



HAL
open science

Asymptotics and fluctuations of largest eigenvalues of empirical covariance matrices associated with long memory stationary processes

Peng Tian

► **To cite this version:**

Peng Tian. Asymptotics and fluctuations of largest eigenvalues of empirical covariance matrices associated with long memory stationary processes. General Mathematics [math.GM]. Université Paris-Est, 2018. English. NNT : 2018PESC1131 . tel-02085884

HAL Id: tel-02085884

<https://theses.hal.science/tel-02085884>

Submitted on 31 Mar 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



THÈSE DE DOCTORAT DE L'UNIVERSITÉ PARIS-EST

Spécialité : Mathématiques

Ecole doctorale MSTIC

Présentée par Peng TIAN

Pour obtenir le grade de

DOCTEUR de l'UNIVERSITÉ PARIS EST

**Asymptotiques et fluctuations des plus grandes valeurs propres de
matrices de covariance empirique associées à des processus
stationnaires à longue mémoire**

Soutenue le 10 décembre devant le jury composé de :

Isabelle Chalendar	Université Marne-la-vallée	Examineur
Romain Couillet	Centrale Supélec, Université ParisSaclay	Examineur
Catherine Donati-Martin	Université de Versailles Saint Quentin	Rapporteur
Walid Hachem	Université Marne-la-vallée	Examineur
Iain Johnstone	Stanford University	Rapporteur
Mylène Maïda	Université des Sciences et Technologies de Lille	Examineur
Florence Merlevède	Université Marne-la-vallée	Directrice de thèse
Jamal Najim	Université Marne-la-vallée	Directeur de thèse

**Asymptotics and fluctuations of largest eigenvalues of
empirical covariance matrices associated with long memory
stationary processes**

Laboratoire LAMA CRS UMR 8050
Université Paris-Est Marne-la-vallée
5, Boulevard Descartes, Champs-sur-Marne
77454 Marne-la-vallée cedex 2, France

ACKNOWLEDGEMENTS

Firstly, I would like to express my sincere gratitude to my advisors Prof. Florence Merleède and Prof. Jamal Najim for their continuous support of my Ph.D study and related research, for their patience, motivation, and immense knowledge, and for their guidance in all the time of research and writing of this thesis. Their advices and the opportunities that they have proposed are also very helpful for my future academical career.

Besides my advisors, I would also like to thank Prof. Isabelle Chalendar, Prof. Walid Hachem, Prof. Iain Johnstone and Prof. Philippe Loubaton for their constructive discussions during my research. My sincere thanks extend to the the rest of my thesis committee: Prof. Romain Couillet, Prof. Catherine Donati-Martin, Prof. Mylne Maïda, for their insightful comments and encouragement, but also for the hard question which incented me to widen my research from various perspectives.

My sincere thanks also goes to our laboratories LAMA and LIGM for their financial and material supports. I would also like to thank our secretories for their warm help in the administrative procedures.

Last but not the least, I would like to thank my family, for supporting me spiritually, and my friends for their accompanies during my stay in France.

RÉSUMÉ

Asymptotiques et fluctuations des plus grandes valeurs propres de matrices de covariance empirique associées à des processus stationnaires à longue mémoire

by

Peng TIAN

Soit $(\mathcal{X}_t)_{t \in \mathbb{Z}}$ un processus stationnaire centré. Nous considérons la matrice d'autocovariance empirique

$$Q_N := \frac{1}{n} X_N X_N^*,$$

où $X_N = (X_{i,j})$ est une $N \times n$ matrice dont les colonnes $X_{\cdot,j} = (X_{1,j}, \dots, X_{N,j})^\top$ sont des copies i.i.d. du vecteur aléatoire $(\mathcal{X}_1, \dots, \mathcal{X}_N)^\top$. Pour N fixé, lorsque n tend vers l'infini, et sous certaines conditions, les valeurs propres de Q_N convergent vers les valeurs propres de la matrice d'autocovariance sous-jacente. Cependant l'hypothèse $N \ll n$ n'est bien souvent pas justifiée, notamment si l'on travaille avec des données de grande dimension. Il est alors plus réaliste de supposer que N est du même ordre que la taille de l'échantillon n , et donc que $N = N_n \rightarrow \infty$ tel que $\lim_{n \rightarrow \infty} N/n =: r \in (0, \infty)$. On s'intéresse alors le comportement asymptotique du spectre de Q_N . Soit μ^{Q_N} la distribution spectrale empirique de Q_N , à savoir

$$\mu^{Q_N} = \frac{1}{N} \sum_{i=1}^N \delta_{\lambda_i(Q_N)}$$

où les $\lambda_i(Q_N)$ sont les valeurs propres de Q_N et δ_x la masse de Dirac au point x . Merlevède et Peligrad [43] ont montré que si le processus $(\mathcal{X}_t)_{t \in \mathbb{Z}}$ est strictement stationnaire centré, dans L^2 satisfaisant des conditions faibles de régularité, alors presque sûrement, μ^{Q_N} convergeait

étroitement vers une distribution non aléatoire μ ne dépendant que de la densité spectrale du processus $(\mathcal{X}_t)_{t \in \mathbb{Z}}$.

Ainsi, pour une très large variété de processus stationnaires, y compris ceux à longue mémoire, la distribution spectrale empirique de la matrice d'autocovariance associée admet toujours une limite ne dépendant que de la densité spectrale du processus sous-jacent. Ce résultat s'applique en particulier aux processus linéaires $\mathcal{X}_t = \sum_{j \geq 0} a_j \varepsilon_{t-j}$ où $(\varepsilon_j)_j$ est une suite de v.a. réelles i.i.d. centrées, dans L^2 et $(a_k) \in \ell^2$.

Ce résultat montre aussi que si la densité spectrale est continue et bornée (ce qui est le cas des processus linéaires dont les coefficients (a_k) sont absolument sommables), alors la distribution spectrale limite a un support compact. Par contre si le processus $(\mathcal{X}_t)_{t \in \mathbb{Z}}$ exhibe de la longue mémoire (en particulier si les covariances ne sont pas absolument sommables), le support de la loi limite n'est plus compact et des études plus fines du comportement des valeurs propres sont alors naturelles afin de pouvoir étudier le "degré" de longue mémoire du processus sous-jacent. Notamment, nous nous intéresserons aux plus grandes valeurs propres de Q_N .

Ainsi, motivé par l'étude des processus stationnaires à mémoire longue, nous étudions, dans cette thèse, les asymptotiques et les fluctuations des m plus grandes valeurs propres $\lambda_1(S_N(T)), \dots, \lambda_m(S_N(T))$ (avec m un entier fixé arbitraire) de la matrice

$$S_N(T) = \frac{1}{n} T_N^{\frac{1}{2}} Z_N Z_N^* T_N^{\frac{1}{2}} \quad (0.1)$$

où $Z_N = (Z_{i,j})$ est une $N \times n$ matrice ayant entrées i.i.d. réelles ou complexes centrées avec la variance unitaire, et T_N est une $N \times N$ matrice de Toeplitz réelle symétrique déterministe, dont la norme spectrale tend vers l'infini. Nous nous intéresserons plus particulièrement au cas où T_N est la N -ième matrice d'autocovariance d'un processus stationnaire à long mémoire, de sorte que S_N peut être considéré comme une représentation linéaire de la matrice d'autocovariance empirique du processus. Nous étudions ce problème dans le régime $N, n \rightarrow \infty$ et $Nn^{-1} \rightarrow r \in (0, \infty)$.

Nous sommes d'abord amenés à étudier les plus grandes valeurs propres ainsi que leurs vecteurs propres associés de matrices de Toeplitz déterministes de dimensions croissantes.

Dans le cas où

$$T_N = (\gamma(i-j))_{i,j=0}^{N-1}$$

avec

$$\gamma(h) := (1 + |h|)^\rho L(|h|)$$

pour un certain $\rho \in (-1, 0)$ et une fonction L à variation lente à l'infini, nous décrivons le comportement des m plus grandes valeurs propres de T_N ainsi que celui des vecteurs propres associés. En particulier, nous prouvons que les m plus grandes valeurs propres tendent vers l'infini et satisfont une propriété de type trou spectral multiple. Nous prouvons également une propriété de délocalisation pour les vecteurs propres associés.

Notre étude ne se limite pas au cas où les T_N sont les matrices de Toeplitz symétriques considérées ci-dessus. Nous considérerons également le modèle générique $S_N(\Gamma)$ défini par

$$S_N(\Gamma) = \frac{1}{n} \Gamma_N^{\frac{1}{2}} Z_N Z_N^* \Gamma_N^{\frac{1}{2}},$$

où Γ_N est une $N \times N$ matrice hermitienne semi-définie positive déterministe dont les plus grandes valeurs propres $\lambda_1(\Gamma_N), \dots, \lambda_m(\Gamma_N)$ tendent vers l'infini. Nous étudions d'abord l'asymptotique des plus grandes valeurs propres et nous montrons que sous de faibles conditions, $\lambda_j(S_N(\Gamma)) \sim \lambda_j(\Gamma_N)$ pour tout $j = 1, \dots, m$. Nous étudions ensuite leurs fluctuations dans différents cas. Dans tous ces cas, le vecteur aléatoire $(\lambda_1(S_N(\Gamma)), \dots, \lambda_m(S_N(\Gamma)))^\top$ est asymptotiquement gaussien sous la normalisation $\lambda_{\max}(\Gamma_N)/\sqrt{n}$, mais la matrice de covariance de la distribution limite dépend fortement de Γ_N .

Nous considérons d'abord le cas où Γ_N est block-diagonale. Dans cette situation, le résultat obtenu peut être appliqué pour étudier le comportement asymptotique des plus grandes valeurs propres des matrices d'autocovariance empiriques associées aux processus stationnaires et gaussiens à longue mémoire.

On considère ensuite le cas où $\Gamma_N = T_N$ avec $\rho \in (-3/4, 0)$. Dans ce cas, les fluctuations ne dépendent que de $\mathbb{E}Z_{i,j}^2$. Ainsi, si les entrées $Z_{i,j}$ sont réelles de variance unitaire, ou complexes avec $\mathbb{E}Z_{i,j}^2 = 0$, les plus grandes valeurs propres fluctuent de façon complètement universelle.

Notons que si $\rho < -1$, le processus $T_N^{1/2} Z_N$ est alors un processus stationnaire à mémoire courte et la plus grande valeur propre fluctue suivant la loi de Tracy-Widom (TW). Le cas $\rho = -1$ représente donc la frontière entre la longue mémoire et la courte mémoire. Ainsi, plus ρ se rapproche de -1 et plus on se rapproche du cadre de la courte mémoire. Nous conjecturons que lorsque $\rho \in (-1, -3/4]$, le résultat asymptotique de fluctuations gaussiennes reste vrai. Nous donnons une preuve partielle dans l'hypothèse (jusqu'ici non prouvée pour T_N) que tous les vecteurs propres de T_N sont uniformément délocalisés.

ABSTRACT

Asymptotics and fluctuations of largest eigenvalues of empirical covariance matrices
associated with long memory stationary processes

by

Peng TIAN

Let $(\mathcal{X}_t)_{t \in \mathbb{Z}}$ be a centered stationary process, we consider the empirical autocovariance matrix

$$Q_N := \frac{1}{n} X_N X_N^*,$$

where $X_N = (X_{i,j})$ is a $N \times n$ matrix whose columns $X_{\cdot,j} = (X_{1,j}, \dots, X_{N,j})^\top$ are i.i.d. copies of the random vector $(\mathcal{X}_1, \dots, \mathcal{X}_N)^\top$. For a fixed N , when n tends to infinity, and under certain conditions, the eigenvalues of Q_N converge to the eigenvalues of the underlying autocovariance matrix. However, the hypothesis $N \ll n$ is not often satisfied, especially when working with large data. It is then more realistic to assume that N is of the same order as the sample size n , and hence $N = N_n \rightarrow \infty$ such that $\lim_{n \rightarrow \infty} N/n =: r \in (0, \infty)$. We are interested in the asymptotic behavior of the spectrum of Q_N . Let μ^{Q_N} be the empirical spectral distribution of Q_N , recall that

$$\mu^{Q_N} = \frac{1}{N} \sum_{i=1}^N \delta_{\lambda_i(Q_N)},$$

where $\lambda_i(Q_N)$ are the eigenvalues of Q_N and δ_x is the Dirac mass at the point x . Merlevède and Peligrad [43] showed that if the process $(\mathcal{X}_t)_{t \in \mathbb{Z}}$ is strictly stationary centered, in L^2 , and satisfies some weak conditions of regularity, then almost surely, μ^{Q_N} converges weakly

towards a non-random distribution μ , depending only on the spectral density of the process $(\mathcal{X}_t)_{t \in \mathbb{Z}}$.

Thus, for a very wide range of stationary processes, including those with long memory, the empirical spectral distribution of the associated autocovariance matrix always has a limit depending only on the spectral density of the underlying process. This result applies in particular to linear processes $\mathcal{X}_t = \sum_{j \geq 0} a_j \varepsilon_{t-j}$ where $(\varepsilon_j)_j$ is a sequence of real i.i.d. centered r.v. in L^2 and $(a_k) \in \ell^2$.

This result also shows that if the spectral density is continuous and bounded (which is the case of linear processes whose coefficients (a_k) are absolutely summable), then the limit spectral distribution has a compact support. To the contrary, if the process $(\mathcal{X}_t)_{t \in \mathbb{Z}}$ exhibits the long memory (in particular if the covariances are not absolutely summable), the support of the limit spectral distribution is no longer compact and finer studies of the behavior of the eigenvalues are then natural, in order to study the "degree" of long memory of the underlying process. In particular, we will be interested in studying the asymptotic behaviour of the largest eigenvalues of Q_N .

Motivated by the study of long memory stationary processes, we then study in this thesis the asymptotics and the fluctuations of the m (with m an arbitrary fixed integer) largest eigenvalues $\lambda_1(S_N(T)), \dots, \lambda_m(S_N(T))$ of the matrix

$$S_N(T) = \frac{1}{n} T_N^{\frac{1}{2}} Z_N Z_N^* T_N^{\frac{1}{2}} \quad (0.2)$$

where $Z_N = (Z_{i,j})$ is a $N \times n$ matrix having i.i.d. centered real or complex entries with variance one, and T_N is a $N \times N$ deterministic real symmetric Toeplitz matrix, whose spectral norm goes to infinity. More precisely, we are particularly interested in considering T_N as the N th autocovariance matrix of a long range dependent stationary process, so that S_N can be viewed as a linear representation of sample autocovariance matrix of the process. We study this problem in the regime $N, n \rightarrow \infty$ and $Nn^{-1} \rightarrow r \in (0, \infty)$.

First we are led to study the largest eigenvalues and the associated eigenvectors of

deterministic Toeplitz matrices of growing dimensions. For the case where

$$T_N = (\gamma(i-j))_{i,j=0}^{N-1}$$

with

$$\gamma(h) := (1 + |h|)^\rho L(|h|)$$

for some $\rho \in (-1, 0)$ and some slowly varying function L at infinity, we give a complete picture of the behavior of the m largest eigenvalues and the associated eigenvectors of T_N . In particular, we prove that the m largest eigenvalues tend to infinity, and satisfy a multiple spectral gap property. We also prove a delocalization property for their associated eigenvectors.

Our study is not limited to the case where T_N 's are the symmetric Toeplitz matrices as considered above. We shall also consider the generic model $S_N(\Gamma)$ defined by

$$S_N(\Gamma) = \frac{1}{n} \Gamma_N^{\frac{1}{2}} Z_N Z_N^* \Gamma_N^{\frac{1}{2}},$$

where Γ_N is a $N \times N$ deterministic positive semi-definite Hermitian matrix whose largest eigenvalues $\lambda_1(\Gamma_N), \dots, \lambda_m(\Gamma_N)$ tend to infinity. We first study the asymptotics of the largest eigenvalues and prove that, under some mild conditions, we have $\lambda_j(S_N(\Gamma)) \sim \lambda_j(\Gamma_N)$ for any $j = 1, \dots, m$. Then we study their fluctuations in different cases. In all these cases, the random vector $(\lambda_1(S_N(\Gamma)), \dots, \lambda_m(S_N(\Gamma)))^\top$ is asymptotically gaussian in the scale $\sqrt{n} \lambda_{\max}^{-1}(\Gamma_N)$, but the covariance matrix of the limiting distribution strongly depends on Γ_N .

We first consider the case where Γ_N is block-diagonal. In this situation, the obtained result can be applied to study the asymptotic behavior of the largest eigenvalues of the empirical autocovariance matrices associated with Gaussian long memory stationary process (LMSP).

Then we consider the case where $\Gamma_N = T_N$ with $\rho \in (-3/4, 0)$. In this case the fluctuations depend only on $\mathbb{E}Z_{i,j}^2$. So if the entries $Z_{i,j}$ are real with variance one, or complex with $\mathbb{E}Z_{i,j}^2 = 0$, the top eigenvalues fluctuate in a completely universal way.

Notice that if $\rho < -1$, then the process $T_N^{1/2}Z_N$ is a short memory stationary process (**SMSP**) and the limiting distribution of the largest eigenvalue suitably renormalized follows the **TW** distribution. The case $\rho = -1$ represents the threshold between **LMSP** and **SMSP**. Hence the extra-difficulty that arises when $\rho \in (-1, 0)$ is close to minus one. We conjecture that when $\rho \in (-1, -3/4]$, the same result holds. We give a partial proof under the assumption (so far unproved) that all eigenvectors of T_N are uniformly delocalized.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS	i
RÉSUMÉ	ii
ABSTRACT	vi
LIST OF FIGURES	xii
LIST OF ABBREVIATIONS	xiii
CHAPTER	
I. Introduction	1
1.1 Long memory stationary process and its sample covariance matrices	1
1.2 Some references on large empirical covariance matrices	6
1.3 Organization and some notations	9
1.4 Main results	12
1.4.1 Asymptotic spectral properties of Toeplitz matrices	12
1.4.2 Asymptotics of largest eigenvalues of $S_N(\Gamma)$	15
1.4.3 CLT for largest eigenvalues of $S_N(\Gamma)$	17
1.4.4 Applications to long memory stationary processes	19
1.4.5 Extension of the results for large covariance matrices close to short memory	21
II. Unbounded largest eigenvalue of large sample covariance matrices: Asymptotics, fluctuations and applications	23
2.1 Abstract	23
2.2 Introduction	24
2.3 Notations and main theorems	32
2.3.1 Notations and assumptions	32
2.3.2 Main results	35
2.3.3 Application to large sample covariance matrices associated with long memory processes	38
2.4 Additional applications and simulations	41
2.4.1 Additional applications to Gaussian stationary processes .	41

2.4.2	Numerical simulations	43
2.4.3	Open questions	46
2.5	Proofs of Proposition II.3 and Theorem II.4	47
2.5.1	A short reminder of results related to large covariance matrices	47
2.5.2	Reduction to the bounded model	50
2.5.3	Proof of Proposition II.11	51
2.5.4	Proof of Theorem II.12	54
2.6	Proof of Theorem II.8	63
2.6.1	Widom-Shampine’s Lemma and convergence of operators	63
2.6.2	Convergence of eigenvalues and simplicity of the largest eigenvalue	68
 III. Joint CLT for top eigenvalues of empirical covariance matrices of long memory stationary processes		71
3.1	Abstract	71
3.2	Introduction	72
3.3	Statement of the main theorems	81
3.3.1	Spectral properties of Toeplitz matrices	81
3.3.2	A general CLT	83
3.3.3	The CLT for covariance matrices of long memory processes	85
3.4	Proofs of the theorems on Toeplitz matrices	87
3.4.1	Some preparation	87
3.4.2	Proof of Theorem III.1	88
3.4.3	Proof of Theorem III.2	92
3.4.4	Proof of Proposition III.3.	94
3.5	Proof of Theorem III.4	95
3.5.1	Some preparation and outline of the proof	95
3.5.2	Proof of Lemma III.8	98
3.5.3	Proof of Lemma III.9	111
 IV. Some extensions in case of large covariance matrices close to short memory		115
4.1	Introduction	115
4.2	Proof of Theorem I.15	116
 APPENDICES		133
A.1	Proof of Proposition I.6	135
A.2	Proof of Proposition I.8	136
 BIBLIOGRAPHY		137

LIST OF FIGURES

Figure

- 2.1 Convergence of $\lambda_{\max}(S_N)/\lambda_{\max}(\Gamma_N)$ to 1. In 2.1(a), the values $\lambda_{\max}(S_N)$ and $\lambda_{\max}(\Gamma_N)$ are plotted as crosses and solid points respectively; in 2.1(b) the values of the ratio $\lambda_{\max}(S_N)/\lambda_{\max}(\Gamma_N)$ are plotted as crosses, compared with the constant 1. 44
- 2.2 Fluctuations of $\lambda_{\max}(S_{1000})$, with Z_{1000} having Gaussian entries in 2.2(a) and standardized exponential entries in 2.2(b). Simulation with 900 samples. 44
- 2.3 Fluctuations of $\lambda_{\max}(S_N)$ in the case of symmetric Bernoulli entries. Γ_N is diagonal in 2.3(a) and nondiagonal in 2.3(b). 45

LIST OF ABBREVIATIONS

ESD empirical spectral distribution

LHS left handside

LMSP long memory stationary process

LSD limiting spectral distribution

RHS right handside

SMSP short memory stationary process

TW Tracy-Widom

CHAPTER I

Introduction

1.1 Long memory stationary process and its sample covariance matrices

A stochastic process $(\mathcal{X}_t)_{t \in \mathbb{Z}}$ is (second order) stationary if the following conditions are satisfied:

$$\mathbb{E}|\mathcal{X}_t|^2 < \infty, \quad \mathbb{E}\mathcal{X}_t = \mu_{\mathcal{X}} \quad \text{and} \quad \text{Cov}(\mathcal{X}_{t+h}, \mathcal{X}_t) = \text{Cov}(\mathcal{X}_h, \mathcal{X}_0) = \gamma(h) \quad \forall t, h \in \mathbb{Z}$$

where $\text{Cov}(\mathcal{X}_{t+h}, \mathcal{X}_t) = \mathbb{E}(\mathcal{X}_{t+h} - \mathbb{E}\mathcal{X}_{t+h})(\overline{\mathcal{X}_t - \mathbb{E}\mathcal{X}_t})$, $\mu_{\mathcal{X}}$ is a constant and $\gamma : \mathbb{Z} \rightarrow \mathbb{C}$ is some positive definite function, usually called the autocovariance function of the process. Note that $\gamma(0)$ is positive and $\gamma(-h) = \overline{\gamma(h)}$ for all $h \in \mathbb{Z}$. By stationarity, the autocovariance matrices T_N of the process

$$T_N := \text{Cov} \begin{pmatrix} \mathcal{X}_{t+1} \\ \vdots \\ \mathcal{X}_{t+N} \end{pmatrix} = \begin{pmatrix} \gamma(0) & \gamma(-1) & \dots & \gamma(-N+1) \\ \gamma(1) & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \gamma(-1) \\ \gamma(N-1) & \dots & \gamma(1) & \gamma(0) \end{pmatrix} \quad (1.1)$$

are positive semidefinite Hermitian Toeplitz matrices.

By Herglotz's Theorem, there exists a finite positive measure α on $(-\pi, \pi]$, whose Fourier coefficients are exactly $\gamma(h)$, i.e.

$$\gamma(h) = \frac{1}{2\pi} \int_{(-\pi, \pi]} e^{-ihx} d\alpha(x), \quad \forall h \in \mathbb{Z}.$$

This measure α is usually called the symbol of the Toeplitz matrices T_N .

The process is usually said to have *short memory* or to be *short range dependent* if $\sum_{h \in \mathbb{Z}} |\gamma(h)| < \infty$. Obviously in this case, the symbol α has a continuous density with respect to the Lebesgue measure. However, there are several definitions of long range dependence, all strongly related but not always equivalent, see for instance [45, Chapter 2]. For further discussion, we state them here. First we have to define the slowly varying functions, and also the quasi-monotone functions.

Definition I.1 (Slowly varying function). A function L is slowly varying at infinity if it is positive and, for any $a > 0$,

$$\lim_{x \rightarrow \infty} \frac{L(ax)}{L(x)} = 1.$$

Definition I.2 (Quasi-monotone function). A positive function f defined on $[0, +\infty)$ and of locally bounded variation is said to be quasi-monotone if for some $\delta > 0$,

$$\int_0^x t^\delta |df(t)| = O(x^\delta f(x)).$$

For a nonnegative function, if it is non-decreasing, then it is quasi-monotone. But the non-increasing property does not implies the quasi-monotonicity. However when a function is slowly varying, then it is quasi-monotone whenever it is monotone. See for example Section 2.7 of [14].

A stationary process $(\mathcal{X}_t)_{t \in \mathbb{Z}}$ is said to be *long range dependent*, or to have *long memory* or *long range dependence*, with a long memory parameter $d \in (0, 1/2)$, if one of the following conditions hold.

CI The time series $(\mathcal{X}_t)_{t \in \mathbb{Z}}$ has a linear representation

$$\mathcal{X}_t = \mu_{\mathcal{X}} + \sum_{k=0}^{\infty} \psi_k Z_{t-k}$$

with $(Z_k)_{k \in \mathbb{Z}}$ a sequence of i.i.d. random variables, and $(\psi_k)_{k \geq 0}$ a sequence of numbers such that

$$\psi_k = (1 + |k|)^{d-1} L_1(|k|)$$

where L_1 is a slowly varying function at infinity.

C II The autocovariance function γ satisfies

$$\gamma(h) = (1 + |h|)^{2d-1} L_2(|h|)$$

with L_2 a slowly varying function at infinity. Notice that $\rho = 2d - 1 \in (-1, 0)$.

C III The autocovariance function γ is not absolutely summable

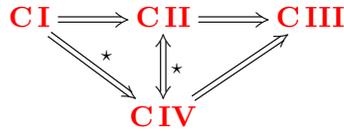
$$\sum_{h \in \mathbb{Z}} |\gamma(h)| = \infty.$$

C IV The symbol α of autocovariance matrices T_N has a density f_α satisfying

$$f_\alpha(x) = |x|^{-2d} L_4(|x|^{-1}), \quad 0 < |x| \leq \pi,$$

where L_4 is a slowly varying function at infinity.

The above conditions have the following relations:



where \Rightarrow denotes the implication and $\overset{*}{\Rightarrow}$ assumes that the slowly varying function is quasi-monotone.

Given a centered stationary process $(\mathcal{X}_t)_{t \in \mathbb{Z}}$, (i.e. with $\mu_{\mathcal{X}} = 0$) one can study the spectral properties of the sample covariance matrix

$$Q_N := \frac{1}{n} X_N X_N^* = \frac{1}{n} \sum_{j=1}^n X_{\cdot,j} X_{\cdot,j}^* \quad (1.2)$$

where X_N is a $N \times n$ random matrix whose columns $(X_{\cdot,j}, 1 \leq j \leq n)$ are i.i.d copies of the random vector $\mathcal{X}_{1:N} = (\mathcal{X}_1, \dots, \mathcal{X}_N)^\top$. In particular, let μ^{Q_N} be the empirical spectral

distribution (**ESD**) of Q_N , defined by

$$\mu^{Q_N} := \frac{1}{N} \sum_{k=1}^N \delta_{\lambda_k(Q_N)},$$

one may be interested in the weak limit of μ^{Q_N} , called the limiting spectral distribution (**LSD**). This has been treated for example in [43]. In particular, the authors proved that if the underlying stationary process is regular then the **LSD** depends only on the spectral density of the process. They characterized this limit in terms of Stieltjes transform via a certain simple equation. No rate of convergence to zero of the covariances is imposed, so, the underlying process can exhibit long memory. Note that if the stationary sequence has trivial left sigma field, the regularity condition is satisfied and then their result holds without any other additional assumptions. In particular, this is always true if the entries are functions of an i.i.d. sequence.

Another problem is the behaviour of individual eigenvalues, especially the asymptotics and fluctuations of the largest eigenvalues. Let

$$\lambda_1(Q_N) \geq \lambda_2(Q_N) \geq \dots \geq \lambda_N(Q_N)$$

be the eigenvalues of Q_N arranged in decreasing order, and let m be a fixed positive integer. We are interested in the study of the asymptotics and fluctuations of the random vector $(\lambda_1(Q_N), \dots, \lambda_m(Q_N))^T$. With this aim, we are lead to investigate the asymptotic locations $\ell_1^{(N)}, \dots, \ell_m^{(N)}$ of these eigenvalues, and to find proper normalizations $a_1^{(N)}, \dots, a_m^{(N)}$ such that for any $j = 1, \dots, m$,

$$a_j^{(N)} \left(\lambda_j(Q_N) - \ell_j^{(N)} \right)$$

converge in law to a non degenerate distribution (if possible) and to determine this distribution.

If the process is Gaussian, then Q_N has the linear form $S_N(T)$ defined by

$$S_N(T) := \frac{1}{n} T_N^{\frac{1}{2}} Z_N Z_N^* T_N^{\frac{1}{2}} \tag{1.3}$$

where T_N is the N th autocovariance matrix of the process, and Z_N is a $N \times n$ random matrix with i.i.d standard gaussian entries. This is a more classical model in the random matrix theory, but here the Toeplitz nature of T_N gives to this model some special perspectives.

It is well-known since Szegő's theorem that if the symbol α has a bounded density $f_\alpha \in L^\infty(-\pi, \pi)$, called the *spectral density* of T_N , then we have

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{k=1}^N \varphi(\lambda_k(T_N)) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \varphi(f_\alpha(x)) dx, \quad (1.4)$$

where $\varphi : \mathbb{R} \mapsto \mathbb{R}$ is a function continuous on the range of f_α . Tyrtyshnikov and Zamarashkin generalized in [55] the above result of Szegő and relaxed the condition on α . They proved that the convergence (1.4) holds for general Radon measure α , and with $f_\alpha \in L^1(-\pi, \pi)$ the density of the absolutely continuous part of α with respect to the Lebesgue measure dx on $(-\pi, \pi]$, and φ is a continuous function with compact support. In this case we can also call f_α the *spectral density* of T_N .

The equality (1.4) can be interpreted as the vague convergence of probability measures μ^{T_N} to the measure ν defined by the integral formula

$$\int \varphi d\nu = \frac{1}{2\pi} \int_{-\pi}^{\pi} \varphi(f_\alpha(x)) dx \quad \forall \varphi \in C_b, \quad (1.5)$$

where C_b denotes the space of all bounded continuous functions. The measure ν being a probability, the sequence μ^{T_N} is tight, and the vague convergence coincides with the weak convergence.

From the above result we can see that in any case of covariance matrix T_N , its **ESD** μ^{T_N} converges weakly. Moreover if its spectral density f_α is not bounded, then the **LSD** of T_N is not compactly supported, thus the largest eigenvalues of T_N tend to infinity. By the definitions of long memory and their relations, we conclude that the above happens under the condition **CIV**. It happens also under conditions **CI** and **CH**, provided that the slowly varying functions L_1, L_2 are quasi-monotone.

Clearly studying the limiting distribution of the largest eigenvalues of the sample covariance matrix Q_N defined by (1.2) via the model (1.3) is restricted to the Gaussian case.

However studying the generic sample covariance matrix $S_N(T)$ defined by (1.3) where (Z_{ij}) are i.i.d. but not necessarily Gaussian has an interest in itself since it corresponds to a linear **LMSP**. We shall also consider the generic sample covariance matrix

$$S_N(\Gamma) := \frac{1}{n} \Gamma_N^{\frac{1}{2}} Z_N Z_N^* \Gamma_N^{\frac{1}{2}} \quad (1.6)$$

where Γ_N is a deterministic positive semidefinite hermitian matrix whose **ESD** μ^{Γ_N} converge weakly and whose largest eigenvalues tend to infinity. We study the asymptotics and fluctuations of the m (with m an arbitrary fixed integer) largest eigenvalues of $S_N(\Gamma)$, as $N, n \rightarrow \infty$ and $N/n \rightarrow r \in (0, \infty)$. In the sequel, this regime will be denoted simply as $N, n \rightarrow \infty$.

1.2 Some references on large empirical covariance matrices

The model S_N defined in (1.6) is a classical model of sample covariance matrices in the random matrix theory, and its spectral properties have been intensively studied in the regime $N, n \rightarrow \infty$ in the last several decades.

At a global scale, the **LSD** of the **ESD** $\mu^{S_N} = N^{-1} \sum_{k=1}^N \delta_{\lambda_k(S_N)}$ has been described in the groundbreaking paper by Marčenko and Pastur [41]. In the important case where $S_N = \frac{1}{n} Z_N Z_N^*$, sometimes referred to as the *white noise* model, the limiting spectral distribution of μ^{S_N} is known as Marčenko-Pastur distribution and admits the following closed-form expression

$$\mathbb{P}_{MP}(d\lambda) := (1 - r^{-1})_+ \delta_0(d\lambda) + \frac{\sqrt{[(\lambda^+ - \lambda)(\lambda - \lambda^-)]_+}}{2\pi r \lambda} d\lambda, \quad \lambda^\pm = (1 \pm \sqrt{r})^2,$$

where $x_+ := \max(x, 0)$. Later, this result was improved by many others, see for instance [56, 34, 59, 49, 48]. In [48], Silverstein proved that for the model $S_N(\Gamma)$ defined in (1.6), if μ^{Γ_N} weakly converges to a certain probability ν supported on \mathbb{R}^+ (not necessarily with compact support), then almost surely, the **ESD** μ^{S_N} weakly converges to a deterministic distribution μ , whose Stieltjes transform g_μ is the unique solution with positive imaginary

part of the equation

$$g_\mu(z) = \int \frac{1}{s(1-r-rzg_\mu(z))-z} d\nu(s) \in \mathbb{C}^+, \quad \forall z \in \mathbb{C}^+. \quad (1.7)$$

Central limit theorems have also been established for linear spectral statistics $\sum_{i=1}^N f(\lambda_i(S_N))$, see for instance [34, 32, 3, 44].

At a local scale, the convergence and fluctuations of individual eigenvalues have been studied, with a special emphasis on the eigenvalues located near each edge of the connected components (bulk) of the **LSD** of S_N . The spiked eigenvalues, that is those which stay away from the bulk of the **LSD**, have also attracted a lot of attention.

For the white noise model, the support of Marčenko-Pastur's **LSD** is $[(1-\sqrt{r})^2, (1+\sqrt{r})^2]$, with $\{0\}$ if $r > 1$. Geman [27] showed that $\lambda_{\max}(S_N) \rightarrow (1+\sqrt{r})^2$ almost surely under moment conditions on the entries. Later, Bai et al. [60, 5, 8] showed that $\lambda_{\max}(S_N)$ almost surely converges to a finite limit if and only if the fourth moment $\mathbb{E}|Z_{1,1}^{(1)}|^4$ of the entries is finite. Concerning the fluctuations of $\lambda_{\max}(S_N)$, they were first studied by Johansson [32] for standard Gaussian complex entries and by Johnstone [33] for standard Gaussian real entries. They both established that

$$\gamma_N n^{2/3} (\lambda_{\max}(S_N) - (1 + \sqrt{r_N})^2) \quad \text{where} \quad r_N = \frac{N}{n} \quad \text{and} \quad \gamma_N = \frac{r_N^{1/6}}{(1 + \sqrt{r_N})^{4/3}} \quad (1.8)$$

converges in distribution to **TW** distributions as $N, n \rightarrow \infty$, introduced in [53, 54] to describe the fluctuations of the largest eigenvalues of GUE and GOE random matrices.

For general sample covariance matrices (1.6), the condition that the spectral norm of Γ_N is uniformly bounded:

$$\sup_{N \geq 1} \|\Gamma_N\| = \sup_{N \geq 1} \lambda_{\max}(\Gamma_N) < \infty$$

implies that the **LSD** μ (defined by its Stieltjes transform g_μ which satisfies (1.7)) has a bounded support. In this case, El Karoui [25] and Lee and Schnelli [39] established Tracy-Widom type fluctuations of the largest eigenvalue in the complex and real Gaussian case respectively. By establishing a local law, Bao et al [11], and Knowles and Yin [37] extended

the fluctuations of the largest eigenvalue for general entries.

The case of spiked models has been addressed by Baik et al [9, 10] where some eigenvalues (the spikes) may separate from the bulk. In [9] where the so-called BBP phase transition phenomenon is described, Baik et al. study the case where Γ_N has exactly m non-unit eigenvalues $\ell_1 \geq \dots \geq \ell_m$. For complex Gaussian entries, they fully describe the fluctuations of $\lambda_{\max}(S_N)$ for different configurations of the ℓ_i 's. Assume for instance that ℓ_1 is simple (cf. the original paper for the general conditions) then (a) if $\ell_1 \leq 1 + \sqrt{r}$, $\lambda_{\max}(S_N)$ has asymptotically TW fluctuations at speed $n^{2/3}$; (b) if $\ell_1 > 1 + \sqrt{r}$, the sequence

$$\frac{\sqrt{n}}{\sqrt{\ell_1^2 - \ell_1^2 r_N / (\ell_1 - 1)^2}} \left(\lambda_{\max}(S_N) - \left(\ell_1 + \frac{\ell_1 r_N}{\ell_1 - 1} \right) \right) \quad (1.9)$$

is asymptotically Gaussian. In [10] Baik and Silverstein consider general entries and prove the strong convergence of the spiked eigenvalues; Bai and Yao [6] consider the spiked model with supercritical spikes (corresponding to the case (b) above) and general entries and establish Gaussian-type fluctuations for the spiked eigenvalues. Other results are, non exhaustively, in [12, 13, 23, 7].

To make a rough conclusion from these results, $\lambda_{\max}(S_N)$ does not in general approach the largest eigenvalue of $\lambda_{\max}(\Gamma_N)$. Moreover, if $\lambda_{\max}(S_N)$ converges to the bulk edge of the LSD of μ^{S_N} , then it often has Tracy-Widom fluctuation at the scale $n^{2/3}$. If $\lambda_{\max}(S_N)$ converges to a point outside the bulk, it often has Gaussian-type fluctuation at the scale $n^{1/2}$.

The previously mentioned results are limited to the case where $\lambda_{\max}(\Gamma_N)$ is uniformly bounded. There are however interesting cases where $\lambda_{\max}(\Gamma_N)$ goes to infinity, see for instance Forni et al. [26] in a context of econometrics.

Recently and mainly fostered by principal component analysis (PCA) in high dimension, there has been a renewed interest in the case where a small number of spiked eigenvalues of the population covariance matrix goes to infinity while the rest of the population eigenvalues remains bounded. Let us mention in growing generality Jung and Marron [35], Shen et al. [47], Wang and Fan [57], Cai et al. [20]. In the latter, a complete description of the various scenarios of the spikes and their multiplicity is considered, and the fluctuations for the first

non-spiked eigenvalue are described. In [38], Ledoit and Wolf consider a similar framework referred to as the ‘‘Arrow model’’.

We complement the general picture by considering population covariance matrices with unbounded limiting spectral distribution. Such a case arises in the context of long memory stationary processes and is not covered by the existing results. In the framework considered here, we are not in the case where a majority of the population eigenvalues remains bounded. In particular, the assumptions in [57, 20] fail to hold.

1.3 Organization and some notations

This thesis is organized as follows. In Section 1.4, we give an overview of all the results we have obtained concerning not only the asymptotics and the fluctuations of the largest eigenvalues of large covariance matrices but also the properties of the eigenvalues and their associated eigenvectors of Toeplitz symmetric deterministic matrices. Chapter II corresponds to the article *Unbounded largest eigenvalue of large sample covariance matrices* by F. Merlevède, J. Najim and P.T. In this article we study the largest eigenvalue of $S_N(\Gamma)$ defined by (1.6) when Γ_N has the block-diagonal form, or in the case where the entries of Z_N are i.i.d. standard Gaussian variables. Chapter III corresponds to the article with the same title *Joint CLT for top eigenvalues of empirical covariance matrices of long memory stationary processes* by P.T. In this article we remove the Gaussian assumption on Z_N and the block-diagonal assumption on Γ_N . But we still have to add a new technical assumption on Γ_N which does not cover all the autocovariance matrices T_N of LMSP. In Chapter IV we try to generalize the result of Chapter III for all the linear models $S_N(T)$ corresponding to LMSP, and we give a partial result.

In Chapters II and III, we keep the original notations of the articles. In other parts of this thesis we also use some of the general notations appearing in the two articles. We list these general notations below.

Given $x \in \mathbb{R}$, denote by $[x]$ the integer satisfying $[x] \leq x < [x] + 1$.

For two real numbers x, y , we denote

$$x \wedge y := \min(x, y) \quad \text{and} \quad x \vee y := \max(x, y).$$

Given two integers $a \leq b$, we write the set of integers between a and b by

$$[[a, b]] := \{x \in \mathbb{Z} : a \leq x \leq b\}.$$

For a function $f \in L^p$ or a vector $v \in l^p$, we denote the L^p or l^p norm by $\|\cdot\|_p$. For a matrix or a linear operator A , the norm of A induced by vector norm $\|\cdot\|_p$ is denoted by $\|A\|_p$, and we recall that $\|A\|_p := \sup_{\|v\|_p=1} \|Av\|_p$. We say that a function f or a vector v is “normalized” or “unit length” when $\|f\|_2 = 1$ or $\|v\|_2 = 1$. When functions or vectors are said to be “orthonormal”, they will be implicitly considered as elements of a Hilbert space. The inner product of two elements u, v of a Hilbert space is denoted by $\langle u, v \rangle$.

For an hermitian operator or matrix A , we denote its real eigenvalues by decreasing order as

$$\lambda_1(A) \geq \lambda_2(A) \geq \dots$$

We also denote the largest eigenvalue of A by $\lambda_{\max}(A)$. The spectrum of A is denoted by $\text{Spec}(A)$. If A is a $N \times N$ hermitian matrix, its Frobenius norm $\|A\|_F$ is defined by

$$\|A\|_F := \sqrt{\sum_{i,j} |A_{i,j}|^2},$$

and we recall that it satisfies

$$\|A\|_F = \sqrt{\sum_{i=1}^N |\lambda_i(A)|^2}.$$

For a matrix or a vector A , we use A^\top to denote the transposition of A , and A^* the conjugate transposition of A ; if A is a $N \times N$ square matrix with real eigenvalues, we use $\lambda_1(A) \geq \dots \geq \lambda_N(A)$ to denote its eigenvalues, and sometimes denote $\lambda_1(A) = \lambda_{\max}(A)$.

The **ESD** μ^A of A is defined as

$$\mu^A := \frac{1}{N} \sum_{k=1}^N \delta_{\lambda_k(A)},$$

where δ_λ is the Dirac measure at λ .

For a $N \times n$ matrix M and integers $a, b \in \{1, \dots, N\}$ and $c, d \in \{1, \dots, n\}$, the following notations are used to deal with submatrices of M :

$$M_{a:b,\cdot} = (M_{i,j})_{a \leq i \leq b, 1 \leq j \leq n}, \quad M_{\cdot,c:d} = (M_{i,j})_{1 \leq i \leq N, c \leq j \leq d}, \quad M_{a:b,c:d} = (M_{i,j})_{a \leq i \leq b, c \leq j \leq d}. \quad (1.10)$$

Similarly we denote $M_{-i,\cdot}$ the submatrix of M obtained by deleting the i th row. And $M_{-i,-j}$ denotes the submatrix of M obtained by deleting the i th row and the j th column. By convention, these subscripts have higher priority than the transposition or conjugate transposition, for example $M_{a:b,c:d}^* := (M_{a:b,c:d})^*$ is the conjugated transposition of the submatrix $M_{a:b,c:d}$.

Given two complex sequences x_n, y_n we denote

$$x_n \sim y_n \Leftrightarrow \lim_{n \rightarrow \infty} \left(\frac{x_n}{y_n} \right) = 1 \quad \text{and} \quad x_n \doteq y_n \Leftrightarrow \lim_{n \rightarrow \infty} (x_n - y_n) = 0. \quad (1.11)$$

The notations $x_n = o(1)$ and $x_n = O(1)$ respectively mean $\lim_{n \rightarrow \infty} x_n = 0$ and $\overline{\lim}_{n \rightarrow \infty} |x_n| < \infty$. These notations are also applicable to functions with continuous arguments. If X_n, X are random variables, the notation $X_n = o_P(1)$ means that $\lim_{n \rightarrow \infty} X_n = 0$ in probability. The notations $X_n \xrightarrow{\mathcal{D}} X$ and $X_n \xrightarrow{\mathcal{P}} X$ respectively denote convergence in distribution and in probability. If μ, μ_n are measures, we denote with a slight abuse of notation $\mu_n \xrightarrow{\mathcal{D}} \mu$ for the weak convergence of μ_n to μ .

For two probability measures P and Q on \mathbb{R}^m , we denote their Lévy-Prokhorov distance by $d_{\mathcal{LP}}(P, Q)$ which is defined by

$$d_{\mathcal{LP}}(P, Q) := \inf\{\varepsilon : P(A) \leq Q(A^\varepsilon) + \varepsilon, Q(A) \leq P(A^\varepsilon) + \varepsilon, \forall A \in \mathcal{B}(\mathbb{R}^m)\}, \quad (1.12)$$

where A^ε is defined by

$$A^\varepsilon := \{x \in \mathbb{R}^m : \exists y \in A, \text{ s.t. } \|x - y\| < \varepsilon\}.$$

It is well known that this distance metrizes the weak convergence. For two random variables $X, Y \in \mathbb{R}^m$ and two probability measures μ_X, μ_Y such that $X \sim \mu_X$ and $Y \sim \mu_Y$, we sometimes write $d_{\mathcal{L}\mathcal{P}}(X, Y)$, $d_{\mathcal{L}\mathcal{P}}(X, \mu_Y)$ or $d_{\mathcal{L}\mathcal{P}}(\mu_X, Y)$, all these notations denote $d_{\mathcal{L}\mathcal{P}}(\mu_X, \mu_Y)$.

Given a random variable Y or a sub-algebra \mathcal{G} , we denote by $\mathbb{E}_Y(X)$ and $\mathbb{E}_{\mathcal{G}}(X)$ the conditional expectation of the random variable X with respect to Y and to \mathcal{G} .

We denote by $\mathcal{N}_{\mathbb{R}}(m, \sigma^2)$ the real Gaussian distribution with mean m and variance σ^2 ; we refer to $\mathcal{N}_{\mathbb{R}}(0, 1)$ as the standard real Gaussian distribution. A complex random variable Z is distributed according to the standard complex Gaussian distribution if $Z = U + iV$ where U, V are independent, each with distribution $\mathcal{N}_{\mathbb{R}}(0, 1/2)$. In this case we denote $Z \sim \mathcal{N}_{\mathbb{C}}(0, 1)$. For a symmetric semidefinite positive matrix T , denote by $\mathcal{N}_{\mathbb{R}}(0, T)$ the distribution of a centered Gaussian vector with covariance matrix T . However we also use the notation \mathcal{N} to denote the Gaussian distribution in two circumstances: we may profit from the ambiguity and we denote the standard Gaussian distribution by $\mathcal{N}(0, 1)$, which can be either real or complex; we may also use $\mathcal{N}(0, T)$ to denote the Gaussian distribution in a case without ambiguity, for example if a vector V can only be real and we write $V \sim \mathcal{N}(0, T)$, then V is real Gaussian.

The cardinal of a set B is denoted by $\#B$. In the proofs we use C to denote a constant that may take different values from one place to another.

In the proofs we use C to denote a constant that may take different values from one place to another.

1.4 Main results

1.4.1 Asymptotic spectral properties of Toeplitz matrices

It is clear that the individual eigenvalues of $S_N(T)$ are tightly related to the corresponding eigenvalues of T_N . During our study, we discovered that sometimes they are also related

to the associated eigenvectors of T_N . So we are led to study the asymptotic behaviours of the largest eigenvalues of T_N as well as of their associated eigenvectors.

We choose the condition **C II** as the definition of long memory stationary process. On one hand this definition is relatively general because the conditions **C I**, **C III** imply **C II** with or without the quasi-monotone condition on the slowly varying function; on the other hand the condition **C II** allows us to describe completely the behaviours of eigenvalues and eigenvectors of T_N .

