Extreme eigenvalues of Random Matrices

Macroscopic and Microscopic Results

Phd Thesis

Michail Louvaris

Advisor: Professor Dimitris Cheliotis

National and Kapodistrian University of Athens

Department of Mathematics

Athens — 2023

Advisory Committee

- **Dimitris Cheliotis**, Professor, National and Kapodistrian University of Athens, Department of Mathematics. (supervisor)
- **Michail Loulakis**, Professor, National Technical University of Athens.
- **Nikolaos Papadatos**, Professor, National and Kapodistrian University of Athens, Department of Mathematics.

Examination Committee

- **Aristides Katavolos**, Emeritus Professor, National and Kapodistrian University of Athens, Department of Mathematics.
- **Ioannis Kontoyiannis**, Professor, University of Cambridge, Department of Pure Mathematics and Mathematical Statistics.
- **Aris Moustakas**, Associate Professor, National and Kapodistrian University of Athens, Department of Physics.
- **Konstantinos Tyros**, Associate Professor, National and Kapodistrian University of Athens, Department of Mathematics.

This thesis was supported by the Hellenic Foundation for Research and Innovation (H. F. R. I.) under the First Call for H. F. R. I. Research Projects to support Facutly members and Researchers and the procurement of high-cost research equipment grant, (Project Number: 1034).

Ευχαριστίες

Αρχικά, θα ήθελα να εκφράσω τις βαθύτατες ευχαριστίες μου στον κ. Χελιώτη για την πολύτιμη καθοδήγησή του κατά τη διάρκεια της διδακτορικής μου διατριβής. Ο χρόνος που αφιέρωσε, καθώς και η εκτενής του γνώση στην περιοχή των Πιθανοτήτων, ήταν καταλυτικοί παράγοντες για την εισαγωγή μου στον ερευνητικό τομέα των Τυχαίων Πινάκων και την επιτυχή ολοκλήρωση της διατριβής αυτής.

Επιθυμώ, επίσης, να ευχαριστήσω τα υπόλοιπα μέλη της τριμελούς συμβουλευτικής επιτροπής για την τιμή που μου έκαναν να με επιβλέπουν κατά τη διάρκεια αυτών των χρόνων, καθώς και τα μέλη της εξεταστικής επιτροπής για την κρίση τους στη διατριβή μου.

Ευχαριστώ θερμά τους κ. Δοδό, Τύρο και Γιαννόπουλο για τις κατά καιρούς συζητήσεις μας, οι οποίες αποδείχθηκαν εξαιρετικά χρήσιμες.

Επιπλέον, η διατριβή αυτή χρηματοδοτήθηκε καθ' όλη τη διάρκεια από το ΕΛΙΔΕΚ, και γι' αυτό θα ήθελα να ευχαριστήσω τα μέλη της ομάδας *Scaling Stochastic Dynamics: From Microscopic Interactions to Macroscopic Phenomena*, που μου παρείχαν αυτή την πολύτιμη ευκαιρία.

Τέλος, θεωρώ ότι η παρούσα διατριβή δεν αποτελεί μόνο αποτέλεσμα της προσωπικής μου προσπάθειας, αλλά και προϊόν της συλλογικής ζωής της ελληνικής μαθηματικής κοινότητας, η οποία με τον τρόπο που αλληλεπιδρά, «γεννά» και «ενηλικιώνει» νέους επιστήμονες. Γι' αυτό τον λόγο, θα ήθελα να ευχαριστήσω θερμά την κοινότητα αυτή, που με δέχθηκε, με διαμόρφωσε και με καθόρισε με τόσους διαφορετικούς τρόπους τα τελευταία δέκα χρόνια. Ιδιαίτερα, ένα ξεχωριστό μέρος αυτής της κοινότητας και της επιρροής της πάνω μου αποτέλεσε ο καθηγητής Δημήτρης Γατζούρας, ο οποίος δυστυχώς έφυγε πολύ νωρίς από κοντά μας. Αφιερώνω, λοιπόν, αυτή τη διδακτορική διατριβή στη μνήμη του.

Επιθυμώ, επίσης, να ευχαριστήσω την οικογένειά μου, που με στήριξε σε αυτό το ταξίδι με τον καλύτερο δυνατό τρόπο. Ιδιαίτερες ευχαριστίες οφείλω στους γονείς μου, Μανώλη και Σοφία, για τις θυσίες και την αγάπη τους όλα αυτά τα χρόνια, καθώς και στον Θάνο και την Ελίνα για τη στήριξη και την κατανόησή τους.

Θα ήθελα να ευχαριστήσω θερμά τους φίλους μου, που με συντρόφευσαν σε αυτή την πορεία. Ιδιαίτερες ευχαριστίες απευθύνω στη Δήμητρα για ό,τι έκανε για μένα αυτά τα χρόνια· η παρουσία και η στήριξή της ήταν ανεκτίμητες.

Θερμές ευχαριστίες οφείλω επίσης στον Απόλλωνα και τον Σταμάτη, των οποίων η παρουσία στο γραφείο 116 και τα εβδομαδιαία μας « meeting» κατά το τελευταίο έτος του διδακτορικού μου, έκαναν την εμπειρία αυτή πιο εύκολη και απολαυστική.

Τέλος, εκφράζω τις ειλικρινείς ευχαριστίες μου στους φίλους μου, Χρήστο, Μήτσο και Αλέκο, για τη συμπαράσταση, την αλληλεγγύη και τη συντροφικότητα που αναπτύχθηκε μεταξύ μας κατά τη διάρκεια αυτών των χρόνων.

Abstract

This thesis consists of two parts and examines the asymptotic behavior of the extreme eigenvalues of some random matrix models. Specifically:

- In the last decades there has been a growing interest on the asymptotic behavior of the smallest singular value of Random Matrix models, see [1],[2], [3] and [4]. A common factor in all of these cases is that the entries of the Random Matrix model under examination have finite variance. So in the first part of this thesis we examine the asymptotic distribution of the smallest singular value of a Random Matrix model with heavy tailed entries, the Lévy non-symmetric Random Matrices. In this model the entries of the matrix are i.i.d. and follow an $\alpha-$stable distribution. We prove that for almost all $\alpha \in (0, 2)$ universality, i.e., the same asymptotic distribution as in the Gaussian case, holds for the least singular value. As a byproduct of our proof, we also prove the complete delocalization of the singular vectors of this model at small energies. The methods are based on the modern techniques, whose heart lie in the three step strategy, an important strategy developed in the last decade in the Random Matrix Theory literature, see [5], [6] and [7]. In order to obtain the universality for the least singular value, we also prove a version of an isotropic local law for a general class of matrices.

- After the seminal works of Wigner in [8] and Marchenko-Pastur in [9], where the limit of the empirical spectral distribution of some class of random matrices has been established, a natural question that emerged is what happens with the extreme eigenvalues of that matrices. The convergence of the operator norm of these matrices to the rightmost element of the limiting spectrum was first proven in [10], under the necessary and sufficient condition that the entries of the matrix have finite 4-th moment. In these "classic" results the entries of the matrices are i.i.d. In the last decade there has been a growing interest on Random Matrix models, whose entries are independent but not necessarily identically distributed and in particular with different variances, see for example [11], [11], [12], [13] and [14]. In particular, in [13] the convergence of the empirical spectral distribution of several classes of matrices to a limiting probability measure is proved. This convergence is proven to heavily

depend only on the properties of the variance profile of these matrices, i.e. the matrix with entries the variances of the entries of the initial matrix. For some models, for which the result of [13] holds, the convergence of the operator norm of the matrix to the rightmost element of the limiting spectrum is also proven in [14] and [12] but under the assumption that the entries of the matrix model have all their moments finite. So in the second part of this thesis we prove that for matrices with a general variance profile and with finite $4 + \epsilon$ moment the convergence of the operator norm holds. The methods we use are based on the comparison of large moments of the matrix model with the rightmost element of the spectrum of the limiting distribution and are based on methods developed in [13], [15], [16] and [10]. Our approach covers the cases of random symmetric or non-symmetric matrices whose variance profile is given by a step or a continuous function, random band matrices whose bandwidth is proportional to their dimension, random Gram triangular matrices and more.

Περίληψη

Η παρούσα διατριβή αποτελείται από δύο μέρη και εξετάζει την ασυμπτωτική συμπεριφορά των ακραίων ιδιοτιμών ορισμένων μοντέλων τυχαίων πινάκων. Συγκεκριμένα:

- Τις τελευταίες δεκαετίες υπάρχει αυξανόμενο ενδιαφέρον για την ασυμπτωτική συμπεριφορά της μικρότερης ιδιάζουσας τιμής μοντέλων τυχαίων πινάκων, βλέπε [1],[2], [3] και [4]. Ένας κοινός παράγοντας στις προαναφερθείσες περιπτώσεις είναι ότι τα στοιχεία των υπό εξέταση μοντέλων τυχαίων πινάκων έχουν πεπερασμένη διασπορά. Έτσι, στο πρώτο μέρος αυτής της διατριβής εξετάζουμε την ασυμπτωτική κατανομή της μικρότερης ιδιάζουσας τιμής ενός μοντέλου τυχαίων πινάκων με στοιχεία με βαριές ουρές, τους Lévy μη συμμετρικούς τυχαίους πίνακες. Σε αυτό το μοντέλο τα στοιχεία του πίνακα είναι ανεξάρτητα και ισόνομα και ακολουθούν μια $\alpha-$ευσταθή κατανομή. Αποδεικνύουμε ότι σχεδόν για όλα τα $\alpha \in (0, 2)$ ότι η καθολικότητα, δηλαδή η ίδια ασυμπτωτική συμπεριφορά όπως στην περίπτωση των πινάκων με στοιχεία που ακολουθούν την τυποποιημένη κανονική κατανομή, ισχύει για την ελάχιστη ιδιάζουσα τιμή των Lévy μη συμμετρικών τυχαίων πινάκων. Ως υποπροϊόν της απόδειξής μας, αποδεικνύουμε επίσης την πλήρη μετατόπιση των ιδιαζόντων διανυσμάτων αυτού του μοντέλου για μικρές ενέργειες. Οι μέθοδοι βασίζονται σε σύγχρονες τεχνικές, οι οποίες βασίζονται στη στρατηγική των τριών βημάτων, μια σημαντική στρατηγική που αναπτύχθηκε την τελευταία δεκαετία στη βιβλιογραφία των τυχαίων πινάκων, βλ. [5], [6] και [7]. Προκειμένου να αποδείξουμε την καθολικότητα για την ελάχιστη ιδιάζουσα τιμή, αποδεικνύουμε επίσης μια εκδοχή ενός ισοτροπικού τοπικού νόμου για μια γενική κλάση τυχαίων πινάκων.

- Μετά τα θεμελιώδη αποτελέσματα του Wigner στο [8] και των Marchenko-Pastur στο [9], όπου καθορίζεται το όριο της εμπειρικής φασματικής κατανομής κάποιας κλάσης τυχαίων πινάκων, ένα φυσικό ερώτημα που προέκυψε είναι τί συμβαίνει με τις ακραίες ιδιοτιμές αυτών των πινάκων. Η σύγκλιση της νόρμας τελεστή αυτών των πινάκων στο δεξιότερο στοιχείο του στηρίγματος του οριακού μέτρου αποδείχθηκε για πρώτη φορά στο [10], υπό την ίκανή και αναγκαία προϋπόθεση ότι τα στοιχεία του πίνακα έχουν πεπερασμένη 4η ροπή. Σε αυτά τα «κλασικά» αποτελέσματα, τα στοιχεία των πινάκων είναι ανεξάρτητες και ισόνομες τυχαίες μεταβλητές. Την τελευταία δεκαετία υπήρξε ένα αυξανόμενο ενδιαφέρον για το μοντέλα τυχαίων πινάκων, των οποίων τα στοιχεία είναι ανεξάρτητες αλλά όχι απαραίτητα ισόνομες τυχαίες μεταβλητές και ειδικότερα έχουν διαφορετικές διασπορές, βλ. για παράδειγμα τα [11], [11], [12], [13] και [14]. Συγκεκριμένα, στο [13] αποδεικνύεται η σύγκλιση της εμπειρικής φασματικής κατανομής μια μεγάλης κλάσης τυχαίων πινάκων σε ένα μέτρο πιθανότητας. Αυτή η σύγκλιση αποδεικνύεται ότι εξαρτάται σε μεγάλο βαθμό μόνο από τις ιδιότητες του προφίλ των διασπορών αυτών των πινάκων, δηλαδή τον πίνακα που έχει για στοιχεία τις διασπορές των στοιχείων του αρχικού πίνακα. Για ορισμένα μοντέλα για τα οποία ισχύουν τα αποτελέσματα του άρθρου [13], η σύγκλιση της νόρμας τελεστή του πίνακα στο δεξιότερο στοιχείο του στηρίγματος του οριακού μέτρου αποδεικνύεται επίσης στα [14] και [12], αλλά υπό την προϋπόθεση ότι οι τα στοιχεία του υπό εξέταση μοντέλου τυχαίων πινάκων έχουν όλες τις ροπές τους πεπερασμένες. Έτσι, στο δεύτερο μέρος αυτής της διατριβής αποδεικνύουμε ότι για πίνακες με γενικό προφίλ διασπορών και με πεπερασμένη την $4 + \epsilon$ ροπή, ισχύει η σύγκλιση στο δεξιότερο μέλος του στηρίγματος του οριακού μέτρου. Οι μέθοδοι που χρησιμοποιούμε βασίζονται στη σύγκριση μεγάλων ροπών του μοντέλου τυχαίων πινάκων με το δεξιότερο στοιχείο του στηρίγματος της οριακής κατανομής και σε μεθόδους που αναπτύχθηκαν στα [13], [15], [16] και [10]. Η προσέγγισή μας καλύπτει τις περιπτώσεις τυχαίων συμμετρικών ή μη πινάκων των οποίων το προφίλ των διασπορών τους δίνεται από κατά τμήματα σταθερή ή συνεχή συνάρτηση, τυχαίων band πινάκων των οποίων το εύρος είναι ανάλογο της διάστασή τους, τυχαίων Gram τριγωνικών πινάκων και άλλα.

Contents

Chapter 1

Introduction

Random matrix theory (RMT) is a branch of mathematics that deals with the study of matrices whose entries are random variables. It provides powerful tools and techniques for understanding the statistical properties of complex systems arising in various fields, including physics, computer science, and statistics. In recent years, there has been a growing interest in the asymptotic behavior of random matrices as their dimensions tend to infinity. Such asymptotic results have proven to be invaluable in analyzing large-scale systems and have found applications in diverse areas such as wireless communications, finance, and quantum information theory.

The study of asymptotic behavior in random matrix theory is motivated by the need to understand the limiting behavior of complex systems involving a large number of random variables. Traditional methods of analysis often fail in such scenarios due to the complexity and interdependence of the variables. Random matrix theory offers a powerful framework to tackle these challenges by providing tractable mathematical models and insightful results. The limiting behavior is often referred to as the "asymptotic regime" and is of great interest.

The asymptotic results in random matrix theory typically involve the analysis of certain limiting distributions or convergence properties as the matrix size tends to infinity. These results provide valuable insights into the behavior of random matrices and can be used to study a wide range of phenomena. For example, in the field of wireless communications, asymptotic results on the eigenvalue distribution of random matrices can be used to analyze the performance of multiple-input multiple-output (MIMO) systems, which are widely used in modern wireless communication systems.

Another important aspect of asymptotic results in random matrix theory is the universality phenomenon. Universality refers to the remarkable observation that the limiting behavior of random matrices often exhibits universal properties that are independent of the specific distribution of the matrix entries. This universality property allows us to make general statements about the behavior of random matrices without detailed knowledge of the underlying distribution. It is a powerful concept that simplifies the analysis and makes random matrix theory applicable in a wide range of contexts.

In this thesis, we aim to explore and contribute to the growing body of literature on asymptotic results in random matrix theory and hope to enhance our understanding of complex systems and contribute to the advancement of several disciplines.

Specifically this thesis is based on the following papers

1. *M.Louvaris*, **Universality of the least singular value and singular vector delocalisation for Lévy non-symmetric random matrices**, Annales de l'Institut Henri Poincaré, Probabilités et Statistiques (accepted), AIHP1413

2. *D.Cheliotis, M.Louvaris*, **The limit of the operator norm for random matrices with a variance profile**, available in arxiv.

1.1 Universality of the least singular value and singular vector delocalisation for Lévy non-symmetric random matrices

1.1.1 General description of the problem

Consider the following problem:

Let $\{Y_N\}_N$ be a sequence of $N \times N$ matrices with i.i.d. entries, $Y_N := (Y_{i,j})_{1 \le i,j \le N}$. Then is there some normalization sequence c_N such that

$$c_N \lambda_{\min}(Y_N^T Y_N) \Rightarrow_{N\to\infty} Z \tag{1.1.1}$$

for some non-degenerate random variable Z?

In (1.1.1) $\Rightarrow_{N\to\infty}$ denotes the convergence in distribution as $N \to \infty$.

For several random matrix models we have:

- (1.1.1) holds, when Y_N has entries all following the $N(0, 1)$ distribution. Firstly proven in [1], by using the properties of the Gaussian random variables.

- (1.1.1) holds, when Y_N has i.i.d. entries with finite moments up to some C (~ 100). Proven in [2], by comparing to Gaussian Matrices and by using a version of the CLT.

- Recently: (1.1.1) holds for sparser random matrices [3] and for the sum of random matrices in [4].

In each of the cases mentioned, it is true that for some $c > 0$

$$\mathbb{P}(|Y_{1,1}| \ge t) \le \frac{c}{t^2} \tag{1.1.2}$$

What if $\mathbb{P}(|Y_{1,1}| \ge t) \sim t^{-\alpha}$, for some $\alpha \in (0, 2)$? That is, the entries are heavy tailed.

So in the first part of this thesis we prove (1.1.1), when $Y_{1,1}$ follows an α-stable law, for almost all $\alpha \in (0, 2)$. In particular $Y_{1,1}$ has heavy tailed entries. This is done in Chapter 2 and specifically in Theorem 2.1.2.

1.1.2 Background

The asymptotic behavior of the spectrum of random matrices has been a crucial topic of studies since Wigner's semicircle law, first proven in [8]. The study of the asymptotic spectral behavior of Wishart matrices was the next important result, firstly investigated in [9], although Wishart matrices actually preceded Wigner matrices.

The Wishart matrices, and more generally the covariance matrices, play a significant role in various scientific fields. See for example [17] and [18] for applications in statistics, [19] for application in economics and [20] for application in population genetics. Several spectral properties of these matrices have been investigated. We focus on the case that the entries of the matrix are identically distributed, independent random variables (i.i.d.). For those matrices some significant results concern the limit of the largest eigenvalue, the asymptotic behavior of the correlation functions and the asymptotic bulk and edge behavior. For example, see [21] or the lecture notes concerning the singular values [22]. These results are proven for matrices whose entries have finite variance.

Besides those results, an important problem in random matrix theory is the asymptotic behavior of the least eigenvalue of covariance matrices, when the matrices' dimensions are equal. Note that the inverse of the least singular value of a matrix is equal to the operator norm of its inverse, so an estimate of the least singular value gives control to the probability that the inverse has large norm and also gives control to its condition number. To name an illustrative application, this estimate of the least singular value for various random matrix models, plays an important role in the analysis of the performance of algorithms, see [23]. In the case that the entries of the matrices are normally distributed, the limiting distribution has been described in Theorem 4.2 of [1], by directly computing the density of the smallest singular value multiplied by N. In the general i.i.d. case, under the assumption of finite moments of sufficiently large order, the least singular value is proven to tend to the same law as the least singular value of a Gaussian random matrix, in Theorem 1.3 of [2]. This phenomenon, the same asymptotic distribution for the least singular value of a matrix as in the Gaussian case, will be called universality of the least singular value for the matrix. Lastly, in the most recent papers [3] and [4] the authors proved that universality of the least singular values holds for more general classes of matrices.

The above results have been focused on the finite variance cases. In the case of infinite second moment, and more specifically in the case of stable entries, there are not so many results concerning the behavior of the spectrum of covariance matrices. There are some results, mostly concerning the limit of the E.S.D. of such matrices ([24],[25]) and the limit of the largest eigenvalues [26] and [27]. Moreover there are also some generalizations, which concern the limit of the largest eigenvalue of heavy tailed autocovariance matrices in [28] and covariance matrices with heavy tailed m-dependent entries in [29]. Despite that, progress has been made concerning the symmetric matrices with heavy tailed entries. In

[30], the authors found the limit of the empirical spectral distribution of such matrices. Next, in [31] and [32] the authors proved some version of local law and examined the localization and delocalization of the eigenvectors in each of these cases. Moreover, in [25] and [33], the authors gave a better understanding of the limiting distribution of the empirical spectral distribution by proving the convergence of resolvent of the matrix to the root of a Poisson weighted infinite tree in some operator space. Recently, in [5] and [34], the authors showed complete delocalization of the eigenvectors whose eigenvalues belong in some interval around 0, GOE statistics for the correlation function and described the precise limit of the eigenvectors respectively.[1]

In this paper we prove universality for the least singular value of random matrices with i.i.d. $\alpha-$stable entries. The methods we use also imply the complete singular vector delocalization for such matrices at small energies. We prove these results using a version of the three step strategy, a strategy developed in the last decade, which is suitable in order to obtain universality results for random matrix models, see [7].

The basic inspiration for this paper is Theorem 2.5 in [5], which proves universality of the correlation functions for symmetric Lévy random matrices at small energies. Both the intermediate local law, Theorem 2.3.14, and the theorem concerning the comparison of the entries of the resolvent, Theorem 2.6.4, are similar to Theorem 3.5 and Theorem 3.15 of [5] respectively, adjusted to our set of matrices. For the intermediate local law we also use a lot of results from [31] and [32].

Results and methods from [6] and [3] had significant influence to this paper as well. In particular the isotropic local law in Sections 2.5 is an analogue of Theorem 2.1 in [6], proven for a different class of matrices. Moreover universality for the least singular value of random matrices after perturbing them by a Brownian motion Matrix can be found in Theorem 3.2 of [3]. So several results from Sections 2.4 and 2.6 are based or influenced by results of [3].

1.2 The limit of the operator norm for random matrices with a variance profile

1.2.1 General description of the problem

Given a sequence of $N \times N$ random matrices A_N, set

$$\mu_{A_N} = \frac{1}{N} \sum_{i=1}^{N} \delta\left(\lambda_i\left(\frac{A_N}{\sqrt{N}}\right)\right) \tag{1.2.1}$$

[1]GOE denotes the Gaussian Orthogonal Ensemble, i.e., symmetric matrix with independent entries (up to symmetry) where the non-diagonal entries have law $N(0,1)$ and the diagonal $N(0,2)$.

to be the empirical spectral distribution of A_N. In (1.2.1) δ denotes the Dirac measure and for any $N \times N$ matrix B, $\{\lambda_i(B)\}_{i\in[N]}$ denotes the set of the eigenvalues of B.

In the case that A_N is symmetric and its entries are i.i.d. (up to symmetry) with mean 0 and variance 1, the almost sure convergence of μ_{A_N} to the semicircle law was first established in the seminal work [8] by Wigner. The support of the semicircle law is the set $[-2, 2]$. Next in [10], the authors also proved that

$$\lim_{N\to\infty} \frac{\max_{i\in[N]} |\lambda_i(A_N)|}{\sqrt{N}} = 2 \quad \text{a.s.} \tag{1.2.2}$$

under the necessary and sufficient condition that the entries of A_N have finite 4-th moment.

Suppose that A_N is symmetric and has independent entries with mean 0, but not necessarily identically distributed. Some sufficient conditions for the almost sure convergence of μ_{A_N} to a non-trivial probability distribution were given in [13]. The limiting distribution is proven to be supported on some set $[-\mu_\infty, \mu_\infty]$ for some $\mu_\infty > 0$.

So in the second part of this thesis we give some sufficient conditions so that

$$\lim_{N\to\infty} \frac{\max_{i\in[N]} |\lambda_i(A_N)|}{\sqrt{N}} = \mu_\infty \quad \text{a.s.} \tag{1.2.3}$$

for a general class of matrices A_N, for which the results of [13] hold and under the extra assumption that the entries of A_N have finite the $4 + \epsilon$ moment, for any small $\epsilon > 0$. This is done in Chapter 3. The main results of this Chapter are Theorems 3.1.8, 3.1.14 and 3.1.10.

Our approach covers several well-known Random Matrix Models and is a generalization of previous results such as Theorem 1.3 of [12] and Corollary 2.3 of [14].

1.2.2 Background

The problem of understanding the operator norm of a large random matrix with independent entries is multidisciplinary, occupying mathematicians, statisticians, physicists. On the mathematical side, tools from classical probability, geometric analysis, combinatorics, free probability and more have been used. The problem dates back to 1981, where in [35] the convergence of the largest eigenvalue of renormalized Wigner matrices (symmetric, i.i.d. entries) to the edge of the limiting distribution was established when the entries of the matrix are bounded. Next, in [10], the authors gave necessary and sufficient conditions for the entries of a Wigner matrix to converge. The crucial condition was that the entries should have finite 4-th moment. Similar bounds have been given to non-symmetric matrices with i.i.d. entries. Then, the difference of the largest eigenvalue and its limit, after re-normalization, was proven to converge to the Tracy-Widow law in [36]. Later, universality results were established for sparse random matrix models, for example in [37] for random graphs and in [38] for random banded matrices. Moreover, sharp non-asymptotic

results for a general class of matrices were established in [39] and in [40].
All the models mentioned above can be considered as random matrices with general variance profile, i.e., random matrices whose entries' variances can depend on the dimension of the matrix and the location of the element in the matrix. These models have also drawn a lot of attention lately, see for example [11], [41], where non-Hermitian models are considered. More specifically, assume that $A_N = (a_{i,j}^{(N)}), N \in \mathbb{N}^+$, is a sequence of symmetric random matrices, with $a_{i,j}^{(N)}$ real valued having mean zero and variance $s_{i,j}^{(N)}$ bounded by a fixed number, say 1. Classically, the first question is whether the empirical spectral distribution of an appropriate normalization of A_N (e.g., $A_N/\sqrt{N}$) converges to a nontrivial probability measure, as in Wigner's theorem. Nothing guarantees that, and one can construct examples where the sequence of the empirical spectral distributions does not converge. The work [13], using the notion of graphons, gave conditions on the variance profile $s_{i,j}^{(N)}, i, j \in [N], N \in \mathbb{N}^+$ so that convergence takes place.

The next, natural, question concerns the convergence of the largest eigenvalue to the largest element of the support of the limiting distribution. Again, this in not automatic but requires additional assumptions. It was established in the recent works [42], [14], [12], [43] (whose focus however is not this question) for some class of random matrices with a general variance profile under the assumption that the entries of the matrices have finite all moments (the first two works assume that each $a_{i,j}^{(N)}$ is sharp sub-Gaussian, the last two assume that for each $k \in \mathbb{N}^+$ there is a constant bounding the $2k$ moment of each $a_{i,j}^{(N)}$). In this paper, we generalize these results, i.e., we establish the convergence of the largest eigenvalue of general variance profile random matrices to the largest element of the support of the limiting empirical spectral distribution under general assumptions for the variance profile of the matrices. Regarding finiteness of moments, we assume only that $\sup_{N\in\mathbb{N}^+, i,j\in[N]} \mathbf{E}|a_{i,j}^{(N)}|^4 < \infty$.

Chapter 2

Universality of the least singular value and singular vector delocalisation for Lévy non-symmetric random matrices

2.1 Main results

Fix a parameter $\alpha \in (0,2)$. A random variable Z is called $(0,\sigma)$ α-stable law if

$$\mathbf{E}(e^{itZ}) = \exp(-\sigma^{\alpha}|t|^{\alpha}), \text{ for all } t \in \mathbb{R}. \tag{2.1.1}$$

Definition 2.1.1. Set

$$\sigma := \left(\frac{\pi}{2\sin(\frac{\pi\alpha}{2})\Gamma(\alpha)}\right)^{1/\alpha} > 0, \tag{2.1.2}$$

and let J be a symmetric random variable with finite variance and let Z be a $(0,\sigma)$ α-stable random variable, independent from J. Then, define the matrix $D_N(\alpha) = \{d_{i,j}\}_{1\leq i,j\leq N}$ to be random matrix with i.i.d. entries, all having the same law as $N^{-1/\alpha}(J+Z)$. In what follows, we may omit explicitly indicating the dependence of the matrices D_N on the parameters α and N, and use the notation D.

Lastly, fix parameters C_1, C_2 such that

$$\frac{C_1}{Nt^{\alpha}+1} \leq \mathbb{P}\Big(|d_{i,j}| \geq t\Big) \leq \frac{C_2}{Nt^{\alpha}+1}. \tag{2.1.3}$$

Such parameters exist due to the tail properties of the stable distribution. See [44], Property 1.2.8.

The parameter σ is chosen in (2.1.2) like so, in order to keep our notation consistent with previous works such as [5],[34],[32] and [31]. This parameter can be altered by a rescaling. Moreover, denote ρ_{sc} the probability density function of the semicircle law, i.e.,

$$\rho_{sc}(x) = \mathbf{1}\{|x| \le 2\} \frac{1}{2\pi} \sqrt{4 - x^2}.$$

Furthermore, set

$$\xi := \frac{\rho_a(0)}{\rho_{sc}(0)}, \tag{2.1.4}$$

where ρ_a is the density of the limiting distribution of the empirical measure of the singular values of D and their negative ones and is described in Proposition 2.2.15.

In what follows we will use the standard Big O notation. Specifically given two functions f, g, we will say $f = O(g)$ if and only if there exists a constant $C > 0$ independent of any other parameter such that

$$\limsup_{x\to\infty} \left|\frac{f(x)}{g(x)}\right| = C < \infty, \tag{2.1.5}$$

where the constant $C > 0$ will be independent of any other parameter. If the constant C depends on some parameter(s) c defined earlier, we will write $f = O_c(g)$. Moreover if the constant $C = 0$ then we will write $f = o(g)$.

Our main result shows that the least singular values of D_N are universal as N tends to infinity. The analogous result for matrices with finite variance entries was proven in [2]. We also prove that the left and right singular vectors of D_N are completely delocalized for small energies, in the following sense.

Theorem 2.1.2. *There exists a countable set $\mathcal{A}$, subset of $(0, 2)$, with no accumulation points in (0,2) such that the following holds. Let $\{D_N(\alpha)\}_{N\in\mathbb{N}}$ be sequences of matrices, where $D_N(\alpha) \in \mathbb{R}^{N\times N}$ with i.i.d. entries all following $N^{-1/\alpha}(Z + J)$, where Z, J as in Definition 2.1.1. Then for every $\alpha \in (0, 2) \setminus \mathcal{A}$:*

1. *Let $s_1(D_N(\alpha))$ denote the least singular value of $D_N(\alpha)$. Then, there exists $c > 0$ such that for all $r \ge 0$*

$$\mathbb{P}\Big(N\xi s_1(D_N(\alpha)) \le r\Big) = 1 - \exp\left(-\frac{r^2}{2} - r\right) + O_r(N^{-c}). \tag{2.1.6}$$

2. *For each $\delta > 0$ and $D > 0$ there exist constants $C = C(\alpha, \delta, D) > 0$ and $c = c(\alpha)$ such that:*

$$\mathbb{P}\Big(\max\Big\{|||u|||_\infty : u \in \mathcal{B}_N\Big\} > N^{\delta-\frac{1}{2}}\Big) \le CN^{-D}. \tag{2.1.7}$$

where $\mathcal{B}_N$ is the set of eigenvectors of $D_N D_N^T$ or $D_N^T D_N$, normalized with the Euclidean norm, whose corresponding eigenvalues belong to the set $[-c, c]$.

The proof of Theorem 2.1.2 can be found in Subsection 2.6.3.

Remark 2.1.3. The set $\mathcal{A}$ for which Theorem 2.1.2 cannot be applied is conjectured to be empty. Its presence is due to some α−dependent fixed point equations in [32], which we use and can be inverted only if $\alpha \notin \mathcal{A}$.

Moreover, we can generalize the proof of Theorem 2.1.2 to the joint distribution of the bottom k singular values in the following sense.

Theorem 2.1.4. *Fix a positive integer k. Let $\mathcal{A} \subseteq (0,2)$ be the countable set of Theorem 2.1.2. Then define, as in Definition 2.1.1, $\{D_N\}_{N\in\mathbb{N}}$ with i.i.d. entries all following $N^{-1/\alpha}(Z+J)$, where Z is $(0,\sigma)$ $\alpha-$stable for $\alpha \in (0,2)\setminus\mathcal{A}$. Also let $\{L_N\}_N$ be a sequence of $N\times N$ i.i.d. matrices, with entries following the same law as a centered normal random variable with variance $\frac{1}{N}$. Also for any matrix A define*

$$\Lambda_k(A) := (Ns_1(A), \cdots, Ns_k(A)),$$

where $\{s_i(A)\}_{i\in[N]}$ are the singular values of A arranged in increasing order. Also denote $1_k = (1,\cdots,1)$ and for all $E\in\mathbb{R}^k$

$$\Omega(E) := \{x\in\mathbb{R}^k : x_i \le E_i \text{ for all } i\in[k]\}.$$

Then there exists $c>0$ such that for all $E\in\mathbb{R}^k$

$$\begin{aligned}&\mathbb{P}\Big(\Lambda_k(L_N)\in\Omega(E-N^{-c}1_k)\Big) - O_E(N^{-c}) \le \mathbb{P}\Big(\Lambda_K(\xi D_N)\in\Omega(E)\Big) \le \\ &\mathbb{P}\Big(\Lambda_K(L_N)\in\Omega(E+N^{-c}1_k)\Big) + O_E(N^{-c}).\end{aligned} \tag{2.1.8}$$

The proof of Theorem 2.1.4 is similar to that of Theorem 2.1.2 and therefore is omitted. Note that the universal limiting distribution of $\Lambda_k(L_N)$ is explicitly given in [2].
Moreover, by the way that we will prove Theorem 2.1.2, we can prove a similar result for the gap probability in the symmetric case. The proof of the following corollary will again be omitted due to its similarity to the proof of Theorem 2.1.2.

Corollary 2.1.5. *Let M_N be an $N\times N$ symmetric matrix with i.i.d. entries (with respect to symmetry) and let all entries follow the same law as $N^{-1/\alpha}(Z+J)$, where Z,J are defined in Definition 2.1.1 for $\alpha\in(0,2)\setminus\mathcal{A}$. Here $\mathcal{A}$ is the set of Theorem 2.1.2. Also let W_N be a GOE matrix ($N\times N$ symmetric, with i.i.d. centered Gaussian entries, with variance N^{-1}). Arrange the eigenvalues of M_N and W_N in increasing order. Then there exists $\delta>0$ such that for any $r>0$,*

$$\left|\mathbb{P}\left(\#\left\{i\in[N] : N\lambda_i(M_N)\in\left(-\frac{r}{2},\frac{r}{2}\right)\right\}=0\right)-\mathbb{P}\left(\#\left\{i\in[N] : N\lambda_i(W_N)\in\left(-\frac{r}{2},\frac{r}{2}\right)\right\}=0\right)\right|\le O_r(N^{-\delta}). \tag{2.1.9}$$

For the Gaussian case, the limiting distribution of the gap probability is given in Theorem 3.12 of [45].

Remark 2.1.6. Note that by Theorem 2.1.2, the least singular value of a random matrix with i.i.d. entries, all following an $\alpha-$stable distribution, are of order $O(N^{\frac{1}{\alpha}-1})$ for $\alpha\in(0,2)\setminus\mathcal{A}$. So for $\alpha\in(0,1)\cap\mathcal{A}^c$ the least singular value, without normalization, tends to ∞, which is different from the finite variance case.

2.2 Preliminaries and sketch of the proof

2.2.1 Preliminaries

In this subsection we present some necessary definitions and lemmas.

Firstly fix parameters a, b, ρ, ν such that

$$a \in (0,2), \quad \nu = \frac{1}{a} - b > 0, \quad 0 < \rho < \nu, \quad \frac{1}{4-a} < \nu < \frac{1}{4-2a}, \quad a\rho < (2-a)\nu. \tag{2.2.1}$$

Note that given $a \in (0,2)$, such parameters will exists. Moreover $\nu > 0$ is the level on which we will truncate the matrix D_N in (2.2.2). This truncation is crucial to our analysis as is explained later in Subsection 2.2.2. The rest of the restrictions for the parameters in (2.2.1), will be explained later in the choice of ϵ_0 in (2.6.19), in the proof of Theorem 2.6.4.

Next we give some preliminaries definitions and lemmas.

Definition 2.2.1. For each $a \in (0,\infty)$ and $u \in \mathbb{C}^N$ we will use the notation

$$\|u\|_a = \left(\sum_{i=1}^{N} |u_i|^a\right)^{1/a}.$$

Moreover if $N = 1$ and $a = 2$, we will use the notation $|u|$ for the Euclidean norm.

Definition 2.2.2. Fix an $N \times N$ matrix Y. Then the empirical spectral distribution of Y is the measure

$$\mu_Y := \frac{1}{N}\sum_{i=1}^{N} \delta_{\lambda_i(Y)},$$

where δ_x is the Dirac measure for $x \in \mathbb{R}$ and $\{\lambda_i(Y)\}_{i\in[N]}$ are the eigenvalues of Y. We will also use the notation $\lambda_{\max}(Y)$ for the largest eigenvalue of Y.

Definition 2.2.3. Let M be an $N \times N$ real matrix. Then the $2N \times 2N$ matrix

$$\begin{bmatrix} 0 & M^T \\ M & 0 \end{bmatrix}$$

is called the symmetrization of M.

Definition 2.2.4. Let H_N be the symmetrization of D_N, i.e.,

$$H_N = \begin{bmatrix} 0 & D_N^T \\ D_N & 0 \end{bmatrix}.$$

Then define the matrix $X_N = \{x_{i,j}\}_{1\le i,j\le 2N}$ such that

$$x_{i,j} := h_{i,j}\mathbf{1}\left\{N^{1/a}|h_{i,j}| \ge N^b\right\}. \tag{2.2.2}$$

The elements of X_N (in the non-diagonal blocks) are called the b−removals of a deformed $(0, \sigma)$ α-stable law. We also define the matrices $A_N := H_N - X_N$, the matrix E_N whose symmetrization is A_N and the matrix K_N whose symmetrization is X_N, i.e.,

$$X_N = \begin{bmatrix} 0 & K_N^T \\ K_N & 0 \end{bmatrix}, \quad A_N = \begin{bmatrix} 0 & E_N^T \\ E_N & 0 \end{bmatrix}$$

Furthermore, define the matrix L_N to be an $N \times N$ matrix with i.i.d. entries all following the law of a normal, centered random variable with variance $\frac{1}{N}$, and its symmetrization W_N. In what follows, we may omit the dependence of the matrices defined here on N, for notational convenience.

Remark 2.2.5. Note that the eigenvalues of H are exactly the singular values of D and their respective negative ones since

$$\det(\hbar \cdot \mathbb{I}_{2N} - H) = \det(\hbar^2 \cdot \mathbb{I}_N - D^T D).$$

Moreover, note that if we prove delocalization for the eigenvectors of H in the sense of the second part of Theorem 2.1.2, then we will have an understanding over the delocalization of the left and right singular vectors of D, because of the following remark.

Remark 2.2.6. If J_1, J_2 are the matrices with columns the normalized left and right singular vectors of D, which by the singular value decomposition gives us that $J_1 D J_2 = \text{diag}(s_1, s_2 \cdots, s_N)$, then one can compute that the matrix

$$\frac{1}{\sqrt{2}} \begin{bmatrix} J_2^* & J_1 \\ J_2^* & -J_1 \end{bmatrix},$$

has columns the normalized eigenvectors of H.

So in what follows, we will focus on proving delocalization for the eigenvectors and universality of the least positive eigenvalue for H.
We will use the notation $\text{Im}(z)$ for the imaginary part of any $z \in \mathbb{C}$ and $\mathbb{C}^+ := \{z \in \mathbb{C} : \text{Im}(z) > 0\}$.
Furthermore we need the following definitions.

Definition 2.2.7. Let M be an $N \times N$ matrix. The matrix $Y = (M - zI)^{-1}$ for $z \in \mathbb{C}^+$ is called the resolvent of M at z.

Definition 2.2.8. (Stieltjes transform) Let M be an $N \times N$ matrix and let μ_M be its empirical spectral distribution. Then for each $z \in \mathbb{C}^+$, we define its Stieltjes transform as the normalized trace of its resolvent, i.e.,

$$m_M^N(z) := \int \frac{1}{x - z} d\mu_M(x) = \frac{1}{N} \text{tr}(M - z\mathbb{I})^{-1}.$$

In what follows, we might omit the dependence on the dimension of the Stieltjes transform or on the matrix, when it is clear to which matrix we refer.

Definition 2.2.9. In what follows we will use the following notation.

$$t := N\,\mathrm{Var}(E_{1,1}). \tag{2.2.3}$$

Moreover, in Corollary 2.2.11 we prove that $t \to 0$ as $N \to \infty$.

In the next Lemma we give an estimate for the entries of A.

Lemma 2.2.10 ([5], Lemma 4.1)**.** *Let $R \geq N^{-1/\alpha}$ and $p > \alpha$. Then there exist a small constant $c = c(\alpha, p, C_1)$ and a large constant $C = C(\alpha, p, C_2)$ such that*

$$cN^{-1}R^{p-\alpha} \leq \mathbf{E}|D_{1,1}|^p \mathbf{1}\{|D_{1,1}| \leq R\} \leq CN^{-1}R^{p-\alpha}.$$

Here $D_{1,1}$ is the (1,1)-entry of D_N. Here C_1, C_2 are the parameters from (2.1.3).

A direct application of the previous result for $R = N^{-\nu}$ and $p = 2$ implies the following.

Corollary 2.2.11. *The entries of E_N satisfy the following*

$$cN^{\nu(\alpha-2)} \leq N\,\mathrm{Var}(E_{1,1}) \leq CN^{\nu(\alpha-2)}.$$

Remark 2.2.12. Note that the convergence of the E.S.D. of a sequence of random matrices, implies that the typical scale of an eigenvalue is $\frac{1}{N}$ (at least in the bulk of the spectrum) of the limiting distribution of the E.S.D.

Definition 2.2.13. Let $F(u)$ be a family of events indexed by some parameter(s) u. We will say that $F(u)$ holds with overwhelming probability, if for any $D > 0$ there exists an $N(D, u)$ such that for all $N \geq N(D, u)$

$$\mathbb{P}(F(u)) \geq 1 - N^{-D}.$$

uniformly in u.

Next we present a measure, for which in Theorem 2.3.14 we will prove that it is the limiting distribution of the E.S.D. of X_N.

Definition 2.2.14. Let M_N be a sequence of symmetric $N \times N$ matrices with i.i.d. entries (up to symmetry) and for each $N \in \mathbb{N}$ let all the entries follow the same law $N^{-1/\alpha}(Z + J)$, where Z, J are defined in Definition 2.1.1. In what follows for any $z \in \mathbb{C}^+$, we will use the notation

$$m_\alpha(z) := \frac{1}{N} \lim_{N\to\infty} \mathrm{tr}(M_N - z\mathbb{I})^{-1}. \tag{2.2.4}$$

So m_α is the Stieltjes transform of the limiting distribution of the E.S.D. of the sequence of matrices M_N, see Theorem 1.4 of [30]. The properties of m_α are described next in Proposition 2.2.15.

Proposition 2.2.15. *The Stieltjes transform $m_a(z)$ of the limiting distribution of the E.S.D. of the matrices M_N satisfies the following equation*

$$m_a(z) = i\psi_{a,z}(y(z)),$$

where

$$\varphi_{a,z}(x) = \frac{1}{\Gamma(\frac{a}{2})}\int_0^\infty t^{\frac{a}{2}-1}e^{itz}e^{-\Gamma(1-\frac{a}{2})t^{\frac{a}{2}}x}dt, \tag{2.2.5}$$

$$\psi_{a,z}(x) = \int_0^\infty e^{itz}e^{-\Gamma(1-\frac{a}{2})t^{a/2}x}dt, \tag{2.2.6}$$

$$y(z) = \varphi_{a,z}(y(z)), \tag{2.2.7}$$

where (2.2.7) is proven to have a unique solution on $\mathbb{C}^+$. Moreover the limiting probability density function ρ_a is bounded, absolutely continuous, analytic except at a possible finite set and with density at 0 given by

$$\rho_a(0) = \frac{1}{\pi}\Gamma(1+\frac{2}{a})\left(\frac{\Gamma(1-\frac{a}{2})}{\Gamma(1+\frac{a}{2})}\right)^{1/a}.$$

These results are proven in Proposition 1.1 of [24] and Theorem 1.6 of [33].

Remark 2.2.16. Later in Theorem 2.3.14, we will prove that the Stieltjes transform of X_N also converges to m_a. So we will refer to the measure whose Stieltjes transform is m_a, as the limiting measure of the E.S.D. of X_N.

2.2.2 Sketch of the proof

Now we are ready to present a sketch of the proof. At this point we will try to avoid as much technicalities as possible. In order to prove universality, meaning the same asymptotic distribution for the least singular value of D_N as in the Gaussian case, we are going to follow the three step strategy, a well known strategy in random matrix theory literature. Some of the most fundamental results concerning this method can be found in [7] and in [46], which focus on proving universality of the correlation function for symmetric matrices. The key idea is that after a slight perturbation of a random matrix by a Brownian Motion matrix, the resulting matrix should behave as a Gaussian one, given that the initial matrix satisfies some mild assumption concerning its Stieltjes transform. This idea is exploited in the study of the evolution of the eigenvalues and the eigenvectors via stochastic differential equations. This method was crucial to the proof of the Wigner-Dyson-Mehta conjecture, see for example [7]. The three step strategy has also been used in establishing universality of the least singular value for random matrices, see for example [3], [4]. Specifically in our case:

- **First step:** We investigate the asymptotic spectral behavior of X at an "intermediate" scale. At this step we prove that the matrix X satisfies the necessary conditions, which insure that after a slight perturbation by a Brownian motion matrix universality will hold. This is done in Section 2.3. Note that by definition, X contains the "big" elements of H. So the first step involves proving two estimates. One comparison of the Stieltjes transform of the E.S.D. of X with the Stieltjes transform of its limiting measure, and one bound for resolvent entries of X. Set $m_X(z)$ the Stieltjes transform of X and $R_{i,j}(z)$ the resolvent of X at z. In particular we wish to show that the following events

$$|m_a(z) - m_X(z)| = o(1), \tag{2.2.8}$$

$$\max_{j\in[2N]} \left|R_{j,j}(z)\right| = O\left(\log^C(N)\right), \text{ for some } C > 0, \tag{2.2.9}$$

hold with overwhelming probability for any $z : \mathrm{Im}(z) \geq N^{\delta-\frac{1}{2}}$ for any small enough $\delta > 0$ and $\mathrm{Re}(z)$ in some $N-$independent interval. These results are called intermediate because the natural scale would be $\mathrm{Im}(z) \geq N^{-1+\delta}$, as is explained Remark 2.2.12.

- **Second step**: We consider the perturbed matrix $X + \sqrt{t}W$, where W is the symmetrization of a full centered Gaussian matrix with i.i.d. entries with variance $\frac{1}{N}$, and t is chosen so that the variances of the entries of $\sqrt{t}W$ and of A match. It can be computed that $t \sim N^{\upsilon(a-2)}$.

  The level of the intermediate scale local law in the previous step, is justified in this part of the proof. In order to apply universality Theorems for the matrices after slightly perturbing by Brownian motion matrices, see for example Theorem 3.2 of [3], we wish the variances of the $\sqrt{t}W$ to be above the intermediate scale of the local law. Since $N^{\delta-\frac{1}{2}} = o(t)$, for small enough $\delta > 0$, this is implied.

  Roughly, what we need to prove at this step is that the desired properties, delocalization of the eigenvectors and universality of the least singular value, hold for the matrix $X + \sqrt{t}W$.

  So for the first part of the second step, we prove universality of the least positive eigenvalue for $X + \sqrt{t}W$. This is based on the regularity of the Stietljes transform of X, proven in the previous step, and some results from [3]. More precisely at the first part of the second step we prove that

$$\lim_{N\to\infty} \mathbb{P}\left(N\xi\lambda_N(X + \sqrt{t}W) \geq r\right) = \lim_{N\to\infty} \mathbb{P}\left(N\lambda_N(W) \geq r\right), \text{ for any } r \in \mathbb{R}^+. \tag{2.2.10}$$

  This result is proven in Section 2.4.

  In most of the universality-type theorems the fact that the entries have finite variances play a significant role, see for example Lemma 15.4 in [7]. In [3] the authors showed universality of the least singular value for sparse random matrix models. Firstly, they prove universality of the least singular value for the sparse models after slightly perturbing them by a Brownian motion matrix and then they remove the Brownian motion matrix. This is done by using results which take advantage of the fact that the entries have finite variance, for example see Lemma 5.14 in [3]. Since our model does not have entries with finite variance, we will need to compare the matrices $X+\sqrt{t}W$ and $X+A$ with a different method. Essential to that method is the fact that the resolvent entries of $X+\sqrt{t}W$ do not grow very fast. Specifically set $T_{i,j}(z)$ to be the resolvent of $X+\sqrt{t}W$ at z. In particular in Section 2.5 we prove that for any small $\delta>0$ the $\delta-$dependent events

$$\sup_{i,j}|T_{i,j}(z)|\le N^{\delta} \tag{2.2.11}$$

hold with overwhelming probability, and for all $z:\mathrm{Im}(z)\ge N^{\epsilon-1}$ for any small $\epsilon>0$, very close to the natural scale in Remark 2.2.12. It is known that bounds as the one in (2.2.11) imply the complete eigenvector delocalization for the matrix $X+\sqrt{t}W$.

In order to establish (2.2.11), we prove something better. A **universal** result which compares the entries of the resolvent of any matrix, which satisfies some mild regularity assumption, Assumption 2.5.1, with the additive free convolution of the matrix with the semicircle law. Thus, the largest part of Section 2.5 is mostly independent for the rest of the paper.

- **Third step:** We first compare the resolvent of $X+A$ and $X+\sqrt{t}W$. During the second step we have proven the desired properties, eigenvector delocalization and universality of the least eigenvalue for the matrix $X+\sqrt{t}W$, so we need to find a way to quantify the transition from the matrix $X+\sqrt{t}W$ to $X+A$ in order to prove the same properties for H. This is done by introducing the matrices

$$H^{\gamma}:=X+\sqrt{t}(1-\gamma^2)^{1/2}W+\gamma A,\ \text{for all}\ \gamma\in[0,1].$$

We manage to prove that the resolvent entries of H^{γ} are asymptotically close for all $\gamma\in[0,1]$, in Theorem 2.6.4. Similarly we study the continuity properties for $\gamma\in[0,1]$ of the functions

$$q\left(\frac{N}{\pi}\int_{\frac{-r}{N}}^{\frac{r}{N}}\mathrm{Im}(m_{\gamma}(E+i\eta)dE\right), \tag{2.2.12}$$

where m_{γ} is the Stieltjes transform of the matrix H^{γ} and η is of order $N^{-\delta-1}$, below the natural scale. Eventually in (2.6.36) we prove that the functions defined in (2.2.12) are asymptotically close for any $\gamma\in[0,1]$.
Next we introduce the functions

$$\iota_N(Y,r):=\#\{i\in[N]:\lambda_i(Y)\in(-r,r)\},$$

where Y is a symmetric $N \times N$ matrix, $\{\lambda_i\}_{i\in[N]}$ are the eigenvalues of Y and r is any positive number. So, it suffices to prove that there exists $c > 0$ such that for any $r > 0$

$$\left|\mathbb{P}\left[\imath_{2N}\left(X + \sqrt{t}G, \frac{r}{N}\right) = 0\right] - \mathbb{P}\left[\imath_{2N}\left(X + A, \frac{r}{N}\right) = 0\right]\right| \leq O_r(N^{-c}). \tag{2.2.13}$$

In order to prove the latter, we approximate the quantities $\mathbb{P}(\imath_{2N}(H^\gamma, \frac{r}{N}) = 0)$ by appropriately choosing functions of the form (2.2.12). This is done in Lemma (2.6.13). After combining the results above, we conclude the proof in Subsection 2.6.3.

