{T}_T > \gamma_T)$ under a fixed level $\alpha \in (0,1)$.

The asymptotic detection threshold as a function of the probability of false alarm is closely related to the distribution of the largest eigenvalue of the sample covariance matrix in the signal-free case given by $\frac{1}{T}W_T W_T^{\mathsf{H}}$. As the test $\mathcal{T}_T$ does not depend on the noise variance, in the following it is assumed that $\sigma^2 = 1$. The largest eigenvalue limiting behavior of the Wishart matrix $\frac{1}{T}W_T W_T^{\mathsf{H}}$ has been already introduced in Chapter 2 (see Theorem 10). It was shown that under H_0 the centered and scaled largest eigenvalue fluctuations follow the Tracy–Widom distribution [35] and, hence, we have, for any real x:

$$\mathbb{P}\left[N^{2/3}\frac{\hat{\lambda}_{0,T} - b_T}{\sigma_T} < x\right] \to F_{TW}(x)$$

where F_{TW} is the Tracy–Widom cumulative distribution function (c.d.f.) and b_T and σ_T are the centering and the scaling parameters, respectively, depending on the ratio $c_T = N/T$ and defined by

$$b_T = (1 + \sqrt{c_T})^2 \tag{3.6}$$

$$\sigma_T = (1 + \sqrt{c_T})^{4/3}\sqrt{c_T}. \tag{3.7}$$

The key point assumed in [14] is that asymptotically the fluctuations of $\frac{1}{N}\operatorname{Tr}\widehat{\Sigma}_T$ are negligible compared to those of $\hat{\lambda}_{0,T}$. Therefore, under H_0, the test $\mathcal{T}_T$ approximately fluctuates as a Tracy–Widom variable:

$$\mathbb{P}\left[N^{2/3}\frac{\mathcal{T}_T - b_T}{\sigma_T} < x\right] \to F_{TW}(x)$$

for any real x. Making use of this result the limiting behavior of the test is given in the following theorem:

Theorem 16 ([14])**.** *Consider a fixed level of the probability of false alarm $\alpha \in (0,1)$ and let γ_T be the threshold for which the power of test (3.5) is maximum. Then the following convergence holds true*

$$N^{2/3}\frac{\mathcal{T}_T - b_T}{\sigma_T} \xrightarrow[T\to\infty]{} \bar{F}_{TW}^{-1}(\alpha)$$

where $\bar{F}_{TW}^{-1} = 1 - F_{TW}$ is the complementary Tracy–Widom c.d.f.. Moreover, the probability of false alarm of the following test

$$\mathcal{T}_T \underset{H_0}{\overset{H_1}{\lessgtr}} b_T + \frac{\sigma_T}{N^{2/3}}\bar{F}_{TW}^{-1}(\alpha)$$

converges to α.

Note that from this result the threshold γ_T can be approximated by $b_T + \frac{\sigma_T}{N^{2/3}} \bar{F}_{TW}^{-1}(\alpha)$. The detection threshold actually depends on the tail probability of the variable $\mathcal{T}_T$. A tuner result on the distribution of the variable $\mathcal{T}_T$ was derived in [48] and, for any real x, given by

$$\mathbb{P}\left[\frac{\mathcal{T}_T - \tilde{b}_T}{\tilde{\sigma_T}} < x\right] = F_{TW}(x) - \frac{1}{NT}\left(\frac{\tilde{b}_T}{\tilde{\sigma}_T}\right)^2 F''_{TW}(x)$$

where $\tilde{b}_T$ and $\tilde{\sigma}_T$ are centering and scaling parameters derived for the real noise model. Note that using this result may lead to a finer detection performance instead of the approximation by the largest eigenvalue which have been not yet studied in the literature.

Assuming $K \geq 1$ and σ^2 unknown, consider now a sequence of hypothesis tests. In this case there is no optimal detector for a specific false alarm rate. Only suboptimal solutions can be derived. A multiple hypothesis testing approach was proposed by Kritchman and Nadler in [40] which considers the following hypothesis testing:

$$\begin{aligned} H_0 &: \text{at most } k \text{ signals present} \\ H_1 &: \text{at least } k+1 \text{ signals present} \end{aligned} \tag{3.8}$$

for $k = 0, \ldots, \min(N, T) - 1$. At each sequence of hypothesis test, the significance level of the kth largest eigenvalue is considered and rejects the hypothesis H_0 if $\hat{\lambda}_{k,T}$ is too large:

$$\hat{\lambda}_{k,T} \underset{H_1}{\overset{H_0}{\lessgtr}} \xi_T. \tag{3.9}$$

The testing is stopped at the smallest index k such that $\hat{\lambda}_{k,T} < \xi_T$. As previously, the detection threshold is set as a function of the probability of false alarm from the distribution of $\hat{\lambda}_{k,T}$. It was shown in [39] that for N and T large enough, $\hat{\lambda}_{k,T}$ approximately follows the Tracy–Widom law. Hence, the detection threshold ξ_T can be approximated by

$$\xi_T \simeq \hat{\sigma}^2(k)\left(b_T + \frac{\bar{F}_{TW}^{-1}(\alpha)\sigma_T}{N^{2/3}}\right)$$

where b_T and σ_T are defined by (3.6) and (3.7), respectively, α is the probability of false alarm, and $\hat{\sigma}^2(k)$ is an consistent estimate of the noise variance

estimated by analyzing the interactions between signal and noise eigenvalues (see [40]). The estimator of the number of signals is then defined by

$$\widehat{K}_{\mathrm{KN}} = \arg\ \min \left\{ \hat{\lambda}_{k,T} < \hat{\sigma}^2(k) \left(b_T + \frac{\bar{F}_{TW}^{-1}(\alpha)\sigma_T}{N^{2/3}} \right) \right\} - 1$$

for $k = 0, \ldots, \min(N, T) - 1$. It was shown in [40] that as $T \to \infty$, $\widehat{K}_{\mathrm{KN}}$ is a consistent estimate of K if the smallest signal eigenvalue ω_{K-1} satisfies the detectability condition, *i.e.* $\omega_{K-1} > \sigma^2\sqrt{c}$. Assume now $K = 1$. Note that the test (3.9) is equivalent to

$$\frac{\hat{\lambda}_{k,T}}{\hat{\sigma}^2(0)} \underset{H_1}{\overset{H_0}{\lessgtr}} b_T + \frac{\bar{F}_{TW}^{-1}(\alpha)\sigma_T}{N^{2/3}}.$$

Hence, the difference with the test $\frac{\hat{\lambda}_{k,T}}{\frac{1}{N}\operatorname{Tr}\widehat{\Sigma}_T}$ is that the noise variance σ^2 is estimated by the normalized trace $\frac{1}{N}\operatorname{Tr}\widehat{\Sigma}_T$ which was shown to be less accurate that the estimate $\hat{\sigma}^2$ in [40]. However, it was noticed in [14] that both tests are asymptotically equivalent in terms of error exponent which is the limit as $T \to \infty$ of a function of the probability of missed detection at a given probability of false alarm.

3.2.2 Correlated noise environment

Regarding the discussions at the beginning of the chapter, we assume now that the noise samples $v_0, \ldots, v_{T-1}$ are independent complex N-variate Gaussian variables with $N \times N$ covariance matrix R_T which is positive semidefinite (see [2, Theorem 2.3.1]) and hence, invertible. Stacking all the observed noise vectors into the matrix $V_T = [v_0, \ldots, v_{T-1}]$, we can write $V_T = R_T^{1/2} W_T$ where $R_T^{1/2}$ is any square root of the matrix R_T and W_T has i.i.d. complex Gaussian with zero mean entries with unit variance. If R_T is known at the receiver we can whiten the received data:

$$R_T^{-1/2} Y_T = R_T^{-1/2} H_T S_T^{\mathsf{H}} + W_T. \tag{3.10}$$

The signal number estimation problem is then equivalent to that in a white noise environment and the methods exposed previously can be applied.

We assume now that the noise covariance R_T is unknown at the receiver. Recall that both the AIC and the MDL algorithms are based on the closeness of the noise eigenvalues and the model order is estimated when a gap between the noise and the signal eigenvalues is displayed. When N is fixed

and $T \to \infty$, it was shown that both the AIC and the MDL detectors tend to overestimate the number of sources [73]. The overestimation of the model order comes from the fact that when the noise is not white, the eigenvalues of the noise covariance matrix are not all equal and when they are not sufficiently clustered an overestimation occurs as was heuristically observed in [43]. Moreover, the probability of overestimation increases as T increases [73].

Pure noise sequence available

Most of the existing methods assume that a noise only sequence is available from which the matrix R_T is estimated. Then the received signal matrix is "whitened" by the estimate $\widehat{R}_T$. One of the first works using this setting belongs to Zhao *et al.* [76] and is based on the information theoretic criteria. They considered the following model at time interval t:

$$y_t = Hs_t + \sigma v_t$$

where $y_t \in \mathbb{C}^{N\times 1}$, σ is an unknown scalar, $s_t \in \mathbb{C}^{K\times 1}$ and $v_t \in \mathbb{C}^{N\times 1}$ are distributed independently as complex multivariate normal with respective covariance matrices Γ and R, and H is a deterministic channel matrix. Since the signal and the noise vectors are independent, the covariance matrix of y_t can be decomposed as

$$\Sigma = H\Gamma H^{\mathsf{H}} + \sigma^2 R.$$

The "whitened" covariance matrix is then defined by

$$\Sigma_R = R^{-1}\Sigma = R^{-1}H\Gamma H^{\mathsf{H}} + I_N.$$

Denote by $\lambda_0 \geq \ldots \geq \lambda_{N-1}$ the ordered eigenvalues of Σ_R. Assuming that the rank of $R^{-1}H\Gamma H^{\mathsf{H}}$ is equal to K, the $N-K$ smallest eigenvalues of Σ_R are all equal to one:

$$\lambda_K = \lambda_{K+2} = \ldots = \lambda_{N-1} = 1.$$

It is further assumed that T_1 independent observations of the noise-only data $v_0, \ldots, v_{T_1-1}$ are available. The noise sample covariance matrix is given by

$$\widehat{R} = \frac{1}{T_1}\sum_{t=0}^{T_1-1} v_t v_t^{\mathsf{H}}.$$

Note that $\widehat{R}$ is the noise-only sample covariance matrix having the Wishart distribution [70] is nonsingular if $T_1 > N$. Hence, $\widehat{R}$ is invertible with probability one [2, Chapter 7.7]. Further, it is assumed that a second independent

sequence of data $y_0, \ldots, y_{T_2-1}$ containing the signal is available. The sample covariance matrix is given by

$$\widehat{\Sigma} = \frac{1}{T_2} \sum_{t=0}^{T_2-1} y_t y_t^{\mathsf{H}}.$$

Denote by $\hat{\lambda}_0 \geq \ldots \geq \hat{\lambda}_{N-1}$ the ordered eigenvalues of the "whitened" sample covariance matrix $\widehat{R}^{-1}\widehat{\Sigma}$. The logarithm of the likelihood-ratio statistic derived by Zhao *et al.* in [76] for $k = 0, \ldots, N-1$ is given by

$$\log L_k = -\frac{1}{2} \log \prod_{i=k}^{N-1} \left[\left(\frac{T_1 \hat{\sigma}_{k_0}^2 + T_2 \hat{\lambda}_i}{T} \right)^T \frac{1}{\hat{\lambda}_i^{T_1} \hat{\sigma}_{k_0}^{2T_1}} \right] \tag{3.11}$$

where $T = T_1 + T_2$, $\hat{\sigma}_{k_0}^2$ is the solution of the equation

$$N - k = \sum_{j=k}^{N-1} \frac{\hat{\lambda}_j}{\frac{T_1}{T}\hat{\sigma}_{k_0}^2 + \frac{T_2}{T}\hat{\lambda}_j}.$$

The estimate of K given in [76] is

$$\widehat{K}_{\text{ZKB}} = \max \{k : 0 \leq k \leq N-1,\ \log L_k - \log L_{k-1} > C_T\}$$

where $\log L_k$ is defined by (3.11) and C_T is a constant satisfying the following

(i) $\lim\limits_{T\to\infty} C_T/T = 0$

(ii) $\lim\limits_{T\to\infty} C_T/\log\ \log T = \infty$.

It was established that the estimate $\widehat{K}_{\text{ZKB}}$ is strongly consistent and the proof is based on the fact that $\lim\limits_{T\to\infty} \hat{\lambda}_i = 1$ for $i > K-1$. However, the problem of this method is that the threshold setting of C_T is heuristic. It is set taking into account the spacing between the eigenvalues of Σ_R and $\widehat{\Sigma}_{\widehat{R}}$ which is of order $\sqrt{\frac{\log\log T}{T}}$ (see [76]).

The large dimensional setting was considered by Nadakuditi and Silverstein in [47]. Similarly to the method of [76], the detection is performed into two steps. First, it is assumed that we have access to the pure noise sequence and the noise sample covariance matrix as above is given by $\widehat{R}_T = \frac{1}{T_1} \sum_{t=0}^{T_1-1} v_t v_t^{\mathsf{H}}$. We consider the sample covariance matrix

$\widehat{\Sigma}_T = \frac{1}{T_2}\sum_{t=0}^{T_2-1} y_t y_t^{\mathsf{H}}$ where the samples y_t contain possibly a signal. Then the matrix $\widehat{\Sigma}_T$ is "whitened" giving rise to

$$\widehat{\Sigma}_{\widehat{R}_T} = \widehat{R}_T^{-1}\widehat{\Sigma}_T.$$

The detection is based on the limiting spectral behavior of $\widehat{\Sigma}_{\widehat{R}_T}$ which refers in the literature to as the so-called multivariate F-matrix whose limiting eigenvalue distribution was derived in [58]. The first step is to consider the signal-free case. The following theorem describes the limiting spectral behavior of $\widehat{\Sigma}_{\widehat{R}_T}$.

Theorem 17 ([58]). *Consider $\widehat{R}_T = \frac{1}{T_1}\sum_{t=0}^{T_1-1} v_t v_t^{\mathsf{H}}$ and assume the signal free received sample covariance matrix $\widehat{\Sigma}_T = \frac{1}{T_1}\sum_{t=0}^{T_1-1} y_t y_t^{\mathsf{H}}$ where $v_t \sim \mathcal{CN}(0, R_T)$ and $y_t \sim \mathcal{CN}(0, R_T)$ are independent. Then the empirical spectral measure of $\widehat{\Sigma}_{\widehat{R}_T} = \widehat{R}_T^{-1}\widehat{\Sigma}_T$ denoted by μ_T converges almost surely to the limit measure μ, as N, $T_1(N) \to \infty$, N, $T_2(N) \to \infty$, $N/T_1 \to c_1 < 1$, $N/T_2 \to c > 0$, for all $x \in \mathbb{R}$, with*

$$d\mu(x) = \frac{(1-c_1)\sqrt{(x-b_1)(b_2-x)}}{2\pi x(xc_1+c)}\mathbb{1}_{[b_1,b_2]}(x)dx + \max\left(0, \left(1-\frac{1}{c}\right)\right)\delta(x)$$

where

$$b_1 = \left(\frac{1-\sqrt{1-(1-c)(1-c_1)}}{1-c_1}\right)^2 \tag{3.12}$$

$$b_2 = \left(\frac{1+\sqrt{1-(1-c)(1-c_1)}}{1-c_1}\right)^2. \tag{3.13}$$

Recall that from the result of [4], for any matrix X_T with i.i.d. zero mean unit variance entries with finite fourth moment, and any deterministic matrix A_T with bounded spectral norm, asymptotically no eigenvalue can be found outside the support of the limiting spectral distribution of $\frac{1}{T_2}X_T^{\mathsf{H}}A_T X_T$. This result was extended in [47] for a random matrix A_T. In our setting A_T is replaced by $R_T^{1/2}\widehat{R}_T^{-1}R_T^{1/2}$ and X_T by W_T with i.i.d. complex Gaussian entries of zero mean and unit variance. The matrix $\frac{1}{T_2}W_T R_T^{1/2}\widehat{R}_T^{-1}R_T^{1/2}W_T^{\mathsf{H}}$ has the same eigenvalues as $\widehat{\Sigma}_{\widehat{R}_T} = \widehat{R}_T^{-1}\widehat{\Sigma}_T$, hence, with probability one for all N, T_1, T_2 large no eigenvalue of $\widehat{\Sigma}_{\widehat{R}_T}$ can be found outside the limiting support $[b_1, b_2]$ with b_1 and b_2 defined by (3.12) and (3.13), respectively. Therefore, if an isolated eigenvalue appears outside the interval $[b_1, b_2]$ it is due to the presence of the signal. The condition of the presence of an isolated eigenvalue is presented in the following theorem:

Theorem 18 ([47]). *Consider now the signal-plus-noise model with the setting of Theorem 17. Let $\lambda_0 \geq \lambda_1 \geq \ldots \geq \lambda_{K-1} > \lambda_K = \ldots = \lambda_{N-1} = 1$ be the eigenvalues of Σ_{R_T}. Denote by $\hat{\lambda}_j$ the $(j+1)$th largest eigenvalue of $\widehat{\Sigma}_{\widehat{R}_T}$. Then as N, $T_1(N) \to \infty$, N, $T_2(N) \to \infty$ and $N/T_1 \to c_1 < 1$, $N/T_2 \to c > 0$, we have*

$$\hat{\lambda}_{j,T} \xrightarrow{\text{a.s.}} \begin{cases} \lambda_j \left(1 - c - c\frac{\sqrt{(c_1-1)^2\lambda_j^2 - 2(c_1+1)\lambda_j + 1} - (c_1+1)\lambda_j + 1}{2c_1\lambda_j}\right) & \text{if } \lambda_j > \tau(c,c_1), \\ b_2 \text{ if } \lambda_j \leq \tau(c,c_1). \end{cases}$$

for $j = 0, \ldots, K-1$ and $\tau(c,c_1)$ is defined by

$$\tau(c,c_1) = \frac{(cc_1 - c - 2c_1)\sqrt{c + c_1 - c_1 c} + cc_1^2 - c - c_1 - c_1^2}{((c_1-1)c - c_1)(c_1-1)^2}.$$

From this theorem, the behavior of the jth largest eigenvalue of $\widehat{\Sigma}_{\widehat{R}_T}$ denoted by $\hat{\lambda}_{j,T}$ and corresponding to the signal is similar to that presented previously. When λ_j is larger then the detectability threshold given by $\tau(c,c_1)$ then $\hat{\lambda}_{j,T}$ converges to the limit outside the interval $[b_1, b_2]$. It is known that the fluctuations of $\hat{\lambda}_{j,T}$ follow the Gaussian law (see [10], [26]), the exact expression of the asymptotic variance was derived in [49]. When λ_j is below the detectability limit then $\hat{\lambda}_{j,T}$ converges to b_2 and fluctuates as a Tracy–Widom variable [10], [26]. An heuristic definition of the effective number of identifiable signals K_{eff} is given by the number of the eigenvalues of Σ_{R_T} greater than $\tau(N/T_2, N/T_1)$. If not all the signal eigenvalues satisfy the condition of detectability then underestimation of the number of signals is unavoidable. Similarly to the method of [40], the test statistic is based on the kth largest eigenvalue of $\widehat{\Sigma}_{\widehat{R}_T}$ and performs the multiple hypothesis testing as defined by (3.8). For $k = 0, \ldots, \min(N, T_2) - 1$, the hypothesis H_1, meaning that at least $k+1$ signals are present, is accepted for large values of the test:

$$\hat{\lambda}_{k,T} \underset{H_1}{\overset{H_0}{\lessgtr}} \chi_T$$

where the detection threshold χ_T is approximated by

$$\chi_T \simeq b_2 + \frac{\bar{F}_{TW}^{-1}(\alpha)\sigma_T'}{N^{2/3}}$$

with the scaling parameter σ_T' different from σ_T defined by (3.7)(see [47] for the exact expression). The first value for which $\hat{\lambda}_{k,T} < \chi_T$ gives the estimate for the number of signal $\widehat{K}_{\text{NS}}$.

When the noise covariance is unknown it was assumed in the above approaches that a pure noise sequence is available in order to estimate the noise covariance matrix. However, in practical systems such assumption is not realistic, especially, in large dimensional regime. Indeed, usually there is not enough time to sense infinitely large number of the noise samples. Therefore, the detection has to be performed directly from the observed signals $y_0, \ldots, y_{T-1}$, *i.e.* without assumption that a pure-noise sequence is available. This is called block detection approach.

Block detection approach

In this thesis we do not make assumption that a pure-noise is available. We start by recalling the transmission model for which it is only assumed that the noise is white in space but correlated in time. The system model is written in the following matrix form

$$Y_T = H_T S_T^{\mathsf{H}} + W_T R_T^{1/2} \tag{3.14}$$

where H_T and S_T are defined as in (3.3), W_T has i.i.d. zero mean complex Gaussian entries with unit variance, and $R_T \in \mathbb{C}^{T\times T}$ is an unknown covariance matrix. The model (3.14) corresponds to the additive spiked model. Note that if the spectral norm of R_T is bounded, then with probability one there is no eigenvalue of $\frac{1}{T}W_T R_T W_T^{\mathsf{H}}$ outside the support of its limiting eigenvalue distribution and the same discussions as in Section 3.2.1 are applicable.

Under the hypothesis where only signal is present, the work of [26] characterizes the limiting behavior of the largest eigenvalue of the empirical noise covariance matrix $\frac{1}{T}V_T V_T^{\mathsf{H}}$. We recall that from Theorem 2.2.4 (see Chapter 2), the centered and scaled largest eigenvalue of $\frac{1}{T}V_T V_T^{\mathsf{H}}$ has the Tracy–Widom distribution:

$$N^{2/3}\frac{\hat{\lambda}_{0,T} - b'_T}{\sigma'_T} \xrightarrow[T\to\infty]{\mathcal{L}} \text{TW}.$$

where b'_T and σ'_T are the centering and the scaling parameters depending not only on the limiting ratio c (as in the white noise case) but also on the spectral measure ν_T of R_T. If R_T is known then the statistics of the largest eigenvalue of the noise covariance matrix can be fully described. Based on this property, the detection threshold can be expressed as a function of the probability of false alarm. However, when R_T is assumed to be unknown the expression of the detection threshold is unknown. It should be noted

that one can imagine to estimate the upperbound of the limiting support corresponding to the noise from the empirical spectrum. However, this is a challenging problem. In this thesis we propose a detection algorithm based on the spectrum of empirical eigenvalues only.

3.3 Direction-of-arrival estimation

In this section we assume that the number of sources K is known. Recall the transmission model

$$Y = H(\Theta)S^{\mathsf{H}} + V$$

where the channel matrix is equal now to $H(\Theta) = [h(\theta_0), \ldots, h(\theta_{K-1})] \in \mathbb{C}^{N\times K}$ with the steering vectors $h(\theta_k) = N^{-1/2}[1\ e^{i\theta_k} \ldots e^{i\theta_k(N-1)}] \in \mathbb{C}^{N\times 1}$, the columns of S^{H} and V are zero mean complex Gaussian vectors, independent of each other with respective covariance matrices $\Gamma \in C^{K\times K}$ and $R = \sigma^2 I_N$ where σ^2 is unknown. The problem is to estimate the directions-of-arrival $\theta_0, \ldots, \theta_{K-1}$ from the sample covariance matrix. The most famous approach is a subspace-based method proposed in [57] and called MUSIC (MUltiple SIgnal Classification) algorithm. It is based on the structure of the covariance matrix of the columns of Y given by

$$\Sigma = H(\Theta)\Gamma H(\Theta)^{\mathsf{H}} + \sigma^2 I_N.$$

Let $u_0, \ldots, u_{N-1}$ be the eigenvectors of Σ corresponding respectively to the decreasingly ordered eigenvalues $\lambda_0 \geq \ldots \geq \lambda_{N-1}$. Let $\Lambda_S = \mathrm{diag}(\lambda_0, \ldots, \lambda_{K-1})$ and $\Lambda_N = \mathrm{diag}(\lambda_K, \ldots, \lambda_{N-1}) = \sigma^2 I_{N-K}$ be the diagonal matrices containing the signal and the noise eigenvalues, respectively. Denoting by $E_S = [u_0, \ldots, u_{K-1}]$ and $E_N = [u_K, \ldots, u_{N-1}]$ the signal and the noise spaces respectively, the covariance matrix Σ can be rewritten as

$$\Sigma = \begin{bmatrix} E_S & E_N \end{bmatrix} \begin{bmatrix} \Lambda_S & 0_{K\times(N-K)} \\ 0_{(N-K)\times K} & \sigma^2 I_N \end{bmatrix} \begin{bmatrix} E_S^{\mathsf{H}} \\ E_N^{\mathsf{H}} \end{bmatrix}.$$

Define the orthogonal projector matrix on the signal subspace generated by the vectors $h(\theta_0), \ldots, h(\theta_{K-1})$ by:

$$\Pi = E_S E_S^{\mathsf{H}} = \sum_{k=0}^{K-1} u_k u_k^{\mathsf{H}}.$$

The MUSIC algorithm is based on the orthogonality of the noise and signal subspaces. Hence, the vectors $h(\theta_k)$ lying in the signal subspace are

orthogonal to the null space of Π and for any $0 \leq k \leq K-1$ we have:

$$h(\theta_k)^{\mathsf{H}}(I_N - \Pi)h(\theta_k) = 0.$$

The arguments of the local maxima of the localization function

$$\gamma(\theta) \triangleq h(\theta)^{\mathsf{H}}\Pi h(\theta) \tag{3.15}$$

correspond to the angles-of-arrival $\theta_0, \ldots, \theta_{K-1}$. In practice we do not have access to the signal eigenvectors $u_0, \ldots, u_{K-1}$ and they are estimated from the sample covariance matrix $\widehat{\Sigma} = \frac{1}{T}YY^{\mathsf{H}}$. Denote by $\hat{u}_0, \ldots, \hat{u}_{N-1}$ the eigenvectors of $\widehat{\Sigma}$ corresponding to the eigenvalues $\hat{\lambda}_0 \geq \ldots \geq \hat{\lambda}_{N-1}$. The localization function is estimated by

$$\widehat{\gamma}(\theta) = h(\theta)^{\mathsf{H}}\widehat{\Pi}h(\theta)$$

where $\widehat{\Pi} = \sum_{k=0}^{K-1} \hat{u}_k\hat{u}_k^{\mathsf{H}}$ is the orthogonal projection matrix on the signal subspace of $\widehat{\Sigma}$.

In large dimensional setting the sample covariance matrix is not a consistent estimate of the true one and, therefore, this conducts to a non accurate eigenvector estimates. It was shown by Mestre and Lagunas in [45] that the classical MUSIC algorithm is not consistent in this regime and presents an asymptotic bias. A so-called G-MUSIC algorithm is provided in [45] making assumption that the observations of the received signal are random multivariate Gaussian variables i.i.d. in the time domain. The method of Vallet *et al.* in [64] deals with a deterministic signal part and the model is referred to as the signal-plus-noise model. We are interested in a different approach based on the results of [11] and proposed by Hachem *et al.* in [32]. It was shown in [64] that as N and $T \to \infty$, the estimator proposed in [32] converges almost surely to that of [64].

We present now the result of [32], the proof of which is based on the convergence of some bilinear forms and with assumption that K is fixed when $T \to \infty$, $N/T \to c > 0$. It is assumed that the signal matrix S_T is deterministic such that $\frac{1}{T}S_T^{\mathsf{H}}S_T$ converges to a diagonal matrix. The noise matrix V_T is assumed to be bi-unitarily invariant. Let μ_T be the spectral measure of $\frac{1}{T}V_TV_T^{\mathsf{H}}$ with Stieltjes transform m_T converging almost surely to the limit measure μ with the Stieltjes transform m. Let $[a, b]$ be the support of μ. Recall that $\widetilde{m}$ is the Stieltjes transform of the limiting measure of $\frac{1}{T}V_T^{\mathsf{H}}V_T$. Define the function $g(x) = xm(x)\widetilde{m}(x)$ which is decreasing on (b, ∞) and denote by $g(b^+) = \lim_{x \to b^+} g(x)$. The following theorem provides an estimator of the localization function $\gamma(\theta)$ defined by (3.15).

Theorem 19 ([32]). *Let $q \in \mathbb{N}$ be the largest integer for which*

$$\omega_q > 1/g(b^+).$$

Denote $\hat{u}_{0,T}, \ldots, \hat{u}_{q-1,T}$ the eigenvectors of $\frac{1}{T} Y_T Y_T^{\mathsf{H}}$ belonging respectively to $\hat{\lambda}_{0,T}, \ldots, \hat{\lambda}_{q-1,T}$. For $\theta \in [-\pi/2, \pi/2]$, let

$$\hat{\gamma}_T(\theta) = \sum_{k=0}^{q-1} \zeta(\hat{\lambda}_{k,T}) h_T(\theta)^{\mathsf{H}} \hat{u}_{k,T} \hat{u}_{k,T}^{\mathsf{H}} h_T(\theta)$$

where

$$\zeta(x) = \frac{(xm(x)\widetilde{m}(x))'}{xm(x)^2 \widetilde{m}(x)}.$$

Then,

$$\gamma(\theta) - \hat{\gamma}_T(\theta) \xrightarrow[T\to\infty]{a.s.} 0.$$

Note that this theorem allows to estimate the angles-of-arrival of the sources whose power is larger than the detectability threshold $1/g(b^+)$ depending on the limiting ratio c and m. When the entries of the matrix V_T are i.i.d. with zero mean and unit variance, then the limiting measure μ is the Marčenko–Pastur law. For other noise models such that V_T is unitarily bi-invariant, this method requires the knowledge of the noise statistics.