In this section we collect our main results on Toeplitz matrices. They appeared initially in our papers [42] and [52], which are the object of Chapter **II** and Chapter **III** respectively.

Let us consider T_N a Toeplitz matrix defined by

$$T_N := (\gamma(i - j))_{i,j=1}^N \tag{1.13}$$

with $\gamma : \mathbb{Z} \rightarrow \mathbb{R}$ a function of the form

$$\gamma(h) = (1 + |h|)^\rho L(|h|), \quad \forall h \in \mathbb{Z}, \tag{1.14}$$

where L is a slowly varying function at infinity and $\rho = 2d - 1 \in (-1, 0)$. Let $\mathcal{K}^{(\rho)}$ be the operator defined on $L^2(0, 1)$ by

$$(\mathcal{K}^{(\rho)} f)(x) = \int_0^1 |x - y|^\rho f(y) \, dy, \quad \text{for } f \in L^2(0, 1). \tag{1.15}$$

We will describe the eigenvalues and eigenvectors of T_N with help of the operator $\mathcal{K}^{(\rho)}$, so we first establish a theorem on the spectral properties of this operator.

Theorem I.3. *The operator $\mathcal{K}^{(\rho)}$ is compact and positive semidefinite. It has infinitely many positive eigenvalues. All its nonzero eigenvalues are simple, and the associated eigenfunctions are continuous on $[0, 1]$.*

We note that $\mathcal{K}^{(\rho)}$ is self-adjoint so for any nonzero eigenvalue λ , its algebraic multiplicity equals to its geometric multiplicity, which is defined as $\dim \ker(\lambda I - \mathcal{K}^{(\rho)})$. For more information about algebraic multiplicity, see [40]. So here we say that a nonzero eigenvalue

λ is simple to mean that

$$\dim \ker \left(\lambda I - \mathcal{K}^{(\rho)} \right) = 1.$$

In the next theorem, we describe the limiting behaviour of the eigenvalues $\lambda_j(T_N)$ and provide a quantitative description of their associated eigenvectors.

Theorem I.4. *For any $j \geq 1$, we have*

$$\lim_{N \rightarrow \infty} \frac{\lambda_j(T_N)}{N\gamma(N)} = \lambda_j(\mathcal{K}^{(\rho)}). \quad (1.16)$$

Moreover let f_j be the normalized eigenfunction of $\mathcal{K}^{(\rho)}$ associated with $\lambda_j(\mathcal{K}^{(\rho)})$, then for any N , we can choose a normalized eigenvector $u_j = (u_{1,j}, \dots, u_{N,j})^\top$ of T_N associated with $\lambda_j(T_N)$, such that

$$\lim_{N \rightarrow \infty} \sup_{1 \leq i \leq N} \left\{ \left| \sqrt{N} u_{i,j} - f_j \left(\frac{i}{N} \right) \right| \right\} = 0. \quad (1.17)$$

Combining Theorem I.3 and I.4 we get the following two important consequences. The first consequence is that for any fixed integer $m \geq 1$, the m largest eigenvalues of T_N tend to infinity at the same speed, and after normalization by $N\gamma(N)$, the m quantities $\lambda_1(T_N)/(N\gamma(N)), \dots, \lambda_m(T_N)/(N\gamma(N))$ converge to m distinct finite limits. In the sequel, such a property will be called *separate limiting property* of these eigenvalues. Precisely, a group of sequences $\left(\alpha_1^{(n)} \right)_n, \dots, \left(\alpha_m^{(n)} \right)_n$ admit the *separate limiting property* if there exists a normalizing sequence $\beta^{(n)} > 0$ such that the sequences

$$\frac{\alpha_1^{(n)}}{\beta^{(n)}}, \dots, \frac{\alpha_m^{(n)}}{\beta^{(n)}}$$

converge to m distinct finite limits.

The second consequence is the delocalization of the eigenvector u_j associated with $\lambda_j(T_N)$ for any fixed $j \geq 1$. Indeed by (1.17), for large N , we have

$$\|u_j\|_\infty \leq \frac{1 + \|f_j\|_\infty}{\sqrt{N}},$$

and by Theorem I.3 we have $\|f_j\|_\infty < \infty$. Thus we conclude that

$$\|u_j\|_\infty = O\left(\frac{1}{\sqrt{N}}\right) \xrightarrow{N \rightarrow \infty} 0.$$

The next proposition provides a comparison between the global behaviour of T_N 's eigenvalues ($\text{tr } T_N$ or $\text{tr } T_N^2$) and $\lambda_{\max}(T_N)$.

Proposition I.5. *Let T_N be defined as above.*

1. *If $\rho \in (-1/2, 0)$ or equivalently $d \in (1/4, 1/2)$, then T_N satisfies*

$$\lim_{N \rightarrow \infty} \frac{\text{tr } T_N}{\sqrt{N} \lambda_{\max}(T_N)} = 0. \quad (1.18)$$

2. *If $\rho \in (-3/4, -1/2]$ or equivalently $d \in (1/8, 1/4]$, then T_N satisfies*

$$\lim_{N \rightarrow \infty} \frac{\text{tr } T_N^2}{\sqrt{N} \lambda_{\max}^2(T_N)} = 0. \quad (1.19)$$

In general, we have

Proposition I.6. *Let T_N be defined as above, for any $\rho \in (-1, 0)$ or equivalently $d \in (0, 1/2)$, there exists an integer $q \geq 1$ such that*

$$\lim_{N \rightarrow \infty} \frac{\text{tr } T_N^q}{\sqrt{N} \lambda_{\max}^q(T_N)} = 0. \quad (1.20)$$

1.4.2 Asymptotics of largest eigenvalues of $S_N(\Gamma)$

Keeping in mind the spectral properties of the Toeplitz matrix T_N provided in Section 1.4.1, we consider now the generic covariance matrix $S_N(\Gamma)$. The assumptions made on Γ_N are directly inspired from T_N 's properties.

Let m be a fixed positive integer. We state our results on sample covariance matrix $S_N(\Gamma)$ under some of the following assumptions:

A1 (Model setting) Let $S_N(\Gamma)$ be $N \times N$ random matrices defined as

$$S_N(\Gamma) = \frac{1}{n} \Gamma_N^{\frac{1}{2}} Z_N Z_N^* \Gamma_N^{\frac{1}{2}}$$

where $Z_N = \left(Z_{i,j}^{(N)} \right)_{1 \leq i \leq N, 1 \leq j \leq n}$ are $N \times n$ matrices whose entries $Z_{i,j}^{(N)}$ are i.d. random variables for all i, j, N , and independent across i, j for each N , satisfying

$$\mathbb{E} Z_{i,j}^{(N)} = 0, \quad \mathbb{E} |Z_{i,j}^{(N)}|^2 = 1 \quad \text{and} \quad \mathbb{E} |Z_{i,j}^{(N)}|^4 < \infty,$$

and Γ_N are $N \times N$ positive semidefinite Hermitian deterministic matrices.

A2 (Asymptotic spectral structure of Γ_N) Given a sequence of $N \times N$ positive semidefinite deterministic matrices Γ_N , the empirical spectral distribution

$$\mu^{\tilde{\Gamma}_N} := \frac{1}{N} \sum_{k=1}^N \delta_{\frac{\lambda_k(\Gamma_N)}{\lambda_{\max}(\Gamma_N)}}$$

of the normalized matrix $\tilde{\Gamma}_N := \Gamma_N / \lambda_{\max}(\Gamma_N)$ converges weakly to δ_0 .

Example I.7. If $\mu^{\Gamma_N} \xrightarrow{\mathcal{D}} \nu$ with ν a non-compactly supported probability on \mathbb{R}^+ , then we have $\lambda_{\max}(\Gamma_N) \rightarrow \infty$ and one can prove that **A2** holds.

A3 The spectral norm $\lambda_1(\Gamma_N)$ is bounded away from zero, and there exists a constant $\kappa > 0$ such that

$$\lambda_m(\Gamma_N) \geq \kappa \lambda_1(\Gamma_N).$$

A4 (Subarray assumption on Z_N) For each N , $Z_N = Z_{1:N,1:n}$ is the top-left submatrix of an infinite matrix $Z = (Z_{i,j})_{i,j \geq 1}$, with $Z_{i,j}$ i.i.d random variables satisfying

$$\mathbb{E} Z_{i,j} = 0, \quad \mathbb{E} |Z_{i,j}|^2 = 1 \quad \text{and} \quad \mathbb{E} |Z_{i,j}|^4 < \infty.$$

Proposition I.8. Under **A1**, **A2** and **A3**, we have

$$\frac{\lambda_j(S_N(\Gamma))}{\lambda_j(\Gamma_N)} \xrightarrow[N, n \rightarrow \infty]{\mathcal{P}} 1.$$

If moreover either the random variables $Z_{i,j}^{(N)}$ are standard (real or complex) Gaussian or Assumption **A4** holds, then the above convergence holds almost surely.

The proof of Proposition **I.8** is in Appendix **A.2**. See also Proposition **II.3** and **II.11** for the result of the particular case where the largest eigenvalue is studied.

1.4.3 CLT for largest eigenvalues of $S_N(\Gamma)$

A5 (Multiple spectral gap conditions) For any $i = 1, \dots, m$, we assume that

$$\overline{\lim}_{N \rightarrow \infty} \frac{\lambda_{i+1}(\Gamma_N)}{\lambda_i(\Gamma_N)} < 1.$$

A6 (Block-diagonal structure of Γ_N) For all N , Γ_N has the block-diagonal form

$$\Gamma_N = \begin{pmatrix} \Xi_m & 0 \\ 0 & \Gamma_{N-m} \end{pmatrix},$$

where Ξ_m is a $m \times m$ diagonal matrix having eigenvalues $\lambda_1(\Gamma_N), \dots, \lambda_m(\Gamma_N)$ and Γ_{N-m} is a $(N-m) \times (N-m)$ matrix having eigenvalues $\lambda_{m+1}(\Gamma_N), \dots, \lambda_N(\Gamma_N)$.

A7

$$\lim_{N \rightarrow \infty} \frac{\text{tr} \Gamma_N}{\sqrt{N} \lambda_{\max}(\Gamma_N)} = 0 \quad \text{and} \quad \mathbb{E}|Z_{i,j}|^4 < \infty.$$

A8

$$\lim_{N \rightarrow \infty} \frac{\text{tr} \Gamma_N^2}{\sqrt{N} \lambda_{\max}^2(\Gamma_N)} = 0 \quad \text{and} \quad \mathbb{E}|Z_{i,j}|^8 < \infty.$$

We define

$$\Lambda_m(\Gamma_N) := \sqrt{n} \begin{pmatrix} \frac{\lambda_1(S_N(\Gamma))}{\lambda_1(\Gamma_N)} - 1 - \frac{1}{n} \sum_{k \neq 1} \frac{\lambda_k(\Gamma_N)}{\lambda_1(\Gamma_N) - \lambda_k(\Gamma_N)} \\ \vdots \\ \frac{\lambda_j(S_N(\Gamma))}{\lambda_j(\Gamma_N)} - 1 - \frac{1}{n} \sum_{k \neq j} \frac{\lambda_k(\Gamma_N)}{\lambda_j(\Gamma_N) - \lambda_k(\Gamma_N)} \\ \vdots \\ \frac{\lambda_m(S_N(\Gamma))}{\lambda_m(\Gamma_N)} - 1 - \frac{1}{n} \sum_{k \neq m} \frac{\lambda_k(\Gamma_N)}{\lambda_m(\Gamma_N) - \lambda_k(\Gamma_N)} \end{pmatrix}. \quad (1.21)$$

We define $\Lambda_m(T_N)$ similarly by replacing Γ_N with T_N and replacing $S_N(\Gamma)$ by $S_N(T)$.

The following theorem is a slight generalization of Theorem II.4 and II.12. By the discussion in Chapter II, Theorem II.4 is a corollary of Theorem II.12. Note that Theorem II.12 studies the CLT of the largest eigenvalue $\lambda_{\max}(S_N(\Gamma))$ in the case where Γ_N is block-diagonal, i.e. the particular case with $m = 1$ of Theorem I.9. With the same method it is easy to generalize Theorem II.12 and get the following result.

Theorem I.9. *We assume A1, A2, A3, A5, and A6. Then we have*

$$\Lambda_m(\Gamma_N) \xrightarrow[N, n \rightarrow \infty]{\mathcal{D}} \mathcal{N}(0, (\mathbb{E}|Z_{1,1}|^4 - 1)I_m). \quad (1.22)$$

From the above result, we can see for example, when $Z_{i,j}$ are i.i.d standard real Gaussian variables and Γ_N are real symmetric matrices, or when $Z_{i,j}$ are i.i.d standard complex Gaussian variables and Γ_N are hermitian matrices, then Γ_N can be diagonalized and we get

$$\Lambda_m(\Gamma_N) \xrightarrow[N, n \rightarrow \infty]{\mathcal{D}} \mathcal{N}(0, \sigma^2 I_m)$$

where $\sigma^2 = 2$ if $Z_{i,j}$ are real, and $\sigma^2 = 1$ if $Z_{i,j}$ are complex.

But if the entries $Z_{i,j}$ are not Gaussian, and if Γ_N 's do not fulfill the assumption A6, the above theorem fails to apply. In order to generalize the above result, and mainly to relax the block-diagonal condition on Γ_N in non-Gaussian case, we developpe the following theorem.

Theorem I.10. *We assume A1, A3, A5, and either A7 or A8. Then we have*

$$d_{\mathcal{LP}}(\Lambda_m(\Gamma_N), \mathcal{N}(0, \Sigma_m^{(N)})) \xrightarrow[N, n \rightarrow \infty]{} 0, \quad (1.23)$$

where $d_{\mathcal{LP}}$ denotes the Lévy-Prokhorov distance, $\Sigma_m^{(N)} = I_m + (\sigma_{i,j}^{(N)})_{i,j=1}^m$ with

$$\sigma_{i,j}^{(N)} = (\mathbb{E}|Z_{1,1}|^4 - |\mathbb{E}Z_{1,1}^2|^2 - 2) \sum_{k=1}^N |u_{k,i}|^2 |u_{k,j}|^2 + |\mathbb{E}Z_{1,1}^2|^2 \left| \sum_{k=1}^N u_{k,i} u_{k,j} \right|^2. \quad (1.24)$$

and $u_j := (u_{j,1}, \dots, u_{j,N})^\top$ is a normalized eigenvector associated with $\lambda_j(\Gamma_N)$.

This theorem corresponds to Theorem III.4. Note that under the assumptions of Theorem I.10, if moreover Γ_N satisfies the block-diagonal condition A6, we recover the result of Theorem I.9. In another case, if Γ_N are real symmetric matrices and the eigenvectors associated to the m largest eigenvalues of Γ_N are delocalized, i.e.

$$\|u_j\|_\infty \leq \frac{K}{\sqrt{N}}$$

for some constant $K > 0$ and for $j = 1, \dots, m$, then

$$\Sigma_m^{(N)} \xrightarrow[N, n \rightarrow \infty]{} (1 + |\mathbb{E}Z_{i,j}^2|)I_m,$$

and thus we get

$$\Lambda_m(\Gamma_N) \xrightarrow[N, n \rightarrow \infty]{\mathcal{D}} \mathcal{N}(0, (1 + |\mathbb{E}Z_{i,j}^2|)I_m).$$

Moreover if $Z_{i,j}$ are real, or if $Z_{i,j}$ are complex with $\mathbb{E}Z_{i,j}^2 = 0$, then $\Lambda_m(\Gamma_N)$ fluctuates in the same way as the Gaussian case.

1.4.4 Applications to long memory stationary processes

In this section, we give the corollaries of the previous results in case where Γ_N is a Toeplitz matrix as defined in (1.13).

Let $S_N(T)$ be defined as (1.3) where T_N is defined as (1.13) and Z_N is a $N \times n$ matrix with i.i.d entries $Z_{i,j}^{(N)}$ such that

$$\mathbb{E}Z_{i,j}^{(N)} = 0, \quad \mathbb{E}|Z_{i,j}^{(N)}|^2 = 1, \quad \text{and} \quad \mathbb{E}|Z_{i,j}^{(N)}|^4 < \infty.$$

Combining the above results we obtain the following theorem.

Theorem I.11. *As $N, n \rightarrow \infty$, for any integer $m \geq 1$, for any $j = 1, \dots, m$, we have*

$$\lim_{N \rightarrow \infty} \frac{\lambda_j(S_N(T))}{\lambda_j(T_N)} = 1$$

in probability. Moreover, if one of the following conditions holds

1. The parameter ρ belongs to $(-1/2, 0)$ or equivalently $d \in (1/4, 1/2)$, and $\mathbb{E}|Z_{i,j}^{(N)}|^4 < \infty$;
2. The parameter ρ belongs to $(-3/4, -1/2]$ or equivalently $d \in (1/8, 1/4]$, and $\mathbb{E}|Z_{i,j}^{(N)}|^8 < \infty$,

then we have

$$\Lambda_m(T_N) \xrightarrow[N, n \rightarrow \infty]{\mathcal{D}} \mathcal{N}(0, (1 + |\mathbb{E}(Z_{i,j}^{(N)})^2|^2)I_m).$$

When the long memory stationary process $(\mathcal{X}_t)_{t \in \mathbb{Z}}$ is Gaussian (real or circularly symmetric complex, which will be defined below), as the sample covariance matrix Q_N defined by (1.2) has the linear representation $S_N(T)$, we can apply the above results on Q_N .

Definition I.12. We say that a complex Gaussian process $(\mathcal{X}_t)_{t \in \mathbb{Z}}$ is circularly symmetric if for any $N \in \mathbb{N}$ the Gaussian vector $\mathcal{X}_{1:N} := (\mathcal{X}_1, \dots, \mathcal{X}_N)^\top$ is circularly symmetric, i.e. for any $\phi \in \mathbb{R}$, the vector $e^{i\phi} \mathcal{X}_{1:N}$ has the same distribution as $\mathcal{X}_{1:N}$.

Corollary I.13. Let $(\mathcal{X}_t)_{t \in \mathbb{Z}}$ be a Gaussian stationary process either real or circularly complex and satisfying satisfying **C II** with long memory parameter $d \in (0, 1/2)$. Let Q_N be defined by (1.2), and T_N be the autocovariance matrix. Then for any $j \geq 1$, we have

$$\lim_{N, n \rightarrow \infty} \frac{\lambda_j(Q_N)}{\lambda_j(T_N)} = 1 \quad a.s.$$

and for any $m \geq 1$, we have

$$\sqrt{n} \begin{pmatrix} \frac{\lambda_1(Q_N)}{\lambda_1(T_N)} - 1 - \frac{1}{n} \sum_{k \neq 1} \frac{\lambda_k(T_N)}{\lambda_1(T_N) - \lambda_k(T_N)} \\ \vdots \\ \frac{\lambda_m(Q_N)}{\lambda_m(T_N)} - 1 - \frac{1}{n} \sum_{k \neq m} \frac{\lambda_k(T_N)}{\lambda_m(T_N) - \lambda_k(T_N)} \end{pmatrix} \xrightarrow[N, n \rightarrow \infty]{\mathcal{D}} \mathcal{N}(0, \sigma^2 I_m)$$

where $\sigma^2 = 1$ if the process is complex gaussian and circularly symmetric, and $\sigma^2 = 2$ if the process is real gaussian.

1.4.5 Extension of the results for large covariance matrices close to short memory

For the model $S_N(T)$ with non-gaussian entries $Z_{i,j}^{(N)}$, when the parameter d approaches to the threshold of short memory, i.e. $d \in (0, 1/8]$, or equivalently $\rho \in (-1, -3/4]$, the fluctuations of largest eigenvalues are not yet solved. However we have the following partial result: if we can prove that the eigenvectors of T_N are uniformly delocalized, then with some other assumptions, the result of Theorem I.11 still holds. We denote $u_j := (u_{j,k})_{1 \leq k \leq N}$ the normalized eigenvector associated with $\lambda_j(\Gamma_N)$.

Definition I.14. We say that two random variables X, Y match to order k if for any integers $\alpha, \beta \geq 0$ s.t $\alpha + \beta \leq k$, we have

$$\mathbb{E}X^\alpha \overline{X}^\beta = \mathbb{E}Y^\alpha \overline{Y}^\beta.$$

A9 The entries $Z_{i,j}^{(N)}$ have all finite moments, i.e.

$$\mathbb{E}|Z_{i,j}^{(N)}|^k < \infty$$

for any $k \geq 1$. Moreover we assume that $Z_{i,j}^{(N)}$ matches with a standard real or complex Gaussian variable G up to order 3.

A10 There exists an integer $q \geq 1$ such that

$$\lim_{N \rightarrow \infty} \frac{\text{tr } \Gamma_N^q}{\sqrt{N} \lambda_{\max}^q(\Gamma_N)} = 0.$$

A11 We assume that the eigenvectors of Γ_N are *uniformly delocalized*, i.e. there exists a constant K independent of j and N , such that

$$\|u_j\|_\infty \leq \frac{K}{\sqrt{N}}.$$

Theorem I.15. *Under **A1**, **A3**, **A5**, and in addition **A9-A11**, then the fluctuations*

$$\sqrt{n} \begin{pmatrix} \frac{\lambda_1(S_N(\Gamma))}{\lambda_1(\Gamma_N)} - 1 - \frac{1}{n} \sum_{k \neq 1} \frac{\lambda_k(\Gamma_N)}{\lambda_1(\Gamma_N) - \lambda_k(\Gamma_N)} \\ \vdots \\ \frac{\lambda_m(S_N(\Gamma))}{\lambda_m(\Gamma_N)} - 1 - \frac{1}{n} \sum_{k \neq m} \frac{\lambda_k(\Gamma_N)}{\lambda_m(\Gamma_N) - \lambda_k(\Gamma_N)} \end{pmatrix} \xrightarrow[N, n \rightarrow \infty]{\mathcal{D}} \mathcal{N}(0, (1 + |\mathbb{E}Z_{1,1}^{(N)2}|^2)I_m)$$

hold.

The proof of this theorem will be given in Chapter **IV**. It is very technical and use a combination of the Lindeberg's method together with a fine analysis of some moments with the help of techniques coming from graph theory and combinatorics.

If Γ_N is the Toeplitz matrix T_N as defined in (1.13) and (1.14) with long memory parameter $d \in (0, 1/8]$ or $\rho \in [-3/4, 0)$, in addition to **A2** and **A3**, T_N satisfies also (A10) (see Proposition I.6). However, to know if T_N does satisfy **A11** remains an open question. From simulations results, we conjecture that it may be true.

Conjecture I.16. *The result of Theorem I.11 holds for all $S_N(T)$ with parameter $d \in (0, 1/2)$ or equivalently $\rho \in (-1, 0)$.*

CHAPTER II

Unbounded largest eigenvalue of large sample covariance matrices: Asymptotics, fluctuations and applications

This chapter corresponds to the article with the same title *Unbounded largest eigenvalue of large sample covariance matrices: Asymptotics, fluctuations and applications* by F. Merlevède, J. Najim and P.T. This article has been submitted to *Linear Algebra and its Applications for publication*. See [42].

2.1 Abstract

Given a large sample covariance matrix $S_N = \frac{1}{n} \Gamma_N^{1/2} Z_N Z_N^* \Gamma_N^{1/2}$, where Z_N is a $N \times n$ matrix with i.i.d. centered entries, and Γ_N is a $N \times N$ deterministic Hermitian positive semidefinite matrix, we study the location and fluctuations of $\lambda_{\max}(S_N)$, the largest eigenvalue of S_N as $N, n \rightarrow \infty$ and $Nn^{-1} \rightarrow r \in (0, \infty)$ in the case where the empirical distribution μ^{Γ_N} of eigenvalues of Γ_N is tight (in N) and $\lambda_{\max}(\Gamma_N)$ goes to $+\infty$. These conditions are in particular met when μ^{Γ_N} weakly converges to a probability measure with unbounded support on \mathbb{R}^+ .

We prove that asymptotically $\lambda_{\max}(S_N) \sim \lambda_{\max}(\Gamma_N)$. Moreover when the Γ_N 's are block-diagonal, and the following *spectral gap condition* is assumed:

$$\limsup_{N \rightarrow \infty} \frac{\lambda_2(\Gamma_N)}{\lambda_{\max}(\Gamma_N)} < 1,$$

where $\lambda_2(\Gamma_N)$ is the second largest eigenvalue of Γ_N , we prove Gaussian fluctuations for

$\lambda_{\max}(S_N)/\lambda_{\max}(\Gamma_N)$ at the scale \sqrt{n} .

In the particular case where Z_N has i.i.d. Gaussian entries and Γ_N is the $N \times N$ autocovariance matrix of a long memory Gaussian stationary process $(\mathcal{X}_t)_{t \in \mathbb{Z}}$, the columns of $\Gamma_N^{1/2} Z_N$ can be considered as n i.i.d. samples of the random vector $(\mathcal{X}_1, \dots, \mathcal{X}_N)^\top$. We then prove that Γ_N is similar to a diagonal matrix which satisfies all the required assumptions of our theorems, hence our results apply to this case.

2.2 Introduction

The model.

In this paper we consider the following model of sample covariance matrix

$$S_N = \frac{1}{n} \Gamma_N^{1/2} Z_N Z_N^* \Gamma_N^{1/2} \quad (2.1)$$

where $Z_N = (Z_{i,j}^{(N)})$ is a $N \times n$ matrix whose entries $Z_{i,j}^{(N)}$ are real or complex random variables identically distributed (i.d.) for all i, j, N and independent across i, j for each N , satisfying

$$\mathbb{E} Z_{i,j}^{(N)} = 0, \quad \mathbb{E} |Z_{i,j}^{(N)}|^2 = 1 \quad \text{and} \quad \mathbb{E} |Z_{i,j}^{(N)}|^4 < \infty, \quad (2.2)$$

and Γ_N is a $N \times N$ deterministic Hermitian positive semidefinite matrix with eigenvalues

$$0 \leq \lambda_N(\Gamma_N) \leq \dots \leq \lambda_1(\Gamma_N) := \lambda_{\max}(\Gamma_N).$$

We consider the case where $\lambda_{\max}(\Gamma_N)$ goes to infinity as $N \rightarrow \infty$ while the empirical spectral distribution (ESD) μ^{Γ_N} associated with Γ_N ,

$$\mu^{\Gamma_N} := \frac{1}{N} \sum_{k=1}^N \delta_{\lambda_k(\Gamma_N)},$$

forms a tight sequence of probabilities on $\mathbb{R}^+ := [0, \infty)$. These conditions encompass the important case where μ^{Γ_N} converges to a limiting distribution with unbounded support on \mathbb{R}^+ .

In this context, our aim is to study the location and fluctuations of the largest eigenvalue $\lambda_{\max}(S_N)$ in the asymptotic regime where

$$N, n \rightarrow \infty \quad \text{and} \quad \frac{N}{n} \rightarrow r \in (0, \infty). \quad (2.3)$$

The regime (2.3) will be simply referred to as $N, n \rightarrow \infty$ in the sequel.

The model S_N defined in (2.1) is a classical model of sample covariance matrices in the random matrix theory, and its spectral properties have been intensively studied in the regime (2.3) in the last several decades.

At a global scale, the LSD of the ESD $\mu^{S_N} = N^{-1} \sum_{k=1}^N \delta_{\lambda_k(S_N)}$ has been described in the groundbreaking paper by Marčenko and Pastur [41]. In the important case where $S_N = \frac{1}{n} Z_N Z_N^*$, sometimes referred to as the *white noise* model, the limiting spectral distribution of μ^{S_N} is known as Marčenko-Pastur distribution and admits the following closed-form expression

$$\mathbb{P}_{MP}(d\lambda) := (1 - r^{-1})_+ \delta_0(d\lambda) + \frac{\sqrt{[(\lambda^+ - \lambda)(\lambda - \lambda^-)]_+}}{2\pi r \lambda} d\lambda, \quad \lambda^\pm = (1 \pm \sqrt{r})^2,$$

where $x_+ := \max(x, 0)$. Later, this result was improved by many others, see for instance [56, 34, 59, 49, 48]. In [48], Silverstein proved that for the model S_N defined in (2.1), if $\mu^{\Gamma N}$ weakly converges to a certain probability ν supported on \mathbb{R}^+ (not necessarily with compact support), then almost surely, the ESD μ^{S_N} weakly converges to a deterministic distribution μ , whose Stieltjes transform g_μ is the unique solution with positive imaginary part of the equation

$$g_\mu(z) = \int \frac{1}{s(1 - r - rzg_\mu(z)) - z} d\nu(s) \in \mathbb{C}^+, \quad \forall z \in \mathbb{C}^+. \quad (2.4)$$

Central limit theorems have also been established for linear spectral statistics $\sum_{i=1}^N f(\lambda_i(S_N))$, see for instance [34, 32, 3, 44].

At a local scale, the convergence and fluctuations of individual eigenvalues have been studied, with a special emphasis on the eigenvalues located near each edge of the connected components (bulk) of the LSD of S_N . The spiked eigenvalues, that is those which stay away from the bulk of the LSD, have also attracted a lot of attention.

For the white noise model, the support of Marčenko-Pastur's LSD is $[(1-\sqrt{r})^2, (1+\sqrt{r})^2]$, with $\{0\}$ if $r > 1$. Geman [27] showed that $\lambda_{\max}(S_N) \rightarrow (1 + \sqrt{r})^2$ almost surely under moment conditions on the entries. Later, Bai et al. [60, 5, 8] showed that $\lambda_{\max}(S_N)$ almost surely converges to a finite limit if and only if the fourth moment $\mathbb{E}|Z_{1,1}^{(1)}|^4$ of the entries is finite. Concerning the fluctuations of $\lambda_{\max}(S_N)$, they were first studied by Johansson [32] for standard Gaussian complex entries and by Johnstone [33] for standard Gaussian real entries. They both established that

$$\gamma_N n^{2/3} (\lambda_{\max}(S_N) - (1 + \sqrt{r_N})^2) \quad \text{where} \quad r_N = \frac{N}{n} \quad \text{and} \quad \gamma_N = \frac{r_N^{1/6}}{(1 + \sqrt{r_N})^{4/3}} \quad (2.5)$$

converges in distribution to Tracy-Widom (TW) distributions as $N, n \rightarrow \infty$, introduced in [53, 54], to describe the fluctuations of the largest eigenvalues of GUE and GOE random matrices.

For general sample covariance matrices (2.1), the condition that the spectral norm of Γ_N is uniformly bounded:

$$\sup_{N \geq 1} \|\Gamma_N\| = \sup_{N \geq 1} \lambda_{\max}(\Gamma_N) < \infty$$

implies that the LSD μ (defined by its Stieltjes transform g_μ which satisfies (2.4)) has a bounded support. In this case, El Karoui [25] and Lee and Schnelli [39] established Tracy-Widom type fluctuations of the largest eigenvalue in the complex and real Gaussian case respectively. By establishing a local law, Bao et al [11], and Knowles and Yin [37] extended the fluctuations of the largest eigenvalue for general entries.

The case of spiked models has been addressed by Baik et al [9, 10] where some eigenvalues (the spikes) may separate from the bulk. In [9] where the so-called BBP phase transition phenomenon is described, Baik et al. study the case where Γ_N has exactly m non-unit eigenvalues $\ell_1 \geq \dots \geq \ell_m$. For complex Gaussian entries, they fully describe the fluctuations of $\lambda_{\max}(S_N)$ for different configurations of the ℓ_i 's. Assume for instance that ℓ_1 is simple (cf. the original paper for the general conditions) then (a) if $\ell_1 \leq 1 + \sqrt{r}$, $\lambda_{\max}(S_N)$ has

asymptotically TW fluctuations at speed $n^{2/3}$; (b) if $\ell_1 > 1 + \sqrt{r}$, the sequence

$$\frac{\sqrt{n}}{\sqrt{\ell_1^2 - \ell_1^2 r_N / (\ell_1 - 1)^2}} \left(\lambda_{\max}(S_N) - \left(\ell_1 + \frac{\ell_1 r_N}{\ell_1 - 1} \right) \right) \quad (2.6)$$

is asymptotically Gaussian. In [10] Baik and Silverstein consider general entries and prove the strong convergence of the spiked eigenvalues; Bai and Yao [6] consider the spiked model with supercritical spikes (corresponding to the case (b) above) and general entries and establish Gaussian-type fluctuations for the spiked eigenvalues. Other results are, non exhaustively, [12, 13, 23, 7].

To make a rough conclusion from these results, $\lambda_{\max}(S_N)$ does not in general approach the largest eigenvalue of $\lambda_{\max}(\Gamma_N)$. Moreover, if $\lambda_{\max}(S_N)$ converges to the bulk edge of the LSD of μ^{S_N} , then it often has Tracy-Widom fluctuation at the scale $n^{2/3}$. If $\lambda_{\max}(S_N)$ converges to a point outside the bulk, it often has Gaussian-type fluctuation at the scale $n^{1/2}$.

The previously mentioned results are limited to the case where $\lambda_{\max}(\Gamma_N)$ is uniformly bounded. There are however interesting cases where $\lambda_{\max}(\Gamma_N)$ goes to infinity, see for instance Forni et al. [26] in a context of econometrics.

Recently and mainly fostered by principal component analysis (PCA) in high dimension, there has been a renewed interest in the case where a small number of spiked eigenvalues of the population covariance matrix goes to infinity while the rest of the population eigenvalues remains bounded. Let us mention in growing generality Jung and Marron [35], Shen et al. [47], Wang and Fan [57], Cai et al. [20]. In the latter, a complete description of the various scenarios of the spikes and their multiplicity is considered, and the first non-spiked eigenvalue's fluctuations is established. In [38], Ledoit and Wolf consider a similar framework referred to as the "Arrow model".

In this article, we complement the general picture by considering population covariance matrices with unbounded limiting spectral distribution. Such a case arises in the context of long memory stationary processes and is not covered by the existing results. In the framework considered here, we are not in the case where a majority of the population eigenvalues remains bounded. In particular, the assumptions in [57, 20] fail to hold.

Description of the main results.

Let S_N be defined in (2.1) and assume that (μ^{Γ_N}) is tight with $\lim_{N \rightarrow \infty} \lambda_{\max}(\Gamma_N) = \infty$, then we establish in Proposition II.3 that

$$\frac{\lambda_{\max}(S_N)}{\lambda_{\max}(\Gamma_N)} \xrightarrow[N, n \rightarrow \infty]{} 1 \quad (2.7)$$

in probability. This convergence can be improved to an almost sure convergence if either the $Z_{i,j}^{(N)}$'s are standard (real or complex) gaussian, or stem from the top left corner of an infinite array $(Z_{i,j}, i, j \in \mathbb{N})$ of i.i.d. random variables.

In order to describe the fluctuations of $\lambda_{\max}(S_N)$, we assume in addition that (Γ_N) satisfies the following *spectral gap condition*

$$\overline{\lim}_{N \rightarrow \infty} \frac{\lambda_2(\Gamma_N)}{\lambda_{\max}(\Gamma_N)} < 1, \quad (2.8)$$

where $\lambda_2(\Gamma_N)$ is the second largest eigenvalue of Γ_N , and that either the $Z_{i,j}^{(N)}$'s are standard Gaussian or the Γ_N 's have a *block-diagonal structure*

$$\Gamma_N = \begin{pmatrix} \lambda_{\max}(\Gamma_N) & 0 \\ 0 & \Gamma_{N-1} \end{pmatrix}. \quad (2.9)$$

In this case, the following fluctuation result, stated in Theorem II.4, holds:

$$\sqrt{n} \left(\frac{\lambda_{\max}(S_N)}{\lambda_{\max}(\Gamma_N)} - 1 - \frac{1}{n} \sum_{k=2}^N \frac{\lambda_k(\Gamma_N)}{\lambda_{\max}(\Gamma_N) - \lambda_k(\Gamma_N)} \right) \xrightarrow[N, n \rightarrow \infty]{\mathcal{D}} \mathcal{N}(0, \sigma^2) \quad (2.10)$$

where “ $\xrightarrow{\mathcal{D}}$ ” denotes the convergence in distribution, $\sigma^2 = \mathbb{E}|Z_{1,1}|^4 - 1$ and \mathcal{N} stands for the real Gaussian distribution.

These results are then applied to long memory stationary processes.

Long memory stationary process.

A process $(\mathcal{X}_t)_{t \in \mathbb{Z}}$ is (second order) stationary if the following conditions are satisfied:

$$\mathbb{E}|\mathcal{X}_t|^2 < \infty, \quad \mathbb{E}\mathcal{X}_t = \mathbb{E}\mathcal{X}_0 \quad \text{and} \quad \text{Cov}(\mathcal{X}_{t+h}, \mathcal{X}_t) = \text{Cov}(\mathcal{X}_h, \mathcal{X}_0) = \gamma(h) \quad \forall t, h \in \mathbb{Z}$$

where $\text{Cov}(\mathcal{X}_{t+h}, \mathcal{X}_t) = \mathbb{E}(\mathcal{X}_{t+h} - \mathbb{E}\mathcal{X}_{t+h})(\overline{\mathcal{X}_t - \mathbb{E}\mathcal{X}_t})$ and $\gamma : \mathbb{Z} \rightarrow \mathbb{C}$ is some positive definite function, usually called the autocovariance function of the process. Note that $\gamma(0)$ is positive and $\gamma(-h) = \overline{\gamma(h)}$ for all $h \in \mathbb{Z}$. By stationarity, the covariance matrices $T_N(\gamma)$ of the process

$$T_N(\gamma) := \text{Cov} \begin{pmatrix} \mathcal{X}_{t+1} \\ \vdots \\ \mathcal{X}_{t+N} \end{pmatrix} = \begin{pmatrix} \gamma(0) & \gamma(-1) & \dots & \gamma(-N+1) \\ \gamma(1) & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \gamma(-1) \\ \gamma(N-1) & \dots & \gamma(1) & \gamma(0) \end{pmatrix} \quad (2.11)$$

are positive semidefinite Hermitian Toeplitz matrices.

By Herglotz's Theorem, there exists a finite positive measure α on $(-\pi, \pi]$, the *symbol* of $T_N(\gamma)$, whose Fourier coefficients are exactly $\gamma(h)$, i.e.

$$\gamma(h) = \frac{1}{2\pi} \int_{(-\pi, \pi]} e^{-ihx} d\alpha(x), \quad \forall h \in \mathbb{Z}.$$

Depending on the context, we may write $T_N(\gamma)$ or $T_N[\alpha]$.

Tyrtshnikov and Zamarashkin generalized in [55] a result of Szegő and proved that the following equality holds

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{k=1}^N \varphi(\lambda_k(T_N(\gamma))) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \varphi(f_\alpha(x)) dx, \quad (2.12)$$

where $\varphi : \mathbb{R} \mapsto \mathbb{R}$ is continuous with compact support and $f_\alpha \in L^1(-\pi, \pi)$ is the density of the absolutely continuous part of α with respect to the Lebesgue measure dx on $(-\pi, \pi]$, called the *spectral density* of $T_N(\gamma)$. The equality (2.12) can be interpreted as the vague convergence of probability measures $\mu^{T_N(\gamma)}$ to the measure ν defined by the integral formula

$$\int \varphi \, d\nu = \frac{1}{2\pi} \int_{-\pi}^{\pi} \varphi(f_\alpha(x)) \, dx \quad \forall \varphi \in C_b, \quad (2.13)$$

where C_b denotes the space of all bounded continuous functions. The measure ν being a probability, the sequence $\mu^{T_N(\gamma)}$ is tight, and the vague convergence coincides with the weak convergence.

The process is usually said to have *short memory* or *short range dependence* if $\sum_{h \in \mathbb{Z}} |\gamma(h)| < \infty$. Otherwise, if

$$\sum_{h \in \mathbb{Z}} |\gamma(h)| = \infty,$$

the process (\mathcal{X}_t) has *long memory* or *long range dependence*¹.

In this article we require that the autocovariance function γ of a long memory stationary process satisfies

$$\gamma(h) = \frac{L(h)}{(1 + |h|)^{1-2d}}, \quad \forall h \in \mathbb{Z} \quad (2.14)$$

for some $d \in (0, 1/2)$ and a function $L : \mathbb{R} \rightarrow \mathbb{R}$ slowly varying at ∞ , that is a function such that $L(y) > 0$ for $|y|$ large enough such that

$$\lim_{y \rightarrow \infty} \frac{L(xy)}{L(y)} = 1 \quad \forall x > 0.$$

In this case γ is real and even and L is an even function as well. Matrix $T_N(\gamma)$ is real symmetric and (\mathcal{X}_t) is a long memory process. In addition, $\lambda_{\max}(\Gamma_N) \xrightarrow{N \rightarrow \infty} \infty$, see for instance Theorem II.8.

The largest eigenvalue associated to a long memory stationary Gaussian processes.

Given a centered stationary process $(\mathcal{X}_t)_{t \in \mathbb{Z}}$ with autocovariance function defined by (2.14), one can study the spectral properties of the sample covariance matrix

$$Q_N := \frac{1}{n} X_N X_N^* = \frac{1}{n} \sum_{j=1}^n X_{\cdot, j} X_{\cdot, j}^* \quad (2.15)$$

¹There are several definitions of long range dependence, all strongly related but not always equivalent, see for instance [45, Chapter 2].

where X_N is a $N \times n$ random matrix whose columns $(X_{\cdot,j}, 1 \leq j \leq n)$ are i.i.d copies of the random vector $\mathcal{X}_{1:N} = (\mathcal{X}_1, \dots, \mathcal{X}_N)^\top$.

Let $T_N(\gamma)$ be the covariance matrix of (\mathcal{X}_t) , it has been recalled that $\mu^{T_N(\gamma)}$ weakly converges. Since the process is Gaussian, Q_N can be written in the form of S_N in (2.1) with $\Gamma_N = T_N(\gamma)$ and the **ESD** μ^{Q_N} weakly converges with probability one to a deterministic probability measure μ by [48, Theorem 1.1].

In order to study the behavior of $\lambda_{\max}(Q_N)$ and to apply the results already presented, note that the process being gaussian, the matrix model (2.1) has the same spectral properties as a model where Γ_N is replaced by the diagonal matrix obtained with Γ_N 's eigenvalues. In particular, the block-diagonal structure condition (2.9) is automatically satisfied. It remains to verify the spectral gap condition (2.8) and that $\lambda_{\max}(T_N(\gamma))$ goes to infinity. In Theorem II.8, we describe the asymptotic behaviour of the individual eigenvalues $\lambda_k(T_N(\gamma))$ and prove that for any $k \geq 1$, there exist nonnegative numbers a_k with $a_1 > 0$ such that

$$\lambda_k(T_N(\gamma)) \sim a_k N \gamma(N) \quad \text{and} \quad \lim_{N \rightarrow \infty} \frac{\lambda_2(T_N(\gamma))}{\lambda_{\max}(T_N(\gamma))} = \frac{a_2}{a_1} < 1,$$

hence the spectral gap condition (2.8) holds. Moreover, standard properties of slowly varying functions [14, Prop. 1.3.6(v)] yield that $N \gamma(N) \rightarrow \infty$ hence $\lambda_{\max}(T_N(\gamma)) \rightarrow \infty$. As a corollary, we obtain the asymptotics and fluctuations of the largest eigenvalue $\lambda_{\max}(Q_N)$ for Gaussian long memory stationary processes with autocovariance function defined in (2.14).

We now point out two references of interest: In the (non-Gaussian) case where the symbol α is absolutely continuous with respect to the Lebesgue measure and under additional regularity conditions on (\mathcal{X}_t) , Merlevde and Peligrad [43] have established the convergence of the **ESD** μ^{Q_N} toward a certain deterministic probability distribution. In a context of a stationary Gaussian field, Chakrabarty et al. [22] studied large random matrices associated to long range dependent processes.

Organization.

Our paper is organized as follows. In Section 2.3 we state the assumptions and main results of the article: Proposition II.3 and Theorem II.4 are devoted to the limiting behaviour

and fluctuations of $\lambda_{\max}(S_N)$; the spectral gap condition for a Toeplitz matrix Γ_N is studied in Theorem II.8; finally Corollary II.10 builds upon the previous results and describes the behaviour and fluctuations of covariance matrices based on samples of stationary long memory Gaussian processes. In Section 2.4, we provide examples, numerical simulations and mention some open questions. Section 2.5 and Section 2.6 are dedicated to the proofs of the main theorems.

Acknowledgement.

The authors would like to thank Walid Hachem for useful discussions.

2.3 Notations and main theorems

2.3.1 Notations and assumptions

Notations.

Given $x \in \mathbb{R}$, denote by $\lfloor x \rfloor$ the integer satisfying $\lfloor x \rfloor \leq x < \lfloor x \rfloor + 1$. For vectors u, v in \mathbb{R}^N or \mathbb{C}^N , $\langle u, v \rangle = \sum_{i=1}^N u_i \bar{v}_i$ denotes the scalar product and $\|u\|$ the Euclidean norm of u .

For a matrix or a vector A , we use A^\top to denote the transposition of A , and A^* the conjugate transposition of A ; if A is a $N \times N$ square matrix with real eigenvalues, we use $\lambda_1(A) \geq \dots \geq \lambda_N(A)$ to denote its eigenvalues, and sometimes denote $\lambda_1(A) = \lambda_{\max}(A)$. The **ESD** μ^A of A is defined as

$$\mu^A := \frac{1}{N} \sum_{k=1}^N \delta_{\lambda_k(A)},$$

where δ_λ is the Dirac measure at λ . For a $N \times n$ matrix M and integers $a, b \in \{1, \dots, N\}$ and $c, d \in \{1, \dots, n\}$, the following notations are used to deal with submatrices of M :

$$M_{a:b,\cdot} = (M_{i,j})_{a \leq i \leq b, 1 \leq j \leq n}, \quad M_{\cdot,c:d} = (M_{i,j})_{1 \leq i \leq N, c \leq j \leq d}, \quad M_{a:b,c:d} = (M_{i,j})_{a \leq i \leq b, c \leq j \leq d}. \quad (2.16)$$

By convention, these subscripts have higher priority than the transposition or conjugate transposition, for example $M_{a:b,c:d}^* := (M_{a:b,c:d})^*$ is the conjugated transposition of the

submatrix $M_{a,b,c,d}$. For a matrix A , we write its operator norm as $\|A\| = \sup_{\|v\|=1} \|Av\|$ and its Frobenius norm $\|A\|_F = \sqrt{\sum_{i,j} |A_{i,j}|^2}$.

If $c = (c_k)_{k \in \mathbb{Z}}$ is a sequence of complex numbers, the $N \times N$ Toeplitz matrix $(c(i-j))$ is denoted by $T_N(c)$. If moreover the sequence (c_k) is a positive-definite function $c : \mathbb{Z} \rightarrow \mathbb{C}$ and admits by Herglotz's theorem the representation

$$c_k = \frac{1}{2\pi} \int_{(-\pi, \pi]} e^{-ikx} d\alpha(x),$$

then α is called the symbol of $T_N(c)$ which will sometimes be written $T_N[\alpha]$. If moreover α admits a density with respect to Lebesgue's measure, i.e. $d\alpha(x) = f(x) dx$, $T_N(c)$ will occasionally be denoted by $T_N[f]$. Notice that if $T_N(c)$ is the covariance matrix of a stationary process as in (2.11) then $f(x)$ (if it exists) is called the spectral density of the process.

Given two complex sequences x_n, y_n we denote

$$x_n \sim y_n \Leftrightarrow \lim_{n \rightarrow \infty} \left(\frac{x_n}{y_n} \right) = 1 \quad \text{and} \quad x_n \doteq y_n \Leftrightarrow \lim_{n \rightarrow \infty} (x_n - y_n) = 0. \quad (2.17)$$

The notations $x_n = o(1)$ and $x_n = O(1)$ respectively mean $\lim_{n \rightarrow \infty} x_n = 0$ and $\overline{\lim}_{n \rightarrow \infty} |x_n| < \infty$. These notations are also applicable to functions with continuous arguments. If X_n, X are random variables, the notations $X_n = o(1)$ and $X_n = o_P(1)$ respectively mean that $\lim_{n \rightarrow \infty} X_n = 0$ almost surely and in probability. The notations $X_n \xrightarrow{\mathcal{D}} X$ and $X_n \xrightarrow{\mathcal{P}} X$ respectively denote convergence in distribution and in probability. If μ, μ_n are measures, we denote with a slight abuse of notation $\mu_n \xrightarrow{\mathcal{D}} \mu$ for the weak convergence of μ_n to μ .