2.3 Intermediate local law for X

Consider the matrices H_N and X_N as they are defined in Definition 2.2.4. In this section we are going to establish the local law (Theorem 2.3.14) for the b-removals of the matrix H, i.e., the matrix X. What we mean by local law is convergence of the Stieltjes transform of X to its asymptotic limit, for complex numbers z that depend on the dimension N in some sense.

We will also use the notation

$$R(z) = (X - (E + i\eta)I)^{-1}, \tag{2.3.1}$$

for $z = E + i\eta$. In what follows we might abbreviate the dependence from the parameter z.

A precise formulation of this result is the following. There exists $C = C(a, b, \delta)$ such that

$$\mathbb{P}\left(\sup_{E\in(-\frac{1}{C},\frac{1}{C})}\ \sup_{\eta\geq N^{\delta-\frac{1}{2}}} |m_a(E + i\eta) - m_X(E + i\eta)| \geq \frac{1}{N^{a\delta/8}}\right) \leq \exp\left(-\frac{(\log(N))^2}{C}\right), \tag{2.3.2}$$

where the properties of $m_a(z)$ are described in Proposition 2.2.15.

We also prove that for all z for which the local law holds, the diagonal entries of the resolvent of X are almost bounded. More specifically for any large enough $N \in \mathbb{N}$ it is true that,

$$\mathbb{P}\left(\sup_{E\in(-\frac{1}{C},\frac{1}{C})}\ \sup_{\eta\geq N^{\delta-\frac{1}{2}}}\ \max_{j\in[2N]} |R_{j,j}| > C\log^C(N)\right) \leq C\exp\left(-\frac{(\log(N))^2}{C}\right). \tag{2.3.3}$$

In order to establish those results we will need to analyze the resolvent of X, in order for us to compare it with m_a. The main influence for this step is Theorem 3.5 of [5], where an intermediate local law is proven for symmetric heavy tailed random matrices. The main difference of the proof of the intermediate local law for our set of matrices from the symmetric case is that, by construction, only half of the minors of the resolvent will participate in the sum of the reductive formula from Schur complement formula. This difference is not crucial since we also prove that each of the diagonal entries of the resolvent of the matrix is identically distributed. Moreover, by Corollary 2.3.22, the sum of half of the diagonal entries of the resolvent is concentrated around its mean, like the sum of all its diagonal entries. The rest of the proof remains almost the same, but we will include most of the proofs for completeness of the paper.

Firstly we need to give a more detailed description of the limiting distribution.

2.3.1 Preliminaries for the intermediate local law

In [30] the authors proved that the E.S.D. of symmetric matrices with heavy tailed entries, converge in distribution to a deterministic measure and they described it. Next in [24], the authors described the limit of the sample covariance matrices. Next the authors in [32] and [31] proved local laws for symmetric heavy tailed matrices at an intermediate scale larger than $N^{\delta-\frac{1}{2}}$. Lastly in [5], the authors proved a local law at the intermediate scale $N^{\delta-\frac{1}{2}}$. All the previously mentioned results, are based on solving a fixed point equation. In the most recent results, these fixed equations are solved more generally, in a metric space which we are going to present in this subsection.

Next we present the metric space in which we will work with in order to prove an intermediate local law for X. The results we present here can be also found in [32].

Definition 2.3.1. For any $u, v \in \mathbb{C}$ define the following "inner product"

$$(u|v) := u\operatorname{Re}(v) + \bar{u}\operatorname{Im}(v) = \operatorname{Re}(u)(\operatorname{Re}(v) + \operatorname{Im}(v)) + i\operatorname{Im}(u)(\operatorname{Re}(v) - \operatorname{Im}(v)).$$

One may compute the following

$$(u|1) = u, \quad (-iu|e^{\pi i/4}) = \operatorname{Im}(u)\sqrt{2}, \quad |(u|v)| \leq 2|u||v|. \tag{2.3.4}$$

Definition 2.3.2. Set $\mathbb{K} = \mathbb{C}^+ \cap \{z \in \mathbb{C} : \operatorname{Re}(z) > 0\}$ and $\mathbb{K}^+ = \bar{\mathbb{K}}$. Let H_w be the space of the C^1, $g : \mathbb{K}^+ \to \mathbb{C}$ such that $g(\lambda u) = \lambda^w u$ for each $\lambda > 0$. Set also $S^1_+ = \overline{S^1 \cap \mathbb{K}^+}$ where S^1 is the unit sphere on $\mathbb{C}$ with respect to the Euclidean norm. Following equation (10) of [32], define for each $r \in [0, 1)$ a norm on H_r

$$|g|_\infty = \sup_{u \in S^1_+} |g(u)|,$$

$$|g|_r = |g|_\infty + \sup_{u \in S^1_+} \sqrt{|(i|u)^r \partial_1 g(u)|^2 + |(i|u)^r \partial_2 g(u)|^2}.$$

Here,

$$\partial_1 g(x + iy) = \frac{dg(x + iy)}{dx}$$

and likewise

$$\partial_2 g(x + iy) = \frac{dg(x + iy)}{dy}.$$

Next, define the spaces $H_{w,r}$ the completion of H_w with respect to the $|g|_r$ norm. Further define $H^\delta_{w,r} \subseteq H_{w,r}$ to be the set $\{g \in H_{w,r} : \inf_{u \in S^1_+} |\operatorname{Re}(g(u))| > \delta\}$. Also define the set

$$H^0_{w,r} = \cup_{\delta > 0} H^\delta_{w,r}.$$

Further, abbreviate $H^{\delta}_{w} := H^{\delta}_{w,0}$

Remark 2.3.3. For any $g \in H_r$, by construction it is true that

$$|g|_\infty \leq |g|_r.$$

Next, we present some lemmas concerning the metric spaces we presented and the fixed point equation we wish to solve.

Lemma 2.3.4 ([32],Lemma 5.2)**.** *Let* $r \in (0,1)$ *and* $u \in S^1_+$ *and* $x_1, x_2 \in \mathbb{K}^+$ *and let* $\eta \in (0,1)$ *such that* $|x_1|, |x_2| \leq \eta^{-1}$. *Set* $F_k(u) = (x_k|u)^r$ *for* $k \in \{1,2\}$. *Then there exists a constant* $C(r)$ *such that for any* $s \in (0,r)$ *we have that*

$$|F_k|_{1-r+s} \leq C|x_k|^r, \quad |F_1 - F_2|_{1-r+s} \leq C\eta^{-r}(|x_1 - x_2|^r + \eta^s|x_1 - x_2|^s). \tag{2.3.5}$$

Furthermore, if we further assume that $\mathrm{Re}(x_1), \mathrm{Re}(x_2) \geq t$ *and set* $G_k(u) = (x_k^{-1}|u)^r,\ k \in \{1,2\}$, *there exists a constant* $C = C(r)$ *such that,*

$$|G_1 - G_2|_{1-r+s} \leq Ct^{r-2}\eta^{2r-1}|x_1 - x_2|. \tag{2.3.6}$$

Definition 2.3.5. Following Section 3.2 of [32], for any numbers $h \in \bar{K}$, $u \in S^1_+$ and $g \in H_{\alpha/2}$ define the functions,

$$F_{h,g}(u) = \int_0^{\pi/2}\int_0^\infty\int_0^\infty [\exp(-r^{\alpha/2}g(e^{i\partial}) - (rh|e^{i\partial}))$$
$$- \exp(-r^{\alpha/2}g(e^{i\partial} + uy) - (urh|u) - (rh|e^{i\partial}))]r^{\alpha/2-1}dr \cdot \cdot y^{-\alpha/2-1}dy(\sin(\partial))^{\alpha/2-1}d\partial \tag{2.3.7}$$

and

$$Y_f(u) = Y_{z,f}(u) = c_\alpha F_{-iz,f}(\tilde{u}),$$

where

$$c_\alpha = \frac{\alpha}{2^{\alpha/2}\Gamma(\alpha/2)^2}.$$

Lemma 2.3.6 ([32],Lemma 4.1)**.** *If* $g \in H^0_{\alpha/2,r}$ *then* $F_{\eta,g} \in H_{\alpha/2,r}$. *Also if* $g \in H^{\bar{0}}_{\alpha/2,r}$ *and* $\mathrm{Re}(h) > 0$ *then* $F_{g,h} \in H^{\bar{0}}_{\alpha/2,r}$.

Next for any $f \in H_{\alpha/2}$ and $p > 0$ define the functions,

- $r_{p,z}(f) = \frac{2^{1-p/2}}{\Gamma(p/2)}\int_0^{\pi/2}\int_0^\infty y^{p-1}\exp((iyz|e^{i\partial}) - y^{\alpha/2}f(e^{i\partial}))\sin(2\partial)^{p/2-1}dyd\partial$
- $s_{p,z}(x) = \frac{1}{\Gamma(p)}\int_0^\infty y^{p-1}\exp(-iyz - xy^{\alpha/2})dy.$

Lemma 2.3.7 ([32],Proposition 3.3)**.** *There exists a countable subset* $\mathcal{A} \subseteq (0,2)$ *with no accumulation points such for any* $r \in (0,1]$ *and* $\alpha \in (0,2) \setminus \mathcal{A}$, *there exists a constant* $c = c(\alpha, r)$ *with the property that:*

There exists a unique function $\Omega_0 \in H_{a/2}$ *such that* $\Omega_0 = Y_{0,\Omega_0}$. *Additionally for* $\mathrm{Im}(z) > 0$ *and* $|z| \le c$, *there exists a unique function* $f = \Omega_z \in H_{a/2,r}$ *that solves* $f = Y_{z,f}$ *with* $\lfloor f - \Omega_0 \rfloor_r \le c$. *Moreover the function satisfies* $\Omega_z(e^{i\pi/4}) \ge c$ *and for any* $p > 0$ *there exists a constant* $C = C(a,p)$ *such that* $|r_{p,z}(\Omega_z)| \le C$.

Lemma 2.3.8. *([32],Proposition 3.4) Adopt the notation of the previous lemma. After decreasing c if necessary, there exists a constant* $C > 0$ *such that the following holds.*
If $\mathrm{Im}(z) > 0, |z| \le c$ *and* $\lfloor f - \Omega_z \rfloor_r \le c$, *then*

$$\lfloor f - \Omega_z \rfloor_r \le C \lfloor f - Y_{z,f} \rfloor_r.$$

Lemma 2.3.9. *([32],Lemma 4.1) Let* $r \in (0,1)$ *and* $p > 0$. *There exists a constant* $C = C(a,p,r) > 0$ *such that, for any* $g \in \bar{H}^0_{a/2,r}$ *and* $h \in \mathbb{K}$, *we have that*

$$|F_\eta(g)|_r \le C(\mathrm{Re}\,\eta)^{-a/2} + C|g|_r(\mathrm{Re}(\eta))^{-a/2},$$

$$|r_{p.i\eta}(g)| \le C(\mathrm{Re}\,\eta)^{-p}, \quad |s_{p,i\eta}(g(1))| \le C(\mathrm{Re}(\eta))^{-p}.$$

Lemma 2.3.10. *([32], Lemma 4.3) For any* $a, r > 0$ *there exists a constant* $C = C(a,a,r) > 0$ *such that for any* $f, g \in H^{\bar{0}}_{a/2,r}$, *and* $z \in \mathbb{C}$

$$\lfloor Y_f - Y_g \rfloor_r \le C \lfloor f - g \rfloor_r + \lfloor f - g \rfloor_\infty (\lfloor f \rfloor_r + \lfloor g \rfloor_r).$$

Furthermore, for any $p > 0$ *there exists a constant* $C' = C'(a,a,r,p)$ *such that for any* $f, g \in H^a_{a/2,r}$ *and for any* $z \in \mathbb{C}$ *and* $x, y \in \mathbb{K}$ *with* $\mathrm{Re}(x), \mathrm{Re}(y) \ge a$ *we have that*

$$|r_{p,z}(f) - r_{p,z}(g)| \le C' \lfloor f - g \rfloor_\infty, \quad |s_{p,z}(x) - s_{p,z}(y)| \le C'|x-y|. \tag{2.3.8}$$

The reason to present all the tools in this subsection is explained in the following Remark.

Remark 2.3.11. Due to Lemma 4.4. of [32], $is_{1,z}(\Omega_z(1))$, which is defined in Lemma 2.3.7, is exactly the limiting Stieltjes transform in Proposition 2.2.15.

2.3.2 Statement of the intermediate local law

In this subsection, we state the local law for the matrix X and state a stronger theorem which will imply the local law.
Before we present the theorem, we give some definitions. Recall the notation from Subsection 2.3.1.

Definition 2.3.12. Define the following quantities

$$\iota_z(u) := \Gamma\left(1 - \frac{a}{2}\right)(-iR_{j,j}|u)^{a/2}, \quad \gamma_z := \mathbf{E}(\iota_z(u)),$$

$$I_p := \mathbf{E}(-iR_{j,j})^p, \quad J_p := \mathbf{E}(|iR_{j,j}|^p),$$

where we have omitted the dependence from the dimension N in the notation we used.

In what follows, keep in mind the definition of the functions $r_{p,z}$ and $s_{p,z}$ in Subsection 2.3.1. Next, we present the theorem which will imply the intermediate local law proved at Subsection 2.3.7.

Theorem 2.3.13. *Let $a \in (0,2)$, $b \in (0, \frac{1}{a})$, $s \in (0, \frac{a}{2})$, $p > 0$, $\epsilon \in (0,1]$ and $N \in \mathbb{N}$. Set $\partial = (\frac{1}{a} - b)(2-a)/10$. Suppose $z = E + i\eta \in \mathbb{C}^+$ with $E, \eta \in \mathbb{R}$. Assume that:*

$$z = E + i\eta, \quad |z| \le \frac{1}{\epsilon}, \quad \eta \ge N^{\epsilon - s/a}, \quad \mathbf{E}(\mathrm{Im}(R_{i,i})^{a/2}) \ge \epsilon, \quad \mathbf{E}|R_{i,i}|^2 \le \epsilon^{-1}, \text{ for all } i \in [2N]. \tag{2.3.9}$$

Then, there exists a constant $C = C(a, \epsilon, b, s, p) > 0$ such that

$$\left|\gamma_z - Y_{\gamma_z}\right|_{1-a/2+s} \le C \log^C(N) \left(\frac{1}{(\eta^2 N)^{a/8}} + \frac{1}{N^{\partial}} + \frac{1}{N^s \eta^{a/2}} \right), \tag{2.3.10}$$

$$\left|I_p - s_{p,z}(\gamma_z(1))\right| \le C \log^C(N) \left(\frac{1}{(\eta^2 N)^{a/8}} + \frac{1}{N^{\partial}} + \frac{1}{N^s \eta^{a/2}} \right), \tag{2.3.11}$$

$$|J_p - r_{p,z}(\gamma_z)| \le C \log^C(N) \left(\frac{1}{(\eta^2 N)^{a/8}} + \frac{1}{N^{\partial}} + \frac{1}{N^s \eta^{a/2}} \right). \tag{2.3.12}$$

Moreover,

$$\inf_{u \in S^1_+} \mathrm{Re}(\gamma_z(u)) > \frac{1}{C} \tag{2.3.13}$$

and

$$\mathbb{P}\left(\max_{j \in [2N]} |R_{j,j}| > C \log^C(N) \right) \le C \exp\left(-\frac{(\log(N))^2}{C} \right). \tag{2.3.14}$$

The proof of Theorem 2.3.13 can be found in Subsection 2.3.7.

Next we present the local law.

Theorem 2.3.14 (Local law)**.** *There exists a countable set $\mathcal{A} \subseteq (0,2)$ with no accumulation points in $(0,2)$ such that for each $a \in (0,2) \setminus \mathcal{A}$ the following holds. Fix $b \in (0, \frac{1}{a})$, $\partial = (\frac{1}{a} - b)(2-a)/10$ and $\delta \in (0, \min\{\partial, \frac{1}{2}\})$. Then there exists a constant $C = C(a, b, \delta, p) > 0$ such that*

$$\mathbb{P}\left(\sup_{z \in D_{C,\delta}} \left| m_N(z) - i s_{1,z}(\Omega_z(1)) \right| > \frac{1}{N^{a\delta/8}} \right) \le C \exp\left(-\frac{\log^2(N)}{C} \right). \tag{2.3.15}$$

Furthermore,

$$\sup_{u \in S^1_+} |\gamma_z(u) - \Omega_z(u)| \le \frac{C}{N^{a\delta/8}} \tag{2.3.16}$$

and

$$\mathbb{P}\left(\sup_{z \in D_{C,\delta}} \max_{j \in [2N]} |R_{j,j}| > C \log^C(N) \right) \le C \exp\left(-\frac{(\log(N))^2}{C} \right). \tag{2.3.17}$$

Where $D_{C,\delta} = \{z = E + i\eta : E \in (-\frac{1}{C}, \frac{1}{C}), \frac{1}{C} \ge \eta \ge N^{\delta - 1/2}\}$, $m_N(z)$ is the Stieltjes transform of X and $\Omega_z(u)$ is defined in Lemma 2.3.7.

Proof Of Theorem 2.3.14 given Theorem 2.3.13. The proof is similar to the proof of Theorem 7.6 given Theorem 7.8 and Lemmas 2.3.7, 2.3.8 (there called Lemma 7.2 and Lemma 7.3) in [5], so it will be omitted. □

2.3.3 General results concerning the resolvent and the eigenvalues of a matrix

Firstly we present a well-known result that compares the eigenvalues of a matrix with the eigenvalues of its minors.

Lemma 2.3.15 (Weyl's inequality)**.** *Let $R, M, Q \in \mathbb{R}^{N^2}$ some symmetric matrices such that*

$$M = R + Q.$$

Let μ_i, ρ_i, q_i be the eigenvalues of M, R, Q respectively arranged in decreasing order. Then

$$q_j + \rho_k \leq \mu_i \leq q_r + \rho_s$$

for any indices such that $j + k - n \geq i \geq r + s - 1$.

In the rest of this subsections we present some general results concerning the resolvent of a matrix. Most of them are known results, but we include them because they will be useful in the proof of Theorem 2.3.14.

Lemma 2.3.16. *Let M_1, M_2 be two invertible, $N \times N$ matrices then the following identity is true*

$$M_1^{-1} - M_2^{-1} = M_1^{-1}(M_2 - M_1)M_2^{-1}. \tag{2.3.18}$$

Moreover if $Y = (M_1 - z\mathbb{I})^{-1}$ such that $z \in \mathbb{C}^+ = \{z \in \mathbb{C} : \mathrm{Im}(z) > 0\}$ then

$$|Y_{i,j}| \leq \frac{1}{\mathrm{Im}(z)}, \quad i, j \in [N]. \tag{2.3.19}$$

Proof. The identity (2.3.18) follows trivially by a right multiplication on both sides by the element M_1 and a left multiplication on both sides by the element M_2. Moreover (2.3.19) follows trivially from the spectral theorem. □

Definition 2.3.17. In what follows in this section we will use the following notation. Consider M to be any $N \times N$ matrix. Let $J \subseteq [N]$. We will use the notation $(M^{(J)} - z\mathbb{I})^{-1}$ for the resolvent of the matrix $M^{(J)}$, where $M^{(J)}$ is the matrix M with the $i - th$ row and column being replaced by zero vectors, for each $i \in J$.

Lemma 2.3.18. *Let M be an $N \times N$ matrix and $z \in \mathbb{C}^+$.*
Then we have the following complements formula

$$\frac{1}{(M - z\mathbb{I})^{-1}_{i,i}} = M_{i,i} - z - \sum_{k,j \in [N] \setminus \{i\}} M_{i,j}(M^{(i)} - z\mathbb{I})^{-1}_{jk} M_{k,i}. \tag{2.3.20}$$

Next, we present the Ward identity. That is, for each $J \subseteq [N]$ and $j \in [N] \setminus J$ it is true that

$$\sum_{k \in [N] \setminus J} |(M^{(J)} - z\mathbb{I})^{-1}_{jk}|^2 = \frac{\mathrm{Im}((M^{(J)} - z\mathbb{I})^{-1}_{j,j})}{\mathrm{Im}(z)}. \tag{2.3.21}$$

Proof. The estimates (2.3.20) and (2.3.21) can be found in (8.8) and (8.3) in [7], respectively. □

Lemma 2.3.19 ([32],Lemma 5.5)**.** *Let M be an $N \times N$ matrix. For any $r \in (0, 1], z \in \mathbb{C}^+$, $\eta =$ $\mathrm{Im}(z)$ and $i \in [N]$ we have the following deterministic bound*

$$\frac{1}{N}\sum_{i=1}^{N}\left|(M - z\mathbb{I})^{-1}_{i,i} - (M^{(i)} - z\mathbb{I})^{-1}_{i,i}\right|^r \le \frac{4}{(N\eta)^r}.$$

Corollary 2.3.20. *([5],Cor 5.7) Let M be an $N\times N$ matrix. For any $r \in [1, 2], z \in \mathbb{C}^+$, $\eta = \mathrm{Im}(z)$ and $i \in [N]$ we have the deterministic estimate,*

$$\frac{1}{N}\sum_{i=1}^{N}\left|(M - z\mathbb{I})^{-1}_{i,i} - (M^{(i)} - z\mathbb{I})^{-1}_{i,i}\right|^r \le \frac{4}{(N\eta)^r} \le \frac{8}{N\eta^r}.$$

2.3.4 Concentration results for the resolvent of a matrix

In this subsection, we present various identities and inequalities concerning the resolvent and the eigenvalues of a matrix.

Next we present some concentration inequalities.

Lemma 2.3.21. *Let N be an even positive integer and let $A = (a_{i,j})_{1\le i,j\le N}$ such that the rows $A_i = (a_{i1}, a_{i2}, \cdots, a_{ii})$ are mutually independent for each $i \in [N]$. Let $B = (A - z\mathbb{I})^{-1}$ and $z = E + i\eta$ where $\eta > 0$. Then for any Lipchitz function f with Lipchitz norm L_f and any $x > 0$,we have that,*

$$\mathbb{P}\left[\left|\frac{2}{N}\sum_{i=1}^{\frac{N}{2}} f(B_{i,i}) - \mathbf{E}\frac{2}{N}\sum_{i=1}^{\frac{N}{2}} f(B_{i,i})\right| \ge x\right] \le 2\exp\left(-\frac{N\eta^2x^2}{8L_f^2}\right), \tag{2.3.22}$$

$$\mathbb{P}\left[\left|\frac{2}{N}\sum_{i=1}^{\frac{N}{2}} f(B_{i+\frac{N}{2},i+\frac{N}{2}}) - \mathbf{E}\frac{2}{N}\sum_{i=1}^{\frac{N}{2}} f(B_{i+\frac{N}{2},i+\frac{N}{2}})\right| \ge x\right] \le 2\exp\left(-\frac{N\eta^2x^2}{8L_f^2}\right). \tag{2.3.23}$$

Proof. The proof is similar to the respective proof for the Stieltjes transform in Lemma C.4 of [31]. We will sketch the proof for the first $N2^{-1}$ diagonal entries. The proof for the remaining $N2^{-1}$ entries is similar. More precisely, for any two deterministic Hermitian matrices C and B, let $R(C)$ and $R(B)$ be their resolvents at z. Then it is proven in equation (91) of Lemma C.4 of [31] that :

$$\frac{1}{N}\left|\sum_{k=1}^{N/2} R_{k,k}(B) - R_{k,k}(C)\right| \le \frac{1}{N}\sum_{k=1}^{N}\left|R_{k,k}(C) - R_{k,k}(B)\right| \le \mathrm{rank}(C - B)2(\mathrm{Im}(z)N)^{-1}. \tag{2.3.24}$$

So if one considers the function

$$F(\{x_i\}_{i=1}^N) = \sum_{k=1}^{N/2}\frac{f(R_{k,k}(X))}{N}, \qquad \{x_i\}_{i=1}^N : x_i \in \mathbb{C}^{i-1}\times\mathbb{R},$$

where X is a Hermitian matrix with the i-th row of X being x_i. Note that it suffices to describe the entries of the i-th row until the i-th column since the remaining elements will be filled by the properties of the Hermitian matrices. So if we consider two elements $X, X' \in \cup_{i=1}^{N} \mathbb{C}^{i-1} \times \mathbb{R}$ with only the i-th vector of X and X' different, then one has:

$$|F(X) - F(X')| \leq \mathrm{rank}(X - X')2(\mathrm{Im}(z)N)^{-1} \leq 4(\mathrm{Im}(z)N)^{-1},$$

since by construction, one has that $\mathrm{rank}(X - X') \leq 2$. Now the desired inequality comes from Azuma{Hoeffding inequality, see Lemma 1.2 in [47]. □

Corollary 2.3.22. *One can apply the previous Lemma to get the following concentration results. Fix an $N \times N$ symmetric random matrix Y with i.i.d. entries (up to symmetry), where N is an even integer, with resolvent matrix $B = (Y - z\mathbb{I})^{-1}$ for $z = E + i\eta$. Then the following bounds are true:*

$$\begin{aligned} \mathbb{P}\left[\frac{2}{N}\left|\sum_{k=1}^{N/2} B_{k,k} - \mathbf{E}B_{k,k}\right| \geq \frac{4\log(N)}{(N\eta^2)^{1/2}}\right] \leq 2\exp\left(-(\log(N))^2\right), \\ \mathbb{P}\left[\frac{2}{N}\left|\sum_{k=1}^{N/2} \mathrm{Im}(B_{k,k}) - \mathbf{E}\,\mathrm{Im}(B_{k,k})\right| \geq \frac{4\log(N)}{(N\eta^2)^{1/2}}\right] \leq 2\exp(-(\log(N))^2). \end{aligned} \tag{2.3.25}$$

Moreover for any $\alpha \in (0, 2)$ there exists a constant $C = C(\alpha)$ such that,

$$\mathbb{P}\left[\frac{2}{N}\left|\sum_{k=1}^{N/2} (-iB_{k,k})^{\alpha/2} - \mathbf{E}(-iB_{k,k})^{\alpha/2}\right| \geq x\right] \leq 2\exp\left(-\frac{N(\eta^{\alpha/2}x)^{4/\alpha}}{C}\right). \tag{2.3.26}$$

The same results hold for the remaining $N/2$ diagonal entries of R.

Proof. The first two inequalities are true by direct application of Lemma 2.3.21 for the functions $f(x) = x$ and $f(x) = \mathrm{Im}(x)$ respectively.

For the third inequality let $c > 0$ and fix $\phi_c : \mathbb{C} \to \mathbb{R}^+$, such that

$$\phi_c(z) = \begin{cases} 0 & |z| \leq c, \\ \frac{1}{c}(|z| - c) & |z| \in (c, 2c), \\ 1 & |z| \geq 2c. \end{cases} \tag{2.3.27}$$

Note that the function ϕ_c is Lipschitz with Lipschitz constant bounded by $\frac{1}{c}$. Then define the function

$$f_c(z) = (-iz)^{\alpha/2}\phi_c(z). \tag{2.3.28}$$

Since $|(1 - \phi_c(z))(-iz)^{\alpha/2}| \leq (2c)^{\alpha/2}$ for all $z \in \mathbb{C}^+$, it is clear that $|(-iz)^{\alpha/2}| \leq f_c(z) + (2c)^{\alpha/2}$. Moreover note that the function $f_c(z)$ is Lipschitz with constant bounded by $2c^{\frac{\alpha}{2}-1}$.

So for any $x \geq 0$ fix c such that $(2c)^{\alpha/2} = x/4$. Then

$$\mathbb{P}\left[\frac{2}{N}\left|\sum_{k=1}^{N/2}(-iB_{k,k})^{\alpha/2} - \mathbf{E}(-(-iB_{k,k})^{\alpha/2}\right| \geq x\right] \tag{2.3.29}$$

$$\leq \mathbb{P}\left[\frac{2}{N}\left|\sum_{k=1}^{N/2} f_c\left(B_{k,k}\right) - \mathbf{E}f_c\left(B_{k,k}\right)\right| \geq \frac{x}{2}\right]. \tag{2.3.30}$$

Now the proof is completed after a direct application of Lemma 2.3.21, after noticing that $2c^{\frac{\alpha}{2}-1} = 2^{\frac{4}{\alpha}-\frac{\alpha}{2}}x^{1-\frac{2}{\alpha}}$. □

The following result is an analogue of Lemma 5.3 in [32] for concentration of only half of the resolvent diagonal entries. The proof is analogous.

Lemma 2.3.23. *Let N be an even and positive integer, $A = \{a_{i,j}\}_{1\leq i,j\leq N}$ a symmetric matrix with independent entries (up to symmetry). Fix $u \in \mathbb{S}^1_+$, $\alpha \in (0,2)$ and $s \in (0, \frac{\alpha}{2})$. Moreover define the resolvent matrix $B = (A - z_0\mathbb{I})^{-1}$ for $z_0 = E + i\eta \in \mathbb{C}^+$.*
Then if we denote $f_u : \mathbb{C} \to \mathbb{C}$ such that $f_u(z) = (iz|u)^{\alpha/2}$, there exists constant $C = C(\alpha) > 0$ such that

$$\mathbb{P}\left[\left\|\frac{2}{N}\sum_{i=1}^{\frac{N}{2}} f_u(B_{i,i}) - \mathbf{E}\frac{2}{N}\sum_{k=1}^{\frac{N}{2}} f_u(B_{k,k})\right\|_{1-\alpha/2+s} \geq x\right] \leq C(\eta^{\alpha/2}x)^{-1/s}\exp\left(-\frac{N(x\eta^{\frac{\alpha}{2}})^{\frac{2}{s}}}{C}\right).$$

A similar estimate is true for the concentration of the second half of the diagonal entries of the resolvent.

Proof. By definition of the norms in Definition 2.3.2 we need to bound the following quantities

$$\mathbb{P}\left[\sup_{u\in\mathbb{S}^1_+}\left|\frac{2}{N}\sum_{k=1}^{\frac{N}{2}} f_u(B_{k,k}) - \mathbf{E}\frac{2}{N}\sum_{k=1}^{\frac{N}{2}} f_u(B_{k,k})\right| \geq x\right] \text{ for any } x > 0, \tag{2.3.31}$$

$$\mathbb{P}\left[\sup_{u\in\mathbb{S}^1_+}\max_{j\in\{1,2\}}\left|(i|u)^{1-\frac{\alpha}{2}+s}\partial_j\left(\frac{2}{N}\sum_{k=1}^{\frac{N}{2}} f_u(B_{k,k}) - \mathbf{E}\frac{2}{N}\sum_{k=1}^{\frac{N}{2}} f_u(B_{k,k})\right)\right| \geq x\right] \text{ for any } x > 0. \tag{2.3.32}$$

Fix $u \in \mathbb{S}^1_+$ and $c > 0$. Then, similarly to the proof of (2.3.26) in Corollary 2.3.22, we can construct a function $\phi_c : \mathbb{C} \to \mathbb{R}^+$, which is $\frac{1}{c}-$ Lipchitz function and for which it is true that if we decompose f_u in the following sense,

$$f_u(z) = \phi_c f_u(z) + (1 - \phi_c)f_u(z) = f_{1,u}(z) + f_{2,u}, \tag{2.3.33}$$

then $f_{2,u}(z)$ is bounded by $(2c)^{1-\frac{\alpha}{2}+s}$ and $f_{1,u}(z)$ is Lipschitz with constant bounded by $c'c^{s-\frac{\alpha}{2}}$, for some other absolute constant c'. So for any $x > 0$, if one fixes c such that

$(2c)^{1-\frac{a}{2}+s} = \frac{x}{4}$ it is implied that

$$\mathbb{P}\left[\left|\frac{2}{N}\sum_{k=1}^{\frac{N}{2}} f_u(B_{k,k}) - \mathbf{E}\frac{2}{N}\sum_{k=1}^{\frac{N}{2}} f_u(B_{k,k})\right| \geq x\right] \leq \mathbb{P}\left[\left|\frac{2}{N}\sum_{k=1}^{\frac{N}{2}} f_{1,u}(B_{k,k}) - \mathbf{E}\frac{2}{N}\sum_{k=1}^{\frac{N}{2}} f_{1,u}(B_{k,k})\right| \geq \frac{x}{2}\right]. \tag{2.3.34}$$

So by a direct application of Lemma 2.3.21 for the function $f_{1,u}$ one can conclude that

$$\mathbb{P}\left[\left|\frac{2}{N}\sum_{k=1}^{\frac{N}{2}} f_u(B_{k,k}) - \mathbf{E}\frac{2}{N}\sum_{k=1}^{\frac{N}{2}} f_u(B_{k,k})\right| \geq x\right] \leq \exp\left(-\frac{N(x\eta^{\frac{a}{2}})^{\frac{2}{s}}}{C}\right), \tag{2.3.35}$$

for some constant $C = C(a)$. Moreover due to the deterministic bounds in (2.3.4) and (2.3.19), we can restrict to the case that $x \leq 4\eta^{-\frac{a}{2}}$. Furthermore, by (4.6) in [48] for any $c \in (0,1)$, any $c-$net of the sphere S^1_+ has cardinality at most $\frac{3}{c}$. Set $\mathcal{F}$ one $c-$net of the sphere. So for any $x \in (0, 4\eta^{-\frac{a}{2}})$, fix c such that $(2^{\frac{1}{2}}c\eta^{-1})^{1-\frac{a}{2}+s} = \frac{x}{4}$. Thus by (2.3.5), we conclude that

$$\mathbb{P}\left[\sup_{u\in S^1_+}\left|\frac{2}{N}\sum_{i=1}^{\frac{N}{2}} f_u(B_{k,k}) - \mathbf{E}\frac{2}{N}\sum_{k=1}^{\frac{N}{2}} f_u(B_{k,k})\right| \geq x\right] \leq \mathbb{P}\left[\sup_{u\in\mathcal{F}}\left|\frac{2}{N}\sum_{k=1}^{\frac{N}{2}} f_u(B_{k,k}) - \mathbf{E}\frac{2}{N}\sum_{k=1}^{\frac{N}{2}} f_u(B_{k,k})\right| \geq \frac{x}{2}\right]. \tag{2.3.36}$$

So we get (2.3.31) after using the union bound and (2.3.35) to bound (2.3.36).

It remains to prove (2.3.32). For this note that

$$\frac{2}{N}\partial_j \sum_{k=1}^{\frac{N}{2}} f_u(B_{k,k}) = \frac{2}{N}\sum_{k=1}^{\frac{N}{2}} (1 - \frac{a}{2} + s)(iB_{k,k}|u)^{\frac{a}{2}-1}(iB_{k,k}|j). \tag{2.3.37}$$

So we can treat the function $g_u(z) = (iz|u)(iz|j)$ analogously $f_u(z)$ in (2.3.33). In particular we have the following decomposition

$$g_u(z) = \phi_c g_u(z) + (1-\phi_c)g_u(z) = g_{1,u}(z) + g_{2,u}, \tag{2.3.38}$$

where $g_{1,u}$ is Lipschitz with constant bounded by $c_0 c^{s-\frac{a}{2}}/|ui|$ and $g_{2,u}$ is bounded by $c_0 c^{1+s-\frac{a}{2}}/|ui|$, for some absolute constant c_0. So for any $x > 0$ let c be a number such that $c_0 c^{1+s-\frac{a}{2}} = x/4$, we get that

$$\mathbb{P}\left[\left|(i|u)^{1-\frac{a}{2}+s}\partial_j\left(\frac{2}{N}\sum_{k=1}^{\frac{N}{2}} f_u(B_{k,k}) - \mathbf{E}\frac{2}{N}\sum_{k=1}^{\frac{N}{2}} f_u(B_{k,k})\right)\right| \geq x\right] \tag{2.3.39}$$

$$\leq \mathbb{P}\left[|(i|u)|^{1-\frac{a}{2}+s}\left|\left(\frac{2}{N}\sum_{k=1}^{\frac{N}{2}} g_{1,u}(B_{k,k}) - \mathbf{E}\frac{2}{N}\sum_{k=1}^{\frac{N}{2}} g_{1,u}(B_{k,k})\right)\right| \geq x/2\right]. \tag{2.3.40}$$

By a direct application of Lemma 2.3.21, we can bound (2.3.40). The last part of the proof is completed by a $c-$net argument, completely analogously to (2.3.36). □

Lemma 2.3.24 ([32],Lemma 5.4)**.** *Let $(y_1, y_2, \cdots, y_N)$ be a Gaussian random vector whose covariance matrix is the Id. Fix $\alpha \in (0,2)$, $s \in (0, \alpha/2)$. Moreover, let $\{h_k\}_{k\in[N]} \in (\mathbb{C}^+)^N$ such that $|h_k| \le \eta^{-1}$, for some $\eta > 0$. Then for each $j \in \mathbb{N}$ define the following quantities*

$$f_j(u) = \left(h_j|u\right)^{\alpha/2} |y_j|^\alpha, \quad g_j(u) = \left(h_j|u\right)^{\alpha/2} \mathbf{E}|y_j|^\alpha.$$

Then there exists a constant $C = C(\alpha)$ such that

$$\mathbb{P}\left[\left\| \frac{1}{N}\left(\sum_{j=1}^{N} f_{j+N} - g_{j+N}\right)\right\|_{1-\alpha/2+s} \ge x\right] \le C(\eta^{\alpha/2}x)^{-1/s} \exp\left(-\frac{N(\eta^{\alpha/2}x)^{2/s}}{C}\right).$$

Remark 2.3.25. Due to the deterministic bound (2.3.19), we can apply Lemma 2.3.24 for any number of the diagonal entries of the resolvent of a matrix.

2.3.5 Gaussian and stable random variables

In this subsection we present several results concerning Gaussian random variables and their interaction with the quantities we study.

Lemma 2.3.26. *([5],Lemma 6.4) Let $N \in \mathbb{N}$ and x be a $b-$removal of a $(0,\sigma)$ α-stable distribution, as is defined in Definition 2.2.4. Then let $\hat{X}$ be an $N-$dimensional vector with independent entries all with law $N^{-1/\alpha}x$. Then for any $u \in \mathbb{R}$ and for A a non-negative symmetric matrix and Y an $N-$dimensional centered Gaussian vector with covariance matrix the Id it is true that,*

$$\mathbf{E}\left[\exp\left(-\frac{u^2}{2}\langle A\hat{X}, \hat{X}\rangle\right)\right] = \tag{2.3.41}$$

$$= \mathbf{E}\exp\left(\frac{-\sigma^\alpha|u|^\alpha\|A^{1/2}Y\|_\alpha^\alpha}{N}\right)\exp\left(O(u^2 N^{(2-\alpha)(b-1/\alpha)-1}\log(N)\operatorname{tr}(A))\right) \tag{2.3.42}$$

$$+N\exp\left(-\frac{\log^2(N)}{2}\right). \tag{2.3.43}$$

Lemma 2.3.27. *([5],Lemma 6.5) Let N be a positive integer and let r, d be positive real numbers such that $0 < r < 2 < d \le 4$. Denote $w = (w_1, w_2 \cdots, w_N)$ to be a centered $N-$dimensional Gaussian random variable with covariance matrix $U_{i,j} = \mathbf{E}(w_i w_j)$ for $i, j \in [N]$. Denote $V_j = \mathbf{E}(w_{j,j}^2)$ for each $j \in [N]$ and define*

$$U = \frac{1}{N^2}\sum_{i,j\in[N]} U_{i,j}^2, \quad V = \frac{1}{N}\sum_{j=1}^{N} V_j, \quad X = \sum_{i=1}^{N} V_j^{d/2}, \quad p = \frac{d-r}{d-2}, \quad q = \frac{d-r}{2-r}.$$

Then if $V > 100\log^{10}(N)U^{1/2}$ there exists a constant $C = C(\alpha, r)$:

$$\mathbb{P}\left(\frac{|w|_r^r}{N} < \frac{V^p}{C((X(\log(N))^8)^{p/q}}\right) \le C\exp\left(-\frac{(\log(N))^2}{2}\right).$$

2.3.6 Bounds for the resolvent of X.

Recall the notation R for the resolvent of X and let $X^{(i)}$ is the matrix X with its i–th row and column replaced by 0 vector, as in Definition 2.3.17.

In what follows we will use the following notation $R^{(i)} = (X^{(i)} - z\mathbb{I})^{-1}$ and

$$S_i(z) = \sum_{j\in[2N]\setminus i} X_{i,j}^2 R_{j,j}^{(i)}(z) \text{ and } T_i(z) = X_{i,i} - U_i(z) \text{ where } U_i(z) = \sum_{j\in 2N]\setminus\{i\}} \sum_{k\in[2N]\setminus\{i,j\}} X_{ij} R_{jk}^{(i)}(z) X_{k,i}, \quad i \in [2N]. \tag{2.3.44}$$

For notational convenience, we will omit the dependence of $S_i(z)$, $T_i(z)$ and $U_i(z)$ from z and N, the dimension of the matrix. By the resolvent equality in Lemma 2.3.18 one has that

$$R_{i,i} = \frac{1}{T_i - z - S_i}. \tag{2.3.45}$$

Moreover for each $i \in [2N]$, one has that $\mathrm{Im}(R^{(i)})$ is positive definite, since it is symmetric and by the spectral theorem its eigenvalues are

$$\frac{\eta}{\left(\lambda_j(X^{(i)}) + E\right)^2 + \eta^2} > 0 \;, j \in [2N], \tag{2.3.46}$$

where $\lambda_j(X^{(i)})$ are the eigenvalues of $X^{(i)}$. So it is true that

$$\mathrm{Im}(S_i) \geq 0 \quad \text{and} \quad \mathrm{Im}(S_i - T_i) \geq 0. \tag{2.3.47}$$

In addition, the diagonal entries of the resolvent $R_{i,i}$ are identically distributed. This is proven in the following Lemma.

Lemma 2.3.28. *The random variables $R_{i,i}$, for each $i \in [2N]$, are identically distributed.*

Proof. Note that due to Schur's complement formula it is true that for any $N \times N$ matrices A, B, C, D , if A, D are invertible then

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} = \begin{bmatrix} (A - BD^{-1}C)^{-1} & * \\ * & * \end{bmatrix} = \begin{bmatrix} * & * \\ * & (D - CA^{-1}B)^{-1} \end{bmatrix}.$$

So if one sets $A = D = -z\mathbb{I}$, $C = K$ and $B = K^T$ it is true that $R_{i,i} = z(K^TK - z^2\mathbb{I})^{-1}_{i,i}$ for $i \in [N]$ and $R_{i,i} = z(KK^T - z^2\mathbb{I})^{-1}_{i,i}$ for $i \in [2N] \setminus [N]$. Thus we can conclude that for each $i \in [N]$ the diagonal term $R_{i,i}$ has the same law as $R_{i+N,i+N}$. Moreover, for $i, j \in [N]$ or $i, j \in [2N] \setminus [N]$ it is easy to see that the matrix X retains its law after the permutation of $i - th$ column and row to the $j - th$. All these imply that the diagonal terms $R_{i,i}$ have the same law for each $i \in [2N]$. ◻

Note that the Lemma above would not be true if the dimensions of the matrix, whose symmetrization is X, were not equal.

Moreover since the matrix X has 0 at its diagonal blocks, one may compute that

$$S_1 = \sum_{j=1}^{N} X_{1,N+j}^2 R_{N+j,N+j}^{(i)}, \quad T_1 = -\sum_{j,k\in[2N]\setminus[N]; j\neq k} X_{1,j} X_{k,1} R_{j,k}^{(1)}. \tag{2.3.48}$$

Keep in mind that we want to prove Theorem 2.3.13, so in what follows in this section we will operate under the assumption that (2.3.9) holds.

The following is the analogue of Proposition 7.9 in [5], adjusted to our set of matrices.

Proposition 2.3.29. *For each $i \in [2N]$ there exists a constant $C = C(a, \epsilon, b) > 1$ such that*

$$\mathbb{P}\left(\operatorname{Im}(S_i) < \frac{1}{C(\log(N))^C}\right) \leq C \exp\left(-\frac{(\log(N))^2}{C}\right).$$

Proof. We will prove the estimate for S_1, since $R_{i,i}$ are identically distributed for $i \in [2N]$ due to Lemma 2.3.28.

Set the event:

$$\mathcal{E} = \left\{ \left| \frac{1}{N} \sum_{i=1}^{N} R_{N+j,N+j}^{(1)} - \frac{1}{2} \mathbf{E} R_{2N,2N} \right| \leq \frac{8\log(N)}{(N\eta^2)^{1/2}} + \frac{16}{N\eta} \right\}.$$

By Corollary 2.3.22 and Lemma 2.3.20, one has that $\mathbb{P}(\mathcal{E}^c) \leq 2\exp(-(\log(N))^2)$.

Observe that $\operatorname{Im}(S_1) = \langle A\tilde{X}, \tilde{X}\rangle$, where A is an $N-$dimensional diagonal matrix with entries $A_{j,j} = \operatorname{Im}(R_{N+j,N+j}^{(1)})$ and $\tilde{X}$ is an $N-$dimensional vector with entries $\tilde{X}_j = X_{1,N+j}$. So we can apply Markov inequality for $u = (\log(N)^{2/\alpha}(2\log(2))^{1/2}$ to get that:

$$\mathbb{P}(\operatorname{Im}(S_1) < \mathbf{1}(\mathcal{E})\log(N)^{-4/\alpha}) \leq 2\mathbf{E}(\mathbf{1}(\mathcal{E})\exp\left(-\frac{u^2}{2}\langle A\tilde{X}, \tilde{X}\rangle\right).$$

Next, we can apply Lemma 2.3.26 and after bounding $\operatorname{tr}(A)$ by $C = C'(a, b, \epsilon)$, which we can do since we work on the set $\mathcal{E}$ and since it is true that $\mathbf{E}\operatorname{Im}(R_{1,1}) \leq (\mathbf{E}|R_{1,1}|^2)^{\frac{1}{2}} \leq \epsilon^{-\frac{1}{2}}$ due to our assumption that $\mathbf{E}|R_{1,1}|^2 \leq \epsilon^{-1}$ in (2.3.9). We conclude that

$$\mathbb{P}\left(\operatorname{Im}(S_1) \leq \frac{1}{\log(N)^{4/\alpha}}\right) \leq C'\mathbf{E}\exp\left(-\frac{-\log^2(N)|||A^{1/2}Y|||_\alpha^\alpha}{C'N}\right) + C'\exp\left(-\frac{\log(N)^2}{C'}\right),$$

where Y is a Gaussian vector with covariance matrix the identical, as is mentioned in Lemma 2.3.26. Thus it remains to prove a lower bound for

$$\frac{|||A^{1/2}Y|||_\alpha^\alpha}{N} = \frac{1}{N}\sum_{j=1}^{N} |\operatorname{Im}(R_{N+j,N+j}^{(1)})|^{\alpha/2}|y_j|^\alpha. \tag{2.3.49}$$

Note that for $s \in (0, \frac{\alpha}{2})$ and by Remark 2.3.3

$$\left|\frac{1}{N}\sum_{j=1}^{N} |\operatorname{Im}(R_{N+j,N+j}^{(1)})|^{\alpha/2}|y_j|^\alpha - \frac{1}{N}\sum_{j=1}^{N} |\operatorname{Im}(R_{N+j,N+j}^{(1)})|^{\alpha/2}\mathbf{E}|y_j|^\alpha\right| \tag{2.3.50}$$

$$\leq \sup_{u\in S_+^1}\left|\frac{1}{N}\sum_{j=1}^{N} |(-iR_{N+j,N+j}^{(1)}|u)|^{\alpha/2}|y_j|^\alpha - \frac{1}{N}\sum_{j=1}^{N} |(-iR_{N+j,N+j}^{(1)}|u)|^{\alpha/2}\mathbf{E}|y_j|^\alpha\right| \tag{2.3.51}$$

$$\leq \left|\frac{1}{N}\sum_{j=1}^{N} |(-iR_{N+j,N+j}^{(1)}|u)|^{\alpha/2}|y_j|^\alpha - \frac{1}{N}\sum_{j=1}^{N} |(-iR_{N+j,N+j}^{(1)}|u)|^{\alpha/2}\mathbf{E}|y_j|^\alpha\right|_{1-\alpha/2+s} \tag{2.3.52}$$

So we can apply Lemma 2.3.24 for

$$x = C\frac{\log^4 N}{N^{\alpha/4}\eta^{\alpha/2}},$$

and $s = \frac{\alpha}{4}$ to get that the inequality

$$\left|\frac{1}{N}\sum_{j=1}^{N} |\operatorname{Im}(R^{(1)}_{N+j,N+j})|^{\alpha/2}|y_j|^{\alpha} - \frac{1}{N}\sum_{j=1}^{N} |\operatorname{Im}(R^{(1)}_{N+j,N+j})|^{\alpha/2}\mathbf{E}|y_j|^{\alpha}\right| \leq x \tag{2.3.53}$$

holds with probability at least $1 - C\exp(-\frac{\log^2 N}{C})$. Thus, it is sufficient to give a lower bound to

$$\frac{1}{N}\sum_{j=1}^{N} |\operatorname{Im}(R^{(1)}_{N+j,N+j})|^{\alpha/2}\mathbf{E}|y_j|^{\alpha}. \tag{2.3.54}$$

in order to obtain a lower bound for the quantity in (2.3.49).

Next we apply again Lemma 2.3.20 for $r = \frac{\alpha}{2}$, and since for any $u_1, u_2 \in \mathbb{R}^+$ and $r \in (0,1]$ it is true that $|u_1^r - u_2^r| \leq |u_1 - u_2|^r$, we obtain that

$$\mathbf{E}|y_1|^{\alpha}\frac{1}{N}\sum_{j=1}^{N} |\operatorname{Im}(R_{N+J,N+J})^{\alpha/2} - \operatorname{Im}(R^{(1)}_{N+j,N+j})^{\alpha/2}| \leq \frac{4\mathbf{E}|y_1|^{\alpha}}{(\eta N)^{\alpha/2}}.$$

So we have concluded that the event that

$$\frac{\||A^{1/2}Y\||_{\alpha}^{\alpha}}{N} \geq O\left(\frac{1}{(\eta N)^{\alpha/2}}\right) + O\left(\frac{1}{(\eta N^{1/2})^{\alpha/2}}\right) + C''\frac{1}{N}\sum_{j=1}^{N}\operatorname{Im}(R_{N+j})^{\alpha/2},$$

holds with probability at least $1 - C\exp\left(-\frac{\log^2(N)}{N}\right)$. Restricting again on the set $\mathcal{E}$ and using the concentration inequality (2.3.25) and our hypothesis 2.3.9, one can conclude that there exists $C = C(\alpha, \epsilon, b)$ such that

$$\mathbb{P}\left(\frac{\|A^{1/2}Y\||_{\alpha}^{\alpha}}{CN} \leq \epsilon\right) \leq C\exp\left(\frac{-\log^2(N)}{C}\right),$$

which finishes the proof. □

The following is the analogue of Proposition 7.10 in [5], adjusted to our set of matrices.