Chapter 4

Detection/estimation of a small rank signal in the presence of correlated noise

4.1 Introduction

In this chapter we consider the general signal-plus-noise model

$$Y_T = A_T + W_T R_T^{1/2} \tag{4.1}$$

where A_T is a random matrix of fixed rank K representing the number of signals, $V_T = W_T R_T^{1/2}$ is the time correlated noise where $W_T \in \mathbb{C}^{N\times T}$ has i.i.d. standard Gaussian entries, and R_T is a Hermitian nonnegative matrix with a bounded spectral norm. Remark that, up to studying Y_T^{H} instead of Y_T, the noise correlation can be either in time or in space. This model is related to the spiked models introduced in Chapter 2, Section 2.3. Define the sample covariance matrix by

$$\widehat{\Sigma}_T = \frac{1}{T} Y_T Y_T^{\mathsf{H}}. \tag{4.2}$$

The results of this chapter are based on the study of the K largest eigenvalues of $\widehat{\Sigma}_T$ under a crucial assumption involving the relationship between the noise and signal eigenspaces. It assumes in some sense that the signal subspace has an "isotropic" behavior towards the noise subspace.

We perform statistical inference on the information part A_T and provide an estimator of the number of signals K, the powers of these signals, and some bilinear forms involving the orthogonal projection matrix on the signal

subspace. The results are used to perform detection/estimation in a concrete application example of narrow band array processing. In this scenario, showing that the model (4.1) satisfies the set of assumptions, estimators of the number of emitting sources, their powers, and the MUSIC-based direction-of-arrival are provided. The fluctuations of the power estimates are also studied.

4.2 System model and assumptions

Considering a sequence of integers $N = N(T)$, $T = 1, 2, \ldots$ and matrices

$$Y_T = A_T + W_T R_T^{1/2} \tag{4.3}$$

we assume the following asymptotic regime:

Assumption 3. *As* $T \to \infty$, $c_T \triangleq N/T \to c > 0$.

In the following we present first the hypothesis on the noise part then on the signal part.

4.2.1 Hypotheses on the noise matrix

Assumptions and some known results

We first characterize the assumptions on the noise matrix $V_T \triangleq W_T R_T^{1/2}$.

Assumption 4. $W_T = [w_{n,t}]_{n,t=0}^{N-1,T-1}$, *with* $(w_{n,t})_{n,t\geq 1}$ *an infinite array of independent* $\mathcal{CN}(0,1)$ *variables.*

Assumption 5. $R_T \in \mathbb{C}^{T\times T}$ *is Hermitian nonnegative with eigenvalues* $\sigma_{0,T}^2, \ldots, \sigma_{T-1,T}^2$ *satisfying:*

(i) $\nu_T = T^{-1} \sum_{t=0}^{T-1} \boldsymbol{\delta}_{\sigma_{t,T}^2} \xrightarrow{\mathcal{L}} \nu$, *a probability measure with support* $\operatorname{supp}(\nu) = [a_\nu, b_\nu] \subset \mathbb{R}_+ \triangleq [0, \infty)$. *Moreover,* $\nu(\{0\}) = 0$.

(ii) The distances from the $\sigma_{t,T}^2$ *to* $\operatorname{supp}(\nu)$ *satisfy:*

$$\max_{t\in\{1,\ldots,T-1\}} \mathbf{d}\left(\sigma_{t,T}^2, \operatorname{supp}(\nu)\right) \xrightarrow[T\to\infty]{} 0.$$

As a consequence of this assumption the matrix R_T has a bounded spectral norm and asymptotically there is no spike due to R_T that can be found outside the limiting support $[a_\nu, b_\nu]$.

The characterization of the limiting spectrum behavior of the noise sample covariance matrix $\frac{1}{T}V_T V_T^{\mathsf{H}} = \frac{1}{T}W_T R_T W_T^{\mathsf{H}}$ is of prime importance. Denote by $\hat{\lambda}_{0,T} \geq \ldots \geq \lambda_{N-1,T}$ the eigenvalues of $\frac{1}{T}V_T V_T^{\mathsf{H}}$ and let $\tau_T = N^{-1}\sum_{i=0}^{N-1} \boldsymbol{\delta}_{\lambda_{i,T}}$ be its spectral measure. The asymptotic characterization of τ_T will be reviewed in the following. We recall some well known results describing this behavior which had been already presented in Chapter 2, Section 2.2.2: see [44, 59] for Items 1)–6), [60] for Item 4), and [4] for Item 5).

Theorem 20. *Under Assumptions 3–5, the following hold true:*

1. *For any $z \in \mathbb{C}_+ \triangleq \{z \in \mathbb{C},\, \Im z > 0\}$, the equation*

$$m = \left(-z + \int \frac{t}{1+cmt}\nu(dt)\right)^{-1} \tag{4.4}$$

   has a unique solution $m \in \mathbb{C}_+$. The function $m(z) = m$ so defined on $\mathbb{C}_+$ is the Stieltjes transform of a probability measure μ.

2. *For every bounded and continuous real function f,*

$$\int f(t)\tau_T(dt) \xrightarrow[T\to\infty]{a.s.} \int f(t)\mu(dt)$$

   and therefore μ, defined by (4.4), is the limiting spectral measure of $\frac{1}{T}V_T V_T^{\mathsf{H}}$.

3. *The function*

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

   is defined on $\mathbb{C}_+$ and is the Stieltjes transform of the probability measure $\tilde{\mu} = c\mu + (1-c)\delta_0$, limiting spectral measure of $\frac{1}{T}V_T^{\mathsf{H}}V_T$. As such, $\tilde{m}(z) = cm(z) - (1-c)/z$.

4. *μ is of the form $\mu(dt) = \max(0, 1-c^{-1})\delta_0 + f(t)dt$ where $f(t)$ is a continuous density on $(0,\infty)$. The support of $f(t)dt$ is a compact interval $[a,b] \subset \mathbb{R}_+$, and $f(t) > 0$ on (a,b).*

5. *For any interval $[x_1, x_2] \subset (0,a) \cup (b,\infty)$,*

$$\sharp\{i : \lambda_{i,T} \in [x_1, x_2]\} = 0 \text{ w.p. } 1 \text{ for all large } T.$$

6. *The function* $\underline{m}_T(x) = N^{-1}\sum_{n=0}^{N-1}(\lambda_{n,T} - x)^{-1}$ *converges w.p. 1 to* $m(x)$*, and uniformly so on the compact subsets of* (b, ∞).

From this theorem it is particularly clear that asymptotically with probability one there is no eigenvalue of $\frac{1}{T}V_T V_T^{\mathsf{H}}$ outside the compact support $[a, b]$ of the limiting spectral measure μ.

Characterization of the right-edge of μ

We are interested here to determine the limiting position of the largest eigenvalue of $\frac{1}{T}V_T V_T^{\mathsf{H}}$ which corresponds to the upper bound b of the support of μ to which $\lambda_{0,T}$ converges. A procedure for determining the interval $[a, b]$ from the knowledge of c and ν is provided in [60] and was introduced in Chapter 2, Section 2.2.3. We recall here the main result which is presented in the following proposition. Observe that $m(z)$ can be extended to $\mathbb{C} - (\{0\} \cup [a, b])$ and that $m(x) = \int (t - x)^{-1}\mu(dt)$, its restriction to $\mathbb{R}$, is negative and increases to zero on (b, ∞). Recall that $\mathrm{supp}(\nu) = [a_\nu, b_\nu] \subset \mathbb{R}_+$.

Proposition 7 (see [60])**.** *The point* b *defined in Theorem 20-4) coincides with the infimum of the function*

$$x(m) = -\frac{1}{m} + \int \frac{t}{1 + cmt}\,\nu(dt)$$

on the interval $(-(cb_\nu)^{-1}, 0)$*. On this interval, there is a unique* m_b $(m_b < 0)$ *such that* $x(m) \to b$ *as* $m \downarrow m_b$*. The restriction of* $x(m)$ *to* $(m_b, 0)$ *coincides with the inverse with respect to composition of the restriction of* $m(x)$ *to* (b, ∞).

From Proposition 7, the edge b is known to exist but does not have a mathematical characterization as a function of ν. In order to easily characterize the value of b, it is necessary to make an assumption on the measure ν. Note that a similar assumption was made by El Karoui [26] to characterize the fluctuations of the largest eigenvalues of $\frac{1}{T}W_T R_T W_T^{\mathsf{H}}$. We will get back to this point in Section 4.4.3.

Assumption 6. *If* $\nu(\{b_\nu\}) = 0$*, then there exists* $\varepsilon > 0$ *and a function* $f_\nu(t) \geq C(b_\nu - t)$ *on* $[b_\nu - \varepsilon, b_\nu]$ *with* $C > 0$ *such that for any Borel set* A *of* $[a_\nu, b_\nu]$*,*

$$\nu(A \cap [b_\nu - \varepsilon, b_\nu]) = \int_{A \cap [b_\nu - \varepsilon, b_\nu]} f_\nu(t)\,dt.$$

This assumption is not restrictive in practice and says that ν either has a mass or a sufficiently sharp density edge at b_ν. Especially, it allows to have an explicit definition for the value b and leads to the following corollary to Proposition 7, proven in Appendix 4.7.1:

Corollary 1. *Under Assumption 6,*

$$b = -\frac{1}{m_b} + \int \frac{t}{1 + cm_b t} \nu(dt)$$

where m_b is the unique solution in $(-(cb_\nu)^{-1}, 0)$ to the equation in m

$$\int \left(\frac{mt}{1+cmt} \right)^2 \nu(dt) = \frac{1}{c}. \tag{4.5}$$

We will see later that Assumption 6 is also important to determine the second order behavior of the estimators close to the signal detectability limit (*cf.* Section 4.4.3).

4.2.2 Hypotheses on the signal matrix

We need now to make two hypotheses on the signal matrix A_T. The first one is on its spectral norm boundness.

Assumption 7. *Let $K \geq 0$ be a fixed integer. The matrix $A_T \in \mathbb{C}^{N \times T}$ is random, independent of W_T, with rank* $\operatorname{rank}(A_T) = K$ *w.p. 1 for all large T. Besides,* $\sup_T \|A_T\|/T < \infty$ *w.p. 1.*

In the following, when $K \leq \min(N, T)$, the notation $A_T = U_T B_T^{\mathsf{H}}$ refers to any factorization of A_T where $U_T \in \mathbb{C}^{N \times K}$ satisfies $U_T^{\mathsf{H}} U_T = I_K$. By Assumption 7, the rank of $B_T \in \mathbb{C}^{T \times K}$ is equal to K, w.p. 1.

The second hypothesis is a fundamental assumption on the relationship between the signal and noise matrices:

Assumption 8. *There exists a factorization $A_T = U_T B_T^{\mathsf{H}}$ such that, for any $z \in \mathbb{C} - \operatorname{supp}(\nu)$,*

$$\frac{1}{T} B_T^{\mathsf{H}} \left(R_T - z I_T \right)^{-1} B_T \xrightarrow[T \to \infty]{a.s.} m_\nu(z) P \tag{4.6}$$

for some

$$P = \begin{bmatrix} p_0 I_{j_0} & & \\ & \ddots & \\ & & p_t I_{j_t} \end{bmatrix}, \; p_0 > \ldots > p_t, \; j_0 + \ldots + j_t = K$$

and where it is recalled that $m_\nu(z)$ is the Stieltjes transform of the probability measure ν.

From this assumption the signal right singular vectors are in some sense isotropic in the basis of eigenvectors of R_T. When a signal is present and is powerful enough, isolated eigenvalues appear at the right of the bulk of the eigenvalues of the sample covariance matrix $\widehat{\Sigma}_T$. With this assumption, the asymptotic locations of the isolated eigenvalues of $\widehat{\Sigma}_T$ depend only on the eigenvalue distribution of R_T and on the limiting ratio c.

Assumption 8 is quite strong in general. However, it holds in many practical situations. Generally, the signal matrix is written as $A_T = H_T S_T^{\mathsf{H}}$ where $H_T \in \mathbb{C}^{N\times K}$ is the channel and $S_T \in \mathbb{C}^{T\times K}$ is the matrix of transmitted signals. In the case of a temporally correlated noise, *i.e.* under the model of the type $Y_T = H_T P^{1/2} S_T^{\mathsf{H}} + W_T R_T^{1/2}$ this assumption holds if the signals S_T are isotropic, particularly, when they are right-unitarily invariant (meaning that S_T can be correlated in space but not in time). In the case of the model with spatially correlated noise given by $Y_T = H_T P^{1/2} S_T^{\mathsf{H}} + R_T^{1/2} W_T$, the assumption holds if H_T is left-unitarily invariant (*e.g.* for Gaussian fading channels) and S_T is isotropic or correlated in time, up to transposition of Y_T. However, this assumption is not valid if H_T contains steering angles (not left-unitarily invariant) and the noise is spatially correlated. Generally speaking Assumption 8 does not hold if the noise and the signal simultaneously time/space correlated.

We give two practical examples for which this assumption holds:

(i) **Array Processing**: Let $A_T = H_T P^{1/2} S_T^{\mathsf{H}}$, with $H_T = [h(\theta_0), \cdots, h(\theta_{K-1})]$ (θ_k distinct) the matrix of steering vectors, $P = \mathrm{diag}(a_0^2, \ldots, a_{K-1}^2)$ the source powers, $S_T \in \mathbb{C}^{T\times K}$ is the signal matrix with i.i.d. of zero mean and unit variance entries. Let $V_T = W_T R_T^{1/2}$ and $[\sqrt{N} h(\theta)]_n = e^{-2\pi \imath n \sin(\theta)}$. Writing $A_T = U_T B_T^{\mathsf{H}}$ with $U_T = H_T (H_T H_T^{\mathsf{H}})^{-1/2}$ and $B_T = S_T P^{1/2} (H_T H_T^{\mathsf{H}})^{1/2}$, we can show $(H_T H_T^{\mathsf{H}})^{-1/2} \to I_K$ while $\frac{1}{T} S_T^{\mathsf{H}} (R_T - z I_T)^{-1} S_T \xrightarrow{\text{a.s.}} m_\nu(z) I_K$ so that Assumption 8 holds. See the proof of Lemma 1 for details.

(ii) **MIMO Communications**: Let $A_T = H_T P^{1/2} S_T^{\mathsf{H}}$, with $H_T = 1/\sqrt{N}[h_0, \ldots, h_{K-1}]$ the wireless channels (i.i.d. zero mean unit variance entries) of K transmitters, P their diagonal power matrix and S_T their matrix of transmitted (i.i.d. zero mean unit variance) signals. Taking $V_T = R_T^{1/2} W_T$, *i.e.* spatially correlated noise, and considering Y_T^{H} instead of Y_T, we may write $A_T^{\mathsf{H}} = U_T B_T^{\mathsf{H}}$ with $U_T = S_T (S_T S_T^{\mathsf{H}})^{-1/2}$

and $B_T = H_T P^{1/2}(S_T S_T^{\mathsf{H}})^{1/2}$ to obtain $\frac{1}{T} B_T^{\mathsf{H}}(R_T - zI_N)^{-1} B_T \xrightarrow{\text{a.s.}} m_\nu(z)P$.

We provide first the theoretical results in the following section. The first example is then used for the application part in Section 4.4 for which the signals are assumed to be uncorrelated and the noise is an ARMA process.

4.3 Results on the information-plus-noise matrix

4.3.1 Preliminary results

The idea is to detect the number of signals and to estimate their powers by studying the locations of the isolated eigenvalues of $\widehat{\Sigma}_T$. Since $\widehat{\Sigma}_T$ is at most a rank $2K$ perturbation of $\frac{1}{T} V_T V_T^{\mathsf{H}}$ with K fixed, Weyl's interlacing inequalities [33, Th. 4.3.6] show, in conjunction with Theorem 20, that the spectral measure of $\widehat{\Sigma}_T$ also converges to μ in the sense of Theorem 20-2). However, a finite number of eigenvalues of $\widehat{\Sigma}_T$ might stay isolated away from the support of μ [21, Th. 2.2]. We need first to recall the following function, for any $x \in (b, \infty)$:

$$g(x) \triangleq x m(x) \tilde{m}(x) \tag{4.7}$$

which is positive and decreases from $g(b^+)$ to zero on (b, ∞).

The following theorem gives an equivalent definition of the detectability condition $p_k g(b^+) > 1$:

Theorem 21. *The detectability condition $p_k g(b^+) > 1$ can be equivalently defined by the following theorem*

$$p_k > \left(\int \frac{-m_b}{1 + c m_b t} \nu(dt) \right)^{-1} \tag{4.8}$$

with m_b the solution in $(-(cb_\nu)^{-1}, 0)$ to Equation (4.5).

Proof. Observe that $g(x) = -\int m(x)(1+cm(x)t)^{-1}\nu(dt)$ from the definition of $\tilde{m}$ in Theorem 20-3) and recall that $m(x) \downarrow m_b$ as $x \downarrow b$. □

From this theorem the detectability condition actually depends on the spectral measure ν and the limiting ratio c. Note that in the white noise setting, *i.e.*, $R_T = I_T$ (hence, $\nu = \delta_1$), μ is the famous Marchenko-Pastur law, and Equation (4.8) boils down to $p_k > \sqrt{c}$ (see, *e.g.*, [32]). Most of source detection in white noise environment approaches rely on this condition (see Chapter 3).

The behavior of the largest eigenvalues of $\widehat{\Sigma}_T$ is described by the following theorem:

Theorem 22 ([21]). *Under Assumptions 3–8, let μ and $[a,b]$ be as in Theorem 20. Let $\hat{\lambda}_{0,T} \geq \cdots \geq \hat{\lambda}_{N-1,T}$ be the eigenvalues of $\widehat{\Sigma}_T$ with spectral measure $\hat{\tau}_T = N^{-1}\sum_{i=0}^{N-1}\hat{\lambda}_{i,T}$. Then:*

1. *For every bounded and continuous real function f,*

$$\int f(t)\hat{\tau}_T(dt) \xrightarrow[T\to\infty]{a.s.} \int f(t)\mu(dt).$$

2. *For any interval $[x_1,x_2] \subset (0,a)$*

$$\sharp\{i : \hat{\lambda}_{i,T} \in [x_1,x_2]\} = 0 \text{ w.p. 1 for all large } T.$$

3. *If $p_0 g(b^+) \leq 1$, then $\hat{\lambda}_{0,T} \xrightarrow{\text{a.s.}} b$. Otherwise, let $s \in \{0,\ldots,t\}$ be the largest index for which $p_s g(b^+) > 1$ where t is defined in Assumption 8. For $k = 0,\ldots,s$, let ρ_k be the unique solution x in (b,∞) of $p_k g(x) = 1$. Then, for $i = 0,\ldots,s$,*

$$\hat{\lambda}_{j_0+\cdots+j_{i-1}+1,T},\ldots,\hat{\lambda}_{j_0+\cdots+j_i,T} \xrightarrow[T\to\infty]{a.s.} \rho_i$$
$$\hat{\lambda}_{j_0+\cdots+j_s+1,T} \xrightarrow[T\to\infty]{a.s.} b.$$

This theorem shows in particular that the number of isolated eigenvalues of $\widehat{\Sigma}_T$ is upper bounded by the rank K of A_T and it reaches this rank if p_t is large enough.

4.3.2 Signal detection

Estimation of the signal dimension in large dimensional regime is closely related to the study of the isolated eigenvalues in spiked models. Recall that if p_k satisfies the detectability condition given by (4.8), then the corresponding isolated eigenvalue converges to a limit outside the support of μ. Otherwise, it converges to b. Based on this we need now to establish a test that allows us to estimate the dimension of the signal from the sample covariance eigenvalues only. Recall from Chapter 3 that in the case of single signal and a white noise with unknown variance, the sub-optimal detector is the GLRT consisting in testing the ratio of the largest eigenvalue of $\widehat{\Sigma}_T$ to its normalized trace [14]. However, when the noise is not white this test cannot be used anymore. We propose rather to test L successive close eigenvalue ratios where L is an upper bound to the number of signals which is a common hypothesis. This is summarized in the following theorem:

Theorem 23. *Under Assumptions 3–8, let $s \geq -1$ be the largest integer for which Equation (4.8) holds. Let $0 < \varepsilon < (\rho_s/b) - 1$ with $\rho_{-1} = \infty$. Given $L \geq K$, define*

$$\hat{k}_T = \arg \max_{k \in \{0,\ldots,L-1\}} \frac{\hat{\lambda}_{k-1,T}}{\hat{\lambda}_{k,T}} > 1 + \varepsilon$$

with $\hat{\lambda}_{-1,T} = \infty$. Then, for all T large, w.p. 1,

$$\hat{k}_T = j_0 + \ldots + j_s \quad .$$

Proof. The result is clear for $s = 0$. Else, writing $k = j_0 + \ldots + j_s$, Items 1) and 3) of Theorem 22 ensure $\hat{\lambda}_{k,T} \xrightarrow{\text{a.s.}} \rho_s > b$ and $\hat{\lambda}_{\ell,T} \xrightarrow{\text{a.s.}} b$ for $\ell = k+1, \ldots, L$. □

The main problem in this detection technique is choosing the detection threshold. Generally, when the noise is white, it is known that under H_0, the centered and scaled largest eigenvalue of $\widehat{\Sigma}_T$ converges to the Tracy–Widom law. The detection threshold can be then fixed theoretically from this law and depends on the probability of false alarm. When the noise is not white, under H_0, it was also shown in [26] that the centered and scaled largest eigenvalue of $\widehat{\Sigma}_T$ converges to the Tracy–Widom law with the scaling and centering factors depending on ν and c. When R_T is unknown, it is not theoretically possible to set the exact threshold and it is set heuristically. Recall that the spacing between the noise eigenvalues is of order $\mathcal{O}(1/N)$ (see *e.g.* [52]) and hence, $1 + \varepsilon$ has to be greater than $\mathcal{O}(1/N)$. However, ε has to be taken such that $\varepsilon < (\rho_s/b) - 1$. Hence, for all large N, one may take ε such that $\varepsilon \to 0$ and $N\varepsilon \to \infty$ as $N \to \infty$.

4.3.3 Parameter estimation

Let now $A_T = U_T B_T^{\mathsf{H}}$ as in Assumption 8. Then $\frac{1}{T} B_T^{\mathsf{H}} B_T \xrightarrow{\text{a.s.}} P$ by multiplying each side of (4.6) by $-z$ and taking z large. Therefore, $p_0, \ldots, p_t$ are the limiting positive eigenvalues of $\frac{1}{T} A_T A_T^{\mathsf{H}}$. We propose now to estimate consistently $p_0, \ldots, p_s$ which will represent the power of the sources in the application section. In the following, for $i \in \{0, \ldots, K-1\}$, let $\mathcal{K}(i) = 0$ if $0 \leq i \leq j_0$, $\mathcal{K}(i) = 1$ if $j_0 + 1 \leq i \leq j_0 + j_1$, ..., $\mathcal{K}(i) = t$ if $j_0 + \cdots + j_{t-1} + 1 \leq i \leq K - 1$. We have the following theorem:

Theorem 24. *In the setting of Theorem 23, let*

$$\begin{aligned}\hat{m}_T(x) &\triangleq \frac{1}{N-\hat{k}_T}\sum_{n=\hat{k}_T}^{N-1}\frac{1}{\hat{\lambda}_{n,T}-x}\\ \hat{g}_T(x) &\triangleq \hat{m}_T(x)(xc_T\hat{m}_T(x)+c_T-1)\\ \hat{p}_{i,T} &\triangleq \frac{1}{\hat{g}_T(\hat{\lambda}_{i,T})},\ i=0,\ldots,\hat{k}_T-1.\end{aligned}$$

Then

$$\hat{p}_{i,T}-p_{\mathcal{K}(i)} \xrightarrow[T\to\infty]{a.s.} 0.$$

Proof. Recall that $\lambda_{0,T}\geq\ldots\geq\lambda_{N-1,T}$ are the eigenvalues of $\frac{1}{T}W_TR_TW_T^{\mathsf{H}}$. In the proof, we restrict the elementary events to belong to the probability one set where $\lambda_{0,T}\to b$, $\underline{m}_T(x)\to m(x)$ uniformly on the compact subsets of (b,∞) (see Theorem 20-6)), $\hat{\lambda}_{i,T}\to\rho_{\mathcal{K}(i)}$ for $i=0,\ldots,j_0+\cdots+j_s$, $\hat{\lambda}_{j_0+\cdots+j_s+1,T}\to b$, and $\hat{k}_T\to j_0+\cdots+j_s$ (Theorems 20–23). Observe that $\widehat{\Sigma}_T$ is at most a (nonnegative) rank $2K$ perturbation of $\frac{1}{T}V_TV_T^{\mathsf{H}}$. In these conditions, Weyl's inequalities [33, Th. 4.3.6] ensure $\hat{\lambda}_{n,T}\leq\lambda_{n-2K,T}$ and $\lambda_{n,T}\leq\hat{\lambda}_{n-2K,T}$ for $n=2K,\ldots,N-1$. Then, for any $x>b$ and T large,

$$\begin{aligned}\hat{m}_T(x) &= \frac{1}{N-\hat{k}_T}\left(\sum_{n=2K}^{N-1}\frac{1}{\hat{\lambda}_{n,T}-x}+\sum_{n=\hat{k}_T}^{2K-1}\frac{1}{\hat{\lambda}_{n,T}-x}\right)\\ &\geq \frac{1}{N-\hat{k}_T}\left(\sum_{n=0}^{N-1-2K}\frac{1}{\lambda_{n,T}-x}+\sum_{n=\hat{k}_T}^{2K-1}\frac{1}{\hat{\lambda}_{n,T}-x}\right)\\ &\triangleq \underline{m}_T(x)+e_T(x)\end{aligned}$$

with $e_T(x)\to 0$ uniformly on compact sets of (b,∞). Similarly we have:

$$\begin{aligned}\hat{m}_T(x) &= \frac{1}{N-\hat{k}_T}\left(\sum_{n=\hat{k}_T}^{N-2K-1}\frac{1}{\hat{\lambda}_{n,T}-x}+\sum_{n=N-2K}^{N-1}\frac{1}{\hat{\lambda}_{n,T}-x}\right)\\ &\leq \frac{1}{N-\hat{k}_T}\left(\sum_{n=\hat{k}_T+2K}^{N-1}\frac{1}{\lambda_{n,T}-x}+\sum_{n=N-2K}^{N-1}\frac{1}{\hat{\lambda}_{n,T}-x}\right)\\ &\triangleq \underline{m}_T(x)+e'_T(x)\end{aligned}$$

with $e'_T(x) \to 0$ uniformly on compact sets of (b, ∞). Hence, we have $\hat{m}_T(x) - m(x) \to 0$ uniformly on compact sets of (b, ∞). Recalling that $g(x) = xm(x)\tilde{m}(x)$ on (b, ∞), we obtain $\hat{g}_T(\hat{\lambda}_{i,T}) - g(\hat{\lambda}_{i,T}) \to 0$ for $i = 0, \ldots, \hat{k}_T - 1$. Clearly, $g(\hat{\lambda}_{i,T}) - g(\rho_{\mathcal{K}(i)}) \to 0$ so that $\hat{g}_T(\hat{\lambda}_{i,T}) - g(\rho_{\mathcal{K}(i)}) \to 0$ which, along with $g(\rho_{\mathcal{K}(i)}) = 1/p_{\mathcal{K}(i)}$, gives the result. □

4.3.4 Subspace estimation

In this section we are interested to estimate bilinear forms involving orthogonal projection matrix on the signal eigenspace. Let $A_T = U_T B_T^{\mathsf{H}}$ following Assumption 8 and write $U_T = [U_{0,T}, \ldots, U_{t,T}]$, $U_{\ell,T} \in \mathbb{C}^{N \times j_\ell}$. Then, the orthogonal projection matrix on the signal subspace is written as $\Pi_{\ell,T} = U_{\ell,T} U_{\ell,T}^{\mathsf{H}} \in \mathbb{C}^{N \times N}$. The goal here is to provide an estimator for the bilinear form $a_T^{\mathsf{H}} \Pi_{\ell,T} b_T$ where a_T and b_T are some N-dimensional vectors. Denote now $\widehat{\Pi}_{\ell,T}$ the orthogonal projection matrix on the eigenspace corresponding to the set of eigenvalues $\{\hat{\lambda}_{j_0+\ldots+j_{\ell-1}+1,T}, \ldots, \hat{\lambda}_{j_0+\ldots+j_\ell}\}$ of $\widehat{\Sigma}_T$, for $\ell = 0, \ldots, t$. With these notations, we have the following estimate:

Theorem 25. *Under Assumptions 3–8, let a_T, $b_T \in \mathbb{C}^N$ be two sequences of deterministic vectors with bounded norms and let $\ell \leq s$ with s the largest integer for which (4.8) holds. Then:*

$$a_T^{\mathsf{H}} \Pi_{\ell,T} b_T - \frac{\hat{g}'_T(\hat{\lambda}_{i,T})}{\hat{m}_T(\hat{\lambda}_{i,T}) \hat{g}_T(\hat{\lambda}_{i,T})} a_T^{\mathsf{H}} \widehat{\Pi}_{\ell,T} b_T \xrightarrow[T \to \infty]{a.s.} 0.$$

Proof. Recall that under Assumption 8, $p_0, \ldots, p_t$ are the limiting positive eigenvalues of $\frac{1}{T} A_T A_T^{\mathsf{H}}$. For $R_T = I_N$, the theorem thus coincides with [32, Theorem 2] since then $V_T = W_T$ is a bi-unitarily invariant (here Gaussian) matrix as requested by [32, Assumption 2]. We now reproduce the steps of [32] under our set of assumptions. Define the resolvents:

$$Q_T(z) = \left(\frac{1}{T} V_T V_T^{\mathsf{H}} - z I_N \right)^{-1},$$
$$\widetilde{Q}_T(z) = \left(\frac{1}{T} V_T^{\mathsf{H}} V_T - z I_T \right)^{-1}$$

and the matrix

$$\underline{Q}_T(z) = \begin{bmatrix} z Q_T(z^2) & \frac{1}{\sqrt{T}} V_T \widetilde{Q}_T(z^2) \\ \frac{1}{\sqrt{T}} \widetilde{Q}_T(z^2) V_T^{\mathsf{H}} & z \widetilde{Q}_T(z^2) \end{bmatrix}.$$

Define also

$$\widehat{M}_T(z) = \begin{bmatrix} zU_T^{\mathsf{H}} Q_T(z^2) U_T & \frac{1}{T} U_T^{\mathsf{H}} V_T \widetilde{Q}_T(z^2) B_T + I_K \\ \frac{1}{T} B_T^{\mathsf{H}} \widetilde{Q}_T(z^2) V_T^{\mathsf{H}} U_T + I_K & \frac{z}{T} B_T^{\mathsf{H}} \widetilde{Q}_T(z^2) B_T \end{bmatrix}.$$

Defining the vectors

$$\begin{aligned} \hat{a}_T &= \begin{bmatrix} zU_T^{\mathsf{H}} Q_T(z^2) \\ \frac{1}{T} B_T^{\mathsf{H}} \widetilde{Q}_T(z^2) V_T^{\mathsf{H}} \end{bmatrix} a_T \\ \hat{b}_T &= \begin{bmatrix} zU_T^{\mathsf{H}} Q_T(z^2) \\ \frac{1}{T} B_T^{\mathsf{H}} \widetilde{Q}_T(z^2) V_T^{\mathsf{H}} \end{bmatrix} b_T \\ \tilde{a}_T^{\mathsf{T}} &= [a_{k,T}^{\mathsf{T}}, 0, \ldots, 0] \\ \tilde{b}_T^{\mathsf{T}} &= [b_{k,T}^{\mathsf{T}}, 0, \ldots, 0] \end{aligned}$$

we have the following equation:

$$a_T^{\mathsf{H}} \widehat{\Pi}_{\ell,T} b_T = -\frac{1}{\imath\pi} \oint_{\mathcal{C}_{\ell,T}} \tilde{a}_T^{\mathsf{H}} \underline{Q}_T(z) \tilde{b}_T dz + \frac{1}{\imath\pi} \oint_{\mathcal{C}_{\ell,T}} \hat{a}_T^{\mathsf{H}} \widehat{M}_T(z)^{-1} \hat{b}_T dz \qquad (4.9)$$

for $\mathcal{C}_{\ell,T}$ a complex positively oriented contour enclosing only the eigenvalues $\hat{\lambda}_{j_0+\ldots+j_{\ell-1}+1,T}, \ldots, \hat{\lambda}_{j_0+\ldots+j_\ell,T}$. Let $\ell \leq s$. From Theorem 22-2), for all large T w.p. 1, the first term on the right-hand side of (4.9) is null (no pole of $\underline{Q}_T$ lies in $\mathcal{C}_{\ell,T}$ for large T), while in the second term $\mathcal{C}_{\ell,T}$ can be replaced by a contour $\mathcal{C}_\ell$ enclosing ρ_ℓ but no ρ_k, $k \neq \ell$. We must now prove $\hat{a}_T^{\mathsf{H}} \widehat{M}_T(z) \hat{b}_T - \bar{a}_T^{\mathsf{H}} \overline{M}_T(z) \bar{b}_T \xrightarrow{\text{a.s.}} 0$ where

$$\begin{aligned} \bar{a}_T &= \begin{bmatrix} zm(z^2) U_T^{\mathsf{H}} \\ 0 \end{bmatrix} a_T, \\ \bar{b}_T &= \begin{bmatrix} zm(z^2) U_T^{\mathsf{H}} \\ 0 \end{bmatrix} b_T \\ \overline{M}_T(z) &= \begin{bmatrix} zm(z^2) I_K & I_K \\ I_K & z\tilde{m}(z^2) P \end{bmatrix}. \end{aligned}$$

By [21, Lemmas 4.1–4.6], $\|\hat{a}_T - \bar{a}_T\| \xrightarrow{\text{a.s.}} 0$, $\|\hat{b}_T - \bar{b}_T\| \xrightarrow{\text{a.s.}} 0$,

$$\left\| \widehat{M}_T(z) - \begin{bmatrix} zm(z^2) I_K & I_K \\ I_K & -\frac{1}{zT} B_T^{\mathsf{H}} \left(I_T + cm(z^2) R_T \right)^{-1} B_T \end{bmatrix} \right\| \xrightarrow{\text{a.s.}} 0.$$

Assumption 8 and the definition of $\tilde{m}(z)$ then imply

$$\left\| -\frac{1}{zT} B_T^{\mathsf{H}} \left(I_T + cm(z^2) R_T \right)^{-1} B_T - z\tilde{m}(z^2) P \right\| \xrightarrow{\text{a.s.}} 0$$

which finally gives

$$\hat{a}_T^{\mathsf{H}} \widehat{M}_T(z) \hat{b}_T - \bar{a}_T^{\mathsf{H}} \overline{M}_T(z) \bar{b}_T \xrightarrow{\text{a.s.}} 0.$$

For $z \in \mathcal{C}_\ell$, $zm(z^2)$ and $z\tilde{m}(z^2)$ are bounded by $[\mathbf{d}(\mathcal{C}_\ell, \mathrm{supp}(\mu))]^{-1}$. Take $0 < \varepsilon < \mathbf{d}(\mathcal{C}_\ell, \mathrm{supp}(\mu))$. Then, for all large T, $zQ_T(z^2)$ and $z\widetilde{Q}(z^2)$ are bounded by ε^{-1} w.p. 1. The dominated convergence theorem therefore ensures that

$$a_T^{\mathsf{H}} \widehat{\Pi}_{\ell,T} b_T - \frac{1}{\imath\pi} \oint_{\mathcal{C}_\ell} \bar{a}_T^{\mathsf{H}} \overline{H}_T(z)^{-1} \bar{b}_T dz \xrightarrow{\text{a.s.}} 0.$$

Residue calculus of the right-hand side integrand as in [32, Equations (10)-(11)] then gives

$$a_T^{\mathsf{H}} \widehat{\Pi}_{\ell,T} b_T - \frac{m(\rho_\ell) g(\rho_\ell)}{g'(\rho_\ell)} a_T^{\mathsf{H}} \Pi_{\ell,T} b_T \xrightarrow[T\to\infty]{\text{a.s.}} 0.$$

Take i such that $\mathcal{K}(i) = \ell$. Using $\hat{\lambda}_{i,T} \xrightarrow{\text{a.s.}} \rho_\ell$, $\hat{m}_T(x) \xrightarrow{\text{a.s.}} m(x)$, $\hat{g}_T(x) \xrightarrow{\text{a.s.}} g(x)$, and $\hat{g}'_T(x) \xrightarrow{\text{a.s.}} g'(x)$ for x outside the support of μ then concludes the proof. □

4.4 Narrow band array processing

4.4.1 System model and assumptions

We consider now a practical example of narrow band array processing. Consider a uniform linear array of N antennas which captures T successive realizations $y_0, \ldots, y_{T-1}$ of the random process:

$$y_t = \sum_{k=0}^{K-1} a_k h(\theta_k) s_{k,t} + v_t \tag{4.10}$$

with $y_t \in \mathbb{C}^{N\times 1}$, $a_0 \geq \ldots \geq a_{K-1} > 0$ the amplitudes of sources $0, \ldots, K-1$, $h(\theta) \in \mathbb{C}^{N\times 1}$ the steering-vector function

$$h(\theta) = \frac{1}{\sqrt{N}} \left[1, e^{-2\imath\pi \sin\theta}, \ldots, e^{-2\imath\pi(N-1)\sin\theta}\right]^{\mathsf{T}} \tag{4.11}$$

with θ_k the angle-of-arrival of the signal from source k (the θ_k are assumed distinct), $s_{k,t} \in \mathbb{C}$ the signal emitted by source k at time t such

that $(s_{t,k})_{t,k=0}^{\infty,K-1}$ is an infinite array of circular complex i.i.d. random variables with $\mathbb{E}s_{0,0} = 0$, $\mathbb{E}|s_{0,0}|^2 = 1$, and $\mathbb{E}|s_{0,0}|^8 < \infty$, and $v_t \in \mathbb{C}^{N\times 1}$ the noise received at the sensor array at time t.

Stacking all the received vectors into the matrix $Y_T = [y_0, \ldots, y_{T-1}] \in \mathbb{C}^{N\times T}$, (4.10) is written as

$$Y_T = H_T P^{1/2} S_T^{\mathsf{H}} + V_T \tag{4.12}$$

where $H_T = [h(\theta_0), h(\theta_1), \ldots, h(\theta_{K-1})] \in \mathbb{C}^{N\times K}$ is the channel matrix, $S_T = [s_{t,k}^*]_{t,k=0}^{T-1,K-1} \in \mathbb{C}^{T\times K}$ is the signal matrix, $P = \mathrm{diag}(a_0^2, \ldots, a_{K-1}^2)$ is the matrix of source powers, and $V_T = [v_0, \ldots, v_{T-1}] \in \mathbb{C}^{N\times T}$ corresponds to the correlated noise matrix. We detail now the noise model. We assume the rows of V_T to be independent snapshots of a complex Gaussian circular causal ARMA(m, n) stationary process. Define the filter transfer function

$$\mathbf{p}(z) = \frac{1 + \alpha_1 z^{-1} + \ldots + \alpha_m z^{-m}}{1 + \beta_1 z^{-1} + \ldots + \beta_n z^{-n}}.$$

The noise process can be represented as the output of a filter with the transfer function $\mathbf{p}(z)$ driven by a standard complex Gaussian circular white noise. For $|z| \geq 1$, we have

$$\mathbf{p}(z) = \sum_{\ell=0}^{\infty} \psi_\ell z^{-\ell}$$

where $\sum |\psi_\ell| < \infty$.

The noise matrix can be hence written as $V_T = W_T R_T^{1/2}$ with W_T as in Assumption 4 and

$$R_T = \begin{bmatrix} r_0 & r_1 & \ldots & r_{T-1} \\ r_{-1} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & r_1 \\ r_{1-T} & \ldots & r_{-1} & r_0 \end{bmatrix}$$

with $r_k = \sum_{\ell\geq 0} \psi_{\ell+k}\psi_\ell^*$ for any $k \in \mathbb{N}$. We note that R_T is a Toeplitz nonnegative matrix.

It should be noted that the assumptions about circularity and the upper moments will be used only in the second order analysis. Nevertheless, second order analysis can be generalized to the case where the $s_{t,k}$ are not circular, and the circularity assumptions have been added only for simplicity.

Lemma 1. *Under Assumption 3, the model* (4.12) *satisfies Assumptions 4–8 with* ν *defined by*

$$\int g(t)\nu(dt) = \int_0^1 g(|\mathbf{p}(\exp(2\imath\pi u))|^2)\, du \tag{4.13}$$

for every positive measurable function g*, and with* P *defined in Assumption 8 the matrix of the source powers* a_k^2*.*

Proof. We start with Assumptions 5 and 6. If $m = n = 0$, then $\nu = \delta_1$ and these assumptions are trivially satisfied. Assume $\max(m, n) > 0$. Then Assumption 5–i) is a well known result on the spectral behavior of large Toeplitz matrices [30, 31]. The support of ν is the compact non-singleton interval $[a_\nu, b_\nu] = [\min_u q(u), \max_u q(u)]$ with $q(u) \triangleq |\mathbf{p}(\exp(2\imath\pi u))|^2$. It is also well known [30, §4.2] that $a_\nu \leq \sigma_{t,T}^2 \leq b_\nu$, so that Assumption 5–ii) is satisfied. Since $\mathbf{p}(z)$ is ARMA, for $g(t)$ the indicator function on a set of Lebesgue measure zero, the right hand side of (4.13) is zero. Hence ν has a density f_ν with respect to the Lebesgue measure. Let us provide the expression of f_ν at a point $s \in (a_\nu, b_\nu)$ such that for any u for which $q(u) = s$, $q'(u) \neq 0$. In a neighborhood of any of these u, q has a local inverse that we denote $q_u^{(-1)}$. Then, for $\varepsilon > 0$ small enough,

$$\nu(s-\varepsilon, s+\varepsilon) = \int_{t\,:\,q(t)\in[s-\varepsilon,s+\varepsilon]} dt = \sum_{u\,:\,q(u)=s} \int_{[s-\varepsilon,s+\varepsilon]} \frac{1}{\left|q'(q_u^{(-1)}(v))\right|}\, dv$$

by the variable change $q(t) = v$. Letting $\varepsilon \downarrow 0$, we obtain

$$\lim_{\varepsilon\downarrow 0} \frac{\nu(s-\varepsilon, s+\varepsilon)}{2\varepsilon} = \sum_{u\,:\,q(u)=s} \frac{1}{|q'(u)|} = f_\nu(s).$$

This proves $f_\nu(s) \to \infty$ as $s \uparrow b_\nu$, implying Assumption 6.
We now turn to Assumptions 7 and 8. Since the θ_i are distinct (modulo π), $H_T^{\mathsf{H}} H_T \to I_K$. By the law of large numbers, $\frac{1}{T} S_T^{\mathsf{H}} S_T \xrightarrow[T\to\infty]{\text{a.s.}} I_K$. Hence $\operatorname{rank}(A_T) = K$ w.p. 1 for all large T, and $\sup_T \frac{\|A_T\|}{T} < \infty$ w.p. 1. Let us write $A_T = U_T B_T^{\mathsf{H}}$ where $U_T = H_T(H_T^{\mathsf{H}} H_T)^{-1/2}$ and where $B_T = S_T P^{1/2}(H_T^{\mathsf{H}} H_T)^{1/2}$. By [4, Lemma 2.7] and $\mathbb{E}|s_{0,0}|^8 < \infty$, for any $z \in \mathbb{C}_+$ and any $0 \leq i, j \leq K-1$,

$$\mathbb{E}\Big|\Big[\frac{1}{T} S_T^{\mathsf{H}} (R_T - zI_T)^{-1} S_T - \frac{\operatorname{Tr}[(R_T - zI_T)^{-1}]}{T} I_K\Big]_{i,j}\Big|^4 \leq \frac{C}{T^2}$$

for some $C > 0$. By Markov's inequality, the argument of $\mathbb{E}|\cdot|^4$ converges to zero w.p. 1, and this convergence can be extended to $\mathbb{C} - \text{supp}(\mu)$. Since $T^{-1}\,\text{Tr}[(R_T - zI_T)^{-1}] \to m_\nu(z)$ for $z \in \mathbb{C} - \text{supp}(\nu)$, Assumption 8 is satisfied. □

4.4.2 Detection, power estimation, and localization

We have previously proved that the model (4.12) satisfies Assumptions (4)–(8). By direct application of Lemma 1 and Theorems 23 and 24 we get the following inference methods of the number of sources and their powers:

Proposition 8. *Consider the model* (4.12). *Let $k \geq 0$ be the largest integer for which*

$$a_k^2 > \left(\int_0^1 \frac{-m_b}{1 + c m_b \, |\mathbf{p}(\exp(2\imath\pi u))|^2} \, du \right)^{-1} \tag{4.14}$$

with $m_b \in (-(c \max_u |\mathbf{p}(\exp(2\imath\pi u))|^2)^{-1}, 0)$ the solution of

$$\int_0^1 \left(\frac{m \, |\mathbf{p}(\exp(2\imath\pi u))|^2}{1 + cm \, |\mathbf{p}(\exp(2\imath\pi u))|^2} \right)^2 du \;=\; \frac{1}{c}.$$

Given $L \geq K$ and $\varepsilon > 0$, define (with $\hat{\lambda}_{-1,T} = \infty$)

$$\hat{k}_T = \arg \max_{m \in \{0,\ldots,L-1\}} \frac{\hat{\lambda}_{m-1,T}}{\hat{\lambda}_{m,T}} > 1 + \varepsilon.$$

Then $\hat{k}_T = k$ w.p. 1 for all large T and ε small enough. Moreover, for $i = 0, \ldots, \hat{k}_T - 1$ let

$$\hat{a}_{i,T}^2 \triangleq \frac{1}{\hat{g}_T(\hat{\lambda}_{i,T})} \tag{4.15}$$

with $\hat{g}_T(\hat{\lambda}_{i,T})$ as in Theorem 24. Then

$$\hat{a}_{i,T}^2 \xrightarrow{\text{a.s.}} a_i^2.$$

We present a source localization method based on the MUSIC approach [57] (see Chapter 3, Section 3.3). Recall that the MUSIC algorithm exploits the fact that $h(\theta_i)^{\mathsf{H}}(I_N - \Pi_{\ell,T})h(\theta_i) = 0$ with $\Pi_{\ell,T}$ a projector on the subspace generated by $h(\theta_0), \ldots, h(\theta_{\ell-1})$ for any $i \leq \ell \leq K-1$. Since $\|h(\theta)\| = 1$, $\theta_0, \ldots, \theta_{\ell-1}$ are the arguments of the local maxima of

$$\gamma_T^\ell(\theta) \triangleq h(\theta)^{\mathsf{H}} \Pi_{\ell,T} h(\theta).$$

Based on Theorem 25, the following proposition provides a localization function estimate:

Proposition 9. *Let k and $\hat{k}_T$ be as in Proposition 8 and denote $\hat{u}_{0,T}, \ldots, \hat{u}_{\hat{k}_T-1,T}$ the eigenvectors of $\widehat{\Sigma}_T$ with respective eigenvalues $\hat{\lambda}_{0,T}, \ldots, \hat{\lambda}_{\hat{k}_T-1,T}$. Then, for $\theta \in [-\pi/2, \pi/2]$,*

$$\gamma_T^k(\theta) - \hat{\gamma}_T^{\hat{k}_T}(\theta) \xrightarrow{\text{a.s.}} 0$$

where

$$\gamma_T^k(\theta) \triangleq h(\theta)^{\mathsf{H}} \Pi_{k,T} h(\theta)$$

$$\hat{\gamma}_T^{\hat{k}_T}(\theta) \triangleq \sum_{j=0}^{\hat{k}_T-1} \frac{\hat{g}_T'(\hat{\lambda}_{j,T})}{\hat{m}_T(\hat{\lambda}_{j,T})\hat{g}_T(\hat{\lambda}_{j,T})} h(\theta)^{\mathsf{H}} \hat{u}_{j,T} \hat{u}_{j,T}^{\mathsf{H}} h(\theta).$$

Proof. From Lemma 1 Assumptions 3–8 are satisfied, so Theorem 25 can be applied for each $i \leq k-1$. Taking $a_T = b_T = h(\theta)$ and $U_T = H_T(H_T^{\mathsf{H}} H_T)^{-1/2}$ as in Theorem 25, we obtain the desired result for $U_T J U_T^{\mathsf{H}}$, $J = \mathrm{diag}(I_k, 0)$, instead of $\Pi_{k,T}$. As $(H_T^{\mathsf{H}} H_T)^{-1/2} J (H_T^{\mathsf{H}} H_T)^{-1/2} \to J$ and $h(\theta)^{\mathsf{H}}(H_T J H_T^{\mathsf{H}} - \Pi_{k,T}) h(\theta) \to 0$, we have $h(\theta)^{\mathsf{H}} \Pi_{k,T} h(\theta) - h(\theta)^{\mathsf{H}} U_T J U_T^{\mathsf{H}} h(\theta) \to 0$, completing the proof. □

Proposition 9 ensures that $\hat{\gamma}_T^{\hat{k}_T}(\theta)$ is a consistent estimator of the localization function $\gamma_T^k(\theta)$. The improved MUSIC algorithm we therefore propose consists in estimating $\theta_0, \ldots, \theta_{k-1}$ as the arguments of the $\hat{k}_T$ highest maxima of $\hat{\gamma}_T^{\hat{k}_T}(\theta)$. Observe that, although the system models differ in comparison to that of [32], the MUSIC estimator proposed here takes the same form as the estimator [32]. This remark would not hold if it were not for Assumption 8. Note also that, as $c \to 0$, $g'(x)m(x)^{-1}g(x)^{-1} \to 1$ for all real $x \neq \int t\nu(dt)$, so that the improved MUSIC algorithm proposed reduces to the standard large T MUSIC approach.

4.4.3 Second order performance analysis

We now provide an asymptotic second order analysis for the power estimators under the model (4.12). We gather the source powers a_k^2 in groups of equal powers $p_0 > \ldots > p_t$ with respective multiplicities $j_0, \ldots, j_t$.

4.4.4 Main results

In order to provide the fluctuations of the source power estimators given by (4.15) in Proposition 4.3.2, first we study the fluctuations of the isolated eigenvalues of $\widehat{\Sigma}_T$. Denote by $\rho_{\mathcal{K}(i),T}$ the "finite horizon" equivalent of the

spike limiting position. More precisely, we provide the fluctuations of $\hat{\lambda}_{i,T} - \rho_{\mathcal{K}(i),T}$, $i \leq s$, with s the largest integer satisfying (4.8).

We start by presenting some important quantities. Recall the definition of ν_T in Assumption 5 and recall that $c_T = N/T$. Let $m_T(z)$ be the Stieltjes transform defined analogously to $m(z)$ defined by (4.4) in Theorem 20 with the difference that the measure ν and the constant c are respectively replaced with their "finite horizon" equivalents ν_T and c_T. We have then

$$m_T(z) = \left(-z + \int \frac{t}{1 + c_T m_T(z) t} \nu_T(dt) \right)^{-1} \tag{4.16}$$

uniquely defines the Stieltjes transform $m_T(z)$ of a probability measure μ_T supported by $\mathbb{R}_+$. In addition, μ_T converges weakly to μ as $T \to \infty$; the Hausdorff distance between the supports of these two measures converges to zero [59, 4] and, for each $b' > b$, $m_T(z)$ is analytic on $\mathbb{C} - [0, b']$ for all large T. Let

$$\begin{aligned} \tilde{m}_T(z) &= \int \frac{-1}{z(1 + c_T m_T(z) t)} \nu_T(dt) \\ &= \frac{-1}{zT} \operatorname{Tr}(I_T + c_T m_T(z) R_T)^{-1}. \end{aligned}$$

Similarly to Theorem 20-3), $\tilde{m}_T(z)$ satisfies $\tilde{m}_T(z) = c_T m_T(z) - (1 - c_T)/z$.

Consequently, for all T large, $g_T(x) \triangleq x m_T(x) \tilde{m}_T(x)$ is defined on (b', ∞), $b' > b$, and, for any k such that $p_k g(b^+) > 1$, $p_k g_T(x) = 1$ has a unique solution $\rho_{k,T}$ in (b, ∞). We introduce then the following important quantity:

Lemma 2. *Consider the model* (4.12). *Then the function*

$$\Delta(x) = 1 - c \int \left(\frac{m(x) t}{1 + c m(x) t} \right)^2 \nu(dt)$$

is defined and positive on (b, ∞). *Furthermore,* $\Delta(x) \to 0$ *as* $x \downarrow b$ *and* $\Delta(x) \to 1$ *as* $x \to \infty$.

Proof. See Appendix 4.7.2. □

We have then the following theorem:

Theorem 26. *Consider the model* (4.12). *Assume in addition* $\mathbb{E}[s_{0,0}^u (s_{0,0}^*)^v] = 0$ *for* $u + v \leq 4$ *and* $u \neq v$, *and let* $\kappa \triangleq \mathbb{E}|s_{0,0}|^4 - 2$. *Let* s *be the largest*

integer (assumed ≥ 0) for which (4.14) holds. For $k = 1, \ldots, s$ and all T large, let $\rho_{k,T}$ be the unique solution in (b, ∞) of $p_k g_T(x) = 1$. Define

$$\eta_{k,T} = \sqrt{T}\left(\begin{bmatrix} \hat{\lambda}_{j_0+\cdots+j_{k-1}+1,T} \\ \vdots \\ \hat{\lambda}_{j_0+\cdots+j_k,T} \end{bmatrix} - \rho_{k,T} \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}\right),$$

$$\begin{aligned}
\alpha_k &= \frac{m^2(\rho_k)}{\Delta(\rho_k)}\left[\int \frac{t^2 + 2p_k t}{(1+cm(\rho_k)t)^2}\nu(dt)\right. \\
&\qquad \left. + c\Big(\int \frac{p_k m(\rho_k) t}{(1+cm(\rho_k)t)^2}\nu(dt)\Big)^2\right], \\
\beta_k &= \int \frac{p_k^2 m(\rho_k)^2}{(1+cm(\rho_k)t)^2}\nu(dt), \quad \textit{and} \\
\phi_k &= \left(\int \frac{p_k m(\rho_k)}{1+cm(\rho_k)t}\nu(dt)\right)^2.
\end{aligned}$$

Let $M_0, \ldots, M_s$, $M_k = [M_{\ell,m,k}]_{0 \leq \ell, m \leq j_k}$, be random independent Hermitian matrices such that $\{M_{\ell,m,k}\}_{\ell \leq m}$ are independent, $M_{\ell,\ell,k} \sim \mathcal{N}(0, \alpha_k + \beta_k + \kappa\phi_k)$, and $M_{\ell,m,k} \sim \mathcal{CN}(0, \alpha_k + \beta_k)$ for $0 \leq \ell < m \leq j_k$. Let χ_k be the $\mathbb{R}^{j_k}-$valued vector of the decreasingly ordered eigenvalues of $(p_k g'(\rho_k))^{-1} M_k$. Then

$$(\eta_{0,T}, \ldots, \eta_{s,T}) \xrightarrow[T\to\infty]{\mathcal{L}} (\chi_0, \ldots, \chi_s).$$

Proof. The proof is provided in Section 4.4.5. □

Theorem 26 shows that, after appropriate centering and scaling, the vector of the isolated eigenvalues of $\widehat{\Sigma}_T$ that converge to $\rho_k > b$ tends to fluctuate like the eigenvalues of a certain Hermitian matrix with Gaussian elements. If $\kappa = 0$, this matrix is a Gaussian Unitary Ensemble (GUE) matrix. Recall that a GUE matrix is such that $M_{ii} \sim \mathcal{N}(0,1)$ and $M_{ij} \sim \mathcal{CN}(0,1)$ for $i < j$, these random variables being independent. When $K = 0$, $T^{2/3}\left(\frac{\hat{\lambda}_{0,T} - b_T}{\sigma_T}\right)$ converges in law to the Tracy–Widom probability distribution $\mathsf{TW}(\cdot)$, where b_T and σ_T are the finite horizon equivalents of b and σ depending on c_T and ν_T [26]. This result can be generalized to show that for any fixed integer r, the vector $T^{2/3}(\hat{\lambda}_{0,T} - b_T, \ldots, \hat{\lambda}_{r,T} - b_T)$ converges in distribution to a multi-dimensional version of the Tracy–Widom law. These results and Theorem 26 can then be used to evaluate the error probabilities of the source detection schemes described in Theorem 23 and Proposition 8. From Theorem 26, we obtain the fluctuations of the source power estimates:

Theorem 27. *Consider the setup of Theorem 26 and let* $\hat{p}_{i,T} = (\hat{g}_T(\hat{\lambda}_{i,T}))^{-1}$ *for* $i = 0, \ldots, j_0 + \cdots + j_s$. *For* $k = 1, \ldots, s$, *define*

$$\xi_{k,T} = \sqrt{T} \left(\begin{bmatrix} \hat{p}_{j_0+\cdots+j_{k-1}+1,T} \\ \vdots \\ \hat{p}_{j_0+\cdots+j_k,T} \end{bmatrix} - p_k \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \right).$$

Let M_k *be defined as in Theorem 26 and let* $\check{\chi}_k$ *be the* $\mathbb{R}^{j_k}-$*valued vector of the decreasingly ordered eigenvalues of* $p_k M_k$. *Then*

$$(\xi_{0,T}, \ldots, \xi_{s,T}) \xrightarrow[T\to\infty]{\mathcal{L}} (\check{\chi}_0, \ldots, \check{\chi}_s).$$

Proof. A sketch of the proof is given in Appendix 4.7.3. □

A straightforward application of the Delta method [66, Th. 3.1] on Theorem 27 implies in particular that, for $k = 0, \ldots, s$,

$$\sqrt{T} \left(\frac{1}{j_k} \sum_{i=1}^{j_k} \hat{p}_{j_0+\ldots+j_{k-1}+i,T} - p_k \right) \xrightarrow[T\to\infty]{\mathcal{L}} \bar{\chi}_k$$

with $\bar{\chi}_k \sim \mathcal{N}(0, j_k^{-1} p_k^2 (\alpha_k + \beta_k + \kappa\phi_k))$, independent across k.