Given a random variable Y or a sub-algebra \mathcal{G} , we denote by $\mathbb{E}_Y(X)$ and $\mathbb{E}_{\mathcal{G}}(X)$ the conditional expectation of the random variable X with respect to Y and to \mathcal{G} .

We denote by $\mathcal{N}_{\mathbb{R}}(m, \sigma^2)$ the real Gaussian distribution with mean m and variance σ^2 ; we refer to $\mathcal{N}_{\mathbb{R}}(0, 1)$ as the standard real Gaussian distribution. A complex random variable Z is distributed according to the standard complex Gaussian distribution if $Z = U + iV$ where U, V are independent, each with distribution $\mathcal{N}_{\mathbb{R}}(0, 1/2)$. In this case we denote $Z \sim \mathcal{N}_{\mathbb{C}}(0, 1)$. For a symmetric semidefinite positive matrix T , denote by $\mathcal{N}_{\mathbb{R}}(0, T)$ the

distribution of a centered Gaussian vector with covariance matrix T .

In the proofs we use C to denote a constant that may take different values from one place to another.

Assumptions.

We state our results under one or several of the following assumptions:

A1 (Model setting) Let S_N be $N \times N$ random matrices defined as

$$S_N = \frac{1}{n} \Gamma_N^{\frac{1}{2}} Z_N Z_N^* \Gamma_N^{\frac{1}{2}}$$

where $Z_N = \left(Z_{i,j}^{(N)} \right)_{1 \leq i \leq N, 1 \leq j \leq n}$ are $N \times n$ matrix whose entries $Z_{i,j}^{(N)}$ are i.d. random variables for all i, j, N , and independent across i, j for each N , satisfying

$$\mathbb{E} Z_{i,j}^{(N)} = 0, \quad \mathbb{E} |Z_{i,j}^{(N)}|^2 = 1 \quad \text{and} \quad \mathbb{E} |Z_{i,j}^{(N)}|^4 < \infty,$$

and Γ_N are $N \times N$ positive semidefinite Hermitian deterministic matrices.

A2 (Asymptotic spectral structure of Γ_N) Given a sequence of $N \times N$ positive semidefinite deterministic matrices Γ_N , the empirical spectral distribution

$$\mu^{\Gamma_N} := \frac{1}{N} \sum_{k=1}^N \delta_{\lambda_k(\Gamma_N)}$$

forms a tight sequence on \mathbb{R}^+ , and the largest eigenvalue $\lambda_{\max}(\Gamma_N)$ tends to ∞ as $N \rightarrow \infty$.

Example II.1. If $\mu^{\Gamma_N} \xrightarrow{\mathcal{D}} \nu$ with ν a non-compactly supported probability on \mathbb{R}^+ , then **A2** holds.

Example II.2. Consider $\Gamma_N = \text{diag}(\ell_1^{(n)}, \dots, \ell_m^{(n)}, 1, \dots, 1)$ where $m = m(n)$ is such that $\frac{m(n)}{n} \rightarrow 0$ and where $\ell_i^{(n)} \nearrow \infty$ ($1 \leq i \leq m$), then **A2** holds. The illustrative and simpler case where $\Gamma_N = \text{diag}(\ell_1^{(n)}, 1, \dots, 1)$ will be used hereafter.

A3 (Subarray assumption on Z_N) For each N , $Z_N = Z_{1:N,1:n}$ is the top-left submatrix of an infinite matrix $Z = (Z_{i,j})_{i,j \geq 1}$, with $Z_{i,j}$ i.i.d random variables satisfying

$$\mathbb{E}Z_{i,j} = 0, \quad \mathbb{E}|Z_{i,j}|^2 = 1 \quad \text{and} \quad \mathbb{E}|Z_{i,j}|^4 < \infty.$$

A4 (Spectral gap condition on Γ_N) The two largest eigenvalues $\lambda_{\max}(\Gamma_N)$ and $\lambda_2(\Gamma_N)$ satisfy

$$\overline{\lim}_{N \rightarrow \infty} \frac{\lambda_2(\Gamma_N)}{\lambda_{\max}(\Gamma_N)} < 1.$$

Notice that this spectral gap condition already appears in [47, 57, 20].

A5 (Block-diagonal structure of Γ_N) For all N , Γ_N has the block-diagonal form

$$\Gamma_N = \begin{pmatrix} \lambda_{\max}(\Gamma_N) & 0 \\ 0 & \Gamma_{N-1} \end{pmatrix},$$

where Γ_{N-1} is a $(N-1) \times (N-1)$ semidefinite positive Hermitian matrix.

2.3.2 Main results

We now present the main results of this article. Proposition II.3 and Theorem II.4 describe the limiting behaviour and fluctuations of $\lambda_{\max}(S_N)$ under generic assumptions. Theorem II.8 and Corollary II.10 specialize the previous results to Toeplitz covariance matrices and Gaussian long memory stationary processes.

Proposition II.3. *Let S_N be a $N \times N$ matrix given by (2.1) and assume that **A1** and **A2** hold. Then*

$$\frac{\lambda_{\max}(S_N)}{\lambda_{\max}(\Gamma_N)} \xrightarrow[N, n \rightarrow \infty]{\mathcal{P}} 1.$$

*If moreover either the random variables $Z_{ij}^{(N)}$ are standard (real or complex) Gaussian or Assumption **A3** holds, then the above convergence holds almost surely.*

This result already appears under different assumptions in [38, Prop. 7.3], [20, Th. 2.1].

Remark 1 (consistency with the bounded case $\|\Gamma_N\| < \infty$). Consider the simple case where $\Gamma_N = \text{diag}(\ell_1, 1, \dots, 1)$, where $\ell_1 > 1 + \sqrt{r}$ is fixed - see for instance (2.6). Then it is well known (cf. [10]) that

$$\lambda_{\max}(S_N) \xrightarrow[N, n \rightarrow \infty]{\mathcal{P}} \ell_1 + \frac{r \ell_1}{\ell_1 - 1}.$$

Notice in particular that $\frac{\lambda_{\max}(S_N)}{\ell_1} = 1 + \frac{r}{\ell_1 - 1} + o_P(1)$, which is heuristically consistent with Proposition II.3 if one lets ℓ_1 go to infinity.

Theorem II.4. *Let S_N be a $N \times N$ matrix given by (2.1) and assume that A1, A2 and A4 hold. Assume moreover that one of the following conditions is satisfied:*

- (i) *Assumption A5 holds,*
- (ii) *The random variables $Z_{ij}^{(N)}$ are standard complex Gaussian,*
- (iii) *The random variables $Z_{ij}^{(N)}$ are standard real Gaussian and matrices Γ_N are real symmetric.*

Consider the quantities

$$\beta_N := \frac{1}{n} \sum_{k=2}^N \frac{\lambda_k(\Gamma_N)}{\lambda_{\max}(\Gamma_N) - \lambda_k(\Gamma_N)} \quad \text{and} \quad F_N := \sqrt{n} \left(\frac{\lambda_{\max}(S_N)}{\lambda_{\max}(\Gamma_N)} - 1 - \beta_N \right). \quad (2.18)$$

Then

$$F_N \xrightarrow[N, n \rightarrow \infty]{\mathcal{D}} \mathcal{N}(0, \sigma^2), \quad (2.19)$$

where $\sigma^2 = \mathbb{E}|Z_{1,1}^{(N)}|^4 - 1$.

Counterparts of Theorem II.4 appear under the assumption that the $(\lambda_i(\Gamma_N))$'s are bounded for $i \geq K$ and $K = o(N)$, see [57, Th. 3.1], [20, Th. 2.2]. In this latter case, the quantity β_N above can be replaced by $n^{-1} \sum_{K+1}^N \lambda_i(\Gamma_N) / (\lambda_{\max}(\Gamma_N) - \lambda_i(\Gamma_N))$. Beware however that under our assumption, the full summation is required because there is no natural threshold K if one does not assume boundedness on the majority of the population eigenvalues.

Remark 2. Notice that if $\mathbb{E}|Z_{1,1}^{(1)}|^4 = 1$ then $\sigma^2 = 0$ in the previous theorem, hence $F_N \xrightarrow[N, n \rightarrow \infty]{\mathcal{P}} 0$. Simulation 3 in Section 2.4.2 (see also Fig. 2.3(b)) supports this fact.

Remark 3. Under **A2** and **A4**, we have $\beta_N \xrightarrow[N, n \rightarrow \infty]{} 0$. Indeed, by the spectral gap condition **A4** and the fact that $N = O(n)$

$$\beta_N = \frac{1}{n} \sum_{k=2}^N \frac{\lambda_k(\Gamma_N)/\lambda_{\max}(\Gamma_N)}{1 - \lambda_k(\Gamma_N)/\lambda_{\max}(\Gamma_N)} \leq \frac{C}{N} \sum_{k=2}^N \frac{\lambda_k(\Gamma_N)}{\lambda_{\max}(\Gamma_N)}.$$

Since μ^{Γ_N} is tight, for any $\varepsilon \in (0, 1)$ there exists $M > 0$ s.t. $|\{k, : \lambda_k(\Gamma_N) > M\}|/N < \varepsilon$ where $|\{\cdot\}|$ denotes the cardinality of a set. Hence

$$\overline{\lim}_{N, n \rightarrow \infty} \beta_N \leq C \overline{\lim}_N \frac{M}{\lambda_{\max}(\Gamma_N)} + C\varepsilon = C\varepsilon,$$

where we use the fact that $\lambda_{\max}(\Gamma_N) \rightarrow \infty$ as $N \rightarrow \infty$ for the last equality. Notice however that $\sqrt{n}\beta_N$ may not go to zero as $N, n \rightarrow \infty$.

Remark 4 (consistency with the bounded case $\|\Gamma_N\| < \infty$, continued). Consider again the case where $\Gamma_N = \text{diag}(\ell_1, 1, \dots, 1)$ with $\ell_1 > 1 + \sqrt{r}$ then

$$\beta_N = \frac{N-1}{n} \frac{1}{\ell_1 - 1} = \frac{r_N}{\ell_1 - 1} + O\left(\frac{1}{n}\right).$$

In the case where $\ell_1 \rightarrow \infty$, F_N in (2.18) writes

$$F_N = \sqrt{n} \left(\frac{\lambda_{\max}(S_N)}{\ell_1} - \left(1 + \frac{r_N}{\ell_1 - 1}\right) \right) + O\left(\frac{1}{\sqrt{n}}\right)$$

and has Gaussian fluctuations. This formula is consistent with (2.6) which can be rewritten

$$\frac{\sqrt{n}}{\sqrt{1 + O(\ell_1^{-2})}} \left(\frac{\lambda_{\max}(S_N)}{\ell_1} - \left(1 + \frac{r_N}{\ell_1 - 1}\right) \right).$$

Example II.5 (various behaviours of $\sqrt{n}\beta_N$). Consider $\Gamma_N = \text{diag}(\ell_1^{(n)}, 1, \dots, 1)$ where $\ell_1^{(n)} \nearrow \infty$, then

$$\sqrt{n}\beta_N = \sqrt{n} \times \left(\frac{N-1}{n}\right) \times \left(\frac{1}{\ell_1 - 1}\right) \xrightarrow[N, n \rightarrow \infty]{} \begin{cases} \infty & \text{if } \ell_1^{(n)} \ll \sqrt{n}, \\ r/a & \text{if } \ell_1^{(n)} = a\sqrt{n}, \\ 0 & \text{if } \ell_1^{(n)} \gg \sqrt{n}. \end{cases}$$

2.3.3 Application to large sample covariance matrices associated with long memory processes

In order to apply the above results to Gaussian stationary processes with long memory, we need to verify the spectral gap condition of their autocovariance matrices. We first recall some definitions.

Definition II.6 (Regularly/Slowly varying functions). A measurable function $R : \mathbb{R} \rightarrow \mathbb{R}$ is regularly varying at infinity if $R(y) > 0$ for $|y|$ large enough and if there exists a real number ρ s.t. for any $x > 0$,

$$\lim_{y \rightarrow \infty} \frac{R(xy)}{R(y)} = x^\rho.$$

The number ρ is called the index of the regular variation. If $\rho = 0$, then we say that the function (often denoted by L in this case) is slowly varying.

A sequence of real numbers $(c_k)_{k \in \mathbb{Z}}$ is regularly (resp. slowly) varying if $y \mapsto c_{\lfloor y \rfloor}$ is a regularly (resp. slowly) varying function.

With Definition II.6, long memory (long range dependence) stationary processes can be defined as follows.

Definition II.7. A stationary process $(\mathcal{X}_t)_{t \in \mathbb{Z}}$ has long memory or long range dependence if its autocovariance function γ is regularly varying with index $\rho \in (-1, 0)$.

Remark 5. Notice that this definition is compatible with the definition of the autocovariance function provided in (2.14). In fact, assume that $\gamma(h)$ is given by (2.14) then it is regularly varying with index $\rho = 2d - 1 \in (-1, 0)$. Conversely, assume that $\gamma(h)$ is an even regularly varying sequence with $\rho \in (-1, 0)$. Set $d = \frac{\rho+1}{2}$, then $L(y) = \gamma(\lfloor y \rfloor)(1 + |y|)^{1-2d}$ is a slowly varying function with $d \in (0, 1/2)$ and

$$\gamma(h) = \frac{L(h)}{(1 + |h|)^{1-2d}}.$$

Remark 6. Notice that definitions II.6 and II.7 enable to consider complex processes (\mathcal{X}_t) , however the associated autocovariance function is necessarily real and cannot be complex.

Remark 7. The above definition coincides with Condition II in [45] where the autocovariance function γ satisfies (2.14).

In this context, the spectral gap condition on autocovariance matrices of a long memory stationary process is ensured by the following theorem:

Theorem II.8. *Suppose that $c = (c_h)_{h \in \mathbb{Z}}$ is an even ($c_h = c_{-h}$ for all $h \in \mathbb{Z}$) regularly varying sequence of index $\rho \in (-1, 0)$, then for any fixed $k \geq 1$ the limit*

$$a_k^{(\rho)} = \lim_{N \rightarrow \infty} \frac{\lambda_k(T_N(c))}{Nc_N}$$

exists and is finite. Moreover, $a_1^{(\rho)} > a_2^{(\rho)} \geq 0$. In particular,

$$\lambda_{\max}(T_N(c)) \sim a_1^{(\rho)} Nc_N \rightarrow \infty \quad \text{and} \quad \lim_{N \rightarrow \infty} \frac{\lambda_2(T_N(c))}{\lambda_{\max}(T_N(c))} = \frac{a_2^{(\rho)}}{a_1^{(\rho)}} < 1.$$

Definition II.9. We say that a complex Gaussian process $(\mathcal{X}_t)_{t \in \mathbb{Z}}$ is circularly symmetric if for any $N \in \mathbb{N}$ the Gaussian vector $\mathcal{X}_{1:N} := (\mathcal{X}_1, \dots, \mathcal{X}_N)^\top$ is circularly symmetric, i.e. for any $\phi \in \mathbb{R}$, the vector $e^{i\phi} \mathcal{X}_{1:N}$ has the same distribution as $\mathcal{X}_{1:N}$.

Remark 8. Notice in particular that such a process is centered and satisfies that $\mathbb{E} \mathcal{X}_{1:N} \mathcal{X}_{1:N}^\top = 0$, for all $N \geq 1$.

As a canonical example, a standard complex Gaussian vector $X = (X_1, \dots, X_N)^\top$ where the X_i 's are i.i.d. and $\mathcal{N}_{\mathbb{C}}(0, 1)$ -distributed is circularly symmetric with

$$\mathbb{E} X X^* = I_N \quad \text{and} \quad \mathbb{E} X X^\top = 0.$$

For such a vector X , we will denote $X \sim \mathcal{N}_{\mathbb{C}}(0, I_N)$.

As a by-product of the above theorems, we have the following result on the largest eigenvalue of sample covariance matrices of a Gaussian long memory stationary process (recall that such a process admits a real autocovariance function).

Corollary II.10. *Suppose that $(\mathcal{X}_t)_{t \in \mathbb{Z}}$ is a real centered (resp. complex circularly symmetric) Gaussian stationary process with long range dependence in the sense of definition II.7.*

Let

$$Q_N = \frac{1}{n} X_N X_N^*$$

where X_N are $N \times n$ random matrices whose columns are i.i.d copies of the random vector $\mathcal{X}_{1:N} = (\mathcal{X}_1, \dots, \mathcal{X}_N)^\top$. Then

$$\frac{\lambda_{\max}(Q_N)}{\lambda_{\max}(T_N)} \xrightarrow[N, n \rightarrow \infty]{a.s.} 1 \quad (2.20)$$

and

$$\sqrt{n} \left(\frac{\lambda_{\max}(Q_N)}{\lambda_{\max}(T_N)} - 1 - \frac{1}{n} \sum_{k=2}^N \frac{\lambda_k(T_N)}{\lambda_{\max}(T_N) - \lambda_k(T_N)} \right) \xrightarrow[N, n \rightarrow \infty]{\mathcal{D}} \mathcal{N}(0, \sigma^2), \quad (2.21)$$

where $T_N = T_N(\gamma)$ is the autocovariance matrix of the process defined in (2.11) and where

$$\sigma^2 = \frac{\mathbb{E}|\mathcal{X}_1^4|}{\gamma^2(0)} - 1 = \begin{cases} 2 & \text{if } \mathcal{X}_1 \text{ is real} \\ 1 & \text{if } \mathcal{X}_1 \text{ is complex} \end{cases}.$$

Corollary II.10 being an easy consequence of Proposition II.3 and Theorems II.4 and II.8, we provide its proof hereafter.

Proof. Let X be a centered N -dimensional random vector either real or circularly symmetric complex gaussian with (real) covariance matrix T . Then X writes $X = T^{1/2}Z$ where $Z \sim \mathcal{N}_{\mathbb{R}}(0, I_N)$ or $\mathcal{N}_{\mathbb{C}}(0, I_N)$ depending on whether X is real or complex. In fact, if T is invertible then $Z = T^{-1/2}X$ has the required properties.

If not, $T = O \text{diag}(d_1^1, \dots, d_p^2, 0 \dots) O^\top$ with O orthogonal and $d_i > 0$. Let $Y = (0, \dots, 0, Y_{p+1}, \dots, Y_N)^\top$ with Y_k i.i.d. standard Gaussian random variables, either real or complex (depending on X), and independent from X . Let

$$Z = O \text{diag}(d_1^{-1}, \dots, d_p^{-1}, 0 \dots) O^\top X + OY$$

then $\text{Cov}(Z) = I_N$ and if X is complex, then $\mathbb{E}ZZ^\top = 0$. In particular, Z is a standard gaussian random vector and a covariance computation yields

$$\text{Cov}(X - T^{1/2}Z) = 0,$$

which implies that $X = T^{1/2}Z$ almost surely.

Then for any N and almost surely, the representation $Q_N = \frac{1}{n}T_N^{1/2}Z_NZ_N^*T_N^{1/2}$ holds with (T_N) the autocovariance matrices of process $(\mathcal{X}_t)_{t \in \mathbb{Z}}$. In Section 2.2 we have noticed that μ^{T_N} converges weakly, and by Theorem II.8, $\lambda_{\max}(T_N) \rightarrow \infty$ and (T_N) satisfy the spectral gap condition. By Proposition II.3 and Theorem II.4, the results follow.

□

2.4 Additional applications and simulations

2.4.1 Additional applications to Gaussian stationary processes

Although Definition II.7 is a common definition of long memory, it can seem restrictive as it requires that the autocovariance function has at most a finite number of nonpositive values. In the first example hereafter we consider Gaussian processes with autocovariance functions either complex or with alternate signs. In the two subsequent examples, we relate our results with other definitions of long memory, via linear representation or via the autocovariance density.

Covariance matrices with alternating signs of entries.

Let $(\mathcal{X}_t)_{t \in \mathbb{Z}}$ be a centered Gaussian stationary process with autocovariance function $\gamma^{\mathcal{X}}$. Let $\theta \in (-\pi, \pi]$, $\theta \neq 0$ be fixed and consider the process $(\mathcal{Y}_t = e^{it\theta} \mathcal{X}_t)_{t \in \mathbb{Z}}$. This process is a Gaussian stationary process with autocovariance function $\gamma^{\mathcal{Y}}(t) = e^{it\theta} \gamma^{\mathcal{X}}(t)$ and

$$\mathcal{Y}_{1:N} = \Sigma^\theta \mathcal{X}_{1:N}$$

where $\Sigma^\theta = \text{diag}(e^{ik\theta}, 1 \leq k \leq N)$ is a unitary matrix. Notice that if (\mathcal{X}_t) is complex circularly symmetric, then so is (\mathcal{Y}_t) but if (\mathcal{X}_t) is real then (\mathcal{Y}_t) is either complex Gaussian but not circularly symmetric if $\theta \neq \pi$ or real Gaussian with alternate signs if $\theta = \pi$.

Let $X_N^{\mathcal{X}}$ (resp. $X_N^{\mathcal{Y}}$) a $N \times n$ matrix whose columns are i.i.d. copies of the vector $\mathcal{X}_{1:N}$ (resp. $\mathcal{Y}_{1:N}$) and assume that the process (\mathcal{X}_t) fulfills the assumptions of Corollary II.10.

Then

$$Q_N^{\mathcal{Y}} = \frac{1}{n} X_n^{\mathcal{Y}} (X_n^{\mathcal{Y}})^* = \frac{1}{n} \Sigma^\theta X_n^{\mathcal{X}} (X_n^{\mathcal{X}})^* (\Sigma^\theta)^* = \Sigma^\theta Q_N^{\mathcal{X}} (\Sigma^\theta)^* \quad \text{where} \quad Q_N^{\mathcal{X}} = \frac{1}{n} X_n^{\mathcal{X}} (X_n^{\mathcal{X}})^*.$$

In particular, $\lambda_{\max}(Q_N^{\mathcal{Y}}) = \lambda_{\max}(Q_N^{\mathcal{X}})$ satisfies (2.20) and (2.21). In this example the positivity constraint of the autocovariance function is relaxed.

Linear processes with long memory.

If the process $(\mathcal{X}_t)_{t \in \mathbb{Z}}$ has a linear representation

$$\mathcal{X}_t = \sum_{j=0}^{\infty} \psi_j \epsilon_{t-j}$$

where $(\epsilon_t)_{t \in \mathbb{Z}}$ is a sequence of i.i.d real valued standard Gaussian r.v.'s and $\psi_j \sim j^{d-1} L(j)$ as $j \rightarrow \infty$ with $d \in (0, 1/2)$ and L a slowly varying function at ∞ , then it is well known (c.f. for example [45, Corollary 2.2.10]) that its autocovariance function γ is regularly varying with index $\rho = 2d - 1$, and more precisely we have

$$\gamma(h) \sim h^{2d-1} L^2(h) B(1 - 2d, d)$$

where $B(1 - 2d, d) = \int_0^1 x^{-2d} (1 - x)^{d-1} dx$ is the beta-function. Corollary II.10 can be applied in this case.

Long range dependence defined through spectral density

Among the various definitions of long range dependence, there is an important one which defines the long range dependence through the spectral density, that is, if the symbol of the autocovariance matrices T_N has a density $f : (-\pi, \pi] \rightarrow \mathbb{R}^+$ satisfying

$$f(x) = |x|^{-2d} L\left(\frac{1}{|x|}\right), \quad x \in (-\pi, 0) \cup (0, \pi], \quad (2.22)$$

with $d \in (0, 1/2)$, and L a slowly varying function defined on $[1/\pi, +\infty)$. (cf. Condition IV in [45]). If a (real centered or complex circularly) Gaussian stationary process $(\mathcal{X}_t)_{t \in \mathbb{Z}}$ has

long memory in this sense, then the **LSD** of the covariance matrices $T_N[f]$ is not compactly supported and in particular $\lambda_{\max}(T_N[f]) \rightarrow \infty$ and (2.20) in Corollary II.10 holds. Moreover if L in (2.22) is quasi-monotone, then the process is also a long memory process in the sense of Definition II.7 (see for instance [45, Corollary 2.2.17]) with index $\rho = 2d - 1$. More precisely we have

$$\gamma(h) \sim 2h^{2d-1}L(h)\Gamma(1 - 2d)\sin(d\pi) \text{ as } h \rightarrow \infty$$

where $\Gamma(t) := \int_0^\infty x^{t-1}e^{-x} dx$ denotes the gamma-function. Hence by Theorem II.8, T_N satisfies the spectral gap condition, and applying Theorem II.4 we get the same result as Corollary II.10.

2.4.2 Numerical simulations

Simulation 1: Limiting behaviour of $\lambda_{\max}(S_N)$.

To illustrate Proposition II.3, we take

$$S_N = \frac{1}{n}\Gamma_N^{\frac{1}{2}}Z_N Z_N^\top \Gamma_N^{\frac{1}{2}}$$

with Z_N a $N \times n$ matrix having i.i.d standard real Gaussian entries, and $\Gamma_N = T_N(\gamma)$ is the Toeplitz matrix determined by $\gamma(h) = \frac{1}{(1+|h|)^{3/4}}$. Let N take all the values in the finite sequence $\{100, 150, 200, \dots, 3000\}$, and let $n = \lfloor 5N/4 \rfloor$. We plot the simulation results in Figure 2.1.

Simulation 2: Fluctuations of $\lambda_{\max}(S_N)$.

To illustrate the fluctuations of $\lambda_{\max}(S_N)$, we fix $N = 1000$ and $n = 1250$ and let $\Gamma_N = \text{diag}(\lambda_k(T_{1000}(\gamma)))$ with γ as in the previous simulation. We take 900 independent samples of S_{1000} , plot the histogram of F_{1000} defined in (2.18) and compare with the density of the theoretical limiting law. In Figure 2.2(a) we simulate the model S_{1000} with Z_{1000} having i.i.d. real Gaussian entries, the limiting law, according to Theorem II.4, is $\mathcal{N}(0, 2)$; while in Figure 2.2(b), Z_{1000} has i.i.d. standardized exponential entries, i.e. $Z_{i,j}^{(1000)} \sim \mathcal{E}(1) - 1$. The limiting law is $\mathcal{N}(0, 8)$.

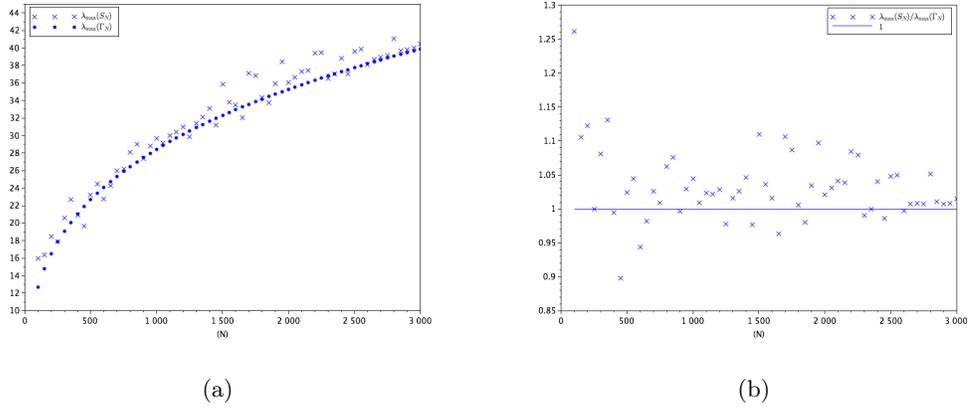


Figure 2.1: Convergence of $\lambda_{\max}(S_N)/\lambda_{\max}(\Gamma_N)$ to 1. In 2.1(a), the values $\lambda_{\max}(S_N)$ and $\lambda_{\max}(\Gamma_N)$ are plotted as crosses and solid points respectively; in 2.1(b) the values of the ratio $\lambda_{\max}(S_N)/\lambda_{\max}(\Gamma_N)$ are plotted as crosses, compared with the constant 1.

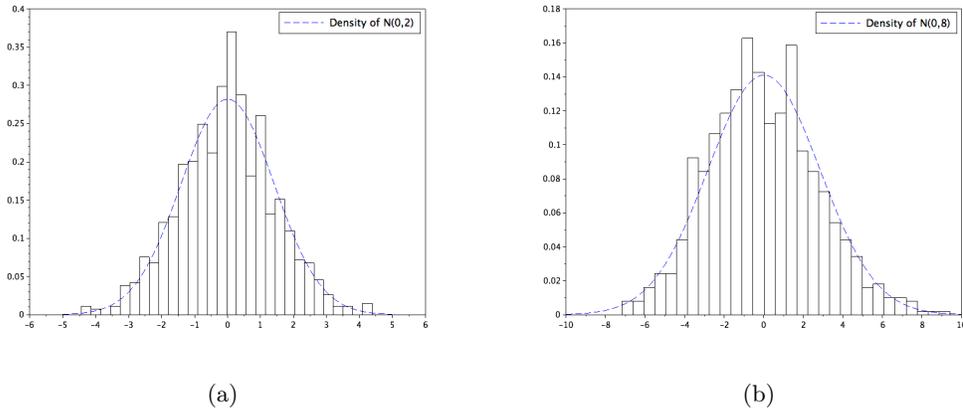


Figure 2.2: Fluctuations of $\lambda_{\max}(S_{1000})$, with Z_{1000} having Gaussian entries in 2.2(a) and standardized exponential entries in 2.2(b). Simulation with 900 samples.

Simulation 3: Concentration.

We now address the case $\mathbb{E}|Z_{1,1}^{(N)}|^4 = 1$. Consider a matrix Z_N with i.i.d. symmetric Bernoulli variables taking values in $\{-1, 1\}$. As previously we take $\Gamma_N = \text{diag}(\lambda_k(T_N(\gamma)))$ with $\gamma(h) = \frac{1}{(1+|h|)^{3/4}}$. In this case, Theorem II.4 asserts that

$$\sqrt{n} \left(\frac{\lambda_{\max}(S_N)}{\lambda_{\max}(\Gamma_N)} - 1 - \frac{1}{n} \sum_{k=2}^N \frac{\lambda_k(\Gamma_N)}{\lambda_{\max}(\Gamma_N) - \lambda_k(\Gamma_N)} \right) \xrightarrow[N, n \rightarrow \infty]{\mathcal{P}} 0.$$

In Figure 2.3(a) we plot the fluctuations of $\lambda_{\max}(S_{1000})$ with $n = 1250$ and notice that the corresponding F_{1000} are far more concentrated around 0 than the previous simulations, as predicted by the theorem.

An interesting phenomenon occurs in Figure 2.3(b), where the same matrix Z_{1000} is considered while we do not diagonalize Γ_{1000} and just take $\Gamma_{1000} = T_{1000}(\gamma)$. In this case, the concentration phenomenon disappears, and the obtained histogram is very close to that in Figure 2.2(a). This simulation suggests that some universality holds in the fluctuations of $\lambda_{\max}(S_N)$ if the (Γ_N) 's are Toeplitz matrices. This will be explored in a forthcoming work.

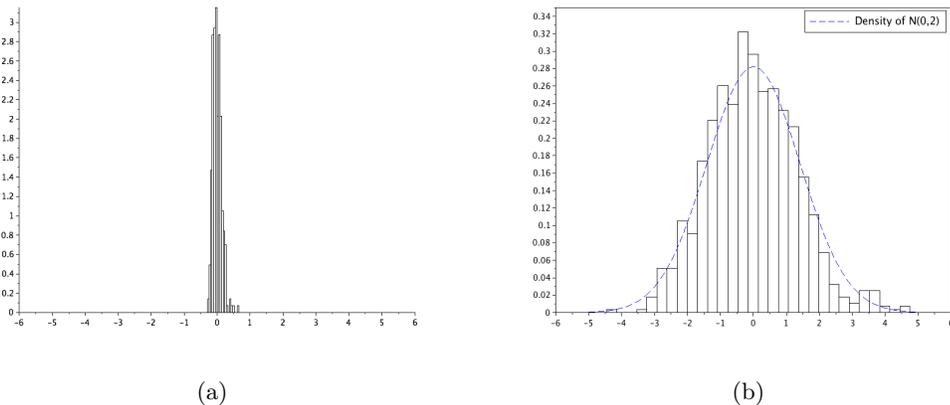


Figure 2.3: Fluctuations of $\lambda_{\max}(S_N)$ in the case of symmetric Bernoulli entries. Γ_N is diagonal in 2.3(a) and nondiagonal in 2.3(b).

2.4.3 Open questions

At the border between long memory and short memory.

An interesting regime is when $\rho = -1$. In this case, the autocovariance function $\gamma(h) = (1 + |h|)^{-1}L(h)$ can be summable or not depending on the slowly varying function L . For example if $L(h) = \log^{-1-\varepsilon}(2 + |h|)$ with $\varepsilon > 0$ then γ is absolutely summable and the process has short memory. If $L(\cdot) = 1$ then one can prove that $\lambda_{\max}(T_N(\gamma)) = O(\log N)$, and that the spectral gap condition no longer holds. In this case, the asymptotics (2.20) remains true as Proposition II.3 does not rely on the spectral gap condition but only on the condition $\lambda_{\max}(T_N(\gamma)) \rightarrow \infty$. The question whether the fluctuations (2.21) together with their normalization and the limiting distribution hold remains open.

Non-Gaussian long memory stationary processes.

A Gaussian long memory stationary process admits a linear representation $\mathcal{X}_{1:N} = T_N^{1/2}(\gamma)Z_{1:N}$, where $T_N(\gamma)$ is a hermitian Toeplitz matrix and $Z_{1:N}$ is a standard Gaussian vector. This representation is key in the analysis of the top eigenvalue of the corresponding large covariance matrix of samples of the process but does not hold anymore if the process is not Gaussian. The question whether it is possible to perform the same eigenvalue analysis in the case of non-Gaussian long memory stationary process is open.

Correlation structure of the top eigenvalues.

Beyond the top eigenvalue $\lambda_{\max}(S_N)$, it would be interesting to understand the asymptotic correlation structure and the fluctuations of the (many) largest eigenvalues $(\lambda_{\max}(S_N), \lambda_2(S_N), \dots, \lambda_k(S_N))$ for a fixed $k \geq 1$.

Behaviour of the eigenvectors associated to the top eigenvalues.

For bounded spiked models (by bounded we mean $\sup_N \|\Gamma_N\| < \infty$) the structure of the eigenvectors associated to the top eigenvalues has been studied and carries interesting information, see for example [13]. A similar study would be interesting in the general context of unbounded population covariance matrices where $\lambda_{\max}(\Gamma_N) \rightarrow \infty$ and of (Gaussian) long

memory stationary processes. In this latter case, one needs to have a good understanding of the Toeplitz matrix' $T_N(\gamma)$ eigenvectors.

Universality for non-Gaussian linear stationary processes with long memory.

In the case where Γ_N is required to be (block-)diagonal, the variance of the limiting distribution depends on the fourth moment of the entries $Z_{i,j}^N$ and may be equal to zero if $\mathbb{E}|Z_{1,1}^{(1)}|^4 = 1$. However when Γ_N is a Toeplitz matrix (2.11) with γ satisfying (2.14), this dependence is weakened and Simulation 3 in Section 2.4.2 strongly suggests that some universality occurs depending on the population covariance matrix Γ_N , see in particular Figures 2.3(a) and 2.3(b). This question will be addressed in a forthcoming work.

2.5 Proofs of Proposition II.3 and Theorem II.4

2.5.1 A short reminder of results related to large covariance matrices

Given a probability measure μ on \mathbb{R} , define its Cauchy-Stieltjes transform as

$$m_\mu(z) := \int \frac{d\mu(s)}{z - s}, \quad \forall z \in \mathbb{C}^+ := \{z \in \mathbb{C} : \Im z > 0\}$$

Notice that $m_\mu(z)$ is the opposite of the Stieltjes transform $g_\mu(z) = \int \frac{d\mu(s)}{s - z}$.

For a random matrix S_N given by (2.1), we will often consider its companion matrix

$$\underline{S}_N = \frac{1}{n} Z_N^* \Gamma_N Z_N, \tag{2.23}$$

which shares the same non-zero eigenvalues with S_N . In particular, $\lambda_{\max}(\underline{S}_N) = \lambda_{\max}(S_N)$.

Recall that $r_N := \frac{N}{n}$ and let $\mu^{S_N}, \mu^{\underline{S}_N}$ be the **ESD** of S_N and \underline{S}_N respectively, then the following relation holds:

$$\mu^{\underline{S}_N} = (1 - r_N)\delta_0 + r_N\mu^{S_N}. \tag{2.24}$$

Limiting spectral distribution.

We recall results from [48, Theorem 1.1]. For any probability ν in \mathbb{R}^+ and any $r \in (0, +\infty)$, there exists a unique probability measure $\mu = \mu(r, \nu)$ whose Cauchy-Stieltjes

transform m_μ satisfies the equation:

$$m_\mu(z) = \int \frac{d\nu(s)}{z - s(1 - r + rz m_\mu(z))} \quad \text{for any } z \in \mathbb{C}^+.$$

If the probability measure ν is the **ESD** μ^A associated with a matrix A , we simply write $\mu = \mu(r, A)$ instead of $\mu = \mu(r, \mu^A)$. Similarly, there exists a unique probability measure $\underline{\mu} = \underline{\mu}(r, \nu)$ with Cauchy-Stieltjes transform $m_{\underline{\mu}}$ satisfying

$$z = \frac{1}{m_{\underline{\mu}}(z)} + r \int \frac{s d\nu(s)}{1 - s m_{\underline{\mu}}(z)} \quad \text{for any } z \in \mathbb{C}^+. \quad (2.25)$$

As previously, we will write $\underline{\mu}(r, A)$ instead of $\underline{\mu}(r, \mu^A)$. If moreover $\mu^{\Gamma_N} \xrightarrow[N \rightarrow \infty]{\mathcal{D}} \nu$, then

$$\mu^{S_N} \xrightarrow[N, n \rightarrow \infty]{\mathcal{D}} \mu \quad \text{a.s.} \quad \text{and} \quad \mu^{\underline{S}_N} \xrightarrow[N; n \rightarrow \infty]{\mathcal{D}} \underline{\mu} \quad \text{a.s.} \quad (2.26)$$

Spectrum confinement.

By "spectrum confinement", we refer to the phenomenon where the empirical spectrum of the eigenvalues "concentrates" near the support of the limiting spectral distribution. In the specific case of model (2.1) under assumption $\sup_N \|\Gamma_N\| < \infty$ and the convergence (2.26), spectrum confinement can be roughly expressed (in the absence of spikes) as: for every $\varepsilon > 0$, almost surely,

$$\text{supp}(\mu^{S_N}) \subset \text{supp}(\mu) + (-\varepsilon, \varepsilon)$$

for N large enough.

A more accurate description of spectrum confinement relies on the deterministic equivalent of μ^{S_N} defined as $\underline{\mu}_N := \underline{\mu}(r_N, \Gamma_N)$ (cf. (2.25) with $\nu = \mu^{\Gamma_N}$). Assume that **A3** holds. By [1, Theorem 1.1], if there exists $\varepsilon > 0$ and an interval $[a, b]$ such that

$$[a, b] \cap \left(\text{supp}(\underline{\mu}_N) + (-\varepsilon, \varepsilon) \right) = \emptyset$$

for N large enough, then almost surely

$$\text{supp}(\mu^{S_N}) \cap [a, b] = \emptyset \quad (2.27)$$

for N large enough. In particular, if $a > 0$ there is no eigenvalue of S_N in $[a, b]$ for N large enough.

The description of the support of a probability distribution defined via a fixed-point equation (2.25) is given in [50, Theorems 4.1 and 4.2]. Based on these results, we now state a necessary and sufficient condition for which a real number x lies outside the support of $\underline{\mu}_N = \underline{\mu}(r_N, \Gamma_N)$. Let

$$B_N := \{y \in \mathbb{R} : y \neq 0, y^{-1} \neq \lambda_k(\Gamma_N), \forall k = 1, \dots, N\}, \quad (2.28)$$

and define

$$x_N(y) := \frac{1}{y} + r_N \int \frac{s \, d\mu^{\Gamma_N}(s)}{1 - sy} \quad \text{for } y \in B_N. \quad (2.29)$$

A real number $x \in \mathbb{R}$ lies outside the support of $\underline{\mu}_N$ if and only if

$$\exists y \in B_N, \quad x = x_N(y) \quad \text{and} \quad x'_N(y) = -\frac{1}{y^2} + r_N \int \frac{s^2 \, d\mu^{\Gamma_N}(s)}{(1 - sy)^2} < 0.$$

Exact separation.

Let $[a, b]$ be an interval eventually outside the support of $\underline{\mu}_N = \underline{\mu}(r_N, \Gamma_N)$, assume that $\mu^{\Gamma_N} \rightarrow \nu$ and let $\underline{\mu} = \underline{\mu}(r, \nu)$. "Exact separation" is a phenomenon that expresses the fact that (almost surely and eventually) the interval $[a, b]$ separates the empirical eigenvalues of matrix S_N exactly in the same proportions as $[1/m_{\underline{\mu}}(a), 1/m_{\underline{\mu}}(b)]$ separates those of matrix Γ_N .

This expression has been coined in the article [2] by Bai and Silverstein, from which we recall the result of interest to us, that is mainly [2, Theorem 1.2(2)]: Assume in addition to the assumptions of [1, Theorem 1.1] (and in particular to assumption **A3**) that the conditions $m_{\underline{\mu}}(b) > 0$ and $r(1 - \nu(\{0\})) \leq 1$ hold. For N large enough, let i_N be an integer

such that

$$\lambda_{i_N}(\Gamma_N) > \frac{1}{m_{\underline{\mu}}(b)} \quad \text{and} \quad \lambda_{i_N+1}(\Gamma_N) < \frac{1}{m_{\underline{\mu}}(a)}.$$

Then almost surely, $\lambda_{i_N}(S_N) > b$ and $\lambda_{i_N+1}(S_N) < a$ for N large enough. This result will be used in the particular case where $\nu = \delta_0$. In this case, $m_{\underline{\mu}}(x) = \frac{1}{x}$ and $r(1 - \delta_0(0)) = 0$.

2.5.2 Reduction to the bounded model

When studying $\lambda_{\max}(S_N)$ under **A2**, the main difficulty is to handle the unboundedness of $\lambda_{\max}(\Gamma_N)$. In order to circumvent this issue, we define

$$\tilde{S}_N = \frac{1}{n} \tilde{\Gamma}_N^{1/2} Z_N Z_N^* \tilde{\Gamma}_N^{1/2} \quad \text{where} \quad \tilde{\Gamma}_N := \frac{\Gamma_N}{\lambda_{\max}(\Gamma_N)}.$$

In particular, notice that $\lambda_{\max}(\tilde{S}_N) = \frac{\lambda_{\max}(S_N)}{\lambda_{\max}(\Gamma_N)}$. Thus, in order to establish the results stated in Proposition **II.3** and Theorem **II.4**, we only need to prove the corresponding results for \tilde{S}_N .

Using the definition of $\tilde{\Gamma}_N$, the tightness of (μ^{Γ_N}) and the fact that $\lambda_{\max}(\Gamma_N) \rightarrow \infty$, we immediatly deduce the following properties for $\tilde{\Gamma}_N$:

$$\lambda_{\max}(\tilde{\Gamma}_N) = 1 \quad \text{and} \quad \mu^{\tilde{\Gamma}_N} \xrightarrow[N \rightarrow \infty]{\mathcal{D}} \delta_0.$$

In particular, the spectral norm of $\tilde{\Gamma}_N$ is bounded and many classical results, for instance those of Bai and Silverstein [48, 1] can be applied to \tilde{S}_N . Considering this fact, we state and prove below Proposition **II.11** and Theorem **II.12** which are the counterparts of Proposition **II.3** and Theorem **II.4**.

A2(b) Given a sequence of $N \times N$ positive semidefinite deterministic matrices Γ_N , the following properties hold:

$$\lambda_{\max}(\Gamma_N) = 1 \quad \forall N \geq 1 \quad \text{and} \quad \mu^{\Gamma_N} \xrightarrow[N \rightarrow \infty]{\mathcal{D}} \delta_0.$$

Proposition II.11. Let S_N be a $N \times N$ matrix given by (2.1) and assume that **A1** and **A2(b)** hold. Then

$$\lambda_{\max}(S_N) \xrightarrow[N, n \rightarrow \infty]{\mathcal{P}} 1. \quad (2.30)$$

If moreover either the random variables $Z_{ij}^{(N)}$ are standard (real or complex) Gaussian or Assumption **A3** holds, then the above convergence holds almost surely.

Theorem II.12. Let S_N be a $N \times N$ matrix given by (2.1) and assume that **A1**, **A2(b)** and **A4** hold. Assume moreover that one of the following conditions is satisfied:

- (i) Assumption **A5** holds,
- (ii) The random variables $Z_{ij}^{(N)}$ are standard complex Gaussian,
- (iii) The random variables $Z_{ij}^{(N)}$ are standard real Gaussian and matrices Γ_N are real symmetric,

then

$$\sqrt{n} \left(\lambda_{\max}(S_N) - 1 - \frac{1}{n} \sum_{k=2}^N \frac{\lambda_k(\Gamma_N)}{1 - \lambda_k(\Gamma_N)} \right) \xrightarrow[N, n \rightarrow \infty]{\mathcal{D}} \mathcal{N}(0, \sigma^2), \quad (2.31)$$

where $\sigma^2 = \mathbb{E}|Z_{1,1}^{(1)}|^4 - 1$.

In order to prove Proposition **II.3** and Theorem **II.4**, we only need to apply the above theorems to \tilde{S}_N .

2.5.3 Proof of Proposition **II.11**

We first prove the theorem under assumption **A3**. We first establish that

$$\overline{\lim}_{N, n \rightarrow \infty} \lambda_{\max}(\underline{S}_N) \leq 1 \text{ a.s.} \quad (2.32)$$

Recall the definition of the set B_N in (2.28). Due to the spectrum confinement property (2.27), we only need to prove that for any $\varepsilon > 0$, the interval $[1 + \varepsilon, +\infty)$ eventually stays outside the support of $\underline{\mu}_N = \underline{\mu}(r_N, \Gamma_N)$. Relying on the characterization of a point x outside $\text{supp}(\underline{\mu}_N)$, this will be the consequence of the following property

$$\forall x \in [1 + \varepsilon, \infty), \quad \exists y \in B_N, \quad x = x_N(y) \quad \text{and} \quad x'_N(y) < 0$$

that we now prove.

Since $\lambda_{\max}(\Gamma_N) = 1$ under **A2(b)**, notice that $(0, 1) \subset B_N$. Consider a real number η such that

$$\eta \in \left(\frac{1}{1 + \varepsilon}, 1 \right) .$$

For $s \leq 1$, we have $|1 - s\eta| \geq 1 - \eta > 0$, therefore by the definition (2.29) of x_N , and by the fact that $\mu^{\Gamma_N} \xrightarrow{N \rightarrow \infty} \delta_0$, we have

$$x_N(\eta) \xrightarrow{N, n \rightarrow \infty} \frac{1}{\eta} < 1 + \varepsilon \quad \text{and} \quad x'_N(\eta) \xrightarrow{N, n \rightarrow \infty} -\frac{1}{\eta^2} < 0 .$$

So for N large enough, we have $x_N(\eta) < 1 + \varepsilon$, and $x'_N(\eta) < 0$. For such N 's, note that x_N is continuous on $(0, \eta)$ and that $x_N(y) \rightarrow +\infty$ as $y \rightarrow 0^+$. We have proved so far that $[1 + \varepsilon, \infty) \subset x_N((0, \eta))$. Notice finally that x'_N is increasing on $(0, \eta)$, in particular $x'_N(y) \leq x'_N(\eta) < 0$ for all $y \in (0, \eta)$. Therefore $x_N((0, \eta))$ and thus $[1 + \varepsilon, \infty)$ eventually lie outside the support of $\underline{\mu}_N$. Equation (2.32) is established.

We now prove that

$$\lim_{N, n \rightarrow \infty} \lambda_{\max}(\underline{S}_N) \geq 1, \text{ a.s.} \tag{2.33}$$

by an exact separation argument.