Proposition 2.3.30. *For each $i \in [2N]$ there exists a constant $C = C(\alpha, \epsilon, b) > 1$ such that*

$$\mathbb{P}\left(\operatorname{Im}(S_i - T_i) < \frac{1}{C(\log(N))^C}\right) \leq C\exp\left(-\frac{(\log(N))^2}{C}\right). \tag{2.3.55}$$

Moreover

$$\mathbb{P}\left(\max_{j\in[2N]}|R_{j,j}| > C\log^C(N)\right) \leq C\exp\left(-\frac{(\log(N))^2}{C}\right). \tag{2.3.56}$$

Proof. By construction, one can prove that for $A = \{\mathrm{Im}(R^{(1)}_{i,j})\}_{i,j\in[2N]\setminus[N]}$ and $\tilde{X} = \{X_{1,N+j}\}_{j\in[N]}$ it is true that,

$$\mathrm{Im}(S_1 - T_1) = \langle A\tilde{X}, \tilde{X}\rangle.$$

So after applying Lemma 2.3.26, like in Proposition 2.3.29, one has that

$$\mathbb{P}\left(\mathrm{Im}(S_1 - T_i) < \frac{1}{\log^{4/\alpha}(N)}\right) \leq C\mathbf{E}\exp\left(-\frac{C\log^2(N)\|A^{1/2}Y\|_\alpha^\alpha}{N}\right) + C\exp\left(-\frac{\log^2(N)}{C}\right),$$

where Y is again a centered $N-$dimensional Gaussian random variable with covariance matrix equal to the identical.

Next, we want to apply Lemma 2.3.27 in order to establish a lower bound for $\frac{1}{N}\|A^{1/2}Y\|_\alpha^\alpha$. Following the notation of Lemma 2.3.27 set

$$w_i = (A^{1/2}Y)_i, \quad V_j = \mathrm{Im}(R^{(1)}_{j,j}), \quad U_{j,k} = \mathrm{Im}\, R^{(1)}_{j,k}(z),$$

$$X' = \frac{1}{N}\sum_{i=1}^{N} V^{\alpha/2}_{N+j,N+j}, \quad U = \frac{1}{N^2}\sum_{i,j\in[2N]\setminus[N]U_{i,j}} U_{i,j}, \quad r = \alpha, \quad d = 2 + \epsilon. \tag{2.3.57}$$

So one may apply Lemma 2.3.16 and Lemma 2.3.18 to get that

$$U \leq \frac{4}{N^2}\sum_{i,j\in[2N]} U_{i,j}^2 \leq \frac{4}{N^2}\sum_{i,j\in[2N]} |\,\mathrm{Im}(R^{(1)}_{ij})|^2 \leq \frac{4}{N^2\eta}\sum_{j=1}^{2N}\mathrm{Im}(R^{(1)}_{j,j}) \leq \frac{4}{N\eta^2}. \tag{2.3.58}$$

Next, we can approximate $V = \frac{1}{N}\sum_{j=1}^{N} V_{N+j}$ by $\frac{1}{N}\sum_{j=1}^{N}\mathrm{Im}(R_{N+j,N+j})$ due to the deterministic bound in Lemma 2.3.20 and then approximate $\frac{1}{N}\sum_{j=1}^{N}\mathrm{Im}(R_{N+j,N+j})$ by $\frac{1}{2}\mathbf{E}\,\mathrm{Im}(R_{1,1})$ due to Corollary 2.3.16, on an event which holds with probability at least $1 - 2\exp\left(-\frac{\log^2(N)}{8}\right)$. The approximation procedure described above is identical to the similar approximation described in Proposition 2.3.29. So after taking into account the Hypothesis (2.3.9), we have that

$$\mathbf{E}\,\mathrm{Im}(R_{1,1}) \geq \left(\mathbf{E}[\mathrm{Im}(R_{1,1})]^{\alpha/2}\right)^{2/\alpha} \geq \epsilon^{2/\alpha}.$$

Thus, it is implied that

$$\mathbb{P}\left(\frac{|V|}{C} < 1\right) < C\exp\left(-\frac{\log^2 N}{C}\right). \tag{2.3.59}$$

So after combining (2.3.58) and (2.3.59), we get that for sufficient large N it is true that

$$\mathbb{P}\left(|V| \leq 100\log^{10}(N)U^{1/2}\right) \leq C'\exp\left(-\frac{\log^2(N)}{C}\right).$$

Next we need to bound X' from (2.3.57). Note that again we can apply Lemma 2.3.20 to get that

$$\begin{aligned}\left|\frac{1}{N}\sum_{j=1}^{N}\mathrm{Im}(R_{j+N,j+N}) - X'\right| &\leq \frac{1}{N}\sum_{j=1}^{N}|\,\mathrm{Im}(R_{j+N,j+N})^{\alpha/2} - \mathrm{Im}(R^{(1)}_{N+j,N+j})^{\alpha/2}| \leq \\ &\leq \frac{1}{N}\sum_{j=1}^{2N}|\,\mathrm{Im}(R_{j,j}) - \mathrm{Im}(R^{(1)}_{j,j})|^{\alpha/2} \leq \sum_{j=1}^{2N}|R_{j,j} - R^{(1)}_{j,j}|^{\alpha/2} \leq \frac{4}{(N\eta)^{\alpha/2}}.\end{aligned} \tag{2.3.60}$$

Moreover, since the function $f(y) = \mathbf{1}\{|\operatorname{Im}(y)| \le \eta\} \operatorname{Im}(y)^{\alpha/2} + \mathbf{1}\{|\operatorname{Im}(y)| \ge \eta\} \eta^{\alpha/2}$ is Lipschitz with Lipschitz-constant $L = \alpha\eta^{1-\alpha/2}$, we can apply Lemma 2.3.21 for $x = N^{-1/2}\eta^{\alpha/2}\log(N)$ to get that

$$\mathbb{P}\left(\left|\frac{1}{N}\sum_{j=1}^{N} |\operatorname{Im}(R_{N+j,N+j})|^{\alpha/2} - \frac{1}{N}\mathbf{E}\,|\operatorname{Im}(R_{N+j,N+j})|^{\alpha/2}\right| \ge \frac{\log(N)}{N^{1/2}\eta^{\alpha/2}}\right) \le 2\exp\left(-\frac{\log^2(N)}{8\alpha^2}\right). \tag{2.3.61}$$

So after combining (2.3.61) ,(2.3.60) with (2.3.9) and specifically with the fact that

$$\mathbf{E}(|R_{j,j}|^{\alpha/2}) \le \mathbf{E}(|R_{j,j}|^2)^{\alpha/4} \le \frac{1}{\epsilon^{\alpha/4}}$$

we get that,

$$\mathbb{P}\left(|X'| > C\right) \le C\exp\left(-\frac{\log^2(N)}{C}\right), \tag{2.3.62}$$

for sufficient large universal constant C. So the bounding for $\frac{1}{N}\|A^{1/2}Y\|_\alpha^\alpha$ comes from a direct application of Lemma 2.3.27 with the bounding for V and X' proven in (2.3.58) and (2.3.62)

Note that (2.3.56) is a corollary of (2.3.55) and (2.3.45). □

The following is the analogue of Proposition 5.9 in [5], adjusted to our set of matrices.

Lemma 2.3.31. *There exists some constant $C = C(\alpha)$ such that for any $x \ge 1$ and for any $i \in [2N]$, it is true that*

$$\mathbb{P}\left[|T_i| \ge \frac{Cx}{(N\eta^2)^{1/2}}\right] \le \frac{C}{x^{\alpha/2}} \tag{2.3.63}$$

Proof. It is sufficient to prove it only for $i = 1$. Recall the definition of T_1 in (2.3.44). So we need to prove that

$$\mathbb{P}\left(|U_1| \ge \frac{x}{(N\eta^2)^{1/2}}\right) \le \frac{C}{x^{\alpha/2}}. \tag{2.3.64}$$

Set the event

$$\Omega_1(s) = \cap_{N+1\le j\le 2N}\{|X_{1,j}| \le s\}. \tag{2.3.65}$$

Then

$$\mathbb{P}\left(|U_1| \ge \frac{x}{(N\eta^2)^{1/2}}\right) \le \mathbb{P}\left(\mathbf{1}(\Omega_1(s))\,|U_1| \ge \frac{x}{(N\eta^2)^{1/2}}\right) + \mathbb{P}(\Omega_1^c(s)) \tag{2.3.66}$$

For the second summand in the right-hand-side of (2.3.66) by the union bound and (2.1.3) one has that

$$\mathbb{P}(\Omega_1^c(s)) \le \sum_{j=N+1}^{2N} \mathbb{P}(|X_{1,j}| \ge s) \le \frac{C}{s^\alpha} \tag{2.3.67}$$

For the first term on the right hand-side of (2.3.66) note that by Markov's inequality, the independence of $\{X_{1,j}\}_{j\in[2N]\setminus[N]}$ and $R^{(1)}$ and the symmetry of the random variables $X_{j,1}$

$$\mathbb{P}\left(\mathbf{1}\left(\Omega_1(s)\right)|U_1| \geq \frac{x}{(N\eta^2)^{1/2}}\right) \leq \frac{N\eta^2}{x^2}\mathbf{E}\left|\sum_{k,j\in 2N]\setminus[N]:k\neq j} X_{1j}R^{(1)}_{jk}(z)X_{k,1}\right|^2 \mathbf{1}\left(\Omega_1(s)\right) \tag{2.3.68}$$

$$= \frac{2N\eta^2}{x^2}\sum_{k,j\in[2N]\setminus[N]:k\neq j} \mathbf{E}|R^{(1)}_{j,k}|^2\mathbf{E}(X^2_{1,j}\mathbf{1}\left(\Omega_1(s)\right))^2. \tag{2.3.69}$$

We will bound each of the terms inside the sum in (2.3.69) individually. Firstly

$$\mathbf{E}X^2_{1,j}\mathbf{1}\left(\Omega_1(s)\right) \leq \mathbf{E}X^2_{1,j}\mathbf{1}\left\{|X_{1,j}| \leq s\right\} \leq \frac{2Cs^{2-\alpha}}{(2-\alpha)N}. \tag{2.3.70}$$

The last inequality in (2.3.70) can be found in the proof of Proposition 5.9 [5].

Moreover, due to (2.3.21) and (2.3.19) one has that

$$\sum_{k,j\in[2N]\setminus[N]:k\neq j} \mathbf{E}|R^{(1)}_{j,k}|^2 < \sum_{j=1}^{N} \frac{\operatorname{Im}(R^{(1)}_{j+N,j+N})}{\eta} \leq \frac{N}{\eta^2}. \tag{2.3.71}$$

Thus combining (2.3.71), (2.3.70) and (2.3.67) we get that for some absolute constant $C = C(\alpha)$ it is true that

$$\mathbb{P}\left(|U_1| \geq \frac{x}{(N\eta^2)^{1/2}}\right) \leq \frac{Cs^{4-2\alpha}}{x^2} + \frac{C}{s^\alpha}. \tag{2.3.72}$$

Setting $s = x^{1/2}$, we get (2.3.64). □

2.3.7 Proof of Theorem 2.3.13

In order to prove Theorem 2.3.13, we wish to replace the entries of X by $\alpha-$stable entries in several quantities, for example in quantities defined in (2.3.48), in order to use the properties of the $\alpha-$stable distribution.

Firstly consider the following

Definition 2.3.32. Define the following quantities:

$$\omega_z(u)^{(i)} = \Gamma\left(1-\frac{\alpha}{2}\right)(iz - i\,S_i\,|u)^{\alpha/2}, \quad \bar{\omega}_z(u) = \mathbf{E}\,\omega_z(u)^{(i)},$$

$$G_i = \sum_{j:|j-i|\geq N} Z_{j,j}R^{(i)}_{j,j}, \quad \Psi^{(i)}_z(u) = \Gamma\left(1-\frac{\alpha}{2}\right)(iz - i\,G_i\,|u)^{\alpha/2}, \quad \psi_z(u) = \mathbf{E}(\Psi_z(u)).$$

Here $Z_{j,j}$ are i.i.d. random variables from the definition of the matrix D_N, all with law $N^{-1/\alpha}Z$ where Z is a $(0,\sigma)$ a-stable random variable as in Definition 2.1.1 .

We start this subsection with a comparison between $(-z - S_i)^{-1}$ and $R_{i,i}$.

Lemma 2.3.33. *For any $p > 0$ there exists a constant $C = C(a, \epsilon, b, s, p)$ such that*

$$\left|\mathbf{E}|R_{i,i}|^p - |(-S_i - z)^{-1}|^p\right| \le \frac{C\log^C(N)}{(N\eta^2)^{a/8}} , \tag{2.3.73}$$

$$\left|\mathbf{E}|(-iR_{i,i})|^p - \mathbf{E}|(-iz - iS_i)|^{-p}\right| \le \frac{C\log^C(N)}{(N\eta^2)^{a/8}}, \tag{2.3.74}$$

$$|\gamma_z - \bar{\omega}_z|_{1-a/2+s} \le \frac{C\log^C(N)}{(N\eta^2)^{a/8}}. \tag{2.3.75}$$

Proof. Let C_1, C_2, C_3 the constants from Propositions 2.3.29, 2.3.30 and Lemma 2.3.31 respectively and set $C = \max\{C_1, C_2, C_3\}$. Moreover let E_1, E_2 the events whose probability we bound in Proposition 2.3.29 and 2.3.30 respectively and set $E = E_1 \cup E_2$.

Note that due to our assumptions in (2.3.9), (2.3.19) and (2.3.47) it is true that

$$\frac{1}{\mathrm{Im}(S_i - T_i + z)} \le N^{1/2} , \frac{1}{\mathrm{Im}(S_i + z)} \le N^{1/2}. \tag{2.3.76}$$

Furthermore by (5.5) in [5] one has that for any $u > 0$

$$\left||R_{i,i}|^p - |(-S_i - z)^{-1}|^p\right| \tag{2.3.77}$$

$$\le \mathbf{1}\{|T_i| < u\}(p-1)u\left(\left|\frac{1}{\mathrm{Im}(S_i - T_i + z)}\right|^{p+1} + \left|\frac{1}{\mathrm{Im}(S_i + z)}\right|^{p+1}\right) \tag{2.3.78}$$

$$+ \mathbf{1}\{|T_i| \ge u\}\left(\left|\frac{1}{\mathrm{Im}(S_i - T_i + z)}\right|^{p} + \left|\frac{1}{\mathrm{Im}(S_i + z)}\right|^{p}\right) \tag{2.3.79}$$

So one by Propositions 2.3.29 and 2.3.30 one has that

$$\mathbf{E}\mathbf{1}(E^c)\left||R_{i,i}|^p - |(-S_i - z)^{-1}|^p\right| \le 2u(p-1)C^{p+1}\log^{C(p+1)} N + 2\mathbb{P}(|T_i| \ge u)C^p\log^{Cp}(N) \tag{2.3.80}$$

$$\mathbf{E}\mathbf{1}(E)\left||R_{i,i}|^p - |(-S_i - z)^{-1}|^p\right| \le 2uN^{(p+1)/2}\exp\left(\frac{-\log^2 N}{C}\right) \tag{2.3.81}$$

So after setting $u = (N\eta^2)^{-1/4}$ and applying Lemma 2.3.31, we get (2.3.73).

The proof of (2.3.74) is analogous and therefore it is omitted.

For the proof of (2.3.75) note that

- By (2.3.5) applied for $x_1 = (iT_i - iS_i - iz)^{-1}, x_2 = (-iS_i - iz)^{-1}$ and for $r = \frac{a}{2}$ and $\eta = (2C\log^{2C} N)^{-1}$, we get that there exists a constant $C' = C'(a) > 0$ such that for any $u > 0$ it is true that

$$\mathbf{1}(E^c)\mathbf{1}\{|T_i| < u\}|i_z - \omega_z|_{1-a/2+s} \tag{2.3.82}$$

$$\le C'\left(2C\log^{2C}(N)\right)^{\frac{a}{2}}\mathbf{1}(E^c)\mathbf{1}\{|T_i| < u\}\left||z - T_i + S_i|^{-1} - |z + S_i|^{-1}\right|^{\frac{a}{2}} \tag{2.3.83}$$

$$\le uC'\left(2C\log^{2C}(N)\right)^{\frac{3a}{2}}\mathbf{1}(E^c), \tag{2.3.84}$$

$$\mathbf{1}(E^c)\mathbf{1}\{|T_i| \ge u\}|i_z - \omega_z|_{1-a/2+s} \tag{2.3.85}$$

$$\le C'\left(2C\log^{2C}(N)\right)^{\frac{a}{2}}\mathbf{1}(E^c)\mathbf{1}\{|T_i| \ge u\}\left||z - T_i + S_i|^{-1} - |z + S_i|^{-\frac{a}{2}}\right| \tag{2.3.86}$$

$$\le 2C'\left(2C\log^{2C}(N)\right)^{\frac{3a}{2}}\mathbf{1}\{|T_i| \ge u\}. \tag{2.3.87}$$

- Moreover again by (2.3.5) for the same x_1, x_2 and r as before and for $\eta = N^{-1/2}$ there exists a constant $C' = C(a)$ such that

$$\mathbf{1}(E)\,|i_z - \omega_z|_{1-a/2+s} \le 2C'\,\mathbf{1}(E)\,N^{a/4} \tag{2.3.88}$$

Note that by definition $\mathbf{E}\omega_z = \bar{\omega}_z$ and $\mathbf{E}i_z = \gamma_z$. So after summing (2.3.85), (2.3.82) and (2.3.88), taking expectation and applying Propositions 2.3.29 and 2.3.30 and Lemma 2.3.31 for $x = (N\eta^2)^{1/4}$, we get (2.3.75).

□

Fixed point equation

In this subsection we establish the asymptotic fixed point equation. Firstly, we show that the quantities in Definition 2.3.32 are approximately equal to the respective quantities of the Stieltjes transform, i.e., the quantities defined in Definition 2.3.12. The latter is proven in the following proposition.

Proposition 2.3.34. *It is true that for any $p \in \mathbb{N}$,*

$$\left|\mathbf{E}|R_{i,i}|^p - \mathbf{E}|(-z - G_i)|^{-p}\right| \le \frac{C\log^C(N)}{(N\eta^2)^{a/8}} + \frac{C\log^C(N)}{N^{4\partial}}, \tag{2.3.89}$$

$$\left|\mathbf{E}|(-iR_{i,i})|^p - \mathbf{E}|(-iz - iG_i)|^{-p}\right| \le \frac{C\log^C(N)}{(N\eta^2)^{a/8}} + \frac{C\log^C(N)}{N^{4\partial}} \tag{2.3.90}$$

and

$$|\gamma_z - \psi_z|_{1-a/2+s} \le \frac{C\log^C(N)}{(N\eta^2)^{a/8}} + \frac{C\log^C(N)}{N^{4\partial}}. \tag{2.3.91}$$

Proof. We first present two facts.

- One can show that there exists $C = C(a) > 0$ such that

  $$\mathbb{P}\left(|S_1 - G_1| \ge N^{-4\partial}\right) \le C(1 + \mathbf{E}(R_{1,1}))N^{-4\partial},$$

  similarly to the proof of Lemma 6.8 in [5]. As a result, by Assumption (2.3.9) we have that

  $$\mathbb{P}(|S_1 - G_1| \ge N^{-4\partial}) \le CN^{-4\partial}, \tag{2.3.92}$$

  for some constant $C = C(a, \epsilon)$.

- For each $i \in [2N]$ there exists a constant $C = C(a, \epsilon, b) > 1$ such that

  $$\mathbb{P}\left(\operatorname{Im}(G_i) < \frac{1}{C(\log(N))^C}\right) \le C\exp\left(-\frac{(\log(N))^2}{C}\right). \tag{2.3.93}$$

  The proof of (2.3.93) is completely analogous to the proof of Proposition 2.3.29, after replacing the usage of Lemma 2.3.26 with Lemma B.1 in [31]. Therefore it is omitted.

Moreover note that due to Lemma 2.3.33, it is sufficient to prove that for any $p \in \mathbb{N}$,

$$\left|\mathbf{E}| - z - S_{i,i}|^{-p} - \mathbf{E}|(-z - G_i)|^{-p}\right| \le \frac{C \log^C(N)}{N^{4\partial}}, \tag{2.3.94}$$

$$\left|\mathbf{E}|(-iz - iS)|^{-p} - \mathbf{E}|(-iz - iG_i)|^{-p}\right| \le \frac{C \log^C(N)}{N^{4\partial}} \tag{2.3.95}$$

$$|\bar{\omega}_z - \psi_z|_{1-\alpha/2+s} \le \frac{C \log^C(N)}{N^{4\partial}}. \tag{2.3.96}$$

Given (2.3.92) and (2.3.93), the proof of (2.3.94), is completely analogous to the proof of Lemma 2.3.33, therefore it is omitted. □

Moreover, we have the following results which will be used in order to establish the limiting fixed point equation. The following Lemma will be the basis for the approximation of the fixed point equation.

Lemma 2.3.35. *Recall Definition 2.3.5. It is true that,*

$$\Psi_z(u) = \mathbf{E}_{\mathcal{D}}(Y_\zeta(u)), \tag{2.3.97}$$

where Y_ζ is as in Definition 2.3.5, $\mathcal{D} = \{y_i\}_{i\in[N]}$ is an $N-$dimensional Gaussian random variable independent from any other quantity with covariance matrix being the identical, $\mathbf{E}_{\mathcal{D}}$ denotes the expectation with respect to the random variable D and

$$\zeta(u) = \frac{1}{N}\sum_{j=1}^{N}\left(-iR^{(1)}_{N+j,N+j}|u\right)^{\alpha/2}\frac{|y_j|^\alpha}{\mathbf{E}|y_j|^\alpha}.$$

Also,

$$\mathbf{E}(-iz - iG_1)^{-p} = \mathbf{E}_{\mathcal{D}} s_{p,z}(\zeta(1)), \quad \mathbf{E}| - z - G_1|^{-p} = \mathbf{E}_{\mathcal{D}} r_{p,z}(\zeta(1)).$$

Proof. This Lemma is a corollary of [[32] Corollary 5.8] □

So in Proposition 2.3.34, we manage to approximate the quantities involving G_1, such as $y_z(u)$, by the analogous quantities involving $R_{1,1}$, such as $\gamma_z(u)$. In order to establish the asymptotic fixed point equation, we will need to approximate the function $\zeta(u)$ mentioned in Lemma 2.3.35 by $\gamma_z(u)$ and then take advantage of (2.3.97). This approximation is done via the following Lemma.

Lemma 2.3.36. *There exists a constant $C = C(\alpha, \epsilon, s) > 1$ such that*

$$\mathbb{P}\left(|\zeta - \gamma_z|_{1-\alpha/2+s} > \frac{C \log^C(N)}{N^{s/2}\eta^{\alpha/2}}\right) \le C \exp\left(-\frac{\log^2(N)}{C}\right). \tag{2.3.98}$$

Proof. Firstly note that ζ is close to $\mathbf{E}_D\zeta$ with high probability due to Lemma 2.3.24 for appropriate x, i.e.,

$$\mathbb{P}\left(|\zeta - \mathbf{E}_D\zeta|_{1-\alpha/2+s} \ge \frac{\log^s(N)}{N^{s/2}\eta^{\alpha/2}}\right) \le C \exp\left(-\frac{\log^2 N}{C}\right), \tag{2.3.99}$$

for some constant $C = C(a)$. Next, note that by Lemma 2.3.23 applied for the matrix $X^{(1)}$ and for appropriate x one has that

$$\mathbb{P}\left(|\mathbf{E}_{X^{(1)}}\mathbf{E}_D\zeta - \mathbf{E}_D\zeta|_{1-a/2+s} \geq \frac{\log^s(N)}{N^{s/2}\eta^{a/2}}\right) \leq C\exp\left(-\frac{\log^2 N}{C}\right), \tag{2.3.100}$$

for some appropriately chosen constant $C = C(a)$. Here $\mathbf{E}_{X^{(1)}}$ denotes the mean value with respect to the law of the matrix $X^{(1)}$. Next by Lemma 2.3.4 one has that

$$\begin{aligned}&\sum_{i=1}^{N}\frac{1}{N}|R_{N+i,N+i} - R^{(1)}_{N+i,N+i}|_{1-a/2+s} \\ &\leq C\eta^{-a/2}\frac{1}{N}\sum_{i=1}^{N}\left(|R_{N+i,N+i} - R^{(1)}_{N+i,N+i|}|^{a/2} + \eta^s|R_{N+i,N+i} - R^{(1)}_{N+i,N+i}|^s\right).\end{aligned} \tag{2.3.101}$$

So after applying Lemma 2.3.20 and since $R_{j,j}$ are identical distributed, one has the deterministic bound

$$|\mathbf{E}_{X^{(1)}}\mathbf{E}_D\zeta - \gamma_z| \leq C'\left(\frac{1}{\eta^a N^{a/2}} + \frac{1}{N^s\eta^{a/2}}\right). \tag{2.3.102}$$

So after combining (2.3.99) ,(2.3.100) and (2.3.102), we get the desired inequality. □

Next, we give some more approximating results.

Corollary 2.3.37. *There exists a constant $C = C(a, \epsilon, s) > 0$ such that*

$$|\gamma_z|_{1-a/2+s} < C, \quad \inf_{u\in S^1_+}\mathrm{Re}(\gamma_z(u)) > \frac{1}{C}, \tag{2.3.103}$$

$$\mathbb{P}\left(\inf_{u\in S^1_+}\zeta(u) < \frac{1}{C}\right) < C\exp\left(-\frac{\log^2(N)}{C}\right), \tag{2.3.104}$$

Proof. By (2.3.98), the estimate in (2.3.104) is a consequence of (2.3.103).

For (2.3.98) note that due to the first estimate in (2.3.5) one has that there exists a constant $C = C(s)$ such that

$$\left|(-iR_{i,i}|u)^{a/2}\right|_{1-\frac{a}{2}+s} \leq C|R_{i,i}|^{a/2}. \tag{2.3.105}$$

By integrating (2.3.105) and by the definition of γ_z in Definition 2.3.12 one has that,

$$|\gamma_z|_{1-\frac{a}{2}+s} \leq C\Gamma\left(1-\frac{a}{2}\right)\mathbf{E}|R_{i,i}|^{a/2} \leq C\Gamma\left(1-\frac{a}{2}\right)\left(\mathbf{E}|R_{i,i}|^2\right)^{a/4} \leq \epsilon^{-a/4}C\Gamma\left(1-\frac{a}{2}\right). \tag{2.3.106}$$

Where in the first inequality in (2.3.106) we used (2.3.105), in the second we used Holder's inequality and in the third we used our Assumption 2.3.9. So the first estimate in (2.3.103) is proven.

For the second estimate in (2.3.103) one has that for any $u \in S^1_+$

$$\mathrm{Re}\,\gamma_z(u) = \Gamma\left(1-\frac{a}{2}\right)\mathbf{E}\,\mathrm{Re}(iR_{i,i}|u)^{a/2} \geq \Gamma\left(1-\frac{a}{2}\right)\mathbf{E}\left((\mathrm{Re}(iR_{i,i}|u))^{a/2}\right) \tag{2.3.107}$$

$$\geq \Gamma\left(1-\frac{a}{2}\right)\mathbf{E}(\mathrm{Im}\,R_{i,i})^{a/2} \geq \Gamma\left(1-\frac{a}{2}\right)\epsilon \tag{2.3.108}$$

where in the first inequality in (2.3.107) we used the fact that $\mathrm{Re}\, c^r \geq (\mathrm{Re}\, c)^r$ for any $c \in \mathbb{K}^+$ and $r \in (0,1)$, see the proof of Lemma 7.18 in [5], in the second inequality we used the fact that $\mathrm{Re}(c|u) \geq \mathrm{Re}(c)$ for any $c \in \mathbb{K}^+$ and $u \in \mathbb{S}^1_+$ and in the third we used our Assumption 2.3.9. Thus the second estimate in (2.3.103) is proven.

□

Before presenting the proof of Theorem 2.3.13, we need a last approximation result.

Lemma 2.3.38. *There exists a constant $C = C(a, \epsilon, s)$ such that*

$$\mathbb{P}\left(|\psi_z - Y_{\gamma_z}|_{1-a/2+s} > \frac{C\log^C(N)}{N^{s/2}\eta^{a/2}}\right) < C\exp\left(-\frac{\log^2(N)}{C}\right). \tag{2.3.109}$$

Proof. The strategy of the proof is firstly to approximate Y_{γ_z} by Y_ζ and then use Lemma 2.3.35.

- For the approximation of Y_{γ_z} and Y_ζ: Let C_1, C_2 be the constants mentioned in Lemma 2.3.36 and Corollary 2.3.37. Set $C = 2\max\{C_1, C_2\}$. Moreover define the following sets

$$E_1 = \left\{|\zeta - \gamma_z|_{1-a/2+s} > \frac{C\log^C(N)}{N^{s/2}\eta^{a/2}}\right\} \tag{2.3.110}$$

$$E_2 = \left\{\inf_{u\in S^1_+} \mathrm{Re}\,\zeta(u) < \frac{1}{C}\right\} \tag{2.3.111}$$

By Lemma 2.3.36 and Corollary 2.3.37 one has that

$$\mathbb{P}(E_1 \cup E_2) \leq C\exp\left(-\frac{\log^2 N}{C}\right) \tag{2.3.112}$$

Set F the complement event of $E_1 \cup E_2$.

So

$$\mathbf{1}(F)\left|Y_\zeta - Y_{\gamma_z}\right|_{1-\frac{a}{2}+s} \leq \mathbf{1}(F)\, C_1|\zeta - \gamma_z|_{1-\frac{a}{2}+s}\left(1 + |\gamma_z|_{1-\frac{a}{2}+s} + |\zeta|_{1-\frac{a}{2}+s}\right) \tag{2.3.113}$$

$$\leq \mathbf{1}(F)\,\frac{C\log^C(N)}{N^{s/2}\eta^{a/2}}\left(1 + \frac{2}{C}\right) \tag{2.3.114}$$

where in the first inequality of (2.3.113) we used Lemma 2.3.10 and Remark 2.3.3 (C_1 is the constant mentioned in Lemma 2.3.10) and the fact that $\gamma_z, \zeta\mathbf{1}(F) \in H^{1/C}_{\frac{a}{2},1-\frac{a}{2}+s}$ by Corollary 2.3.37 and the definition of the set F. For the second inequality we used again the definition of F, Corollary 2.3.37 and Lemma 2.3.36.

Now working on the event $E_1 \cup E_2$ we get that by Lemma 2.3.9 and Corollary 2.3.37 we there exists a constant $C' > 0$ such that

$$\mathbf{1}(E_1 \cup E_2)|Y_{\gamma_z}|_{1-\frac{a}{2}+s} \leq C'\eta^{-\frac{a}{2}}(1 + C)\mathbf{1}(E_1 \cup E_2) \tag{2.3.115}$$

- Note that similarly to the proof of (2.3.76) one can prove that

$$\left|\frac{1}{G_i + z}\right| \leq \frac{1}{\eta} \tag{2.3.116}$$

Thus, we can apply (2.3.5) to get that there exists a constant $C = C(a)$ such that

$$|\Psi_z|_{1-\frac{a}{2}+s} \leq C\eta^{-a/2} \tag{2.3.117}$$

So by Lemma 2.3.35, one has that

$$\begin{aligned} &|\psi_z - Y_{\gamma_z}|_{1-\frac{a}{2}+s} \leq \\ &\mathbf{E1}(F)\,|Y_\zeta - Y_{\gamma_z}|_{1-\frac{a}{2}} + \mathbf{E1}(E_1 \cup E_2)\,|\Psi_z|_{1-\frac{a}{2}+s} + \mathbf{E1}(E_1 \cup E_2)\,|Y_{\gamma_z}|_{1-\frac{a}{2}+s} \end{aligned} \tag{2.3.118}$$

Now (2.3.109) is proven by combining (2.3.112), (2.3.117), (2.3.115), (2.3.118) and (2.3.113). □

Next, the proof of the main theorem of this subsection is presented.

Proof of Theorem 2.3.13. Note that, (2.3.10) is a consequence of (2.3.90) and (2.3.109). Additionally (2.3.14) is already proven in (2.3.56). Lastly, note that (2.3.13) is a consequence of (2.3.103). So all that remains is to establish (2.3.11) and (2.3.12) in order to complete the proof. We will prove only (2.3.12). The proof of (2.3.11) is similar and will be omitted.

To that end, define the sets E_1 and E_2 as in (2.3.112) and F the complement event of $E_1 \cup E_2$. So by the first estimate in (2.3.8) and Remark 2.3.3 one has that

$$\mathbf{1}(F)\,|r_{p,z}(\zeta) - r_{p,z}(\gamma_z)| \leq \mathbf{1}(F)\,C'|\gamma_z - \zeta| \tag{2.3.119}$$

for some constant C'. By the definition of the event F and Lemma 2.3.35, we get the bound in (2.3.12) on the event F.

On the event $E_1 \cup E_2$ we can use the deterministic bound in Lemma 2.3.9 to get that

$$\mathbf{1}(E_1 \cup E_2)\,|r_{p,z}(\zeta) - r_{p,z}(\gamma_z)| \leq 2C''\eta^{-p}\mathbf{1}(E_1 \cup E_2) \tag{2.3.120}$$

for some other constant C''. Now the bound in (2.3.12) on the event $\mathbf{1}(E_1 \cup E_2)$ is a consequence of (2.3.112) and (2.3.120). □

2.4 Universality for the least singular value after short time

At this section, universality of the least eigenvalue for the matrices $X + \sqrt{t}W$ is proven. More precisely :

Theorem 2.4.1. *Let L_N be an $N \times N$ matrix with i.i.d. entries all following the Gaussian distribution with mean 0 and variance $\frac{1}{N}$, independent from H_N. Then denote W be the symmetrization of L_N. Let $\tilde{W}$ be an independent copy of W. Moreover for every matrix Y*

denote $\lambda_N(Y)$ to be the smallest positive eigenvalue of Y. Then for all $a \in (0,2)$ for which local law, Theorem 2.3.14, holds there exists $\delta_{2.4.1} = \delta_{2.4.1}(a) > 0$ such that for all $r > 0$

$$\left|\mathbb{P}(N\xi\lambda_N(X + \sqrt{s}W) \geq r) - \mathbb{P}(N\lambda_N(\tilde{W})| \geq r)\right| \leq \frac{1}{N^{\delta_{2.4.1}}}, \tag{2.4.1}$$

for all $s \in (N^{2\delta - \frac{1}{2}}, N^{-2\delta})$. Note that ξ is the constant defined in (2.1.4).

The proof of Theorem 2.4.1 can be found in paragraph 2.4.

In order to begin the proof we need the following definition.

Definition 2.4.2. For an $N \times N$ matrix J with eigenvalues $\{\lambda_i(J)\}_{i\in[N]}$ we define the free additive convolution of J, with s times the semicircle law, to be the probability measure with Stieltjes transform $\mathrm{m}_{\mathrm{s,fc}}$, such that

$$\mathrm{m}_{\mathrm{s,fc}}(z) = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{\lambda_i(J) - z - sm_{s,fc}(z)}.$$

It can be proven that the equation above has a unique solution. Moreover we denote by $\rho_{\mathrm{s,fc}}(\mathrm{E})$ the density of the free convolution given by $\rho_{\mathrm{s,fc}}(E) = \frac{1}{\pi}\lim_{\epsilon\to 0}\mathrm{Im}(\mathrm{m}_{\mathrm{s,fc}}(E + i\epsilon))$.

Remark 2.4.3. For $z \in D_{C_a,\delta}$, the set for which the local law holds in Theorem 2.3.14, and $s \in (N^{\delta - \frac{1}{2} + \sigma}, N^{-2\delta})$, one has that $|\mathrm{m}_{\mathrm{s,fc}}(z) - m_{N,s}(z)| \leq \frac{1}{N\eta}$ with overwhelming probability, as is proven in Theorem 4.5 of [3]. Here $\mathrm{m}_{\mathrm{s,fc}}$ is the Stieltjes transform of the free additive convolution of X with s times the semicircle law and $m_{N,s}$ is the Stieltjes transform of the E.S.D of the matrix $X + \sqrt{s}W$, where W is the symmetrization of a matrix with i.i.d. entries all following the Gaussian distribution with mean 0 and variance $\frac{1}{N}$. Moreover the following stability result is true, due to Lemma 4.1 of [3],

$$c \leq \mathrm{Im}(\mathrm{m}_{\mathrm{s,fc}}(z)) \leq C. \tag{2.4.2}$$

In order to establish Theorem 2.4.1, we wish to apply Theorem 3.2 in [3] but we need to take into account Remark 7.6 of [4]. So firstly we state the following.

Lemma 2.4.4. *Fix $s \in (N^{2\delta - \frac{1}{2}}, N^{-2\delta})$ for appropriate small δ. Then*

$$|\rho_a(x) - \rho_{\mathrm{s,fc}}(x)| \leq N^{-\frac{a\delta}{8}},$$

for $x \in (-\frac{1}{C_a}, \frac{1}{C_a})$ and a, δ are parameters satisfying the assumptions of Theorem 2.3.14 and C_a the constant mentioned in the statement of Theorem 2.3.14.

Proof. The proof of the lemma is due to the local law Theorem 2.3.14 and similar to the proof of [49], Lemma 3.4, so it is omitted. □

Proof of Theorem 2.4.1. Firstly we apply Theorem 3.2 of [3] to the sequence of matrices $\rho_{sc}(0)X_N$. Note that due to Theorem 2.3.14, the matrix X satisfies the assumptions of Theorem 3.2 for $g = N^{\delta-\frac{1}{2}}$ and $G = N^{-\delta}$ with overwhelming probability, for any small enough $\delta > 0$. So for all $s_0, s_1 \in (N^{2\delta-\frac{1}{2}}, N^{-2\delta})$ such that $s_0 = \frac{N^{\omega_0}}{N}$, $s_1 = \frac{N^{\omega_1}}{N}$ with $\omega_1 < \frac{\omega_0}{2} < \frac{1}{2}$, there exists $\delta_{2.4.1} > 0$ and a coupling of $\lambda_N(X + \sqrt{s_1 + s_0}W)$ and $\lambda_N(\tilde{W} + \sqrt{s_1 + s_0}\tilde{W}')$ such that,

$$\left|\frac{\rho_{sc}(0)}{\rho_{s_0,fc}(0)}\lambda_N^o(X + \sqrt{s_1 + s_0}W) - \lambda_N^o(\tilde{W} + \sqrt{s_1 + s_0}\tilde{W}')\right| \leq \frac{1}{N^{\delta_{2.4.1}+1}}, \tag{2.4.3}$$

where $\tilde{W}, \tilde{W}'$ are independent copies of W. Moreover, by the properties of the Gaussian law, one has that $\tilde{W} + \sqrt{s_1 + s_0}\tilde{W}'$ has the same law as $\sqrt{1 + s_1 + s_0}\tilde{W}''$, where $\tilde{W}''$ is again an independent copy of W. But by Slutsky's theorem one has that

$$\lim_{N\to\infty} N\lambda_N(\sqrt{1 + s_1 + s_0}\tilde{W}'') \overset{d}{=} \lim_{N\to\infty} N\lambda_N(\tilde{W}'').$$

So one has that for each $r > 0$,

$$\left|\mathbb{P}(N\frac{\rho_{sc}(0)}{\rho_{s_0,fc}(0)}\lambda_N(X + \sqrt{s_1 + s_0}W) \geq r) - \mathbb{P}(N\lambda_N(\tilde{W}) \geq r)\right| \leq \frac{1}{N^{\delta_{2.4.1}}}, \tag{2.4.4}$$

where we have violated the notation in (2.4.4) by keeping the same constant $\delta_{2.4.1}$. Next, since Remark 2.4.3 and Lemma 2.4.4 are true, one has that

$$\left|\mathbb{P}\left(N\xi\lambda_N(X + \sqrt{s_1 + s_0}W) \geq r\right) - \mathbb{P}(N\lambda_N(\tilde{W}) \geq r)\right| \leq \frac{1}{N^{\delta_{2.4.1}}}. \tag{2.4.5}$$

Moreover for $s_1, s_2 \in (N^{2\delta-\frac{1}{2}}, N^{-2\delta})$, such that $s_1 < s_2$, one can apply Weyl's inequality, Lemma 2.3.15, to get that

$$\lambda_N(X + \sqrt{s_1}W) - \lambda_N(X + \sqrt{s_2}W) \geq (s_1 - s_2)\lambda_{\min}(W) \geq 0. \tag{2.4.6}$$

The first inequality of (2.4.6) comes from the bottom of Weyl's inequality, for the $\frac{N}{2} + 1$−th eigenvalues of $X + \sqrt{s_1}W$ and $X + \sqrt{s_2}W$ when the eigenvalues are arranged in decreasing order. Note that in the notation we normally use, we have arranged the eigenvalues in decreasing order with respect to their absolute values. The second inequality comes from the fact that $\lambda_{\min}(W)$ is the negative of the maximum singular value of L.
So (2.4.6) implies that if $s_1 \leq s_2$ then

$$\lambda_N(X + \sqrt{s_1}W) \geq \lambda_N(X + \sqrt{s_2}W). \tag{2.4.7}$$

Finally, fix $s \in (N^{2\delta-\frac{1}{2}}, N^{-2\delta})$ and $s_1 = \frac{N^{\omega_1}}{N}, s_2 = \frac{N^{\omega_2}}{N}$ parameters such that

$$s_1 - \frac{N^{\omega/2}}{N} > N^{2\delta-\frac{1}{2}}, \quad s_1 < s, \quad s_2 - \frac{N^{\omega_2/2}}{N} \geq s \quad \text{and} \quad s_2 < N^{-2\delta}.$$

So by construction, one has that $\lambda_N(X + \sqrt{s_1}W)$ and $\lambda_N(X + \sqrt{s_2}W)$ are both universal in the sense of (2.4.5) and $s_1 < s < s_2$. So by (2.4.7),

$$N\xi\lambda_N(X + \sqrt{s_2}W) \leq N\xi\lambda_N(X + \sqrt{s}W) \leq N\xi\lambda_N(X + \sqrt{s_1}W),$$

which implies Theorem 2.4.1. □

Corollary 2.4.5. *The least singular value of $X + \sqrt{t}W$ is universal in the sense of Theorem 2.4.1 , where t is defined in Definition 2.2.9.*

Proof. We just need to show that t belongs to the interval $(N^{2\delta-\frac{1}{2}}, N^{-2\delta})$, for any small enough $\delta > 0$, and then apply Theorem 2.4.1. Note that the latter claim is true due to the way υ is chosen in (2.2.1), i.e.,

$$0 < \upsilon(2-\alpha) < \frac{1}{2},$$

and since t is of order $N^{-\upsilon(2-\alpha)}$. □

2.5 Isotropic local law for the perturbed matrices at the optimal scale

At this point we have proven, in Theorem 2.3.14, that some kind of regularity holds for the matrix X. Specifically we have proven that with high probability, the Stieltjes transform of X converges to its deterministic limit, and its diagonal entries of its resolvent are logarithmically bounded, for complex numbers with imaginary parts of order just above $N^{-\frac{1}{2}}$. So, at this section we "justify" the reason why we have splitted the matrix H into its "big" and "small" elements, i.e., the matrices X and A, in Definition 2.2.4. More precisely, we prove that given the regularity properties of X and after perturbing it by a Gaussian component, then the matrix becomes even more regular in some sense. Thus, what will remain to investigate is whether the "small" elements of H preserve this regularity, which will be proven in the next section.

Specifically, at this section we show that for any small $\delta > 0$, the event δ–dependent events

$$\left\{ \sup_{\mathbb{D}_{C_\alpha,\delta}} \sup_{i,j} |T_{i,j}(z)| \leq N^{\delta} \right\}, \tag{2.5.1}$$

hold with overwhelming probability. Here

$$\mathbb{D}_{C_\alpha,\delta} = \left\{ E + i\eta : E \in \left(-\frac{1}{2C_\alpha}, \frac{1}{2C_\alpha} \right), \eta \in \left[N^{\delta-1}, \frac{1}{4C_\alpha} \right] \right\}$$

and C_α is the constant mentioned in Theorem 2.3.13. This is stated in Corollary 2.5.15.

In order to prove the latter, we will show a general result which can be used for a general class of matrices. So except from Corollary 2.5.15, the rest of this section is independent from the rest of the paper. The general result we prove in Theorem 2.5.6 is an approximation of the resolvent of the symmetrization of a slightly perturbed by a Gaussian component matrix, which initially satisfies some regularity assumptions, Assumption 2.5.1. This resolvent is approximated by a quantity which involves the free additive convolution of the initial matrix with the semicircle law and the eigenvectors of the initial matrix. This approximation is achieved at any direction on the sphere, so it is called isotropic local law.

The isotropic local law is an analogue of Theorem 2.1 in [6] for our set of matrices, i.e., matrices perturbed by Gaussian factors with 0 at the diagonal blocks. In [6] an isotropic local law is proven for matrices after perturbing them by a symmetric Brownian motion matrix.

This kind of results demands precise computations for the resolvent entries. In our case the "target" matrix, with which we compare the resolvent, is a diagonal matrix who lives in $M_N(M_2(\mathbb{C}))$. This increases the complexity of the calculations from the symmetric case where the "target matrix" is diagonal, but eventually this increase is not that significant.

2.5.1 Terminology

Firstly we introduce the terminology of [6].

For any $N-$dependent random variables Y_1, Y_2 we denote

1. $Y_1 \preceq Y_2$ if there exists a universal constant $C > 0$ such that $|Y_1| \leq CY_2$.
2. $Y_1 \preceq_k Y_2$ if there exists a constant C_k (which depends on some k) such that $|Y_1| \leq C_k Y_2$.
3. $Y_1 \ll Y_2$ if there a positive constant c such that $Y_1 N^c \leq Y_2$.

2.5.2 Statement of the main result of this section

Assumption 2.5.1. Let V be a deterministic $N \times N$ matrix. Denote $\tilde{V}$ the symmetrization of V and $m_{\tilde{V}}$ the Stieltjes transform of $\tilde{V}$. Assume that there exists a large constant $a > 1$ such that

1. $\| \tilde{V} \|_{op} \leq N^a$.
2. $a^{-1} \leq \operatorname{Im}(m_{\tilde{V}}(z)) \leq a$, for all $z \in \{E + i\eta, \quad E \in (E_0 - r, E_0 + r) \quad , h_* \leq \eta \leq 1\}$ for some $N-$dependent constants r, h_* such that: $\frac{1}{N} \ll h_* \ll r \leq 1$.

Moreover, fix $c > 0$ some arbitrary small constant and set

$$\psi = \frac{N^c}{N}. \tag{2.5.2}$$

Remark 2.5.2. Note that the matrix X satisfies with high probability the Assumptions 2.5.1 due to Theorem 2.3.14, for $E_0 = 0$, $h_* = N^{\delta - \frac{1}{2}}$ for arbitrary small constant $\delta > 0$, $r = \frac{1}{C}$ where C is the constant mentioned in Theorem 2.3.14 and since for fixed large $D > 0$, one can compute by (2.1.3) that any given entry of X has magnitude greater than $N^{\frac{2D+1}{a}}$ with probability less than CN^{-2D-2}, which implies that

$$\mathbb{P}\left(\| X \|_{op} \geq N^{\frac{2(D+3)}{a}} \right) \leq CN^{-2D}. \tag{2.5.3}$$

Remark 2.5.3. Let V be a deterministic $N \times N$ matrix. Due to the singular value decomposition of V, there exist J_1, J_2 two orthogonal $N \times N$ matrices, such that $\Sigma = J_2 V J_1$, where Σ is a diagonal matrix with diagonal entries the singular values of V. Then denote $\tilde{V}$ the symmetrization of V and set

$$U = \begin{bmatrix} J_1^T & 0 \\ 0 & J_2 \end{bmatrix}.$$

Then it is true that

$$U\tilde{V}U^T = \begin{bmatrix} 0 & \Sigma \\ \Sigma & 0 \end{bmatrix}.$$

Moreover, note that U is orthogonal.

Definition 2.5.4. Suppose V is a deterministic matrix which satisfies the Assumption 2.5.1 for some N-dependent constants h_*, r. Then for any $k \in (0,1)$, define the set

$$\mathbb{D}_k = \left\{ z = E + i\eta : E \in (E_0 - (1-k)r, E_0 + (1-k)r), \frac{\psi^4}{N} \le \eta \le 1 - k\,r \right\}.$$

The parameter ψ is defined in (2.5.2).

Definition 2.5.5. Recall the definition of the the Stieltjes transform of the Empirical spectral distribution of $\tilde{V}$ with s–times the semicircle law in Definition 2.4.2. We will use the following notation

$$\mathrm{m}_{s,\mathrm{fc}}(z) = \frac{1}{N} \sum_{\{i \in [N]\} \cup \{-i \in [N]\}} g_i(s,z), \quad \text{with} \quad g_i(s,z) = \frac{1}{\hbar_i - z - s\mathrm{m}_{s,\mathrm{fc}}(z)}$$

and $\hbar_i$ are the eigenvalues of $\tilde{V}$ arranged in increasing order so that $\hbar_i = -\hbar_{-i}$.

Theorem 2.5.6. *Let V be a deterministic matrix that satisfies the Assumptions 2.5.1. Denote the matrix*

$$G(z,s) = (\tilde{V} + \sqrt{s}W - z\mathbb{I})^{-1}.$$

Here W is the symmetrization of a matrix with i.i.d. entries, all following the Gaussian law with mean 0 and variance $\frac{1}{N}$. Moreover fix U to be the orthogonal matrix constructed in Remark 2.5.3 for V. Moreover fix $k \in (0,1)$, $s : h_* \ll s \ll r$ and $q \in \mathbb{R}^N : \|q\|_2 = 1$. Then it is true that,

$$\left| \langle q, G(s,z)q \rangle - \sum_{i=-N}^{N} \frac{1}{2}(g_i + g_{-i})(s,z)\langle u_i, q\rangle^2 - \sum_{i=1}^{N}(g_i - g_{-i})(s,z)\langle u_i, q\rangle\langle u_{i+N}, q\rangle \right| \tag{2.5.4}$$

$$\le \frac{\psi^2}{\sqrt{N\eta}} \operatorname{Im}\left(\sum_{i=1}^{N} (\langle u_i, q\rangle^2 + \langle u_{i+N}, q\rangle^2)(g_i(s,z) + g_{-i}(s,z)) \right), \tag{2.5.5}$$

with overwhelming probability, uniformly for all $z \in \mathbb{D}_k$. Here u_i denote the columns of U.

Set C_j, for $j \in \{1, 2\}$, to be the $N \times N$ diagonal matrices with their $i - th$ diagonal element equal to $g_i + (-1)^{j+1} g_{-i}$. Fix the $2N \times 2N$ matrix,

$$C = \frac{1}{2} \begin{bmatrix} C_1 & C_2 \\ C_2 & C_1 \end{bmatrix}.$$

In general, what Theorem 2.5.6 states is that the matrix $G(z, s)$ can be well approximated by UCU^*, since

$$\langle q, UCU^* q \rangle = \sum_{i=-N}^{N} \frac{1}{2}(g_i + g_{-i})(s, z)\langle u_i, q \rangle^2 + \sum_{i=1}^{N} (g_i - g_{-i})(s, z\langle u_i, q \rangle \langle u_{i+N}, q \rangle.$$

Moreover we can reduce the proof of Theorem 2.5.6 to the diagonal case.

Theorem 2.5.7. *Fix* $V = diag(v_1, \cdots, v_N)$ *a diagonal matrix which satisfies Assumption 2.5.1. Moreover set* W *to be the symmetrization of a matrix* L *with i.i.d. entries, all following the Gaussian law with 0 mean and* $\frac{1}{N}$ *variance. Define the resolvent* $G(z, s) = (\tilde{V} + \sqrt{s}W - z\mathbb{I})^{-1}$. *Fix* $k \in (0, 1)$, $h_* \ll s \ll r$ *and* $q : \|q\|_2 = 1$. *Then*

$$\left| \langle q, G(s, z)q \rangle - \sum_{i=-N}^{N} \frac{1}{2}(g_i + g_{-i})(s, z)q_{i+N}^2 - \sum_{i=1}^{N} (g_i - g_{-i})(s, z)q_i q_{i+N} \right| \tag{2.5.6}$$

$$\leq \frac{\psi^2}{\sqrt{N}\eta} \operatorname{Im} \left(\sum_{i=1}^{N} (q_i^2 + q_{i+N}^2)(g_{-i}(s, z) + g_i(s, z) \right), \tag{2.5.7}$$

holds with overwhelming probability uniformly for all $z \in \mathbb{D}_k$.