As a corollary of Theorem 27, the following proposition provides the behavior of the power estimates for extreme values of p_k, *i.e.* for $p_k \to \infty$ and for p_k close to the detectability limit given by (4.14):

Proposition 10. *Consider the setting of Theorem 27. Let* p_{lim} *be the infimum of the* p_k *satisfying* (4.14), M_k *be defined as in Theorem 26, and* $\psi_k \triangleq \alpha_k + \beta_k + \kappa\phi_k$, $\check{\psi}_k \triangleq \alpha_k + \beta_k$. *Then*

$$\psi_k \xrightarrow[p_k \downarrow p_{\text{lim}}]{} \infty, \quad \check{\psi}_k \xrightarrow[p_k \downarrow p_{\text{lim}}]{} \infty$$
$$\psi_k \xrightarrow[p_k \to \infty]{} 1 + \kappa, \quad \check{\psi}_k \xrightarrow[p_k \to \infty]{} 1.$$

Proof. See Appendix 4.7.4. □

From this result when the power is close to p_{lim}, then the variance of the estimator converges to infinity confirming that it is not possible to estimate the signal power close to the detectability limit.

4.4.5 Proof of Theorem 26

The proof relies on two ingredients: an adaptation of [21, Th. 2.3] and a result on fluctuations of quadratic forms. Let $A_T = U_T B_T^{\mathsf{H}}$ with $U_T = H_T(H_T^{\mathsf{H}} H_T)^{-1/2}$ and $B_T = S_T P^{1/2}(H_T^{\mathsf{H}} H_T)^{1/2} = [B_{0,T}, \ldots, B_{t,T}]$, $B_{k,T} \in \mathbb{C}^{T \times j_k}$. Recall the quantity

$$\eta_{k,T} = \sqrt{T}\left(\begin{bmatrix} \hat{\lambda}_{j_0+\cdots+j_{k-1}+1,T} \\ \vdots \\ \hat{\lambda}_{j_0+\cdots+j_k,T} \end{bmatrix} - \rho_{k,T} \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \right).$$

In [21], it is shown that the $\eta_{k,T}$ fluctuate like the ordered eigenvalues of the matrices $(p_k g(\rho_k)')^{-1}(\sqrt{\alpha_k} G_k + \sqrt{T} F_{k,T})$ where

$$F_{k,T} = \frac{m_T(\rho_{k,T})}{T} B_{k,T}^{\mathsf{H}} (I_T + c_T m_T(\rho_{k,T}) R_T)^{-1} B_{k,T} + I_{j_k} \tag{4.17}$$

and the G_k are GUE matrices independent of the $F_{k,T}$. This is formalized by Proposition 11 below. Using $H_T^{\mathsf{H}} H_T \xrightarrow{\text{a.s.}} I_K$, the law of large numbers and the definition of $\rho_{k,T}$ informally give

$$F_{k,T} \simeq \left(\frac{p_k}{T} \operatorname{Tr}\left[m_T(\rho_{k,T})(I_T + c_T m_T(\rho_{k,T}) R_T)^{-1} \right] + 1 \right) I_{j_k} = 0.$$

We thus need to study the fluctuations of $\sqrt{T} F_{k,T}$, which is the purpose of the three following lemmas. Lemma 3 is a Central Limit Theorem characterizing the fluctuations of random matrices of the type $S_T^{\mathsf{H}} D_T S_T$ where D_T is a sequence of $T \times T$ deterministic matrices. Lemma 4 particularizes the results of Lemma 3 to the case where $D_T = p_k m_T(\rho_{k,T})(I_T + c_T m_T(\rho_{k,T}) R_T)^{-1}$. In Lemma 5 these results are used to characterize the fluctuations of $F_{k,T}$. Essentially, it is shown there that the matrices $B_{k,T}$ can be replaced with $\sqrt{p_k} S_{k,T}$.

Lemma 3. *Let $D_T \in \mathbb{C}^{T \times T}$ be a sequence of deterministic Hermitian matrices with $\sup_T \|D_T\| < \infty$. Assume that*

$$\frac{1}{T} \operatorname{Tr} D_T^2 \xrightarrow[T\to\infty]{} \beta \quad \text{and} \quad \frac{1}{T} \operatorname{Tr}(\operatorname{diag}(D_T))^2 \xrightarrow[T\to\infty]{} \phi.$$

Consider the matrices S_T defined by (4.12). Then

$$\sqrt{T}\Big(\frac{1}{T} S_T^{\mathsf{H}} D_T S_T - \frac{\operatorname{Tr} D_T}{T} I_K \Big) \xrightarrow[T\to\infty]{\mathcal{L}} G$$

where $G = [G_{ij}]_{0 \le i,j \le K-1}$ is random Hermitian such that $\{G_{ij}\}_{i \le j}$ are independent, $G_{ii} \sim \mathcal{N}(0, \beta + \kappa\phi)$ for $0 \le i \le K-1$, and $G_{ij} \sim \mathcal{CN}(0, \beta)$ for $0 \le i < j \le K-1$.

Proof. See Appendix 4.7.5. □

Lemma 4. *Let* $0 \leq k \leq s$ *and*

$$D_T = p_k m_T(\rho_{k,T})(I_T + c_T m_T(\rho_{k,T})R_T)^{-1}.$$

Then $\limsup_T \|D_T\| < \infty$,

$$\frac{1}{T}\operatorname{Tr}(D_T^2) \xrightarrow[T\to\infty]{} \beta_k \ , \text{ and } \frac{1}{T}\operatorname{Tr}(\operatorname{diag}(D_T))^2 \xrightarrow[T\to\infty]{} \phi_k$$

where β_k *and* ϕ_k *are given in Theorem 26.*

Proof. See Appendix 4.7.6. □

Lemma 5. *Let* $M_0, \ldots, M_t$, $M_k = [M_{\ell,m,k}]_{0\leq\ell,m\leq j_k}$, *be random independent Hermitian matrices such that the* $\{M_{\ell,m,k}\}_{\ell\leq m}$ *are independent,* $M_{\ell,\ell,k} \sim \mathcal{N}(0, \beta_k + \kappa\phi_k)$, *and* $M_{\ell,m,k} \sim \mathcal{CN}(0, \beta_k)$ *for* $0 \leq \ell < m \leq j_k$. *Then*

$$(\sqrt{T}F_{k,T})_{k=0,\ldots,t} \xrightarrow[T\to\infty]{\mathcal{L}} (M_k)_{k=0,\ldots,t}.$$

Proof. See Appendix 4.7.7. □

Theorem 2.3 of [21] can be adapted to obtain the following result:

Proposition 11. *In the setting of Theorem 26, let* $G_0, \ldots, G_s$, $G_k \in \mathbb{C}^{j_k\times j_k}$, *be independent GUE matrices. Then, for any bounded and continuous* $f : \mathbb{R}^{j_0+\cdots+j_s} \to \mathbb{R}$,

$$\mathbb{E}[f(\eta_{0,T}, \ldots, \eta_{s,T})] - \mathbb{E}[f(\zeta_0, \ldots, \zeta_s)] \to 0$$

where ζ_k *is the random vector of the decreasingly ordered eigenvalues of* $(p_k g(\rho_k)')^{-1}(\sqrt{\alpha_k}G_k + \sqrt{T}F_{k,T})$.

Remark 1. *In fact, [21, Th. 2.3] characterizes the asymptotic fluctuations of the random variables* $\sqrt{T}(\hat{\lambda}_{i,T} - \rho_{\mathcal{K}(i)})$ *instead of the* $\sqrt{T}(\hat{\lambda}_{i,T} - \rho_{\mathcal{K}(i),T})$, *so that the speed of convergence of* ν_T *towards* ν *and of* c_T *towards* c *had to be controlled through [21, Assumption 7]. By replacing* ρ_k *with* $\rho_{k,T}$, *the proof of [21, Th. 2.3] goes on without the need for that assumption. Replacing* ρ_k *by* $\rho_{k,T}$ *is enough for the present purpose.*

By Lemma 5, the $(s+1)$-uple of matrices $(\sqrt{\alpha_k}G_k + \sqrt{T}F_{k,T})_{k=0}^s$ converges in distribution to the $(s+1)$-uple $(M_0, \ldots, M_s)$ provided in the statement of Theorem 26. Applying Proposition 11, this theorem is proven.

4.5 Numerical results

We consider the setting of Section 4.4, with signals $s_{t,k}$ drawn from a QPSK constellation for which $\kappa = -1$. The signal power a_k^2 defines the signal-to-noise ratio (SNR). The noise is issued from an autoregressive (AR) process of order 1 and parameter a, so that $[R_T]_{k,l} = a^{|k-l|}$. All other parameters are given in the figure captions.

In Figure 4.1, the probability of correct order estimation of the estimator proposed in Proposition 8 is compared against the MDL and AIC criteria, for $K = 2$ equal power sources, for growing N, and for $c_T = 0.5$ fixed. We observe that the proposed estimator outperforms the MDL and the AIC methods, consistently with the known inappropriateness of the latter. Note that the AIC particularly fails to detect any source, irrespective of N.

In Figure 4.2, the false alarm rate (FAR) and correct detection rate (CDR) for single source detection is evaluated for different values of ε and for growing ratios c_T. We observe here the impact of an appropriate choice of ε which, if too small, generates a high FAR when the noise eigenvalues tend to spread (*i.e.* for c_T large) while, if too large, does not allow for correct source detection close to the detectability threshold (*i.e.* for c_T large).

Figure 4.3 depicts the normalized mean square error (NMSE) $\mathbb{E}[(\hat{a}_0^2 - a_0^2)^2 a_0^{-4}]$ of the power estimation of Proposition 8 against its theoretical value obtained from Theorem 26. For the purpose of analysis, we assume that the source is always detected, *i.e.* $\hat{k}_T = 1$, irrespective of the SNR. As confirmed by Proposition 10, the theoretical variance diverges as $p_k \downarrow p_{\lim}$. We however observe that in the finite N, T regime, the power estimator errors remain bounded at low SNR. This is explained by the fact that, while the theoretical error diverges due to $\Delta \downarrow 0$ (see Lemma 2) as $p_k \downarrow p_{\lim}$, its estimator for each N, T (obtained by replacing m by $\hat{m}_T$) is always non-zero even for $p_k = p_{\lim}$. In the high SNR regime, here with $\kappa = -1$, the NMSE becomes linear (in dB scale) with slope -10 dB/decade. It is easily shown that the limiting SNR gap between the proposed and oracle estimators is exactly

$$10 \log \Big(\int_0^1 |\mathbf{p}(\exp(2\imath\pi u))|^2 du \cdot \int_0^1 |\mathbf{p}(\exp(2\imath\pi u))|^{-2} du \Big) \text{ dB}$$

which is merely due to a gain in SNR after whitening. In particular, the larger the correlation parameter a, the bigger the limiting gap.

In Figure 4.4, the mean square error $\mathbb{E}[(\hat{\gamma}(\theta_0) - \gamma(\theta_0))^2]$ of the localization function at position $\theta_0 = 10^\circ$ is compared against the performances of the oracle estimator (which performs pre-whitening prior to using the estimator

of [32] or equivalently that of Proposition 9) and of the traditional MUSIC estimator with localization function $\hat{\gamma}_{\text{trad},T}(\theta) \triangleq \sum_{k=0}^{\hat{k}_T-1} h(\theta)^{\mathsf{H}} \hat{u}_{k,T} \hat{u}_{k,T}^{\mathsf{H}} h(\theta)$ in the notations of Proposition 8. The source is again supposed always detected so that $\hat{k}_T = 1$ throughout the experiment. The proposed estimator outperforms greatly the traditional MUSIC approach here, which is both due to the large N, T regime improvement and to the consideration of the non-white noise setting. The oracle estimator shows a huge performance improvement in the low SNR regime, which translates the fact that condition (4.8) (which needs to be fulfilled for either method to be valid) is extremely demanding when $\beta_1 = 0.6$ (due to $\text{supp}(\mu)$ being large). In the large SNR regime, a constant gap is maintained which, although we do not provide theoretical support, appears as a similar SNR-gap phenomenon as observed in Figure 4.3.

In Figure 4.5, we now take $K = 2$ sources, with $a_0 = a_1$ the amplitude of which define the SNR, and again assuming $\hat{k}_T = 2$. Here are compared the performances of resolution of two close sources located at $\theta_0 = 10°$ and $\theta_0 = 12°$ for the localization method proposed in Proposition 9, for the oracle estimator, and for the traditional MUSIC estimator. The figure of merit, referred to as resolution probability, is the probability of identifying exactly two local minima of the localization function in the window $[5°, 17°]$. We observe that the proposed algorithm performs significantly better than the traditional MUSIC method, confirming the results of [32] for the current model.

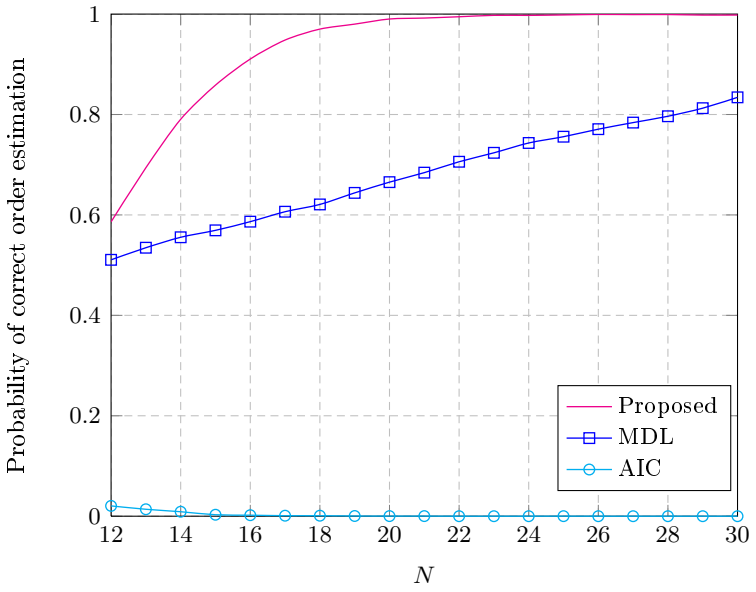

Figure 4.1: Probability of correct order estimation versus N with $K = 2$, SNR= 10 dB (same power for each source), $L = 5$, $\varepsilon = 0.75$, $c_T = 0.5$, and $a = 0.6$.

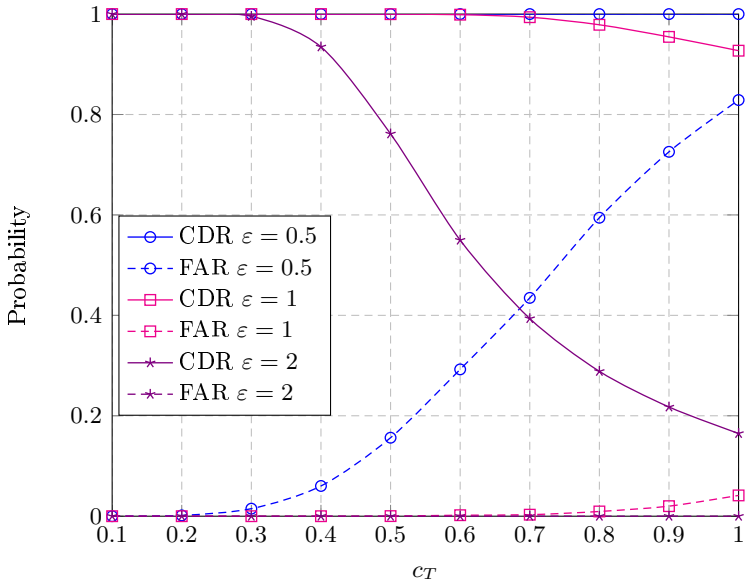

Figure 4.2: CDR (plain curve) and FAR (dashed curves) versus c_T with $K = 1$, $N = 20$, SNR= 10 dB, $L = 5$, and $a = 0.6$.

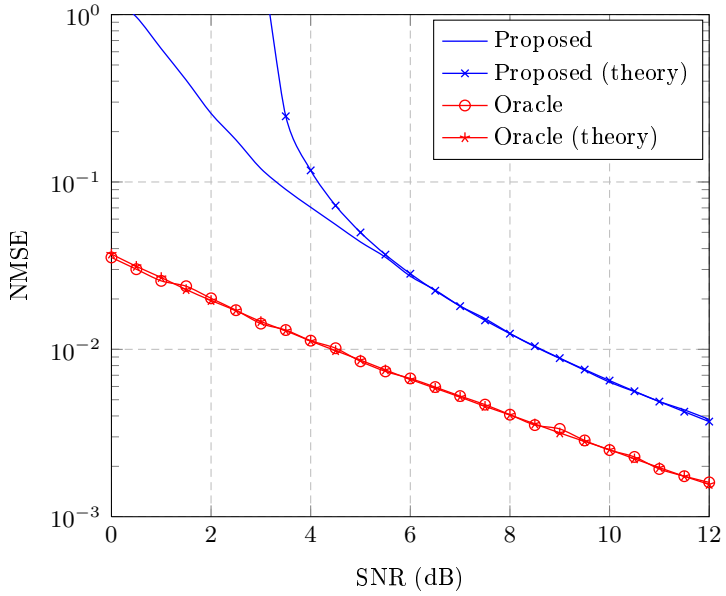

Figure 4.3: NMSE of the estimated power versus SNR with $K = 1$, $N = 20$, $c_T = 0.5$, and $a = 0.6$.

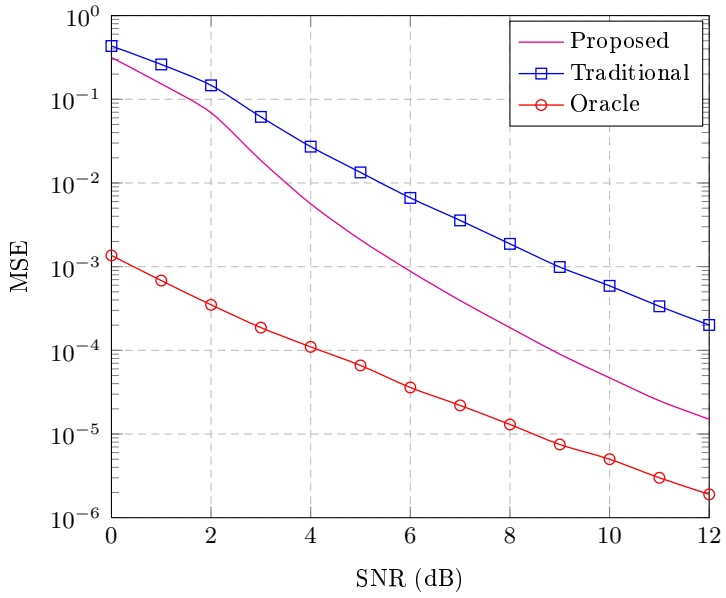

Figure 4.4: MSE of the localization function versus SNR with $K = 1$, $N = 20$, $c_T = 0.2$, and $a = 0.6$.

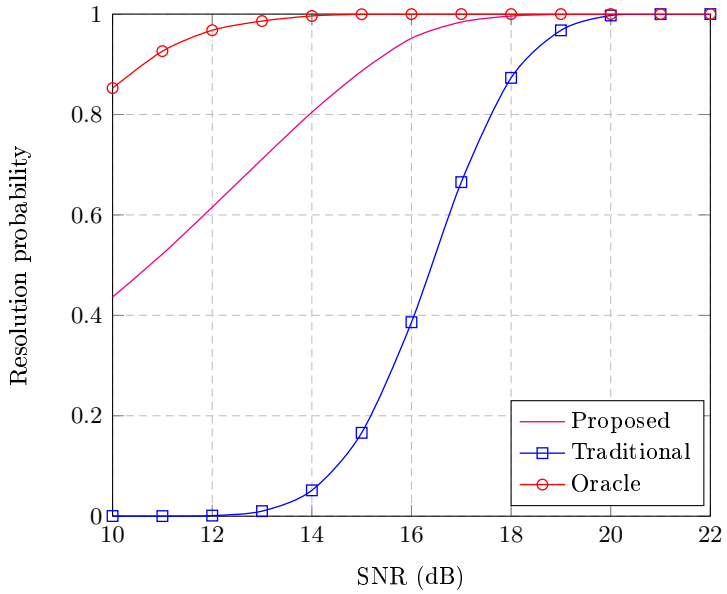

Figure 4.5: Resolution probability versus SNR with $K = 2$, $N = 20$, $c_T = 0.2$, and $a = 0.6$.

4.6 Conclusions

This chapter proposed a set of statistical inference methods for large dimensional information-plus-noise models with multiple sources, when the noise is correlated in time while the information is correlated in space (or vice-versa), without any assumption on the structure of the noise correlation matrix. The estimators were proved consistent in the limiting regime where both the system size and the number of observations go large. The proposed approach relies on the asymptotic spectral separation between noise and signal in the observed sample covariance matrix. The problem of signal detection is the choice of the eigenvalue "gap parameter" ε. It seems natural to be able to evaluate the right-edge of $\mathrm{supp}(\mu)$ from the noise eigenvalues, thus resulting in a test to compare $\hat{\lambda}_{i,T}$, $i = 0, \ldots, L-1$, to the estimated edge. To finely tune the test, one can then use the results from [26] which proves Tracy–Widom fluctuations at the edge with scaling coefficient $\underline{x}''(m_b)$ (m_b given by Corollary 1). However, estimating both the edge and this coefficient remains a challenging problem.

Under the same hypotheses, using instead prior information on the noise structure, an alternative approach consists in estimating the noise covariance

in the presence of signals. This is the main purpose of Chapter 5.

4.7 Appendix

4.7.1 Proof of Corollary 1

The derivative

$$x'(m) = \frac{1}{m^2} - c \int \left(\frac{t}{1+cmt} \right)^2 \nu(dt)$$

of $x(m)$ is continuous and increasing on $(-(cb_\nu)^{-1}, 0)$, and $x'(m) \to \infty$ as $m \uparrow 0$. To establish the proposition, it will be enough to show that $x'(m) \to -\infty$ as $m \downarrow -(cb_\nu)^{-1}$. This is obvious when $\nu(b_\nu) > 0$. Assume then $\nu(b_\nu) = 0$. When $m \downarrow -(cb_\nu)^{-1}$, by the monotone convergence theorem

$$\begin{aligned} \int \frac{t^2}{(1+cmt)^2} \nu(dt) &\uparrow \int \frac{t^2}{(1-t/b)^2} \nu(dt) \\ &\geq \int_{[b_\nu - \varepsilon, b_\nu]} \frac{b_\nu^2 t^2}{(b_\nu - t)^2} f_\nu(t)\, dt = \infty \end{aligned}$$

from the behavior of $f_\nu(t)$ near b_ν, which proves the result.

4.7.2 Proof of Lemma 2

Considering Equation (4.4), we obtain after some calculus that $m'(x) = m^2(x)/\Delta(x)$ on (b, ∞). Since $m(x)$ is negative and increasing on (b, ∞), both $m'(x)$ and $m^2(x)$ are positive on this interval so that $\Delta(x) > 0$ on (b, ∞).

Proposition 7 shows that b coincides with the minimum of $x(m)$ on $((-cb_\nu)^{-1}, 0)$. Moreover, when Assumption 6 is satisfied (which is the case for the model (4.12) by Lemma 1), the proof of Corollary 1 shows that $x(m)$ attains its minimum at a unique point $m_b \in ((-cb_\nu)^{-1}, 0)$, and $x'(m_b) = 0$. Finally, Proposition 7 shows that $x(m)$ is the inverse of $m(x)$ on (b, ∞). It results that $m(x) \to m_b$ and $m'(x) = 1/x'(m(x)) \to \infty$ as $x \downarrow b$. This proves $\Delta(x) \to 0$ as $x \downarrow b$.

When $x \to \infty$, both $(xm(x))^2 = (\int x(t-x)^{-1} \mu(dt))^2$ and $x^2 m'(x) = \int x^2 (t-x)^{-2} \mu(dt)$ converge to 1. Hence, $\Delta(x) = (xm(x))^2 (x^2 m'(x))^{-1} \to 1$, concluding the proof.

4.7.3 Proof of Theorem 27

For simplicity, we focus on the fluctuations of $\sqrt{T}(\hat{p}_{0,T} - p_0)$. Recall that $\hat{p}_{0,T} = \hat{g}_T(\hat{\lambda}_{0,T})^{-1}$ and $p_0 = g_T(\rho_{0,T})^{-1}$. Define $\underline{g}_T(x) = \underline{m}_T(x)(xc_T\underline{m}_T(x) + c_T - 1)$ with $\underline{m}_T(x)$ defined in Theorem 20-6). We have

$$\begin{aligned}
\sqrt{T}(\hat{p}_{0,T} - p_0) &= \sqrt{T}(\hat{g}_T(\hat{\lambda}_{0,T})^{-1} - g_T(\rho_{0,T})^{-1}) \\
&= \sqrt{T}(\hat{g}_T(\hat{\lambda}_{0,T})^{-1} - \underline{g}_T(\hat{\lambda}_{0,T})^{-1}) \\
&\quad + \sqrt{T}(\underline{g}_T(\hat{\lambda}_{0,T})^{-1} - g_T(\hat{\lambda}_{0,T})^{-1}) \\
&\quad + \sqrt{T}(g_T(\hat{\lambda}_{0,T})^{-1} - g_T(\rho_{0,T})^{-1}) \\
&\triangleq f_{1,T}(\hat{\lambda}_{0,T}) + f_{2,T}(\hat{\lambda}_{0,T}) + f_{3,T}(\hat{\lambda}_{0,T}).
\end{aligned}$$

As $\lambda_{0,T} \xrightarrow{\text{a.s.}} \rho_0$, we can replace $f_{1,T}(\hat{\lambda}_{0,T})$ by $f_{1,T}(\hat{\lambda}_{0,T})\mathbb{1}_I(\lambda_{0,T})$ where $\mathbb{1}_I$ is the indicator function on a small compact interval I in a neighborhood of ρ_0. Mimicking the proof of Theorem 24, we can show that $\sup_{x\in I} f_{1,T}(x) \xrightarrow{\mathcal{P}} 0$. We similarly restrict $f_{2,T}$ to I. On this set, it is possible to show that the random process $T(\underline{m}_T(x) - m_T(x))$ valued in the set $C(I)$ of the continuous functions on I, converges in distribution towards a Gaussian process in $C(I)$. This result was shown in [6] for I a compact path of $\mathbb{C}_+$; this can be generalized to the interval I of interest in this proof by using the Gaussian tools used in *e.g.* [21]. As a result, $\sup_{x\in I} f_{2,T}(x) \xrightarrow{\mathcal{P}} 0$. To deal with $f_{3,T}$, we start by observing that $g_T(\rho_{k,T}) \to g(\rho_k)$ and $(1/g_T(\rho_{k,T}))' \to -g'(\rho_k)/g^2(\rho_k) = -p_k^2 g'(\rho_k)$. Using the result of Theorem 26 and applying the Delta method [17, Prop. 6.1.6], we can show that $f_{3,T}(\hat{\lambda}_{0,T}) \xrightarrow{\mathcal{L}} p_0[M_0]_{11}$. The generalization to the vectors $\xi_{k,T}$ defined in the theorem shows no major difficulty.

4.7.4 Proof of Proposition 10

From Theorem 22, $\rho_k \downarrow b$ as $p_k \downarrow p_{\lim}$. Hence, by Lemma 2, $\Delta(\rho_k) \to 0$ as $p_k \downarrow p_{\lim}$. Moreover, the proof of this lemma shows that $|m(\rho_k)|$ remains bounded as $\rho_k \downarrow b$. Hence, since $\nu \neq \delta_0$ by Assumption 5, the integrals in the expression of α_k are lower bounded by a positive number as $p_k \downarrow p_{\lim}$. Thus, $\alpha_k \to \infty$ which proves the first part of the lemma.
When $p_k \to \infty$, $\rho_k/p_k \to 1$ and $\rho_k m(\rho_k) \to -1$. Taking $p_k \to \infty$ into the expressions of the integrals on the right hand sides of the expressions of α_k, β_k, and ϕ_k and recalling that $\Delta(\rho_k) \to 1$, we get $\alpha_k \to 0$, $\beta_k \to 1$, and $\phi_k \to 1$, which proves the lemma.