As $\underline{\mu} = \delta_0$, we have $1/m_{\underline{\mu}}(a) = a$, $1/m_{\underline{\mu}}(b) = b$ for any $a, b > 0$. We intend to find some constant interval of the form $[a, 1 - \varepsilon]$, for small $\varepsilon > 0$ which separates the eigenvalues of matrix Γ_N into two non-empty parts. This is not always possible because even if $\lambda_{\max}(\Gamma_N) = 1$ and $\mu^{\Gamma_N} \rightarrow \delta_0$, there might be some intermediate eigenvalues among the $(\lambda_i(\Gamma_N))$'s for $i \geq 2$ eventually lying in $(0, 1)$. In order to circumvent this issue, we introduce the auxiliary matrices

$$\hat{\underline{S}}_N := \frac{1}{n} Z_N^* \Theta_N Z_N \quad \text{and} \quad \hat{S}_N := \frac{1}{n} \Theta_N^{\frac{1}{2}} Z_N Z_N^* \Theta_N^{\frac{1}{2}} ,$$

where Θ_N is obtained from the spectral decomposition of Γ_N as:

$$\Theta_N := U_N \text{diag}(1, 0, \dots) U_N^* \quad \text{where} \quad \Gamma_N = U_N \text{diag}(1, \lambda_2(\Gamma_N), \dots) U_N^* .$$

Using [50, Theorems 4.1 and 4.2], we conclude that for any $0 < \varepsilon < 1/2$, the interval $[\varepsilon, 1 - \varepsilon]$ is eventually outside the support of probability $\underline{\mu}(r_N, \Theta_N)$, obtained from (2.25) with parameters r_N and Θ_N . Notice in particular that

$$\lambda_{\max}(\Theta_N) = 1 \quad \text{and} \quad \lambda_i(\Theta_N) = 0 \quad \text{for} \quad i = 2 : N .$$

Applying [2, Theorem 1.2] to $\hat{\underline{S}}_N$ with separating interval $[\varepsilon, 1 - \varepsilon]$ for arbitrary $\varepsilon \in (0, 1/2)$, we conclude that almost surely, $\lambda_{\max}(\hat{\underline{S}}_N) > 1 - \varepsilon$ for N large enough. We have proved so far that

$$\underline{\lim}_{N, n \rightarrow \infty} \lambda_{\max}(\hat{\underline{S}}_N) \geq 1$$

almost surely. Now, since

$$\underline{S}_N - \hat{\underline{S}}_N = \frac{1}{n} Z_N^* U_N \text{diag}(0, \lambda_2(\underline{S}_N), \dots, \lambda_N(\underline{S}_N)) U_N^* Z_N$$

is nonnegative definite, we have $\underline{\lim}_{N, n \rightarrow \infty} \lambda_{\max}(\underline{S}_N) \geq \underline{\lim}_{N, n \rightarrow \infty} \lambda_{\max}(\hat{\underline{S}}_N) \geq 1$. Therefore Proposition II.11 with assumption A3 is proved.

As a byproduct of the above proof, we can easily prove $\lambda_{\max}(S_N) \xrightarrow{\mathcal{P}} 1$ without imposing A3. Suppose that $S_N = \frac{1}{n} \Gamma_N^{1/2} Z_N Z_N^* \Gamma_N^{1/2}$ satisfies A1 and A2(b) and construct $\check{Z} = (\check{Z}_{i,j})_{i,j \geq 1}$ with $\check{Z}_{i,j}$ i.i.d random variables identically distributed as the entries of Z_N . Let $S'_N = \frac{1}{n} \Gamma_N^{1/2} Z'_N Z_N^* \Gamma_N^{1/2}$ with Z'_N the top-left $N \times n$ submatrix of \check{Z} . Then according to the above proof, $\lambda_{\max}(S'_N)$ converges to 1 almost surely, hence in probability. Since $\lambda_{\max}(S_N)$ and $\lambda_{\max}(S'_N)$ have the same distribution, we have also $\lambda_{\max}(S_N) \xrightarrow{\mathcal{P}} 1$.

Finally, we prove that if the entries $Z_{i,j}^{(N)}$ are i.d standard Gaussian variables, and i.i.d for all $1 \leq i, j \leq N$, the convergence (2.30) holds almost surely without the need of assumption A3. This mainly relies on a concentration argument. Recall that we already have $\lambda_{\max}(S_N) \xrightarrow{\mathcal{P}} 1$. Using [19, Theorem 5.6], we prove the following concentration inequality: for all $N \geq 1$ and all $\varepsilon > 0$,

$$\mathbb{P}(|\sqrt{\lambda_{\max}(S_N)} - \mathbb{E}\sqrt{\lambda_{\max}(S_N)}| > \varepsilon) < 2e^{-CN\varepsilon^2} \quad (2.34)$$

where $C > 0$ is a proper fixed constant. Indeed it suffices to show that the function s defined by

$$s : Z_N \mapsto \sqrt{\lambda_{\max}(S_N)} = \sqrt{\lambda_{\max}\left(\frac{1}{n}\Gamma_N^{\frac{1}{2}}Z_NZ_N^*\Gamma_N^{\frac{1}{2}}\right)}$$

is $\frac{1}{\sqrt{CN}}$ -lipschitz, where we consider the $N \times n$ matrix Z_N as a vector in Euclidean space \mathbb{R}^{Nn} when $Z_{i,j}^{(N)}$ are real Gaussian, and in \mathbb{R}^{2Nn} when the entries are complex Gaussian. Note that the Euclidean norm of the vector Z_N is the same as the Frobenius norm $\|Z_N\|_F$ of the matrix Z_N . So for any two matrices Z_N and \hat{Z}_N , we have

$$\begin{aligned} |s(Z_N) - s(\hat{Z}_N)| &= \frac{1}{\sqrt{n}} \left| \|\Gamma_N^{1/2}Z_N\| - \|\Gamma_N^{1/2}\hat{Z}_N\| \right| \leq \frac{1}{\sqrt{n}} \|\Gamma_N^{1/2}(Z_N - \hat{Z}_N)\| \\ &\leq \frac{1}{\sqrt{n}} \|\Gamma_N^{1/2}(Z_N - \hat{Z}_N)\|_F \stackrel{(a)}{\leq} \frac{1}{\sqrt{n}} \|\Gamma_N^{1/2}\| \|Z_N - \hat{Z}_N\|_F \stackrel{(b)}{=} \frac{1}{\sqrt{n}} \|Z_N - \hat{Z}_N\|_F, \end{aligned}$$

where (a) follows from the Frobenius norm inequality $\|AB\|_F \leq \|A\| \|B\|_F$, and (b) from the fact that $\|\Gamma_N^{1/2}\| = \sqrt{\lambda_{\max}(\Gamma_N)} = 1$. Thus s is $1/\sqrt{n}$ -lipschitz, and the concentration inequality (2.34) is proved. Using Borel-Cantelli lemma, we have

$$\sqrt{\lambda_{\max}(S_N)} - \mathbb{E}\sqrt{\lambda_{\max}(S_N)} \rightarrow 0 \text{ a.s.} \quad (2.35)$$

Together with $\sqrt{\lambda_{\max}(S_N)} \xrightarrow{\mathcal{P}} 1$, we then obtain that $\mathbb{E}\sqrt{\lambda_{\max}(S_N)} \rightarrow 1$. By (2.35) again, it follows that

$$\lambda_{\max}(S_N) \xrightarrow[N, n \rightarrow \infty]{} 1 \text{ a.s.}$$

The proof of Proposition II.11 is complete.

2.5.4 Proof of Theorem II.12

We first prove the fluctuation of $\lambda_{\max}(S_N)$ under **A1**, **A2(b)**, **A4** and **A5**. Under these assumptions, Γ_N is of the form

$$\Gamma_N = \begin{pmatrix} 1 & 0 \\ 0 & \Gamma_{N-1} \end{pmatrix},$$

where (Γ_{N-1}) is a sequence of semidefinite positive Hermitian matrices satisfying $\mu^{\Gamma_{N-1}} \xrightarrow{\mathcal{D}} \delta_0$, and

$$\overline{\lim}_{N \rightarrow \infty} \lambda_{\max}(\Gamma_{N-1}) = \overline{\lim}_{N \rightarrow \infty} \lambda_2(\Gamma_N) = \overline{\lim}_{N \rightarrow \infty} \frac{\lambda_2(\Gamma_N)}{\lambda_1(\Gamma_N)} < 1$$

by assumption **A4**. We set $d = \overline{\lim}_{N \rightarrow \infty} \lambda_{\max}(\Gamma_{N-1})$. For convenience, in this section we omit all the subscript N of matrices, for example we write $S = n^{-1} \Gamma^{\frac{1}{2}} Z Z^* \Gamma^{\frac{1}{2}}$. In the following of this section we write $\lambda_{\max}(S_N)$ as λ_{\max} if it does not cause any ambiguity.

Recall the submatrix notations introduced in (2.16) and consider the following block decomposition of matrix S :

$$S = \begin{pmatrix} S_{1,1} & S_{1,2:N} \\ S_{2:N,1} & S_{2:N,2:N} \end{pmatrix} = \frac{1}{n} \begin{pmatrix} Z_{1,\cdot} Z_{1,\cdot}^* & Z_{1,\cdot} Z_{2:N,\cdot}^* \Gamma^{\frac{1}{2}} \\ \Gamma^{\frac{1}{2}} Z_{2:N,\cdot} Z_{1,\cdot}^* & \Gamma^{\frac{1}{2}} Z_{2:N,\cdot} Z_{2:N,\cdot}^* \Gamma^{\frac{1}{2}} \end{pmatrix}. \quad (2.36)$$

Denote

$$\underline{S}_{2:N,2:N} := \frac{1}{n} Z_{2:N,\cdot}^* \Gamma Z_{2:N,\cdot} \quad \text{and} \quad r_N = \frac{N-1}{n}.$$

Analog to $\underline{\mu}(r_N, \Gamma_N)$ defined in (2.25), we define the probability measure $\underline{\mu}(r_N, \Gamma_{N-1})$ whose Cauchy-Stieltjes transform $\underline{\mathbf{m}}$ satisfies the equation

$$z = \frac{1}{\underline{\mathbf{m}}(z)} + \frac{1}{n} \sum_{k=2}^N \frac{\lambda_k(\Gamma_N)}{1 - \lambda_k(\Gamma_N) \underline{\mathbf{m}}(z)}, \quad \forall z \in \mathbb{C}^+.$$

Also, for all $y \in \mathbf{B}_N := \{y \in \mathbb{R} : y \neq 0, y^{-1} \neq \lambda_k(\Gamma_N), \forall k = 2, \dots, N\}$, we define

$$\mathbf{x}_N(y) := \frac{1}{y} + \frac{1}{n} \sum_{k=2}^N \frac{\lambda_k(\Gamma_N)}{1 - \lambda_k(\Gamma_N) y} = \frac{1}{y} + r_N \int \frac{s}{1 - sy} d\mu^{\Gamma_{N-1}}(s). \quad (2.37)$$

Consider in particular

$$\theta_N = \mathbf{x}_N(1) = 1 + \frac{1}{n} \sum_{k=2}^N \frac{\lambda_k(\Gamma_N)}{1 - \lambda_k(\Gamma_N)}. \quad (2.38)$$

Let $\varepsilon > 0$ be small enough. Thanks to the assumption $\mu^{\Gamma_{N-1}} \xrightarrow{\mathcal{D}} \delta_0$ and $d = \overline{\lim} \lambda_{\max}(\Gamma_{N-1}) <$

1, one can adapt the first part of the proof of Proposition II.11 to obtain that eventually

$$\sup \text{supp } \underline{\mu}(r_N, \Gamma_{N-1}) < d + \varepsilon .$$

Let $\varepsilon < \frac{1-d}{2}$ so that $d + \varepsilon < 1 - \varepsilon$ and consider the family of events Ω_N defined as

$$\Omega_N := \{\lambda_{\max}(S_{2:N,2:N}) < d + \varepsilon < 1 - \varepsilon < \lambda_{\max}\}. \quad (2.39)$$

According to the spectrum confinement property [1, Theorem 1.1] and to Proposition II.11, one has

$$\mathbb{P}(\Omega_N) \xrightarrow[N, n \rightarrow \infty]{} 1 .$$

In particular, for any sequence of events A_N , we have $\mathbb{P}(A_N) - \mathbb{P}(A_N \cap \Omega_N) \rightarrow_{N \rightarrow \infty} 0$, which can be written

$$\mathbb{P}(A_N) \doteq \mathbb{P}_\Omega(A_N)$$

if one writes $\mathbb{P}_\Omega(\cdot)$ for $\mathbb{P}(\cdot \cap \Omega_N)$ and recall the notation $x \doteq y$ for $x - y \rightarrow 0$. Hence, with no loss of generality, we will assume below that Ω_N holds.

Let $\lambda_{\max} \in \Omega_N$. Using the bloc decomposition (2.36) of S together with the determinantal formula based on Schur complements (see for instance [30, Section 0.8.5]), the eigenvalue λ_{\max} satisfies the equation:

$$\begin{aligned} & \det(\lambda_{\max} I - S) \\ &= (\lambda_{\max} I - S_{1,1} - S_{1,2:N}(\lambda_{\max} I - S_{2:N,2:N})^{-1} S_{2:N,1}) \det(\lambda_{\max} I - S_{2:N,2:N}) = 0 . \end{aligned}$$

Since $\det(\lambda_{\max} I - S_{2:N,2:N}) \neq 0$ on Ω_N , we have

$$\begin{aligned} \lambda_{\max} &= S_{1,1} + S_{1,2:N}(\lambda_{\max} I - S_{2:N,2:N})^{-1} S_{2:N,1} , \\ &= \frac{Z_{1,\cdot} Z_{1,\cdot}^*}{n} + \frac{Z_{1,\cdot} Z_{2:N,\cdot}^* \Gamma^{\frac{1}{2}}}{n} \left(\lambda_{\max} I - \frac{1}{n} \Gamma^{\frac{1}{2}} Z_{2:N,\cdot} Z_{2:N,\cdot}^* \Gamma^{\frac{1}{2}} \right) \frac{\Gamma^{\frac{1}{2}} Z_{2:N,\cdot} Z_{1,\cdot}^*}{n} , \\ &= \frac{Z_{1,\cdot}}{n} \left(I + A^* (\lambda_{\max} I - A A^*) A \right) Z_{1,\cdot}^* . \quad \left(A = n^{-1/2} \Gamma^{\frac{1}{2}} Z_{2:N,\cdot} \right) \quad (2.40) \end{aligned}$$

Using the equality $I + A^*(\lambda I - AA^*)^{-1}A = \lambda(\lambda I - A^*A)^{-1}$ for all scalar λ and all matrix A such that $\lambda I - AA^*$ and $\lambda I - A^*A$ are invertible, the equation (2.40) is equivalent to

$$1 = \frac{1}{n} Z_{1,\cdot} (\lambda_{\max} I - \underline{S}_{2:N,2:N})^{-1} Z_{1,\cdot}^* . \quad (2.41)$$

As $\theta_N = x_N(1) \geq 1$ lies outside the support of $\underline{\mu}(r_N, \Gamma_{N-1})$ for large N , [50, Theorem 4.2] yields

$$\underline{\mathbf{m}}(\theta_N) = \underline{\mathbf{m}}(x_N(1)) = 1 .$$

This can be regarded as a “deterministic” version of (2.41), which indicates that λ_{\max} and θ_N are comparable.

In order to prove the Gaussian fluctuations of λ_{\max} , we need to prove that for all $b \in \mathbb{R}$

$$\mathbb{P}(\lambda_{\max} \leq \eta_N) \xrightarrow[N, n \rightarrow \infty]{} \Phi(b, \sigma) \quad (2.42)$$

where

$$\eta_N := \theta_N + \frac{b}{\sqrt{n}}, \quad \Phi(x, \sigma) := \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2\sigma^2}} dt \quad \text{and} \quad \sigma^2 = \mathbb{E} \left| Z_{1,1}^{(1)} \right|^4 - 1 .$$

Note that on Ω_N the function

$$\Upsilon(\lambda) := \frac{1}{n} Z_{1,\cdot} (\lambda I - \underline{S}_{2:N,2:N})^{-1} Z_{1,\cdot}^* .$$

is decreasing on $(d + \varepsilon, +\infty)$. Let N large enough so that $\eta_N > d + \varepsilon$.

Taking into account the fact that $\Upsilon(\lambda_{\max}) = 1$ due to (2.41), we have

$$\mathbb{P}_{\Omega}(\lambda_{\max} \leq \eta_N) = \mathbb{P}_{\Omega}(\Upsilon(\eta_N) \leq 1) = \mathbb{P}_{\Omega}(\sqrt{n}(\Upsilon(\eta_N) - \underline{\mathbf{m}}(\eta_N)) \leq \sqrt{n}(1 - \underline{\mathbf{m}}(\eta_N))) . \quad (2.43)$$

We first prove that

$$\sqrt{n}(1 - \underline{\mathbf{m}}(\eta_N)) = b + o(1) . \quad (2.44)$$

Taking into account the fact that $\underline{\mathbf{m}}(\theta_N) = 1$ and performing a Taylor expansion on $\underline{\mathbf{m}}$

around θ_N yields

$$\sqrt{n}(1 - \underline{\mathbf{m}}(\eta_N)) = \sqrt{n}(\underline{\mathbf{m}}(\theta_N) - \underline{\mathbf{m}}(\eta_N)) = -b \underline{\mathbf{m}}'(\theta_N) - \underline{\mathbf{m}}''(\xi_N) \frac{b^2}{\sqrt{n}}$$

where ξ_N is between $\theta_N = x_N(1)$ and η_N . The assumptions $\mu^{\Gamma_{N-1}} \xrightarrow{\mathcal{D}} \delta_0$ and $d = \overline{\lim} \lambda_{\max}(\Gamma_{N-1}) < 1$ yield

$$\theta_N, \eta_N \xrightarrow[N, n \rightarrow \infty]{} 1 .$$

Similarly, one proves that $x'_N(1) \xrightarrow[N, n \rightarrow \infty]{} -1$. By [50, Theorem 4.2], equality $\underline{\mathbf{m}}(x_N(y)) = y$ holds for any $y \notin \text{supp } \mu^{r_N, \Gamma_{N-1}}$. Differentiating, we get

$$\underline{\mathbf{m}}'(x_N(y)) x'_N(y) = 1 \quad \text{and} \quad \underline{\mathbf{m}}'(x_N(1)) = \frac{1}{x'_N(1)} \xrightarrow[N, n \rightarrow \infty]{} -1 .$$

Finally, for large N , we have $\sup_N \text{supp } \underline{\mu}(r_N, \Gamma_{N-1}) < d + \varepsilon < 1 - \varepsilon < \min(\eta_N, \theta_N)$ which implies

$$|\underline{\mathbf{m}}''(\xi_N)| = 2 \left| \int \frac{d\underline{\mu}(r_N, \Gamma_{N-1})(s)}{(\xi_N - s)^3} \right| \leq \frac{2}{(1 - d - 2\varepsilon)^3} .$$

Plugging this into the Taylor expansion finally yields (2.44).

We now go back to (2.43) and handle the quantity $\sqrt{n}(\Upsilon(\eta_N) - \underline{\mathbf{m}}(\eta_N))$. More precisely, we prove in the sequel that

$$\sqrt{n}(\Upsilon(\eta_N) - \underline{\mathbf{m}}(\eta_N)) = \sqrt{n} \left(\frac{1}{n} Z_{1, \cdot} Z_{1, \cdot}^* - 1 \right) + o_P(1) . \quad (2.45)$$

In order to proceed, we need the following estimates, valid under the assumptions of Theorem II.12.

Proposition II.13. *Assume that A1, A2(b), A4 and A5 hold, then*

$$(a) \quad \sqrt{n} \left(\frac{1}{n} \text{tr}(\eta_N I - \underline{\mathcal{S}}_{2:N, 2:N})^{-1} - \underline{\mathbf{m}}(\eta_N) \right) \xrightarrow[N, n \rightarrow \infty]{\mathcal{P}} 0 ,$$

$$(b) \quad \frac{\sqrt{n}}{\eta_N} \left(\frac{1}{n} Z_{1, \cdot} (\eta_N I - \underline{\mathcal{S}}_{2:N, 2:N})^{-1} \underline{\mathcal{S}}_{2:N, 2:N} Z_{1, \cdot}^* - \frac{1}{n} \text{tr}(\eta_N I - \underline{\mathcal{S}}_{2:N, 2:N})^{-1} \underline{\mathcal{S}}_{2:N, 2:N} \right) \xrightarrow[N, n \rightarrow \infty]{\mathcal{P}} 0 .$$

Proof of Proposition II.13 is postponed to Section 2.5.4.1. We have

$$\begin{aligned}
 & \sqrt{n}(\Upsilon(\eta_N) - \underline{\mathbf{m}}(\eta_N)) \\
 &= \sqrt{n} \left(\Upsilon(\eta_N) - \frac{1}{n} \operatorname{tr}(\eta_N I - \underline{\mathcal{S}}_{2:N,2:N})^{-1} + \frac{1}{n} \operatorname{tr}(\eta_N I - \underline{\mathcal{S}}_{2:N,2:N})^{-1} - \underline{\mathbf{m}}(\eta_N) \right) \\
 &= \sqrt{n} \left(\Upsilon(\eta_N) - \frac{1}{n} \operatorname{tr}(\eta_N I - \underline{\mathcal{S}}_{2:N,2:N})^{-1} \right) + o_P(1)
 \end{aligned} \tag{2.46}$$

by the first part of Proposition II.13. We now apply the resolvent identity $A^{-1} - B^{-1} = A^{-1}(B - A)B^{-1}$ to $A = \eta_N - \underline{\mathcal{S}}_{2:N,2:N}$ and $B = \eta_N I$ and obtain

$$\begin{aligned}
 & \sqrt{n} \left(\Upsilon(\eta_N) - \frac{1}{n} \operatorname{tr}(\eta_N I - \underline{\mathcal{S}}_{2:N,2:N})^{-1} \right) \\
 &= \frac{\sqrt{n}}{\eta_N} \left(\frac{1}{n} Z_{1,\cdot} (\eta_N I - \underline{\mathcal{S}}_{2:N,2:N})^{-1} \underline{\mathcal{S}}_{2:N,2:N} Z_{1,\cdot}^* - \frac{1}{n} \operatorname{tr}(\eta_N I - \underline{\mathcal{S}}_{2:N,2:N})^{-1} \underline{\mathcal{S}}_{2:N,2:N} \right) \\
 &\quad + \frac{\sqrt{n}}{\eta_N} \left(\frac{1}{n} Z_{1,\cdot} Z_{1,\cdot}^* - 1 \right) \\
 &= \frac{\sqrt{n}}{\eta_N} \left(\frac{1}{n} Z_{1,\cdot} Z_{1,\cdot}^* - 1 \right) + o_P(1)
 \end{aligned} \tag{2.47}$$

where the last equality follows from the second estimate of Proposition II.13. Notice that by the standard Central Limit theorem,

$$\sqrt{n} \left(\frac{1}{n} Z_{1,\cdot} Z_{1,\cdot}^* - 1 \right) = \sqrt{n} \left(\frac{1}{n} \sum_{j=1}^n |Z_{1,j}|^2 - 1 \right) \xrightarrow[n \rightarrow \infty]{\mathcal{D}} \mathcal{N}(0, \operatorname{Var} |Z_{1,1}|^2)$$

where $\operatorname{Var} |Z_{1,1}|^2 = \mathbb{E}|Z_{1,1}|^4 - 1$. Since $\eta_N \rightarrow 1$, one has

$$\frac{\sqrt{n}}{\eta_N} \left(\frac{1}{n} Z_{1,\cdot} Z_{1,\cdot}^* - 1 \right) = \sqrt{n} \left(\frac{1}{n} Z_{1,\cdot} Z_{1,\cdot}^* - 1 \right) + o_P(1).$$

Plugging this last estimate into (2.47) and (2.46) finally yields (2.45). We can now conclude

the proof of the CLT:

$$\begin{aligned}
\mathbb{P}(\lambda_{\max} \leq \eta_N) &\doteq \mathbb{P}_\Omega(\lambda_{\max} \leq \eta_N) \\
&\stackrel{(a)}{=} \mathbb{P}_\Omega(\sqrt{n}(\Upsilon(\eta_N) - \underline{\mathbf{m}}(\eta_N)) \leq b + o(1)) \\
&\stackrel{(b)}{=} \mathbb{P}_\Omega\left(\sqrt{n}\left(\frac{1}{n}Z_{1,\cdot}Z_{1,\cdot}^* - 1\right) + o_P(1) \leq b\right) \\
&\doteq \mathbb{P}\left(\sqrt{n}\left(\frac{1}{n}Z_{1,\cdot}Z_{1,\cdot}^* - 1\right) + o_P(1) \leq b\right) \tag{2.48}
\end{aligned}$$

where (a) follows from (2.43) and (2.44) and (b) follows from (2.45). We can now get rid of the term $o_P(1)$ in (2.48) by Slutsky's theorem and finally obtain the desired result:

$$\mathbb{P}(\sqrt{n}(\lambda_{\max} - \theta_N) \leq b) = \mathbb{P}(\lambda_{\max} \leq \eta_N) \xrightarrow{N, n \rightarrow \infty} \Phi(b, \sigma), \quad \sigma^2 = \mathbb{E}|Z_{1,1}|^4 - 1.$$

This completes the proof of Theorem II.12 under condition (i).

Assume now that $Z_{ij}^{(N)} \sim \mathcal{N}_{\mathbb{C}}(0, 1)$ and consider the eigen-decomposition $\Gamma_N = U_N D_N U_N^*$, where U_N is unitary and $D_N = \text{diag}(\lambda_1(\Gamma_N), \dots, \lambda_N(\Gamma_N))$. Then S_N can be written as

$$S_N = \frac{1}{n} U_N D_N^{\frac{1}{2}} (U_N^* Z_N) (Z_N^* U_N) D_N^{\frac{1}{2}} U_N^* = \frac{1}{n} U_N D_N^{\frac{1}{2}} \tilde{Z}_N \tilde{Z}_N^* D_N^{\frac{1}{2}} U_N^* \quad \text{where} \quad \tilde{Z}_N = U_N^* Z_N$$

and has the same eigenvalues as the matrix $R_N = n^{-1} D_N^{\frac{1}{2}} \tilde{Z}_N \tilde{Z}_N^* D_N^{\frac{1}{2}}$. It remains to notice that \tilde{Z}_N has i.i.d. $\mathcal{N}_{\mathbb{C}}(0, 1)$ entries. In particular, R_N satisfies **A1**, **A2(b)**, **A4** and **A5**, and the desired result follow for S_N . Theorem II.12 is established under condition (ii).

Assume now that $Z_{ij}^{(N)} \sim \mathcal{N}_{\mathbb{R}}(0, 1)$ and that Γ_N is real symmetric. In this case, Γ_N 's eigen-decomposition writes $\Gamma_N = O_N D_N O_N^\top$, where matrix O_N is orthogonal. It remains to notice that $O_N^\top Z_N$ has i.i.d $\mathcal{N}_{\mathbb{R}}(0, 1)$ entries and to proceed as in the complex case to prove Theorem II.12 under condition (iii).

Proof of Theorem II.12 is completed.

2.5.4.1 Proof of Proposition II.13

We first establish item (a). Denote by

$$\Delta_N(x) = \frac{1}{n} \operatorname{tr} (xI - \underline{S}_{2:N,2:N})^{-1} - \underline{\mathbf{m}}(x) .$$

We will first establish that $n\Delta_n(\eta_N)$ is tight and then, as an easy consequence, we will deduce the desired convergence: $\sqrt{n}\Delta_N(\eta_N) \xrightarrow[N, n \rightarrow \infty]{\mathcal{P}} 0$.

If $x \geq 1 - \varepsilon$ is fixed with $1 - \varepsilon > d + \varepsilon$, then the tightness of $n\Delta_N(x)$ is a consequence of Bai and Silverstein's peripheral results of their CLT paper [3], see also [4, Chapter 9]. In fact,

$$\Delta_N(x) = \int f(x, \lambda) \mu^{\underline{S}_{2:N,2:N}}(d\lambda) - \int f(x, \lambda) \underline{\mu}(r_N, \Gamma_{N-1})(d\lambda) \quad \text{where} \quad f(x, \lambda) = \frac{1}{x - \lambda} .$$

Notice that for any $x \geq 1 - \varepsilon$, $\lambda \mapsto f(x, \lambda)$ is analytic in a neighbourhood of $[0, d + \varepsilon]$ which contains the support of $\underline{\mu}(r_N, \Gamma_{N-1})$. According to [4, Theorem 9.10(1)] and to the remark at the end of page 265 in [4] which tightens the interval where the function $f(x, \cdot)$ needs to be analytic, we immediatly obtain the tightness of $(n\Delta_N(x))$.

The case where $x = \eta_N \geq 1 - \varepsilon$ for N large necessitates some adaptation. We closely follow [4, Chapter 9]. Denote by

$$M_N(z) = n \left(m^{\underline{S}_{2:N,2:N}}(z) - \underline{\mathbf{m}}(z) \right)$$

and by \mathcal{C}^+ the contour defined by ($\delta, u > 0$ fixed)

$$\mathcal{C}^+ = \mathcal{C}_\ell \cup \mathcal{C}_{\text{up}} \cup \mathcal{C}_r \quad \text{where} \quad \begin{cases} \mathcal{C}_\ell &= \{z = (-\delta, y), y \in [0, u]\} \\ \mathcal{C}_{\text{up}} &= \{z = (x, u), x \in [-\delta, d + \varepsilon]\} \\ \mathcal{C}_r &= \{z = (d + \varepsilon, y), y \in [0, u]\} \end{cases} .$$

Consider the truncated version $\hat{M}_N(z)$ of $M_N(z)$ as defined in [4, (9.8.2)] then

$$\int_{\mathcal{C}} \frac{1}{\eta_N - z} \left(\hat{M}_N(z) - M_N(z) \right) dz \xrightarrow[N, n \rightarrow \infty]{\text{a.s.}} 0 \quad \text{where} \quad \mathcal{C} = \mathcal{C}^+ \cup \overline{\mathcal{C}^+}$$

and $\{\hat{M}_N(\cdot)\}$ forms a tight sequence on \mathcal{C} . Consider now the mapping

$$\Gamma_N : \hat{M}_N(\cdot) \longmapsto \frac{1}{2i\pi} \int_{\mathcal{C}} \frac{1}{\eta_N - z} \hat{M}_N(z) dz .$$

Γ_N is a continuous mapping from $C(\mathcal{C}, \mathbb{R}^2)$ to \mathbb{C} . Applying Prohorov's theorem (see for instance [36, Theorem 16.3]) and the continuous mapping theorem [36, Theorem 4.27], we conclude that $\Gamma_N(\hat{M}_N)$ is tight. It remains to notice that

$$n\Delta_N(\eta_N) = \Gamma_N(\hat{M}_N) + \underbrace{\left(\Gamma_N(\hat{M}_N) - \Gamma_N(M_N) \right)}_{\rightarrow 0 \text{ a.s.}}$$

to conclude that $n\Delta_N(\eta_N)$ is tight. Now let $\delta > 0$ be fixed, then

$$\mathbb{P}(|\sqrt{n}\Delta_N(\eta_N)| > \delta) = \mathbb{P}(|n\Delta_N(\eta_N)| > \sqrt{n}\delta) \xrightarrow[N, n \rightarrow \infty]{} 0$$

by tightness, hence the convergence of $\sqrt{n}\Delta_N(\eta_N)$ to zero in probability. Part (a) of Proposition II.13 is proved.

We now prove part (b) of Proposition II.13 and rely on the lemma on quadratic forms [4, Lemma B.26]. Denote by

$$P_N = \sqrt{n} \left(\frac{1}{n} Z_{1,\cdot} (\eta_N I - \underline{S}_{2:N,2:N})^{-1} \underline{S}_{2:N,2:N} Z_{1,\cdot}^* - \frac{1}{n} \text{tr} \{ (\eta_N I - \underline{S}_{2:N,2:N})^{-1} \underline{S}_{2:N,2:N} \} \right)$$

and apply the lemma on quadratic forms with $p = 2$: There exists a constant C such that

$$\mathbb{E}_{\mathbf{Z}_{2:N,\cdot}} (|P_N|^2) \leq \frac{C}{n} \mathbb{E}(|Z_{1,1}^{(1)}|^4) \text{tr} \{ (\eta_N I - \underline{S}_{2:N,2:N})^{-2} \underline{S}_{2:N,2:N}^2 \} .$$

Taking into account the facts that

$$\lim_{n \rightarrow \infty} \eta_N = 1 , \quad \overline{\lim}_N \lambda_{\max}(\underline{S}_{2:N,2:N}) \stackrel{\mathcal{P}}{\leq} d \quad \text{and} \quad \mu^{\underline{S}_{2:N,2:N}} \xrightarrow[N, n \rightarrow \infty]{\mathcal{D}} \delta_0 \text{ a.s.} ,$$

we obtain that

$$\frac{1}{n} \operatorname{tr} \{ (\eta_N - \underline{S}_{2:N,2:N})^{-2} \underline{S}_{2:N,2:N}^2 \} = \int \frac{s^2}{(\eta_N - s)^2} \mu^{\underline{S}_{2:N,2:N}}(ds) \xrightarrow[N, n \rightarrow \infty]{\mathcal{P}} 0.$$

Thus $\mathbb{E}_{Z_{2:N,\cdot}}(|P_N|^2)$ converges to zero in probability, from which we deduce that for $\delta > 0$,

$$\mathbb{E}_{Z_{2:N,\cdot}}(\mathbb{1}_{|P_N|^2 > \delta}) \leq \frac{1}{\delta} \mathbb{E}_{Z_{2:N,\cdot}}(|P_N|^2 \mathbb{1}_{|P_N|^2 > \delta}) \xrightarrow[N, n \rightarrow \infty]{\mathcal{P}} 0.$$

Finally

$$\mathbb{P}(|P_N|^2 > \delta) = \mathbb{E} \mathbb{E}_{Z_{2:N,\cdot}}(\mathbb{1}_{|P_N|^2 > \delta}) \xrightarrow[N, n \rightarrow \infty]{} 0,$$

which completes the proof of Proposition [II.13](#).

2.6 Proof of Theorem [II.8](#)

In order to study the spectral gap associated to the family of Toeplitz matrices and to prove Theorem [II.8](#), we follow the method used in [\[16\]](#). The main idea is to interpret the eigenvalues of the Toeplitz matrix T_N as eigenvalues of an operator \mathcal{K}_N using Widom-Shampine's Lemma, and then analyse the convergence of this operator, correctly normalized.

In this section, for $p \in [1, \infty]$, the L^p norm of a function f is denoted by $\|f\|_p$, and the $L^p \rightarrow L^p$ norm of an operator \mathcal{K} is denoted by $\|\mathcal{K}\|_p$. Recall that $\|\mathcal{K}\|_p := \sup_{\|f\|_p=1} \|\mathcal{K}f\|_p$.

2.6.1 Widom-Shampine's Lemma and convergence of operators

We first recall Widom-Shampine's Lemma, see [\[16\]](#) for a proof.

Lemma II.14 (Widom-Shampine). *Let $A = (a_{i,j})_{i,j=0}^{N-1}$ be a matrix with complex entries $a_{i,j}$, and let G be the integral operator on $L^2(0, 1)$ defined by*

$$(Gf)(x) = \int_0^1 a_{\lfloor Nx \rfloor \lfloor Ny \rfloor} f(y) dy, \quad x \in (0, 1).$$

Then a nonzero complex number λ is an eigenvalue of A of a certain algebraic multiplicity if and only if λ/N is an eigenvalue of G of the same algebraic multiplicity.

Let $c = (c_k)_{k \in \mathbb{Z}}$ be the sequence in Theorem II.8, and $\rho \in (-1, 0)$ be the index, then the function $R(h) := c_{\lfloor h \rfloor}$ is even and regularly varying and $R(k) = c_k$. By Definition II.6, $R(N) \neq 0$ for large enough $N \in \mathbb{N}$, for convenience we can suppose that $R(N) \neq 0$ for all $N \in \mathbb{N}$ without loss of generality. By Widom-Shampine's Lemma, for each N , the matrix $T_N(c)/(NR(N))$ has the same nonzero eigenvalues (with the same multiplicities) as the integral operator $\mathcal{K}_N^{(\rho)}$ defined on $L^2(0, 1)$ by

$$(\mathcal{K}_N^{(\rho)} f)(x) = \int_0^1 \frac{R(\lfloor Nx \rfloor - \lfloor Ny \rfloor)}{R(N)} f(y) dy. \quad (2.49)$$

We will prove that the operators $\mathcal{K}_N^{(\rho)}$ converge in the operator norm to the operator $\mathcal{K}^{(\rho)}$ defined on $L^2(0, 1)$ by

$$(\mathcal{K}^{(\rho)} f)(x) = \int_0^1 |x - y|^\rho f(y) dy. \quad (2.50)$$

For this we need the following Lemma II.15 which is a special case of the uniform convergence theorem of regularly varying functions.

Lemma II.15 ([14, Theorem 1.5.2]). *If R is regularly varying with index $\rho < 0$, then for every $a > 0$*

$$\sup_{x > a} \left| \frac{R(xy)}{R(y)} - x^\rho \right| \xrightarrow{y \rightarrow \infty} 0.$$

The following description of the asymptotic integral of regularly varying functions will also be useful in the sequel.

Lemma II.16 ([14, Proposition 1.5.8]). *If R is regularly varying with index $\rho > -1$, and suppose that R is locally bounded, then*

$$\int_0^y R(x) dx \sim \frac{yR(y)}{1 + \rho} \quad (y \rightarrow +\infty).$$

Recall that for an operator defined by $(\mathcal{K}f)(x) = \int_0^1 K(x, y)f(y) dy$, we have

$$\|\mathcal{K}\|_1 \leq M_1 \quad \text{and} \quad \|\mathcal{K}\|_\infty \leq M_\infty,$$

where

$$M_1 := \operatorname{ess\,sup}_{y \in [0,1]} \int_0^1 |K(x, y)| \, dx, \quad M_\infty := \operatorname{ess\,sup}_{x \in [0,1]} \int_0^1 |K(x, y)| \, dy. \quad (2.51)$$

If the kernel K is symmetric for x and y , then $M_1 = M_\infty$. In this case and if $M_1 = M_\infty < \infty$, then by the Riesz-Thorin interpolation theorem (cf. [24, Theorem 2.2.14], taking $p_0 = q_0 = 1, p_1 = q_1 = \infty$), for all $p \in [1, +\infty]$, we have

$$\|\mathcal{K}\|_p \leq M_1 = M_\infty. \quad (2.52)$$

We are now ready to prove the theorem. As mentioned above, we first prove the following convergence of operators.

Lemma II.17. *Let $\rho \in (-1, 0)$, then for any $p \in [1, \infty]$ and any $N \in \mathbb{N}$, the formulas (2.49) and (2.50) define bounded operators $\mathcal{K}_N^{(\rho)}$ and $\mathcal{K}^{(\rho)}$ on $L^p(0, 1)$. Moreover we have*

$$\lim_{N \rightarrow \infty} \|\mathcal{K}_N^{(\rho)} - \mathcal{K}^{(\rho)}\|_p = 0$$

for any $p \in [1, \infty]$.

Proof. Let $K_N^{(\rho)} : [0, 1]^2 \rightarrow \mathbb{R}$ and $K_\rho : [0, 1]^2 \rightarrow \mathbb{R}$ be the integral kernels defining respectively $\mathcal{K}_N^{(\rho)}$ and $\mathcal{K}^{(\rho)}$, that is,

$$K_N^{(\rho)}(x, y) = \frac{R(\lfloor Nx \rfloor - \lfloor Ny \rfloor)}{R(N)}, \quad K^{(\rho)}(x, y) = |x - y|^\rho.$$

Recall that since R is even, the considered kernels are symmetric and the two essential supremums in (2.51) of each kernel are equal. Moreover, for each N , $K_N^{(\rho)}$ is bounded on $[0, 1]^2$ as it takes only a finite number of values, hence

$$\operatorname{ess\,sup}_{x \in [0,1]} \int_0^1 |K_N^{(\rho)}(x, y)| \, dy = \operatorname{ess\,sup}_{y \in [0,1]} \int_0^1 |K_N^{(\rho)}(x, y)| \, dx < \infty.$$

For $\rho \in (-1, 0)$, easy calculations yield

$$\operatorname{ess\,sup}_{x \in [0,1]} \int_0^1 |x - y|^\rho \, dy = \operatorname{ess\,sup}_{y \in [0,1]} \int_0^1 |x - y|^\rho \, dx = \frac{2^{-\rho}}{(1 + \rho)} < \infty.$$

So by (2.52), for all $p \in [1, +\infty]$ we have

$$\|\mathcal{K}_N^{(\rho)}\|_p < \infty \quad \text{and} \quad \|\mathcal{K}^{(\rho)}\|_p \leq \frac{2^{-\rho}}{(1+\rho)}.$$

Also by (2.52), we have

$$\|\mathcal{K}_N^{(\rho)} - \mathcal{K}^{(\rho)}\|_p \leq \operatorname{ess\,sup}_{y \in [0,1]} \int_0^1 \left| \frac{R(\lfloor Nx \rfloor - \lfloor Ny \rfloor)}{R(N)} - |x - y|^\rho \right| dx. \quad (2.53)$$

We now prove that $\|\mathcal{K}_N^{(\rho)} - \mathcal{K}^{(\rho)}\|_p \rightarrow 0$ by showing that the right handside (RHS) of (2.53) goes to 0 as $N \rightarrow \infty$. Taking an arbitrary $\varepsilon \in (0, 1)$, we set $A_\varepsilon := \{(x, y) \in [0, 1]^2 : |x - y| > \varepsilon\}$ and for $y \in [0, 1]$, we set $A_\varepsilon(y) := \{x \in [0, 1] : (x, y) \in A_\varepsilon\} = \{x \in [0, 1] : |x - y| > \varepsilon\}$.

By the inequality

$$|x - y| - \frac{1}{N} \leq \frac{|\lfloor Nx \rfloor - \lfloor Ny \rfloor|}{N} \leq |x - y| + \frac{1}{N} \quad (2.54)$$

and the uniform continuity of the function $x \mapsto x^\rho$ on $[\frac{\varepsilon}{2}, +\infty)$, we can take $N_1 = N_1(\varepsilon) \in \mathbb{N}$ such that for $N > N_1$ and $(x, y) \in A_\varepsilon$ we have

$$\frac{|\lfloor Nx \rfloor - \lfloor Ny \rfloor|}{N} > \frac{\varepsilon}{2} \quad (2.55)$$

and

$$\left| \frac{|\lfloor Nx \rfloor - \lfloor Ny \rfloor|^\rho}{N^\rho} - |x - y|^\rho \right| < \frac{\varepsilon}{2}. \quad (2.56)$$

Then applying Lemma II.15 with $a = \varepsilon/2$, we can find $N_2 = N_2(\varepsilon) \in \mathbb{N}$ such that for $N > N_2$ and for all c satisfying $|c| \geq \varepsilon/2$, we have

$$\left| \frac{R(cN)}{R(N)} - |c|^\rho \right| < \frac{\varepsilon}{2}. \quad (2.57)$$

For all $N > \max(N_1, N_2)$ and $(x, y) \in A_\varepsilon$, let $c = \frac{\lfloor Nx \rfloor - \lfloor Ny \rfloor}{N}$ then by (2.55) we have

$|c| > \varepsilon/2$. Moreover

$$\left| \frac{R(\lfloor Nx \rfloor - \lfloor Ny \rfloor)}{R(N)} - \frac{|\lfloor Nx \rfloor - \lfloor Ny \rfloor|^\rho}{N^\rho} \right| < \frac{\varepsilon}{2} \quad (2.58)$$

by (2.57). Combining (2.56), (2.58) and the triangle inequality, we have

$$\left| \frac{R(\lfloor Nx \rfloor - \lfloor Ny \rfloor)}{R(N)} - |x - y|^\rho \right| < \varepsilon$$

for all $N > \max(N_1, N_2)$ and $(x, y) \in A_\varepsilon$. Then for N large enough, we have

$$\operatorname{ess\,sup}_{y \in [0,1]} \int_{A_\varepsilon(y)} \left| \frac{R(\lfloor Nx \rfloor - \lfloor Ny \rfloor)}{R(N)} - |x - y|^\rho \right| dx < \varepsilon. \quad (2.59)$$

On the other hand, for all $y \in [0, 1]$, we have

$$\int_{[0,1] \setminus A_\varepsilon(y)} |x - y|^\rho dx \leq \int_{-\varepsilon}^\varepsilon |x|^\rho dx = \frac{2\varepsilon^{1+\rho}}{1+\rho}. \quad (2.60)$$

Hence we just need to control the integral

$$\int_{[0,1] \setminus A_\varepsilon(y)} \left| \frac{R(\lfloor Nx \rfloor - \lfloor Ny \rfloor)}{R(N)} \right| dx.$$

Notice that both R and $|R|$ are even, locally bounded and regularly varying with index ρ .

By Lemma II.16, we have

$$\begin{aligned} \int_{[0,1] \setminus A_\varepsilon(y)} \left| \frac{R(\lfloor Nx \rfloor - \lfloor Ny \rfloor)}{R(N)} \right| dx &\leq \int_{y-\varepsilon}^{y+\varepsilon} \left| \frac{R(\lfloor Nx \rfloor - \lfloor Ny \rfloor)}{R(N)} \right| dx \\ &\stackrel{(a)}{=} \int_{-\varepsilon}^\varepsilon \left| \frac{R(\lfloor Nx + (Ny - \lfloor Ny \rfloor)\rfloor)}{R(N)} \right| dx \\ &= \int_{-N\varepsilon}^{N\varepsilon} \left| \frac{R(\lfloor x + (Ny - \lfloor Ny \rfloor)\rfloor)}{NR(N)} \right| dx \\ &\leq \int_{-N\varepsilon-1}^{N\varepsilon+1} \left| \frac{R(\lfloor x \rfloor)}{NR(N)} \right| dx \\ &\stackrel{(b)}{\approx} 2 \left| \frac{R(N\varepsilon+1)}{R(N)} \right| \frac{\varepsilon}{1+\rho} \stackrel{(c)}{\approx} \frac{2\varepsilon^{1+\rho}}{1+\rho} \quad (2.61) \end{aligned}$$

as $N \rightarrow \infty$, where (a) follows from a change of variable and the fact that $\lfloor x \rfloor + h = \lfloor x + h \rfloor$

for every $h \in \mathbb{Z}$, (b) follows from Lemma II.16 and (c) from Lemma II.15. Notice that the controls (2.60) and (2.61) are independent of y , hence for N large enough, we have

$$\operatorname{ess\,sup}_{y \in [0,1]} \int_{[0,1] \setminus A_\varepsilon(y)} \left| \frac{R(\lfloor Nx \rfloor - \lfloor Ny \rfloor)}{R(N)} - |x - y|^\rho \right| dx < \frac{5\varepsilon^{1+\rho}}{1 + \rho}. \quad (2.62)$$

Combining (2.59) and (2.62), and taking $\varepsilon \rightarrow 0$, we finally obtain

$$\|\mathcal{K}_N^{(\rho)} - \mathcal{K}^{(\rho)}\|_p \xrightarrow{N \rightarrow \infty} 0$$

for all $p \in [1, \infty]$. □

As a consequence of Lemma II.17, we conclude that $\mathcal{K}^{(\rho)}$ is compact on $L^p(0, 1)$ for all $p \in [1, \infty]$, because it is the limit in operator norm of finite dimensional operators $\mathcal{K}_N^{(\rho)}$.

We will complete the proof of Theorem II.8 in the next section.