The proof of Theorem 2.5.7 can be found in paragraph 2.5.3.

Proof of Theorem 2.5.6 assuming Theorem 2.5.7. Let V be a general deterministic matrix with singular value decomposition $\Sigma = J_2 V J_1$ where J_1 and J_2 are orthogonal matrices. Define U as in Remark 2.5.3. Then

$$U(\tilde{V} + \sqrt{s}W)U^T = \begin{bmatrix} 0 & \Sigma + J_1^T \sqrt{s} L^T J_2^T \\ \Sigma + J_2 \sqrt{s} L J_1 & 0 \end{bmatrix}.$$

But L is invariant under orthogonal transformation, so $J_2 L J_1$ has the same law as L. This implies that $U(\tilde{V} + \sqrt{s}W)U^T$ has the same law as $U\tilde{V}U^T + \sqrt{s}W$. Next, by the properties of the inner product, one has that

$$\langle q, (\tilde{V} + \sqrt{s}W - z\mathbb{I})^{-1} q \rangle = \langle q, U(U\tilde{V}U^T + UWU^T - z\mathbb{I})^{-1} U^T q \rangle = \langle U^T q, (U\tilde{V}U^T + UWU^T - z\mathbb{I})^{-1} U^T q \rangle.$$

By a similar computation for $\langle q, UCU^* q \rangle$, one reduces the problem in bounding

$$\left| \langle q, (U\tilde{V}U^T + \sqrt{s}W - z\mathbb{I})^{-1} q \rangle - \left(\sum_{i=1}^{N} q_i^2 g_i(t, z) + q_{i+N}^2 g_{-i} \right) - \sum_{i \in [N]} q_i q_{i+N} (g_i - g_{-i})(s, z) \right|,$$

which is true by a direct application of Theorem 2.5.7.

□

So it suffices to prove Theorem 2.5.7, i.e., to consider V to be diagonal. Moreover we have the following identities.

Remark 2.5.8. Let V be a deterministic diagonal matrix which satisfies Assumptions 2.5.1. Adopt the notation of Theorem 2.5.7. Then consider the following matrix $F = \{F_{i,j}\}_{i,j\in[N]}$, where

$$F_{i,j} := \begin{bmatrix} 0 & [V+\sqrt{s}L]_{i,j} \\ [V+\sqrt{s}L]_{j,i} & 0 \end{bmatrix}, \text{ for all } i,j \in [N].$$

Note that there exists a unitary matrix S, the product of permutation matrices, such that if we set $F = S^T(V+\sqrt{s}W)S$ then $G(s,z) = S(F-z\mathbb{I})^{-1}S^T$ and

$$(F-z\mathbb{I})^{-1}_{i,j} = \begin{bmatrix} G_{i,j} & G_{i+N,j} \\ G_{i,j+N} & G_{i+N,j+N} \end{bmatrix},$$

where $G_{i,j}$ are the entries of $G(s,z)$.

It is more convenient to work with the matrix F and its resolvent as it can be thought as a full symmetric matrix in $\mathcal{M}_N(\mathcal{M}_2(\mathbb{C}))$, instead of a symmetric matrix with 0 at the diagonal blocks in $\mathcal{M}_{2N}(\mathbb{C})$.

2.5.3 Proof of Theorem 2.5.7

In this subsection we will prove Theorem 2.5.7. First, we present some results from [3], necessary for the proof.

Proposition 2.5.9 ([3],Theorem 4.5)**.** *Fix s as in Theorem 2.5.7, the parameter ψ defined in* (2.5.2) *and $\kappa \in (0,1)$. Then it is true that,*

$$|m_s(z) - \mathrm{m}_{s,fc}(z)| \leq \frac{\psi}{N\eta},$$

holds with overwhelming probability uniformly for all $z \in \mathbb{D}_\kappa$. Here $m_s(z)$ is the Stieltjes transform of $\tilde{V} + \sqrt{s}W$.

Lemma 2.5.10. *Fix s and κ as in Theorem 2.5.7. Then uniformly for all $z \in \mathbb{D}_\kappa$, there exists a constant $C > 1$ such that:*

$$C^{-1} \leq |\mathrm{m}_{s,fc}(z)| \leq C, \tag{2.5.8}$$

$$|\mathrm{m}_{s,fc}(z)| \leq \frac{1}{N}\sum_{i=1}^{N} |g_i| + |g_{-i}| \leq C\log(N). \tag{2.5.9}$$

Proof. These estimates can be found in [3] Lemma 4.1 and Lemma 4.12. □

Moreover the following estimates hold.

Lemma 2.5.11. *Fix $\mathbb{T} \subseteq [2N]$ such that $|\mathbb{T}| \leq \log(N)$, which consists of pairs of indeces $\{k, k+N\}$ for $k \in [N]$. Moreover set $(H + \sqrt{s}G)^{\mathbb{T}}$ the sub matrix of $H + \sqrt{s}G$ with the $i-th$ columns and row removed for all $i \in \mathbb{T}$ and $G^{\mathbb{T}}(s,z) = \left((H+\sqrt{s}G)^{\mathbb{T}} - z\mathbb{I}_{2N-|\mathbb{T}|}\right)^{-1}$. Then the following estimates hold with overwhelming probability.*

$$\left|G^{\mathbb{T}}_{i,i} - \frac{1}{2}(g_i + g_{-i})\right| \leq (|g_i| + |g_{-i}|)^2 \frac{\psi s}{\sqrt{N\eta}} \text{ for all } i \in [2N] \setminus \mathbb{T}, \tag{2.5.10}$$

$$\left|G^{\mathbb{T}}_{i,N+i} - \frac{1}{2}(g_i - g_{-i})\right| \leq (|g_i| + |g_{-i}|)^2 \frac{\psi s}{\sqrt{N\eta}} \text{ for all } i \in [2N] \setminus \mathbb{T}, \tag{2.5.11}$$

$$|G^{\mathbb{T}}_{i,j}| \leq \frac{\min(|g_i| + |g_{-i}|, |g_j| + |g_{-j}|)\psi}{\sqrt{N\eta}} \leq \frac{\left((|g_i| + |g_{-i}|)(|g_j| + |g_{-j}|)\right)^{1/2}\psi}{\sqrt{N\eta}} \text{ for all } i,j \in [2N] \setminus \mathbb{T}. \tag{2.5.12}$$

Proof. The first two estimates are proven by the Schur Complement formula and the bounds (4.69) and (4.89) from [3]. The last bound is given in [3], equation (4.70) and (4.79). □

Next, we present a bound for the diagonal and the anti-diagonal entries of $G(s,z)$.

Lemma 2.5.12. *Adopt the notation of Theorem 2.5.7. Then it is true that with overwhelming probability*

$$\left|\langle q, G(s,z)q\rangle - \frac{1}{2}\Big(\sum_{i=-N}^{N} q_i^2(g_i(s,z) + g_{-i}(s,z)) - \sum_{i=1}^{N} q_i q_{i+N}(g_i - g_{-i})\right| \tag{2.5.13}$$

$$\leq \frac{\psi^2}{\sqrt{N\eta}} \operatorname{Im}\left(\sum_{i=-N}^{N} q_i^2\,(g_i(s,z) + g_{-i}(s,z))\right) + \frac{\psi^2}{\sqrt{N\eta}} \operatorname{Im}\left(\sum_{i=1}^{N}(g_i + g_{-i})|(q_{i+N}^2 + q_i^2)\right) \tag{2.5.14}$$

$$+ \left|\sum_{i\neq j, i\neq N+j} G_{i,j} q_i q_j\right| \tag{2.5.15}$$

Proof. One has that,

$$\langle q, G(s,z)q\rangle = \sum_{i=1}^{2N} q_i^2 G_{i,i} + 2\sum_{i=1}^{N} G_{i,i+N} q_i q_{i+N} + \sum_{i\neq j, i\neq N+j} q_i q_j G_{i,j}. \tag{2.5.16}$$

So for the first part on the right side of the equality in (2.5.16), one can apply (2.5.10)) to get that,

$$\sum_{i=-N}^{N} q_{i+N}^2 \left|G_{i,i} - \frac{1}{2}(g_i + g_{-i})\right| \leq \tag{2.5.17}$$

$$\frac{s\psi}{\sqrt{N\eta}} \sum_{i=-N}^{N} q_{i+N}^2\,(|g_i| + |g_{-i}|)^2 \leq \frac{2s\psi}{\sqrt{N\eta}} \sum_{i=-N}^{N} q_{i+N}^2 |g_i|^2 + \frac{2s\psi}{\sqrt{N\eta}} \sum_{i=-N}^{N} q_{i+N} |g_{-i}|^2. \tag{2.5.18}$$

Next, we can apply Proposition 2.8 from [6] to get that with overwhelming probability,

$$\sum_{i=-N}^{N} q_{i+N}^2 \left|G_{i,i} - \frac{1}{2}(g_i + g_{-i})\right| \leq \frac{2\psi}{\sqrt{N\eta}} \operatorname{Im}\left(\sum_{i=-N}^{N}(g_i + g_{-i})q_{i+N}^2\right).$$

Similarly, by (2.5.11) one has that with overwhelming probability,

$$\sum_{i=1}^{N} |q_i q_{i+N}| \left| G_{i,N+i} - \frac{1}{2}(g_i - g_{-i}) \right| \leq \frac{2\psi}{\sqrt{N\eta}} \operatorname{Im}\left(\sum_{i=1}^{N} (g_i + g_{-i}) |q_{i+N}^2 + q_i^2| \right), \tag{2.5.19}$$

$$\sum_{i=1}^{N} |q_i q_{i+N}| \left| G_{N+i,i} - \frac{1}{2}(g_i - g_{-i}) \right| \leq \frac{2\psi}{\sqrt{N\eta}} \operatorname{Im}\left(\sum_{i=1}^{N} (g_i + g_{-i}) |q_{i+N}^2 + q_i^2| \right). \tag{2.5.20}$$

□

So in order to prove Theorem 2.5.7, it suffices to prove that

$$\mathbf{E}\, Z^{2k} \preceq_k Y^{2k} \quad \text{for all } k \in \mathbb{N}, \tag{2.5.21}$$

where,

$$Z := \left| \sum_{i \neq j \mod N} q_i q_j G_{i,j} \right| \quad \text{and} \quad Y = \frac{\log Ny}{\sqrt{N\eta}} \operatorname{Im}\left(\sum_{i=1}^{N} (g_i + g_{-i}) q_i^2 + q_{i+N}^2 \right). \tag{2.5.22}$$

By (2.5.21), one can obtain Theorem 2.5.7 by Markov's inequality, which will imply

$$\left| \sum_{i \neq j \mod N} q_i q_j G_{i,j} \right| \leq \frac{\psi^2}{\sqrt{N\eta}} \operatorname{Im}\left(\sum_{i=1}^{N} (g_i + g_{-i})(q_i^2 + q_{-i}^2) \right) \tag{2.5.23}$$

with overwhelming probability. More precisely for any $D > 0$ if we fix $k : ck \geq D$ and sufficient large N such that $N^{ck-D} \geq C_k$. Here c is the constant in the definition of ψ in (2.5.2) and C_k is implied in (2.5.21). Thus, one can apply Markov's inequality in order to get that:

$$\mathbb{P}\left(Z \geq \frac{\psi}{\log(N)} Y \right) = \mathbb{P}\left(Z^{2k} \geq \left(\frac{\psi}{\log(N)} Y \right)^{2k} \right) \leq \frac{C_k \log^{2k}(N)}{N^{c2k}} \leq \frac{C_k}{N^{ck}} \leq \frac{1}{N^D}.$$

Next, we give an analysis for the moments of Z. Firstly, note that

$$\mathbf{E}|Z|^{2k} = \sum_{\mathbf{b}} q_{b_1} q_{b_2} q_{b_3} \cdots q_{b_{4k}} \mathbb{E} X_{b_1,b_2} X_{b_3,b_4} \cdots X_{b_{4k-1},b_{4k}}, \tag{2.5.24}$$

where the sum is taken over all $\mathbf{b} \subseteq [2N]^{4k}$ such that $b_{2i-1} \neq b_{2i} \mod N$ and $X_{b_{2i-1},b_{2i}} = G_{b_{2i-1},b_{2i}}$ for $i \in [k]$ and $X_{b_{2i-1},b_{2i}} = \bar{G}_{b_{2i-1},b_{2i}}$ and $i \in [2k] \setminus [k]$. Furthermore, we can continue the analysis of the sum such that,

$$\mathbf{E}|Z|^{2k} = \sum_{\mathbf{B}} \sum_{b_i = \{B_i, B_i + N\}} q_{b_1} q_{b_2} q_{b_3} \cdots q_{b_{4k}} \mathbb{E} X_{b_1,b_2} X_{b_3,b_4} \cdots X_{b_{4k-1},b_{4k}}. \tag{2.5.25}$$

Now the sum is considered, firstly over all $\mathbf{B} \subseteq [i \mod N]^{4k}$ with the restriction that $B_{2i-1} \neq B_{2i}$ and then over the possible $b_i = k$ such that $k \in [N]$ or $k \in B_i$ or $b_i = k + N$ for $k \in [N]$ and $k \in B_i$. Next, for every summand in (2.5.25) set $\mathbb{T} = \cup_{b_i \in \mathbf{B}, b_i \in [N]} \{b_i, b_i + N\}$. Moreover set the diagonal block matrices

$$\{\mathrm{M}_{\mathrm{fs,s}}^{(\mathbb{T})}\}_{ii \in \mathbb{T}} = \begin{bmatrix} \mathrm{m}_{\mathrm{s,fc}} & 0 \\ 0 & \mathrm{m}_{\mathrm{s,fc}} \end{bmatrix}$$

and

$$M^{\mathbb{T}}_{ii\in[T]} = \begin{bmatrix} m^{(\mathbb{T})} & 0 \\ 0 & m^{(\mathbb{T})} \end{bmatrix},$$

where $m^{(\mathbb{S})}(z)$ is the trace of the resolvent of $(V+\sqrt{s}W)^{(\mathbb{S})}$ divided by $2N$. Here $\mathbb{S}$ is any subset of $[2N]$ and $(V+\sqrt{s}W)^{(\mathbb{S})}$ is the minor of $(V+\sqrt{s}W)$ with rows and columns not included in $\mathbb{S}$. Moreover set

$$\Phi_{i,i} = \begin{bmatrix} -z & \mathcal{n}_i \\ \mathcal{n}_i & -z \end{bmatrix}$$

and

$$W'_{i,j} = \begin{bmatrix} 0 & w_{i,j} \\ w_{j,i} & 0 \end{bmatrix}.$$

Adopting the notation of Proposition 2.5.8 and since $G(s,z) = S(F - z\mathbb{I})^{-1}S^T$, one can apply Schur complements formula to get that

$$(F - z\mathbb{I})^{-1}_{i,j\in\mathbb{T}} = (\Phi_{i\in\mathbb{T}} + \sqrt{s}W'_{i,j\in\mathbb{T}} - s(W')^*_{i,j:i\in\mathbb{T},j\in[2N]\setminus\mathbb{T}}(S^TG(s,z)S)^{(\mathbb{T})}W'_{i\in[2N]\setminus\mathbb{T},j\in\mathbb{T}})^{-1} \tag{2.5.26}$$

$$= (D - E^1 - E^2 - E^3)^{-1}, \tag{2.5.27}$$

where

$$D = \Phi_{i\in\mathbb{T}} - s\mathrm{M}^{(\mathbb{T})}_{\mathrm{fc,s}},\ E^1 = s(M^{\mathbb{T}} - \mathrm{M}^{\mathbb{T}}_{\mathrm{fs,s}}),\ E^2 = -\sqrt{s}W', \tag{2.5.28}$$

$$E^3 = s(W')^*_{i,j:i\in\mathbb{T},j\in[2N]\setminus\mathbb{T}}(S^TG(s,z)S)^{(\mathbb{T})}W_{i\in[2N]\setminus\mathbb{T},j\in\mathbb{T}} - M^{\mathbb{T}}). \tag{2.5.29}$$

Next, we wish to estimate the operator norm of the matrix ED^{-1}. We will show that,

$$|ED^{-1}|_{op} \le C_k \frac{\psi}{\sqrt{N\eta}}, \tag{2.5.30}$$

with overwhelming probability. Here $E = \sum_{i=1}^3 E^i$.

More precisely, firstly note that D is a $2N \times 2N$ dimensional matrix with 0 at the all non-diagonal 2×2 blocks and with diagonal blocks equal to

$$D_{i,i} = \begin{bmatrix} -z - \mathrm{m}_{\mathrm{s,fc}}(z) & \mathcal{n}_i \\ \mathcal{n}_i & -z - \mathrm{m}_{\mathrm{s,fc}}(z) \end{bmatrix}.$$

So the inverse of D will preserve the same structure. Thus, we can compute that:

$$D^{-1}_{i,i} = \frac{1}{2}\begin{bmatrix} g_i + g_{-i} & g_i - g_{-i} \\ g_i - g_{-i} & g_i + g_{-i} \end{bmatrix}.$$

Moreover since $\mathrm{Im}(z + s\mathrm{m}_{\mathrm{s,fc}}(z)) \ge (s+\eta)$, we get that $|g_i| \le \frac{1}{s+\eta}$, for all $i : |i| \in [N]$. All these imply that all the entries of D^{-1} are bounded by $\frac{1}{s+\eta}$ up to some universal constant. So it is implied that

$$|D^{-1}|_{op} \preceq_k \frac{1}{s+\eta}. \tag{2.5.31}$$

Next, similarly to the proof of (2.16) in [6] one can prove that

$$|E|_{op} \preceq_k (s+\eta)\frac{\psi}{\sqrt{N\eta}}. \tag{2.5.32}$$

So after combining (2.5.31) and (2.5.32), we get (2.5.30). Set $\mathcal{A}$ to be the event where (2.5.30) holds with overwhelming probability. Then it is true that for appropriately large N

$$\mathbb{P}(\mathcal{A}^c) \le N^{-(4\,a+6)k}, \tag{2.5.33}$$

where a is given in the Assumptions 2.5.1. Next by Taylor's expansion on the event $\mathcal{A}$ one has

$$(F-z\mathbb{I})^{-1}_{i,j\in\mathbb{T}} = (D-E)^{-1} = \sum_{l=0}^{f-1} D^{-1}(ED^{-1})^l + (D-E)^{-1}(ED)^{-f}, \tag{2.5.34}$$

where f can be chosen to be arbitrary large. We choose $f = \left\lceil \frac{8k(a+1)}{c} \right\rceil$, where c is mentioned in the definition of ψ in (2.5.2) and a is mentioned in the Assumption 2.5.1. Moreover since all the non diagonal 2×2 blocks of D^{-1} are 0, we can ignore the case of $l=0$ in (2.5.34), since we are interested in the elements of $G_{i,j\in\mathbb{T}}$, such that $b_i \neq b_{i+1} \bmod N$. Moreover set $X^l_{b_i,b_{i+1}} = (D^{-1}(ED^{-1})^l)_{b_i,b_{i+1}}$ and $X^\infty_{b_i,b_{i+1}} = ((D-E)^{-1}(ED)^f)_{b_i,b_{i+1}}$.

So in order to prove (2.5.21), firstly we need to bound Y from below. Note that similarly to [6] (2.13) one has

$$\operatorname{Im}\left(\sum_{i=-N}^{N} q^2_{i+N} g_i\right) = \sum_{i=-N}^{N} \frac{(\eta + s\operatorname{Im}(m_{s,fc}(z)))\,q_i^2}{|\mathcal{h}_i(0) - z - m_{s,fc}(z)|^2} \ge \frac{\eta}{N^{2\,a}}, \tag{2.5.35}$$

due to the fact that $z\in\mathbb{D}_k$, Assumption 2.5.1 and (2.5.9). So it is easily implied that

$$Y \ge \frac{\eta\psi\log(N)}{N^{2\,a}\sqrt{N\eta}}. \tag{2.5.36}$$

Returning to the analysis of equation (2.5.25), one has that

$$\sum_{\mathbf{B}} \sum_{b_i=\{B_i,B_i+N\}} q_{b_1}q_{b_2}q_{b_3}\cdots q_{b_{4k}}\mathbb{E}X_{b_1,b_2}X_{b_3,b_4}\cdots X_{b_{4k-1},b_{4k}} \tag{2.5.37}$$

$$= \sum_{\mathbf{B}} \sum_{b_i\{B_i,B_i+N\}} q_{b_1}q_{b_2}q_{b_3}\cdots q_{b_{4k}}\mathbb{E}\prod_{i=1}^{2k} X_{b_{2i-1},b_{2i}}\mathbf{1}(\mathcal{A}) + O(N^{2k}\eta^{-2k})\mathbb{P}(\mathcal{A}^c), \tag{2.5.38}$$

where the second part on the right hand side of the equation comes from the fact that $X_{b_i,b_{i+N}}$ are uniformly bounded by η^{-1} and the fact that $|q|_1 \le N^{1/2}$ since $|q|_2 = 1$. So one has that

$$\eta^{-2k}\mathbb{P}(\mathcal{A}^c)\left|\sum_{i\neq j} q_iq_j\right|^{2k} = \eta^{-2k}\mathbb{P}(\mathcal{A}^c)\left(\sum_{i=1}^{2N}|q_i|\sum_{j\neq i}|q_j|\right)^{2k}$$
$$\le \eta^{-2k}\mathbb{P}(\mathcal{A}^c)\left(\sum_{i\in[2N]} N^{1/2}|q_i|\right)^{2k} \le N^{2k}\eta^{-2k}\mathbb{P}(\mathcal{A}^c). \tag{2.5.39}$$

Next by (2.5.33),(2.5.36) and the fact that $z \in \mathbb{D}_k$ one has that

$$N^{2k}\eta^{-2k}\mathbb{P}(\mathcal{A}^c) \leq N^{-2k-2}\eta^{-2k} \leq \frac{\eta^{2k}}{N^{2k}} \leq Y^{2k}.$$

So, we have proven that we can restrain to the event that $\mathcal{A}$ holds. Returning again to the analysis of the sum in the form (2.5.24) one has that

$$\sum_{\mathbf{b}} q_{b_1}q_{b_2}q_{b_3}\cdots q_{b_{4k}}\mathbb{E}X_{b_1,b_2}X_{b_3,b_4}\cdots X_{b_{4k-1},b_{4k}}\mathbf{1}(\mathcal{A}) = \tag{2.5.40}$$

$$= \sum_{\mathbf{b}} q_{b_1}q_{b_2}q_{b_3}\cdots q_{b_{4k}}\mathbf{E}\mathbf{1}(\mathcal{A})\prod_{i=1}^{2k}\sum_{l=1}^{f-1}X^{(l)}_{b_{2i-1},b_{2i}} \tag{2.5.41}$$

$$+ \sum_{\mathbf{b}} q_{b_1}q_{b_2}q_{b_3}\cdots q_{b_{4k}}\sum_{i=1}^{2k}\mathbf{E}\mathbf{1}(\mathcal{A})X^{(\infty)}_{2i-1,2i}\prod_{j\leq i-1}X_{b_{2j-1},b_{2j}}\prod_{j\geq i+1}\left(X_{b_{2j-1},b_{2j}} - X^{(\infty)}_{b_{2j-1},b_{2j}}\right). \tag{2.5.42}$$

We will show that the second part of the right hand side of the equation is negligible on the event $\mathcal{A}$. Note that since $|(D-E)^{-1}|_{op} \preceq_k \frac{1}{\eta}$ and since (2.5.30) holds in $\mathcal{A}$ we get that

$$|X^{(\infty)}_{b_i,b_{i+1}}| \preceq_k \frac{1}{\eta}\left(\frac{\psi}{\sqrt{N\eta}}\right)^f.$$

All these, imply that

$$\left|\sum_{\mathbf{b}} q_{b_1}q_{b_2}q_{b_3}\cdots q_{b_{4k}}\sum_{i=1}^{2k}\mathbf{E}\mathbf{1}(\mathcal{A})X^{(\infty)}_{2i-1,2i}\prod_{j\leq i-1}X_{b_{2j-1},b_{2j}}\prod_{j\geq i+1}(X_{b_{2j-1},b_{2j}} - X^{(\infty)}_{b_{2j-1},b_{2j}})\right| \preceq_k \frac{N^{2k}}{\eta}\left(\frac{\psi}{\sqrt{N\eta}}\right)^f. \tag{2.5.43}$$

The N^{2k} factor in (2.5.43), comes from bounding the quantity $|\sum_{i\neq j} q_i q_j|$. By the way f is chosen, we get that

$$\frac{N^{2k}}{\eta}\left(\frac{\psi}{\sqrt{N\eta}}\right)^f \preceq_k \left(\frac{\eta}{N^{2a}}\right)^{2k}\left(\frac{\psi}{\sqrt{N\eta}}\right)^{2k} \preceq_k Y^{2k}.$$

So, the remaining quantity in the sum we need to bound is

$$\sum_{1\leq l_1,l_2,\cdots,l_{2k}\leq f-1}\sum_{\mathbf{b}} q_{b_1}q_{b_2}q_{b_3}\cdots q_{b_{4k}}\mathbf{E}\mathbf{1}(\mathcal{A})\prod_{i=1}^{2k}X^{(l_i)}_{b_{2i-1},b_{2i}}.$$

Moreover due to Cauchy–Schwarz inequality one can show that

$$\left|\mathbf{E}\mathbf{1}(\mathcal{A})\prod_{i=1}^{2k}X^{(l_i)}_{b_{2i-1},b_{2i}}\right| \leq \left|\mathbf{E}\prod_{i=1}^{2k}X^{(l_i)}_{b_{2i-1},b_{2i}}\right| + \mathbb{P}(\mathcal{A}^c)\left|\mathbf{E}\prod_{i=1}^{2k}X^{(l_i)}_{b_{2i-1},b_{2i}}\right|^2.$$

So, we will work with the right hand side of the last inequality, meaning we won't focus anymore on the event $\mathcal{A}$. Moreover we will focus on the first summand of the right hand side of the inequality, since the second one can be treated analogously.

Next, we can transform the previously mentioned quantity in a more appropriate form. Firstly, note for each $i \in [k]$:

$$X^{(l_i)}_{b_{2i-1},b_{2i}} = \sum_{\mathbf{a^{(i)}}} (D^{-\mathbf{1}_{j=1)}} ED^{-1})_{a^{(i)}_j, a^{(i)}_{j+1}}$$

and similarly for $i \in [2k] \setminus [k]$

$$\bar{X}^{(l_i)}_{b_{2i-1},b_{2i}} = \sum_{\mathbf{a^{(i)}}} (D^{-\mathbf{1}_{j=1)}} ED^{-1})_{a^{(i)}_j, a^{(i)}_{j+1}},$$

where the sum is taken over all $\mathbf{a^{(i)}} \subseteq \mathbb{T}^{l_i+1}$, i.e., all the l_i-tuples with the restriction that $a^{(i)}_1 = b_{2i}$ and $a^{(i)}_{l_i+1} = b_{2i+1}$. So, since this is true for all $i \in [2k]$ one can show:

$$\prod_{i=1}^{2k} X^{(l_i)}_{b_{2i-1},b_{2i}} = \sum_{\mathbf{a}} \prod (D^{-\mathbf{1}_{(j=1)}} ED^{-1})'_{a^{(i)}_j, a^{(i)}_{j+1}},$$

where the sum is taken over all $\mathbf{a} = (\mathbf{a^1}, \mathbf{a^2} \cdots \mathbf{a^{2k}})$ and for $i \in [2k] \setminus [k]$, the $'$ denotes the conjugate.

Next set

$$\mathbf{E}_{[a^{(i)}_j],[a^{(i)}_{j+1}]} = \begin{bmatrix} E_{[a^{(i)}_j],[a^{(i)}_{j+1}]} & E_{[a^{(i)}_j]+N,[a^{(i)}_{j+1}]} \\ E_{[a^{(i)}_j],[a^{(i)}_{j+1}]+N} & E_{[a^{(i)}_j]+N,[a^{(i)}_{j+1}]+N} \end{bmatrix},$$

where $[a^{(i)}_j]$ is the least positive integer which is equal to $a^{(i)}_j \mod (N)$. Moreover set

$$x(a^{(i)}_j) = \mathbf{1}\left\{a^{(i)}_j = [a^{(i)}_j] + N\right\} + 1.$$

Furthermore, since D^{-1} consists of zero at the non diagonal 2×2 blocks, one has that for $j \neq 1$

$$(ED^{-1})_{a^{(i)}_j, a^{(i)}_{j+1}} = \left(\mathbf{E}_{[a^{(i)}_j],[a^{(i)}_{j+1}]]} D^{-1}_{[a^{(i)}_{j+1}],[a^{(i)}_{j+1}]}\right)_{x(a^{(i)}_j), x(a^{(i)}_{j+1})} \tag{2.5.44}$$

$$= \left(\mathbf{E}_{[a^{(i)}_j],[[a^{(i)}_{j+1}]}\right)_{x(a^{(i)}_j),1} \left(D^{-1}_{[a^{(i)}_{j+1}],[a^{(i)}_{j+1}]}\right)_{1,x(a^{(i)}_{j+1})} + \left(\mathbf{E}_{[a^{(i)}_j],[[a^{(i)}_{j+1}]}\right)_{x(a^{(i)}_j),2} \left(D^{-1}_{a^{(i)}_{j+1}, a^{(i)}_{j+1}}\right)_{2,x(a^{(i)}_{j+1})} \tag{2.5.45}$$

and similarly for $j = 1$

$$(D^{-1}ED^{-1})_{a^{(i)}_1, a^{(i)}_2} = \sum_{l=1}^{2} \sum_{m=1}^{2} \left(D^{-1}_{[a^{(i)}_1],[a^{(i)}_1]}\right)_{[a^{(i)}_1],l} \left(\mathbf{E}_{[a_1]^{(i)},[a_2]^2}\right)_{l,m} \left(D^{-1}_{[a^{(i)}_2],[a_2]^{(i)}}\right)_{m,[a^{(1)}_2]}.$$

So it is implied that

$$\mathbf{E} \prod_{i=1}^{2k} X^{(l_i)}_{b_{2i-1},b_{2i}} = \sum_{\mathbf{a}} \sum_{\mathbf{c}} \mathbf{E} \prod_{i=1}^{2k} \left(D^{-1}_{[a^{(i)}_1],[a^{(i)}_1]}\right)'_{x(a^{(i)}_1),c^i_1} \left(\mathbf{E}_{[a^{(i)}_1],[a^{(i)}_2]}\right)'_{c^{(i)}_1,c^{(i)}_2} \left(D^{-1}_{[a^{(i)}_2],[a^{(i)}_2]}\right)'_{c^{(i)}_2,x(a^{(i)}_2)} \cdot \tag{2.5.46}$$

$$\cdot \prod_{j\neq 1} \left(\mathbf{E}_{[a^{(i)}_j],[[a^{(i)}_{j+1}]}\right)'_{x(a^{(i)}_j),c^{(i)}_{j+1}} \left(D^{-1}_{[a^{(i)}_{j+1}],[a^{(i)}_{j+1}]}\right)'_{c^{(i)}_{j+1},x(a^{(i)}_{j+1})}. \tag{2.5.47}$$

Here $\mathbf{c}$ is any subset of $\{1,2\}^{(\sum_{i=1}^{2k} l_i)+1}$. So since the entries of D^{-1} are deterministic and for all $a_j^{(i)}$ the entries of $D^{-1}_{[a_j^{(i)}],[a_j^{[i]}]}$ are bounded by $|g_{[a_j^{(i)}]}| + |g_{-[a_j^{(i)}]}|$, it is true that

$$\left|\mathbf{E}\prod_{i=1}^{2k} X^{(l_i)}_{b_{2i-1},b_{2i}}\right| \leq \left|\sum_{\mathbf{a}}\prod_{i,j}\left(|g_{[a_j^{(i)}]}| + |g_{-[a_j^{(i)}]}|\right)\sum_{\mathbf{c}}\mathbf{E}\prod_{i=1}^{2k}\left(\mathbf{E}_{[a_1^{(i)}],[a_2^{(i)}]}\right)'_{c_1^{(i)},c_2^{(i)}}\prod_{j\neq 1}\left(\mathbf{E}_{[a_j^{(i)}],[[a_{j+1}^{(i)}]}\right)'_{x(a_j^{(i)}),c_{j+1}^{(i)}}\right|.$$

Next we will show an important inequality, necessary to estimate the expectation of the products in the previous equations.

Lemma 2.5.13. *It is true that for each array (a_j^i) with entries in $\mathbb{T}$,*

$$\left|\sum_{\mathbf{c}}\mathbf{E}\prod_{i=1}^{2k}\left(\mathbf{E}_{[a_1^{(i)}],[a_2^{(i)}]}\right)'_{c_1^{(i)},c_2^{(i)}}\prod_{j\neq 1}\left(\mathbf{E}_{[a_j^{(i)}],[[a_{j+1}^{(i)}]}\right)'_{x(a_j^{(i)}),c_{j+1}^{(i)}}\right| \leq_k \frac{(\psi\log(N))^{\sum l_i}(s+\eta)^{\sum_i l_i}}{(N\eta)^{\sum_{i=1} l_i/2}} \tag{2.5.48}$$

$$\cdot X\left([a_1^{(1)}],[a_2^{(1)}],[a_2^1],[a_3^1]\cdots,[a_{l_1+1}^{(1)}],[a_1^{(2)}]\cdots,[a_{l_2+1}^{(2)}]\cdots,[a_{l_{2k}}^{(2k)}]\right), \tag{2.5.49}$$

where $X(\cdot)$ is the indicator function that indicates if every element in the array appears an even number of times.

Proof. Note that by the definition of the matrices, one has that

$$\mathbf{E}_{[a_j^{(i)}],[a_{j+1}^{(i)}]} = \mathbf{E}^1_{[a_j^{(i)}],[a_{j+1}^{(i)}]} + \mathbf{E}^2_{[a_j^{(i)}],[a_{j+1}^{(i)}]} + \mathbf{E}^3_{[a_j^{(i)}],[a_{j+1}^{(i)}]}.$$

Moreover, after conditioning on the matrix $W'_{\mathbb{T},\mathbb{T}}$, the matrices $\mathbf{E}^1,\mathbf{E}^2,\mathbf{E}^3$ are independent since $\mathbf{E}^1$ is dependent only on $(F-z\mathbb{I})^{(\mathbb{T})}$ and is diagonal and deterministic, $\mathbf{E}^2$ depends only on $W'_{\mathbb{T},\mathbb{T}}$ and $\mathbf{E}^3$ depends on $G^{\mathbb{T}}$ and $W'_{[2N]\setminus\mathbb{T},\mathbb{T}}$. We will use the notation $\mathbf{E}_{\mathbb{T}}$ for the conditional expected value. So in order to prove Lemma 2.5.13 it suffices to show that

$$\begin{aligned}\left|\sum_{\mathbf{c}}\mathbf{E}_{\mathbb{T}}\prod_{i=1}^{2k}\left(\mathbf{E}^1_{[a_1^{(i)}],[a_2^{(i)}]}\right)'_{c_1^{(i)},c_2^{(i)}}\prod_{j\neq 1}\left(\mathbf{E}^1_{[a_j^{(i)}],[[a_{j+1}^{(i)}]}\right)'_{x(a_j^{(i)}),c_{j+1}^{(i)}}\right| &\leq_k \left(\frac{\psi s}{N\eta}\right)^{\sum l_i} X\left(\left([a_j^i]\right)_{i\in[2k],j\in[2l_i]}\right) \leq \\ &\leq X\left(\left([a_j^i]\right)_{i\in[2k],j\in[l_i]}\right)\frac{(\psi\log(N))^{\sum l_i}(s+\eta)^{\sum_i l_i}}{(N\eta)^{\sum_{i=1} l_i/2}},\end{aligned} \tag{2.5.50}$$

$$\begin{aligned}\left|\sum_{\mathbf{c}}\mathbf{E}_{\mathbb{T}}\prod_{i=1}^{2k}\left(\mathbf{E}^2_{[a_1^{(i)}],[a_2^{(i)}]}\right)'_{c_1^{(i)},c_2^{(i)}}\prod_{j\neq 1}\left(\mathbf{E}^2_{[a_j^{(i)}],[[a_{j+1}^{(i)}]}\right)'_{x(a_j^{(i)}),c_{j+1}^{(i)}}\right| &\leq_k \left(\frac{s}{N}\right)^{\sum l_i/2} X\left(\left([a_j^i]\right)_{i\in[2k],j\in[l_i]}\right) \leq \\ &\leq X(\left([a_j^i]\right)_{i\in[2k],j\in[l_i]})\frac{(\psi\log(N))^{\sum l_i}(s+\eta)^{\sum_i l_i}}{(N\eta)^{\sum_{i=1} l_i/2}},\end{aligned} \tag{2.5.51}$$

$$\begin{aligned}\left|\sum_{\mathbf{c}}\mathbf{E}_{\mathbb{T}}\prod_{i=1}^{2k}\left(\mathbf{E}^3_{[a_1^{(i)}],[a_2^{(i)}]}\right)'_{c_1^{(i)},c_2^{(i)}}\prod_{j\neq 1}\left(\mathbf{E}^3_{[a_j^{(i)}],[[a_{j+1}^{(i)}]}\right)'_{x(a_j^{(i)}),c_{j+1}^{(i)}}\right| &\leq_k \left(\frac{s\psi\log(N)}{\sqrt{N\eta}}\right)^{\sum l_i/2} X(\left([a_j^i]\right)_{i\in[2k],j\in[l_i]}) \leq \\ &X\left(\left([a_j^i]\right)_{i\in[2k],j\in[l_i]}\right)\frac{(\psi\log(N))^{\sum l_i}(s+\eta)^{\sum_i l_i}}{(N\eta)^{\sum_{i=1} l_i/2}}.\end{aligned} \tag{2.5.52}$$

This is true, since if we assume (2.5.50),(2.5.51),(2.5.52) hold for any array then

$$\left|\sum_{\mathbf{c}} \mathbf{E}_{\mathbb{T}} \prod_{i=1}^{2k} \left(\mathbf{E}_{[a_1^{(i)}],[a_2^{(i)}]}\right)'_{c_1^{(i)},c_2^{(i)}} \prod_{j\neq 1} \left(\mathbf{E}_{[a_j^{(i)}],[[a_{j+1}^{(i)}]}\right)'_{x(a_j^{(i)}),c_{j+1}^{(i)}}\right| \tag{2.5.53}$$

$$\leq \sum_{P^1,P^2,P^3} \left|\sum_{\mathbf{c}} \prod_{y\in[3]} \mathbf{E}_{\mathbb{T}} \prod_{i=1}^{2k} \mathbf{1}\left\{a_1^{(i)}, a_2^{(i)} \in P^y\right\} \left(\mathbf{E}^y_{[a_1^{(i)}],[a_2^{(i)}]}\right)'_{c_1^{(i)},c_2^{(i)}}\right| \cdot \tag{2.5.54}$$

$$\cdot \left|\prod_{j\neq 1} \mathbf{1}\left\{a_j^i, a_j^i \in P^y\right\} \cdot \left(\mathbf{E}^y_{[a_j^{(i)}],[[a_{j+1}^{(i)}]}\right)'_{x(a_j^{(i)}),c_{j+1}^{(i)}}\right| \tag{2.5.55}$$

$$\preceq_k \sum_{P^1,P^2,P^3} \prod_{y\in[3]} X\left([a_j^i] : a_j^i \in P^y\right)\left((s+\eta)\frac{\psi \log(N)}{\sqrt{N\eta}}\right)^{|P^y|} \leq \sum_{P^1,P^2,P^3} X([a_j^i])\left((s+\eta)\frac{\psi \log(N)}{\sqrt{N\eta}}\right)^{\sum l_i} \tag{2.5.56}$$

$$\preceq_k X([a_j^i])\left((s+\eta)\frac{\psi \log(N)}{\sqrt{N\eta}}\right)^{\sum l_i|}, \tag{2.5.57}$$

where the sum is taken over all 3-partitions P^1, P^2, P^3 of the set $\{i,j : i \in [2k], j \in [l_i+1]\}$ such that if $a_j^i \in P^y$ then $a_{j+1}^i \in P^y$ for each $j \in [l_i] \cap (2\mathbb{N}+1)$, $i \in [2k]$ and $y \in [3]$. So the number of these partitions depends only on k which implies the last inequality.

For the first inequality (2.5.50) note that the matrix $\mathbf{E}^1$ is diagonal and its diagonal entries are bounded by $s|\mathbb{T}|\frac{\psi}{N\eta}$ due to Theorem 4.5 of [3] and the interlacing properties of the minors of the eigenvalues. So it is implied that $|\mathbf{E}^1_{[a_j^i],a_{j+1}^i]}|_{op} \leq \mathbf{1}\left\{a_i^j = a_{j+1}^i\right\} s\frac{\psi}{N\eta}$. So

$$\sum_{\mathbf{c}} \left|\mathbf{E}_{\mathbb{T}} \prod_{i=1}^{2k} \left(\mathbf{E}^1_{[a_1^{(i)}],[a_2^{(i)}]}\right)'_{c_1^{(i)},c_2^{(i)}} \prod_{j\neq 1} \left(\mathbf{E}^1_{[a_j^{(i)}],[[a_{j+1}^{(i)}]}\right)'_{x(a_j^{(i)}),c_{j+1}^{(i)}}\right| \leq \sum_{\mathbf{c}} \mathbf{E}_{\mathbb{T}} \prod_{i=1}^{2k} \left|\mathbf{E}^1_{[a_1^{(i)}],[a_2^{(i)}]} \prod_{j\neq 1} \mathbf{E}^1_{[a_j^{(i)}],[[a_{j+1}^{(i)}]}\right|_{op} \tag{2.5.58}$$

$$\leq \sum_{\mathbf{c}} \prod_{i,j} \mathbf{1}\left\{a_j^i = a_{j+1}^i\right\} s\frac{\psi}{N\eta} \leq \sum_{\mathbf{c}} \prod_{i,j} \mathbf{1}\left\{[a_j^i] = [a_{j+1}^i]\right\} s\frac{\psi}{N\eta} \preceq_k \left(\frac{\psi s}{N\eta}\right)^{\sum l_i} X\left(\left([a_j^i]\right)_{i\in[2k],j\in[l_i+1]}\right). \tag{2.5.59}$$

For the second inequality (2.5.51) by the way $\mathbf{E}^2$ was defined, one can compute that

$$\left|\sum_{\mathbf{c}} \mathbf{E}_{\mathbb{T}} \prod_{i=1}^{2k} \left(\mathbf{E}^2_{[a_1^{(i)}],[a_2^{(i)}]}\right)'_{c_1^{(i)},c_2^{(i)}} \prod_{j\neq 1} \left(\mathbf{E}^2_{[a_j^{(i)}],[[a_{j+1}^{(i)}]}\right)'_{x(a_j^{(i)}),c_{j+1}^{(i)}}\right| \tag{2.5.60}$$

$$= \left|\sum_{c_1^i\in\{1,2\}} \mathbf{E}_{\mathbb{T}} \prod_{i=1}^{2k} \left(\mathbf{E}^2_{[a_1^{(i)}],[a_2^{(i)}]}\right)_{c_1^{(i)},[c_1^{(i)}+1]_2} \prod_{j\neq 1} \left(\mathbf{E}^2_{[a_j^{(i)}],[[a_{j+1}^{(i)}]}\right)_{x(a_j^{(i)}),[x(a_j^{(i)})+1]_2}\right|, \tag{2.5.61}$$

where $[a]_2$ the least positive integer such that it is equal to $a \mod 2$. Note that

$$\left(\mathbf{E}^2_{[a_j^i],[a_{j+1}^i]}\right)_{1,2} = -\sqrt{s} w_{a_j^i,a_{j+1}^i}$$

and $\left(\mathbf{E}^2_{[a^i_j],[a^i_{j+1}]}\right)_{2,1} = -\sqrt{s}w_{a^i_{j+1},a^i_j}$. Moreover, since the non zero entries of W' are independent, symmetric, normal random variables with variance $\frac{1}{N}$, the product is non-zero only if every pair $([a^i_j],[a^i_{j+1}])$ in the product appears an even number of times. All these imply that

$$\sum_{c^i_1\in\{1,2\}}\left|\mathbf{E}_{\mathbb{T}}\prod_{i=1}^{2k}\left(\mathbf{E}^2_{[a^{(i)}_1],[a^{(i)}_2]}\right)_{c^{(i)}_1,[c^{(i)}_1+1]_2}\prod_{j\neq 1}\left(\mathbf{E}^2_{[a^{(i)}_j],[[a^{(i)}_{j+1}]}\right)_{x(a^{(i)}_j),[x(a^{(i)}_j)+1]_2}\right| \preceq_k \sum_{c^i_1\in\{1,2\}}(\frac{s}{N})^{\sum l_i/2}X([a^i_j]),$$

which implies (2.5.51). The constant which is implied in the last inequality can be chosen to be $2k\prod_{j=1}^{\sum l_i}\mathbf{E}(\sqrt{N}w_{1,1})^{2j}$, which is a large constant depending only on k since $\sqrt{N}w_{1,1}\sim N(0,1)$.

For the third inequality, (2.5.52), one can show that

$$\sum_{\mathbf{c}}\mathbf{E}_{\mathbb{T}}\prod_{i=1}^{2k}\left(\mathbf{E}^3_{[a^{(i)}_1],[a^{(i)}_2]}\right)'_{c^{(i)}_1,c^{(i)}_2}\prod_{j\neq 1}\left(\mathbf{E}^3_{[a^{(i)}_j],[[a^{(i)}_{j+1}]}\right)'_{x(a^{(i)}_j),c^{(i)}_{j+1}} \tag{2.5.62}$$

$$= \sum_{\{(c^i_{2j-1},c^i_{2j})\in\mathbf{c}:a^i_{2j-1}\neq a^i_{2j}\}}\mathbf{E}_{\mathbb{T}}\left(\prod_{i=1}^{2k}\left(\mathbf{E}^3_{[a^{(i)}_1],[a^{(i)}_2]}\right)'_{c^{(i)}_1,c^{(i)}_2}+\mathbf{1}\left\{a^{(i)}_1=a^i_2\right\}\left(\mathbf{E}^3_{[a^{(i)}_1],[a^{(i)}_2]}\right)'_{[c^{(i)}_1+1]_2,[c^{(i)}_2+1]_2}\right)\cdot \tag{2.5.63}$$

$$\cdot\left(\prod_{j\neq 1}\left(\mathbf{E}^3_{[a^{(i)}_j],[[a^{(i)}_{j+1}]}\right)_{x(a^{(i)}_j),c^i_{j+1}}+\mathbf{1}\left\{a^i_j=a^i_{j+1}\right\}\left(\mathbf{E}^3_{[a^{(i)}_j],[[a^{(i)}_{j+1}]}\right)_{[x(a^{(i)}_j)+1]_2,[c^i_{j+1}+1]_2}\right) \tag{2.5.64}$$

$$= \sum_{\mathbf{c}}\mathbf{E}_{\mathbb{T}}\prod_{i=1}^{2k}\left(\tilde{\mathbf{E}}^3_{[a^{(i)}_1],[a^{(i)}_2]}\right)'_{c^{(i)}_1,c^{(i)}_2}\prod_{j\neq 1}\left(\tilde{\mathbf{E}}^3_{[a^{(iF)}_j],[[a^{(i)}_{j+1}]}\right)'_{x(a^{(i)}_j),c^{(i)}_{j+1}}, \tag{2.5.65}$$

where

$$\left(\tilde{\mathbf{E}}^3_{[i],[j]}\right)_{l,m} = \begin{cases}\left(\mathbf{E}^3_{[i],[j]}\right)_{l,m}\mathbf{1}\{[i]\neq[j]\text{ or }l\neq m\} \\ \left(\mathbf{E}^3_{[i],[j]}\right)_{l,m}+\frac{s}{N}\mathbf{1}\{l=1\}\sum_{i\in[N]\cap[2N]\setminus\mathbb{T}}G^{\mathbb{T}}_{ii}+\frac{s}{N}\mathbf{1}\{l=2\}\sum^N_{i\in[N]\cap[2N]\setminus\mathbb{T}}G^{\mathbb{T}}_{i+N,i+N}, & \text{else}\end{cases}$$

So by construction one can compute that

$$\left(\tilde{\mathbf{E}}^3\right)_{[i],[j]} =$$
$$s\begin{bmatrix}\sum_{f,k\notin\mathbb{T}\cup 2N\setminus[N]}(w_{j,f}w_{k,i}-\mathbf{1}_{\{[i]=[j]\}}\mathbf{1}_{\{f=k\}}\frac{1}{N})G^{\mathbb{T}}_{k+N,f+N} & \sum_{f,k\notin\mathbb{T}\cup 2N\setminus[N]}w_{f,j}w_{k,i}G^{\mathbb{T}}_{k,f+N} \\ \sum_{f,k\notin\mathbb{T}\cup 2N\setminus[N]}w_{j,f}w_{i,k}G^{\mathbb{T}}_{k+N,f} & \sum_{f,k\notin\mathbb{T}\cup 2N\setminus[N]}(w_{f,j}w_{i,k}-\mathbf{1}_{\{[i]=[j]\}}\mathbf{1}_{\{f=k\}}\frac{1}{N})G^{\mathbb{T}}_{k,f}\end{bmatrix} \tag{2.5.66}$$

As a result

$$\mathbf{E}_{\mathbb{T}}\sum_{\mathbf{c}}\prod_{i=1}^{2k}\left(\tilde{\mathbf{E}}^{3}_{[a_1^{(i)}],[a_2^{(i)}]}\right)'_{c_1^{(i)},c_2^{(i)}}\prod_{j\neq 1}\left(\tilde{\mathbf{E}}^{3}_{[a_j^{(i)}],[[a_{j+1}^{(i)}]}\right)'_{x(a_j^{(i)}),c_{j+1}^{(i)}} = \tag{2.5.67}$$

$$s^{\sum l_i}\sum_{\mathbf{c}}\sum_{\beta_1^1,\beta_2^1\cdots..\beta_{2l_{2k}}^{2k}\notin\mathbb{T}}\mathbf{E}_{\mathbb{T}}\prod_{i=1}^{2k}\left[\left(W'_{[a_1^i],[\beta_1^i]}\right)_{[c_1^i+1]_2,c_1^i}\left(W'_{[a_2^i],[\beta_2^i]}\right)_{c_2^i,[c_2^i+1]_2}-\frac{\mathbf{1}_{a_1^i=\beta_1^i}\mathbf{1}_{a_2^i=\beta_2^i}}{N}\right]\cdot \tag{2.5.68}$$

$$\cdot\prod_{j\neq 1}\left[\left(W'_{[a_j^i],[\beta_j^i]}\right)_{[x(a_j^i)+1]_2,x(a_j^i)}\left(W'_{[a_{j+1}^i],[\beta_{j+1}^i]}\right)_{c_j^i,[c_{j+1}^i+1]_2}-\frac{\mathbf{1}_{a_j^i=\beta_j^i}\mathbf{1}_{a_{j+1}^i=\beta_{j+1}^i}}{N}\right]\cdot \tag{2.5.69}$$

$$\cdot\prod_{i=1}^{2k}\left(G^{\mathbb{T}}_{[\beta_1^i],[\beta_2^i]}\right)_{[c_1^i+1]_2,[c_2^i+1]_2}\prod_{j\neq 1}\left(G^{\mathbb{T}}_{[\beta_j^i],[\beta_{j+1}^i]}\right)_{[x(a_j^i)+1]_2,[c_j^i+1]_2}. \tag{2.5.70}$$

Next set $\mathcal{G}$ to be the graph with vertices $\{[\beta_1^1],[\beta_2^1],\cdots\cdots[\beta_{2l_{2k}}^{2k}]\}$ and with edges the successive terms $([\beta_{2j-1}^i],[\beta_{2j}^i])$. Set $\rho(\mathcal{G})$ the indicator function that every vertex of $\mathcal{G}$ is adjacent to at least two edges, $v=\{[\beta_1^1],[\beta_2^1],\cdots[\beta_{2l_{2k}}^{2k}]\}$, $\gamma_{r\in[v]}$ the non-repeating vertices of $\mathcal{G}$, d_r the multiplicity of γ_r and o the number of self loops in $\mathcal{G}$. So, by (2.5.11), (2.5.10) and (2.5.12) one has that with overwhelming probability

$$\rho(\mathcal{G})\left|\prod_{i=1}^{2k}\left(G^{\mathbb{T}}_{[\beta_1^i],[\beta_2^i]}\right)_{[c_1^i+1]_2,[c_2^i+1]_2}\prod_{j\neq 1}\left(G^{\mathbb{T}}_{[\beta_j^i],[\beta_{j+1}^i]}\right)_{[x(a_j^i)+1]_2,[c_j^i+1]_2}\right| \tag{2.5.71}$$

$$\preceq_k \rho(\mathcal{G})\frac{\psi^{\sum l_i}}{\sqrt{N}\eta^{\sum l_i - o}}\prod_{r\in[v]}\left(|g_{\gamma_r}|^{d_r/2}+|g_{-\gamma_r}|^{d_r/2}\right). \tag{2.5.72}$$

Thus one can show similarly to the proof of (2.40) in [6] that the following holds with overwhelming probability

$$\rho(\mathcal{G})\sum_{\beta_j^i\notin\mathbb{T}}\left|\prod_{i=1}^{2k}\left(G^{\mathbb{T}}_{[\beta_1^i],[\beta_2^i]}\right)_{[c_1^i+1]_2,[c_2^i+1]_2}\prod_{j\neq 1}\left(G^{\mathbb{T}}_{[\beta_j^i],[\beta_{j+1}^i]}\right)_{[x(a_j^i)+1]_2,[c_j^i+1]_2}\right| \preceq_k \frac{(\psi\log(N))^{\sum l_i}N^{\sum l_i/2}}{\eta^{\sum l_i/2}}. \tag{2.5.73}$$

Next we need to bound the quantity

$$\mathbf{E}_{\mathbb{T}}\prod_{i=1}^{2k}\left[\left(W'_{[a_1^i],[\beta_1^i]}\right)_{[c_1^i+1]_2,c_1^i}\left(W'_{[a_2^i],[\beta_2^i]}\right)_{c_2^i,[c_2^i+1]_2}-\frac{\mathbf{1}_{a_1^i=\beta_1^i}\mathbf{1}_{a_2^i=\beta_2^i}}{N}\right]\cdot \tag{2.5.74}$$

$$\cdot\prod_{j\neq 1}\left[\left(W'_{[a_j^i],[\beta_j^i]}\right)_{[x(a_j^i)+1]_2,x(a_j^i)}\left(W'_{[a_{j+1}^i],[\beta_{j+1}^i]}\right)_{c_j^i,[c_{j+1}^i+1]_2}-\frac{\mathbf{1}_{a_j^i=\beta_j^i}\mathbf{1}_{a_{j+1}^i=\beta_{j+1}^i}}{N}\right]. \tag{2.5.75}$$

Note, that in order for the product to be different than 0, every pair $([a_i^j],[b_i^j])$ must appear an even number of times. Moreover in order for the product to be different than 0, for each i,j the number of consecutive pairs $([a_r^m],[\beta_r^m])$ and $([a_{r+1}^m],[\beta_{r+1}^m])$ for $m\in[2k]$ and $r\in[l_m]$, such that exactly one of them is equal to $[a_j^i,\beta_j^i]$, must be also even. Furthermore, for each i,j, if such pairs do not exist, then the number of consecutive pairs which are both equal to $[a_i^j,\beta_i^j]$ must be at least 2, or else the product would be 0. The latter is true since either the square of a centered Gaussian random variable minus its variance would appear, either the product of two independent centered Gaussian random variables would appear.