4.7.5 Proof of Lemma 3

The fluctuations of quadratic forms of the type $T^{-1}\boldsymbol{s}_T^H D_T \boldsymbol{s}_T$ where $\boldsymbol{s}_T \in \mathbb{C}^{T\times 1}$ has i.i.d. entries have been well studied (*e.g.* [13, Th. 2.1], [37, Th. 3]). Here, the vector s_T is replaced by the matrix $S_T \in \mathbb{C}^{T\times K}$ which introduces some differences in the proof. We follow here the lines of the proof of [37, Th. 3] and stress the main differences. The following lemma is of main importance being a direct application of the Cramér-Wold device:

Lemma 6. *Cramer-Wold device: application*
Let X_T be a sequence of $T\times T$ Hermitian matrices. Then

$$X_T \xrightarrow[T\to\infty]{\mathcal{L}} X$$

with X taken from the GUE if and only if for all $C \in \mathbb{H}^{T\times 1}$

$$\operatorname{Tr} CX_T \xrightarrow[T\to\infty]{\mathcal{L}} \operatorname{Tr} CX. \tag{4.18}$$

Let $S_T^{\mathsf{H}} = [\boldsymbol{s}_0, \cdots, \boldsymbol{s}_{T-1}]$ where $\boldsymbol{s}_t = [s^*_{t,0}, \ldots, s^*_{t,K-1}]^{\mathsf{T}}$ and let $C = [c_{ij}] \in \mathbb{C}^{K\times K}$ Hermitian matrix. Showing that

$$\sqrt{T}\operatorname{Tr} C\Big(\frac{1}{T}S_T^{\mathsf{H}} D_T S_T - \frac{1}{T}\operatorname{Tr} D_T I_K\Big) \xrightarrow[T\to\infty]{\mathcal{L}} \mathcal{N}\left(0, \beta\operatorname{Tr}(C^2) + \kappa\alpha\operatorname{Tr}[(\operatorname{diag}(C))^2]\right)$$

and invoking Lemma 6 establishes the lemma.

Consider the sequence of increasing σ-fields $\mathcal{F}_t = \sigma(\boldsymbol{s}_0, \ldots, \boldsymbol{s}_t)$, $t = 0, \ldots, T-1$, and denote $\mathbb{E}_t$ the expectation conditional to $\mathcal{F}_t$. Then, with $\mathbb{E}_0 = \mathbb{E}$,

$$\sqrt{T}\operatorname{Tr} C\left(\frac{1}{T}S_T^{\mathsf{H}} D_T S_T - \frac{1}{T}\operatorname{Tr} D_T I_K\right) = \frac{\sqrt{T}}{T}\sum_{t=0}^{T-1}(\mathbb{E}_{t+1} - \mathbb{E}_t)\operatorname{Tr} C S_T^{\mathsf{H}} D_T S_T$$

which is a sum of martingale increments, so that the key tool for establishing Lemma 3 is martingale CLT [16, Th. 35.12]. Writing $Z_t = \frac{1}{T}(\mathbb{E}_{t+1} - \mathbb{E}_t)\operatorname{Tr} C S_T^{\mathsf{H}} D_T S_T$, we need to show:

(i) *Lyapunov's condition* : there exists $\delta > 0$ for which

$$T^{1+\delta/2}\sum_{t=0}^{T-1}\mathbb{E}Z_t^{2+\delta} \xrightarrow[T\to\infty]{} 0.$$

(ii) The following convergence holds

$$T\sum_{t=0}^{T-1}\mathbb{E}_t Z_t^2 \xrightarrow[T\to\infty]{\mathcal{P}} \beta\,\mathrm{Tr}(C^2)+\kappa\alpha\,\mathrm{Tr}[(\mathrm{diag}(C))^2].$$

Proof of (i)

The proof is based on mimicking the calculus of [37, page 5058] (based on Burkholder's inequality and $\mathbb{E}|s_{0,0}|^8<\infty$). Taking $\delta=2$, we need to show:

$$\sup_T T^2\sum_{t=0}^{T-1}\mathbb{E}Z_t^4 \xrightarrow[T\to\infty]{\text{a.s.}} 0. \tag{4.19}$$

Denoting $D_T=[d_{ij}]$ and developing the conditional expectations, we obtain

$$TZ_t = d_{t+1,t+1}\,\mathrm{Tr}\,C(\boldsymbol{s}_{t+1}\boldsymbol{s}_{t+1}^{\mathsf{H}}-I_K)+2\Re\Big(\sum_{i,j=0}^{K-1}c_{i,j}\sum_{k=0}^{t}s_{k,j}^*\,s_{t+1,i}\,d_{k,t+1}\Big). \tag{4.20}$$

The Burkholder's inequality is defined by the following lemma:

Lemma 7. *Burkholder's inequality*
Let X_k be a complex martingale difference sequence with respect to the increasing sequence of σ-fields $\mathcal{F}_k$. Then for $p\geq 2$, there exists a constant α_p for which

$$\mathbb{E}\left|\sum_k X_k\right|^p \leq \alpha_p\left(\mathbb{E}\left(\sum_k \mathbb{E}_{k-1}|X_k|^2\right)^{p/2}+\mathbb{E}\sum_k|X_k|^p\right).$$

Developing (4.20) and applying Burkholder's inequality with the filtrations $\mathcal{F}_0,\ldots,\mathcal{F}_{t-1}$, after some calculations it is shown that

$$T^2\sum_{t=0}^{T-1}\mathbb{E}Z_t^4\leq\frac{\alpha}{T}$$

with α a constant depending on $\mathbb{E}|s_{0,0}|^8$. From $\mathbb{E}|s_{0,0}|^8<\infty$ Lyapunov's condition is proven.

Proof of (ii)

From (4.20), using the independence of the $s_{i,j}$ and the moments $\mathbb{E}s_{0,0} = 0$, $\mathbb{E}|s_{0,0}|^2 = 1$, and $\mathbb{E}[s_{0,0}^u (s_{0,0}^*)^v] = 0$ for $u \neq v$, after some calculations we obtain

$$T^2 \mathbb{E}_t Z_t^2 = d_{t+1,t+1}^2 \Big(\operatorname{Tr} C^2 + \kappa \sum_{k=0}^{K-1} c_{kk}^2 \Big) + 2 \sum_{i,j,n=0}^{K-1} c_{i,j} c_{n,i} \sum_{k,\ell=0}^{t} s_{k,j}^* \, s_{\ell,n} \, d_{k,t+1} d_{t+1,\ell}.$$

Let $\check{D}_T = [d_{ij} \mathbb{1}_{i>j}]$ be the matrix consisting in the strictly lower triangular part of D_T. Then we have

$$T \sum_{t=0}^{T-1} \mathbb{E}_t Z_t^2 = \Big(\operatorname{Tr} C^2 + \kappa \sum_{k=0}^{K-1} c_{kk}^2 \Big) \frac{1}{T} \operatorname{Tr}(\operatorname{diag}(D_T))^2 + \frac{2}{T} \operatorname{Tr} C S_T^{\mathsf{H}} \check{D}_T^{\mathsf{H}} \check{D}_T S_T C.$$

Using [4, Lemma 2.7] and [37, Lemma 3] (or [51, P. 278]), we then get

$$\frac{1}{T} \operatorname{Tr} C S_T^{\mathsf{H}} \check{D}_T^{\mathsf{H}} \check{D}_T S_T C - \operatorname{Tr} C^2 \frac{1}{T} \operatorname{Tr} \check{D}_T^{\mathsf{H}} \check{D}_T \xrightarrow[T\to\infty]{\mathcal{P}} 0.$$

We finally get the result by observing that

$$\frac{2}{T} \operatorname{Tr} \check{D}_T^{\mathsf{H}} \check{D}_T = \frac{1}{T} \operatorname{Tr} D_T^2 - \frac{1}{T} \operatorname{Tr}(\operatorname{diag}(D_T))^2.$$

4.7.6 Proof of Lemma 4

We recall first that from [21, Lemma 3.1], for any compact $K \subset \mathbb{R} - \operatorname{supp}(\mu)$, there exists $C > 0$ such that

$$\forall T \text{ large}, \ \forall t \in \operatorname{supp}(\nu_T), \ \inf_{x \in K} |1 + c_T m_T(x) t| > C.$$

From this result we then have

$$\liminf_T \inf_{t \in \operatorname{supp}(\nu_T)} |1 + c_T m_T(\rho_{k,T}) t| > 0.$$

Hence, $\limsup_T \|D_T\| < \infty$. Furthermore, since

$$\frac{1}{T} \operatorname{Tr}(D_T^2) = \int \frac{p_k^2 m_T(\rho_{k,T})^2}{(1 + c_T m_T(\rho_{k,T}) t)^2} \nu_T(dt)$$

the first convergence in the statement of Lemma 4 holds true.
As for the second convergence, recall that $R_T = [r_{t-n}]_{1 \leq t,n \leq T}$, with $\sum_t |r_t| < \infty$. Define the Toeplitz matrix

$$\Gamma_T \triangleq [\gamma_{t-n}]_{1 \leq t,n \leq T}$$

where $\gamma_\ell = \boldsymbol{\delta}_\ell + cm(\rho_k)r_\ell$. Observe that $D_T = p_k m_T(\rho_{k,T})\Gamma_T^{-1}$. Let $[\cdot]_T$ be the modulo-T operator, and let $\widetilde{\Gamma}_T = [\gamma_{[t-n]_T}]_{1\leq t,n\leq T}$ be a circulant matrix associated with Γ_T. By [21, Lemma 3.1] again,

$$\liminf_T \inf_{u\in[0,1]} (1 + c_T m_T(\rho_{k,T})|\mathbf{p}(\exp(2\imath\pi u))|^2) > 0. \tag{4.21}$$

Hence, $\sup_T \|\widetilde{\Gamma}_T\| < \infty$. It results that

$$\frac{1}{T}\|\Gamma_T^{-1} - \widetilde{\Gamma}_T^{-1}\|_{\text{fro}}^2 \to 0 \tag{4.22}$$

with $\|\cdot\|_{\text{fro}}$ the Frobenius norm [30, Th. 5.2]. On the other hand, since $\widetilde{\Gamma}_T$ is circulant, its eigenvector matrix is the Fourier $T \times T$ matrix, so that we can show

$$\text{diag}(\widetilde{\Gamma}_T^{-1}) = \left(\frac{1}{T}\sum_{t=0}^{T-1} \frac{1}{1 + c_T m_T(\rho_{k,T})|\mathbf{p}(\exp(2\imath\pi t/T))|^2}\right) I_T.$$

The lemma is obtained by combining these last two results.

4.7.7 Proof of Lemma 5

We essentially show that we can replace the $B_{k,T}$ by $\sqrt{p_k}S_{k,T}$ with $S_T = [S_{0,T}, \ldots, S_{t,T}]$, similar to B_T. Since $\theta_i \neq \theta_j$ if $i \neq j$, from the definition of the vector function $a(\theta)$, we have

$$[H_T^{\mathsf{H}} H_T]_{k,\ell} - \boldsymbol{\delta}_{k\ell} = a_T(\theta_k)^{\mathsf{H}} a_T(\theta_\ell) - \boldsymbol{\delta}_{k\ell} = \mathcal{O}(1/T).$$

Hence,

$$(H_T^{\mathsf{H}} H_T)^{1/2} \triangleq I_K + E_T$$

where $\|E_T\| = \mathcal{O}(1/T)$. Given any sequence D_T of deterministic matrices such that $\sup_T \|D_T\| < \infty$, it can be seen by a moment derivation with respect to the law of S_T that

$$\mathbb{E}\left|\left[\frac{1}{T}B_T^{\mathsf{H}} D_T B_T - P^{1/2} S_T^{\mathsf{H}} D_T S_T P^{1/2}\right]_{k,\ell}\right| = \mathcal{O}(1/T)$$

for any $k, \ell \leq K-1$. Hence, by Markov's inequality,

$$\sqrt{T}\left(\frac{1}{T}B_T^{\mathsf{H}} D_T B_T - P^{1/2} S_T^{\mathsf{H}} D_T S_T P^{1/2}\right) \xrightarrow{\mathcal{P}} 0$$

Replacing D_T with any of the matrices $p_k m_T(\rho_{k,T})(I_T + c_T m_T(\rho_{k,T})R_T)^{-1}$, we get from Lemma 4 that $\sup_T \|D_T\| < \infty$. Therefore, the $B_{k,T}$ can be replaced with the $\sqrt{p_k}S_{k,T}$. The result is then obtained upon applying Lemmas 3 and 4 and recalling that, for $k = 0, \ldots, t$, the $S_{k,T}$ are independent.

Chapter 5

Estimation of Toeplitz covariance matrices and application to source detection

5.1 Introduction

Let $(v_t)_{t\in\mathbb{Z}}$ be a complex circularly symmetric Gaussian stationary process with zero mean and covariance function $(r_k)_{k\in\mathbb{Z}}$ with $r_k = \mathbb{E}[v_{t+k}v_t^*]$ and $r_k \to 0$ as $k \to \infty$. We observe N independent copies of $(v_t)_{t\in\mathbb{Z}}$ over the time window $t \in \{0,\dots,T-1\}$, and stack the observations in a matrix $V_T = [v_{n,t}]_{n,t=0}^{N-1,T-1}$. This matrix can be written as $V_T = W_T R_T^{1/2}$, where $W_T \in \mathbb{C}^{N\times T}$ has independent standard circularly symmetric complex Gaussian entries and $R_T^{1/2}$ is any square root of the Hermitian nonnegative definite Toeplitz $T \times T$ matrix

$$R_T \triangleq [r_{i-j}]_{0\le i,j\le T-1} = \begin{bmatrix} r_0 & r_1 & \dots & r_{T-1} \\ r_{-1} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & r_1 \\ r_{1-T} & \dots & r_{-1} & r_0 \end{bmatrix}.$$

Recently this estimation problem has drawn a renewed attention considering the high dimensional setting for which both N and T are large. Generally estimation methods of R_T rely on the classical biased and unbi-

ased estimates $\hat{r}^b_{k,T}$ and $\hat{r}^u_{k,T}$ of r_k, respectively defined by

$$\hat{r}^b_{k,T} = \frac{1}{NT}\sum_{n=0}^{N-1}\sum_{t=0}^{T-1} v_{n,t+k}v^*_{n,t}\mathbb{1}_{0\leq t+k\leq T-1}$$
$$\hat{r}^u_{k,T} = \frac{1}{N(T-|k|)}\sum_{n=0}^{N-1}\sum_{t=0}^{T-1} v_{n,t+k}v^*_{n,t}\mathbb{1}_{0\leq t+k\leq T-1}$$

where $\mathbb{1}_A$ is the indicator function on the set A. Depending on the relative rate of growth of N and T, the estimates $\widehat{R}^b_T = [\hat{r}^b_{i-j,T}]_{0\leq i,j\leq T-1}$ and $\widehat{R}^u_T = [\hat{r}^u_{i-j,T}]_{0\leq i,j\leq T-1}$ may not be consistent. The estimation approaches developed during the last decade propose all to build banded or tapered versions of the estimated matrix $\widehat{R}_T$ by down-scaling estimates of entries sufficiently away from the diagonal [71, 15, 72, 19, 18]. These give rise to the consistent estimate $\widehat{R}_{\gamma,T} = [[\widehat{R}_T]_{i,j}\mathbb{1}_{|i-j|\leq\gamma}]$ for some well-chosen functions $\gamma(T)$ usually satisfying $\gamma(T)\to\infty$ and $\gamma(T)/T\to 0$. These methods however suffer from the following main limitations:

(i) they assume the *a priori* knowledge of the rate of decrease of r_k (and restrict these rates to specific classes);

(i) the results are asymptotic in nature and do not provide explicit rules for selecting $\gamma(T)$ for practical finite values of N and T;

(i) the operations of banding and tapering do not guarantee the positive definiteness of the resulting covariance estimate.

The propose of this chapter is to consider non banded estimates of R_T. The only assumption on r_k is that $\sum_{k=-\infty}^{\infty}|r_k|<\infty$. The consistency of the non banded estimates of R_T given by $\widehat{R}^b_T$ and $\widehat{R}^u_T$ is obtained thanks to the choice $N,T\to\infty$ with $N/T\to c\in(0,\infty)$. This setting is more practical in applications as long as both the finite values N and T are large and of the same order of magnitude. The contribution of this work consists in the establishment of concentration inequalities for the errors in spectral norm $\|R_T-\widehat{R}^b_T\|$ and $\|R_T-\widehat{R}^u_T\|$. The results are then generalized to the case where V_T is replaced by V_T+P_T for a rank-one matrix P_T and we show that the concentration inequalities remain identical. As an application of the latter, we study a single source detection (modeled through P_T) by an array of N sensors embedded in a temporally correlated noise (modeled by V_T) performed in two steps. First, the matrix R_T is estimated from V_T+P_T giving $\widehat{R}^b_T$ or $\widehat{R}^u_T$ which are both nonnegative definite with probability one.

Then this estimate is used as a whitening matrix, before applying a GLRT procedure on the whitened observation.

5.2 Performance of the estimators

5.2.1 Model and assumptions

Let $(r_k)_{k\in\mathbb{Z}}$ be a doubly infinite sequence of covariance coefficients. For any $T \in \mathbb{N}$, let

$$R_T = \mathcal{T}(r_{-(T-1)}, \ldots, r_{T-1}) = \begin{bmatrix} r_0 & r_1 & \ldots & r_{T-1} \\ r_{-1} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & r_1 \\ r_{1-T} & \ldots & r_{-1} & r_0 \end{bmatrix}.$$

a Hermitian nonnegative definite matrix. Given $N = N(T) > 0$, consider the matrix model

$$V_T = [v_{n,t}]_{n,t=0}^{N-1,T-1} = W_T R_T^{1/2} \tag{5.1}$$

where $W_T = [w_{n,t}]_{n,t=0}^{N-1,T-1}$ has independent $\mathcal{CN}(0,1)$ entries. It is clear that $r_k = \mathbb{E}[v_{n,t+k} v_{n,t}^*]$ for any t, k, and $n \in \{0, \ldots, N-1\}$.

Define the covariance function for $\lambda \in [0, 2\pi)$

$$\boldsymbol{\Upsilon}(\lambda) \triangleq \sum_{k=-\infty}^{\infty} r_k e^{-\imath k\lambda}. \tag{5.2}$$

We assume in the remainder of this chapter that the covariance coefficients r_k are absolutely summable and $r_0 \neq 0$. Hence, $\boldsymbol{\Upsilon}(\lambda)$ is continuous on the interval $[0, 2\pi]$. Since $\|R_T\| \leq \|\boldsymbol{\Upsilon}\|_\infty$ (see *e.g.* [30, Lemma 4.1]), the absolute summability of r_k implies that $\sup_T \|R_T\| < \infty$. We moreover assume the asymptotic regime denoted as "$T \to \infty$" for which $N/T \to c > 0$ as $T \to \infty$.

The following section presents the main results presented in Theorems 28 and 29.

5.2.2 Main results

The aim is to study the performances of the estimators of the covariance matrices. These are based on two frequently considered estimates of r_k

defined respectively by

$$\hat{r}^b_{k,T} = \frac{1}{NT} \sum_{n=0}^{N-1} \sum_{t=0}^{T-1} v_{n,t+k} v^*_{n,t} \mathbb{1}_{0 \leq t+k \leq T-1} \tag{5.3}$$

$$\hat{r}^u_{k,T} = \frac{1}{N(T-|k|)} \sum_{n=0}^{N-1} \sum_{t=0}^{T-1} v_{n,t+k} v^*_{n,t} \mathbb{1}_{0 \leq t+k \leq T-1}. \tag{5.4}$$

Since $\mathbb{E}\hat{r}^b_{k,T} = (1-|k|/T) r_k$ and $\mathbb{E}\hat{r}^u_{k,T} = r_k$, the estimate $\hat{r}^b_{k,T}$ is biased while $\hat{r}^u_{k,T}$ is unbiased.

Define also the estimates of R_T which are Toeplitz matrices formed from $\hat{r}^b_{k,T}$ and $\hat{r}^u_{k,T}$, respectively:

$$\widehat{R}^b_T \triangleq \mathcal{T}\left(\hat{r}^b_{-(T-1),T}, \ldots, \hat{r}^b_{(T-1),T}\right) \tag{5.5}$$

$$\widehat{R}^u_T \triangleq \mathcal{T}\left(\hat{r}^u_{-(T-1),T}, \ldots, \hat{r}^u_{(T-1),T}\right). \tag{5.6}$$

A well known advantage of $\widehat{R}^b_T$ over $\widehat{R}^u_T$ as an estimate of R_T is its structural nonnegative definiteness.

The results on the spectral behavior of the matrices $\widehat{R}^b_T$ and $\widehat{R}^u_T$ are provided under the form of concentration inequalities on $\|\widehat{R}^b_T - R_T\|$ and $\|\widehat{R}^u_T - R_T\|$ and given in the following two theorems:

Theorem 28. *Assume r_k are absolutely summable and $r_0 \neq 0$. Let $T \to \infty$ and $N/T \to c > 0$. Let $\widehat{R}^b_T$ be defined as in (5.5). Then, for any $x > 0$,*

$$\mathbb{P}\left[\left\|\widehat{R}^b_T - R_T\right\| > x\right] \leq \exp\left(-cT\left(\frac{x}{\|\boldsymbol{\Upsilon}\|_\infty} - \log\left(1 + \frac{x}{\|\boldsymbol{\Upsilon}\|_\infty}\right) + o(1)\right)\right)$$

where $o(1)$ is with respect to T and depends on x.

Theorem 29. *Assume r_k are absolutely summable and $r_0 \neq 0$. Let $T \to \infty$ and $N/T \to c > 0$. Let $\widehat{R}^u_T$ be defined as in (5.6). Then, for any $x > 0$,*

$$\mathbb{P}\left[\left\|\widehat{R}^u_T - R_T\right\| > x\right] \leq \exp\left(-\frac{cTx^2}{4\|\boldsymbol{\Upsilon}\|^2_\infty \log T}(1 + o(1))\right)$$

where $o(1)$ is with respect to T and depends on x.

A consequence of these theorems, obtained by the Borel-Cantelli lemma, is that $\|\widehat{R}^b_T - R_T\| \to 0$ and $\|\widehat{R}^u_T - R_T\| \to 0$ almost surely as $T \to \infty$. The

slower rate of decrease of $T/\log(T)$ in the unbiased estimator exponent may be interpreted by the increased inaccuracy in the estimates of r_k for values of k close to $T-1$.

We now turn to the proofs of Theorems 28 and 29, starting with some basic mathematical results that will be needed throughout the proofs.

5.2.3 Some basic mathematical results

Lemma 8. *For $x, y \in \mathbb{C}^{m\times 1}$ and $A \in \mathbb{C}^{m\times m}$,*

$$\left|x^{\mathsf{H}}Ax - y^{\mathsf{H}}Ay\right| \leq \|A\| \left(\|x\| + \|y\|\right) \|x - y\| .$$

Proof.

$$\begin{aligned}\left|x^{\mathsf{H}}Ax - y^{\mathsf{H}}Ay\right| &= \left|x^{\mathsf{H}}Ax - y^{\mathsf{H}}Ax + y^{\mathsf{H}}Ax - y^{\mathsf{H}}Ay\right| \\ &\leq \left|(x-y)^{\mathsf{H}}Ax\right| + \left|y^{\mathsf{H}}A(x-y)\right| \\ &\leq \|A\| \left(\|x\| + \|y\|\right) \|x - y\| .\end{aligned}$$

□

Lemma 9. *Let $X_0, \ldots, X_{M-1}$ be independent $\mathcal{CN}(0,1)$ random variables. Then, for any $x > 0$,*

$$\mathbb{P}\left[\frac{1}{M}\sum_{m=0}^{M-1}(|X_m|^2 - 1) > x\right] \leq \exp\left(-M(x - \log(1+x))\right) .$$

Proof. This is a classical Chernoff bound. Indeed, given $\xi \in (0,1)$, we have by the Markov inequality

$$\begin{aligned}\mathbb{P}\Big[\frac{1}{M}\sum_{m=0}^{M-1}(|X_m|^2 - 1) > x\Big] &= \mathbb{P}\left[\exp\left(\xi\sum_{m=0}^{M-1}|X_m|^2\right) > \exp\xi M(x+1)\right] \\ &\leq \exp(-\xi M(x+1))\mathbb{E}\left[\exp\left(\xi\sum_{m=0}^{M-1}|X_m|^2\right)\right] \\ &= \exp\left(-M\left(\xi(x+1) + \log(1-\xi)\right)\right)\end{aligned}$$

since $\mathbb{E}\left[\exp(\xi|X_m|^2)\right] = 1/(1-\xi)$. The result follows upon minimizing this expression with respect to ξ. □

We turn now to the proof of Theorem 28 given in the following section.

5.2.4 Biased estimator: proof of Theorem 28

Define the following functions

$$\widehat{\Upsilon}_T^b(\lambda) \triangleq \sum_{k=-(T-1)}^{T-1} \hat{r}_{k,T}^b e^{\imath k\lambda}$$
$$\Upsilon_T(\lambda) \triangleq \sum_{k=-(T-1)}^{T-1} r_k e^{\imath k\lambda}.$$

Since $\widehat{R}_T^b - R_T$ is a Toeplitz matrix, from [30, Lemma 4.1],

$$\begin{aligned}\left\|\widehat{R}_T^b - R_T\right\| &\leq \sup_{\lambda\in[0,2\pi)} \left|\widehat{\Upsilon}_T^b(\lambda) - \Upsilon_T(\lambda)\right| \\ &\leq \sup_{\lambda\in[0,2\pi)} \left|\widehat{\Upsilon}_T^b(\lambda) - \mathbb{E}\widehat{\Upsilon}_T^b(\lambda)\right| + \sup_{\lambda\in[0,2\pi)} \left|\mathbb{E}\widehat{\Upsilon}_T^b(\lambda) - \Upsilon_T(\lambda)\right|.\end{aligned}$$

By Kronecker's lemma ([36, Lemma 3.21]), the rightmost term at the right-hand side satisfies

$$\left|\mathbb{E}\widehat{\Upsilon}_T^b(\lambda) - \Upsilon_T(\lambda)\right| \leq \sum_{k=-(T-1)}^{T-1} \frac{|kr_k|}{T} \xrightarrow[T\to\infty]{} 0. \tag{5.7}$$

In order to deal with the term $\sup_{\lambda\in[0,2\pi)} |\widehat{\Upsilon}_T^b(\lambda) - \mathbb{E}\widehat{\Upsilon}_T^b(\lambda)|$, two ingredients will be used. The first one is based on the important fact that the terms $\widehat{\Upsilon}_T^b(\lambda)$ and $\mathbb{E}\widehat{\Upsilon}_T^b(\lambda)$ can be written in a quadratic form. This is given in the following lemma (proven in Appendix 5.7.1):

Lemma 10. *The following facts hold:*

$$\widehat{\Upsilon}_T^b(\lambda) = d_T(\lambda)^{\mathsf{H}} \frac{V_T^{\mathsf{H}} V_T}{N} d_T(\lambda)$$
$$\mathbb{E}\widehat{\Upsilon}_T^b(\lambda) = d_T(\lambda)^{\mathsf{H}} R_T d_T(\lambda)$$

where $d_T(\lambda) = 1/\sqrt{T}\left[1, e^{-\imath\lambda}, \ldots, e^{-\imath(T-1)\lambda}\right]^{\mathsf{T}}$.