2.6.2 Convergence of eigenvalues and simplicity of the largest eigenvalue

For a fixed $k = 1, 2, \dots$, let $a_k^{(\rho)}$ be the k -th largest eigenvalue of $\mathcal{K}^{(\rho)}$ and $a_{N,k}^{(\rho)} = \lambda_k(T_N(c))/(NR(N))$ be the k -th largest eigenvalue of $\mathcal{K}_N^{(\rho)}$. From the convergence $\|\mathcal{K}_N^{(\rho)} - \mathcal{K}^{(\rho)}\|_2 \rightarrow 0$ we deduce that $a_{N,k}^{(\rho)} \rightarrow a_k^{(\rho)}$ as $N \rightarrow \infty$. In fact, as $\mathcal{K}_N^{(\rho)}$ and $\mathcal{K}^{(\rho)}$ are compact and self-adjoint, by the Min-Max Formula (see also [51, Theorem 4.12]) we have

$$\begin{aligned} a_{N,k}^{(\rho)} &= \min_{\dim U = k-1} \max_{\substack{u \in U^\perp \\ \|u\|_2 = 1}} \langle u, \mathcal{K}_N^{(\rho)} u \rangle \\ &\leq \min_{\dim U = k-1} \max_{\substack{u \in U^\perp \\ \|u\|_2 = 1}} \langle u, \mathcal{K}^{(\rho)} u \rangle + \|\mathcal{K}_N^{(\rho)} - \mathcal{K}^{(\rho)}\|_2 = a_k^{(\rho)} + \|\mathcal{K}_N^{(\rho)} - \mathcal{K}^{(\rho)}\|_2. \end{aligned}$$

Symmetrically, we also have $a_k^{(\rho)} \leq a_{N,k}^{(\rho)} + \|\mathcal{K}_N^{(\rho)} - \mathcal{K}^{(\rho)}\|_2$, from which we deduce

$$|a_{N,k}^{(\rho)} - a_k^{(\rho)}| \leq \|\mathcal{K}_N^{(\rho)} - \mathcal{K}^{(\rho)}\|_2$$

for all N and k . This implies the convergence of each eigenvalue $a_{N,k}^{(\rho)}$ toward $a_k^{(\rho)}$.

We now prove that $a_1^{(\rho)}$ is a simple eigenvalue of $\mathcal{K}^{(\rho)}$. Let $u \in L^2(0, 1)$ be an eigenfunc-

tion of $a_1^{(\rho)}$, then by the mini-max formula $a_1^{(\rho)} = \max_{f \in L^2, \|f\|_2=1} \langle f, \mathcal{K}^{(\rho)} f \rangle$, we have

$$a_1^{(\rho)} = \int_0^1 \int_0^1 |x-y|^\rho u(x) \overline{u(y)} \, dx \, dy \leq \int_0^1 \int_0^1 |x-y|^\rho |u(x)u(y)| \, dx \, dy \leq a_1^{(\rho)}$$

which implies

$$\int_0^1 \int_0^1 |x-y|^\rho (|u(x)u(y)| - u(x)\overline{u(y)}) \, dx \, dy = 0.$$

Hence $|u(x)u(y)| = u(x)\overline{u(y)}$ for $(x, y) \in [0, 1]^2$ $dx \, dy$ -a.e. This implies that for almost all $y \in [0, 1]$, the equality $|u(x)u(y)| = u(x)\overline{u(y)}$ holds for almost all $x \in [0, 1]$. Let y_0 be such that $u(y_0) \neq 0$ and $c = u(y_0)/|u(y_0)| = e^{i\varphi_0}$. Then for almost every $x \in [0, 1]$, we have

$$u(x) = \frac{|u(x)u(y_0)|}{\overline{u(y_0)}} = c|u(x)|.$$

So up to a nonzero constant multiplier we can suppose that $u \geq 0$ on $[0, 1]$. Therefore

$$a_1^{(\rho)} u(x) = \int_0^1 |x-y|^\rho u(y) \, dy \geq \int_0^1 u(y) \, dy > 0.$$

This implies that $a_1^{(\rho)} > 0$ and $u(x) > 0$ for all $x \in [0, 1]$. Then for any other function $v \in L^2(0, 1)$ s.t. $\langle u, v \rangle = 0$, v cannot be an eigenfunction of the eigenvalue $a_1^{(\rho)}$. Otherwise following the same line of reasoning as previously we may write $v = |v|e^{i\tilde{\varphi}_0}$ where $|v|$ is also an eigenfunction associated to $a_1^{(\rho)}$, $|v| > 0$ on $[0, 1]$ and $\langle u, |v| \rangle = e^{-i\tilde{\varphi}_0} \langle u, v \rangle = 0$. But $u, |v| > 0$ contradict the orthogonality

$$\langle u, |v| \rangle = \int_0^1 u(x)|v(x)| \, dx = 0.$$

Finally recall that if R is regularly varying of index $\rho \in (-1, 0)$, then $NR(N) \rightarrow \infty$ as $N \rightarrow \infty$ by [14, Prop. 1.3.6(v)]. Combining Widom-Shampine's lemma and the convergence of eigenvalues, we obtain

$$a_{N,1}^{(\rho)} = \frac{\lambda_{\max}(T_N(c))}{NR(N)} \xrightarrow{N \rightarrow \infty} a_1^{(\rho)} \quad \text{and} \quad a_{N,2}^{(\rho)} = \frac{\lambda_2(T_N(c))}{NR(N)} \xrightarrow{N \rightarrow \infty} a_2^{(\rho)}.$$

Hence $\lambda_{\max}(T_N(c)) \sim a_1^{(\rho)} NR(N) \rightarrow \infty$ and

$$\frac{\lambda_2(T_N(c))}{\lambda_{\max}(T_N(c))} \xrightarrow{N \rightarrow \infty} \frac{a_2^{(\rho)}}{a_1^{(\rho)}} < 1.$$

Proof of Theorem [II.8](#) is now completed.

CHAPTER III

Joint CLT for top eigenvalues of empirical covariance matrices of long memory stationary processes

This chapter corresponds to the article with the same title *Joint CLT for top eigenvalues of empirical covariance matrices of long memory stationary processes* by P.T. This article has been submitted to *Annals of applied probability*.

3.1 Abstract

In this paper we study the joint CLT of the m largest eigenvalues $\lambda_1(S_N(T)), \dots, \lambda_m(S_N(T))$ of the matrix

$$S_N(T) = \frac{1}{n} T_N^{\frac{1}{2}} Z_N Z_N^* T_N^{\frac{1}{2}}$$

where $Z_N = (Z_{i,j})$ is a $N \times n$ matrix having i.i.d. centered entries with variance one, and T_N is a $N \times N$ deterministic real symmetric Toeplitz matrix. More precisely, T_N is the N th autocovariance matrix of a long range dependent stationary process, so that S_N can be viewed as a linear representation of sample autocovariance matrix of the process. As $N, n \rightarrow \infty$ and $Nn^{-1} \rightarrow r \in (0, \infty)$, we prove the convergence in distribution of the vector $(\lambda_1(S_N(T)), \dots, \lambda_m(S_N(T)))^\top$ to a Gaussian vector at scale $\sqrt{n} \lambda_{\max}^{-1}(T_N)$, whose covariance matrix only depends on $\mathbb{E} Z_{i,j}^2$, the second non-absolute moment of the entries $Z_{i,j}$. This result substantially extends our previous result in [42], where we studied the fluctuations of $\lambda_1(S_N(T))$ in the case where the $Z_{i,j}$ s were gaussian.

In order to establish this CLT, we are led to study the asymptotics of the m largest

eigenvalues (and their associated eigenvectors) of deterministic Toeplitz matrices of growing dimensions. In particular, we prove multiple spectral gap properties for these largest eigenvalues and a delocalization property for their associated eigenvectors. During our studies, we also establish a CLT for the top eigenvalues of a generic model of a large empirical covariance matrix

$$S_N(\Gamma) = \frac{1}{n} \Gamma_N^{\frac{1}{2}} Z_N Z_N^* \Gamma_N^{\frac{1}{2}},$$

where Γ_N is a $N \times N$ deterministic positive semi-definite Hermitian matrix with unbounded spectral norm. This general CLT is of interest on its own.

3.2 Introduction

In this article we study the joint CLT of the m (with m a fixed positive integer) largest eigenvalues of

$$S_N(T) := \frac{1}{n} T_N^{\frac{1}{2}} Z_N Z_N^* T_N^{\frac{1}{2}}, \quad (3.1)$$

in the regime $N, n \rightarrow \infty$ and $Nn^{-1} \rightarrow r \in (0, \infty)$ (in the sequel, we simply denote this regime as $N, n \rightarrow \infty$), where $Z_N = (Z_{i,j})$ is a $N \times n$ random matrix with i.i.d. centered entries $Z_{i,j}$ with variance one, and T_N is a $N \times N$ deterministic real symmetric Toeplitz matrix. More precisely, T_N is the N th autocovariance matrix of a long range dependent stationary process.

Recall that a process $(\mathcal{X}_t)_{t \in \mathbb{Z}}$ is (second order) stationary if

$$\mathbb{E}|\mathcal{X}_t|^2 < \infty, \quad \mathbb{E}\mathcal{X}_t = \mathbb{E}\mathcal{X}_0 \quad \text{and} \quad \text{Cov}(\mathcal{X}_{t+h}, \mathcal{X}_t) = \text{Cov}(\mathcal{X}_h, \mathcal{X}_0) = \gamma(h) \quad \forall t, h \in \mathbb{Z}$$

where $\gamma : \mathbb{Z} \rightarrow \mathbb{C}$ is a positive definite function, called the autocovariance function of the process, which satisfies $\gamma(-h) = \overline{\gamma(h)}$. In this paper the process will be said to have long memory, or to be long range dependent, if its autocovariance function γ is real and of the form

$$\gamma(h) = \frac{L(|h|)}{(1 + |h|)^{1-2d}}, \quad \forall h \in \mathbb{Z}, \quad (3.2)$$

where $d \in (0, 1/2)$ is called the long memory parameter and L is a slowly varying function

at infinity (c.f. for example Chapter 2 of [45]). Recall that a real function L is said to be slowly varying at infinity, if it is asymptotically positive, and for any $c > 0$,

$$\lim_{x \rightarrow \infty} \frac{L(cx)}{L(x)} = 1.$$

Note that in this case, γ satisfies

$$\sum_{h \in \mathbb{Z}} |\gamma(h)| = \infty.$$

If to the contrary a stationary process satisfies $\sum |\gamma(h)| < \infty$, then it is a short memory process.

Given a long memory process $(\mathcal{X}_t)_{t \in \mathbb{Z}}$, assumed to be centered, i.e. $\mathbb{E}\mathcal{X}_t = 0$ for all $t \in \mathbb{Z}$, its autocovariance matrices T_N are defined as

$$T_N := \text{Cov} \begin{pmatrix} \mathcal{X}_{t+1} \\ \vdots \\ \mathcal{X}_{t+N} \end{pmatrix} = \begin{pmatrix} \gamma(0) & \gamma(-1) & \dots & \gamma(-N+1) \\ \gamma(1) & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \gamma(-1) \\ \gamma(N-1) & \dots & \gamma(1) & \gamma(0) \end{pmatrix}. \quad (3.3)$$

Note that T_N are real symmetric Toeplitz matrices. From the same process, we can also construct its sample autocovariance matrix

$$Q_N := \frac{1}{n} \sum_{j=1}^n X_j X_j^*, \quad (3.4)$$

where X_1, \dots, X_n are i.i.d. observations of the random vector $(\mathcal{X}_1, \dots, \mathcal{X}_N)^\top$. We are interested in the spectral properties of Q_N . After considering the global behaviour of the **ESD** $\mu^{Q_N} := \frac{1}{N} \sum_{k=1}^N \delta_{\lambda_k(Q_N)}$ (see for example [43]), it is natural to study the individual eigenvalues, especially the largest ones. If the process is Gaussian, then Q_N has the linear form S_N given in (3.1). In this case, a full description of the asymptotic behaviour and the fluctuations of $\lambda_{\max}(Q_N)$ has been provided in [42]. In the present paper, we extend these results in the following two directions:

1. We extend our previous result to nongaussian linear models.

2. We extend the fluctuations of one top eigenvalue to the joint fluctuations of several top eigenvalues, and describe in particular the asymptotic correlation structure of these top eigenvalues.

In the non-Gaussian case, a sample autocovariance model Q_N in (3.4) may not admit the linear representation $S_N(T)$ defined in (3.1). In this article, we restrict our attention to the study of $S_N(T)$ in the non-Gaussian case, and prove that, for any fixed integer $m \geq 1$, the m largest eigenvalues of T_N tend to infinity at the same speed, and after normalization by $N\gamma(N)$, the m quantities $\lambda_1(T_N)/(N\gamma(N)), \dots, \lambda_m(T_N)/(N\gamma(N))$ converge to m distinct finite limits. In the sequel, such a property will be called *separate limiting property* of these eigenvalues. Precisely, a group of sequences $(\alpha_1^{(n)})_n, \dots, (\alpha_m^{(n)})_n$ admit the *separate limiting property* if there exists a normalizing sequence $\beta^{(n)} > 0$ such that the sequences

$$\frac{\alpha_1^{(n)}}{\beta^{(n)}}, \dots, \frac{\alpha_m^{(n)}}{\beta^{(n)}}$$

converge to m distinct finite limits.

On this basis, we prove that for the T_N 's with parameter $d \in (1/8, 1/2)$, the m largest eigenvalues of $S_N(T)$ satisfy the following CLT

$$\sqrt{n} \begin{pmatrix} \frac{\lambda_1(S_N(T))}{\lambda_1(T_N)} - 1 - \frac{1}{n} \sum_{k \neq 1} \frac{\lambda_k(T_N)}{\lambda_1(T_N) - \lambda_k(T_N)} \\ \vdots \\ \frac{\lambda_m(S_N(T))}{\lambda_m(T_N)} - 1 - \frac{1}{n} \sum_{k \neq m} \frac{\lambda_k(T_N)}{\lambda_m(T_N) - \lambda_k(T_N)} \end{pmatrix} \xrightarrow[N, n \rightarrow \infty]{\mathcal{D}} \mathcal{N}(0, (1 + |\mathbb{E}Z_{1,1}^2|^2)I_m) \quad (3.5)$$

where $\xrightarrow{\mathcal{D}}$ denotes the convergence in distribution. The case where $d \in (0, 1/8]$ is still an open question by the time of submission of this article. Note that the value $d^* = 0$ represents the threshold between short memory process (where one may expect Tracy-Widom fluctuations) and long memory process, and the more d gets closer to d^* the harder it is to establish Gaussian fluctuations.

From the above results, we can see that the model $S_N(T)$ unveils a number of original properties with respect to more standard large empirical covariance matrices. Among these properties, we mention the following:

1. For any $m \geq 1$, with high probability, the m largest eigenvalues $\lambda_1(S_N(T)), \dots, \lambda_m(S_N(T))$ satisfy the separate limiting property with normalizing sequence $\lambda_{\max}(T_N)$. This is a corollary of (3.5) together with the separate limiting property of largest eigenvalues of T_N . Note that by the same argument of Remark 3 in [42], one can prove that

$$\frac{1}{n} \sum_{k \neq i} \frac{\lambda_k(T_N)}{\lambda_i(T_N) - \lambda_k(T_N)} \xrightarrow{N, n \rightarrow \infty} 0, \quad \text{for } i = 1, \dots, m.$$

2. For any $m \geq 1$, the m largest eigenvalues of S_N , when properly normalized, are asymptotically independent Gaussian random variables. They fluctuate in an analogous way as spiked eigenvalues in classical sample covariance matrix models.
3. The fluctuation result is universal: Unlike the model with block-diagonal population covariance matrix described in our previous paper [42], where the fluctuations depend on $\mathbb{E}|Z_{1,1}|^4$, we can see that here they depend only on $\mathbb{E}Z_{1,1}^2$. So if the entries $Z_{i,j}$ are real with variance one, or complex with $\mathbb{E}Z_{i,j}^2 = 0$, the top eigenvalues fluctuate in a completely universal way. For a numerical illustration of this difference see Simulation 3 in Section 3.2 of [42].

We hope that the techniques developed to establish these properties could inspire further research on the general model Q_N .

The proof of the CLT in (3.5) divides into two parts. In the first part, we study the asymptotic behaviour of the top eigenvalues of Toeplitz matrices T_N and the associated eigenvectors. We prove the separate limiting property for the top eigenvalues of T_N , and the delocalization of the associated eigenvectors. In the second part, we develop a CLT for the m largest eigenvalues of generic sample covariance matrix

$$S_N(\Gamma) = \frac{1}{n} \Gamma_N^{\frac{1}{2}} Z_N Z_N^* \Gamma_N^{\frac{1}{2}} \tag{3.6}$$

where Γ_N is a deterministic hermitian matrix whose assumptions are designed to fulfill T_N 's properties.

For almost a century the Toeplitz matrices have been studied by many mathematicians.

The early pioneers include G. Szegő, U. Grenander, M. Kac, H. Widom. In recent decades this topic has been greatly developed by many others like S. M. Grudsky, J. M. Bogoya, A. Bttcher etc. Some excellent textbooks can be found, for example [28, 15, 18]. Their results have been widely applied in many areas. As we have seen, this kind of matrices appear naturally in the stationary processes. However, If the underlying process exhibits long memory, the associated Toeplitz matrices usually have unbounded symbols, so their top eigenvalues are unbounded. In this case, much less is known about their behaviour.

In [42], we established the asymptotics of the j th largest eigenvalue $\lambda_j(T_N)$ for any fixed j , and proved that

$$\frac{\lambda_j(T_N)}{N\gamma(N)} \rightarrow \lambda_j(\mathcal{K}^{(\rho)}) \quad (3.7)$$

where we denote $\rho = 2d - 1$, and $\mathcal{K}^{(\rho)}$ is a compact operator acting on $L^2(0, 1)$ defined by

$$(\mathcal{K}^{(\rho)} f)(x) = \int_0^1 |x - y|^\rho f(y) dy. \quad (3.8)$$

We also proved that $\lambda_1(\mathcal{K}^{(\rho)})$ is simple, so the two largest eigenvalues of T_N have separate limiting property.

In the present paper, we prove that $\lambda_j(\mathcal{K}^{(\rho)})$ is simple for all $j \geq 1$, which implies the separate limiting property for any finite number of top eigenvalues of T_N . Moreover, we prove that the eigenfunction of $\mathcal{K}^{(\rho)}$ associated with $\lambda_j(\mathcal{K}^{(\rho)})$ is continuous on $[0, 1]$, and that the eigenvectors of T_N associated with $\lambda_j(T_N)$ approximate the eigenfunction of $\mathcal{K}^{(\rho)}$ in some sense, see Theorem III.2. This implies the delocalization of eigenvectors of T_N .

Using the above properties of Toeplitz matrices T_N , we can study the fluctuations of top eigenvalues of $S_N(T)$. The key tool is a CLT for the m largest eigenvalues of generic sample covariance matrix $S_N(\Gamma)$ defined in (3.6), assuming that Γ_N , the population covariance matrix, satisfies some crucial properties of Toeplitz matrices, including

1. The m largest eigenvalues of Γ_N tend to ∞ at the same speed.
2. The $m + 1$ largest eigenvalues of Γ_N satisfy multiple spectral gap conditions, i.e. there

exists a constant $c < 1$ such that for any $j = 1, \dots, m$,

$$\frac{\lambda_{j+1}(\Gamma_N)}{\lambda_j(\Gamma_N)} \leq c.$$

3. Γ_N satisfies the following condition:

$$\frac{\text{tr } \Gamma_N^2}{\sqrt{N} \lambda_{\max}^2(\Gamma_N)} \rightarrow 0$$

(See Remark 11 for some comments on this condition).

Under these assumptions, we prove that the m largest eigenvalues of the generic model (3.6) fluctuate as Gaussian variables. More precisely, let

$$\Lambda_m := \sqrt{n} \begin{pmatrix} \frac{\lambda_1(S_N(\Gamma))}{\lambda_1(\Gamma_N)} - 1 - \frac{1}{n} \sum_{k \neq 1} \frac{\lambda_k(\Gamma_N)}{\lambda_1(\Gamma_N) - \lambda_k(\Gamma_N)} \\ \vdots \\ \frac{\lambda_m(S_N(\Gamma))}{\lambda_m(\Gamma_N)} - 1 - \frac{1}{n} \sum_{k \neq m} \frac{\lambda_k(\Gamma_N)}{\lambda_m(\Gamma_N) - \lambda_k(\Gamma_N)} \end{pmatrix}$$

then we prove that

$$d_{\mathcal{LP}}(\Lambda_m, \mathcal{N}(0, \Sigma_m^{(N)})) \rightarrow 0, \quad (3.9)$$

as $N, n \rightarrow \infty$, where $d_{\mathcal{LP}}$ is the Lévy-Prokhorov distance, and

$$\Sigma_m^{(N)} = I_m + (\sigma_{i,j}^{(N)})_{i,j=1}^m$$

with

$$\sigma_{i,j}^{(N)} = (\mathbb{E}|Z_{1,1}|^4 - |\mathbb{E}Z_{1,1}^2|^2 - 2) \sum_{k=1}^N |u_{k,i}|^2 |u_{k,j}|^2 + |\mathbb{E}Z_{1,1}^2|^2 \left| \sum_{k=1}^N u_{k,i} u_{k,j} \right|^2 \quad (3.10)$$

where $u_j := (u_{1,j}, \dots, u_{N,j})^\top$ the eigenvector of Γ_N associated with $\lambda_j(\Gamma_N)$. This formula, in the case where $Z_{i,j}$ and Γ_N are real with convergent entries $\sigma_{i,j}^{(N)}$, has already appeared in the context of principal component analysis (PCA) in other papers, see for example [20]. Here we do not assume that $\sigma_{i,j}^{(N)}$ converges, so we state our result in term of a distance

metrizing the weak convergence. Note that our generic CLT takes simpler forms in the following two situations:

1. If the eigenvectors $(u_j)_{1 \leq j \leq m}$ are real and delocalized, i.e. $u_j \in \mathbb{R}^m$ and $\|u_j\|_\infty \rightarrow 0$ as $N \rightarrow \infty$, then we have

$$\Sigma_m^{(N)} \xrightarrow{N \rightarrow \infty} (1 + |\mathbb{E}Z_{1,1}^2|^2)I_m$$

2. If Γ_N is diagonal, then we have

$$\Sigma_m^{(N)} = (\mathbb{E}|Z_{1,1}|^4 - 1)I_m,$$

which is consistent with [42].

This general CLT, although primarily designed to address the study of the top eigenvalues of the model $S_N(T)$ in (3.1), also provides new perspectives in the study of generic sample covariance matrices, with population covariance matrices with unbounded spectral norm.

The spectral properties of generic sample covariance matrix $S_N(\Gamma)$ defined by (3.6) has been intensively studied in the last several decades. In the groundbreaking work of Marčenko and Pastur [41], the authors studied the **LSD** of $S_N(\Gamma)$ in the case where Γ_N is a diagonal matrix with i.i.d. diagonal entries which are independent of the entries of Z_N , and they discovered the so-called Marčenko-Pastur Distribution. After that, many works on the **LSD** of $S_N(\Gamma)$ with different configuration of Γ_N have appeared, see for instance [56, 34, 59, 49, 48]. We mention that Silversteins result [48] applies to our model $S_N(T)$ and combined with variations of Szegő's theorem [55] describes the **LSD** of $S_N(T)$.

Apart from the behaviour of **LSD**, people are also interested in individual eigenvalues of $S_N(\Gamma)$, especially the extreme eigenvalues. Unlike the global scale, the extreme eigenvalues are more sensitive to the eigenstructure of Γ_N . For example, Baik *et al.* [9] studied $\lambda_{\max}(S_N(\Gamma))$ with several different configurations of Γ_N and discovered the so-called BBP phase transition. Even for the white noise model where $\Gamma_N = I_N$, the fluctuations of $\lambda_{\max}(S_N(\Gamma))$ are different between the cases where Z_N has real entries or complex entries

(see Johansson [32] and Johnstone [33]). All these works show that it is not realistic to build a universal result for the fluctuations of individual eigenvalues for general Γ_N . People have to study case by case. For instance El Karoui [25], Lee and Schnelli [39], Bao *et al.* [11], and Knowles and Yin [37] established the Tracy-Widom fluctuation of the largest eigenvalue in the case where $\lambda_{\max}(\Gamma_N)$ is bounded and sticks to the upper bound of the bulk of the **LSD** of Γ_N , while Bai and Yao [6, 7] studied the spiked eigenvalues and established the Gaussian-type fluctuations.

Recently several models of matrices $S_N(\Gamma)$ with Γ_N having a small number of divergent eigenvalues have been considered, in the context of principal component analysis (PCA) [35, 47, 57, 20] and long memory processes [42]. Although the assumptions in these various works differ, many results coincide (see Remark 13).

Organizations. This paper is organized as follows. In Section 3.3 we state our main theorems. This section is divided in three parts. In 3.3.1, we state the results on Toeplitz matrices T_N , whose proofs are given in Section 3.4. In 3.3.2, we state the general CLT, whose proof is in Section 3.5. In 3.3.3 we state the CLT for top eigenvalues of long memory sample covariance matrix $S_N(T)$ which is a direct corollary of the previous results.

Notations. For an hermitian operator or matrix A , we denote its real eigenvalues by decreasing order as

$$\lambda_1(A) \geq \lambda_2(A) \geq \dots$$

We also denote the largest eigenvalue of A by $\lambda_{\max}(A)$. For a matrix or a vector A , we use A^\top to denote the transpose of A , and A^* the conjugate transpose of A .

The kernel of a linear operator $A : X \rightarrow X$, is denoted by $\ker A$ and defined by

$$\ker A := \{u \in X : Au = 0\}.$$

The spectrum of A is denoted by $\text{Spec}(A)$.

For a matrix A , we denote A_i , the i th row of A , and A_{-i} , the submatrix of A obtained by deleting the i th row. Similarly $A_{-i,-j}$ denotes the submatrix of A obtained by deleting

the i th row and the j th column. When using these subscripts, we do not indicate the dependence on N . By convention, these subscripts have higher priority than the transpose or conjugate transpose, for example $M_{i,\cdot}^* := (M_{i,\cdot})^*$ is the conjugated transposition of the submatrix $M_{i,\cdot}$.

We denote the L^p or l^p norm by $\|\cdot\|_p$. For a matrix or a linear operator A , the norm of A induced by vector norm $\|\cdot\|_p$ is denoted by $\|A\|_p$, and we recall that $\|A\|_p := \sup_{\|v\|_p=1} \|Av\|_p$. We say that a function f or a vector v is “normalized” or “unit length” when $\|f\|_2 = 1$ or $\|v\|_2 = 1$. When functions or vectors are said to be “orthonormal”, they will be implicitly considered as elements of a Hilbert space. The inner product of two elements u, v of a Hilbert space is denoted by $\langle u, v \rangle$.

For two probability measures P and Q on \mathbb{R}^m , we denote their Lévy-Prokhorov distance by $d_{\mathcal{LP}}(P, Q)$ which is defined by

$$d_{\mathcal{LP}}(P, Q) := \inf\{\varepsilon : P(A) \leq Q(A^\varepsilon) + \varepsilon, Q(A) \leq P(A^\varepsilon) + \varepsilon, \forall A \in \mathcal{B}(\mathbb{R}^m)\}, \quad (3.11)$$

where A^ε is defined by

$$A^\varepsilon := \{x \in \mathbb{R}^m : \exists y \in A, \text{ s.t. } \|x - y\| < \varepsilon\}.$$

It is well known that this distance metrizes the weak convergence. For two random variables $X, Y \in \mathbb{R}^m$ and two probability measures μ_X, μ_Y such that $X \sim \mu_X$ and $Y \sim \mu_Y$, we sometimes write $d_{\mathcal{LP}}(X, Y)$, $d_{\mathcal{LP}}(X, \mu_Y)$ or $d_{\mathcal{LP}}(\mu_X, Y)$, all these notations denote $d_{\mathcal{LP}}(\mu_X, \mu_Y)$.

Given $x \in \mathbb{R}$, we denote by $\lfloor x \rfloor$ the integer satisfying $\lfloor x \rfloor \leq x < \lfloor x \rfloor + 1$. For two real numbers x, y , we denote

$$x \wedge y := \min(x, y) \quad \text{and} \quad x \vee y := \max(x, y).$$

Given two sequences of nonnegative numbers x_n, y_n , we denote

$$x_n \sim y_n \Leftrightarrow \lim_{n \rightarrow \infty} \left(\frac{x_n}{y_n} \right) = 1, \quad x_n \ll y_n \Leftrightarrow \lim_{n \rightarrow \infty} \frac{x_n}{y_n} = 0 \quad \text{and} \quad x_n \gg y_n \Leftrightarrow \lim_{n \rightarrow \infty} \frac{y_n}{x_n} = 0.$$

The notations $x_n = o(1)$ and $x_n = O(1)$ respectively mean $\lim_{n \rightarrow \infty} x_n = 0$ and $\overline{\lim}_{n \rightarrow \infty} |x_n| < \infty$. If X_n, X are random variables, the notation $X_n = o_P(1)$ means that $\lim_{n \rightarrow \infty} X_n = 0$ in probability. The notations $X_n \xrightarrow{\mathcal{D}} X$ and $X_n \xrightarrow{\mathcal{P}} X$ respectively denote convergence in distribution and in probability. If μ_n, μ are measures, we denote with a slight abuse of notation $\mu_n \xrightarrow{\mathcal{D}} \mu$ for the weak convergence of μ_n to μ .

The cardinal of a set B is denoted by $\#B$. In the proofs we use C to denote a constant that may take different values from one place to another.

Acknowledgement. This work is a part of my Ph-D research, and is supported financially by Université Paris-Est. I would like to thank gratefully my Ph-D advisers Professor Florence MERLEVEDE and Professor Jamal NAJIM, for all useful discussion and advises during completion of this work.

3.3 Statement of the main theorems

3.3.1 Spectral properties of Toeplitz matrices

Let T_N be a Toeplitz matrix defined by

$$T_N := (\gamma(i - j))_{i,j=1}^N \tag{3.12}$$

with $\gamma : \mathbb{Z} \rightarrow \mathbb{R}$ a function in the form

$$\gamma(h) = (1 + |h|)^\rho L(|h|), \quad \forall h \in \mathbb{Z}, \tag{3.13}$$

where L is a slowly varying function at infinity and $\rho = 2d - 1 \in (-1, 0)$. Let $\mathcal{K}^{(\rho)}$ be the operator defined on $L^2(0, 1)$ by

$$(\mathcal{K}^{(\rho)} f)(x) = \int_0^1 |x - y|^\rho f(y) dy, \quad \text{for } f \in L^2(0, 1). \tag{3.14}$$

We will describe the eigenvalues and eigenvectors of T_N with help of the operator $\mathcal{K}^{(\rho)}$, so we first establish a theorem on the spectral properties of this operator.

Theorem III.1. *The operator $\mathcal{K}^{(\rho)}$ is compact and positive semidefinite. It has infinitely many positive eigenvalues. All its nonzero eigenvalues are simple, and the associated eigenfunctions are continuous on $[0, 1]$.*

We note that $\mathcal{K}^{(\rho)}$ is self-adjoint so for any nonzero eigenvalue λ , its algebraic multiplicity equals to its geometric multiplicity, which is defined as $\dim \ker(\lambda I - \mathcal{K}^{(\rho)})$. For more information about algebraic multiplicity, see [40]. So here we say that a nonzero eigenvalue λ is simple, it means that

$$\dim \ker \left(\lambda I - \mathcal{K}^{(\rho)} \right) = 1.$$

In the next theorem, we describe the limiting behaviour of the eigenvalues $\lambda_j(T_N)$ and provide a quantitative description of their associated eigenvectors. Note that the formula (3.15) has already appeared in our previous article [42].

Theorem III.2. *For any $j \geq 1$, we have*

$$\lim_{N \rightarrow \infty} \frac{\lambda_j(T_N)}{N\gamma(N)} = \lambda_j(\mathcal{K}^{(\rho)}). \quad (3.15)$$

Moreover let f_j be the normalized eigenfunction of $\mathcal{K}^{(\rho)}$ associated with $\lambda_j(\mathcal{K}^{(\rho)})$, then for any N , we can choose a normalized eigenvector $u_j = (u_{1,j}, \dots, u_{N,j})^\top$ of T_N associated with $\lambda_j(T_N)$, such that

$$\lim_{N \rightarrow \infty} \sup_{1 \leq i \leq N} \left\{ \left| \sqrt{N} u_{i,j} - f_j \left(\frac{i}{N} \right) \right| \right\} = 0. \quad (3.16)$$

Combining Theorem III.1 and III.2 we get the following two important consequences. The first consequence is the separate limiting property of any m (with $m \geq 1$) largest eigenvalues $\lambda_1(T_N), \dots, \lambda_m(T_N)$, with normalizing sequence $N\gamma(N)$. The second consequence is the delocalization of the eigenvector u_j associated with $\lambda_j(T_N)$ for any fixed $j \geq 1$. Indeed by (3.16), for large N , we have

$$\|u_j\|_\infty \leq \frac{1 + \|f_j\|_\infty}{\sqrt{N}},$$

and by Theorem III.1 we have $\|f_j\|_\infty < \infty$. Thus we conclude that

$$\|u_j\|_\infty = O\left(1/\sqrt{N}\right) \rightarrow 0.$$

The next proposition provides a comparison between the global behaviour of T_N s eigenvalues ($\text{tr } T_N$ or $\text{tr } T_N^2$) and $\lambda_{\max}(T_N)$.

Proposition III.3. *Let T_N be defined as above.*

1. *If $\rho \in (-1/2, 0)$ or equivalently $d \in (1/4, 1/2)$, then T_N satisfies*

$$\lim_{N \rightarrow \infty} \frac{\text{tr } T_N}{\sqrt{N} \lambda_{\max}(T_N)} = 0. \tag{3.17}$$

2. *If $\rho \in (-3/4, -1/2]$ or equivalently $d \in (1/8, 1/4]$, then T_N satisfies*

$$\lim_{N \rightarrow \infty} \frac{\text{tr } T_N^2}{\sqrt{N} \lambda_{\max}^2(T_N)} = 0. \tag{3.18}$$

3.3.2 A general CLT

Let

$$S_N(\Gamma) = \frac{1}{n} \Gamma_N^{\frac{1}{2}} Z_N Z_N^* \Gamma_N^{\frac{1}{2}}$$

where Z_N is a $N \times n$ random matrix with i.i.d. entries $Z_{i,j}$ such that

$$\mathbb{E}Z_{i,j} = 0, \quad \mathbb{E}|Z_{i,j}|^2 = 1, \quad \text{for all } 1 \leq i \leq N \quad \text{and} \quad 1 \leq j \leq n,$$

and Γ_N is a $N \times N$ deterministic positive semidefinite hermitian matrix. For $1 \leq i \leq N$, let $u_i = (u_{1,i}, \dots, u_{N,i})^\top$ be an eigenvector of Γ_N associated with $\lambda_i(\Gamma_N)$.

Given an integer $m \geq 1$, we define

$$\Lambda_m(\Gamma_N) := \sqrt{n} \begin{pmatrix} \frac{\lambda_1(S_N(\Gamma))}{\lambda_1(\Gamma_N)} - 1 - \frac{1}{n} \sum_{k \neq 1} \frac{\lambda_k(\Gamma_N)}{\lambda_1(\Gamma_N) - \lambda_k(\Gamma_N)} \\ \vdots \\ \frac{\lambda_j(S_N(\Gamma))}{\lambda_j(\Gamma_N)} - 1 - \frac{1}{n} \sum_{k \neq j} \frac{\lambda_k(\Gamma_N)}{\lambda_j(\Gamma_N) - \lambda_k(\Gamma_N)} \\ \vdots \\ \frac{\lambda_m(S_N(\Gamma))}{\lambda_m(\Gamma_N)} - 1 - \frac{1}{n} \sum_{k \neq m} \frac{\lambda_k(\Gamma_N)}{\lambda_m(\Gamma_N) - \lambda_k(\Gamma_N)} \end{pmatrix}. \quad (3.19)$$

We establish the CLT for the vector $\Lambda_m(\Gamma_N)$ under the following assumptions:

A1 The spectral norm $\lambda_1(\Gamma_N)$ is bounded away from zero, and there exists a constant $\kappa > 0$ such that

$$\lambda_m(\Gamma_N) \geq \kappa \lambda_1(\Gamma_N).$$

A2 (Multiple spectral gap conditions) For any $i = 1, \dots, m$, we assume that

$$\overline{\lim}_{N \rightarrow \infty} \frac{\lambda_{i+1}(\Gamma_N)}{\lambda_i(\Gamma_N)} < 1.$$

Remark 9. Note that the multiple spectral gap conditions assumed in **A2** ensures that the m eigenvalues $\lambda_1(\Gamma_N), \dots, \lambda_m(\Gamma_N)$ are asymptotically simple, so that the denominators in the definition of $\Lambda_m(\Gamma_N)$ are nonzero for large N , thus $\Lambda_m(\Gamma_N)$ is asymptotically well defined.

Remark 10. Note also that if the $m + 1$ nonnegative eigenvalues $\lambda_1(\Gamma_N), \dots, \lambda_{m+1}(\Gamma_N)$ satisfy the so-called separate limiting property, then **A2** holds. In fact if there exists a sequence β_N such that $\lambda_j(\Gamma_N)/\beta_N \rightarrow \ell_j$ for $j = 1, \dots, m + 1$ with ℓ_j distinct with each other, then $\ell_1 > \dots > \ell_{m+1} \geq 0$ and $\lambda_{j+1}(\Gamma_N)/\lambda_j(\Gamma_N) \rightarrow \ell_{j+1}/\ell_j < 1$ for $j = 1, \dots, m$.

A3

$$\lim_{N \rightarrow \infty} \frac{\text{tr} \Gamma_N}{\sqrt{N} \lambda_{\max}(\Gamma_N)} = 0 \quad \text{and} \quad \mathbb{E}|Z_{i,j}|^4 < \infty.$$

A4

$$\lim_{N \rightarrow \infty} \frac{\text{tr} \Gamma_N^2}{\sqrt{N} \lambda_{\max}^2(\Gamma_N)} = 0 \quad \text{and} \quad \mathbb{E}|Z_{i,j}|^8 < \infty.$$

Remark 11. For a positive semidefinite hermitian matrix Γ_N , if we consider the normalized matrix $\tilde{\Gamma}_N := \Gamma_N/\lambda_{\max}(\Gamma_N)$ and its **ESD** $\mu^{\tilde{\Gamma}_N} = N^{-1} \sum_k \delta_{\lambda_k(\tilde{\Gamma}_N)}$, then the condition **A3** (resp. **A4**) says exactly that the first (resp. second) moment of $\mu^{\tilde{\Gamma}_N}$ is of order $o(1/\sqrt{N})$. Note that the eigenvalues of $\tilde{\Gamma}_N$ are in $[0, 1]$, so the general inequality

$$\lim_{N \rightarrow \infty} \frac{\text{tr} \Gamma_N^2}{\sqrt{N} \lambda_{\max}^2(\Gamma_N)} \leq \lim_{N \rightarrow \infty} \frac{\text{tr} \Gamma_N}{\sqrt{N} \lambda_{\max}(\Gamma_N)}$$

holds. Note also that by Chebyshev's inequality, if the first or second moment of $\mu^{\tilde{\Gamma}_N}$ converges to 0, then $\mu^{\tilde{\Gamma}_N} \xrightarrow{\mathcal{D}} \delta_0$.

Theorem III.4. *Under **A1**, **A2**, and either **A3** or **A4**, we have*

$$d_{\mathcal{LP}}(\Lambda_m(\Gamma_N), \mathcal{N}(0, \Sigma_m^{(N)})) \xrightarrow[N, n \rightarrow \infty]{} 0, \quad (3.20)$$

where $\Sigma_m^{(N)} = I_m + (\sigma_{i,j}^{(N)})_{i,j=1}^m$ with

$$\sigma_{i,j}^{(N)} = (\mathbb{E}|Z_{1,1}|^4 - |\mathbb{E}Z_{1,1}^2|^2 - 2) \sum_{k=1}^N |u_{k,i}|^2 |u_{k,j}|^2 + |\mathbb{E}Z_{1,1}^2|^2 \left| \sum_{k=1}^N u_{k,i} u_{k,j} \right|^2. \quad (3.21)$$

and $u_j := (u_{j,1}, \dots, u_{j,N})^\top$ is a normalized eigenvector associated with $\lambda_j(\Gamma_N)$.

Remark 12. In view of the expression (3.21), it is not clear that the covariance matrix $\Sigma_m^{(N)}$ converges. In order to avoid any cumbersome assumption enforcing this convergence, we express the CLT with the help of Lévy-Prokhorovs distance. If however it happens that $\Sigma_m^{(N)}$ converges to some matrix Σ_m , then we conclude the CLT in the following usual form

$$\Lambda_m(\Gamma_N) \xrightarrow[N, n \rightarrow \infty]{\mathcal{D}} \mathcal{N}(0, \Sigma_m).$$

3.3.3 The CLT for covariance matrices of long memory processes

Let T_N be defined as in Section 3.3.1, assume in addition that γ is positive definite, this implies that T_N is positive semidefinite for all N . By Polya Theorem (c.f. [31, Theorem 3.5.22]), a simple example of positive definite function is $\gamma : h \mapsto (1 + |h|)^\rho$ with $\rho \in (-1, 0)$. Let $S_N(T) = n^{-1} T_N^{\frac{1}{2}} Z_N Z_N^* T_N^{\frac{1}{2}}$ with $Z_N = (Z_{i,j})$ the same as in $S_N(\Gamma)$.

For an arbitrary integer $m \geq 1$, we define analogously $\Lambda_m(T_N)$ as in (3.19) by replacing $S_N(\Gamma)$ with $S_N(T)$, and Γ_N with T_N . And by Theorem III.2, any finite number of top eigenvalues of T_N have separate limiting property, thus $\Lambda_m(T_N)$ is asymptotically well defined.

Theorem III.5. *Let T_N and $S_N(T)$ be defined as before. If one of the following is satisfied:*

1. *The parameter ρ belongs to $(-1/2, 0)$ or equivalently $d \in (1/4, 1/2)$, and $\mathbb{E}|Z_{i,j}|^4 < \infty$;*
2. *The parameter ρ belongs to $(-3/4, -1/2]$ or equivalently $d \in (1/8, 1/4]$, and $\mathbb{E}|Z_{i,j}|^8 < \infty$,*

then we have

$$\Lambda_m(T_N) \xrightarrow[N, n \rightarrow \infty]{\mathcal{D}} \mathcal{N}(0, (1 + |\mathbb{E}Z_{i,j}^2|^2)I_m).$$

Theorem III.5 is a direct corollary of Theorem III.1, III.2, Proposition III.3 and Theorem III.4. Note that by Theorem III.1, III.2 the Toeplitz matrices T_N satisfy the assumptions A1 and A2, and Proposition III.3 ensures that T_N satisfies A3 or A4 for different ρ . Thus Theorem III.4 applies to the model $S_N(T)$. The limit of covariance matrices $\Sigma_m^{(N)}$ can be determined by the asymptotics of eigenvectors of T_N in Theorem III.2, from which we have

$$\Sigma_m^{(N)} \rightarrow (1 + |\mathbb{E}Z_{1,1}^2|^2)I_m,$$

and the result follows.

Remark 13. Comparing the assumptions of Theorem III.4 and those of [20], we can see that [20] requires that all the eigenvalues of Γ_N , except for K of them with $K \ll N^{1/6}$, are uniformly bounded by a fixed constant C . This condition cannot be fulfilled by most of T_N . Indeed, let us take the third example in Section 3.1 of [42], where

$$T_N := \left(\frac{1}{2\pi} \int_{-\pi}^{\pi} |x|^{-2d} e^{i(j-k)x} dx \right)_{j,k=1}^N$$

with $d \in (0, 1/2)$, $\mathbf{i} = \sqrt{-1}$. This Toeplitz matrix has an unbounded spectral density $f(x) = |x|^{-2d}$, so by a result of Tyrtshnikov and Zamarashkin [55], for any fixed bound

$C > 0$, the number of eigenvalues of T_N larger than C is asymptotically $(\pi^{-1}C^{-\frac{1}{2a}})N$. Note that this is always of order $O(N)$ no matter how large C is.

Inversely, we can also find some models which satisfy the assumptions in [20] but not the assumptions in the present paper. For instance let Γ_N be a $N \times N$ hermitian matrix such that $\lambda_{\max}(\Gamma_N) = \log N$ but all other eigenvalues equal to 1. Then $S_N(\Gamma)$ satisfies all assumptions in [20] but not ours.

3.4 Proofs of the theorems on Toeplitz matrices

3.4.1 Some preparation

Recall that the Toeplitz matrix T_N is defined in Section 3.3.1 as

$$T_N := (\gamma(i-j))_{i,j=1}^N \quad \text{where} \quad \gamma(h) = L(|h|)(1+|h|)^\rho$$

with L a slowly varying function at infinity and $\rho \in (-1, 0)$. Note that by the definition of slowly varying function, $\gamma(N)$ is positive for N sufficiently large.

For $p \in [1, \infty]$, and for N sufficient such that $\gamma(N) \neq 0$, we define a finite-rank operator $\mathcal{K}_N^{(\gamma)}$ acting on $L^p(0, 1)$ by

$$(\mathcal{K}_N^{(\gamma)} f)(x) = \int_0^1 \frac{\gamma(\lfloor Nx \rfloor - \lfloor Ny \rfloor)}{\gamma(N)} f(y) dy. \quad (3.22)$$

The operator $\mathcal{K}^{(\rho)}$ in (3.14) is also well-defined for any $f \in L^p(0, 1)$ by the integral formula:

$$(\mathcal{K}^{(\rho)} f)(x) = \int_0^1 |x-y|^\rho f(y) dy. \quad (3.23)$$

The operators $\mathcal{K}_N^{(\gamma)}$ and $\mathcal{K}^{(\rho)}$ acting on $L^p(0, 1)$ are bounded, see [42, Lemma 5.4]. Moreover, from Lemma 5.4 of [42], we have the convergence

$$\|\mathcal{K}_N^{(\gamma)} - \mathcal{K}^{(\rho)}\|_p \rightarrow 0 \quad \text{as} \quad N \rightarrow \infty, \quad \forall p \in [1, \infty]. \quad (3.24)$$

The above convergence (3.24) has many useful consequences in this proof. The first conse-

quence is that the operator $\mathcal{K}^{(\rho)}$ is compact on $L^p(0, 1)$ for any $p \in [1, \infty]$.

For each $p \in [1, \infty]$, $\mathcal{K}^{(\rho)}$ (resp. $\mathcal{K}_N^{(\gamma)}$) has its spectrum as an operator acting on $L^p(0, 1)$. The following proposition shows that its nonzero eigenvalues and the associated eigenfunctions are invariant as p changes.

Proposition III.6. *The nonzero eigenvalues and the associated eigenfunctions of $\mathcal{K}^{(\rho)} : L^p(0, 1) \rightarrow L^p(0, 1)$ and $\mathcal{K}_N^{(\gamma)} : L^p(0, 1) \rightarrow L^p(0, 1)$ do not change when p runs across $[1, \infty]$.*

Proof. We only prove the result for $\mathcal{K}^{(\rho)}$. For the other operator $\mathcal{K}_N^{(\gamma)}$ one can prove in the same way.

We only need to prove that, for any $p \in [1, \infty)$, the operator $\mathcal{K}^{(\rho)} : L^p(0, 1) \rightarrow L^p(0, 1)$ has the same nonzero eigenvalues and associated eigenfunctions as $\mathcal{K}^{(\rho)} : L^\infty(0, 1) \rightarrow L^\infty(0, 1)$. As we have already noticed that $\mathcal{K}^{(\rho)}$ is compact on $L^p(0, 1)$ and on $L^\infty(0, 1)$, the desired result is a direct application of Theorem 4.2.15 in [24].

Indeed, we recall that two Banach spaces B_1 and B_2 or their associated norms are said to be compatible if $B = B_1 \cap B_2$ is dense in each of them, and the following condition is satisfied: if $f_n \in B$, $\|f_n - f\|_{B_1} \rightarrow 0$ and $\|f_n - g\|_{B_2} \rightarrow 0$, then $f = g \in B$. The operators $A_i : B_i \rightarrow C_i$ with $i = 1, 2$ and C_1, C_2 two Banach spaces, are said to be consistent if $A_1 f = A_2 f$ for all $f \in B_1 \cap B_2$. Then we can verify that $L^p(0, 1)$ and $L^\infty(0, 1)$ are compatible, and $\mathcal{K}^{(\rho)}$ defined by an integral formula is obviously consistent. Then Theorem 4.2.15 in [24] applies. \square

According to the above proposition, when we talk about the nonzero eigenvalues and the associated eigenfunctions of these operators, we do not need to specify the space.