So it is implied that in order for the product above to not be zero, it is demanded that $\rho(\mathcal{G}) = 1$ and $X([a_i^j]) = 1$. Here $\mathcal{G}$ is the graph which is associated with β_i^j. So by a trivial bounding in the moments of Gaussian random variables one can show that

$$\left|\mathbf{E}_{\mathbb{T}} \prod_{i=1}^{2k} \left[\left(W'_{[a_1^i],[\beta_1^i]}\right)_{[c_1^i+1]_2, c_1^i} \left(W'_{[a_2^i],[\beta_2^i]}\right)_{c_2^i,[c_2^i+1]_2} - \frac{\mathbf{1}_{a_1^i=\beta_1^i}\mathbf{1}_{a_2^i=\beta_2^i}}{N}\right]\cdot \tag{2.5.76}$$

$$\cdot \prod_{j\neq 1} \left[\left(W'_{[a_j^i],[\beta_j^i]}\right)_{[x(a_j^i)+1]_2, x(a_j^i)} \left(W'_{[a_{j+1}^i],[\beta_{j+1}^i]}\right)_{c_j^i,[c_{j+1}^i+1]_2} - \frac{\mathbf{1}_{a_j^i=\beta_j^i}\mathbf{1}_{a_{j+1}^i=\beta_{j+1}^i}}{N}\right]\right| \lesssim_k N^{-\sum l_i}\rho(\mathcal{G})X([a_j^i]). \tag{2.5.77}$$

Thus, the proof of the lemma is complete after combining (2.5.76) and (2.5.73). □

We are now ready to present the proof of Theorem 2.5.7.

Proof of Theorem 2.5.7. Note that Lemma 2.5.13 holds for every sequence of indexes. In our case though, by construction, every term in $[a_j^i]$ appears a non-zero even number of times since they appear consecutive times for $j \neq 1, l_i+1$. So one has that $X([a_i^j]_{i\in[2k],j\in[l_i+1]}) = X([a_i^j]_{i\in[2k],j\in\{1,l_i+1\}}) = X([\mathbf{B}])$. So by a direct application of Lemma 2.5.13

$$\sum_{\mathbf{a}}\sum_{\mathbf{c}}\left|\mathbf{E}\prod_{i=1}^{2k}\left(D^{-1}_{[a_1^{(i)}],[a_1^{(i)}]}\right)'_{x(a_1^{(i)}),c_1^{(i)}}\left(\mathbf{E}_{[a_1^{(i)}],[a_2^{(i)}]}\right)'_{c_1^{(i)},c_2^{(i)}}\left(D^{-1}_{[a_2^{(i)}],[a_2^{(i)}]}\right)'_{c_2^{(i)},x(a_2^{(i)})}\cdot\right. \tag{2.5.78}$$

$$\left.\cdot\prod_{j\neq 1}\left(\mathbf{E}_{[a_j^{(i)}],[[a_{j+1}^{(i)}]}\right)'_{x(a_j^{(i)}),c_{j+1}^{(i)}}\left(D^{-1}_{[a_{j+1}^{(i)}],[a_{j+1}^{(i)}]}\right)'_{c_{j+1}^{(i)},x(a_{j+1}^{(i)})}\right| \tag{2.5.79}$$

$$\lesssim_k \sum_{\mathbf{a}}\prod_{i,j}\left(|g_{[a_j^{(i)}]}|+|g_{-[a_j^{(i)}]}|\right)\frac{(\psi\log(N))^{\sum l_i}(s+\eta)^{\sum_i l_i}}{(N\eta)^{\sum_{i=1} l_i/2}}X(\mathbf{B}). \tag{2.5.80}$$

As a result

$$\sum_{1\leq l_1,l_2,\cdots,l_{2k}\leq f-1}\sum_{\mathbf{B}}\sum_{b_i=\{B_i,B_i+N\}}|q_{b_1}||q_{b_2}||q_{b_3}|\cdots|q_{b_{4k}}|\left|\mathbf{E}\prod_{i=1}^{2k}X^{(l_i)}_{b_{2i-1},b_{2i}}\right| \tag{2.5.81}$$

$$\lesssim_k \sum_{1\leq l_1,l_2,\cdots,l_{2k}\leq f-1}\sum_{\mathbf{B}}\frac{(\psi\log(N))^{\sum l_i}(s+\eta)^{\sum_i l_i}}{(N\eta)^{\sum_{i=1} l_i/2}}X(\mathbf{B})\cdot \tag{2.5.82}$$

$$\cdot\sum_{b_i=\{B_i,B_i+N\}}|q_{b_1}||q_{b_2}||q_{b_3}|\cdots|q_{b_{4k}}|\sum_{\mathbf{a}}\prod_{i,j}\left(|g_{[a_j^{(i)}]}|+|g_{-[a_j^{(i)}]}|\right) \tag{2.5.83}$$

$$\lesssim \sum_{1\leq l_1,l_2,\cdots,l_{2k}\leq f-1}\sum_{\mathbf{B}}\frac{(\psi\log(N))^{\sum l_i}(s+\eta)^{\sum_i l_i}}{(N\eta)^{\sum_{i=1} l_i/2}}X(\mathbf{B})\prod_{i=1}^{4k}(|q_{B_i}|+|q_{B_i+N}|)\sum_{\mathbf{A}}\prod_{i,j}\left(|g_{[a_j^{(i)}]}|+|g_{-[a_j^{(i)}]}|\right), \tag{2.5.84}$$

where the sum now is considered over all $\mathbf{A}\subseteq\mathbf{B}^{\sum l_i+2k}$ with the restriction that $[a_1^i] = B_{2i-1}$ and $[a_{l_i+1}^i] = B_{2i}$.

Moreover note that the array $[a^i_j]$ defines a partition on the set $\{(i,j) : i \in [2k], j \in [l_i + 1]\}$ such that (i,j) belongs to the same block of the partition with (i',j') if and only if $a^i_j = a^{j'}_{i'}$. Furthermore, denote $n = |\mathbf{B}|$, d_i the number of times the $i-th$ element of $\mathbf{B}$, which we denote with γ_i, appears without repetition and r_i such that $r_i + d_i$ is the number of times the $i-th$ element of $\mathbf{B}$ appears in $\mathbf{A}$. Note that since we are interested in the sequences that $X(\mathbf{B}) = 1$, it is implied that d_i are all even. So it is true that

$$\sum d_i = 2k, \quad 2k + \sum r_i = \sum l_i.$$

Moreover, notice that each induced partition mentioned before, uniquely determines the quantities d_i, l_i and each block of the partition has at least two elements, since $X(\mathbf{B}) = 1$. So we can modify the sum, into first summing over all partitions P and then over all $\mathbf{A}$-possible choices in the partition. Note that $\mathbf{B}$ is completely described by the set $\mathbf{A}$. So one has that

$$(2.5.84) = \sum_{1 \le l_1, l_2 \cdots, l_{2k} \le f-1} \frac{(\psi \log(N))^{\sum l_i} (s+\eta)^{\sum_i l_i}}{(N\eta)^{\sum_{i=1} l_i/2}} \sum_P \sum_{\mathbf{A} \sim P} X\left(\{[a^i_j]\}_{i \in [2k], j \in \{1, l_i+1\}}\right) \cdot \tag{2.5.85}$$

$$\cdot \prod_{i=1}^{2k} \left(|q_{[a^i_1]}| + |q_{[a^i_1]+N}|\right)\left(|q_{[a^i_{l_i+1}]}| + |q_{[a^i_{l_i+1}]+N}\right) \prod_{i,j} \left(|g_{[a^{(i)}_j]}| + |g_{-[a^{(i)}_j]}|\right) \lesssim_k \tag{2.5.86}$$

$$\lesssim_k \sum_{1 \le l_1, l_2 \cdots, l_{2k} \le f-1} \frac{(\psi \log(N))^{\sum l_i} (s+\eta)^{\sum_i l_i}}{(N\eta)^{\sum_{i=1} l_i/2}} \cdot \tag{2.5.87}$$

$$\sum_P \sum_{1 \le \gamma_1, \gamma_2 \cdots, \gamma_n \le N} \prod_{i=1}^{n} (|q_{\gamma_i}|^{d_i} + |q_{\gamma_i+N}|^{d_i})(|g_{-\gamma_i}|^{d_i+r_i} + |g_{\gamma_i}|^{d_i+r_i}) \tag{2.5.88}$$

$$\lesssim_k \sum_{1 \le l_1, l_2 \cdots, l_{2k} \le f-1} \sum_P \frac{(\psi \log(N))^{\sum l_i} (s+\eta)^{\sum_i l_i}}{(N\eta)^{\sum_{i=1} l_i/2}} \frac{\operatorname{Im}\left(\sum (q_i^2 + q_{i+N}^2)(g_i + g_{-i})^{2k}\right)}{(s+\eta)^{\sum l_i}} \lesssim_k Y^{2k}, \tag{2.5.89}$$

where in the last inequality we used the fact that $\psi \log(N)(\sqrt{N\eta})^{-1} \le 1$, the fact that $\sum l_i \ge 2k$, and the fact that both the number of partitions and the number of possible l_i are bounded by constants depending only on k. For the second to last inequality, we used Proposition 2.18-inequality (2.38) in [6], the facts that $d_i \ge 2$ and that $\sum d_i = 2k$.
This finishes the proof of Theorem 2.5.7 □

2.5.4 Bounding the pertubed matrices at the optimal scale

At this subsection we are going to essentially bound the entries of the resolvent $G(s,z)$ at the optimal scale $\operatorname{Im}(z) = N^{\epsilon-1}$, for all matrices $\tilde{V}$ that are initially bounded by an $N-$dependent parameter. Next, we will apply this result to the matrix X, which is initially bounded due to (2.3.17) with high probability. Thus we will prove that the matrix X, after slightly perturbing it, has essentially bounded resolvent entries at the optimal scale $N^{\delta-1}$, for any small enough, positive δ.

Proposition 2.5.14. *Let V be an $N \times N$ matrix and consider $\tilde{V}$ the symmetrization of D. Suppose that $\tilde{V}$ satisfies Assumption 2.5.1 for some parameters h_*, r at energy level $E_0 = 0$ and that there exists an $N-$dependent parameter $B \in (0, \frac{1}{h_*})$ such that $\max_j |(\tilde{V} - zI)^{-1}_{j,j}| \leq B$. Then for any $\delta > 0$ and $s : N^\delta h_* \leq s \leq rN^{-\delta}$, it is true that for any $D > 1$ there exists $C = C(\delta, D)$ such that*

$$\mathbb{P}\left(\sup_{\mathbb{D}} \sup_{i,j} |G(s, z)| \geq BN^\delta\right) \leq CN^{-D},$$

where $G(s, z) = (\tilde{V} + \sqrt{s}W - z\mathbb{I})^{-1}$, W is the symmetrization of an i.i.d. Gaussian matrix with centered entries and variance $\frac{1}{N}$ and $\mathbb{D} = \{E + i\eta : E \in (-\frac{r}{2}, \frac{r}{2}), \eta \in [N^{\delta-1}, 1 - \frac{r}{2}]\}$.

Proof. By a direct application of Theorem 2.5.6 for $q_k = \mathbf{1}\{i = k\}$ for any $k \in [2N]$ (without loss of generality suppose $k \in [N]$), one has that with overwhelming probability uniformly on $\mathbb{D}$ it is true that,

$$|G_{k,k}(s, z)| \leq \sum_{i=-N}^{N} (|g_i| + |g_{-i}|) \langle u_{i+N}(0), q_k\rangle^2 + \sum_{i=1}^{2N} (|g_i| + |g_{-i}|) |\langle u_i(0), q_k\rangle| |\langle u_{i+N}(0), q_k\rangle| \tag{2.5.90}$$

$$+ \frac{N^{\delta/2}}{\sqrt{N\eta}} \operatorname{Im}(\sum_{i=1}^{N} (g_i + g_{-i})(\langle u_i, q_k\rangle + \langle u_{i+N}, q_k\rangle)). \tag{2.5.91}$$

Note that by definition the $k - th$ element of each of the columns/rows of U is 0 for all the columns/rows with index larger that N. Moreover by definition $N^\delta \leq N\eta$. So it is implied that the above bound becomes

$$|G_{k,k}(s, z)| \leq \sum_{i=1}^{N} (|g_i + |g_{-i}|)|u_{k,i}|.$$

Furthermore, due to Schur's complement formula, one can prove, as in Lemma 2.3.28, that

$$G_{k,k} = z\tilde{G}_{k,k}(z^2),$$

where $\tilde{G}$ is the resolvent of the matrix $D^T D$. Moreover, one may compute that

$$D^T D = \left((U)_{i\in[N],j\in[N]}\right)^T \Sigma^2 (U)_{i\in[N],j\in[N]},$$

where Σ is the diagonal matrix with the singular values of D. So it is true that

$$\left|\sum_{i=1}^{N} u_{k,i}^2 \frac{1}{\hat{\lambda}_i^2 - z^2}\right| = \left|\tilde{G}_{k,k}(z^2)\right| = \left|\frac{1}{z} G_{k,k}(z)\right| \leq \frac{B}{|z|},$$

with overwhelming probability uniformly on $z \in \{z = E + i\eta : E \in (-r, r), h_* \leq \eta \leq 1\}$. Thus if we consider the sets $\mathcal{A}_m(0) = (2^{m-1}h_*, 2^m h_*) \cup (-2^m h_*, -2^{m-1}h_*)$ and set $z = i\eta_*$ it is true that

$$\max_{j\in[N]} \left|\sum_{\hat{\lambda}_i \in \mathcal{A}_m} u_{k,i}^2\right| \leq \min\{B\eta_*^2 2^{4m}, 1\}, \tag{2.5.92}$$

where the bounding by 1 is true due to the fact that the eigenvectors are considered normalized. After this observation the proof continues in a completely analogous way to the Proof of Proposition 3.9 in [5] and so it is omitted.

□

Corollary 2.5.15. *Adopt the notation of Section 2.3. Let $\mathcal{A}$ be the set mentioned in Theorem 2.3.13. For all $a \in (0,2) \setminus \mathcal{A}$ consider the matrix $X + \sqrt{t}W$, where $t = t(N)$ is defined in Definition 2.2.9. Set $\{T\}_{i,j \in [2N]}(z)$ the resolvent of $X + \sqrt{t}W$ at z. Then it is true that for any $D > 0$ and $\delta > 0$, there exists a constant $C' = C'(a, \nu, \rho, \delta, D)$ such that*

$$\mathbb{P}\left(\sup_{\mathbb{D}_\delta} \sup_{i,j} |T_{i,j}(z)| \geq N^{\delta}\right) \leq C' N^{-D}, \tag{2.5.93}$$

where $\mathbb{D}_{C_a,\delta} = \{E + i\eta : E \in (-\frac{1}{2C_a}, \frac{1}{2C_a}), \eta \in [N^{\delta-1}, \frac{1}{4C_a}]\}$ where C_a is the constant mentioned in Theorem 2.3.13.

Proof. Due to (2.3.17), (2.3.15) and since t belongs to the desired interval $(N^{2\delta-\frac{1}{2}}, N^{-2\delta})$, as is mentioned in the proof of Corollary 2.4.5, the proof of Corollary 2.5.15 is just an application of Proposition 2.5.14 to our set of matrices. □

Remark 2.5.16. Note that bounding the entries of the resolvent of $X + \sqrt{t}W$ as we did in Corollary 2.5.15 at scale $N^{\delta-1}$, implies the complete eigenvector delocalization in the sense of Theorem 2.1.2. The proof of the latter claim is well-known and can be found in the proof of Theorem 6.3 in [49].

2.6 Establishing universality of the least singular value and eigenvector delocalization

Thus far, we have proven both universality of the least singular value, Corollary 2.4.5, and complete eigenvector de-localization, Remark 2.5.16, for the matrix $X + \sqrt{t}W$ in the sense of Theorem 2.1.2. What we need to prove next, is that the transition from $X + \sqrt{t}W$ to $X + A$ is smooth enough to preserve both the eigenvector delocalization and universality of the least singular value. A first step to that direction is Theorem 2.6.4, whose proof is more or less the same as its symmetric counterpart in [5]. Furthermore what we manage, is to extend Theorem 3.15 of [5] to its "integrating analogue" in Proposition 2.6.7, which is not very difficult given Theorem 2.6.4. Proposition 2.6.7 is the milestone for the comparison of the least positive eigenvalues of $X + A$ and $X + \sqrt{t}W$.

Firstly, we will use a convenient decomposition of the elements of H in order to express the dependence of the "small" and the "large" entries of H with Bernoulli random variables.

Definition 2.6.1. Define the following random variables for $i,j \in [2N] : |i-j| \geq N$

$$\psi_{i,j} = \mathbb{P}\left(|h_{i,j}| \geq N^{-\rho}\right), \qquad x_{i,j} = \frac{\mathbb{P}\left(|h_{i,j}| \in (N^{-\nu}, N^{-\rho})\right)}{\mathbb{P}\left(|h_{i,j}| \leq N^{-\rho}\right)}$$

and

$$\mathbb{P}[a_{i,j} \in I) = \frac{\mathbb{P}(h_{i,j} \in I \cap (-N^{-\nu}, N^{-\nu})]}{\mathbb{P}(|H_{i,j}| < N^{-\nu}))}, \quad \mathbb{P}(c_{i,j} \in I) = \frac{\mathbb{P}(h_{i,j} \in (-\infty, N^{-\rho}) \cup (N^{\rho}, \infty) \cap I)}{\mathbb{P}(|h_{i,j}| \geq N^{-\rho})},$$

$$\mathbb{P}(b_{i,j} \in I) = \frac{\mathbb{P}(|h_{i,j}| \in (N^{-\nu}, N^{-\rho}) \cap h_{i,j} \in I)}{\mathbb{P}(|h_{i,j}| \in [N^{-\nu}, N^{-\rho}])},$$

for any interval I, subset of $\mathbb{R}$.

Moreover we define each bunch of

$$\{x_{i,j}\}_{i,j \in 2N:|i-j| \geq N}, \{\psi_{i,j}\}_{i,j \in [2N]:|i-j| \geq N}, \{a_{i,j}\}_{i,j \in 2N:|i-j| \geq N}, \tag{2.6.1}$$

$$\{b_{i,j}\}_{i,j \in 2N:|i-j| \geq N}, \{c_{i,j}\}_{i,j \in 2N:|i-j| \geq N} \tag{2.6.2}$$

to be independent up to symmetry and independent amongst them for different indexes i,j.

Definition 2.6.2. Define the following matrices

$$A_{i,j} = \begin{cases} (1-\psi_{i,j})(1-x_{i,j})a_{i,j}, & i,j : |i-j| \geq N \\ 0, & \text{otherwise} \end{cases} \tag{2.6.3}$$

$$B_{i,j} = \begin{cases} (1-\psi_{i,j})x_{i,j}b_{i,j}, & i,j : |i-j| \geq N \\ 0, & \text{otherwise} \end{cases} \tag{2.6.4}$$

$$C_{i,j} = \begin{cases} \psi_{i,j}c_{i,j}, & i,j : |i-j| \geq N \\ 0, & \text{otherwise} \end{cases} \tag{2.6.5}$$

$$\Psi_{i,j} = \begin{cases} \psi_{i,j}, & i,j : |i-j| \geq N \\ 0, & \text{otherwise} \end{cases} \tag{2.6.6}$$

Note that by definition $H = A + B + C$ and $X = B + C$.

Next we define the way to quantify the transition from $X + \sqrt{t}W$ to $X + A$.

Definition 2.6.3. Define the matrices

$$H^{\gamma} = \gamma A + \sqrt{t}(1-\gamma^2)^{1/2}W + X, \qquad \gamma \in [0,1]$$

and $G^{\gamma}(z) = (H^{\gamma} - z\mathbb{I})^{-1}$.

2.6.1 Green function Comparison

Next we present a comparison theorem for the resolvent entries of H^γ.

Theorem 2.6.4. *Let a, b, ρ, ν be constants that satisfy (2.2.1). Additionally suppose that $a \in (0,2) \setminus \mathcal{A}$ as in Theorem 2.3.14. Moreover let $F : \mathbb{R} \to \mathbb{R}$ such that*

$$\sup_{|x| \le 2N^{\epsilon}} \left| F^{(\mu)}(x) \right| \le N^{C_0 \epsilon}, \qquad \sup_{|x| \le 2N^2} \left| F^{(\mu)}(x) \right| \le N^{C_0}, \tag{2.6.7}$$

for some absolute constant $C_0 > 0$, for some integer $n = n(a, b, \rho, \nu, C_0)$ sufficiently large and any $\epsilon > 0$ and $\mu \in [n]$. Furthermore fix $z = E + i\eta$ for $E \in \mathbb{R}$ and $\eta \ge N^{-2}$. Moreover for any matrix Ψ denote $\mathbf{E}_\Psi$ the conditional expectation with respect to Ψ. Set

$$\Xi(z) = \sup_{\gamma \in [0,1]} \max_{\mu \in [n]} \max_{i,j \in [2N]} \mathbf{E}_\Psi \left| F^{(\mu)} \operatorname{Im}(G^{\gamma}_{i,j}(z)) \right|, \tag{2.6.8}$$

$$\Omega_0(z, \epsilon) = \left\{ \sup_{i,j} |G^{\gamma}_{i,j}(z)| \le N^{\epsilon} \right\}, \ \mathcal{Q}_0(z, \epsilon) = 1 - \mathbb{P}_\Psi \left(\Omega_0(z, \epsilon) \right). \tag{2.6.9}$$

Then there exist $\epsilon = \epsilon(a, b, \rho, \nu)$ and $\omega = \omega(a, b, \rho, \nu)$ such that for any matrix Ψ with at most $N^{1+a\rho+\epsilon}$ non-zero entries, there exists a constant $C = C(a, \nu, \rho)$ so that

$$\sup_{\gamma \in [0,1]} \left| \mathbf{E}_\Psi F\left(\operatorname{Im}(G^{\gamma}_{i,j}(z)) \right) - \mathbf{E}_\Psi F\left(\operatorname{Im}(G^{0}_{i,j}(z)) \right) \right| \le CN^{-\omega} \left(\Xi(z) + 1 \right) + C\mathcal{Q}_0(z, \epsilon) N^{C + C_0}, \forall i, j \in [2N]. \tag{2.6.10}$$

A similar bound to (2.6.10) can be proven, if one replaces $\operatorname{Im}(G^{\gamma}_{i,j}(z))$ and $\operatorname{Im}(G^{0}_{i,j}(z))$ with $\operatorname{Re}(G^{\gamma}_{i,j}(z)))$ and $\operatorname{Re}(G^{0}_{i,j}(z)))$ respectively.

Proof. The proof is similar to the proof of Theorem 3.15 in [5]. Next we give a short description of the main ideas behind the proof. We will do so only for the imaginary parts $\operatorname{Im}(G^{\gamma}_{i,j}(z))$. The proof for the real parts $\operatorname{Re}(G^{\gamma}_{i,j}(z))$ is completely analogous.
Fix $z \in \mathbb{C}$ and $F : \mathbb{R} \to \mathbb{R}$ satisfying the hypothesis of Theorem 2.6.4.
Firstly note that since $G^\gamma = G^\gamma (H^\gamma - z\mathbb{I}) G^\gamma$, it is true that

$$\frac{d}{d\gamma} G^\gamma = \frac{d}{d\gamma} G^\gamma (H^\gamma - z\mathbb{I}) G^\gamma + G^\gamma (H^\gamma - z\mathbb{I}) \frac{d}{d\gamma} G^\gamma + G^\gamma \frac{d}{d\gamma} (H^\gamma - z\mathbb{I}) G^\gamma$$

So it is implied that

$$-\frac{d}{d\gamma} G^\gamma = G^\gamma \frac{d}{d\gamma} (H^\gamma) G^\gamma, \tag{2.6.11}$$

where the derivative $\frac{d}{d\gamma}$ is considered in every entry. So by (2.6.11) and Leibniz integral rule, it is true that

$$\left| \frac{d}{d\gamma} \mathbf{E}_\Psi G^{\gamma}_{i,j} \right| = \left| \sum_{p,q \in [2N] : |p-q| \ge N} \mathbf{E}_\Psi G^{\gamma}_{i,p} \left(A_{p,q} - \frac{\gamma t^{1/2} w_{p,q}}{(1-\gamma^2)^{1/2}} \right) G^{\gamma}_{q,j} \right|.$$

Thus, in order to prove (2.6.10) it is sufficient to show that there exists a constant $C = C(a,\upsilon,\rho) > 0$ such that for all $\gamma \in (0,1)$

$$\begin{aligned}&\sum_{p,q\in[2N]:|p-q|\geq N}\left|\mathbf{E}_{\Psi}\operatorname{Im}(G^{\gamma}_{i,p}G^{\gamma}_{q,j})\left(A_{p,q}-\frac{\gamma t^{1/2}w_{p,q}}{(1-\gamma^2)^{1/2}}\right)F'(\operatorname{Im}(G^{\gamma}_{i,j})\right|\\&\leq \frac{C}{(1-\gamma^2)^{1/2}}(N^{-\omega}(\Xi+1)+\mathcal{Q}_0N^{C+C_0})\end{aligned} \tag{2.6.12}$$

and then integrate over any interval of the form $(0,\gamma')$ with $\gamma' \in (0,1]$. The proof of (2.6.12) is completely analogous to the proof of Proposition 4.4 in [5]. So we will give a sketch of the proof. Firstly fix $p,q \in [2N] : |p-q| \geq N$ and set the matrices

$$D_{a,b} = \begin{cases} H^{\gamma}_{a,b}, & (i,j) \notin \{(p,q),(q,p)\} \\ X_{p,q}, & \text{else} \end{cases} \tag{2.6.13}$$

$$E_{a,b} = \begin{cases} H^{\gamma}_{a,b}, & (a,b) \notin \{(p,q),(q,p)\} \\ C_{p,q}, & \text{else} \end{cases}. \tag{2.6.14}$$

Moreover set

$$\Gamma = H^{\gamma} - D, \qquad \Lambda = D - E \tag{2.6.15}$$

$$R = (D - z\mathbb{I})^{-1}, \qquad U = (E - z\mathbb{I})^{-1}. \tag{2.6.16}$$

So by Lemma 2.3.16, and as we have mentioned in the proof of Theorem 2.5.7, one can apply Taylor's Theorem for matrices to get that

$$G^{\gamma} - R = -R\Gamma R + (R\Gamma)^2 R - (R\Gamma)^3 G^{\gamma}. \tag{2.6.17}$$

Moreover, by a Taylor expansion for the function F', it is true that for some

$$\zeta_0 \in [\operatorname{Im}(G^{\gamma}_{i,j}), \operatorname{Im}(R_{i,j})]$$

and $\zeta = \operatorname{Im}(R_{i,j}) - \operatorname{Im}(G^{\gamma}_{i,j})$,

$$F'(\operatorname{Im}(G^{\gamma}_{i,j}) = F'(\operatorname{Im} R_{i,j}) + \zeta F^{(2)}(\operatorname{Im} R_{i,j}) + \frac{\zeta^2}{2}F^{(3)}(\operatorname{Im} R_{i,j}) + \frac{\zeta^3}{6}F^{(4)}(\zeta_0), \tag{2.6.18}$$

where we have denoted $F^{(l)}(x) = \frac{d^l}{dx^l}F(x)$ for all $l \in \mathbb{N}$.

So by combining (2.6.17) and (2.6.18), one can notice that each of the (p,q)-summand in (2.6.12) can be viewed as a sum of finite number of monomials of $A_{p,q}$ and $t^{1/2}w_{p,q}$ with coefficients depending on the matrices R and G^{γ}. These monomials can be categorized into the following cases:

1. The product of even degree of terms, i.e., $\prod_{r=1}^{s}\xi^{k_r}_{i,j}$ such that $\sum k_r$ is even and $\xi^{k_r}_{i_r,j_r}$ is equal either to $\left((R\Gamma)^{k_r}R\right)_{i_r,j_r}$, either equal to $\operatorname{Im}(\left((R\Gamma)^{k_r}R\right)_{i_r,j_r})$, either to $\operatorname{Re}(\left((R\Gamma)^{k_r}R\right)_{i_r,j_r})$ for some $s \in \mathbb{N}$ and $k_r \in \{0,1,2\}$. Then for any $m \in \{1,2,3\}$ it is true that

$$\mathbf{E}_{\Psi}F^{(m)}\left(\operatorname{Im}(R_{i,j})\right)\left(A_{p,q} - \frac{\gamma t^{1/2}w_{p,q}}{(1-\gamma^2)^{1/2}}\right)\prod_{r=1}^{s}\xi_{i_r,j_r} = 0,$$

which is a consequence to the independence of the matrix R from $A_{p,q}, w_{p,q}$ after further conditioning on the matrix X, and the symmetry of the random variables $A_{p,q}, w_{p,q}$, which has a consequence that every odd moment of them is 0.

2. The terms that contain $F^{(4)}(\zeta_0)$ can be bounded by a Taylor expansion similarly to Lemma 4.7 in [5]. More precisely one can show that

$$\left|\mathbf{E}_\Psi \operatorname{Im}(G^\gamma_{i,p} G^\gamma_{q,j}) \left(A_{p,q} - \frac{\gamma t^{1/2} w_{p,q}}{(1-\gamma^2)^{1/2}}\right) \zeta^3 F^{(4)}(\zeta_0)\right| \le N^{-2} \frac{C}{(1-\gamma^2)^{1/2}} (N^{-\omega}(\Xi+1) + Q_0 N^{11+C_0}),$$

for parameters $\omega > \epsilon_0 > 0$, such that

$$\epsilon_0 := \frac{a}{100} \min\{(4-a)v - 1, (2-a)v - a\rho, v - \rho, \frac{\rho}{2}, 1\}, \tag{2.6.19}$$

$$\omega := \min\{(a - 2\epsilon_0)\rho - 15\epsilon_0, (2-a)v - a\rho - 15\epsilon_0, (4-a)v - 1 - 10\epsilon_0, (4-2a)v - 15\epsilon_0\} \tag{2.6.20}$$

These parameters also appear in (4.25) of [5].

3. Analogously to the previous bound, one can prove that for the $s-$ products of $\xi^{k_r}_{i,j}$, when $s \in \{1,2,3,4\}$, $k_r \in \{1,2,3\}$ and $\sum k_r \ge 3$, it holds that for any $m \in \{1,2,3\}$,

$$\left|\mathbf{E}_\Psi F^{(m)}(\operatorname{Im}(R_{i,j})) \left(A_{p,q} - \frac{\gamma t^{1/2} w_{p,q}}{(1-\gamma^2)^{1/2}}\right) \prod_{r=1}^{s} \xi_{i_r, j_r}\right| \le N^{-2} \frac{C}{(1-\gamma^2)^{1/2}} (N^{-\omega}(\Xi+1) + Q_0 N^{11+C_0}). \tag{2.6.21}$$

4. The remaining terms are the monomials of 2− degree. So it can be proven that,

$$\left|\mathbf{E}_\Psi (\operatorname{Im}(R\Gamma R)_{i,p} R_{q,j}) F'(\operatorname{Im}(R_{i,j}) \left(A_{p,q} - \frac{\gamma t^{1/2} w_{p,q}}{(1-\gamma^2)^{1/2}}\right)\right| \tag{2.6.22}$$

$$\le N^{-2} \frac{C}{(1-\gamma^2)^{1/2}} \left[(N^{-\omega}(\Xi+1) + Q_0 N^{11+C_0}) + N^{a\rho + 3\epsilon_0 - 1} t \Xi(\psi_{p,q} + \mathbf{1}\{p=q\})\right], \tag{2.6.23}$$

$$\left|\mathbf{E}_\Psi \operatorname{Im}(R_{i,p} R_{q,j}) \operatorname{Im}(R\Gamma R)_{i,j} F^{(2)}(\operatorname{Im} R_{i,j}) \left(A_{p,q} - \frac{\gamma t^{1/2} w_{p,q}}{(1-\gamma^2)^{1/2}}\right)\right| \tag{2.6.24}$$

$$\le N^{-2} \frac{C}{(1-\gamma^2)^{1/2}} \left[(N^{-\omega}(\Xi+1) + Q_0 N^{11+C_0}) + N^{a\rho + 3\epsilon_0 - 1} t \Xi(\psi_{p,q} + \mathbf{1}\{p=q\})\right]. \tag{2.6.25}$$

The proof of these inequalities is a consequence of further comparison between the entries of the matrices R and U, similar to the one which was done for the matrices G^γ and R before.

So after summing over all possible (p,q) and taking into account that $t \sim N^{(a-2)v}$ and that there are at most $N^{1+a\rho+\epsilon}$ non-zero entries of Ψ with overwhelming probability, see the proof of Corollary 2.6.6, one has that (2.6.12) holds, which finishes the proof. □

In what follows, set C_a the constant mentioned in Theorem 2.3.14. So due to Theorem 2.6.4 one can prove the following.

Proposition 2.6.5. *Let a, b, v, ρ as in (2.2.1). Moreover fix $\varsigma > 0$ arbitrary small. Then for each $\delta > 0$ and $D > 0$ there exists a constant $C = C(a, \rho, v, b)$ such that*

$$\mathbb{P}\left(\sup_{\gamma\in[0,1]} \sup_{E\in[-\frac{1}{2C_a},\frac{1}{2C_a}]} \sup_{\eta\geq N^{\varsigma}-1} \max_{i,j} |G^{\gamma}_{i,j}(E + i\eta)| \geq N^{\delta} \right) \leq CN^{-D}. \tag{2.6.26}$$

The constant C_a in (2.6.26) is the constant mentioned in Theorem 2.3.14.

Proof. The proof is based on Theorem 2.6.4 and is similar to the proof of Proposition 3.17 in [5], so we will just describe the key ideas behind the proof. The proof is done in steps. Set $p = \left\lceil \frac{D+30}{\delta} \right\rceil$ and consider the function $F_{2p}(x) = |x|^{2p}+1$. Note that F_{2p} satisfies the hypothesis of Theorem 2.6.4. Moreover by Corollary 2.5.15 there exists a constant $C' = C'(a, b, v, \rho)$ such that

$$\mathbb{P}\left(\sup_{E\in[-\frac{1}{2C_a},\frac{1}{2C_a}]} \sup_{\eta\geq N^{\varsigma}-1} \max_{i,j} |G^{0}_{i,j}(E + i\eta)| \right) \leq C'N^{-D}.$$

Fix ϵ_0 and ω, the constants from the application of Theorem 2.6.4 for the function F_p. Moreover define the quantities

$$\mathcal{B}(\delta, \eta) = \mathbb{P}\left(\sup_{\gamma\in[0,1]} \max_{i,j} |G^{\gamma}(E + i\eta)| \geq N^{\delta} \right),$$

for $E \in \left[-\frac{1}{2C_a}, \frac{1}{2C_a}\right]$ and $\eta \geq N^{\varsigma-1}$. Set $s = \frac{\epsilon}{4}$. Then one can show that there exists a constant $A = A(\delta, D)$ such that,

$$\mathcal{B}(\delta, \eta) \leq AN^{A}\mathcal{B}(\frac{\epsilon_0}{2}, N^{\sigma}\eta) + AN^{-D}, \tag{2.6.27}$$

which can be proven by (i) integrating over Ψ in the conclusion of Theorem 2.6.4 for F_p after using (2.6.28), (ii) Corollary 2.5.15, (iii) Markov's Inequality applied for $\gamma \in N^{-20}\mathbb{Z} \cap (0, 1)$ and (iv) the deterministic estimates in the end of the proof of Lemma 4.3 in [5].

Thus in order to conclude, one can use induction over all $k \in \left[-1, \left\lceil \frac{1-\varsigma}{s} \right\rceil\right]$ to show that

$$\mathcal{B}\left(\frac{\epsilon_0}{2}, N^{-k\sigma}\right) \leq AN^{-D}$$

and then extend to all $E \in \left[-\frac{1}{3C}, \frac{1}{3C}\right]$ and $\eta \geq N^{\varsigma-1}$ by deterministic estimates of the form $|G^{\gamma}(z) - G^{\gamma}(z')| \leq N^{6}|z - z'|$ for an appropriately chosen grid. □

Corollary 2.6.6. *Fix $F : \mathbb{R} \to \mathbb{R}$ such that it satisfies the assumption of Theorem 2.6.4 and $E \in \left[\frac{-1}{3C}, \frac{1}{3C}\right]$ and $\eta \geq N^{\varsigma-1}$, for an arbitrary small $\varsigma > 0$. Then there exists a constant $c = c(a, b, v, \rho, C_0)$ and a large constant $C = C(a, b, v, \rho)$ such that*

$$\sup_{\gamma\in[0,1]} \left| \mathbf{E}F(\mathrm{Im}(G^{\gamma}_{i,j}(z))) - \mathbf{E}F(\mathrm{Im}(G^{0}_{i,j}(z))) \right| \leq CN^{-c}, \text{ for all } i, j \in [2N].$$

Proof. Firstly note that due to Chernoff bound there exists a constant C' such that

$$\mathbb{P}\left(|(i, j) : H_{ij} \in [N^{-\rho}, \infty)| \notin \left(\frac{N^{1+a\rho}}{C'}, C'N^{1+a\rho} \right) \right) \leq C' \exp\left(\frac{-N}{C'} \right). \tag{2.6.28}$$

Set $\Omega = \left\{(i,j) : |H_{i,j}| \in [N^{-\rho}, \infty)| \in \left(\frac{N^{1+a\rho}}{C'}, C'N^{1+a\rho}\right)\right\}$. Moreover by the deterministic estimate $|G^{\gamma}_{i,j}| \leq \eta^{-1} \leq N$ and the hypothesis for F one has that $|F(\mathrm{Im}(G^{\gamma}_{i,j})| \leq N^{C_0}$ and hence,

$$\left|\mathbf{E}F(\mathrm{Im}(G^{\gamma}_{i,j}(z))) - F(\mathrm{Im}(G^{0}_{i,j}(z)))\right| \leq \left|\mathbf{E}\mathbf{1}(\Omega) F(\mathrm{Im}(G^{\gamma}_{i,j}(z))) - F(\mathrm{Im}(G^{0}_{i,j}(z)))\right| + N^{C_0}C' \exp\left(-\frac{N}{C'}\right).$$

Note that on the set Ω we can apply Theorem 2.6.4. Moreover by Proposition 2.6.5, one has that $Q_0(z,\epsilon) \leq CN^{-D}$ for any $D > 0$ and similarly show that

$$\Xi \leq N^{C_0}Q_0(z) + CN^{C_0\epsilon}.$$

So the proof is complete after choosing an appropriately large $D > 0$. □

Next, we extend the comparison result in such way that we can use in order to approximate the gap probability.

Proposition 2.6.7. *Fix parameters a, b, ρ, v as in (2.2.1). Let $q : \mathbb{R} \to \mathbb{R}$ a C^{∞} function with all its derivatives bounded by an absolute constant M greater than 1. Then for any $\eta \geq N^{-2}$ and any positive sequence $r(N)$ such that $\lim r(N) = r > 0$ there exist constants $\omega = \omega(a,\rho,v,b,r)$, $\epsilon = \epsilon(a,\rho,v,b,r)$ and $C = C(a,\rho,v,b,r)$ such that*

$$\begin{aligned}\sup_{\gamma\in[0,1]}&\left|\mathbf{E}q\left(\int_{-\frac{r(N)}{N}}^{\frac{r(N)}{N}}\sum_{i=1}^{2N}\mathrm{Im}\, G^{\gamma}_{i,i}(y+i\eta)dy\right) - \mathbf{E}q\left(\int_{-\frac{r(N)}{N}}^{\frac{r(N)}{N}}\sum_{i=1}^{2N}\mathrm{Im}\, G^{0}_{i,i}(y+i\eta)dy\right)\right| \leq \\ &C\left(MN^{-\omega} + MN^{C}Q(\epsilon,\eta)\right),\end{aligned} \tag{2.6.29}$$

where

$$Q(\epsilon,\eta) = \mathbb{P}\left(\sup_{E\in[-\frac{1}{3C},\frac{1}{3C}]}\max_{i,j}|G^{\gamma}_{i,j}(E+i\eta)| \geq N^{\epsilon}\right).$$

Moreover if we suppose that $\eta \geq N^{\varsigma-1}$, for arbitrary small $\varsigma > 0$, then there exists a constant $c = c(a,\rho,v,b)$ such that,

$$\sup_{\gamma\in[0,1]}\left|\mathbf{E}q\left(\int_{-\frac{r(N)}{N}}^{\frac{r(N)}{N}}\sum_{i=1}^{2N}\mathrm{Im}\, G^{\gamma}_{i,i}(y+i\eta)dy\right) - \mathbf{E}q\left(\int_{-\frac{r(N)}{N}}^{\frac{r(N)}{N}}\sum_{i=1}^{2N}\mathrm{Im}\, G^{0}_{i,i}(y+i\eta)dy\right)\right| \leq CN^{-c}. \tag{2.6.30}$$

Proof. For simplicity we will assume r is a constant. The proof of (2.6.29) is similar to the proof of Theorem 2.6.4. Next we highlight the differences.

Note that similarly to the proof of Corollary 2.6.6, it is sufficient to prove that for any matrix Ψ with at most $N^{1+a\rho}$ non-zero entries it is true that,

$$\begin{aligned}\sup_{\gamma\in[0,1]}&\left|\mathbf{E}_{\Psi}q\left(\int_{-\frac{r}{N}}^{\frac{r}{N}}\sum_{i=1}^{2N}\mathrm{Im}\, G^{\gamma}_{i,i}(y+i\eta)dy\right) - \mathbf{E}_{\Psi}q\left(\int_{-\frac{r}{N}}^{\frac{r}{N}}\sum_{i=1}^{2N}\mathrm{Im}\, G^{0}_{i,i}(y+i\eta)dy\right)\right| \\ &\leq C\left(MN^{-\omega} + MN^{C}Q(\epsilon)\right).\end{aligned} \tag{2.6.31}$$

Furthermore one can compute the derivative of the previous quantity with respect to γ, as in Theorem 2.6.4. Thus by Leibniz integral rule, Fubini Theorem and (2.6.11) it is true that

$$\left|\frac{d}{d\gamma}\mathbf{E}_{\Psi}q\left(\int_{-\frac{r}{N}}^{\frac{r}{N}}\sum_{i=1}^{2N}\mathrm{Im}\, G^{\gamma}_{i,i}(y+i\eta)dy\right)\right| \leq$$

$$\leq \int_{-\frac{r}{N}}^{\frac{r}{N}} \sum_{i=1}^{2N} \sum_{p,q\in[2N]:|p-q|\geq N} \left| \mathbf{E}_{\psi} q' \left(\int_{-\frac{r}{N}}^{\frac{r}{N}} \sum_{i=1}^{2N} \operatorname{Im} G_{i,i}^{\gamma}(y+i\eta)dy \right) \operatorname{Im}(G_{i,p}^{\gamma} G_{q,i}^{\gamma}) \left(A_{p,q} - \frac{\gamma t^{1/2} w_{p,q}}{(1-\gamma^2)^{1/2}} \right) dy \right|.$$

As a result it is sufficient to prove that for any $y \in \left(-\frac{r}{N}, \frac{r}{N}\right)$,

$$\sum_{p,q\in[2N]:|p-q|\geq N} \left| \mathbf{E}_{\Psi} q' \left(\int_{-\frac{r}{N}}^{\frac{r}{N}} \sum_{i=1}^{2N} \operatorname{Im} G_{i,i}^{\gamma}(y+i\eta)dy \right) \operatorname{Im}(G_{i,p}^{\gamma} G_{q,i}^{\gamma}) \left(A_{p,q} - \frac{\gamma t^{1/2} w_{p,q}}{(1-\gamma^2)^{1/2}} \right) \right| \tag{2.6.32}$$

$$\leq \frac{C}{(1-\gamma^2)^{1/2}} M(N^{-\omega} + Q(\epsilon,\eta)N^C). \tag{2.6.33}$$

But the proof of (2.6.32) is similar to the proof of (2.6.12), with the main difference located in the Taylor expansion which now instead of being applied as in (2.6.18), it will be applied for the quantities $\int_{-\frac{r}{N}}^{\frac{r}{N}} \sum_{i=1}^{2N} \operatorname{Im} G_{i,i}^{\gamma}(y+i\eta)dy$ and $\int_{-\frac{r}{N}}^{\frac{r}{N}} \sum_{i=1}^{2N} \operatorname{Im} R_{i,i}(y+i\eta)dy$ for fixed p,q. But eventually, this does not affect the proof since each (p,q)-summand can again be expressed into monomials of $A_{p,q}, w_{p,q}$, which do not depend on the parameters η and y, and since the quantity $Q_0(\epsilon,\eta)$ is replaced by $Q(\epsilon,\eta)$ in the bound.

Moreover, if we assume $\eta \geq N^{\varsigma-1}$, then $Q(\epsilon,\eta)$ is smaller than N^{-D} for any D and for sufficient large N. Thus similarly to the proof of (2.6.6), one can prove (2.6.30) by (2.6.29). □

Moreover, we wish to prove that the righthand side of (2.6.29) tends to 0 as N tends to infinity for $\eta = O(\frac{1}{N^{1+\varsigma}})$, below the natural scale. This is achieved via the following lemma.

Lemma 2.6.8 ([50],Lemma 2.1). *Let Y be an $N\times N$ matrix. Set the following quantity*

$$\Gamma(Y, E+i\eta) = \max\{1, \max_{i,j} |(Y-(E+i\eta)\mathbb{I})_{i,j}^{-1}|)\}.$$

Then for any $M \geq 1$ and $\eta > 0$ the following deterministic inequality holds

$$\Gamma\left(Y, E + i\frac{\eta}{M}\right) \leq M\Gamma(Y, E+i\eta). \tag{2.6.34}$$

Corollary 2.6.9. *Fix ς and δ arbitrary small positive numbers. Set $\eta_1 = N^{-\varsigma/2-1}$. Then by (2.6.34) and (2.6.26) one has that for any $D > 0$ and for sufficient large N, it is true that*

$$\mathbb{P}\left(\sup_{\gamma\in[0,1]} \sup_{E\in[-\frac{1}{3C},\frac{1}{3C}]} \max_{i,j} |G_{i,j}^{\gamma}(E+i\eta_1)| \geq N^{\delta+\varsigma} \right) \leq CN^{-D}. \tag{2.6.35}$$

So in the setting of Proposition 2.6.7, it is implied that there exist two positive constants $C = C(a,b,v,\rho,r)$, $c = c(a,b,\rho,v)$ such that

$$\sup_{\gamma\in[0,1]} \left| \mathbf{E}q\left(\int_{-\frac{r(N)}{N}}^{\frac{r(N)}{N}} \sum_{i=1}^{2N} \operatorname{Im} G_{i,i}^{\gamma}(y+i\eta_1)dy \right) - \mathbf{E}q\left(\int_{-\frac{r(N)}{N}}^{\frac{r(N)}{N}} \sum_{i=1}^{2N} \operatorname{Im} G_{i,i}^{0}(y+i\eta_1)dy \right) \right| \leq CN^{-c}. \tag{2.6.36}$$

2.6.2 Approximation of the gap probability

The goal of this subsection is to approximate the gap probability, i.e., the probability that there are no eigenvalues in an interval, by C^∞ functions of the Stieltjes transform as in Proposition 2.6.7. In order to prove the latter, we use similar tools as in Section 5 of [5]. First, define the following quantities for any $r > 0, \gamma > 0$ and $\eta \geq 0$.

$$\begin{aligned} X_x(t) &= \mathbf{1}\{t \in (-x, x)\}, \text{ for all } x \in \mathbb{R} \\ \vartheta_\eta(x) &= \frac{1}{\pi}\operatorname{Im}\left(\frac{1}{x - i\eta}\right) = \frac{1}{\pi}\frac{\eta}{\eta^2 + x^2}, \text{ for all } x \in \mathbb{R} \\ \operatorname{Tr} X_r * \vartheta_\eta(H^\gamma) &= \frac{1}{\pi}\int_{-\frac{r}{2N}}^{\frac{r}{2N}} \sum_{i=1}^{2N} \operatorname{Im} G^\gamma_{i,i}(x + i\eta)dx = \frac{1}{\pi}\int_{-\frac{r}{2N}}^{\frac{r}{2N}} \sum_{i=1}^{2N} \frac{\eta}{(\lambda_i(H^\gamma) - x)^2 + \eta^2}dx. \end{aligned} \tag{2.6.37}$$

Moreover for any $N \times N$ matrix Y with eigenvalues denoted by $\lambda_i(Y)$ and for any $E_1, E_2, E \in \mathbb{R}$ such that $E_1 \leq E_2$ and $E > 0$, we denote

$$\begin{aligned} \imath_N(Y, E_1, E_2) &= \#\{i \in [N] : \lambda_i(Y) \in (E_1, E_2)\}, \\ \imath_N(Y, E) &= \#\{i \in [N] : \lambda_i(Y) \in (-E, E)\}. \end{aligned} \tag{2.6.38}$$

Moreover set $\{\lambda_i^\gamma\}_{i \in [2N]}$ the eigenvalues of H^γ arranged in increasing order.