The second ingredient is a Lipschitz property of the function $\|d_T(\lambda) - d_T(\lambda')\|$ seen as a function of λ. From the inequality $|e^{-\imath t\lambda} - e^{-\imath t\lambda'}| \leq t|\lambda - \lambda'|$, we indeed have

$$\|d_T(\lambda) - d_T(\lambda')\| = \sqrt{\frac{1}{T}\sum_{t=0}^{T-1} |e^{-\imath t\lambda} - e^{-\imath t\lambda'}|^2} \leq \frac{T|\lambda - \lambda'|}{\sqrt{3}}. \tag{5.8}$$

In order to control the term $\sup_{\lambda\in[0,2\pi)} \left|\widehat{\Upsilon}_T^b(\lambda) - \mathbb{E}\widehat{\Upsilon}_T^b(\lambda)\right|$ which is a function of λ we need to discretize the interval $[0, 2\pi]$. Denote by $\lfloor\cdot\rfloor$ the floor function and choose $\beta > 2$. Define the interval $\mathcal{I} = \left\{0, \ldots, \lfloor T^\beta \rfloor - 1\right\}$. Let $\lambda_i = 2\pi\frac{i}{\lfloor T^\beta\rfloor}$, $i \in \mathcal{I}$, be a regular discretization of the interval $[0, 2\pi]$. We write

$$\begin{aligned}
&\sup_{\lambda\in[0,2\pi)} \left|\widehat{\Upsilon}_T^b(\lambda) - \mathbb{E}\widehat{\Upsilon}_T^b(\lambda)\right| \\
&\leq \max_{i\in\mathcal{I}} \sup_{\lambda\in[\lambda_i,\lambda_{i+1}]} \Big(\left|\widehat{\Upsilon}_T^b(\lambda) - \widehat{\Upsilon}_T^b(\lambda_i)\right| + \left|\widehat{\Upsilon}_T^b(\lambda_i) - \mathbb{E}\widehat{\Upsilon}_T^b(\lambda_i)\right| \\
&+ \left|\mathbb{E}\widehat{\Upsilon}_T^b(\lambda_i) - \mathbb{E}\widehat{\Upsilon}_T^b(\lambda)\right|\Big) \leq \max_{i\in\mathcal{I}} \sup_{\lambda\in[\lambda_i,\lambda_{i+1}]} \left|\widehat{\Upsilon}_T^b(\lambda) - \widehat{\Upsilon}_T^b(\lambda_i)\right| \\
&+ \max_{i\in\mathcal{I}} \left|\widehat{\Upsilon}_T^b(\lambda_i) - \mathbb{E}\widehat{\Upsilon}_T^b(\lambda_i)\right| + \max_{i\in\mathcal{I}} \sup_{\lambda\in[\lambda_i,\lambda_{i+1}]} \left|\mathbb{E}\widehat{\Upsilon}_T^b(\lambda_i) - \mathbb{E}\widehat{\Upsilon}_T^b(\lambda)\right| \\
&\triangleq \chi_1 + \chi_2 + \chi_3.
\end{aligned}$$

With the help of Lemma 10 and (5.8), we need to derive concentration inequalities on the random terms χ_1 and χ_2 and to provide a bound on the deterministic term χ_3. This is the purpose of the three following lemmas. Herein and in the remainder of this chapter, C denotes a positive constant independent of T. This constant can change from an expression to another.

Lemma 11. *There exists a constant $C > 0$ such that for any $x > 0$ and any T large enough,*

$$\mathbb{P}\left[\chi_1 > x\right] \leq \exp\left(-cT^2\left(\frac{xT^{\beta-2}}{C\|\boldsymbol{\Upsilon}\|_\infty} - \log\frac{xT^{\beta-2}}{C\|\boldsymbol{\Upsilon}\|_\infty} - 1\right)\right).$$

Proof. Using Lemmas 10 and 8 along with (5.8), we have

$$\begin{aligned}
\left|\widehat{\Upsilon}_T^b(\lambda) - \widehat{\Upsilon}_T^b(\lambda_i)\right| &= \left|d_T(\lambda)^{\mathsf{H}}\frac{V_T^{\mathsf{H}}V_T}{N}d_T(\lambda) - d_T(\lambda_i)^{\mathsf{H}}\frac{V_T^{\mathsf{H}}V_T}{N}d_T(\lambda_i)\right| \\
&\leq 2N^{-1}\left\|d_T(\lambda) - d_T(\lambda_i)\right\|\left\|R_T\right\|\left\|W_T^{\mathsf{H}}W_T\right\| \\
&\leq C|\lambda - \lambda_i|\|\boldsymbol{\Upsilon}\|_\infty\left\|W_T^{\mathsf{H}}W_T\right\|.
\end{aligned}$$

From $\|W_T^{\mathsf{H}}W_T\| \leq \mathrm{Tr}(W_T^{\mathsf{H}}W_T)$ and Lemma 9, assuming T large enough so

that $f(x,T) \triangleq xT^{\beta-1}/(CN\|\boldsymbol{\Upsilon}\|_\infty)$ satisfies $f(x,T) \geq 1$, we then obtain

$$\begin{aligned}\mathbb{P}\left[\chi_1 > x\right] &\leq \mathbb{P}\left[C\|\boldsymbol{\Upsilon}\|_\infty T^{-\beta}\sum_{t=0}^{T-1}\sum_{n=0}^{N-1}|w_{n,t}|^2 > x\right] \\ &= \mathbb{P}\left[\frac{1}{NT}\sum_{t=0}^{T-1}\sum_{n=0}^{N-1}(|w_{n,t}|^2-1) > f(x,T)-1\right] \\ &\leq \exp(-NT(f(x,T)-\log f(x,T)-1)).\end{aligned}$$

□

Lemma 12. *The following inequality holds*

$$\mathbb{P}\left[\chi_2 > x\right] \leq 2T^\beta \exp\left(-cT\left(\frac{x}{\|\boldsymbol{\Upsilon}\|_\infty} - \log\Big(1+\frac{x}{\|\boldsymbol{\Upsilon}\|_\infty}\Big)\right)\right).$$

Proof. From the union bound we obtain:

$$\mathbb{P}\left[\chi_2 > x\right] \leq \sum_{i=0}^{\lfloor T^\beta \rfloor - 1} \mathbb{P}\left[\left|\widehat{\Upsilon}_T^b(\lambda_i) - \mathbb{E}\widehat{\Upsilon}_T^b(\lambda_i)\right| > x\right].$$

We shall bound each term of the sum separately. Since

$$\begin{aligned}\mathbb{P}\left[\left|\widehat{\Upsilon}_T^b(\lambda_i) - \mathbb{E}\widehat{\Upsilon}_T^b(\lambda_i)\right| > x\right] &= \mathbb{P}\left[\widehat{\Upsilon}_T^b(\lambda_i) - \mathbb{E}\widehat{\Upsilon}_T^b(\lambda_i) > x\right] \\ &\quad + \mathbb{P}\left[-\left(\widehat{\Upsilon}_T^b(\lambda_i) - \mathbb{E}\widehat{\Upsilon}_T^b(\lambda_i)\right) > x\right]\end{aligned}$$

it will be enough to deal with the first right-hand side term as the second one is treated similarly. Let $\eta_T(\lambda_i) \triangleq W_T q_T(\lambda_i) = [\eta_{0,T}(\lambda_i), \ldots, \eta_{N-1,T}(\lambda_i)]^\mathsf{T}$ where $q_T(\lambda_i) \triangleq R_T^{1/2} d_T(\lambda_i)$. Observe that $\eta_{k,T}(\lambda_i) \sim \mathcal{CN}(0, \|q_T(\lambda_i)\|^2 I_N)$. We know from Lemma 10 that

$$\widehat{\Upsilon}_T^b(\lambda_i) - \mathbb{E}\widehat{\Upsilon}_T^b(\lambda_i) = \frac{1}{N}\left(\|\eta_T(\lambda_i)\|^2 - \mathbb{E}\|\eta_T(\lambda_i)\|^2\right). \tag{5.9}$$

From (5.9) and Lemma 9, we therefore get

$$\mathbb{P}\left[\widehat{\Upsilon}_T^b(\lambda_i) - \mathbb{E}\widehat{\Upsilon}_T^b(\lambda_i) > x\right] \leq \exp\left(-N\left(\frac{x}{\|q_T(\lambda_i)\|^2} - \log\Big(1+\frac{x}{\|q_T(\lambda_i)\|^2}\Big)\right)\right).$$

Noticing that $\|q_T(\lambda_i)\|^2 \leq \|\boldsymbol{\Upsilon}\|_\infty$ and that the function $f(x) = x - \log\Big(1+x\Big)$ is increasing for $x > 0$, we get the result. □

Finally, the bound for the deterministic term χ_3 is provided by the following lemma:

Lemma 13. $\chi_3 \leq C\|\boldsymbol{\Upsilon}\|_\infty T^{-\beta+1}$.

Proof. From Lemmas 10 and 8 along with (5.8), we obtain

$$\begin{aligned}\left|\mathbb{E}\widehat{\Upsilon}_T^b(\lambda) - \mathbb{E}\widehat{\Upsilon}_T^b(\lambda_i)\right| &= \left|d_T(\lambda)^{\mathsf{H}} R_T d_T(\lambda) - d_T(\lambda_i)^{\mathsf{H}} R_T d_T(\lambda_i)\right| \\ &\leq 2\,\|R_T\|\,\|d_T(\lambda) - d_T(\lambda_i)\| \\ &\leq C\|\boldsymbol{\Upsilon}\|_\infty |\lambda - \lambda_i| T.\end{aligned}$$

From $\max\limits_{i\in\mathcal{I}} \sup\limits_{\lambda\in[\lambda_i,\lambda_{i+1}]} |\lambda - \lambda_i| = \lambda_{i+1} - \lambda_i = T^{-\beta}$ we get the result. □

We now complete the proof of Theorem 28. From (5.7) and Lemma 13, we get

$$\mathbb{P}\left[\left\|\widehat{R}_T^b - R_T\right\| > x\right] = \mathbb{P}\left[\chi_1 + \chi_2 > x + o(1)\right].$$

Given a parameter $\epsilon_T \in [0,1]$, we can write (with some slight notation abuse)

$$\mathbb{P}\left[\chi_1 + \chi_2 > x + o(1)\right] \leq \mathbb{P}\left[\chi_1 > x\epsilon_T\right] + \mathbb{P}\left[\chi_2 > x(1-\epsilon_T) + o(1)\right].$$

With the results of Lemmas 11 and 12, setting $\epsilon_T = 1/T$, we get

$$\begin{aligned}\mathbb{P}\left[\chi_1 + \chi_2 > x + o(1)\right] &\leq \mathbb{P}\left[\chi_1 > \frac{x}{T}\right] + \mathbb{P}\left[\chi_2 > x(1-\frac{x}{T}) + o(1)\right] \\ &\leq \exp\Big(-cT^2\Big(\frac{xT^{\beta-3}}{C\|\boldsymbol{\Upsilon}\|_\infty} - \log\frac{xT^{\beta-3}}{C\|\boldsymbol{\Upsilon}\|_\infty} - 1\Big)\Big) \\ &+ \exp\Big(-cT\Big(\frac{x\left(1-\frac{1}{T}\right)}{\|\boldsymbol{\Upsilon}\|_\infty} - \log\Big(1 + \frac{x\left(1-\frac{1}{T}\right)}{\|\boldsymbol{\Upsilon}\|_\infty}\Big) + o(1)\Big)\Big) \\ &= \exp\Big(-cT\Big(\frac{x}{\|\boldsymbol{\Upsilon}\|_\infty} - \log\Big(1 + \frac{x}{\|\boldsymbol{\Upsilon}\|_\infty}\Big) + o(1)\Big)\Big)\end{aligned}$$

since $\beta > 2$.

5.2.5 Unbiased estimator: proof of Theorem 29

The proof follows basically the same main steps as for Theorem 28 with an additional difficulty due to the scaling terms $1/(T-|k|)$.

Defining the function

$$\widehat{\Upsilon}_T^u(\lambda) \triangleq \sum_{k=-(T-1)}^{T-1} \hat{r}_{k,T}^u e^{ik\lambda}$$

we have

$$\begin{aligned}\left\|\widehat{R}_T^u - R_T\right\| &\leq \sup_{\lambda\in[0,2\pi)} \left|\widehat{\Upsilon}_T^u(\lambda) - \Upsilon_T(\lambda)\right| \\ &= \sup_{\lambda\in[0,2\pi)} \left|\widehat{\Upsilon}_T^u(\lambda) - \mathbb{E}\widehat{\Upsilon}_T^u(\lambda)\right|\end{aligned}$$

since $\Upsilon_T(\lambda) = \mathbb{E}\widehat{\Upsilon}_T^u(\lambda)$, the estimates $\hat{r}_{k,T}^u$ being unbiased.

In order to deal with the right-hand side of this expression, we need the following analogue of Lemma 10, borrowed from [65] and proven here in Appendix 5.7.2.

Lemma 14. *The following fact holds:*

$$\widehat{\Upsilon}_T^u(\lambda) = d_T(\lambda)^{\mathsf{H}} \left(\frac{V_T^{\mathsf{H}} V_T}{N} \odot B_T\right) d_T(\lambda)$$

where $\odot$ is the Hadamard product of matrices and where

$$B_T \triangleq \left[\frac{T}{T-|i-j|}\right]_{0\leq i,j\leq T-1} = \begin{bmatrix} 1 & \frac{T}{T-1} & \dots & T \\ \frac{T}{T-1} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \frac{T}{T-1} \\ T & \dots & \frac{T}{T-1} & 1 \end{bmatrix}.$$

In order to make $\widehat{\Upsilon}_T^u(\lambda)$ more tractable, we rely on the following lemma which can be proven by direct calculation.

Lemma 15. *Let $x,\ y \in \mathbb{C}^m$ and $A, B \in C^{m\times m}$. Then*

$$x^{\mathsf{H}}(A \odot B)y = \mathrm{Tr}(D_x^{\mathsf{H}} A D_y B^{\mathsf{T}})$$

where we recall $D_x = \mathrm{diag}(x)$ and $D_y = \mathrm{diag}(y)$.

Denoting

$$\begin{aligned} D_T(\lambda) &\triangleq \mathrm{diag}(d_T(\lambda)) = \frac{1}{\sqrt{T}}\,\mathrm{diag}(1, e^{i\lambda}, \dots, e^{i(T-1)\lambda}) \\ Q_T(\lambda) &\triangleq R_T^{1/2} D_T(\lambda) B_T D_T(\lambda)^{\mathsf{H}} (R_T^{1/2})^{\mathsf{H}} \end{aligned}$$

we get from Lemmas 14 and 15

$$\begin{aligned}\widehat{\Upsilon}_T^u(\lambda) &= \frac{1}{N}\,\mathrm{Tr}(D_T(\lambda)^{\mathsf{H}} (R_T^{1/2})^{\mathsf{H}} W_T^{\mathsf{H}} W_T R_T^{1/2} D_T(\lambda) B_T) \\ &= \frac{1}{N}\,\mathrm{Tr}(W_T Q_T(\lambda) W_T^{\mathsf{H}}) \\ &= \frac{1}{N}\sum_{n=0}^{N-1} w_n^{\mathsf{H}} Q_T(\lambda) w_n \end{aligned} \tag{5.10}$$

where w_i^{H} is such that $W_T = [w_0^{\mathsf{H}}, \ldots, w_{N-1}^{\mathsf{H}}]$.

Compared to the biased case, the main difficulty lies here in the fact that the matrices B_T/T and $Q_T(\lambda)$ have unbounded spectral norm as $T \to \infty$. Note, that Lemma 15 allows us to deal with Q_T with a reduced spectral norm by a factor of order T instead of B_T. The following lemma, proven in Appendix 5.7.3, provides some information on the spectral behavior of these two matrices that will be used subsequently.

Lemma 16. *The matrix B_T satisfies*

$$\|B_T\| \leq \sqrt{2}T(\sqrt{\log T} + C). \tag{5.11}$$

For any $\lambda \in [0, 2\pi)$, the eigenvalues $\sigma_0, \ldots, \sigma_{T-1}$ of the matrix $Q(\lambda)$ satisfy the following inequalities:

$$\sum_{t=0}^{T-1} \sigma_t^2 \leq 2\|\boldsymbol{\Upsilon}\|_\infty^2 \log T + C \tag{5.12}$$

$$\max_t |\sigma_t| \leq \sqrt{2}\|\boldsymbol{\Upsilon}\|_\infty (\log T)^{1/2} + C \tag{5.13}$$

$$\sum_{t=0}^{T-1} |\sigma_t|^3 \leq C((\log T)^{3/2} + 1) \tag{5.14}$$

where the constant C is independent of λ.

We shall also need the following easily shown Lipschitz property of the function $\|D_T(\lambda) - D_T(\lambda')\|$:

$$\|D_T(\lambda) - D_T(\lambda')\| \leq \sqrt{T}|\lambda - \lambda'|. \tag{5.15}$$

We now enter the core of the proof of Theorem 29. Choosing $\beta > 2$, let $\lambda_i = 2\pi \frac{i}{\lfloor T^\beta \rfloor}$, $i \in \mathcal{I}$, be a regular discretization of the interval $[0, 2\pi]$ with $\mathcal{I} = \{0, \ldots, \lfloor T^\beta \rfloor - 1\}$. We write

$$\begin{aligned}
&\sup_{\lambda \in [0,2\pi)} \left| \widehat{\Upsilon}_T^u(\lambda) - \mathbb{E}\widehat{\Upsilon}_T^u(\lambda) \right| \\
&\leq \max_{i \in \mathcal{I}} \sup_{\lambda \in [\lambda_i, \lambda_{i+1}]} \left| \widehat{\Upsilon}_T^u(\lambda) - \widehat{\Upsilon}_T^u(\lambda_i) \right| + \max_{i \in \mathcal{I}} \left| \widehat{\Upsilon}_T^u(\lambda_i) - \mathbb{E}\widehat{\Upsilon}_T^u(\lambda_i) \right| \\
&+ \max_{i \in \mathcal{I}} \sup_{\lambda \in [\lambda_i, \lambda_{i+1}]} \left| \mathbb{E}\widehat{\Upsilon}_T^u(\lambda_i) - \mathbb{E}\widehat{\Upsilon}_T^u(\lambda) \right| \\
&\triangleq \chi_1 + \chi_2 + \chi_3.
\end{aligned}$$

Our task is now to provide concentration inequalities on the random terms χ_1 and χ_2 and a bound on the deterministic term χ_3.

Lemma 17. *There exists a constant $C > 0$ such that, if T is large enough, the following inequality holds:*

$$\mathbb{P}\left[\chi_1 > x\right] \leq \exp\left(-cT^2\left(\frac{xT^{\beta-2}}{C\sqrt{\log T}} - \log\frac{xT^{\beta-2}}{C\sqrt{\log T}} - 1\right)\right).$$

Proof. From Equation (5.10), we have

$$\begin{aligned}\left|\widehat{\Upsilon}_T^u(\lambda) - \widehat{\Upsilon}_T^u(\lambda_i)\right| &= \frac{1}{N}\left|\sum_{n=0}^{N-1} w_n^{\mathsf{H}}\left(Q_T(\lambda) - Q_T(\lambda_i)\right) w_n\right| \\ &\leq \frac{1}{N}\sum_{n=0}^{N-1}\left|w_n^{\mathsf{H}}\left(Q_T(\lambda) - Q_T(\lambda_i)\right) w_n\right| \\ &\leq \frac{1}{N}\left\|Q_T(\lambda) - Q_T(\lambda_i)\right\|\sum_{n=0}^{N-1}\left\|w_n\right\|^2.\end{aligned}$$

The norm above further develops as

$$\begin{aligned}\left\|Q_T(\lambda) - Q_T(\lambda_i)\right\| &\leq \left\|R_T\right\|\left\|D_T(\lambda)B_T D_T(\lambda)^{\mathsf{H}} - D_T(\lambda_i)B_T D_T(\lambda)^{\mathsf{H}}\right. \\ &\quad \left. + D_T(\lambda_i)B_T D_T(\lambda)^{\mathsf{H}} - D_T(\lambda_i)B_T D_T(\lambda_i)^{\mathsf{H}}\right\| \\ &\leq 2\left\|D_T(\lambda)\right\|\left\|R_T\right\|\left\|B_T\right\|\left\|D_T(\lambda) - D_T(\lambda_i)\right\| \\ &\leq CT(\sqrt{\log T} + 1)\left|\lambda - \lambda_i\right|\end{aligned}$$

where we used (5.11), (5.15), and $\|D_T(\lambda)\| = 1/\sqrt{T}$. Up to a change in C, we can finally write $\|Q_T(\lambda) - Q_T(\lambda_i)\| \leq CT^{1-\beta}\sqrt{\log T}$. Assume that $f(x,T) \triangleq xT^{\beta-2}/\left(C\sqrt{\log T}\right)$ satisfies $f(x,T) > 1$ (always possible for every fixed x by taking T large). Then we get by Lemma 9

$$\begin{aligned}\mathbb{P}\left[\chi_1 > x\right] &\leq \mathbb{P}\left(CN^{-1}T^{1-\beta}\sqrt{\log T}\sum_{t=0}^{T-1}\sum_{n=0}^{N-1}\left|w_{n,t}\right|^2 > x\right) \\ &= \mathbb{P}\left(\frac{1}{NT}\sum_{t=0}^{T-1}\sum_{n=0}^{N-1}\left(\left|w_{n,t}\right|^2 - 1\right) > f(x,T) - 1\right) \\ &\leq \exp\left(-NT\left(f(x,T) - \log\left(f(x,T)\right) - 1\right)\right).\end{aligned}$$

□

The most technical part of the proof is to control the term χ_2, which we handle hereafter.

Lemma 18. *The following inequality holds:*

$$\mathbb{P}\left[\chi_2 > x\right] \leq \exp\left(-\frac{cx^2T}{4\left\|\mathbf{\Upsilon}\right\|_\infty^2 \log T}(1+o(1))\right).$$

Proof. From the union bound we obtain:

$$\mathbb{P}\left[\chi_2 > x\right] \leq \sum_{i=0}^{\lfloor T^\beta \rfloor - 1} \mathbb{P}\left[\left|\widehat{\Upsilon}_T^u(\lambda_i) - \mathbb{E}\widehat{\Upsilon}_T^u(\lambda_i)\right| > x\right]. \tag{5.16}$$

Each term of the sum can be written

$$\begin{aligned}\mathbb{P}\left[\left|\widehat{\Upsilon}_T^u(\lambda_i) - \mathbb{E}\widehat{\Upsilon}_T^u(\lambda_i)\right| > x\right] &= \mathbb{P}\left[\widehat{\Upsilon}_T^u(\lambda_i) - \mathbb{E}\widehat{\Upsilon}_T^u(\lambda_i) > x\right] \\ &\quad + \mathbb{P}\left[-\left(\widehat{\Upsilon}_T^u(\lambda_i) - \mathbb{E}\widehat{\Upsilon}_T^u(\lambda_i)\right) > x\right].\end{aligned}$$

We will deal with the term $\psi_i = \mathbb{P}\left[\widehat{\Upsilon}_T^u(\lambda_i) - \mathbb{E}\widehat{\Upsilon}_T^u(\lambda_i) > x\right]$, the term $\mathbb{P}\left[-\left(\widehat{\Upsilon}_T^u(\lambda_i) - \mathbb{E}\widehat{\Upsilon}_T^u(\lambda_i)\right) > x\right]$ being treated similarly. Let $Q_T(\lambda_i) = U_T \Sigma_T U_T^{\mathsf{H}}$ be a spectral factorization of the Hermitian matrix $Q_T(\lambda_i)$ with $\Sigma_T = \mathrm{diag}(\sigma_0, \ldots, \sigma_{T-1})$. Since U_T is unitary and W_T has independent $\mathcal{CN}(0,1)$ elements, we get from Equation (5.10)

$$\begin{aligned}\widehat{\Upsilon}_T^u(\lambda_i) &\overset{\mathcal{L}}{=} \frac{1}{N}\sum_{n=0}^{N-1} w_n^{\mathsf{H}} \Sigma_T(\lambda_i) w_n \\ &= \frac{1}{N}\sum_{n=0}^{N-1}\sum_{t=0}^{T-1} |w_{n,t}|^2 \sigma_t \end{aligned} \tag{5.17}$$

where $\overset{\mathcal{L}}{=}$ denotes equality in law. Since $\mathbb{E}\left[e^{a|X|^2}\right] = 1/(1-a)$ when $X \sim \mathcal{CN}(0,1)$ and $0 < a < 1$, we have by Markov's inequality and from the

independence of the variables $|w_{n,t}|^2$

$$\begin{aligned}
\psi_i &= \mathbb{P}\left(\frac{1}{N}\sum_{n=0}^{N-1}\sum_{t=0}^{T-1}|w_{n,t}|^2\sigma_t - \operatorname{Tr} Q_T(\lambda_i) > x\right) \\
&\leq \mathbb{E}\left[\exp\Big(\frac{\tau}{N}\sum_{t=0}^{T-1}\sum_{n=0}^{N-1}|w_{n,t}|^2\sigma_t\Big)\right]\exp\Big(-\tau\Big(x+\sum_{t=0}^{T-1}\sigma_t\Big)\Big) \\
&= \exp\Big(-\tau\Big(x+\sum_{t=0}^{T-1}\sigma_t\Big)\Big)\prod_{t=0}^{T-1}\Big(1-\frac{\sigma_t\tau}{N}\Big)^{-N} \qquad (5.18)\\
&= \exp\Big(-\tau\Big(x+\sum_{t=0}^{T-1}\sigma_t\Big) - N\sum_{t=0}^{T-1}\log\Big(1-\frac{\sigma_t\tau}{N}\Big)\Big)
\end{aligned}$$

for any τ such that $0 \leq \tau < \min_{0\leq t\leq T-1}\frac{N}{\sigma_t}$. Writing $\log(1-x) = -x-\frac{x^2}{2}+R_3(x)$ with $|R_3(x)| \leq \frac{|x|^3}{3(1-\epsilon)^3}$ when $|x| < \epsilon < 1$, we get

$$\begin{aligned}
\psi_i &\leq \exp\Big(-\tau x + N\sum_{t=0}^{T-1}\Big(\frac{\sigma_t^2\tau^2}{2N^2} + R_3\Big(\frac{\sigma_t\tau}{N}\Big)\Big)\Big) \\
&\leq \exp\Big(-N\Big(\frac{\tau x}{N} - \frac{\tau^2}{2N^2}\sum_{t=0}^{T-1}\sigma_t^2\Big)\Big)\exp\Big(N\sum_{t=0}^{T-1}\Big|R_3\Big(\frac{\sigma_t\tau}{N}\Big)\Big|\Big). \qquad (5.19)
\end{aligned}$$

We shall manage this expression by using Lemma 16. In order to control the term $\exp(N\sum|R_3(\cdot)|)$, we make the choice

$$\tau = \frac{axT}{\log T}$$

where a is a parameter of order one to be optimized later. From (5.13) we get $\max_t \frac{\sigma_t\tau}{N} = \mathcal{O}\left((\log T)^{-1/2}\right)$. Hence, for all T large, $\tau < \min_t \frac{N}{\sigma_t}$. Therefore, (5.18) is valid for this choice of τ and for T large. Moreover, for ϵ fixed and T large, $\frac{\sigma_t\tau}{N} < \epsilon < 1$ so that for these T

$$\begin{aligned}
N\sum_{t=0}^{T-1}\Big|R_3\Big(\frac{\sigma_t\tau}{N}\Big)\Big| &\leq \frac{a^3T^3x^3}{3N^2(1-\epsilon)^3(\log T)^3}\sum_{t=0}^{T-1}|\sigma_t|^3 \\
&= \mathcal{O}\left(T(\log T)^{-3/2}\right)
\end{aligned}$$

from (5.14). Plugging the expression of τ in (5.19), we get

$$\psi_i \leq \exp\Big(-N\Big(\frac{aTx^2}{(\log T)N} - \frac{a^2T^2x^2}{2N^2(\log T)^2}\sum_{t=0}^{T-1}\sigma_t^2\Big)\Big)\exp\Big(C\Big(T(\log T)^{-3/2}\Big)\Big).$$

Using (5.12), we have

$$\psi_i \leq \exp\Big(-\frac{x^2 T}{\log T}\Big(a - \frac{\|\boldsymbol{\Upsilon}\|_\infty^2 a^2 T}{N}\Big)\Big) \exp\Big(\frac{CT}{(\log T)^{3/2}}\Big).$$

The right hand side term is minimized for $a = \frac{N}{2T\|\boldsymbol{\Upsilon}\|_\infty^2}$ which finally gives

$$\psi_i \leq \exp\Big(-\frac{Nx^2}{4\,\|\boldsymbol{\Upsilon}\|_\infty^2 \log T}(1 + o(1))\Big).$$

Combining the above inequality with (5.16) (which induces additional $o(1)$ terms in the argument of the exponential) concludes the lemma. □

Lemma 19. $\chi_3 \leq CT^{-\beta+2}\sqrt{\log T}$.

Proof. From Lemma 14, $\|R_T \odot B_T\| \leq \|R_T\|\,\|B_T\|$ (see [33, Theorem 5.5.1]), and (5.8), we get:

$$\begin{aligned}\Big|\mathbb{E}\widehat{\Upsilon}_T^u(\lambda_i) - \mathbb{E}\widehat{\Upsilon}_T^u(\lambda)\Big| &\leq 2\,\|d_T(\lambda) - d_T(\lambda_i)\|\,\|R_T\|\,\|B_T\| \\ &\leq CT^2\,|\lambda - \lambda_i|\,\|\boldsymbol{\Upsilon}\|_\infty \sqrt{\log T}.\end{aligned}$$

□

Lemmas 17–19 show that $\mathbb{P}[\chi_2 > x]$ dominates the term $\mathbb{P}[\chi_1 > x]$ and that the term χ_3 is vanishing. Mimicking the end of the proof of Theorem 28, we obtain Theorem 29.