3.4.2 Proof of Theorem III.1

Compactity and semidefinite positivity, infinitely many positive eigenvalues. The compactity of $\mathcal{K}^{(\rho)}$ is already proved by the convergence (3.24). Also by this convergence, we prove the positivity semidefinite of $\mathcal{K}^{(\rho)}$ on $L^2(0, 1)$. Indeed, taking $L \equiv 1$, by Polya Theorem, $\frac{T_N}{N\gamma(N)}$ are positive semidefinite for any N , it does not have negative eigenvalues. Hence $\mathcal{K}_N^{(\gamma)}$ does not have negative eigenvalues, since by Widom-Shampine Lemma [42, Lemma 5.1], $T_N/(N\gamma(N))$ has the same nonzero eigenvalues as $\mathcal{K}_N^{(\gamma)}$. Then for any $f \in$

$L^2(0, 1)$, we have

$$\langle \mathcal{K}_N^{(\gamma)} f, f \rangle \geq 0.$$

Let $N \rightarrow \infty$, from (3.24) we have

$$\langle \mathcal{K}^{(\rho)} f, f \rangle \geq 0.$$

The fact that $\mathcal{K}^{(\rho)}$ has infinitely many positive eigenvalues is implicitly implied by Widom's formula in [58], which writes (after some calculation and rearrangement)

$$\lambda_j(\mathcal{K}^{(\rho)}) \sim 4\pi^{-\rho} \frac{\Gamma(\rho)}{\Gamma(\frac{\rho}{2}) \Gamma(-\frac{\rho}{2})} \frac{1}{j^{1+\rho}} \quad \text{as } j \rightarrow \infty.$$

Continuity of eigenfunctions associated with nonzero eigenvalues. Let $\lambda \neq 0$ be an eigenvalue of $\mathcal{K}^{(\rho)}$ and f be an associated eigenfunction. We now prove that f is continuous on $[0, 1]$.

Note that f satisfies the equation

$$f(x) = \lambda^{-1} \int_0^1 |y - x|^\rho f(y) \, dy,$$

and from Proposition III.6, f also belongs to $L^\infty(0, 1)$. So for any $x_0 \in [0, 1]$, one has

$$|f(x) - f(x_0)| \leq \lambda^{-1} \|f\|_\infty \int_0^1 \left| |y - x|^\rho - |y - x_0|^\rho \right| \, dy \leq \lambda^{-1} \|f\|_\infty \int_{-1}^1 \left| |y - x + x_0|^\rho - |y|^\rho \right| \, dy,$$

and the integral on the RHS tends to 0 when $|x - x_0| \rightarrow 0$.

Simplicity of nonzero eigenvalues. We prove that all nonzero eigenvalues of $\mathcal{K}^{(\rho)}$ are simple. We need the following key lemma. It says that any normalized eigenfunction of $\mathcal{K}^{(\rho)}$ associated with a nonzero eigenvalue, taken at $x = 1$, has the absolute value $\sqrt{1 + \rho}$.

Lemma III.7. *Let $\lambda > 0$ be a nonzero eigenvalue of $\mathcal{K}^{(\rho)}$, and let f be a normalized eigenfunction associated with λ . Then f satisfies*

$$|f(1)| = \sqrt{1 + \rho}. \tag{3.25}$$

A result similar to the above lemma first appeared in [46] for a general but square integrable kernel $k(x - y)$, see Theorem 3 of [46]. Note that thanks to the explicit formula of $\mathcal{K}^{(\rho)}$, the result of Lemma III.7 is stronger than [46]. Directly using Theorem 3 of [46], we can only conclude that for $\rho \in (-1/2, 0)$, for any nonzero eigenvalue λ of $\mathcal{K}^{(\rho)}$, there exists a group of orthonormal eigenfunctions $f_{\lambda,1}, \dots, f_{\lambda,m}$ associated with λ , where m is the multiplicity of λ , such that

$$|f_{\lambda,i}(1)| = \sqrt{1 + \rho}.$$

However we will notice later that this result is not sufficient to our need.

Whenever Lemma III.7 is proved, we can prove the simplicity of any nonzero eigenvalue λ of $\mathcal{K}^{(\rho)}$ by contradiction. Assume to the contrary that $\lambda > 0$ had multiplicity $m \geq 2$, then we could choose two orthonormal eigenfunctions $f_{\lambda,1}, f_{\lambda,2}$ associated with λ . From Lemma 3.25, without loss of generality we can assume that $f_{\lambda,1}(1) = f_{\lambda,2}(1) = \sqrt{1 + \rho}$. Then the function

$$f_\lambda := \frac{1}{\sqrt{2}}f_{\lambda,1} + \frac{1}{\sqrt{2}}f_{\lambda,2}$$

is also a normalized eigenfunction of $\mathcal{K}^{(\rho)}$. But this function satisfies

$$f_\lambda(1) = \sqrt{2(1 + \rho)} \neq \sqrt{1 + \rho},$$

which is a contradiction to Lemma III.7.

Proof of Lemma III.7. We follow the outline of the proof in [46]. For any $\tau > 0$, we define $\mathcal{K}_\tau^{(\rho)}$ the operator on $L^2(0, \tau)$ by

$$(\mathcal{K}_\tau^{(\rho)} f)(x) = \int_0^\tau |x - y|^\rho f(y) dy. \tag{3.26}$$

By a change of variable, it is easy to see that a function $f \in L^2(0, 1)$ is an eigenfunction of $\mathcal{K}^{(\rho)}$ associated with an eigenvalue λ if and only if $f(\frac{\cdot}{\tau})$ is an eigenfunction of $\mathcal{K}_\tau^{(\rho)}$ associated with the eigenvalue $\lambda\tau^{1+\rho}$. By this fact, a positive number λ is an eigenvalue of $\mathcal{K}^{(\rho)}$ with multiplicity m if and only if $\lambda\tau^{1+\rho}$ is an eigenvalue of $\mathcal{K}_\tau^{(\rho)}$ with the same multiplicity m for all $\tau > 0$.

Suppose that f is a normalized eigenfunction of $\mathcal{K}^{(\rho)}$ associated with nonzero eigenvalue $\lambda > 0$. Then for any $\tau > 1$ we have the following two equations

$$\lambda\tau^{1+\rho}f\left(\frac{x}{\tau}\right) = \int_0^\tau |x-y|^\rho f\left(\frac{y}{\tau}\right) dy, \text{ for } x \in (0, \tau) \quad (3.27)$$

and

$$\lambda\overline{f(y)} = \int_0^1 |x-y|^\rho \overline{f(x)} dx, \text{ for } y \in (0, 1). \quad (3.28)$$

We define the function g on $[0, \infty)$ by

$$g(y) = \frac{1}{\lambda} \int_0^1 |x-y|^\rho f(x) dx, \text{ for } y \in [0, \infty),$$

then g is a continuous extension of f on $[0, \infty)$. Multiply the two sides of (3.27) by $\overline{f(x)}$, and integrate for $x \in [0, 1]$, we get

$$\lambda\tau^{1+\rho} \int_0^1 f\left(\frac{x}{\tau}\right) \overline{f(x)} dx = \int_0^1 \int_0^\tau |x-y|^\rho f\left(\frac{y}{\tau}\right) \overline{f(x)} dy dx. \quad (3.29)$$

Note that by the boundedness of f , Fubini Theorem applies to the RHS, thus changing the order of two integrations and taking into account the definition of g , we get

$$\tau^{1+\rho} \int_0^1 f\left(\frac{x}{\tau}\right) \overline{f(x)} dx = \int_0^\tau \overline{g(y)} f\left(\frac{y}{\tau}\right) dy. \quad (3.30)$$

It is easy to see from (3.30) that

$$\left(\frac{\tau^{1+\rho} - 1}{\tau - 1}\right) \int_0^1 f\left(\frac{x}{\tau}\right) \overline{f(x)} dx = \frac{\int_1^\tau \overline{g(y)} f\left(\frac{y}{\tau}\right) dy}{\tau - 1}. \quad (3.31)$$

Letting $\tau \rightarrow 1^+$ on the two sides of (3.31), and noting that the continuity of f on $[0, 1]$ implies the uniform convergence of $f\left(\frac{x}{\tau}\right)$ to $f(x)$, we get

$$1 + \rho = |f(1)|^2 \quad (3.32)$$

and the result follows. □

3.4.3 Proof of Theorem III.2

For any $j \geq 1$, the convergence (3.15) was already proved in [42, Theorem 2.3]. So we only have to prove the asymptotic behaviour (3.16) of eigenvectors.

By Theorem III.1, we have $\lambda_j(\mathcal{K}^{(\rho)}) > 0$. Also according to the simplicity of nonzero eigenvalues, we have $\lambda_{j-1}(\mathcal{K}^{(\rho)}) > \lambda_j(\mathcal{K}^{(\rho)}) > \lambda_{j+1}(\mathcal{K}^{(\rho)})$. Let f_j be a normalized eigenfunction associated with $\lambda_j(\mathcal{K}^{(\rho)})$. In the sequel, we shall rely on the spectral projections (to be defined later) to construct an eigenvector of T_N associated with $\lambda_j(T_N)$ and prove that such an eigenvector approximates f_j in the sense of (3.16).

Let $\varepsilon = \frac{1}{2} \min(\lambda_{j-1}(\mathcal{K}^{(\rho)}) - \lambda_j(\mathcal{K}^{(\rho)}), \lambda_j(\mathcal{K}^{(\rho)}) - \lambda_{j+1}(\mathcal{K}^{(\rho)}))$ and \mathcal{C} be the circle centered at $\lambda_j(\mathcal{K}^{(\rho)})$ and of radius ε on complex plane. We take N sufficiently large such that $\|\mathcal{K}_N^{(\gamma)} - \mathcal{K}^{(\rho)}\|_\infty < \varepsilon$. So we have $|\lambda_k(\mathcal{K}_N^{(\gamma)}) - \lambda_k(\mathcal{K}^{(\rho)})| < \varepsilon$ for all $1 \leq k \leq N$, which implies that only the eigenvalues $\lambda_j(\mathcal{K}_N^{(\gamma)})$ and $\lambda_j(\mathcal{K}^{(\rho)})$ are enclosed by \mathcal{C} and all the other eigenvalues are outside \mathcal{C} . We define the spectral projections

$$P_N := \frac{1}{2\pi i} \int_{\mathcal{C}} (z - \mathcal{K}_N^{(\gamma)})^{-1} dz \quad \text{and} \quad P := \frac{1}{2\pi i} \int_{\mathcal{C}} (z - \mathcal{K}^{(\rho)})^{-1} dz. \quad (3.33)$$

By Riesz decomposition Theorem (c.f. for example [24, Theorem 1.5.4 and Theorem 4.3.19]), P_N (resp. P) is a projection onto the eigenspace of $\mathcal{K}_N^{(\gamma)}$ (resp. $\mathcal{K}^{(\rho)}$) corresponding to $\lambda_j(\mathcal{K}_N^{(\gamma)})$ (resp. $\lambda_j(\mathcal{K}^{(\rho)})$). To those who are unfamiliar with Riesz' Theorem, we explain the arguments with $\mathcal{K}^{(\rho)}$ and P . Indeed, from Riesz' Theorem, P is a finite rank projection which commutes with $\mathcal{K}^{(\rho)}$. Let $\mathcal{R}(P)$ be the range of P , then $\mathcal{R}(P)$ is an invariant space of $\mathcal{K}^{(\rho)}$ (due to the commutativity of the projection P and $\mathcal{K}^{(\rho)}$), and the restriction of $\mathcal{K}^{(\rho)}$ to $\mathcal{R}(P)$ is self-adjoint (because $\mathcal{K}^{(\rho)}$ is self-adjoint) and has spectrum $\{\lambda_j(\mathcal{K}^{(\rho)})\}$, then from the finite dimensional linear algebra, $\mathcal{R}(P)$ is spanned by the eigenfunctions of $\mathcal{K}^{(\rho)}$ associated with $\lambda_j(\mathcal{K}^{(\rho)})$. Therefore, recall that f_j is a normalized eigenfunction of $\mathcal{K}^{(\rho)}$ associated with $\lambda_j(\mathcal{K}^{(\rho)})$, we have $Pf_j = f_j$. The same argument shows that P_N is a projection to the eigenspace of $\mathcal{K}_N^{(\gamma)}$ and thus $P_N f_j$ is an eigenfunction of $\mathcal{K}_N^{(\gamma)}$ associated with $\lambda_j(\mathcal{K}_N^{(\gamma)})$, in condition that $P_N f_j \neq 0$.

We prove that $\|P_N - P\|_\infty \rightarrow 0$. Indeed we have

$$\|P_N - P\|_\infty \leq \frac{1}{2\pi} \int_{\mathcal{C}} \|(z - \mathcal{K}_N^{(\gamma)})^{-1} - (z - \mathcal{K}^{(\rho)})^{-1}\|_\infty |dz|.$$

Thus the main task is to uniformly control $\|(z - \mathcal{K}_N^{(\gamma)})^{-1} - (z - \mathcal{K}^{(\rho)})^{-1}\|_\infty$ in term of $\|\mathcal{K}_N^{(\gamma)} - \mathcal{K}^{(\rho)}\|_\infty$ for $z \in \mathcal{C}$. As $(z - \mathcal{K}^{(\rho)})^{-1}$ is analytic outside of $\text{Spec}(\mathcal{K}^{(\rho)})$, there exists $C > 0$ such that $\sup_{z \in \mathcal{C}} \|(z - \mathcal{K}^{(\rho)})^{-1}\|_\infty \leq C$. Let N be sufficiently large such that $\|\mathcal{K}_N^{(\gamma)} - \mathcal{K}^{(\rho)}\|_\infty < 1/(2C)$. Then we have

$$\begin{aligned} \|(z - \mathcal{K}_N^{(\gamma)})^{-1} - (z - \mathcal{K}^{(\rho)})^{-1}\|_\infty &= \|(z - \mathcal{K}^{(\rho)} - \mathcal{K}_N^{(\gamma)} + \mathcal{K}^{(\rho)})^{-1} - (z - \mathcal{K}^{(\rho)})^{-1}\|_\infty \\ &= \left\| (z - \mathcal{K}^{(\rho)})^{-1} \left[(I - (\mathcal{K}_N^{(\gamma)} - \mathcal{K}^{(\rho)})(z - \mathcal{K}^{(\rho)})^{-1})^{-1} - I \right] \right\|_\infty \\ &= \left\| (z - \mathcal{K}^{(\rho)})^{-1} \sum_{k=1}^{\infty} ((\mathcal{K}_N^{(\gamma)} - \mathcal{K}^{(\rho)})(z - \mathcal{K}^{(\rho)})^{-1})^k \right\|_\infty \\ &\leq \sum_{k=1}^{\infty} \frac{C^{k+1}}{2^{k-1} C^{k-1}} \|\mathcal{K}_N^{(\gamma)} - \mathcal{K}^{(\rho)}\|_\infty^k \\ &= 2C^2 \|\mathcal{K}_N^{(\gamma)} - \mathcal{K}^{(\rho)}\|_\infty. \end{aligned}$$

Thus as $N \rightarrow \infty$ we have

$$\|P_N - P\|_\infty \leq \frac{1}{2\pi} \int_{\mathcal{C}} \|(z - \mathcal{K}_N^{(\gamma)})^{-1} - (z - \mathcal{K}^{(\rho)})^{-1}\|_\infty |dz| \leq 2\varepsilon C^2 \|\mathcal{K}_N^{(\gamma)} - \mathcal{K}^{(\rho)}\|_\infty \rightarrow 0.$$

From this convergence we have

$$\|P_N f_j - f_j\|_\infty = \|P_N f_j - P f_j\|_\infty \xrightarrow{N \rightarrow \infty} 0. \quad (3.34)$$

Then from (3.34) we obtain

$$\|P_N f_j\|_2 \xrightarrow{N \rightarrow \infty} \|f_j\|_2 = 1. \quad (3.35)$$

Combining (3.34) and (3.35) we conclude

$$\left\| \frac{P_N f_j}{\|P_N f_j\|_2} - f_j \right\|_\infty \xrightarrow{N \rightarrow \infty} 0. \quad (3.36)$$

Notice that the range of $\mathcal{K}_N^{(\gamma)}$ consists of step functions

$$f(x) = \sum_{k=1}^N v_k \mathbb{1}_{[\frac{k-1}{N}, \frac{k}{N})}(x),$$

so the eigenfunctions of $\mathcal{K}_N^{(\gamma)}$ must also have this form. Notice also that a N -dimensional normalized vector $v = (v_k)_{k=1}^N$ is an eigenvector of T_N associated with $\lambda_j(T_N)$ if and only if the normalized function

$$f(x) = \sqrt{N} \sum_{k=1}^N v_k \mathbb{1}_{[\frac{k-1}{N}, \frac{k}{N})}(x) \quad (3.37)$$

is an eigenfunction of $\mathcal{K}_N^{(\gamma)}$ associated with $\lambda_j(\mathcal{K}_N^{(\gamma)}) = \lambda_j(T_N)/(N\gamma(N))$. Since $Pf_j/\|Pf_j\|_2$ is a normalized eigenfunction of $\mathcal{K}_N^{(\gamma)}$, we construct the eigenvector of T_N from $Pf_j/\|Pf_j\|_2$ by the relation (3.37). So we define

$$u_{k,j} := \frac{(Pf_j)_{(k-1)}}{\sqrt{N}\|Pf_j\|_2} \quad (3.38)$$

and let $u_j = (u_{1,j}, \dots, u_{N,j})^\top$, then u_j is a normalized eigenvector of T_N associated with $\lambda_j(T_N)$. From (3.36) we get the desired result (3.16).

3.4.4 Proof of Proposition III.3.

First we prove Item 1. Assume $\rho \in (-1/2, 0)$, from Theorem III.2, we have

$$\lambda_{\max}(T_N) \sim \lambda_{\max}(\mathcal{K}^{(\rho)})N^{1+\rho}L(N).$$

Since $N^{1/2+\rho}L(N) \rightarrow \infty$, we have

$$\frac{\text{tr } T_N}{\sqrt{N}\lambda_{\max}(T_N)} = \frac{\sqrt{N}\gamma(0)}{\lambda_{\max}(T_N)} \sim \frac{\gamma(0)}{\lambda_{\max}(\mathcal{K}^{(\rho)})N^{1/2+\rho}L(N)} \rightarrow 0.$$

Then we prove Item 2. Assume $\rho \in (-3/4, -1/2]$. Note that

$$\text{tr } T_N^2 = \|T_N\|_F^2 \leq 2N \sum_{k=0}^N |\gamma(k)|^2.$$

Also from Theorem III.2, we have $\lambda_{\max}^2(T_N) \sim \lambda_{\max}^2(\mathcal{K}(\rho))N^{2+2\rho}L^2(N) \gg N^{1/2+\epsilon}$ for some $\epsilon > 0$. We then have

$$\frac{\text{tr } T_N^2}{\sqrt{N}\lambda_{\max}^2(T_N)} \ll \frac{\sum_{k=0}^N |\gamma(k)|^2}{N^\epsilon} \leq \frac{1}{N^{\epsilon/2}} \left(\gamma(0) + \sum_{k=1}^N \frac{|\gamma(k)|^2}{k^{\epsilon/2}} \right).$$

Since $|\gamma(k)|^2 \sim L^2(k)k^{2\rho} \leq L^2(k)k^{-1}$, it follows that

$$\sum_{k=1}^{\infty} \frac{|\gamma(k)|^2}{k^{\epsilon/2}} \leq \sum_{k=1}^{\infty} \frac{L^2(k)}{k^{1+\epsilon/2}} < \infty.$$

Hence

$$\frac{\text{tr } T_N^2}{\sqrt{N}\lambda_{\max}^2(T_N)} \ll \frac{1}{N^{\epsilon/2}} \left(\gamma(0) + \sum_{k=1}^{\infty} \frac{L^2(k)}{k^{1+\epsilon/2}} \right) \rightarrow 0.$$

The proof is complete.

3.5 Proof of Theorem III.4

3.5.1 Some preparation and outline of the proof

In order to prove that the sequence $\left(d_{\mathcal{LP}}(\Lambda_m(\Gamma_N), \mathcal{N}(0, \Sigma_m^{(N)})) \right)_{N \geq 1}$ converges to 0, it suffices to prove that from any subsequence we can extract a subsequence converging to 0. Since by the definition of $\Sigma_m^{(N)}$ its components are uniformly bounded, from any subsequence of the $m \times m$ matrix sequence $\left(\Sigma_m^{(N)} \right)_{N \geq 1}$ we can extract a subsequence whose components converge. Without loss of generality we can assume that $\Sigma_m^{(N)} \rightarrow \Sigma_m$ and prove that

$$\Lambda_m(\Gamma_N) \xrightarrow[N, n \rightarrow \infty]{\mathcal{D}} \mathcal{N}(0, \Sigma_m), \quad (3.39)$$

which is equivalent to

$$d_{\mathcal{LP}}(\Lambda_m(\Gamma_N), \mathcal{N}(0, \Sigma_m)) \rightarrow 0,$$

and because $d_{\mathcal{LP}}(\mathcal{N}(0, \Sigma_m^{(N)}), \mathcal{N}(0, \Sigma_m)) \rightarrow 0$, then the convergence (3.20) follows.

Now to prove the convergence (3.39), it suffices to prove that for any fixed vector

$(b_1, \dots, b_m)^\top \in \mathbb{R}^m$, we have

$$\mathbb{P}\left(\sqrt{n}(\tilde{\lambda}_i - \theta_i) < b_i \quad \text{for } 1 \leq i \leq m\right) \xrightarrow{N, n \rightarrow \infty} \mathbb{P}(G_i < b_i \quad \text{for } 1 \leq i \leq m), \quad (3.40)$$

where we define

$$\tilde{\lambda}_i = \frac{\lambda_i(S_N(\Gamma))}{\lambda_i(\Gamma_N)}, \quad \text{and} \quad \theta_i = 1 + \frac{1}{n} \sum_{k \neq i} \frac{\lambda_k(\Gamma_N)}{\lambda_i(\Gamma_N) - \lambda_k(\Gamma_N)}, \quad (3.41)$$

and

$$(G_1, \dots, G_m)^\top \sim \mathcal{N}(0, \Sigma_m).$$

Let $E_{i,N}$ be the event

$$E_{i,N} := \{\sqrt{n}(\tilde{\lambda}_i - \theta_i) < b_i\}.$$

Then the event $E_N := \{\sqrt{n}(\tilde{\lambda}_i - \theta_i) < b_i \quad \text{for } 1 \leq i \leq m\}$ can be written as

$$E_N = \bigcap_{i=1}^m E_{i,N}.$$

Let

$$\mathfrak{Y}_{i,N} := \sqrt{n} \left(\frac{u_i^* Z_N Z_N^* u_i}{n} - 1 \right)$$

where we recall that u_1, \dots, u_m are orthonormal eigenvectors of Γ_N associated with its m largest eigenvalues, and let

$$\mathfrak{Y}_N := (\mathfrak{Y}_{1,N}, \dots, \mathfrak{Y}_{m,N})^\top. \quad (3.42)$$

In order to prove (3.40) we will prove the following two lemmas. The first lemma allows to rewrite the event E_N in term of \mathfrak{Y}_N , and the second one establishes a CLT for \mathfrak{Y}_N .

Lemma III.8. *Let $1 \leq i \leq m$ be a fixed number. Under the assumptions of Theorem III.4, as $N, n \rightarrow \infty$, the events $E_{i,N}$ can be rewritten as*

$$E_{i,N} = \{\mathfrak{Y}_{i,N} + o_P(1) < b_i\}. \quad (3.43)$$

Lemma III.8 allows to transform the fluctuations of eigenvalues $(\tilde{\lambda}_1, \dots, \tilde{\lambda}_m)$ to the fluctuations of quadratic forms $(\mathfrak{Y}_{1,N}, \dots, \mathfrak{Y}_{m,N})$, i.e.

$$\mathbb{P}\left(\sqrt{n}(\tilde{\lambda}_i - \theta_i) < b_i, \quad 1 \leq i \leq m\right) = \mathbb{P}\left(\mathfrak{Y}_{i,N} + o_P(1) < b_i, 1 \leq i \leq m\right).$$

Then we only need to study the fluctuations of quadratic forms $(\mathfrak{Y}_{1,N}, \dots, \mathfrak{Y}_{m,N})$. This is the goal of the following lemma.

Note that the proof of the above lemma is the main difficulty of Theorem III.4. Precisely, we have to study the fluctuations of a quadratic form $\Upsilon(\lambda) = n^{-1}Y_{i,\cdot}(\lambda I - \widehat{\underline{S}}_{-i,-i})^{-1}Y_{i,\cdot}^*$ defined in (3.49), where $Y_{i,\cdot}$ is a one-row matrix, and $\widehat{\underline{S}}_{-i,-i}$ is a $(N-1) \times (N-1)$ matrix. We will see that the two matrices are dependent thus the method that we have used in [42] no longer works.

Lemma III.9. *Let \mathfrak{Y}_N be defined by (3.42) with $Z = (Z_{i,j})$ and u_i the same meaning as before. Let $\Sigma_m^{(N)}$ be defined as in Theorem III.4. If*

$$\Sigma_m^{(N)} \xrightarrow[N \rightarrow \infty]{} \Sigma_m$$

and if

$$\mathbb{E}Z_{i,j} = 0, \quad \mathbb{E}|Z_{i,j}|^2 = 1, \quad \mathbb{E}|Z_{i,j}|^4 < \infty,$$

then we have

$$\mathfrak{Y}_N \xrightarrow[N, n \rightarrow \infty]{\mathcal{D}} \mathcal{N}(0, \Sigma_m). \quad (3.44)$$

This will complete the proof because from the above two lemmas we have

$$\begin{aligned} \mathbb{P}\left(\sqrt{n}(\tilde{\lambda}_i - \theta_i) < b_i, \quad 1 \leq i \leq m\right) &= \mathbb{P}\left(\mathfrak{Y}_{i,N} + o_P(1) < b_i, 1 \leq i \leq m\right) \\ &\rightarrow \mathbb{P}(G_i < b_i, 1 \leq i \leq m), \end{aligned}$$

where in the last convergence we used Slutsky's Theorem.

Thus in the following we prove Lemma III.8 and III.9.

3.5.2 Proof of Lemma III.8

The method used here is based on the proof of Proposition 2.1 and Theorem 2.2 in [42] with some necessary adaptations. All through this proof i will denote a constant integer fixed in the statement of lemma. We mention that some notations in the proof may depend on i . To simplify the notation we keep the dependence on i implicit if it does not cause any ambiguity.

We normalize Γ_N and $S_N(\Gamma)$ with $\lambda_i(\Gamma_N)$, and set

$$\tilde{\Gamma}_N := \frac{\Gamma_N}{\lambda_i(\Gamma_N)}, \quad \text{and} \quad S_N(\tilde{\Gamma}) := \frac{1}{n} \tilde{\Gamma}_N^{\frac{1}{2}} Z_N Z_N^* \tilde{\Gamma}_N^{\frac{1}{2}} = \frac{S_N(\Gamma)}{\lambda_i(\Gamma_N)}.$$

Then $\lambda_i(\tilde{\Gamma}_N) \equiv 1$ and $\lambda_i(S_N(\tilde{\Gamma})) = \tilde{\lambda}_i$. Furthermore the combination of **A1** and either **A3** or **A4** gives

$$\mu^{\tilde{\Gamma}_N} \xrightarrow[N \rightarrow \infty]{\mathcal{D}} \delta_0, \quad \text{and} \quad \lambda_j(\tilde{\Gamma}_N) \in [c, C] \text{ for some } 0 < c < C < \infty \text{ and all } 1 \leq j \leq m.$$

Now we write $d_j := \lambda_j(\tilde{\Gamma}_N)$ for $j = 1, \dots, N$ and let $D_N := \text{diag}(d_1, \dots, d_N)$. Recall that $U_N = (u_1, \dots, u_N)$ with u_j the eigenvectors of Γ_N , then we have $\tilde{\Gamma}_N = U_N D_N U_N^*$. Let $Y_N := U_N^* Z_N$, and

$$\hat{S}_N := \frac{1}{n} D_N^{\frac{1}{2}} Y_N Y_N^* D_N^{\frac{1}{2}} = U_N^* S_N(\tilde{\Gamma}) U_N,$$

then \hat{S}_N has the same eigenvalues as $S_N(\tilde{\Gamma})$, so $\lambda_i(\hat{S}_N) = \tilde{\lambda}_i$.

Taking into account the assumption **A2**, there exists a positive number $\epsilon > 0$ such that for N sufficiently large,

$$d_{i+1} < 1 - 3\epsilon < d_i = 1 < 1 + 3\epsilon < d_{i-1},$$

where if $i = 1$ we do not have the last inequality. By a spectrum confinement argument similar to the proof of Proposition 2.1 in [42] (see especially Section 4.1-4.3 of [42]), we can prove that $\tilde{\lambda}_i \rightarrow 1$ in probability; and the event

$$\Omega'_N := \{\lambda_{i+1}(\hat{S}_N) < 1 - 2\epsilon < 1 - \epsilon < \tilde{\lambda}_i < 1 + \epsilon < 1 + 2\epsilon < \lambda_{i-1}(\hat{S}_N)\}$$

occurs with high probability. Similarly, setting

$$\widehat{S}_{-i,-i} := \frac{1}{n} D_{-i,-i}^{\frac{1}{2}} Y_{-i,\cdot} Y_{-i,\cdot}^* D_{-i,-i}^{\frac{1}{2}},$$

which is a matrix obtained by removing the i th row and i th column from the matrix \widehat{S}_N , then the event

$$\Omega_N'' := \{\lambda_i(\widehat{S}_{-i,-i}) < 1 - 2\epsilon < 1 + 2\epsilon < \lambda_{i-1}(\widehat{S}_{-i,-i})\}$$

occur with high probability. This is so because $\widehat{S}_{-i,-i}$ has the same nonzero eigenvalues as the matrix

$$\frac{1}{n} \text{diag}(d_1, \dots, d_{i-1}, 0, d_{i+1}, \dots)^{\frac{1}{2}} U_N^* Z_N Z_N^* U_N \text{diag}(d_1, \dots, d_{i-1}, 0, d_{i+1}, \dots)^{\frac{1}{2}}.$$

Then we set

$$\Omega_N := \Omega_N' \cap \Omega_N''$$

and we have $\mathbb{P}(\Omega_N) \rightarrow 1$.

Now we prove that there exists a random variable $R_{i,N} = o_P(1)$, such that

$$E_{i,N} = \{\mathfrak{Y}_{i,N} + R_{i,N} + o(1) < b_i\}.$$

Outside Ω_N , we can attribute to $R_{i,N}$ the value $\sqrt{n}(\tilde{\lambda}_i - \theta_i) - \mathfrak{Y}_{i,N}$. Let us now assume that we are in Ω_N . $\tilde{\lambda}_i$ being an eigenvalue of \widehat{S}_N , it satisfies the equation

$$\det(\tilde{\lambda}_i I - \widehat{S}_N) = 0. \tag{3.45}$$

A similar decomposition and reformulation of determinant as in the proof of Theorem 4.2 in [42] (see Section 4.4 of that paper, especially the equations (40) and (41)) shows that for all $\lambda \in \mathbb{C} \setminus (\text{Spec}(\widehat{S}_{-i,-i}) \cup \{0\})$, we have

$$\det(\lambda I - \widehat{S}_N) = \lambda \left(1 - \frac{1}{n} Y_{i,\cdot} (\lambda I - \widehat{S}_{-i,-i})^{-1} Y_{i,\cdot}^* \right) \det(\lambda I - \widehat{S}_{-i,-i}), \tag{3.46}$$

where

$$\widehat{\underline{S}}_{-i,-i} := \frac{1}{n} Y_{-i,\cdot}^* D_{-i,-i} Y_{-i,\cdot}.$$

From (3.45) and (3.46), we know that $\tilde{\lambda}_i$ satisfies the equation

$$1 - \frac{1}{n} Y_{i,\cdot} (\tilde{\lambda}_i I - \widehat{\underline{S}}_{-i,-i})^{-1} Y_{i,\cdot}^* = 0 \quad (3.47)$$

because $\tilde{\lambda}_i$ is not an eigenvalue of $\widehat{S}_{-i,-i}$ if Ω_N occurs.

Let

$$\eta_i := \theta_i + b_i/\sqrt{n} \quad (3.48)$$

and

$$\Upsilon(\lambda) := \frac{1}{n} Y_{i,\cdot} (\lambda I - \widehat{\underline{S}}_{-i,-i})^{-1} Y_{i,\cdot}^*, \quad (3.49)$$

Note that for λ sufficiently large, all the three terms $\det(\lambda I - \widehat{S}_N)$, $(1 - \Upsilon(\lambda))$ and $\det(\lambda I - \widehat{S}_{-i,-i})$ in (3.46) are positive. Note also that $\det(\lambda I - \widehat{S}_N)$ and $\det(\lambda I - \widehat{S}_{-i,-i})$ change sign $i - 1$ times respectively on $(1 + \epsilon, \infty)$, so we have

$$1 - \Upsilon(1 + \epsilon) > 0,$$

and $1 - \Upsilon(\lambda)$ changes sign in $(1 - \epsilon, 1 + \epsilon)$ exactly at $\tilde{\lambda}_i$. Let N, n sufficiently large such that $\theta_i, \eta_i \in (1 - \epsilon, 1 + \epsilon)$, then we have

$$E_{i,N} \cap \Omega_N = \{\Upsilon(\eta_i) < 1\} \cap \Omega_N. \quad (3.50)$$

We define

$$\mathfrak{x}_{i,N}(y) := \frac{1}{y} + \frac{1}{n} \sum_{k \neq i} \frac{d_k}{1 - y d_k},$$

and $\underline{\mu}_i$ the probability measure whose Cauchy-Stieljes transform $\underline{\mathfrak{m}}_i$ satisfies

$$\mathfrak{x}_{i,N}(\underline{\mathfrak{m}}_i(z)) = z, \quad \forall z \in \mathbb{C}^+. \quad (3.51)$$

Then we have $\theta_i = \mathfrak{x}_{i,N}(1)$. From (3.50) and a similar argument as in the proof of Theo-

rem 4.2 in [42] shows that for N and n sufficiently large, we have

$$E_{i,N} \cap \Omega_N = \{\sqrt{n}(\Upsilon(\eta_i) - \underline{\mathbf{m}}_i(\eta_i)) + o(1) < b_i\} \cap \Omega_N. \quad (3.52)$$

Using the resolvent formula $A^{-1} - B^{-1} = A^{-1}(B - A)B^{-1}$ for matrices or scalars, we have

$$\begin{aligned} \sqrt{n}(\Upsilon(\eta_i) - \underline{\mathbf{m}}_i(\eta_i)) &= \sqrt{n}\eta_i^{-1} \left(\frac{1}{n} Y_{i,\cdot} Y_{i,\cdot}^* - 1 \right) \\ &\quad + \sqrt{n}\eta_i^{-1} \left(\frac{1}{n} Y_{i,\cdot} (\eta_i I - \widehat{\underline{S}}_{-i,-i})^{-1} \widehat{\underline{S}}_{-i,-i} Y_{i,\cdot}^* - \int \frac{s}{\eta_i - s} d\underline{\mu}_i(s) \right). \end{aligned}$$

Therefore we set

$$R_{i,N} = \begin{cases} \sqrt{n} \left(\frac{1}{n} Y_{i,\cdot} (\eta_i I - \widehat{\underline{S}}_{-i,-i})^{-1} \widehat{\underline{S}}_{-i,-i} Y_{i,\cdot}^* - \int \frac{s}{\eta_i - s} d\underline{\mu}_i(s) \right) & \text{in } \Omega_N \\ \sqrt{n}(\tilde{\lambda}_i - \theta_i) - \mathfrak{Y}_{i,N} & \text{outside } \Omega_N. \end{cases} \quad (3.53)$$

Because $\eta_i \rightarrow 1$, we have

$$E_{i,N} = \{\mathfrak{Y}_{i,N} + R_{i,N} + o(1)\eta_i < b_i\eta_i\} = \{\mathfrak{Y}_{i,N} + R_{i,N} + o(1) < b_i\},$$

where we have multiplied the two sides of inequalities by η_i , and noticed that $o(1)\eta_i + b_i(1 - \eta_i) = o(1)$.

To complete the proof of (3.43) we only need to prove that $R_{i,N} = o_P(1)$. In [42] we proved and used a Proposition 4.3, to which the independence between $Y_{i,\cdot}$ and $\widehat{\underline{S}}_{-i,-i}$ is crucial. In our present case, the entries of $Y_{i,\cdot}$ and $\widehat{\underline{S}}_{-i,-i}$ may be correlated because we do not assume any structure feature on the eigenvectors. We have to develop some other ways to handle this difficulty, and have to add some other technical conditions to ensure the result.

Proposition III.10. *Under the conditions of Theorem III.4, we have*

$$R_{i,N} \rightarrow 0$$

in probability as $N, n \rightarrow \infty$.

Proof. As Ω_N occurs with high probability, we only need to prove that $R_{i,N} \rightarrow 0$ in probability in Ω_N , under **A3** and **A4** respectively.

We assume **A3**, i.e. assume that

$$\lim_{N \rightarrow \infty} \frac{1}{\sqrt{N}} \sum_{k=1}^N d_k = 0, \quad \text{and} \quad \mathbb{E}|Z_{i,j}|^4 < \infty,$$

and prove that

$$\sqrt{n} \left(\frac{1}{n} Y_{i,\cdot} (\eta_i I - \widehat{\underline{S}}_{-i,-i})^{-1} \widehat{\underline{S}}_{-i,-i} Y_{i,\cdot}^* - \int \frac{s}{\eta_i - s} d\mu_i(s) \right) = o_P(1). \quad (3.54)$$

Note that η_i defined in (3.48) has positive distances to the support of $\underline{\mu}_i(s)$ and to the spectrum of $\widehat{\underline{S}}_{-i,-i}$ for N sufficiently large. Therefore there exists $C > 0$ such that

$$\left| \int \frac{s}{\eta_i - s} d\mu_i(s) \right| \leq C \int s d\mu_i(s),$$

and such that the following two matrices

$$C \widehat{\underline{S}}_{-i,-i} - (\eta_i I - \widehat{\underline{S}}_{-i,-i})^{-1} \widehat{\underline{S}}_{-i,-i}, \quad C \widehat{\underline{S}}_{-i,-i} + (\eta_i I - \widehat{\underline{S}}_{-i,-i})^{-1} \widehat{\underline{S}}_{-i,-i}$$

are all positive semidefinite. The last fact shows that

$$\left| Y_{i,\cdot} (\eta_i I - \widehat{\underline{S}}_{-i,-i})^{-1} \widehat{\underline{S}}_{-i,-i} Y_{i,\cdot}^* \right| \leq C Y_{i,\cdot} \widehat{\underline{S}}_{-i,-i} Y_{i,\cdot}^* = C Y_{i,\cdot} \frac{1}{n} Y_{-i,\cdot}^* D_{-i,-i} Y_{-i,\cdot} Y_{i,\cdot}^*.$$

We will prove under **A3** that

$$\frac{1}{n^{3/2}} Y_{i,\cdot} Y_{-i,\cdot}^* D_{-i,-i} Y_{-i,\cdot} Y_{i,\cdot}^* = o_P(1), \quad \text{and} \quad \sqrt{n} \int s d\mu_i(s) = o(1).$$

Lemma III.11. *The first moment of $\underline{\mu}_i$ is*

$$\int s d\mu_i(s) = \frac{1}{n} \sum_{k \neq i} d_k. \quad (3.55)$$

Proof of Lemma III.11. Note that the complex function

$$\frac{1}{z} \underline{\mathbf{m}}_i \left(\frac{1}{z} \right) = \int \frac{d\underline{\mu}_i(s)}{1 - zs}$$

is analytic near 0. So for any z such that $|z|$ is small enough, we have

$$\frac{1}{z} \underline{\mathbf{m}}_i \left(\frac{1}{z} \right) = 1 + \sum_{j=1}^{\infty} z^j \int s^j d\underline{\mu}_i(s).$$

Then by the equation (3.51) satisfied by $\underline{\mathbf{m}}_i$, we have

$$1 + \sum_{j=1}^{\infty} z^j \int s^j d\underline{\mu}_i(s) = 1 + \frac{1}{n} \sum_{k \neq i} \frac{\underline{\mathbf{m}}_i \left(\frac{1}{z} \right) d_k}{1 - \underline{\mathbf{m}}_i \left(\frac{1}{z} \right) d_k}.$$

Noting that $\underline{\mathbf{m}}_i \left(\frac{1}{z} \right) \rightarrow 0$ as $z \rightarrow 0$, and expanding the fraction on the RHS, we have

$$\sum_{j=1}^{\infty} z^j \int s^j d\underline{\mu}_i(s) = \sum_{j=1}^{\infty} z^j \left[\left(\frac{1}{z} \underline{\mathbf{m}}_i \left(\frac{1}{z} \right) \right)^j \frac{1}{n} \sum_{k \neq i} d_k^j \right] = \sum_{j=1}^{\infty} z^j \left[\left(\int \frac{d\underline{\mu}_i(s)}{1 - zs} \right)^j \frac{1}{n} \sum_{k \neq i} d_k^j \right]. \quad (3.56)$$

Dividing the two sides by z , and letting $z \rightarrow 0$, we obtain (3.55). This ends the proof of Lemma III.11.

From Lemma III.11 it follows that under **A3**,

$$\sqrt{n} \int s d\underline{\mu}_i(s) = \frac{1}{\sqrt{n}} \sum_{k \neq i} d_k \rightarrow 0$$

as $N, n \rightarrow \infty$ and $N = O(n)$.

To prove that the random part of $R_{i,N}$ tends to 0 in probability, we compute its expectation. Note that the columns of $Y_N = U_N^* Z_N$ are i.i.d. with covariance matrix equal to identity, but the rows are not necessarily independent. We have

$$\frac{1}{n^{3/2}} \mathbb{E} Y_{i, \cdot} Y_{-i, \cdot}^* D_{-i, -i} Y_{-i, \cdot} Y_{i, \cdot}^* = \frac{1}{\sqrt{n}} \sum_{k \neq i} d_k \mathbb{E} (|Y_{i,1}|^2 |Y_{k,1}|^2). \quad (3.57)$$

By simple algebra, or by the Remark 14 below, $\mathbb{E} (|Y_{i,1}|^2 |Y_{k,1}|^2)$ is uniformly bounded. So

under **A3** we have indeed

$$\frac{1}{n^{3/2}} \mathbb{E} Y_{i,\cdot} Y_{-i,\cdot}^* D_{-i,-i} Y_{-i,\cdot} Y_{i,\cdot}^* \rightarrow 0.$$

Note that this random part is always nonnegative. Hence by Chebyshev's inequality we have

$$\frac{1}{n^{3/2}} Y_{i,\cdot} Y_{-i,\cdot}^* D_{-i,-i} Y_{-i,\cdot} Y_{i,\cdot}^* = o_P(1).$$

Remark 14. Here after we have to verify the uniform boundedness of expectations in the form

$$\mathbb{E}(|Y_{i_1,1}|^{\alpha_1} \dots |Y_{i_m,1}|^{\alpha_m}).$$

To prevent complicated calculations, we can use Hölder's inequality and obtain

$$\mathbb{E}(|Y_{i_1,1}|^{\alpha_1} \dots |Y_{i_m,1}|^{\alpha_m}) \leq (\mathbb{E}|Y_{i_1,1}|^\alpha)^{\frac{\alpha_1}{\alpha}} \dots (\mathbb{E}|Y_{i_m,1}|^\alpha)^{\frac{\alpha_m}{\alpha}},$$

where $\alpha = \alpha_1 + \dots + \alpha_m$. And then we use Lemma 2.1 in [1] to obtain the uniform boundedness of each $\mathbb{E}|Y_{i_k,1}|^\alpha$, under the condition that $\mathbb{E}|Z_{i,j}|^\alpha < \infty$.

We now assume **A4**, i.e.

$$\lim_{N \rightarrow \infty} \frac{1}{\sqrt{N}} \sum_{k=1}^N d_k^2 = 0 \quad \text{and} \quad \mathbb{E}|Z_{i,j}|^8 < \infty,$$

and prove (3.54). This time we can no longer separate the random part and the deterministic part of $R_{i,N}$, and we have to prove that the difference between these two parts is small.

Notice that, by the resolvent formula again, we have

$$\begin{aligned} & \frac{1}{\sqrt{n}} Y_{i,\cdot} (\eta_i I - \widehat{\underline{S}}_{-i,-i})^{-1} \widehat{\underline{S}}_{-i,-i} Y_{i,\cdot}^* - \sqrt{n} \int \frac{s}{\eta_i - s} d\underline{\mu}_i(s) \\ &= \frac{1}{\eta_i} \left(\frac{1}{\sqrt{n}} Y_{i,\cdot} \widehat{\underline{S}}_{-i,-i} Y_{i,\cdot}^* - \sqrt{n} \int s d\underline{\mu}_i(s) \right) \\ &+ \frac{1}{\eta_i} \left(\frac{1}{\sqrt{n}} Y_{i,\cdot} (\eta_i I - \widehat{\underline{S}}_{-i,-i})^{-1} \widehat{\underline{S}}_{-i,-i}^2 Y_{i,\cdot}^* - \sqrt{n} \int \frac{s^2}{\eta_i - s} d\underline{\mu}_i(s) \right). \end{aligned}$$

Under **A4** we can still prove that

$$\frac{1}{\sqrt{n}}Y_{i,\cdot}(\eta_i I - \widehat{\underline{S}}_{-i,-i})^{-1}\widehat{\underline{S}}_{-i,-i}^2 Y_{i,\cdot}^* = o_P(1), \quad \text{and} \quad \sqrt{n} \int \frac{s^2}{\eta_i - s} d\mu_i(s) = o(1).$$

This is done analogously as the first part of proof. It suffices to prove that

$$\frac{1}{\sqrt{n}}Y_{i,\cdot}\widehat{\underline{S}}_{-i,-i}^2 Y_{i,\cdot}^* = o_P(1), \quad \text{and} \quad \sqrt{n} \int s^2 d\mu_i(s) = o(1).$$

Lemma III.12. *The second moment of μ_i is*

$$\int s^2 d\mu_i(s) = \frac{1}{n} \sum_{k \neq i} d_k^2 + \left(\frac{1}{n} \sum_{k \neq i} d_k \right)^2. \quad (3.58)$$

Proof of Lemma III.12. By (3.56) again, we have

$$\begin{aligned} \sum_{j=1}^{\infty} z^j \int s^{j+1} d\mu_i(s) &= \int \frac{d\mu_i(s)}{1-zs} \left(\frac{1}{n} \sum_{k \neq i} d_k \right) - \int s d\mu_i(s) \\ &\quad + \sum_{j=1}^{\infty} z^j \left[\left(\int \frac{d\mu_i(s)}{1-zs} \right)^{j+1} \frac{1}{n} \sum_{k \neq i} d_k^{j+1} \right]. \end{aligned}$$

Taking into account (3.55), we have

$$\sum_{j=1}^{\infty} z^j \int s^{j+1} d\mu_i(s) = \int \frac{zs d\mu_i(s)}{1-zs} \left(\frac{1}{n} \sum_{k \neq i} d_k \right) + \sum_{j=1}^{\infty} z^j \left[\left(\int \frac{d\mu_i(s)}{1-zs} \right)^{j+1} \frac{1}{n} \sum_{k \neq i} d_k^{j+1} \right].$$

Dividing the two sides by z and letting $z \rightarrow 0$, we obtain (3.58). This ends the proof of Lemma III.12.