Lemma 2.6.10. *For any $\gamma \in [0, 1]$ and $I \subseteq \left(-\frac{1}{3C}, \frac{1}{3C}\right)$ such that $|I| = N^{\varsigma/2-1}$, it is true that*

$$\left|\{i \in [2N] : \lambda_i^\gamma \in I\}\right| \leq 2|I|N^{1+\varsigma/2}$$

with overwhelming probability.

Proof. For the convenience of notation, suppose that

$$I = (E - \eta, E + \eta)$$

.

Moreover set the event

$$\Omega_\eta = \left\{ \sup_{\gamma \in [0,1]} \sup_{E \in [-\frac{1}{3C}, \frac{1}{3C}]} \max_{i,j} |G^\gamma_{i,j}(E + i\eta)| \geq N^{\varsigma/2} \right\}. \tag{2.6.39}$$

By (2.6.5), Ω_η^c holds with overwhelming probability. Then

$$\mathbf{1}\left(\Omega_\eta^c\right) N^{\varsigma/2} \geq \frac{\mathbf{1}\left(\Omega_\eta^c\right)}{2N} \sum_{i=1}^{2N} \operatorname{Im}(G^\gamma_{i,i}(E + i\eta)) \geq \frac{\mathbf{1}\left(\Omega_\eta^c\right)}{2N} \sum_{\lambda_i \in I} \operatorname{Im}(G^\gamma_{i,i}(E + i\eta)) \tag{2.6.40}$$

$$= \frac{\mathbf{1}\left(\Omega_\eta^c\right)}{2N} \sum_{\lambda_i \in I} \frac{\eta}{(\lambda_i - E)^2 + \eta^2} \geq \frac{\mathbf{1}\left(\Omega_\eta^c\right)}{2N|I|} |\{i \in [2N] : \lambda_i^\gamma \in I\}| \tag{2.6.41}$$

□

Next fix $\epsilon > 0$ arbitrary small and $r \in \mathbb{R}$. Set

$$\eta_1 = N^{-1-99\epsilon}, \qquad l = N^{-1-3\epsilon}, \qquad l_1 = lN^{2\epsilon}.$$

Lemma 2.6.11. *For any $\gamma \in [0,1]$, it is true that there exists an absolute constant C such that with overwhelming probability*

$$\left| i_{2N}\left(H^{\gamma}, \frac{r}{N}\right) - \operatorname{Tr}\, X_r * \vartheta_{\eta_1}(H^{\gamma}) \right| \leq C\left(N^{-2\epsilon} + i_{2N}\left(H^{\gamma}, -\frac{r}{N} - l, -\frac{r}{N} + l\right) + i_{2N}\left(H^{\gamma}, \frac{r}{N} - l, \frac{r}{N} + l\right)\right).$$

Proof. Firstly note that by elementary computation as in (6.10) of [51] one has that,

$$|X_{\frac{r}{N}}(x) - X_r * \vartheta_{\eta_1}(x)| \leq C\eta_1 \left(\frac{2r}{Nd_1(x)d_2(x)} + \frac{X_{r/N}(x)}{d_1(x) + d_2(x)} \right), \tag{2.6.42}$$

where C is some absolute constant, $d_1(x) = \left|\frac{r}{N} + x\right| + \eta_1$ and $d_2 = \left|\frac{r}{N} - x\right| + \eta_1$. Moreover note that the right hand side of (2.6.42) is always bounded by an absolute constant and is $O(\eta_1/l)$ if $\min d_i \geq l$. Thus by Lemma 2.6.10 one has that with overwhelming probability

$$\left| i_{2N}\left(H^{\gamma}, \frac{r}{N}\right) - \operatorname{Tr}\, X_r * \vartheta_{\eta_1}(H^{\gamma}) \right| \leq C\left(\operatorname{Tr}\,(f_1(H^{\gamma}) + \operatorname{Tr}(f_2(H^{\gamma})) + \frac{\eta}{l} i_{2N}\left(H^{\gamma}, \frac{-r}{N} + l, \frac{r}{N} - l\right)\right) \tag{2.6.43}$$

$$+ C\left(i_{2N}\left(H^{\gamma}, -\frac{r}{N} - l, -\frac{r}{N} + l\right) + i_{2N}\left(H^{\gamma}, \frac{r}{N} - l, \frac{r}{N} + l\right)\right), \tag{2.6.44}$$

where

$$f_1(x) = \mathbf{1}\{x \leq -E - l\} \frac{2r\eta_1}{Nd_1(x)d_2(x)}, \qquad f_2(x) = \mathbf{1}\{x \geq E + l\} \frac{2r\eta_1}{Nd_1(x)d_2(x)}.$$

So in order to complete the proof we need to show that the first term on the right side of the inequality is of order $N^{-2\epsilon}$. Note that due to Lemma 2.6.10 and the fact that the length of the interval $\left(-\frac{r}{N} - l, \frac{r}{N} + l\right)$ is smaller than $N^{\varsigma-1}$ for any $\varsigma > 0$ one has that $\frac{\eta_1}{l} i_{2N}\left(H^{\gamma}, -\frac{r}{N} - l, \frac{r}{N} + l\right) \leq N^{-2\epsilon}$ with overwhelming probability.

Moreover after splitting the interval $\left(-\frac{1}{3C}, -\frac{r}{N} - l\right)$ into intervals with length $O(N^{-1})$, like in [3] (5.61) and since

$$f_1(x) \leq \frac{\eta_1}{|E - x|} \mathbf{1}\left\{x \in \left(\frac{-1}{3C}, -\frac{r}{N} - l\right)\right\} + \frac{\eta_1}{|E - x|} \mathbf{1}\left\{x \in \left(-\infty, \frac{-1}{3C}\right)\right\},$$

one can show that $\operatorname{Tr} f_1(H^{\gamma}) \leq N^{-2\epsilon}$. Similar bound can be proven for $\operatorname{Tr} f_2(H^{\gamma})$. □

Lemma 2.6.12. *For any $\gamma \in [0,1]$ there exists an absolute constant C such that*

$$\operatorname{Tr}\, X_r * \vartheta_{\eta_1 - l_1}(H^{\gamma}) - CN^{-\epsilon} \leq i_{2N}\left(H^{\gamma}, \frac{r}{N}\right) \leq \operatorname{Tr}\, X_r * \vartheta_{\eta_1 + l_1}(H^{\gamma}) + CN^{-\epsilon}. \tag{2.6.45}$$

Proof. We will prove the second inequality of (2.6.45). The proof of the first inequality is similar.

First note that by definition, one has that for any $y_1 \geq y_2 > 0$,

$$\mathfrak{i}_{2N}(H^\gamma, y_2) \leq \mathfrak{i}_{2N}(H^\gamma, y_1) \tag{2.6.46}$$

$$\text{Tr } X_{y_2} * \vartheta_{\eta_1}(H^\gamma) \leq \text{Tr } X_{y_1} * \vartheta_{\eta_1}(H^\gamma)) \tag{2.6.47}$$

So we get that with overwhelming probability

$$\mathfrak{i}_{2N}\left(H^\gamma, \frac{r}{N}\right) \leq \frac{1}{l_1}\int_{\frac{r}{N}}^{\frac{r}{N}+l_1} \mathfrak{i}_{2N}\left(H^\gamma, \frac{r}{N}+y\right) d_y \tag{2.6.48}$$

$$\leq \frac{1}{l_1}\left(\int_{\frac{r}{N}}^{\frac{r}{N}+l_1} \text{Tr } X_{Ny} * \vartheta_{\eta_1}(H^\gamma) + C\left(N^{-2\epsilon} + \mathfrak{i}_{2N}(H^\gamma, y-l, -y+l) + \mathfrak{i}_{2N}(H^\gamma, -y-l, -y+l)\right) dy\right) \tag{2.6.49}$$

$$\leq \text{Tr } X_r * \vartheta_{\eta_1+l_1}(H^\gamma) + CN^{-\epsilon} \tag{2.6.50}$$

In the first inequality of (2.6.48) we used (2.6.46), in the second we used Lemma 2.6.11 and in the third we used (2.6.47) for the first term in the sum and Lemma 2.6.10 for the second. □

Next we proceed as in Lemma 5.13 of [3]. Set $\tilde{q}(x) : \mathbb{R} \to \mathbb{R}_+$ be a C^∞, even function, with all its derivatives bounded by a constant M, such that

- $\tilde{q}(x) = 0$ for $x \in \left(-\infty, \frac{-2}{9}\right) \cup \left(\frac{2}{9}, \infty\right)$
- $\tilde{q}(x) = 1$ for $x \in \left(\frac{-1}{9}, \frac{1}{9}\right)$
- $\tilde{q}(x)$ is decreasing on $\left(\frac{1}{9}, \frac{2}{9}\right)$.

In the following Lemma we prove the approximation of the gap probability of H^γ by function of the form appearing in (2.6.36).

Lemma 2.6.13. *For any $\gamma \in [0, 1]$ and $D > 0$ it is true that*

$$\mathbf{E}\tilde{q}\left(\text{Tr } X_r * \vartheta_{\eta_1+l_1}(H^\gamma)\right) - N^{-D} \leq \mathbb{P}\left(\mathfrak{i}_{2N}\left(H^\gamma, \frac{r}{N}\right) = 0\right) \leq \mathbf{E}\tilde{q}\left(\text{Tr } X_r * \vartheta_{\eta_1-l_1}(H^\gamma)\right) + N^{-D}. \tag{2.6.51}$$

Proof. By Lemma 2.6.12 and for large enough N, it is true that if $\mathfrak{i}_{2N}\left(H^\gamma, \frac{r}{N}\right) = 0$ then $\text{Tr } X_r * \vartheta_{\eta_1-l_1}(H^\gamma) \leq \frac{1}{9}$ with overwhelming probability. This implies that for any large $D > 0$ and for N sufficiently large one has that,

$$\mathbb{P}\left(\mathfrak{i}_{2N}\left(H^\gamma, \frac{r}{N}\right) = 0\right) \leq \mathbb{P}\left(\text{Tr } X_r * \vartheta_{\eta_1-l_1}(H^\gamma) \leq \frac{1}{9}\right) + N^{-D} \leq \mathbb{P}\left(\text{Tr } X_r * \vartheta_{\eta_1-l_1}(H^\gamma) \leq \frac{2}{9}\right) + N^{-D} \tag{2.6.52}$$

$$= \mathbb{P}\left[\tilde{q}\left(\text{Tr } X_r * \vartheta_{\eta_1-l_1}(H^\gamma)\right) \geq 1\right] + N^{-D} \leq \mathbf{E}\tilde{q}\left(\text{Tr } X_r * \vartheta_{\eta_1-l_1}(H^\gamma)\right) + N^{-D} \tag{2.6.53}$$

In (2.6.53), we used the Markov inequality for the random variable $\tilde{q}\left(\text{Tr } X_r * \vartheta_{\eta_1-l_1}(H^\gamma)\right)$. So we have proven the second inequality of (2.6.51).

For the first, note that again by Lemma 2.6.12, with overwhelming probability it is true that if $\tilde{q}\Big(\text{Tr } X_r * \partial_{\eta_1+l_1}(H^\gamma)\Big) \le \frac{2}{9}$ then $i_{2N}(H^\gamma, \frac{r}{N}) \le CN^{-\epsilon} + \frac{2}{9}$. Thus,

$$\mathbf{E}\tilde{q}\Big(\text{Tr } X_r * \partial_{\eta_1+l_1}(H^\gamma)\Big) \le \mathbb{P}\left[\tilde{q}\Big(\text{Tr } X_r * \partial_{\eta_1+l_1}(H^\gamma)\Big) \le \frac{2}{9}\right] \tag{2.6.54}$$

$$\le \mathbb{P}\left[i_{2N}\left(H^\gamma, \frac{r}{N}\right) \le CN^{-\epsilon} + \frac{2}{9}\right] + N^{-D} = \mathbb{P}\left[i_{2N}\left(H^\gamma, \frac{r}{N}\right) = 0\right] + N^{-D}. \tag{2.6.55}$$

□

2.6.3 Proof of Theorem 2.1.2

At this subsection we prove Theorem 2.1.2. Fix $r \in (0, \infty)$ and $\epsilon > 0$ small enough. Set $\eta_1 = N^{-1-\epsilon}$ and $l = N^{-1-99\epsilon}$. Furthermore $r \in (0, \infty)$. Let $\tilde{q}(x)$ denote the function defined before Lemma 2.6.13.

- For the first part of Theorem 2.1.2 note that due to (2.6.36) and Lemma 2.6.13 one has that there exist constants $C = C(r) > 0$ and $c > 0$, such that for large enough $D > 0$ it is true that

$$\mathbf{E}\tilde{q}\Big(\text{Tr } X_r * \partial_{\eta_1+l}(H^0)\Big) - N^{-D} - CN^{-c} \le \mathbf{E}\tilde{q}\Big(\text{Tr } X_r * \partial_{\eta_1+l}(H^1)\Big) - N^{-D} \tag{2.6.56}$$

$$\le \mathbb{P}\left(i_{2N}\left(H^1, \frac{r}{N}\right) = 0\right) \le \mathbf{E}\tilde{q}\Big(\text{Tr } X_r * \partial_{\eta_1-l}(H^1)\Big) + N^{-D} \tag{2.6.57}$$

$$\le \mathbf{E}\tilde{q}\Big(\text{Tr } X_r * \partial_{\eta_1-l}(H^0)\Big) + CN^{-c} + N^{-D}. \tag{2.6.58}$$

  Next note that by the definition of the symmetrization of a matrix, the gap probability is actually the tail distribution of the smallest singular value, i.e.,

$$\mathbb{P}\left(i_{2N}\left(H^1, \frac{r}{N}\right) = 0\right) = \mathbb{P}\left(s_1(D_N) \ge \frac{r}{N}\right).$$

  Moreover note that the limiting distribution of the least singular value of a Gaussian matrix is $1 - \exp(-r^2/2 - r)$ as mentioned in Theorem 1.3. of [2]. Let L_N be a matrix with i.i.d. entries all following the Gaussian law with mean 0 and variance N^{-1}. Set $s_1(L_N)$ the least singular value of L_N. Let W_N be the symmetrization of L_N. As before one can notice that

$$\mathbb{P}\left(i_{2N}\left(E_N, \frac{r}{N}\right) = 0\right) = \mathbb{P}\left(s_1(L_N) \ge \frac{r}{N}\right).$$

  So after another application of Lemma 2.6.13 for the matrix H^0 and Corollary 2.4.5 for $r' = r\xi^{-1}$, where ξ is defined in (2.1.4), one has that there exists a small constant $\tilde{c} > 0$ and a large constant such that

$$\mathbb{P}\left(Ns_1(L_N) \ge r - N^{-\epsilon}\right) - CN^{-\tilde{c}} \le \mathbb{P}\left(\xi N s_1(D_N) \ge r\right) \le \mathbb{P}\left(Ns_1(L_N) \ge r + N^{-\epsilon}\right) + CN^{-\tilde{c}},$$

  which implies universality of the least singular value for D_N multiplied by $N\xi$.

- For the proof of the second part, it is well-known that bounding the entries of the resolvent implies the complete eigenvector delocalization. So by (2.6.26), one can prove the complete eigenvector delocalization as in Theorem 6.3 in [49].

Chapter 3

The limit of the operator norm for random matrices with general variance profile

3.1 Statement of the results

Notation. For any $N \times N$ matrix $A = (a_{i,j})_{i,j \in [N]} \in \mathbb{R}^{N \times N}$ with eigenvalues $\{\lambda_i(A)\}_{i \in [N]}$, the measure

$$\mu_A := \frac{1}{N} \sum_{i \in [N]} \delta_{\lambda_i(A)}$$

will be called the Empirical Spectral Distribution (E.S.D.) of A. When the eigenvalues are real, write $\lambda_{\max}(A)$ for the maximum among them. We will use the following two norms on square matrices. For $A \in \mathbb{R}^{N \times N}$,

$$|A|_{\text{op}} := \max_{x \in \mathbb{R}^N : \|x\|_2 = 1} \|Ax\|_2 = \sqrt{\lambda_{\max}(AA^T)} \tag{3.1.1}$$

$$\|A\|_{\max} := \max_{i,j \in [N]} |a_{i,j}|. \tag{3.1.2}$$

It is easy to see that $|A|_{\text{op}} \le N\|A\|_{\max}$ and if the matrix A is symmetric, then

$$|A|_{\text{op}} := \max_{i \in [N]} |\lambda_i(A)|. \tag{3.1.3}$$

Throughout this section, $(A_N)_{N \in \mathbb{N}^+}$ is a sequence of symmetric random matrices with independent entries (up to symmetry), $A_N = (a_{i,j}^{(N)})_{i,j \in [N]}$ is an $N \times N$ matrix, and all $\{a_{i,j}^{(N)} : N \in \mathbb{N}^+, i, j \in [N]\}$ are defined on the same probability space and take real values.

A standard assumption for the sequence is the following (see relation (2.2.1) in [52]).

Assumption 3.1.1.

1. $\mathbf{E}a_{i,j}^{(N)} = 0$ for all $N \in \mathbb{N}^+, i, j \in [N]$, and $\sup_{N \in \mathbb{N}^+, i,j \in [N]} \mathbf{E}|a_{i,j}^{(N)}|^2 < \infty$.

2. For any $\varepsilon > 0$,

$$\lim_{N\to\infty} \frac{1}{N^2} \sum_{i,j\in[N]} \mathbf{E}\Big\{|a_{i,j}^{(N)}|^2 \mathbf{1}_{|a_{i,j}^{(N)}|\geq \varepsilon\sqrt{N}}\Big\} = 0. \tag{3.1.4}$$

This is satisfied in the case that $\{a_{i,j}^{(N)} : N \in \mathbb{N}^+, i, j \in [N], i \leq j\}$ are i.i.d. with mean 0 and finite variance. But it is not enough to guarantee that the ESD of the appropriately normalized A_N converges to a nontrivial limit. To state a sufficient condition for this, we introduce some notation that will be used throughout the work. We let

$$s_{i,j}^{(N)} := \mathbf{E}\{|a_{i,j}^{(N)}|^2\} \tag{3.1.5}$$

for all $N \in \mathbb{N}^+, i, j \in [N]$ and $V_0 := \sup_{N\in\mathbb{N}^+,i,j\in[N]} s_{i,j}^{(N)} \in [0, \infty)$.

Also, let $\mathbf{C}_k$ be the set of ordered rooted trees with k edges (where $k \in \mathbb{N}$) of all non-isomorphic plane rooted trees with $k + 1$ vertices, i.e. all trees with $k + 1$ vertices, a vertex distinguished as a root and an ordering amongst the children of any vertex. The number of such trees is the $k-$th Catalan number, i.e.,

$$|\mathbf{C}_k| = \frac{1}{k+1}\binom{2k}{k}, \tag{3.1.6}$$

and a trivial bound that we will use is $|\mathbf{C}_k| \leq 2^{2k}$. For each such tree, we consider its vertices ordered $v_0 < v_1 < \cdots < v_k$ so that v_0 is the root, each parent is smaller than its children, and the children keep the order they have as vertices of an ordered tree. A labeling of such a tree is an ordered $k+1$-tuple $(\ell_0, \ell_1, \cdots, \ell_k)$ of different objects, the object ℓ_i is the label of vertex v_i.

A quantity of fundamental importance for the sequel is the following sum

$$M_N(k) := \sum_{T\in\mathbf{C}_k} \sum_{\substack{\mathbf{i}\in[N]^{k+1} \\ \text{labeling of } T}} \prod_{\{i,j\}\in E(T)} s_{i,j}^{(N)}. \tag{3.1.7}$$

$E(T)$ denotes the set of edges of the tree T. Note that $M_N(0) = N$ since by convention the product over an empty index set equals 1.

Assumption 3.1.2. There is a probability measure μ on $\mathbb{R}$ such that for each $k \in \mathbb{N}$ it holds

$$\lim_{N\to\infty} \frac{M_N(k)}{N^{k+1}} = \int x^{2k}\, d\mu(x). \tag{3.1.8}$$

A tool for checking this assumption is explained in Remark 3.1.13 below.

If the sequence $(A_N)_{N\in\mathbb{N}^+}$ satisfies both Assumptions 3.1.1 and 3.1.2, then $\mu_{A_N/\sqrt{N}} \Rightarrow \mu$ with probability one (see the proof of Theorem 3.2 of [13]). The measure μ is symmetric with compact support contained in $[-2\sqrt{V_0}, 2\sqrt{V_0}]$. The compactness of the support follows from (3.1.8), $M_N(k) \leq |\mathbf{C}_k| N^{k+1} V_0^k$, and $|\mathbf{C}_k| \leq 2^{2k}$. Let

$$\mu_\infty := \sup \operatorname{suppt} \mu. \tag{3.1.9}$$

We seek conditions under which the maximum eigenvalue of $A_N/\sqrt{N}$ converges to μ_∞ in probability. An easy argument will give us the lower bound, and since $\lambda_{\max}(A) \le \|A\|_{\text{op}}$ for any symmetric matrix $A \in \mathbb{R}^{N\times N}$, it will be enough to prove the upper bound for the operator norm of $A_N/\sqrt{N}$.

For this purpose, we need stronger assumptions. The following is stronger than Assumption 3.1.1.

Assumption 3.1.3.

(a) $\mathbf{E}a_{i,j}^{(N)} = 0$ for all $N \in \mathbb{N}^+$, $i,j \in [N]$, $\sup_{N\in\mathbb{N}^+, i,j\in[N]} \mathbf{E}|a_{i,j}^{(N)}|^2 \le 1$, and $\sup_{N\in\mathbb{N}^+, i,j\in[N]} \mathbf{E}|a_{i,j}^{(N)}|^4 < \infty$.

(b) For any $\varepsilon > 0$ it is true that

$$\lim_{N\to\infty} \sum_{i,j} \mathbf{P}(|a_{i,j}^{(N)}| \ge \varepsilon\sqrt{N}) = 0. \tag{3.1.10}$$

Note that condition (3.1.10) is satisfied if we assume that all $\{a_{i,j}^{(N)} : N \in \mathbb{N}^+, i,j \in [N]\}$ have the same distribution with finite 4-th moment.

We gain control over $\|A_N\|_{\text{op}}$ through the traces of high moments of A_N, and the main difficulty, which the next conditions (Assumption 3.1.4 and Assumption 3.1.6) try to address, is how to connect these traces with μ_∞, which emerges out of $\{\lambda_i(A_N) : i \in [N]\}$ only after we take $N \to \infty$.

Assumption 3.1.4. For every $N \in \mathbb{N}^+$ and $i,j \in [N]$ it is true that

$$s_{i,j}^{(N)} \le \min\{s_{2i,2j}^{(2N)}, s_{2i-1,2j}^{(2N)}, s_{2i-1,2j-1}^{(2N)}\}. \tag{3.1.11}$$

For example, this assumption is satisfied if $s_{i,j}^{(N)} = h(i/N, j/N)$ for all $N \in \mathbb{N}^+$, $i,j \in [N]$, where $h : [0,1]^2 \to [0,\infty)$ is a function decreasing separately in each variable.

In order to give the next sufficient condition, we first give some definitions.

Definition 3.1.5. (i) We call graphon any Borel measurable function $W : [0,1]\times[0,1] \to \mathbb{R}$ which is symmetric and integrable.
(ii) For any bounded graphon W and any multigraph $G = (V, E)$, we call isomorphism density from G to W the quantity

$$t(G, W) := \int_{[0,1]^{|V|}} \prod_{\{i,j\}\in E} W(x_i, x_j) \prod_{i\in V} dx_i. \tag{3.1.12}$$

Now, let $(A_N)_{N\in\mathbb{N}^+}$ be a sequence of random matrices with elements having finite second moment. Each A_N defines a graphon, W_N, through the relation

$$W_N(x, y) := s_{\lceil Nx\rceil, \lceil Ny\rceil}^{(N)} \tag{3.1.13}$$

for each $(x, y) \in [0,1]\times[0,1]$. For this relation, $\lceil 0\rceil$ denotes 1.

Assumption 3.1.6. There exists a graphon W such that the W_N of (3.1.13) satisfies

$$\lim_{N\to\infty} t(T, W_N) = t(T, W) \tag{3.1.14}$$

for any finite tree T. Moreover, for any $D > 0$ there exists some $C = C(D) \in (0, \infty)$ and $N_0 = N_0(D) \in \mathbb{N}^+$ such that for any $N \geq N_0$ it holds

$$\int_{[0,1]^2} |W_N(x, y) - W(x, y)| \, dx \, dy \leq CN^{-D}. \tag{3.1.15}$$

This assumption together with Assumption 3.1.3 implies Assumption 3.1.2 (This will be explained in Lemma 3.3.2). Again, we denote by μ_∞ the maximum of the support of μ.

The assumptions we made so far will lead to convergence in probability of the largest eigenvalue. Next we give some extra condition, which will lead to the almost sure convergence of the largest eigenvalue.

Assumption 3.1.7. $(A_N)_{N\in\mathbb{N}^+}$ is a sequence of symmetric random matrices, the entries of each A_N are independent (up to symmetry), and there exists a random variable X with mean 0, variance 1, and finite $4 + \delta$ moment for some $\delta > 0$, which stochastically dominates the entries of A_N in the following sense

$$\mathbf{P}(|\{A_N\}_{i,j}| \geq t) \leq \mathbf{P}(|X| \geq t), \text{ for all } t \in [0, \infty), N \in \mathbb{N}^+, i, j \in [N]. \tag{3.1.16}$$

We are now ready to present our first main result.

Theorem 3.1.8. *Let $(A_N)_{N\in\mathbb{N}^+}$ be a sequence of matrices satisfying Assumption 3.1.3. Then if either Assumptions 3.1.2 and 3.1.4 hold or Assumption 3.1.6 holds, it is true that*

$$\lim_{N\to\infty} \frac{\|A_N\|_{\mathrm{op}}}{\sqrt{N}} = \mu_\infty \text{ in probability} \tag{3.1.17}$$

where μ_∞ is defined in (3.1.9). Moreover, if the sequence $(A_N)_{N\in\mathbb{N}^+}$ satisfies Assumption 3.1.7, the convergence in (3.1.17) holds in the almost sure sense.

Note that Assumption 3.1.4 is restrictive and does not cover several of the well-known and studied models. Thus, in what follows, we try to extend the domain of validity of Theorem 3.1.8. We first give two definitions.

For $N \in \mathbb{N}^+$ and $U \subset [N]^2$:

• We call a $(x, y) \in U$ *internal point* of U if $\{(x + d_1, y + d_2) : d_1, d_2 \in \{-1, 0, 1\}\} \subset U$. We denote by U^o the set of internal points of U.

• We say that U is *axially convex* if $(i, j) \in U, (i, j') \in U, r \in [N], (r - j)(r - j') < 0$ imply $(i, r) \in U$ and $(i, j) \in U, (i', j) \in U, r \in [N], (r - i)(r - i') < 0$ imply $(r, j) \in U$.

Definition 3.1.9 (Generalized step function variance profile)**.** Let $(A_N)_{N\in\mathbb{N}^+}$ be a sequence of symmetric random matrices, A_N of dimension $N\times N$, with each element having zero mean and finite second moment. Moreover, suppose that there exists an $\mathbb{N}^+$−valued sequence $(d_N)_{N\in\mathbb{N}^+}$ with $\lim_{N\to\infty} d_N/N = 0$ and such that for each N there is a partition $\mathcal{P}_N := \{\mathcal{B}_i^{(N)} : i = 1, 2, \ldots, d_N\}$ of the grid $[N]^2$ consisting of d_N axially convex sets with the following properties.

(a) If $A \in \mathcal{P}_N$ then $R(A) := \{(i, j) : (j, i) \in A\} \in \mathcal{P}_N$.

(b) For any $m \in [d_N]$ there exists $f \in [d_{2N}]$ such that

$$2\mathcal{B}_m^{(N)} \subset \mathcal{B}_f^{(2N)}. \tag{3.1.18}$$

(c) For any $N \in \mathbb{N}$, $m \in [d_N]$ and $i \in [N]$ the line segment $x = i$ intersects $\mathcal{B}_m^{(N)} \setminus (\mathcal{B}_m^{(N)})^\circ$ at most 2 times.

Then if for all $(i, j) \in [N]^2$ the variance of the (i, j)-entry of A_N is given by

$$s_{i,j}^{(N)} := \sum_{m\in[d_N]} s_m^{(N)} 1_{(i,j)\in\mathcal{B}_m^{(N)}} \tag{3.1.19}$$

for some set of numbers $\{s_i\}_{i\in[d_N]}$ so that $s_m^{(N)} = s_k^{(N)}$ if $R(\mathcal{B}_m^{(N)}) = \mathcal{B}_k^{(N)}$, we will call the sequence of matrices $(A_N)_{N\geq 1}$ random matrix model whose variance profile is given by a generalized step function.

The following Theorem is a corollary of Theorem 3.1.8 and gives results of the type (3.1.17) for the operator norm of the matrix

• A_N when A_N is a non-periodic band matrix with band size proportional to N or has a step or continuous profile.

• $A_N A_N^T$ (i.e., Gram matrix) when A_N is a rectangular matrix with step or continuous variance profile.

Details are given after the next theorem and in subsection 3.7.2.

Theorem 3.1.10. *Let $(A_N)_{N\in\mathbb{N}^+}$ be a random matrix model whose variance profile is given by a generalized step function. If it also satisfies Assumptions 3.1.2, 3.1.3, and for every $N \in \mathbb{N}$ and $(i, j) \in [N]^2$ it is true that*

$$s_{i,j}^{(N)} \leq s_{2i,2j}^{(2N)}, \tag{3.1.20}$$

then

$$\lim_{N\to\infty} \frac{\|A_N\|_{\text{op}}}{\sqrt{N}} = \mu_\infty \quad \textit{in probability}, \tag{3.1.21}$$

where μ_∞ is defined in (3.1.9). Moreover, if the sequence $(A_N)_{N\in\mathbb{N}^+}$ satisfies Assumption 3.1.7 the convergence in (3.1.21) holds in the almost sure sense.

For any $N \in \mathbb{N}^+$ and any two $N \times N$ matrices A, B we will denote by $A \odot B$ their Hadamard product, which is the entry-wise product of A, B,i.e., the $N \times N$ matrix with entries

$$\{A \odot B\}_{i,j} = \{A\}_{i,j}\{B\}_{i,j} \text{ for all } i, j \in [N]. \tag{3.1.22}$$

Note that Assumption 3.1.7 is satisfied if A_N can be written as

$$A_N = \Sigma_N \odot A'_N, \tag{3.1.23}$$

where A'_N is a sequence of symmetric random matrices with i.i.d. entries all following the same law, with 0 mean, unit variance and finite $4 + \delta$ moment for some $\delta > 0$ and for each N the entries of Σ_N are elements of $[0, 1]$.

Next, we study the operator norm of two widespread random matrix models.

Definition 3.1.11 (Step function variance profile)**.** Consider

a) $m \in \mathbb{N}^+$ and numbers $\{\sigma_{p,q}\}_{p,q \in [m]} \in [0, 1]^{m \times m}$ with $\sigma_{p,q} = \sigma_{q,p}$ for all $p, q \in [m]$.

b) For each $N \in \mathbb{N}^+$, a partition of $[N]$ into m intervals $\{I_p^{(N)}\}_{p \in [m]}$. The numbering of the intervals is such that $x < y$ whenever $x \in I_p^{(N)}, y \in I_q^{(N)}$ and $p < q$. Let $L_p^{(N)}$ and $R_p^{(N)}$ be the left and right endpoint respectively of $I_p^{(N)}$.

c) Numbers $0 = a_0 < a_1 < \cdots < a_{m-1} < a_m := 1$. We assume that $\lim_{N \to \infty} R_p^{(N)}/N = a_p$ for each $p \in [m]$.

d) A random variable X_0 with $\mathbf{E}(X_0) = 0, \mathbf{E}(X_0^2) = 1$.

For each $N \in \mathbb{N}^+$, define the matrix $\Sigma_N \in \mathbb{R}^{N \times N}$ by $(\Sigma_N)_{i,j} = \sigma_{p,q}$ if $i \in I_p^{(N)}, j \in I_q^{(N)}$, and let $\{A_N\}_{N \in \mathbb{N}^+}$ be the sequence of symmetric random matrices defined by

$$A_N = \Sigma_N \odot A'_N \tag{3.1.24}$$

where A'_N is symmetric and its entries are independent (up to symmetry) random variables all with distribution the same as X_0. Then $(A_N)_{N \in \mathbb{N}^+}$ will be called *symmetric random matrix model whose variance profile is given by a step function.*

Let $\hat{I}_p := [a_{p-1}, a_p)$ for $p \in [m-1]$, and $\hat{I}_m := [a_{m-1}, 1]$. These intervals together with the numbers from a) determine a function $\sigma : [0, 1]^2 \to [0, 1]$ as follows

$$\sigma(x, y) := \sigma_{p,q} \text{ if } x \in \hat{I}_p, y \in \hat{I}_q. \tag{3.1.25}$$

We call the function σ^2 the variance profile of the model.

Definition 3.1.12 (Continuous function variance profile)**.** For

a) a continuous and symmetric function $\sigma : [0,1]^2 \to [0,1]$ (i.e, $\sigma(x,y) = \sigma(y,x)$ for all $x, y \in [0,1]$),

b) a sequence $(\Sigma_N)_{N\in\mathbb{N}^+}$ of symmetric matrices, $\Sigma_N \in [0,1]^{N\times N}$, with the property

$$\lim_{N\to\infty} \sup_{1\le i,j\le N} \left|(\Sigma_N)_{i,j} - \sigma(i/N, j/N)\right| = 0, \tag{3.1.26}$$

c) a random variable X_0 with $\mathbf{E}(X_0) = 0, \mathbf{E}(X_0^2) = 1$,

consider the sequence $\{A_N\}_{N\in\mathbb{N}^+}$ of symmetric random matrices, A_N of dimension $N \times N$, defined by

$$A_N = \Sigma_N \odot A'_N \tag{3.1.27}$$

where the entries of A'_N are independent (up to symmetry) random variables all with distribution the same as X_0. Then we say that $(A_N)_{N\in\mathbb{N}^+}$ is a *random matrix model whose variance profile is given by a continuous function*. Again, we call the function σ^2 the variance profile.

Remark 3.1.13 (Checking Assumption 3.1.2). A sufficient condition for the validity of Assumption 3.1.2 is that $(A_N)_{N\in\mathbb{N}^+}$ satisfies Assumption 3.1.1 and there is a graphon W such that $W_N \to W$ almost everywhere in $[0,1]\times[0,1]$.

Indeed, the bounded convergence theorem gives that $t(T, W_N) \to t(T, W)$ for all trees. Then Theorem 3.2 (a) of [13] shows that the ESD of $A_N/\sqrt{N}$ converges almost surely weakly to a probability measure $\mu^{\sqrt{W}}$ whose $2k$ moment equals

$$\lim_{N\to\infty} \sum_{T\in\mathbf{C}_k} t(T, W_N) \tag{3.1.28}$$

while the moments of odd order are 0. Then, for each $T \in \mathbf{C}_k$,

$$0 \le t(T, W_N) - \sum_{\substack{\mathbf{i}\in[N]^{K+1} \\ \text{labeling of } T}} N^{-k-1} \prod_{\{i,j\}\in E(T)} s_{i,j}^{(N)} = O(1/N), \tag{3.1.29}$$

because $t(T, W_N)$ is simply the same as the sum in the previous relation with the only difference that $\mathbf{i}$ is not required to be a labeling, i.e., it can have repetitions. It follows that Assumption 3.1.2 holds. As we remarked after (3.1.8), $\mu^{\sqrt{W}}$ is symmetric and has bounded support. Denote by $\mu_\infty^{\sqrt{W}}$ the largest element of the support.

If, in the two models above, X_0 has finite $4+\delta$ moment for some small $\delta > 0$, then it is easy to see that the sequence $(A_N)_{N\in\mathbb{N}^+}$ satisfies Assumptions 3.1.3. It also satisfies Assumption 3.1.2 because it satisfies Assumption 3.1.1 and, in both cases, $W_N(x,y)$ converges to $\sigma^2(x,y)$ for almost all $(x,y) \in [0,1]\times[0,1]$, thus the preceding remark applies.

Our result for the model (3.1.24) is the following.

Theorem 3.1.14. *Let $(A_N)_{N\in\mathbb{N}^+}$ be a random matrix model whose variance profile is given by a step function as above. Assume that X_0 has mean value 0, variance 1, and finite $4+\delta$ moment, for some small $\delta > 0$. Then it is true that*

$$\lim_{N\to\infty} \frac{\|A_N\|_{\text{op}}}{\sqrt{N}} = \mu_\infty^\sigma \quad a.s. \tag{3.1.30}$$

The previous theorem together with an approximation result that we prove in Section 3.6 (Proposition 3.6.1) has the following consequence for the model (3.1.27).

Corollary 3.1.15. *Let $(A_N)_{N\in\mathbb{N}^+}$ be a sequence of matrices whose variance profile is given by a continuous function. If X_0 has mean zero, variance one, and finite $4+\delta$ moment, then*

$$\lim_{N\to\infty} \frac{\|A_N\|_{\text{op}}}{\sqrt{N}} = \mu_\infty^\sigma \quad a.s. \tag{3.1.31}$$

Remark 3.1.16. 1) Theorem 3.1.14 covers the cases in the Wigner matrix model [i. e., $A_N := (a_{i,j})_{i,j\in[N]}$ with $\{a_{i,j} : 1 \le i \le j \le N, N \in \mathbb{N}^+\}$ i.i.d. with $\mathbf{E}(a_{1,1}) = 0, \mathbf{E}(a_{1,1}^2) = 1$] where $\mathbf{E}(|a_{1,1}|^{4+\delta}) < \infty$ for some $\delta > 0$. Recall that the necessary and sufficient condition for the validity of (3.1.30) in that model is $\mathbf{E}(|a_{1,1}|^4) < \infty$.

2) Corrolary 3.1.15 holds also in the case that the function σ of Definition 3.1.12 is piecewise continuous in a sense explained in the end of Section 3.6.

3.2 Analysis of high order moments

Assume at the moment that the entries of A_N have finite moments of all orders.

We will relate the largest eigenvalue with a high moment of the measure μ_N and at the same time this moment will be controlled by μ_∞. In general, for $k \in \mathbb{N}$, it is true that

$$\mathbf{E}\,\text{tr}(A_N^{2k}) = \sum_{i_1,i_2,\dots\dots\dots,i_{2k}\in[N]} \mathbf{E}\left(\prod_{l=1}^{2k} a_{i_l,i_{l+1}}^{(N)}\right) \tag{3.2.1}$$

with the conventions that $i_{2k+1} = i_1$, when $k = 0$ the sum is only over $i_1 \in [N]$, and the product over an empty set equals 1.

Now, for a term with indices $i_1, i_2, \dots, i_{2k}$, we let $\mathbf{i} := (i_1, i_2, \dots, i_{2k})$ and $X(\mathbf{i}) := \prod_{l=1}^{2k} a_{i_l,i_{l+1}}^{(N)}$. For such an $\mathbf{i}$ we also use the term cycle. Then consider the graph $G(\mathbf{i})$ with vertex set

$$V(\mathbf{i}) = \{i_1, i_2, \dots, i_{2k}\}$$

and set of edges

$$\{\{i_r, i_{r+1}\} : r = 1, 2, \dots, 2k\}. \tag{3.2.2}$$

As explained in [52] (in the proof of relation (3.1.6) there, pages 49, 50 or in Theorem 3.2 of [13]), the limit

$$\lim_{N\to\infty} \frac{1}{N^{k+1}} \mathbf{E}\,\text{tr}(A^{2k})$$

remains the same if in the sum of (3.2.1) we keep only the summands such that

$$\text{the graph } G(\mathbf{i}) \text{ is a tree with } k+1 \text{ vertices} \tag{3.2.3}$$

Then, necessarily, the path $i_1 \to i_2 \to \cdots \to i_{2k} \to i_1$ traverses each edge of the tree exactly twice, in opposite directions of course. Such a $G(\mathbf{i})$ becomes an ordered rooted tree if we mark i_1 as the root and order children of the same vertex according to the order they appear in the cycle.

Cycles $\mathbf{i}$ that don't satisfy (3.2.3) we call bad cycles. So, for $k \in \mathbb{N}$, the sum in (3.2.1) can be written as

$$\mathbf{E}\,\mathrm{tr}(A^{2k}) = M_N(k) + B_N(k), \tag{3.2.4}$$

where

$$M_N(k) := \sum_{T \in \mathbf{C}_k} \sum_{\mathbf{i} \in [N]^{(2k)\vee 1}: G(\mathbf{i}) \sim T} \prod_{\{i,j\} \in E(G(\mathbf{i}))} s_{i,j}^{(N)}, \tag{3.2.5}$$

$$B_N(k) := \sum_{\mathbf{i} \in [N]^{2k}: \text{bad cycle}} \mathbf{E}X(\mathbf{i}). \tag{3.2.6}$$

Recall that $\mathbf{C}_k$ are the ordered rooted trees with k edges and $G(\mathbf{i}) \sim T$ means that the graphs are isomorphic as ordered rooted trees. Note also that $M_N(k)$ has already been defined in (3.1.7) but the two definitions for it agree. Also, $M_N(0) = N, B_N(0) = 0$.

The plan is to control the expectation of the trace in (3.2.4) through an appropriate bound involving various $M_N(j)'s$. To control the term $B_N(k)$, we adopt the analysis of Section 2.3 of [16].

Proposition 3.2.1. *Let A_N be an $N \times N$ symmetric random matrix with independent entries (up to symmetry) and with $\mathbf{E}(a_{i,j}^{(N)}) = 0$, $s_{i,j}^{(N)} \le 1$ for all $N \in \mathbb{N}$, $i,j \in [N]$. Assume additionally that the absolute value of the entries of the matrix are all supported in $[0, CN^{\frac{1}{2}-\epsilon}]$ for some $\epsilon > 0$. Then for all N large enough and all integers $1 \le k < N$ it is true that*

$$|B_N(k)| \le \sum_{s=1}^{k} (4k^5)^{2k-2s} \left(CN^{\frac{1}{2}-\epsilon}\right)^{2k-2s} \sum_{t=1}^{(s+1)\wedge k} (4k^4)^{4(s+1-t)} M_N(t-1). \tag{3.2.7}$$

Proof. We bound each term of the sum defining $B_N(k)$. Take a bad cycle $\mathbf{i}$ and let

- t: the number of vertices of $G(\mathbf{i})$,
- s: the number of the edges of $G(\mathbf{i})$,
- $e_1, e_2, \dots, e_s$: the edges of $G(\mathbf{i})$ in order of appearance in the cycle,
- $a_1, a_2, \dots, a_s$: the multiplicities of $e_1, e_2, \dots, e_s$ in the cycle.

That is, a_q is the number of times the (undirected) edge e_q appears in the cycle. Note that $t \le s+1$ (true for all graphs) and $t \le k$ because the cycle is bad.

Additionally, in case $t \ge 2$, we let $T(\mathbf{i})$ be the rooted ordered tree obtained from $G(\mathbf{i})$ by keeping only edges that lead to a new vertex at the time of their appearance in the cycle. The root is i_1 and we declare a child of a vertex smaller than another if it appears earlier in the cycle. In case $t = 1$, $T(\mathbf{i})$ is the graph with one vertex, i_1, and one edge (loop) with end vertices i_1, i_1. Thus, $T(\mathbf{i})$ has t vertices and $1 \vee (t-1)$ edges.

To bound $|\mathbf{E}X(\mathbf{i})|$, notice that if any of $a_1, a_2, \ldots, a_s$ is 1, we have $\mathbf{E}X(\mathbf{i}) = 0$ by the independence of the elements of A_N and the zero mean assumption. We assume therefore that all multiplicities are at least 2. Using the information about the mean, variance, and support of $|a_{i,j}^{(N)}|$, we get that for any integer $a \ge 2$ it holds $\mathbf{E}(|a_{i,j}^{(N)}|^a) \le (C_1 N^{1/2-\varepsilon})^{a-2} s_{i,j}^{(N)}$. Thus

$$\mathbf{E}|X(\mathbf{i})| = \prod_{q=1}^{s} \mathbf{E}|X_{e_q}|^{a_q} \le (CN^{1/2-\varepsilon})^{a_1+\cdots+a_s-2s} \prod_{\{i,j\}\in E(G(\mathbf{i}))} s_{i,j}^{(N)} \le (CN^{1/2-\varepsilon})^{2k-2s} \prod_{\{i,j\}\in E(T(\mathbf{i}))} s_{i,j}^{(N)}. \tag{3.2.8}$$

In the second inequality, we used the fact that $s_{i,j}^{(N)} \in [0,1]$ for all i, j, N. For integers $s, t \ge 1, a_1, \ldots, a_s \ge 2$ and $T \in \mathbf{C}_{t-1}$ let

$$N_{T,a_1,a_2,\ldots,a_s} = \begin{array}{l} \text{the number of bad cycles with } T(\mathbf{i}) \sim T, \text{ indices } 1,2,\ldots,t, \text{ appearing in this order,} \\ \text{and edge multiplicities } a_1, a_2, \ldots, a_s. \end{array} \tag{3.2.9}$$

Using the bound on $N_{T,a_1,a_2,\ldots,a_s}$ provided by Lemma 3.8.1, we obtain

$$|B_N(k)| \le \sum_{s=1}^{k} \sum_{t=1}^{k\wedge(s+1)} \sum_{a_1,a_2,\ldots,a_s} (CN^{1/2-\varepsilon})^{2k-2s} \left\{ \mathbf{1}_{t=1} \sum_{i\in[N]} s_{i,i}^{(N)} + \mathbf{1}_{t\ge 2} \sum_{T\in\mathbf{C}_{t-1}} N_{T,a_1,a_2,\ldots,a_s} \sum_{\mathbf{i}\in[N]^{2k}: T(\mathbf{i})\sim T} \prod_{\{i,j\}\in E(T(\mathbf{i}))} s_{i,j}^{(N)} \right\} \tag{3.2.10}$$

$$\le \sum_{s=1}^{k} \sum_{t=1}^{k\wedge(s+1)} \sum_{a_1,a_2,\ldots,a_s} (CN^{1/2-\varepsilon})^{2k-2s} \left\{ \mathbf{1}_{t=1} \sum_{i\in[N]} s_{i,i}^{(N)} + \mathbf{1}_{t\ge 2} \sum_{T\in\mathbf{C}_{t-1}} N_{T,a_1,a_2,\ldots,a_s} \sum_{\mathbf{i}\in[N]^{2(t-1)}: T(\mathbf{i})\sim T} \prod_{\{i,j\}\in E(T(\mathbf{i}))} s_{i,j}^{(N)} \right\} \tag{3.2.11}$$

$$\le \sum_{s=1}^{k} \sum_{t=1}^{k\wedge(s+1)} \sum_{a_1,a_2,\ldots,a_s} (CN^{1/2-\varepsilon})^{2k-2s} (4k^4)^{4(s+1-t)+2(k-s)} M_N(t-1). \tag{3.2.12}$$

We used here the fact that $s_{i,i}^{(N)} \le 1$, so that $\sum_{i\in[N]} s_{i,i}^{(N)} \le N = M_N(0)$. The inside sum in (3.2.12) is over all s-tuples of integers $a_1, a_2, \ldots, a_s$ greater than or equal to 2 with sum $2k$. By subtracting 2 from each a_i, we get an s-tuple of non-negative integers with sum $2k-2s$. The number of such s-tuples is $\left(\!\binom{s}{2k-2s}\!\right)$(combinations with repetition), which is at most $s^{2(k-s)} \le k^{2(k-s)}$. Thus the above sum is bounded by

$$\sum_{s=1}^{k} (4k^5)^{2(k-s)} (CN^{1/2-\varepsilon})^{2k-2s} \sum_{t=1}^{k\wedge(s+1)} (4k^4)^{4(s+1-t)} M_N(t-1). \tag{3.2.13}$$

□

Proposition 3.2.2. *Let $(A_N)_{N\in\mathbb{N}^+}$ be a sequence of symmetric matrices and $R > 0$ so that the sequence satisfies Assumption 3.1.3 and the following condition $\Sigma(R)$:*
For each $C_1 > 0$ there are $C_2 > 0$ and $N_0 \in \mathbb{N}^+$ such that

$$M_N(k) \leq C_2 N^{k+1} R^{2k} \tag{3.2.14}$$

for all $N, k \in \mathbb{N}^+$ with $N \geq N_0$ and $1 \leq k \leq C_1 \log N$.
Then for each $\epsilon > 0$, it holds

$$\lim_{N\to\infty} \mathbb{P}\left(\frac{\|A_N\|_{\text{op}}}{\sqrt{N}} \geq R(1+\epsilon)\right) = 0. \tag{3.2.15}$$

Proof. Fix $\eta \in (0, 1/8)$ and define the $N \times N$ matrices $A_N^{\leq}, A_N^{>}$ by

$$(A_N^{\leq})_{i,j} := a_{i,j}^{(N)} 1_{|a_{i,j}|\leq N^{\frac{1}{2}-\eta}}, \tag{3.2.16}$$

$$(A_N^{>})_{i,j} := a_{i,j}^{(N)} 1_{|a_{i,j}|> N^{\frac{1}{2}-\eta}} \tag{3.2.17}$$

for all $i, j \in [N]$. For a random matrix $H := (h_{i,j})$, $\mathbf{E}H$ denotes the matrix whose (i, j) element is $\mathbf{E}h_{i,j}$ provided that the mean value of $h_{i,j}$ can be defined. Note that

$$\frac{1}{\sqrt{N}}\|A_N\|_{\text{op}} \leq \frac{1}{\sqrt{N}}\left(\|A_N^{\leq} - \mathbf{E}A_N^{\leq}\|_{\text{op}} + \|\mathbf{E}A_N^{\leq}\|_{\text{op}} + \|A_N^{>}\|_{\text{op}}\right). \tag{3.2.18}$$

We will bound the three terms in the right hand side of the last inequality. For the first two, we use only Assumption 3.1.3 and the arguments in the proof of Theorem 2.3.23 in [16].