We conclude this section by an empirical evaluation by Monte Carlo simulations of $\mathbb{P}[\|\widehat{R}_T - R_T\| > x]$ (curves labeled Biased and Unbiased), with $\widehat{R}_T \in \{\widehat{R}_T^b, \widehat{R}_T^u\}$, $T = 2N$, $x = 2$. This is shown in Figure 5.1 against the theoretical exponential bounds of Theorems 28 and 29 (curves labeled Biased theory and Unbiased theory). We observe that the rates obtained in Theorems 28 and 29 are asymptotically close to optimal.

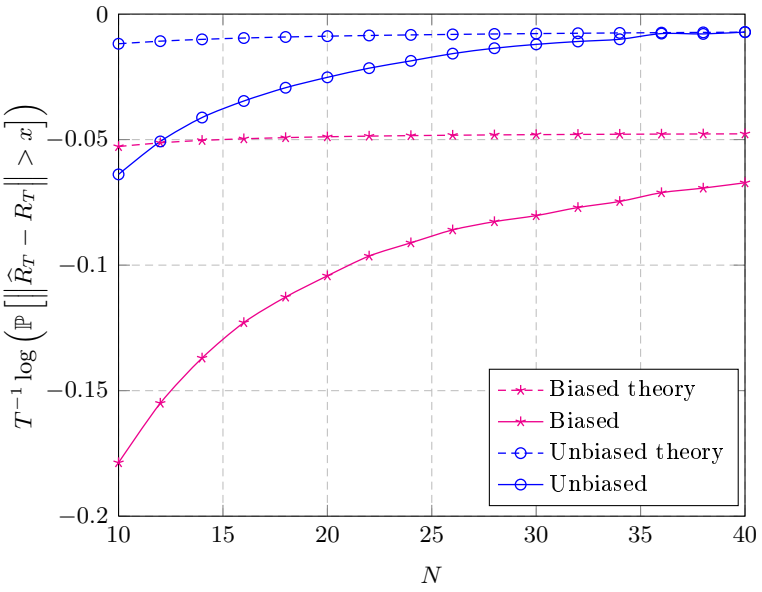

Figure 5.1: Error probability of the spectral norm for $x = 2$, $c = 0.5$, $[R_T]_{k,l} = a^{|k-l|}$ with $a = 0.6$.

5.3 Estimators for "signal-plus-noise" model

5.3.1 Model, assumptions, and results

Consider now the following signal-plus-noise model:

$$Y_T = [y_{n,t}]_{\substack{0\leq n\leq N-1\\ 0\leq t\leq T-1}} = P_T + V_T \tag{5.20}$$

where the $N \times T$ matrix V_T is defined in (5.1) and where P_T satisfies the following assumption:

Assumption 9. *$P_T \triangleq \boldsymbol{h}_T \boldsymbol{s}_T^{\mathsf{H}} \Gamma_T^{1/2}$ where $\boldsymbol{h}_T \in \mathbb{C}^N$ is a deterministic vector such that $\sup_T \|\boldsymbol{h}_T\| < \infty$, the vector $\boldsymbol{s}_T = (s_0, \ldots, s_{T-1})^{\mathsf{T}} \in \mathbb{C}^T$ is a random vector independent of W_T with the distribution $\mathcal{CN}(0, I_T)$, and $\Gamma_T = [\gamma_{ij}]_{i,j=0}^{T-1}$ is Hermitian nonnegative such that $\sup_T \|\Gamma_T\| < \infty$.*

We have here a model for a rank-one signal corrupted with a Gaussian spatially white and temporally correlated noise with stationary temporal correlations. Observe that the signal can also be temporally correlated. Our purpose is still to estimate the noise correlation matrix R_T. To that end, we

use one of the estimators (5.3) or (5.4) with the difference that the samples $v_{n,t}$ are simply replaced with the samples $y_{n,t}$. It turns out that these estimators are still consistent in spectral norm. Intuitively, P_T does not break the consistency of these estimators as it can be seen as a rank-one perturbation of the noise term V_T in which the subspace spanned by $(\Gamma^{1/2})^{\mathsf{H}}\boldsymbol{s}_T$ is "delocalized" enough so as not to perturb much the estimators of R_T. In fact, we even have the following strong result.

Theorem 30. *Let Y_T be defined as in (5.20). Assume r_k are absolutely summable and $r_0 \neq 0$. Let $T \to \infty$ and $N/T \to c > 0$. Define the estimates*

$$\hat{r}^{bp}_{k,T} = \frac{1}{NT}\sum_{n=0}^{N-1}\sum_{t=0}^{T-1} y_{n,t+k}y^*_{n,t}\mathbb{1}_{0\leq t+k\leq T-1}$$
$$\hat{r}^{up}_{k,T} = \frac{1}{N(T-|k|)}\sum_{n=0}^{N-1}\sum_{t=0}^{T-1} y_{n,t+k}y^*_{n,t}\mathbb{1}_{0\leq t+k\leq T-1}$$

and let

$$\widehat{R}^{bp}_T = \mathcal{T}(\hat{r}^{bp}_{-(T-1),T},\ldots,\hat{r}^{bp}_{(T-1),T})$$
$$\widehat{R}^{up}_T = \mathcal{T}(\hat{r}^{up}_{-(T-1),T},\ldots,\hat{r}^{up}_{(T-1),T}).$$

Then for any $x > 0$,

$$\mathbb{P}\left[\left\|\widehat{R}^{bp}_T - R_T\right\| > x\right] \leq \exp\Big(-cT\Big(\frac{x}{\|\boldsymbol{\Upsilon}\|_\infty} - \log\Big(1+\frac{x}{\|\boldsymbol{\Upsilon}\|_\infty}\Big) + o(1)\Big)\Big)$$

and

$$\mathbb{P}\left[\left\|\widehat{R}^{up}_T - R_T\right\| > x\right] \leq \exp\Big(-\frac{cTx^2}{4\|\boldsymbol{\Upsilon}\|^2_\infty \log T}(1+o(1))\Big).$$

Before proving this theorem, some remarks are in order.

Remark 2. *Theorem 30 generalizes without difficulty to the case where P_T has a fixed rank $K > 1$. This captures the situation of $K \ll \min(N,T)$ sources.*

Remark 3. *Similar to the proofs of Theorems 28 and 29, the proof of Theorem 30 uses concentration inequalities for functionals of Gaussian random variables based on the moment generating function and the Chernoff bound. Exploiting instead McDiarmid's concentration inequality [41], it is possible to adapt Theorem 30 to $\boldsymbol{s}_T$ with bounded (instead of Gaussian) entries. This adaptation may account for discrete sources met in digital communication signals.*

5.3.2 Main elements of the proof of Theorem 30

We restrict the proof to the more technical part that concerns $\widehat{R}_T^{up}$. Defining the function

$$\widehat{\Upsilon}_T^{up}(\lambda) \triangleq \sum_{k=-(T-1)}^{T-1} \hat{r}_{k,T}^{up} e^{ik\lambda}$$

and recalling that $\Upsilon_T(\lambda) = \sum_{k=-(T-1)}^{T-1} r_k e^{ik\lambda}$, we need to establish a concentration inequality on $\mathbb{P}\left[\sup_{\lambda\in[0,2\pi)} |\widehat{\Upsilon}_T^{up}(\lambda) - \Upsilon_T(\lambda)| > x\right]$. For any $\lambda \in [0, 2\pi)$, the term $\widehat{\Upsilon}_T^{up}(\lambda)$ can be written as (see Lemma 14)

$$\begin{aligned}
\widehat{\Upsilon}_T^{up}(\lambda) &= d_T(\lambda)^{\mathsf{H}} \left(\frac{Y_T^{\mathsf{H}} Y_T}{N} \odot B_T\right) d_T(\lambda) \\
&= d_T(\lambda)^{\mathsf{H}} \left(\frac{V_T^{\mathsf{H}} V_T}{N} \odot B_T\right) d_T(\lambda) \\
&+ d_T(\lambda)^{\mathsf{H}} \left(\frac{P_T^{\mathsf{H}} V_T + V_T^{\mathsf{H}} P_T}{N} \odot B_T\right) d_T(\lambda) \\
&+ d_T(\lambda)^{\mathsf{H}} \left(\frac{P_T^{\mathsf{H}} P_T}{N} \odot B_T\right) d_T(\lambda) \\
&\triangleq \widehat{\Upsilon}_T^{u}(\lambda) + \widehat{\Upsilon}_T^{cross}(\lambda) + \widehat{\Upsilon}_T^{sig}(\lambda)
\end{aligned}$$

where B_T is the matrix defined in the statement of Lemma 14. We know from the proof of Theorem 29 that

$$\mathbb{P}\left[\sup_{\lambda\in[0,2\pi)} |\widehat{\Upsilon}_T^{u}(\lambda) - \Upsilon_T(\lambda)| > x\right] \leq \exp\left(-\frac{cTx^2}{4\,\|\boldsymbol{\Upsilon}\|_\infty^2 \log T}(1 + o(1))\right). \tag{5.21}$$

We then need only handle the terms $\widehat{\Upsilon}_T^{cross}(\lambda)$ and $\widehat{\Upsilon}_T^{sig}(\lambda)$.

We start with a simple lemma.

Lemma 20. *Let X and Y be two independent $\mathcal{N}(0,1)$ random variables. Then for any $\tau \in (-1, 1)$,*

$$\mathbb{E}[\exp(\tau XY)] = (1 - \tau^2)^{-1/2}.$$

Proof.

$$\begin{aligned}\mathbb{E}[\exp(\tau XY)] &= \frac{1}{2\pi}\int_{\mathbb{R}^2} e^{\tau xy} e^{-x^2/2} e^{-y^2/2}\,dx\,dy \\ &= \frac{1}{2\pi}\int_{\mathbb{R}^2} e^{-(x-\tau y)^2/2} e^{-(1-\tau^2)y^2/2}\,dx\,dy \\ &= (1-\tau^2)^{-1/2}.\end{aligned}$$

□

With this result, we now have

Lemma 21. *There exists a constant $a > 0$ such that*

$$\mathbb{P}\left[\sup_{\lambda\in[0,2\pi)} |\widehat{\Upsilon}_T^{cross}(\lambda)| > x\right] \leq \exp\Big(-\frac{axT}{\sqrt{\log T}}(1+o(1))\Big).$$

Proof. We only sketch the proof of this lemma. We show that for any $\lambda \in [0, 2\pi]$,

$$\mathbb{P}[|\widehat{\Upsilon}_T^{cross}(\lambda)| > x] \leq \exp\Big(-\frac{axT}{\sqrt{\log T}} + C\Big)$$

where C does not depend on $\lambda \in [0, 2\pi]$. The lemma is then proven by a discretization argument of the interval $[0, 2\pi]$ analogous to what was done in the proofs of Section 5.2. We shall bound $\mathbb{P}[\widehat{\Upsilon}_T^{cross}(\lambda) > x]$, the term $\mathbb{P}[\widehat{\Upsilon}_T^{cross}(\lambda) < -x]$ being bounded similarly. From Lemma 15, we get

$$\begin{aligned}\widehat{\Upsilon}_T^{cross}(\lambda) &= \mathrm{Tr}\Big(D_T(\lambda)^{\mathsf{H}} \frac{P_T^{\mathsf{H}} V_T + V_T^{\mathsf{H}} P_T}{N} D_T(\lambda) B_T\Big) \\ &= \mathrm{Tr}\,\frac{D_T(\lambda)^{\mathsf{H}} (\Gamma_T^{1/2})^{\mathsf{H}} \boldsymbol{s}_T \boldsymbol{h}_T^{\mathsf{H}} W_T R_T^{1/2} D_T(\lambda) B_T}{N} \\ &\quad + \mathrm{Tr}\,\frac{D_T(\lambda)^{\mathsf{H}} (R_T^{1/2})^{\mathsf{H}} W_T^{\mathsf{H}} \boldsymbol{h}_T \boldsymbol{s}_T^{\mathsf{H}} \Gamma_T^{1/2} D_T(\lambda) B_T}{N} \\ &= \frac{2}{N}\Re(\boldsymbol{h}_T^{\mathsf{H}} W_T G_T(\lambda) \boldsymbol{s}_T)\end{aligned}$$

where $G_T(\lambda) = R_T^{1/2} D_T(\lambda) B_T D_T(\lambda)^{\mathsf{H}} (\Gamma_T^{1/2})^{\mathsf{H}}$. Let $G_T(\lambda) = U_T \Omega_T \widetilde{U}_T^{\mathsf{H}}$ be a singular value decomposition of $G_T(\lambda)$ where $\Omega = \mathrm{diag}(\omega_0, \ldots, \omega_{T-1})$. Observe that the vector $\boldsymbol{x}_T \triangleq W_T^{\mathsf{H}} \boldsymbol{h}_T = (x_0, \ldots, x_{T-1})^{\mathsf{T}}$ has the distribution $\mathcal{CN}(0, \|\boldsymbol{h}_T\|^2 I_T)$. We can then write

$$\widehat{\Upsilon}_T^{cross}(\lambda) \stackrel{\mathcal{L}}{=} \frac{2}{N}\Re\left(\boldsymbol{x}_T^{\mathsf{H}} \Omega_T \boldsymbol{s}_T\right) = \frac{2}{N}\sum_{t=0}^{T-1} \omega_t (\Re x_t \Re s_t + \Im x_t \Im s_t).$$

Notice that $\{\Re x_t, \Im x_t, \Re s_t, \Im s_t\}_{t=0}^{T-1}$ are independent with $\Re x_t$, $\Im x_t \sim \mathcal{N}(0, \|\boldsymbol{h}_T\|^2/2)$ and $\Re s_t, \Im s_t \sim \mathcal{N}(0, 1/2)$. Letting

$$0 < \tau < (\sup_T \|\boldsymbol{h}_T\|)^{-1} (\sup_\lambda \|G_T(\lambda)\|)^{-1}$$

and using Markov's inequality and Lemma 20, we get

$$\begin{aligned}
\mathbb{P}\left[\widehat{\Upsilon}_T^{cross}(\lambda) > x\right] &= \mathbb{P}\left[e^{N\tau\widehat{\Upsilon}_T^{cross}(\lambda)} > e^{N\tau x}\right] \\
&\leq e^{-N\tau x}\mathbb{E}\left[e^{2\tau\sum_t \omega_t(\Re x_t \Re s_t + \Im x_t \Im s_t)}\right] \\
&= e^{-N\tau x}\prod_{t=0}^{T-1}\left(1-\tau^2\omega_t^2\|\boldsymbol{h}_T\|^2\right)^{-1} \\
&= \exp\left(-N\tau x - \sum_{t=0}^{T-1}\log(1-\tau^2\omega_t^2\|\boldsymbol{h}_T\|^2)\right).
\end{aligned}$$

Mimicking the proof of Lemma 16, we can establish that $\sum_t \omega_t^2 = O(\log T)$ and $\max_t \omega_t = O(\sqrt{\log T})$ uniformly in $\lambda \in [0, 2\pi]$. Set $\tau = b/\sqrt{\log T}$ where $b > 0$ is small enough so that $\sup_{T,\lambda}(\tau\|\boldsymbol{h}_T\|\,\|G_T(\lambda)\|) < 1$. Observing that $\log(1-x) = O(x)$ for x small enough, we get

$$\mathbb{P}[\widehat{\Upsilon}_T^{cross}(\lambda) > x] \leq \exp\bigl(-Nbx/\sqrt{\log T} + \mathcal{E}(\lambda, T)\bigr)$$

where $|\mathcal{E}(\lambda, T)| \leq (C/\log T)\sum_t \omega_t^2 \leq C$. This establishes Lemma 21. □

Lemma 22. *There exists a constant $a > 0$ such that*

$$\mathbb{P}\left[\sup_{\lambda\in[0,2\pi)} |\widehat{\Upsilon}_T^{sig}(\lambda)| > x\right] \leq \exp\Bigl(-\frac{axT}{\sqrt{\log T}}(1+o(1))\Bigr).$$

Proof. By Lemma 15,

$$\begin{aligned}
\widehat{\Upsilon}_T^{sig}(\lambda) &= N^{-1}\operatorname{Tr}(D_T^{\mathsf{H}} P_T^{\mathsf{H}} P_T D_T B_T) \\
&= \frac{\|\boldsymbol{h}_T\|^2}{N}\boldsymbol{s}_T^{\mathsf{H}} G_T(\lambda)\boldsymbol{s}_T
\end{aligned}$$

where $G_T(\lambda) = \Gamma_T^{1/2} D_T(\lambda) B_T D_T(\lambda)^{\mathsf{H}} (\Gamma_T^{1/2})^{\mathsf{H}}$. By the spectral factorization $G_T(\lambda) = U_T \Sigma_T U_T^{\mathsf{H}}$ with $\Sigma_T = \operatorname{diag}(\sigma_0, \ldots, \sigma_{T-1})$, we get

$$\widehat{\Upsilon}_T^{sig}(\lambda) \stackrel{\mathcal{L}}{=} \frac{\|\boldsymbol{h}_T\|^2}{N}\sum_{t=0}^{T-1}\sigma_t|s_t|^2$$

and

$$\begin{aligned}\mathbb{P}[\widehat{\Upsilon}_T^{sig}(\lambda) > x] &\leq e^{-N\tau x}\mathbb{E}\Big[e^{\tau\|\boldsymbol{h}_T\|^2\sum_t \sigma_t|s_t|^2}\Big] \\ &= \exp\Big(-N\tau x - \sum_{t=0}^{T-1}\log(1-\sigma_t\tau\|\boldsymbol{h}_T\|^2)\Big)\end{aligned}$$

for any $\tau \in (0, 1/(\|\boldsymbol{h}_T\|^2 \sup_\lambda \|G_T(\lambda)\|))$. Let us show that

$$|\operatorname{Tr} G_T(\lambda)| \leq C\sqrt{\frac{\log T + 1}{T}}.$$

Indeed, we have

$$\begin{aligned}|\operatorname{Tr} G_T(\lambda)| = N^{-1}|\operatorname{Tr} D_T B_T D_T^{\mathsf{H}}\Gamma_T| &= \frac{1}{N}\left|\sum_{k,\ell=0}^{T-1}\frac{e^{-\imath(k-\ell)\lambda}\gamma_{\ell,k}}{T-|k-\ell|}\right| \\ &\leq \Big(\frac{1}{N}\sum_{k,\ell=0}^{T-1}|\gamma_{k,\ell}|^2\Big)^{1/2}\Big(\frac{1}{N}\sum_{k,\ell=0}^{T-1}\frac{1}{(T-|k-\ell|)^2}\Big)^{1/2} \\ &= \Big(\frac{\operatorname{Tr}\Gamma_T\Gamma_T^{\mathsf{H}}}{N}\Big)^{1/2}\Big(\frac{2}{N}(\log T + C)\Big)^{1/2} \leq C\sqrt{\frac{\log T+1}{T}}.\end{aligned}$$

Moreover, similar to the proof of Lemma 16, we can show that $\sum_t \sigma_t^2 = O(\log T)$ and $\max_t |\sigma_t| = O(\sqrt{\log T})$ uniformly in λ. Taking $\tau = b/\sqrt{\log T}$ for $b > 0$ small enough, and recalling that $\log(1-x) = 1 - x + O(x^2)$ for x small enough, we get that

$$\mathbb{P}[\widehat{\Upsilon}_T^{sig}(\lambda) > x] \leq \exp\Big(-\frac{Nbx}{\sqrt{\log T}} + \frac{b\|\boldsymbol{h}_T\|^2}{\sqrt{\log T}}\operatorname{Tr} G_T(\lambda) + \mathcal{E}(T,\lambda)\Big)$$

where $|\mathcal{E}(T,\lambda)| \leq (C/\log T)\sum_t \sigma_t^2 \leq C$. We therefore get

$$\mathbb{P}[\widehat{\Upsilon}_T^{sig}(\lambda) > x] \leq \exp\Big(-\frac{Nbx}{\sqrt{\log T}} + C\Big)$$

where C is independent of λ. Lemma 22 is then obtained by the discretization argument of the interval $[0, 2\pi]$. □

Gathering Inequality (5.21) with Lemmas 21 and 22, we get the second inequality of the statement of Theorem 30.

5.4 Application to source detection

From an application point of view, the model (5.20) corresponds to a general transmission model where the matrix P_T represents the signal matrix and $V_T = W_T R_T^{1/2}$ corresponds to a temporally correlated noise. We consider an example of a linear sensor network composed from one possibly emitting source and N sensors. Recall the null hypothesis H_0 for which the signal is absent and the hypothesis H_1 under which the source signal is present. By stacking the T observations into the matrix $Y_T = [y_0, \ldots, y_{T-1}] \in \mathbb{C}^{N\times T}$ we have

Consider a sensor network composed of N sensors impinged by zero (hypothesis H_0) or one (hypothesis H_1) source signal. The stacked signal matrix $Y_T = [y_0, \ldots, y_{T-1}] \in \mathbb{C}^{N\times T}$ from time $t = 0$ to $t = T-1$ is modeled as

$$Y_T = \begin{cases} V_T, & H_0 \\ \boldsymbol{h}_T \boldsymbol{s}_T^{\mathsf{H}} + V_T, & H_1 \end{cases} \tag{5.22}$$

where $\boldsymbol{s}_T^{\mathsf{H}} = [s_0^*, \ldots, s_{T-1}^*]$ are (hypothetical) independent $\mathcal{CN}(0,1)$ signals transmitted through the constant channel $\boldsymbol{h}_T \in \mathbb{C}^{N\times 1}$, and $V_T = W_T R_T^{1/2} \in \mathbb{C}^{N\times T}$ models a stationary noise matrix as in (5.1).

As opposed to standard procedures where preliminary pure noise data are available , we shall proceed here to an online signal detection test solely based on Y_T, by exploiting the consistency established in Theorem 30. The approach consists precisely in estimating R_T by $\widehat{R}_T \in \{\widehat{R}_T^{bp}, \widehat{R}_T^{up}\}$, which is then used as a whitening matrix for Y_T. The binary hypothesis (5.22) can then be equivalently written

$$Y_T \widehat{R}_T^{-1/2} = \begin{cases} W_T R_T^{1/2} \widehat{R}_T^{-1/2}, & H_0 \\ \boldsymbol{h}_T \boldsymbol{s}_T^{\mathsf{H}} \widehat{R}_T^{-1/2} + W_T R_T^{1/2} \widehat{R}_T^{-1/2}, & H_1. \end{cases} \tag{5.23}$$

Since $\|R_T \widehat{R}_T^{-1} - I_T\| \to 0$ almost surely (by Theorem 30 as long as $\inf_{\lambda\in[0,2\pi)} \boldsymbol{\Upsilon}(\lambda) > 0$), for T large, the decision on the hypotheses (5.23) can be handled by the GLRT [14] by approximating $W_T R_T^{1/2} \widehat{R}_T^{-1/2}$ as a purely white noise. We then have the following result.

Theorem 31. *Let $\widehat{R}_T$ be any of $\widehat{R}_T^{bp}$ or $\widehat{R}_T^{up}$ strictly defined in Theorem 30 for Y_T now following model (5.22). Assume $\inf_{\lambda\in[0,2\pi)} \boldsymbol{\Upsilon}(\lambda) > 0$ and $\liminf_T \|\boldsymbol{h}_T\|^2 \operatorname{Tr}\left(R_T^{-1}\right)/T \geq \sqrt{c}$ and define the test*

$$\alpha = \frac{N \left\| Y_T \widehat{R}_T^{-1} Y_T^{\mathsf{H}} \right\|}{\operatorname{Tr}\left(Y_T \widehat{R}_T^{-1} Y_T^{\mathsf{H}} \right)} \underset{H_1}{\overset{H_0}{\lessgtr}} \gamma \tag{5.24}$$

where $\gamma \in \mathbb{R}^+$ *satisfies* $\gamma > (1+\sqrt{c})^2$. *Then, as* $T \to \infty$,

$$\mathbb{P}\left[\alpha \geq \gamma\right] \to \begin{cases} 0, & H_0 \\ 1, & H_1. \end{cases}$$

Recall from [14] that the decision threshold $(1+\sqrt{c})^2$ corresponds to the almost sure limiting largest eigenvalue of $\frac{1}{T} W_T W_T^{\mathsf{H}}$, that is the right-edge of the support of the Marčenko–Pastur law.

5.5 Numerical results

Simulations are performed hereafter to assess the performance of the test (5.24) under several system settings. We take here h_T to be the following steering vector $h_T = \sqrt{p/T}[1, \ldots, e^{2i\pi\theta(T-1)}]$ with $\theta = 10°$ and p a power parameter. The matrix R_T models an autoregressive process of order 1 with parameter a, so the noise covariance matrix is expressed as $[R_T]_{k,l} = a^{|k-l|}$.

In Figure 5.2, the detection error $1 - \mathbb{P}[\alpha \geq \gamma | H_1]$ of the test (5.24) for a false alarm rate (FAR) $\mathbb{P}[\alpha \geq \gamma | H_0] = 0.05$ under $\widehat{R}_T = \widehat{R}_T^{up}$ (Unbiased) or $\widehat{R}_T = \widehat{R}_T^{bp}$ (Biased) is compared against the estimator that assumes R_T perfectly known (Oracle), *i.e.* that sets $\widehat{R}_T = R_T$ in (5.24), and against the GLRT test that wrongly assumes temporally white noise (White), *i.e.* that sets $\widehat{R}_T = I_T$ in (5.24). The source signal power is set to $p = 1$, that is a signal-to-noise ratio (SNR) of 0 dB, N is varied from 10 to 50 and $T = N/c$ for $c = 0.5$ fixed. In the same setting as Figure 5.2, the number of sensors is now fixed to $N = 20$, $T = N/c = 40$ and the SNR (hence p) is varied from -10 dB to 4 dB. The powers of the various tests are displayed in Figure 5.3 and compared to the detection methods which estimate R_T from a pure noise sequence called Biased PN (pure noise) and Unbiased PN. The results of the proposed online method are close to that of Biased/Unbiased PN, this last presenting the disadvantage to have at its disposal a pure noise sequence at the receiver. Both figures suggest a close match in performance between Oracle and Biased, while Unbiased shows weaker performance. The gap evidenced between Biased and Unbiased confirms the theoretical conclusions.

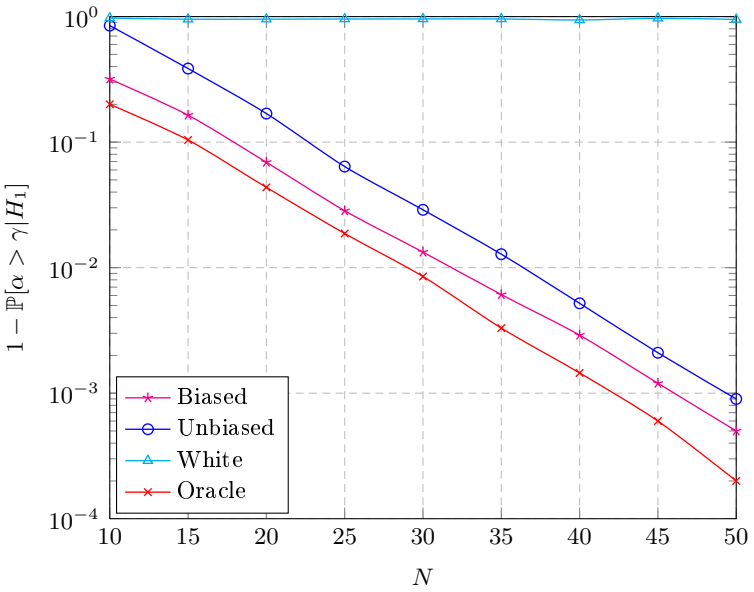

Figure 5.2: Detection error versus N with FAR= 0.05, $p = 1$, SNR= 0 dB, $c = 0.5$, and $a = 0.6$.

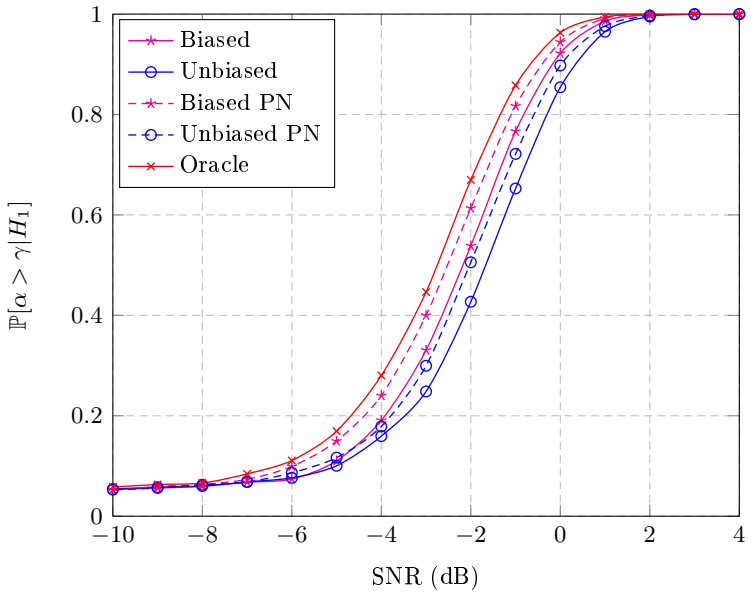

Figure 5.3: Power of detection tests versus SNR (dB) with FAR= 0.05, $N = 20$, $c = 0.5$, and $a = 0.6$.