From Lemma III.12 and the inequality

$$\left(\frac{1}{n} \sum_{k \neq i} d_k \right)^2 \leq \left(\frac{N-1}{n} \right) \frac{1}{n} \sum_{k \neq i} d_k^2,$$

under **A4** we have

$$\sqrt{n} \int s^2 d\mu_i(s) = o(1)$$

as $N, n \rightarrow \infty$ and $N = O(n)$.

By simple algebra, we have

$$\begin{aligned} \mathbb{E} \frac{1}{\sqrt{n}} Y_{i,\cdot} \widehat{\underline{S}}_{-i,-i}^2 Y_{i,\cdot}^* &= \frac{n-1}{n^{3/2}} \sum_{k \neq i} \mathbb{E} |Y_{i,1}|^2 |Y_{k,1}|^2 d_k^2 \\ &+ \frac{1}{n^{3/2}} \sum_{k_1, k_2 \neq i} \mathbb{E} |Y_{i,1}|^2 |Y_{k_1,1}|^2 |Y_{k_2,1}|^2 d_{k_1} d_{k_2}. \end{aligned} \quad (3.59)$$

Again we verify that $\mathbb{E} |Y_{i,1}|^2 |Y_{k,1}|^2$ and $\mathbb{E} |Y_{i,1}|^2 |Y_{k_1,1}|^2 |Y_{k_2,1}|^2$ are uniformly bounded. Thus under the assumption **A4** the **RHS** of (3.59) tends to 0, and we have

$$\frac{1}{\sqrt{n}} Y_{i,\cdot} \widehat{\underline{S}}_{-i,-i}^2 Y_{i,\cdot}^* = o_P(1).$$

It remains to prove that

$$\frac{1}{\sqrt{n}} Y_{i,\cdot} \widehat{\underline{S}}_{-i,-i} Y_{i,\cdot}^* - \sqrt{n} \int s d\mu_i(s) = o_P(1).$$

First, by (3.57) and Lemma III.11, we have

$$\begin{aligned} &\mathbb{E} \frac{1}{\sqrt{n}} Y_{i,\cdot} \widehat{\underline{S}}_{-i,-i} Y_{i,\cdot}^* - \sqrt{n} \int s d\mu_i(s) \\ &= \frac{1}{\sqrt{n}} \sum_{k \neq i} d_k (\mathbb{E} |Y_{i,1}|^2 |Y_{k,1}|^2 - 1) \\ &= \frac{1}{\sqrt{n}} \sum_{k \neq i} d_k \left[\sum_{j=1}^N |u_{j,i}|^2 |u_{j,k}|^2 (\mathbb{E} |Z_{1,1}|^4 - \mathbb{E} |Z_{1,1}^2|^2 - 2) + \left| \sum_{j=1}^N u_{j,i} u_{j,k} \right|^2 \mathbb{E} |Z_{1,1}^2|^2 \right]. \end{aligned}$$

To prove that this is small, it suffices to prove that $\frac{1}{\sqrt{n}} \sum_{k \neq i} d_k \sum_{j=1}^N |u_{j,i}|^2 |u_{j,k}|^2 = o(1)$

and $\frac{1}{\sqrt{n}} \sum_{k \neq i} d_k \left| \sum_{j=1}^N u_{j,i} u_{j,k} \right|^2 = o(1)$. Changing the order of summations, we have

$$\begin{aligned} \sum_{k \neq i} d_k \sum_{j=1}^N |u_{j,i}|^2 |u_{j,k}|^2 &= \sum_{j=1}^N |u_{j,i}|^2 \sum_{k \neq i} d_k |u_{j,k}|^2 \\ &\leq d_1 \sum_{j=1}^N |u_{j,i}|^2 \sum_{k \neq i} |u_{j,k}|^2 \leq d_1, \end{aligned}$$

and

$$\begin{aligned} \sum_{k \neq i} d_k \left| \sum_{j=1}^N u_{j,i} u_{j,k} \right|^2 &\leq d_1 \sum_{k=1}^N \left| \sum_{j=1}^N u_{j,i} u_{j,k} \right|^2 \\ &= d_1 \sum_{k=1}^N \left(\sum_{j=1}^N |u_{j,i}|^2 |u_{j,k}|^2 + \sum_{1 \leq j_1 \neq j_2 \leq N} u_{j_1,i} u_{j_1,k} \overline{u_{j_2,i} u_{j_2,k}} \right) \\ &= d_1 + d_1 \sum_{1 \leq j_1 \neq j_2 \leq N} \left(u_{j_1,i} \overline{u_{j_2,i}} \sum_{k=1}^N u_{j_1,k} \overline{u_{j_2,k}} \right) = d_1. \end{aligned}$$

As $d_1 := \lambda_{\max}(\Gamma_N) / \tilde{\lambda}_i(\Gamma_N)$ is uniformly bounded according to **A1**, we have proved that

$$\mathbb{E} \frac{1}{\sqrt{n}} Y_{i,\cdot} \widehat{\underline{S}}_{-i,-i} Y_{i,\cdot}^* - \sqrt{n} \int s d\mu_{-i}(s) = O\left(\frac{1}{\sqrt{n}}\right) \rightarrow 0.$$

Finally, to complete the proof, we need to prove that $\text{Var} \frac{1}{\sqrt{n}} Y_{i,\cdot} \widehat{\underline{S}}_{-i,-i} Y_{i,\cdot}^* = o(1)$ so that

$$\frac{1}{\sqrt{n}} Y_{i,\cdot} \widehat{\underline{S}}_{-i,-i} Y_{i,\cdot}^* - \mathbb{E} \frac{1}{\sqrt{n}} Y_{i,\cdot} \widehat{\underline{S}}_{-i,-i} Y_{i,\cdot}^* = o_P(1). \quad (3.60)$$

By simple algebra, we have

$$\begin{aligned}
 \text{Var} \left(\frac{1}{\sqrt{n}} Y_{i,\cdot}, \frac{1}{n} Y_{-i,\cdot}^* \cdot D_{-i,-i} Y_{-i,\cdot}, Y_{i,\cdot}^* \right) &= \frac{1}{n^3} \left(\mathbb{E}(Y_{i,\cdot}, Y_{-i,\cdot}^* \cdot D_{-i,-i} Y_{-i,\cdot}, Y_{i,\cdot}^*)^2 - (\mathbb{E} Y_{i,\cdot}, Y_{-i,\cdot}^* \cdot D_{-i,-i} Y_{-i,\cdot}, Y_{i,\cdot}^*)^2 \right) \\
 &= \frac{1}{n^3} \left(\mathbb{E}(Y_{i,\cdot}, Y_{-i,\cdot}^* \cdot D_{-i,-i} Y_{-i,\cdot}, Y_{i,\cdot}^*)^2 - n^2 \left(\sum_{k \neq i} d_k \mathbb{E}(|Y_{i,1}|^2 |Y_{k,1}|^2) \right)^2 \right) \\
 &\leq \frac{1}{n^2} \sum_{k_1, k_2 \neq i} d_{k_1} d_{k_2} \mathbb{E}|Y_{i,1}|^4 |Y_{k_1,1}|^2 |Y_{k_2,1}|^2 \\
 &\quad + \frac{1}{n} \sum_{k_1, k_2 \neq i} d_{k_1} d_{k_2} |\mathbb{E} Y_{i,1}^2 \overline{Y_{k_1,1} Y_{k_2,1}}|^2 \\
 &\quad + \frac{1}{n} \sum_{k_1, k_2 \neq i} d_{k_1} d_{k_2} |\mathbb{E}|Y_{i,1}|^2 Y_{k_1,1} \overline{Y_{k_2,1}}|^2.
 \end{aligned}$$

As $\mathbb{E}|Y_{i,1}|^4 |Y_{k_1,1}|^2 |Y_{k_2,1}|^2$ is uniformly bounded and $\frac{1}{n^2} \sum_{k_1, k_2 \neq i} d_{k_1} d_{k_2} = \left(\frac{1}{n} \sum_{k \neq i} d_k \right)^2 \rightarrow 0$, we only need to prove that

$$\frac{1}{n} \sum_{k_1, k_2 \neq i} d_{k_1} d_{k_2} |\mathbb{E} Y_{i,1}^2 \overline{Y_{k_1,1} Y_{k_2,1}}|^2 \rightarrow 0, \quad \text{and} \quad \frac{1}{n} \sum_{k_1, k_2 \neq i} d_{k_1} d_{k_2} |\mathbb{E}|Y_{i,1}|^2 Y_{k_1,1} \overline{Y_{k_2,1}}|^2 \rightarrow 0. \tag{3.61}$$

We now prove the first convergence. By simple algebra, we have

$$\mathbb{E} Y_{i,1}^2 \overline{Y_{k_1,1} Y_{k_2,1}} = \sum_{j=1}^N \overline{u_{j,i}}^2 u_{j,k_1} u_{j,k_2} (\mathbb{E}|Z_{1,1}|^4 - |\mathbb{E} Z_{1,1}^2|^2 - 2) + \left(\sum_{j=1}^N \overline{u_{j,i}}^2 \right) \left(\sum_{j=1}^N u_{j,k_1} u_{j,k_2} \right) |\mathbb{E} Z_{1,1}^2|^2.$$

By the elementary inequality $|a + b|^2 \leq 2|a|^2 + 2|b|^2$, we only need to prove

$$\frac{1}{n} \sum_{k_1, k_2 \neq i} d_{k_1} d_{k_2} \left| \sum_{j=1}^N \overline{u_{j,i}}^2 u_{j,k_1} u_{j,k_2} \right|^2 \rightarrow 0, \quad \text{and} \quad \frac{1}{n} \sum_{k_1, k_2 \neq i} d_{k_1} d_{k_2} \left| \sum_{j=1}^N u_{j,k_1} u_{j,k_2} \right|^2 \rightarrow 0. \tag{3.62}$$

To prove the first convergence in (3.62), we have

$$\begin{aligned}
 \frac{1}{n} \sum_{k_1, k_2 \neq i} d_{k_1} d_{k_2} \left| \sum_{j=1}^N \overline{u_{j,i}}^2 u_{j,k_1} u_{j,k_2} \right|^2 &\leq \frac{d_1^2}{n} \sum_{k_1, k_2} \left| \sum_{j=1}^N \overline{u_{j,i}}^2 u_{j,k_1} u_{j,k_2} \right|^2 \\
 &= \frac{d_1^2}{n} \sum_{k_1, k_2} \sum_{j_1, j_2} \overline{u_{j_1,i}}^2 u_{j_2,i}^2 u_{j_1,k_1} \overline{u_{j_2,k_1}} u_{j_1,k_2} \overline{u_{j_2,k_2}} \\
 &= \frac{d_1^2}{n} \sum_{j_1, j_2} \overline{u_{j_1,i}}^2 u_{j_2,i}^2 \left(\sum_{k=1}^N u_{j_1,k} \overline{u_{j_2,k}} \right)^2 \\
 &= \frac{d_1^2}{n} \sum_{j=1}^N |u_{j,i}|^4 \rightarrow 0.
 \end{aligned} \tag{3.63}$$

To prove the second in (3.62), we take an arbitrary small $\varepsilon > 0$. Then by the assumption **A4**, we have

$$\#\{d_k > \varepsilon\} = \frac{\sqrt{N}o(1)}{\varepsilon^2}. \tag{3.64}$$

We write

$$d_{k,>} := \begin{cases} d_k & \text{if } d_k > \varepsilon \\ 0 & \text{otherwise.} \end{cases} \quad \text{and} \quad d_{k,\leq} := \begin{cases} d_k & \text{if } d_k \leq \varepsilon \\ 0 & \text{otherwise.} \end{cases}$$

Then we have

$$\frac{1}{n} \sum_{k_1, k_2 \neq i} d_{k_1} d_{k_2} \left| \sum_{j=1}^N u_{j,k_1} u_{j,k_2} \right|^2 = \frac{1}{n} \sum_{k_1, k_2 \neq i} (d_{k_1,>} + d_{k_1,\leq})(d_{k_2,>} + d_{k_2,\leq}) \left| \sum_{j=1}^N u_{j,k_1} u_{j,k_2} \right|^2.$$

By (3.64), we have

$$\frac{1}{n} \sum_{k_1, k_2 \neq i} d_{k_1,>} d_{k_2,>} \leq \frac{N d_1^2 o(1)}{n \varepsilon^4} \rightarrow 0.$$

We also have

$$\begin{aligned}
 \frac{1}{n} \sum_{k_1, k_2 \neq i} (2d_{k_1,>} d_{k_2,\leq} + d_{k_1,\leq} d_{k_2,\leq}) \left| \sum_{j=1}^N u_{j,k_1} u_{j,k_2} \right|^2 &\leq \frac{3d_1 \varepsilon}{n} \sum_{k_1, k_2} \left| \sum_{j=1}^N u_{j,k_1} u_{j,k_2} \right|^2 \\
 &= \frac{3d_1 \varepsilon}{n} \sum_{j_1, j_2} \left(\sum_{k=1}^N u_{j_1,k} \overline{u_{j_2,k}} \right)^2 \\
 &= 3d_1 \varepsilon.
 \end{aligned}$$

Thus we have proved (3.62), implying the first part of (3.61).

We then prove the second part of (3.61). We have

$$\begin{aligned} & \mathbb{E}|Y_{i,1}|^2 Y_{k_1,1} \overline{Y_{k_2,1}} \\ &= \sum_{j=1}^N |u_{j,i}|^2 \overline{u_{j,k_1}} u_{j,k_2} (\mathbb{E}|Z_{1,1}|^4 - |\mathbb{E}Z_{1,1}^2|^2 - 2) + \delta_{k_1,k_2} \\ & \quad + \left(\sum_{j=1}^N \overline{u_{j,i}} u_{j,k_1} \right) \left(\sum_{j=1}^N u_{j,i} u_{j,k_2} \right) |\mathbb{E}Z_{1,1}^2|^2, \end{aligned}$$

where δ_{k_1,k_2} is the Kronecker symbol. Using the inequality $|a+b+c|^2 \leq 3(|a|^2 + |b|^2 + |c|^2)$ we only need to prove

$$\frac{1}{n} \sum_{k_1, k_2 \neq i} d_{k_1} d_{k_2} \left| \sum_{j=1}^N |u_{j,i}|^2 \overline{u_{j,k_1}} u_{j,k_2} \right|^2 \rightarrow 0, \quad \frac{1}{n} \sum_{k \neq i} d_k^2 \rightarrow 0, \quad (3.65)$$

and

$$\frac{1}{n} \sum_{k_1, k_2 \neq i} d_{k_1} d_{k_2} \left| \left(\sum_{j=1}^N \overline{u_{j,i}} u_{j,k_1} \right) \left(\sum_{j=1}^N u_{j,i} u_{j,k_2} \right) \right|^2 \rightarrow 0. \quad (3.66)$$

The first of (3.65) can be proved similarly as (3.63), the second of (3.65) is a consequence of **A4**. To prove (3.66), we have

$$\begin{aligned} & \frac{1}{n} \sum_{k_1, k_2 \neq i} d_{k_1} d_{k_2} \left| \left(\sum_{j=1}^N \overline{u_{j,i}} u_{j,k_1} \right) \left(\sum_{j=1}^N u_{j,i} u_{j,k_2} \right) \right|^2 \\ & \leq \frac{d_1^2}{n} \sum_{k_1, k_2} \sum_{j_1, j_2, j_3, j_4} \overline{u_{j_1, i} u_{j_1, k_1}} u_{j_2, i} u_{j_2, k_1} u_{j_3, i} u_{j_3, k_2} \overline{u_{j_4, i} u_{j_4, k_2}} \\ & = \frac{d_1^2}{n} \sum_{j_1, j_2, j_3, j_4} \overline{u_{j_1, i} u_{j_2, i} u_{j_3, i} u_{j_4, i}} \left(\sum_{k=1}^N \overline{u_{j_1, k} u_{j_2, k}} \right) \left(\sum_{k=1}^N u_{j_3, k} u_{j_4, k} \right) \\ & = \frac{d_1^2}{n} \left(\sum_{j=1}^N |u_{j,i}|^2 \right)^2 = \frac{d_1^2}{n} \rightarrow 0. \end{aligned}$$

The second part of (3.61) is proved. So we have

$$\text{Var} \left(\frac{1}{\sqrt{n}} Y_i, \frac{1}{n} Y_{-i}^*, DY_{-i}, Y_i^* \right) = o(1).$$

Thus we proved (3.60) and the proof of the proposition is complete. \square

3.5.3 Proof of Lemma III.9

We use Cramr-Wold device to prove the CLT of the m -dimensional vector \mathfrak{Y}_N . By a direct calculation, the covariance matrix of \mathfrak{Y}_N is exactly $\Sigma_m^{(N)}$ which tends to Σ_m as we have assumed. Then we fix a deterministic vector $a \in \mathbb{R}^m$, and prove that

$$\langle a, \mathfrak{Y}_N \rangle \xrightarrow{\mathcal{D}} \mathcal{N}(0, a^\top \Sigma_m a). \quad (3.67)$$

If $a^\top \Sigma_m a = 0$, it means that $\text{Var}\langle a, \mathfrak{Y}_N \rangle = a^\top \Sigma_m^{(N)} a \rightarrow 0$. Then $\langle a, \mathfrak{Y}_N \rangle = o_P(1)$ and hence (3.67) holds. Now we assume that $a^\top \Sigma_m a \neq 0$ and prove that

$$\frac{\langle a, \mathfrak{Y}_N \rangle}{\sqrt{a^\top \Sigma_m^{(N)} a}} \xrightarrow[N, n \rightarrow \infty]{\mathcal{D}} \mathcal{N}(0, 1). \quad (3.68)$$

Because

$$\langle a, \mathfrak{Y}_N \rangle = \sum_{i=1}^m a_i \mathfrak{Y}_{i,N} = \sum_{j=1}^n \frac{\sum_{i=1}^m a_i (|Y_{i,j}|^2 - 1)}{\sqrt{n}}$$

with

$$Y_{i,j} = \sum_{k=1}^N \overline{u_{k,i}} Z_{k,j},$$

for each N , $\langle a, \mathfrak{Y}_N \rangle$ is a sum of n i.i.d. random variables. We can use Lindeberg's CLT to prove (3.68). To do so, we need to verify the Lindeberg condition

$$\mathbb{E} \left| \sum_{i=1}^m a_i (|Y_{i,1}|^2 - 1) \right|^2 \mathbb{1}_{\left| \sum_{i=1}^m a_i (|Y_{i,1}|^2 - 1) \right| > \varepsilon \sqrt{n}} \rightarrow 0$$

as $N, n \rightarrow \infty$ for any $\varepsilon > 0$.

Since $\text{Var} Y_{i,1} = 1$, for any $\varepsilon > 0$, the events

$$A_N := \left\{ \left| \sum_{i=1}^m a_i (|Y_{i,1}|^2 - 1) \right| > \varepsilon \sqrt{n} \right\}$$

occur with low probability. By Minkowski's inequality, we have

$$\left(\mathbb{E} \left| \sum_{i=1}^m a_i (|Y_{i,1}|^2 - 1) \right|^2 \mathbb{1}_{A_N} \right)^{\frac{1}{2}} \leq \sum_{i=1}^m |a_i| \left(\mathbb{E} \left| |Y_{i,1}|^2 - 1 \right|^2 \mathbb{1}_{A_N} \right)^{\frac{1}{2}}.$$

So we only need to prove that

$$\mathbb{E} \left| |Y_{i,1}|^2 - 1 \right|^2 \mathbb{1}_{A_N} = \mathbb{E} |Y_{i,1}|^4 \mathbb{1}_{A_N} - 2\mathbb{E} |Y_{i,1}|^2 \mathbb{1}_{A_N} + \mathbb{P}(A_N) \xrightarrow{N, n \rightarrow \infty} 0. \quad (3.69)$$

Since $\mathbb{P}(A_N) \rightarrow 0$ and from the uniform boundedness of $\mathbb{E} |Y_{i,1}|^4$ we have $\mathbb{E} |Y_{i,1}|^2 \mathbb{1}_{A_N} \rightarrow 0$, then (3.69) is equivalent to

$$\mathbb{E} |Y_{i,1}|^4 \mathbb{1}_{A_N} \xrightarrow{N, n \rightarrow \infty} 0.$$

This is a corollary of the following lemma.

Lemma III.13. *Let $u = (u_1, \dots, u_N) \in \mathbb{C}^N$ such that $\|u\|_2 \leq 1$. Let $(Z_k)_{k \geq 1}$ be a sequence of centered, reduced, i.i.d. random variables, satisfying $\mathbb{E} |Z_k|^4 < \infty$. Let A_N be a sequence of events such that $\mathbb{P}(A_N) \rightarrow 0$, then*

$$\lim_{N \rightarrow \infty} \mathbb{E} \left| \sum_{k=1}^N u_k Z_k \right|^4 \mathbb{1}_{A_N} = 0.$$

Proof. We only prove the case where u and Z_k are real. For the complex case, it can be easily proved from the real case by separating real and complex parts, and then using Minkowski's inequality.

As all the random variables $|Z_k|^4$ are identically distributed and integrable, they are uniformly integrable. Thus we have

$$\max_k \mathbb{E} |Z_k|^4 \mathbb{1}_{A_N} \rightarrow 0.$$

Let (e_N) be a sequence of positive numbers tending to 0 such that

$$\frac{\max_k (\mathbb{E} |Z_k|^4 \mathbb{1}_{A_N})^{\frac{1}{4}}}{e_N^2} \rightarrow 0 \quad \text{and} \quad N e_N^2 \rightarrow \infty \quad \text{as} \quad N \rightarrow \infty.$$

Note that from $\|u\|_2 \leq 1$ we have

$$\#\{k : |u_k| > e_N\} \leq \frac{1}{e_N^2}.$$

We write

$$\sum_{k=1}^N u_k Z_k = \sum_{k=1}^N u_k \mathbb{1}_{|u_k| > e_N} Z_k + \sum_{k=1}^N u_k \mathbb{1}_{|u_k| \leq e_N} Z_k =: P_1 + P_2.$$

By Minkowski's inequality, we have

$$(\mathbb{E}|P_1 + P_2|^4 \mathbb{1}_{A_N})^{\frac{1}{4}} \leq (\mathbb{E}|P_1|^4 \mathbb{1}_{A_N})^{\frac{1}{4}} + (\mathbb{E}|P_2|^4 \mathbb{1}_{A_N})^{\frac{1}{4}}.$$

For the first part, using again Minkowski's inequality and noting that $|u_k| \leq 1$, we get

$$(\mathbb{E}|P_1|^4 \mathbb{1}_{A_N})^{\frac{1}{4}} \leq \#\{i : |u_k| > e_N\} \max_k (\mathbb{E}|Z_k|^4 \mathbb{1}_{A_N})^{\frac{1}{4}} \rightarrow 0.$$

For the second part, it suffices to prove that from any subsequence of $(\mathbb{E}|P_2|^4 \mathbb{1}_{A_N})$ we can extract a subsequence tending to 0. From any subsequence of $(\sum_{k=1}^N |u_k|^2 \mathbb{1}_{|u_k| < e_N})$, there exists a convergent subsequence. So we can assume that

$$\sum_{k=1}^N |u_k|^2 \mathbb{1}_{|u_k| < e_N} \rightarrow \sigma^2 \leq 1.$$

Then if we can prove that

$$\lim_{M \rightarrow \infty} \overline{\lim}_{N \rightarrow \infty} \mathbb{E}|P_2|^4 \mathbb{1}_{|P_2| > M} = 0, \quad (3.70)$$

the proof of the lemma will be complete due to the inequality

$$\mathbb{E}|P_2|^4 \mathbb{1}_{A_N} \leq \mathbb{E}|P_2|^4 \mathbb{1}_{|P_2| > M} + \mathbb{E}|P_2|^4 \mathbb{1}_{A_N \cap \{|P_2| \leq M\}} \leq \mathbb{E}|P_2|^4 \mathbb{1}_{|P_2| > M} + M^4 \mathbb{P}(A_N).$$

To prove (3.70), by the equality $x^4 = x^4 \wedge M^4 + x^4 \mathbb{1}_{|x| > M} - M^4 \mathbb{1}_{|x| > M}$ for any $x \in \mathbb{R}$

and $M > 0$, we have

$$\mathbb{E}|P_2|^4 \mathbb{1}_{|P_2| > M} = \mathbb{E}|P_2|^4 - \mathbb{E}|P_2|^4 \wedge M^4 + M^4 \mathbb{P}(|P_2| > M).$$

Let $N \rightarrow \infty$, since $\max_k \{|u_k| \mathbb{1}_{|u_k| \leq e_N}\} \leq e_N \rightarrow 0$, by Lindeberg's Theorem we have $P_2 \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma^2)$. Let $G \sim \mathcal{N}(0, \sigma^2)$, then we have

$$\mathbb{E}|P_2|^4 \rightarrow 3\sigma^4 = \mathbb{E}|G|^4, \quad \mathbb{E}|P_2|^4 \wedge M^4 \rightarrow \mathbb{E}|G|^4 \wedge M^4, \quad \text{and} \quad \mathbb{P}(|P_2| > M) \rightarrow \mathbb{P}(|G| > M),$$

where the first convergence is from direct calculation, the second and the third are from the fact that $P_2 \xrightarrow{\mathcal{D}} G$ and that the function $x \mapsto x^4 \wedge M^4$ is continuous and bounded. So we have

$$\lim_{N \rightarrow \infty} \mathbb{E}|P_2|^4 \mathbb{1}_{|P_2| > M} = \mathbb{E}|G|^4 \mathbb{1}_{|G| > M}.$$

Finally we take $M \rightarrow \infty$ and see that (3.70) holds. □

The proof of Theorem III.4 is finished.

CHAPTER IV

Some extensions in case of large covariance matrices close to short memory

4.1 Introduction

Notice that if the parameter $d < 0$, or equivalently $\rho < -1$, then the process $T_N^{1/2}Z_N$ is a **SMSP** and **TW** fluctuations for the largest eigenvalue are expected. The case $\rho = -1$ represents the threshold between **LMSP** and **SMSP**. Hence the extra-difficulty appears when ρ is close to -1 .

In Chapters **II** and **III** we have established the fluctuations for the m largest eigenvalues of $S_N(T)$ with parameter $d \in (1/8, 1/2)$, or equivalently $\rho \in (-3/4, 0)$. Note also that if the entries of Z_N are i.i.d. standard Gaussian, then Theorem **I.9** also gives the result for the case where $d \in (0, 1/2)$ or $\rho \in (-1, 0)$.

In this chapter we try to study the non-Gaussian case with parameter close to the threshold of short memory, that is, $d \in (0, 1/8)$ or $\rho \in (-1, -3/4)$. In this case, we prove Proposition **I.6** for T_N . This proposition can be viewed as a generalization of Proposition **I.5**.

Proposition IV.1 (Proposition **I.6**). *Let T_N be defined as in Section **1.4**, for any $\rho \in (-1, 0)$ or equivalently $d \in (0, 1/2)$. Then there exists an integer $q \geq 1$ such that*

$$\lim_{N \rightarrow \infty} \frac{\text{tr } T_N^q}{\sqrt{N} \lambda_{\max}^q(T_N)} = 0. \quad (4.1)$$

The proof of this proposition is postponed to the appendix, see Appendix **A.1**.

Now the following fluctuations' result holds:

Theorem IV.2 (Theorem I.15). *Under A1, A3, A5, and in addition A9-A11, then the fluctuations*

$$\sqrt{n} \begin{pmatrix} \frac{\lambda_1(S_N(\Gamma))}{\lambda_1(\Gamma_N)} - 1 - \frac{1}{n} \sum_{k \neq 1} \frac{\lambda_k(\Gamma_N)}{\lambda_1(\Gamma_N) - \lambda_k(\Gamma_N)} \\ \vdots \\ \frac{\lambda_m(S_N(\Gamma))}{\lambda_m(\Gamma_N)} - 1 - \frac{1}{n} \sum_{k \neq m} \frac{\lambda_k(\Gamma_N)}{\lambda_m(\Gamma_N) - \lambda_k(\Gamma_N)} \end{pmatrix} \xrightarrow[N, n \rightarrow \infty]{\mathcal{D}} \mathcal{N}(0, (1 + |\mathbb{E}Z_{1,1}^{(N)}|^2)I_m)$$

hold.

With Proposition I.6 and other results on T_N in hand, we can verify that all conditions on Γ_N in Theorem I.15 are satisfied by T_N except A11. From simulation we believe that A11 also holds for T_N , so we raise the conjecture I.16 and hope that it will be solved afterwards.

Conjecture IV.3 (Conjecture I.16). *The result of Theorem I.11 holds for all $S_N(T)$ with parameter $d \in (0, 1/2)$ or equivalently $\rho \in (-1, 0)$.*

4.2 Proof of Theorem I.15

The proof of this theorem is very technical and in addition to some arguments developed in the proof of Theorem III.4, it uses a combination of the Lindeberg's method together with a fine analysis of some moments with the help of techniques coming from graph theory and combinatorics.

We take the notations of that theorem and prove that $R_{i,N} = o_P(1)$, where $R_{i,N}$ is defined in (3.53). More precisely we prove

$$\sqrt{n} \left(\frac{1}{n} Y_{i,\cdot} (\eta_i I - \widehat{\mathcal{S}}_{-i,-i})^{-1} \widehat{\mathcal{S}}_{-i,-i} Y_{i,\cdot}^* - \int \frac{s}{\eta_i - s} d\mu_i(s) \right) = o_P(1) \quad (4.2)$$

under the restriction on Ω_N .

Because of the lack of symbols for subscription we prove the above for $i = 1$ so that we can release i for the use of a general subscription symbol. The general case can be proved without any changes.

Using the resolvent formula, and the same argument as in the proof of Theorem III.4,

this leads to prove that

$$\sqrt{n} \left(\frac{1}{n} Y_{1,\cdot} \widehat{\underline{S}}_{-1,-1}^k Y_{1,\cdot}^* - \int s^k d\underline{\mu}_1(s) \right) = o_P(1) \quad (4.3)$$

for $1 \leq k \leq q$, and to prove that

$$\sqrt{n} \int s^q d\underline{\mu}_1(s) = o(1). \quad (4.4)$$

Note that in the case of Theorem III.4 we have $q = 2$. Note also that the combination of (4.3) and (4.4) implies

$$\frac{1}{\sqrt{n}} Y_{1,\cdot} \widehat{\underline{S}}_{-1,-1}^q Y_{1,\cdot}^* = o_P(1).$$

We use the Gaussian interpolation trick to achieve the proof of (4.3). Precisely, let $G = G_N$ be a $N \times n$ matrix whose entries $G_{i,j}$ are standard Gaussian random variables, real or complex, and suppose that the entries of Z and G match to order 3. We rewrite the left handside (LHS) of (4.3) as

$$\frac{1}{\sqrt{n}} Y_{1,\cdot} \left(\frac{1}{n} Y_{2:N,\cdot}^* D_{2:N,2:N} Y_{2:N,\cdot} \right)^k Y_{1,\cdot}^* - \sqrt{n} \int s^k d\underline{\mu}_1(s) = P_1 + P_2 + P_3 + P_4,$$

where

$$\begin{aligned} P_1 &:= \frac{1}{\sqrt{n}} Y_{1,\cdot} \left(\frac{1}{n} Y_{2:N,\cdot}^* D_{2:N,2:N} Y_{2:N,\cdot} \right)^k Y_{1,\cdot}^* - \mathbb{E} \frac{1}{\sqrt{n}} Y_{1,\cdot} \left(\frac{1}{n} Y_{2:N,\cdot}^* D_{2:N,2:N} Y_{2:N,\cdot} \right)^k Y_{1,\cdot}^*; \\ P_2 &:= \mathbb{E} \frac{1}{\sqrt{n}} Y_{1,\cdot} \left(\frac{1}{n} Y_{2:N,\cdot}^* D_{2:N,2:N} Y_{2:N,\cdot} \right)^k Y_{1,\cdot}^* - \mathbb{E} \frac{1}{\sqrt{n}} G_{1,\cdot} \left(\frac{1}{n} G_{2:N,\cdot}^* D_{2:N,2:N} G_{2:N,\cdot} \right)^k G_{1,\cdot}^*; \\ P_3 &:= \mathbb{E} \frac{1}{\sqrt{n}} \operatorname{tr} \left(\frac{1}{n} G_{2:N,\cdot}^* D_{2:N,2:N} G_{2:N,\cdot} \right)^k - \frac{1}{\sqrt{n}} \operatorname{tr} \left(\frac{1}{n} G_{2:N,\cdot}^* D_{2:N,2:N} G_{2:N,\cdot} \right)^k; \\ P_4 &:= \frac{1}{\sqrt{n}} \operatorname{tr} \left(\frac{1}{n} G_{2:N,\cdot}^* D_{2:N,2:N} G_{2:N,\cdot} \right)^k - \sqrt{n} \int s^k d\underline{\mu}_1(s). \end{aligned}$$

Note that by independence we have

$$\mathbb{E} \frac{1}{\sqrt{n}} G_{1,\cdot} \left(\frac{1}{n} G_{2:N,\cdot}^* D_{2:N,2:N} G_{2:N,\cdot} \right)^k G_{1,\cdot}^* = \mathbb{E} \frac{1}{\sqrt{n}} \operatorname{tr} \left(\frac{1}{n} G_{2:N,\cdot}^* D_{2:N,2:N} G_{2:N,\cdot} \right)^k.$$

To prove $P_1 = o_P(1)$ we prove

$$\text{Var} \frac{1}{\sqrt{n}} Y_{1,\cdot} \left(\frac{1}{n} Y_{2:N,\cdot}^*, D_{2:N,2:N} Y_{2:N,\cdot} \right)^k Y_{1,\cdot}^* - \text{Var} \frac{1}{\sqrt{n}} G_{1,\cdot} \left(\frac{1}{n} G_{2:N,\cdot}^*, D_{2:N,2:N} G_{2:N,\cdot} \right)^k G_{1,\cdot}^* = o(1). \quad (4.5)$$

and

$$\text{Var} \frac{1}{\sqrt{n}} G_{1,\cdot} \left(\frac{1}{n} G_{2:N,\cdot}^*, D_{2:N,2:N} G_{2:N,\cdot} \right)^k G_{1,\cdot}^* = o(1). \quad (4.6)$$

Then using Lindeberg replacement trick we prove that

$$P_2 = \mathbb{E} \frac{1}{\sqrt{n}} Y_{1,\cdot} \left(\frac{1}{n} Y_{2:N,\cdot}^*, D_{2:N,2:N} Y_{2:N,\cdot} \right)^k Y_{1,\cdot}^* - \mathbb{E} \frac{1}{\sqrt{n}} G_{1,\cdot} \left(\frac{1}{n} G_{2:N,\cdot}^*, D_{2:N,2:N} G_{2:N,\cdot} \right)^k G_{1,\cdot}^* = o(1). \quad (4.7)$$

The fact that $P_3 = o_P(1)$ is ensured by proving

$$\text{Var} \frac{1}{\sqrt{n}} \text{tr} \left(\frac{1}{n} G_{2:N,\cdot}^*, D_{2:N,2:N} G_{2:N,\cdot} \right)^k = o(1). \quad (4.8)$$

And by [3, Theorem 1.1(i)] we have already $P_4 = o_P(1)$.

The particular convergence (4.4) will be proved by verifying

$$\mathbb{E} \frac{1}{\sqrt{n}} \text{tr} \left(\frac{1}{n} G_{2:N,\cdot}^*, D_{2:N,2:N} G_{2:N,\cdot} \right)^q = o(1) \quad (4.9)$$

and by considering $P_3 = o_P(1)$ and $P_4 = o_P(1)$. So we need to prove (4.5)-(4.9).

We have to expand the expectations and variations in these formulas as in the proof of Theorem III.4. Since k is arbitrary, these expansions are arbitrarily complicated. Hence we will need some tools from the graph theory.

First we need the following lemma, which is the core of the Lindeberg comparison.

Lemma IV.4. *For each $N \in \mathbb{N}$ let $\mathcal{Z} = (\mathcal{Z}_1, \dots, \mathcal{Z}_N)^\top$ and $\mathcal{G} = (\mathcal{G}_1, \dots, \mathcal{G}_N)^\top$ be N -dimensional random vectors whose entries \mathcal{Z}_i and \mathcal{G}_i are respectively i.i.d random variables and i.i.d standard Gaussian variables. Suppose that \mathcal{Z} and \mathcal{G} are independent between them. Suppose also that the entries \mathcal{Z}_i are i.d across i, N and match with \mathcal{G}_i to order 3, and all the moments of \mathcal{Z}_i are finite. Let U be a $N \times N$ orthonormal matrix, or unitary if \mathcal{G}_i are complex gaussian. Suppose that there exists a constant K s.t. $|U_{i,j}| \leq K/\sqrt{N}$ for any N*

and $1 \leq i, j \leq N$. Let $\mathcal{Y} = U\mathcal{Z}$.

Then for any integer $k \geq 1$, there is a constant $C_k > 0$ such that for any function

$$f(x_1, x_2, \dots, x_s) := x_1^{\alpha_1} \overline{x_1}^{\beta_1} \dots x_s^{\alpha_s} \overline{x_s}^{\beta_s}$$

with α_l, β_l nonnegative integers s.t. $\alpha_1 + \beta_1 + \dots + \alpha_s + \beta_s \leq k$, we have

$$|\mathbb{E}f(\mathcal{Y}_{i_1}, \dots, \mathcal{Y}_{i_s}) - \mathbb{E}f(\mathcal{G}_{i_1}, \dots, \mathcal{G}_{i_s})| \leq \frac{C_k}{N},$$

with any $1 \leq i_1, \dots, i_s \leq N$.

Proof. For any k , as the number of monomials of degree no greater than k with coefficient 1 is finite, we just need to verify the result for an arbitrary monomial of degree k . Without loss of generality we suppose that $f(x_1, \dots, x_k) = x_1 \dots x_k$, for the other cases the principal is the same. We will prove that there exists a constant $C_k > 0$ s.t.

$$|\mathbb{E}\mathcal{Y}_{i_1} \dots \mathcal{Y}_{i_k} - \mathbb{E}\mathcal{G}_{i_1} \dots \mathcal{G}_{i_k}| \leq \frac{C_k}{N}.$$

For any $1 \leq i \leq N$ and any $1 \leq j \leq N$, we set

$$\mathcal{Y}_i^{(j)} := U_{i,1}\mathcal{Z}_1 + \dots + U_{i,j}\mathcal{Z}_j + U_{i,j+1}\mathcal{G}_{j+1} + \dots + U_{i,N}\mathcal{G}_N$$

and

$$\mathcal{S}_i^{(j)} := U_{i,1}\mathcal{Z}_1 + \dots + U_{i,j-1}\mathcal{Z}_{j-1} + U_{i,j+1}\mathcal{G}_{j+1} + \dots + U_{i,N}\mathcal{G}_N = \mathcal{Y}_i^{(j)} - U_{i,j}\mathcal{Z}_j.$$

Then we have $\mathcal{Y}_i^{(0)} = (U\mathcal{G})_i$ and $\mathcal{Y}_i^{(N)} = \mathcal{Y}_i$. As $U\mathcal{G}$ has the same distribution as \mathcal{G} , we have

$$|\mathbb{E}\mathcal{Y}_{i_1} \dots \mathcal{Y}_{i_k} - \mathbb{E}\mathcal{G}_{i_1} \dots \mathcal{G}_{i_k}| \leq \sum_{j=1}^N |\mathbb{E}\mathcal{Y}_{i_1}^{(j)} \dots \mathcal{Y}_{i_k}^{(j)} - \mathbb{E}\mathcal{Y}_{i_1}^{(j-1)} \dots \mathcal{Y}_{i_k}^{(j-1)}|. \quad (4.10)$$

To control each term of the sum, we write $\mathcal{Y}_{i_l}^{(j)} = \mathcal{S}_{i_l}^{(j)} + U_{i_l,j}\mathcal{Z}_j$ and $\mathcal{Y}_{i_l}^{(j-1)} = \mathcal{S}_{i_l}^{(j)} + U_{i_l,j}\mathcal{G}_j$,

then we have

$$\mathbb{E}\mathcal{Y}_{i_1}^{(j)} \dots \mathcal{Y}_{i_k}^{(j)} = \mathbb{E}(\mathcal{S}_{i_1}^{(j)} + U_{i_1,j}\mathcal{Z}_j) \dots (\mathcal{S}_{i_k}^{(j)} + U_{i_k,j}\mathcal{Z}_j) = \mathbb{E}P_j(\mathcal{Z}_j)$$

and

$$\mathbb{E}\mathcal{Y}_{i_1}^{(j-1)} \dots \mathcal{Y}_{i_k}^{(j-1)} = \mathbb{E}(\mathcal{S}_{i_1}^{(j)} + U_{i_1,j}\mathcal{G}_j) \dots (\mathcal{S}_{i_k}^{(j)} + U_{i_k,j}\mathcal{G}_j) = \mathbb{E}P_j(\mathcal{G}_j)$$

where P_j is the polynomial $P_j(x) := (\mathcal{S}_{i_1}^{(j)} + U_{i_1,j}x) \dots (\mathcal{S}_{i_k}^{(j)} + U_{i_k,j}x)$ with random coefficients. We expand the polynomial P_j and get

$$P_j(\mathcal{Z}_j) = A_0 + A_1\mathcal{Z}_j + A_2\mathcal{Z}_j^2 + A_3\mathcal{Z}_j^3 + A_4\mathcal{Z}_j^4 + \dots + A_k\mathcal{Z}_j^k.$$

Note that the coefficient A_l is a sum of $\binom{k}{l}$ terms, each of which is typically a product of l U -entries and $k-l$ random variables $\mathcal{S}_i^{(j)}$'s, i.e.

$$A_l = \sum_{1 \leq k_1 \leq \dots \leq k_l \leq k} U_{i_{k_1},j} \dots U_{i_{k_l},j} \prod_{s \neq k_1, \dots, k_l} \mathcal{S}_{i_s}^{(j)}.$$

Note also that the random variables $\mathcal{S}_i^{(j)}$ are independent of \mathcal{Z}_j and \mathcal{G}_j . Thus because \mathcal{Z}_j and \mathcal{G}_j match to order 3 we have

$$\mathbb{E}A_l\mathcal{Z}_j^l = \mathbb{E}A_l\mathbb{E}\mathcal{Z}_j^l = \mathbb{E}A_l\mathbb{E}\mathcal{G}_j^l = \mathbb{E}A_l\mathcal{Z}_j^l, \quad \text{for } l = 0, 1, 2, 3,$$

and for $l = 4, \dots, k$, we have

$$|\mathbb{E}A_l\mathcal{Z}_j^l - \mathbb{E}A_l\mathcal{G}_j^l| = |\mathbb{E}A_l| |\mathbb{E}\mathcal{Z}_j^l - \mathbb{E}\mathcal{G}_j^l| \leq \frac{KC_{k,l}}{N^{l/2}} |\mathbb{E}\mathcal{Z}_j^l - \mathbb{E}\mathcal{G}_j^l|, \quad \text{for } l \geq 4, \quad (4.11)$$

where $C_{k,l}$ depends only on k and l . This is because of the uniform decay $|U_{j,l}| \leq K/N^{1/2}$, and the uniform boundedness of

$$\mathbb{E} \prod_{s \in \{1, \dots, k\} \setminus \{k_1, \dots, k_l\}} \mathcal{S}_{i_s}^{(j)} \quad (4.12)$$

by a constant depending only on k and l . The latter fact is obtained from Hölder's inequality

and Lemma 2.1 in [1]. From (4.11) and (4.12) we have

$$|\mathbb{E}P_j(\mathcal{Z}_j) - \mathbb{E}P_j(\mathcal{G}_j)| \leq \frac{K \sum_{l=4}^k C_{k,l} |\mathbb{E}\mathcal{Z}_j^l - \mathbb{E}\mathcal{G}_j^l|}{N^2}.$$

Finally we have

$$\begin{aligned} |\mathbb{E}\mathcal{Y}_{i_1} \dots \mathcal{Y}_{i_k} - \mathbb{E}\mathcal{G}_{i_1, \dots, i_k}| &\leq \frac{K \sum_{j=1}^N \sum_{l=4}^k C_{k,l} |\mathbb{E}\mathcal{Z}_j^l - \mathbb{E}\mathcal{G}_j^l|}{N^2} \\ &= \frac{K \sum_{l=4}^k C_{k,l} |\mathbb{E}\mathcal{Z}_j^l - \mathbb{E}\mathcal{G}_j^l|}{N} = \frac{C_k}{N} \end{aligned}$$

where we note $C_k = K \sum_{l=4}^k C_{k,l} |\mathbb{E}\mathcal{Z}_j^l - \mathbb{E}\mathcal{G}_j^l|$. \square

We will first prove the gaussian estimates (4.9), (4.8) and (4.6), and then we prove the other two nongaussian estimates (4.7) and (4.5).

We begin with proving (4.9). The method used here is very similar with the classical argument in the proof of Wigner's Theorem (c.f. [29, Theorem 1.3] or [17, Lemma 2.5]).

For any integer $q \geq 1$, we expand the trace and we have

$$\frac{1}{\sqrt{n}} \mathbb{E} \operatorname{tr} \left(\frac{1}{n} G_{2:N}^* D_{2:N, 2:N} G_{2:N} \right)^q = \frac{1}{n^{q+1/2}} \sum_{\substack{i_1 \dots i_q \in [[2, N]] \\ j_1 \dots j_q \in [[1, n]]}} d_{i_1} \dots d_{i_q} P(\mathbf{j}), \quad (4.13)$$

where we recall that d_i is the normalized eigenvalue $d_i = \lambda_i(\Gamma_N)/\lambda_1(\Gamma_N)$, and we denote $\mathbf{j} := (j_1, i_1, j_2, i_2, \dots, j_q, i_q)$ and $P(\mathbf{j}) := \mathbb{E} \overline{G_{i_1, j_1}} G_{i_1, j_2} \overline{G_{i_2, j_2}} G_{i_2, j_3} \dots \overline{G_{i_q, j_q}} G_{i_q, j_1}$.

Note that \mathbf{j} is a sequence of $2q$ integers, the odd positions are integers in $[[1, n]]$, and the even positions are integers in $[[2, N]]$. We will call such a sequence a *ji-word*. We say that two *ji-words* \mathbf{j} and \mathbf{j}' are *equivalent* iff for any integer $l_1, l_2 \in [[1, q]]$, $i_{l_1} = i_{l_2} \Leftrightarrow i'_{l_1} = i'_{l_2}$ and $j_{l_1} = j_{l_2} \Leftrightarrow j'_{l_1} = j'_{l_2}$. For example if $q = 2$, the *ji-words* $\mathbf{j} = (1, 2, 1, 3)$ and $\mathbf{j}' = (1, 3, 1, 2)$ are equivalent, but they are not equivalent to $\mathbf{j}'' = (1, 2, 1, 2)$. For this example we have

$$P(\mathbf{j}) = \mathbb{E} \overline{G_{2,1}} G_{2,1} \overline{G_{3,1}} G_{3,1} = 1,$$

$$P(\mathbf{j}') = \mathbb{E} \overline{G_{3,1}} G_{3,1} \overline{G_{2,1}} G_{2,1} = 1,$$

and

$$P(\mathbf{j}\mathbf{i}'') = \mathbb{E}\overline{G_{2,1}}G_{2,1}\overline{G_{2,1}}G_{2,1} = \mathbb{E}|G_{1,2}|^4.$$

We can see that if $\mathbf{j}\mathbf{i}$ is equivalent to $\mathbf{j}\mathbf{i}'$, then $P(\mathbf{j}\mathbf{i}) = P(\mathbf{j}\mathbf{i}')$.