1) The term $N^{-\frac{1}{2}}\|\mathbf{E}(A_N^{\leq})\|_{\text{op}}$ is a deterministic sequence that converges to 0 because, since $a_{i,j}^{(N)}$ is centered, we have

$$|(\mathbf{E}A_N^{\leq})_{i,j}| = |(\mathbf{E}A_N^{>})_{i,j}| \leq N^{-3(\frac{1}{2}-\eta)} \sup_N \max_{i,j} \mathbf{E}|a_{i,j}^{(N)}|^4. \tag{3.2.19}$$

And using the inequality $\|C\|_{\text{op}} \leq N\|C\|_{\max}$, we get that

$$\|\mathbf{E}(A_N^{\leq})\|_{\text{op}} \leq N^{3\eta-\frac{1}{2}} \sup_N \max_{i,j} \mathbf{E}|a_{i,j}^{(N)}|^4 \overset{N\to\infty}{\to} 0.$$

2) The term $N^{-\frac{1}{2}}\|A_N^{>}\|_{\text{op}}$ converges to 0 in probability. Indeed, for any $\delta_1 > 0$,

$$\mathbf{P}(\|A_N^{>}\|_{\text{op}} > \delta_1\sqrt{N}) \leq \mathbf{P}(|a_{i,j}^{(N)}| > \delta_1\sqrt{N} \text{ for some } i, j \in [N]) \tag{3.2.20}$$

$$+\mathbf{P}(\|A_N^{>}\|_{\text{op}} > \delta_1\sqrt{N} \text{ and } |a_{i,j}^{(N)}| \leq \delta_1\sqrt{N} \text{ for all } i, j \in [N]). \tag{3.2.21}$$

The first quantity goes to zero as $N \to \infty$ because of (3.1.10). For the second, it is an easy exercise to show that if each entry of a matrix M has absolute value at most a and each row and column of M has at most one non-zero element then $|M|_{\mathrm{op}} \le a$ (use the expression $|M|_{\mathrm{op}} = \sup_{x:\|x\|_2=1} \|Mx\|_2$). Consequently, the probability in (3.2.21) is at most

$$\sum_{i=1}^{N} \sum_{1\le j_1<j_2\le N} \mathbf{P}(|a_{i,j_1}| > N^{\frac{1}{2}-\eta}, |a_{i,j_2}| > N^{\frac{1}{2}-\eta}) + \sum_{j=1}^{N} \sum_{1\le i_1<i_2\le N} \mathbf{P}(|a_{i_1,j}| > N^{\frac{1}{2}-\eta}, |a_{i_2,j}| > N^{\frac{1}{2}-\eta}) \tag{3.2.22}$$

$$\le 2N\binom{N}{2}\frac{\sup_{N\in\mathbb{N}^+,i,j\in[N]}(\mathbf{E}\{|a_{i,j}^{(N)}|^4\})^2}{N^{4-8\eta}} \le \frac{1}{N^{1-8\eta}} \sup_{N\in\mathbb{N}^+,i,j\in[N]} (\mathbf{E}\{|a_{i,j}^{(N)}|^4\})^2 \overset{N\to\infty}{\to} 0. \tag{3.2.23}$$

We used the independence of the entries in each row or column and Markov's inequality.

3) To deal with $|A_N^{\le} - \mathbf{E}A_N^{\le}|_{\mathrm{op}}$, we will use Proposition 3.2.1. Let

$$\tilde{A}_N := A_N^{\le} - \mathbf{E}A_N^{\le}, \tag{3.2.24}$$

$$s_{i,j}^{(N),\le} := \mathbf{E}\{(\tilde{A}_N)_{i,j}^2\}. \tag{3.2.25}$$

Proposition 3.2.1 applies to $\tilde{A}_N$ because any element of the matrix, say $(\tilde{A}_N)_{i,j}$, has zero mean and variance $s_{i,j}^{(N),\le} \le \mathbf{E}\{(A_N^{\le})^2\} \le \mathbf{E}(A_N^2) = s_{i,j}^{(N)} \le 1$. Thus, if we denote by $\tilde{M}_N(m)$ the terms (3.1.7) for $m \in [N]$ and for the matrix $\tilde{A}_N$, we will have $\tilde{M}_N(m) \le M_N(m)$ for all $m \in \mathbb{N}$, and Proposition 3.2.1, gives that for any $1 \le k < N$,

$$\mathbf{E}\,\mathrm{tr}(\tilde{A}_N^{2k}) \le M_N(k) + \sum_{s=1}^{k}(4k^5)^{2k-2s}\left(2N^{\frac{1}{2}-\eta}\right)^{2k-2s} \sum_{t=1}^{(s+1)\wedge k} (4k^4)^{4(s+1-t)} M_N(t-1) \tag{3.2.26}$$

Now fix $C_1 > 0$, its value will be determined in (3.2.32) below. For $1 \le k \le C_1 \log N$,

$$\mathbf{E}\,\mathrm{tr}(\tilde{A}_N^{2k}) \le C_2 N^{k+1} R^{2k} + C_2 \sum_{s=1}^{k}(4k^5)^{2k-2s}\left(2N^{\frac{1}{2}-\eta}\right)^{2k-2s} \sum_{t=1}^{(s+1)\wedge k} (4k^4)^{4(s+1-t)} N^t R^{2(t-1)}. \tag{3.2.27}$$

Next, we focus on the second summand in the right hand side of the previous inequality for N large enough. In the sum in t we factor out $(4k^4)^{4(s+1)}R^{-2}$, and in the resulting sum of geometric progression with ratio a larger than 1 we use the bound $a + a^2 + \cdots + a^{(s+1)\wedge k} \le ka^{(s+1)\wedge k}$. Thus the sum in (3.2.27) is bounded by

$$C_2 \frac{k}{R^2} \sum_{s=1}^{k}(4k^5)^{2k-2s}\left(2N^{\frac{1}{2}-\eta}\right)^{2k-2s} (4k^4)^{4(s+1)} \left(\frac{NR^2}{(4k^4)^4}\right)^{(s+1)\wedge k} \tag{3.2.28}$$

$$= 2^8 C_2 k^{17} N^k (R^2)^{k-1} + C_2 \frac{k}{R^2}(NR^2)^{k+1} \sum_{s=1}^{k-1}\left(\frac{4(4k^5)^2}{N^{2\eta}R^2}\right)^{k-s} \tag{3.2.29}$$

$$\le 2^8 C_2 k^{17} N^k (R^2)^{k-1} + 2^7 C_2 k^{11} (R^2)^{(k-1)} N^{k+1-2\eta} \le 2^9 C_2 k^{17} N^{k+1-2\eta} R^{2k-2} \tag{3.2.30}$$

[in summing the geometric series in (3.2.29), we used the bound $c + c^2 + \cdots + c^r \le 2c$ if $0 \le c < 1/2$]. Thus, returning to (3.2.27),

$$\mathbf{E}\,\mathrm{tr}(\tilde{A}_N^{2k}) \le N^{k+1} R^{2k}\{1 + o(1)\}^{2k}. \tag{3.2.31}$$

with $o(1)$ depending on R, C_2, η.

Fix $\epsilon > 0$, pick

$$C_1 > \frac{2+\epsilon}{\log(1+\epsilon)}, \tag{3.2.32}$$

and apply the above for $k := [C_1 \log N]$. Relation (3.2.31) implies

$$\begin{aligned}\mathbb{P}\left(\frac{\|\tilde{A}_N\|_{\text{op}}}{\sqrt{N}} \geq R(1+\epsilon)\right) &\leq \mathbb{P}\left(\frac{\|\tilde{A}_N\|_{\text{op}}^{2k}}{N^k} \geq R^{2k}(1+\epsilon)^{2k}\right) \leq \frac{1}{R^{2k}(1+\epsilon)^{2k}} \frac{1}{N^k}\mathbf{E}\|\tilde{A}_N\|_{\text{op}}^{2k} \\ &\leq N\left(\frac{1+o(1)}{1+\epsilon}\right)^{2k} = O\left(\frac{1}{N^{1+\epsilon}}\right),\end{aligned} \tag{3.2.33}$$

for any N large enough. The last equality is true because of the choice of k and C_1. □

A tool for proving almost sure convergence of the sequence $\|A_N\|_{\text{op}}/\sqrt{N}$ is the following lemma.

Lemma 3.2.3. *Let $(A_N)_{N\in\mathbb{N}^+}$ be a sequence of matrices, A_N is $N \times N$, and $R > 0$ so that the sequence satisfies Assumption 3.1.3(a), condition $\Sigma(R)$, and Assumption 3.1.7. Then*

$$\limsup_{N\to\infty} \frac{\|A_N\|_{\text{op}}}{\sqrt{N}} \leq R \ \ a.s. \tag{3.2.34}$$

Proof. Pick $\eta \in (0, 1/8)$, its exact value will be determined below, and define the matrices $A_N^{\leq}, \mathbf{E}A_N^{\leq}$ as in the proof of Proposition 3.2.2. The proof will be accomplished once we show that

$$\limsup_{N} \frac{\|A_N^{\leq}\|_{\text{op}}}{\sqrt{N}} \leq R, \quad \text{a.s., and} \tag{3.2.35}$$

$$\mathbf{P}\left(A_N \neq A_N^{\leq} \text{ for infinitely many } N\right) = 0. \tag{3.2.36}$$

Proof of (3.2.35). Since $\Sigma(R)$ holds for the sequence $(A_N)_{N\geq 1}$, the proof of Proposition 3.2.2 (the part with heading 3) shows that

$$\limsup_{N} \frac{\|A_N^{\leq} - \mathbf{E}A_N^{\leq}\|_{\text{op}}}{\sqrt{N}} \leq \mu_\infty \ \ \text{a.s.} \tag{3.2.37}$$

because the upper bound in (3.2.33) is summable with respect to N. In the same proof it is shown that

$$\limsup_{N} \frac{\|\mathbf{E}A_N^{\leq}\|_{\text{op}}}{\sqrt{N}} = 0 \text{ a.s.}$$

Using these two facts and the triangle inequality we get (3.2.35).

Proof of (3.2.36). Let X be the random variable that stochastically dominates the entries of A_N in the sense of (3.1.16). Let X_N be a sequence of symmetric random matrices after an appropriate coupling such that for all $N \in \mathbb{N}$ and $i, j \in [N]$ it is true that

$$|a_{i,j}^{(N)}| \leq |(X_N)_{i,j}| \tag{3.2.38}$$

and the entries of X_N are independent up to symmetry and all following the same law as X. It is an easy exercise to show that for any $a, c > 1$ and Y real valued random variable we have

$$\sum_{k=1}^{\infty} a^k \mathbf{P}(|Y| \geq c^k) \leq \frac{1}{a-1} \mathbf{E}\{|Y|^{\frac{\log a}{\log c}}\}. \tag{3.2.39}$$

Using this inequality and the fact that the random variable X has finite $4+\delta$ moment, we get that all $\eta \leq \frac{\delta}{\delta+4}$ satisfy

$$\sum_{m=1}^{\infty} 2^{2m} \mathbf{P}(|X| \geq 2^{\frac{m}{2}(1-\eta)}) < \infty. \tag{3.2.40}$$

Thus, picking in the beginning of the proof an arbitrary η with $0 < \eta < (1/8) \wedge (\delta/(4+\delta))$, we have

$$\mathbf{P}\Big(A_N \neq A_N^{\leq} \text{ for infinitely many } N\Big) = \mathbf{P}\Big(\text{ for infinitely many } N \text{ there are } i,j \in [N] : |a_{i,j}^{(N)}| > CN^{\frac{1}{2}-\eta}\Big) \tag{3.2.41}$$

$$\leq \mathbf{P}\Big(\text{ for infinitely many } N \text{ there are } i,j \in [N] : |X_{i,j}^{(N)}| > CN^{\frac{1}{2}-\eta}\Big) = \mathbf{P}\Big(X_N \neq X_N^{\leq} \text{ for infinitely many } N\Big). \tag{3.2.42}$$

In the second line, the inequality is a consequence of (3.2.38), and the matrix $X_N^{\leq}$ is the matrix whose (i,j) element is $(X_N)_{i,j}\mathbf{1}_{|(X_N)_{i,j}| \leq CN^{\frac{1}{2}-\eta}}$. The convergence of the series in (3.2.40). implies that the probability in the right hand side of (3.2.42) is 0 (see [52], pages 94 and 95) and finishes the proof of (3.2.36). □

3.3 Proof of Theorem 3.1.8

The convergence $\mu_{A_N/\sqrt{N}} \Rightarrow \mu$ in probability implies that

$$\liminf_N \frac{\|A_N\|_{\text{op}}}{\sqrt{N}} \geq \mu_\infty \quad \text{in probability,} \tag{3.3.1}$$

that is, for all $\epsilon > 0$, $\lim_{N\to\infty} \mathbf{P}(\|A_N\|_{\text{op}}/\sqrt{N} < \mu_\infty - \epsilon) = 0$. So in order to prove Theorem 3.1.8 one needs to prove that

$$\limsup_N \frac{\|A_N\|_{\text{op}}}{\sqrt{N}} \leq \mu_\infty \tag{3.3.2}$$

in probability. By Proposition 3.2.2, it is enough to prove that condition $\Sigma(\mu_\infty)$ is satisfied.

We will prove condition $\Sigma(\mu_\infty)$ separately for each of the Assumptions 3.1.4 and 3.1.6 in the next two lemmas.

Lemma 3.3.1. *Let $(A_N)_{N\in\mathbb{N}^+}$ be a sequence of matrices that satisfies Assumptions 3.1.2, 3.1.3, and 3.1.4. Then for every $k, N \in \mathbb{N}^+$ such that $k < N$ it is true that*

$$M_N(k) \leq N^{k+1}\mu_\infty^{2k}.$$

In case $\mu_\infty > 0$, the inequality is true (as equality) for $k = 0$ also.

Proof. Fix $N, k \in \mathbb{N}^+$ with $k < N$ and a tree $T \in \mathbf{C}_k$. Then, for each $\mathbf{d} := (d_1, d_2, \dots, d_{k+1}) \in \{-1, 0\}^{k+1}$ consider the function

$$\varphi_{\mathbf{d}} : [N]^{k+1} \to [2N]^{k+1}$$

with

$$\varphi_{\mathbf{d}}(i_1, i_2, \cdots, i_{k+1}) = 2(i_1, i_2, \cdots, i_{k+1}) + (d_1, d_2, \cdots, d_{k+1})$$

for all $i_1, i_2, \dots, i_{k+1} \in [N]$. Each $\varphi_{\mathbf{d}}$ is one to one and, for different vectors $\mathbf{d}, \mathbf{d}' \in \{-1, 0\}^{k+1}$, the image of $\varphi_{\mathbf{d}}$ is disjoint from that of $\varphi_{\mathbf{d}'}$. If G is a plane rooted tree whose vertices in order of appearance in a depth first search are $(i_1, i_2, \dots, i_{k+1}) \in [N]^{k+1}$, and $\varphi_{\mathbf{d}}(i_1, i_2, \dots, i_{k+1}) = (j_1, j_2, \dots, j_{k+1})$, we denote by $\varphi_{\mathbf{d}}(G)$ the plane rooted tree with vertex set $\{j_1, j_2, \dots, j_{k+1}\}$, root j_1, and edges $\{\{j_a, j_b\} : \{a, b\}$ is an edge of $G\}$. Note that if all coordinates of $\mathbf{i} \in [N]^{k+1}$ are different, the same is true for the coordinates of $\varphi_{\mathbf{d}}(\mathbf{i})$.

Lastly, by assumption 3.1.4, for any $T \in \mathbf{C}_k, \mathbf{i} \in [N]^{2k}$ such that $G(\mathbf{i}) \sim T$ and $\mathbf{d} \in \{-1, 0\}^{k+1}$, it is true that

$$\prod_{\{i,j\} \in E(G(\mathbf{i}))} s_{i,j}^{(N)} \leq \prod_{\{i,j\} \in E(\varphi_{\mathbf{d}}(G(\mathbf{i})))} s_{i,j}^{(2N)}. \tag{3.3.3}$$

So if one sums over all possible trees in $\mathbf{C}_k$ and $d \in \{-1, 1\}^{k+1}$, (3.3.3) implies that

$$2^{k+1} M_N(k) = \sum_{\mathbf{d} \in \{-1,0\}^{k+1}} \sum_{T \in \mathbf{C}_k} \sum_{\mathbf{i} \in [N]^{2k} : G(\mathbf{i}) \sim T} \prod_{\{i,j\} \in E(G(\mathbf{i}))} s_{i,j}^{(N)} \leq \sum_{\mathbf{d} \in \{-1,0\}^{k+1}} \sum_{T \in \mathbf{C}_k} \sum_{\mathbf{i} \in [N]^{2k} : G(\mathbf{i}) \sim T} \prod_{\{i,j\} \in E(\varphi_{\mathbf{d}}(G(\mathbf{i})))} s_{i,j}^{(2N)} \leq M_{2N}(k). \tag{3.3.4}$$

By applying (3.3.4) inductively, one can prove that for fixed $N, k \in \mathbb{N}$ the sequence

$$q_m := M_{2^m N}(k) / (2^m N)^{k+1}, m \in \mathbb{N}$$

is increasing in the variable m. So by (3.1.8) it is true that

$$\sup_m q_m = \lim_{m \to \infty} q_m = \int x^{2k} \, d\mu(x) \leq \mu_\infty^{2k}.$$

In particular, $q_0 \leq \mu_\infty^{2k}$, completing the proof. □

Lemma 3.3.2. *Suppose $(A_N)_{N \in \mathbb{N}^+}$ is a sequence of matrices such that Assumptions 3.1.3,3.1.6 hold. Then for each $C_1 > 0$ there is $C_2 > 0$ such that*

$$M_N(k) \leq C_2 N^{k+1} \mu_\infty^{2k} \tag{3.3.5}$$

for all $N \in \mathbb{N}^+$ and $1 \leq k \leq C_1 \log N$.

Proof. Note that for $1 \leq k < N$,

$$\frac{1}{N^{k+1}} M_N(k) \leq \sum_{T \in \mathbf{C}_k} \int_{[0,1]^{k+1}} \Big(\prod_{\{i,j\} \in E(T)} W_N(x_i, x_j) \Big) dx_1 dx_2 \cdots dx_{k+1} =: \Xi_N(k). \tag{3.3.6}$$

The inequality holds because the left hand side results if on the right hand side we restrict the domain of integration to the union of the sets $\prod_{r=1}^{k+1}((i_r-1)/N, i_r/N]$ where all $i_1, i_2, \ldots, i_{k+1} \in [N]$ are different. Thus, it is enough to show (3.3.5) with the left hand side replaced by $N^{k+1}\Xi_N(k)$.

Fix $T \in \mathbf{C}_k$ and enumerate the edges of T in the order of first appearance during a depth first search algorithm. For $\{i,j\} \in E(T)$, let $\{i,j\}_{\rm ord}$ be its enumeration. Then for any integer $l \in [0, k]$ define the following quantities.

$$\mu_N^{(l)}(k,T) = \int_{[0.1]^{k+1}} \prod_{\{i,j\}\in E(T):\{i,j\}_{\rm ord}\le l} W_N(x_i,x_j) \prod_{\{i,j\}\in E(T):\{i,j\}_{\rm ord}\ge l+1} W(x_i,x_j) dx_1 dx_2 \cdots dx_{k+1}. \quad (3.3.7)$$

Note that

$$\sum_{T\in\mathbf{C}_k} \mu_N^{(0)}(k,T) = \mu_{2k}, \quad \sum_{T\in\mathbf{C}_k} \mu_N^{(k)}(k,T) = \Xi_N(k) \quad (3.3.8)$$

Fix $D > 0$. Since all the variances are uniformly bounded by 1, Assumption (3.1.15) implies that there exists some $N_0(D)$ and $C > 0$ such that for $N \ge N_0(D)$ and any $1 \le l \le k < N$ it is true that

$$|\mu_N^{(l)}(k,T) - \mu_N^{(l-1)}(k,T)| \le \int_{[0,1]^2} |W_N(x,y) - W(x,y)|\, dx\, dy \le C\frac{1}{N^D}. \quad (3.3.9)$$

Consequently, since $|\mathbf{C}_k| \le 2^{2k}$, for $k < N$ we have

$$|\Xi_N(k) - \mu_{2k}| \le \sum_{T\in\mathbf{C}} \left| \sum_{l=1}^{k} \{\mu_N^{(l)}(k,T) - \mu_N^{(l-1)}(k,T)\} \right| \le \sum_{T\in\mathbf{C}_k} \sum_{l=1}^{k} |\mu_N^{(l)}(k,T) - \mu_N^{(l-1)}(k,T)| \le \frac{Ck2^{2k}}{N^D}. \quad (3.3.10)$$

Pick any $D > -2C_1 \log(\mu_\infty/2)$. Then there is $N_0' \in \mathbb{N}^+, N_0' > N_0(D)$ such that $Ck2^{2k}/N^D \le \mu_\infty^{2k}$ for all $N > N_0'$ and $1 \le k \le C_1 \log N$. And since $\mu_{2k} \le \mu_\infty^{2k}$, we will have $\Xi_N(k) \le 2\mu_\infty^{2k}$ for the same N and k. If we choose a constant $C_2 \ge 2$ so that (3.3.5) is satisfied for $N \in [N_0']$ and $1 \le k \le C_1 \log N$, then we will have (3.3.5) for all N, k claimed. □

3.3.1 Proof of almost sure convergence under the additional Assumption 3.1.7

The convergence in probability that we have proven so far gives

$$\liminf_{N\to\infty} \frac{\|A_N\|_{\rm op}}{\sqrt{N}} \ge \mu_\infty \text{ a.s.} \quad (3.3.11)$$

The opposite inequality follows from Lemma 3.2.3 whose assumptions are satisfied, with $R = \mu_\infty$, because, under both scenarios of the Theorem, assumption $\Sigma(\mu_\infty)$ holds.

3.4 Proof of Theorem 3.1.10

The plan is to write the matrix A_N as $A_N^{(1)} + A_N^{(2)}$ so that for the sequence $\{A_N^{(1)}\}_{N\ge1}$ we can apply Theorem 3.1.8 while for $\{A_N^{(2)}/\sqrt{N}\}_{N\ge1}$ the operator norm will tend to zero.

Let

$$\mathcal{D}_N := \{(i,j) \in [N]^2 : \text{ there exists } m \in [d_N] : (i,j) \in (\mathcal{B}_m^{(N)})^\circ\}. \tag{3.4.1}$$

Then define the matrices

$$\{A_N^{(1)}\}_{i,j} := \mathbf{1}_{(i,j)\in\mathcal{D}_N}\{A_N\}_{i,j}, \tag{3.4.2}$$

$$\{A_N^{(2)}\}_{i,j} = \mathbf{1}_{(i,j)\notin\mathcal{D}_N}\{A_N\}_{i,j}. \tag{3.4.3}$$

The proof follows from the triangle inequality for the operator norm and the following two statements, which we are going to prove next.

$$\lim_{N\to\infty} \frac{\|A_N^{(2)}\|_{\text{op}}}{\sqrt{N}} = 0 \quad \text{in probability.} \tag{3.4.4}$$

$$\lim_{N\to\infty} \frac{\|A_N^{(1)}\|_{\text{op}}}{\sqrt{N}} = \mu_\infty. \tag{3.4.5}$$

Proof of (3.4.4). For any $k \in \mathbb{N}$ denote by $M_N^{(2)}(k)$ the quantity (3.1.7) but with the role of A_N played by $A_N^{(2)}$, i.e., $s_{i,j}^{(N)}$ is replaced by $s_{i,j}^{(N)}\mathbf{1}_{(i,j)\notin\mathcal{D}_N}$ By Proposition 3.2.2 it is sufficient to prove that for any constant $C_1 > 0$ it is true that for any $k \le C_1 \log N$,

$$M_N^{(2)}(k) \le N(8d_N)^k. \tag{3.4.6}$$

This is true because each product in (3.1.7) is at most 1, then the inner sum has at most $N(2d_N)^k$ non zero terms [there are N choices for i_1, and then, for each choice of i_1 there are at most $2d_N$ choices for i_2 that have $s_{i_1,i_2}^{(N)} \ne 0$ due to condition (c) of Definition 3.1.9, and the same restriction holds for $i_3, \dots, i_{k+1}$] and the outer sum has $|\mathbf{C}_k| \le 4^k$ terms.

Proof of (3.4.5). We will show that Theorem 3.1.8 can be applied to the sequence $\{A_N^{(1)}\}_{N\ge1}$. First we prove that

$$\mu_{A_N^{(1)}/\sqrt{N}} \Rightarrow \mu \text{ in probability as } N \to \infty. \tag{3.4.7}$$

As remarked after relation (3.1.8), $\mu_{A_N/\sqrt{N}} \Rightarrow \mu$ in probability as $N \to \infty$. Then, from a well known inequality (Corollary A.41 in [52]), the Levy distance between $\mu_{A_N/\sqrt{N}}$ and $\mu_{A_N^{(1)}/\sqrt{N}}$ is bounded as follows.

$$L^3(\mu_{A_N/\sqrt{N}}, \mu_{A_N^{(1)}/\sqrt{N}}) \le \frac{1}{N}\text{tr}\left\{\left(\frac{1}{\sqrt{N}}A_N - \frac{1}{\sqrt{N}}A_N^{(1)}\right)^2\right\} = \frac{1}{N^2}\sum_{i,j\in[N]}\{(A_N^{(2)})_{i,j}\}^2. \tag{3.4.8}$$

The expectation of the rightmost quantity is at most $N^{-2}N2d_N$ (since each row of $A_N^{(2)}$ has at most $2d_N$ elements that are not identically zero random variables and these random variables have second moment at most 1), which tends to 0 as $N \to 0$ because of the assumption on d_N.

Then the sequence $\{A_N^{(1)}\}_{N\ge1}$ satisfies:

- Assumption 3.1.2 with the same measure as $\{A_N\}_{N\ge1}$. This follows from Lemma 3.8.2.

Assumption (c) of that lemma is satisfied because of (3.4.7).

• Assumption 3.1.3, this is clear,

• Assumption 3.1.4. Indeed, fix $(i,j) \in [N]^2$. If $(i,j) \in \mathcal{D}_N$, there exists some $m \in [d_N]$ such that

$$\{(i+d_1, j+d_2) : d_1, d_2 \in \{-1, 0, 1\}\} \subseteq \mathcal{B}_m^{(N)}.$$

Then Assumption (3.1.18) implies that there exists some $f \in [d_{2N}]$ such that

$$\{(2i+d_1, 2j+d_2) : d_1, d_2 \in \{-2, 0, 2\}\} \subseteq \mathcal{B}_f^{(2N)}.$$

But since $\mathcal{B}_f^{(2N)}$ is axially convex (see before Definition 3.1.9), one can conclude that

$$\{(2i+d_1, 2j+d_2) : d_1, d_2 \in \{-2, -1, 0, 1, 2\}\} \subseteq \mathcal{B}_f^{(2N)}.$$

Now since $(k, \ell) \mapsto s_{k,\ell}^{(2N)}$ is constant in $\mathcal{B}_f^{(2N)}$ [see (3.1.19)] and we assumed (3.1.20), our claim follows.

Thus, all the Assumptions of Theorem 3.1.8 hold for $A_N^{(1)}$, and hence (3.4.5) holds.

Almost sure convergence under the additional Assumption 3.1.7. Using Lemma 3.2.3, we will prove that

$$\limsup_{N\to\infty} \frac{|A_N^{(1)}|_{\mathrm{op}}}{\sqrt{N}} \le \mu_\infty \quad \text{a.s.} \tag{3.4.9}$$

$$\limsup_{N\to\infty} \frac{|A_N^{(2)}|_{\mathrm{op}}}{\sqrt{N}} \le \epsilon \;\; \text{a.s. for any } \epsilon > 0. \tag{3.4.10}$$

And these are enough to prove our claim.

Notice that the validity of Assumptions 3.1.3(a) and 3.1.7 for the sequence $(A_N)_{N\in\mathbb{N}^+}$ implies the validity of the same assumptions for the sequences $(A_N^{(1)})_{N\in\mathbb{N}^+}$ and $(A_N^{(2)})_{N\in\mathbb{N}^+}$. As was mentioned above, the sequence $\{A_N^{(1)}\}_{N\ge 1}$ satisfies Assumption 3.1.2 with the same measure as $\{A_N\}_{N\ge 1}$. And then Lemma 3.3.1 implies that the sequence $(A_N^{(1)})_{N\in\mathbb{N}^+}$ satisfies condition $\Sigma(\mu_\infty)$, while (3.4.6) and $\lim_{N\to\infty} d_N/n = 0$ imply that, for any $\epsilon > 0$, the sequence $(A_N^{(2)})_{N\in\mathbb{N}^+}$ satisfies condition $\Sigma(\epsilon)$. Thus, Lemma 3.2.3 applies and gives the desired inequalities.

3.5 Step function profile. Proof of Theorem 3.1.14

Proof of Theorem 3.1.14. The inequality $\liminf_{N\to\infty} |A_N|_{\mathrm{op}}/\sqrt{N} \ge \mu_\infty^\sigma$ almost surely is justified with the same argument as (3.3.1) with the only difference that here we have $\mu_{A_N/\sqrt{N}} \Rightarrow \mu^\sigma$ a.s., and so the inequality will be true in the a.s. sense.

For the reverse inequality, we will apply Lemma 3.2.3. To check Assumptions 3.1.3(a) and 3.1.7, required by that lemma, note that the (i,j) element of A_N is of the form $\sigma_{p,q} X_0'$ for a constant $\sigma_{p,q} \in [0,1]$ and $X_0' \overset{d}{=} X_0$, and clearly X_0 can play the role of X in relation

(3.1.16). We will prove that $(A_N)_{N\in\mathbb{N}^+}$ satisfies condition $\Sigma(c)$ for all $c > \mu_\infty$, and this will finish the proof.

Define the matrix $\hat{\Sigma}_N \in \mathbb{R}^{N\times N}$ by

$$(\hat{\Sigma}_N)_{i,j} = \sigma_{p,q} \qquad \text{if} \quad \begin{array}{l} a_{p-1} + (1/N) \le i/N < a_p \text{ and} \\ a_{q-1} + (1/N) \le j/N < a_q, \end{array} \tag{3.5.1}$$

and $\hat{A}_N := \hat{\Sigma}_N \odot A'_N$.

Also, let

$$\varepsilon_N := \max\left\{\left|\frac{R_p^{(N)}}{N} - a_p\right| : p \in [m-1]\right\}. \tag{3.5.2}$$

By Definition 3.1.11, it holds $\lim_{N\to\infty} \varepsilon_N = 0$.

Claim 1: a) With probability one, $\hat{A}_N/\sqrt{N}$ has the same limiting ESD as $A_N/\sqrt{N}$.

b) For $\hat{A}_N$, Lemma 3.3.1 applies.

Consequently,

$$\hat{M}_N(k) \le N^{k+1}(\mu_\infty^\sigma)^{2k} \tag{3.5.3}$$

for all $1 \le k < N$.

Proof of Claim 1:

a) This is true because by Theorem A.43 in the Appendix A of [52], the Kolmogorov distance between $\mu_{A_N/\sqrt{N}}$ and $\mu_{\hat{A}_N/\sqrt{N}}$ is at most

$$\frac{1}{N}\operatorname{rank}(A_N - \hat{A}_N) \le \frac{m}{N}\max_{p\in[m]}(\max\{R_p, Na_p\} - \min\{R_p, Na_p\}) = m\varepsilon_N \overset{N\to\infty}{\to} 0. \tag{3.5.4}$$

b) Assumption 3.1.3 is satisfied because $\mathbf{E}(|X_0|^{4+\delta})$ and $\sigma_{p,q} \le 1$ for all $p, q \in [m]$. To show that Assumption 3.1.2 is satisfied, we repeat the argument just before the statement of the Theorem. For the sequence $(\hat{A}_N)_{N\in\mathbb{N}^+}$, the corresponding $W_N(x,y)$, as $N \to \infty$, converges to $\sigma^2(x,y)$ for almost all $(x,y) \in [0,1]^2$. Assumption 3.1.4 is satisfied because if for some i,j we have $\operatorname{Var}[(\hat{A}_N)_{i,j}] > 0$, then this equals $\sigma_{p,q}^2$ for the unique p,q as in (3.5.1). Then

$$\frac{2i-1}{2N} \in [a_{p-1} + \frac{1}{2N}, a_p - \frac{1}{2N}), \frac{2i}{2N} \in [a_{p-1} + \frac{1}{N}, a_p), \tag{3.5.5}$$
$$\frac{2j-1}{2N} \in [a_{q-1} + \frac{1}{2N}, a_q - \frac{1}{2N}), \frac{2j}{2N} \in [a_{q-1} + \frac{1}{N}, a_q). \tag{3.5.6}$$

Thus, (3.1.11) holds as equality.

Claim 2: There is $\partial \in (0,\infty)$ so that $M_N(k) \le e^{\partial(k+1)\varepsilon_N}\hat{M}_N(k)$ for all $k < N$.

Proof of Claim 2: Define the following sets of indices.

$$\Delta_p^{(N)} := I_p^{(N)} \cap [a_{p-1}N + 1, a_pN), \tag{3.5.7}$$
$$\Delta^{(N)} := \cup_{p=1}^m \Delta^{(N)}. \tag{3.5.8}$$

When $p = m$, the interval in the intersection becomes closed on the right also. Then

$$M_N(k) \leq \hat{M}_N(k) + \sum_{\emptyset \neq J \subset [k+1]} \sum_{T \in \mathbf{C}_k} \sum_{i_1 \cdots i_{k+1}} \mathbf{1}\left(i_l \notin \Delta^{(N)} \text{ if } l \in J \text{ and } i_l \in \Delta^{(N)} \text{ if } l \notin J\right) \prod_{\{i,j\} \in E(T)} s_{i,j}^{(N)} \tag{3.5.9}$$

$$= \hat{M}_N(k) + \sum_{\emptyset \neq J \subset [k+1]} \sum_{T \in \mathbf{C}_k} \sum_{m_1, \ldots, m_{k+1} \in [m]^{k+1}} a(T, J, m_1, m_1, \ldots, m_{k+1}) \tag{3.5.10}$$

where

$$a(T, J, m_1 \cdots m_{k+1}) := \sum_{i_1 \in I_{m_1}^{(N)}, \cdots i_{k+1} \in I_{m_{k+1}}^{(N)}} \mathbf{1}(i_l \notin \Delta^{(N)} \text{ if } l \in J, i_l \in \Delta^{(N)} \text{ if } l \notin J, (i_\ell)_{\ell \in [k+1]} \text{distinct}) \prod_{\{i,j\} \in E(T)} s_{i,j}^{(N)}. \tag{3.5.11}$$

Note that

$$\sum_{T \in \mathbf{C}_k} \sum_{m_1, m_2 \cdots m_{k+1} \in [m]^{k+1}} a(T, \emptyset, m_1 \cdots m_{k+1}) \leq \hat{M}_N(k). \tag{3.5.12}$$

We will show that for some constant $\eth = \eth(I_1, I_2, \ldots, I_m) \in (0, \infty)$ we have

$$a(T, J, m_1 \cdots m_{k+1}) \leq (\eth \varepsilon_N)^{|J|} a(T, \emptyset, m_1, \cdots m_{k+1}). \tag{3.5.13}$$

In the definition of $a(T, J, m_1, m_2, \ldots, m_{k+1})$, the product is common to all summands [recall the rectangles of constancy of the map $(i,j) \mapsto s_{i,j}^{(N)}$]. We write $a(T, J, m_1 \cdots m_{k+1})$ and $a(T, \emptyset, m_1, \ldots m_{k+1})$ as

$$\sum_{i_\ell \in I_{m_\ell}^{(N)} \text{ for } \ell \notin J} \mathbf{1}(i_l \in \Delta^{(N)} \text{ for } l \notin J, (i_\ell)_{\ell \in [k+1] \setminus J} \text{ distinct}) \sum_{i_\ell \in I_{m_\ell}^{(N)} \text{ for all } \ell \in J} \mathbf{1}(i_l \notin \Delta^{(N)} \text{ for all } l \in J, (i_\ell)_{\ell \in [k+1]} \text{distinct}) \prod_{\{i,j\} \in E(T)} \tag{3.5.14}$$

$$\sum_{i_\ell \in I_{m_\ell}^{(N)} \text{ for } \ell \notin J} \mathbf{1}(i_l \in \Delta^{(N)} \text{ for } l \notin J, (i_\ell)_{\ell \in [k+1] \setminus J} \text{ distinct}) \sum_{i_\ell \in I_{m_\ell}^{(N)} \text{ for all } \ell \in J} \mathbf{1}(i_l \in \Delta^{(N)} \text{ for all } l \in J, (i_\ell)_{\ell \in [k+1]} \text{distinct}) \prod_{\{i,j\} \in E(T)} \tag{3.5.15}$$

We will compare the inner sums in the two expressions. Notice that there are $C_1, C_2 > 0$ that depend on $a_1, a_2, \ldots, a_m$ only so that

$$|\Delta_p^{(N)}| \geq C_1 N, \tag{3.5.16}$$

$$|I_p^{(N)} \setminus \Delta^{(N)}| \leq C_2 \varepsilon_N N \tag{3.5.17}$$

for all $p \in [m]$. For each fixed collection $(i_\ell)_{\ell \notin J}$, the inner sum in (3.5.14) is at most $(C_2 \varepsilon_N N)^{|J|}$ while the inner sum in (3.5.15) is at least $(C_1 N/2)^{|J|}$. The ratio of the first over the second bound is $(2C_2 \varepsilon_N / C_1)^{|J|}$. Thus, we get (3.5.13) with $\eth := 2C_2/C_1$.

Taking into account (3.5.13) and (3.5.12), we get that the second summand in (3.5.10) is bounded above by

$$\sum_{t=1}^{k+1} \sum_{J\subset[k+1]:|J|=t} \hat{M}_N(k)(\partial\varepsilon_N)^t = M_N^{(1)}(k) \sum_{t=1}^{k+1} \binom{k+1}{t} (\partial\varepsilon_N)^t = \hat{M}_N(k) \left\{(1+\partial\varepsilon_N)^{k+1} - 1\right\} \tag{3.5.18}$$

Consequently, $M_N(k) \le (1+\partial\varepsilon_N)^{k+1}\hat{M}_N(k) \le e^{\partial(k+1)\varepsilon_N}\hat{M}_N(k)$, and this proves Claim 2.

Now, combining this with (3.5.3), we get that condition $\Sigma((1+\varepsilon)\mu_\infty)$ is satisfied for each $\varepsilon > 0$. □

3.6 An approximation result and proof of Corollary 3.1.15

Proposition 3.6.1. *Let $(A_N)_{N\in\mathbb{N}^+}$ be a sequence of symmetric random matrices, A_N of dimension $N \times N$, of the form*

$$A_N = \Sigma_N \odot A'_N, \tag{3.6.1}$$

where $\Sigma_N \in [0,\infty)^{N\times N}$ and A'_N is a random $N \times N$ symmetric matrix with independent entries (up to symmetry) all with zero mean and unit variance.

For every $n \in \mathbb{N}^+$ consider a sequence $(\Sigma_N^{(n)})_{N\in\mathbb{N}^+}$ of matrices, with $\Sigma_N^{(n)} \in [0,\infty)^{N\times N}$, and define

$$A_N^{(n)} := \Sigma_N^{(n)} \odot A'_N \quad \textit{for each } N \in \mathbb{N}^+. \tag{3.6.2}$$

(a) Assume that

(i) the sequence $(A'_N)_{N\in\mathbb{N}^+}$ satisfies Assumption 3.1.3,

(ii) for each $n \in \mathbb{N}^+$ it holds

$$\lim_{N\to\infty} \frac{|A_N^{(n)}|_{\mathrm{op}}}{\sqrt{N}} = \mu_\infty^{(n)} \textit{ in probability}, \tag{3.6.3}$$

where $\mu_\infty^{(n)}$ is a finite constant,

(iii)

$$\lim_{n\to\infty} \mu_\infty^{(n)} =: \mu_\infty \in \mathbb{R}, \tag{3.6.4}$$

(iv)

$$\lim_{n\to\infty} \limsup_{N\to\infty} |\Sigma_N - \Sigma_N^{(n)}|_{\max} = 0. \tag{3.6.5}$$

Then

$$\lim_{N\to\infty} \frac{|A_N|_{\mathrm{op}}}{\sqrt{N}} = \mu_\infty \qquad \textit{in probability}. \tag{3.6.6}$$

(b) Assume that, in addition to the assumptions of (a), the convergence in (3.6.3) holds in the a.s. sense and Assumption 3.1.7 holds for the sequence $(A'_N)_{N\in\mathbb{N}^+}$. Then the limit in (3.6.6) holds in the a.s. sense.

Proof. (a) Fix $\epsilon \in (0, 1/2)$ and n_0 large enough such that for every $n \geq n_0$ it is true that

$$|\mu_\infty - \mu_\infty^{(n)}| \leq \epsilon \text{ and } \limsup_{N\to\infty} |\Sigma_N - \Sigma_N^{(n)}|_{\max} < \epsilon.$$

Fix an $n \geq n_0$. There is an $N_0 = N_0(n) \in \mathbb{N}^+$ so that $|\Sigma_N - \Sigma_N^{(n)}|_{\max} < \epsilon^2$ for all $N \geq N_0$. Then for $N \geq N_0$ we have

$$\mathbf{P}\left(\left|\frac{|A_N|_{\text{op}}}{\sqrt{N}} - \mu_\infty\right| \geq 5\epsilon\right) \leq \mathbf{P}\left(\left|\frac{|A_N|_{\text{op}}}{\sqrt{N}} - \mu_\infty^{(n)}\right| \geq 4\epsilon\right) \leq \mathbf{P}\left(\frac{|A_N - A_N^{(n)}|_{\text{op}}}{\sqrt{N}} \geq 3\epsilon\right) + \mathbf{P}\left(\left|\frac{|A_N^{(n)}|_{\text{op}}}{\sqrt{N}} - \mu_\infty^{(n)}\right| \geq \epsilon\right). \tag{3.6.7}$$

The last term in (3.6.7) converges to zero as $N \to \infty$ due to (3.6.3). For the previous term we will apply Proposition 3.2.2. Notice that the sequence $(A_N - A_N^{(n)})_{N\in\mathbb{N}^+}$ satisfies

- Assumption 3.1.3 because $(A_N - A_N^{(n)})_{i,j} = ((\Sigma_N)_{i,j} - (\Sigma_N^{(n)})_{i,j})(A'_N)_{i,j}$ and $|(A_N - A_N^{(n)})_{i,j}| \leq |(A'_N)_{i,j}|$ (for all $N \in \mathbb{N}^+, i, j \in [N]$) and we assumed that $(A'_N)_{N\in\mathbb{N}^+}$ satisfies Assumption 3.1.3

- condition $\Sigma(2\epsilon)$ because if, for $t \in \mathbb{N}^+$ with $t < N$, we call $M'_N(t)$ the quantity defined in (3.1.7) for the matrix $A_N - A_N^{(n)}$, and note that the (i, j) element of $A_N - A_N^{(n)}$ has mean zero and variance $\{(\Sigma_N)_{i,j} - (\Sigma_N^{(n)})_{i,j}\}^2$, we obtain that

$$M'_N(t) \leq N^{t+1} 2^{2t} (|\Sigma_N - \Sigma_N^{(n)}|_{\max})^{2t} < N^{t+1} (2\epsilon)^{2t}. \tag{3.6.8}$$

Since $3\epsilon > 2\epsilon(1 + \epsilon)$, Proposition 3.2.2 implies that the penultimate term in (3.6.7) goes to zero as $N \to \infty$.

(b) It is enough to prove that with probability 1 it holds $\overline{\lim}_{N\to\infty} \frac{|A_N|_{\text{op}}}{\sqrt{N}} \leq \mu_\infty$. Because of (3.6.3) (holding a.s.) and (3.6.4), it is enough to prove that for all $\epsilon > 0$ and all n large enough, with probability 1, it holds

$$\overline{\lim_{N\to\infty}} \frac{|A_N - A_N^{(n)}|_{\text{op}}}{\sqrt{N}} \leq 2\epsilon. \tag{3.6.9}$$

To prove this, we will apply Lemma 3.2.3. Take n_0 so that for all $n \geq n_0$ it holds $\limsup_{N\to\infty} |\Sigma_N - \Sigma_N^{(n)}|_{\max} < \epsilon^2$. Now fix $n \geq n_0$. There is $N_0 = N_0(n) \in \mathbb{N}^+$ so that $|\Sigma_N - \Sigma_N^{(n)}|_{\max} < \epsilon^2$ for all $N \geq N_0$. Then the sequence $(A_N - A_N^{(n)})_{N\geq N_0}$ satisfies Assumption 3.1.3(a) as we saw in part a) of the proposition, Assumption 3.1.7 (because A'_N does so and $|\Sigma_N - \Sigma_N^{(n)}|_{\max} < 1$), and assumption $\Sigma(2\epsilon)$ because of (3.6.8). Then Lemma 3.2.3 gives the desired inequality. □

Proof of Corollary 3.1.15. We will apply Proposition 3.6.1(b) for the sequence $(A_N)_{N\in\mathbb{N}^+}$. The sequence $(A'_N)_{N\in\mathbb{N}^+}$ mentioned in that Proposition is exactly the sequence $(A'_N)_{N\in\mathbb{N}^+}$ of relation (3.1.27) and it satisfies Assumption 3.1.3 because for it the discussion following Assumption 3.1.3 applies (X_0 has finite fourth moment).

For each $n \in \mathbb{N}^+$, we define the following obvious approximation to σ.

$$\sigma^{(n)}(x, y) := n^2 \int_{I_k} \int_{I_\ell} \sigma(a, b)\, da\, db \text{ if } (x, y) \in I_k \times I_\ell \text{ for } k, \ell \in [n], \tag{3.6.10}$$

where $I_k := [\frac{k-1}{n}, \frac{k}{n})$ for $k \in [n-1]$ and $I_n := [(n-1)/n, 1]$. Then, we define the matrices $\Sigma_N^{(n)}$ through the relation $(\Sigma_N^{(n)})_{i,j} := \sigma^{(n)}(i/N, j/N)$.

For each $n \in \mathbb{N}^+$, the sequence of matrices $(\Sigma_N^{(n)} \odot A'_N)_{N\geq 1}$ satisfies the assumptions of Theorem 3.1.14. Consequently, as $N \to \infty$, the sequence $(\mu_{A_N^{(n)}/\sqrt{N}})_{N\in\mathbb{N}^+}$ converges almost surely weakly to a symmetric measure, say $\mu^{(n)}$, with support contained in $[-\mu_\infty^{(n)}, \mu_\infty^{(n)}]$ and (3.6.3) holds in the a.s. sense. In a claim below we prove that condition (3.6.4) is satisfied. Finally, to check (3.6.5), note that

$$\left|(\Sigma_N)_{i,j} - (\Sigma_N^{(n)})_{i,j}\right| \leq \left|(\Sigma_N)_{i,j} - \sigma(i/N, j/N)\right| + \left|\sigma(i/N, j/N) - \sigma^{(n)}(i/N, j/N)\right|. \tag{3.6.11}$$

In the right hand side of the last inequality, the first term converges to zero as $N \to \infty$ due to (3.1.26), and the second term is at most the supremum norm of $\sigma - \sigma^{(n)}$, which goes to zero as $n \to \infty$ because σ is uniformly continuous in $[0, 1]^2$. Thus, Proposition 3.6.1(b) applies and completes the proof.

Claim: Condition (3.6.4) is satisfied.

We modify the proof of Lemma 6.4 of [12]. Call μ the weak limit as $N \to \infty$ of $\mu_{A_N/\sqrt{N}}$, then $F_N, F_N^{(n)}$ the distribution function of $\mu_{A_N/\sqrt{N}}$ and $\mu_{A_N^{(n)}/\sqrt{N}}$ respectively, and $F, F^{(n)}$ the distribution function of $\mu, \mu^{(n)}$ respectively. Let

$$\begin{gathered} \lambda_{N,1} \leq \lambda_{N,2} \leq \cdots \leq \lambda_{N,N}, \\ \lambda_{N,1}^{(n)} \leq \lambda_{N,2}^{(n)} \leq \cdots \leq \lambda_{N,N}^{(n)} \end{gathered}$$

the eigenvalues of $A_N/\sqrt{N}, A_N^{(n)}/\sqrt{N}$ respectively.

Let $\epsilon \in (0, 1/2)$. There is $n_0 = n_0(\epsilon)$ so that for all $n \geq n_0$ it holds $\limsup_{N\to\infty} |\Sigma_N - \Sigma_N^{(n)}|_{\max} < \epsilon^2$. Take now $n \geq n_0$ fixed. There is $N_0 = N_0(n) \in \mathbb{N}^+$ so that $|\Sigma_N - \Sigma_N^{(n)}|_{\max} < \epsilon^2$ for all $N \geq N_0$. As explained in the proof of Proposition 3.6.1, $\lim_{N\to\infty} \mathbf{P}\left(|A_N - A_N^{(n)}|_{\mathrm{op}} \geq 3\epsilon\sqrt{N}\right) = 0$. There is sequence $(N_k)_{k\geq 1}$ so that in a set Ω_ϵ of probability 1, eventually for all k we have $|A_{N_k} - A_{N_k}^{(n)}|_{\mathrm{op}} < 3\epsilon\sqrt{N_k}$. Since

$$\max_{i\in[N]} |\lambda_{N_k,i}^{(n)} - \lambda_{N_k,i}| \leq |A_{N_k} - A_{N_k}^{(n)}|_{\mathrm{op}}/\sqrt{N_k},$$

in Ω_ϵ (the inequality is true by Theorem A46 in [52]), we will have eventually for all $k \in \mathbb{N}^+$ that

$$F_{N_k}^{(n)}(a - 3\epsilon) \leq F_{N_k}(a) \leq F_{N_k}^{(n)}(a + 3\epsilon) \tag{3.6.12}$$

for all $a \in \mathbb{R}$. From here, using the convergence as $N \to \infty$ of F_N to F and of $F_N^{(n)}$ to $F^{(n)}$, we have that for all $a \in \mathbb{R}$ it holds

$$F^{(n)}(a - 3\epsilon) \leq F(a) \leq F^{(n)}(a + 3\epsilon). \tag{3.6.13}$$

[First we get this for all a outside a countable subset of $\mathbb{R}$ and then using the right continuity of $F, F^{(n)}$ we get it for all $a \in \mathbb{R}$.] This implies that $|\mu_\infty^{(n)} - \mu_\infty| \le 3\epsilon$ and finishes the proof of the claim. □

Remark 3.6.2. The above proof easily generalizes to the case that the function σ is piecewise continuous in the following sense. There are $m \in \mathbb{N}^+$, $0 = a_0 < a_1 < \cdots < a_{m-1} < a_m = 1$ so that letting $I_p := [a_{p-1}, a_p)$ for $p = 1, 2, \dots, m-1$, and $I_m = [a_{m-1}, 1]$ the function $\sigma | I_p \times I_q$ is uniformly continuous for all $p, q \in [m]$ (i.e., when σ extends continuously in the closure of each rectangle $I_p \times I_q$. Recall that to handle the last term in (3.6.11) all we needed was the uniform continuity of σ.

3.7 Examples

3.7.1 Random Gram matrices

Let $(X_N)_{N \in \mathbb{N}^+}$ be a sequence of matrices so that X_N is an $M(N) \times N$ matrix with independent, centered entries with unit variance, and $M : \mathbb{N}^+ \to \mathbb{N}^+$ a function with $\lim_{N\to\infty} \frac{M(N)}{N} = c \in (0, \infty)$. It is known that the empirical spectral distribution of XX^T, after rescaling, converges to the Marchenko-Pastur law μ_{MP} [9]. Moreover, the convergence of the rescaled largest eigenvalue to the largest element of the support of μ_{MP} has been established in [53] under the assumption of finite fourth moment for the entries. However, some applications in wireless communication require understanding the spectrum of XX^T, where X has a variance profile, see for example [54] or [55]. Such matrices are called random Gram matrices. In this subsection, we establish the convergence of the largest eigenvalue of random Gram matrices to the largest element of the support of its limiting empirical spectral distribution for specific variance profiles. Firstly we give some definitions.