In the same setting as Figure 5.3, in Figure 5.4 we propose to compare the powers of the proposed detector (with the biased estimate) with different banding parameters γ_0 going from 4 to 20. The estimated covariance matrix used for the different detectors is given by $\widehat{R}^{bp}_{\gamma_0,T} = [[\widehat{R}^{bp}_T]_{i,j}\mathbb{1}_{|i-j|\leq\gamma_0}]$. From the zoom of Figure 5.4 the best detection performance is obtained for $\gamma_0 = 8$ in this setting. Observe that the detection performance gain obtained by using a banded estimator in this setup is quite poor. Moreover, as its theoretic characterization is not known for finite N and T, the banding is arbitrary. We recall that the coefficients of the tapered estimate of R_T are defined as in [18]

$$[\widehat{R}^{bp}_{\tau_0,T}]_{i,j} = \begin{cases} [\widehat{R}^{bp}_T]_{i,j} \text{ when } |i-j| \leq \tau_0, \\ (2 - |i-j|/\tau_0)[\widehat{R}^{bp}_T]_{i,j} \text{ when } \tau_0 < |i-j| \leq 2\tau_0, \\ 0 \text{ otherwise.} \end{cases}$$

The same conclusions as previously are drawn for Figure 5.5 where the proposed detector is compared to the detectors with different tapering parameters τ_0 going from 4 to 20.

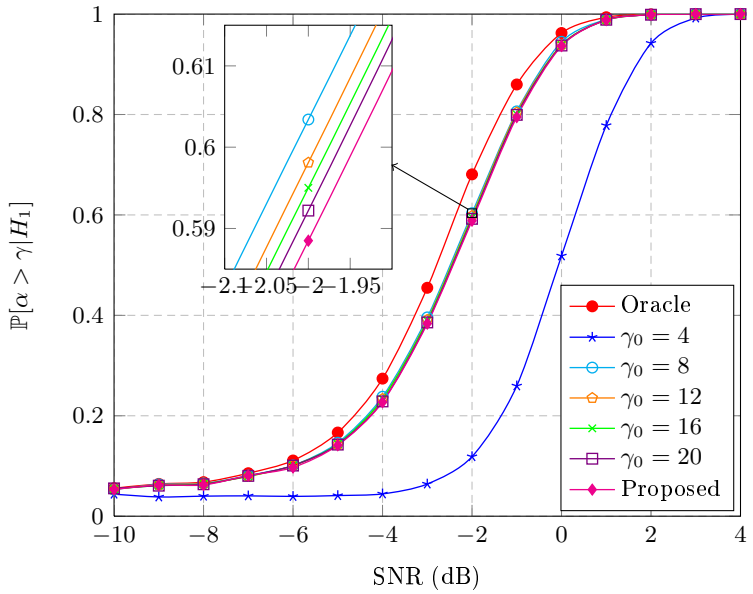

Figure 5.4: Power of detection tests versus SNR (dB) for different banding parameters γ_0 with FAR$= 0.05$, $N = 20$, $c = 0.5$, and $a = 0.6$.

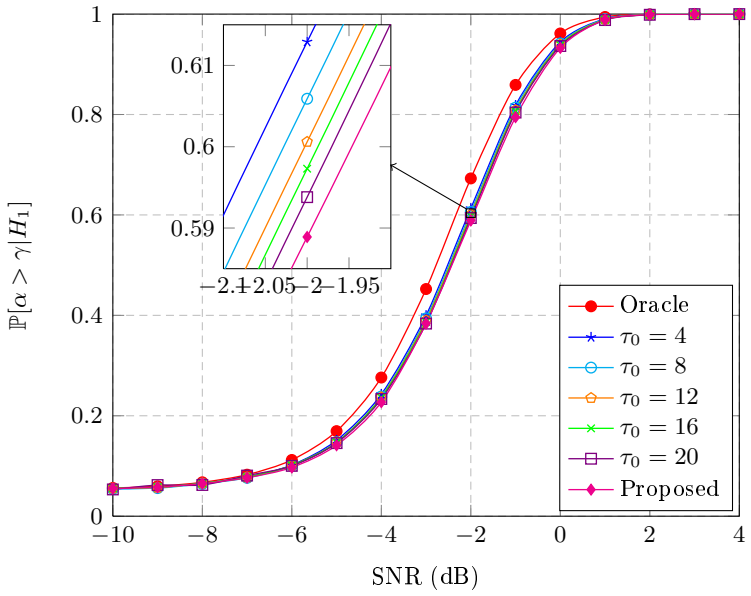

Figure 5.5: Power of detection tests versus SNR (dB) for different tapering parameters τ_0 with FAR$= 0.05$, $N = 20$, $c = 0.5$, and $a = 0.6$.

5.6 Conclusions

In this chapter concentration inequalities were derived for two types of covariance matrix estimates. One of the main advantages of the proposed methods is that we do not need to apply banding procedure in comparison to all existing methods. However, the estimated coefficients which are too far from the main diagonal do not bring much information, they rather degrade the estimate of the covariance matrix. Hence, intuitively, it would be beneficial to perform banding/tapering. The choice of the banding window is an open problem, the size of such window depends on the decrease of the coefficients r_k and we do not assume any knowledge on their decrease.

The application to source detection turns out to be an alternative approach to the method proposed in Chapter 4. Indeed, assuming that more prior information is available on R_T, *i.e.* its Toeplitz structure, we extend the applicability of the theoretical model, allowing in particular to the noise and the signal to have the same type of correlations (both temporal or both spatial). A second advantage is that by whitening, the performance of the proposed detector will outperform the performances of the detector of Chap-

ter 4 at low SNR regime. In the high SNR regime, the covariance estimation will instead be too degraded for this method to be beneficial. We have therefore a trade-off between the both approaches.

5.7 Appendix

5.7.1 Proofs for Theorem 28: proof of Lemma 10

Developing the quadratic forms given in the statement of the lemma, we get

$$\begin{aligned} d_T(\lambda)^{\mathsf{H}} \frac{V_T^{\mathsf{H}} V_T}{N} d_T(\lambda) &= \frac{1}{NT} \sum_{l,l'=0}^{T-1} e^{-\imath(l'-l)\lambda} [V_T^{\mathsf{H}} V_T]_{l,l'} \\ &= \frac{1}{NT} \sum_{l,l'=0}^{T-1} e^{-\imath(l'-l)\lambda} \sum_{n=0}^{N-1} v_{n,l}^* v_{n,l'} \\ &= \sum_{k=-(T-1)}^{T-1} e^{-\imath k\lambda} \frac{1}{NT} \sum_{n=0}^{N-1} \sum_{t=0}^{T-1} v_{n,t}^* v_{n,t+k} \mathbb{1}_{0 \le t+k \le T-1} \\ &= \sum_{k=-(T-1)}^{T-1} \hat{r}_k^b e^{-\imath k\lambda} = \widehat{\Upsilon}_T^b(\lambda), \end{aligned}$$

and

$$\begin{aligned} &\mathbb{E}\left[d_T(\lambda)^{\mathsf{H}} \frac{V_T^{\mathsf{H}} V_T}{N} d_T(\lambda) \right] \\ &= d_T(\lambda)^{\mathsf{H}} (R_T^{1/2})^{\mathsf{H}} \frac{\mathbb{E}[W_T^{\mathsf{H}} W_T]}{N} R_T^{1/2} d_T(\lambda) \\ &= d_T(\lambda)^{\mathsf{H}} R_T d_T(\lambda). \end{aligned}$$

5.7.2 Proofs for Theorem 29: proof of Lemma 14

We have

$$\begin{aligned}
&d_T(\lambda)^{\mathsf{H}} \left(\frac{V_T^{\mathsf{H}} V_T}{N} \odot B_T \right) d_T(\lambda) \\
&= \frac{1}{NT} \sum_{l,l'=-(T-1)}^{T-1} e^{i(l-l')\lambda} [V_T^{\mathsf{H}} V_T]_{l,l'} \frac{T}{T-|l-l'|} \\
&= \sum_{k=-(T-1)}^{T-1} e^{ik\lambda} \frac{1}{N(T-|k|)} \sum_{n=0}^{N-1} \sum_{t=0}^{T-1} v_{n,t}^* v_{n,t+k} \mathbb{1}_{0 \le t+k \le T-1} \\
&= \sum_{k=-(T-1)}^{T-1} \hat{r}_k^u e^{ik\lambda} = \widehat{\Upsilon}_T^u(\lambda).
\end{aligned}$$

5.7.3 Proofs for Theorem 29: proof of Lemma 16

We start by observing that

$$\begin{aligned}
\operatorname{Tr} B_T^2 = \sum_{i,j=0}^{T-1} [B_T]_{i,j}^2 &= \sum_{i,j=0}^{T-1} \left(\frac{T}{T-|i-j|} \right)^2 \\
&= 2 \sum_{i>j}^{T-1} \left(\frac{T}{T-|i-j|} \right)^2 + T \\
&= 2 \sum_{k=1}^{T-1} \left(\frac{T}{T-k} \right)^2 (T-k) + T \\
&= 2T^2 \sum_{k=1}^{T-1} \frac{1}{T-k} + T = 2T^2 \left(\log T + C \right).
\end{aligned}$$

Inequality (5.11) is then obtained upon noticing that $\|B_T\| \le \sqrt{\operatorname{Tr} B_T^2}$.

We now show (5.12). Using twice the inequality $\operatorname{Tr}(FG) \le \|F\| \operatorname{Tr}(G)$

when $F, G \in \mathbb{C}^{m\times m}$ and G is nonnegative definite [33], we get

$$\begin{aligned}\sum_{t=0}^{T-1} \sigma_t^2(\lambda_i) &= \operatorname{Tr} Q_T(\lambda_i)^2 \\ &= \operatorname{Tr} R_T D_T(\lambda_i) B_T D_T(\lambda_i)^{\mathsf{H}} R_T D_T(\lambda_i) B_T D_T(\lambda_i)^{\mathsf{H}} \\ &\le \|R_T\| \operatorname{Tr} R_T (D_T(\lambda_i) B_T D_T(\lambda_i)^{\mathsf{H}})^2 \\ &\le T^{-2} \|R_T\|^2 \operatorname{Tr}(B_T^2) \\ &\le 2 \|\boldsymbol{\Upsilon}\|_\infty^2 \log T + C.\end{aligned}$$

Inequality (5.13) is immediate since $\|Q_T\|^2 \le \operatorname{Tr} Q_T^2$.
As regards (5.14), by the Cauchy–Schwarz inequality,

$$\begin{aligned}\sum_{t=0}^{T-1} |\sigma_t^3(\lambda_i)| = \sum_{t=0}^{T-1} \sigma_t^2(\lambda_i)|\sigma_t(\lambda_i)| &\le \sqrt{\sum_{t=0}^{T-1} \sigma_t^4(\lambda_i) \sum_{t=0}^{T-1} \sigma_t^2(\lambda_i)} \\ &\le \sqrt{\left(\sum_{t=0}^{T-1} \sigma_t^2(\lambda_i)\right)^2 \sum_{t=0}^{T-1} \sigma_t^2(\lambda_i)} \\ &= \left(\sum_{t=0}^{T-1} \sigma_t^2(\lambda_i)\right)^{3/2} \\ &= C((\log T)^{3/2} + 1).\end{aligned}$$

Conclusion and perspectives

In this thesis, we dealt with two difficulties often occurring in modern sensor networks: (i) the number of sensors and the number of observations are both large and of the same order of magnitude and (ii) the sensors are embedded in a correlated noise with unknown covariance matrix. Regarding the assumption on the noise covariance matrix two approaches were elaborated. The first one estimates the parameters of the system without making any assumption on the statistics of the noise samples. The second one assumes that the noise is a stationary process and proposes consistent estimates for the noise covariance matrix leading then to the "whitening" procedure. We have seen that the first approach cannot be applied to the scenarios where the noise and the signal are not simultaneously temporally or spatially correlated. Making an assumption on the noise covariance matrix structure allows to avoid this restriction.

In parallel with the structure of this thesis two main directions of research can be carried out.

Non structured noise covariance matrices

As regards the behavior of outliers of large random matrices and associated detection and estimation algorithms more sophisticated matrix models than the ones considered in this thesis can be explored. One of such models is the so-called doubly correlated model with the noise matrix presenting space and time correlations. The noise matrix for this model is written as $V_T = \widetilde{R}_T^{1/2} W_T R_T^{1/2}$ where W_T has i.i.d. entries and $\widetilde{R}_T$ and R_T are covariance matrices. This model is of particular interest in radar systems where the noise may present simultaneous temporal and spatial correlations.

A recent subject of particular practical interest is the sparse component analysis as it finds applications in many fields of multivariate analysis and signal processing. An interesting perspective consists in exploring the link

between the spiked models and compressive sensing methods in the case where the eigenvectors associated with the spikes have a sparse structure. In this framework, detection in sparse principal component has been for instance explored by Berthet and Rigolet [12] who shedded some light on the link between the spiked models and sparse small rank perturbations of the sample covariance matrix.

Structured noise covariance matrices

In our work we estimated the covariance matrix of a stationary process without banding or tapering. Finding the right banding parameter can lead to a more efficient estimator. However, *a priori* knowledge on the decrease of the Toeplitz coefficients is required in order to construct an optimal banding. One of the directions to this problem is an iterative method.

The technique developed in the second part of the thesis based on the estimation of the noise covariance matrix when the signal is considered as a nuisance can be further developed. A more thorough analysis can be performed, as for instance, a study of the fluctuations of the estimates.

In the case where the probability law of the samples of the stationary process is heavy tailed, robust estimation methods can be used. In the framework of the work of Couillet *et al.* [23], a robust estimation of Toeplitz covariance matrices is under current study. The estimation is performed in two steps. First, we obtain an estimate based on the robust M-estimation giving rise to an instructed first estimate. Then, this estimate is improved by performing a Toeplitzification of the latter. The consistency can be then extended to the case where the aggregated matrix of time samples is corrupted by a rank one (or more generally, low rank) matrix.

Finally, from the application side, the isolated eigenvalue theory in large random matrices can be applied to other problems than in array processing. We can cite, among others, principal component analysis [35], failure detection/diagnosis in large networks [22], or chemometrics [50].

Bibliography

[1] H. Akaike, "A new look at the statistical model identification," *IEEE Transactions on Automatic Control*, vol. 19, no. 6, pp. 716–723, 1974.

[2] T. W. Anderson, *An Introduction to Multivariate Statistical Analysis*. John Wiley & Sons Inc., 2003.

[3] Z. D. Bai, "Methodologies in spectral analysis of large-dimensional random matrices, a review," *Statististica Sinica*, vol. 9, pp. 611–677, 1999.

[4] Z. D. Bai and J. W. Silverstein, "No eigenvalues outside the support of the limiting spectral distribution of large-dimensional sample covariance matrices," *The Annals of Probability*, vol. 26, no. 1, pp. 316–345, 1998.

[5] ——, *Spectral analysis of large dimensional random matrices*, 2nd ed., ser. Springer Series in Statistics. New York: Springer, 2009.

[6] ——, "CLT for linear spectral statistics of large-dimensional sample covariance matrices," *The Annals of Probability*, vol. 32, no. 1A, pp. 553–605, 2004.

[7] Z. D. Bai and J. F. Yao, "Limit theorems for sample eigenvalues in a generalized spiked population model," 2008. [Online] arXiv:0806.1141.

[8] ——, "Central limit theorems for eigenvalues in a spiked population model," *Annales de l'Institut Henri Poincaré–Probabilités et Statistiques*, vol. 44, no. 3, pp. 447–474, 2008.

[9] J. Baik and J. W. Silverstein, "Eigenvalues of large sample covariance matrices of spiked population models," *Journal of Multivariate Analysis*, vol. 97, no. 6, pp. 1382–1408, 2006.

[10] J. Baik, G. Ben Arous, and S. Péché, "Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices," *The Annals of Probability*, vol. 33, no. 5, pp. 1643–1697, 2005.

[11] F. Benaych-Georges and R. R. Nadakuditi, "The eigenvalues and eigenvectors of finite, low rank perturbations of large random matrices," *Advances in Mathematics*, vol. 227, no. 1, pp. 494 – 521, 2011.

[12] Q. Berthet and P. Rigollet, "Optimal detection of sparse principal components in high dimension," *The Annals of Statistics*, vol. 41, no. 1, pp. 1780—-1815, 2013.

[13] R. J. Bhansali, L. Giraitis, and P. S. Kokoszka, "Convergence of quadratic forms with nonvanishing diagonal," *Statistics & Probability Letters*, vol. 77, no. 7, pp. 726–734, 2007.

[14] P. Bianchi, M. Debbah, M. Maida, and J. Najim, "Performance of statistical tests for single-source detection using random matrix theory," *IEEE Transactions on Information Theory*, vol. 57, no. 4, pp. 2400–2419, 2011.

[15] P. J. Bickel and E. Levina, "Regularized estimation of large covariance matrices," *The Annals of Statistics*, vol. 36, pp. 199—227, 2008.

[16] P. Billingsley, *Probability and measure*, 3rd ed., ser. Wiley Series in Probability and Mathematical Statistics. New York: John Wiley & Sons Inc., 1995.

[17] P. J. Brockwell and R. A. Davis, *Time series: theory and methods*, ser. Springer Series in Statistics. New York: Springer, 2006, reprint of the second (1991) edition.

[18] T. T. Cai, Z. Ren, and H. H. Zhou, "Optimal rates of convergence for estimating Toeplitz covariance matrices," *Probability Theory and Related Fields*, vol. 156, pp. 101–143, 2013.

[19] T. T. Cai, C.-H. Zhang, and H. H. Zhou, "Optimal rates of convergence for covariance matrix estimation," *The Annals of Statistics*, vol. 38, pp. 2118—2144, 2010.

[20] L. S. Cardoso, M. Debbah, P. Bianchi, and J. Najim, "Cooperative spectrum sensing using random matrix theory," in *IEEE 3rd International Symposium on Wireless Pervasive Computing*, 2008, pp. 334–338.

[21] F. Chapon, R. Couillet, W. Hachem, and X. Mestre, "The outliers among the singular values of large rectangular random matrices with additive fixed rank deformation," *Markov Processes and Related Fields*, vol. 20, pp. 183–228, 2014.

[22] R. Couillet and W. Hachem, "Fluctuations of spiked random matrix models and failure diagnosis in sensor networks," *IEEE Transactions on Information Theory*, vol. 59, no. 1, pp. 509–525, 2013.

[23] R. Couillet, F. Pascal, and J. W. Silverstein, "Robust estimates of covariance matrices in the large dimensional regime," *to appear in IEEE Transactions on Information Theory*, 2014.

[24] F. Digham, M.-S. Alouini, and M. K. Simon, "On the energy detection of unknown signals over fading channels," *IEEE Transactions on Communications*, vol. 55, no. 1, pp. 21–24, 2007.

[25] C. Donati-Martin, M. Capitaine, and D. Féral, "Central limit theorems for eigenvalues of deformations of Wigner matrices," *Annales de l'IHP - Probabilités et Statistiques*, vol. 48, no. 1, pp. 107–133, 2012.

[26] N. El Karoui, "Tracy-Widom limit for the largest eigenvalue of a large class of complex sample covariance matrices," *The Annals of Probability*, vol. 35, no. 2, pp. 663–714, 2007.

[27] O. Feldheim and S. Sodin, "A universality result for the smallest eigenvalues of certain sample covariance matrices," *Geometric and Functional Analysis*, vol. 20, no. 1, pp. 88–123, 2010.

[28] D. Féral and S. Péché, "The largest eigenvalue of rank one deformation of large Wigner matrices," *Journal of Mathematical Physics*, vol. 272, pp. 185–228, 2007.

[29] ——, "The largest eigenvalues of sample covariance matrices for a spiked population: diagonal case," *Journal of Mathematical Physics*, 2009.

[30] R. M. Gray, *Toeplitz and circulant matrices: A review.* Now Publisher, 2006.

[31] U. Grenander and G. Szegő, *Toeplitz forms and their applications*, 2nd ed. New York: Chelsea Publishing Co., 1984.

[32] W. Hachem, P. Loubaton, X. Mestre, J. Najim, and P. Vallet, "A subspace estimator for fixed rank perturbations of large random matrices," *Journal of Multivariate Analysis*, vol. 114, pp. 427–447, 2013.

[33] R. A. Horn and C. R. Johnson, *Matrix analysis.* Cambridge: Cambridge University Press, 1990.

[34] K. Johansson, "Shape fluctuations and random matrices," *Communications in Mathematical Physics*, vol. 209, pp. 437–476, 2000.

[35] M. Johnstone, "On the distribution of the largest eigenvalue in principal components analysis," *The Annals of Statistics*, vol. 29, pp. 295–327, 2001.

[36] O. Kallenberg, *Foundations of modern probability*, ser. Probability and its Applications (New York). New York: Springer-Verlag, 1997.

[37] A. Kammoun, M. Kharouf, W. Hachem, and J. Najim, "A Central Limit Theorem for the SINR at the LMMSE estimator output for large-dimensional signals," *IEEE Transactions on Information Theory*, vol. 55, no. 11, pp. 5048–5063, 2009.

[38] V. Kostylev, "Energy detection of a signal with random amplitude," in *IEEE International Conference on Communications*, vol. 3, 2002, pp. 1606–1610.

[39] S. Kritchman and B. Nadler, "Determining the number of components in a factor model from limited noisy data," *Chemometrics and Intelligent Laboratory Systems*, vol. 94, no. 1, pp. 19–32, 2008.

[40] ——, "Non-parametric detection of the number of signals: hypothesis testing and random matrix theory," *IEEE Transactions on Signal Processing*, vol. 57, no. 10, pp. 3930–3941, 2009.

[41] M. Ledoux, *The concentration of measure phenomenon*, ser. Mathematical Surveys and Monographs. Providence, RI: American Mathematical Society, 2001, vol. 89.

[42] E. L. Lehman and J. P. Romano, *Testing Statistical Hypotheses*, ser. Springer Texts in Statistics. New York: Springer, 2006.

[43] A. P. Liavas and P. A. Regalia, "On the behavior of information theoretic criteria for model order selection," *IEEE Transactions on Signal Processing*, vol. 49, no. 8, pp. 1689–1695, 2001.

[44] V. A. Marčenko and L. A. Pastur, "Distribution of eigenvalues for some sets of random matrices," *Mathematics of the USSR-Sbornik*, vol. 1, no. 4, pp. 457–483, 1967.

[45] X. Mestre and M. A. Lagunas, "Modified subspace algorithms for DoA estimation with large arrays," *IEEE Transactions on Signal Processing*, vol. 56, no. 2, pp. 598–614, 2008.

[46] R. R. Nadakuditi and A. Edelman, "Sample eigenvalue based detection of high-dimensional signals in white noise using relatively few samples," *IEEE Transactions on Signal Processing*, vol. 56, no. 7, pp. 2625–2638, 2008.

[47] R. R. Nadakuditi and J. W. Silverstein, "Fundamental limit of sample generalized eigenvalue based detection of signals in noise using relatively few signal-bearing and noise-only samples," *IEEE Journal of Selected Topics in Signal Processing*, vol. 4, no. 3, pp. 468–480, 2010.

[48] B. Nadler, "On the distribution of the ratio of the largest eigenvalue to the trace of a Wishart matrix," *Journal of Multivariate Analysis*, vol. 102, no. 2, pp. 363–371, 2011.

[49] B. Nadler and I. M. Johnstone, "Detection performance of Roy's largest root test when the noise covariance matrix is arbitrary," in *IEEE Statistical Signal Processing Workshop*, 2011, pp. 681–684.

[50] T. Næs, T. Isaksson, T. Fearn, and T. Davies, *A User-Friendly Guide to Multivariate Calibration and Classification.* Chichester, UK: NIR Publications, 2002.

[51] N. K. Nikolski, *Operators, functions, and systems: an easy reading. Vol. 2*, ser. Mathematical Surveys and Monographs. Providence, RI: American Mathematical Society, 2002, vol. 93.

[52] L. Pastur and M. Ŝerbina, *Eigenvalue distribution of large random matrices.* American Mathematical Society, 2011, vol. 171.

[53] D. Paul, "Asymptotics of sample eigenstructure for a large dimensional spiked covariance model," *Statististica Sinica*, vol. 17, no. 4, pp. 1617–1642, 2007.

[54] S. Péché, "The largest eigenvalue of small rank perturbations of Hermitian random matrices," *Probability Theory and Related Fields*, vol. 134, pp. 127–173, 2006.

[55] F. Penna, R. Garello, and M. Spirito, "Cooperative spectrum sensing based on the limiting eigenvalue ratio distribution in Wishart matrices," *IEEE Communications Letters*, vol. 13, no. 7, pp. 507–509, 2009.

[56] J. Rissanen, "Modeling by shortest data description," *Automatica*, vol. 14, pp. 465–471, 1978.

[57] R. O. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Transactions on Antennas and Propagation*, vol. 34, no. 3, pp. 276–280, 1986.

[58] J. W. Silverstein, "The limiting eigenvalue distribution of a multivariate f-matrix," *SIAM Journal of Mathematical Analysis*, vol. 16, no. 3, pp. 641–646, 1985.

[59] J. W. Silverstein and Z. D. Bai, "On the empirical distribution of eigenvalues of a class of large-dimensional random matrices," *Journal of Multivariate Analysis*, vol. 54, no. 2, pp. 175–192, 1995.

[60] J. W. Silverstein and S. Choi, "Analysis of the limiting spectral distribution of large-dimensional random matrices," *Journal of Multivariate Analysis*, vol. 54, no. 2, pp. 295–309, 1995.

[61] J. W. Silverstein and P. L. Combettes, "Large dimensional random matrix theory for signal detection and estimation in array processing," in *IEEE Sixth SP Workshop on Statistical Signal and Array Processing*, 1992, pp. 276–279.

[62] C. A. Tracy and H. Widom, "On orthogonal and symplectic matrix ensembles," *Communications in Mathematical Physics*, vol. 177, pp. 727–754, 1996.

[63] H. Urkowitz, "Energy detection of unknown deterministic signals," *Proceedings of the IEEE*, vol. 55, no. 4, pp. 523–531, 1967.

[64] P. Vallet, P. Loubaton, and X. Mestre, "Improved subspace estimation for multivariate observations of high dimension: the deterministic signal case," *IEEE Transactions on Information Theory*, vol. 58, no. 2, pp. 1043–1068, 2012.

[65] P. Vallet and P. Loubaton, "Toeplitz rectification and DoA estimation with MUSIC," in *IEEE ICASSP'14, Florence, Italy*. IEEE, 2014, pp. 2237–2241.

[66] A. W. Van der Vaart, *Asymptotic Statistics*. New York: Cambridge University Press, 2000.

[67] M. Wax and T. Kailath, "Detection of signals by information theoretic criteria," *IEEE Transanstions on Acoustics, Speech, Signal Processing*, vol. 33, no. 2, pp. 387–392, 1985.

[68] E. Wigner, "On the distribution of the roots of certain symmetric matrices," *The Annals of Mathematics*, vol. 67, pp. 325–327, 1958.

[69] E. P. Wigner, "Characteristic Vectors of Bordered Matrices With Infinite Dimensions," *The Annals of Mathematics*, vol. 62, no. 3, pp. 548–564, 1955.

[70] J. Wishart, "The generalized product moment distribution in samples from a normal multivariate population," *Biometrika*, vol. 20, no. 1–2, pp. 32–52, 1928.

[71] W. B. Wu and M. Pourahmadi, "Banding sample covariance matrices of stationary processes," *Statististica Sinica*, vol. 19, pp. 1755—-1768, 2009.

[72] H. Xiao and W. B. Wu, "Covariance matrix estimation for stationary time series," *The Annals of Statistics*, vol. 40, pp. 466—493, 2012.

[73] W. Xu and M. Kaveh, "Analysis of the performance and sensitivity of eigendecomposition-based detectors," *IEEE Transactions on Signal Processing*, vol. 43, no. 6, pp. 1413–1426, 1995.

[74] Y. Zeng and Y.-C. Liang, "Eigenvalue-based spectrum sensing algorithms for cognitive radio," *IEEE Transactions on Communications*, vol. 57, no. 6, pp. 1784–1793, 2009.

[75] L. C. Zhao, P. R. Krishnaiah, and Z. D. Bai, "On detection of the number of signals in presence of white noise," *Journal of Multivariate Analysis*, vol. 20, no. 1, pp. 1–25, 1986.

[76] ——, "On detection of the number of signals when the noise covariance matrix is arbitrary," *Journal of Multivariate Analysis*, vol. 20, no. 1, pp. 26–49, 1986.

List of Figures