With each ji -word $\mathbf{j}\mathbf{i}$ we associate an eulerien circuit $\mathbf{Eu}(\mathbf{j}\mathbf{i}) = (V(\mathbf{j}\mathbf{i}), C(\mathbf{j}\mathbf{i}))$ whose set of vertices is

$$V(\mathbf{j}\mathbf{i}) := \{(j_1, \mathbf{b}), (i_1, \mathbf{w}), (j_2, \mathbf{b}), (i_2, \mathbf{w}), \dots, (j_q, \mathbf{b}), (i_q, \mathbf{w})\},$$

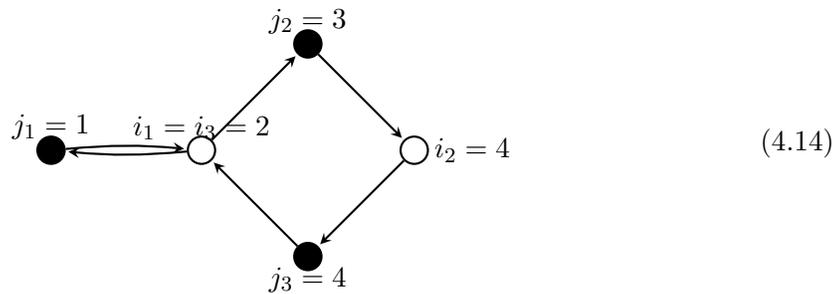
and whose steps of the circuit $C(\mathbf{j}\mathbf{i})$ is defined by

$$C(\mathbf{j}\mathbf{i}) := ((\mathbf{j}_l, \mathbf{i}_l), (\mathbf{i}_l, \mathbf{j}_{l+1}))_{l=1}^q.$$

Here we color all the column labels j_l in black (\mathbf{b}) and all the row labels i_l in white (\mathbf{w}).

We denote the colored vertices as $\mathbf{j}_l := (j_l, \mathbf{b})$ and $\mathbf{i}_l := (i_l, \mathbf{w})$. We use the convention $\mathbf{j}_{q+1} := \mathbf{j}_1$.

For example if $q = 3$, if $\mathbf{j}\mathbf{i} = (1, 2, 3, 4, 4, 2)$, then $i_1 = i_3 \neq i_2$, and j_1, j_2, j_3 are distinct, the circuit $(V(\mathbf{j}\mathbf{i}), C(\mathbf{j}\mathbf{i}))$ can be represented by the following graph:



which begins with \mathbf{j}_1 , and passes through the paths $\mathbf{j}_1 \rightarrow \mathbf{i}_1 \rightarrow \mathbf{j}_2 \rightarrow \mathbf{i}_2 \rightarrow \mathbf{j}_3 \rightarrow \mathbf{i}_3 \rightarrow \mathbf{j}_1$.

We define the multiplicity m_v of a vertex $v \in V(\mathbf{j}\mathbf{i})$ by

$$m_v := \sum_{l \in [1, q]} \mathbb{1}_{v=\mathbf{i}_l \text{ or } v=\mathbf{j}_l}.$$

For example, in the above graph the vertex \mathbf{i}_1 and \mathbf{i}_3 have multiplicity 2, and all others have multiplicity 1. In other words, the multiplicity of a vertex is the number of visits (where

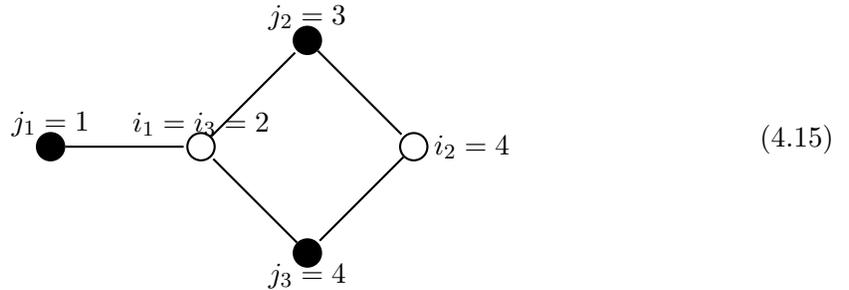
a departure and an arrival are considered as a visit) to this vertex by the circuit. We call every element of the sequence $C(\mathbf{j}\mathbf{i})$ a (directed) edge. We say that an edge $e = (v_1, v_2) \in C(\mathbf{j}\mathbf{i})$ is repeated if there is another edge $e' = (v'_1, v'_2) \in C(\mathbf{j}\mathbf{i})$ s.t. $v_1 = v'_1, v_2 = v'_2$ or $v_1 = v'_2, v_2 = v'_1$. We say that two eulerian circuits $\mathbf{Eu}(\mathbf{j}\mathbf{i}) = (V(\mathbf{j}\mathbf{i}), C(\mathbf{j}\mathbf{i}))$ and $\mathbf{Eu}(\mathbf{j}\mathbf{i}') = (V(\mathbf{j}\mathbf{i}'), C(\mathbf{j}\mathbf{i}'))$ are *congruent* iff there exists a bijection $\varphi : V(\mathbf{j}\mathbf{i}) \rightarrow V(\mathbf{j}\mathbf{i}')$ s.t. $C(\mathbf{j}\mathbf{i}') = ((\varphi(\mathbf{j}_l), \varphi(\mathbf{i}_l)), (\varphi(\mathbf{i}_l), \varphi(\mathbf{j}_{l+1})))_{l=1}^q$. In this case we denote $\mathbf{Eu}(\mathbf{j}\mathbf{i}) \cong \mathbf{Eu}(\mathbf{j}\mathbf{i}')$. Obviously two eulerian circuits engendered by $\mathbf{j}\mathbf{i}$ and $\mathbf{j}\mathbf{i}'$ are congruent iff the two $j\mathbf{i}$ -words $\mathbf{j}\mathbf{i}$ and $\mathbf{j}\mathbf{i}'$ are equivalent.

Inversely, a $2q$ -step eulerian circuit who begins and ends with a black vertex, and visits alternatively white and black vertices is said *eligible* to the set of $j\mathbf{i}$ -words $\{\mathbf{j}\mathbf{i}\}$. The set of all the eligible eulerian circuits is denoted by $\mathcal{E}_{j\mathbf{i},q}$. Here we do not distinguish two congruent eulerian circuits. We can see easily that $\#\mathcal{E}_{j\mathbf{i},q} < \infty$ for any q .

Given an eulerian circuit $\mathbf{Eu}(\mathbf{j}\mathbf{i}) = (V(\mathbf{j}\mathbf{i}), C(\mathbf{j}\mathbf{i}))$ we define its skeleton as a graph $S(\mathbf{j}\mathbf{i}) = (V(\mathbf{j}\mathbf{i}), E(\mathbf{j}\mathbf{i}))$, where the set of edges $E(\mathbf{j}\mathbf{i})$ are defined as follows:

$$E(\mathbf{j}\mathbf{i}) = \{\{\mathbf{j}_l, \mathbf{i}_l\}, \{\mathbf{i}_l, \mathbf{j}_{l+1}\}\}_{l=1}^q.$$

In other words, the skeleton of an eulerian circuit is a graph obtained by removing all the directions and repetitions of the edges. The skeleton of the eulerian circuit (4.14) is as follows:



After associating a $j\mathbf{i}$ -word $\mathbf{j}\mathbf{i}$ with an eulerian circuit $\mathbf{Eu}(\mathbf{j}\mathbf{i})$, we can rewrite the expansion (4.13) as

$$\frac{1}{\sqrt{n}} \mathbb{E} \operatorname{tr} \left(\frac{1}{n} G_{2:N, \cdot}^* D_{2:N, 2:N} G_{2:N, \cdot} \right)^q = \frac{1}{n^{q+1/2}} \sum_{\mathbf{Eu} \in \mathcal{E}_{j\mathbf{i},q}} \sum_{\mathbf{Eu}(\mathbf{j}\mathbf{i}) \cong \mathbf{Eu}} \left(\prod_{v=(i, \mathbf{w}) \in V(\mathbf{j}\mathbf{i})} d_i^{m_v} \right) P(\mathbf{j}\mathbf{i}),$$

where in the product $\prod_{v=(i,\mathbf{w}) \in V(\mathbf{j}\mathbf{i})} d_i^{m_v}$ the variable $v = (i, \mathbf{w})$ runs across the distinct white vertices in $V(\mathbf{j}\mathbf{i})$. Recall that by the definition of multiplicity, we have $\sum_{v=(i,\mathbf{w}) \in V(\mathbf{j}\mathbf{i})} m_v = q$. As the set $\mathcal{E}_{\mathbf{j}\mathbf{i},q}$ has a finite number of elements, we just need to prove, for each possible $\mathbf{E}\mathbf{u} \in \mathcal{E}_{\mathbf{j}\mathbf{i},q}$, the sum $\frac{1}{n^{q+1/2}} \sum_{\mathbf{E}\mathbf{u}(\mathbf{j}\mathbf{i}) \cong \mathbf{E}\mathbf{u}} \left(\prod_{v=(i,\mathbf{w}) \in V(\mathbf{j}\mathbf{i})} d_i^{m_v} \right) P(\mathbf{j}\mathbf{i})$ tends to 0.

By independence, $P(\mathbf{j}\mathbf{i})$ is zero unless all the edges of the eulerian circuit $\mathbf{E}\mathbf{u}(\mathbf{j}\mathbf{i})$ are repeated, so we only need to consider this kind of eulerian circuits. We fix such an eulerian circuit $\mathbf{E}\mathbf{u} = (V, C) \in \mathcal{E}_{\mathbf{j}\mathbf{i},q}$ and let $S = (V, E)$ be its skeleton. Because all the edges are repeated, we have $\#E \leq q$. By [29, Lemma 1.1], we have $\#V \leq q + 1$. So in V if there are w white vertices, there are at most $q + 1 - w$ black ones. Suppose without loss of generality that the w distinct white vertices are visited after the steps $1, 3, 5, \dots, 2w - 1$, and their multiplicities are m_1, m_2, \dots, m_w , then for any $\mathbf{E}\mathbf{u}(\mathbf{j}\mathbf{i}) \cong \mathbf{E}\mathbf{u}$, the distinct white vertices are $\mathbf{i}_1, \dots, \mathbf{i}_w$ with multiplicities m_1, \dots, m_w . Since $P(\mathbf{j}\mathbf{i})$ are uniformly bounded, and since j_1, \dots, j_q take values in $[[1, n]]$, we have then

$$\begin{aligned} \frac{1}{n^{q+1/2}} \sum_{\mathbf{E}\mathbf{u}(\mathbf{j}\mathbf{i}) \cong \mathbf{E}\mathbf{u}} d_{i_1}^{m_1} \dots d_{i_w}^{m_w} P(\mathbf{j}\mathbf{i}) &= O \left(\frac{\sqrt{n}}{n^q} \sum_{i_1 \dots i_w \in [[2, N]] \text{ distinct}} d_{i_1}^{m_1} \dots d_{i_w}^{m_w} \right) \\ &\leq O \left(\frac{\sqrt{n}}{n^q} \sum_{i_1 \dots i_w \in [[2, N]]} d_{i_1}^{m_1} \dots d_{i_w}^{m_w} \right), \end{aligned}$$

By Hölder's inequality and the assumption **A10** we have

$$O \left(\frac{\sqrt{n}}{n^q} \sum_{i_1, \dots, i_w} d_{i_1}^{m_1} \dots d_{i_w}^{m_w} \right) = O \left(\sqrt{n} \left(\frac{\sum_{i=2}^N d_i^{m_1}}{n} \right) \dots \left(\frac{\sum_{i=2}^N d_i^{m_w}}{n} \right) \right) \leq O \left(\frac{\sum_{i=2}^N d_i^q}{\sqrt{n}} \right) \rightarrow 0$$

as $N, n \rightarrow \infty$. Therefore (4.9) is proved.

Remark 15. A consequence of the above proof is that for any integer $k \geq 1$, we have in fact

$$\frac{1}{n} \mathbb{E} \operatorname{tr} \left(\frac{1}{n} G_{2:N}^* D_{2:N, 2:N} G_{2:N} \right)^k \leq O \left(\frac{\sum_{i=2}^N d_i^k}{n} \right) \rightarrow 0$$

as $N, n \rightarrow \infty$. This result will be used soon after.

Next we prove that (4.8) holds for any $k \geq 1$. We also use the combinatorial tools as in the proof of (4.9). Analogous ideas can be found in [29, Theorem 1.13] (part 4 of the

proof), or [17, Lemma 2.7].

We expand the trace and the variance and we get

$$\begin{aligned}
 & \text{Var} \frac{1}{\sqrt{n}} \text{tr} \left(\frac{1}{n} G_{2:N, \cdot}^* D_{2:N, 2:N} G_{2:N, \cdot} \right)^k \\
 &= \frac{1}{n^{2k+1}} \left(\mathbb{E} (\text{tr} (G_{2:N, \cdot}^* D_{2:N, 2:N} G_{2:N, \cdot})^k)^2 - (\mathbb{E} \text{tr} (G_{2:N, \cdot}^* D_{2:N, 2:N} G_{2:N, \cdot})^k)^2 \right) \\
 &= \frac{1}{n^{2k+1}} \sum_{\substack{i_1 \dots i_k, i'_1, \dots, i'_k \in [[2, N]] \\ j_1 \dots j_k, j'_1, \dots, j'_k \in [[1, n]]}} d_{i_1} \dots d_{i_k} d_{i'_1} \dots d_{i'_k} (P(\mathbf{j}\mathbf{i}, \mathbf{j}\mathbf{i}') - P(\mathbf{j}\mathbf{i})P(\mathbf{j}\mathbf{i}')) \\
 &\leq \frac{1}{n^{2k+1}} \sum_{\substack{i_1 \dots i_k, i'_1, \dots, i'_k \in [[2, N]] \\ j_1 \dots j_k, j'_1, \dots, j'_k \in [[1, n]]}} |P(\mathbf{j}\mathbf{i}, \mathbf{j}\mathbf{i}') - P(\mathbf{j}\mathbf{i})P(\mathbf{j}\mathbf{i}')|,
 \end{aligned}$$

where $P(\mathbf{j}\mathbf{i}, \mathbf{j}\mathbf{i}') := \mathbb{E} \overline{G_{i_1, j_1}} G_{i_1, j_2} \dots \overline{G_{i_k, j_k}} G_{i_k, j_1} \overline{G_{i'_1, j'_1}} G_{i'_1, j'_2} \dots \overline{G_{i'_k, j'_k}} G_{i'_k, j'_1}$.

Similarly as the proof of (4.9), we associate the $2k$ -long j *i*-words $\mathbf{j}\mathbf{i}$ and $\mathbf{j}\mathbf{i}'$ with eulerian circuits $\mathbf{Eu}(\mathbf{j}\mathbf{i})$ and $\mathbf{Eu}(\mathbf{j}\mathbf{i}')$. Let the set of all the eligible eulerian circuits be $\mathcal{E}_{\mathbf{j}\mathbf{i}, k}$. Then we have

$$\begin{aligned}
 & \text{Var} \frac{1}{\sqrt{n}} \text{tr} \left(\frac{1}{n} G_{2:N, \cdot}^* D_{2:N, 2:N} G_{2:N, \cdot} \right)^k \\
 &\leq \frac{1}{n^{2k+1}} \sum_{\mathbf{Eu}, \mathbf{Eu}' \in \mathcal{E}_{\mathbf{j}\mathbf{i}, k}} \sum_{\mathbf{Eu}(\mathbf{j}\mathbf{i}) \cong \mathbf{Eu}, \mathbf{Eu}(\mathbf{j}\mathbf{i}') \cong \mathbf{Eu}'} |P(\mathbf{j}\mathbf{i}, \mathbf{j}\mathbf{i}') - P(\mathbf{j}\mathbf{i})P(\mathbf{j}\mathbf{i}')|.
 \end{aligned}$$

For every pair $\mathbf{Eu} = (V, C)$, $\mathbf{Eu}' = (V', C')$, let $S = (V, E)$, $S' = (V', E')$ be their skeletons. We define the union of S and S' by

$$S \cup S' = (V \cup V', E \cup E').$$

By the independence, if S and S' do not have edges in common, we will have $P(\mathbf{j}\mathbf{i}, \mathbf{j}\mathbf{i}') = P(\mathbf{j}\mathbf{i})P(\mathbf{j}\mathbf{i}')$; if there is an edge which is not in common, i.e. $e \in (E - E') \cup (E' - E)$, and the corresponding edge in \mathbf{Eu} or \mathbf{Eu}' is simple, then we will have $P(\mathbf{j}\mathbf{i}, \mathbf{j}\mathbf{i}') = 0 = P(\mathbf{j}\mathbf{i})P(\mathbf{j}\mathbf{i}')$. Hence we may restrict to the case where $S \cup S'$ is connected, and each edge of $S \cup S'$ is either in common, or repeated in the original eulerian circuit. So we have $\#(E \cup E') \leq 2k$. By [29, Lemma 1.1] again, we have $\#(V \cup V') \leq 2k + 1$. But the case where $\#(V \cup V') = 2k + 1$ is impossible, because if it were true, $S \cup S'$ would be a tree, and thus there should not be

any cycle in S or S' , so S and S' should be trees; then let $e \in E \cap E'$, the corresponding edges of e in \mathbf{Eu} and \mathbf{Eu}' should be both repeated, implying that e should be repeated at least 4 times in total in \mathbf{Eu} and \mathbf{Eu}' , and thus $\#(E \cup E') \leq 2k - 1$, this would imply $\#(V \cup V') \leq 2k$ by [29, Lemma 1.1] and contradict with $\#(V \cup V') = 2k + 1$. So in fact we have $\#(V \cup V') \leq 2k$. Hence by the uniform bound of $P(\mathbf{ji}, \mathbf{ji}')$, $P(\mathbf{ji})$ and $P(\mathbf{ji}')$, we have

$$\frac{1}{n^{2k+1}} \sum_{\mathbf{Eu}(\mathbf{ji}) \cong \mathbf{Eu}, \mathbf{Eu}(\mathbf{ji}') \cong \mathbf{Eu}'} |P(\mathbf{ji}, \mathbf{ji}') - P(\mathbf{ji})P(\mathbf{ji}')| \leq O(1/n) \rightarrow 0$$

as $N, n \rightarrow \infty$ and $N/n \rightarrow r \in (0, \infty)$. Therefore (4.8) is proved for any $k \geq 1$.

To prove (4.6), we write $Q_G := G_{1,\cdot} (\frac{1}{n} G_{2:N,\cdot}^* D_{2:N,2:N} G_{2:N,\cdot})^k G_{1,\cdot}^*$. Then we have

$$\text{Var} \frac{1}{\sqrt{n}} Q_G = \frac{1}{n} \mathbb{E} \mathbb{E}_{G_{2:N,\cdot}} |Q_G - \mathbb{E}_{G_{2:N,\cdot}} Q_G|^2 + \frac{1}{n} \mathbb{E} |\mathbb{E}_{G_{2:N,\cdot}} Q_G - \mathbb{E} Q_G|^2.$$

Then applying [1, Lemma 2.7] with $p = 2$, and use the result in Remark 15, we have

$$\frac{1}{n} \mathbb{E} \mathbb{E}_{G_{2:N,\cdot}} |Q_G - \mathbb{E}_{G_{2:N,\cdot}} Q_G|^2 \leq C \frac{\mathbb{E} |G_{1,1}|^4}{n} \mathbb{E} \text{tr} \left(\frac{1}{n} G_{2:N,\cdot}^* D_{2:N,2:N} G_{2:N,\cdot} \right)^{2k} = o(1).$$

For the other part, note that $\mathbb{E}_{G_{2:N,\cdot}} Q_G = \text{tr} \left(\frac{1}{n} G_{2:N,\cdot}^* D_{2:N,2:N} G_{2:N,\cdot} \right)^k$, so we have

$$\frac{1}{n} \mathbb{E} |\mathbb{E}_{G_{2:N,\cdot}} Q_G - \mathbb{E} Q_G|^2 = \text{Var} \frac{1}{\sqrt{n}} \text{tr} \left(\frac{1}{n} G_{2:N,\cdot}^* D_{2:N,2:N} G_{2:N,\cdot} \right)^k = o(1)$$

according to (4.8).

In the following we focus on the proof of (4.7) and (4.5). In fact we will prove that under the assumptions of Theorem I.15, for any $k \geq 1$, we have

$$\mathbb{E} Y_{1,\cdot} \left(\frac{1}{n} Y_{2:N,\cdot}^* D_{2:N,2:N} Y_{2:N,\cdot} \right)^k Y_{1,\cdot}^* - \mathbb{E} G_{1,\cdot} \left(\frac{1}{n} G_{2:N,\cdot}^* D_{2:N,2:N} G_{2:N,\cdot} \right)^k G_{1,\cdot}^* = o(1). \quad (4.16)$$

We write

$$Q_Y = Y_{1,\cdot} \left(\frac{1}{n} Y_{2:N,\cdot}^* D_{2:N,2:N} Y_{2:N,\cdot} \right)^k Y_{1,\cdot}^* \quad \text{and} \quad Q_G = G_{1,\cdot} \left(\frac{1}{n} G_{2:N,\cdot}^* D_{2:N,2:N} G_{2:N,\cdot} \right)^k G_{1,\cdot}^* .$$

Expanding the matrix multiplication we have

$$\mathbb{E}Q_Y = \frac{1}{n^k} \sum_{\substack{i_1, \dots, i_k \in [[2, N]] \\ j_1, \dots, j_{k+1} \in [[1, n]]}} d_{i_1} \dots d_{i_k} P_Y(\mathbf{ij})$$

where $P_Y(\mathbf{ij}) := \mathbb{E}Y_{1,j_1} \overline{Y_{i_1,j_1}} Y_{i_1,j_2} \dots \overline{Y_{i_k,j_k}} Y_{i_k,j_{k+1}} \overline{Y_{1,j_{k+1}}}$. We call $\mathbf{ij} := 1j_1i_1j_2 \dots i_kj_{k+1}1$ a ij -word. The difference between an ij -word and a ji -word defined in the proof of (4.9) is that an ij -word begins and ends with a row-label, while a ji -word begins and ends with a column-label. With each ij -word we associate an eulerian circuits $\mathbf{Eu}(\mathbf{ij}) = (V(\mathbf{ij}), C(\mathbf{ij}))$ defined as follows:

$$V(\mathbf{ij}) := \{(1, \mathbf{r}), (i_1, \mathbf{w}), (j_1, \mathbf{b}), \dots, (i_k, \mathbf{w}), (j_{k+1}, \mathbf{b})\}$$

is the set of vertices, where we color the fixed row label “1” on red, the other vertices are colored as before. The vertices are denoted as $\mathbf{1}$, \mathbf{i}_l and \mathbf{j}_l . We use the convention $\mathbf{i}_0 = \mathbf{i}_{k+1} = \mathbf{1}$, then we define the sequence of steps

$$E(\mathbf{ij}) := ((\mathbf{i}_l, \mathbf{j}_l), (\mathbf{j}_l, \mathbf{i}_{l+1}))_{l=0}^k.$$

Inversely we say a $(2k + 2)$ -step eulerian circuit eligible to the set of ij -words $\{\mathbf{ij}\}$ if it begins with a red vertex, then visits a black vertex, and then pass white and black vertices alternatively, and it returns to the red vertex from a black vertex. The set of all eligible eulerian circuits is denoted by $\mathcal{E}_{ij,k}$. For an eligible eulerian circuit $\mathbf{Eu} \in \mathcal{E}_{ij,k}$, a black vertex is said to be in the first class if all the edges out and into it are repeated; otherwise it is said to be in the second class. Let w, c_1, c_2 denote the number of white vertices, first class black vertices and second class vertices respectively. Then we use Lemma IV.5 (page 130) to conclude that, if $c_2 = 0$, we have $w + c_1 \leq k + 1$, and if $c_2 \neq 0$, we have $w + c_1 \leq k$. Indeed when we apply Lemma IV.5 to \mathbf{Eu} , we temporarily color the beginning and ending vertex, i.e. the red vertex, on white. In this new graph we have $w + 1$ white vertices. Then from Lemma IV.5 we conclude that if $c_2 = 0$, we have $1 + w + c_1 \leq k + 2$, and if $c_2 \neq 0$, we have $1 + w + c_1 \leq k + 1$. Recoloring the beginning and ending vertex to red, we get the

mentioned result.

Returning to the proof of (4.7), we have

$$\mathbb{E}Q_Y - \mathbb{E}Q_G = \frac{1}{n^k} \sum_{\mathbf{Eu} \in \mathcal{E}_{ij,k}} \sum_{\mathbf{Eu}(\mathbf{ij}) \cong \mathbf{Eu}} d_{i_1} \dots d_{i_k} (P_Y(\mathbf{ij}) - P_G(\mathbf{ij})).$$

For a fixed $\mathbf{Eu} \in \mathcal{E}_{ij}$ and any $\mathbf{Eu}(\mathbf{ij}) \cong \mathbf{Eu}$, we assume without loss of generality that the distinct white vertices are $\mathbf{i}_1, \dots, \mathbf{i}_w$ with multiplicities m_1, \dots, m_w and recall that $m_1 + \dots + m_w = k$. As the columns of Y are independent, and each distinct black vertex corresponds a column, we have

$$P_Y(\mathbf{ij}) = \prod_{(j,\mathbf{b}) \in V(\mathbf{ij})} \mathbb{E} \left[\prod_{((i,\mathbf{w}), (j,\mathbf{b})) \in C(\mathbf{ij})} Y_{i,j} \prod_{((j,\mathbf{b}), (i,\mathbf{w})) \in C(\mathbf{ij})} \overline{Y_{i,j}} \right].$$

We can decompose similarly $P_G(\mathbf{ij})$ as above. For each $\mathbf{j} := (j, \mathbf{b}) \in V(\mathbf{ij})$, denote $E_Y(\mathbf{j}) := \mathbb{E} \left[\prod_{((i,\mathbf{w}), (j,\mathbf{b})) \in C(\mathbf{ij})} Y_{i,j} \prod_{((j,\mathbf{b}), (i,\mathbf{w})) \in C(\mathbf{ij})} \overline{Y_{i,j}} \right]$ and similarly $E_G(\mathbf{j})$ is the expectation obtained from E_Y by replacing Y with G . By Lemma IV.4, we have $E_Y(\mathbf{j}) - E_G(\mathbf{j}) = O(1/N)$. Now we estimate $P_Y(\mathbf{ij}) - P_G(\mathbf{ij})$. According to the case of \mathbf{Eu} , we consider respectively two cases:

- Case 1: $c_2 \neq 0$. Then we have $P_G(\mathbf{ij}) = 0$, and from the decomposition above, we have

$$P_Y(\mathbf{ij}) = O(1/N^{c_2}).$$

and

$$\begin{aligned} n^{-k} \sum_{\mathbf{Eu}(\mathbf{ij}) \cong \mathbf{Eu}} d_{i_1}^{m_1} \dots d_{i_w}^{m_w} |P_Y(\mathbf{ij}) - P_G(\mathbf{ij})| &= O(n^{c_1-k} \sum d_{i_1}^{m_1} \dots d_{i_w}^{m_w}) \\ &= O(n^{c_1+w-k} (\frac{1}{n} \sum d_i^{m_1}) \dots (\frac{1}{n} \sum d_i^{m_w})). \end{aligned}$$

From Lemma IV.5, in this case we have $w + c_1 \leq k$, so by Assumption A10, the above sum converges to 0.

- Case 2: $c_2 = 0$. Suppose that the distinct black vertices are $\mathbf{j}_1, \dots, \mathbf{j}_{c_1}$. We have then

$$P_Y(\mathbf{ij}) - P_G(\mathbf{ij}) = [E_G(\mathbf{j}_1) + O(1/N)] \dots [E_G(\mathbf{j}_{c_1}) + O(1/N)] - E_G(\mathbf{j}_1) \dots E_G(\mathbf{j}_{c_1}) = O(1/N).$$

Hence

$$\begin{aligned} n^{-k} \sum_{\mathbf{Eu}(\mathbf{ij}) \cong \mathbf{Eu}} d_{i_1}^{m_1} \dots d_{i_w}^{m_w} |P_Y(\mathbf{ij}) - P_G(\mathbf{ij})| &= O\left(\frac{1}{n^{k-c_1} N} \sum d_{i_1}^{a_1} \dots d_{i_w}^{a_w}\right) \\ &= O(n^{c_1+w-k-1} (n^{-1} \sum d_i^{a_1}) \dots (n^{-1} \sum d_i^{a_w})). \end{aligned}$$

From Lemma IV.5, in this case we have $w + c_1 \leq k + 1$, so by Assumption A2, the above sum converges to 0.

The proof of (4.7) is now complete.

For (4.5), we write

$$\text{Var} \frac{1}{\sqrt{n}} Q_Y - \text{Var} \frac{1}{\sqrt{n}} Q_G = \frac{1}{n} (\mathbb{E} Q_Y^2 - \mathbb{E} Q_G^2) - (\mathbb{E} Q_Y - \mathbb{E} Q_G) \left(\frac{1}{n} (\mathbb{E} Q_Y - \mathbb{E} Q_G) + \frac{2}{n} \mathbb{E} Q_G \right).$$

From (4.16) and Remark 15, we have $(\mathbb{E} Q_Y - \mathbb{E} Q_G) \left(\frac{1}{n} (\mathbb{E} Q_Y - \mathbb{E} Q_G) + \frac{2}{n} \mathbb{E} Q_G \right) \rightarrow 0$. Now we prove that $\frac{1}{n} (\mathbb{E} Q_Y^2 - \mathbb{E} Q_G^2) \rightarrow 0$. We write

$$\frac{1}{n} \mathbb{E} Q_Y^2 = \frac{1}{n^{2k+1}} \sum_{\substack{i_1 \dots i_{2k} \in \llbracket 2, N \rrbracket \\ j_1 \dots j_{2k+2} \in \llbracket 1, n \rrbracket}} d_{i_1} \dots d_{i_{2k}} P_Y(\mathbf{ij}_1 \mathbf{ij}_2)$$

where

$$P_Y(\mathbf{ij}_1 \mathbf{ij}_2) := \mathbb{E} \overline{Y_{1,j_1}} \overline{Y_{i_1,j_1}} Y_{i_1,j_2} \dots Y_{i_k,j_{k+1}} \overline{Y_{1,j_{k+1}}} Y_{1,j_{k+2}} \overline{Y_{i_{k+1},j_{k+2}}} \dots \overline{Y_{i_{2k},j_{2k+1}}} Y_{i_{2k},j_{2k+2}} \overline{Y_{1,j_{2k+2}}}.$$

We argue in the same way as the proof of (4.7). We consider the set of ij -words $\{\mathbf{ij}_1 \mathbf{ij}_2\}$, and consider the set of eulerian circuits $\mathcal{E}_{\mathbf{ij},k} := \{\mathbf{Eu}(\mathbf{ij}_1 \mathbf{ij}_2)\} = \{(V(\mathbf{ij}_1 \mathbf{ij}_2), C(\mathbf{ij}_1 \mathbf{ij}_2))\}$, where

$$V(\mathbf{ij}_1 \mathbf{ij}_2) = \{(1, \mathbf{r}), (i_1, \mathbf{w}), \dots, (i_{2k}, \mathbf{w}), (j_1, \mathbf{b}), \dots, (j_{2k+2}, \mathbf{b})\},$$

and

$$C(\mathbf{ij}_1 \mathbf{ij}_2) = ((\mathbf{1}, \mathbf{j}_1), ((\mathbf{j}_l, \mathbf{i}_l), (\mathbf{i}_l, \mathbf{j}_{l+1}))_{l=1}^k, (\mathbf{j}_{k+1}, \mathbf{1}), (\mathbf{1}, \mathbf{j}_{k+2}), ((\mathbf{j}_l, \mathbf{i}_{l+1}), (\mathbf{i}_{l+1}, \mathbf{j}_{l+2}))_{l=k+1}^{2k}, (\mathbf{j}_{2k+2}, \mathbf{1})).$$

Let w, c_1, c_2 denote the number of white vertices, first class black vertices and second class vertices respectively, where the definitions of first and second class black vertices is in the proof of (4.7). Using Lemma IV.5, we conclude that if $c_2 = 0$, we have $w + c_1 \leq 2k + 2$, and if $c_2 \neq 0$, we have $w + c_1 \leq 2k + 1$. Then an argument analog to the proof of (4.7) leads to the result. This complete the proof of Theorem I.15.

Lemma IV.5. *Let \mathcal{E}_{2k} be the set of all $2k$ -step eulerian circuits beginning and ending with a white vertex, and passing across black and white vertices alternatively. For an eulerian circuit $\mathbf{Eu} \in \mathcal{E}_{2k}$, we define first class and second class of black vertices as in the proof of (4.7), and let w, c_1, c_2 be the numbers of white, first class and second class black vertices. Then if $c_2 = 0$, we have $w + c_1 \leq k + 1$; if $c_2 \neq 0$, we have $w + c_1 \leq k$.*

Proof. If $c_2 = 0$, i.e. all the edges are repeated, then the skeleton has at most k edges, so there are at most $k + 1$ distinct vertices. We obviously have $w + c_1 \leq k + 1$.

If $c_2 \neq 0$, i.e. there is at least one simple edge, we prove the inequality $w + c_1 \leq k$ by induction on k . For $k = 1$, we have nothing to prove because in this case $c_2 = 0$. Suppose we have proved the inequality for all $2r$ -step eulerian circuits with $r < k$. Then for an eulerian circuit $\mathbf{Eu} = (V, C) \in \mathcal{E}_{2k}$ such that $c_2 \neq 0$, if all the edges are simple, we have $c_1 = 0$ and $w + c_1 = w = k$; if there is at least one repeated edge, say the edge connecting \mathbf{i}_1 and \mathbf{j}_1 . We remove two repetitions of this edge and we denote the remaining part as $\mathbf{Eu}^{\{\mathbf{i}_1, \mathbf{j}_1\}}$. There are three possibilities:

- Case 1: $\mathbf{Eu}^{\{\mathbf{i}_1, \mathbf{j}_1\}}$ is connected. In this case, as the degree of each vertex is still even. By the famous Euler Theorem, we can still draw $\mathbf{Eu}^{\{\mathbf{i}_1, \mathbf{j}_1\}}$ in a single trajectory without lifting up the pen (but with changing the order and directions of edges if necessary). And because each edge connects still white and black vertices, we have $\mathbf{Eu}^{\{\mathbf{i}_1, \mathbf{j}_1\}} \in \mathcal{E}_{2k-2}$. Let w', c'_1, c'_2 are the numbers of white vertices, first class black vertices and second class black vertices respectively of $\mathbf{Eu}^{\{\mathbf{i}_1, \mathbf{j}_1\}}$. We have assumed

that \mathbf{Eu} has at least one simple edge, and this simple edge is obviously left in $\mathbf{Eu}^{\{i_1, j_1\}}$, so $c'_2 \neq 0$. Moreover, we have $w' = w$, and all the black vertices except j_1 remain in the same class as in \mathbf{Eu} , so we have $c'_1 \geq c_1 - 1$. By induction we have $w + c'_1 \leq k - 1$, thus $w + c_1 \leq k$.

- Case 2: i_1 or j_1 is disconnected from the other vertices. We remove this isolate vertex. After doing that w or c_1 decreases by one. Using Euler Theorem as in the first case, it is easy to conclude from the induction.
- Case 3: $\mathbf{Eu}^{\{i_1, j_1\}}$ is disconnected and is separated into two connected components. Using Euler Theorem again on the two components and inducing from these two components we obtain the inequality for \mathbf{Eu} .

□

APPENDICES

APPENDIX A

Some complementary proofs

A.1 Proof of Proposition I.6

Let $\gamma(h) = (1+|h|)^\rho L(k)$ be the serie of first row entries, with L a slowly varying function and $\rho \in (0, 1)$. For the case where $\rho \in (-3/4, 0)$, the result is prooved by Proposition I.5 and the proof of that proposition can be found in Chapter III, Proposition III.3. We only need to verify the result for $\rho \in (-1, -3/4]$. For these ρ , the length of the interval $(\frac{1}{2(1+\rho)}, \frac{1}{(1+\rho)})$ is greater than 1, so there must be an integer $q \geq 2$ in this interval. For this q , let $p = \frac{q}{q-1}$. Then because $-p\rho > 1$, we have $\gamma \in l^p$. And we have also $p \leq 2$, from Hausdorff-Young Theorem (Theorem 2.3-ii of [61]), the symbol of these Toeplitz matrices f is in L^q . By Theorem 2 of [21], we have

$$\lim_{N \rightarrow \infty} \frac{\text{tr } T_N^q}{N} = \frac{1}{2\pi} \int_{(-\pi, \pi]} |f(x)|^q dx < \infty.$$

So we have

$$\frac{\text{tr } T_N^q}{\sqrt{N} \lambda_{\max}^q(T_N)} = O\left(\frac{\sqrt{N}}{(N^{1+\rho} L(N))^q}\right) \rightarrow 0$$

since $(1 + \rho)q > \frac{1}{2}$.

A.2 Proof of Proposition I.8

Proposition II.11 has proved the case of the largest one eigenvalue. With some adaptations the same method can prove the general case.

For a fixed $j = 1, \dots, m$, let

$$S_{N,(j)} := \frac{S_N(\Gamma)}{\lambda_j(\Gamma_N)}.$$

The goal is to prove that $\lambda_j(S_{N,(j)}) \rightarrow 1$ in probability or almost surely. We denote

$$d_1 \geq \dots \geq d_N$$

the eigenvalues of $\Gamma_N/\lambda_j(\Gamma_N)$. let U_N be a unitary matrix such that

$$\Gamma_N = \lambda_j(\Gamma_N) U_N \text{diag}(d_1, \dots, d_N) U_N^*.$$

Let also

$$\widehat{S}_{N,(j)} := \frac{1}{n} Z_N^* U_N \text{diag}(d_1, \dots, d_j, 0, \dots) U_N^* Z_N,$$

and

$$\widetilde{S}_{N,(j)} := \frac{1}{n} Z_N^* U_N \text{diag}(d_1 \vee 2, \dots, d_{j-1} \vee 2, d_j, d_{j+1}, \dots, d_N) U_N^* Z_N.$$

Then by the same argument of spectral confinement used in Proposition II.11, we have

$$1 = \underline{\lim}_{N \rightarrow \infty} \lambda_j(\widehat{S}_{N,(j)}) \leq \underline{\lim}_{N \rightarrow \infty} \lambda_j(S_{N,(j)}) \leq \overline{\lim}_{N \rightarrow \infty} \lambda_j(S_{N,(j)}) \leq \overline{\lim}_{N \rightarrow \infty} \lambda_j(\widetilde{S}_{N,(j)}) = 1$$

in probability or almost surely. The result follows.

BIBLIOGRAPHY

BIBLIOGRAPHY

- [1] Z.D. Bai and J.W. Silverstein. No eigenvalues outside the support of the limiting spectral distribution of large-dimensional sample covariance matrices. *Annals of probability*, pages 316–345, 1998.
- [2] Z.D. Bai and J.W. Silverstein. Exact separation of eigenvalues of large dimensional sample covariance matrices. *Annals of probability*, pages 1536–1555, 1999.
- [3] Z.D. Bai and J.W. Silverstein. CLT for linear spectral statistics of large-dimensional sample covariance matrices. *The Annals of Probability*, 32(1A):553–605, 2004.
- [4] Z.D. Bai and J.W. Silverstein. *Spectral analysis of large dimensional random matrices*, volume 20. Springer, 2010.
- [5] Z.D. Bai, J.W. Silverstein, and Y.Q. Yin. A note on the largest eigenvalue of a large dimensional sample covariance matrix. *Journal of Multivariate Analysis*, 26(2):166–168, 1988.
- [6] Z.D. Bai and J.F. Yao. Central limit theorems for eigenvalues in a spiked population model. In *Annales de l'IHP Probabilités et statistiques*, volume 44, pages 447–474, 2008.
- [7] Z.D. Bai and J.F. Yao. On sample eigenvalues in a generalized spiked population model. *Journal of Multivariate Analysis*, 106:167–177, 2012.
- [8] Z.D. Bai and Y.Q. Yin. Limit of the smallest eigenvalue of a large dimensional sample covariance matrix. *The annals of Probability*, pages 1275–1294, 1993.
- [9] J. Baik, G. Ben Arous, and S. Péché. Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices. *Annals of Probability*, pages 1643–1697, 2005.
- [10] J. Baik and J.W. Silverstein. Eigenvalues of large sample covariance matrices of spiked population models. *Journal of Multivariate Analysis*, 97(6):1382–1408, 2006.
- [11] Z.G. Bao, G.M. Pan, and W. Zhou. Universality for the largest eigenvalue of sample covariance matrices with general population. *The Annals of Statistics*, 43(1):382–421, 2015.
- [12] F. Benaych-Georges, A. Guionnet, and M. Maida. Fluctuations of the extreme eigenvalues of finite rank deformations of random matrices. *Electronic Journal of Probability*, 16:1621–1662, 2011.

- [13] F. Benaych-Georges and R. R. Nadakuditi. The singular values and vectors of low rank perturbations of large rectangular random matrices. *Journal of Multivariate Analysis*, 111:120–135, 2012.
- [14] N.H. Bingham, C.M. Goldie, and J.L. Teugels. *Regular variation*, volume 27. Cambridge university press, 1989.
- [15] A. Boeóttcher and S.M. Grudsky. *Spectral properties of banded Toeplitz matrices*, volume 96. Siam, 2005.
- [16] J.M. Bogoya, A. Böttcher, and S.M. Grudsky. Eigenvalues of hermitian toeplitz matrices with polynomially increasing entries. *Journal of Spectral Theory*, 2(3):267–292, 2012.
- [17] C. Bordenave. A short course on random matrices (preliminary draft). <https://pdfs.semanticscholar.org/08b0/20431a2e740e9e19b536ef8532f15d10c3f9.pdf>, 2014. [Online; accessed 1-January-2017].
- [18] A. Böttcher and B. Silbermann. *Introduction to large truncated Toeplitz matrices*. Springer Science & Business Media, 2012.
- [19] S. Boucheron, G. Lugosi, and P. Massart. *Concentration inequalities: A nonasymptotic theory of independence*. Oxford university press, 2013.
- [20] T. Cai, X. Han, and G. Pan. Limiting laws for divergent spiked eigenvalues and largest non-spiked eigenvalue of sample covariance matrices. *arXiv preprint arXiv:1711.00217*, 2017.
- [21] S Serra Capizzano. Test functions, growth conditions and toeplitz matrices. In *Proceedings of the Fourth International Conference on Functional Analysis and Approximation Theory*, volume 2, pages 791–795, 2002.
- [22] A. Chakrabarty, R. S. Hazra, and D. Sarkar. From random matrices to long range dependence. *Random Matrices: Theory and Applications*, 5(02):1650008, 2016.
- [23] R. Couillet and W. Hachem. Fluctuations of spiked random matrix models and failure diagnosis in sensor networks. *IEEE Transactions on Information Theory*, 59(1):509–525, 2013.
- [24] E.B. Davies. *Linear operators and their spectra*, volume 106. Cambridge University Press, 2007.
- [25] N. El Karoui. Tracy-widom limit for the largest eigenvalue of a large class of complex sample covariance matrices. *The Annals of Probability*, pages 663–714, 2007.
- [26] M. Forni, M. Hallin, M. Lippi, and L. Reichlin. The generalized dynamic-factor model: Identification and estimation. *The review of Economics and Statistics*, 82(4):540–554, 2000.
- [27] S. Geman. A limit theorem for the norm of random matrices. *The Annals of Probability*, pages 252–261, 1980.

- [28] U. Grenander and G. Szegö. *Toeplitz forms and their applications*, volume 321. Univ of California Press, 2001.
- [29] A Guionnet. Large random matrices: Lectures on macroscopic asymptotics, 2008.
- [30] R.A. Horn and C.R. Johnson. *Matrix analysis*. Cambridge university press, 2012.
- [31] N. Jacob. *Pseudo Differential Operators & Markov Processes: Markov Processes And Applications*, volume 1. Imperial College Press, 2001.
- [32] K. Johansson. Shape fluctuations and random matrices. *Communications in mathematical physics*, 209(2):437–476, 2000.
- [33] I.M. Johnstone. On the distribution of the largest eigenvalue in principal components analysis. *Annals of statistics*, pages 295–327, 2001.
- [34] D. Jonsson. Some limit theorems for the eigenvalues of a sample covariance matrix. *Journal of Multivariate Analysis*, 12(1):1–38, 1982.
- [35] S. Jung and J. S. Marron. Pca consistency in high dimension, low sample size context. *The Annals of Statistics*, 37(6B):4104–4130, 2009.
- [36] O. Kallenberg. *Foundations of modern probability*. Springer Science & Business Media, 2006.
- [37] A. Knowles and J. Yin. Anisotropic local laws for random matrices. *Probability Theory and Related Fields*, 169(1-2):257–352, 2017.
- [38] O. Ledoit and M. Wolf. Optimal estimation of a large-dimensional covariance matrix under Stein’s loss. *Accepted for publication in Bernoulli*, 2017.
- [39] J.O. Lee and K. Schnelli. Tracy–widom distribution for the largest eigenvalue of real sample covariance matrices with general population. *The Annals of Applied Probability*, 26(6):3786–3839, 2016.
- [40] J. López-Gómez and C. Mora-Corral. *Algebraic Multiplicity of Eigenvalues of Linear Operators*. Operator Theory: Advances and Applications. Birkhäuser Basel, 2007.
- [41] V.A. Marčenko and L.A. Pastur. Distribution of eigenvalues for some sets of random matrices. *Mathematics of the USSR-Sbornik*, 1(4):457, 1967.
- [42] F. Merlevède, J. Najim, and P. Tian. Unbounded largest eigenvalue of large sample covariance matrices: Asymptotics, fluctuations and applications. *arXiv preprint arXiv:1802.01874*, 2018.
- [43] F. Merlevède and M. Peligrad. On the empirical spectral distribution for matrices with long memory and independent rows. *Stochastic Processes and their Applications*, 126(9):2734–2760, 2016.
- [44] J. Najim and J.F. Yao. Gaussian fluctuations for linear spectral statistics of large random covariance matrices. *The Annals of Applied Probability*, 26(3):1837–1887, 2016.
- [45] V. Pipiras and M.S. Taqqu. *Long-Range Dependence and Self-Similarity*. Cambridge university press, 2017.

- [46] R. V. Rao. On the eigenvalues of integral operators with difference kernels. *Journal of Mathematical Analysis and Applications*, 53(3):554–566, 1976.
- [47] D. Shen, H. Shen, H. Zhu, and J.S. Marron. Surprising asymptotic conical structure in critical sample eigen-directions. *arXiv preprint arXiv:1303.6171*, 2013.
- [48] J.W. Silverstein. Strong convergence of the empirical distribution of eigenvalues of large dimensional random matrices. *Journal of Multivariate Analysis*, 55(2):331–339, 1995.
- [49] 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*, 54(2):175–192, 1995.
- [50] J.W. Silverstein and Sang-II Choi. Analysis of the limiting spectral distribution of large dimensional random matrices. *Journal of Multivariate Analysis*, 54(2):295–309, 1995.
- [51] G. Teschl. *Mathematical methods in quantum mechanics*, volume 157. American Mathematical Soc., 2014.
- [52] P. Tian. Joint clt for top eigenvalues of empirical covariance matrices of long memory stationary processes. *Submitted to Annals of applied probability*, 2018.
- [53] C.A. Tracy and H. Widom. Level-spacing distributions and the airy kernel. *Physics Letters B*, 305(1-2):115–118, 1993.
- [54] C.A. Tracy and H. Widom. On orthogonal and symplectic matrix ensembles. *Communications in Mathematical Physics*, 177(3):727–754, 1996.
- [55] E.E. Tyrtyshnikov and N.L. Zamarashkin. Toeplitz eigenvalues for radon measures. *Linear algebra and its applications*, 343:345–354, 2002.
- [56] K.W. Wachter. The strong limits of random matrix spectra for sample matrices of independent elements. *The Annals of Probability*, pages 1–18, 1978.
- [57] W. Wang and J. Fan. Asymptotics of empirical eigenstructure for high dimensional spiked covariance. *Annals of statistics*, 45(3):1342, 2017.
- [58] Harold Widom. Asymptotic behavior of the eigenvalues of certain integral equations. *Transactions of the American Mathematical Society*, 109(2):278–295, 1963.
- [59] Y.Q. Yin. Limiting spectral distribution for a class of random matrices. *Journal of multivariate analysis*, 20(1):50–68, 1986.
- [60] Y.Q. Yin, Z.D. Bai, and P.R. Krishnaiah. On the limit of the largest eigenvalue of the large dimensional sample covariance matrix. *Probability theory and related fields*, 78(4):509–521, 1988.
- [61] A. Zygmund. *Trigonometric series*, volume 2. Cambridge university press, 2002.