Definition 3.7.1 (Step function variance profile)**.** Consider

a) $m, n \in \mathbb{N}^+$ and numbers $\{\sigma_{p,q}\}_{p\in[m], q\in[n]} \in [0, \infty)^{mn}$.

b) For each $K \in \mathbb{N}^+$, two partitions $\{I_p^{(K)}\}_{p\in[m]}, \{J_p^{(K)}\}_{p\in[n]}$ of $[K]$ in m and n intervals respectively. The numbering of the intervals is such that $x < y$ whenever $x \in I_p^{(K)}, y \in I_q^{(K)}$ or $x \in J_p^{(K)}, y \in J_q^{(K)}$ with $p < q$. Let $LI_p^{(K)}$ and $RI_p^{(K)}$ be the left and right endpoint respectively of $I_p^{(K)}$ and similarly $LJ_p^{(K)}$ and $RJ_p^{(K)}$ for $J_p^{(K)}$.

c) Numbers $0 = a_0 < a_1 < \cdots < a_{m-1} < a_m := 1$. We assume that $\lim_{M\to\infty} RI_p^{(M)}/M = a_p$ for each $p \in [m]$.

d) Numbers $0 = \beta_0 < \beta_1 < \cdots < \beta_{n-1} < \beta_n := 1$. We assume that $\lim_{N\to\infty} RJ_q^{(N)}/N = \beta_q$ for each $q \in [n]$.

e) $M : \mathbb{N}^+ \to \mathbb{N}^+$ a function,

f) A random variable X_0 with $\mathbf{E}(X_0) = 0, \mathbf{E}(X_0^2) = 1$.

For each $M, N \in \mathbb{N}^+$, define the matrix $\Sigma_{M,N} \in \mathbb{R}^{M\times N}$ by $(\Sigma_{M,N})_{i,j} = \sigma_{p,q}$ if $i \in I_p^{(M)}, j \in I_q^{(N)}$, and let $\{A_N\}_{N\in\mathbb{N}^+}$ be the sequence of random matrices defined by

$$A_N = \Sigma_{M(N),N} \odot A'_{M(N),N} \tag{3.7.1}$$

where $A'_{M(N),N}$ is an $M(N) \times N$ matrix whose elements are independent random variables all with distribution the same as X_0. We say that A_N in (3.7.1) is a *random matrix model whose variance profile is given by a step function.*

Definition 3.7.2 (Continuous function variance profile)**.** For

a) a continuous function $\sigma : [0, 1]^2 \to [0, 1]$,

b) $M : \mathbb{N}^+ \to \mathbb{N}^+$ a function

c) a sequence $(\Sigma_{M(N),N})_{N\in\mathbb{N}^+}$ of matrices, $\Sigma_{M(N),N} \in [0, 1]^{M(N)\times N}$, with the property

$$\lim_{N\to\infty} \sup_{i\in[M(N)],j\in[N]} \left|(\Sigma_{M(N),N})_{i,j} - \sigma(i/M(N), j/N)\right| = 0, \tag{3.7.2}$$

d) a random variable X_0 with $\mathbf{E}(X_0) = 0, \mathbf{E}(X_0^2) = 1$,

consider the sequence $\{A_N\}_{N\in\mathbb{N}^+}$ of random matrices, $A_N \in \mathbb{R}^{M(N)\times N}$, defined by

$$A_N = \Sigma_{M(N),N} \odot A'_N \tag{3.7.3}$$

where the entries of A'_N are independent random variables all with distribution the same as X_0. Then we say that $(A_N)_{N\in\mathbb{N}^+}$ is a *random matrix model whose variance profile is given by a continuous function.* Again, we call σ the variance profile.

Symmetrization To study the eigenvalues of $A_N A_N^T$, where A_N falls in one of the cases of the two last definitions, we use the trick of symmetrization. If A is an $M \times N$ matrix, where $M, N \in \mathbb{N}^+$, we call symmetrization of A the $(M + N) \times (M + N)$ symmetric matrix $\tilde{A}$ defined by

$$\tilde{A} := \begin{bmatrix} \mathbf{O}_{M,M} & A \\ A^T & \mathbf{O}_{N,N} \end{bmatrix} \tag{3.7.4}$$

where, for any $k, l \in \mathbb{N}^+$, $\mathbf{O}_{k,l}$ denotes the $k \times l$ matrix with all of its entries equal to 0. The characteristic polynomials of $AA^T, \tilde{A}$ are connected through the relation

$$\lambda^M \det(\lambda \mathbb{I}_{M+N} - \tilde{A}) = \lambda^N \det(\lambda^2 \mathbb{I}_M - AA^T) \tag{3.7.5}$$

for all $\lambda \in \mathbb{C}$. Thus, in the case $M \leq N$, if we call $(t_1, t_2, \ldots, t_{M+N})$ the eigenvalues of the symmetric matrix $\tilde{A}$, then the vector $(t_1^2, t_2^2, \ldots, t_{M+N}^2)$ contains twice each eigenvalue of

AA^T and $N - M$ times the eigenvalue 0 (multiple eigenvalues appear in the previous vectors according to their multiplicities). Thus, the empirical spectral distributions of $AA^T, \tilde{A}$ are related through

$$\mu_{\tilde{A}} \circ T^{-1} = \frac{2M}{M+N}\mu_{AA^T} + \frac{N-M}{M+N}\delta_0 \tag{3.7.6}$$

with $T : \mathbb{R} \to [0, \infty)$ being the map $x \mapsto x^2$.

Step function profile: If $(A_N)_{N \in \mathbb{N}^+}$ is as in Definition 3.7.1 with $M(N) := \lceil cN \rceil$ for some $c \in (0, 1]$, then the sequence $(\tilde{A}_N)_{N \in \mathbb{N}^+}$ is of the form given in Definition 3.1.11 with the following modification. We require that there is some $\gamma : \mathbb{N}^+ \to \mathbb{N}^+$ with $\lim_{N\to\infty} \gamma(N) = \infty$ so that the N-th matrix is of dimension $\gamma(N) \times \gamma(N)$ and, for each $N \in \mathbb{N}^+$, the family $(I_p^{(N)})_{p \in [m]}$ is a partition of $[\gamma(N)]$. The numbers a_p satisfy $\lim_{N\to\infty} R_p^{(N)}/\gamma(N) = a_p$. With this modification, Theorem 3.1.14 holds if the denominator in (3.1.30) is replaced by $\sqrt{\gamma(N)}$.

The sequence $(\tilde{A}_N)_{N \in \mathbb{N}^+}$ fits into this framework. We have $\gamma(N) = \lceil cN \rceil + N$, the role of m (of Definition 3.1.11) is played by $m + n$ (m, n from Definition 3.7.1), the $(m + n)^2$ constants are

$$\tilde{\sigma}_{p,q} := \begin{cases} 0 & \text{if } p \in [m], q \in [m], \\ \sigma_{p,q-m} & \text{if } p \in [m], q \in [m+n]\backslash[m], \\ \sigma_{q,p-m} & \text{if } p \in [m+n]\backslash[m], q \in [m], \\ 0 & \text{if } p \in [m+n]\backslash[m], q \in [m+n]\backslash[m] \end{cases} \tag{3.7.7}$$

for each N, and the partition of $[\gamma(N)]$ into $m+n$ intervals consists of the intervals (we write M instead of $\lceil cN \rceil$)

$$\Big\{[Ma_{p-1}, Ma_p) \cap \mathbb{N}^+ : p \in [m]\Big\}, \tag{3.7.8}$$

$$\Big\{[M + N\beta_{q-1}, M + N\beta_q) \cap \mathbb{N}^+ : q \in [n]\Big\}. \tag{3.7.9}$$

Dividing the right endpoints of the intervals by $\gamma(N)$ and taking $N \to \infty$, we get the $m + n$ numbers

$$\frac{c}{1+c}a_1 < \frac{c}{1+c}a_2 < \cdots < \frac{c}{1+c}a_m < \frac{c}{1+c} + \frac{1}{1+c}\beta_1 < \cdots < \frac{c}{1+c} + \frac{1}{1+c}\beta_n. \tag{3.7.10}$$

If we feed these data to the recipe of Definition 3.1.11, relation (3.1.24) will give as A_N the matrix $\tilde{A}_N$ where A_N is given by (3.7.1). The discussion preceding Theorem 3.1.14 applied to the sequence $(\tilde{A}_N)_{N\ge1}$ gives that the ESD of $\tilde{A}_N/\sqrt{\gamma(N)}$ converges almost surely weakly to a symmetric probability measure $\tilde{\mu}^\sigma$ with compact support. Call $\tilde{\mu}^\sigma_\infty$ the largest element of the support. Relation (3.7.6) implies that the ESD of $A_N A_N^T/N$ converges to a measure with compact support contained in $[0, \infty)$ and the largest element of this support is $\mu_\infty = (1 + c)(\tilde{\mu}^\sigma_\infty)^2$. Then Theorem 3.1.14 has the following corollary.

Corollary 3.7.3. *Assume that* $(A_N)_{N\geq 1}$ *is as in Definition 3.7.1 with* $M := \lceil cN \rceil$ *for some* $c \in (0,1]$ *and* $\mathbf{E}(|X_0|^{4+\delta}) < \infty$ *for some* $\delta > 0$*. Then it is true that*

$$\lim_{N\to\infty} \frac{\|A_N A_N^T\|_{\mathrm{op}}}{N} = \mu_\infty \quad a.s. \tag{3.7.11}$$

CONTINUOUS FUNCTION PROFILE: If $(A_N)_{N\in\mathbb{N}^+}$ is as in Definition 3.7.2 with $M(N) := \lceil cN \rceil$ for some $c \in (0,1]$, then we apply the discussion preceding Theorem 3.1.14 to the sequence $(\tilde{A}_N)_{N\in\mathbb{N}^+}$. The graphon, W_N, corresponding to $\tilde{A}_N$ converges pointwise in $[0,1]^2$ to the graphon $\tilde{\sigma}$ with

$$\tilde{\sigma}(x,y) := \begin{cases} 0 & \text{if } (x,y) \in [0, c/(1+c)]^2 \cup (c/(1+c), 1]^2, \\ \sigma(x(1+c)/c, (1+c)y - c) & \text{if } (x,y) \in [0, c/(1+c)] \times (c/(1+c), 1], \\ \sigma(y(1+c)/c, (1+c)x - c) & \text{if } (x,y) \in (c/(1+c), 1] \times [0, x/(1+c)]. \end{cases} \tag{3.7.12}$$

We used (3.7.2) and the continuity of σ. Since $(\tilde{A}_N)_{N\in\mathbb{N}^+}$ also satisfies Assumption 3.1.1, we get that the ESD of $\tilde{A}_N/\sqrt{\gamma(N)}$ converges almost surely weakly to a symmetric probability measure $\tilde{\mu}^\sigma$ with compact support. Call $\tilde{\mu}^\sigma_\infty$ the largest element of the support. As above, the ESD of $A_N A_N^T/N$ converges to a measure with compact support contained in $[0,\infty)$, and the largest element of this support is $\mu_\infty = (1+c)(\tilde{\mu}^\sigma_\infty)^2$.

Corollary 3.7.4. *Assume that* $(A_N)_{N\geq 1}$ *is as in Definition 3.7.2 with* $M := \lceil cN \rceil$ *for some* $c \in (0,1]$ *and* $\mathbf{E}(|X_0|^{4+\delta}) < \infty$ *for some* $\delta > 0$*. Then it is true that*

$$\lim_{N\to\infty} \frac{\|A_N A_N^T\|_{\mathrm{op}}}{N} = \mu_\infty \quad a.s. \tag{3.7.13}$$

Proof. The proof does not follow directly from Corollary 3.1.15 because the sequence $(\tilde{A}_N)_{N\in\mathbb{N}^+}$ does not necessarily have a continuous variance profile in the sense of Definition 3.1.12. Instead, we mimic the proof of that corollary. We define $\sigma^{(n)}$ as in (3.6.10), and the $M \times N$ matrix $\Sigma_N^{(n)}$ as $(\Sigma_N^{(n)})_{i,j} := \sigma^{(n)}(i/M, j/N)$ for all $i \in [M], j \in [N]$. Then we apply an obvious modification of Proposition 3.6.1 (the N-th matrix is of dimension $\gamma(N) \times \gamma(N)$, with $\gamma(N) = \lceil cN \rceil + N$) with the role of Σ_N and $\Sigma_N^{(n)}$ played by $\tilde{\Sigma}_{M,N}, \tilde{\Sigma}_N^{(n)}$ (the symmetrizations of $\Sigma_{M,N}$ and $\Sigma_N^{(n)}$, defined in (3.7.4). The proof continues by adopting the proof of Corollary 3.1.15 to this setting. Note that $\|\tilde{\Sigma}_{M,N} - \tilde{\Sigma}_N^{(n)}\|_{\max} = \|\Sigma_{M,N} - \Sigma_N^{(n)}\|_{\max}$, which has $\lim_{n\to\infty} \limsup_{N\to\infty} \|\Sigma_{M,N} - \Sigma_{M,N}^{(n)}\|_{\max} = 0$. □

Remark 3.7.5. In [54] the authors showed that if the variances of the entries of $A_{N,M}$ are given by the values of a continuous function (and some extra assumptions such as bounded $4+\epsilon$ moments of the entries) the limiting distribution of the E.S.D. of $A_N A_N^T$ does exist. So in Theorem 3.7.3 we prove the convergence of the largest eigenvalue of these models as well. The authors in [54] also studied the non-centered version of these models, i.e. when the entries of the matrix do not have 0 mean, but we do not cover this case with our result.

3.7.2 Further applications of Theorem 3.1.10

In the Random Matrix Theory literature what are commonly described as Random matrices with variance-profile given by a step function are more or less what we describe in Theorem 3.1.14. In this subsection we give some examples which are covered by the generalized version of this variance-profile matrices (Definition 3.1.9) but not from the "standard" step functions.

Let $\{a_{i,j}^{(N)} : N \in \mathbb{N}^+, i,j \in [N]\}$ identically distributed random variables, $a_{i,j}^{(N)} = a_{j,i}^{(N)}$ for all $N \in \mathbb{N}^+, i,j \in [N]$, $\{a_{i,j}^{(N)} : 1 \le j \le i \le N\}$ independent for each N, and $a_{1,1}^{(1)}$ has mean 0 and variance 1. Fix $p \in (0,1]$ and let A_N be the matrix with entries

$$\{A_N\}_{i,j} = a_{i,j}^{(N)} 1_{|i-j| \le pN}, \quad i,j \in [N], \tag{3.7.14}$$

The sequence $(A_N)_{N\in\mathbb{N}^+}$ satisfies Assumption 3.1.1 (easy to check) and also Assumption 3.1.2. To see the last point, we follow Remark 3.1.13. The graphon corresponding to A_N is $W_N(x,y) = \mathbf{1}_{|\lceil Nx\rceil - \lceil Ny\rceil| \le pN}$ which converges to the graphon $W(x,y) = \mathbf{1}_{|x-y|\le p}$ at least on the set $\{(x,y) \in [0,1]^2 : |x-y| \neq pn\}$, which has measure 1. Thus, with probability one, the ESD of $A_N/\sqrt{N}$ converges weakly to a symmetric measure μ with compact support. Call μ_∞ the supremum of the support of μ.

Corollary 3.7.6 (Non-Periodic Band Matrices with Bandwidth proportional to the dimension)**.** *Assume that for the matrix defined in* (3.7.14) *we have that* $a_{1,1}^{(N)}$ *has* 0 *mean, unit variance and finite* $4+\delta$ *moment for some* $\delta > 0$*. Then*

$$\lim_{N\to\infty} \frac{|A_N|_{\text{op}}}{\sqrt{N}} = \mu_\infty \quad a.s.$$

Proof. The sequence $(A_N)_{N\in\mathbb{N}^+}$ satisfies Assumption 3.1.2, as we saw above, and also Assumption 3.1.3 because $\{a_{i,j}^{(N)} : N \in \mathbb{N}^+, i,j \in [N]\}$ are identically distributed and $a_{1,1}$ has mean zero, variance one, and finite fourth moment. The corollary then is a straightforward application of Theorem 3.1.10, where $d_N = 3$, the partition of $[N]^2$ required by Definition 3.1.9 consists of the sets

$$\mathcal{B}_1^{(N)} := \{(i,j) \in [N]^2 : |(i/N) - (j/N)| \le p\}, \tag{3.7.15}$$

$$\mathcal{B}_2^{(N)} := \{(i,j) \in [N]^2 : (i/N) > (j/N) + p\}, \tag{3.7.16}$$

$$\mathcal{B}_3^{(N)} := \{(i,j) \in [N]^2 : (j/N) > (i/N) + p\}, \tag{3.7.17}$$

and $s_1^{(N)} = 1, s_2^{(N)} = s_3^{(N)} = 0$. Condition (b) of that definition is satisfied by $f := m$ for each $m \in [3]$. ◻

Remark 3.7.7. The random band matrix models have been extensively studied after the novel work in [56] and have tremendous application in various research areas. When the bandwidth of the matrices is periodic, i.e., the distance from the diagonal outside which the entries are 0 is periodic, the operator norm has been extensively studied, see for example [38] or the survey [57]. Moreover when the bandwidth of such matrices is non-periodic but the bandwidth (the maximum number of non identically zero entries per row) is $o(N)$ but also tends to infinity has also been examined in [58]. To the best of our knowledge, the convergence of the largest eigenvalue of non-periodic Band Matrices with bandwidth proportional to the dimension has not been established.

Our next result concerns the singular values of triangular random matrices. It is well known under various assumptions for the entries, but we record it here as another application of our main theorem.

Let $\{a_{i,j}^{(N)} : N \in \mathbb{N}^+, 1 \le i \le j \in [N]\}$ identically distributed random variables, $\{a_{i,j}^{(N)} : 1 \le j \le i \le N\}$ independent for each N, and $a_{1,1}^{(1)}$ has mean 0 and variance 1. Let A_N be the matrix with entries

$$\{A_N\}_{i,j} = a_{i,j}^{(N)} 1_{i \le j}, \quad i, j \in [N], \tag{3.7.18}$$

Corollary 3.7.8 (Triangular matrices)**.** *Assume that for the matrix defined in* (3.7.18) *we have that* $a_{1,1}^{(N)}$ *has* 0 *mean, unit variance and finite* $4 + \delta$ *moment for some* $\delta > 0$*. Then*

$$\lim_{N\to\infty} \frac{|A_N A_N^T|_{\text{op}}}{N} = e \quad a.s.$$

Proof. As in the case of Gram matrices, we denote by $\tilde{A}_N$ the symmetrization of A_N, defined in (3.7.4). We have $|A_N A_N^T|_{\text{op}} = |\tilde{A}_N|_{\text{op}}^2$. We will apply Theorem 3.1.10 to the sequence $(\tilde{A}_N)_{N\in\mathbb{N}^+}$. The partition of $[2N]^2$ required by Definition 3.1.9 consists of the following three sets (i.e., $d_N = 3$)

$$\mathcal{B}_1^{(N)} := \{(i, j) \in [2N]^2 : |i - j| \le N - 1\}, \tag{3.7.19}$$

$$\mathcal{B}_2^{(N)} := \{(i, j) \in [2N]^2 : i \ge N + j\}, \tag{3.7.20}$$

$$\mathcal{B}_3^{(N)} := \{(i, j) \in [2N]^2 : j \ge N + i\}, \tag{3.7.21}$$

and the corresponding values of the variance are $s_1^{(N)} = 0$, $s_2^{(N)} = s_3^{(N)} = 1$. Assumption 3.1.2 follows as an application of Remark 3.1.13, in the same way as in the previous corollary. The measure μ of that assumption satisfies $\mu \circ T^{-1} = \nu$, where ν is the limit of the E.S.D of $N^{-1} A_N A^T$ [recall (3.7.6)]. It was shown in [15] that ν has support $[0, e]$. It follows that μ has support $[-\sqrt{e}, \sqrt{e}]$ [See Remark 2.2 of [59] for a more detailed discussion of this phenomenon].

Assumption 3.1.3 is satisfied because the elements of $\tilde{A}_N$ with indices in $\mathcal{B}_2^{(N)} \cup \mathcal{B}_3^{(N)}$ are identically distributed with zero mean, unit variance and finite fourth moment (the remaining elements of the matrix are identically zero random variables). Finally, condition (3.1.20) is satisfied as equality.

Thus, the corollary follows from Theorem 3.1.10. □

3.8 Two technical lemmas

In the next lemma, we prove the crucial estimate we invoked in the proof of Proposition 3.2.1. We adopt and present the terminology of Section 5.1.1 of [52].

Lemma 3.8.1. *$N_{T,a_1,a_2,\dots,a_s} \le (4k^4)^{4(s+1-t)+2(k-s)}$ if $t \ge 2$ and $N_{T,a_1,a_2,\dots,a_s} = 1$ if $t = 1$.*

Proof. When $t = 1$, since the cycle is bad, we have $s = 1$ and $a_1 = 2k$, and there is only one cycle with these s, t and vertex set $\{1\}$.

For the case $t \ge 2$, take a cycle $\mathbf{i} := (i_1, i_2, \dots, i_{2k})$ as in (3.2.9) and assume that it has edge multiplicities $a_1, a_2, \dots, a_s \ge 2$. Each step in the cycle we call a *leg*. More formally, legs are the elements of the set $\{(r, (i_r, i_{r+1})) : r = 1, 2, \dots, 2k\}$. Edges of the cycle we call the edges of $G(\mathbf{i})$, and the multiplicity of each edge is computed from $\mathbf{i}$. The graph $G(\mathbf{i})$ does not have multiple edges.

For $1 \le a < b$, we say that the leg $(a, (i_a, i_{a+1}))$ is *single* up to b if $\{i_a, i_{a+1}\} \ne \{i_c, i_{c+1}\}$ for every $c \in \{1, 2, \dots, b-1\}, c \ne a$. We classify the $2k$ legs of the cycle into 4 sets T_1, T_2, T_3, T_4. The leg $(a, (i_a, i_{a+1}))$ belongs to

T_1: if $i_{a+1} \notin \{i_1, \dots, i_a\}$. I. e., the leg leads to a new vertex.

T_3: if there is a T_1 leg $(b, (i_b, i_{b+1}))$ with $b < a$ so that $a = \min\{c > b : \{i_c, i_{c+1}\} = \{i_b, i_{b+1}\}\}$. I. e., at the time of its appearance, it increases the multiplicity of a T_1 edge of $G(\mathbf{i})$ from 1 to 2.

T_4: if it is not T_1 or T_3.

T_2: if it is T_4 and there is no $b < a$ with $\{i_a, i_{a+1}\} = \{i_b, i_{b+1}\}$.

I.e., at the time of its appearance, it creates a new edge but leads to a vertex that has

appeared already.

Moreover, a T_3 leg $(a, (i_a, i_{a+1}))$ is called *irregular* if there is exactly one T_1 leg $(b, (i_b, i_{b+1}))$ which has $b < a$, $v_a \in \{i_b, i_{b+1}\}$, and is single up to a. Otherwise the leg is called *regular*. It is immediate that a T_4 leg is one of the following three kinds.

a) It is a T_2 leg.

b) Its appearance increases the multiplicity of a T_2 edge from 1 to 2.

c) Its edge marks the third or higher order appearance of an edge.

The number of edges of $G(\mathbf{i})$ is s and the number of its vertices is t (since $T(\mathbf{i}) \sim T \in \mathbf{C}_{t-1}$). Call

ℓ: the number of edges of $G(\mathbf{i})$ that have multiplicity at least 3.

m: the number of T_2 legs.

r: the number of regular T_3 legs.

We have for r, t, and $|T_4|$ the following bounds

$$r \leq 2m, \tag{3.8.1}$$

$$t = s + 1 - m \leq k, \tag{3.8.2}$$

$$|T_4| = 2m + 2(k - s). \tag{3.8.3}$$

The first relation is Lemma 5.6 in [52]. The second is true because if we remove the m edges traveled by T_2 legs, we get a tree with $s - m$ edges and t vertices, and in any tree the number or vertices equals the number of edges plus one. Then the inequality is true because $s \leq k$ (all edges of $G(\mathbf{i})$ have multiplicity at least 2) and if $s = k$, then $m \geq 1$ since the cycle is bad. For the last relation, note that $|T_3| = |T_1| = t - 1$ and thus, using (3.8.2) too, we have $|T_4| = 2k - 2(t - 1) = 2k - 2(s - m)$.

Now back to the task of bounding $N_{T,a_1,\ldots,a_s}$. We fix a cycle as in the beginning of the proof and we record

- for each T_4 leg, a) its order in the cycle, b) the index of its initial vertex, c) the index of its final vertex, and d) the index of the final vertex of the next leg in case that leg is T_1. This gives a $Q_1 \subset \{1, 2, \ldots, 2k\} \times (\{1, 2, \ldots t\}^2 \cup \{1, 2, \ldots t\}^3)$ with $|T_4|$ elements.
- for each regular T_3 leg, a) its order in the cycle, b) the index of its initial vertex, and c) the index of its final vertex. This gives a $Q_2 \subset \{1, 2, \ldots, 2k\} \times \{1, 2, \ldots, t\}^2$ with r elements.

We call U the set of all indices that appear as fourth coordinate in elements of Q_1. These are indices of final vertices of T_1 legs.

We claim that, having Q_1, Q_2 and knowing that $T(\mathbf{i}) = T$, we can reconstruct the cycle $\mathbf{i}$.

We determine what kind each leg of the cycle is and what the index of its initial and its final vertex is. These data are known for the T_4 and T_3 regular legs. The remaining legs are T_1 or T_3 irregular. We discover the nature of each of them by traversing the cycle from the beginning as follows. The first leg is T_4 (if $i_2 = i_1$) or T_1. The set Q_1 will tell us if we are in the first case and will give us all we want. If we are in the second case, the initial vertex is 1 and the final 2. Assume that we have arrived at a vertex v_i in the cycle with the smallest i for which the nature of the leg $\ell_i := (i, (v_i, v_{i+1}))$ is not known yet. If the vertex v_i has no neighbors in $G(\mathbf{i})$ that we haven't encountered up to the leg ℓ_{i-1}, then ℓ_i is T_3 irregular, and by the defining property of T_3 irregular legs, we can determine the index of its final vertex. If the vertex v_i does have such neighbors, call z the one that appears earlier in the cycle.

- If $z \in U$, then in case it was included in U because of ℓ_{i-1} (this can be read off from Q_1. Note that z could not have been included because of an earlier leg because z has not appeared earlier than v_i), we have that ℓ_i is T_1 with $v_{i+1} = z$, while in case it was included with a leg $\ell_{i'}$ with index $i' \geq i$, we have that ℓ_i can't be T_1 (because then v_{i+1} would be a neighbor of v_i appearing earlier than z, contradicting the choice of z), thus ℓ_i is T_3 irregular.

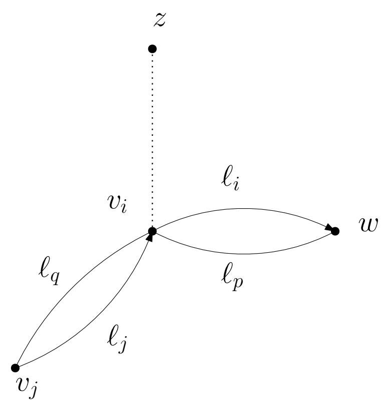

Figure 3.1: The case $z \notin U$.The legs $\ell_i, \ell_j (i < j)$ are T_3, while ℓ_p, ℓ_q are T_1.

• If $z \notin U$, we will show that $\ell_i = (i, (v_i, w))$ is T_1. Assume on the contrary that it is T_3 irregular. Clearly $z \neq w$, and call ℓ_p ($p < i$) the T_1 leg that has vertices v_i, w and is single up to $i - 1$. The cycle will visit the vertex v_i at a later point, with a leg $\ell_j = (j, (v_j, v_i))$ with $j > i$ and $v_j \neq z, v_j \neq v_i$, in order to create the edge that connects v_i with z (that is, $\ell_{j+1} = (j+1, (v_i, z))$ will be T_1), see Figure 3.1. The leg ℓ_j is not T_1 because v_i has been visited by an earlier leg, and it is not T_4 because we assumed that $z \notin U$. It has then to be T_3. Thus, there is a leg ℓ_q connecting vertices v_i, v_j that is T_1.

If $q < i$, then we consider two cases. If $v_j = w$, then ℓ_j is T_4, because the edge v_i, w has been traveled already by ℓ_p, ℓ_i (recall that $p < i < j$), and this would force $z \in U$, a contradiction. If $v_j \neq w$, then ℓ_i would have been T_3 regular as there are at least two T_1 legs (i.e., ℓ_p, ℓ_q) with order less than i with one vertex v_i, traveling different edges, and single up to $i - 1$, again a contradiction because ℓ_i is T_1 or T_3 irregular.

If $q > i$, then $v_j(\neq z)$ is a neighbor of v_i (that is, the T_1 leg ℓ_q goes from v_i to v_j) that appears after leg ℓ_i but earlier than z, which contradicts the definition of z. We conclude that ℓ_i is T_1.

Thus, having T, Q_1, Q_2 allows to determine $\mathbf{i}$.

The above imply that the number of bad cycles with given T, t, r is at most

$$(2kt^2(t+1))^{|T_4|}(2kt^2)^r \leq (4k^4)^{r+|T_4|}. \tag{3.8.4}$$

Then (3.8.1) and (3.8.3) give $r + |T_4| \leq 4m + 2(k - s)$, and finally using (3.8.2), we get the desired bound. □

The next lemma is used in the proof of Theorem 3.1.10.

Lemma 3.8.2. *Let $(A_N)_{N\in\mathbb{N}^+}$ be a sequence of matrices, A_N of dimension $N \times N$, that satisfies Assumption 3.1.1 and Assumption 3.1.2 with measure μ. Suppose that there are two sequences of matrices $(A_N^{(1)})_{N\in\mathbb{N}^+}$ and $(A_N^{(2)})_{N\in\mathbb{N}^+}$ such that*

(a) $A_N = A_N^{(1)} + A_N^{(2)}$,

(b) *For all* $N \in \mathbb{N}^+$ *and* $i,j \in [N]$*, at least one of* $(A_N^{(1)})_{i,j}, (A_N^{(2)})_{i,j}$ *is identically zero random variable.*

(c) $\mu_{A_N^{(1)}/\sqrt{N}} \Rightarrow \mu$ *in probability as* $N \to \infty$*.*

Then $(A_N^{(1)})_{N\in\mathbb{N}^+}$ *also satisfies Assumptions 3.1.1, 3.1.2 with the measure* μ*.*

Proof. We only need to check the validity of Assumption 3.1.2 as the validity of Assumption 3.1.1 is immediate.

Because $A_N^{(1)}$ satisfies Assumption 3.1.1, there is a decreasing sequence $(\eta_N)_{N\in\mathbb{N}^+}$ of positive reals converging to 0 so that

$$\lim_{N\to\infty} \frac{1}{N^2} \sum_{i,j\in[N]} \mathbf{E}\left[(\{A_N^{(1)}\}_{i,j})^2 \mathbf{1}\left(|\{A_N^{(1)}\}_{i,j}| > \eta_N N^{\frac{1}{2}}\right)\right] = 0. \tag{3.8.5}$$

Set $A_N^{(1),\leq}$ to be the matrix whose (i,j) entry is

$$\{A_N^{(1)}\}_{i,j}\mathbf{1}\left(|\{A_N^{(1)}\}_{i,j}| \leq \eta_N N^{\frac{1}{2}}\right) - \mathbf{E}\left[\{A_N^{(1)}\}_{i,j}\mathbf{1}\left(|\{A_N^{(1)}\}_{i,j}| \leq \eta_N N^{\frac{1}{2}}\right)\right] \tag{3.8.6}$$

and $\mu_{N,i,j} := \mathbf{E}\left[\{A_N^{(1)}\}_{i,j}\mathbf{1}\left(|\{A_N^{(1)}\}_{i,j}| \leq \eta_N N^{\frac{1}{2}}\right)\right]$.

Claim:

$$\mu_{N^{-1/2}A_N^{(1),\leq}} \Rightarrow \mu \quad \text{in probability as } N \to \infty \tag{3.8.7}$$

The Levy distance between $\mu_{A_N^{(1)}/\sqrt{N}}$ and $\mu_{A_N^{(1),\leq}/\sqrt{N}}$ is bounded as follows.

$$L^3(\mu_{A_N^{(1)}/\sqrt{N}}, \mu_{A_N^{(1),\leq}/\sqrt{N}}) \leq \frac{1}{N}\mathrm{tr}\left\{\left(\frac{1}{\sqrt{N}}A_N^{(1)} - \frac{1}{\sqrt{N}}A_N^{(1),\leq}\right)^2\right\} \tag{3.8.8}$$

$$= \frac{1}{N^2}\sum_{i,j\in[N]}\{\mu_{N,i,j}^2\mathbf{1}\left(|\{A_N^{(1)}\}_{i,j}| \leq \eta_N N^{\frac{1}{2}}\right) + (\{A_N^{(1)}\}_{i,j} + \mu_{N,i,j})^2\mathbf{1}\left(|\{A_N^{(1)}\}_{i,j}| > \eta_N N^{\frac{1}{2}}\right)\} \tag{3.8.9}$$

$$\leq \frac{3}{N^2}\sum_{i,j\in[N]}\mu_{N,i,j}^2 + \frac{2}{N^2}\sum_{i,j\in[N]}(\{A_N^{(1)}\}_{i,j})^2\mathbf{1}\left(|\{A_N^{(1)}\}_{i,j}| > \eta_N N^{\frac{1}{2}}\right). \tag{3.8.10}$$

Since the entries of $A_N^{(1)}$ have mean 0, we have

$$\mu_{N,i,j}^2 = \left(\mathbf{E}\left[\{A_N^{(1)}\}_{i,j}\mathbf{1}\left(|\{A_N^{(1)}\}_{i,j}| > \eta_N N^{\frac{1}{2}}\right)\right]\right)^2 \leq \mathbf{E}\left[(\{A_N^{(1)}\}_{i,j})^2\mathbf{1}\left(|\{A_N^{(1)}\}_{i,j}| > \eta_N N^{\frac{1}{2}}\right)\right].$$

Thus, the expectation of the expression in (3.8.10) is at most

$$\frac{5}{N^2}\sum_{i,j\in[N]}\mathbf{E}\left\{(\{A_N^{(1)}\}_{i,j})^2\mathbf{1}\left(|\{A_N^{(1)}\}_{i,j}| > \eta_N N^{\frac{1}{2}}\right)\right\},$$

which tends to zero as $N \to \infty$ due to (3.8.5). This, combined with assumption (c), proves the claim.

Fix $k \in \mathbb{N}^+$ and set $M_N(k), M_N^{(1),\leq}(k)$ the asymptotic contributing terms (see (3.1.7)) of A_N and $A_N^{(1),\leq}$ respectively. Notice that

$$M_N^{(1),\leq}(k) \leq M_N^{(1)}(k) \leq M_N(k). \tag{3.8.11}$$

The rightmost inequality is true because the variance of $\{A_N^{(1)}\}_{i,j}$ is either zero or $s_{i,j}^{(N)}$ due to assumption (b) of the lemma. The leftmost inequality is true because if W is a real valued random variable with mean 0 and finite variance and $\tilde{W}$ is a variable with $|\tilde{W}| \le |W|$, then $\mathrm{Var}(\tilde{W}) \le \mathrm{Var}(W)$.

Lemma 3.6 of [13] implies that

$$\frac{M_N^{(1),\le}(k)}{N^{k+1}} = \frac{1}{N^{k+1}} \mathbf{E}\,\mathrm{tr}\{(A_N^{(1),\le})^{2k}\} + o(1) \tag{3.8.12}$$

as $N \to \infty$. We will prove that the right hand side converges to $\int x^{2k} d\mu$ as $N \to \infty$. It will be convenient to let $B_N := A_N^{(1),\le}/\sqrt{N}$ and $\{\lambda_i(B_N) : i \in [N]\}$ its eigenvalues.

Pick some $C > \mu_\infty$ and consider the function $g_C(x) = (|x| \wedge C)^{2k}$, which is bounded and continuous. Then,

$$\frac{1}{N^{k+1}} \mathbf{E}\,\mathrm{tr}\{(A_N^{(1),\le})^{2k}\} = \frac{1}{N}\sum_{i=1}^{N} \mathbf{E}\{(\lambda_i\{B_N\})^{2k}\} \tag{3.8.13}$$

and the right hand side can be estimated as follows.

$$\begin{aligned}
&\left|\frac{1}{N}\sum_{i=1}^{N} \mathbf{E}\{(\lambda_i\{B_N\})^{2k}\} - \frac{1}{N}\sum_{i=1}^{N} \mathbf{E} g_C(\lambda_i\{B_N\})\right| \le \frac{1}{N}\sum_{i=1}^{N} \mathbf{E}\{\lambda_i^{2k}(B_N)\mathbf{1}_{|\lambda_i(B_N)|\ge C}\} \\
&\le \frac{1}{N}\sum_{i=1}^{N} \sqrt{\mathbf{E}\lambda_i^{4k}(B_N)}\sqrt{\mathbb{P}(|\lambda_i(B_N)| \ge C)} \le \sqrt{\sum_{i=1}^{N} \frac{\mathbf{E}\lambda_i^{4k}(B_N)}{N}}\sqrt{\frac{\mathbf{E}\sum_{i=1}^{N} \mathbf{1}_{|\lambda_i(B_N)|\ge C}}{N}} \overset{N\to\infty}{\to} 0.
\end{aligned} \tag{3.8.14}$$

To justify the convergence to zero, note that the quantity in the second square root converges to zero by our choice of $C > \mu_\infty$ and the in probability weak convergence of the E.S.D. of B_N to μ. The quantity in the first square root is bounded in N because, due to (3.8.12), its difference from $M_N^{(1),\le}(2k)/N^{2k+1}$ is bounded and the latter is less than $M_N(2k)/N^{2k+1}$ which is bounded in N since it converges to $\int x^{4k}\, d\mu$.

The in probability weak convergence (3.8.7) implies that

$$\frac{1}{N}\sum_{i=1}^{N} g_C(\lambda_i\{B_N\}) \to \int x^{2k} d\mu \text{ in probability,} \tag{3.8.15}$$

and the boundedness of g_C allows to conclude that

$$\lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} \mathbf{E} g_C(\lambda_i\{B_N\}) = \int x^{2k} d\mu. \tag{3.8.16}$$

Thus, relations (3.8.12), (3.8.13), (3.8.14),(3.8.16) show that

$$\lim_{N\to\infty} \frac{M_N^{(1),\le}(k)}{N^{k+1}} = \int x^{2k} d\mu. \tag{3.8.17}$$

And this combined with (3.8.11) concludes the proof. □

Bibliography

[1] A. Edelman, "Eigenvalues and condition numbers of random matrices," *SIAM journal on matrix analysis and applications*, vol. 9, no. 4, pp. 543–560, 1988.

[2] T. Tao and V. Vu, "Random matrices: The distribution of the smallest singular values," *Geometric And Functional Analysis*, vol. 20, no. 1, pp. 260–297, 2010.

[3] Z. Che and P. Lopatto, "Universality of the least singular value for sparse random matrices," *Electronic Journal of Probability*, vol. 24, pp. 1–53, 2019.

[4] ——, "Universality of the least singular value for the sum of random matrices," *Electronic Journal of Probability*, vol. 26, pp. 1–38, 2021.

[5] A. Aggarwal, P. Lopatto, and H.-T. Yau, "Goe statistics for lévy matrices," *Journal of the European Mathematical Society*, 2021.

[6] P. Bourgade, J. Huang, and H.-T. Yau, "Eigenvector statistics of sparse random matrices," *Electronic Journal of Probability*, vol. 22, pp. 1–38, 2017.

[7] L. Erdos and H.-T. Yau, "A dynamical approach to random matrix theory," *Courant Lecture Notes in Mathematics*, vol. 28, 2017.

[8] E. P. Wigner, "On the distribution of the roots of certain symmetric matrices," *Annals of Mathematics*, pp. 325–327, 1958.

[9] V. A. Marčenko and L. A. Pastur, "Distribution of eigenvalues for some sets of random matrices," *Mathematics of the USSR-Sbornik*, vol. 1, no. 4, p. 457, 1967.

[10] Z.-D. Bai and Y.-Q. Yin, "Necessary and sufficient conditions for almost sure convergence of the largest eigenvalue of a wigner matrix," *The Annals of Probability*, pp. 1729–1741, 1988.

[11] N. Cook, W. Hachem, J. Najim, and D. Renfrew, "Non-hermitian random matrices with a variance profile (i): deterministic equivalents and limiting esds," *Electronic Journal of Probability*, vol. 23, pp. 1–61, 2018.

[12] J. Husson, “Large deviations for the largest eigenvalue of matrices with variance profiles,” *Electronic Journal of Probability*, vol. 27, pp. 1–44, 2022.

[13] Y. Zhu, “A graphon approach to limiting spectral distributions of wigner-type matrices,” *Random Structures & Algorithms*, vol. 56, no. 1, pp. 251–279, 2020.

[14] L. Erdős, T. Krüger, and D. Schröder, “Random matrices with slow correlation decay,” in *Forum of Mathematics, Sigma*, vol. 7. Cambridge University Press, 2019.

[15] D. Cheliotis, “Singular values of triangular random matrices through the moment method,” 2022.

[16] T. Tao, *Topics in random matrix theory*. American Mathematical Soc., 2012, vol. 132.

[17] M. Dieng and C. A. Tracy, “Application of random matrix theory to multivariate statistics,” in *Random Matrices, Random Processes and Integrable Systems*. Springer, 2011, pp. 443–507.

[18] I. M. Johnstone, “On the distribution of the largest eigenvalue in principal components analysis,” *The Annals of statistics*, vol. 29, no. 2, pp. 295–327, 2001.

[19] A. Onatski, “Testing hypotheses about the number of factors in large factor models,” *Econometrica*, vol. 77, no. 5, pp. 1447–1479, 2009.

[20] N. Patterson, A. L. Price, and D. Reich, “Population structure and eigenanalysis,” *PLoS genetics*, vol. 2, no. 12, p. e190, 2006.

[21] N. S. Pillai and J. Yin, “Universality of covariance matrices,” *The Annals of Applied Probability*, vol. 24, no. 3, pp. 935–1001, 2014.

[22] D. Chafaı, D. Chafä, O. Guédon, G. Lecue, and A. Pajor, “Singular values of random matrices,” *Lecture Notes*, 2009.

[23] D. A. Spielman and S.-H. Teng, “Smoothed analysis of algorithms: Why the simplex algorithm usually takes polynomial time,” *Journal of the ACM (JACM)*, vol. 51, no. 3, pp. 385–463, 2004.

[24] S. Belinschi, A. Dembo, and A. Guionnet, “Spectral measure of heavy tailed band and covariance random matrices,” *Communications in Mathematical Physics*, vol. 289, no. 3, pp. 1023–1055, 2009.

[25] C. Bordenave, P. Caputo, D. Chafaï, and D. Piras, “Spectrum of large random markov chains: heavy-tailed weights on the oriented complete graph,” *Random Matrices: Theory and Applications*, vol. 6, no. 02, p. 1750006, 2017.

[26] A. Auffinger, G. Ben Arous, and S. Péché, "Poisson convergence for the largest eigenvalues of heavy tailed random matrices," in *Annales de l'IHP Probabilités et statistiques*, vol. 45, 2009, pp. 589–610.

[27] J. Heiny and T. Mikosch, "Eigenvalues and eigenvectors of heavy-tailed sample covariance matrices with general growth rates: The iid case," *Stochastic Processes and their Applications*, vol. 127, no. 7, pp. 2179–2207, 2017.

[28] ——, "Large sample autocovariance matrices of linear processes with heavy tails," *Stochastic Processes and their Applications*, vol. 141, pp. 344–375, 2021.

[29] B. Basrak, Y. Cho, J. Heiny, and P. Jung, "Extreme eigenvalue statistics of m-dependent heavy-tailed matrices," in *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, vol. 57, no. 4. Institut Henri Poincaré, 2021, pp. 2100–2127.

[30] G. B. Arous and A. Guionnet, "The spectrum of heavy tailed random matrices," *Communications in Mathematical Physics*, vol. 278, no. 3, pp. 715–751, 2008.

[31] C. Bordenave and A. Guionnet, "Localization and delocalization of eigenvectors for heavy-tailed random matrices," *Probability Theory and Related Fields*, vol. 157, no. 3-4, pp. 885–953, 2013.

[32] ——, "Delocalization at small energy for heavy-tailed random matrices," *Communications in Mathematical Physics*, vol. 354, no. 1, pp. 115–159, 2017.

[33] C. Bordenave, P. Caputo, and D. Chafai, "Spectrum of large random reversible markov chains: heavy-tailed weights on the complete graph," *The Annals of Probability*, vol. 39, no. 4, pp. 1544–1590, 2011.

[34] A. Aggarwal, P. Lopatto, and J. Marcinek, "Eigenvector statistics of lévy matrices," *arXiv preprint arXiv:2002.09355*, 2020.

[35] Z. Füredi and J. Komlós, "The eigenvalues of random symmetric matrices," *Combinatorica*, vol. 1, no. 3, pp. 233–241, 1981.

[36] C. A. Tracy and H. Widom, "Level-spacing distributions and the airy kernel," *Communications in Mathematical Physics*, vol. 159, no. 1, pp. 151–174, 1994.

[37] J. Huang and H.-T. Yau, "Edge universality of sparse random matrices," *arXiv preprint arXiv:2206.06580*, 2022.

[38] S. Sodin, "The spectral edge of some random band matrices," *Annals of mathematics*, pp. 2223–2251, 2010.

[39] A. S. Bandeira and R. Van Handel, "Sharp nonasymptotic bounds on the norm of random matrices with independent entries," *The Annals of Probability*, vol. 44, no. 4, pp. 2479–2506, 2016.

[40] A. S. Bandeira, M. T. Boedihardjo, and R. van Handel, "Matrix concentration inequalities and free probability," *arXiv preprint arXiv:2108.06312*, 2021.

[41] N. Cook, W. Hachem, J. Najim, and D. Renfrew, "Non-hermitian random matrices with a variance profile (ii): Properties and examples," *Journal of Theoretical Probability*, vol. 35, no. 4, pp. 2343–2382, 2022.

[42] J. Alt, L. Erdős, T. Krüger, and Y. Nemish, "Location of the spectrum of kronecker random matrices," 2019.

[43] R. Ducatez, J. Husson, and A. Guionnet, "Large deviation principle for the largest eigenvalue of random matrices with a variance profile," *arXiv preprint arXiv:2403.05413*, 2024.

[44] G. Samorodnitsky, *Stable non-Gaussian random processes: stochastic models with infinite variance*. Routledge, 2017.

[45] G. W. Anderson, A. Guionnet, and O. Zeitouni, *An introduction to random matrices*, ser. 1. Cambridge university press, 2010, no. 118.

[46] B. Landon, P. Sosoe, and H.-T. Yau, "Fixed energy universality of dyson brownian motion," *Advances in Mathematics*, vol. 346, pp. 1137–1332, 2019.

[47] C. McDiarmid *et al.*, "On the method of bounded differences," *Surveys in combinatorics*, vol. 141, no. 1, pp. 148–188, 1989.

[48] R. Vershynin, *High-dimensional probability: An introduction with applications in data science*. Cambridge university press, 2018, vol. 47.

[49] J. Huang, B. Landon, and H.-T. Yau, "Bulk universality of sparse random matrices," *Journal of Mathematical Physics*, vol. 56, no. 12, p. 123301, 2015.

[50] R. Bauerschmidt, A. Knowles, and H.-T. Yau, "Local semicircle law for random regular graphs," *Communications on Pure and Applied Mathematics*, vol. 70, no. 10, pp. 1898–1960, 2017.

[51] L. Erdős, H.-T. Yau, and J. Yin, "Rigidity of eigenvalues of generalized wigner matrices," *Advances in Mathematics*, vol. 229, no. 3, pp. 1435–1515, 2012.

[52] Z. Bai and J. W. Silverstein, *Spectral analysis of large dimensional random matrices*. Springer, 2010, vol. 20.

[53] Y. Yin, Z. Bai, and Krishnaiah, "On limit of the largest eigenvalue of the large dimensional sample covariance matrix," *PR*, 1984.

[54] W. Hachem, P. Loubaton, and J. Najim, "A clt for information-theoretic statistics of gram random matrices with a given variance profile," *The Annals of Applied Probability*, vol. 18, no. 6, pp. 2071–2130, 2008.

[55] R. Couillet and M. Debbah, *Random matrix methods for wireless communications*. Cambridge University Press, 2011.

[56] L. V. Bogachev, S. A. Molchanov, and L. A. Pastur, "On the level density of random band matrices," *Mathematical notes of the Academy of Sciences of the USSR*, vol. 50, no. 6, pp. 1232–1242, 1991.

[57] P. Bourgade, "Random band matrices," in *Proceedings of the International Congress of Mathematicians: Rio de Janeiro 2018*. World Scientific, 2018, pp. 2759–2783.

[58] F. Benaych-Georges and S. Péché, "Largest eigenvalues and eigenvectors of band or sparse random matrices," *Electronic Communications in Probability*, vol. 19, pp. 1–9, 2014.

[59] A. Bose and P. Sen, "XX^t matrices with independent entries," *arXiv preprint arXiv:2205.10811*, 2022.

[60] A. Soshnikov *et al.*, "Poisson statistics for the largest eigenvalues of wigner random matrices with heavy tails," *Electronic Communications in Probability*, vol. 9, pp. 82–91, 2004.

[61] 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*, vol. 78, pp. 509–521, 1988.

[62] R. A. Davis, O. Pfaffel, and R. Stelzer, "Limit theory for the largest eigenvalues of sample covariance matrices with heavy-tails," *Stochastic Processes and their Applications*, vol. 124, no. 1, pp. 18–50, 2014.

[63] I. Zakharevich, "A generalization of wigner's law," *Communications in mathematical physics*, vol. 268, no. 2, pp. 403–414, 2006.

[64] C. Bordenave, D. Chafaï *et al.*, "Around the circular law," *Probability surveys*, vol. 9, pp. 1–89, 2012.

[65] J. D. Murray, *Asymptotic analysis*. Springer, 1984.

[66] F. W. J. Olver, *Introduction to asymptotics and special functions.* Academic Press, 1974.

[67] S. Mittnik, S. T. Rachev, and J.-R. Kim, “Chi-square-type distributions for heavy-tailed variates,” *Econometric Theory*, pp. 339–354, 1998.

[68] J. Y. Hwang, J. O. Lee, and K. Schnelli, “Local law and tracy–widom limit for sparse sample covariance matrices,” *The Annals of Applied Probability*, vol. 29, no. 5, pp. 3006–3036, 2019.

[69] J. W. Silverstein, “The smallest eigenvalue of a large dimensional wishart matrix,” *The Annals of Probability*, pp. 1364–1368, 1985.

[70] D. C. Lay, *Linear algebra and it’s applications.* Pearson, Addison-Wesley, 2006.

[71] M. Rudelson and R. Vershynin, “Non-asymptotic theory of random matrices: extreme singular values,” in *Proceedings of the International Congress of Mathematicians 2010 (ICM 2010) (In 4 Volumes) Vol. I: Plenary Lectures and Ceremonies Vols. II–IV: Invited Lectures.* World Scientific, 2010, pp. 1576–1602.

[72] A. Bose, K. Saha, A. Sen, and P. Sen, “Random matrices with independent entries: beyond non-crossing partitions,” *Random Matrices: Theory and Applications*, vol. 11, no. 02, p. 2250021, 2022.

[73] X. Ding, “Spectral analysis of large block random matrices with rectangular blocks,” *Lithuanian Mathematical Journal*, vol. 54, no. 2, pp. 115–126, 2014.

[74] R. R. Far, T. Oraby, W. Bryc, and R. Speicher, “On slow-fading mimo systems with nonseparable correlation,” *IEEE Transactions on Information Theory*, vol. 54, no. 2, pp. 544–553, 2008.

[75] R. A. Horn and C. R. Johnson, *Matrix analysis.* Cambridge university press, 2012.

[76] P. Billingsley, *Convergence of probability measures*, 2nd ed., 1999.